CHEM 2141: SURVEY OF PHYSICAL CHEMISTRY



Lauren Woods
University of North Carolina Charlotte



CHEM 2141: Survey of Physical Chemistry

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden borne from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and were supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 10/25/2026

TABLE OF CONTENTS

Licensing

1: Introduction

1.1: Units

1.2: Quantity Calculus

1.3: Dimensional Analysis

1.4: Chapter 1 Problem

2: General Chemistry Review

2.1: Gases

2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law

2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions

2.1.3: The Kinetic-Molecular Theory

2.1.4: Non-Ideal Gas Behavior

2.2: Thermodynamics

2.2.1: Energy Basics

2.2.2: Enthalpy

2.2.3: Spontaneity

2.2.4: Entropy

2.2.5: The Second and Third Laws of Thermodynamics

2.2.6: Free Energy

2.3: Kinetics

2.3.1: Chemical Reaction Rates

2.3.2: Factors Affecting Reaction Rates

2.3.3: Rate Laws

2.3.4: Integrated Rate Laws

2.3.5: Collision Theory

2.3.6: Reaction Mechanisms

2.3.7: Catalysis

2.4: Using logarithms - Log vs. Ln

3: The First Law of Thermodynamics

3.1: Overview of Classical Thermodynamics

3.2: Pressure-Volume Work

3.3: Work and Heat are not State Functions

3.4: Energy is a State Function

3.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred

3.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion

3.7: Pressure-Volume Work

3.8: Heat Capacity is a Path Function
3.9: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition
3.10: Enthalpy Changes for Chemical Equations are Additive
3.11: Heats of Reactions Can Be Calculated from Tabulated Heats of Formation
3.12: The Temperature Dependence of ΔH
3.13: Enthalpy is a State Function
3.E: The First Law of Thermodynamics (Exercises)

4: Entropy and The Second and 3rd Law of Thermodynamics

4.1: Energy Does not Determine Spontaneity
4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal
4.3: Unlike heat, Entropy is a State Function
4.4: The Second Law of Thermodynamics
4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes
4.6: Entropy Increases With Increasing Temperature
4.7: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale
4.8: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition
4.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions
4.E: Entropy and The Second Law of Thermodynamics (Exercises)

5: Helmholtz and Gibbs Energies

5.1: Free Energy Functions
5.2: ΔA, ΔG, and Maximum Work
5.3: Pressure Dependence of Gibbs Energy
5.4: Temperature Dependence of A and G
5.5: When Two Variables Change at Once
5.6: Thermodynamic Functions have Natural Variables
5.7: The Gibbs-Helmholtz Equation
5.E: Gibbs Energy (Exercises)

6: Chemical Kinetics I

6.1: Reaction Rate
6.2: Rate Laws
6.3: 0th order Rate Law
6.4: 1st order rate law
6.5: 2nd order Rate Laws
6.6: The Method of Initial Rates
6.7: The Method of Half-Lives
6.8: Temperature Dependence
6.E: Chemical Kinetics I (Exercises)
6.S: Chemical Kinetics I (Summary)

7: Chemical Kinetics II

7.1: Reaction Mechanisms

7.2: Concentration Profiles for Some Simple Mechanisms

7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws

7.4: The Rate Determining Step Approximation

7.5: The Steady-State Approximation

7.6: The Equilibrium Approximation

7.7: The Lindemann Mechanism

7.8: The Michaelis-Menten Mechanism

7.9: Catalysis

7.10: Theories of Reaction Rates

7.E: Chemical Kinetics II (Exercises)

7.S: Chemical Kinetics II (Summary)

8: Optional- Special topics

8.1: Chemistry of Cooking (Rodriguez-Velazquez)

8.1.1: Thickening and Concentrating Flavors

8.1.1.1: Viscosity

8.1.1.2: Thickening Agents

8.1.1.3: Types of Thickening Agents

8.1.1.4: Coagulation

8.1.1.5: Gelatinization

8.1.1.6: Crystallization

8.1.1.7: Non-traditional thickeners

8.1.1.8: Sauces

8.1.1.9: Low-temperature and sous-vide

8.1.2: Flour

8.1.2.1: Introduction - Understanding Ingredients

8.1.2.2: The History of Wheat Flour

8.1.2.3: Milling of Wheat

8.1.2.4: Flour Streams and Types of Wheat Flour

8.1.2.5: Flour Terms and Treatments

8.1.2.6: Flour Additives

8.1.2.7: Whole Grain and Artisan Milling

8.1.2.8: Flour in Baking

8.1.2.9: Rye Flour

8.1.2.10: Other Grains and Flours

8.1.3: Fat

8.1.3.1: Understanding Fats and Oils

8.1.3.2: Sources of Bakery Fats and Oils

8.1.3.3: Major Fats and Oils Used in Bakeries
8.1.3.4: Functions of Fat in Baking

8.1.4: Sugar
8.1.4.1: Sugar Chemistry (ADD US)
8.1.4.2: Sugar Refining
8.1.4.3: The Application of Sugar
8.1.4.4: Agave
8.1.4.5: Honey
8.1.4.6: Malt
8.1.4.7: Maple Syrup (ADD US)
8.1.4.8: Sugar Substitutes (ADD US)
8.1.4.9: Glucose/Dextrose

8.1.5: Leavening Agents
8.1.5.1: Introduction to Leavening Agents
8.1.5.2: Yeast
8.1.5.3: The Functions of Yeast
8.1.5.4: Using Yeast in Baking
8.1.5.5: Baking Powder
8.1.5.6: Sodium Bicarbonate
8.1.5.7: Ammonium Bicarbonate
8.1.5.8: Water Hardness and pH

8.1.6: Dairy Products
8.1.6.1: Introduction to Dairy Products
8.1.6.2: Milk
8.1.6.3: Milk Products ADD US
8.1.6.4: Milk in bread baking
8.1.6.5: Yogurt
8.1.6.6: Lactose
8.1.6.7: Cheese

8.1.7: Eggs
8.1.7.1: Eggs Grade
8.1.7.2: Composition and Nutrition
8.1.7.3: Egg Products
8.1.7.4: The Function of Eggs
8.1.7.5: Storing Eggs

8.1.8: Chocolate
8.1.8.1: From the Cocoa Bean to the Finished Chocolate
8.1.8.2: Chocolate Produced for the Baking Industry
8.1.8.3: Couverture

8.1.8.4: Definitions and Regulations (ADD US)

8.1.9: Spices

8.1.9.1: Elements of Taste

8.1.9.2: Introduction to Salt

8.1.9.3: Origins of Salt

8.1.9.4: Functions of Salt in Baking

8.1.9.5: Using Salt in Fermented Doughs

8.1.9.6: Storing Salt

8.1.9.7: Introduction to Spices and Other Flavorings

8.1.9.8: Seasoning and Flavoring

8.1.9.9: Herbs

8.1.9.10: Spices

8.1.9.11: Flavorings in Baking

8.2: Aquatic Chemistry

8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems

8.2.2: 2. Carbonate Equilibria in Natural Waters

8.2.3: 3. Redox Equilibria in Natural Waters

8.2.4: 4. Solids in Contact With Natural Waters

8.2.5: Fundamental Characteristics of Water

8.2.6: Natural Water

8.2.7: Water Biology

8.2.8: Water Chemistry

8.2.9: Water Physics

8.2.10: Water Treatment

8.3: Biochemical Cycles

8.3.1: Biochemical Cycles

8.3.2: Carbon Cycle

8.3.3: Nitrogen Cycle

8.4: Toxicology MSDT

8.4.1: Pathophysiology

8.4.1.1: What is Pathophysiology?

8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity

8.4.1.3: Cellular Response to Toxicant-Induced Injury

8.4.1.4: Repair and Adaptation

8.4.1.5: Patterns of Toxic Injury

8.4.1.6: Final Evaluation

8.4.2: Biochemistry and Molecular Genetics

8.4.2.1: Introduction to Biomolecules and Cell Components

8.4.2.2: Cell Structure and Subcellular Compartments

8.4.2.3: DNA and RNA Metabolism
8.4.2.4: Epigenetic Mechanisms
8.4.2.5: Section 2 Final Evaluation

8.4.3: Principles of Genetic Toxicology
8.4.3.1: Introduction to Genetic-toxicology Assay
8.4.3.2: Different Genetic Damages or Mutations
8.4.3.3: Different Genetic-Toxicology Assays
8.4.3.4: Different Cytotoxicity Assays
8.4.3.5: Epigenetics Assay
8.4.3.6: Section 3 Final Evaluation

8.4.4: Applied Systems Toxicology
8.4.4.1: Systems Toxicology
8.4.4.2: Dose Level and Applied Toxicology
8.4.4.3: Tools and Technologies in Systems Toxicology
8.4.4.4: Other Approaches for Predictive Toxicity Modeling
8.4.4.5: Technologies Used In Systems Biology/Toxicology
8.4.4.6: Takeaways Summary
8.4.4.7: Section 4 Final Evaluation

8.4.5: Regulatory Toxicology
8.4.5.1: Introduction to Regulatory Toxicology
8.4.5.2: Global Regulatory Toxicology
8.4.5.3: Topic 3- Regional Regulatory Toxicology
8.4.5.4: National Regulatory Toxicology
8.4.5.5: State Regulatory Toxicology
8.4.5.6: Non-Governmental Regulatory Toxicology
8.4.5.7: Section 5 Final Evaluation

8.4.6: Principles of Toxicology

8.4.6.1: Introduction to Toxicology
8.4.6.1.1: What is toxicology?
8.4.6.1.2: Basic Terminology

8.4.6.2: Dose and Dose Response
8.4.6.2.1: Dose and It's Impact on Toxicity
8.4.6.2.2: The Dose Response Relationship
8.4.6.2.3: Dose Estimates of Toxic Effects
8.4.6.2.4: Determining the Safety of a Drug
8.4.6.2.5: NOAEL and LOAEL

8.4.6.3: Toxic Effects
8.4.6.3.1: Types of Toxic Effects
8.4.6.3.2: Factors Affecting Toxicity

8.4.6.3.3: Systemic Toxic Effects
8.4.6.3.4: Organ Specific Toxic Effects

8.4.6.4: Interactions
8.4.6.4.1: Interactions

8.4.6.5: Toxicity Testing Methods
8.4.6.5.1: Testing and Assessing Toxicity
8.4.6.5.2: Clinical Investigations and Other Types of Human Data
8.4.6.5.3: Epidemiology Studies

8.4.6.6: Risk Assessment
8.4.6.6.1: Risk Assessment
8.4.6.6.2: Hazard Identification
8.4.6.6.3: Dose-Response Assessment
8.4.6.6.4: Exposure Assessment
8.4.6.6.5: Risk Characterization

8.4.6.7: Exposure Standards and Guidelines
8.4.6.7.1: Exposure Standards and Guidelines
8.4.6.7.2: Regulation of Consumer Products and Drug Safety
8.4.6.7.3: Environmental Exposure Standards/Guidelines
8.4.6.7.4: Occupational (Workplace) Exposure Standards/Guidelines/Approaches

8.4.6.8: Basic Physiology
8.4.6.8.1: Introduction to Basic Physiology
8.4.6.8.2: Homeostasis
8.4.6.8.3: Organs and Organ Systems
8.4.6.8.4: Tissues
8.4.6.8.5: Cells
8.4.6.8.6: Chemicals

8.4.6.9: Introduction to Toxicokinetics
8.4.6.9.1: What is Toxicokinetics

8.4.6.10: Absorption
8.4.6.10.1: Introduction to Absorption
8.4.6.10.2: Gastrointestinal Tract
8.4.6.10.3: Respiratory Tract
8.4.6.10.4: Dermal Route
8.4.6.10.5: Other Routes of Exposure

8.4.6.11: Distribution
8.4.6.11.1: Introduction to Distribution
8.4.6.11.2: Influence of Route of Exposure
8.4.6.11.3: Disposition Models
8.4.6.11.4: Structural Barriers to Distribution
8.4.6.11.5: Storage Sites

8.4.6.12: Biotransformation
8.4.6.12.1: Introduction to Biotransformation
8.4.6.12.2: Chemical Reactions
8.4.6.12.3: Biotransformation Sites
8.4.6.12.4: Modifiers of Biotransformation

8.4.6.13: Excretion
8.4.6.13.1: Introduction to Secretion
8.4.6.13.2: Urinary Excretion
8.4.6.13.3: Fecal Excretion
8.4.6.13.4: Exhaled Air
8.4.6.13.5: Other Routes

8.4.6.14: Cellular Toxicology
8.4.6.14.1: Adaptation
8.4.6.14.2: Cell Damage and Tissue Repair
8.4.6.14.3: Cancer
8.4.6.14.4: Neurotoxicity

8.4.6.15: Intuitive Toxicology and Risk Communication
8.4.6.15.1: Intuitive Toxicology
8.4.6.15.2: Risk Communication

8.4.6.16: Environmental Toxicology, Environmental Health, and One Health
8.4.6.16.1: Environmental Toxicology
8.4.6.16.2: Environmental Health
8.4.6.16.3: One Health

8.4.6.17: Conclusion

8.5: Food to energy metabolic pathways
8.5.1: Basics of metabolism
8.5.2: Digestion of food
8.5.3: Oxidation of glucose -the glycolysis
8.5.4: Citric acid cycle
8.5.5: Oxidative Phosphorylation
8.5.6: Oxidation of fatty acids
8.5.7: Degradation of amino acids

9: Quantum Basics
9.1: Atoms and Photons- Origin of the Quantum Theory
9.2: Waves and Particles
9.3: The Free Particles
9.4: The Particle in a Box
9.5: Quantum Mechanics of Some Simple Systems
9.6: Principles of Quantum Mechanics

9.7: Harmonic Oscillator
9.8: Angular Momentum
9.9: Hydrogen Atom
9.10: Helium Atom
9.11: Atomic Structure and The Periodic Law
9.12: The Chemical Bond
9.13: Molecular Orbital Theory
9.14: Molecular Symmetry
9.15: Molecular Spectroscopy
9.16: Nuclear Magnetic Resonance

10: Molecular Spectroscopy
10.1: The Electromagnetic Spectrum
10.2: Rotations Accompany Vibrational Transitions
10.3: Unequal Spacings in Vibration-Rotation Spectra
10.4: Unequal Spacings in Pure Rotational Spectra
10.5: Vibrational Overtones
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
10.7: Electronic Spectroscopy
10.7.1: Spectroscopy of the Particle-in-a-Box Model
10.7.2: Cyanine Dyes
10.7.3: Fluorescence and Phosphorescence
10.8: Rotational Spectra of Polyatomic Molecules
10.9: Normal Modes in Polyatomic Molecules
10.10: Irreducible Representation of Point Groups
10.11: The Selection Rule for the Rigid Rotor
10.12: The Harmonic Oscillator Selection Rule
10.13: Group Theory Determines Infrared Activity
10.E: Molecular Spectroscopy (Exercises)

Index
Glossary
Detailed Licensing
Detailed Licensing

This page titled CHEM 2141: Survey of Physical Chemistry is shared under a license and was authored, remixed, and/or curated by LibreTexts.

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

This page titled Licensing is shared under a license and was authored, remixed, and/or curated by LibreTexts.

1: INTRODUCTION

Thermodynamics is a quantitative subject. It allows us to derive relations between the values of numerous physical quantities. Some physical quantities, such as a mole fraction, are dimensionless; the value of one of these quantities is a pure number. Most quantities, however, are not dimensionless and their values must include one or more *units*. This chapter reviews the SI system of units, which are the preferred units in science applications. The chapter then discusses some useful mathematical manipulations of physical quantities using quantity calculus, and certain general aspects of dimensional analysis.

1.1: Units

1.2: Quantity Calculus

1.3: Dimensional Analysis

1.4: Chapter 1 Problem

This page titled 1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Howard DeVoe via source content that was edited to the style and standards of the LibreTexts platform.

- 1: Introduction by Howard DeVoe is licensed CC BY 4.0. Original source: https://www2.chem.umd.edu/thermobook.

1.1: Units

There is international agreement that the units used for physical quantities in science and technology should be those of the International System of Units, or SI (standing for the French **Système International d'Unités**). The Physical Chemistry Division of the International Union of Pure and Applied Chemistry, or IUPAC, produces a manual of recommended symbols and terminology for physical quantities and units based on the SI. The manual has become known as the Green Book (from the color of its cover) and is referred to here as the IUPAC Green Book. This e-book will, with a few exceptions, use symbols recommended in the third edition (2007) of the IUPAC Green Book (E. Richard Cohen et al, *Quantities, Units and Symbols in Physical Chemistry*, 3rd edition. RSC Publishing, Cambridge, 2007). These symbols are listed for convenient reference in Appendices C and D.

Any of the symbols for units listed in Tables 1.1–1.3, except kg and $^{\circ}\mathrm{C}$, may be preceded by one of the prefix symbols of Table 1.4 to construct a decimal fraction or multiple of the unit. (The symbol g may be preceded by a prefix symbol to construct a fraction or multiple of the gram.) The combination of prefix symbol and unit symbol is taken as a new symbol that can be raised to a power without using parentheses, as in the following examples:

- The physical quantity formally called **amount of substance** is a counting quantity for particles, such as atoms or molecules, or for other chemical entities. The counting unit is invariably the **mole**, defined as the amount of substance containing as many particles as the number of atoms in exactly 12 grams of pure carbon-12 nuclide, $^{12}\mathrm{C}$. See Appendix A for the wording of the official IUPAC definition. This definition is such that one mole of H_2O molecules, for example, has a mass of 18.0153 grams (where 18.0153 is the relative molecular mass of H_2O) and contains 6.02214×10^{23} molecules (where $6.02214 \times 10^{23}\,\mathrm{mol}^{-1}$ is the *Avogadro constant* to six significant digits). The same statement can be made for any other substance if 18.0153 is replaced by the appropriate atomic mass or molecular mass value.

  The symbol for amount of substance is n. It is admittedly awkward to refer to $n(H_2O)$ as "the amount of substance of water." This e-book simply shortens "amount of substance" to **amount**. An alternative name suggested for n is "chemical amount." Thus, "the amount of water in the system" refers not to the mass or volume of water, but to the *number* of H_2O molecules in the system expressed in a counting unit such as the mole.

- ADAPT 1.1.1

  This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

  View this Interactive Content online

This page titled 1.1: Units is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Howard DeVoe via source content that was edited to the style and standards of the LibreTexts platform.

- 1.1: Units by Howard DeVoe is licensed CC BY 4.0. Original source: https://www2.chem.umd.edu/thermobook.

1.2: Quantity Calculus

This section gives examples of how we may manipulate physical quantities by the rules of algebra. The method is called *quantity calculus*, although a better term might be "quantity algebra."

This shows the new text

Quantity calculus is based on the concept that a physical quantity, unless it is dimensionless, has a value equal to the product of a *numerical value* (a pure number) and one or more *units*:

$$\text{physical quantity} = \text{numerical value} \times \text{units} \tag{1.2.1}$$

(If the quantity is dimensionless, it is equal to a pure number without units.) The physical property may be denoted by a symbol, but the symbol does *not* imply a particular choice of units. For instance, this e-book uses the symbol ρ for density, but ρ can be expressed in any units having the dimensions of mass divided by volume.

A simple example illustrates the use of quantity calculus. We may express the density of water at $25\ ^\circ\mathrm{C}$ to four significant digits in SI base units by the equation

$$\rho = 9.970 \times 10^2\ \mathrm{kg\ m^{-3}} \tag{1.2.2}$$

and in different density units by the equation

$$\rho = 0.9970\ \mathrm{g\ cm^{-3}} \tag{1.2.3}$$

We may divide both sides of the last equation by $1\ \mathrm{g\ cm^{-3}}$ to obtain a new equation

$$\rho/\mathrm{g\ cm^{-3}} = 0.9970 \tag{1.2.4}$$

Now the pure number 0.9970 appearing in this equation is the number of grams in one cubic centimeter of water, so we may call the ratio $\rho/\mathrm{g\ cm^{-3}}$ "the number of grams per cubic centimeter." By the same reasoning, $\rho/\mathrm{kg\ m^{-3}}$ is the number of kilograms per cubic meter. In general, a physical quantity divided by particular units for the physical quantity is a pure number representing the number of those units.

> Just as it would be incorrect to call ρ "the number of grams per cubic centimeter," because that would refer to a particular choice of units for ρ, the common practice of calling n "the number of moles" is also strictly speaking not correct. It is actually the ratio n/mol that is the number of moles.

In a table, the ratio $\rho/\mathrm{g\ cm^{-3}}$ makes a convenient heading for a column of density values because the column can then show pure numbers. Likewise, it is convenient to use $\rho/\mathrm{g\ cm^{-3}}$ as the label of a graph axis and to show pure numbers at the grid marks of the axis. You will see many examples of this usage in the tables and figures in this e-book.

A major advantage of using SI base units and SI derived units is that they are *coherent*. That is, values of a physical quantity expressed in different combinations of these units have the same numerical value.

For example, suppose we wish to evaluate the pressure of a gas according to the ideal gas equation

$$p = \frac{nRT}{V} \tag{1.2.5}$$
$$(\text{ideal gas})$$

This is the first equation that, like many others to follow, shows *conditions of validity* in parentheses immediately below the equation number at the right. Thus, Eq. 1.2.5 is valid for an ideal gas. In this equation, p, n, T, and V are the symbols for the physical quantities pressure, amount (amount of substance), thermodynamic temperature, and volume, respectively, and R is the gas constant.

The calculation of p for 5.000 moles of an ideal gas at a temperature of 298.15 kelvins, in a volume of 4.000 cubic meters, is

$$p = \frac{(5.000\ \mathrm{mol})(8.3145\ \mathrm{J\ K^{-1}\ mol^{-1}})(298.15\ \mathrm{K})}{4.000\ \mathrm{m^3}} = 3.099 \times 10^3\ \mathrm{J\ m^{-3}} \tag{1.2.6}$$

The mole and kelvin units cancel, and we are left with units of $\mathrm{J\ m^{-3}}$, a combination of an SI derived unit (the joule) and an SI base unit (the meter). The units $\mathrm{J\ m^{-3}}$ must have dimensions of pressure, but are not commonly used to express pressure.

To convert J m^{-3} to the SI derived unit of pressure, the pascal (Pa), we can use the following relations from Table 1.2:

$$1\,\mathrm{J} = 1\,\mathrm{N\ m} \qquad 1\,\mathrm{Pa} = 1\,\mathrm{N\ m^{-2}} \tag{1.2.7}$$

When we divide both sides of the first relation by $1\,\mathrm{J}$ and divide both sides of the second relation by $1\,\mathrm{N\ m^{-2}}$, we obtain the two new relations

$$1 = (1\,\mathrm{N\ m/J}) \qquad (1\,\mathrm{Pa/N\ m^{-2}}) = 1 \tag{1.2.8}$$

The ratios in parentheses are *conversion factors*. When a physical quantity is multiplied by a conversion factor that, like these, is equal to the pure number 1, the physical quantity changes its units but not its value. When we multiply Eq. 1.2.6 by both of these conversion factors, all units cancel except Pa:

$$p = (3.099 \times 10^3\,\mathrm{J\ m^{-3}}) \times (1\,\mathrm{N\ m/J}) \times (1\,\mathrm{Pa/N\ m^{-2}}) = 3.099 \times 10^3\,\mathrm{Pa} \tag{1.2.9}$$

This example illustrates the fact that to calculate a physical quantity, we can simply enter into a calculator numerical values expressed in SI units, and the result is the numerical value of the calculated quantity expressed in SI units. In other words, as long as we use only SI base units and SI derived units (without prefixes), *all conversion factors are unity*.

Of course we do not have to limit the calculation to SI units. Suppose we wish to express the calculated pressure in torrs, a non-SI unit. In this case, using a conversion factor obtained from the definition of the torr in Table 1.3, the calculation becomes

$$p = (3.099 \times 10^3\,\mathrm{Pa}) \times (760\,\mathrm{Torr}/101,325\,\mathrm{Pa}) = 23.24\,\mathrm{Torr} \tag{1.2.10}$$

This page titled 1.2: Quantity Calculus is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Howard DeVoe via source content that was edited to the style and standards of the LibreTexts platform.

- 1.2: Quantity Calculus by Howard DeVoe is licensed CC BY 4.0. Original source: https://www2.chem.umd.edu/thermobook.

1.3: Dimensional Analysis

Sometimes you can catch an error in the form of an equation or expression, or in the dimensions of a quantity used for a calculation, by checking for dimensional consistency. Here are some rules that must be satisfied:

- In this e-book the *differential* of a function, such as $\mathrm{d}\ f$, refers to an *infinitesimal* quantity. If one side of an equation is an infinitesimal quantity, the other side must also be. Thus, the equation $\mathrm{d}\ f = a\,\mathrm{d}\ x + b\,\mathrm{d}y$ (where ax and by have the same dimensions as f) makes mathematical sense, but $\mathrm{d}\ f = ax + b\,\mathrm{d}y$ does not.

  Derivatives, partial derivatives, and integrals have dimensions that we must take into account when determining the overall dimensions of an expression that includes them. For instance:

  - Some examples of applying these principles are given here using symbols described in Sec. 1.2.

    Example 1. Since the gas constant R may be expressed in units of J K^{-1} mol^{-1}, it has dimensions of energy divided by thermodynamic temperature and amount. Thus, RT has dimensions of energy divided by amount, and nRT has dimensions of energy. The products RT and nRT appear frequently in thermodynamic expressions.

    Example 3. Find the dimensions of the constants a and b in the van der Waals equation

    $$p = \frac{nRT}{V - nb} - \frac{n^2 a}{V^2} \tag{1.3.1}$$

    Dimensional analysis tells us that, because nb is subtracted from V, nb has dimensions of volume and therefore b has dimensions of volume/amount. Furthermore, since the right side of the equation is a difference of two terms, these terms have the same dimensions as the left side, which is pressure. Therefore, the second term $n^2 a/V^2$ has dimensions of pressure, and a has dimensions of pressure $\times$ volume^2 $\times$ amount^{-2}.

    Example 4. Consider an equation of the form

    $$\left(\frac{\partial \ln x}{\partial T}\right)_p = \frac{y}{R} \tag{1.3.2}$$

    What are the SI units of y? $\ln x$ is dimensionless, so the left side of the equation has the dimensions of $1/T$, and its SI units are K^{-1}. The SI units of the right side are therefore also K^{-1}. Since R has the units J K^{-1} mol^{-1}, the SI units of y are J K^{-2} mol^{-1}.

This page titled 1.3: Dimensional Analysis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Howard DeVoe via source content that was edited to the style and standards of the LibreTexts platform.

- 1.3: Dimensional Analysis by Howard DeVoe is licensed CC BY 4.0. Original source: https://www2.chem.umd.edu/thermobook.

1.4: Chapter 1 Problem

1.1
Consider the following equations for the pressure of a real gas. For each equation, find the dimensions of the constants a and b and express these dimensions in SI units.

(a) The Dieterici equation:

$$p = \frac{RTe^{-(an/VRT)}}{(V/n) - b} \tag{1.4.1}$$

(b) The Redlich–Kwong equation:

$$p = \frac{RT}{(V/n) - b} - \frac{an^2}{T^{1/2}V(V + nb)} \tag{1.4.2}$$

This page titled 1.4: Chapter 1 Problem is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Howard DeVoe via source content that was edited to the style and standards of the LibreTexts platform.

- 1.4: Chapter 1 Problem by Howard DeVoe is licensed CC BY 4.0. Original source: https://www2.chem.umd.edu/thermobook.

2: GENERAL CHEMISTRY REVIEW

2.1: Gases

2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law

2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions

2.1.3: The Kinetic-Molecular Theory

2.1.4: Non-Ideal Gas Behavior

2.2: Thermodynamics

2.2.1: Energy Basics

2.2.2: Enthalpy

2.2.3: Spontaneity

2.2.4: Entropy

2.2.5: The Second and Third Laws of Thermodynamics

2.2.6: Free Energy

2.3: Kinetics

2.3.1: Chemical Reaction Rates

2.3.2: Factors Affecting Reaction Rates

2.3.3: Rate Laws

2.3.4: Integrated Rate Laws

2.3.5: Collision Theory

2.3.6: Reaction Mechanisms

2.3.7: Catalysis

2.4: Using logarithms - Log vs. Ln

This page titled 2: General Chemistry Review is shared under a license and was authored, remixed, and/or curated by LibreTexts.

2.1: GASES

2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions
2.1.3: The Kinetic-Molecular Theory
2.1.4: Non-Ideal Gas Behavior

This page titled 2.1: Gases is shared under a license and was authored, remixed, and/or curated by LibreTexts.

2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law

Learning Objectives

By the end of this section, you will be able to:

- Identify the mathematical relationships between the various properties of gases
- Use the ideal gas law, and related gas laws, to compute the values of various gas properties under specified conditions

During the seventeenth and especially eighteenth centuries, driven both by a desire to understand nature and a quest to make balloons in which they could fly (Figure 2.1.1.1), a number of scientists established the relationships between the macroscopic physical properties of gases, that is, pressure, volume, temperature, and amount of gas. Although their measurements were not precise by today's standards, they were able to determine the mathematical relationships between pairs of these variables (e.g., pressure and temperature, pressure and volume) that hold for an *ideal* gas—a hypothetical construct that real gases approximate under certain conditions. Eventually, these individual laws were combined into a single equation—the *ideal gas law*—that relates gas quantities for gases and is quite accurate for low pressures and moderate temperatures. We will consider the key developments in individual relationships (for pedagogical reasons not quite in historical order), then put them together in the ideal gas law.

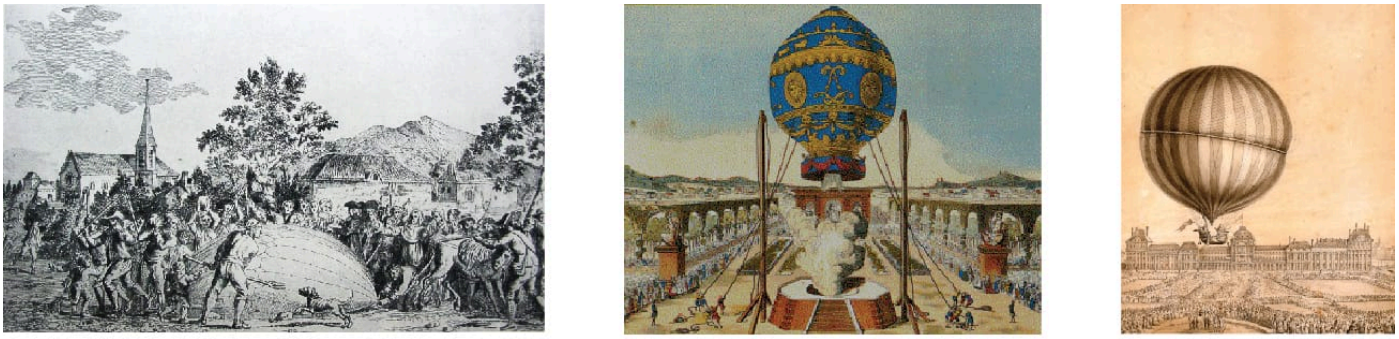


Figure 2.1.1.1: In 1783, the first (a) hydrogen-filled balloon flight, (b) manned hot air balloon flight, and (c) manned hydrogen-filled balloon flight occurred. When the hydrogen-filled balloon depicted in (a) landed, the frightened villagers of Gonesse reportedly destroyed it with pitchforks and knives. The launch of the latter was reportedly viewed by 400,000 people in Paris.

Pressure and Temperature: Amontons's Law

Imagine filling a rigid container attached to a pressure gauge with gas and then sealing the container so that no gas may escape. If the container is cooled, the gas inside likewise gets colder and its pressure is observed to decrease. Since the container is rigid and tightly sealed, both the volume and number of moles of gas remain constant. If we heat the sphere, the gas inside gets hotter (Figure 2.1.1.2) and the pressure increases.

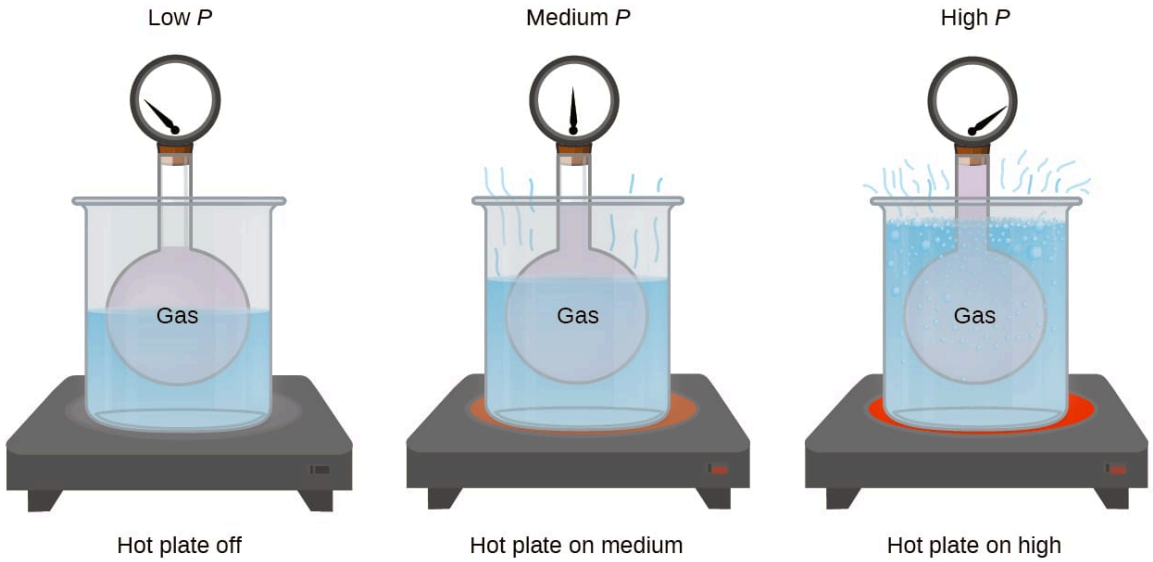


Figure 2.1.1.2: The effect of temperature on gas pressure: When the hot plate is off, the pressure of the gas in the sphere is relatively low. As the gas is heated, the pressure of the gas in the sphere increases.

This relationship between temperature and pressure is observed for any sample of gas confined to a constant volume. An example of experimental pressure-temperature data is shown for a sample of air under these conditions in Figure 2.1.1.3. We find that temperature and pressure are linearly related, and if the temperature is on the kelvin scale, then *P* and *T* are directly proportional (again, when *volume and moles of gas are held constant*); if the temperature on the kelvin scale increases by a certain factor, the gas pressure increases by the same factor.

Temperature (°C)	Temperature (K)	Pressure (kPa)
−100	173	36.0
−50	223	46.4
0	273	56.7
50	323	67.1
100	373	77.5
150	423	88.0

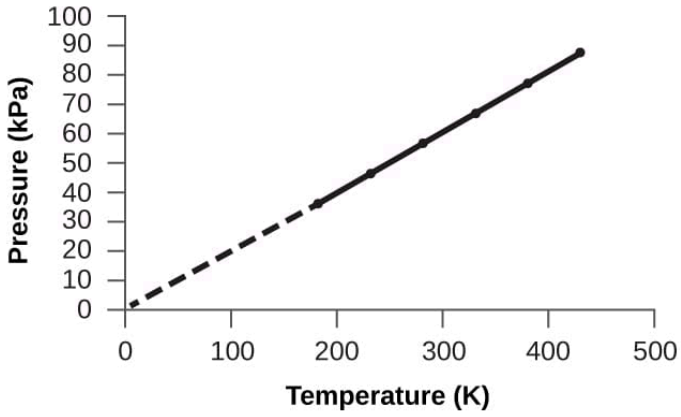


Figure 2.1.1.3: For a constant volume and amount of air, the pressure and temperature are directly proportional, provided the temperature is in kelvin. (Measurements cannot be made at lower temperatures because of the condensation of the gas.) When this line is extrapolated to lower pressures, it reaches a pressure of 0 at –273 °C, which is 0 on the kelvin scale and the lowest possible temperature, called absolute zero.

Guillaume Amontons was the first to empirically establish the relationship between the pressure and the temperature of a gas (~1700), and Joseph Louis Gay-Lussac determined the relationship more precisely (~1800). Because of this, the *P*-*T* relationship for gases is known as either **Amontons's law** or **Gay-Lussac's law**. Under either name, it states that *the pressure of a given amount of gas is directly proportional to its temperature on the kelvin scale when the volume is held constant*. Mathematically, this can be written:

$$P \propto T \text{ or } P = \text{ constant } \times T \text{ or } P = k \times T$$

where $\propto$ means "is proportional to," and k is a proportionality constant that depends on the identity, amount, and volume of the gas.

For a confined, constant volume of gas, the $\frac{P}{T}$ ratio is therefore constant (i.e., $\frac{P}{T} = k$). If the gas is initially in "Condition 1" (with *P*

= P_1 and T = T_1), and then changes to "Condition 2" (with P = P_2 and T = T_2), we have that $\frac{P_1}{T_1} = k$ and $\frac{P_2}{T_2} = k$ which reduces to

$$\frac{P_1}{T_1} = \frac{P_2}{T_2}.$$

This equation is useful for pressure-temperature calculations for a confined gas at constant volume. Note that temperatures must be on the kelvin scale for any gas law calculations (0 on the kelvin scale and the lowest possible temperature is called **absolute zero**). (Also note that there are at least three ways we can describe how the pressure of a gas changes as its temperature changes: We can use a table of values, a graph, or a mathematical equation.)

Example 2.1.1.1: Predicting Change in Pressure with Temperature

A can of hair spray is used until it is empty except for the propellant, isobutane gas.

a. On the can is the warning "Store only at temperatures below 120 °F (48.8 °C). Do not incinerate." Why?
b. The gas in the can is initially at 24 °C and 360 kPa, and the can has a volume of 350 mL. If the can is left in a car that reaches 50 °C on a hot day, what is the new pressure in the can?

Solution

(a) The can contains an amount of isobutane gas at a constant volume, so if the temperature is increased by heating, the pressure will increase proportionately. High temperature could lead to high pressure, causing the can to burst. (Also, isobutane is combustible, so incineration could cause the can to explode.)

(b) We are looking for a pressure change due to a temperature change at constant volume, so we will use Amontons's/Gay-Lussac's law. Taking P_1 and T_1 as the initial values, T_2 as the temperature where the pressure is unknown and P_2 as the unknown pressure, and converting °C to K, we have:

$$\frac{P_1}{T_1} = \frac{P_2}{T_2} \text{ which means that } \frac{360 \text{ kPa}}{297 \text{ K}} = \frac{P_2}{323 \text{ K}}$$

Rearranging and solving gives:

$$P_2 = \frac{360 \text{ kPa} \times 323 \cancel{\text{K}}}{297 \cancel{\text{K}}} = 390 \text{ kPa}$$

Exercise 2.1.1.1

A sample of nitrogen, N_2, occupies 45.0 mL at 27 °C and 600 torr. What pressure will it have if cooled to –73 °C while the volume remains constant?

Answer available at the end of this section.

Volume and Temperature: Charles's Law

If we fill a balloon with air and seal it, the balloon contains a specific amount of air at atmospheric pressure, let's say 1 atm. If we put the balloon in a refrigerator, the gas inside gets cold and the balloon shrinks (although both the amount of gas and its pressure remain constant). If we make the balloon very cold, it will shrink a great deal, and it expands again when it warms up.

Link to Learning

This video shows how cooling and heating a gas causes its volume to decrease or increase, respectively.

These examples of the effect of temperature on the volume of a given amount of a confined gas at constant pressure are true in

general: The volume increases as the temperature increases, and decreases as the temperature decreases. Volume-temperature data for a 1-mole sample of methane gas at 1 atm are listed and graphed in Figure 9.12.

Temperature (°C)	Temperature (K)	Volume (L)
−3	270	22
−23	250	21
−53	220	18
−162	111	9

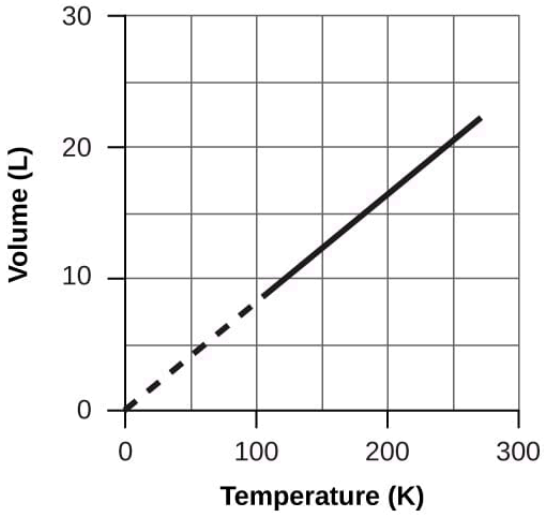


Figure 9.12 The volume and temperature are linearly related for 1 mole of methane gas at a constant pressure of 1 atm. If the temperature is in kelvin, volume and temperature are directly proportional. The line stops at 111 K because methane liquefies at this temperature; when extrapolated, it intersects the graph's origin, representing a temperature of absolute zero.

The relationship between the volume and temperature of a given amount of gas at constant pressure is known as Charles's law in recognition of the French scientist and balloon flight pioneer Jacques Alexandre César Charles. **Charles's law** states that *the volume of a given amount of gas is directly proportional to its temperature on the kelvin scale when the pressure is held constant*.

Mathematically, this can be written as:

$$V \alpha T \text{ or } V = \text{ constant } \cdot T \text{ or } V = k \cdot T \text{ or } V_1/T_1 = V_2/T_2$$

with k being a proportionality constant that depends on the amount and pressure of the gas.

For a confined, constant pressure gas sample, $\frac{V}{T}$ is constant (i.e., the ratio = *k*), and as seen with the *P*-*T* relationship, this leads to another form of Charles's law:

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}. \tag{2.1.1.1}$$

Example 2.1.1.2: Predicting Change in Volume with Temperature

A sample of carbon dioxide, CO_2, occupies 0.300 L at 10 °C and 750 torr. What volume will the gas have at 30 °C and 750 torr?

Solution

Because we are looking for the volume change caused by a temperature change at constant pressure, this is a job for Charles's law. Taking V_1 and T_1 as the initial values, T_2 as the temperature at which the volume is unknown and V_2 as the unknown volume, and converting °C into K we have:

$$\frac{V_1}{T_1} = \frac{V_2}{T_2} \text{ which means that } \frac{0.300L}{283K} = \frac{V_2}{303K}$$

Rearranging and solving gives:

$$V_2 = \frac{0.300L \times 303 \cancel{K}}{283 \cancel{K}} = 0.321 \text{ L}$$

This answer supports our expectation from Charles's law, namely, that raising the gas temperature (from 283 K to 303 K) at a constant pressure will yield an increase in its volume (from 0.300 L to 0.321 L).

Exercise 2.1.1.2

A sample of oxygen, O_2, occupies 32.2 mL at 30 °C and 452 torr. What volume will it occupy at –70 °C and the same pressure?

Answer available at the end of this section.

Example 2.1.1.3: Measuring Temperature with a Volume Change

Temperature is sometimes measured with a gas thermometer by observing the change in the volume of the gas as the temperature changes at constant pressure. The hydrogen in a particular hydrogen gas thermometer has a volume of 150.0 cm^3 when immersed in a mixture of ice and water (0.00 °C). When immersed in boiling liquid ammonia, the volume of the hydrogen, at the same pressure, is 131.7 cm^3. Find the temperature of boiling ammonia on the kelvin and Celsius scales.

Solution

A volume change caused by a temperature change at constant pressure means we should use Charles's law. Taking V_1 and T_1 as the initial values, T_2 as the temperature at which the volume is unknown and V_2 as the unknown volume, and converting °C into K we have:

$$\frac{V_1}{T_1} = \frac{V_2}{T_2} \text{ which means that } \frac{150.0cm^3}{273.15K} = \frac{131.7cm^3}{T_2}$$

Rearrangement gives

$$T_2 = \frac{131.7 \ \cancel{cm^3} \times 273.15 \text{ K}}{150.0 \ \cancel{cm^3}} = 239.8 \text{ K}$$

Subtracting 273.15 from 239.8 K, we find that the temperature of the boiling ammonia on the Celsius scale is –33.4 °C.

Exercise 2.1.1.3

What is the volume of a sample of ethane at 467 K and 1.1 atm if it occupies 405 mL at 298 K and 1.1 atm?

Answer available at the end of this section.

Volume and Pressure: Boyle's Law

If we partially fill an airtight syringe with air, the syringe contains a specific amount of air at constant temperature, say 25 °C. If we slowly push in the plunger while keeping temperature constant, the gas in the syringe is compressed into a smaller volume and its pressure increases; if we pull out the plunger, the volume increases and the pressure decreases. This example of the effect of volume on the pressure of a given amount of a confined gas is true in general. Decreasing the volume of a contained gas will increase its pressure, and increasing its volume will decrease its pressure. In fact, if the volume increases by a certain factor, the pressure decreases by the same factor, and vice versa. Volume-pressure data for an air sample at room temperature are graphed in Figure 2.1.1.4.

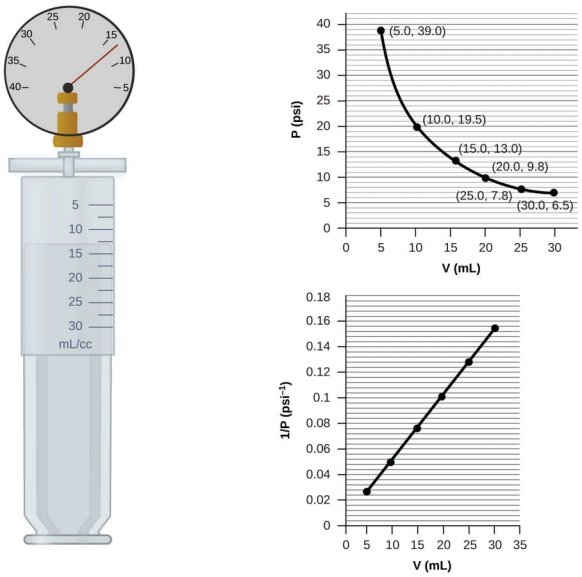


Figure 2.1.1.4: When a gas occupies a smaller volume, it exerts a higher pressure; when it occupies a larger volume, it exerts a lower pressure (assuming the amount of gas and the temperature do not change). Since *P* and *V* are inversely proportional, a graph of $\frac{1}{P}$ vs. *V* is linear.

Unlike the *P*-*T* and *V*-*T* relationships, pressure and volume are not directly proportional to each other. Instead, *P* and *V* exhibit inverse proportionality: Increasing the pressure results in a decrease of the volume of the gas. Mathematically this can be written:

$$P \propto \frac{1}{V} \tag{2.1.1.2}$$

or

$$P = k \cdot \frac{1}{V} \tag{2.1.1.3}$$

or

$$P \cdot V = k \tag{2.1.1.4}$$

or

$$P_1 V_1 = P_2 V_2 \tag{2.1.1.5}$$

with *k* being a constant. Graphically, this relationship is shown by the straight line that results when plotting the inverse of the pressure

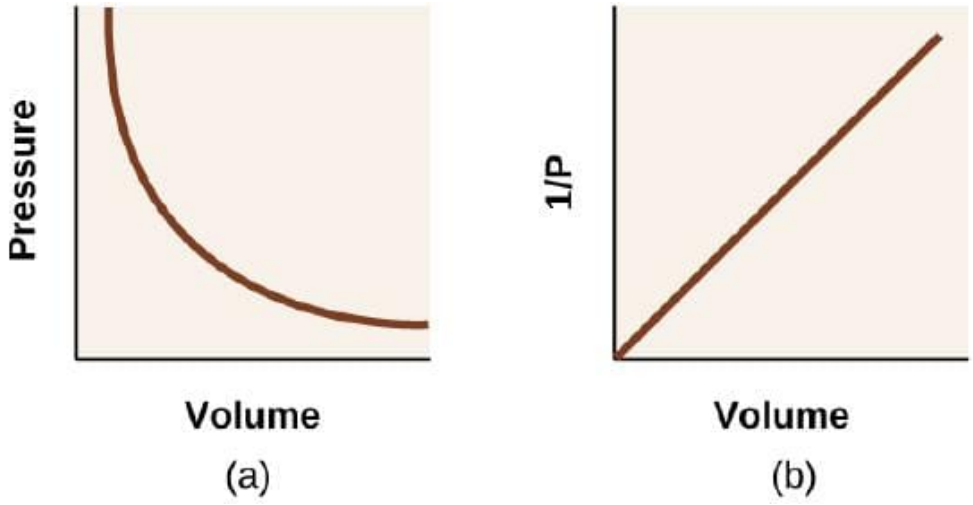


Figure 2.1.1.5: The relationship between pressure and volume is inversely proportional. (a) The graph of *P* vs. *V* is a hyperbola, whereas (b) the graph of $\frac{1}{P}$ vs. *V* is linear.

The relationship between the volume and pressure of a given amount of gas at constant temperature was first published by the English natural philosopher Robert Boyle over 300 years ago. It is summarized in the statement now known as **Boyle's law**: *The volume of a given amount of gas held at constant temperature is inversely proportional to the pressure under which it is measured.*

Example 2.1.1.4: Volume of a Gas Sample

The sample of gas in Figure 2.1.1.4: has a volume of 15.0 mL at a pressure of 13.0 psi. Determine the pressure of the gas at a volume of 7.5 mL, using:

a. the *P*-*V* graph in Figure 2.1.1.5
b. the $\frac{1}{P}$ vs. V graph in Figure 2.1.1.5
c. the Boyle's law equations (Equations 2.1.1.2 - 2.1.1.5)

Comment on the likely accuracy of each method.

Solution

(a) Estimating from the *P*-*V* graph gives a value for *P* somewhere around 27 psi.

(b) Estimating from the \frac{1}{P}\) versus V graph give a value of about 26 psi.

(c) From Boyle's law, we know that the product of pressure and volume (*PV*) for a given sample of gas at a constant temperature is always equal to the same value. Therefore we have $P_1V_1 = k$ and $P_2V_2 = k$ which means that $P_1V_1 = P_2V_2$.

Using P_1 and V_1 as the known values 13.0 psi and 15.0 mL, P_2 as the pressure at which the volume is unknown, and V_2 as the unknown volume, we have:

$$P_1\ V_1 = P_2\ V_2$$
$$(13.0\ \text{psi}) \times (15.0\ \text{mL} = P_2 \times (7.5\ \text{mL})$$

Solving:

$$P_2 = \frac{13.0\ \text{psi} \times 15.0\ \cancel{\text{mL}}}{7.5\ \cancel{\text{mL}}} = 26\ \text{psi}$$

It was more difficult to estimate well from the *P*-*V* graph, so (a) is likely more inaccurate than (b) or (c). The calculation will be as accurate as the equation and measurements allow.

Exercise 2.1.1.4

The sample of gas in Figure 2.1.1.4 has a volume of 30.0 mL at a pressure of 6.5 psi. Determine the volume of the gas at a pressure of 11.0 psi, using:

a. the *P*-*V* graph in Figure 2.1.1.5
b. the $\frac{1}{P}$ vs. \(*V*\) graph in Figure 2.1.1.5
c. the Boyle's law equations (2.1.1.2 - 2.1.1.5)

Comment on the likely accuracy of each method.

Answer available at the end of this section.

Chemistry in Everyday Life: Breathing and Boyle's Law

What do you do about 20 times per minute for your whole life, without break, and often without even being aware of it? The answer, of course, is respiration, or breathing. How does it work? It turns out that the gas laws apply here. Your lungs take in gas that your body needs (oxygen) and get rid of waste gas (carbon dioxide). Lungs are made of spongy, stretchy tissue that expands and contracts while you breathe. When you inhale, your diaphragm and intercostal muscles (the muscles between your ribs) contract, expanding your chest cavity and making your lung volume larger. The increase in volume leads to a decrease in pressure (Boyle's law). This causes air to flow into the lungs (from high pressure to low pressure). When you exhale, the process reverses: Your diaphragm and rib muscles relax, your chest cavity contracts, and your lung volume decreases, causing the pressure to increase (Boyle's law again), and air flows out of the lungs (from high pressure to low pressure). You then breathe in and out again, and again, repeating this Boyle's law cycle for the rest of your life (Figure 2.1.1.6).

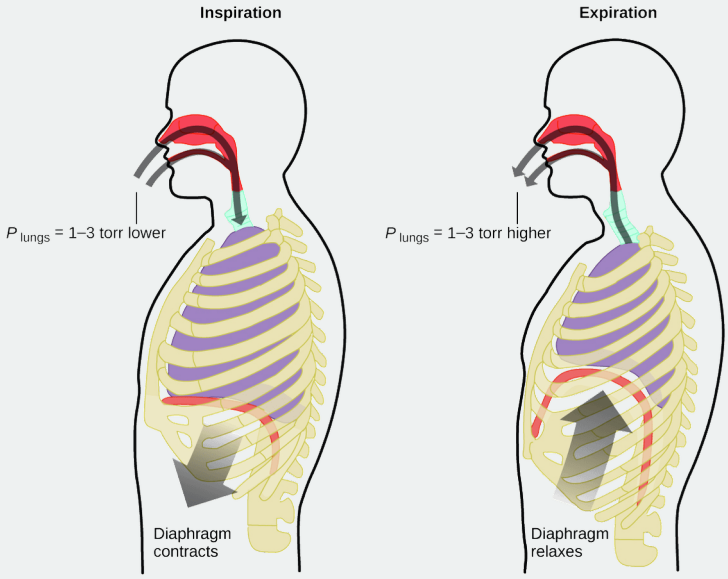


Figure 2.1.1.6: Breathing occurs because expanding and contracting lung volume creates small pressure differences between your lungs and your surroundings, causing air to be drawn into and forced out of your lungs.

Moles of Gas and Volume: Avogadro's Law

The Italian scientist Amedeo Avogadro advanced a hypothesis in 1811 to account for the behavior of gases, stating that equal volumes of all gases, measured under the same conditions of temperature and pressure, contain the same number of molecules. Over time, this relationship was supported by many experimental observations as expressed by **Avogadro's law**: *For a confined gas, the volume (V) and number of moles (n) are directly proportional if the pressure and temperature both remain constant*.

In equation form, this is written as:

$$V \propto n \tag{2.1.1.6}$$

or

$$V = k \times n \tag{2.1.1.7}$$

or

$$\frac{V_1}{n_1} = \frac{V_2}{n_2} \tag{2.1.1.8}$$

Mathematical relationships can also be determined for the other variable pairs, such as *P* versus *n*, and *n* versus *T*.

Link to Learning

Visit this interactive PhET simulation to investigate the relationships between pressure, volume, temperature, and amount of gas. Use the simulation to examine the effect of changing one parameter on another while holding the other parameters constant (as described in the preceding sections on the various gas laws).

The Ideal Gas Law

To this point, four separate laws have been discussed that relate pressure, volume, temperature, and the number of moles of the gas:

- Boyle's law: PV = constant at constant *T* and *n*
- Amontons's law: $\frac{P}{T}$ = constant at constant *V* and *n*
- Charles's law: $\frac{V}{T}$ = constant at constant *P* and *n*
- Avogadro's law: $\frac{V}{n}$ = constant at constant *P* and *T*

Combining these four laws yields the **ideal gas law**, a relation between the pressure, volume, temperature, and number of moles of a gas:

$$PV = nRT$$

where *P* is the pressure of a gas, *V* is its volume, *n* is the number of moles of the gas, *T* is its temperature on the kelvin scale, and *R* is a constant called the **ideal gas constant** or the universal gas constant. The units used to express pressure, volume, and temperature will determine the proper form of the gas constant as required by dimensional analysis, the most commonly encountered values being 0.08206 L atm mol^{-1} K^{-1} and 8.314 kPa L mol^{-1} K^{-1}.

Gases whose properties of *P*, *V*, and *T* are accurately described by the ideal gas law (or the other gas laws) are said to exhibit *ideal behavior* or to approximate the traits of an **ideal gas**. An ideal gas is a hypothetical construct that may be used along with *kinetic molecular theory* to effectively explain the gas laws as will be described in a later module of this chapter. Although all the calculations presented in this module assume ideal behavior, this assumption is only reasonable for gases under conditions of relatively low pressure and high temperature. In the final module of this chapter, a modified gas law will be introduced that accounts for the *non-ideal* behavior observed for many gases at relatively high pressures and low temperatures.

The ideal gas equation contains five terms, the gas constant *R* and the variable properties *P*, *V*, *n*, and *T*. Specifying any four of these terms will permit use of the ideal gas law to calculate the fifth term as demonstrated in the following example exercises.

Example 2.1.1.5: Using the Ideal Gas Law

Methane, CH_4, is being considered for use as an alternative automotive fuel to replace gasoline. One gallon of gasoline could be replaced by 655 g of CH_4. What is the volume of this much methane at 25 °C and 745 torr?

Solution

We must rearrange *PV* = *nRT* to solve for *V*:

$$V = \frac{nRT}{P}$$

If we choose to use *R* = 0.08206 L atm mol^{-1} K^{-1}, then the amount must be in moles, temperature must be in kelvin, and pressure must be in atm.

Converting into the "right" units:

$$n = 655 \text{ g } CH_4 \times \frac{1 \text{ mol}}{16.043 \text{ g } CH_4} = 40.8 \text{ mol}$$

$$T = 25^{\circ}C + 273 = 298 \text{ K}$$

$$P = 745 \text{ torr} \times \frac{1 \text{ atm}}{760 \text{ torr}} = 0.980 \text{ atm}$$

$$V = \frac{nRT}{P} = \frac{(40.8 \text{ mol})\left(0.08206 \text{ L atm mol}^{-1}\text{K}^{-1}\right)(298 \text{ K})}{0.980 \text{ atm}} = 1.02 \times 10^3 \text{ L}$$

It would require 1020 L (269 gal) of gaseous methane at about 1 atm of pressure to replace 1 gal of gasoline. It requires a large container to hold enough methane at 1 atm to replace several gallons of gasoline.

Exercise 2.1.1.5

Calculate the pressure in bar of 2520 moles of hydrogen gas stored at 27 °C in the 180-L storage tank of a modern hydrogen-powered car.

Answer available at the end of this section.

If the number of moles of an ideal gas are kept constant under two different sets of conditions, a useful mathematical relationship called the **combined gas law** is obtained:

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$$

using units of atm, L, and K. Both sets of conditions are equal to the product of *n R* (where *n* = the number of moles of the gas and *R* is the ideal gas law constant).

Example 2.1.1.6: Using the Combined Gas Law

When filled with air, a typical scuba tank with a volume of 13.2 L has a pressure of 153 atm (Figure 2.1.1.7). If the water temperature is 27 °C, how many liters of air will such a tank provide to a diver's lungs at a depth of approximately 70 feet in the ocean where the pressure is 3.13 atm?

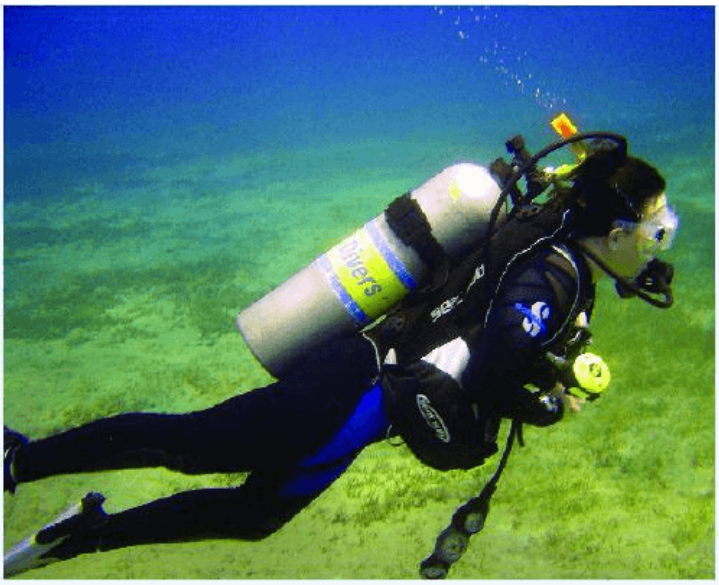


Figure 2.1.1.7: Scuba divers use compressed air to breathe while underwater. (credit: modification of work by Mark Goodchild)

Letting *1* represent the air in the scuba tank and *2* represent the air in the lungs, and noting that body temperature (the temperature the air will be in the lungs) is 37 °C, we have:

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2} \rightarrow \frac{(153atm)(13.2L)}{(300K)} = \frac{(3.13atm)\,(V_2)}{(310K)}$$

Solving for V_2:

$$V_2 = \frac{(153\text{ atm }\,)(13.2L)(310K)}{(300K)(3.13\text{ atm }\,)} = 667L$$

(Note: Be advised that this particular example is one in which the assumption of ideal gas behavior is not very reasonable, since it involves gases at relatively high pressures and low temperatures. Despite this limitation, the calculated volume can be viewed as a good “ballpark” estimate.)

Exercise 2.1.1.6

A sample of ammonia is found to occupy 0.250 L under laboratory conditions of 27 °C and 0.850 atm. Find the volume of this sample at 0 °C and 1.00 atm.

Answer available at the end of this section.

Chemistry in Everyday Life: The Interdependence between Ocean Depth and Pressure in Scuba Diving

Whether scuba diving at the Great Barrier Reef in Australia (shown in Figure 2.1.1.8) or in the Caribbean, divers must understand how pressure affects a number of issues related to their comfort and safety.



Figure 2.1.1.8: Scuba divers, whether at the Great Barrier Reef or in the Caribbean, must be aware of buoyancy, pressure equalization, and the amount of time they spend underwater, to avoid the risks associated with pressurized gases in the body. (credit: Kyle Taylor)

Pressure increases with ocean depth, and the pressure changes most rapidly as divers reach the surface. The pressure a diver experiences is the sum of all pressures above the diver (from the water and the air). Most pressure measurements are given in units of atmospheres, expressed as “atmospheres absolute” or ATA in the diving community: Every 33 feet of salt water represents 1 ATA of pressure in addition to 1 ATA of pressure from the atmosphere at sea level. As a diver descends, the increase in pressure causes the body’s air pockets in the ears and lungs to compress; on the ascent, the decrease in pressure causes these air pockets to expand, potentially rupturing eardrums or bursting the lungs. Divers must therefore undergo equalization by adding air to body airspaces on the descent by breathing normally and adding air to the mask by breathing out of the nose or adding air to the ears and sinuses by equalization techniques; the corollary is also true on ascent, divers must release air from the body to maintain equalization. Buoyancy, or the ability to control whether a diver sinks or floats, is controlled by the buoyancy compensator (BCD). If a diver is ascending, the air in their BCD expands because of lower pressure according to Boyle’s law (decreasing the pressure of gases increases the volume). The expanding air increases the buoyancy of the diver, and they begin to ascend. The diver must vent air from the BCD or risk an uncontrolled ascent that could rupture the lungs. In descending, the increased pressure causes the air in the BCD to compress and the diver sinks much more quickly; the diver must add air to the BCD or risk an uncontrolled descent, facing much higher pressures near the ocean floor. The pressure also impacts how long a diver can stay underwater before ascending. The deeper a diver dives, the more compressed the air that is breathed because of increased pressure: If a diver dives 33 feet, the pressure is 2 ATA and the air would be compressed to one-half of its original volume. The diver uses up available air twice as fast as at the surface.

Standard Conditions of Temperature and Pressure

We have seen that the volume of a given quantity of gas and the number of molecules (moles) in a given volume of gas vary with changes in pressure and temperature. Chemists sometimes make comparisons against a **standard temperature and pressure (STP)** for reporting properties of gases: 273.15 K and 1 atm (101.325 kPa).[1] At STP, one mole of an ideal gas has a volume of about 22.4 L—this is referred to as the **standard molar volume** (Figure 2.1.1.9).

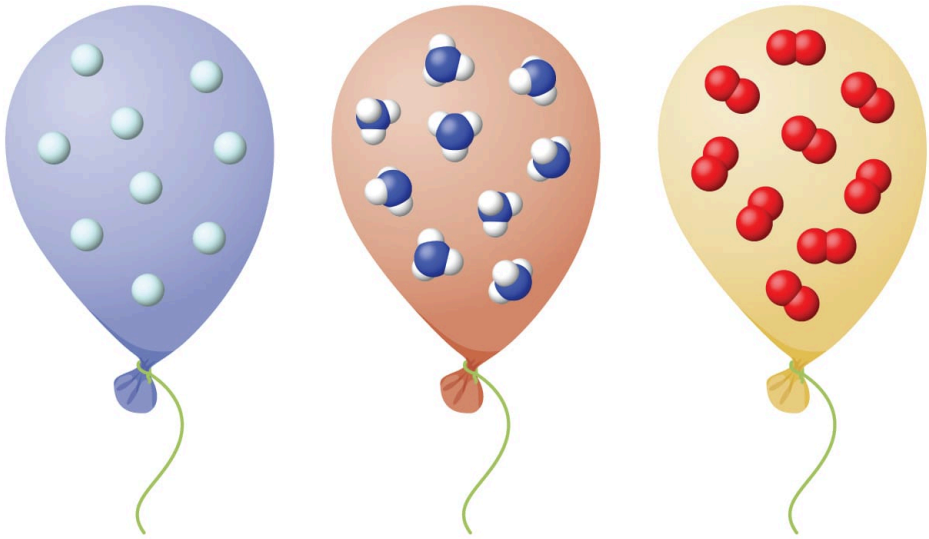


Figure 2.1.1.9: Regardless of its chemical identity, one mole of gas behaving ideally occupies a volume of ~22.4 L at STP.

Footnotes

- 1The IUPAC definition of standard pressure was changed from 1 atm to 1 bar (100 kPa) in 1982, but the prior definition remains in use by many literature resources and will be used in this text.

Answers

Exercise 2.1.1.1

400 torr

Exercise 2.1.1.2

21.6 mL

Exercise 2.1.1.3

635 mL

Exercise 2.1.1.4

(a) about 17–18 mL; (b) ~18 mL; (c) 17.7 mL; it was more difficult to estimate well from the *P*-*V* graph, so (a) is likely more inaccurate than (b); the calculation will be as accurate as the equation and measurements allow

Exercise 2.1.1.5

350 bar

Exercise 2.1.1.6

0.193 L

This page titled 2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 9.2: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions

Learning Objectives

By the end of this section, you will be able to:

- Use the ideal gas law to compute gas densities and molar masses
- Perform stoichiometric calculations involving gaseous substances
- State Dalton's law of partial pressures and use it in calculations involving gaseous mixtures

The study of the chemical behavior of gases was part of the basis of perhaps the most fundamental chemical revolution in history. French nobleman Antoine Lavoisier, widely regarded as the "father of modern chemistry," changed chemistry from a qualitative to a quantitative science through his work with gases. He discovered the law of conservation of matter, discovered the role of oxygen in combustion reactions, determined the composition of air, explained respiration in terms of chemical reactions, and more. He was a casualty of the French Revolution, guillotined in 1794. Of his death, mathematician and astronomer Joseph-Louis Lagrange said, "It took the mob only a moment to remove his head; a century will not suffice to reproduce it."[2] Much of the knowledge we do have about Lavoisier's contributions is due to his wife, Marie-Anne Paulze Lavoisier, who worked with him in his lab. A trained artist fluent in several languages, she created detailed illustrations of the equipment in his lab, and translated texts from foreign scientists to complement his knowledge. After his execution, she was instrumental in publishing Lavoisier's major treatise, which unified many concepts of chemistry and laid the groundwork for significant further study.

As described in an earlier chapter of this text, we can turn to chemical stoichiometry for answers to many of the questions that ask "How much?" The essential property involved in such use of stoichiometry is the amount of substance, typically measured in moles (*n*). For gases, molar amount can be derived from convenient experimental measurements of pressure, temperature, and volume. Therefore, these measurements are useful in assessing the stoichiometry of pure gases, gas mixtures, and chemical reactions involving gases. This section will not introduce any new material or ideas, but will provide examples of applications and ways to integrate concepts already discussed.

Gas Density and Molar Mass

The ideal gas law described previously in this chapter relates the properties of pressure *P*, volume *V*, temperature *T*, and molar amount *n*. This law is universal, relating these properties in identical fashion regardless of the chemical identity of the gas:

$$PV = nRT$$

The density *d* of a gas, on the other hand, is determined by its identity. As described in another chapter of this text, the density of a substance is a characteristic property that may be used to identify the substance.

$$d = \frac{m}{V}$$

Rearranging the ideal gas equation to isolate *V* and substituting into the density equation yields

$$d = \frac{mP}{nRT} = \left(\frac{m}{n}\right)\frac{P}{RT}$$

The ratio *m*/*n* is the definition of molar mass, M:

$$\mathcal{M} = \frac{m}{n}$$

The density equation can then be written

$$d = \frac{\mathscr{M}P}{RT}$$

This relation may be used for calculating the densities of gases of known identities at specified values of pressure and temperature as demonstrated in Example 2.1.2.1.

Example 2.1.2.1: Measuring Gas Density

What is the density of molecular nitrogen gas at STP?

Solution

The molar mass of molecular nitrogen, N_2, is 28.01 g/mol. Substituting this value along with standard temperature and pressure into the gas density equation yields

$$d = \frac{\mathscr{M}P}{RT} = \frac{(28.01\ \text{g/mol}\ )(1.00\ \text{atm})}{\left(0.0821\ \text{L}\cdot\text{atm}\cdot\text{mol}^{-1}\text{K}^{-1}\right)(273\ \text{K})} = 1.25\ \text{g}\ /\ \text{L}$$

Exercise 2.1.2.1

What is the density of molecular hydrogen gas at 17.0 °C and a pressure of 760 torr?

Answer available at the end of this section.

When the identity of a gas is unknown, measurements of the mass, pressure, volume, and temperature of a sample can be used to calculate the molar mass of the gas (a useful property for identification purposes). Combining the ideal gas equation

$$PV = nRT$$

and the definition of molar mass

$$\mathscr{M} = \frac{m}{n}$$

yields the following equation:

$$\mathscr{M} = \frac{mRT}{PV}$$

Determining the molar mass of a gas via this approach is demonstrated in Example 2.1.2.2.

Example 2.1.2.2: Determining the Molecular Formula of a Gas from its Molar Mass and Empirical Formula

Cyclopropane, a gas once used with oxygen as a general anesthetic, is composed of 85.7% carbon and 14.3% hydrogen by mass. Find the empirical formula. If 1.56 g of cyclopropane occupies a volume of 1.00 L at 0.984 atm and 50 °C, what is the molecular formula for cyclopropane?

Solution

First determine the empirical formula of the gas. Assume 100 g and convert the percentage of each element into grams. Determine the number of moles of carbon and hydrogen in the 100-g sample of cyclopropane. Divide by the smallest number of moles to relate the number of moles of carbon to the number of moles of hydrogen.

$$85.7\ \text{g}\ C \times \frac{1\ \text{mol}\ C}{12.01\ \text{g}\ C} = 7.136\ \text{mol}\ C$$

$$14.3\ \text{g}\ H \times \frac{1\ \text{mol}\ H}{1.01\ \text{g}\ H} = 14.158\ \text{mol}\ H$$

In the last step, realize that the smallest whole number ratio is the empirical formula:

$$\frac{7.136}{7.136} = 1.00\ \text{mol}\ C$$

$$\frac{14.158}{7.136} = 1.98\ \text{mol}\ H$$

Empirical formula is CH_2 [empirical mass (EM) of 14.03 g/empirical unit].

Next, use the provided values for mass, pressure, temperature and volume to compute the molar mass of the gas:

$$\mathscr{M} = \frac{mRT}{PV} = \frac{(1.56\ \text{g})\left(0.0821\ \text{L}\cdot\text{atm}\cdot\text{mol}^{-1}\text{K}^{-1}\right)(323\ \text{K})}{(0.984\ \text{atm})(1.00\ \text{L})} = 42.0\ \text{g/mol}$$

Comparing the molar mass to the empirical formula mass shows how many empirical formula units make up a molecule:

$$\frac{\mathscr{M}}{EM} = \frac{42.0\ \text{g/mol}}{14.0\ \text{g/mol}} = 3$$

The molecular formula is thus derived from the empirical formula by multiplying each of its subscripts by three:

$$(\text{CH}_2)_3 = \text{C}_3\text{H}_6$$

Exercise 2.1.2.2

Acetylene, a fuel used welding torches, is composed of 92.3% C and 7.7% H by mass. Find the empirical formula. If 1.10 g of acetylene occupies of volume of 1.00 L at 1.15 atm and 59.5 °C, what is the molecular formula for acetylene?

Answer available at the end of this section.

Example 2.1.2.3: Determining the Molar Mass of a Volatile Liquid

The approximate molar mass of a volatile liquid can be determined by:

1. Heating a sample of the liquid in a flask with a tiny hole at the top, which converts the liquid into gas that may escape through the hole
2. Removing the flask from heat at the instant when the last bit of liquid becomes gas, at which time the flask will be filled with only gaseous sample at ambient pressure
3. Sealing the flask and permitting the gaseous sample to condense to liquid, and then weighing the flask to determine the sample's mass (see Figure 2.1.2.1)

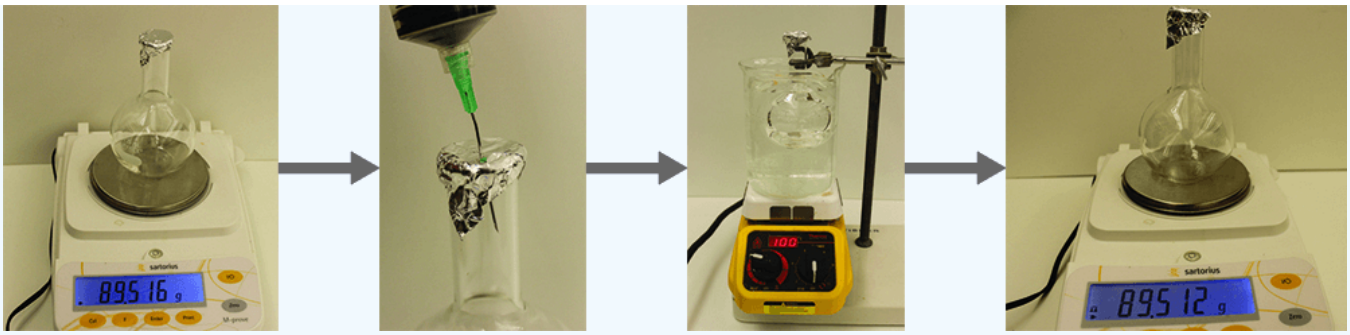


Figure 2.1.2.1: When the volatile liquid in the flask is heated past its boiling point, it becomes gas and drives air out of the flask. At the flask is filled with volatile liquid gas at the same pressure as the atmosphere. If the flask is then cooled to room temperature, the gas condenses and the mass of the gas that filled the flask, and is now liquid, can be measured. (credit: modification of work by Mark Ott)

Using this procedure, a sample of chloroform gas weighing 0.494 g is collected in a flask with a volume of 129 cm^3 at 99.6 °C when the atmospheric pressure is 742.1 mm Hg. What is the approximate molar mass of chloroform?

Solution

Since $\mathscr{M} = \frac{m}{n}$ and $n = \frac{PV}{RT}$ substituting and rearranging gives $\mathscr{M} = \frac{mRT}{PV}$,

then

$$\mathscr{M} = \frac{mRT}{PV} = \frac{(0.494\ \text{g}) \times 0.08206\ \text{L} \cdot \text{atm/mol K}\ \times 372.8\ \text{K}}{0.976\ \text{atm} \times 0.129\ \text{L}} = 120\ \text{g/mol}$$

Exercise 2.1.2.3

A sample of phosphorus that weighs 3.243×10^{-2} g exerts a pressure of 31.89 kPa in a 56.0-mL bulb at 550 °C. What are the molar mass and molecular formula of phosphorus vapor?

Answer available at the end of this section.

The Pressure of a Mixture of Gases: Dalton's Law

Unless they chemically react with each other, the individual gases in a mixture of gases do not affect each other's pressure. Each individual gas in a mixture exerts the same pressure that it would exert if it were present alone in the container (Figure 2.1.2.2). The pressure exerted by each individual gas in a mixture is called its **partial pressure**. This observation is summarized by **Dalton's law of partial pressures**: *The total pressure of a mixture of ideal gases is equal to the sum of the partial pressures of the component gases*:

$$P_{\text{Total}} = P_A + P_B + P_C + \ldots = \sum_i P_i$$

In the equation P_{Total} is the total pressure of a mixture of gases, P_A is the partial pressure of gas A; P_B is the partial pressure of gas B; P_C is the partial pressure of gas C; and so on.

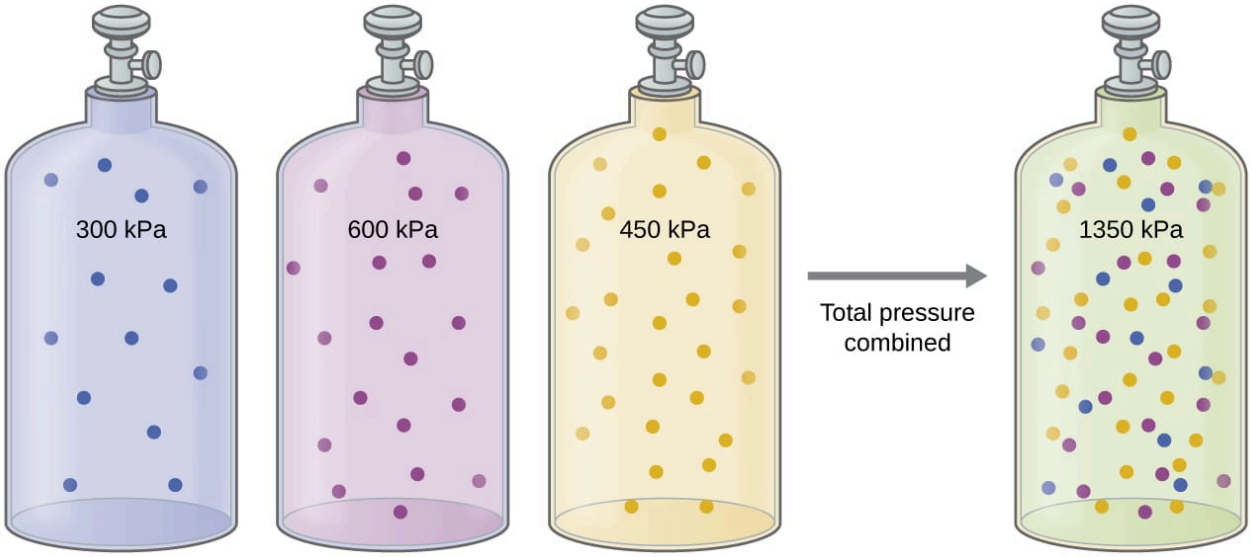


Figure 2.1.2.2: If equal-volume cylinders containing gasses at pressures of 300 kPa, 450 kPa, and 600 kPa are all combined in the same-size cylinder, the total pressure of the gas mixture is 1350 kPa.

The partial pressure of gas A is related to the total pressure of the gas mixture via its **mole fraction** (X), a unit of concentration defined as the number of moles of a component of a solution divided by the total number of moles of all components:

$$P_A = X_A \times P_{\text{Total}}$$

where

$$X_A = \frac{n_A}{n_{\text{Total}}}$$

where P_A, X_A, and n_A are the partial pressure, mole fraction, and number of moles of gas A, respectively, and n_{Total} is the number of moles of all components in the mixture.

Example 2.1.2.4: The Pressure of a Mixture of Gases

A 10.0-L vessel contains $2.50\ 10^{-3}$ mol of H_2, $1.00\ 10^{-3}$ mol of He, and $3.00\ 10^{-4}$ mol of Ne at 35 °C.

a. What are the partial pressures of each of the gases?
b. What is the total pressure in atmospheres?

Solution

The gases behave independently, so the partial pressure of each gas can be determined from the ideal gas equation, using $P = \frac{nRT}{V}$:

$$P_{H_2} = \frac{\left(2.50 \times 10^{-3}\ \cancel{\text{mol}}\right)\left(0.08206\ \cancel{\text{L}}\ \text{atm}\ \cancel{\text{mol}^{-1}\ \text{K}^{-1}}\right)(308\ \cancel{\text{K}})}{10.0\ \cancel{\text{L}}} = 6.32 \times 10^{-3}\ \text{atm}$$

$$= \frac{\left(1.00 \times 10^{-3}\ \cancel{\text{mol}}\right)\left(0.08206\ \cancel{\text{L}}\ \text{atm}\ \cancel{\text{mol}^{-1}\ \text{K}^{-1}}\right)(308\ \cancel{\text{K}})}{10.0\ \cancel{\text{L}}} = 2.53 \times 10^{-3}\ \text{atm}$$

$$= \frac{\left(3.00 \times 10^{-4}\ \cancel{\text{mol}}\right)\left(0.08206\ \cancel{\text{L}}\ \text{atm}\ \cancel{\text{mol}^{-1}\ \text{K}^{-1}}\right)(308\ \cancel{\text{K}})}{10.0\ \cancel{\text{L}}} = 7.58 \times 10^{-4}\ \text{atm}$$

The total pressure is given by the sum of the partial pressures:

$$P_T = P_{H_2} + P_{He} + P_{Ne} = (0.00632 + 0.00253 + 0.00076)\ \text{atm} = 9.61 \times 10^{-3}\ \text{atm}$$

Exercise 2.1.2.4

A 5.73-L flask at 25 °C contains 0.0388 mol of N_2, 0.147 mol of CO, and 0.0803 mol of H_2. What is the total pressure in the flask in atmospheres?

Answer available at the end of this section.

Here is another example of this concept, but dealing with mole fraction calculations.

Example 2.1.2.5: The Pressure of a Mixture of Gases

A gas mixture used for anesthesia contains 2.83 mol oxygen, O_2, and 8.41 mol nitrous oxide, N_2O. The total pressure of the mixture is 192 kPa.

a. What are the mole fractions of O_2 and N_2O?
b. What are the partial pressures of O_2 and N_2O?

Solution

The mole fraction is given by $X_A = \frac{n_A}{n_{\text{Total}}}$ and the partial pressure is $P_A = X_A P_{Total}$.

For O_2,

$$X_{O_2} = \frac{n_{O_2}}{n_{\text{Total}}} = \frac{2.83mol}{(2.83 + 8.41)mol} = 0.252$$

and $P_{O_2} = X_{O_2} \times P_{\text{Total}} = 0.252 \times 192kPa = 48.4kPa$

For N_2O,

$$X_{N_2O} = \frac{n_{N_2O}}{n_{\text{Total}}} = \frac{8.41mol}{(2.83 + 8.41)mol} = 0.748$$

and

$$P_{N_2O} = X_{N_2O} \times P_{\text{Total}} = 0.748 \times 192kPa = 144kPa$$

Exercise 2.1.2.5

What is the pressure of a mixture of 0.200 g of H_2, 1.00 g of N_2, and 0.820 g of Ar in a container with a volume of 2.00 L at 20 °C?

Answer available at the end of this section.

Collection of Gases over Water

A simple way to collect gases that do not react with water is to capture them in a bottle that has been filled with water and inverted into a dish filled with water. The pressure of the gas inside the bottle can be made equal to the air pressure outside by raising or lowering the bottle. When the water level is the same both inside and outside the bottle (Figure 2.1.2.3), the pressure of the gas is equal to the atmospheric pressure, which can be measured with a barometer.

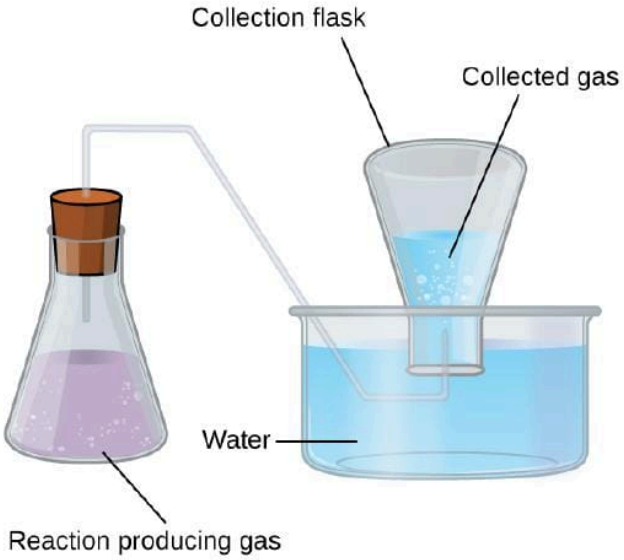


Figure 2.1.2.3: When a reaction produces a gas that is collected above water, the trapped gas is a mixture of the gas produced by the reaction and water vapor. If the collection flask is appropriately positioned to equalize the water levels both within and outside the flask, the pressure of the trapped gas mixture will equal the atmospheric pressure outside the flask (see the earlier discussion of manometers).

However, there is another factor we must consider when we measure the pressure of the gas by this method. Water evaporates and there is always gaseous water (water vapor) above a sample of liquid water. As a gas is collected over water, it becomes saturated with water vapor and the total pressure of the mixture equals the partial pressure of the gas plus the partial pressure of the water vapor. The pressure of the pure gas is therefore equal to the total pressure minus the pressure of the water vapor—this is referred to as the "dry" gas pressure, that is, the pressure of the gas only, without water vapor. The **vapor pressure of water**, which is the pressure exerted by water vapor in equilibrium with liquid water in a closed container, depends on the temperature (Figure 2.1.2.4); more detailed information on the temperature dependence of water vapor can be found in Table 2.1.2.1, and vapor pressure will be discussed in more detail in the next chapter on liquids.

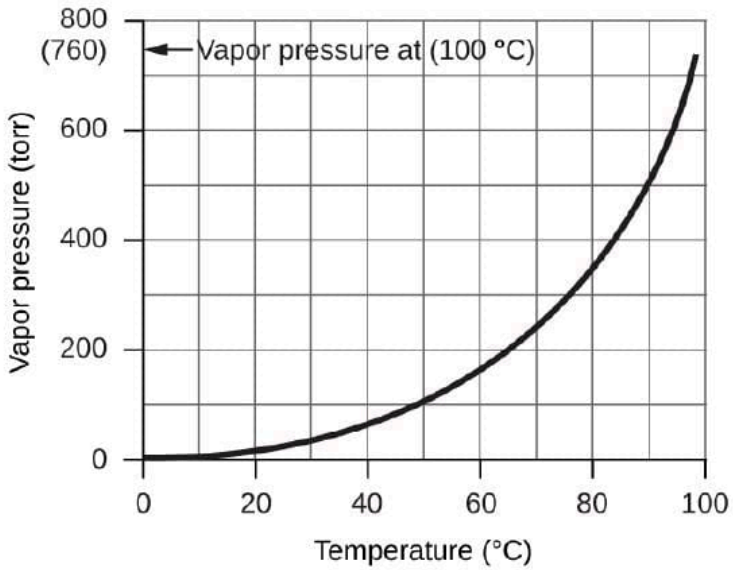


Figure 2.1.2.4: This graph shows the vapor pressure of water at sea level as a function of temperature.

Table 2.1.2.1: Vapor Pressure of Ice and Water in Various Temperatures at Sea Level

Temperature (°C)	Pressure (torr)	Temperature (°C)	Pressure (torr)	Temperature (°C)	Pressure (torr)
–10	1.95	18	15.5	30	31.8
–5	3.0	19	16.5	35	42.2

Temperature (°C)	Pressure (torr)	Temperature (°C)	Pressure (torr)	Temperature (°C)	Pressure (torr)
–2	3.9	20	17.5	40	55.3
0	4.6	21	18.7	50	92.5
2	5.3	22	19.8	60	149.4
4	6.1	23	21.1	70	233.7
6	7.0	24	22.4	80	355.1
8	8.0	25	23.8	90	525.8
10	9.2	26	25.2	95	633.9
12	10.5	27	26.7	99	733.2
14	12.0	28	28.3	100.0	760.0
16	13.6	29	30.0	101.0	787.6

Example 2.1.2.6: Pressure of a Gas Collected Over Water

If 0.200 L of argon is collected over water at a temperature of 26 °C and a pressure of 750 torr in a system like that shown in Figure 2.1.2.3, what is the partial pressure of argon?

Solution

According to Dalton's law, the total pressure in the bottle (750 torr) is the sum of the partial pressure of argon and the partial pressure of gaseous water:

$$P_T = P_{Ar} + P_{H_2O}$$

Rearranging this equation to solve for the pressure of argon gives:

$$P_{Ar} = P_T - P_{H_2O}$$

The pressure of water vapor above a sample of liquid water at 26 °C is 25.2 torr (Appendix E), so:

$$P_{Ar} = 750 \text{ torr } - 25.2 \text{ torr } = 725 \text{ torr}$$

Exercise 2.1.2.6

A sample of oxygen collected over water at a temperature of 29.0 °C and a pressure of 764 torr has a volume of 0.560 L. What volume would the dry oxygen from this sample have under the same conditions of temperature and pressure?

Answer available at the end of this section.

Chemical Stoichiometry and Gases

Chemical stoichiometry describes the quantitative relationships between reactants and products in chemical reactions.

We have previously measured quantities of reactants and products using masses for solids and volumes in conjunction with the molarity for solutions; now we can also use gas volumes to indicate quantities. If we know the volume, pressure, and temperature of a gas, we can use the ideal gas equation to calculate how many moles of the gas are present. If we know how many moles of a gas are involved, we can calculate the volume of a gas at any temperature and pressure.

Avogadro's Law Revisited

Sometimes we can take advantage of a simplifying feature of the stoichiometry of gases that solids and solutions do not exhibit: All gases that show ideal behavior contain the same number of molecules in the same volume (at the same temperature and pressure). Thus, the ratios of volumes of gases involved in a chemical reaction are given by the coefficients in the equation for the reaction, provided that the gas volumes are measured at the same temperature and pressure.

We can extend Avogadro's law (that the volume of a gas is directly proportional to the number of moles of the gas) to chemical reactions with gases: Gases combine, or react, in definite and simple proportions by volume, provided that all gas volumes are measured at the same temperature and pressure. For example, since nitrogen and hydrogen gases react to produce ammonia gas according to

$$\mathrm{N_2\,(g) + 3\,H_2\,(g) \longrightarrow 2\,NH_3\,(g)}$$

a given volume of nitrogen gas reacts with three times that volume of hydrogen gas to produce two times that volume of ammonia gas, if pressure and temperature remain constant.

The explanation for this is illustrated in Figure 2.1.2.5. According to Avogadro's law, equal volumes of gaseous N_2, H_2, and NH_3, at the same temperature and pressure, contain the same number of molecules. Because one molecule of N_2 reacts with three molecules of H_2 to produce two molecules of NH_3, the volume of H_2 required is three times the volume of N_2, and the volume of NH_3 produced is two times the volume of N_2.

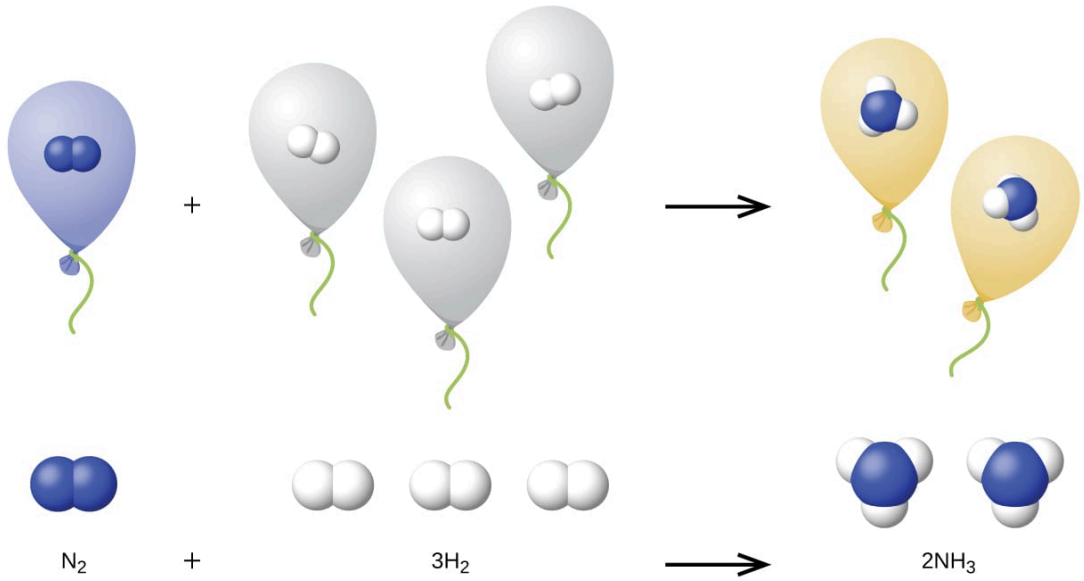


Figure 2.1.2.5: One volume of N_2 combines with three volumes of H_2 to form two volumes of NH_3.

Example 2.1.2.7: Reaction of Gases

Propane, $\mathrm{C_3H_8\,(g)}$, is used in gas grills to provide the heat for cooking. What volume of $O_2(g)$ measured at 25 °C and 760 torr is required to react with 2.7 L of propane measured under the same conditions of temperature and pressure? Assume that the propane undergoes complete combustion.

Solution

The ratio of the volumes of C_3H_8 and O_2 will be equal to the ratio of their coefficients in the balanced equation for the reaction:

$$\underset{\mathrm{C_3H_8\,(g)+5\,O_2\,(g)}}{1 \text{ volume} + 5 \text{ volumes}} \longrightarrow \underset{\mathrm{3\,CO_2\,(g)+4\,H_2O\,(l)}}{3 \text{ volumes} + 4 \text{ volumes}}$$

From the equation, we see that one volume of $\mathrm{C_3H_8}$ will react with five volumes of $\mathrm{O_2}$:

$$2.7 \cancel{\text{ L C}_3\text{H}_8} \times \frac{5 \text{ L O}_2}{1 \cancel{\text{ L C}_3\text{H}_8}} = 13.5 \text{ L O}_2$$

A volume of 13.5 L of O_2 will be required to react with 2.7 L of C_3H_8.

Exercise 2.1.2.7

An acetylene tank for an oxyacetylene welding torch provides 9340 L of acetylene gas, C_2H_2, at 0 °C and 1 atm. How many tanks of oxygen, each providing 7.00×10^3 L of O_2 at 0 °C and 1 atm, will be required to burn the acetylene?

$$2C_2H_2 + 5O_2 \longrightarrow 4CO_2 + 2H_2O$$

Answer available at the end of this section.

Example 2.1.2.8: Volumes of Reacting Gases

Ammonia is an important fertilizer and industrial chemical. Suppose that a volume of 683 billion cubic feet of gaseous ammonia, measured at 25 °C and 1 atm, was manufactured. What volume of $H_2(g)$, measured under the same conditions, was required to prepare this amount of ammonia by reaction with N_2?

$$\text{N}_2\,(\text{g}) + 3\,\text{H}_2\,(\text{g}) \longrightarrow 2\,\text{NH}_3\,(\text{g})$$

Solution

Because equal volumes of H_2 and NH_3 contain equal numbers of molecules and each three molecules of H_2 that react produce two molecules of NH_3, the ratio of the volumes of H_2 and NH_3 will be equal to 3:2. Two volumes of NH_3, in this case in units of billion ft^3, will be formed from three volumes of H_2:

$$683 \text{ billion } ft^3NH_3 \times \frac{3 \text{ billion } ft^3H_2}{2 \text{ billion } ft^3NH_3} = 1.02 \times 10^3 \text{ billion } ft^3H_2$$

The manufacture of 683 billion ft^3 of NH_3 required 1020 billion ft^3 of H_2. (At 25 °C and 1 atm, this is the volume of a cube with an edge length of approximately 1.9 miles.)

Exercise 2.1.2.8

What volume of $O_2(g)$ measured at 25 °C and 760 torr is required to react with 17.0 L of ethylene, $C_2H_4(g)$, measured under the same conditions of temperature and pressure? The products are CO_2 and water vapor.

Answer available at the end of this section.

Example 2.1.2.9: Volume of Gaseous Product

What volume of hydrogen at 27 °C and 723 torr may be prepared by the reaction of 8.88 g of gallium with an excess of hydrochloric acid?

$$2\,\text{Ga}\,(\text{s}) + 6\,\text{HCl}\,(\text{aq}) \longrightarrow 2\,\text{GaCl}_3\,(\text{aq}) + 3\,\text{H}_2\,(\text{g})$$

Solution

Convert the provided mass of the limiting reactant, Ga, to moles of hydrogen produced:

$$8.88 \ \cancel{\text{g Ga}} \times \frac{1 \ \cancel{\text{mol Ga}}}{69.723 \ \cancel{\text{g Ga}}} \times \frac{3 \text{ mol H}_2}{2 \ \cancel{\text{mol Ga}}} = 0.191 \text{ mol H}_2$$

Convert the provided temperature and pressure values to appropriate units (K and atm, respectively), and then use the molar amount of hydrogen gas and the ideal gas equation to calculate the volume of gas:

$$V = \left(\frac{nRT}{P}\right) = \frac{0.191 \ \cancel{\text{mol}} \times 0.08206 \text{ L } \cancel{\text{atm mol}}^{-1}\text{K}^{-1} \times 300 \text{ K}}{0.951 \ \cancel{\text{atm}}} = 4.94 \text{ L}$$

Exercise 2.1.2.9

Sulfur dioxide is an intermediate in the preparation of sulfuric acid. What volume of SO_2 at 343 °C and 1.21 atm is produced by burning 1.00 kg of sulfur in excess oxygen?

Answer available at the end of this section.

How Sciences Interconnect: Greenhouse Gases and Climate Change

The thin skin of our atmosphere keeps the earth from being an ice planet and makes it habitable. In fact, this is due to less than 0.5% of the air molecules. Of the energy from the sun that reaches the earth, almost $\frac{1}{3}$ is reflected back into space, with the rest absorbed by the atmosphere and the surface of the earth. Some of the energy that the earth absorbs is re-emitted as infrared (IR) radiation, a portion of which passes back out through the atmosphere into space. Most if this IR radiation, however, is absorbed by certain atmospheric gases, effectively trapping heat within the atmosphere in a phenomenon known as the greenhouse effect. This effect maintains global temperatures within the range needed to sustain life on earth. Without our atmosphere, the earth's average temperature would be lower by more than 30 °C (nearly 60 °F). The major greenhouse gases (GHGs) are water vapor, carbon dioxide, methane, and ozone. Since the Industrial Revolution, human activity has been increasing the concentrations of GHGs, which have changed the energy balance and are significantly altering the earth's climate (Figure 2.1.2.6).

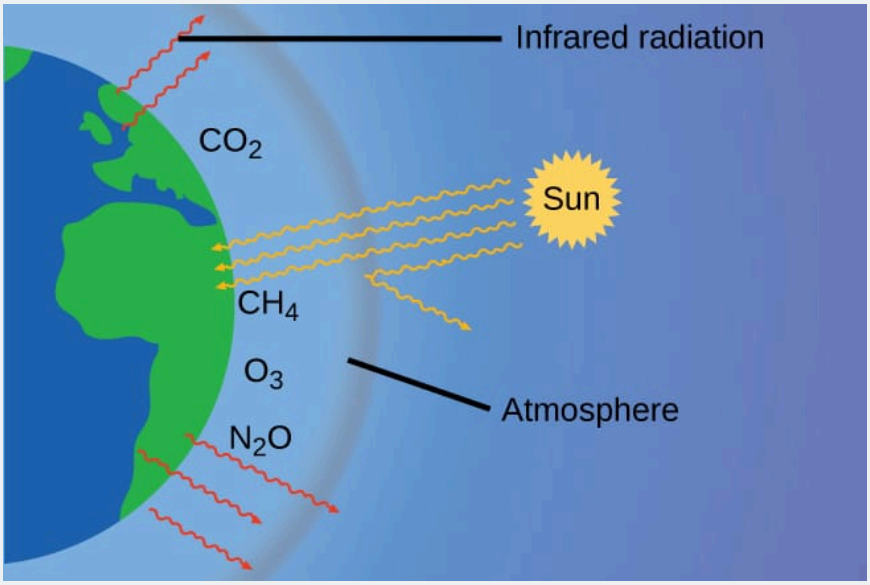


Figure 2.1.2.6: Greenhouse gases trap enough of the sun's energy to make the planet habitable—this is known as the greenhouse effect. Human activities are increasing greenhouse gas levels, warming the planet and causing more extreme weather events.

There is strong evidence from multiple sources that higher atmospheric levels of CO2 are caused by human activity, with fossil fuel burning accounting for about $\frac{3}{4}$ of the recent increase in CO2. Reliable data from ice cores reveals that CO2 concentration in the atmosphere is at the highest level in the past 800,000 years; other evidence indicates that it may be at its highest level in 20 million years. In recent years, the CO2 concentration has increased preindustrial levels of ~280 ppm to more than 400 ppm today (Figure 2.1.2.7).

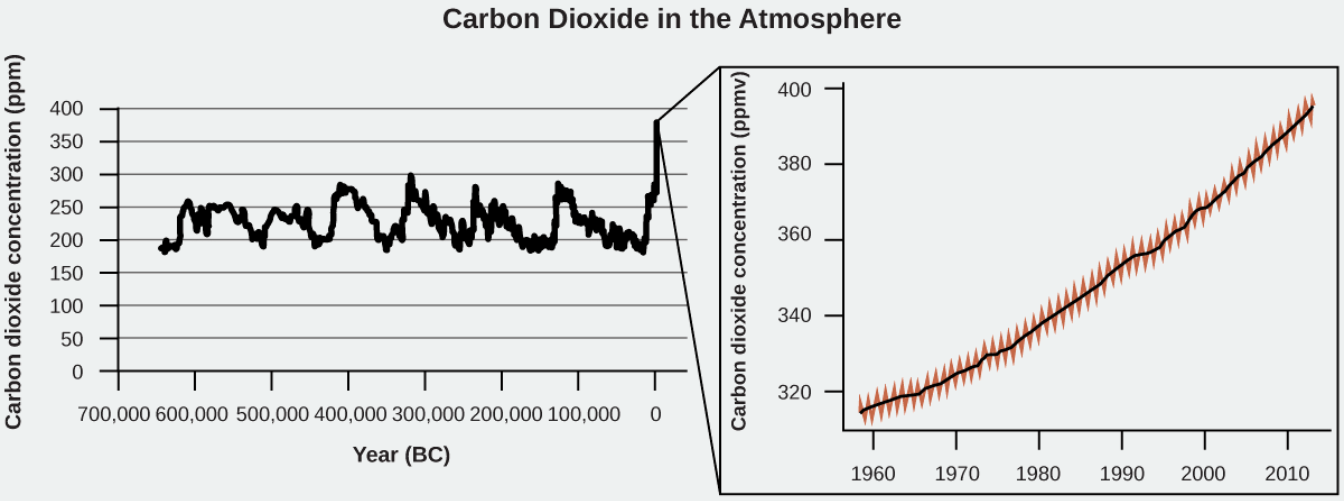


Figure 2.1.2.7: CO_2 levels over the past 700,000 years were typically from 200–300 ppm, with a steep, unprecedented increase over the past 50 years.

Link to Learning

Click here to see a 2-minute video explaining greenhouse gases and global warming.

Portrait of a Chemist: Susan Solomon

Atmospheric and climate scientist Susan Solomon (Figure 2.1.2.8) is the author of one of *The New York Times* books of the year (*The Coldest March*, 2001), one of Time magazine's 100 most influential people in the world (2008), and a working group leader of the Intergovernmental Panel on Climate Change (IPCC), which was the recipient of the 2007 Nobel Peace Prize. She helped determine and explain the cause of the formation of the ozone hole over Antarctica, and has authored many important papers on climate change. She has been awarded the top scientific honors in the US and France (the National Medal of Science and the Grande Medaille, respectively), and is a member of the National Academy of Sciences, the Royal Society, the French Academy of Sciences, and the European Academy of Sciences. Formerly a professor at the University of Colorado, she is now at MIT, and continues to work at NOAA.



Figure 2.1.2.8: Susan Solomon's research focuses on climate change and has been instrumental in determining the cause of the ozone hole over Antarctica. (credit: National Oceanic and Atmospheric Administration)

For more information, watch this video about Susan Solomon.

Footnotes

- 2"Quotations by Joseph-Louis Lagrange," last modified February 2006, accessed February 10, 2015, www-history.mcs.st-andrews.ac.../Lagrange.html

Answers

Exercise 2.1.2.1

d = 0.0847 g/L

Exercise 2.1.2.2

Empirical formula, CH; Molecular formula, C_2H_2

Exercise 2.1.2.3

124 g/mol P_4

Exercise 2.1.2.4

1.137 atm

Exercise 2.1.2.5

1.87 atm

Exercise 2.1.2.6

0.537 L

Exercise 2.1.2.7

3.34 tanks ($2.34\ 10^4$ L)

Exercise 2.1.2.8

51.0 L

Exercise 2.1.2.9

$1.30\ 10^3$ L

This page titled 2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 9.3: Stoichiometry of Gaseous Substances, Mixtures, and Reactions by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.1.3: The Kinetic-Molecular Theory

Learning Objectives

- State the postulates of the kinetic-molecular theory
- Use this theory's postulates to explain the gas laws

The gas laws that we have seen to this point, as well as the ideal gas equation, are empirical, that is, they have been derived from experimental observations. The mathematical forms of these laws closely describe the macroscopic behavior of most gases at pressures less than about 1 or 2 atm. Although the gas laws describe relationships that have been verified by many experiments, they do not tell us why gases follow these relationships.

The **kinetic molecular theory** (KMT) is a simple microscopic model that effectively explains the gas laws described in previous modules of this chapter. This theory is based on the following five postulates described here. (Note: The term "molecule" will be used to refer to the individual chemical species that compose the gas, although some gases are composed of atomic species, for example, the noble gases.)

1. Gases are composed of molecules that are in continuous motion, travelling in straight lines and changing direction only when they collide with other molecules or with the walls of a container.
2. The molecules composing the gas are negligibly small compared to the distances between them.
3. The pressure exerted by a gas in a container results from collisions between the gas molecules and the container walls.
4. Gas molecules exert no attractive or repulsive forces on each other or the container walls; therefore, their collisions are *elastic* (do not involve a loss of energy).
5. The average kinetic energy of the gas molecules is proportional to the kelvin temperature of the gas.

The test of the KMT and its postulates is its ability to explain and describe the behavior of a gas. The various gas laws can be derived from the assumptions of the KMT, which have led chemists to believe that the assumptions of the theory accurately represent the properties of gas molecules. We will first look at the individual gas laws (Boyle's, Charles's, Amontons's, Avogadro's, and Dalton's laws) conceptually to see how the KMT explains them. Then, we will more carefully consider the relationships between molecular masses, speeds, and kinetic energies with temperature, and explain Graham's law.

The Kinetic-Molecular Theory Explains the Behavior of Gases, Part I

Recalling that gas pressure is exerted by rapidly moving gas molecules and depends directly on the number of molecules hitting a unit area of the wall per unit of time, we see that the KMT conceptually explains the behavior of a gas as follows:

- *Amontons's law.* If the temperature is increased, the average speed and kinetic energy of the gas molecules increase. If the volume is held constant, the increased speed of the gas molecules results in more frequent and more forceful collisions with the walls of the container, therefore increasing the pressure (Figure 2.1.3.1).
- *Charles's law.* If the temperature of a gas is increased, a constant pressure may be maintained only if the volume occupied by the gas increases. This will result in greater average distances traveled by the molecules to reach the container walls, as well as increased wall surface area. These conditions will decrease the both the frequency of molecule-wall collisions and the number of collisions per unit area, the combined effects of which balance the effect of increased collision forces due to the greater kinetic energy at the higher temperature.
- *Boyle's law.* If the gas volume volume of a given amount of gas at a given temperature is decreased (that is, if the gas is *compressed*), the molecules will be exposed to a decreased container wall area. Collisions with the container wall will therefore occur more frequently and the pressure exerted by the gas will increase (Figure 2.1.3.1).
- *Avogadro's law.* At constant pressure and temperature, the frequency and force of molecule-wall collisions are constant. Under such conditions, increasing the number of gaseous molecules will require a proportional increase in the container volume in order to yield a decrease in the number of collisions per unit area to compensate for the increased frequency of collisions (Figure 2.1.3.1).
- *Dalton's Law.* Because of the large distances between them, the molecules of one gas in a mixture bombard the container walls with the same frequency whether other gases are present or not, and the total pressure of a gas mixture equals the sum of the (partial) pressures of the individual gases.

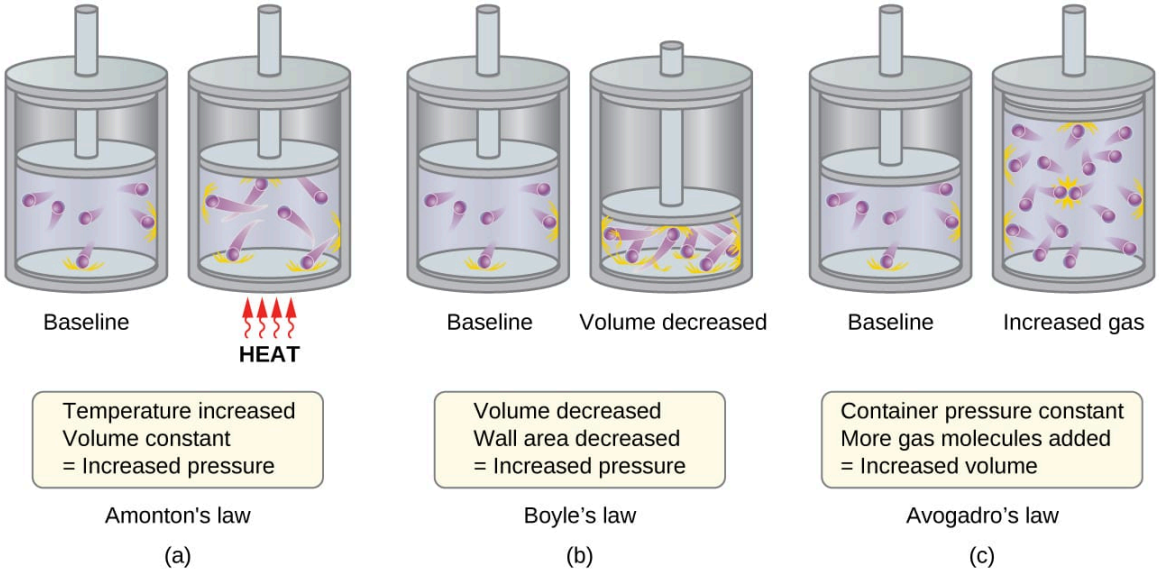


Figure 2.1.3.1: (a) When gas temperature increases, gas pressure increases due to increased force and frequency of molecular collisions. (b) When volume decreases, gas pressure increases due to increased frequency of molecular collisions. (c) When the amount of gas increases at a constant pressure, volume increases to yield a constant number of collisions per unit wall area per unit time.

molecular speeds and Kinetic Energy

The previous discussion showed that the KMT qualitatively explains the behaviors described by the various gas laws. The postulates of this theory may be applied in a more quantitative fashion to derive these individual laws. To do this, we must first look at speeds and kinetic energies of gas molecules, and the temperature of a gas sample.

In a gas sample, individual molecules have widely varying speeds; however, because of the *vast* number of molecules and collisions involved, the molecular speed distribution and average speed are constant. This molecular speed distribution is known as a Maxwell-Boltzmann distribution, and it depicts the relative numbers of molecules in a bulk sample of gas that possesses a given speed (Figure 2.1.3.2).

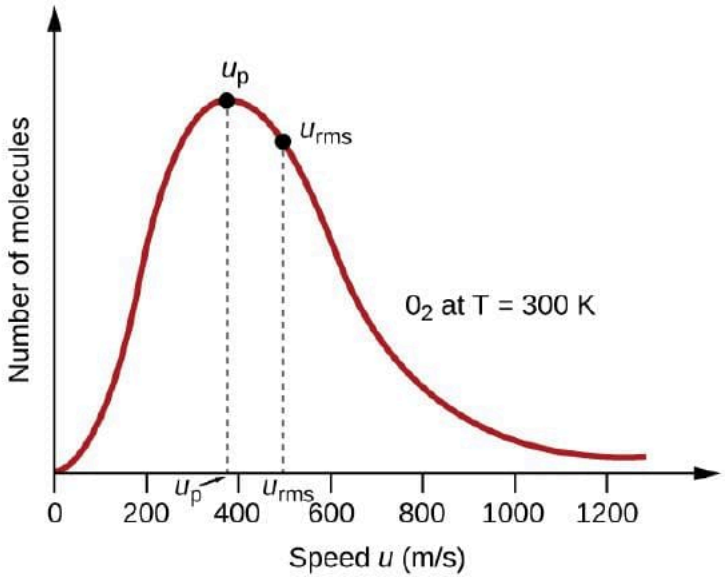


Figure 2.1.3.2: The molecular speed distribution for oxygen gas at 300 K is shown here. Very few molecules move at either very low or very high speeds. The number of molecules with intermediate speeds increases rapidly up to a maximum, which is the most probable speed, then drops off rapidly. Note that the most probable speed, v_p, is a little less than 400 m/s, while the root mean square speed, u_{rms}, is closer to 500 m/s.

The kinetic energy (KE) of a particle of mass (m) and speed (u) is given by:

$$KE = \frac{1}{2}mu^2$$

Expressing mass in kilograms and speed in meters per second will yield energy values in units of joules (J = kg m^2 s^{-2}). To deal with a large number of gas molecules, we use averages for both speed and kinetic energy. In the KMT, the root mean square speed of a particle, u_{rms}, is defined as the square root of the average of the squares of the speeds with n = the number of particles:

$$u_{rms} = \sqrt{\overline{u^2}} = \sqrt{\frac{u_1^2 + u_2^2 + u_3^2 + u_4^2 + ...}{n}}$$

The average kinetic energy for a mole of particles, KE_{avg}, is then equal to:

$$KE_{avg} = \frac{1}{2}Mu_{rms}^2$$

where M is the molar mass expressed in units of kg/mol. The KE_{avg} of a mole of gas molecules is also directly proportional to the temperature of the gas and may be described by the equation:

$$KE_{avg} = \frac{3}{2}RT$$

where R is the gas constant and T is the kelvin temperature. When used in this equation, the appropriate form of the gas constant is 8.314 J/mol · K (8.314 kg $m^2s^{-2}mol^{-1}K^{-1}$). These two separate equations for KE_{avg} may be combined and rearranged to yield a relation between molecular speed and temperature:

$$\frac{1}{2}Mu_{rms}^2 = \frac{3}{2}RT$$

with

$$u_{rms} = \sqrt{\frac{3RT}{M}}$$

Example 2.1.3.1: Calculation of u_{rms}

Calculate the root-mean-square speed for a nitrogen molecule at 30 °C.

Solution

Convert the temperature into Kelvin:

$$30^{\circ}C + 273 = 303K$$

Determine the molar mass of nitrogen in kilograms:

Extra close brace or missing open brace

Replace the variables and constants in the root-mean-square speed equation, replacing Joules with the equivalent kg m^2s^{-2}:

$$u_{rms} = \sqrt{\frac{3RT}{M}}$$

$$u_{rms} = \sqrt{\frac{3(8.314J/molK)(303K)}{(0.028kg/mol)}} = \sqrt{2.70 \times 10^5 m^2 s^{-2}} = 519m/s$$

Exercise 2.1.3.1

Calculate the root-mean-square speed for a mole of oxygen molecules at –23 °C.

Answer available at the end of this section.

If the temperature of a gas increases, its KE_{avg} increases, more molecules have higher speeds and fewer molecules have lower speeds, and the distribution shifts toward higher speeds overall, that is, to the right. If temperature decreases, KE_{avg} decreases, more molecules have lower speeds and fewer molecules have higher speeds, and the distribution shifts toward lower speeds overall, that is, to the left. This behavior is illustrated for nitrogen gas in Figure 2.1.3.3.

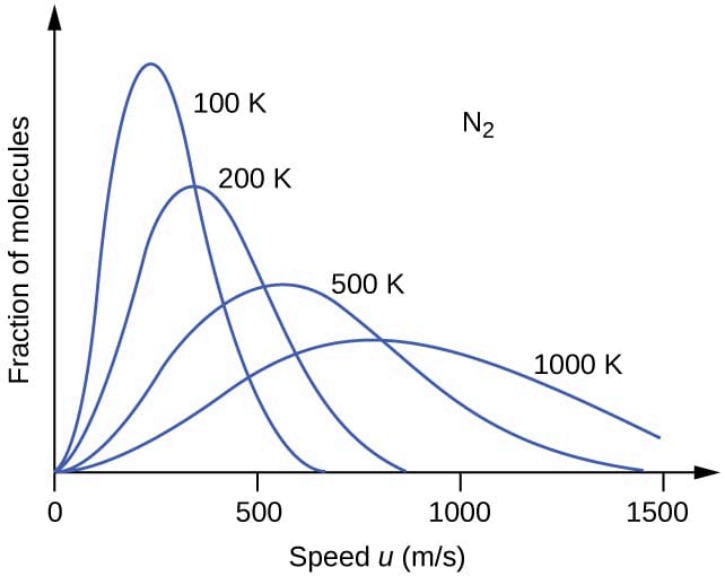


Figure 2.1.3.3: The molecular speed distribution for nitrogen gas (N_2) shifts to the right and flattens as the temperature increases; it shifts to the left and heightens as the temperature decreases.

At a given temperature, all gases have the same KE_{avg} for their molecules. Gases composed of lighter molecules have more high-speed particles and a higher u_{rms}, with a speed distribution that peaks at relatively higher speeds. Gases consisting of heavier molecules have more low-speed particles, a lower u_{rms}, and a speed distribution that peaks at relatively lower speeds. This trend is demonstrated by the data for a series of noble gases shown in Figure 2.1.3.4.

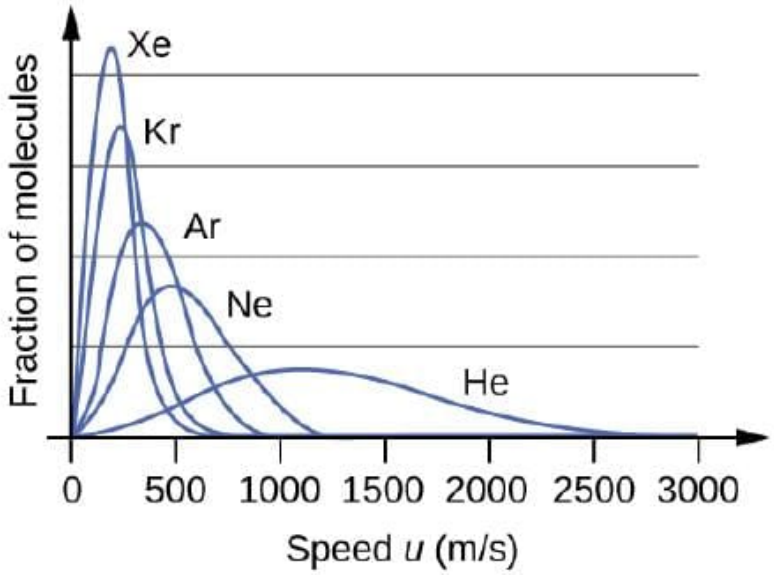


Figure 2.1.3.4: molecular speed is directly related to molecular mass. At a given temperature, lighter molecules move faster on average than heavier molecules.

Link to Learning

The gas simulator may be used to examine the effect of temperature on molecular speeds. Examine the simulator's "energy histograms" (molecular speed distributions) and "species information" (which gives average speed values) for molecules of different masses at various temperatures.

The Kinetic-Molecular Theory Explains the Behavior of Gases, Part II

According to Graham's law, the molecules of a gas are in rapid motion and the molecules themselves are small. The average distance between the molecules of a gas is large compared to the size of the molecules. As a consequence, gas molecules can move past each other easily and diffuse at relatively fast rates.

The rate of effusion of a gas depends directly on the (average) speed of its molecules:

$$\text{effusion rate } \propto u_{\text{rms}}$$

Using this relation, and the equation relating molecular speed to mass, Graham's law may be easily derived as shown here:

$$u_{rms} = \sqrt{\frac{3RT}{M}}$$

$$M = \frac{3RT}{u_{rms}^2} = \frac{3RT}{\bar{u}^2}$$

$$\frac{\text{effusion rate } A}{\text{effusion rate } B} = \frac{u_{rmsA}}{u_{rmsB}} = \frac{\sqrt{\frac{3RT}{M_A}}}{\sqrt{\frac{3RT}{M_B}}} = \sqrt{\frac{M_B}{M_A}}$$

The ratio of the rates of effusion is thus derived to be inversely proportional to the ratio of the square roots of their masses. This is the same relation observed experimentally and expressed as Graham's law.

Answers

Exercise 2.1.3.1

441 m/s

This page titled 2.1.3: The Kinetic-Molecular Theory is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 9.5: The Kinetic-Molecular Theory by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.1.4: Non-Ideal Gas Behavior

Learning Objectives

By the end of this section, you will be able to:

- Describe the physical factors that lead to deviations from ideal gas behavior
- Explain how these factors are represented in the van der Waals equation
- Define compressibility (Z) and describe how its variation with pressure reflects non-ideal behavior
- Quantify non-ideal behavior by comparing computations of gas properties using the ideal gas law and the van der Waals equation

Thus far, the ideal gas law, $PV = nRT$, has been applied to a variety of different types of problems, ranging from reaction stoichiometry and empirical and molecular formula problems to determining the density and molar mass of a gas. As mentioned in the previous modules of this chapter, however, the behavior of a gas is often non-ideal, meaning that the observed relationships between its pressure, volume, and temperature are not accurately described by the gas laws. In this section, the reasons for these deviations from ideal gas behavior are considered.

One way in which the accuracy of *PV* = *nRT* can be judged is by comparing the actual volume of 1 mole of gas (its molar volume, V_m) to the molar volume of an ideal gas at the same temperature and pressure. This ratio is called the **compressibility factor** (Z) with:

$$Z = \frac{\text{molar volume of gas at same } T \text{ and } P}{\text{molar volume of ideal gas at same } T \text{ and } P} = \left(\frac{PV_m}{RT}\right)_{\text{measured}}$$

Ideal gas behavior is therefore indicated when this ratio is equal to 1, and any deviation from 1 is an indication of non-ideal behavior. Figure 2.1.4.1 shows plots of Z over a large pressure range for several common gases.

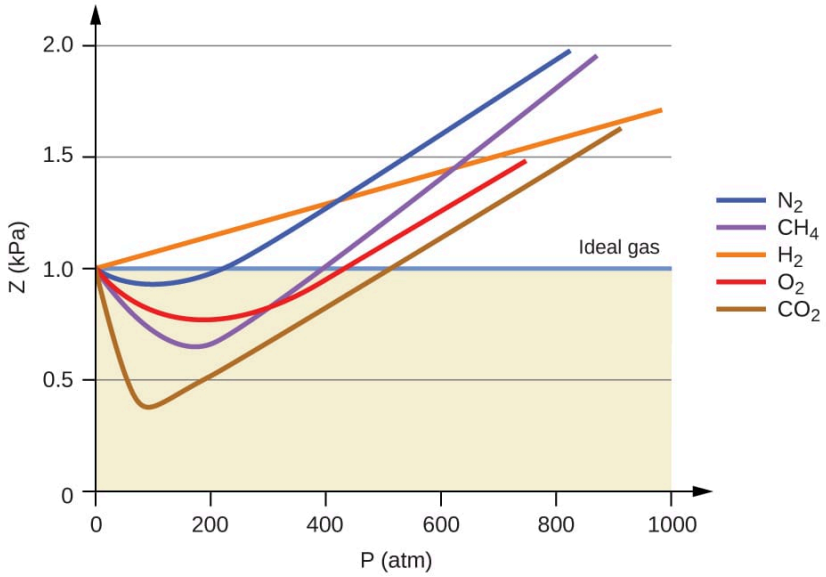


Figure 2.1.4.1: A graph of the compressibility factor (Z) vs. pressure shows that gases can exhibit significant deviations from the behavior predicted by the ideal gas law.

As is apparent from Figure 2.1.4.1, the ideal gas law does not describe gas behavior well at relatively high pressures. To determine why this is, consider the differences between real gas properties and what is expected of a hypothetical ideal gas.

Particles of a hypothetical ideal gas have no significant volume and do not attract or repel each other. In general, real gases approximate this behavior at relatively low pressures and high temperatures. However, at high pressures, the molecules of a gas

are crowded closer together, and the amount of empty space between the molecules is reduced. At these higher pressures, the volume of the gas molecules themselves becomes appreciable relative to the total volume occupied by the gas. The gas therefore becomes less compressible at these high pressures, and although its volume continues to decrease with increasing pressure, this decrease is not *proportional* as predicted by Boyle's law.

At relatively low pressures, gas molecules have practically no attraction for one another because they are (on average) so far apart, and they behave almost like particles of an ideal gas. At higher pressures, however, the force of attraction is also no longer insignificant. This force pulls the molecules a little closer together, slightly decreasing the pressure (if the volume is constant) or decreasing the volume (at constant pressure) (Figure 2.1.4.2). This change is more pronounced at low temperatures because the molecules have lower KE relative to the attractive forces, and so they are less effective in overcoming these attractions after colliding with one another.

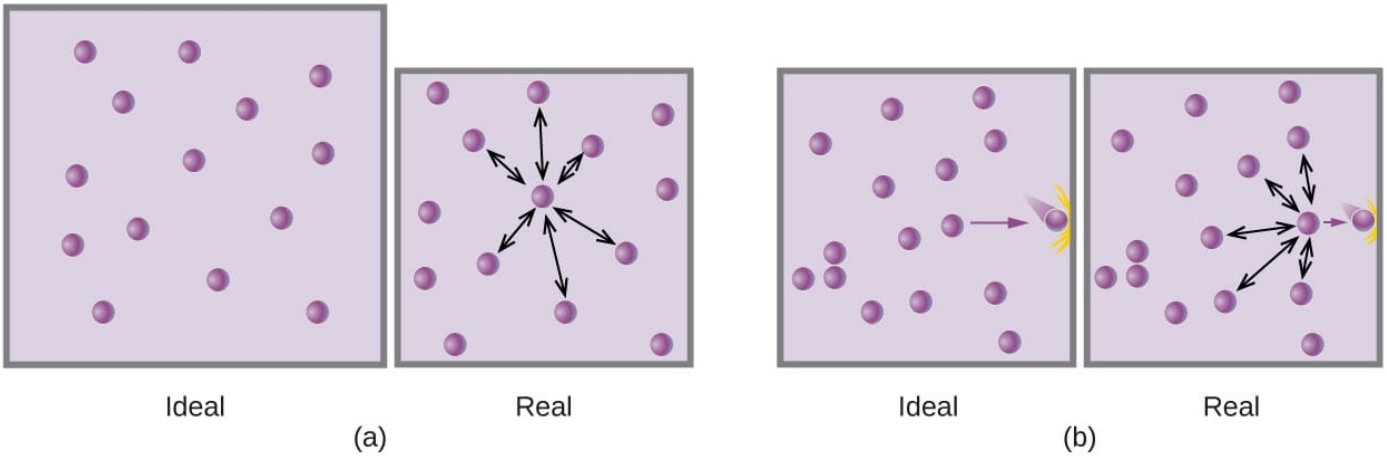


Figure 2.1.4.2: (a) Attractions between gas molecules serve to decrease the gas volume at constant pressure compared to an ideal gas whose molecules experience no attractive forces. (b) These attractive forces will decrease the force of collisions between the molecules and container walls, therefore reducing the pressure exerted at constant volume compared to an ideal gas.

There are several different equations that better approximate gas behavior than does the ideal gas law. The first, and simplest, of these was developed by the Dutch scientist Johannes van der Waals in 1879. The **van der Waals equation** improves upon the ideal gas law by adding two terms: one to account for the volume of the gas molecules and another for the attractive forces between them.

$$PV = nRT \longrightarrow \left(P + \frac{an^2}{V^2}\right)(V - nb) = nRT$$

Correction for molecular attraction

Correction for volume of molecules

The constant *a* corresponds to the strength of the attraction between molecules of a particular gas, and the constant *b* corresponds to the size of the molecules of a particular gas. The "correction" to the pressure term in the ideal gas law is

Table 2.1.4.1: Values of van der Waals Constants for Some Common Gases

Gas	*a* (L^2 atm/mol^2)	*b* (L/mol)
N_2	1.39	0.0391
O_2	1.36	0.0318
CO_2	3.59	0.0427
H_2O	5.46	0.0305

Gas	*a* (L^2 atm/mol^2)	*b* (L/mol)
He	0.0342	0.0237
CCl_4	20.4	0.1383

At low pressures, the correction for intermolecular attraction, *a*, is more important than the one for molecular volume, *b*. At high pressures and small volumes, the correction for the volume of the molecules becomes important because the molecules themselves are incompressible and constitute an appreciable fraction of the total volume. At some intermediate pressure, the two corrections have opposing influences and the gas appears to follow the relationship given by *PV* = *nRT* over a small range of pressures. This behavior is reflected by the "dips" in several of the compressibility curves shown in Figure 2.1.4.1. The attractive force between molecules initially makes the gas more compressible than an ideal gas, as pressure is raised (Z decreases with increasing *P*). At very high pressures, the gas becomes less compressible (Z increases with *P*), as the gas molecules begin to occupy an increasingly significant fraction of the total gas volume.

Strictly speaking, the ideal gas equation functions well when intermolecular attractions between gas molecules are negligible and the gas molecules themselves do not occupy an appreciable part of the whole volume. These criteria are satisfied under conditions of *low pressure and high temperature*. Under such conditions, the gas is said to behave ideally, and deviations from the gas laws are small enough that they may be disregarded—this is, however, very often not the case.

Example 2.1.4.1: Comparison of Ideal Gas Law and van der Waals Equation

A 4.25-L flask contains 3.46 mol CO_2 at 229 °C. Calculate the pressure of this sample of CO_2:

a. from the ideal gas law
b. from the van der Waals equation
c. Explain the reason(s) for the difference.

Solution

(a) From the ideal gas law:

$$P = \frac{nRT}{V} = \frac{3.46 \text{ mol} \times 0.08206 Latm mol^{-1} K^{-1} \times 502 K}{4.25 I} = 33.5 atm$$

(b) From the van der Waals equation:

$$\left(P + \frac{n^2 a}{V^2}\right) \times (V - nb) = nRT \longrightarrow P = \frac{nRT}{(V - nb)} - \frac{n^2 a}{V^2}$$

$$P = \frac{3.46 \text{ mol} \times 0.08206 \text{ L atm mol}^{-1} \text{ K}^{-1} \times 502 \text{ K}}{\left(4.25 \text{ L} - 3.46 \text{ mol} \times 0.0427 \text{ L mol}^{-1}\right)} - \frac{(3.46 \text{ mol})^2 \times 3.59 \text{ L}^2 \text{ atm mol}^2}{(4.24 \text{ L})^2}$$

This finally yields *P* = 32.4 atm.

(c) This is not very different from the value from the ideal gas law because the pressure is not very high and the temperature is not very low. The value is somewhat different because CO_2 molecules do have some volume and attractions between molecules, and the ideal gas law assumes they do not have volume or attractions.

Exercise 2.1.4.1

A 560-mL flask contains 21.3 g N_2 at 145 °C. Calculate the pressure of N_2:

a. from the ideal gas law
b. from the van der Waals equation
c. Explain the reason(s) for the difference.

Answer available at the end of this section.

Answers

Exercise 2.1.4.1

(a) 46.562 atm; (b) 46.594 atm; (c) The van der Waals equation takes into account the volume of the gas molecules themselves as well as intermolecular attractions.

This page titled 2.1.4: Non-Ideal Gas Behavior is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 9.6: Non-Ideal Gas Behavior by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.2: THERMODYNAMICS

2.2.1: Energy Basics
2.2.2: Enthalpy
2.2.3: Spontaneity
2.2.4: Entropy
2.2.5: The Second and Third Laws of Thermodynamics
2.2.6: Free Energy

This page titled 2.2: Thermodynamics is shared under a license and was authored, remixed, and/or curated by LibreTexts.

2.2.1: Energy Basics

Learning Objectives

By the end of this section, you will be able to:

- Define energy, distinguish types of energy, and describe the nature of energy changes that accompany chemical and physical changes
- Distinguish the related properties of heat, thermal energy, and temperature
- Define and distinguish specific heat and heat capacity, and describe the physical implications of both
- Perform calculations involving heat, specific heat, and temperature change

Chemical changes and their accompanying changes in energy are important parts of our everyday world (Figure 2.2.1.1). The macronutrients in food (proteins, fats, and carbohydrates) undergo metabolic reactions that provide the energy to keep our bodies functioning. We burn a variety of fuels (gasoline, natural gas, coal) to produce energy for transportation, heating, and the generation of electricity. Industrial chemical reactions use enormous amounts of energy to produce raw materials (such as iron and aluminum). Energy is then used to manufacture those raw materials into useful products, such as cars, skyscrapers, and bridges.



Figure 2.2.1.1: The energy involved in chemical changes is important to our daily lives: (a) A cheeseburger for lunch provides the energy you need to get through the rest of the day; (b) the combustion of gasoline provides the energy that moves your car (and you) between home, work, and school; and (c) coke, a processed form of coal, provides the energy needed to convert iron ore into iron, which is essential for making many of the products we use daily. (credit a: modification of work by "Pink Sherbet Photography"/Flickr; credit b: modification of work by Jeffery Turner)

Over 90% of the energy we use comes originally from the sun. Every day, the sun provides the earth with almost 10,000 times the amount of energy necessary to meet all of the world's energy needs for that day. Our challenge is to find ways to convert and store incoming solar energy so that it can be used in reactions or chemical processes that are both convenient and nonpolluting. Plants and many bacteria capture solar energy through photosynthesis. We release the energy stored in plants when we burn wood or plant products such as ethanol. We also use this energy to fuel our bodies by eating food that comes directly from plants or from animals that got their energy by eating plants. Burning coal and petroleum also releases stored solar energy: These fuels are fossilized plant and animal matter.

This chapter will introduce the basic ideas of an important area of science concerned with the amount of heat absorbed or released during chemical and physical changes—an area called thermochemistry. The concepts introduced in this chapter are widely used in almost all scientific and technical fields. Food scientists use them to determine the energy content of foods. Biologists study the energetics of living organisms, such as the metabolic combustion of sugar into carbon dioxide and water. The oil, gas, and transportation industries, renewable energy providers, and many others endeavor to find better methods to produce energy for our commercial and personal needs. Engineers strive to improve energy efficiency, find better ways to heat and cool our homes, refrigerate our food and drinks, and meet the energy and cooling needs of computers and electronics, among other applications. Understanding thermochemical principles is essential for chemists, physicists, biologists, geologists, every type of engineer, and just about anyone who studies or does any kind of science.

Energy

Energy can be defined as the capacity to supply heat or do work. One type of work (*w*) is the process of causing matter to move against an opposing force. For example, we do work when we inflate a bicycle tire—we move matter (the air in the pump) against the opposing force of the air already in the tire.

Like matter, energy comes in different types. One scheme classifies energy into two types: potential energy, the energy an object has because of its relative position, composition, or condition, and kinetic energy, the energy that an object possesses because of its motion. Water at the top of a waterfall or dam has potential energy because of its position; when it flows downward through generators, it has kinetic energy that can be used to do work and produce electricity in a hydroelectric plant (Figure 2.2.1.2). A battery has potential energy because the chemicals within it can produce electricity that can do work.



Figure 2.2.1.2: (a) Water at a higher elevation, for example, at the top of Victoria Falls, has a higher potential energy than water at a lower elevation. As the water falls, some of its potential energy is converted into kinetic energy. (b) If the water flows through generators at the bottom of a dam, such as the Hoover Dam shown here, its kinetic energy is converted into electrical energy. (credit a: modification of work by Steve Jurvetson; credit b: modification of work by "curimedia"/Wikimedia commons)

Energy can be converted from one form into another, but all of the energy present before a change occurs always exists in some form after the change is completed. This observation is expressed in the law of conservation of energy: during a chemical or physical change, energy can be neither created nor destroyed, although it can be changed in form. (This is also one version of the first law of thermodynamics, as you will learn later.)

When one substance is converted into another, there is always an associated conversion of one form of energy into another. Heat is usually released or absorbed, but sometimes the conversion involves light, electrical energy, or some other form of energy. For example, chemical energy (a type of potential energy) is stored in the molecules that compose gasoline. When gasoline is combusted within the cylinders of a car's engine, the rapidly expanding gaseous products of this chemical reaction generate mechanical energy (a type of kinetic energy) when they move the cylinders' pistons.

According to the law of conservation of matter (seen in an earlier chapter), there is no detectable change in the total amount of matter during a chemical change. When chemical reactions occur, the energy changes are relatively modest and the mass changes are too small to measure, so the laws of conservation of matter and energy hold well. However, in nuclear reactions, the energy changes are much larger (by factors of a million or so), the mass changes are measurable, and matter-energy conversions are significant. This will be examined in more detail in a later chapter on nuclear chemistry.

Thermal Energy, Temperature, and Heat

Thermal energy is kinetic energy associated with the random motion of atoms and molecules. Temperature is a quantitative measure of "hot" or "cold." When the atoms and molecules in an object are moving or vibrating quickly, they have a higher average kinetic energy (KE), and we say that the object is "hot." When the atoms and molecules are moving slowly, they have lower average

KE, and we say that the object is “cold” (Figure 2.2.1.3). Assuming that no chemical reaction or phase change (such as melting or vaporizing) occurs, increasing the amount of thermal energy in a sample of matter will cause its temperature to increase. And, assuming that no chemical reaction or phase change (such as condensation or freezing) occurs, decreasing the amount of thermal energy in a sample of matter will cause its temperature to decrease.

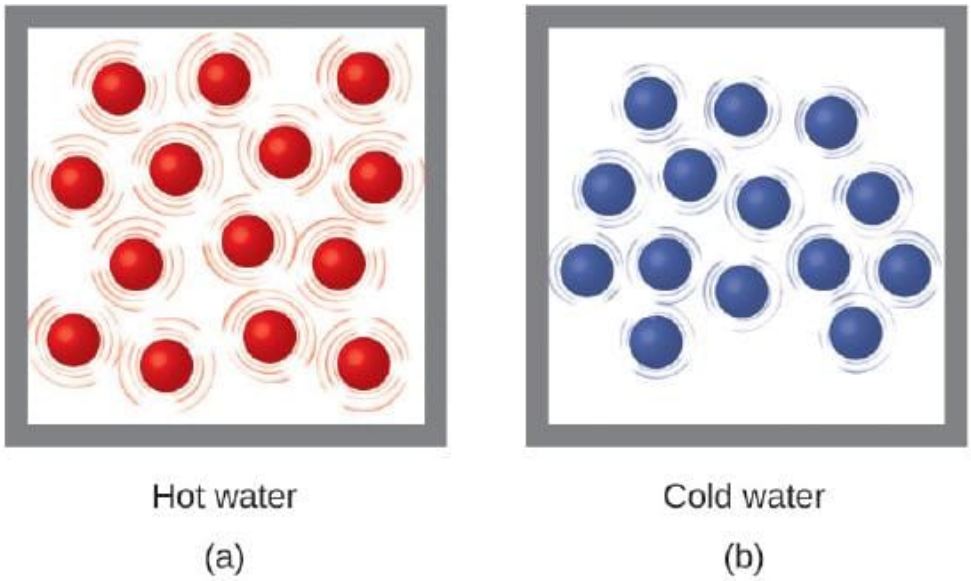


Figure 2.2.1.3: (a) The molecules in a sample of hot water move more rapidly than (b) those in a sample of cold water.

Link to Learning

Click on this interactive simulation to view the effects of temperature on molecular motion.

Most substances expand as their temperature increases and contract as their temperature decreases. This property can be used to measure temperature changes, as shown in Figure 2.2.1.4. The operation of many thermometers depends on the expansion and contraction of substances in response to temperature changes.

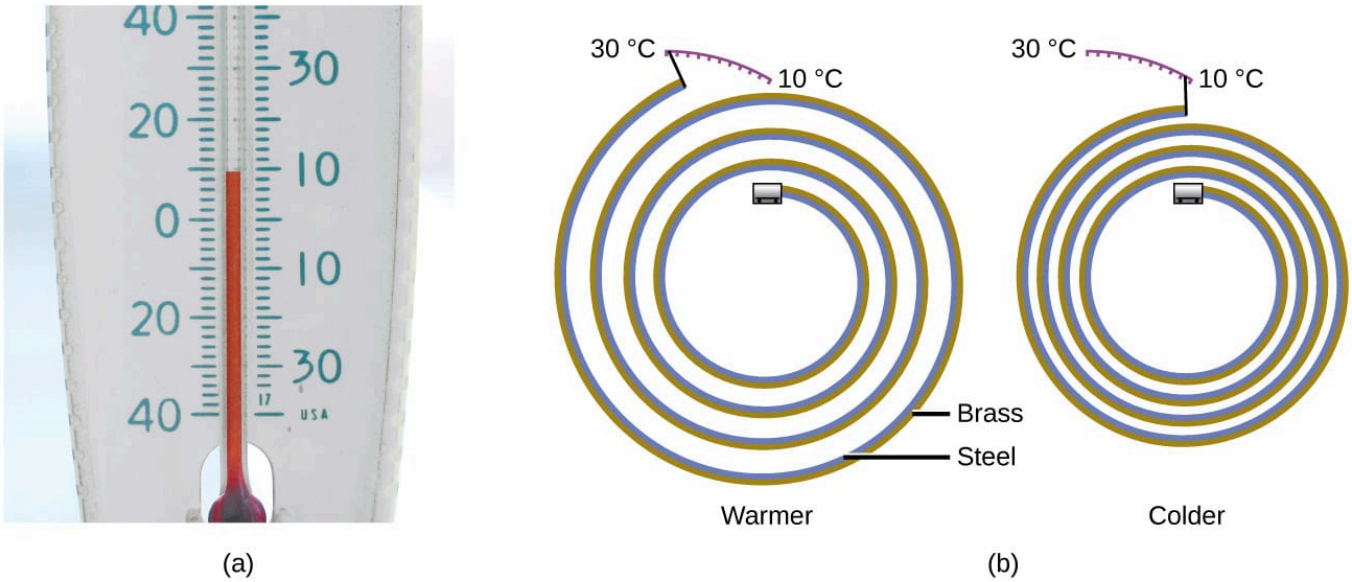


Figure 2.2.1.4: (a) In an alcohol or mercury thermometer, the liquid (dyed red for visibility) expands when heated and contracts when cooled, much more so than the glass tube that contains the liquid. (b) In a bimetallic thermometer, two different metals (such as brass and steel) form a two-layered strip. When heated or cooled, one of the metals (brass) expands or contracts more than the other metal (steel), causing the strip to coil or uncoil. Both types of thermometers have a calibrated scale that indicates the temperature. (credit a: modification of work by “dwstucke”/Flickr)

Link to Learning

The following demonstration allows one to view the effects of heating and cooling a coiled bimetallic strip.

Heat (q) is the transfer of thermal energy between two bodies at different temperatures. Heat flow (a redundant term, but one commonly used) increases the thermal energy of one body and decreases the thermal energy of the other. Suppose we initially have a high temperature (and high thermal energy) substance (H) and a low temperature (and low thermal energy) substance (L). The atoms and molecules in H have a higher average KE than those in L. If we place substance H in contact with substance L, the thermal energy will flow spontaneously from substance H to substance L. The temperature of substance H will decrease, as will the average KE of its molecules; the temperature of substance L will increase, along with the average KE of its molecules. Heat flow will continue until the two substances are at the same temperature (Figure 2.2.1.5).

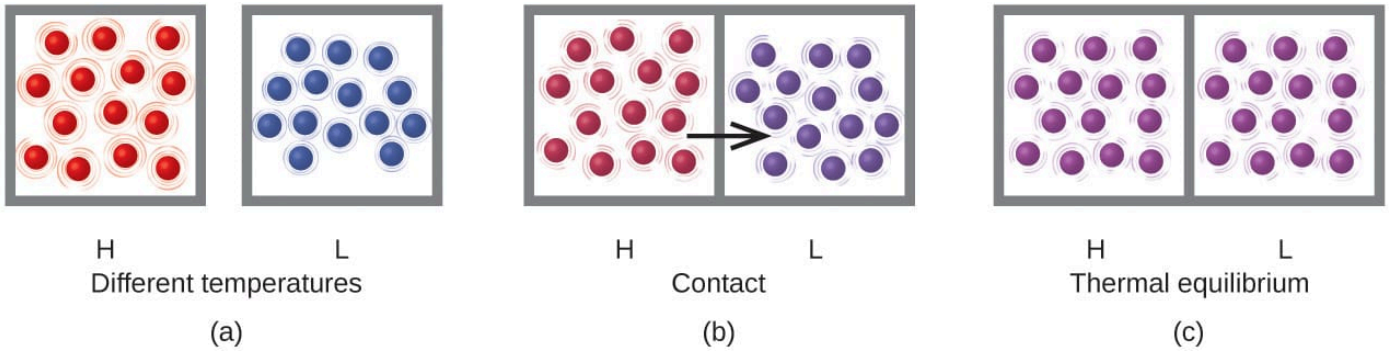


Figure 2.2.1.5: (a) Substances H and L are initially at different temperatures, and their atoms have different average kinetic energies. (b) When they contact each other, collisions between the molecules result in the transfer of kinetic (thermal) energy from the hotter to the cooler matter. (c) The two objects reach "thermal equilibrium" when both substances are at the same temperature and their molecules have the same average kinetic energy.

Link to Learning

Click on the PhET simulation to explore energy forms and changes. Visit the Energy Systems tab to create combinations of energy sources, transformation methods, and outputs. Click on Energy Symbols to visualize the transfer of energy.

Matter undergoing chemical reactions and physical changes can release or absorb heat. A change that releases heat is called an exothermic process. For example, the combustion reaction that occurs when using an oxyacetylene torch is an exothermic process—this process also releases energy in the form of light as evidenced by the torch's flame (Figure 2.2.1.6). A reaction or change that absorbs heat is an endothermic process. A cold pack used to treat muscle strains provides an example of an endothermic process. When the substances in the cold pack (water and a salt like ammonium nitrate) are brought together, the resulting process absorbs heat, leading to the sensation of cold.



Figure 2.2.1.6: (a) An oxyacetylene torch produces heat by the combustion of acetylene in oxygen. The energy released by this exothermic reaction heats and then melts the metal being cut. The sparks are tiny bits of the molten metal flying away. (b) A cold pack uses an endothermic process to create the sensation of cold. (credit a: modification of work by "Skatebiker"/Wikimedia commons)

Historically, energy was measured in units of calories (cal). A calorie is the amount of energy required to raise one gram of water by 1 degree C (1 kelvin). However, this quantity depends on the atmospheric pressure and the starting temperature of the water. The ease of measurement of energy changes in calories has meant that the calorie is still frequently used. The Calorie (with a capital C), or large calorie, commonly used in quantifying food energy content, is a kilocalorie. The SI unit of heat, work, and energy is the joule. A joule (J) is defined as the amount of energy used when a force of 1 newton moves an object 1 meter. It is named in honor of the English physicist James Prescott Joule. One joule is equivalent to 1 kg m^2/s^2, which is also called 1 newton–meter. A kilojoule (kJ) is 1000 joules. To standardize its definition, 1 calorie has been set to equal 4.184 joules.

We now introduce two concepts useful in describing heat flow and temperature change. The heat capacity (C) of a body of matter is the quantity of heat (q) it absorbs or releases when it experiences a temperature change (ΔT) of 1 degree Celsius (or equivalently, 1 kelvin):

$$C = \frac{q}{\Delta T}$$

Heat capacity is determined by both the type and amount of substance that absorbs or releases heat. It is therefore an extensive property—its value is proportional to the amount of the substance.

For example, consider the heat capacities of two cast iron frying pans. The heat capacity of the large pan is five times greater than that of the small pan because, although both are made of the same material, the mass of the large pan is five times greater than the mass of the small pan. More mass means more atoms are present in the larger pan, so it takes more energy to make all of those atoms vibrate faster. The heat capacity of the small cast iron frying pan is found by observing that it takes 18,150 J of energy to raise the temperature of the pan by 50.0 °C:

$$C_{\text{small pan}} = \frac{18,140J}{50.0°C} = 363J/°C$$

The larger cast iron frying pan, while made of the same substance, requires 90,700 J of energy to raise its temperature by 50.0 °C. The larger pan has a (proportionally) larger heat capacity because the larger amount of material requires a (proportionally) larger amount of energy to yield the same temperature change:

$$C_{\text{large pan}} = \frac{90,700J}{50.0°C} = 1814J/°C$$

The specific heat capacity (c) of a substance, commonly called its "specific heat," is the quantity of heat required to raise the temperature of 1 gram of a substance by 1 degree Celsius (or 1 kelvin):

$$c = \frac{q}{m\Delta T}$$

Specific heat capacity depends only on the kind of substance absorbing or releasing heat. It is an intensive property—the type, but not the amount, of the substance is all that matters. For example, the small cast iron frying pan has a mass of 808 g. The specific heat of iron (the material used to make the pan) is therefore:

$$c_{\text{iron}} = \frac{18,140J}{(808g)\,(50.0^{\circ}C)} = 0.449J/g^{\circ}C$$

The large frying pan has a mass of 4040 g. Using the data for this pan, we can also calculate the specific heat of iron:

$$c_{\text{iron}} = \frac{90,700J}{(4040g)\,(50.0^{\circ}C)} = 0.449J/g^{\circ}C$$

Although the large pan is more massive than the small pan, since both are made of the same material, they both yield the same value for specific heat (for the material of construction, iron). Note that specific heat is measured in units of energy per temperature per mass and is an intensive property, being derived from a ratio of two extensive properties (heat and mass). The molar heat capacity, also an intensive property, is the heat capacity per mole of a particular substance and has units of J/mol °C (Figure 2.2.1.7).



Figure 2.2.1.7: Because of its larger mass, a large frying pan has a larger heat capacity than a small frying pan. Because they are made of the same material, both frying pans have the same specific heat. (credit: Mark Blaser)

Water has a relatively high specific heat (about 4.2 J/g °C for the liquid and 2.09 J/g °C for the solid); most metals have much lower specific heats (usually less than 1 J/g °C). The specific heat of a substance varies somewhat with temperature. However, this variation is usually small enough that we will treat specific heat as constant over the range of temperatures that will be considered in this chapter. Specific heats of some common substances are listed in Table 2.2.1.1.

Table 2.2.1.1: Specific Heats of Common Substances at 25 °C and 1 bar

Substance	Symbol (*state*)	Specific Heat (J/g °C)
helium	He(*g*)	5.193
water	$H_2O(l)$	4.184
ethanol	$C_2H_6O(l)$	2.376
ice	$H_2O(s)$	2.093 (at –10 °C)
water vapor	$H_2O(g)$	1.864
nitrogen	$N_2(g)$	1.040
air		1.007

Substance	Symbol (*state*)	Specific Heat (J/g °C)
oxygen	$O_2(g)$	0.918
aluminum	Al(s)	0.897
carbon dioxide	$CO_2(g)$	0.853
argon	Ar(*g*)	0.522
iron	Fe(s)	0.449
copper	Cu(s)	0.385
lead	Pb(s)	0.130
gold	Au(s)	0.129
silicon	Si(s)	0.712

If we know the mass of a substance and its specific heat, we can determine the amount of heat, *q*, entering or leaving the substance by measuring the temperature change before and after the heat is gained or lost:

$$\begin{aligned} q &= (\text{ specific heat }) \times (\text{ mass of substance }) \times (\text{ temperature change }) \\ &= c \times m \times \Delta T \\ &= c \times m \times (T_{\text{final}} - T_{\text{initial}}) \end{aligned} \tag{2.2.1.1}$$

In this equation, *c* is the specific heat of the substance, *m* is its mass, and ΔT (which is read "delta T") is the temperature change, $T_{\text{final}} - T_{\text{initial}}$. If a substance gains thermal energy, its temperature increases, its final temperature is higher than its initial temperature, $T_{\text{final}} - T_{\text{initial}}$ has a positive value, and the value of *q* is positive. If a substance loses thermal energy, its temperature decreases, the final temperature is lower than the initial temperature, $T_{\text{final}} - T_{\text{initial}}$ has a negative value, and the value of *q* is negative.

Example 2.2.1.1: Measuring Heat

A flask containing $8.0\ 10^2$ g of water is heated, and the temperature of the water increases from 21 °C to 85 °C. How much heat did the water absorb?

Solution

To answer this question, consider these factors:

- the specific heat of the substance being heated (in this case, water)
- the amount of substance being heated (in this case, 8.0×10^2 g)
- the magnitude of the temperature change (in this case, from 21 °C to 85 °C).

The specific heat of water is 4.184 J/g °C, so to heat 1 g of water by 1 °C requires 4.184 J. We note that since 4.184 J is required to heat 1 g of water by 1 °C, we will need *800 times as much* to heat 8.0×10^2 g of water by 1 °C. Finally, we observe that since 4.184 J are required to heat 1 g of water by 1 °C, we will need *64 times as much* to heat it by 64 °C (that is, from 21 °C to 85 °C).

This can be summarized using Equation 2.2.1.1:

$$\begin{aligned} q = c \times m \times \Delta T &= c \times m \times (T_{\text{final}} - T_{\text{initial}}) \\ &= \left(4.184 J/\frac{g}{^\circ}{}^\circ C\right) \times \left(8.0 \times 10^2 g\right) \times (85 - 21)^\circ C \\ &= (4.184 J/g^\circ C) \times \left(8.0 \times 10^2 g\right) \times (64)^\circ C \\ &= 210,000 J \left(= 2.1 \times 10^2 kJ\right) \end{aligned} \tag{2.2.1.2}$$

Because the temperature increased, the water absorbed heat and q is positive.

Exercise 2.2.1.1

How much heat, in joules, must be added to a 502 g iron skillet to increase its temperature from 25 °C to 250 °C? The specific heat of iron is 0.449 J/g °C.

Answer available at the end of this section.

Note that the relationship between heat, specific heat, mass, and temperature change can be used to determine any of these quantities (not just heat) if the other three are known or can be deduced.

Example 2.2.1.2: Determining Other Quantities

A piece of unknown metal weighs 348 g. When the metal piece absorbs 6.64 kJ of heat, its temperature increases from 22.4 °C to 43.6 °C. Determine the specific heat of this metal (which might provide a clue to its identity).

Solution

Since mass, heat, and temperature change are known for this metal, we can determine its specific heat using the relationship in Equation 2.2.1.1:

$$q = c \times m \times \Delta T = c \times m \times (T_{\text{final}} - T_{\text{initial}})$$

Substituting the known values:

$$6640J = c \times (348g) \times (43.6 - 22.4)^{\circ}C$$

Solving:

$$c = \frac{6640J}{(348g) \times (21.2^{\circ}C)} = 0.900J/g^{\circ}C$$

Comparing this value with the values in Table 2.2.1.1, this value matches the specific heat of aluminum, which suggests that the unknown metal may be aluminum.

Exercise 2.2.1.2

A piece of unknown metal weighs 217 g. When the metal piece absorbs 1.43 kJ of heat, its temperature increases from 24.5 °C to 39.1 °C. Determine the specific heat of this metal, and predict its identity.

Answer available at the end of this section.

Chemistry in Everyday Life: Solar Thermal Energy Power Plants

The sunlight that reaches the earth contains thousands of times more energy than we presently capture. Solar thermal systems provide one possible solution to the problem of converting energy from the sun into energy we can use. Large-scale solar thermal plants have different design specifics, but all concentrate sunlight to heat some substance; the heat “stored” in that substance is then converted into electricity.

The Solana Generating Station in Arizona’s Sonora Desert produces 280 megawatts of electrical power. It uses parabolic mirrors that focus sunlight on pipes filled with a heat transfer fluid (HTF) (Figure 2.2.1.8). The HTF then does two things: It turns water into steam, which spins turbines, which in turn produces electricity, and it melts and heats a mixture of salts,

which functions as a thermal energy storage system. After the sun goes down, the molten salt mixture can then release enough of its stored heat to produce steam to run the turbines for 6 hours. Molten salts are used because they possess a number of beneficial properties, including high heat capacities and thermal conductivities.

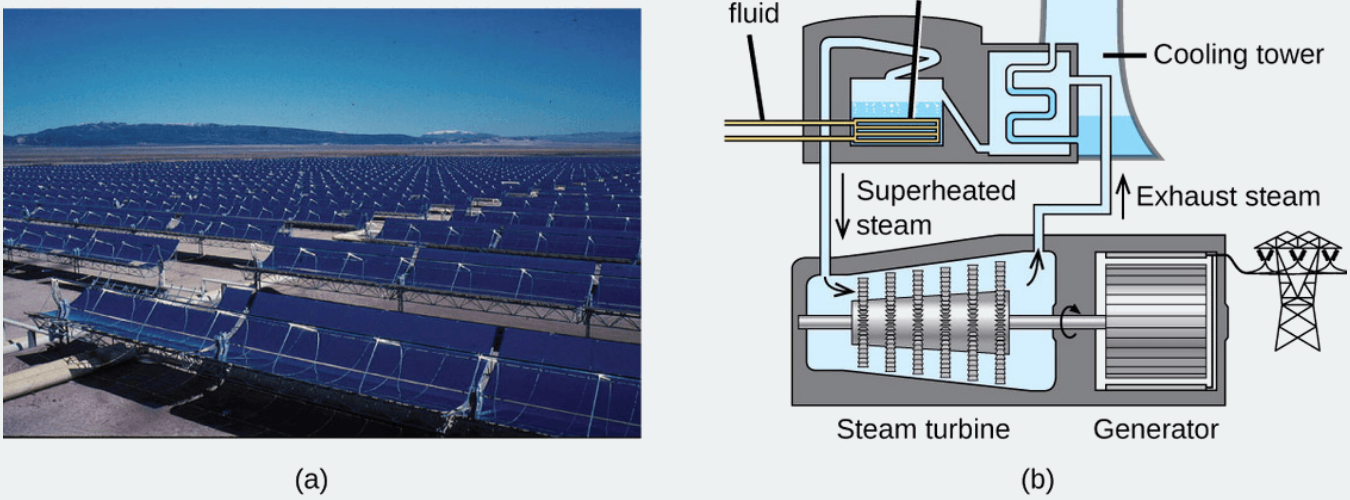


Figure 2.2.1.8: This solar thermal plant uses parabolic trough mirrors to concentrate sunlight. (credit a: modification of work by Bureau of Land Management)

The 377-megawatt Ivanpah Solar Generating System, located in the Mojave Desert in California, is the largest solar thermal power plant in the world (Figure 2.2.1.9). Its 170,000 mirrors focus huge amounts of sunlight on three water-filled towers, producing steam at over 538 °C that drives electricity-producing turbines. It produces enough energy to power 140,000 homes. Water is used as the working fluid because of its large heat capacity and heat of vaporization.



Figure 2.2.1.9: (a) The Ivanpah solar thermal plant uses 170,000 mirrors to concentrate sunlight on water-filled towers. (b) It covers 4000 acres of public land near the Mojave Desert and the California-Nevada border. (credit a: modification of work by Craig Dietrich; credit b: modification of work by “USFWS Pacific Southwest Region”/Flickr)

Answers

Exercise 2.2.1.1

$5.07\ 10^{4}$ J

Exercise 2.2.1.2

c = 0.451 J/g °C; the metal is likely to be iron

This page titled 2.2.1: Energy Basics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.1: Energy Basics by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.2.2: Enthalpy

Learning Objectives

By the end of this section, you will be able to:

- State the first law of thermodynamics
- Define enthalpy and explain its classification as a state function
- Write and balance thermochemical equations
- Calculate enthalpy changes for various chemical reactions
- Explain Hess's law and use it to compute reaction enthalpies

Thermochemistry is a branch of chemical thermodynamics, the science that deals with the relationships between heat, work, and other forms of energy in the context of chemical and physical processes. As we concentrate on thermochemistry in this chapter, we need to consider some widely used concepts of thermodynamics.

Substances act as reservoirs of energy, meaning that energy can be added to them or removed from them. Energy is stored in a substance when the kinetic energy of its atoms or molecules is raised. The greater kinetic energy may be in the form of increased translations (travel or straight-line motions), vibrations, or rotations of the atoms or molecules. When thermal energy is lost, the intensities of these motions decrease and the kinetic energy falls. The total of all possible kinds of energy present in a substance is called the internal energy (*U*), sometimes symbolized as *E*.

As a system undergoes a change, its internal energy can change, and energy can be transferred from the system to the surroundings, or from the surroundings to the system. Energy is transferred into a system when it absorbs heat (*q*) from the surroundings or when the surroundings do work (*w*) on the system. For example, energy is transferred into room-temperature metal wire if it is immersed in hot water (the wire absorbs heat from the water), or if you rapidly bend the wire back and forth (the wire becomes warmer because of the work done on it). Both processes increase the internal energy of the wire, which is reflected in an increase in the wire's temperature. Conversely, energy is transferred out of a system when heat is lost from the system, or when the system does work on the surroundings.

The relationship between internal energy, heat, and work can be represented by the equation:

$$\Delta U = q + w$$

as shown in Figure $2.2.2.1$. This is one version of the first law of thermodynamics, and it shows that the internal energy of a system changes through heat flow into or out of the system (positive *q* is heat flow in; negative *q* is heat flow out) or work done on or by the system. The work, *w*, is positive if it is done on the system and negative if it is done by the system.

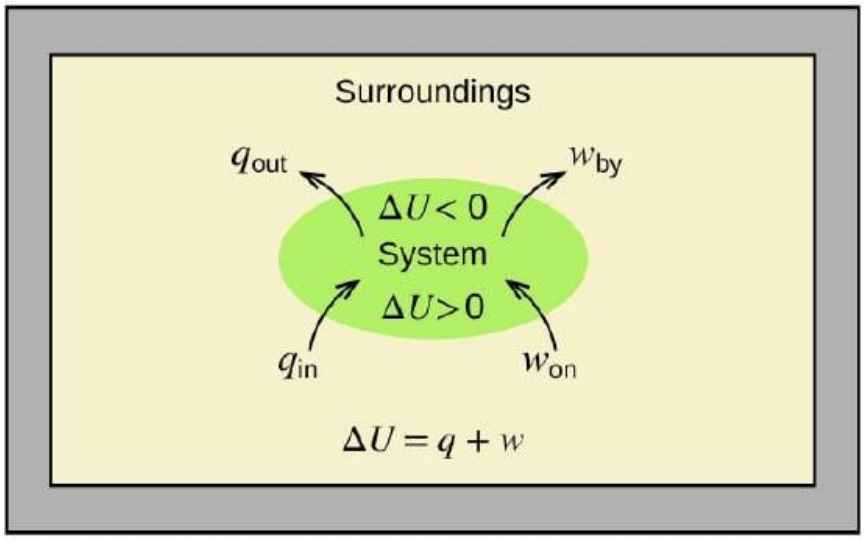


Figure 2.2.2.1: The internal energy, U, of a system can be changed by heat flow and work. If heat flows into the system, q_{in}, or work is done on the system, w_{on}, its internal energy increases, $\Delta U > 0$. If heat flows out of the system, q_{out}, or work is done by the system, w_{by}, its internal energy decreases, $\Delta U < 0$.

A type of work called expansion work (or pressure-volume work) occurs when a system pushes back the surroundings against a restraining pressure, or when the surroundings compress the system. An example of this occurs during the operation of an internal combustion engine. The reaction of gasoline and oxygen is exothermic. Some of this energy is given off as heat, and some does work pushing the piston in the cylinder. The substances involved in the reaction are the system, and the engine and the rest of the universe are the surroundings. The system loses energy by both heating and doing work on the surroundings, and its internal energy decreases. (The engine is able to keep the car moving because this process is repeated many times per second while the engine is running.) We will consider how to determine the amount of work involved in a chemical or physical change in the chapter on thermodynamics.

Link to Learning

This view of an internal combustion engine illustrates the conversion of energy produced by the exothermic combustion reaction of a fuel such as gasoline into energy of motion.

As discussed, the relationship between internal energy, heat, and work can be represented as $\Delta U = q + w$. Internal energy is an example of a state function (or state variable), whereas heat and work are not state functions. The value of a state function depends only on the state that a system is in, and not on how that state is reached. If a quantity is not a state function, then its value *does* depend on how the state is reached. An example of a state function is altitude or elevation. If you stand on the summit of Mt. Kilimanjaro, you are at an altitude of 5895 m, and it does not matter whether you hiked there or parachuted there. The distance you traveled to the top of Kilimanjaro, however, is not a state function. You could climb to the summit by a direct route or by a more roundabout, circuitous path (Figure 2.2.2.2). The distances traveled would differ (distance is not a state function) but the elevation reached would be the same (altitude is a state function).

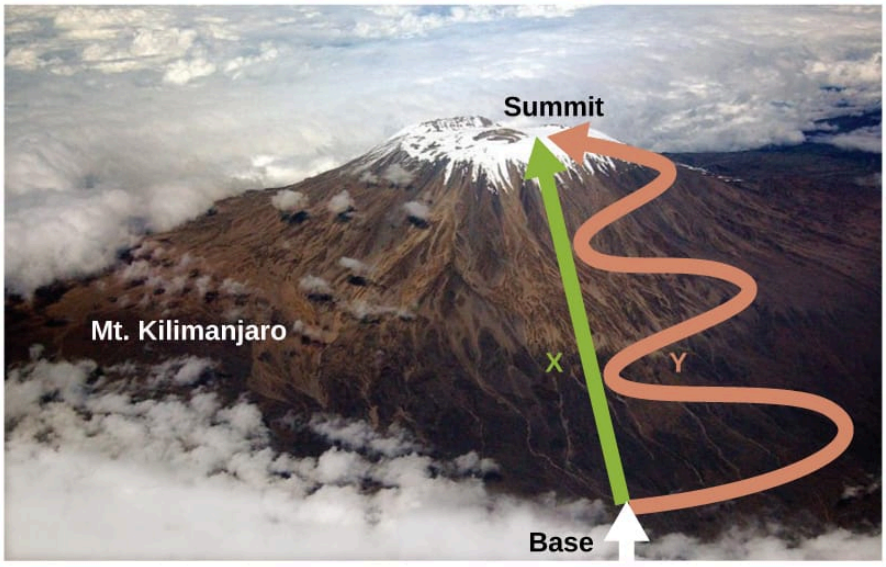


Figure 2.2.2.2: Paths X and Y represent two different routes to the summit of Mt. Kilimanjaro. Both have the same change in elevation (altitude or elevation on a mountain is a state function; it does not depend on path), but they have very different distances traveled (distance walked is not a state function; it depends on the path). (credit: modification of work by Paul Shaffner)

Chemists ordinarily use a property known as enthalpy (*H*) to describe the thermodynamics of chemical and physical processes. Enthalpy is defined as the sum of a system's internal energy (*U*) and the mathematical product of its pressure (*P*) and volume (*V*):

$$H = U + PV$$

Enthalpy is also a state function. Enthalpy values for specific substances cannot be measured directly; only enthalpy *changes* for chemical or physical processes can be determined. For processes that take place at constant pressure (a common condition for many chemical and physical changes), the enthalpy change (Δ*H*) is:

$$\Delta H = \Delta U + P\Delta V$$

The mathematical product $P\Delta V$ represents work (w), namely, expansion or pressure-volume work as noted. By their definitions, the arithmetic signs of ΔV and w will **always** be opposite:

$$P\Delta V = -w$$

Substituting this equation and the definition of internal energy into the enthalpy-change equation yields:

$$\begin{aligned}\Delta H &= \Delta U + P\Delta V \\ &= q_p + w - w \\ &= q_p\end{aligned}$$

where q_p is the **heat of reaction** under conditions of constant pressure.

And so, if a chemical or physical process is carried out at constant pressure with the only work done caused by expansion or contraction, then the heat flow (q_p) and enthalpy change (Δ*H*) for the process are equal.

The heat given off when you operate a Bunsen burner is equal to the enthalpy change of the methane combustion reaction that takes place, since it occurs at the essentially constant pressure of the atmosphere. On the other hand, the heat produced by a reaction measured in a bomb calorimeter is not equal to ΔH because the closed, constant-volume metal container prevents the pressure from remaining constant (it may increase or decrease if the reaction yields increased or decreased amounts of gaseous species). Chemists usually perform experiments under normal atmospheric conditions, at constant external pressure with $q = \Delta H$, which makes enthalpy the most convenient choice for determining heat changes for chemical reactions.

The following conventions apply when using Δ*H*:

- A negative value of an enthalpy change, Δ*H* < 0, indicates an exothermic reaction; a positive value, Δ*H* > 0, indicates an

endothermic reaction. If the direction of a chemical equation is reversed, the arithmetic sign of its ΔH is changed (a process that is endothermic in one direction is exothermic in the opposite direction).

- Chemists use a thermochemical equation to represent the changes in both matter and energy. In a thermochemical equation, the enthalpy change of a reaction is shown as a ΔH value following the equation for the reaction. This ΔH value indicates the amount of heat associated with the reaction involving the number of moles of reactants and products *as shown in the chemical equation*. For example, consider this equation:

$$\mathrm{H_2\,(g)} + \tfrac{1}{2}\,\mathrm{O_2\,(g)} \longrightarrow \mathrm{H_2O\,(l)} \quad \Delta H = -286kJ$$

This equation indicates that when 1 mole of hydrogen gas and mole of oxygen gas at some temperature and pressure change to 1 mole of liquid water at the same temperature and pressure, 286 kJ of heat are released to the surroundings. If the coefficients of the chemical equation are multiplied by some factor, the enthalpy change must be multiplied by that same factor (ΔH is an extensive property):

  - two-fold increase in amounts

$$2\,\mathrm{H_2\,(g)} + \mathrm{O_2\,(g)} \longrightarrow 2\,\mathrm{H_2O\,(l)} \quad \Delta H = 2 \times (-286kJ) = -572kJ$$

  - two-fold decrease in amounts

$$\tfrac{1}{2}\,\mathrm{H_2\,(g)} + \tfrac{1}{4}\,\mathrm{O_2\,(g)} \longrightarrow \tfrac{1}{2}\,\mathrm{H_2O\,(l)} \quad \Delta H = \frac{1}{2} \times (-286kJ) = -143kJ$$

- The enthalpy change of a reaction depends on the physical states of the reactants and products, so these must be shown. For example, when 1 mole of hydrogen gas and mole of oxygen gas change to 1 mole of liquid water at the same temperature and pressure, 286 kJ of heat are released. If gaseous water forms, only 242 kJ of heat are released.

$$\mathrm{H_2\,(g)} + \tfrac{1}{2}\,\mathrm{O_2\,(g)} \longrightarrow \mathrm{H_2O\,(g)} \quad \Delta H = -242kJ$$

Example 2.2.2.1: Writing Thermochemical Equations

When 0.0500 mol of HCl(*aq*) reacts with 0.0500 mol of NaOH(*aq*) to form 0.0500 mol of NaCl(*aq*), 2.9 kJ of heat are produced. Write a balanced thermochemical equation for the reaction of one mole of HCl.

$$\mathrm{HCl\,(aq)} + \mathrm{NaOH\,(aq)} \longrightarrow \mathrm{NaCl\,(aq)} + \mathrm{H_2O\,(l)}$$

Solution

For the reaction of 0.0500 mol acid (HCl), q = −2.9 kJ. The reactants are provided in stoichiometric amounts (same molar ratio as in the balanced equation), and so the amount of acid may be used to calculate a molar enthalpy change. Since ΔH is an extensive property, it is proportional to the amount of acid neutralized:

$$\Delta H = 1\ \cancel{\text{mol HCl}} \times \frac{-2.9kJ}{0.0500\ \cancel{\text{mol HCl}}} = -58kJ$$

The thermochemical equation is then

$$\mathrm{HCl\,(aq)} + \mathrm{NaOH\,(aq)} \longrightarrow \mathrm{NaCl\,(aq)} + \mathrm{H_2O\,(l)} \quad \Delta H = -58kJ$$

Exercise 2.2.2.1

When 1.34 g Zn(s) reacts with 60.0 mL of 0.750 M HCl(*aq*), 3.14 kJ of heat are produced. Determine the enthalpy change per mole of zinc reacting for the reaction:

$$\mathrm{Zn\,(s)} + 2\,\mathrm{HCl\,(aq)} \longrightarrow \mathrm{ZnCl_2\,(aq)} + \mathrm{H_2\,(g)}$$

Answer available at the end of this section.

Be sure to take both stoichiometry and limiting reactants into account when determining the ΔH for a chemical reaction.

Example 2.2.2.2: Writing Thermochemical Equations

A gummy bear contains 2.67 g sucrose, $C_{12}H_{22}O_{11}$. When it reacts with 7.19 g potassium chlorate, $KClO_3$, 43.7 kJ of heat are produced. Write a thermochemical equation for the reaction of one mole of sucrose:

$$\mathrm{C_{12}H_{22}O_{11}\,(aq) + 8\,KClO_3\,(aq) \longrightarrow 12\,CO_2\,(g) + 11\,H_2O\,(l) + 8\,KCl\,(aq)}$$

Solution

Unlike the previous example exercise, this one does not involve the reaction of stoichiometric amounts of reactants, and so the *limiting reactant* must be identified (it limits the yield of the reaction and the amount of thermal energy produced or consumed).

The provided amounts of the two reactants are

$$(2.67\,\mathrm{g})\left(\frac{1\,\mathrm{mol}}{342.3\,\mathrm{g}}\right) = 0.00780\,\mathrm{mol\,C_{12}H_{22}O_{11}}$$

$$(7.19\,\mathrm{g})\left(\frac{1\,\mathrm{mol}}{122.5\,\mathrm{g}}\right) = 0.0587\,\mathrm{mol\,KCIO}$$

The provided molar ratio of perchlorate-to-sucrose is then

$$\frac{0.0587\,\mathrm{mol\ KClO_3}}{0.00780\,\mathrm{mol\,C_{12}H_{22}O_{11}}} = 7.52$$

The balanced equation indicates 8 mol $KClO_3$ are required for reaction with 1 mol $C_{12}H_{22}O_{11}$. Since the provided amount of $KClO_3$ is less than the stoichiometric amount, it is the limiting reactant and may be used to compute the enthalpy change:

$$\Delta H = \frac{-43.7\,\mathrm{kJ}}{0.0587\,\mathrm{molKCIO_3}} = \frac{744\,\mathrm{kJ}}{\mathrm{mol\,KCIO_3}}$$

Because the equation, as written, represents the reaction of 8 mol $KClO_3$, the enthalpy change is

$$\left(\frac{744\,\mathrm{kJ}}{1\ \cancel{\mathrm{mol\,KCIO_3}}}\right)\left(8\ \cancel{\mathrm{mol\,KCIO_3}}\right) = 5960\,\mathrm{kJ}$$

The enthalpy change for this reaction is –5960 kJ, and the thermochemical equation is:

$$\mathrm{C_{12}H_{22}O_{11} + 8\,KClO_3 \longrightarrow 12\,CO_2 + 11\,H_2O + 8\,KCl} \quad \Delta H = -5960\,\mathrm{kJ}$$

Exercise 2.2.2.1

When 1.42 g of iron reacts with 1.80 g of chlorine, 3.22 g of $FeCl_2\,(s)$ and 8.60 kJ of heat is produced. What is the enthalpy change for the reaction when 1 mole of $FeCl_2\,(s)$ is produced?

Answer available at the end of this section.

Enthalpy changes are typically tabulated for reactions in which both the reactants and products are at the same conditions. A standard state is a commonly accepted set of conditions used as a reference point for the determination of properties under other different conditions. For chemists, the IUPAC standard state refers to materials under a pressure of 1 bar and solutions at 1 M, and does not specify a temperature. Many thermochemical tables list values with a standard state of 1 atm. Because the ΔH of a reaction changes very little with such small changes in pressure (1 bar = 0.987 atm), ΔH values (except for the most precisely measured values) are essentially the same under both sets of standard conditions. We will include a superscripted "o" in the enthalpy change symbol to designate standard state. Since the usual (but not technically standard) temperature is 298.15 K, this temperature will be assumed unless some other temperature is specified. Thus, the symbol is used to indicate an enthalpy change for a process occurring under these conditions. (The symbol ΔH is used to indicate an enthalpy change for a reaction occurring under nonstandard conditions.)

The enthalpy changes for many types of chemical and physical processes are available in the reference literature, including those for combustion reactions, phase transitions, and formation reactions. As we discuss these quantities, it is important to pay attention to the *extensive* nature of enthalpy and enthalpy changes. Since the enthalpy change for a given reaction is proportional to the amounts of substances involved, it may be reported on that basis (i.e., as the ΔH for specific amounts of reactants). However, we often find it more useful to divide one extensive property (ΔH) by another (amount of substance), and report a per-amount *intensive* value of ΔH, often "normalized" to a per-mole basis. (Note that this is similar to determining the intensive property specific heat from the extensive property heat capacity, as seen previously.)

Standard Enthalpy of Combustion

Standard enthalpy of combustion is the enthalpy change when 1 mole of a substance burns (combines vigorously with oxygen) under standard state conditions; it is sometimes called "heat of combustion." For example, the enthalpy of combustion of ethanol, –1366.8 kJ/mol, is the amount of heat produced when one mole of ethanol undergoes complete combustion at 25 °C and 1 atmosphere pressure, yielding products also at 25 °C and 1 atm.

$$\mathrm{C_2H_5OH\,(l) + 3\,O_2\,(g) \longrightarrow 2\,CO_2 + 3\,H_2O\,(l)} \quad \Delta H^\circ = -1366.8\,\mathrm{kJ}$$

Enthalpies of combustion for many substances have been measured; a few of these are listed in Table 2.2.2.1. Many readily available substances with large enthalpies of combustion are used as fuels, including hydrogen, carbon (as coal or charcoal), and hydrocarbons (compounds containing only hydrogen and carbon), such as methane, propane, and the major components of gasoline.

Table 2.2.2.1: Standard Molar Enthalpies of Combustion

Substance	Combustion Reaction	Enthalpy of Combustion, ΔH_c° (kJ/mol) at 25 °C
carbon	$C(s) + O_2(g) \longrightarrow CO_2(g)$	–393.5
hydrogen	$H_2(g) + \frac{1}{2}O_2(g) \longrightarrow H_2O(l)$	–285.8
magnesium	$Mg(s) + \frac{1}{2}O_2(g) \longrightarrow MgO(s)$	–601.6
sulfur	$S(s) + O_2(g) \longrightarrow SO_2(g)$	–296.8
carbon monoxide	$CO(g) + \frac{1}{2}O_2(g) \longrightarrow CO_2(g)$	–283.0
methane	$CH_4(g) + 2O_2(g) \longrightarrow CO_2(g) + 2H_2O(l)$	–890.8
acetylene	$C_2H_2(g) + \frac{5}{2}O_2(g) \longrightarrow 2CO_2(g) + H_2O(l)$	–1301.1
ethanol	$C_2H_5OH(l) + 3O_2(g) \longrightarrow 2CO_2(g) + 3H_2O(l)$	–1366.8
methanol	$CH_3OH(l) + \frac{3}{2}O_2(g) \longrightarrow CO_2(g) + 2H_2O(l)$	–726.1
isooctane	$C_8H_{18}(l) + \frac{25}{2}O_2(g) \longrightarrow 8CO_2(g) + 9H_2O(l)$	–5461

Example 2.2.2.3: Using Enthalpy of Combustion

As Figure 2.2.2.3: suggests, the combustion of gasoline is a highly exothermic process. Let us determine the approximate amount of heat produced by burning 1.00 L of gasoline, assuming the enthalpy of combustion of gasoline is the same as that of isooctane, a common component of gasoline. The density of isooctane is 0.692 g/mL.



Figure 2.2.2.3: The combustion of gasoline is very exothermic. (credit: modification of work by "AlexEagle"/Flickr)

Solution

Starting with a known amount (1.00 L of isooctane), we can perform conversions between units until we arrive at the desired amount of heat or energy. The enthalpy of combustion of isooctane provides one of the necessary conversions. Table 2.2.2.1 gives this value as –5460 kJ per 1 mole of isooctane (C_8H_{18}).

Using these data,

$$1.00\ \cancel{\text{L C}_8\text{H}_{18}} \times \left(\frac{1000\ \cancel{\text{mL C}_8\text{H}_{18}}}{1\ \cancel{\text{L C}_8\text{H}_{18}}}\right) \times \left(\frac{0.692\ \cancel{\text{g C}_8\text{H}_{18}}}{1\ \cancel{\text{mL C}_8\text{H}_{18}}}\right) \times \left(\frac{1\ \cancel{\text{mol C}_8\text{H}_{18}}}{114\ \cancel{\text{g C}_8\text{H}_{18}}}\right) \times \frac{-5460\,\text{kJ}}{1\ \cancel{\text{mol C}_8\text{H}_{18}}} = -3.31 \times 10^4\,\text{kJ}$$

The combustion of 1.00 L of isooctane produces 33,100 kJ of heat. (This amount of energy is enough to melt 99.2 kg, or about 218 lbs, of ice.)

Note: If you do this calculation one step at a time, you would find:

$$1.00\,\text{L C}_8\text{H}_{18} \longrightarrow 1.00 \times 10^3\,\text{mL C}_8\text{H}_{18}$$
$$1.00 \times 10^3\,\text{mL C}_8\text{H}_{18} \longrightarrow 692\,\text{g C}_8\text{H}_{18}$$
$$692\,\text{g C}_8\text{H}_{18} \longrightarrow 6.07\,\text{mol C}_8\text{H}_{18}$$
$$6.07\,\text{mol C}_8\text{H}_{18} \longrightarrow -3.31 \times 10^4 kJ$$

Exercise 2.2.2.3

How much heat is produced by the combustion of 125 g of acetylene?

Answer available at the end of this section.

Chemistry in Everyday Life: Emerging Algae-Based Energy Technologies (Biofuels)

As reserves of fossil fuels diminish and become more costly to extract, the search is ongoing for replacement fuel sources for the future. Among the most promising biofuels are those derived from algae (Figure 5.22). The species of algae used are nontoxic, biodegradable, and among the world's fastest growing organisms. About 50% of algal weight is oil, which can be

readily converted into fuel such as biodiesel. Algae can yield 26,000 gallons of biofuel per hectare—much more energy per acre than other crops. Some strains of algae can flourish in brackish water that is not usable for growing other crops. Algae can produce biodiesel, biogasoline, ethanol, butanol, methane, and even jet fuel.

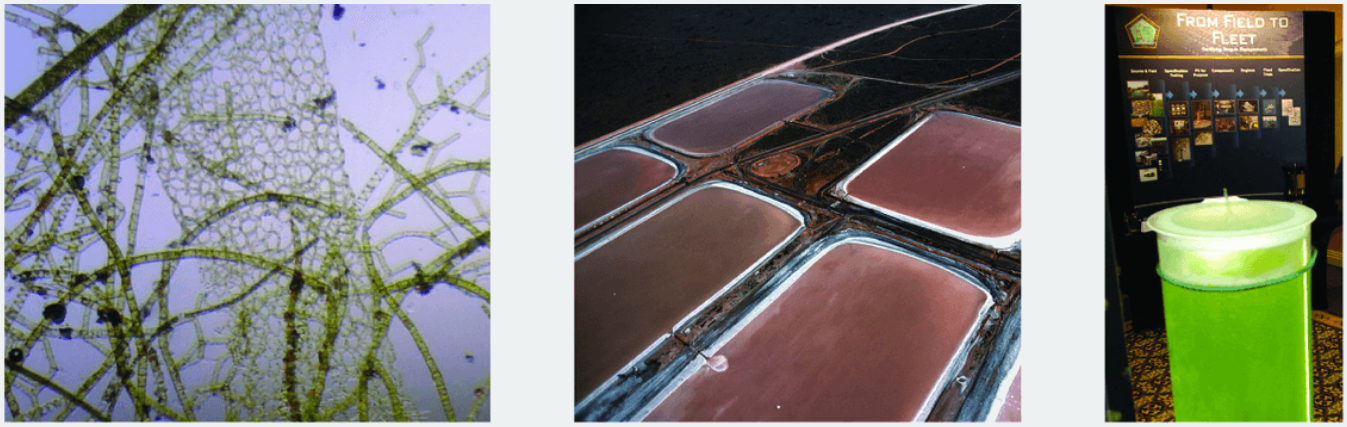


Figure 5.22 (a) Tiny algal organisms can be (b) grown in large quantities and eventually (c) turned into a useful fuel such as biodiesel. (credit a: modification of work by Micah Sittig; credit b: modification of work by Robert Kerton; credit c: modification of work by John F. Williams)

According to the US Department of Energy, only 39,000 square kilometers (about 0.4% of the land mass of the US or less than of the area used to grow corn) can produce enough algal fuel to replace all the petroleum-based fuel used in the US. The cost of algal fuels is becoming more competitive—for instance, the US Air Force is producing jet fuel from algae at a total cost of under $5 per gallon.[3] The process used to produce algal fuel is as follows: grow the algae (which use sunlight as their energy source and CO_2 as a raw material); harvest the algae; extract the fuel compounds (or precursor compounds); process as necessary (e.g., perform a transesterification reaction to make biodiesel); purify; and distribute (Figure 2.2.2.4).

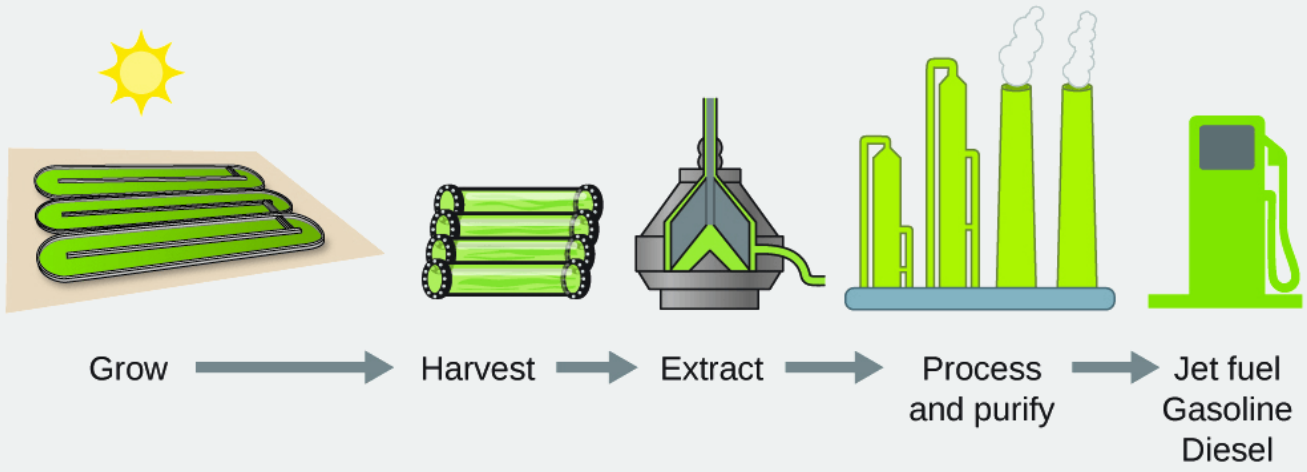


Figure 2.2.2.4: Algae convert sunlight and carbon dioxide into oil that is harvested, extracted, purified, and transformed into a variety of renewable fuels.

Link to Learning

Click here to learn more about the process of creating algae biofuel.

Standard Enthalpy of Formation

A standard enthalpy of formation is an enthalpy change for a reaction in which exactly 1 mole of a pure substance is formed from free elements in their most stable states under standard state conditions. These values are especially useful for computing or predicting enthalpy changes for chemical reactions that are impractical or dangerous to carry out, or for processes for which it is difficult to make measurements. If we have values for the appropriate standard enthalpies of formation, we can determine the enthalpy change for any reaction, which we will practice in the next section on Hess's law.

The standard enthalpy of formation of $CO_2(g)$ is –393.5 kJ/mol. This is the enthalpy change for the exothermic reaction:

$$\mathrm{C(s) + O_2(g) \longrightarrow CO_2(g) \quad \Delta H_f^\circ = \Delta H^\circ = -393.5\,kJ}$$

starting with the reactants at a pressure of 1 atm and 25 °C (with the carbon present as graphite, the most stable form of carbon under these conditions) and ending with one mole of CO_2, also at 1 atm and 25 °C. For nitrogen dioxide, $NO_2(g)$, is 33.2 kJ/mol. This is the enthalpy change for the reaction:

$$\mathrm{\tfrac{1}{2}\,N_2(g) + O_2(g) \longrightarrow NO_2(g) \quad \Delta H_f^\circ = \Delta H^\circ = +33.2\,kJ}$$

A reaction equation with mole of $\mathrm{N_2}$ and 1 mole of $\mathrm{O_2}$ is correct in this case because the standard enthalpy of formation always refers to 1 mole of product, $\mathrm{NO_2(g)}$.

You will find a table of standard enthalpies of formation of many common substances in Appendix G. These values indicate that formation reactions range from highly exothermic (such as –2984 kJ/mol for the formation of $\mathrm{P_4O_{10}}$) to strongly endothermic (such as +226.7 kJ/mol for the formation of acetylene, $\mathrm{C_2H_2}$). By definition, the standard enthalpy of formation of an element in its most stable form is equal to zero under standard conditions, which is 1 atm for gases and 1 M for solutions.

Example 2.2.2.4: Evaluating an Enthalpy of Formation

Ozone, $\mathrm{O_3(g)}$, forms from oxygen, $\mathrm{O_2(g)}$, by an endothermic process. Ultraviolet radiation is the source of the energy that drives this reaction in the upper atmosphere. Assuming that both the reactants and products of the reaction are in their standard states, determine the standard enthalpy of formation, of ozone from the following information:

$$\mathrm{3\,O_2(g) \longrightarrow 2\,O_3(g) \quad \Delta H^\circ = +286\,kJ}$$

Solution

ΔH_f° is the enthalpy change for the formation of one mole of a substance in its standard state from the elements in their standard states. Thus, for $\mathrm{O_3(g)}$ is the enthalpy change for the reaction:

$$\mathrm{\tfrac{3}{2}\,O_2(g) \longrightarrow O_3(g)}$$

For the formation of 2 mol of $\mathrm{O_3(g)}$, $\Delta H^\circ = +286$ kJ. This ratio, can be used as a conversion factor to find the heat produced when 1 mole of $O_3(g)$ is formed, which is the enthalpy of formation for $\mathrm{O_3(g)}$:

$$\Delta H^\circ \text{ for 1 mole of } \mathrm{O_3}(g) = 1\ \cancel{\text{mol O}_3} \times \frac{286\,\text{kJ}}{2\,\text{mol O}_3} = 143\,\text{kJ}$$

Therefore,

$$\Delta H_f^\circ\,[O_3(g)] = +143\,\text{kJ/mol}.$$

Exercise 2.2.2.4

Hydrogen gas, $\mathrm{H_2}$, reacts explosively with gaseous chlorine, $\mathrm{Cl_2}$, to form hydrogen chloride, $\mathrm{HCl(g)}$. What is the enthalpy change for the reaction of 1 mole of $\mathrm{H_2(g)}$ with 1 mole of $\mathrm{Cl_2(g)}$ if both the reactants and products are at standard state conditions? The standard enthalpy of formation of $\mathrm{HCl(g)}$ is –92.3 kJ/mol.

Answer available at the end of this section.

Example 2.2.2.5: Writing Reaction Equations for

Write the heat of formation reaction equations for:

a. $C_2H_5OH\,(l)$
b. $Ca_3(PO_4)_2\,(s)$

Solution

Remembering that reaction equations are for forming 1 mole of the compound from its constituent elements under standard conditions, we have:

a. $$2\,C\,(s, \text{graphite}) + 3\,H_2\,(g) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow C_2H_5OH\,(l)$$
b. $$3\,Ca\,(s) + \tfrac{1}{2}\,P_4\,(s) + 4\,O_2\,(g) \longrightarrow Ca_3(PO_4)_2\,(s)$$

Note: The standard state of carbon is graphite, and phosphorus exists as P_4.

Exercise 2.2.2.5

Write the heat of formation reaction equations for:

a. $C_2H_5OC_2H_5\,(l)$
b. $Na_2CO_3\,(s)$

Answer available at the end of this section.

Hess's Law

There are two ways to determine the amount of heat involved in a chemical change: measure it experimentally, or calculate it from other experimentally determined enthalpy changes. Some reactions are difficult, if not impossible, to investigate and make accurate measurements for experimentally. And even when a reaction is not hard to perform or measure, it is convenient to be able to determine the heat involved in a reaction without having to perform an experiment.

This type of calculation usually involves the use of Hess's law, which states: *If a process can be written as the sum of several stepwise processes, the enthalpy change of the total process equals the sum of the enthalpy changes of the various steps*. Hess's law is valid because enthalpy is a state function: Enthalpy changes depend only on where a chemical process starts and ends, but not on the path it takes from start to finish. For example, we can think of the reaction of carbon with oxygen to form carbon dioxide as occurring either directly or by a two-step process. The direct process is written:

$$C\,(s) + O_2\,(g) \longrightarrow CO_2\,(g) \quad \Delta H^\circ = -394kJ$$

In the two-step process, first carbon monoxide is formed:

$$C\,(s) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow CO\,(g) \quad \Delta H^\circ = -111kJ$$

Then, carbon monoxide reacts further to form carbon dioxide:

$$CO\,(g) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow CO_2\,(g) \quad \Delta H^\circ = -283kJ$$

The equation describing the overall reaction is the sum of these two chemical changes:

$$\begin{aligned}
&\text{step 1} && \mathrm{C(s) + \tfrac{1}{2}\,O_2(g) \longrightarrow CO(g)}\\
&\text{step 2} && \mathrm{CO(g) + \tfrac{1}{2}\,O_2(g) \longrightarrow CO_2(g)}\\
&\text{sum} && \mathrm{C(s) + \cancel{CO(g)} + O_2(g) \longrightarrow \cancel{CO(g)} + CO_2(g)}
\end{aligned}$$

Because the $\mathrm{CO(g)}$ produced in Step 1 is consumed in Step 2, the net change is:

$$\mathrm{C(s) + O_2(g) \longrightarrow CO_2(g)}$$

According to Hess's law, the enthalpy change of the reaction will equal the sum of the enthalpy changes of the steps.

$$\begin{aligned}
&\text{step 1} && \mathrm{C(s) + \tfrac{1}{2}\,O_2(g) \longrightarrow \cancel{CO(g)}} && \Delta H^\circ = -111\,\mathrm{kJ}\\
&\text{step 2} && \mathrm{\cancel{CO(g)} + \tfrac{1}{2}\,O_2(g) \longrightarrow CO_2(g)} && \Delta H^\circ = -283\,\mathrm{kJ}\\
&\text{sum} && \mathrm{C(s) + O_2(g) \longrightarrow CO_2(g)} && \Delta H^\circ = -394\,\mathrm{kJ}
\end{aligned}$$

The result is shown in Figure 2.2.2.5. We see that Δ*H* of the overall reaction is the same whether it occurs in one step or two. This finding (overall Δ*H* for the reaction = sum of Δ*H* values for reaction "steps" in the overall reaction) is true in general for chemical and physical processes.

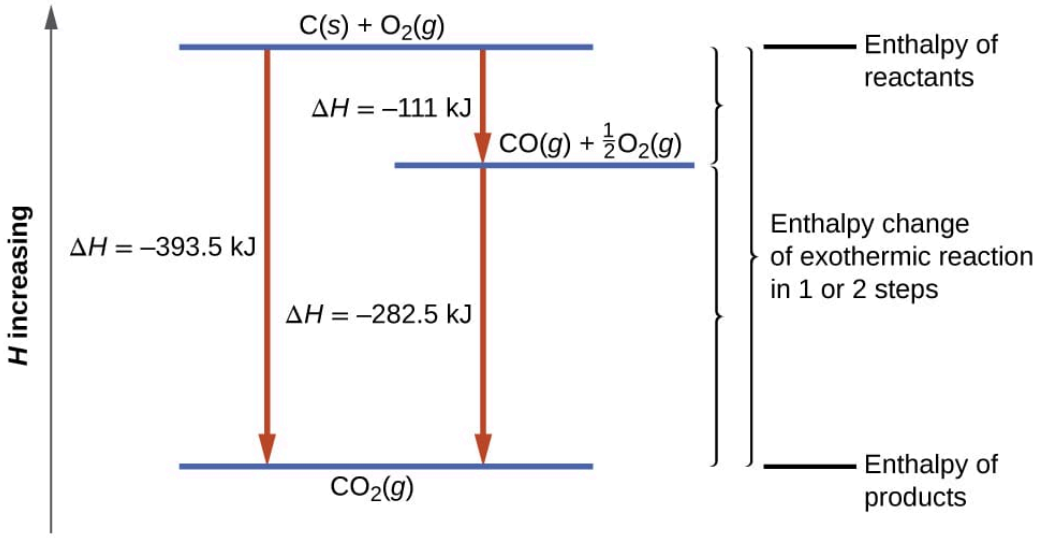


Figure 2.2.2.5: The formation of $\mathrm{CO_2(g)}$ from its elements can be thought of as occurring in two steps, which sum to the overall reaction, as described by Hess's law. The horizontal blue lines represent enthalpies. For an exothermic process, the products are at lower enthalpy than are the reactants.

Before we further practice using Hess's law, let us recall two important features of ΔH.

1. ΔH is directly proportional to the quantities of reactants or products. For example, the enthalpy change for the reaction forming 1 mole of $NO_2(g)$ is +33.2 kJ:

   $$\mathrm{\tfrac{1}{2}\,N_2(g) + O_2(g) \longrightarrow NO_2(g)} \quad \Delta H = +33.2\,\mathrm{kJ}$$

   When 2 moles of $\mathrm{NO_2}$ (twice as much) are formed, the Δ*H* will be twice as large:

   $$\mathrm{N_2(g) + 2\,O_2(g) \longrightarrow 2\,NO_2(g)} \quad \Delta H = +66.4\,\mathrm{kJ}$$

   In general, if we multiply or divide an equation by a number, then the enthalpy change should also be multiplied or divided by the same number.

2. ΔH for a reaction in one direction is equal in magnitude and opposite in sign to $\Delta H$for the reaction in the reverse direction. For example, given that:

$$\mathrm{H_2\,(g) + Cl_2\,(g) \longrightarrow 2\,HCl\,(g)} \quad \Delta H = -184.6\,\mathrm{kJ}$$

Then, for the "reverse" reaction, the enthalpy change is also "reversed":

Example 2.2.2.6: Stepwise Calculation of Heat of Formation Using Hess's Law

Determine the enthalpy of formation, of $\mathrm{FeCl_3\,(s)}$ from the enthalpy changes of the following two-step process that occurs under standard state conditions:

$$\mathrm{Fe\,(s) + Cl_2\,(g) \longrightarrow FeCl_2\,(s)} \quad \Delta H^\circ = -341.8\,\mathrm{kJ}$$

$$\mathrm{FeCl_2\,(s) + \tfrac{1}{2}\,Cl_2\,(g) \longrightarrow FeCl_3\,(s)} \quad \Delta H^\circ = -57.7\,\mathrm{kJ}$$

Solution

We are trying to find the standard enthalpy of formation of $\mathrm{FeCl_3\,(s)}$, which is equal to ΔH° for the reaction:

$$\mathrm{Fe\,(s) + \tfrac{3}{2}\,Cl_2\,(g) \longrightarrow FeCl_3\,(s)} \quad \Delta H_f^\circ = ?$$

Looking at the reactions, we see that the reaction for which we want to find ΔH° is the sum of the two reactions with known ΔH values, so we must sum their ΔH values:

step 1	$\mathrm{Fe\,(s) + Cl_2\,(g) \longrightarrow \cancel{FeCl2(s)}}$	$\Delta H^\circ = -341.8\,\mathrm{kJ}$
step 2	$\mathrm{\cancel{FeCl2(s)} + \tfrac{1}{2}\,Cl_2\,(g) \longrightarrow FeCl_3\,(s)}$	$\Delta H^\circ = -57.7\,\mathrm{kJ}$
sum	$\mathrm{Fe\,(s) + \tfrac{3}{2}\,Cl_2\,(g) \longrightarrow FeCl_3\,(s)}$	$\Delta H^\circ = -399.5\,\mathrm{kJ}$

The enthalpy of formation, ΔH_f° of $\mathrm{FeCl_3\,(s)}$ is –399.5 kJ/mol.

Exercise 2.2.2.6

Calculate ΔH for the process:

$$\mathrm{N_2\,(g) + 2\,O_2\,(g) \longrightarrow 2\,NO_2\,(g)}$$

from the following information:

$$\mathrm{N_2\,(g) + O_2\,(g) \longrightarrow 2\,NO\,(g)} \quad \Delta H = 180.5\,\mathrm{kJ}$$

$$\mathrm{NO\,(g) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow NO_2\,(g)} \quad \Delta H = -57.06\,\mathrm{kJ}$$

Answer available at the end of this section.

Here is a less straightforward example that illustrates the thought process involved in solving many Hess's law problems. It shows how we can find many standard enthalpies of formation (and other values of ΔH) if they are difficult to determine experimentally.

Example 2.2.2.7: A More Challenging Problem Using Hess's Law

Chlorine monofluoride can react with fluorine to form chlorine trifluoride:

i. $$\mathrm{ClF\,(g) + F_2\,(g) \longrightarrow ClF_3\,(g)} \quad \Delta H^\circ = ?$$

Use the reactions here to determine the ΔH^o for reaction (i):

ii. $$\mathrm{2\,OF_2\,(g) \longrightarrow O_2\,(g) + 2\,F_2\,(g)} \quad \Delta H^\circ_{(ii)} = -49.4\,\mathrm{kJ}$$

iii. $$\mathrm{2\,ClF\,(g) + O_2\,(g) \longrightarrow Cl_2O\,(g) + OF_2\,(g)} \quad \Delta H^\circ_{(iii)} = +214.0\,\mathrm{kJ}$$

iv. $$\mathrm{ClF_3\,(g) + O_2\,(g) \longrightarrow \tfrac{1}{2}\,Cl_2O\,(g) + \tfrac{3}{2}\,OF_2\,(g)} \quad \Delta H^\circ_{(in)} = +236.2\,\mathrm{kJ}$$

Solution

Our goal is to manipulate and combine reactions (ii), (iii), and (iv) such that they add up to reaction (i). Going from left to right in (i), we first see that ClF(g) is needed as a reactant. This can be obtained by multiplying reaction (iii) by $\frac{1}{2}$ which means that the ΔH^o change is also multiplied by $\frac{1}{2}$:

$$\mathrm{ClF\,(g) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow \tfrac{1}{2}\,Cl_2O\,(g) + \tfrac{1}{2}\,OF_2\,(g)} \quad \Delta H^\circ = \frac{1}{2}(214.0) = +107.0\,\mathrm{kJ}$$

Next, we see that F_2 is also needed as a reactant. To get this, reverse and halve reaction *(ii)*, which means that the ΔH^o changes sign and is halved:

$$\mathrm{\tfrac{1}{2}\,O_2\,(g) + F_2\,(g) \longrightarrow OF_2\,(g)} \quad \Delta H^\circ = +24.7\,\mathrm{kJ}$$

To get $\mathrm{ClF_3}$ as a product, reverse (iv), changing the sign of ΔH^o:

$$\mathrm{\tfrac{1}{2}\,Cl_2O\,(g) + \tfrac{3}{2}\,OF_2\,(g) \longrightarrow ClF_3\,(g) + O_2\,(g)} \quad \Delta H^\circ = -236.2kJ$$

Now check to make sure that these reactions add up to the reaction we want:

step 1	$\mathrm{ClF\,(g) + \cancel{1/2O2(g)} \longrightarrow \cancel{1/2Cl2O(g)} + \cancel{1/2OF2(g)}}$	$\Delta H^\circ = +107.0\,\mathrm{kJ}$
step 2	$\mathrm{\cancel{1/2O2(g)} + F_2\,(g) \longrightarrow \cancel{OF2(g)}}$	$\Delta H^\circ = +24.7\,\mathrm{kJ}$
step 3	$\mathrm{\cancel{1/2Cl2O(g)} + \cancel{3/2OF2(g)} \longrightarrow ClF_3\,(g) + \cancel{O2(g)}}$	$\Delta H^\circ = -236.2\,\mathrm{kJ}$
sum	$\mathrm{ClF\,(g) + F_2\,(g) \longrightarrow ClF_3\,(g)}$	$\Delta H^\circ = -104.5\,\mathrm{kJ}$

Reactants $\frac{1}{2}\,O_2$ and $\frac{1}{2}\,O_2$ cancel out the product O_2; the $\frac{1}{2}\,Cl_2O$ product cancels $\frac{1}{2}\,Cl_2O$ reactant and $\frac{3}{2}\,Cl_2O$ reactant is cancelled by products $\frac{1}{2}\,OF_2$ and OF_2. This leaves only reactants $\mathrm{ClF\,(g)}$ and $\mathrm{F_2\,(g)}$ and product $\mathrm{ClF_3\,(g)}$, which are what we want. Since summing these three modified reactions yields the reaction of interest, summing the three modified ΔH^o values will give the desired ΔH^o:

$$\Delta H^\circ = (+107.0kJ) + (24.7kJ) + (-236.2kJ) = -104.5\,\mathrm{kJ}$$

Exercise 2.2.2.7

Aluminum chloride can be formed from its elements:

i. $$\mathrm{2\,Al\,(s) + 3\,Cl_2\,(g) \longrightarrow 2\,AlCl_3\,(s)} \quad \Delta H^\circ = ?$$

Use the reactions here to determine the ΔH^o for reaction *(i)*:

ii. $$\mathrm{HCl\,(g) \longrightarrow HCl\,(aq)} \quad \Delta H^\circ_{(ii)} = -74.8\,\mathrm{kJ}$$

iii. $$\mathrm{H_2\,(g) + Cl_2\,(g) \longrightarrow 2\,HCl\,(g)} \quad \Delta H^\circ_{(iii)} = -185\,\mathrm{kJ}$$

iv. $$\mathrm{AlCl_3\,(aq) \longrightarrow AlCl_3\,(s)} \quad \Delta H^\circ_{(iv)} = +323\,\mathrm{kJ}$$

v. $$\mathrm{2\,Al\,(s) + 6\,HCl\,(aq) \longrightarrow 2\,AlCl_3\,(aq) + 3\,H_2\,(g)} \quad \Delta H^\circ_{(v)} = -1049\,\mathrm{kJ}$$

Answer available at the end of this section.

We also can use Hess's law to determine the enthalpy change of any reaction if the corresponding enthalpies of formation of the reactants and products are available. The stepwise reactions we consider are:

i. decompositions of the reactants into their component elements (for which the enthalpy changes are proportional to the negative of the enthalpies of formation of the reactants), followed by
ii. re-combinations of the elements to give the products (with the enthalpy changes proportional to the enthalpies of formation of the products).

The standard enthalpy change of the overall reaction is therefore equal to: (ii) the sum of the standard enthalpies of formation of all the products plus (i) the sum of the negatives of the standard enthalpies of formation of the reactants. This is usually rearranged slightly to be written as follows, with ∑ representing "the sum of" and *n* standing for the stoichiometric coefficients:

$$\Delta H^\circ_{\text{reaction}} = \sum n \times \Delta H^\circ_f(\text{ products }) - \sum n \times \Delta H^\circ_f(\text{ reactants })$$

The following example shows in detail why this equation is valid, and how to use it to calculate the enthalpy change for a reaction of interest.

Example 2.2.2.8: Using Hess's Law

What is the standard enthalpy change for the reaction:

$$3\,NO_2\,(g) + H_2O\,(l) \longrightarrow 2\,HNO_3\,(aq) + NO\,(g) \quad \Delta H^\circ = ?$$

Solution: Using the Equation

Use the special form of Hess's law given previously, and values from Appendix G:

$$\begin{aligned}
\Delta H^\circ_{\text{reaction}} &= \sum n \times \Delta H^\circ_f(\text{ products }) - \sum n \times \Delta H^\circ_f(\text{ reactants }) \\
&= \left[2\ \cancel{\text{mol HNO}_3\,(aq)} \times \frac{-207.4\,\text{kJ}}{1\ \cancel{\text{mol HNO}_3\,(aq)}} + 1\ \cancel{\text{mol NO}\,(g)} \times \frac{+90.2\,\text{kJ}}{1\ \cancel{\text{mol NO}\,(g)}}\right] \\
&\quad - \left[3\ \cancel{\text{mol NO}_2\,(g)} \times \frac{+33.2\,\text{kJ}}{1\ \cancel{\text{mole NO}_2\,(g)}} + 1\ \cancel{\text{mol H}_2\text{O}\,(l)} \times \frac{-285.8\,\text{kJ}}{1\ \cancel{\text{mol H}_2\text{O}\,(l)}}\right] \\
&= [2 \times (-206.64) + 90.25] - [3 \times 33.2 + -(-285.83)] \\
&= -323.03 + 186.23 \\
&= -136.80\,\text{kJ}
\end{aligned}$$

Solution: Supporting Why the General Equation is Valid

Alternatively, we can write this reaction as the sum of the decompositions of $3\,NO_2\,(g)$ and $H2O(l)$ into their constituent elements, and the formation of $2\,HNO_3\,(aq)$ and $1\,NO\,(g)$ from their constituent elements. Writing out these reactions, and noting their relationships to the ΔH° values for these compounds (from Appendix G), we have:

$$\begin{aligned}
3\,NO_2\,(g) &\longrightarrow \tfrac{3}{2}\,N_2\,(g) + 3\,O_2\,(g) && \Delta H^\circ_1 = -99.6\,\text{kJ} \\
H_2O\,(l) &\longrightarrow H_2\,(g) + \tfrac{1}{2}\,O_2\,(g) && \Delta H^\circ_2 = +285.8\,\text{kJ}\ \left[-1 \times \Delta H^\circ_f\,(H_2O)\right] \\
H_2\,(g) + N_2\,(g) + 3\,O_2\,(g) &\longrightarrow 2\,HNO_3\,(aq) && \Delta H^\circ_3 = -414.8\,\text{kJ}\ \left[2 \times \Delta H^\circ_f\,(HNO_3)\right] \\
\tfrac{1}{2}\,N_2\,(g) + \tfrac{1}{2}\,O_2\,(g) &\longrightarrow NO\,(g) && \Delta H^\circ_4 = +90.2\,\text{kJ}\ \left[1 \times \Delta H^\circ_f(NO)\right]
\end{aligned}$$

Summing these reaction equations gives the reaction we are interested in:

$$3\,NO_2\,(g) + H_2O\,(l) \longrightarrow 2\,HNO_3\,(aq) + NO\,(g)$$

Summing their enthalpy changes gives the value we want to determine:

$$\begin{aligned}
\Delta H^\circ_{rxn} &= \Delta H^\circ_1 + \Delta H^\circ_2 + \Delta H^\circ_3 + \Delta H^\circ_4 \\
&= (-99.6\,\text{kJ}) + (+285.8\,\text{kJ}) + (-414.8\,\text{kJ}) + (+90.2\,\text{kJ}) \\
&= -138.4\,\text{kJ}
\end{aligned}$$

So the standard enthalpy change for this reaction is ΔH° = –138.4 kJ.

Note that this result was obtained by:

a. multiplying the of each product by its stoichiometric coefficient and summing those values,
b. multiplying the of each reactant by its stoichiometric coefficient and summing those values, and then
c. subtracting the result found in (2) from the result found in (1).

This is also the procedure in using the general equation, as shown.

Exercise 2.2.2.8

Calculate the heat of combustion of 1 mole of ethanol, $C_2H_5OH\,(l)$, when $H_2O\,(l)$ and $CO_2\,(g)$ are formed. Use the following enthalpies of formation: $C_2H_5OH\,(l)$, –278 kJ/mol; $H_2O\,(l)$, –286 kJ/mol; and $CO_2\,(g)$, –394 kJ/mol.

Answer available at the end of this section.

Footnotes

- 3For more on algal fuel, see http://www.theguardian.com/environme...n-fuel-problem.

Answers

Exercise 2.2.2.1
ΔH = –153 kJ

Exercise 2.2.2.1
ΔH = –338 kJ

Exercise 2.2.2.3
6.25 10^3 kJ

Exercise 2.2.2.4
For the reaction

$$H_2\,(g) + Cl_2\,(g) \longrightarrow 2\,HCl\,(g) \quad \Delta H^\circ = -184.6\,kJ$$

Exercise 2.2.2.5

a. $4\,C\,(s, \text{graphite}) + 5\,H_2\,(g) + \frac{1}{2}\,O_2\,(g) \longrightarrow C_2H_5OC_2H_5\,(l)$

b. $2\,Na\,(s) + C\,(s, \text{graphite}) + \frac{3}{2}\,O_2\,(g) \longrightarrow Na_2CO_3\,(s)$

Exercise 2.2.2.6
66.4 kJ

Exercise 2.2.2.7
–1407 kJ

Exercise 2.2.2.8
–1368 kJ/mol

This page titled 2.2.2: Enthalpy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 5.3: Enthalpy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.2.3: Spontaneity

Learning Objectives

By the end of this section, you will be able to:

- Distinguish between spontaneous and nonspontaneous processes
- Describe the dispersal of matter and energy that accompanies certain spontaneous processes

Processes have a natural tendency to occur in one direction under a given set of conditions. Water will naturally flow downhill, but uphill flow requires outside intervention such as the use of a pump. Iron exposed to the earth's atmosphere will corrode, but rust is not converted to iron without intentional chemical treatment. A **spontaneous process** is one that occurs naturally under certain conditions. A **nonspontaneous process**, on the other hand, will not take place unless it is "driven" by the continual input of energy from an external source. A process that is spontaneous in one direction under a particular set of conditions is nonspontaneous in the reverse direction. At room temperature and typical atmospheric pressure, for example, ice will spontaneously melt, but water will not spontaneously freeze.

The spontaneity of a process is *not* correlated to the speed of the process. A spontaneous change may be so rapid that it is essentially instantaneous or so slow that it cannot be observed over any practical period of time. To illustrate this concept, consider the decay of radioactive isotopes, a topic more thoroughly treated in the chapter on nuclear chemistry. Radioactive decay is by definition a spontaneous process in which the nuclei of unstable isotopes emit radiation as they are converted to more stable nuclei. All the decay processes occur spontaneously, but the rates at which different isotopes decay vary widely. Technetium-99m is a popular radioisotope for medical imaging studies that undergoes relatively rapid decay and exhibits a half-life of about six hours. Uranium-238 is the most abundant isotope of uranium, and its decay occurs much more slowly, exhibiting a half-life of more than four billion years (Figure 2.2.3.1).

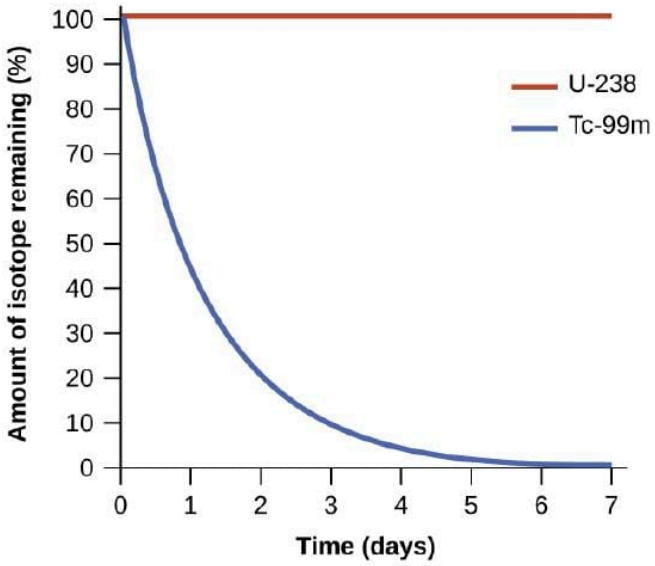


Figure 2.2.3.1: Both U-238 and Tc-99m undergo spontaneous radioactive decay, but at drastically different rates. Over the course of one week, essentially all of a Tc-99m sample and none of a U-238 sample will have decayed.

As another example, consider the conversion of diamond into graphite (Figure 2.2.3.2).

$$C(s, \text{ diamond }) \longrightarrow C(s, \text{ graphite })$$

The phase diagram for carbon indicates that graphite is the stable form of this element under ambient atmospheric pressure, while diamond is the stable allotrope at very high pressures, such as those present during its geologic formation. Thermodynamic calculations of the sort described in the last section of this chapter indicate that the conversion of diamond to graphite at ambient

pressure occurs spontaneously, yet diamonds are observed to exist, and persist, under these conditions. Though the process is spontaneous under typical ambient conditions, its rate is extremely slow; so, for all practical purposes diamonds are indeed “forever.” Situations such as these emphasize the important distinction between the thermodynamic and the kinetic aspects of a process. In this particular case, diamonds are said to be *thermodynamically unstable* but *kinetically stable* under ambient conditions.

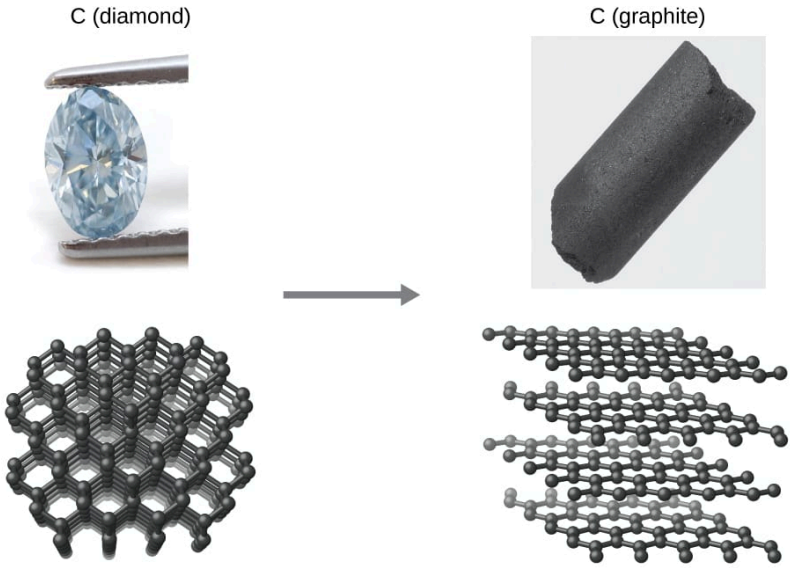


Figure 2.2.3.2: The conversion of carbon from the diamond allotrope to the graphite allotrope is spontaneous at ambient pressure, but its rate is immeasurably slow at low to moderate temperatures. This process is known as *graphitization*, and its rate can be increased to easily measurable values at temperatures in the 1000–2000 K range. (credit "diamond" photo: modification of work by "Fancy Diamonds"/Flickr; credit "graphite" photo: modification of work by images-of-elements.com/carbon.php)

Dispersal of Matter and Energy

Extending the discussion of thermodynamic concepts toward the objective of predicting spontaneity, consider now an isolated system consisting of two flasks connected with a closed valve. Initially there is an ideal gas in one flask and the other flask is empty ($P = 0$) (Figure 2.2.3.3). When the valve is opened, the gas spontaneously expands to fill both flasks equally. Recalling the definition of pressure-volume work from the chapter on thermochemistry, note that no work has been done because the pressure in a vacuum is zero.

\[w=-P \Delta V=0 \quad \quad (P=0 \text { in a vacuum })|]

Note as well that since the system is isolated, no heat has been exchanged with the surroundings (q = 0). The *first law of thermodynamics* confirms that there has been no change in the system’s internal energy as a result of this process.

$$\Delta U = q + w = 0 + 0 = 0$$

The spontaneity of this process is therefore not a consequence of any change in energy that accompanies the process. Instead, the driving force appears to be related to the *greater, more uniform dispersal of matter* that results when the gas is allowed to expand. Initially, the system was comprised of one flask containing matter and another flask containing nothing. After the spontaneous expansion took place, the matter was distributed both more widely (occupying twice its original volume) and more uniformly (present in equal amounts in each flask).

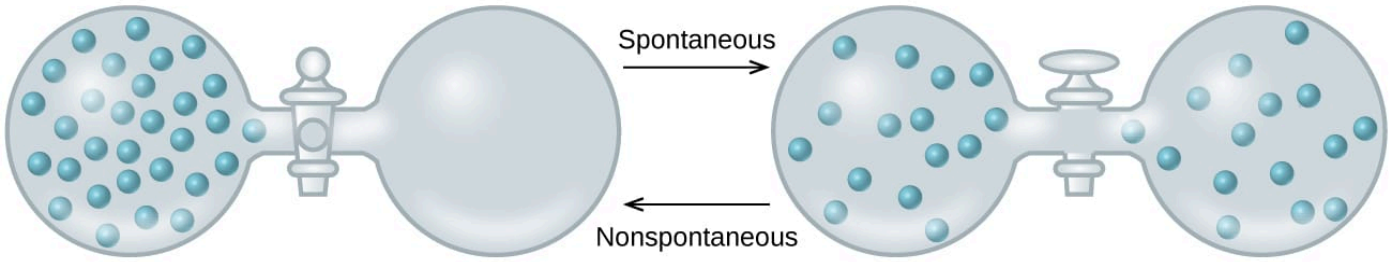


Figure 2.2.3.3: An isolated system consists of an ideal gas in one flask that is connected by a closed valve to a second flask containing a vacuum. Once the valve is opened, the gas spontaneously becomes evenly distributed between the flasks.

Now consider two objects at different temperatures: object X at temperature T_X and object Y at temperature T_Y, with $T_X > T_Y$ (Figure 2.2.3.4). When these objects come into contact, heat spontaneously flows from the hotter object (X) to the colder one (Y). This corresponds to a loss of thermal energy by X and a gain of thermal energy by Y.

$$q_X < 0 \quad \text{and} \quad q_Y = -q_X > 0$$

From the perspective of this two-object system, there was no net gain or loss of thermal energy, rather the available thermal energy was redistributed among the two objects. This spontaneous process resulted in a *more uniform dispersal of energy*.



Figure 2.2.3.4: When two objects at different temperatures come in contact, heat spontaneously flows from the hotter to the colder object.

As illustrated by the two processes described, an important factor in determining the spontaneity of a process is the extent to which it changes the dispersal or distribution of matter and/or energy. In each case, a spontaneous process took place that resulted in a more uniform distribution of matter or energy.

Example 2.2.3.1: Redistribution of Matter during a Spontaneous Process

Describe how matter is redistributed when the following spontaneous processes take place:

a. A solid sublimes.
b. A gas condenses.
c. A drop of food coloring added to a glass of water forms a solution with uniform color.

Solution



Figure 2.2.3.5: (credit a: modification of work by Jenny Downing; credit b: modification of work by "Fuzzy Gerdes"/Flickr; credit c: modification of work by Paul A. Flowers)

a. Sublimation is the conversion of a solid (relatively high density) to a gas (much lesser density). This process yields a much greater dispersal of matter, since the molecules will occupy a much greater volume after the solid-to-gas transition.
b. Condensation is the conversion of a gas (relatively low density) to a liquid (much greater density). This process yields a much lesser dispersal of matter, since the molecules will occupy a much lesser volume after the gas-to-liquid transition.
c. The process in question is diffusion. This process yields a more uniform dispersal of matter, since the initial state of the system involves two regions of different dye concentrations (high in the drop of dye, zero in the water), and the final state of the system contains a single dye concentration throughout.

Exercise 2.2.3.1

Describe how energy is redistributed when a spoon at room temperature is placed in a cup of hot coffee.

Answer available at the end of this section.

Answers

Exercise 2.2.3.1

Heat will spontaneously flow from the hotter object (coffee) to the colder object (spoon), resulting in a more uniform distribution of thermal energy as the spoon warms and the coffee cools.

This page titled 2.2.3: Spontaneity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.1: Spontaneity by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

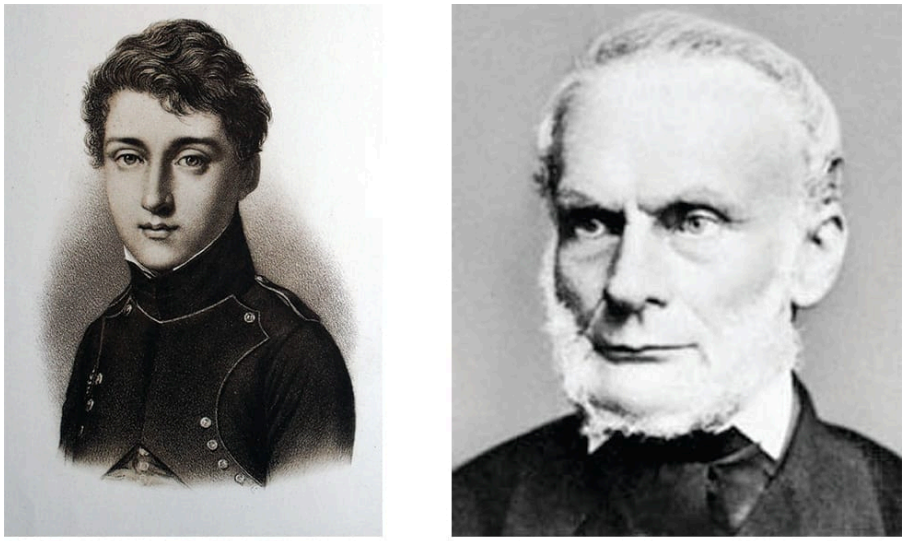
2.2.4: Entropy

Learning Objectives

By the end of this section, you will be able to:

- Define entropy
- Explain the relationship between entropy and the number of microstates
- Predict the sign of the entropy change for chemical and physical processes

In 1824, at the age of 28, Nicolas Léonard Sadi Carnot (Figure 2.2.4.1) published the results of an extensive study regarding the efficiency of steam heat engines. A later review of Carnot's findings by Rudolf Clausius introduced a new thermodynamic property that relates the spontaneous heat flow accompanying a process to the temperature at which the process takes place. This new property was expressed as the ratio of the *reversible* heat (q_{rev}) and the kelvin temperature (T). In thermodynamics, a **reversible process** is one that takes place at such a slow rate that it is always at equilibrium and its direction can be changed (it can be "reversed") by an infinitesimally small change in some condition. Note that the idea of a reversible process is a formalism required to support the development of various thermodynamic concepts; no real processes are truly reversible, rather they are classified as *irreversible*.



(a) (b)

Figure 2.2.4.1: (a) Nicholas Léonard Sadi Carnot's research into steam-powered machinery and (b) Rudolf Clausius's later study of those findings led to groundbreaking discoveries about spontaneous heat flow processes.

Similar to other thermodynamic properties, this new quantity is a state function, so its change depends only upon the initial and final states of a system. In 1865, Clausius named this property entropy (S) and defined its change for any process as the following:

$$\Delta S = \frac{q_{rev}}{T}$$

The entropy change for a real, irreversible process is then equal to that for the theoretical reversible process that involves the same initial and final states.

Entropy and Microstates

Following the work of Carnot and Clausius, Ludwig Boltzmann developed a molecular-scale statistical model that related the entropy of a system to the *number of microstates* (*W*) possible for the system. A **microstate** is a specific configuration of all the locations and energies of the atoms or molecules that make up a system. The relation between a system's entropy and the

number of possible microstates is

$$S = k \ln W$$

where k is the Boltzmann constant, 1.38 10^{-23} J/K.

As for other state functions, the change in entropy for a process is the difference between its final (S_f) and initial (S_i) values:

$$\Delta S = S_f - S_i = k \ln W_f - k \ln W_i = k \ln \frac{W_f}{W_i}$$

For processes involving an increase in the number of microstates, $W_f > W_i$, the entropy of the system increases and $\Delta S > 0$. Conversely, processes that reduce the number of microstates, $W_f < W_i$, yield a decrease in system entropy, $\Delta S < 0$. This molecular-scale interpretation of entropy provides a link to the probability that a process will occur as illustrated in the next paragraphs.

Consider the general case of a system comprised of N particles distributed among n boxes. The number of microstates possible for such a system is n^N. For example, distributing four particles among two boxes will result in 2^4 = 16 different microstates as illustrated in Figure 2.2.4.2. Microstates with equivalent particle arrangements (not considering individual particle identities) are grouped together and are called *distributions*. The probability that a system will exist with its components in a given distribution is proportional to the number of microstates within the distribution. Since entropy increases logarithmically with the number of microstates, *the most probable distribution is therefore the one of greatest entropy*.

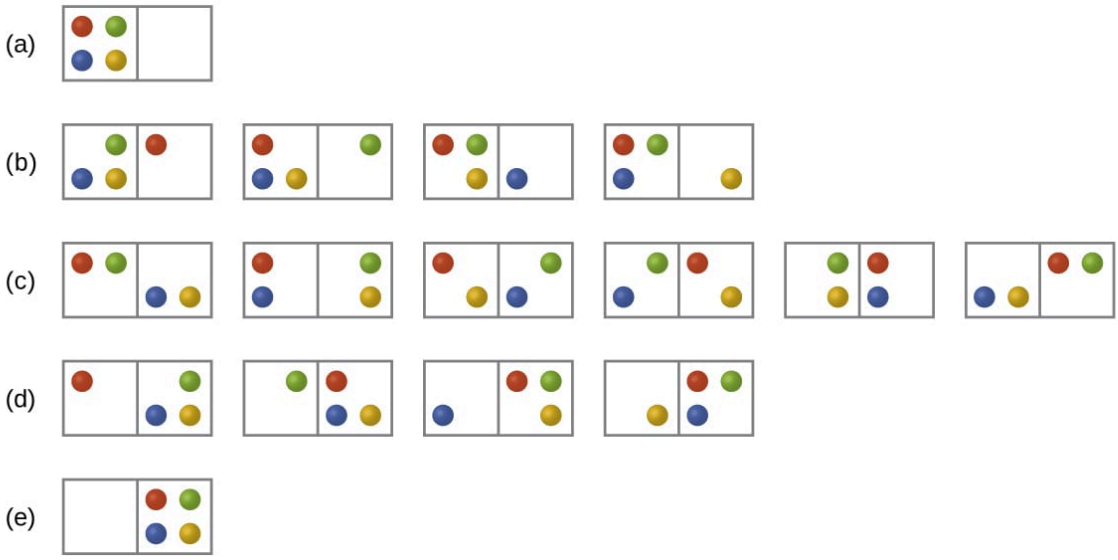


Figure 2.2.4.2: The sixteen microstates associated with placing four particles in two boxes are shown. The microstates are collected into five distributions—(a), (b), (c), (d), and (e)—based on the numbers of particles in each box.

For this system, the most probable configuration is one of the six microstates associated with distribution (c) where the particles are evenly distributed between the boxes, that is, a configuration of two particles in each box. The probability of finding the system in this configuration is $\frac{6}{16}$ or $\frac{3}{8}$ The least probable configuration of the system is one in which all four particles are in one box, corresponding to distributions (a) and (e), each with a probability of The probability of finding all particles in only one box (either the left box or right box) is then or

$$\left(\frac{1}{16} + \frac{1}{16}\right) = \frac{2}{16} \text{ or } \frac{1}{8}$$

As you add more particles to the system, the number of possible microstates increases exponentially (2^N). A macroscopic (laboratory-sized) system would typically consist of moles of particles ($N \sim 10^{23}$), and the corresponding number of microstates would be staggeringly huge. Regardless of the number of particles in the system, however, the distributions in which roughly equal numbers of particles are found in each box are always the most probable configurations.

This matter dispersal model of entropy is often described qualitatively in terms of the *disorder* of the system. By this description,

microstates in which all the particles are in a single box are the most ordered, thus possessing the least entropy. Microstates in which the particles are more evenly distributed among the boxes are more disordered, possessing greater entropy.

The previous description of an ideal gas expanding into a vacuum (Figure 16.4) is a macroscopic example of this particle-in-a-box model. For this system, the most probable distribution is confirmed to be the one in which the matter is most uniformly dispersed or distributed between the two flasks. Initially, the gas molecules are confined to just one of the two flasks. Opening the valve between the flasks increases the volume available to the gas molecules and, correspondingly, the number of microstates possible for the system. Since $W_f > W_i$, the expansion process involves an increase in entropy ($\Delta S > 0$) and is spontaneous.

A similar approach may be used to describe the spontaneous flow of heat. Consider a system consisting of two objects, each containing two particles, and two units of thermal energy (represented as "*") in Figure 2.2.4.3. The hot object is comprised of particles **A** and **B** and initially contains both energy units. The cold object is comprised of particles **C** and **D**, which initially has no energy units. Distribution (a) shows the three microstates possible for the initial state of the system, with both units of energy contained within the hot object. If one of the two energy units is transferred, the result is distribution (b) consisting of four microstates. If both energy units are transferred, the result is distribution (c) consisting of three microstates. Thus, we may describe this system by a total of ten microstates. The probability that the heat does not flow when the two objects are brought into contact, that is, that the system remains in distribution (a), is $\frac{3}{10}$. More likely is the flow of heat to yield one of the other two distribution, the combined probability being $\frac{7}{10}$. The most likely result is the flow of heat to yield the uniform dispersal of energy represented by distribution (b), the probability of this configuration being $\frac{4}{10}$. This supports the common observation that placing hot and cold objects in contact results in spontaneous heat flow that ultimately equalizes the objects' temperatures. And, again, this spontaneous process is also characterized by an increase in system entropy.

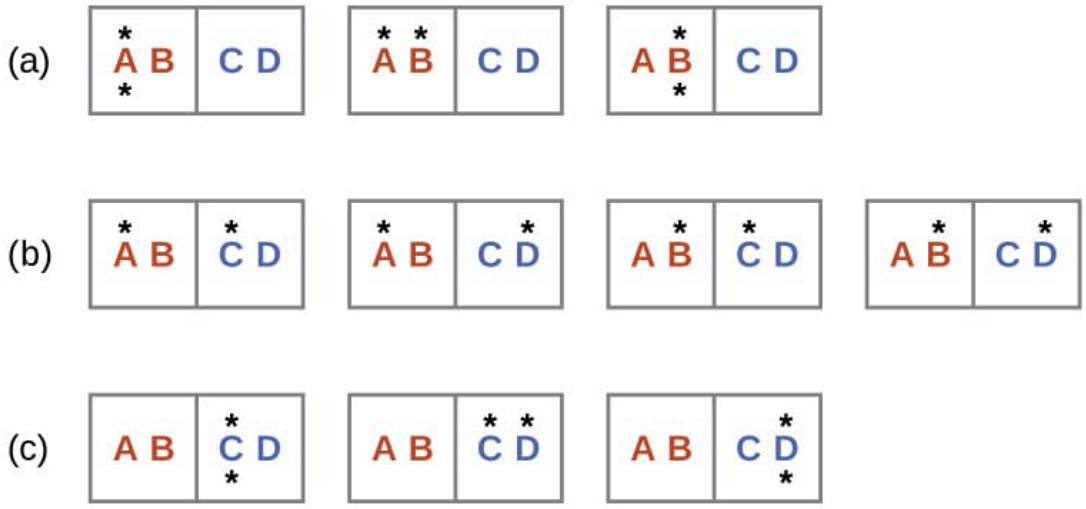
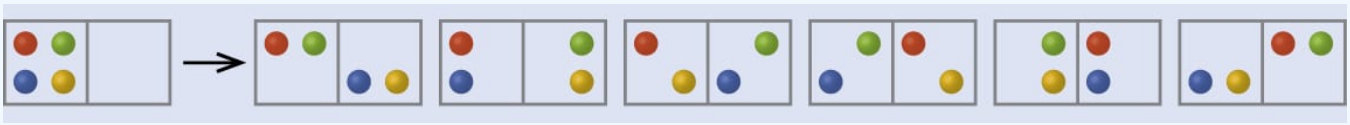


Figure 2.2.4.3: This shows a microstate model describing the flow of heat from a hot object to a cold object. (a) Before the heat flow occurs, the object comprised of particles **A** and **B** contains both units of energy and as represented by a distribution of three microstates. (b) If the heat flow results in an even dispersal of energy (one energy unit transferred), a distribution of four microstates results. (c) If both energy units are transferred, the resulting distribution has three microstates.

Example 2.2.4.1: Determination of ΔS

Calculate the change in entropy for the process depicted below.



Solution

The initial number of microstates is one, the final six:

$$\Delta S = k \ln \frac{W_c}{W_a} = 1.38 \times 10^{-23} J/K \times \ln \frac{6}{1} = 2.47 \times 10^{-23} J/K$$

The sign of this result is consistent with expectation; since there are more microstates possible for the final state than for the initial state, the change in entropy should be positive.

Exercise 2.2.4.1

Consider the system shown in Figure 2.2.4.3. What is the change in entropy for the process where *all* the energy is transferred from the hot object (**AB**) to the cold object (**CD**)?

Answer available at the end of this section.

Predicting the Sign of ΔS

The relationships between entropy, microstates, and matter/energy dispersal described previously allow us to make generalizations regarding the relative entropies of substances and to predict the sign of entropy changes for chemical and physical processes. Consider the phase changes illustrated in Figure 2.2.4.4. In the solid phase, the atoms or molecules are restricted to nearly fixed positions with respect to each other and are capable of only modest oscillations about these positions. With essentially fixed locations for the system's component particles, the number of microstates is relatively small. In the liquid phase, the atoms or molecules are free to move over and around each other, though they remain in relatively close proximity to one another. This increased freedom of motion results in a greater variation in possible particle locations, so the number of microstates is correspondingly greater than for the solid. As a result, $S_{liquid} > S_{solid}$ and the process of converting a substance from solid to liquid (melting) is characterized by an increase in entropy, $\Delta S > 0$. By the same logic, the reciprocal process (freezing) exhibits a decrease in entropy, $\Delta S < 0$.

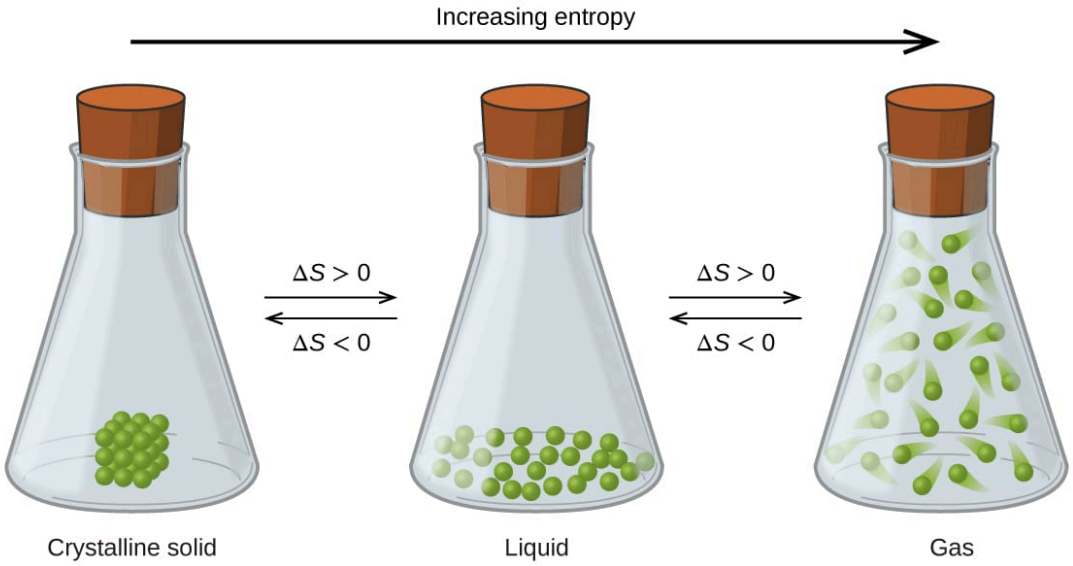


Figure 2.2.4.4: The entropy of a substance increases ($\Delta S > 0$) as it transforms from a relatively ordered solid, to a less-ordered liquid, and then to a still less-ordered gas. The entropy decreases ($\Delta S < 0$) as the substance transforms from a gas to a liquid and then to a solid.

Now consider the gaseous phase, in which a given number of atoms or molecules occupy a *much* greater volume than in the liquid phase. Each atom or molecule can be found in many more locations, corresponding to a much greater number of microstates. Consequently, for any substance, $S_{gas} > S_{liquid} > S_{solid}$, and the processes of vaporization and sublimation likewise involve increases in entropy, $\Delta S > 0$. Likewise, the reciprocal phase transitions, condensation and deposition, involve decreases in entropy, $\Delta S < 0$.

According to kinetic-molecular theory, the temperature of a substance is proportional to the average kinetic energy of its particles. Raising the temperature of a substance will result in more extensive vibrations of the particles in solids and more rapid

translations of the particles in liquids and gases. At higher temperatures, the distribution of kinetic energies among the atoms or molecules of the substance is also broader (more dispersed) than at lower temperatures. Thus, the entropy for any substance increases with temperature (Figure 2.2.4.5).

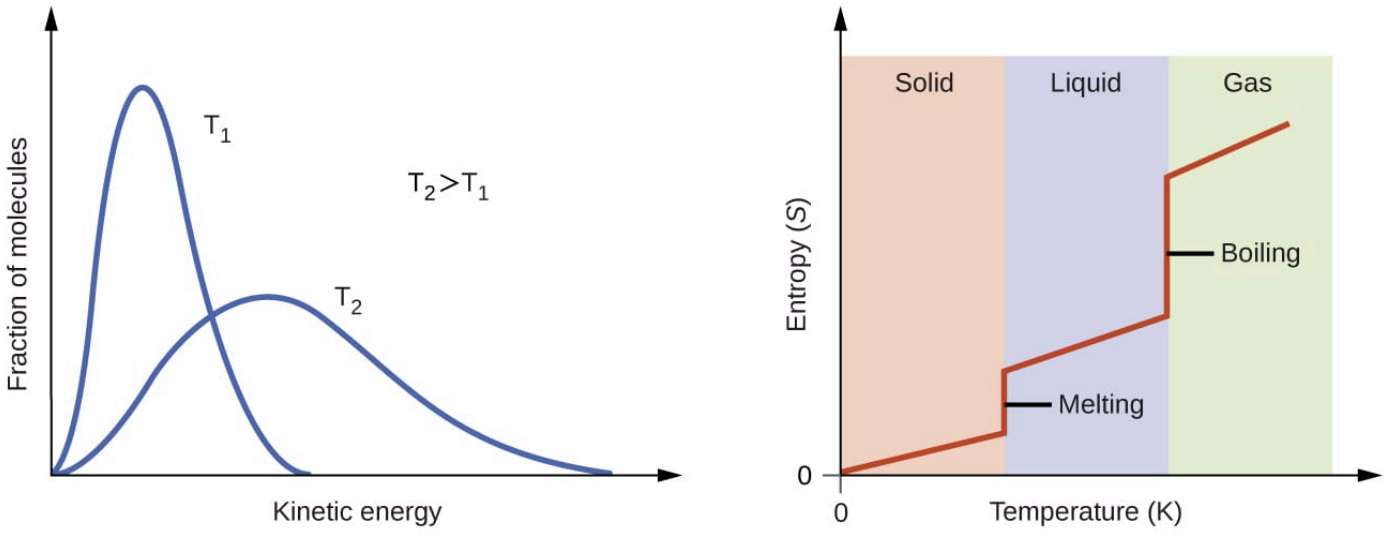


Figure 2.2.4.5: Entropy increases as the temperature of a substance is raised, which corresponds to the greater spread of kinetic energies. When a substance undergoes a phase transition, its entropy changes significantly.

Link to Learning

Try this simulator with interactive visualization of the dependence of particle location and freedom of motion on physical state and temperature.

The entropy of a substance is influenced by the structure of the particles (atoms or molecules) that comprise the substance. With regard to atomic substances, heavier atoms possess greater entropy at a given temperature than lighter atoms, which is a consequence of the relation between a particle's mass and the spacing of quantized translational energy levels (a topic beyond the scope of this text). For molecules, greater numbers of atoms increase the number of ways in which the molecules can vibrate and thus the number of possible microstates and the entropy of the system.

Finally, variations in the types of particles affects the entropy of a system. Compared to a pure substance, in which all particles are identical, the entropy of a mixture of two or more different particle types is greater. This is because of the additional orientations and interactions that are possible in a system comprised of nonidentical components. For example, when a solid dissolves in a liquid, the particles of the solid experience both a greater freedom of motion and additional interactions with the solvent particles. This corresponds to a more uniform dispersal of matter and energy and a greater number of microstates. The process of dissolution therefore involves an increase in entropy, $\Delta S > 0$.

Considering the various factors that affect entropy allows us to make informed predictions of the sign of ΔS for various chemical and physical processes as illustrated in Example 2.2.4.2.

Example 2.2.4.2: Predicting the Sign of ΔS

Predict the sign of the entropy change for the following processes. Indicate the reason for each of your predictions.

a. One mole liquid water at room temperature one mole liquid water at 50 °C
b. $Ag^+(aq) + Cl^-(aq) \longrightarrow AgCl(s)$
c. \(C_6 H_6(l)+\frac{15}{2} O_2(g) \longrightarrow 6 CO_2(g)+3 H_2O(l)
d. $NH_3(s) \longrightarrow NH_3(l)$

Solution

a. positive, temperature increases
b. negative, reduction in the number of ions (particles) in solution, decreased dispersal of matter
c. negative, net decrease in the amount of gaseous species
d. positive, phase transition from solid to liquid, net increase in dispersal of matter

Exercise 2.2.4.1

Predict the sign of the entropy change for the following processes. Give a reason for your prediction.

a. $NaNO_3(s) \longrightarrow Na^+(aq) + NO_3{}^-(aq)$
b. the freezing of liquid water
c. $CO_2(s) \longrightarrow CO_2(g)$
d. $CaCO_3(s) \longrightarrow CaO(s) + CO_2(g)$

Answer available at the end of this section.

Answers

Exercise 2.2.4.1

0 J/K

Exercise 2.2.4.1

(a) Positive; The solid dissolves to give an increase of mobile ions in solution.
(b) Negative; The liquid becomes a more ordered solid.
(c) Positive; The relatively ordered solid becomes a gas.
(d) Positive; There is a net increase in the amount of gaseous species.

This page titled 2.2.4: Entropy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.2: Entropy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.2.5: The Second and Third Laws of Thermodynamics

Learning Objectives

By the end of this section, you will be able to:

- State and explain the second and third laws of thermodynamics
- Calculate entropy changes for phase transitions and chemical reactions under standard conditions

The Second Law of Thermodynamics

In the quest to identify a property that may reliably predict the spontaneity of a process, a promising candidate has been identified: entropy. Processes that involve an increase in entropy *of the system* ($\Delta S > 0$) are very often spontaneous; however, examples to the contrary are plentiful. By expanding consideration of entropy changes to include *the surroundings*, we may reach a significant conclusion regarding the relation between this property and spontaneity. In thermodynamic models, the system and surroundings comprise everything, that is, the universe, and so the following is true:

$$\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}}$$

To illustrate this relation, consider again the process of heat flow between two objects, one identified as the system and the other as the surroundings. There are three possibilities for such a process:

1. The objects are at different temperatures, and heat flows from the hotter to the cooler object. *This is always observed to occur spontaneously.* Designating the hotter object as the system and invoking the definition of entropy yields the following:

   $$\Delta S_{sys} = \frac{-q_{rev}}{T_{sys}} \quad \text{and} \quad \Delta S_{surr} = \frac{q_{rev}}{T_{surr}}$$

   The magnitudes of $-q_{rev}$ and q_{rev} are equal, their opposite arithmetic signs denoting loss of heat by the system and gain of heat by the surroundings. Since $T_{sys} > T_{surr}$ in this scenario, the entropy *decrease* of the system will be less than the entropy *increase* of the surroundings, and so *the entropy of the universe will increase*:

   $$|\Delta S_{\text{sys}}| < |\Delta S_{\text{surr}}|$$
   $$\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}} > 0$$

2. The objects are at different temperatures, and heat flows from the cooler to the hotter object. *This is never observed to occur spontaneously.* Again designating the hotter object as the system and invoking the definition of entropy yields the following:

   $$\Delta S_{sys} = \frac{q_{rev}}{T_{sys}} \quad \text{and} \quad \Delta S_{surr} = \frac{-q_{rev}}{T_{surr}}$$

   The arithmetic signs of q_{rev} denote the gain of heat by the system and the loss of heat by the surroundings. The magnitude of the entropy change for the surroundings will again be greater than that for the system, but in this case, the signs of the heat changes (that is, *the direction of the heat flow*) will yield a negative value for ΔS_{univ}. *This process involves a decrease in the entropy of the universe.*
3. The objects are at essentially the same temperature, $T_{sys} \approx T_{surr}$, and so the magnitudes of the entropy changes are essentially the same for both the system and the surroundings. In this case, the entropy change of the universe is zero, and the system is *at equilibrium*.

   $$|\Delta S_{\text{sys}}| \approx |\Delta S_{\text{surr}}|$$
   $$\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}} = 0$$

These results lead to a profound statement regarding the relation between entropy and spontaneity known as the second law of thermodynamics: *all spontaneous changes cause an increase in the entropy of the universe.* A summary of these three relations

is provided in Table 2.2.5.1.

Table 2.2.5.1: The Second Law of Thermodynamics

$\Delta S_{univ} > 0$	spontaneous
$\Delta S_{univ} < 0$	nonspontaneous (spontaneous in opposite direction)
$\Delta S_{univ} = 0$	at equilibrium

For many realistic applications, the surroundings are vast in comparison to the system. In such cases, the heat gained or lost by the surroundings as a result of some process represents a very small, nearly infinitesimal, fraction of its total thermal energy. For example, combustion of a fuel in air involves transfer of heat from a system (the fuel and oxygen molecules undergoing reaction) to surroundings that are infinitely more massive (the earth's atmosphere). As a result, q_{surr} is a good approximation of q_{rev}, and the second law may be stated as the following:

$$\Delta S_{univ} = \Delta S_{sys} + \Delta S_{surr} = \Delta S_{sys} + \frac{q_{surr}}{T}$$

We may use this equation to predict the spontaneity of a process as illustrated in Example 2.2.5.1.

Example 2.2.5.1: Will Ice Spontaneously Melt?

The entropy change for the process

$$H_2O(s) \longrightarrow H_2O(l)$$

is 22.1 J/K and requires that the surroundings transfer 6.00 kJ of heat to the system. Is the process spontaneous at –10.00 °C? Is it spontaneous at +10.00 °C?

Solution

We can assess the spontaneity of the process by calculating the entropy change of the universe. If ΔS_{univ} is positive, then the process is spontaneous. At both temperatures, ΔS_{sys} = 22.1 J/K and q_{surr} = –6.00 kJ.

At –10.00 °C (263.15 K), the following is true:

$$\begin{aligned}\Delta S_{\text{univ}} &= \Delta S_{\text{sys}} + \Delta S_{\text{surr}} = \Delta S_{\text{sys}} + \frac{q_{\text{surr}}}{T} \\ &= 22.1 J/K + \frac{-6.00 \times 10^3 J}{263.15 K} = -0.7 J/K\end{aligned}$$

$S_{univ} < 0$, so melting is nonspontaneous (*not* spontaneous) at –10.0 °C.

At 10.00 °C (283.15 K), the following is true:

$$\begin{aligned}\Delta S_{\text{univ}} &= \Delta S_{\text{sys}} + \frac{q_{\text{surr}}}{T} \\ &= 22.1 J/K + \frac{-6.00 \times 10^3 J}{283.15 K} = +0.9 J/K\end{aligned}$$

$S_{univ} > 0$, so melting *is* spontaneous at 10.00 °C.

Exercise 2.2.5.1

Using this information, determine if liquid water will spontaneously freeze at the same temperatures. What can you say about the values of S_{univ}?

Answer available at the end of this section.

The Third Law of Thermodynamics

The previous section described the various contributions of matter and energy dispersal that contribute to the entropy of a system. With these contributions in mind, consider the entropy of a pure, perfectly crystalline solid possessing no kinetic energy (that is, at a temperature of absolute zero, 0 K). This system may be described by a single microstate, as its purity, perfect crystallinity and complete lack of motion means there is but one possible location for each identical atom or molecule comprising the crystal (W = 1). According to the Boltzmann equation, the entropy of this system is zero.

$$S = k \ln W = k \ln(1) = 0$$

This limiting condition for a system's entropy represents the third law of thermodynamics: *the entropy of a pure, perfect crystalline substance at 0 K is zero.*

Careful calorimetric measurements can be made to determine the temperature dependence of a substance's entropy and to derive absolute entropy values under specific conditions. **Standard entropies** ($S°$) are for one mole of substance under standard conditions (a pressure of 1 bar and a temperature of 298.15 K; see details regarding standard conditions in the thermochemistry chapter of this text). The **standard entropy change** ($\Delta S°$) for a reaction may be computed using standard entropies as shown below:

$$\Delta S^\circ = \sum \nu S^\circ(\text{ products }) - \sum \nu S^\circ(\text{ reactants })$$

where ν represents stoichiometric coefficients in the balanced equation representing the process. For example, $\Delta S°$ for the following reaction at room temperature

$$mA + nB \longrightarrow xC + yD$$

is computed as:

$$= [xS^\circ(C) + yS^\circ(D)] - [mS^\circ(A) + nS^\circ(B)]$$

A partial listing of standard entropies is provided in Table 2.2.5.2, and additional values are provided in Appendix G. The example exercises that follow demonstrate the use of $S°$ values in calculating standard entropy changes for physical and chemical processes.

Table 2.2.5.2: Standard entropies for selected substances measured at 1 atm and 298.15 K. (Values are approximately equal to those measured at 1 bar, the currently accepted standard state pressure.)

Substance	($J\ mol^{-1}\ K^{-1}$)
carbon	
C(s, graphite)	5.740
C(s, diamond)	2.38
CO(*g*)	197.7
CO_2(*g*)	213.8
CH_4(*g*)	186.3
C_2H_4(*g*)	219.5
C_2H_6(*g*)	229.5
CH_3OH(*l*)	126.8
C_2H_5OH(*l*)	160.7
hydrogen	
H_2(*g*)	130.57
H(*g*)	114.6

Substance	($J\ mol^{-1}\ K^{-1}$)
$H_2O(g)$	188.71
$H_2O(l)$	69.91
HCI(g)	186.8
$H_2S(g)$	205.7
oxygen	
$O_2(g)$	205.03

Example 2.2.5.2: Determination of ΔS°

Calculate the standard entropy change for the following process:

$$H_2O(g) \longrightarrow H_2O(l)$$

Solution

Calculate the entropy change using standard entropies as shown above:

$$\Delta S^\circ = (1mol)\left(70.0Jmol^{-1}K^{-1}\right) - (1mol)\left(188.8Jmol^{-1}K^{-1}\right) = -118.8J/K$$

The value for ΔS° is negative, as expected for this phase transition (condensation), which the previous section discussed.

Exercise 2.2.5.2

Calculate the standard entropy change for the following process:

$$H_2(g) + C_2H_4(g) \longrightarrow C_2H_6(g)$$

Answer available at the end of this section.

Example 2.2.5.3: Determination of ΔS°

Calculate the standard entropy change for the combustion of methanol, CH_3OH:

$$2CH_3OH(l) + 3O_2(g) \longrightarrow 2CO_2(g) + 4H_2O(l)$$

Solution

Calculate the entropy change using standard entropies as shown above:

$$\Delta S^\circ = \sum \nu S^\circ(\text{ products }) - \sum \nu S^\circ(\text{ reactants })$$

$$[2mol \times S^\circ(CO_2(g)) + 4mol \times S^\circ(H_2O(l))] - [2mol \times S^\circ(CH_3OH(l)) + 3mol \times S^\circ(O_2(g))]$$

$$= \{[2(213.8) + 4 \times 70.0] - [2(126.8) + 3(205.2)]\} = -161.6J/K$$

Exercise 2.2.5.3

Calculate the standard entropy change for the following reaction:

$$Ca(OH)_2(s) \longrightarrow CaO(s) + H_2O(l)$$

Answer available at the end of this section.

Answers

Exercise 2.2.5.1

Entropy is a state function, so $\Delta S_{freezing} = -\Delta S_{melting} = -22.1$ J/K and $q_{surr} = +6.00$ kJ. At −10.00 °C spontaneous, +0.7 J/K; at +10.00 °C nonspontaneous, −0.9 J/K.

Exercise 2.2.5.2

$-120.6\ J\ K^{-1}\ mol^{-1}$

Exercise 2.2.5.3

24.7 J/K

This page titled 2.2.5: The Second and Third Laws of Thermodynamics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.3: The Second and Third Laws of Thermodynamics by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.2.6: Free Energy

Learning Objectives

By the end of this section, you will be able to:

- Define Gibbs free energy, and describe its relation to spontaneity
- Calculate free energy change for a process using free energies of formation for its reactants and products
- Calculate free energy change for a process using enthalpies of formation and the entropies for its reactants and products
- Explain how temperature affects the spontaneity of some processes
- Relate standard free energy changes to equilibrium constants

One of the challenges of using the second law of thermodynamics to determine if a process is spontaneous is that it requires measurements of the entropy change for the system *and* the entropy change for the surroundings. An alternative approach involving a new thermodynamic property defined in terms of system properties only was introduced in the late nineteenth century by American mathematician Josiah Willard Gibbs. This new property is called the **Gibbs free energy** (G) (or simply the *free energy*), and it is defined in terms of a system's enthalpy and entropy as the following:

$$G = H - TS$$

Free energy is a state function, and at constant temperature and pressure, the free energy change (ΔG) may be expressed as the following:

$$\Delta G = \Delta H - T\Delta S$$

(For simplicity's sake, the subscript "sys" will be omitted henceforth.)

The relationship between this system property and the spontaneity of a process may be understood by recalling the previously derived second law expression:

$$\Delta S_{\text{univ}} = \Delta S + \frac{q_{\text{surr}}}{T}$$

The first law requires that $q_{surr} = -q_{sys}$, and at constant pressure $q_{sys} = \Delta H$, so this expression may be rewritten as:

$$\Delta S_{\text{univ}} = \Delta S - \frac{\Delta H}{T}$$

Multiplying both sides of this equation by $-T$, and rearranging yields the following:

$$-T\Delta S_{\text{univ}} = \Delta H - T\Delta S$$

Comparing this equation to the previous one for free energy change shows the following relation:

$$\Delta G = -T\Delta S_{\text{univ}}$$

The free energy change is therefore a reliable indicator of the spontaneity of a process, being directly related to the previously identified spontaneity indicator, ΔS_{univ}. Table 2.2.6.1 summarizes the relation between the spontaneity of a process and the arithmetic signs of these indicators.

Table 2.2.6.1: Relation between Process Spontaneity and Signs of Thermodynamic Properties

$\Delta S_{univ} > 0$	$\Delta G <$ 0	moves spontaneously in the forward direction, as written, to reach equilibrium

$\Delta S_{univ} < 0$	$\Delta G > 0$	nonspontaneous in the forward direction, as written, but moves spontaneously in the reverse direction, as written, to reach equilibrium
$\Delta S_{univ} = 0$	$\Delta G = 0$	reversible (at equilibrium)

What's "Free" about ΔG?

In addition to indicating spontaneity, the free energy change also provides information regarding the amount of useful work (w) that may be accomplished by a spontaneous process. Although a rigorous treatment of this subject is beyond the scope of an introductory chemistry text, a brief discussion is helpful for gaining a better perspective on this important thermodynamic property.

For this purpose, consider a spontaneous, exothermic process that involves a decrease in entropy. The free energy, as defined by

$$\Delta G = \Delta H - T\Delta S$$

may be interpreted as representing the difference between the energy produced by the process, Δ*H*, and the energy lost to the surroundings, *T*Δ*S*. The difference between the energy produced and the energy lost is the energy available (or "free") to do useful work by the process, Δ*G*. If the process somehow could be made to take place under conditions of thermodynamic reversibility, the amount of work that could be done would be maximal:

$$\Delta G = w_{max}$$

where w_{max} refers to all types of work except expansion (pressure-volume) work.

However, as noted previously in this chapter, such conditions are not realistic. In addition, the technologies used to extract work from a spontaneous process (e.g., batteries) are never 100% efficient, and so the work done by these processes is always less than the theoretical maximum. Similar reasoning may be applied to a nonspontaneous process, for which the free energy change represents the *minimum* amount of work that must be done *on* the system to carry out the process.

Calculating Free Energy Change

Free energy is a state function, so its value depends only on the conditions of the initial and final states of the system. A convenient and common approach to the calculation of free energy changes for physical and chemical reactions is by use of widely available compilations of standard state thermodynamic data. One method involves the use of standard enthalpies and entropies to compute **standard free energy changes**, ΔG^o, according to the following relation.

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

It is important to understand that for phase changes, ΔG^o tells you if the phase change is spontaneous or not; will it happen, or not happen. For chemical reactions, ΔG^o tells you the **extent** of a reaction. In other words, ΔG^o for a reaction tells you how much product will be present at equilibrium. A reaction with ΔG^o < 0 is considered **product-favored at equilibrium**; there will be more products than reactants when the reaction reaches equilibrium. A reaction with ΔG^o > 0 is considered **reactant-favored at equilibrium**; there will be more reactants than products when the reaction reaches equilibrium.

Example 2.2.6.1: Using Standard Enthalpy and Entropy Changes to Calculate ΔG°

Use standard enthalpy and entropy data from Appendix G to calculate the standard free energy change for the vaporization of water at room temperature (298 K). What does the computed value for ΔG^o say about the spontaneity of this process?

Solution

The process of interest is the following:

$$\mathrm{H_2O\,(l)} \longrightarrow \mathrm{H_2O\,(g)}$$

The standard change in free energy may be calculated using the following equation:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

From Appendix G:

Substance	$\Delta H_f^\circ (\mathrm{kJ/mol})$	$S^\circ (\mathrm{kJ/K\ mol})$
$H_2O(l)$	−285.83	70.0
$H_2O(g)$	−241.82	188.8

Using the appendix data to calculate the standard enthalpy and entropy changes yields:

$$\begin{aligned}
\Delta H^\circ &= \Delta H_f^\circ\,(\mathrm{H_2O\,(g)}) - \Delta H_f^\circ\,(\mathrm{H_2O\,(l)}) \\
&= [-241.82\ \mathrm{kJ/mol} - (-285.83)]\ \mathrm{kJ/mol} = 44.01\ \mathrm{kJ/mol} \\
\Delta S^\circ &= 1\ \mathrm{mol} \times S^\circ\,(\mathrm{H_2O\,(g)}) - 1\ \mathrm{mol} \times S^\circ\,(\mathrm{H_2O\,(l)}) \\
&= (1\ \mathrm{mol})\ 188.8 J/mol \cdot K - (1\ \mathrm{mol})\ 70.0\ \mathrm{J/mol\ K} = 118.8\ \mathrm{J\ /\ K} \\
\Delta G^\circ &= \Delta H^\circ - T\Delta S^\circ
\end{aligned}$$

Substitution into the standard free energy equation yields:

$$\begin{aligned}
\Delta G^\circ &= \Delta H^\circ - T\Delta S^\circ \\
&= 44.01\ \mathrm{kJ} - (298K \times 118.8\ \mathrm{J/K}) \times \frac{1\ \mathrm{kJ}}{1000\ \mathrm{J}} \\
&= 44.01\ \mathrm{kJ} - 35.4\ \mathrm{kJ} \\
&= 8.6\ \mathrm{kJ}
\end{aligned}$$

At 298 K (25 °C) so boiling is nonspontaneous (*not* spontaneous).

Exercise 2.2.6.1

Use standard enthalpy and entropy data from Appendix G to calculate the standard free energy change for the reaction shown here (298 K). What does the computed value for ΔG° say about the spontaneity of this process?

$$\mathrm{C_2H_6\,(g)} \longrightarrow \mathrm{H_2\,(g) + C_2H_4\,(g)}$$

Answer available at the end of this section.

The standard free energy change for a reaction may also be calculated from **standard free energy of formation** ΔG_f^o values of the reactants and products involved in the reaction. The standard free energy of formation is the free energy change that accompanies the formation of one mole of a substance from its elements in their standard states. Similar to the standard enthalpy of formation, ΔG_f^o is by definition zero for elemental substances in their standard states. The approach used to calculate ΔG^o for a reaction from ΔG_f^o values is the same as that demonstrated previously for enthalpy and entropy changes. For the reaction

$$mA + nB \longrightarrow xC + yD$$

the standard free energy change at room temperature may be calculated as

$$\begin{aligned}
\Delta G^\circ &= \sum \nu \Delta G^\circ(\text{ products }) - \sum \nu \Delta G^\circ(\text{ reactants }) \\
&= \left[x\Delta G_f^\circ(C) + y\Delta G_f^\circ(D)\right] - \left[m\Delta G_f^\circ(A) + n\Delta G_f^\circ(B)\right].
\end{aligned}$$

Example 2.2.6.2: Using Standard Free Energies of Formation to Calculate ΔG°

Consider the decomposition of yellow mercury(II) oxide.

$$\mathrm{HgO}(s,\ \text{yellow}\ ) \rightarrow \mathrm{Hg}(l) + \frac{1}{2}\mathrm{O_2}(g)$$

Calculate the standard free energy change at room temperature, using (a) standard free energies of formation and (b) standard enthalpies of formation and standard entropies. Do the results indicate the reaction to be spontaneous or nonspontaneous under standard conditions?

Solution

The required data are available in Appendix G and are shown here.

Compound	ΔG_f°(kJ / mol)	ΔH_f°(kJ / mol)	S°(kJ / K mol)
HgO (s, yellow)	−58.43	−90.46	71.13
Hg(*l*)	0	0	75.9
O_2(*g*)	0	0	205.2

(a) Using free energies of formation:

$$\begin{aligned}\Delta G^\circ &= \sum \nu G_f^\circ(\ \text{products}\ ) - \sum \nu \Delta G_f^\circ(\ \text{reactants}\ ) \\ &= \left[1\Delta G_f^\circ\ \mathrm{Hg\,(l)} + \frac{1}{2}\Delta G_f^\circ\ \mathrm{O_2\,(g)}\right] - 1\Delta G_f^\circ\ \mathrm{HgO\,(s, yellow)} \\ &= \left[1\ \text{mol}(0\ \text{kJ / mol}) + \frac{1}{2}\ \text{mol}(0\ \text{kJ / mol})\right] - 1\ \text{mol}(-58.43\ \text{kJ / mol}) = 58.43\ \text{kJ / mol}\end{aligned}$$

(b) Using enthalpies and entropies of formation:

$$\begin{aligned}\Delta H^\circ &= \sum \nu \Delta H_f^\circ(\ \text{products}\ ) - \sum \nu \Delta H_f^\circ(\ \text{reactants}\ ) \\ &= \left[1\Delta H_f^\circ\ \mathrm{Hg\,(l)} + \frac{1}{2}\Delta H_f^\circ\ \mathrm{O_2\,(g)}\right] - 1\Delta H_f^\circ\ \mathrm{HgO\,(s, yellow)} \\ &= \left[1\ \text{mol}(0\ \text{kJ / mol}) + \frac{1}{2}\ \text{mol}(0\ \text{kJ / mol})\right] - 1\ \text{mol}(-90.46\ \text{kJ / mol}) = 90.46\ \text{kJ / mol} \\ \Delta S^\circ &= \sum \nu \Delta S^\circ(\ \text{products}\ ) - \sum \nu \Delta S^\circ(\ \text{reactants}\ ) \\ &= \left[1\ \Delta S^\circ\ \mathrm{Hg\,(l)} + \frac{1}{2}\Delta S^\circ\ \mathrm{O_2\,(g)}\right] - 1\Delta S^\circ\ \mathrm{HgO\,(s, yellow)} \\ &= \left[1\ \text{mol}(75.9\ \text{J / mol K}) + \frac{1}{2}\ \text{mol}(205.2\ \text{J / mol K})\right] - 1\text{mol}(71.13\ \text{J / mol K}) = 107.4\ \ \text{J / mol K}\end{aligned}$$

$$\begin{aligned}\Delta G^\circ &= \Delta H^\circ - T\Delta S^\circ = 90.46\ \text{kJ} - 298.15\ \text{K} \times 107.4\ \text{J / K} \cdot \text{mol} \times \frac{1\ \ \text{kJ}}{1000\ \text{J}} \\ &= (90.46 - 32.01)\ \text{kJ / mol} = 58.45\ \text{kJ / mol}\end{aligned}$$

Both ways to calculate the standard free energy change at 25 °C give the same numerical value (to three significant figures), and both predict that the process is nonspontaneous (*not* spontaneous) at room temperature.

Exercise 2.2.6.1

Calculate ΔG° using (a) free energies of formation and (b) enthalpies of formation and entropies (Appendix G). Do the results indicate the reaction to be spontaneous or nonspontaneous at 25 °C?

$$C_2H_4\,(g) \longrightarrow H_2\,(g) + C_2H_2\,(g)$$

Answer available at the end of this section.

Free Energy Changes for Coupled Reactions

The use of free energies of formation to compute free energy changes for reactions as described above is possible because ΔG is a state function, and the approach is analogous to the use of Hess' Law in computing enthalpy changes (see the chapter on thermochemistry). Consider the vaporization of water as an example:

$$H_2O\,(l) \longrightarrow H_2O\,(g)$$

An equation representing this process may be derived by adding the formation reactions for the two phases of water (necessarily reversing the reaction for the liquid phase). The free energy change for the sum reaction is the sum of free energy changes for the two added reactions:

$$\begin{aligned} H_2\,(g) + \tfrac{1}{2}\,O_2\,(g) &\longrightarrow H_2O\,(g) && \Delta G_f^\circ \text{ gas} \\ H_2O\,(l) &\longrightarrow H_2\,(g) + \tfrac{1}{2}\,O_2\,(g) && -\Delta G_f^o \text{ liquid} \\ H_2O\,(l) &\longrightarrow H_2O\,(g) && \Delta G^\circ = \Delta G_f^\circ \text{ gas} - \Delta G_f^\circ \text{ liquid} \end{aligned}$$

This approach may also be used in cases where a nonspontaneous reaction is enabled by coupling it to a spontaneous reaction. For example, the production of elemental zinc from zinc sulfide is thermodynamically unfavorable, as indicated by a positive value for ΔG^o:

$$ZnS\,(s) \rightarrow Zn\,(s) + S\,(s) \quad \Delta G_1^\circ = 201.3kJ$$

The industrial process for production of zinc from sulfidic ores involves coupling this decomposition reaction to the thermodynamically favorable oxidation of sulfur:

$$S\,(s) + O_2\,(g) \rightarrow SO_2\,(g) \quad \Delta G_2^\circ = -300.1kJ$$

The coupled reaction exhibits a negative free energy change and is spontaneous:

$$ZnS\,(s) + O_2\,(g) \rightarrow Zn\,(s) + SO_2\,(g)$$

$$\Delta G^\circ = 201.3\text{ kJ} + -300.1\text{ kJ} = -98.8\text{ kJ}$$

This process is typically carried out at elevated temperatures, so this result obtained using standard free energy values is just an estimate. The gist of the calculation, however, holds true.

Example 2.2.6.3: Calculating Free Energy Change for a Coupled Reaction

Is a reaction coupling the decomposition of ZnS to the formation of H_2S expected to be spontaneous under standard conditions?

Solution

Following the approach outlined above and using free energy values from Appendix G:

Decomposition of zinc sulfide:	$ZnS\,(s) \rightarrow Zn\,(s) + S\,(s)$	$\Delta G_1^\circ = 201.3\text{ kJ}$
Formation of hydrogen sulfide:	$S\,(s) + H_2\,(g) \rightarrow H_2S\,(g)$	$\Delta G_1^\circ = -33.4\text{ kJ}$
Coupled reaction:	$ZnS\,(s) + H_2\,(g) \rightarrow Zn\,(s) + H_2S\,(g)$	$\Delta G^\circ = 201.3\text{ kJ} + -33.4\text{ kJ} = 167.9\text{ kJ}$

The coupled reaction exhibits a positive free energy change and is thus nonspontaneous.

Exercise 2.2.6.3

What is the standard free energy change for the reaction below? Is the reaction expected to be spontaneous under standard conditions?

$$\mathrm{FeS(s)} + \mathrm{O_2(g)} \rightarrow \mathrm{Fe(s)} + \mathrm{SO_2(g)}$$

Answer available at the end of this section.

Temperature Dependence of Spontaneity and Extent of Reaction

As was previously demonstrated in this chapter's section on entropy, the spontaneity of a process may depend upon the temperature of the system. Phase transitions, for example, will proceed spontaneously in one direction or the other depending upon the temperature of the substance in question. In a similar, but not identical fashion, some chemical reactions can switch from being product-favored at equilibrium, to being reactant-favored at equilibrium, depending on the temperature.

Note

The numerical value of ΔG° is always dependent on the temperature. In this section we are determining whether or not the **sign** of ΔG° is dependent on the temperature.

To illustrate this concept, the equation relating free energy change to the enthalpy and entropy changes for the process is considered:

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}$$

The extent of a process, as reflected in the arithmetic sign of its standard free energy change, is then determined by the signs of the enthalpy and entropy changes and, in some cases, the absolute temperature. Since T is the absolute (Kelvin) temperature, it can only have positive values. Four possibilities therefore exist with regard to the signs of the enthalpy and entropy changes:

1. **Both Δ*H*° and Δ*S*° are positive.** This condition describes an endothermic process that involves an increase in system entropy. In this case, Δ*G*° will be negative if the magnitude of the *T*Δ*S*° term is greater than Δ*H*°. If the *T*Δ*S*° term is less than Δ*H*°, the free energy change will be positive. Such a process is *product-favored at equilibrium at high temperatures and reactant-favored at equilibrium at low temperatures*.
2. **Both Δ*H*° and Δ*S*° are negative.** This condition describes an exothermic process that involves a decrease in system entropy. In this case, Δ*G*° will be negative if the magnitude of the *T*Δ*S*° term is less than Δ*H*°. If the *T*Δ*S*° term's magnitude is greater than Δ*H*°, the free energy change will be positive. Such a process is *product-favored at equilibrium at low temperatures* and *reactant-favored at equilibrium at high temperatures*.
3. **Δ*H*° is positive and Δ*S*° is negative.** This condition describes an endothermic process that involves a decrease in system entropy. In this case, Δ*G*° will be positive regardless of the temperature. Such a process is *reactant-favored at equilibrium at all temperatures*.
4. **Δ*H*° is negative and Δ*S*° is positive.** This condition describes an exothermic process that involves an increase in system entropy. In this case, Δ*G*° will be negative regardless of the temperature. Such a process is *product-favored at equilibrium at all temperatures*.

These four scenarios are summarized in Table 2.2.6.1

Table 2.2.6.1

Sign of ΔH^o	Sign of ΔS^o	Sign of ΔG^o	Temperature Dependence of ΔG^o
-	+	-	The sign of ΔG^o does not depend on the temperature.The reaction is product-favored at equilibrium at all temperatures.
+	-	+	The sign of ΔG^o does not depend on the temperature.The reaction is reactant-favored at equilibrium at all temperatures.
-	-	- or +	The sign of ΔG^o **does** depend on the temperature. The reaction will be product-favored at equilibrium at lower temperatures.
+	+	- or +	The sign of ΔG^o **does** depend on the temperature. The reaction will be product-favored at equilibrium at lower temperatures.

Example 2.2.6.3: Predicting the Temperature Dependence of Spontaneity

The incomplete combustion of carbon is described by the following equation:

$$2\,\mathrm{C}(s)+\mathrm{O}_2(g) \longrightarrow 2\,\mathrm{CO}(g)$$

Does the sign of ΔG° of this process depend upon temperature?

Solution

Combustion processes are exothermic ($\Delta H^{\circ} < 0$). This particular reaction involves an increase in entropy due to the accompanying increase in the amount of gaseous species (net gain of one mole of gas, $\Delta S^{\circ} > 0$). The reaction is therefore product-favored at equilibrium ($\Delta G^{\circ} < 0$) at all temperatures.

Exercise 2.2.6.3

Popular chemical hand warmers generate heat by the air-oxidation of iron:

$$4\,\mathrm{Fe}(s)+3\,\mathrm{O}_2(g) \longrightarrow 2\,\mathrm{Fe}_2\mathrm{O}_3(s)$$

Does the sign of ΔG^o of this process depend upon temperature?

Answer available at the end of this section.

When considering the conclusions drawn regarding the temperature dependence of the sign of $\Delta G°$, it is important to keep in mind what the terms “high” and “low” mean. Since these terms are adjectives, the temperatures in question are deemed high or low relative to some reference temperature. A process that is reactant-favored at equilibrium at one temperature but product-favored at equilibrium at another temperature will necessarily undergo a change in “extent” (as reflected by its $\Delta G°$) as temperature varies. This is clearly illustrated by a graphical presentation of the free energy change equation, in which $\Delta G°$ is plotted on the *y* axis versus *T* on the *x* axis:

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ}$$

$$y = b + mx$$

Such a plot is shown in Figure 2.2.6.2. A process whose enthalpy and entropy changes are of the same arithmetic sign will exhibit a temperature-dependence for the sign of $\Delta G°$ as depicted by the two yellow lines in the plot. Each line crosses from one spontaneity domain (positive or negative $\Delta G°$) to the other at a temperature that is characteristic of the process in question. This

temperature is represented by the *x*-intercept of the line, that is, the value of *T* for which ΔG° is zero:

$$\Delta G^\circ = 0 = \Delta H^\circ - T\Delta S^\circ$$

$$T = \frac{\Delta H^\circ}{\Delta S^\circ}$$

Thus, saying a process is product-favored at equilibrium at “high” or “low” temperatures means the temperature is above or below, respectively, that temperature at which ΔG° for the process is zero.

Note

In this discussion, we have used two different descriptions for the meaning of the sign of ΔG°. You should be aware of the meaning of each description.

a) **Extent of Reaction:** This description is used to predict the ratio of the product and reactant concentrations at equilibrium. In this description, we use the thermodynamic term ΔG° to tell us the same information as the equilibrium constant, K. When $\Delta G^\circ < 0$, $K > 1$, and the reaction will be product-favored at equilibrium. When $\Delta G^\circ > 0$, $K < 1$, and the reaction is reactant-favored at equilibrium. When $\Delta G^\circ = 0$, $K = 1$, and the reaction will have roughly equal amounts of products and reactants at equilibrium. In all cases, the reaction will form a mixture of products and reactants at equilibrium. We use the sign and magnitude of ΔG° to tell us how much product will be made if the reaction is allowed to reach equilibrium.

b) **Spontaneity:** This description is much more complicated because it involves two different interpretations of how a reaction at standard state occurs. One interpretation involves the hypothetical process in which the reaction proceeds from a starting point of pure reactants to a finishing point of pure products, with all substances isolated in their own containers under standard state conditions. In the second, more realistic interpretation, the reaction starts with all reactants and all products in their standard state in one container. We then allow this specific mixture to react an infinitesimally small amount so that we can obtain a rate of change in free energy with respect to the extent of reaction when all reactants and products are mixed and (essentially) in their standard states. Although each interpretation describes a different reaction scenario, the value of the difference in free energy and the value of the rate of change in free energy are the same number. If $\Delta G^\circ < 0$, we say that the reaction is spontaneous, meaning that the reaction would proceed in the forward direction, as written, to form pure products in their standard state. If $\Delta G^\circ > 0$, we say that the reaction is nonspontaneous, meaning that the reaction would proceed in the reverse direction, as written, to form pure reactants in their standard state. If $\Delta G^\circ = 0$, we say that the neither the reactants nor the products are favored to be formed.

A detailed treatment of the meaning of ΔG° can be found in the paper, "Free Energy versus Extent of Reaction" by Richard S. Treptow, Journal of Chemical Education, **1996**, Volume 73 (1), 51-54.

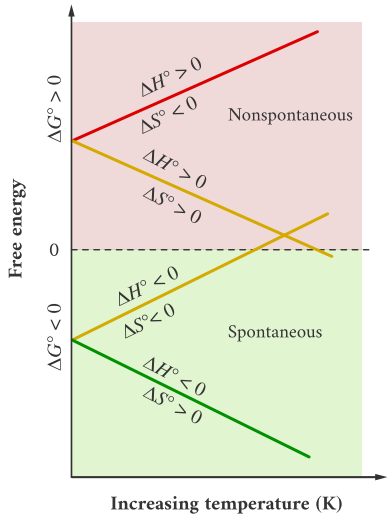


Figure 2.2.6.2: These plots show the variation in ΔG° with temperature for the four possible combinations of arithmetic sign for ΔH° and ΔS°. Note that in this graph of ΔG°, "spontaneous" is synonymous with "product-favored at equilibrium", and "nonspontaneous" is synonymous with "reactant-favored at equilibrium".

Example 2.2.6.4: Equilibrium Temperature for a Phase Transition

As defined in the chapter on liquids and solids, the boiling point of a liquid is the temperature at which its solid and liquid phases are in equilibrium (that is, when vaporization and condensation occur at equal rates). Use the information in Tables T1 or T2 to estimate the boiling point of water.

Solution

The process of interest is the following phase change:

$$\mathrm{H_2O}(l) \longrightarrow \mathrm{H_2O}(g)$$

When this process is at equilibrium, $\Delta G = 0$, so the following is true:

$$0 = \Delta H^\circ - T\Delta S^\circ \qquad \text{or} \qquad T = \frac{\Delta H^\circ}{\Delta S^\circ}$$

Using the standard thermodynamic data from Tables T1 or T2,

$$\begin{aligned}\Delta H^\circ &= \Delta H_\mathrm{f}^\circ(\mathrm{H_2O}(g)) - \Delta H_\mathrm{f}^\circ(\mathrm{H_2O}(l)) \\ &= -241.82\ \mathrm{kJ/mol} - (-285.83\ \mathrm{kJ/mol}) = 44.01\ \mathrm{kJ/mol}\end{aligned}$$

$$\begin{aligned}\Delta S^\circ &= \Delta S_{298}^\circ(\mathrm{H_2O}(g)) - \Delta S_{298}^\circ(\mathrm{H_2O}(l)) \\ &= 188.8\ \mathrm{J/K \cdot mol} - 70.0\ \mathrm{J/K \cdot mol} = 118.8\ \mathrm{J/K \cdot mol}\end{aligned}$$

$$T = \frac{\Delta H^\circ}{\Delta S^\circ} = \frac{44.01 \times 10^3\ \mathrm{J/mol}}{118.8\ \mathrm{J/K \cdot mol}} = 370.5\ \mathrm{K} = 97.3\ ^\circ\mathrm{C}$$

The accepted value for water's normal boiling point is 373.2 K (100.0 °C), and so this calculation is in reasonable agreement. Note that the values for enthalpy and entropy changes data used were derived from standard data at 298 K (Tables T1 or T2.). If desired, you could obtain more accurate results by using enthalpy and entropy changes determined at (or at least closer to) the actual boiling point.

Exercise 2.2.6.4

Use the information in Tables T1 or T2 to estimate the boiling point of CS_2.

Answer available at the end of this section.

Free Energy and Equilibrium

The free energy change for a process may be viewed as a measure of its driving force. A negative value for ΔG represents a driving force for the process in the forward direction, while a positive value represents a driving force for the process in the reverse direction. When ΔG is zero, the forward and reverse driving forces are equal, and the process occurs in both directions at the same rate (the system is at equilibrium).

In the chapter on equilibrium the *reaction quotient*, Q, was introduced as a convenient measure of the status of an equilibrium system. Recall that Q is the numerical value of the mass action expression for the system, and that you may use its value to identify the direction in which a reaction will proceed in order to achieve equilibrium. When Q is lesser than the equilibrium constant, *K*, the reaction will proceed in the forward direction until equilibrium is reached and $Q = K$. Conversely, if $Q > K$, the process will proceed in the reverse direction until equilibrium is achieved.

The free energy change for a process taking place with reactants and products present under *nonstandard conditions* (pressures other than 1 bar; concentrations other than 1 M) is related to the standard free energy change according to this equation:

$$\Delta G = \Delta G^\circ + RT \ln Q$$

R is the gas constant (8.314 J/K mol), *T* is the kelvin or absolute temperature, and Q is the reaction quotient. For gas phase equilibria, the pressure-based reaction quotient, Q_P, is used. The concentration-based reaction quotient, Q_C, is used for condensed phase equilibria. This equation may be used to predict the spontaneity for a process under any given set of conditions as illustrated in Example 2.2.6.6.

Example 2.2.6.6: Calculating ΔG under Nonstandard Conditions

What is the free energy change for the process shown here

$$2NH_3(g) \longrightarrow 3H_2(g) + N_2(g)$$

under the specified conditions?

- $T = 25^\circ C$,
- $P_{N_2} = 0.870$ atm
- $P_{H_2} = 0.250$ atm
- $P_{NH_3} = 12.9$ atm

$$\Delta G^\circ = 33.0 kJ/mol$$

Solution

The equation relating free energy change to standard free energy change and reaction quotient may be used directly:

$$\begin{aligned}\Delta G &= \Delta G^\circ + RT \ln Q \\ &= 33.0 \frac{\text{kJ}}{\text{mol}} + \left(8.314 \frac{\text{J}}{\text{mol K}} \times 298 K \times \ln \frac{\left(0.250^3\right) \times 0.870}{12.9^2}\right) \\ &= 9680 \frac{\text{J}}{\text{mol}} \text{ or } 9.68 \text{ kJ/mol}\end{aligned}$$

Since the computed value for ΔG is positive, the reaction is nonspontaneous under these conditions.

Exercise 2.2.6.6

Calculate the free energy change for this same reaction at 875 °C in a 5.00 L mixture containing 0.100 mol of each gas. Is the reaction spontaneous under these conditions?

Answer available at the end of this section.

For a system at equilibrium, Q = *K* and Δ*G* = 0, and the previous equation may be written as

$$0 = \Delta G^\circ + RT \ln K \quad \text{(at equilibrium)}$$

$$\Delta G^\circ = -RT \ln K$$

or

$$K = e^{-\frac{\Delta G}{RT}}$$

This form of the equation provides a useful link between these two essential thermodynamic properties, and it can be used to derive equilibrium constants from standard free energy changes and vice versa. The relations between standard free energy changes and equilibrium constants are summarized in Table 2.2.6.2.

Table 2.2.6.2: Relations between Standard Free Energy Changes and Equilibrium Constants

K	Δ*G*°	Composition of an Equilibrium Mixture
> 1	< 0	Products are more abundant
< 1	> 0	Reactants are more abundant
= 1	= 0	Reactants and products are comparably abundant

Example 2.2.6.7: Calculating an Equilibrium Constant using Standard Free Energy Change

Given that the standard free energies of formation of $Ag^+(aq)$, $Cl^-(aq)$, and AgCl(s) are 77.1 kJ/mol, –131.2 kJ/mol, and –109.8 kJ/mol, respectively, calculate the solubility product, K_{sp}, for AgCl.

Solution

The reaction of interest is the following:

$$AgCl(s) \rightleftharpoons Ag^+(aq) + Cl^-(aq) \quad K_{sp} = [Ag^+]\,[Cl^-]$$

The standard free energy change for this reaction is first computed using standard free energies of formation for its reactants and products:

$$\begin{aligned}\Delta G^\circ &= \left[\Delta G_f^\circ\left(\mathrm{Ag^+\,(aq)}\right) + \Delta G_f^\circ\left(\mathrm{Cl^-\,(aq)}\right)\right] - \left[\Delta G_f^\circ(\mathrm{AgCl\,(s)})\right] \\ &= [77.1\ \mathrm{kJ/mol} - 131.2\ \mathrm{kJ/mol}] - [-109.8\ \mathrm{kJ/mol}] = 55.7\ \mathrm{kJ/mol}\end{aligned}$$

The equilibrium constant for the reaction may then be derived from its standard free energy change:

$$K_{sp} = e^{-\frac{\Delta G^\circ}{RT}} = \exp\left(-\frac{\Delta G^\circ}{RT}\right) = \exp\left(-\frac{55.7 \times 10^3 J/mol}{8.314 J/mol \cdot K \times 298.15 K}\right) = \exp(-22.470) = e^{-22.470} = 1.74 \times 10^{-10}$$

This result is in reasonable agreement with the value provided in Appendix J.

Exercise 2.2.6.7

Use the thermodynamic data provided in Appendix G to calculate the equilibrium constant for the dissociation of dinitrogen tetroxide at 25 °C.

$$2NO_2(g) \rightleftharpoons N_2O_4(g)$$

Answer available at the end of this section.

To further illustrate the relation between these two essential thermodynamic concepts, consider the observation that reactions spontaneously proceed in a direction that ultimately establishes equilibrium. As may be shown by plotting the free energy versus the extent of the reaction (for example, as reflected in the value of *Q*), equilibrium is established when the system's free energy is minimized (Figure 2.2.6.3). If a system consists of reactants and products in nonequilibrium amounts ($Q \neq K$), the reaction will proceed spontaneously in the direction necessary to establish equilibrium.

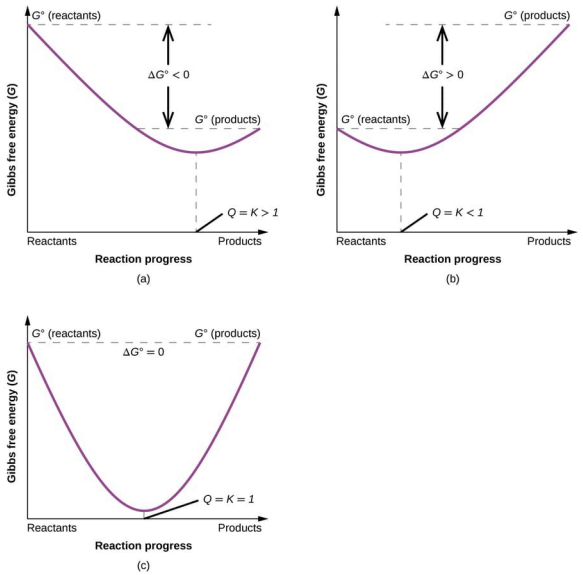


Figure 2.2.6.3: These plots show the free energy versus reaction progress for systems whose standard free energy changes are (a) negative, (b) positive, and (c) zero. Nonequilibrium systems will proceed spontaneously in whatever direction is necessary to minimize free energy and establish equilibrium.

Answers

Exercise 2.2.6.1

the reaction is nonspontaneous (*not* spontaneous) at 25 °C.

Exercise 2.2.6.1

a. 140.8 kJ/mol, nonspontaneous
b. 141.5 kJ/mol, nonspontaneous

Exercise 2.2.6.3

–199.7 kJ; spontaneous

Exercise 2.2.6.3

Δ*H*° and Δ*S*° are both negative; the reaction is product-favored at equilibrium at low temperatures.

Exercise 2.2.6.4

313 K (accepted value 319 K).

Exercise 2.2.6.6

Δ*G* = –123.5 kJ/mol; yes

Exercise 2.2.6.7

K = 3.1

This page titled 2.2.6: Free Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.4: Free Energy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.
- 16.4: Gibbs Energy by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.3: KINETICS

2.3.1: Chemical Reaction Rates
2.3.2: Factors Affecting Reaction Rates
2.3.3: Rate Laws
2.3.4: Integrated Rate Laws
2.3.5: Collision Theory
2.3.6: Reaction Mechanisms
2.3.7: Catalysis

This page titled 2.3: Kinetics is shared under a license and was authored, remixed, and/or curated by LibreTexts.

2.3.1: Chemical Reaction Rates

12.01: Chemical Reaction Rates

This page titled 2.3.1: Chemical Reaction Rates is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

2.3.2: Factors Affecting Reaction Rates

Learning Objectives

By the end of this section, you will be able to:

- Describe the effects of chemical nature, physical state, temperature, concentration, and catalysis on reaction rates

The rates at which reactants are consumed and products are formed during chemical reactions vary greatly. Five factors typically affecting the rates of chemical reactions will be explored in this section: the chemical nature of the reacting substances, the state of subdivision (one large lump versus many small particles) of the reactants, the temperature of the reactants, the concentration of the reactants, and the presence of a catalyst.

The Chemical Nature of the Reacting Substances

The rate of a reaction depends on the nature of the participating substances. Reactions that appear similar may have different rates under the same conditions, depending on the identity of the reactants. For example, when small pieces of the metals iron and sodium are exposed to air, the sodium reacts completely with air overnight, whereas the iron is barely affected. The active metals calcium and sodium both react with water to form hydrogen gas and a base. Yet calcium reacts at a moderate rate, whereas sodium reacts so rapidly that the reaction is almost explosive.

The Physical States of the Reactants

A chemical reaction between two or more substances requires intimate contact between the reactants. When reactants are in different physical states, or phases (solid, liquid, gaseous, dissolved), the reaction takes place only at the interface between the phases. Consider the heterogeneous reaction between a solid phase and either a liquid or gaseous phase. Compared with the reaction rate for large solid particles, the rate for smaller particles will be greater because the surface area in contact with the other reactant phase is greater. For example, large pieces of iron react more slowly with acids than they do with finely divided iron powder (Figure 2.3.2.1). Large pieces of wood smolder, smaller pieces burn rapidly, and saw dust burns explosively.

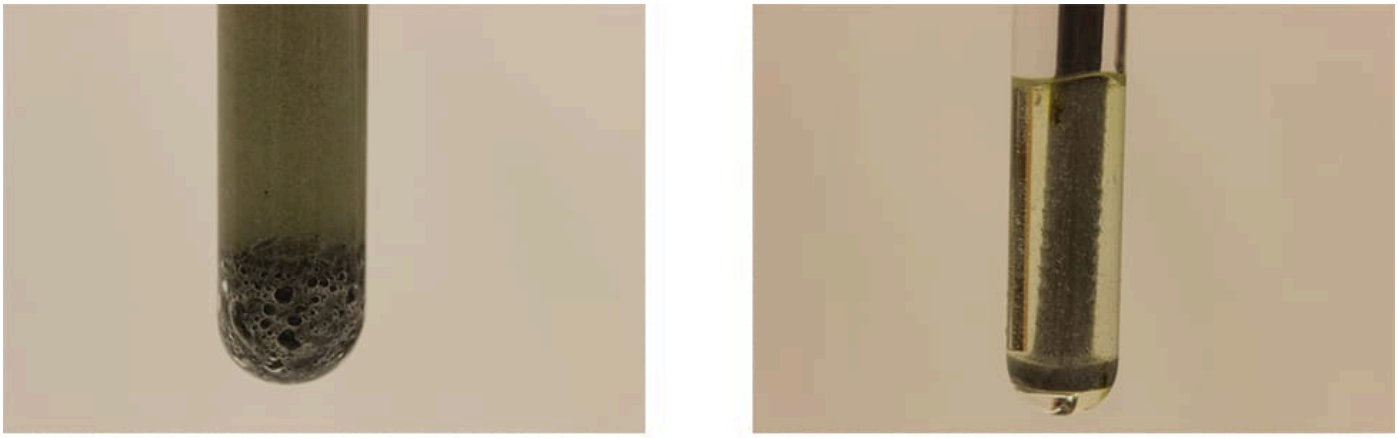


Figure 2.3.2.1: (a) Iron powder reacts rapidly with dilute hydrochloric acid and produces bubbles of hydrogen gas: $2Fe(s) + 6HCl(aq)\ 2FeCl_3(aq) + 3H_2(g)$. (b) An iron nail reacts more slowly because the surface area exposed to the acid is much less.

Link to Learning

Watch this video to see the reaction of cesium with water in slow motion and a discussion of how the state of reactants and particle size affect reaction rates.

Temperature of the Reactants

Chemical reactions typically occur faster at higher temperatures. Food can spoil quickly when left on the kitchen counter. However, the lower temperature inside of a refrigerator slows that process so that the same food remains fresh for days. Gas burners, hot plates, and ovens are often used in the laboratory to increase the speed of reactions that proceed slowly at ordinary temperatures. For many chemical processes, reaction rates are approximately doubled when the temperature is raised by 10 °C.

Concentrations of the Reactants

The rates of many reactions depend on the concentrations of the reactants. Rates usually increase when the concentration of one or more of the reactants increases. For example, calcium carbonate ($CaCO_3$) deteriorates as a result of its reaction with the pollutant sulfur dioxide. The rate of this reaction depends on the amount of sulfur dioxide in the air (Figure 2.3.2.2). An acidic oxide, sulfur dioxide combines with water vapor in the air to produce sulfurous acid in the following reaction:

$$\mathrm{SO_2\,(g) + H_2O\,(g) \longrightarrow H_2SO_3\,(aq)}$$

Calcium carbonate reacts with sulfurous acid as follows:

$$\mathrm{CaCO_3\,(s) + H_2SO_3\,(aq) \longrightarrow CaSO_3\,(aq) + CO_2\,(g) + H_2O\,(l)}$$

In a polluted atmosphere where the concentration of sulfur dioxide is high, calcium carbonate deteriorates more rapidly than in less polluted air. Similarly, phosphorus burns much more rapidly in an atmosphere of pure oxygen than in air, which is only about 20% oxygen.

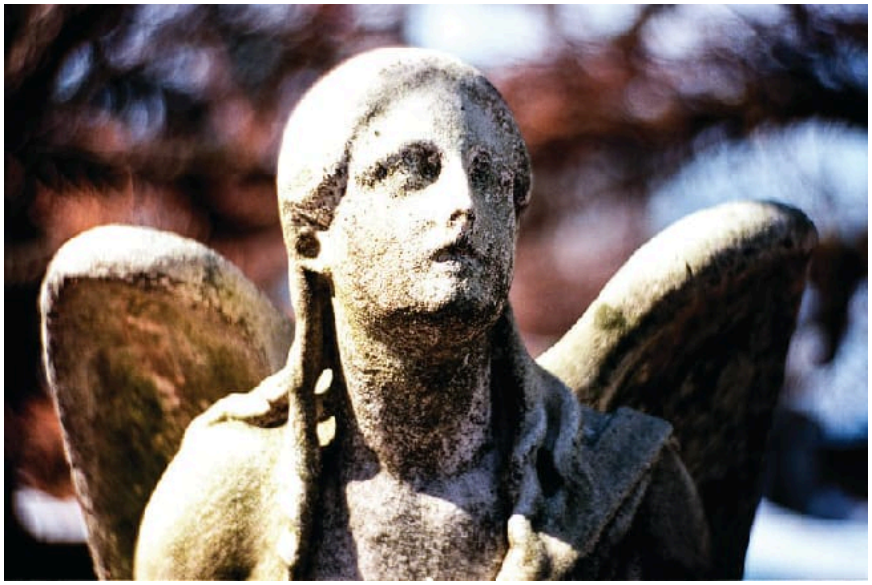


Figure 2.3.2.2: Statues made from carbonate compounds such as limestone and marble typically weather slowly over time due to the actions of water, and thermal expansion and contraction. However, pollutants like sulfur dioxide can accelerate weathering. As the concentration of air pollutants increases, deterioration of limestone occurs more rapidly. (credit: James P Fisher III)

Link to Learning

Phosphorus burns rapidly in air, but it will burn even more rapidly if the concentration of oxygen is higher. Watch this video to see an example.

The Presence of a Catalyst

Relatively dilute aqueous solutions of hydrogen peroxide, $\mathrm{H_2O_2}$, are commonly used as topical antiseptics. Hydrogen peroxide decomposes to yield water and oxygen gas according to the equation:

$$2\,\mathrm{H_2O_2\,(aq)} \longrightarrow 2\,\mathrm{H_2O\,(l)} + \mathrm{O_2\,(g)}$$

Under typical conditions, this decomposition occurs very slowly. When dilute $\mathrm{H_2O_2\,(aq)}$ is poured onto an open wound, however, the reaction occurs rapidly and the solution foams because of the vigorous production of oxygen gas. This dramatic difference is caused by the presence of substances within the wound's exposed tissues that accelerate the decomposition process. Substances that function to increase the rate of a reaction are called **catalysts**, a topic treated in greater detail later in this chapter.

Link to Learning

Chemical reactions occur when molecules collide with each other and undergo a chemical transformation. Before physically performing a reaction in a laboratory, scientists can use molecular modeling simulations to predict how the parameters discussed earlier will influence the rate of a reaction. Use the PhET Reactions & Rates interactive to explore how temperature, concentration, and the nature of the reactants affect reaction rates.

This page titled 2.3.2: Factors Affecting Reaction Rates is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.2: Factors Affecting Reaction Rates by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.3.3: Rate Laws

Learning Objectives

By the end of this section, you will be able to:

- Explain the form and function of a rate law
- Use rate laws to calculate reaction rates
- Use rate and concentration data to identify reaction orders and derive rate laws

As described in the previous module, the rate of a reaction is often affected by the concentrations of reactants. Rate laws (sometimes called *differential rate laws*) or rate equations are mathematical expressions that describe the relationship between the rate of a chemical reaction and the concentration of its reactants. As an example, consider the reaction described by the chemical equation

$$aA + bB \longrightarrow \text{ products}$$

where a and b are stoichiometric coefficients. The rate law for this reaction is written as:

$$\text{rate } = k[A]^m[B]^n$$

in which [*A*] and [*B*] represent the molar concentrations of reactants, and *k* is the **rate constant**, which is specific for a particular reaction at a particular temperature. The exponents *m* and *n* are the **reaction orders** and are typically positive integers, though they can be fractions, negative, or zero. The rate constant *k* and the reaction orders *m* and *n* must be determined experimentally by observing how the rate of a reaction changes as the concentrations of the reactants are changed. The rate constant *k* is independent of the reactant concentrations, but it does vary with temperature.

The reaction orders in a rate law describe the mathematical dependence of the rate on reactant concentrations. Referring to the generic rate law above, the reaction is *m* order with respect to *A* and *n* order with respect to *B*. For example, if *m* = 1 and *n* = 2, the reaction is first order in *A* and second order in *B*. The overall reaction order is simply the sum of orders for each reactant. For the example rate law here, the reaction is third order overall (1 + 2 = 3). A few specific examples are shown below to further illustrate this concept.

The rate law:

$$\text{rate } = k\left[H_2O_2\right]$$

describes a reaction that is first order in hydrogen peroxide and first order overall. The rate law:

$$\text{rate } = k\left[\mathrm{C_4H_6}\right]^2$$

describes a reaction that is second order in $\mathrm{C_4H_6}$ and second order overall. The rate law:

$$\text{rate } = k\left[\mathrm{H^+}\right]\left[\mathrm{OH^-}\right]$$

describes a reaction that is first order in $\mathrm{H^+}$, first order in $\mathrm{OH^-}$, and second order overall.

Example 2.3.3.1: Writing Rate Laws from Reaction Orders

An experiment shows that the reaction of nitrogen dioxide with carbon monoxide:

$$\mathrm{NO_2\,(g) + CO\,(g) \longrightarrow NO\,(g) + CO_2\,(g)}$$

is second order in $\mathrm{NO_2}$ and zero order in CO at 100 °C. What is the rate law for the reaction?

Solution

The reaction will have the form:

$$\text{rate } = k\left[\text{NO}_2\right]^m\left[\text{CO}\right]^n$$

The reaction is second order in NO_2; thus $m = 2$. The reaction is zero order in CO; thus $n = 0$. The rate law is:

$$\text{rate } = k\left[\text{NO}_2\right]^2\left[\text{CO}\right]^0 = k\left[\text{NO}_2\right]^2$$

Remember that a number raised to the zero power is equal to 1, thus $[CO]^0 = 1$, which is why the CO concentration term may be omitted from the rate law: the rate of reaction is solely dependent on the concentration of NO_2. A later chapter section on reaction mechanisms will explain how a reactant's concentration can have no effect on a reaction rate despite being involved in the reaction.

Exercise 2.3.3.1A

The rate law for the reaction:

$$\text{H}_2\left(\text{g}\right) + 2\,\text{NO}\left(\text{g}\right) \longrightarrow \text{N}_2\text{O}\left(\text{g}\right) + \text{H}_2\text{O}\left(\text{g}\right)$$

has been determined to be rate = $k[NO]^2[H_2]$. What are the orders with respect to each reactant, and what is the overall order of the reaction?

Answer available at the end of this section.

Exercise 2.3.3.1B

In a transesterification reaction, a triglyceride reacts with an alcohol to form an ester and glycerol. Many students learn about the reaction between methanol (CH_3OH) and ethyl acetate ($\text{CH}_3\text{CH}_2\text{OCOCH}_3$) as a sample reaction before studying the chemical reactions that produce biodiesel:

$$\text{CH}_3\text{OH} + \text{CH}_3\text{CH}_2\text{OCOCH}_3 \longrightarrow \text{CH}_3\text{OCOCH}_3 + \text{CH}_3\text{CH}_2\text{OH}$$

The rate law for the reaction between methanol and ethyl acetate is, under certain conditions, determined to be:

$$\text{rate } = k\left[\text{CH}_3\text{OH}\right]$$

What is the order of reaction with respect to methanol and ethyl acetate, and what is the overall order of reaction?

Answer available at the end of this section.

A common experimental approach to the determination of rate laws is the **method of initial rates**. This method involves measuring reaction rates for multiple experimental trials carried out using different initial reactant concentrations. Comparing the measured rates for these trials permits determination of the reaction orders and, subsequently, the rate constant, which together are used to formulate a rate law. This approach is illustrated in the next two example exercises.

Example 2.3.3.2: Determining a Rate Law from Initial Rates

Ozone in the upper atmosphere is depleted when it reacts with nitrogen oxides. The rates of the reactions of nitrogen oxides with ozone are important factors in deciding how significant these reactions are in the formation of the ozone hole over Antarctica (Figure 2.3.3.1). One such reaction is the combination of nitric oxide, NO, with ozone, O_3:

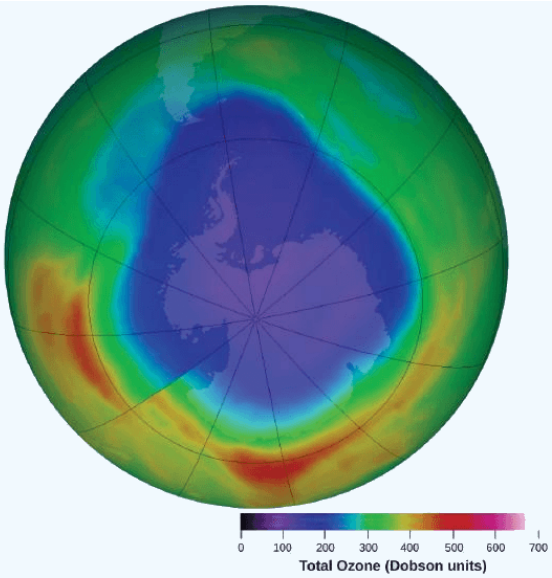


Figure 2.3.3.1: A contour map showing stratospheric ozone concentration and the “ozone hole” that occurs over Antarctica during its spring months. (credit: modification of work by NASA)

$$\mathrm{NO\,(g)} + \mathrm{O_3\,(g)} \longrightarrow \mathrm{NO_2\,(g)} + \mathrm{O_2\,(g)}$$

This reaction has been studied in the laboratory, and the following rate data were determined at 25 °C.

Trial	[NO] (mol/L)	[O_3] (mol/L)	$\frac{\Delta[\mathrm{NO_2}]}{\Delta t}(\mathrm{mol\,L^{-1}\,s^{-1}})$
1	1.00×10^{-6}	3.00×10^{-6}	6.60×10^{-5}
2	1.00×10^{-6}	6.00×10^{-6}	1.32×10^{-4}
3	1.00×10^{-6}	9.00×10^{-6}	1.98×10^{-4}
4	2.00×10^{-6}	9.00×10^{-6}	3.96×10^{-4}
5	3.00×10^{-6}	9.00×10^{-6}	5.94×10^{-4}

Determine the rate law and the rate constant for the reaction at 25 °C.

Solution

The rate law will have the form:

$$\text{rate } = k[NO]^m\,[O_3]^n$$

Determine the values of *m*, *n*, and *k* from the experimental data using the following three-part process:

Step 1:

Determine the value of m *from the data in which [NO] varies and [O_3] is constant.* In the last three experiments, [NO] varies while [O_3] remains constant. When [NO] doubles from trial 3 to 4, the rate doubles, and when [NO] triples from trial 3 to 5, the rate also triples. Thus, the rate is also directly proportional to [NO], and *m* in the rate law is equal to 1.

Step 2:

Determine the value of n *from data in which [O_3] varies and [NO] is constant.* In the first three experiments, [NO] is constant and [O_3] varies. The reaction rate changes in direct proportion to the change in [O_3]. When [O_3] doubles from trial 1 to 2, the rate doubles; when [O_3] triples from trial 1 to 3, the rate increases also triples. Thus, the rate is directly proportional to [O_3], and *n* is equal to 1.The rate law is thus:

$$\text{rate } = k[NO]^1\,[O_3]^1 = k[NO]\,[O_3]$$

Step 3:

Determine the value of k *from one set of concentrations and the corresponding rate*. The data from trial 1 are used below:

$$\begin{aligned} k &= \frac{\text{rate}}{[\text{NO}]\,[\text{O}_3]} \\ &= \frac{6.60 \times 10^{-5}\ \text{mol L}^{-1}\text{s}^{-1}}{\left(1.00 \times 10^{-6}\ \text{mol L}^{-1}\right)\left(3.00 \times 10^{-6}\ \text{mol L}^{-1}\right)} \\ &= 2.20 \times 10^{7}\ \text{L mol}^{-1}\text{s}^{-1} \end{aligned}$$

Exercise 2.3.3.2

Acetaldehyde decomposes when heated to yield methane and carbon monoxide according to the equation:

$$\text{CH}_3\text{CHO}\,(\text{g}) \longrightarrow \text{CH}_4\,(\text{g}) + \text{CO}\,(\text{g})$$

Determine the rate law and the rate constant for the reaction from the following experimental data:

Trial	[CH_3CHO] (mol/L)	$-\frac{\Delta[\text{CH}_3\text{CHO}]}{\Delta t}$ (mol $L^{-1}s^{-1}$)
1	1.75×10^{-3}	2.06×10^{-11}
2	3.50×10^{-3}	8.24×10^{-11}
3	7.00×10^{-3}	3.30×10^{-10}

Answer available at the end of this section.

Example 2.3.3.3: Determining Rate Laws from Initial Rates

Using the initial rates method and the experimental data, determine the rate law and the value of the rate constant for this reaction:

$$2NO(g) + Cl_2(g) \longrightarrow 2NOCl(g)$$

Trial	[NO] (mol/L)	[Cl_2] (mol/L)	$-\frac{\Delta[\text{CO}]}{\Delta t}$ (mol $L^{-1}s^{-1}$)
1	0.10	0.10	0.00300
2	0.10	0.15	0.00450
3	0.15	0.10	0.00675

Solution

The rate law for this reaction will have the form:

$$\text{rate } = k[NO]^m\,[Cl_2]^n$$

As in Example 2.3.3.2, approach this problem in a stepwise fashion, determining the values of *m* and *n* from the experimental data and then using these values to determine the value of *k*. In this example, however, an explicit algebraic approach (vs. the implicit approach of the previous example) will be used to determine the values of *m* and *n*:

Step 1

Determine the value of m from the data in which $[\mathrm{NO}]$ varies and $[\mathrm{Cl_2}]$ is constant. Write the ratios with the subscripts x and y to indicate data from two different trials:

$$\frac{\text{rate }_x}{\text{rate }_y} = \frac{k[NO]_x^m\,[Cl_2]_x^n}{k[NO]_y^m\,[Cl_2]_y^n}$$

Using the third trial and the first trial, in which $[\mathrm{Cl_2}]$ does not vary, gives:

$$\frac{\text{rate } 3}{\text{rate } 1} = \frac{0.00675}{0.00300} = \frac{k(0.15)^m(0.10)^n}{k(0.10)^m(0.10)^n}$$

Canceling equivalent terms in the numerator and denominator leaves:

$$\frac{0.00675}{0.00300} = \frac{(0.15)^m}{(0.10)^m}$$

which simplifies to:

$$2.25 = (1.5)^m$$

Use logarithms to determine the value of the exponent *m*:

$$\ln(2.25) = m\ln(1.5)$$
$$\frac{\ln(2.25)}{\ln(1.5)} = m$$
$$2 = m$$

Confirm the result

$$1.5^2 = 2.25$$

Step 2

Determine the value of n from data in which $[\mathrm{Cl_2}]$ varies and $[\mathrm{NO}]$ is constant.

$$\frac{\text{rate } 2}{\text{rate } 1} = \frac{0.00450}{0.00300} = \frac{k0.10)^m(0.15)^n}{k(0.10)^m(0.10)^n}$$

Cancelation gives:

$$\frac{0.0045}{0.0030} = \frac{(0.15)^n}{(0.10)^n}$$

which simplifies to:

$$1.5 = (1.5)^n$$

Thus n must be 1, and the form of the rate law is:

$$\text{rate } = k[NO]^m\,[Cl_2]^n = k[NO]^2\,[Cl_2]$$

Step 3.

Determine the numerical value of the constant with appropriate units. The units for the rate of a reaction are mol/L/s. The units for are whatever is needed so that substituting into the rate law expression affords the appropriate units for the rate. In this example, the concentration units are mol^3/L^3. The units for *k* should be mol^{-} L/s so that the rate is in terms of mol/L/s.

To determine the value of *k* once the rate law expression has been solved, simply plug in values from the first experimental trial and solve for *k*:

$$0.00300 \text{ mol } L^{-1}s^{-1} = k\left(0.10 \text{ mol } L^{-1}\right)^2\left(0.10 \text{ mol } L^{-1}\right)^1$$
$$k = 3.0 \text{ mol}^{-2}L^2s^{-1}$$

Exercise 2.3.3.3

Use the provided initial rate data to derive the rate law for the reaction whose equation is:

$$OCl^-(aq) + I^-(aq) \longrightarrow OI^-(aq) + Cl^-(aq)$$

Trial	[OCl^-] (mol/L)	[I^-] (mol/L)	Initial Rate (mol/L/s)
1	0.0040	0.0020	0.00184
2	0.0020	0.0040	0.00092
3	0.0020	0.0020	0.00046

Determine the rate law expression and the value of the rate constant *k* with appropriate units for this reaction.

Answer available at the end of this section.

Reaction Order and Rate Constant Units

In some of our examples, the reaction orders in the rate law happen to be the same as the coefficients in the chemical equation for the reaction. This is merely a coincidence and very often not the case.

Rate laws may exhibit fractional orders for some reactants, and negative reaction orders are sometimes observed when an increase in the concentration of one reactant causes a decrease in reaction rate. A few examples illustrating these points are provided:

$$\begin{aligned}
NO_2 + CO &\longrightarrow NO + CO_2 & \text{rate} &= k\left[NO_2\right]^2\\
CH_3\ CHO &\longrightarrow CH_4 + CO & \text{rate} &= k\left[CH_3CHO\right]^2\\
2\,N_2\ O_5 &\longrightarrow NO_2 + O_2 & \text{rate} &= k\left[N_2O_5\right]\\
2\,NO_2 + F_2 &\longrightarrow 2\,NO_2\ F & \text{rate} &= k\left[NO_2\right]\left[F_2\right]\\
2\,NO_2\ Cl &\longrightarrow 2\,NO_2 + Cl_2 & \text{rate} &= k\left[NO_2Cl\right]
\end{aligned}$$

Note

It is important to note that *rate laws are determined by experiment only and are not reliably predicted by reaction stoichiometry.*

The units for a rate constant will vary as appropriate to accommodate the overall order of the reaction. The unit of the rate constant for the second-order reaction described in Example 2.3.3.2 was determined to be k was derived to be $L\ mol^{-1}s^{-1}$. For the third-order reaction, k was derived to be $L^2mol^{-2}s^{-1}$. Dimensional analysis requires the rate constant unit for a reaction whose overall order is x to be $L^{x-1}\ mol^{1-x}s^{-1}$. Table 2.3.3.1 summarizes the rate constant units for common reaction orders.

Table 2.3.3.1: Rate Constant Units for Common Reaction Orders

Overall Reaction Order (x)	Rate Constant Unit (L^{x-1} mol^{1-x} s^{-1})
0 (zero)	mol L^{-1} s^{-1}
1 (first)	s^{-1}
2 (second)	L mol^{-1} s^{-1}
3 (third)	L^2 mol^{-2} s^{-1}

Note that the units in this table were derived using specific units for concentration (mol/L) and time (s), though any valid units for these two properties may be used.

Answers

Exercise 2.3.3.1*A*

order in NO = 2; order in H_2 = 1; overall order = 3

Exercise 2.3.3.1*B*

order in CH_3OH = 1; order in $CH_3CH_2OCOCH_3$ = 0; overall order = 1

Exercise 2.3.3.2

$\text{rate} = k\left[CH_3CHO\right]^2$ with k = 6.7310^{-6} L/mol/s

Exercise 2.3.3.3

$$\frac{\text{rate 2}}{\text{rate 3}} = \frac{0.00092}{0.00046} = \frac{k(0.0020)^x(0.0040)^y}{k(0.0020)^x(0.0020)^y}$$

$$2.00 = 2.00^y$$

$$y = 1$$

$$\frac{\text{rate 1}}{\text{rate 2}} = \frac{0.00184}{0.00092} = \frac{k(0.0040)^x(0.0020)^y}{k(0.0020)^x(0.0040)^y}$$

$$2.00 = \frac{2^x}{2^y}$$

$$2.00 = \frac{2^x}{2^1}$$

$$4.00 = 2^x$$

$$x = 2$$

Substituting the concentration data from trial 1 and solving for k yields:

$$\text{rate} = k\left[OCl^-\right]^2\left[I^-\right]^1$$

$$0.00184 = k(0.0040)^2(0.0020)^1$$

$$k = 5.75 \times 10^4 mol^{-2}L^2s^{-1}$$

This page titled 2.3.3: Rate Laws is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.3: Rate Laws by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.3.4: Integrated Rate Laws

Learning Objectives

By the end of this section, you will be able to:

- Explain the form and function of an integrated rate law
- Perform integrated rate law calculations for zero-, first-, and second-order reactions
- Define half-life and carry out related calculations
- Identify the order of a reaction from concentration/time data

The rate laws discussed thus far relate the rate and the concentrations of reactants. We can also determine a second form of each rate law that relates the concentrations of reactants and time. These are called **integrated rate laws**. We can use an integrated rate law to determine the amount of reactant or product present after a period of time or to estimate the time required for a reaction to proceed to a certain extent. For example, an integrated rate law is used to determine the length of time a radioactive material must be stored for its radioactivity to decay to a safe level.

Using calculus, the differential rate law for a chemical reaction can be integrated with respect to time to give an equation that relates the amount of reactant or product present in a reaction mixture to the elapsed time of the reaction. This process can either be very straightforward or very complex, depending on the complexity of the differential rate law. For purposes of discussion, we will focus on the resulting integrated rate laws for first-, second-, and zero-order reactions.

First-Order Reactions

Integration of the rate law for a simple first-order reaction (rate = $k[A]$) results in an equation describing how the reactant concentration varies with time:

$$[A]_t = [A]_0 e^{-kt}$$

where $[A]t$ is the concentration of A at any time t, $[A]_0$ is the initial concentration of A, and k is the first-order rate constant.

For mathematical convenience, this equation may be rearranged to other formats, including direct and indirect proportionalities:

$$\ln\left(\frac{[A]_t}{[A]_0}\right) = -kt$$

or

$$\ln\left(\frac{[A]_0}{[A]_t}\right) = kt$$

and a format showing a linear dependence of concentration in time:

$$\ln [A]_t = \ln [A]_0 - kt$$

Example 2.3.4.1: The Integrated Rate Law for a First-Order Reaction

The rate constant for the first-order decomposition of cyclobutane, C_4H_8 at 500 °C is 9.2 10^{-3} s^{-1}:

$$C_4H_8 \longrightarrow 2\,C_2H_4$$

How long will it take for 80.0% of a sample of C_4H_8 to decompose?

Solution

Since the relative change in reactant concentration is provided, a convenient format for the integrated rate law is:

$$\ln\left(\frac{[A]_0}{[A]_t}\right) = kt$$

The initial concentration of C_4H_8, $[A]_0$, is not provided, but the provision that 80.0% of the sample has decomposed is enough information to solve this problem. Let *x* be the initial concentration, in which case the concentration after 80.0% decomposition is 20.0% of *x* or 0.200*x*. Rearranging the rate law to isolate *t* and substituting the provided quantities yields:

$$\begin{aligned} t &= \ln\frac{[x]}{[0.200x]} \times \frac{1}{k} \\ &= \ln 5 \times \frac{1}{9.2 \times 10^{-3}\ \text{s}^{-1}} \\ &= 1.609 \times \frac{1}{9.2 \times 10^{-3}\ \text{s}^{-1}} \\ &= 1.7 \times 10^{2}\ \text{s} \end{aligned}$$

Exercise 2.3.4.1

Iodine-131 is a radioactive isotope that is used to diagnose and treat some forms of thyroid cancer. Iodine-131 decays to xenon-131 according to the equation:

$$\text{I-131} \longrightarrow \text{Xe-131} + \text{electron}$$

The decay is first-order with a rate constant of 0.138 d^{-1}. How many days will it take for 90% of the iodine–131 in a 0.500 *M* solution of this substance to decay to Xe-131?

Answer available at the end of this section.

In the next example exercise, a linear format for the integrated rate law will be convenient:

$$\begin{aligned} \ln[A]_t &= (-k)(t) + \ln[A]_0 \\ y &= mx + b \end{aligned}$$

A plot of $\ln[A]_t$ versus *t* for a first-order reaction is a straight line with a slope of –*k* and a *y*-intercept of $\ln[A]_0$. If a set of rate data are plotted in this fashion but do *not* result in a straight line, the reaction is not first order in *A*.

Example 2.3.4.2: Graphical Determination of Reaction Order and Rate Constant

Show that the data in Figure 2.3.4.1 can be represented by a first-order rate law by graphing $\ln[H_2O_2]$ versus time. Determine the rate constant for the decomposition of H_2O_2 from these data.

Solution

The data from Figure 2.3.4.1 are tabulated below, and a plot of $\ln[H_2O_2]$ is shown in Figure 2.3.4.1.

Time (h)	$[H_2O_2]$ (*M*)	$\ln[H_2O_2]$
0.00	1.000	0.000
6.00	0.500	–0.693
12.00	0.250	–1.386
18.00	0.125	–2.079

Time (h)	$[H_2O_2]$ (*M*)	$\ln[H_2O_2]$
24.00	0.0625	−2.772

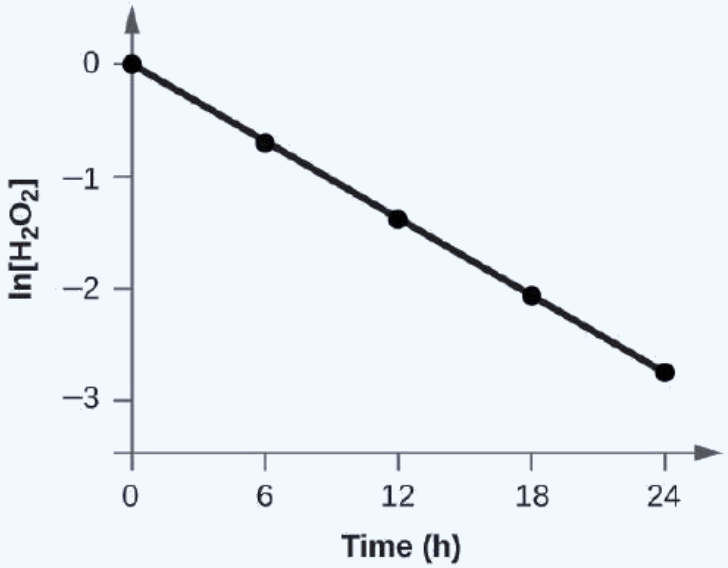


Figure 2.3.4.1: A linear relationship between $\ln[H_2O_2]$ and time suggests the decomposition of hydrogen peroxide is a first-order reaction.

The plot of $\ln[H_2O_2]$ versus time is linear, indicating that the reaction may be described by a first-order rate law.

According to the linear format of the first-order integrated rate law, the rate constant is given by the negative of this plot's slope.

$$\text{slope} = \frac{\text{change in } y}{\text{change in } x} = \frac{\Delta y}{\Delta x} = \frac{\Delta \ln [H_2O_2]}{\Delta t}$$

The slope of this line may be derived from two values of $\ln[H_2O_2]$ at different values of *t* (one near each end of the line is preferable). For example, the value of $\ln[H_2O_2]$ when *t* is 0.00 h is 0.000; the value when *t* = 24.00 h is −2.772

$$\begin{aligned} \text{slope} &= \frac{-2.772 - 0.000}{24.00 - 0.00h} \\ &= \frac{-2.772}{24.00h} \\ &= -0.116h^{-1} \\ k = -\text{ slope} &= -\left(-0.116h^{-1}\right) = 0.116h^{-1} \end{aligned}$$

Exercise 2.3.4.2

Graph the following data to determine whether the reaction $A \longrightarrow B + C$ is first order.

Time (s)	[*A*]
4.0	0.220
8.0	0.144
12.0	0.110
16.0	0.088
20.0	0.074

Answer available at the end of this section.

Second-Order Reactions

The equations that relate the concentrations of reactants and the rate constant of second-order reactions can be fairly complicated. To illustrate the point with minimal complexity, only the simplest second-order reactions will be described here, namely, those whose rates depend on the concentration of just one reactant. For these types of reactions, the differential rate law is written as:

$$\text{rate } = k[A]^2$$

For these second-order reactions, the integrated rate law is:

$$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$$

where the terms in the equation have their usual meanings as defined earlier.

Example 2.3.4.3: The Integrated Rate Law for a Second-Order Reaction

The reaction of butadiene gas (C_4H_6) to yield C_8H_{12} gas is described by the equation:

$$2C_4H_6(g) \longrightarrow C_8H_{12}(g)$$

This "dimerization" reaction is second order with a rate constant equal to 5.76 10^{-2} L mol^{-1} min^{-1} under certain conditions. If the initial concentration of butadiene is 0.200 *M*, what is the concentration after 10.0 min?

Solution

For a second-order reaction, the integrated rate law is written

$$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$$

We know three variables in this equation: $[A]_0$ = 0.200 mol/L, k = 5.76 10^{-2} L/mol/min, and t = 10.0 min. Therefore, we can solve for [*A*], the fourth variable:

$$\begin{aligned}\frac{1}{[A]_t} &= \left(5.76 \times 10^{-2} L mol^{-1} min^{-1}\right)(10 min) + \frac{1}{0.200 mol^{-1}} \\ &= \left(5.76 \times 10^{-1} L mol^{-1}\right) + 5.00 L mol^{-1} \\ &= 5.58 L mol^{-1} \\ [A]_t &= 1.79 \times 10^{-1} mol L^{-1}\end{aligned}$$

Therefore 0.179 mol/L of butadiene remain at the end of 10.0 min, compared to the 0.200 mol/L that was originally present.

Exercise 2.3.4.3

If the initial concentration of butadiene is 0.0200 *M*, what is the concentration remaining after 20.0 min?

Answer available at the end of this section.

The integrated rate law for second-order reactions has the form of the equation of a straight line:

$$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$$

$$y = mx + b$$

A plot of $\frac{1}{[A]_t}$ versus t for a second-order reaction is a straight line with a slope of k and a y-intercept of $\frac{1}{[A]_0}$. If the plot is not a straight line, then the reaction is not second order.

Example 2.3.4.4: Graphical Determination of Reaction Order and Rate Constant

The data below are for the same reaction described in Example 2.3.4.3. Prepare and compare two appropriate data plots to identify the reaction as being either first or second order. After identifying the reaction order, estimate a value for the rate constant.

Solution

Time (s)	$[C_4H_6]$ (M)
0	1.00×10^{-2}
1600	5.04×10^{-3}
3200	3.37×10^{-3}
4800	2.53×10^{-3}
6200	2.08×10^{-3}

In order to distinguish a first-order reaction from a second-order reaction, prepare a plot of $\ln[C_4H_6]_t$ versus t and compare it to a plot of $\frac{1}{[C_4H_6]_t}$ versus t. The values needed for these plots follow.

Time (s)	$\frac{1}{[C_4H_6]}(M^{-1})$	$\ln[C_4H_6]$
0	100	−4.605
1600	198	−5.289
3200	296	−5.692
4800	395	−5.978
6200	481	−6.175

The plots are shown in Figure 2.3.4.2, which clearly shows the plot of $\ln [C_4H_6]_t$ versus t is not linear, therefore the reaction is not first order. The plot of $\frac{1}{[C_4H_6]_t}$ versus t is linear, indicating that the reaction is second order.

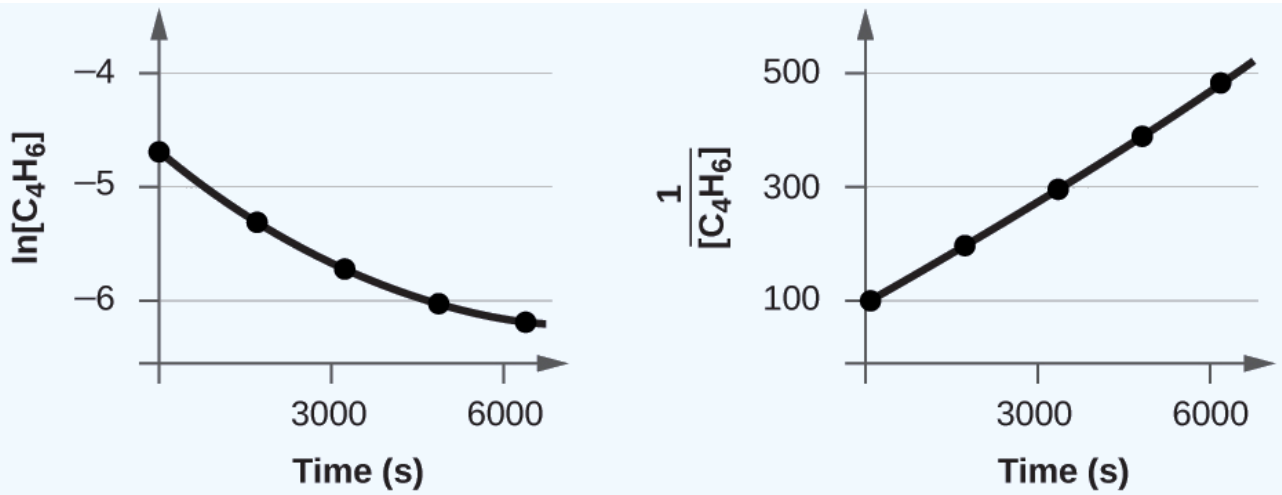


Figure 2.3.4.2: These two graphs show first- and second-order plots for the dimerization of C_4H_6. The linear trend in the second-order plot (right) indicates that the reaction follows second-order kinetics.

According to the second-order integrated rate law, the rate constant is equal to the slope of the versus *t* plot. Using the data for *t* = 0 s and *t* = 6200 s, the rate constant is estimated as follows:

$$k = \text{ slope } = \frac{(481M^{-1} - 100M^{-1})}{(6200s - 0s)} = 0.0614M^{-1}s^{-1}$$

Exercise 2.3.4.4

Do the following data fit a second-order rate law?

Time (s)	[*A*] (*M*)
5	0.952
10	0.625
15	0.465
20	0.370
25	0.308
35	0.230

Answer available at the end of this section.

Zero-Order Reactions

For zero-order reactions, the differential rate law is:

$$\text{rate } = k$$

A zero-order reaction thus exhibits a constant reaction rate, regardless of the concentration of its reactant(s). This may seem counterintuitive, since the reaction rate certainly can't be finite when the reactant concentration is zero. For purposes of this introductory text, it will suffice to note that zero-order kinetics are observed for some reactions only under certain specific conditions. These same reactions exhibit different kinetic behaviors when the specific conditions aren't met, and for this reason the more prudent term *pseudo-zero-order* is sometimes used.

The integrated rate law for a zero-order reaction is a linear function:

$$[A]_t = -kt + [A]_0$$
$$y = mx + b$$

A plot of [*A*] versus *t* for a zero-order reaction is a straight line with a slope of –*k* and a *y*-intercept of $[A]_0$. Figure 2.3.4.3: shows a plot of $[NH_3]$ versus *t* for the thermal decomposition of ammonia at the surface of two different heated solids. The decomposition reaction exhibits first-order behavior at a quartz (SiO_2) surface, as suggested by the exponentially decaying plot of concentration versus time. On a tungsten surface, however, the plot is linear, indicating zero-order kinetics.

Example 2.3.4.5: Graphical Determination of Zero-Order Rate Constant

Use the data plot in Figure 2.3.4.3 to graphically estimate the zero-order rate constant for ammonia decomposition at a tungsten surface.

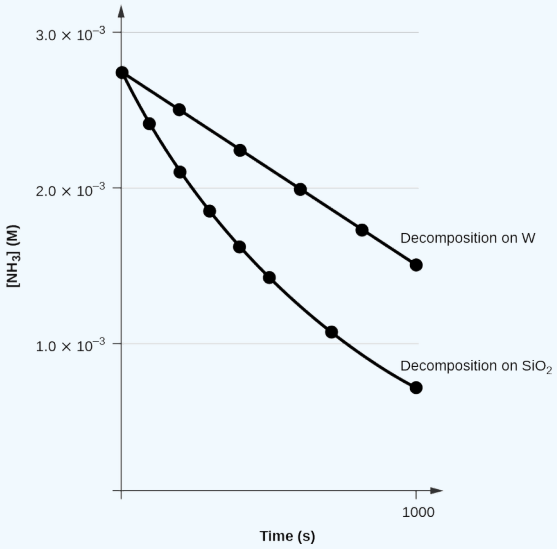


Figure 2.3.4.3: The decomposition of NH_3 on a tungsten (W) surface is a zero-order reaction, whereas on a quartz (SiO_2) surface, the reaction is first order.

Solution

The integrated rate law for zero-order kinetics describes a linear plot of reactant concentration, $[A]_t$, versus time, *t*, with a slope equal to the negative of the rate constant, –*k*. Following the mathematical approach of previous examples, the slope of the linear data plot (for decomposition on W) is estimated from the graph. Using the ammonia concentrations at *t* = 0 and *t* = 1000 s:

$$k = -\text{ slope } = -\frac{(0.0015 molL^{-1} - 0.0028 molL^{-1})}{(1000s - 0s)} = 1.3 \times 10^{-6} molL^{-1}s^{-1}$$

Exercise 2.3.4.5

The zero-order plot in Figure 2.3.4.3 shows an initial ammonia concentration of 0.0028 mol L^{-1} decreasing linearly with time for 1000 s. Assuming no change in this zero-order behavior, at what time (min) will the concentration reach 0.0001 mol L^{-1}?

Answer available at the end of this section.

The Half-Life of a Reaction

The half-life of a reaction ($t_{1/2}$) is the time required for one-half of a given amount of reactant to be consumed. In each succeeding

half-life, half of the remaining concentration of the reactant is consumed. Using the decomposition of hydrogen peroxide (Figure 2.3.4.1) as an example, we find that during the first half-life (from 0.00 hours to 6.00 hours), the concentration of H_2O_2 decreases from 1.000 *M* to 0.500 *M*. During the second half-life (from 6.00 hours to 12.00 hours), it decreases from 0.500 *M* to 0.250 *M*; during the third half-life, it decreases from 0.250 *M* to 0.125 *M*. The concentration of H_2O_2 decreases by half during each successive period of 6.00 hours. The decomposition of hydrogen peroxide is a first-order reaction, and, as can be shown, the half-life of a first-order reaction is independent of the concentration of the reactant. However, half-lives of reactions with other orders depend on the concentrations of the reactants.

First-Order Reactions

An equation relating the half-life of a first-order reaction to its rate constant may be derived from the integrated rate law as follows:

$$\ln \frac{[A]_0}{[A]_t} = kt$$

$$t = \ln \frac{[A]_0}{[A]_t} \times \frac{1}{k}$$

Invoking the definition of half-life, symbolized requires that the concentration of *A* at this point is one-half its initial concentration: $t = t_{1/2}$, $[A]_t = \frac{1}{2}[A]_0$

Substituting these terms into the rearranged integrated rate law and simplifying yields the equation for half-life:

$$t_{1/2} = \ln \frac{[A]_0}{\frac{1}{2}[A]_0} \times \frac{1}{k}$$

$$= \ln 2 \times \frac{1}{k} = 0.693 \times \frac{1}{k}$$

$$t_{1/2} = \frac{0.693}{k}$$

This equation describes an expected inverse relation between the half-life of the reaction and its rate constant, *k*. Faster reactions exhibit larger rate constants and correspondingly shorter half-lives. Slower reactions exhibit smaller rate constants and longer half-lives.

Example 2.3.4.6: Calculation of a First-order Rate Constant using Half-Life

Calculate the rate constant for the first-order decomposition of hydrogen peroxide in water at 40 °C, using the data given in Figure 2.3.4.4.

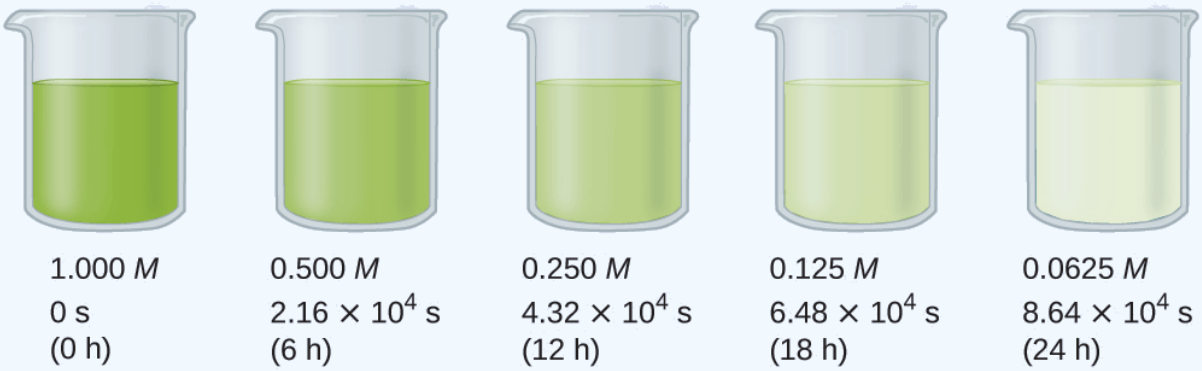


Figure 2.3.4.4: The decomposition of H_2O_2 at 40 °C is illustrated. The intensity of the color symbolizes the concentration of H_2O_2 at the indicated times; H_2O_2 is actually colorless.

Solution

Inspecting the concentration/time data in Figure 2.3.4.4 shows the half-life for the decomposition of H_2O_2 is 2.16 × 10^4 s:

$$t_{1/2} = \frac{0.693}{k}$$

$$k = \frac{0.693}{t_{1/2}} = \frac{0.693}{2.16 \times 10^4 s} = 3.21 \times 10^{-5} s^{-1}$$

Exercise 2.3.4.1

The first-order radioactive decay of iodine-131 exhibits a rate constant of 0.138 d^{-1}. What is the half-life for this decay?

Answer available at the end of this section.

Second-Order Reactions

Following the same approach as used for first-order reactions, an equation relating the half-life of a second-order reaction to its rate constant and initial concentration may be derived from its integrated rate law:

$$\frac{1}{[A]_t} = kt + \frac{1}{[A]_0}$$

or

$$\frac{1}{[A]} - \frac{1}{[A]_0} = kt$$

Restrict t to $t_{1/2}$

$$t = t_{1/2}$$

define $[A]_t$ as one-half $[A]_0$

$$[A]_t = \frac{1}{2}[A]_0$$

and then substitute into the integrated rate law and simplify:

$$\frac{1}{\frac{1}{2}[A]_0} - \frac{1}{[A]_0} = kt_{1/2}$$

$$\frac{2}{[A]_0} - \frac{1}{[A]_0} = kt_{1/2}$$

$$\frac{1}{[A]_0} = kt_{1/2}$$

$$t_{1/2} = \frac{1}{k[A]_0}$$

For a second-order reaction, $t_{1/2}$ is inversely proportional to the concentration of the reactant, and the half-life increases as the reaction proceeds because the concentration of reactant decreases. Unlike with first-order reactions, the rate constant of a second-order reaction cannot be calculated directly from the half-life unless the initial concentration is known.

Zero-Order Reactions

As for other reaction orders, an equation for zero-order half-life may be derived from the integrated rate law:

$$[A] = -kt + [A]_0$$

Restricting the time and concentrations to those defined by half-life: and Substituting these terms into the zero-order integrated rate law yields:

$$\frac{[A]_0}{2} = -kt_{1/2} + [A]_0$$

$$kt_{1/2} = \frac{[A]_0}{2}$$

$$t_{1/2} = \frac{[A]_0}{2k}$$

As for all reaction orders, the half-life for a zero-order reaction is inversely proportional to its rate constant. However, the half-life of a zero-order reaction increases as the initial concentration increases.

Equations for both differential and integrated rate laws and the corresponding half-lives for zero-, first-, and second-order reactions are summarized in Table 2.3.4.1.

Table 2.3.4.1: Summary of Rate Laws for Zero-, First-, and Second-Order Reactions

	Zero-Order	First-Order	Second-Order
rate law	rate = k	rate = $k[A]$	rate = $k[A]^2$
units of rate constant	M s^{-1}	s^{-1}	M^{-1} s^{-1}
integrated rate law	$[A] = -kt + [A]_0$	$\ln[A] = -kt + \ln[A]_0$	$\frac{1}{[A]} = kt + \left(\frac{1}{[A]_0}\right)$
plot needed for linear fit of rate data	$[A]$ vs. t	ln$[A]$ vs. t	vs. t
relationship between slope of linear plot and rate constant	k = –slope	k = –slope	k = slope
half-life	$t_{1/2} = \frac{[A]_0}{2k}$	$t_{1/2} = \frac{0.693}{k}$	$t_{1/2} = \frac{1}{[A]_0 k}$

Example 2.3.4.7: Half-Life for Zero-Order and Second-Order Reactions

What is the half-life for the butadiene dimerization reaction described in Example 2.3.4.3?

Solution

The reaction in question is second order, is initiated with a 0.200 mol L^{-1} reactant solution, and exhibits a rate constant of 0.0576 L mol^{-1} min^{-1}. Substituting these quantities into the second-order half-life equation:

$$t_{1/2} = \frac{1}{[(0.0576 L mol^{-1} min^{-1})\,(0.200 mol L^{-1})]} = 86.8 \text{ min}$$

Exercise 2.3.4.7

What is the half-life (min) for the thermal decomposition of ammonia on tungsten (see Figure 2.3.4.3)?

Answer available at the end of this section.

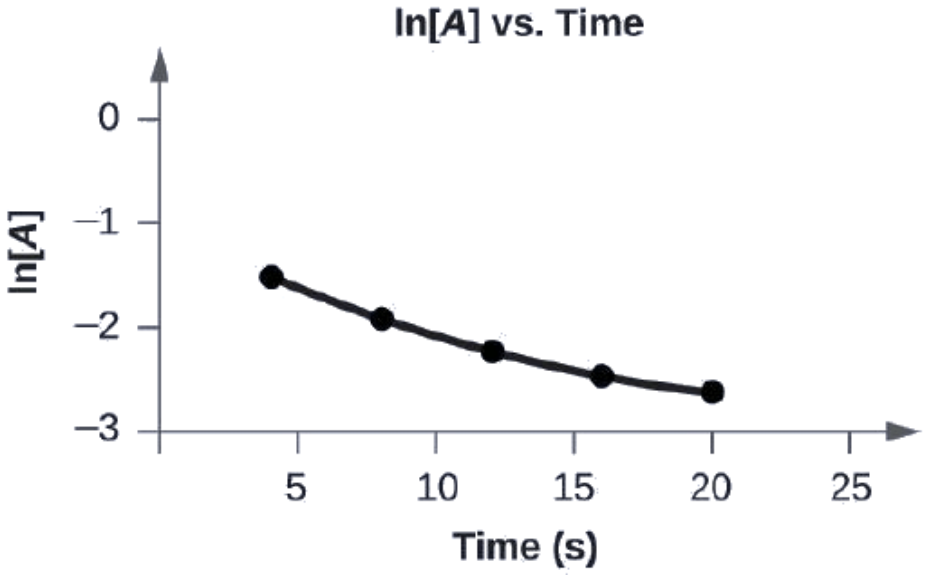
Answers

Exercise 2.3.4.1

16.7 days

Exercise 2.3.4.2

The plot of ln$[A]_t$ vs. t is not linear, indicating the reaction is not first order:

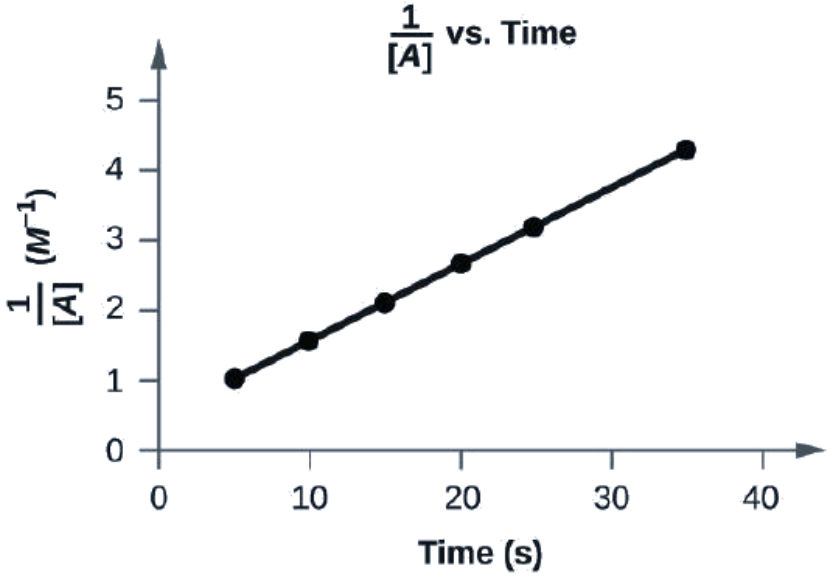


Exercise 2.3.4.3

0.0195 mol/L

Exercise 2.3.4.4

Yes. The plot of $\frac{1}{[A]_t}$ vs. t is linear:



Exercise 2.3.4.5

35 min

Exercise 2.3.4.1

5.02 d.

Exercise 2.3.4.7

87 min

This page titled 2.3.4: Integrated Rate Laws is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.4: Integrated Rate Laws by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.3.5: Collision Theory

Learning Objectives

By the end of this section, you will be able to:

- Use the postulates of collision theory to explain the effects of physical state, temperature, and concentration on reaction rates
- Define the concepts of activation energy and transition state
- Use the Arrhenius equation in calculations relating rate constants to temperature

We should not be surprised that atoms, molecules, or ions must collide before they can react with each other. Atoms must be close together to form chemical bonds. This simple premise is the basis for a very powerful theory that explains many observations regarding chemical kinetics, including factors affecting reaction rates.

Collision theory is based on the following postulates:

1. The rate of a reaction is proportional to the rate of reactant collisions:

$$\text{reaction rate} \propto \frac{\#\text{ collisions}}{\text{time}}$$

2. The reacting species must collide in an orientation that allows contact between the atoms that will become bonded together in the product.
3. The collision must occur with adequate energy to permit mutual penetration of the reacting species' valence shells so that the electrons can rearrange and form new bonds (and new chemical species).

We can see the importance of the two physical factors noted in postulates 2 and 3, the orientation and energy of collisions, when we consider the reaction of carbon monoxide with oxygen:

$$2\,\mathrm{CO\,(g)} + \mathrm{O_2\,(g)} \longrightarrow 2\,\mathrm{CO_2\,(g)}$$

Carbon monoxide is a pollutant produced by the combustion of hydrocarbon fuels. To reduce this pollutant, automobiles have catalytic converters that use a catalyst to carry out this reaction. It is also a side reaction of the combustion of gunpowder that results in muzzle flash for many firearms. If carbon monoxide and oxygen are present in sufficient amounts, the reaction will occur at high temperature and pressure.

The first step in the gas-phase reaction between carbon monoxide and oxygen is a collision between the two molecules:

$$\mathrm{CO\,(g)} + \mathrm{O_2\,(g)} \longrightarrow \mathrm{CO_2\,(g)} + \mathrm{O\,(g)}$$

Although there are many different possible orientations the two molecules can have relative to each other, consider the two presented in Figure 2.3.5.1. In the first case, the oxygen side of the carbon monoxide molecule collides with the oxygen molecule. In the second case, the carbon side of the carbon monoxide molecule collides with the oxygen molecule. The second case is clearly more likely to result in the formation of carbon dioxide, which has a central carbon atom bonded to two oxygen atoms (\ce{O=C=O}\). This is a rather simple example of how important the orientation of the collision is in terms of creating the desired product of the reaction.

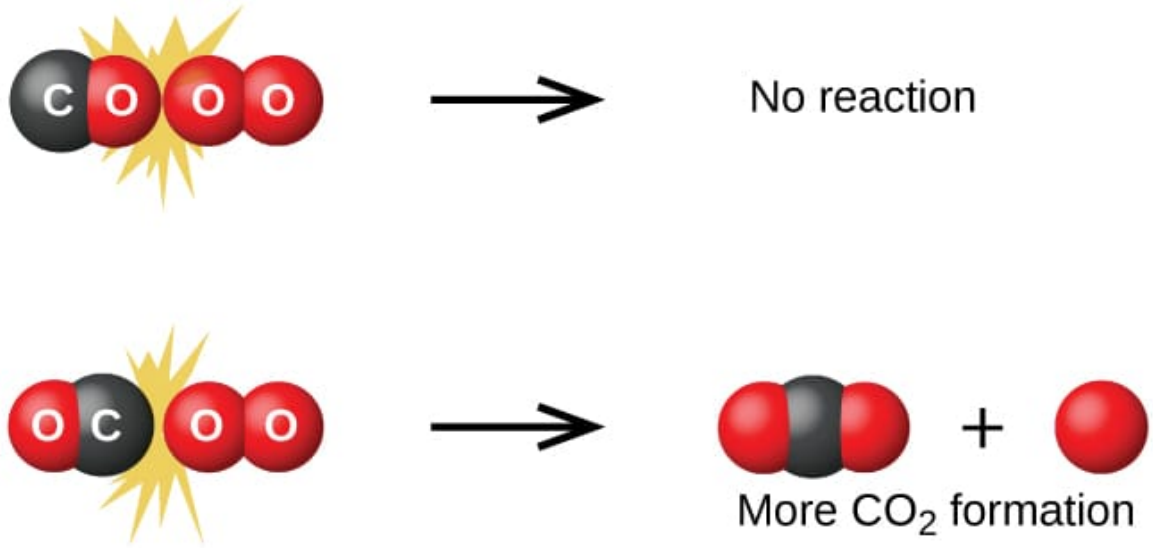


Figure 2.3.5.1: Illustrated are two collisions that might take place between carbon monoxide and oxygen molecules. The orientation of the colliding molecules partially determines whether a reaction between the two molecules will occur.

If the collision does take place with the correct orientation, there is still no guarantee that the reaction will proceed to form carbon dioxide. In addition to a proper orientation, the collision must also occur with sufficient energy to result in product formation. When reactant species collide with both proper orientation and adequate energy, they combine to form an unstable species called an **activated complex** or a **transition state**. These species are very short lived and usually undetectable by most analytical instruments. In some cases, sophisticated spectral measurements have been used to observe transition states.

Collision theory explains why most reaction rates increase as concentrations increase. With an increase in the concentration of any reacting substance, the chances for collisions between molecules are increased because there are more molecules per unit of volume. More collisions mean a faster reaction rate, assuming the energy of the collisions is adequate.

Activation Energy and the Arrhenius Equation

The minimum energy necessary to form a product during a collision between reactants is called the **activation energy** (E_a). How this energy compares to the kinetic energy provided by colliding reactant molecules is a primary factor affecting the rate of a chemical reaction. If the activation energy is much larger than the average kinetic energy of the molecules, the reaction will occur slowly since only a few fast-moving molecules will have enough energy to react. If the activation energy is much smaller than the average kinetic energy of the molecules, a large fraction of molecules will be adequately energetic and the reaction will proceed rapidly.

Figure 2.3.5.2 shows how the energy of a chemical system changes as it undergoes a reaction converting reactants to products according to the equation

$$\mathrm{A^{+}B} \longrightarrow \mathrm{C^{+}D}$$

These reaction diagrams are widely used in chemical kinetics to illustrate various properties of the reaction of interest. Viewing the diagram from left to right, the system initially comprises reactants only, $\mathrm{A} + \mathrm{B}$. Reactant molecules with sufficient energy can collide to form a high-energy activated complex or transition state. The unstable transition state can then subsequently decay to yield stable products, *C* + *D*. The diagram depicts the reaction's activation energy, E_a, as the energy difference between the reactants and the transition state. Using a specific energy, the *enthalpy* (see chapter on thermochemistry), the enthalpy change of the reaction, ΔH, is estimated as the energy difference between the reactants and products. In this case, the reaction is exothermic ($\Delta H < 0$) since it yields a decrease in system enthalpy.

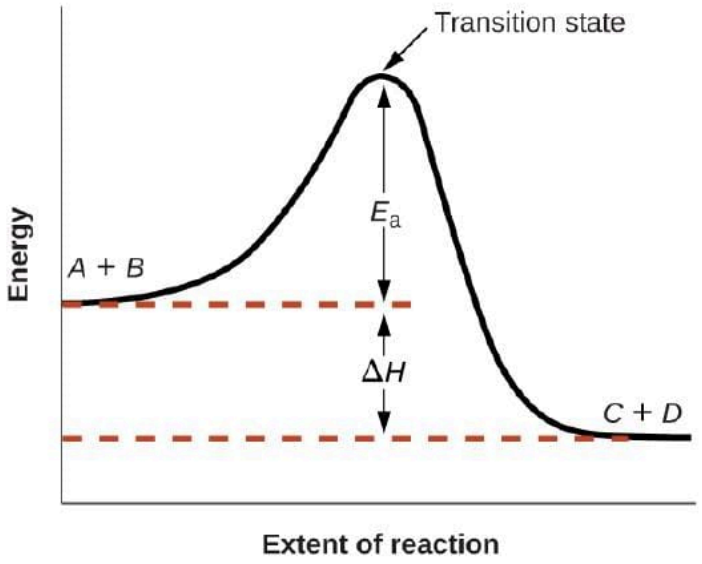


Figure 2.3.5.2: Reaction diagram for the exothermic reaction

The **Arrhenius equation** relates the activation energy and the rate constant, *k*, for many chemical reactions:

$$k = Ae^{-E_a/RT}$$

In this equation, R is the ideal gas constant, which has a value 8.314 J/mol/K, T is temperature on the Kelvin scale, E_a is the activation energy in joules per mole, *e* is the constant 2.7183, and *A* is a constant called the **frequency factor**, which is related to the frequency of collisions and the orientation of the reacting molecules.

Postulates of collision theory are nicely accommodated by the Arrhenius equation. The frequency factor, *A*, reflects how well the reaction conditions favor properly oriented collisions between reactant molecules. An increased probability of effectively oriented collisions results in larger values for *A* and faster reaction rates.

The exponential term, $e^{-Ea/RT}$, describes the effect of activation energy on reaction rate. According to kinetic molecular theory (see chapter on gases), the temperature of matter is a measure of the average kinetic energy of its constituent atoms or molecules. The distribution of energies among the molecules composing a sample of matter at any given temperature is described by the plot shown in Figure 2.3.5.3*a*. Two shaded areas under the curve represent the numbers of molecules possessing adequate energy (*RT*) to overcome the activation barriers (E_a). A lower activation energy results in a greater fraction of adequately energized molecules and a faster reaction.

The exponential term also describes the effect of temperature on reaction rate. A higher temperature represents a correspondingly greater fraction of molecules possessing sufficient energy (*RT*) to overcome the activation barrier (E_a), as shown in Figure 2.3.5.3*b*. This yields a greater value for the rate constant and a correspondingly faster reaction rate.

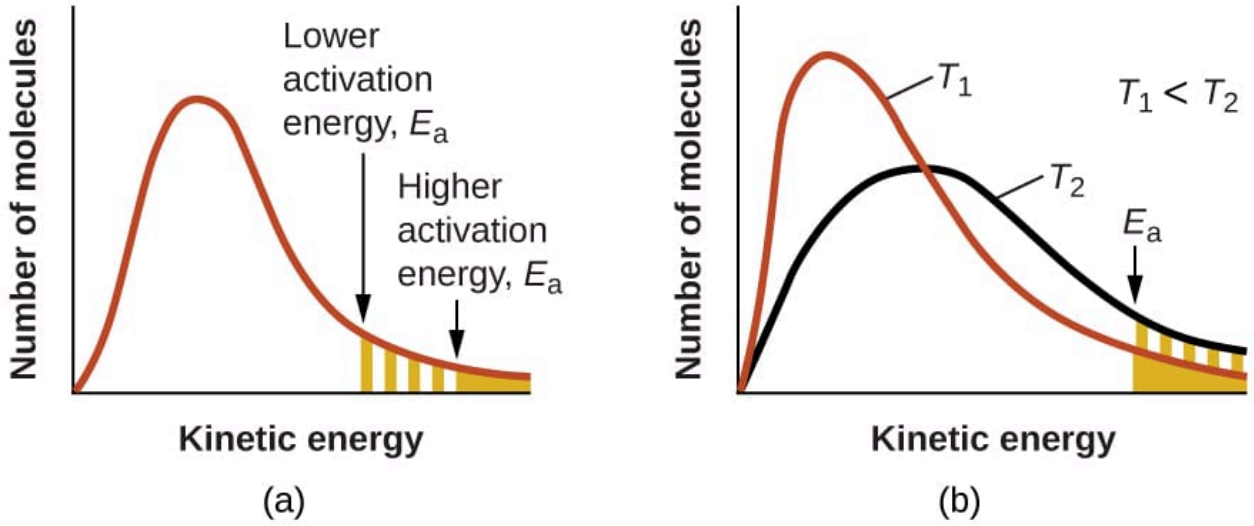


Figure 2.3.5.3: Molecular energy distributions showing numbers of molecules with energies exceeding (a) two different activation energies at a given temperature, and (b) a given activation energy at two different temperatures.

A convenient approach for determining E_a for a reaction involves the measurement of *k* at two or more different temperatures and using an alternate version of the Arrhenius equation that takes the form of a linear equation

$$\ln k = \left(\frac{-E_a}{R}\right)\left(\frac{1}{T}\right) + \ln A$$

$$y = mx + b$$

A plot of $\ln k$ versus $\frac{1}{T}$ is linear with a slope equal to $\frac{-E_a}{R}$ and a y-intercept equal to $\ln A$.

Example 2.3.5.1: Determination of E_a

The variation of the rate constant with temperature for the decomposition of $\mathrm{HI(g)}$ to $\mathrm{H_2(g)}$ and $\mathrm{I_2(g)}$ is given here. What is the activation energy for the reaction?

$$2\,\mathrm{HI(g)} \longrightarrow \mathrm{H_2(g) + I_2(g)}$$

T (K)	*k* (L/mol/s)
555	3.52×10^{-7}
575	1.22×10^{-6}
645	8.59×10^{-5}
700	1.16×10^{-3}
781	3.95×10^{-2}

Solution

Use the provided data to derive values of $\frac{1}{T}$ and $\ln k$:

$\frac{1}{T}$	$\ln k$
1.80×10^{-3}	−14.860
1.74×10^{-3}	−13.617

$\frac{1}{T}$	$\ln k$
1.55×10^{-3}	−9.362
1.43×10^{-3}	−6.759
1.28×10^{-3}	−3.231

Figure 2.3.5.4 is a graph of $\ln k$ versus $\frac{1}{T}$. In practice, the equation of the line (slope and *y*-intercept) that best fits these plotted data points would be derived using a statistical process called regression. This is helpful for most experimental data because a perfect fit of each data point with the line is rarely encountered. For the data here, the fit is nearly perfect and the slope may be estimated using any two of the provided data pairs. Using the first and last data points permits estimation of the slope.

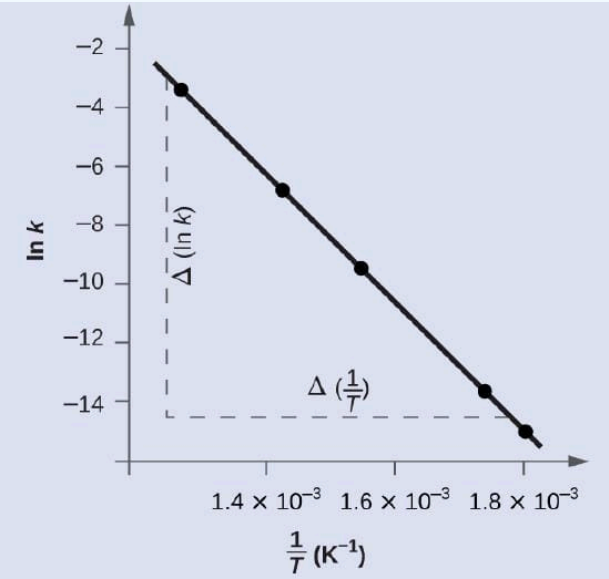


Figure 2.3.5.4: This graph shows the linear relationship between $\ln k$ versus $\frac{1}{T}$ for the reaction $2HI \longrightarrow H_2 + I_2$ according to the Arrhenius equation.

$$\begin{aligned}
\text{Slope} &= \frac{\Delta(\ln k)}{\Delta\left(\frac{1}{T}\right)} \\
&= \frac{(-14.860) - (-3.231)}{\left(1.80 \times 10^{-3} K^{-1}\right) - \left(1.28 \times 10^{-3} K^{-1}\right)} \\
&= \frac{-11.629}{0.52 \times 10^{-3} K^{-1}} = -2.2 \times 10^{4} K \\
&= -\frac{E_a}{R} \\
E_a &= -\ \text{slope}\ \times R = -\left(-2.2 \times 10^{4} K \times 8.314 J mol^{-1} K^{-1}\right) \\
&\quad 1.8 \times 10^{5} J mol^{-1} \text{ or } 180 k J mol^{-1}
\end{aligned}$$

Alternative approach: A more expedient approach involves deriving activation energy from measurements of the rate constant at just two temperatures. In this approach, the Arrhenius equation is rearranged to a convenient two-point form:

$$\ln \frac{k_1}{k_2} = \frac{E_a}{R}\left(\frac{1}{T_2} - \frac{1}{T_1}\right)$$

Rearranging this equation to isolate activation energy yields:

$$E_a = -R\left(\frac{\ln k_2 - \ln k_1}{\left(\frac{1}{T_2}\right) - \left(\frac{1}{T_1}\right)}\right)$$

Any two data pairs may be substituted into this equation—for example, the first and last entries from the above data table:

$$E_a = -8.314 J mol^{-1} K^{-1}\left(\frac{-3.231 - (-14.860)}{1.28 \times 10^{-3} K^{-1} - 1.80 \times 10^{-3} K^{-1}}\right)$$

and the result is E_a = 1.8 10^5 J mol^{-1} or 180 kJ mol^{-1}

This approach yields the same result as the more rigorous graphical approach used above, as expected. In practice, the graphical approach typically provides more reliable results when working with actual experimental data.

Exercise 2.3.5.1

The rate constant for the rate of decomposition of $\mathrm{N_2O_5}$ to NO and $\mathrm{O_2}$ in the gas phase is 1.66 L/mol/s at 650 K and 7.39 L/mol/s at 700 K:

$$2\,\mathrm{N_2O_5\,(g)} \longrightarrow 4\,\mathrm{NO\,(g)} + 3\,\mathrm{O_2\,(g)}$$

Assuming the kinetics of this reaction are consistent with the Arrhenius equation, calculate the activation energy for this decomposition.

Answer available at the end of this section.

Answers

Exercise 2.3.5.1

1.1 10^5 J mol^{-1} or 110 kJ mol^{-1}

This page titled 2.3.5: Collision Theory is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.5: Collision Theory by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.3.6: Reaction Mechanisms

Learning Objectives

By the end of this section, you will be able to:

- Distinguish net reactions from elementary reactions (steps)
- Identify the molecularity of elementary reactions
- Write a balanced chemical equation for a process given its reaction mechanism
- Derive the rate law consistent with a given reaction mechanism

Chemical reactions very often occur in a step-wise fashion, involving two or more distinct reactions taking place in sequence. A balanced equation indicates what is reacting and what is produced, but it reveals no details about how the reaction actually takes place. The reaction mechanism (or reaction path) provides details regarding the precise, step-by-step process by which a reaction occurs.

The decomposition of ozone, for example, appears to follow a mechanism with two steps:

$$O_3(g) \longrightarrow O_2(g) + O$$
$$O + O_3(g) \longrightarrow 2O_2(g)$$

Each of the steps in a reaction mechanism is an **elementary reaction**. These elementary reactions occur precisely as represented in the step equations, and they must sum to yield the balanced chemical equation representing the overall reaction:

$$2O_3(g) \longrightarrow 3O_2(g)$$

Notice that the oxygen atom produced in the first step of this mechanism is consumed in the second step and therefore does not appear as a product in the overall reaction. Species that are produced in one step and consumed in a subsequent step are called intermediates.

While the overall reaction equation for the decomposition of ozone indicates that two molecules of ozone react to give three molecules of oxygen, the mechanism of the reaction *does not involve the direct collision and reaction of two ozone molecules*. Instead, one O_3 decomposes to yield O_2 and an oxygen atom, and a second O_3 molecule subsequently reacts with the oxygen atom to yield two additional O_2 molecules.

Unlike balanced equations representing an overall reaction, the equations for elementary reactions are explicit representations of the chemical change taking place. The reactant(s) in an elementary reaction's equation undergo only the bond-breaking and/or making events depicted to yield the product(s). For this reason, *the rate law for an elementary reaction may be derived directly from the balanced chemical equation describing the reaction*. This is not the case for typical chemical reactions, for which rate laws may be reliably determined only via experimentation.

Unimolecular Elementary Reactions

The molecularity of an elementary reaction is the number of reactant species (atoms, molecules, or ions). For example, a unimolecular reaction involves the reaction of a *single* reactant species to produce one or more molecules of product:

$$A \longrightarrow \text{ products}$$

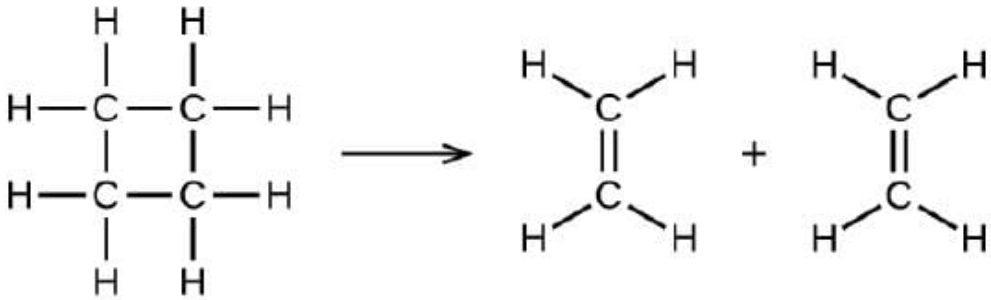
The rate law for a unimolecular reaction is first order:

$$\text{rate } = k[A]$$

A unimolecular reaction may be one of several elementary reactions in a complex mechanism. For example, the reaction:

$$O_3 \longrightarrow O_2 + O$$

illustrates a unimolecular elementary reaction that occurs as one part of a two-step reaction mechanism as described above. However, some unimolecular reactions may be the only step of a single-step reaction mechanism. (In other words, an "overall" reaction may also be an elementary reaction in some cases.) For example, the gas-phase decomposition of cyclobutane, C_4H_8, to ethylene, C_2H_4, is represented by the following chemical equation:



This equation represents the overall reaction observed, and it might also represent a legitimate unimolecular elementary reaction. The rate law predicted from this equation, assuming it is an elementary reaction, turns out to be the same as the rate law derived experimentally for the overall reaction, namely, one showing first-order behavior:

$$\text{rate } = -\frac{\Delta\left[C_4H_8\right]}{\Delta t} = k\left[C_4H_8\right]$$

This agreement between observed and predicted rate laws is interpreted to mean that the proposed unimolecular, single-step process is a reasonable mechanism for the butadiene reaction.

Bimolecular Elementary Reactions

A bimolecular reaction involves two reactant species, for example:

$$A + B \longrightarrow \text{ products}$$

and

$$2A \longrightarrow \text{ products}$$

For the first type, in which the two reactant molecules are different, the rate law is first-order in *A* and first order in *B* (second-order overall):

$$\text{rate } = k[A][B]$$

For the second type, in which two identical molecules collide and react, the rate law is second order in *A*:

$$\text{rate } = k[A][A] = k[A]^2$$

Some chemical reactions occur by mechanisms that consist of a single bimolecular elementary reaction. One example is the reaction of nitrogen dioxide with carbon monoxide:

$$NO_2(g) + CO(g) \longrightarrow NO(g) + CO_2(g)$$

(see Figure 2.3.6.1)



Figure 2.3.6.1: The probable mechanism for the reaction between NO_2 and CO to yield NO and CO_2.

Bimolecular elementary reactions may also be involved as steps in a multistep reaction mechanism. The reaction of atomic oxygen with ozone is the second step of the two-step ozone decomposition mechanism discussed earlier in this section:

$$O(g) + O_3(g) \longrightarrow 2O_2(g)$$

Termolecular Elementary Reactions

An elementary termolecular reaction involves the simultaneous collision of three atoms, molecules, or ions. Termolecular elementary reactions are uncommon because the probability of three particles colliding simultaneously is less than one one-thousandth of the probability of two particles colliding. There are, however, a few established termolecular elementary reactions. The reaction of nitric oxide with oxygen appears to involve termolecular steps:

$$2NO + O_2 \longrightarrow 2NO_2$$
$$\text{rate } = k[NO]^2 [O_2]$$

Likewise, the reaction of nitric oxide with chlorine appears to involve termolecular steps:

$$2NO + Cl_2 \longrightarrow 2NOCl$$
$$\text{rate } = k[NO]^2 [Cl_2]$$

Relating Reaction Mechanisms to Rate Laws

It's often the case that one step in a multistep reaction mechanism is significantly slower than the others. Because a reaction cannot proceed faster than its slowest step, this step will limit the rate at which the overall reaction occurs. The slowest step is therefore called the rate-limiting step (or rate-determining step) of the reaction Figure 2.3.6.2.



Figure 2.3.6.2: A cattle chute is a nonchemical example of a rate-determining step. Cattle can only be moved from one holding pen to another as quickly as one animal can make its way through the chute. (credit: Loren Kerns)

As described earlier, rate laws may be derived directly from the chemical equations for elementary reactions. This is not the case, however, for ordinary chemical reactions. The balanced equations most often encountered represent the overall change for some chemical system, and very often this is the result of some multistep reaction mechanisms. In every case, the rate law must be determined from experimental data and the reaction mechanism subsequently deduced from the rate law (and sometimes from other data). The reaction of NO_2 and CO provides an illustrative example:

$$NO_2(g) + CO(g) \longrightarrow CO_2(g) + NO(g)$$

For temperatures above 225 °C, the rate law has been found to be:

$$\text{rate } = k\,[NO_2]\,[CO]$$

The reaction is first order with respect to NO_2 and first-order with respect to CO. This is consistent with a single-step bimolecular mechanism and it is *possible* that this is the mechanism for this reaction at high temperatures.

At temperatures below 225 °C, the reaction is described by a rate law that is second order with respect to NO_2:

$$\text{rate } = k\,[NO_2]^2$$

This rate law is not consistent with the single-step mechanism, but is consistent with the following two-step mechanism:

$$NO_2(g) + NO_2(g) \longrightarrow NO_3(g) + NO(g)\ \text{(slow)}$$
$$NO_3(g) + CO(g) \longrightarrow NO_2(g) + CO_2(g)\ \text{(fast)}$$

The rate-determining (slower) step gives a rate law showing second-order dependence on the NO_2 concentration, and the sum of the two equations gives the net overall reaction.

In general, when the rate-determining (slower) step is the first step in a mechanism, the rate law for the overall reaction is the same as the rate law for this step. However, when the rate-determining step is preceded by a step involving a rapidly reversible reaction the rate law for the overall reaction may be more difficult to derive.

As discussed in several chapters of this text, a reversible reaction is at *equilibrium* when the rates of the forward and reverse processes are equal. Consider the reversible elementary reaction in which NO dimerizes to yield an intermediate species N_2O_2.

$$NO + NO \rightleftharpoons N_2O_2$$

When this reaction is at equilibrium:

$$\text{rate}_{\text{ forward}} = \text{rate}_{\text{ reverse}}$$
$$k_1[NO]^2 = k_{-1}\,[N_2O_2]$$

This expression may be rearranged to express the concentration of the intermediate in terms of the reactant NO:

$$\left(\frac{k_1[NO]^2}{k_{-1}}\right) = [N_2O_2]$$

Since intermediate species concentrations are not used in formulating rate laws for overall reactions, this approach is sometimes necessary, as illustrated in the following example exercise.

Example 2.3.6.1: Deriving a Rate Law from a Reaction Mechanism

The two-step mechanism below has been proposed for a reaction between nitrogen monoxide and molecular chlorine:

$$\text{Step 1:}\quad \mathrm{NO\,(g) + Cl_2\,(g) \rightleftharpoons NOCl_2\,(g)} \qquad \text{(fast)}$$
$$\text{Step 2:}\quad \mathrm{NOCl_2\,(g) + NO\,(g) \rightleftharpoons 2\,NOCl\,(g)} \qquad \text{(slow)}$$

Use this mechanism to derive the equation and predicted rate law for the overall reaction.

Solution

The equation for the overall reaction is obtained by adding the two elementary reactions:

$$2NO(g) + Cl_2(g) \longrightarrow 2NOCl(g)$$

To derive a rate law from this mechanism, first write rates laws for each of the two steps.

$$\text{rate}_{\,1} = k_1[NO]\,[Cl_2] \text{ for the forward reaction of step 1}$$
$$\text{rate}_{\,-1} = k_{-1}\,[NOCl_2] \text{ for the reverse reaction of step 1}$$
$$\text{rate}_{\,2} = k_2\,[NOCl_2]\,[NO] \text{ for step 2}$$

Step 2 is the rate-determining step, and so the rate law for the overall reaction should be the same as for this step. However, the step 2 rate law, as written, contains an intermediate species concentration, [$NOCl_2$]. To remedy this, use the first step's rate laws to derive an expression for the intermediate concentration in terms of the reactant concentrations.

Assuming step 1 is at equilibrium:

$$\text{rate}_{\,1} = \text{rate}_{\,-1}$$
$$k_1[NO]\,[Cl_2] = k_{-1}\,[NOCl_2]$$
$$[NOCl_2] = \left(\frac{k_1}{k_{-1}}\right)[NO]\,[Cl_2]$$

Substituting this expression into the rate law for step 2 yields:

$$\text{rate}_{\,2} = \text{rate}_{\,\text{overall}} = \left(\frac{k_2 k_1}{k_{-1}}\right)[NO]^2\,[Cl_2]$$

Exercise 2.3.6.1

The first step of a proposed multistep mechanism is:

\[\ce{F2(g) <=> 2 F(g) \tag{fast} \]

Derive the equation relating atomic fluorine concentration to molecular fluorine concentration.

Answer available at the end of this section.

Answers

Exercise 2.3.6.1

$$[\mathrm{F}] = \left(\frac{k_1\,[F_2]}{k_{-1}}\right)^{1/2}$$

This page titled 2.3.6: Reaction Mechanisms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.6: Reaction Mechanisms by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.3.7: Catalysis

Learning Objectives

By the end of this section, you will be able to:

- Explain the function of a catalyst in terms of reaction mechanisms and potential energy diagrams
- List examples of catalysis in natural and industrial processes

Among the factors affecting chemical reaction rates discussed earlier in this chapter was the presence of a *catalyst*, a substance that can increase the reaction rate without being consumed in the reaction. The concepts introduced in the previous section on reaction mechanisms provide the basis for understanding how catalysts are able to accomplish this very important function.

Figure 2.3.7.1 shows reaction diagrams for a chemical process in the absence and presence of a catalyst. Inspection of the diagrams reveals several traits of these reactions. Consistent with the fact that the two diagrams represent the same overall reaction, both curves begin and end at the same energies (in this case, because products are more energetic than reactants, the reaction is endothermic). The reaction mechanisms, however, are clearly different. The uncatalyzed reaction proceeds via a one-step mechanism (one transition state observed), whereas the catalyzed reaction follows a two-step mechanism (two transition states observed) with *a notably lesser activation energy*. This difference illustrates the means by which a catalyst functions to accelerate reactions, namely, by providing an alternative reaction mechanism with a lower activation energy. Although the catalyzed reaction mechanism for a reaction needn't necessarily involve a different number of steps than the uncatalyzed mechanism, it must provide a reaction path whose rate determining step is faster (lower E_a).

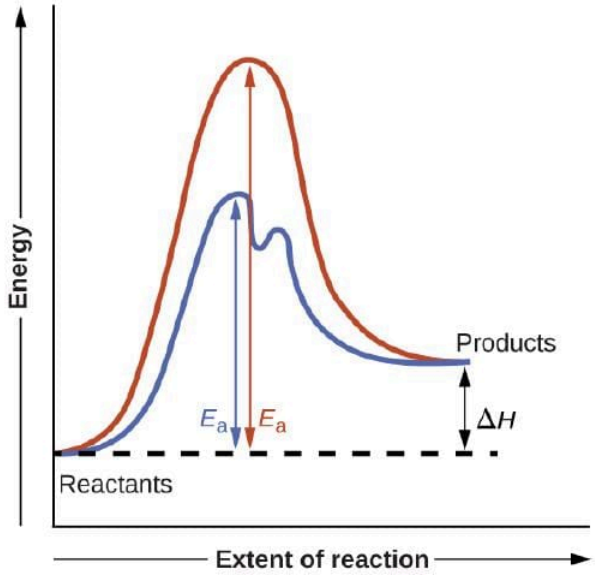
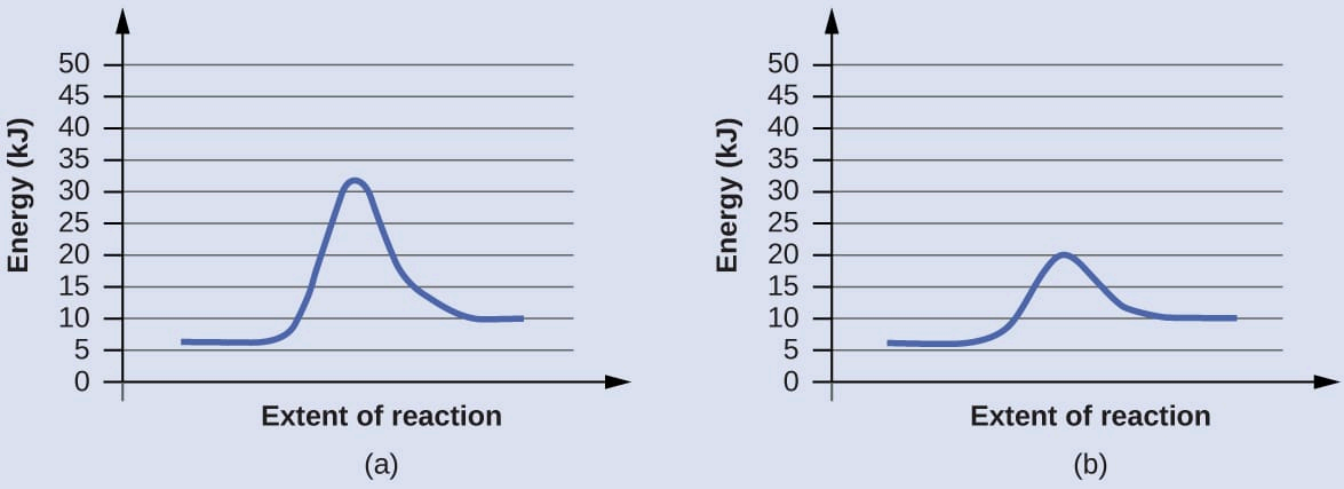


Figure 2.3.7.1: Reaction diagrams for an endothermic process in the absence (red curve) and presence (blue curve) of a catalyst. The catalyzed pathway involves a two-step mechanism (note the presence of two transition states) and an intermediate species (represented by the valley between the two transitions states).

Example 2.3.7.1: Reaction Diagrams for Catalyzed Reactions

The two reaction diagrams here represent the same reaction: one without a catalyst and one with a catalyst. Estimate the activation energy for each process, and identify which one involves a catalyst.



Solution

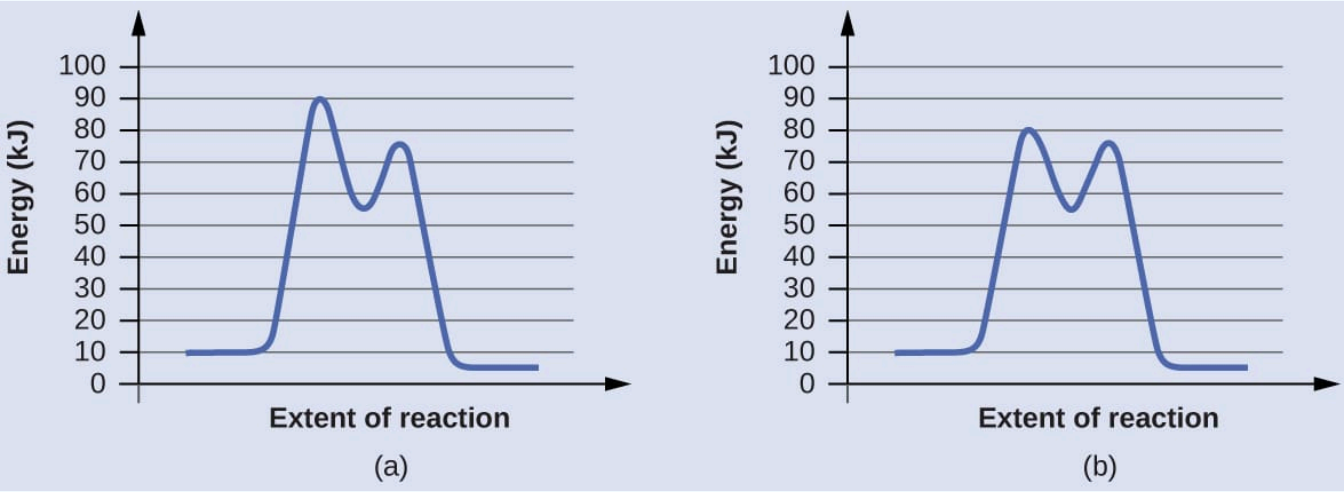
Activation energies are calculated by subtracting the reactant energy from the transition state energy.

$$\text{diagram}(a): E_a = 32kJ - 6kJ = 26kJ$$
$$\text{diagram}(b): E_a = 20kJ - 6kJ = 14kJ$$

The catalyzed reaction is the one with lesser activation energy, in this case represented by diagram b.

Exercise 2.3.7.1

Reaction diagrams for a chemical process with and without a catalyst are shown below. Both reactions involve a two-step mechanism with a rate-determining first step. Compute activation energies for the first step of each mechanism, and identify which corresponds to the catalyzed reaction. How do the second steps of these two mechanisms compare?



Answer available at the end of this section.

Homogeneous Catalysts

A homogeneous catalyst is present in the same phase as the reactants. It interacts with a reactant to form an intermediate substance, which then decomposes or reacts with another reactant in one or more steps to regenerate the original catalyst and

form product.

As an important illustration of homogeneous catalysis, consider the earth's ozone layer. Ozone in the upper atmosphere, which protects the earth from ultraviolet radiation, is formed when oxygen molecules absorb ultraviolet light and undergo the reaction:

$$3O_2(g) \xrightarrow{h\nu} 2O_3(g)$$

Ozone is a relatively unstable molecule that decomposes to yield diatomic oxygen by the reverse of this equation. This decomposition reaction is consistent with the following two-step mechanism:

$$O_3 \longrightarrow O_2 + O$$
$$O + O_3 \longrightarrow 2O_2$$

A number of substances can catalyze the decomposition of ozone. For example, the nitric oxide–catalyzed decomposition of ozone is believed to occur via the following three-step mechanism:

$$NO(g) + O_3(g) \longrightarrow NO_2(g) + O_2(g)$$
$$O_3(g) \longrightarrow O_2(g) + O(g)$$
$$NO_2(g) + O(g) \longrightarrow NO(g) + O_2(g)$$

As required, the overall reaction is the same for both the two-step uncatalyzed mechanism and the three-step NO-catalyzed mechanism:

$$2O_3(g) \longrightarrow 3O_2(g)$$

Notice that NO is a reactant in the first step of the mechanism and a product in the last step. This is another characteristic trait of a catalyst: Though it participates in the chemical reaction, it is not consumed by the reaction.

Portrait of a Chemist: Mario J. Molina

The 1995 Nobel Prize in Chemistry was shared by Paul J. Crutzen, Mario J. Molina (Figure 2.3.7.2), and F. Sherwood Rowland "for their work in atmospheric chemistry, particularly concerning the formation and decomposition of ozone."[1] Molina's initial discovery of stratospheric depletion of ozone by chlorine atoms was done at the University of California Irvine, he later moved to the NASA/Jet Propulsion Laboratory, MIT and the University of California, San Diego. At the time of his death in 2020 Molina, a Mexican citizen had established a new center for the study of energy and the environment in Mexico City.

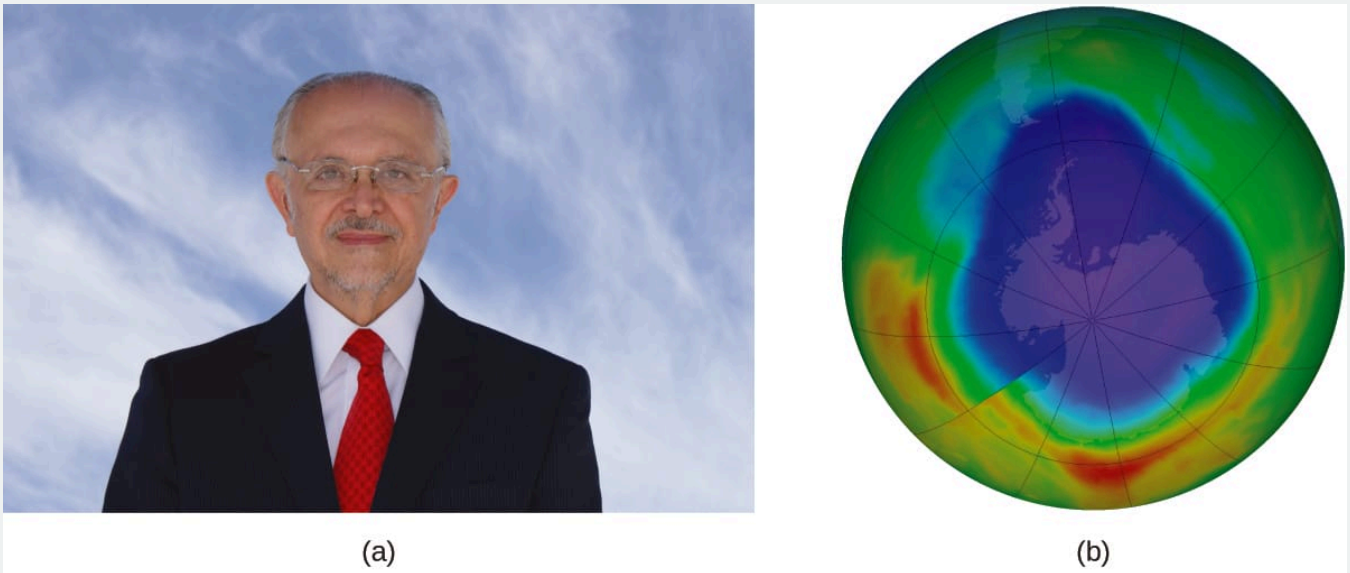


Figure 2.3.7.2: (a) Mexican chemist Mario Molina (1943 –) shared the Nobel Prize in Chemistry in 1995 for his (b) early research on stratospheric depletion of ozone and later the Antarctic ozone hole. (credit a: courtesy of Mario Molina; credit b: modification of work by NASA)

In 1974, Molina and Rowland published a paper in the journal *Nature* detailing the threat of chlorofluorocarbon gases to the stability of the ozone layer in earth's upper atmosphere. The ozone layer protects earth from solar radiation by absorbing ultraviolet light. As chemical reactions deplete the amount of ozone in the upper atmosphere, a measurable "hole" forms above Antarctica, and an increase in the amount of solar ultraviolet radiation— strongly linked to the prevalence of skin cancers—reaches earth's surface. The work of Molina and Rowland was instrumental in the adoption of the Montreal Protocol, an international treaty signed in 1987 that successfully began phasing out production of chemicals linked to ozone destruction.

Molina and Rowland demonstrated that chlorine atoms from human-made chemicals can catalyze ozone destruction in a process similar to that by which NO accelerates the depletion of ozone. Chlorine atoms are generated when chlorocarbons or chlorofluorocarbons—once widely used as refrigerants and propellants—are photochemically decomposed by ultraviolet light or react with hydroxyl radicals. A sample mechanism is shown here using methyl chloride:

$$CH_3Cl + OH \longrightarrow Cl + \text{ other products} \tag{2.3.7.1}$$

Chlorine radicals break down ozone and are regenerated by the following catalytic cycle:

$$Cl + O_3 \longrightarrow ClO + O_2$$
$$ClO + O \longrightarrow Cl + O_2$$

overall Reaction: $O_3 + O \longrightarrow 2O_2$ \nonumber \]

A single monatomic chlorine can break down thousands of ozone molecules. Luckily, the majority of atmospheric chlorine exists as the catalytically inactive forms Cl_2 and $ClONO_2$.

Enzymes in the human body act as catalysts for important chemical reactions in cellular metabolism. As such, a deficiency of a particular enzyme can translate to a life-threatening disease. G6PD (glucose-6-phosphate dehydrogenase) deficiency, a genetic condition that results in a shortage of the enzyme glucose-6-phosphate dehydrogenase, is the most common enzyme deficiency in humans. This enzyme, shown in Figure 12.21, is the rate-limiting enzyme for the metabolic pathway that supplies NADPH to cells (Figure $2.3.7.4$).

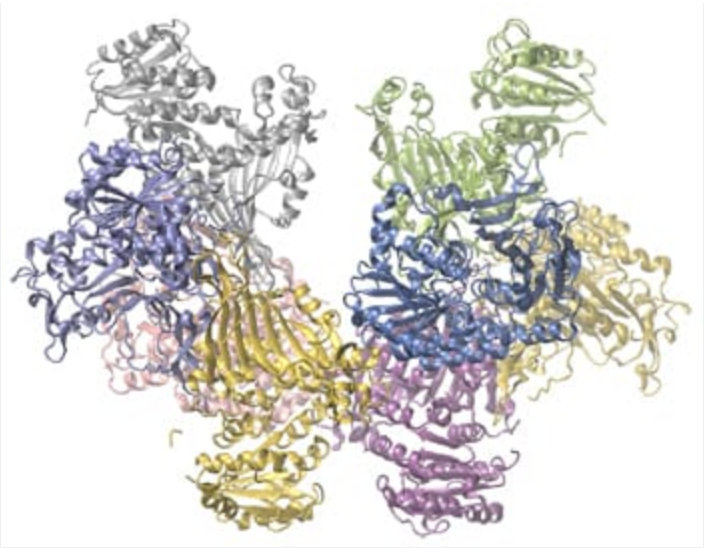


Figure 12.21 Glucose-6-phosphate dehydrogenase is a rate-limiting enzyme for the metabolic pathway that supplies NADPH to cells.

A disruption in this pathway can lead to reduced glutathione in red blood cells; once all glutathione is consumed, enzymes and other proteins such as hemoglobin are susceptible to damage. For example, hemoglobin can be metabolized to bilirubin, which leads to jaundice, a condition that can become severe. People who suffer from G6PD deficiency must avoid certain foods and medicines containing chemicals that can trigger damage their glutathione-deficient red blood cells.

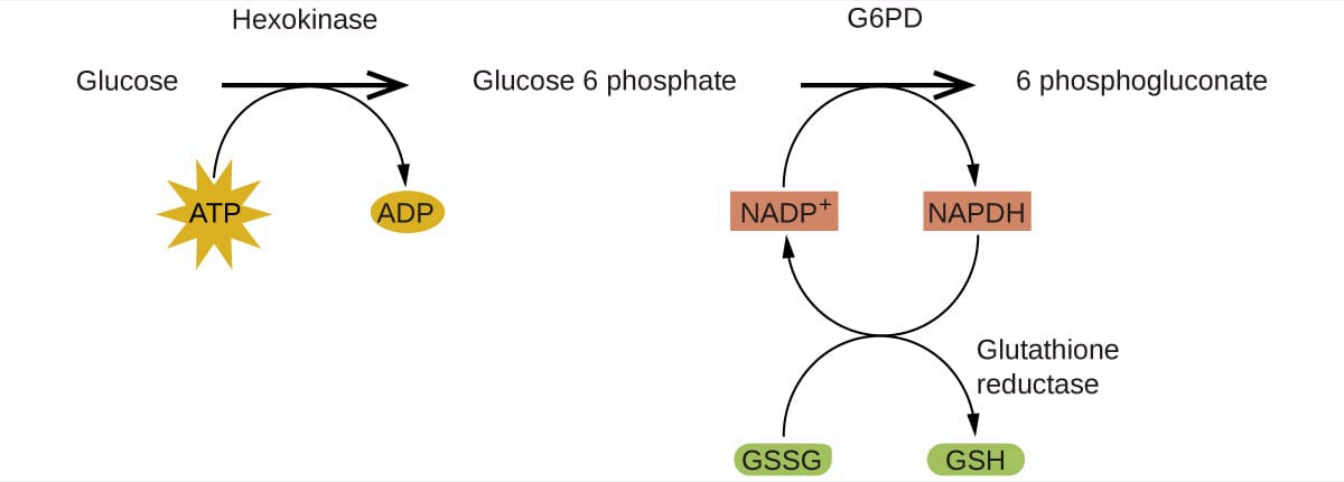


Figure 2.3.7.4: In the mechanism for the pentose phosphate pathway, G6PD catalyzes the reaction that regulates NADPH, a co-enzyme that regulates glutathione, an antioxidant that protects red blood cells and other cells from oxidative damage.

Heterogeneous Catalysts

A heterogeneous catalyst is a catalyst that is present in a different phase (usually a solid) than the reactants. Such catalysts generally function by furnishing an active surface upon which a reaction can occur. Gas and liquid phase reactions catalyzed by heterogeneous catalysts occur on the surface of the catalyst rather than within the gas or liquid phase.

Heterogeneous catalysis typically involves the following processes:

1. Adsorption of the reactant(s) onto the surface of the catalyst
2. Activation of the adsorbed reactant(s)
3. Reaction of the adsorbed reactant(s)

Figure 2.3.7.5: illustrates the steps of a mechanism for the reaction of compounds containing a carbon–carbon double bond with

hydrogen on a nickel catalyst. Nickel is the catalyst used in the hydrogenation of polyunsaturated fats and oils (which contain several carbon–carbon double bonds) to produce saturated fats and oils (which contain only carbon–carbon single bonds).

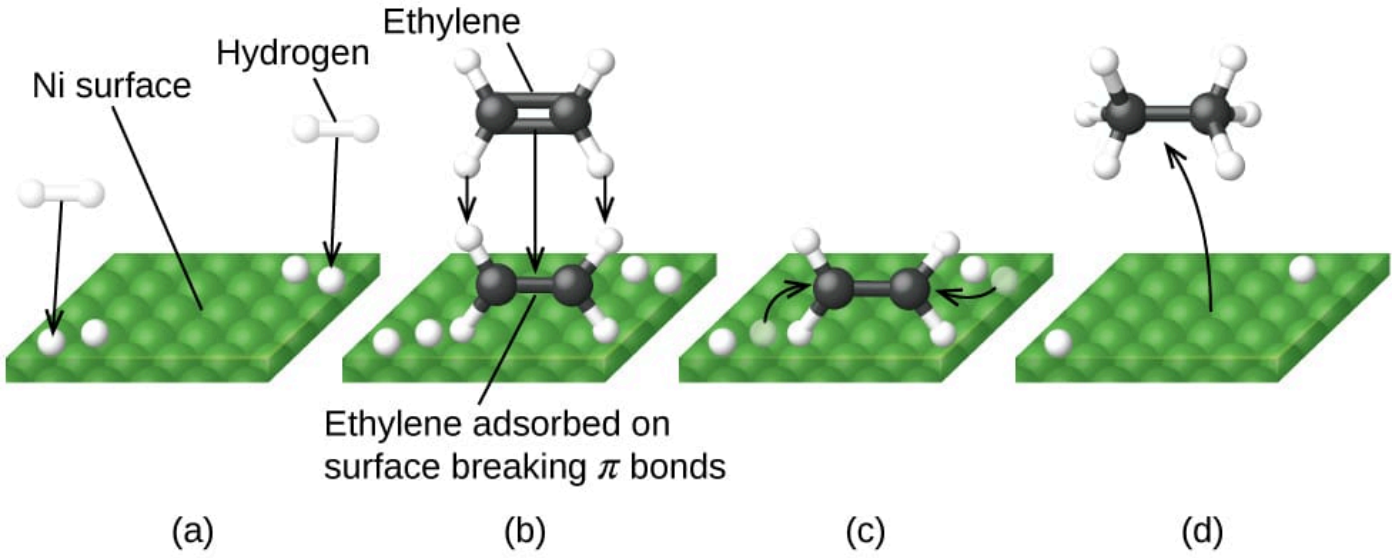


Figure 2.3.7.5: Mechanism for the Ni-catalyzed reaction $C_2H_4 + H_2 \longrightarrow C_2H_6$. (a) Hydrogen is adsorbed on the surface, breaking the H–H bonds and forming Ni–H bonds. (b) Ethylene is adsorbed on the surface, breaking the C–C π-bond and forming Ni–C bonds. (c) Atoms diffuse across the surface and form new C–H bonds when they collide. (d) C_2H_6 molecules desorb from the Ni surface.

Many important chemical products are prepared via industrial processes that use heterogeneous catalysts, including ammonia, nitric acid, sulfuric acid, and methanol. Heterogeneous catalysts are also used in the catalytic converters found on most gasoline-powered automobiles (Figure 2.3.7.6).

Chemistry in Everyday Life: Automobile Catalytic Converters

Scientists developed catalytic converters to reduce the amount of toxic emissions produced by burning gasoline in internal combustion engines. By utilizing a carefully selected blend of catalytically active metals, it is possible to effect complete combustion of all carbon-containing compounds to carbon dioxide while also reducing the output of nitrogen oxides. This is particularly impressive when we consider that one step involves adding more oxygen to the molecule and the other involves removing the oxygen (Figure 2.3.7.6).

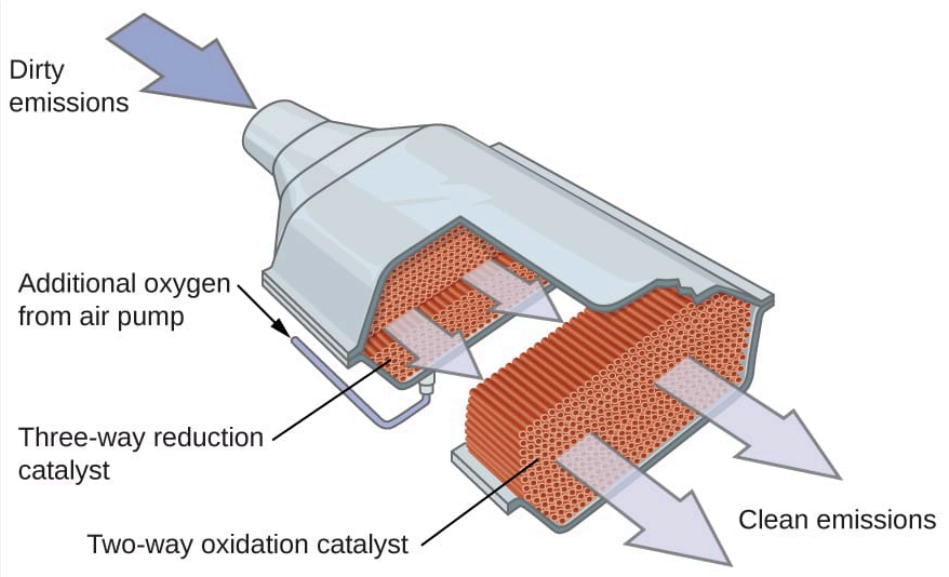


Figure 2.3.7.6: A catalytic converter allows for the combustion of all carbon-containing compounds to carbon dioxide, while at the same time reducing the output of nitrogen oxide and other pollutants in emissions from gasoline-burning engines.

Most modern, three-way catalytic converters possess a surface impregnated with a platinum-rhodium catalyst, which catalyzes the conversion of nitric oxide into dinitrogen and oxygen as well as the conversion of carbon monoxide and hydrocarbons such as octane into carbon dioxide and water vapor:

$$2NO_2(g) \longrightarrow N_2(g) + 2O_2(g)$$
$$2CO(g) + O_2(g) \longrightarrow 2CO_2(g)$$
$$2C_8H_{18}(g) + 25O_2(g) \longrightarrow 16CO_2(g) + 18H_2O(g)$$

In order to be as efficient as possible, most catalytic converters are preheated by an electric heater. This ensures that the metals in the catalyst are fully active even before the automobile exhaust is hot enough to maintain appropriate reaction temperatures.

Link to Learning

The LibreTexts provides a thorough explanation of how catalytic converters work.

How Sciences Interconnect: Enzyme Structure and Function

The study of enzymes is an important interconnection between biology and chemistry. Enzymes are usually proteins (polypeptides) that help to control the rate of chemical reactions between biologically important compounds, particularly those that are involved in cellular metabolism. Different classes of enzymes perform a variety of functions, as shown in Table 2.3.7.1.

Table 2.3.7.1: Classes of Enzymes and Their Functions

Class	Function
oxidoreductases	redox reactions
transferases	transfer of functional groups
hydrolases	hydrolysis reactions
lyases	group elimination to form double bonds

Class	Function
isomerases	isomerization
ligases	bond formation with ATP hydrolysis

Enzyme molecules possess an active site, a part of the molecule with a shape that allows it to bond to a specific substrate (a reactant molecule), forming an enzyme-substrate complex as a reaction intermediate. There are two models that attempt to explain how this active site works. The most simplistic model is referred to as the lock-and-key hypothesis, which suggests that the molecular shapes of the active site and substrate are complementary, fitting together like a key in a lock. The induced fit hypothesis, on the other hand, suggests that the enzyme molecule is flexible and changes shape to accommodate a bond with the substrate. This is not to suggest that an enzyme's active site is completely malleable, however. Both the lock-and-key model and the induced fit model account for the fact that enzymes can only bind with specific substrates, since in general a particular enzyme only catalyzes a particular reaction (Figure 2.3.7.7).

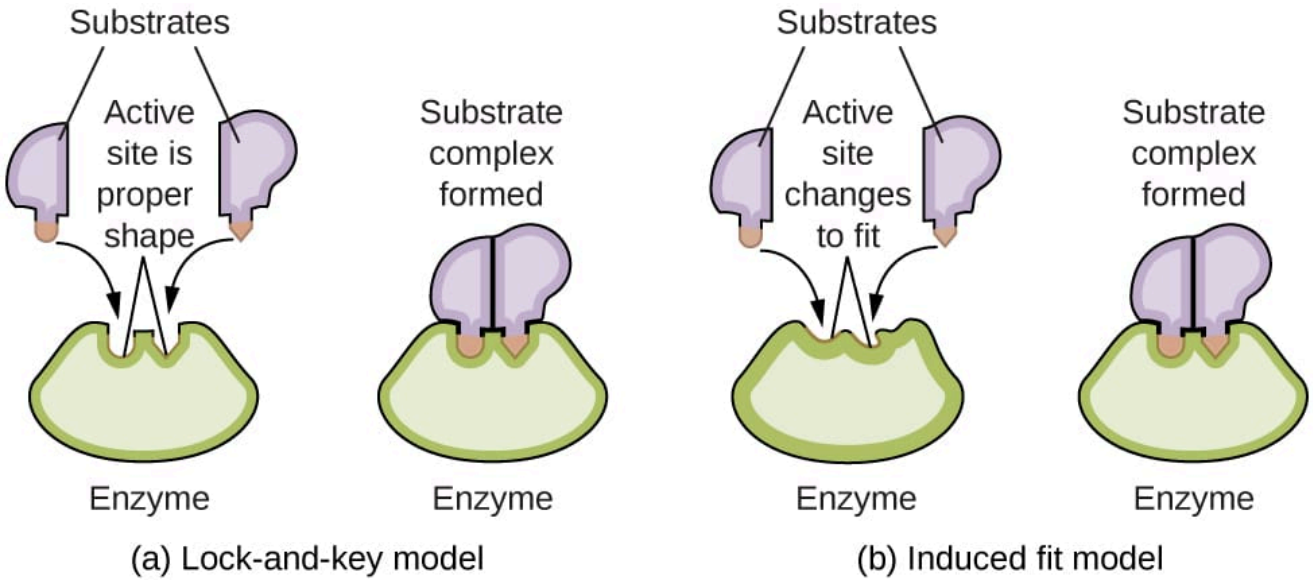


Figure 2.3.7.7: (a) According to the lock-and-key model, the shape of an enzyme's active site is a perfect fit for the substrate. (b) According to the induced fit model, the active site is somewhat flexible, and can change shape in order to bond with the substrate.

Link to Learning

The Royal Society of Chemistry provides an excellent introduction to enzymes for students and teachers.

Footnotes

- 1"The Nobel Prize in Chemistry 1995," Nobel Prize.org, accessed February 18, 2015, http://www.nobelprize.org/nobel_priz...aureates/1995/.

Answers

Exercise 2.3.7.1

For the first step, E_a = 80 kJ for (a) and 70 kJ for (b), so diagram (b) depicts the catalyzed reaction. Activation energies for the second steps of both mechanisms are the same, 20 kJ.

This page titled 2.3.7: Catalysis is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 12.7: Catalysis by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/chemistry-2e.

2.4: Using logarithms - Log vs. Ln

A common question exists regarding the use of logarithm base 10 ($\log$ or $\log_{10}$) vs. logarithm base e ($\ln$). The logarithm base e is called the **natural logarithm** since it arises from the integral:

$$\ln(a) = \int_1^a \frac{dx}{x}$$

Of course, one can convert from $\ln$ to $\log$ with a constant multiplier.

$$\ln(10^{\log a}) = \log(a)\ln(10) \approx 2.3025\log(a)$$

but $10^{\log a} = a$ so

$$\ln a \approx 2.4025 \log a$$

The analysis of the reaction order and rate constant using the method of initial rates is performed using the $\log_{10}$ function. This could have been done using the $\ln$ function just as well. The initial rate is given by

$$r_o = k'[A]_0^a$$

The analysis can proceed by taking the logarithm base 10 of each side of the equation

$$\log r_o = \log k' + a \log [A]_0$$

or the $\ln$ of each side of the equation

$$\ln r_o = \ln k' + a \ln [A]_0$$

as long as one is **consistent**.

Once can think of the $\log$ or the $\ln$ as a way to '*linearize data*' that has some kind of power law dependence. The only difference between these two functions is a scaling factor ($\ln 10 \approx 2.3025$) in the slope. Obviously, if you multiply both sides of the equation by the same number the relative values of the constants remains the same on both sides.

Contributors and Attributions

- Stefan Franzen (North Carolina State University)

This page titled 2.4: Using logarithms - Log vs. Ln is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 10: Using logarithms - Log vs. Ln is licensed CC BY-NC-SA 4.0.

3: THE FIRST LAW OF THERMODYNAMICS

3.1: Overview of Classical Thermodynamics

3.2: Pressure-Volume Work

3.3: Work and Heat are not State Functions

3.4: Energy is a State Function

3.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred

3.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion

3.7: Pressure-Volume Work

3.8: Heat Capacity is a Path Function

3.9: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition

3.10: Enthalpy Changes for Chemical Equations are Additive

3.11: Heats of Reactions Can Be Calculated from Tabulated Heats of Formation

3.12: The Temperature Dependence of ΔH

3.13: Enthalpy is a State Function

3.E: The First Law of Thermodynamics (Exercises)

This page titled 3: The First Law of Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19: The First Law of Thermodynamics is licensed CC BY-NC-SA 4.0.

3.1: Overview of Classical Thermodynamics

This page titled 3.1: Overview of Classical Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

3.2: Pressure-Volume Work

Work in general is defined as a product of a force $\mathbf{F}$ and a path element $\mathbf{ds}$. Both are vectors and work is computed by integrating over their inner product:

$$w = \int \mathbf{F}\cdot\mathbf{ds}$$

Moving an object against the force of friction as done in the above dissipation experiment is but one example of work:

$$w_{friction} = \int \mathbf{F}_{\text{friction}}\cdot\mathbf{ds}$$

We could also think of *electrical* work. In that case we would be moving a charge *e* (e.g. the negative charge of an electron) against an electrical (vector) field $\mathbf{E}$. The work would be:

$$w_{electical} = \int e\mathbf{E}\cdot\mathbf{ds}$$

Other examples are the stretching of a rubber band against the elastic force or moving a magnet in a magnetic field etc, etc.

Pressure-volume (PV) work

In the case of a cylinder with a piston, the pressure of gas molecules on the inside of the cylinder, P, and the gas molecules external to the piston, P_{ext} both exert a force against each other. Pressure, (\P\), is the force, F, being exerted by the particles per area, A:

$$P = \frac{F}{A}$$

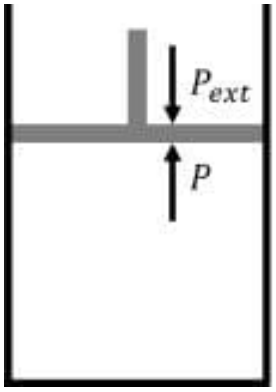


Figure 19.2.1 : A piston with internal pressure P and external pressure P_{ext}.

We can assume that all the forces generated by the pressure of the particles operate parallel to the direction of motion of the piston. That is, the force moves the piston up or down as the movement of the piston is constrained to one direction. The piston moves as the molecules of the gas rapidly equilibrate to the applied pressure such that the internal and external pressures are the same. The result of this motion is work:

$$w_{volume} = \int \left(\frac{F}{A}\right)(A\,ds) = \int P\,dV \tag{3.2.1}$$

This particular form of work is called **pressure-volume (PV) work** and will play an important role in the development of our theory. Notice however that volume work is only ***one form*** of work.

Sign Conventions

It is important to create a sign convention at this point: positive heat, positive work is always energy you put in into the system. If the system decides to remove energy by giving off heat or work, that gets a minus sign.

In other words: ***you pay the bill***.

To comply with this convention we need to rewrite volume work (Equation 3.2.1) as

$$w_{PV} = -\int \left(\frac{F}{A}\right)(A\,ds) = -\int P\,dV$$

Hence, to decrease the volume of the gas (ΔV is negative), we must put in (positive) work.

Thermodynamics would not have come very far without cylinders to hold gases, in particular steam. The following figure shows when the external pressure, P_{ext}, is greater than and less than the internal pressure, P, of the piston.

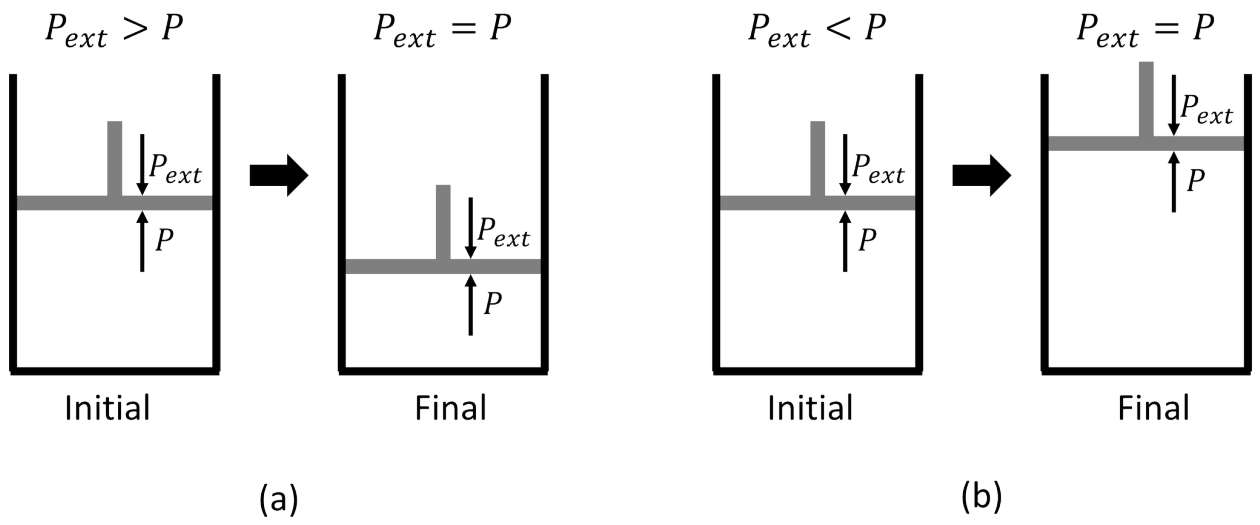


Figure 19.2.2 : Pistons showing a compression (left) and an expansion (right).

If the pressure, P_{ext}, being exerted on the system is constant, then the integral becomes:

$$w = -P_{\text{ext}} \int_{V_{initial}}^{V_{final}} dV = -P_{\text{ext}} \Delta V \tag{3.2.2}$$

Since the system pressure (inside the piston) is not the same as the pressure exerted on the system, the system is not in a state of equilibrium and cannot be shown directly on and PV diagram. This type of process is called an irreversible process. For a system that undergoes irreversible work at constant external pressure, we can show the amount of work being done on a PV diagram despite not being able to show the process itself.

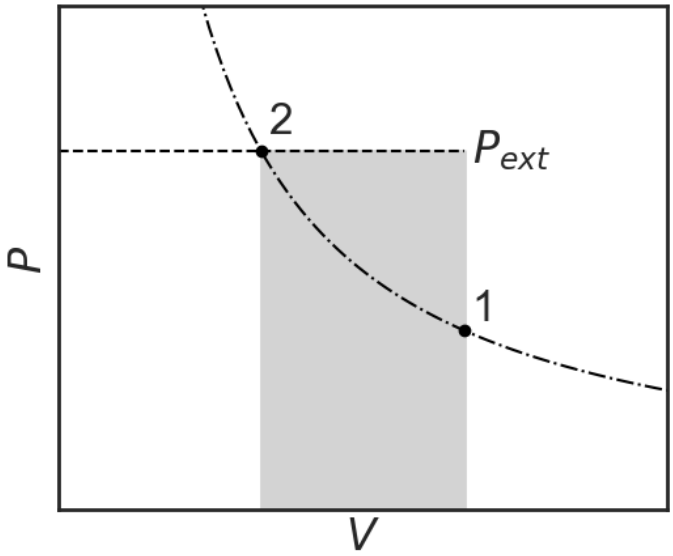


Figure 19.2.3 : A system is compressed under constant external pressure, P_{ext}, from state 1 to state 2. The shaded area shows the amount of work being done for the compression. The dash-dot line is an isotherm, a path of constant temperature, showing that the initial and final temperature of the irreversible compression are the same.

Note that the external pressure, P_{ext}, exerted on the system is constant. If the external pressure changes during the compression, we must *integrate* over the whole range:

$$w = -\int_{V_{initial}}^{V_{final}} P_{\text{ext}}(V)\, dV$$

This page titled 3.2: Pressure-Volume Work is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Jerry LaRue.

- 19.2: Pressure-Volume Work by Jerry LaRue is licensed CC BY 4.0.

3.3: Work and Heat are not State Functions

Heat and work are path functions

Heat (q) and work (w) are path functions, not state functions:

1. They are path dependent.
2. They are energy transfer → they are not intrinsic to the system.

Path Functons

Functions that depend on the path taken, such as work (w) and heat (q), are referred to as **path functions**.

Reversible versus irreversible

Let's consider a piston that is being compressed at constant temperature (isothermal) to half of its initial volume:

1. Start with cylinder 1 liter, both external and internal pressure 1 bar.
2. ***Peg*** the piston in a fixed position.
3. Put cylinder in a pressure chamber with $P_{ext} = 2$ bar.
4. Suddenly pull the peg.

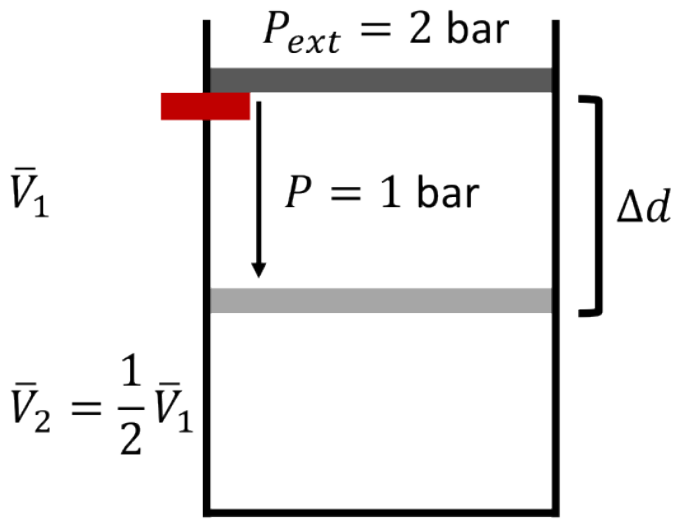


Figure 19.3.1 : A piston is being irreversibly and isothermally compressed under constant external pressure from state 1 to state 2.

The piston will shoot down till the internal and external pressures balance out again and the volume is 1/2 L. Notice that the external pressure was ***maintained*** constant at 2 bar during the peg-pulling and that the internal and external pressures were ***not*** balanced at all time. In a $P - V$ diagram of an ideal gas, P is a hyperbolic function of V under constant temperature (isothermal), but this refers to the *internal* pressure of the gas. It is the external one that counts when computing work and they are not necessarily the same. As long as $P_{external}$ is constant, work is represented by a rectangle.

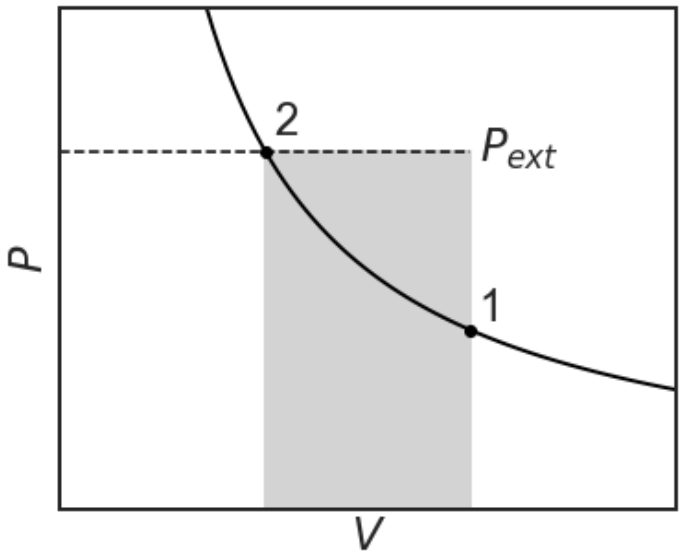


Figure 19.3.1 : The system is being irreversibly and isothermally compressed from state 1 to 2. The amount of work being done on the system is shown by the shaded area.

The amount of work being done is equal to the shaded region and in equation:

$$w = -\int_{V_1}^{V_2} PdV = -P_{ext}(V_2 - V_1) = -P\Delta V$$

This represents the maximum amount of work that can be done for an isothermal compression. Work is being done on the system, so the overall work being done is positive. Let's repeat the experiment, but this time the piston will compress reversibly over infinitesimally small steps where the $P_{ext} = P_{system}$:

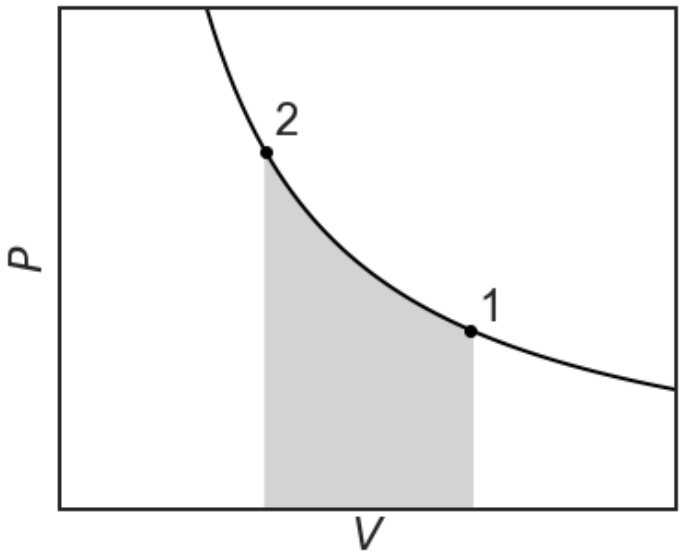


Figure 19.3.1 : The system is being reversibly and isothermally compressed from state 1 to 2. The amount of work being done on the system is shown by the shaded area. Note that for a compression, a reversible process does less work than an irreversible process.

For an ideal gas, the amount of work being done along the reversible compression is:

$$w = -\int_{V_1}^{V_2} PdV = -nRT\int_{V_1}^{V_2} \frac{1}{V} = -nRT\ln\left(\frac{V_2}{V_1}\right)$$

The amount of work being done to the two systems are not the same in the two diagrams (see the gray areas). Work is not a state, but a path function, as it depends on the path taken. You may say, what's the big difference. In both cases, the system is compressed from state 1 to state 2. The key is the word ***suddenly***. By pegging the position in place for the first compression, we have created a situation where the external pressure is higher than the internal pressure ($P_{ext} > P$). Because work is done suddenly by pulling the peg out, the internal pressure is struggling to catch up with the external one. During the second compression, we have $P_{ext} = P$ at all times. It's a bit like falling off a cliff versus gently sliding down a hill. Path one is called an irreversible path, the second a reversible path.

Reversible vs. Irreversible Processes

A reversible path is a path that follows a series of states at rest (i.e., the forces are allowed to balance at all times). In an irreversible one the forces only balance at the very end of the process.

Notice that less work is being done on the reversible isothermal compression than the one-step irreversible isothermal compression. In fact, the minimum amount of work that can be done during a compression always occurs along the reversible path.

Isothermal Expansion

Let's consider a piston that is being expanded at constant temperature (isothermal) to twice of its initial volume:

1. Start with cylinder 1 liter in a pressure chamber with both an external and internal pressure of 2 bar.
2. ***Peg*** the piston in a fixed position.
3. Take the cylinder out of the pressure chamber with P_{ext}= 1 bar.
4. Suddenly pull the peg.

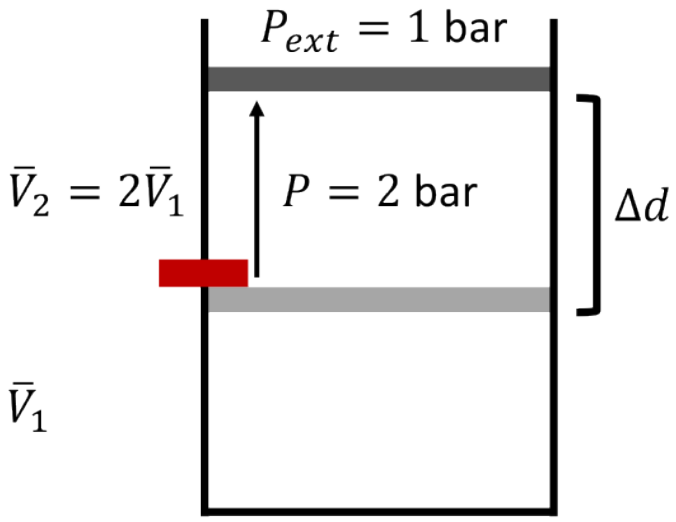


Figure 19.3.1 : A piston is being irreversibly and isothermally expanded under constant external pressure from state 1 to state 2.

The piston will shoot up till the internal and external pressures balance out again and the volume is 2 L. Notice that the external pressure was ***maintained*** constant at 1 bar during the peg-pulling and that the internal and external pressures were ***not*** balanced at all time.

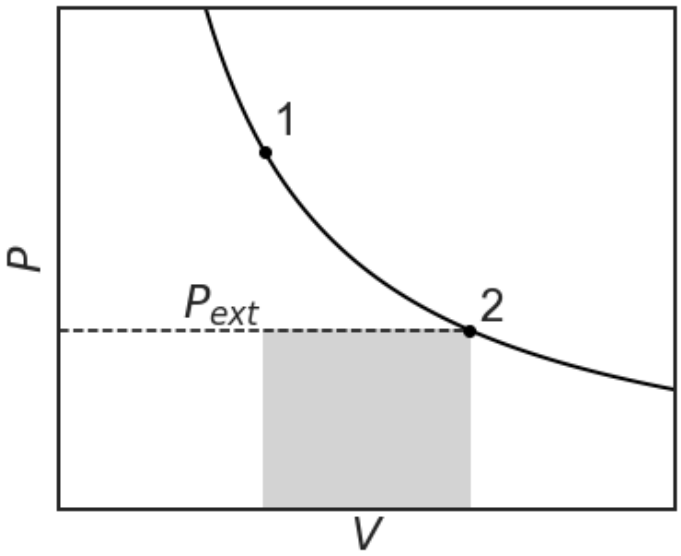


Figure 19.3.1 : The system is being irreversibly and isothermally expanded from state 1 to 2. The amount of work being done by the system is shown by the shaded area.

The amount of irreversible work being done is again equal to the shaded region and the equation:

$$w = -P\Delta V = -P_{ext}(V_2 - V_1) = -P\Delta V$$

This represents the minimum amount of work that can be done for an isothermal expansion. Work is being done on the system, so the overall work being done is negative. Let's repeat the experiment, but this time the piston will compress reversibly over infinitesimally small steps where the $P_{ext} = P_{system}$:

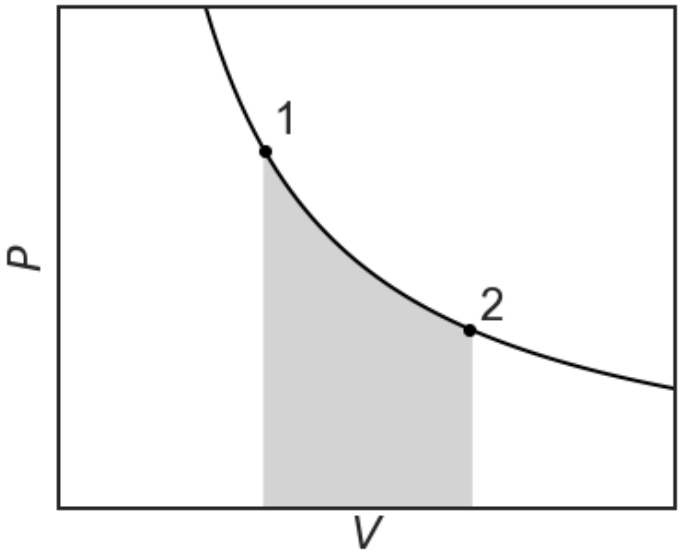


Figure 19.3.1 : The system is being reversibly and isothermally expanded from state 1 to 2. The amount of work being done by the system is shown by the shaded area. Note that for an expansion, a reversible process does more work than an irreversible process.

Notice that not only is more work is being done than the one-step irreversible isothermal expansion, but it is the same amount of work being done as the reversible isothermal compression. This is the maximum amount of work that can be done during an expansion.

This page titled 3.3: Work and Heat are not State Functions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.3: Work and Heat are not State Functions is licensed CC BY-NC-SA 4.0.

3.4: Energy is a State Function

Work and heat are not state functions

A Better Definition of the First law of thermodynamics

The change in internal energy of a system is the sum of w and q, which is a state function.

The realization that work and heat are both forms of energy transfer undergoes quite an extension by saying that internal energy is a state function. It means that although heat and work can be produced and destroyed (and transformed into each other), energy is conserved. This allows us to do some serious bookkeeping! We can write the law as:

$$\Delta U = w + q$$

But the (important!) bit about the state function is better represented if we talk about small changes of the energy:

$$dU = \delta w + \delta q$$

We write a straight Latin d for U to indicate when the change in a state function, where as the changes in work and heat are path-dependent. This is indicated by the 'crooked' δ. We can represent changes as integrals, but only for U can we say that regardless of path we get $\Delta U = U_2 - U_1$ if we go from state one to state two. (I.e. it only depends on the end points, not the path).

Notice that when we write dU or δq, we always mean infinitesimally small changes, i.e. we are implicitly taking a limit for the change approaching zero. To arrive at a macroscopic difference like ΔU or a macroscopic (finite) amount of heat q or work w **we need to integrate**.

We will now invoke the **first law of thermodynamics**:

- $dU = \delta q + \delta w$
- $\oint dU = 0$
- Internal energy is conserved

These are all ways of saying that internal energy is a state function.

This page titled 3.4: Energy is a State Function is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.4: Energy is a State Function is licensed CC BY-NC-SA 4.0.

3.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred

Isothermal expansion of an ideal gas

For a monatomic ideal gas we have seen that energy $\langle E \rangle$ observed as $U = 3/2nRT$. This means that energy is only dependent on temperature and if a gas is compressed **isothermally**, then the internal energy does not change:

$$\Delta U_{isothermal-idealgas} = 0$$

This means that the reversible work must ***cancel*** the reversible heat:

$$\Delta U_{rev} = w_{rev} + q_{rev} = 0$$

Therefore

$$w_{rev} = -q_{rev}$$

so from the expression of the reversible work for expansion in the last section

$$q_{rev} = nRT \ln \frac{V_2}{V_1}$$

If $V_2 > V_1$ (expansion), then you (or the environment) **must** put heat into the system because this is a positive number.

Adiabatic expansion of an ideal gas

Now suppose you make sure that no heat can enter the cylinder. (Put it in styrofoam or so). Then the path can still be reversible (slow pulling) but the process is then adiabatic.

> This bat- part comes from a Greek verb *βαινω* (baino) that means walking, compare acro**bat**, someone who goes high places (acro-). The *δια* (dia) part means 'through' (cf. diagram, diorama, diagonal etc.) and the prefix *α*- (a-) denies it all (compare *atypical* versus *typical*).

So the styrofoam prevents the heat from walking through the wall. When expanding the gas from V_1 to V_2 it still does reversible work but where does that come from? It can only come from the internal energy itself. So in this case any energy change should consist of work (adiabatic means: $\delta q = 0$).

$$dU = \delta w_{rev}$$

This implies that the temperature must drop, because if U changes, then T must change.

The change of energy with temperature at constant volume is known as the heat capacity (at constant volume) C_v

$$C_v = \left(\frac{\partial U}{\partial T} \right)_V$$

For an ideal gas U only changes with temperature, so that

Image:CH431_Image61.gif

or:

Image:CH431_Image62.gif

We can now compare two paths to go from state P_1, V_1, T_1 to state P_2, V_2, T_1:

1. Reversible isothermal expansion A
2. Reversible adiabatic expansion B followed by reversible isochoric heating C

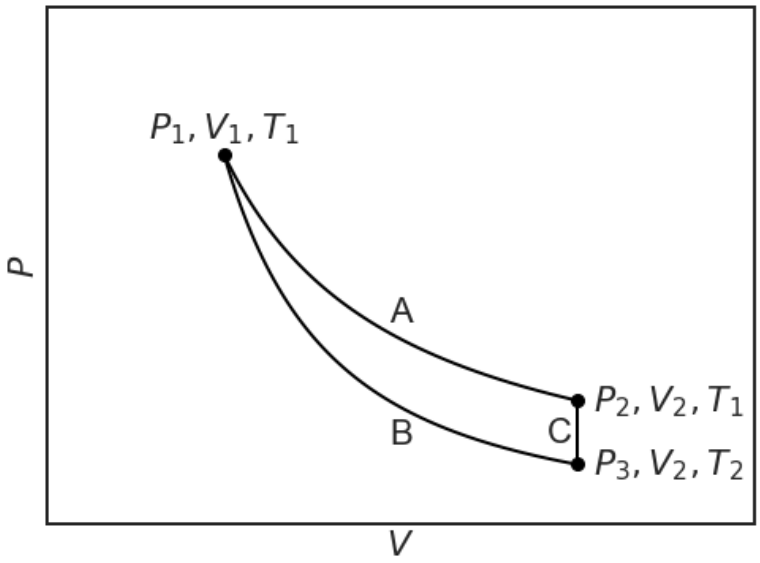


Figure __:

Notice that the temperature remains T_1 for path A (isotherm!), but that it drops to T_2 on the adiabat B, so that the cylinder has to be isochorically warmed up, C, to regain the same temperature.

ΔU_{tot} should be the same for both path A and the combined path B+C, because the end points are the same (U is a state function!). As the and points are at the same temperature and U only depends on T:

$$\Delta U_{tot} = 0$$

Along adiabat B:

$$q_{rev} = 0$$

Along isochoric heating C, there is no volume work because the volume is kept constant, so that:

Image:CH431_Image63.gif

This is the only reversible heat involved in path B+C. However, we know that ΔU_{tot} for path A is zero (isothermal!). This means that the volume work along B must cancel the heat along C:

The book keeping looks as follows, all paths are reversible:

$$\Delta U_{B+C} = \Delta U_A = 0 = q_B + w_B + q_C + w_C$$

We know that $q_B = 0$ since it is an adiabat and $w_C = 0$ since it is an isochore:

$$\Delta U_{B+C} = \Delta U_A = 0 = 0 + w_B + q_C + 0$$

Therefore:

$$w_B = -q_C$$

We had already seen before that along the isotherm A:

$$w_A = -q_A = -nRT \ln \frac{V_1}{V_2}$$

As you can see \w_A\) and w_B are ***not*** the same. Work is a path function, even if reversible. As we are working with an ideal gas

we can be more precise about w_B and q_c as well. The term w_B along the adiabat is reversible volume work. Since there is no heat along B we can write a straight d instead of δ for the work contributions (It is the only contribution and must be identical to the state function dU):

Image:CH431_Image64.gif

This page titled 3.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred is licensed CC BY-NC-SA 4.0.

3.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion

We can make the same argument for the heat along C. If we do the three processes A and B+C only to a tiny extent we can write:

Image:CH431_Image65.gif

And now we can integrate from V_1 to V_2 over the reversible adiabatic work along B and from T_1 to T_2 for the reversible isochoric heat along C. To separate the variables we do need to bring the temperature to the right side of the equation.:

Image:CH431_Image66.gif

The latter expression is valid for a reversible adiabatic expansion of a monatomic ideal gas (say Argon) because we used the C_v expression for such a system. We can use the gas law $PV = nRT$ to translate this expression in one that relates pressure and volume see Eq 19.23

We can mathematically show that the temperature of a gas decreases during an adiabatic expansion. Assuming an ideal gas, the internal energy along an adiabatic path is:

$$\begin{aligned} d\bar{U} &= \delta q + \delta w \\ &= 0 - Pd\bar{V} \\ &= -Pd\bar{V} \end{aligned}$$

The constant volume heat capacity is defined as:

$$\bar{C}_V = \left(\frac{\partial \bar{U}}{\partial T}\right)_V$$

We can rewrite this for internal energy:

$$d\bar{U} = \bar{C}_V dT$$

Combining these two expressions for internal energy, we obtain:

$$\bar{C}_V dT = -Pd\bar{V}$$

Using the ideal gas law for pressure of an ideal gas:

$$\bar{C}_V dT = -\frac{RT}{\bar{V}} d\bar{V}$$

Separating variables:

$$\frac{\bar{C}_V}{T} dT = -\frac{R}{\bar{V}} d\bar{V}$$

This is an expression for an ideal path along a reversible, adiabatic path that relates temperature to volume. To find our path along a PV surface for an ideal gas, we can start in TV surface and convert to a PV surface. Let's go from (T_1, V_1) to (T_2, V_2).

$$\int_{T1}^{T_2} \frac{\bar{C}_V}{T} dT = -\int_{\bar{V}_1}^{\bar{V}_2} \frac{R}{\bar{V}} d\bar{V}$$

$$\bar{C}_V \ln\left(\frac{T_2}{T_1}\right) = -R \ln\left(\frac{\bar{V}_2}{\bar{V}_1}\right) = R \ln\left(\frac{\bar{V}_1}{\bar{V}_2}\right)$$

$$\ln\left(\frac{T_2}{T_1}\right) = \frac{R}{\bar{C}_V}\ln\left(\frac{\bar{V}_1}{\bar{V}_2}\right)$$

$$\left(\frac{T_2}{T_1}\right) = \left(\frac{\bar{V}_1}{\bar{V}_2}\right)^{\frac{R}{\bar{C}_V}}$$

We know that:

$$R = \bar{C}_P - \bar{C}_V$$

$$\frac{R}{\bar{C}_V} = \frac{\bar{C}_P - \bar{C}_V}{\bar{C}_V} = \frac{\bar{C}_P}{\bar{C}_V} - 1$$

$$\frac{R}{\bar{C}_V} = \gamma - 1$$

Therefore:

$$\left(\frac{T_2}{T_1}\right) = \left(\frac{\bar{V}_1}{\bar{V}_2}\right)^{\gamma-1}$$

This expression shows that volume and temperature are inversely related. That is, as the volume increase from V_1 to V_2, the temperature must decrease from T_1 to T_2.

This page titled 3.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion is licensed CC BY-NC-SA 4.0.

3.7: Pressure-Volume Work

Enthalpy

An important point is that pressure-volume work $-PdV$ is only one kind of work. It is the important one for gases but for most other systems we are interested in other kinds of work (e.g. electrical work in a battery).

A good way to measure ΔU's is to make sure there are no work terms at all. If so:

$$\Delta U_{nowork} = q + w = q + 0 = q$$

However, this means that the $-PdV$ volume work term should also be zero and this implies we must keep volumes the same. That can actually be hard. Therefore we define at new state function ENTHALPY

$$H \equiv U + PV$$

(The $\equiv$ symbol is used to show that this equality is actually a definition.)

If we differentiate we get:

$$dH = dU + d(PV) = dU + PdV + VdP$$

We know that under reversible conditions we have

$$dU = \delta w + \delta q = -PdV + \delta q$$

(+ other work terms that we assume zero)

Thus,

$$dH = -PdV + \delta q + PdV + VdP$$

$$dH = \delta q + VdP$$

That means that as long as there is no other work and we keep the pressure constant:

$$\Delta H = q_P$$

instead of

$$\Delta U = q_V$$

Working at constant P is a lot easier to do than at constant V. This means that the enthalpy is a much easier state function to deal with than the energy U.

For example when we melt ice volumes change whether we like or not, but at long as the weather does not change too much pressure is constant. So if we measure how much heat we need to add to melt a mole of ice we get the molar heat of fusion:

Image:CH431 Image71.gif

Such enthalpies are measured and tabulated.

In this case the volume change is actually quite small, as it usually is for condensed matter. Only if we are dealing with gases is the difference between enthalpy and energy really important

So, U H for condensed matter, but U and H differ for gases.

A good example of this is the difference between the heat capacity at constant V and at constant P. For most materials there is

not much of a difference, but for an ideal gas we have

$$C_p = C - V + nR$$

Needless to say that the heat capacity is a path function: it depends on what you keep constant.

This page titled 3.7: Pressure-Volume Work is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.8: Pressure-Volume Work is licensed CC BY-NC-SA 4.0.

3.8: Heat Capacity is a Path Function

Determining enthalpies from heat capacities.

The functions H and C_p are related by differentiation:

$$\left(\frac{\partial H}{\partial T}\right)_P = C_P$$

This means that we can:

1. measure C_p as a function of temperature
2. integrate this function and find $H(T)$

However, there are problems with this approach:

1. Reference point: we have to deal with the lower limit of integration.
   Ideally we start at zero Kelvin (but we cannot get there), but how do we compare one compound to the other?
2. At temperatures where there is a phase transition there is a sudden jump in enthalpy. E.g. when ice melts we have to first add the heat of fusion until all ice is gone before the temperature can go up again (assuming all is done under reversible well-equilibrated conditions).
3. At the jumps in H, the C_p is infinite.

It should be stressed that there **are no absolute enthalpies**.

This page titled 3.8: Heat Capacity is a Path Function is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.9: Heat Capacity is a Path Function is licensed CC BY-NC-SA 4.0.

3.9: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition

It should be stressed that there **are no absolute enthalpies**. All that is properly defined are differences in enthalpy ΔH and these are only defined for ***processes***

When dealing with enthalpies

Define the Process

For example for the process of:

Process 1: ***heating ice from -20°C to -0°C***

We could write $\Delta H_1 = \int C_{P,ice}$. dT from T=253K to T=273K = H(273) -H(253). But before moving beyond the melting point first a different process needs to take place, that of

Process 2: ***melting***

This gives us $\Delta_{fus}H = H_{liquid} - H_{solid}$ (both at 273K!). When we heat the liquid water further to say +20°C we would have to integrate over the heat capacity of the liquid.

Process 3: ***heating water from 0°C to +20°C***

We could write

$$\Delta H_3 = \int C_{P,water} dT$$

from T=273K to T=293K = H(293)-H(273).

The total change in enthalpy between -20 and +20 would be the sum of the three enthalpy changes.

$$\Delta H_{totalprocess} = \Delta H_1 + \Delta_{fus}H + \Delta H_3$$

Of course we could consider doing the same calculation for any temperature between -20 and +20 and summarize all our results in a graph. The three processes can thus schematically be shown in Figure 19.10.1 .

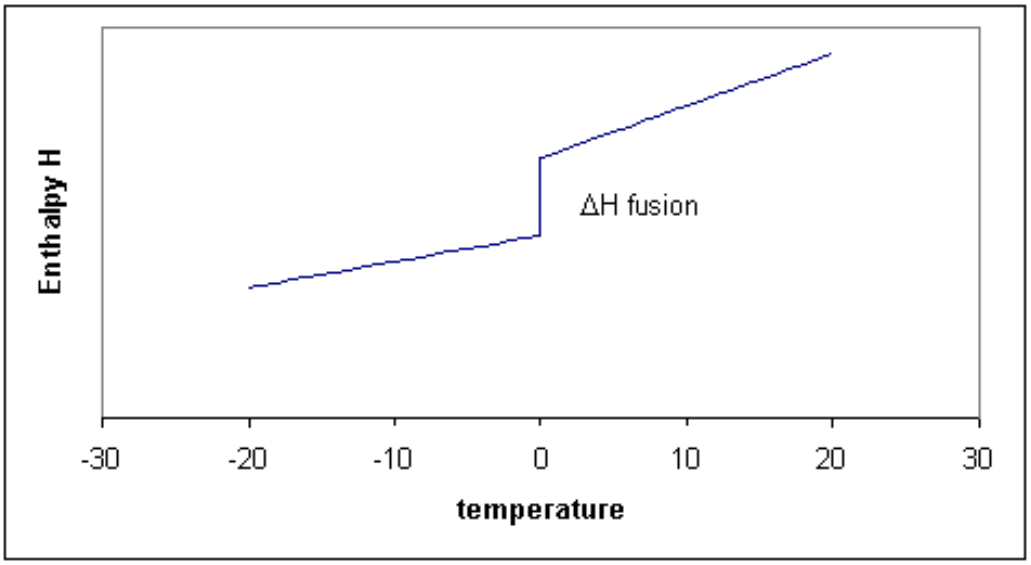


Figure 19.10.1 *: Schematic enthalpy function showing the jump at the melting point*

Notice that the slopes (i.e. the heat capacities!) before and after the melting point differ. The slope for the liquid is a little steeper because the liquid has more degrees of freedom and therefore the heat capacity of the liquid tends to be higher than of the solid. In the figure the enthlapy curves are shown as straight lines. This would be the case if the heat capacities are *constant* over the temperature interval. Although C_p is typically a 'slow' or 'weak' function of temperature it usually does change a bit, which means that the straight lines for H become curves.

Although C_p is typically a 'slow' or 'weak' function of temperature and is well approximated as a constant.

Notice that for process two, the temperature is constant, that means that ΔT or dT is zero, but ΔH is finite, consequently ΔH/ΔT is infinitly large. Taking the limit for ΔT going to zero, we get a derivative:

$$\left(\frac{\partial H}{\partial T}\right)_p = C_p$$

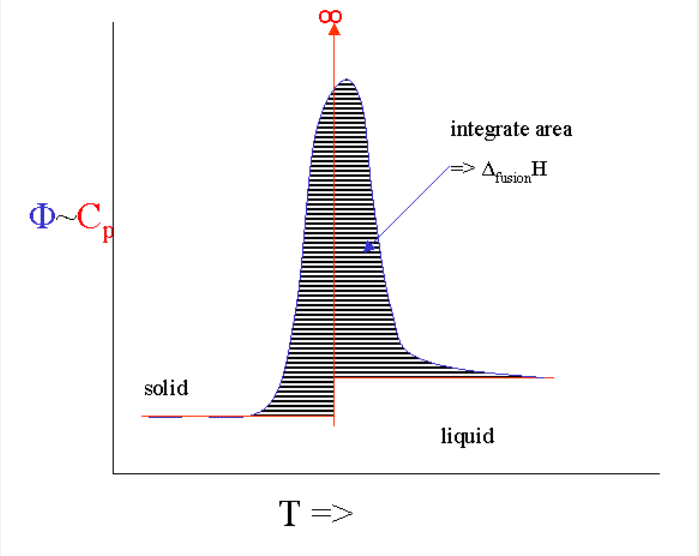
This derivative, the heat capacity must undergo a singularity: the slope is infinitely large (i.e, the H curve goes straight up). When there are more phase transitions, more discontinuities in H and singularities in C_p result (Figure 19.10.1). Note that $H(T) - H(0)$, not $H(T)$ is plotted to avoid the question what the absolute enthalpy is.

Scanningcalorimetry

There is technique that allows us to measure the heat capacity as a function of temperature fairly directly. It is called Differential Scanning Calorimetry (DSC). You put a sample in a little pan and put the pan plus an empty reference pan in the calorimeter. The instrument heats up both pans with a constant heating rate. Both pans get hotter by ***conduction***, but the heat capacity of the filled pan is obviously bigger. This means that the heat flow into the sample pan must be a bit bigger than into the empty one. This differential heat flow induces a tiny temperature difference ΔT between the two pans that can be measured. This temperature difference is proportional to the heat flow difference which is proportional to the heat capacity difference.

$$\Delta T\ \Delta\Phi\ \Delta_{betweenpans} C_p = C_p^{sample}\ (if the pans cancel)$$

However, there are number of serious broadening issues with the technique. If you melt something you will never get to see the infinite singularity of the heat capacity. Instead it broadens out into a peak. If you integrate the peak you get the $\Delta_{fusion}H$ and the onset is calibrated to give you the melting point.



The ideal heat capacity signal and its broadened DSC signal

It is even possible to heat the sample with a rate that fluctuates with a little sine wave. This "Modulated DSC" version can even give you the (small) difference in C_p before and after the melting event.

This page titled 3.9: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.10: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition is licensed CC BY-NC-SA 4.0.

3.10: Enthalpy Changes for Chemical Equations are Additive

Hess's law

As enthalpy and energy are state functions we should expect ***additivity*** of U and H when we study chemical reactions. This additivity is expressed in **Hess's Law**. The additivity has important consequences and the law finds wide spread application in the prediction of heats of reaction.

1. The reverse reaction has the negative enthalpy of the forward one.
2. If we can do a reaction in two steps we can calculate the enthalpy of the combined reaction by adding up:

Reaction	Enthalpy
$C_{(s)}$ + ½ $O_{2(g)}$ -> $CO_{(g)}$	$\Delta_r H$ = -110.5 kJ
$CO_{(g)}$ + ½ $O_{2(g)}$ -> $CO_{2(g)}$	$\Delta_r H$ = -283.0 kJ
This means that	----------+
$C_{(s)}$ + O2(g) -> CO2(g)	$\Delta_r H$ = -393.5 kJ

By this mechanism it is often possible to calculate the heat of a reaction even if this reaction is hard to carry out. E.g. we could burn both graphite and diamond and measure the heats of combustion for both. The difference would give us the heat of the transformation reaction from graphite to diamond.

Reaction-as-written convention (caution!)

The enthalpy is for the ***reaction-as-written***. That means that if we write:

$$2C(s) + O_2(g) \rightarrow 2CO(g)$$

with $\Delta_r H$ = -221 kJ (***not***: -110.5 kJ)

Reverse reactions

Because H is a state function the reverse reaction has the same enthalpy but with opposite sign

$$2CO(g) \rightarrow 2C(s) + O_2(g)$$

with $\Delta_r H$ = +221 kJ

Combining values

It is quite possible that you cannot really do a certain reaction in practice. For many reactions we can arrive at enthalpy values by doing some bookkeeping. For example, we can calculate the enthalpy for the reaction of PCl_3 with chlorine if we know the two reactions that the elements phosphorous and chlorine can undergo.

> You do have to make sure you balance your equations properly!

This page titled 3.10: Enthalpy Changes for Chemical Equations are Additive is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 19.11: Enthalpy Changes for Chemical Equations are Additive is licensed CC BY-NC-SA 4.0.

3.11: Heats of Reactions Can Be Calculated from Tabulated Heats of Formation

Reaction enthalpies are important, but difficult to tabulate. However, because enthalpy is a state function, it is possible to use **Hess' Law** to simplify the tabulation of reaction enthalpies. Hess' Law is based on the addition of reactions. By knowing the reaction enthalpy for constituent reactions, the enthalpy of a reaction that can be expressed as the sum of the constituent reactions can be calculated. The key lies in the canceling of reactants and products that °Ccur in the "data" reactions but not in the "target reaction.

Example 3.11.1:

Find ΔH_{rxn} for the reaction

$$2CO(g) + O_2(g) \rightarrow 2CO_2(g)$$

Given

$$C(gr) + ½O_2(g) \rightarrow CO(g)$$

with $\Delta H_1 = -110.53\ kJ$

$$C(gr) + O_2(g) \rightarrow CO_2(g)$$

with $\Delta H_2 = -393.51\ kJ$

Solution

The target reaction can be generated from the data reactions.

$$2\times\ [CO(g) \rightarrow C(gr) + O_2(g)]$$

plus

$$2\times\ [C(gr) + 2O_2(g) \rightarrow 2CO_2(g)]$$

equals

$$2CO(g) + O_2(g) \rightarrow 2CO_2(g)$$

so

$$2\times\Delta H_1 = -787.02\,kJ$$

$$2\times\Delta H_2 = 221.06\,kJ$$

$$2\times\Delta H_1 + 2\times\Delta H_2 = -565.96\,kJ$$

Standard Enthalpy of Formation

One of the difficulties with many thermodynamic state variables (such as enthalpy) is that while it is possible to measure changes, it is impossible to measure an absolute value of the variable itself. In these cases, it is necessary to define a zero to the scale defining the variable. For enthalpy, the definition of a zero is that the standard enthalpy of formation of a pure element in its standard state is zero. All other enthalpy changes are defined relative to this standard. Thus it is essential to very carefully define a standard state.

Definition: the Standard State

The standard state of a substance is the most stable form of that substance at 1 atmosphere pressure and the specified temperature.

Using this definition, a convenient reaction for which enthalpies can be measured and tabulated is the **standard formation reaction**. This is a reaction which forms one mole of the substance of interest in its standard state from elements in their standard states. The enthalpy of a standard formation reaction is the **standard enthalpy of formation** (ΔH_{f^o}). Some examples are

- $NaCl(s)$:

  $$Na(s) + ½Cl_2(g) \rightarrow NaCl(s)$$

  with $\Delta H_f^o = -411.2\ kJ/mol$
- $C_3H_8(g)$:

  $$3C(gr) + 4H_2(g) \rightarrow C_3H_8(g)$$

  with $\Delta H_f^o = -103.8\ kJ/mol$

It is important to note that the standard state of a substance is **temperature dependent**. For example, the standard state of water at -10 °C is solid, whereas the standard state at room temperature is liquid. Once these values are tabulated, calculating reaction enthalpies becomes a snap. Consider the heat combustion (ΔH_c) of methane (at 25 °C) as an example.

$$CH_4(g) + 2O_2(g) \rightarrow CO_2(g) + 2H_2O(l)$$

The reaction can expressed as a sum of a combination of the following standard formation reactions.

$$C(gr) + 2H_2(g) \rightarrow CH_4(g)$$

with $\Delta H_f^o = -74.6\ kJ/mol$

$$C(gr) + O_2(g) \rightarrow CO_2(g)$$

with $\Delta H_f^o = -393.5\ kJ/mol$

$$H_2(g) + ½O_2(g) \rightarrow H_2O(l)$$

with $\Delta H_f^o = -285.8\ kJ/mol$

The target reaction can be generated from the following combination of reactions

$$-1\times [C(gr) + 2H_2(g) \rightarrow CH_4(g)]$$

$$CH_4(g) \rightarrow C(gr) + 2H_2(g)$$

with $\Delta H_f^o = -1\times [-74.6\,kJ/mol] = 74.6\ kJ/mol$

$$C(gr) + O_2(g) \rightarrow CO_2(g)$$

with $\Delta H_f^o = -393.5\ kJ/mol$

$$2\times [H_2(g) + ½O_2(g) \rightarrow H_2O(l)]$$

$$2H_2(g) + O_2(g) \rightarrow 2H_2O(l)$$

with $\Delta H_f^o = 2\times [-285.8\,kJ/mol] = -571.6\ kJ/mol$.

$$CH_4(g) + 2O_2(g) \rightarrow CO2_(g) + 2H_2O(l)$$

with $\Delta H_c^o = -890.5\ kJ/mol$

Alternately, the reaction enthalpy could be calculated from the following relationship

$$\Delta H_{rxn} = \sum_{products} \nu \cdot \Delta H_f^o - \sum_{reactants} \nu \cdot \Delta H_f^o$$

where ν is the stoichiometric coefficient of a species in the balanced chemical reaction. For the combustion of methane, this calculation is

$$\begin{aligned} \Delta_{rxn} &= (1\,mol)\left(\Delta H_f^o(CO_2)\right) + (2\,mol)\left(\Delta H_f^o(H_2O)\right) - (1\,mol)\left(\Delta H_f^o(CH_4)\right) && (3.11.1)\\ &= (1\,mol)(-393.5\,kJ/mol) + (2\,mol)\,(-285.8\,kJ/mol) - (1\,mol)\,(-74.6\,kJ/mol) && (3.11.2)\\ &= -890.5\,kJ/mol && (3.11.3) \end{aligned}$$

A note about units is in order. Note that reaction enthalpies have units of kJ, whereas enthalpies of formation have units of kJ/mol. The reason for the difference is that enthalpies of formation (or for that matter enthalpies of combustion, sublimation, vaporization, fusion, etc.) refer to specific substances and/or specific processes involving those substances. As such, the total enthalpy change is scaled by the amount of substance used. General reactions, on the other hand, have to be interpreted in a very specific way. When examining a reaction like the combustion of methane

$$CH_4(g) + 2O_2(g) \rightarrow CO_2(g) + 2H_2O(l)$$

with $\Delta H_{rxn} = -890.5\ kJ$. The correct interpretation is that the reaction of one mole of CH_4(g) with two moles of O_2(g) to form one mole of CO_2(g) and two moles of H_2O(l) releases 890.5 kJ at 25 °C.

Ionization Reactions

Ionized species appear throughout chemistry. The energy changes involved in the formation of ions can be measured and tabulated for several substances. In the case of the formation of positive ions, the enthalpy change to remove a single electron at 0 K is defined as the **ionization potential**.

$$M(g) \rightarrow M^+(g) + e^-$$

with $\Delta H(0K) \equiv 1^{st}\ \text{ionization potential (IP)}$

The removal of subsequent electrons requires energies called the 2nd Ionization potential, 3rd ionization potential, and so on.

$$M^+(g) \rightarrow M^{2+}(g) + e^-$$

with $\Delta H(0K) \equiv 2^{nd} IP$

$$M^{2+}(g) \rightarrow M^{3+}(g) + e^-$$

with $\Delta H(0K) \equiv 3^{rd} IP$

An atom can have as many ionization potentials as it has electrons, although since very highly charged ions are rare, only the first few are important for most atoms.

Similarly, the **electron affinity** can be defined for the formation of negative ions. In this case, the first electron affinity is defined by

$$X(g) + e^- \rightarrow X^-(g)$$

with $-\Delta H(0K) \equiv 1^{st}\ \text{electron affinity (EA)}$

The minus sign is included in the definition in order to make electron affinities mostly positive. Some atoms (such as noble gases) will have negative electron affinities since the formation of a negative ion is very unfavorable for these species. Just as in the case of ionization potentials, an atom can have several electron affinities.

$$X^-(g) + e^- \rightarrow X^{2-}(g)$$

with $-\Delta H(0K) \equiv 2^{nd} EA$.

$$X^{2-}(g) + e^- \rightarrow X^{3-}(g)$$

with $-\Delta H(0K) \equiv 3^{rd} EA$.

Average Bond Enthalpies

In the absence of standard formation enthalpies, reaction enthalpies can be estimated using average bond enthalpies. This method is not perfect, but it can be used to get ball-park estimates when more detailed data is not available. A **bond dissociation energy** D is defined by

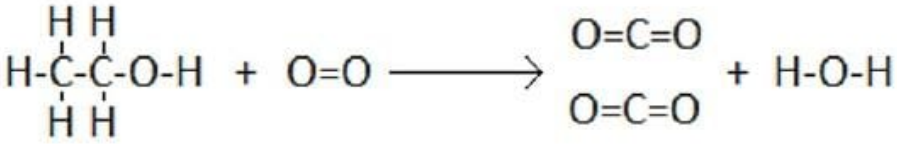
$$XY(g) \rightarrow X(g) + Y(g)$$

with $\Delta H \equiv D(X-Y)$

In this process, one adds energy to the reaction to break bonds, and extracts energy for the bonds that are formed.

$$\Delta H_{rxn} = \sum(\text{bonds broken}) - \sum(\text{bonds formed})$$

As an example, consider the combustion of ethanol:



In this reaction, five C-H bonds, one C-C bond, and one C-O bond, and one O=O bond must be broken. Also, four C=O bonds, and one O-H bond are formed.

Bond	Average Bond Energy (kJ/mol)
C-H	413
C-C	348
C-O	358
O=O	495
C=O	799
O-H	463

The reaction enthalpy is then given by

$$\begin{aligned}\Delta H_c &= 5(413\,kJ/mol) + 1(348\,kJ/mol) + 1(358\,kJ/mol) \\ &\quad + 1(495\,kJ/mol) - 4(799\,kJ/mol) - 2(463\,kJ/mol) \\ &= -856\,kJ/mol\end{aligned} \tag{3.11.4}$$

Because the bond energies are defined for gas-phase reactants and products, this method does not account for the enthalpy change of condensation to form liquids or solids, and so the result may be off systematically due to these differences. Also, since the bond enthalpies are averaged over a large number of molecules containing the particular type of bond, the results may deviate due to the variance in the actual bond enthalpy in the specific molecule under consideration. Typically, reaction enthalpies derived by this method are only reliable to within ± 5-10%.

This page titled 3.11: Heats of Reactions Can Be Calculated from Tabulated Heats of Formation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 3.6: Reaction Enthalpies by Patrick Fleming is licensed CC BY-NC-SA 4.0.

3.12: The Temperature Dependence of ΔH

It is often required to know thermodynamic functions (such as enthalpy) at temperatures other than those available from tabulated data. Fortunately, the conversion to other temperatures is not difficult.

At constant pressure

$$dH = C_p\, dT$$

And so for a temperature change from T_1 to T_2

$$\Delta H = \int_{T_2}^{T_2} C_p\, dT \tag{3.12.1}$$

Equation 3.12.1 is often referred to as *Kirchhoff's Law*. If C_p is independent of temperature, then

$$\Delta H = C_p\, \Delta T \tag{3.12.2}$$

If the temperature dependence of the heat capacity is known, it can be incorporated into the integral in Equation 3.12.1. A common *empirical* model used to fit heat capacities over broad temperature ranges is

$$C_p(T) = a + bT + \frac{c}{T^2} \tag{3.12.3}$$

After combining Equations 3.12.3 and 3.12.1, the enthalpy change for the temperature change can be found obtained by a simple integration

$$\Delta H = \int_{T_1}^{T_2} \left(a + bT + \frac{c}{T^2}\right) dT \tag{3.12.4}$$

Solving the definite integral yields

$$\Delta H = \left[aT + \frac{b}{2}T^2 - \frac{c}{T}\right]_{T_1}^{T_2} \tag{3.12.5}$$

$$= a(T_2 - T_1) + \frac{b}{2}(T_2^2 - T_1^2) - c\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \tag{3.12.6}$$

This expression can then be used with experimentally determined values of a, b, and c, some of which are shown in the following table.

Table 3.12.1: Empirical Parameters for the temperature dependence of C_p

Substance	a (J mol^{-1} K^{-1})	b (J mol^{-1} K^{-2})	c (J mol^{-1} K)
C(gr)	16.86	4.77 x 10^{-3}	-8.54 x 10^{5}
CO_2(g)	44.22	8.79 x 10^{-3}	-8.62 x 10^{5}
H_2O(l)	75.29	0	0
N_2(g)	28.58	3.77 x 10^{-3}	-5.0 x 10^{4}
Pb(s)	22.13	1.172 x 10^{-2}	9.6 x 10^{4}

Example 3.12.1: Heating Lead

What is the molar enthalpy change for a temperature increase from 273 K to 353 K for Pb(s)?

Solution

The enthalpy change is given by Equation 3.12.1 with a temperature dependence C_p given by Equation 3.12.1 using the parameters in Table 3.12.1. This results in the integral form (Equation 3.12.6):

$$\Delta H = a(T_2 - T_1) + \frac{b}{2}(T_2^2 - T_1^2) - c\left(\frac{1}{T_2} - \frac{1}{T_1}\right)$$

when substituted with the relevant parameters of Pb(s) from Table 3.12.1.

$$\begin{aligned}\Delta H =& (22.14\,\frac{J}{mol\,K}(353\,K - 273\,K) \\ &+ \frac{1.172 \times 10^{-2}\,\frac{J}{mol\,K^2}}{2}\left((353\,K)^2 - (273\,K)^2\right) \\ &- 9.6 \times 10^4\,\frac{J\,K}{mol}\left(\frac{1}{(353\,K)} - \frac{1}{(273\,K)}\right) \\ \Delta H =& 1770.4\,\frac{J}{mol} + 295.5\,\frac{J}{mol} + 470.5\,\frac{J}{mol} \\ =& 2534.4\,\frac{J}{mol}\end{aligned}$$

For chemical reactions, the reaction enthalpy at differing temperatures can be calculated from

$$\Delta H_{rxn}(T_2) = \Delta H_{rxn}(T_1) + \int_{T_1}^{T_2} \Delta C_p \Delta T$$

Example 3.12.2: Enthalpy of Formation

The enthalpy of formation of NH_3(g) is -46.11 kJ/mol at 25 °C. Calculate the enthalpy of formation at 100 °C.

Solution

$$\mathrm{N_2\,(g) + 3\,H_2\,(g)} \rightleftharpoons \mathrm{2\,NH_3\,(g)}$$

with $\Delta H\ (298\ K) = -46.11\ kJ/mol$

Compound	Cp (J mol^{-1} K^{-1})
N_2(g)	29.12
H_2(g)	28.82
NH_3(g)	35.06

$$\begin{aligned}\Delta H(373\,K) &= \Delta H(298\,K) + \Delta C_p \Delta T \\ &= -46110 + \frac{J}{mol}\left[2\left(35.06\frac{J}{mol\,K}\right) - \left(29.12\,\frac{J}{mol\,K}\right) - 3\left(28.82\,\frac{J}{mol\,K}\right)\right](373\,K - 298\,K) \\ &= -49.5\,\frac{kJ}{mol}\end{aligned}$$

This page titled 3.12: The Temperature Dependence of ΔH is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 3.5: Temperature Dependence of Enthalpy by Patrick Fleming is licensed CC BY-NC-SA 4.0.

3.13: Enthalpy is a State Function

Our expression for internal energy at constant pressure:

$$\Delta U = q_P + w = q_P - P\Delta V$$

Rearrange:

$$q_P = \Delta U + P\Delta V = U_2 - U_1 + P(V_2 - V_1)$$
$$q_P = (U_2 + PV_2) - (U_1 + PV_1)$$

We can define this term as enthalpy:

$$H \equiv U + PV$$

This is a new state function.

This page titled 3.13: Enthalpy is a State Function is shared under a license and was authored, remixed, and/or curated by LibreTexts.

- 19.14: Enthalpy is a State Function has no license indicated.

3.E: The First Law of Thermodynamics (Exercises)

Gas Law Practice problems with answers:

Exercise 3.E.1

An ideal gas originally at 0.85 atm and 66°C was allowed to expand until its final volume, pressure, and temperature were 94 mL, 0.60 atm, and 45°C, respectively. What was its initial volume?

Answer available at the end of this section.

Exercise 3.E.2

When you drink something from a straw, how does the liquid get from the glass up through the straw into your mouth?

Answer available at the end of this section.

First Law of Thermodynamics Practice problems with answers:

Exercise 3.E.3

A sports bottle containing Gatorade is held in a man's hand while running on a treadmill. Suppose the Gatorade is the system in question. Answer the subsequent questions.

1. Will the temperature rise due to the movement caused by the man's running?
2. Was any heat actually added to the system? If so, what was the source?
3. Was work done onto the system?
4. In what ways has the internal energy of the system been affected, if affected at all?

Answer available at the end of this section.

Exercise 3.E.4

For a constant-pressure process, the heat capacity per mole is given by the expression:

C_p per mole = $(16.0 + 4.35 \times 10^{-3}\ T + 1.15\ T^2)$ J K^{-1} mol^{-1}

Find the molar enthalpy of oxygen gas from 38°C to 115°C

Answer available at the end of this section.

Exercise 3.E.5

You notice your system changes volume from 10 L to 30 L with a constant pressure at 770 torr. Then, there is a decrease in pressure to 740 torr with the volume staying constant at 30 L. Calculate the total amount of work done.

Answer available at the end of this section.

Exercise $3.E.6$

Determine the $\Delta_{vap}\overline{H}^{\circ}$ at 343.15 K for H_2O. (Note: The heat of vaporization for water is known to be 40.79 kJ mol^{-1} at 273.15 K, the $\overline{C}_P$ of $H_2O(l)$ was determined to be 85.6 J K^{-1} mol^{-1} and the $\overline{C}_P$ of $H_2O(s)$ was determined to be 54.35 J K^{-1} mol^{-1}). Use the reaction below and give your final answer in kJ mol^{-1}.

$$H_2O(s, 273.15K, 0.1atm) \rightarrow H_2O(l, 273.15K, 0.1atm)$$

Answer available at the end of this section.

Additional Practice problems

1. 1.00 mol of an ideal gas, initially occupying 12.2 L at 298 K, expands isothermally against a constant external pressure of 1.00 atm until the pressure of the gas is equal to the external pressure. Calculate Δp, q, w, ΔU, and ΔH for the expansion.
2. Some ice cubes are placed inside a sealed plastic bag on a hot day, what is the sign of w, q, ΔH, ΔU? What does this mean about the system?
3. Consider 1.00 mol of an ideal gas expanding isothermally at 298 K from an initial volume of 12.2 L to a final volume of 22.4 L. Calculate Δp, q, w, ΔU, and ΔH for the expansion.
4. Use the following data [Huff, Squitieri, and Snyder, J. Am. Chem. Soc., **70**, 3380 (1948)] to calculate the standard enthalpy of formation of tungsten carbide, WC(s).

Reaction	ΔH^{o} (kJ)
$C(gr)+O_2(g) \rightarrow CO_2(g)$	-393.51
$WC(s)+5/2O^2(g) \rightarrow WO_3(s)+CO_2(g)$	-1195.79
$W(s)+3/2O_{2(g)} \rightarrow WO_3(s)$	-837.42

5. You are given a gas at 750 torr with a volume of 0.62 Liters at 25°C. A while later, you measure the temperature and notice a 5°C temperature increase in the system. Assuming there is no pressure change and that this is an ideal gas, what is the new volume?

6. Consider an ice cube melting.

a. Is the change of enthalpy positive or negative for this process?

b. Is the change of entropy positive or negative for this process?

7. 2.5 moles of CO_2 expands against an external pressure of 1.5 atm. It was initially at 9.0 atm and 30°C. The final volume is 35.0 L.

1. What is the final temperature?
2. Calculate q, w, and ΔU for the process

8. A 10.0 g sheet of gold with a temperature of 18.0°C is laid flat on a sheet of iron that weights 20.0 g and has a temperature of 55.6°C. Given that the specific heats of Au and Fe are 0.129 J g^{-1} $°C^{-1}$ and 0.444 J g^{-1} $°C^{-1}$, respectively, what is the final temperature of the combined metals? (Hint: The heat gained by the gold must be equal to the heat lost by the iron.)

9. What is ΔH if of 3.00 moles of $O_2(g)$ are cooled from 175°C to 35.0°C and

$$Cp^{-}=(7.50\times10^{-3}T-2.30\times10^{-6}T^{2}\ Jmol^{-1}K^{-1}$$

10.0 1.52 g of toluene (C_7H_8) is combusted in a constant volume bomb calorimeter at 25°C. Calculate the ΔrU^{-} and ΔrH^{-} for this reaction given that 39.183 kJ of heat was released.

ADAPT 3.E.1

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this Interactive Content online

Answers

Exercise 3.E.1

71 mL

Exercise 3.E.2

When you close your mouth around the straw, you close off the atmosphere from outside to inside. During suction, you increase the volume inside your mouth, thus lowering the internal pressure and the atmospheric pressure exerted onto the liquid over comes the internal pressure and so the liquid flows into your mouth.

Exercise 3.E.3

1. The man's running motion causes the Gatorade to be shaken. The shaking motion adds energy to the system and increases the random motion of the molecules. Since the motion of the molecules rises, so does the temperature of the system.
2. Yes, heat exchange did occur between the system and its surroundings (the heat given off from the man's hand).
3. Yes, work was done onto the system via the shaking motion.
4. The internal energy has been increased

Exercise 3.E.3

$$\begin{aligned}\Delta H &= \int_{38}^{115}(16+4.35\times10^{-3}T+1.15T^2)dT \\ &= (16*115+(4.35\times10^{-3})/2*115^2+1.15/3*115^3)-(16*38+(4.35\times10^{-3})/2*38^2+1.15/3*38^3) \\ &= 5.63J/mol\end{aligned}$$

Exercise 3.E.5

We split up the solution into two parts.

Work done during the volume change and the work done during the pressure change.

First part:

$$w = P_{ex}\Delta V$$

$$770torr\cdot\frac{1atm}{760torr}=1.01atm$$

$$w=(-1.01atm)(20L)(101.3\frac{J}{1L\cdot atm})=-2046.3J$$

Second part:

Convert pressure to atm:

$$740torr\cdot\frac{1atm}{760torr}=0.974atm$$

$$w=-P_2V_2ln\frac{P_1}{P_2}$$

$$w=-(0.974atm)(30L)ln\frac{0.974}{1.0}\cdot\frac{101.3J}{1L\cdot atm}$$

=107.4 J

Then add up the two works

w_{tot}= -2046.3+107.4=-1938.9 J

Exercise 3.E.6

$$\Delta_{vap}\overline{H}_{343.15K}=\Delta\overline{H}_1+\Delta\overline{H}_2+\Delta\overline{H}_3$$

$$\Delta\overline{H}_1 = \overline{C}_P(l)\Delta T = 85.6 \times (273.15 - 343.15) = 3.852\text{kJmol}^{-1}$$
$$\Delta\overline{H}_2 = \Delta_{\text{vap}}\overline{H}^{\circ}_{273.15\text{K}} = +40.79\text{kJmol}^{-1}$$
$$\Delta\overline{H}_3 = \overline{C}_P(g)\Delta T = 54.35 \times (343.15 - 273.15) = -2.446\text{kJmol}^{-1}$$
$$\Delta H = 3.825 + 40.79 + (-2.446) = 42.17\text{kJmol}^{-1}$$
$$\Delta_{\text{V}}\overline{H}_{343.15\text{K}} - \Delta_{\text{V}}\overline{H}_{273.15\text{K}} = \Delta\overline{C}_P(T_{343.15} - T_{273.15})$$
$$\Delta_{\text{V}}\overline{H}_{343.15\text{K}} = \Delta_{\text{V}}\overline{H}_{273.15\text{K}} + \Delta\overline{C}_P(T_{343.15} - T_{273.15})$$
$$= 40.79 + (54.35 - 85.6) \times (343.15 - 273.15) = 40.79\text{kJmol}^{-1}$$

This page titled 3.E: The First Law of Thermodynamics (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4: ENTROPY AND THE SECOND AND 3RD LAW OF THERMODYNAMICS

4.1: Energy Does not Determine Spontaneity

4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal

4.3: Unlike heat, Entropy is a State Function

4.4: The Second Law of Thermodynamics

4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes

4.6: Entropy Increases With Increasing Temperature

4.7: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale

4.8: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition

4.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions

4.E: Entropy and The Second Law of Thermodynamics (Exercises)

This page titled 4: Entropy and The Second and 3rd Law of Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 20: Entropy and The Second Law of Thermodynamics is licensed CC BY-NC-SA 4.0.

4.1: Energy Does not Determine Spontaneity

There are many spontaneous events in nature. If you open the valve in both cases a spontaneous event occurs. In the first case the gas fills the evacuated chamber, in the second the gases will mix. The state functions U and H do not give us a clue what will happen. You might think that only those events are spontaneous that produce heat.

Not so:

- If you dissolve $\mathrm{KNO_3}$ in water, it does so spontaneously, but the solution gets ***cold***.
- If you dissolve KOH in water, it does so spontaneously, but the solution gets ***hot***.

Clearly the first law is not enough to describe nature.

Two items left on our wish list

The development of the new state function entropy has brought us much closer to a complete understanding of how heat and work are related:

1. the spontaneity problem
   we now have a criterion for spontaneity for isolated systems

2. the asymmetry between work->heat (dissipation) and heat->work (power generation)
   at least we can use the new state law to predict the limitations on the latter.

Two problems remain:

1. we would like a spontaneity criterion *for all systems* (not just isolated ones)
2. we have a new state function S, but what is it?

Causality vs. Correlation

In Europe nosey little children who are curious to find out where their newborn little brother or sister came from, often get told that the stork brought it during the night. When you look at the number of breeding pairs of this beautiful bird in e.g. Germany since 1960 you see a long decline to about 1980 when the bird almost got extinct. After that the numbers go up again due to breeding programs mostly. The human birth rate in the country follows a pretty much identical curve and the correlation between the two is very high (>0.98 or so). (Dr H. Sies in Nature (volume 332 page 495; 1988). Does this prove that storks indeed bring babies?



An adult White Stork. from Wikipedia (credit: Frank Vassen).

Answer:

No, it does ***not*** show causality, just a correlation due a common underlying factor. In this case that is the choices made by the German people. First they concentrated on working real hard and having few children to get themselves out of the poverty WWII had left behind and neglected the environment, then they turned to protect the environment and opened the doors to immigration of people, mostly from Muslim countries like Turkey or Morocco, that usually have larger families. The lesson from this is that you can only conclude causality ***if*** you are sure that there are no other intervening factors.

This page titled 4.1: Energy Does not Determine Spontaneity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 20.1: Energy Does not Determine Spontaneity is licensed CC BY-NC-SA 4.0.

4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal

In spontaneous processes for an isolated system, there is a competition between minimizing the energy of the system and increasing the dispersal of energy within the system. If energy is constant, then the system will evolve to maximize energy dispersal. If energy dispersal is not a factor, the system will evolve to minimize its energy. We already have a quantitative basis for the energy of a system and we need to do the same for energy dispersal. Suppose we have a small reversible change dU in the energy of an ideal gas. We know that U only depends on temperature:

$$dU = C_v dT$$

We also know that any reversible work would be volume work.

$$\delta w_{rev} = -PdV$$

This means that we can write:

$$\begin{aligned} \delta q_{rev} &= dU - \delta w_{rev} \\ &= C_v dT + PdV \end{aligned}$$

Let us examine if this represents an **exact differential**. If δq were an exact differential, we could write the total differential:

$$\delta q_{rev} = \left(\frac{\partial q}{dT}\right)_V dT + \left(\frac{\partial q}{dV}\right)_T dV$$

And the following would be true:

$$\frac{\partial^2 q_{rev}}{\partial T \partial V} = \frac{\partial^2 q_{rev}}{\partial V \partial T}$$

From our equation above, we know that:

$$\frac{\partial q_{rev}}{\partial T} = C_V$$

$$\frac{\partial q_{rev}}{\partial V} = P$$

Therefore, the following should be true:

$$\frac{\partial C_V}{\partial V} = \frac{\partial P}{\partial T}$$

However,

$$\frac{\partial C_v}{\partial V} = 0$$

Because C_v does not depend on volume (only T, just like U: it is its derivative). And:

$$\frac{\partial P}{\partial T} = \frac{\partial nRT}{\partial T} = \frac{nR}{V}$$

Which is not zero!! Clearly, δq_{rev} is **not a state function**, but look what happens if we multiply everything with an 'integration factor' $1/T$:

$$\frac{\delta q_{rev}}{T} = \frac{C_v}{T} dT + \frac{P}{T} dV$$

$$\frac{\partial C_v/T}{\partial V} = 0$$

Because $\frac{C_v}{T}$ does not depend on volume. However,

$$\frac{\partial (P/T)}{\partial T} = \frac{\partial (nR/V)}{\partial T} = 0$$

Thus, the quantity $dS = \frac{\delta q_{rev}}{T}$ is an exact differential, so S is a state function and it is called **entropy**. Entropy is the dispersal of energy, or equivalently, the measure of the number of possible microstates of a system.

This page titled 4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 20.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal is licensed CC BY-NC-SA 4.0.

4.3: Unlike heat, Entropy is a State Function

Circular integrals

Because entropy is a state function, it integrates to zero over any circular path going back to initial conditions, just like U and H:

$$\oint dS = 0$$

$$\oint dH = 0$$

$$\oint dU = 0$$

As discussed previously, we can use this fact to revisit the isotherm + isochore + adiabat circular path (Figure 20.3.1).

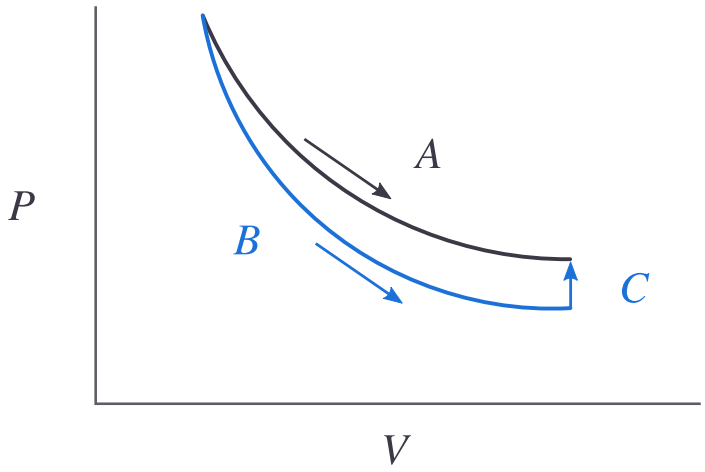


Figure 20.3.1 : A: isothermal expansion, B:adiabat, C:isochore (CC BY-NC; Ümit Kaya via LibreTexts)

Along adiabat B and isochore C:

- There is no heat transfer along adiabat B:

$$q_{rev,B} = 0$$

- There is no work along isochore C:

$$\delta w = 0$$

- But the temperature changes from T_2 back to T_1. This requires heat:

$$q_{rev,C} = C_V \Delta T$$

Along the isotherm A:

- We have seen that

$$q_{rev,A} = nRT \ln \frac{V_2}{V_1}$$

The quantities $q_{rev,A}$, $q_{rev,B}$, and $q_{rev,C}$ are ***not*** the same, which once again underlines that heat is a path function. How about entropy?

First, consider the combined paths of B and C:

$$q_{rev,B+C} = \int_{T_2}^{T_1} C_v dT$$

$$\int dS_{B+C} = \int \frac{dq_{rev,B+C}}{T} = \int_{T_2}^{T_1} \frac{C_v}{T} dT$$

We had seen this integral before from Section 19-6, albeit from T_1 to T_2:

$$\Delta S_{B+C} = nR \ln \frac{V_2}{V_1} \tag{4.3.1}$$

(Notice sign in Equation 4.3.1 is positive)

Along the isotherm A:

$$q_{rev,A} = nRT \ln \frac{V_2}{V_1}$$

T is a constant so we can just divide $q_{rec,A}$ by T to get ΔS_A:

$$\Delta S_A = nR \ln \frac{V_2}{V_1}$$

We took two different paths to get start and end at the same points. Both paths had the same change in entropy. Clearly entropy is a *state function* while q_{rev} is not.

This page titled 4.3: Unlike heat, Entropy is a State Function is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

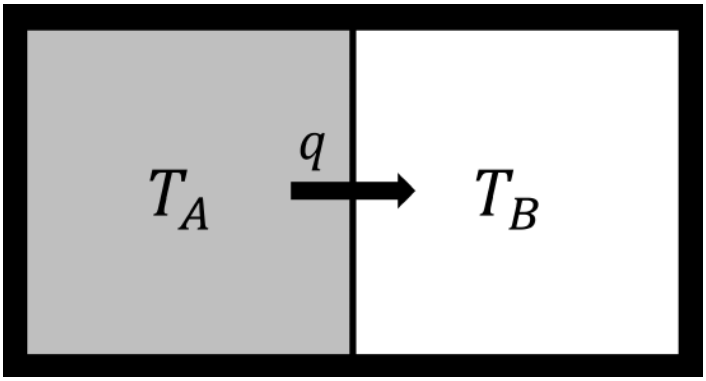
- 20.3: Unlike heat, Entropy is a State Function is licensed CC BY-NC-SA 4.0.

4.4: The Second Law of Thermodynamics

Spontaneity of an isolated system

> An **isolated system** is a little more than just adiabatic. In the latter heat cannot get in or out. In an isolated system *nothing* gets in or out, neither heat nor mass nor even any radiation, such as light. The isolated system is like a little universe all to itself.

Let us consider a zero law process. We have two identical blocks of metal, say aluminum. They are each at thermal equilibrium, but at different temperatures. They are brought into contact with each other but isolated from the rest of the universe.



- ***Zeroth law***: *Heat will flow from hot to cold*
- ***First law***: *There is no change in total energy*

so:

$$dU_A = -dU_B$$

There is also no work so:

$$dU_A = \delta q_A + 0$$

Because U is a state function this makes q a state function as well, otherwise this equality does not hold. As there is only one term on the right there is only one path (along q). So we could write:

$$dU_A = dq_A$$

This implies that we do not need to worry about reversible and irreversible paths as there is only one path. Since:

$$dS = \frac{\delta q_{rev}}{T}$$

In this particular case:

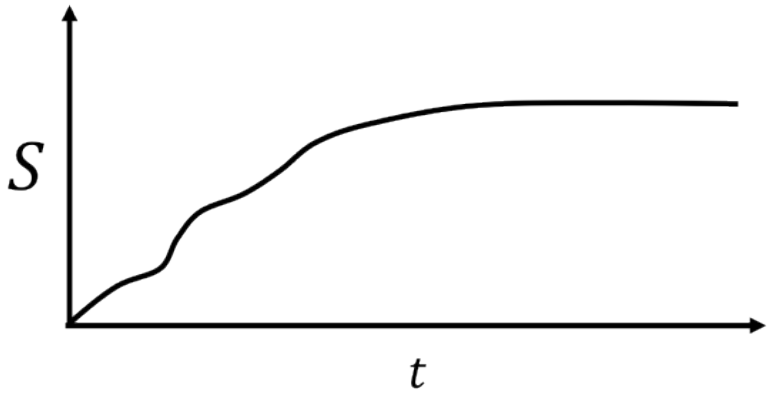
$$TdS = \delta q_{rev} = dU$$

Thus we get:

$$dS = \frac{dU_A}{T_A} + \frac{dU_B}{T_B} = \frac{dU_A}{T_A} - \frac{dU_A}{T_B} = dU_A \left(\frac{1}{T_A} - \frac{1}{T_B} \right)$$

Clearly as long as the two temperatures are not the same dS is not zero and entropy is not conserved. Instead it is **increasing**.

Over time, the temperatures will become the same (if the blocks are identical, the final temperature is the average of T_A and T_B) and the entropy will reach a maximum.



For our two identical blocks of metal (with same heat capacity, C_V), we can, in fact, derive that the entropy change:

$$\Delta S = C_V \ln \left[\frac{T_A^2 + T_B^2}{4 T_A T_B} \right].$$

This is indeed a positive quantity. In general, we can say for an isolated system:

$$dS \geq 0$$

Thus if we are dealing with a spontaneous (and isolated) process $dS > 0$ and ***entropy is being produced***. This gives us a ***criterion for spontaneity***.

Entropy exchange of an open system

In an isolated system dS represents the produced entropy dS_{prod} and this is a good criterion for spontaneity. Of course the requirement that the system is isolated is very restrictive and makes the criterion as good as useless... What happens in a system that can exchange heat with the rest of the universe? We do have entropy changes in that case, but part of them may have nothing to do with production, because we also have to consider the heat that is exchanged.

$$dS = dS_{prod} + dS_{exchange}$$

If the process is reversible (that is completely non-spontaneous) we are dealing with δq_{rev} so that $dS_{exchange} = \delta q_{rev}/T$, but that is also what dS_{tot} is equal to (by definition). This leaves no room for entropy production.

So we have:

Isolated: $dS = dS_{prod} + 0$

Reversible $dS = 0 + \delta q_{rev}/T$

Notice that this demonstrates that for non-isolated systems entropy change is *not* a good criterion for spontaneity at all... In the case the heat exchange is ***irreversible*** *part* of the entropy is entropy production by the system:

Irreversible: $dS = dS_{prod} + \delta q_{irrev}/T$

$dS > \delta q_{irrev}/T$ in this case.

Generalizing the isolated, irreversible and reversible cases we may say:

$$dS \geq \frac{\delta q}{T}$$

This is the ***Clausius inequality***.

This page titled 4.4: The Second Law of Thermodynamics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 20.4: The Second Law of Thermodynamics is licensed CC BY-NC-SA 4.0.

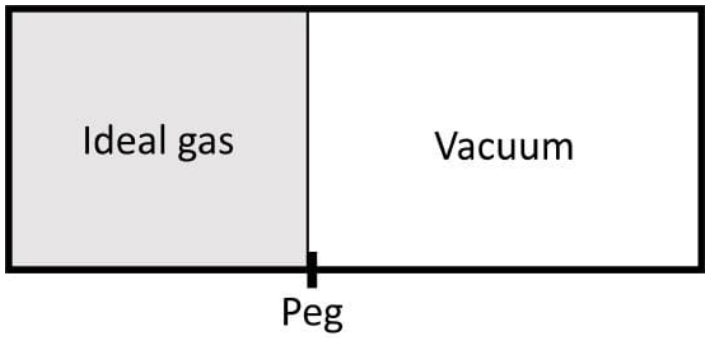
4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes

The second law of thermodynamics can be formulated in many ways, but in one way or another they are all related to the fact that the state function entropy, S, tends to increase over time in isolated systems. For a long time, people have looked at the entire universe as an example of an isolated system and concluded that its entropy must be steadily increasing until $\delta S_{universe}$ becomes zero. As we will see below, the second law has important consequences for the question of how we can use heat to do useful work.

> Of late, cosmologists like the late Richard Hawkins have begun to question the assumption that the entropy of the universe is steadily increasing. The key problem is the role that gravity and relativity play in creating black holes.

Vacuum Expansion

Let's compare two expansions from V_1 to V_2 for an ideal gas, both are isothermal. The first is an irreversible one, where we pull a peg an let the piston move against vacuum:



The second one is a reversible isothermal expansion from V_1 to V_2 (and P_1 to P_2) that we have examined before. In both cases, the is no change in internal energy since T does not change. During the irreversible expansion, however, there is also no volume work because the piston is expanding against a vacuum and the following integral:

$$\int -P_{ext} dV = 0$$

integrates to zero. The piston has nothing to perform work against until it slams into the right hand wall. At this point $V = V_2$ and then dV becomes zero. This is not true for the reversible isothermal expansion as the external pressure must always equal the internal pressure.

> No energy and no work means no heat!

Clearly the zero heat is irreversible heat ($q_{irr} = 0$) and this makes it hard to calculate the entropy of this spontaneous process. But then this process ends in the same final state as the reversible expansion from V_1 to V_2. We know that dU is still zero, but now $\delta w_{rev} = -\delta q_{rev}$ is nonzero. We calculated its value before:

$$q_{rev} = nRT \ln \left(\frac{V_2}{V_1} \right) \tag{4.5.1}$$

The Claussius definition of entropy change can be used to find ΔS (under constant temperature).

$$\Delta S = \frac{q_{rev}}{T} \tag{4.5.2}$$

Substituting Equation 4.5.1 into Equation 4.5.2 results in

$$\Delta S = nR \ln \left(\frac{V_2}{V_1} \right)$$

As S is a state function this equation also holds for the irreversible expansion against vacuum.

> Always calculate the entropy difference between two points along a **reversible** path.

For the irreversible expansion into vacuum we see that

$$\begin{aligned} \Delta S_{\text{total}} &= \Delta S_{\text{sys}} + \Delta S_{\text{surr}} \\ &= nR \ln \left(\frac{V_2}{V_1} \right) + 0 \\ &= nR \ln \left(\frac{V_2}{V_1} \right) \end{aligned}$$

For the reversible expansion, heat is transferred to the system while the system does work on the surroundings in order to keep the process isothermal:

$$\Delta S_{\text{sys}} = nR \ln \left(\frac{V_2}{V_1} \right)$$

The entropy change for the surrounding is the opposite of the system:

$$\Delta S_{\text{surr}} = -nR \ln \left(\frac{V_2}{V_1} \right)$$

This is because the amount of heat transferred to the system is the same as the heat transferred from the surroundings and this process is reversible so the system and surroundings are at the same temperature (equilibrium). Heat is related to entropy by the following equation:

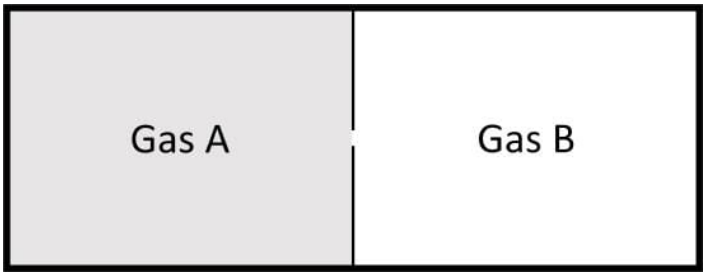
$$dS = \frac{\delta q}{T}$$

Therefore, the total entropy change for the reversible process is zero:

$$\begin{aligned} \Delta S_{\text{total}} &= \Delta S_{\text{sys}} + \Delta S_{\text{surr}} \\ &= nR \ln \left(\frac{V_2}{V_1} \right) - nR \ln \left(\frac{V_2}{V_1} \right) \\ &= 0 \end{aligned}$$

The Mixing of Two Gases

Consider two ideal gases at same pressure separated by a thin wall that is punctured. Both gases behave as if the other one is not there and again we get a spontaneous process, mixing in this case.



If the pressure is the same the number of moles of each gas should be proportional to the original volumes, V_A and V_B, and the total number of moles to the total volume V_{tot}.

For gas A we can write:

$$\Delta S_A = n_A R \ln \frac{V_{tot}}{V_A} = n_A R \ln \frac{n_{tot}}{n_A}$$

and similarly for gas B we can write:

$$\Delta S_B = n_B R \ln \frac{V_{tot}}{V_B} = n_B R \ln \frac{n_{tot}}{n_B}$$

The total entropy change is therefore the sum of constituent entropy changes:

$$\Delta S = \Delta S_A + \Delta S_B$$

and the entropy change total per mole of gas is:

$$\frac{\Delta S}{n_{tot}} = R \frac{\left[n_B \ln \frac{n_{tot}}{n_B} + n_A \ln \frac{n_{tot}}{n_A} \right]}{n_{tot}} \tag{4.5.3}$$

Equation 4.5.3 can be simplified using mole fractions:

$$\chi_A = \frac{n_A}{n_{tot}}$$

and the mathematical relationship of logarithms that:

$$\ln \left(\frac{x}{y} \right) = - \ln \left(\frac{y}{x} \right)$$

to:

$$\Delta \bar{S} = -R \left[\chi_A \ln \chi_A + \chi_B \ln \chi_B \right] \tag{4.5.4}$$

In the case of mixing of more than two gases, Equation 4.5.4 can be expressed as:

$$\Delta \bar{S} = -R \sum \chi_i \ln \chi_i \tag{4.5.5}$$

This entropy expressed in Equations 4.5.4 and 4.5.5 is known as the **entropy of mixing**; its existence is the major reason why there is such a thing as diffusion and mixing when gases, and also solutions (even solid ones), are brought into contact with each other.

This page titled 4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jerry LaRue.

- 20.6: We Must Always Devise a Reversible Process to Calculate Entropy Changes by Jerry LaRue is licensed CC BY-NC-SA 4.0.

4.6: Entropy Increases With Increasing Temperature

Learning Objectives

- Define entropy and its relation to energy flow.

Entropy versus temperature

We can put together the first and the second law for a reversible process with no other work than volume (PV) work and obtain:

$$dU = \delta q_{rev} + \delta w_{rev}$$

Entropy is the dispersal of energy and is related to heat:

$$\delta q_{rev} = TdS$$

Work is related to the change in volume:

$$\delta w_{rev} = -PdV$$

Plugging these into our expression for dU for reversible changes:

$$dU = TdS - PdV$$

We no longer have any path functions in the expression, as U, S and V are ***all state functions***. This means this expression must be an exact differential. We can generalize the expression to hold for irreversible processes, but then the expression becomes an inequality:

$$dU \leq TdS - PdV$$

This equality expresses U as a function of two variables, entropy and volume: $U(S,V)$. S and V are the ***natural variables*** of U.

Entropy and heat capacity

At constant volume, dU becomes:

$$dU = TdS$$

Recall that internal energy is related to constant volume heat capacity, C_V:

$$C_V = \left(\frac{dU}{dT}\right)_V$$

Combining these two expressions, we obtain:

$$dS = \frac{C_V}{T}dT$$

Integrating:

$$\Delta S = \int_{T_1}^{T_2} \frac{C_V(T)}{T}dT$$

If we know how C_V changes with temperature, we can calculate the change in entropy, ΔS. Since heat capacity is always a positive value, entropy must increase as the temperature increases. There is nothing to stop us from expressing U in ***other*** variables, e.g.

T and V. In fact, we can derive some interesting relationships if we do.

Example 21.1.1

1. Write U as a function of T and V.
2. Write U as a function of its natural variables.
3. Rearrange (2) to find an expression for dS.
4. Substitute (1) into (3) and rearrange. This is the definition of C_V.
5. Write out S as a function of T and V.

We can also derive an expression for the change in entropy as a function of constant pressure heat capacity, C_P. To start, we need to change from internal energy, U, to enthalpy, H:

$$\begin{aligned} H &= U + PV \\ dH &= dU + d(PV) \\ &= dU + PdV + VdP \end{aligned}$$

For reversible processes:

$$\begin{aligned} dH &= dU + PdV + VdP \\ &= TdS - PdV + PdV + VP \\ &= TdS + VdP \end{aligned}$$

The natural variables of the enthalpy are S and P (not: V). A similar derivation as above shows that the temperature change of entropy is related to the constant pressure heat capacity:

$$dH = TdS + VdP$$

At constant pressure:

$$dH = TdS + VdP$$

Recall that:

$$C_P = \frac{dH}{dT}$$

Combining, we obtain:

$$dS = \frac{C_P}{T} dT$$

Integrating:

$$\Delta S = \int_{T_1}^{T_2} \frac{C_P(T)}{T} dT$$

This means that if we know the heat capacities as a function of temperature we can calculate how the entropy changes with temperature. Usually it is easier to obtain data under constant P conditions than for constant V, so that the route with C_p is the more common one.

This page titled 4.6: Entropy Increases With Increasing Temperature is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 21.1: Entropy Increases with Increasing Temperature is licensed CC BY-NC-SA 4.0.

4.7: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale

Learning Objectives

- Recall the third law of thermodynamics.

In the unlikely case that we have C_p data all the way from 0 K we could write:

$$S(T) = S(0) + \int_0^T C_p(T)\ dT$$

It is tempting to set the $S(0)$ to zero and make the entropy thus an absolute quantity. As we have seen with enthalpy, it was not really possible set $H(0)$ to zero. All we did was define ΔH for a particular process, although if C_p data are available we could construct an enthalpy function (albeit with a floating zero point) by integration of C_P (instead of C_p/T!). The situation is different for entropy S than for H. Here we can actually put things on an ***absolute*** scale. Both Nernst and Planck have proposed to do so. Nernst postulated that for a pure and perfect crystal S should indeed to go to zero as T goes to zero (Figure 4.7.1).

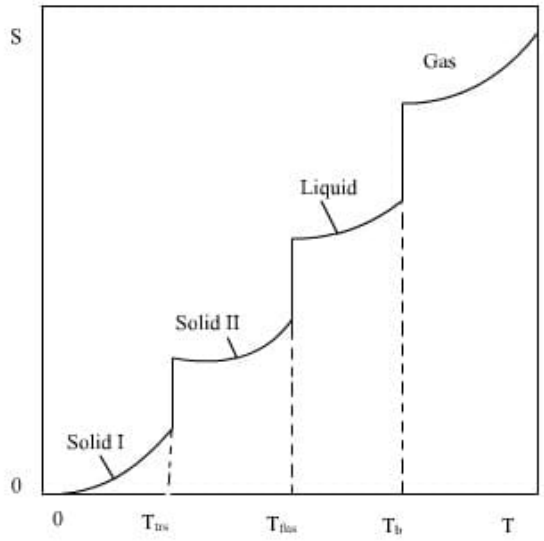


Figure 4.7.1: Entropy as a Function of Temperature. The entropy (S) of a substance (compound or element) as a function of temperature (T). Note that there are three phase transitions in this system with evolution through two solid phases. These result in positive jumps in the entropy.

Example 4.7.1: Sulfur

Sulfur has two solid crystal structure, rhombic and monoclinic. At 368.5 K, the entropy of the phase transition from rhombic to monoclinic, $S_{(rh)} \rightarrow S_{(mono)}$, is:

$$\Delta_{\text{trs}} S(368.5\ \text{K}) = 1.09\ \frac{\text{J} \cdot \text{mol}}{\text{K}}$$

As the temperature is lowered to 0 K, the entropy of the phase transition approaches zero:

$$\lim_{T \to 0} \Delta_r S = 0$$

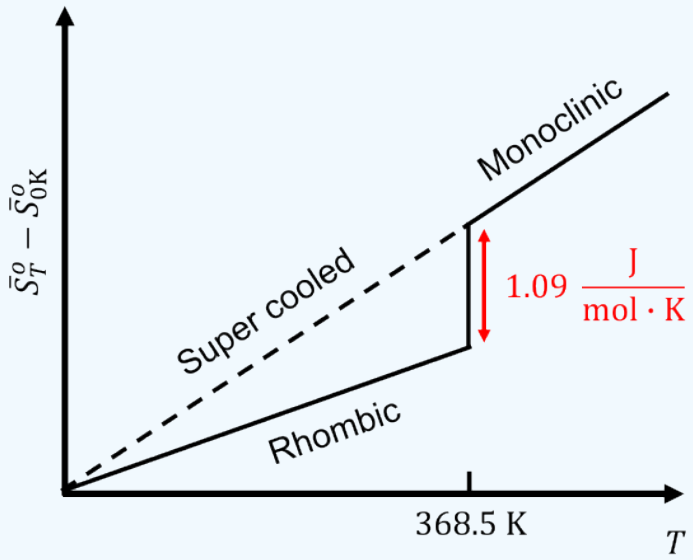


Figure 4.7.2: Schematic showing the entropy of the rhombic and monoclinic structures of sulfur as a function of temperature.

This shows that the entropy of the two crystalline forms are the same. The only way is if all species have the same absolute entropy at 0 K. For energy dispersal at 0 K:

1. Energy is as concentrated as it can be.
2. Only the very lowest energy states can be occupied.

For a pure, perfect crystalline solid at 0 K, there is only one energetic arrangement to represent its lowest energy state. We can use this to define a natural zero, giving entropy an absolute scale. The **third Law of Thermodynamics** states that the entropy of a pure substance in a perfect crystalline form is 0 $\frac{J}{mol \cdot K}$ at 0 K:

$$\bar{S}^\circ_{0\text{ K}} = 0$$

This is consistent with our molecular formula for entropy:

$$S = k \ln W$$

For a perfect crystal at 0 K, the number of ways the total energy of a system can be distributed is one ($W = 1$). The $\ln W$ term goes to zero, resulting in the perfect crystal at 0 K having zero entropy (Figure $4.7.3a$).

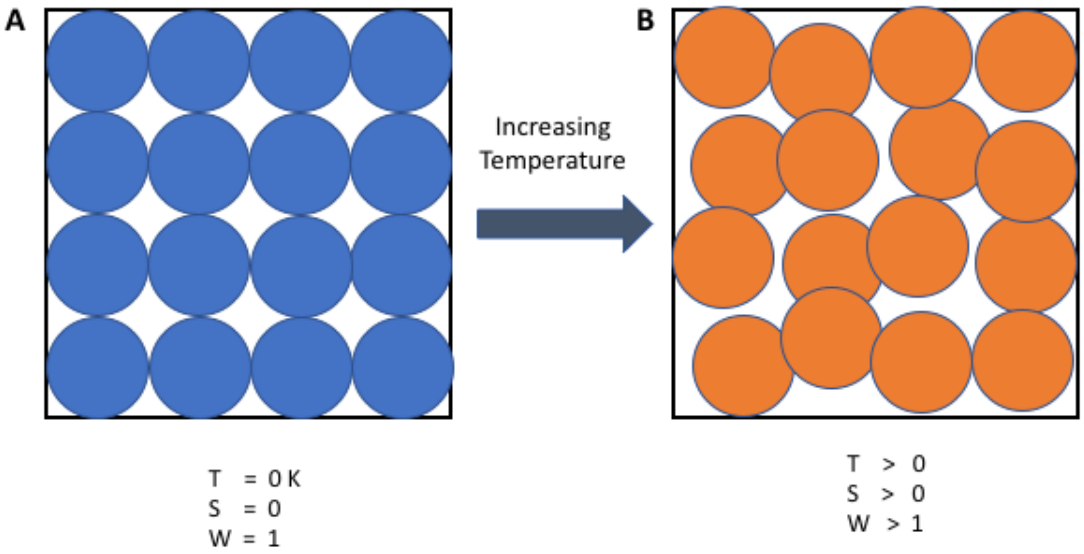


Figure 4.7.3: (a) Single possible configuration for a system at absolute zero, i.e., only one microstate is accessible. Thus $S = k \ln W = 0$. (b) At temperatures greater than absolute zero, multiple microstates are accessible due to atomic dislocation (exaggerated in the figure). Since the number of accessible microstates is greater than 1, $S = k \ln W > 0$. (CC BY-SA 4.0; BlyumJ via Wikipedia)

It is certainly true that for the great majority of materials we end up with a crystalline material at sufficiently low materials (although there are odd exceptions like liquid helium). However, it should be mentioned that a completely *perfect* crystal can only be grown at zero Kelvin! It is not possible to grow anything at 0 K however. At any finite temperature the crystal always incorporates defects, more so if grown at higher temperatures. When cooled down very slowly the defects tend to be ejected from the lattice for the crystal to reach a new equilibrium with less defects. This tendency towards less and less disorder upon cooling is what the third law is all about.

However, the ordering process often becomes impossibly slow, certainly when approaching absolute zero. This means that real crystals always have *some* frozen in imperfections. Thus there is always *some* residual entropy. Fortunately, the effect is often too small to be measured. This is what allows us to ignore it in many cases (but not all). We could state the Third law of thermodynamics as follows:

Definition: Third law of thermodynamics

The entropy of a perfect crystal approaches zero when T approaches zero (but perfect crystals do not exist).

$$\lim_{T \to 0} S^{\circ}(T) = 0 \tag{4.7.1}$$

Another complication arises when the system undergoes a phase transition, e.g. the melting of ice. As we can write:

$$\Delta_{fus} H = q_P$$

If ice and water are in equilibrium with each other the process is quite reversible and so we have:

$$\Delta_{fus} S = \frac{q_{rev}}{T} = \frac{\Delta_{fus} H}{T}$$

This means that at the melting point the curve for S makes a sudden jump by this amount because all this happens at one an the same temperature. Entropies are typically calculated from calorimetric data (e.g., C_P measurements) and tabulated in standard molar form.

There are some clear trends in the entropy values of elements and molecules, e.g. when the noble gas gets heavier this induces more entropy. This is a direct consequence of the particle in the box formula. It has mass in the denominator and therefore the energy levels get more crowded together when m increases: more energy levels, more entropy.

The energy levels get more crowded together when m increases: more energy levels, more entropy

There are tables for $H^\circ(T) - H^\circ(0)$, $S^\circ(T)$ and $G^\circ(T) - H^\circ(0)$ as a function of temperature for numerous substances. As we discussed before the plimsoll defines are standard state in terms of pressure (1 bar) and of concentration reference states, if applicable, but temperature is the one of interest.

Residual Entropy in Carbon Monoxide

For most substances, the assumption made in Equation 4.7.1 regarding the Third Law is generally valid; however, there are notable exceptions, such as carbon monoxide. In its solid state, carbon monoxide molecules would ideally be fully ordered at absolute zero. However, due to the similar sizes of carbon and oxygen atoms and the molecule's weak dipole, it is possible for some molecules to be positioned "upside down"—with oxygen occupying a carbon site and vice versa (i.e., the $\mathrm{C{\equiv}O}$ and $\mathrm{O{\equiv}C}$ orientations are nearly isoenergetic in the lattice). When the crystal forms at higher temperatures, this disorder increases entropy, thereby lowering the Gibbs energy (G). Essentially, if each molecule can be placed into the lattice in two different orientations, the total number of possible configurations (W_{disorder}) for N molecules is 2^N.

This leads to an additional contribution to the entropy of

$$S_{disorder} = k\,\ln W_{disorder} = k\,\ln 2^N = Nk\ln 2 = 5.7\,\mathrm{J/(mol\ K)}$$

At lower temperatures, the entropy term in $\Delta G = \Delta H - T\Delta S$ diminishes in significance, and the crystal should theoretically transition to a more ordered state with lower entropy. However, in practice, the kinetics of this process are extremely slow, preventing the ordering from occurring. As a result, solid CO retains **a nonzero entropy** even as it approaches absolute zero (Equation 4.7.1).

In principle *all* crystalline materials have this effect to some extent, but CO is unusual because the concentration of 'wrongly aligned' entities is of the order of 50% rather than say 1 in 10^{13} (a typical defect concentration in say single crystal silicon). The experimental residual entropy of CO is approximately 4.4 J/(mol K); this indicates that the two orientations of CO are not exactly identical in energy (i.e., equal probability of forming), which would give a residulate entropy of 5.7 J/(mol K).

Exercise 4.7.1

The entropy of a crystal composed of one mole of identical molecules is measured to be 9.134 J/K at 0 K. Determine the number of possible orientations each molecule can adopt within the crystal. Additionally, illustrate a molecule that would be expected to exhibit this behavior.

Summary

Residual entropy quantifies the remaining disorder or randomness in a system even at absolute zero temperature (0 K). According to the Third Law of Thermodynamics, the entropy of a perfect crystal approaches zero as the temperature nears absolute zero. However, some systems retain a nonzero entropy at 0 K, exhibiting residual entropy. This phenomenon represents an exception to the Third Law rather than a contradiction. The law assumes a perfect crystal with a singular, unique ground state, whereas systems with residual entropy, such as carbon monoxide, possess multiple degenerate ground states.

This page titled 4.7: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 21.2: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale is licensed CC BY-NC-SA 4.0.

4.8: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition

Phase transitions (e.g. melting) often occur under equilibrium conditions. We have seen that both the H and the S curves undergo a discontinuity at constant temperature during melting, because there is an enthalpy of fusion to overcome. For a general phase transition at equilibrium at constant T and P, we can say that:

$$\begin{aligned}\Delta_{trs}G &= \Delta_{trs}H - T_{trs}\Delta_{trs}S \\ &= 0 \\ \Delta_{trs}H &= T_{trs}\Delta_{trs}S \\ \frac{\Delta_{trs}H}{T_{trs}} &= \Delta_{trs}S\end{aligned}$$

For melting of a crystalline solid, we now see *why* there is a sudden jump in enthalpy. The reason is that the solid has a much more ordered structure than the crystalline solid. The decrease in order implies a finite $\Delta_{trs}S$. We should stress at this point that we are talking about ***first order*** transitions here. The reason for this terminology is that the discontinuity is in a function like S, that is a first order derivative of G (or A):

$$\left(\frac{\partial \bar{G}}{\partial T}\right)_P = -\bar{S}$$

Second order derivatives (e.g. the heat capacity) will display a singularity (+∞) at the transition point.

Every phase transition will have a change in entropy associated with it. The different types of phase transitions that can occur are:

$l \rightarrow g$	Vaporization / boiling
$g \rightarrow l$	Condensation
$s \rightarrow l$	Fusion / melting
$l \rightarrow s$	Freezing
$s \rightarrow g$	Sublimation
$g \rightarrow s$	Deposition
$s \rightarrow s$	Solid to solid phase transition

This page titled 4.8: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 21.3: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition is licensed CC BY-NC-SA 4.0.

4.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions

Entropy is a state function, so we can calculate values for a process using any path. This allows us to calculate the entropy change of a chemical reaction using standard entropies. Specifically, we sum the entropies of the products and subtract the entropies of the reactions:

$$\Delta_{rxn}S^\circ = \sum_{\text{Products}} v_i S_i^\circ - \sum_{\text{Reactants}} v_i S_i^\circ$$

Where v_i is the stoichiometric coefficient. Let's look at the combustion of methane:

$$\mathrm{CH_4}\,(g) + 2\mathrm{O_2}\,(g) \rightarrow 2\mathrm{H_2O} + \mathrm{CO_2}\,(g)$$

The standard entropies are:

Molecule	Entropy $\left(\frac{\mathrm{J}}{\mathrm{mol \cdot K}}\right)$
$\mathrm{CH_4}$	186.25[1]
$\mathrm{O_2}$	205.15[1]
$\mathrm{H_2O}$	188.84[1]
$\mathrm{CO_2}$	213.79[1]

The entropy for the combustion of methane is:

$$\Delta S^\circ = [2\,(188.84) + 1\,(213.70)] - [1\,(186.25) + 2\,(205.15)] = -5.17\,\frac{\mathrm{J}}{\mathrm{mol \cdot K}}$$

References

1. Chase, M.W., Jr., *NIST-JANAF Themochemical Tables, Fourth Edition*, **J. Phys. Chem. Ref. Data, Monograph 9**, 1998, 1-1951.

This page titled 4.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 21.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions is licensed CC BY-NC-SA 4.0.

4.E: Entropy and The Second Law of Thermodynamics (Exercises)

This page titled 4.E: Entropy and The Second Law of Thermodynamics (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5: HELMHOLTZ AND GIBBS ENERGIES

5.1: Free Energy Functions

5.2: ΔA, ΔG, and Maximum Work

5.3: Pressure Dependence of Gibbs Energy

5.4: Temperature Dependence of A and G

5.5: When Two Variables Change at Once

5.6: Thermodynamic Functions have Natural Variables

5.7: The Gibbs-Helmholtz Equation

5.E: Gibbs Energy (Exercises)

This page titled 5: Helmholtz and Gibbs Energies is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 22: Helmholtz and Gibbs Energies is licensed CC BY-NC-SA 4.0.

5.1: Free Energy Functions

In the previous chapter, we saw that for a spontaneous process, $\Delta S_{universe} > 0$. While this is a useful criterion for determining whether or not a process is spontaneous, it is rather cumbersome, as it requires one to calculate not only the entropy change for the system, but also that of the surroundings. It would be much more convenient if there was a single criterion that would do the job and focus only on the system. As it turns out, there is.

Since we know that

$$\Delta S_{univ} \geq 0$$

for any natural process, and

$$\Delta S_{univ} = \Delta S_{sys} + \Delta S_{surr}$$

all we need to do is to find an expression for ΔS_{sys} that can be determined by the changes in the system itself. Fortunately, we have already done that! Recalling that at constant temperature

$$\Delta S = -\frac{q_{rev}}{T}$$

and at constant pressure

$$\Delta H = q_p$$

it follows that at constant temperature and pressure

$$\Delta S_{surr} = -\frac{\Delta H_{sys}}{T}$$

Substitution into the above equations yields an expression for the criterion of spontaneity that depends only on variables describing the changes in the system!

$$\Delta S_{univ} \geq \Delta S_{sys} - \frac{\Delta H_{sys}}{T}$$

so

$$\Delta S_{sys} - \frac{\Delta H_{sys}}{T} \geq 0$$

Multiplying both sides by $-T$ yields

$$\Delta H - T\Delta S \leq 0 \tag{5.1.1}$$

A similar derivation for constant volume processes results in the expression (at constant volume and temperature)

$$\Delta U - T\Delta S \leq 0 \tag{5.1.2}$$

> Equation 5.1.1 is of great use to chemists, as most of chemistry occurs at constant pressure. For geologists, however, who are interested in processes that occur at very high pressures (say, under the weight of an entire mountain) and expansion is not a possibility, the constant volume expression of Equation 5.1.1 may be of greater interest.

All of the above arguments can be made for systems in which the temperature is not constant by considering infinitesimal changes. The resulting expressions are

$$dH - TdS \leq 0 \tag{5.1.3}$$

and

$$dU - TdS \leq 0 \tag{5.1.4}$$

The Gibbs and Helmholtz Functions

Equation 5.1.3 suggests a very convenient thermodynamic function to help keep track of both the effects of entropy and enthalpy changes. This function, the **Gibbs function** (or **Gibbs Free Energy**) is defined by

$$G \equiv H - TS$$

A change in the Gibbs function can be expressed

$$\Delta G = \Delta H - \Delta(TS)$$

Or at constant temperature

$$\Delta G = \Delta H - T\Delta S$$

And the criterion for a process to be spontaneous is the DG < 0. As such, it should be clear spontaneity is not merely a function the enthalpy change (although exothermic processes tend to be spontaneous) but also a function of the entropy change, weighted by the temperature. Going back to an earlier example,

$$NaOH(s) \rightarrow Na^{+}(aq) + OH^{-}(aq)$$

with $\Delta H < 0$.

and

$$NaHCO_3(s) \rightarrow Na^{+}(aq) + HCO_3^{-}(aq)$$

with $\Delta H > 0$.

It is easy to see why both processes are spontaneous. In the first case, the process is exothermic (favorable) and proceeds with an increase in entropy (also favorable) due to the formation of fragments in the liquid phase (more chaotic) from a very ordered solid (more ordered). The second reaction is endothermic (unfavorable) but proceeds with an increase in entropy (favorable). So, so long as the temperature is high enough, the entropy term will overwhelm the enthalpy term and cause the process to be spontaneous. The conditions for spontaneous processes at constant temperature and pressure can be summarized in Table 6.1.1.

Table 6.1.1: Spontaneity Conditions for a Process under Constant Temperature and Pressure

ΔH	ΔS	Spontaneous
> 0	> 0	At high T
> 0	< 0	At no T
< 0	> 0	At all T
< 0	< 0	At low T

Similarly to the Gibbs function, the **Helmholtz function** is defined by

$$A \equiv U - TS$$

and provides another important criterion for spontaneous processes at constant value and temperature. *At constant temperature*, the Helmholtz function can be expressed by

$$\Delta A \equiv \Delta U - T\Delta S$$

Based on similar arguments used for the Gibbs function, the Helmholtz function also can be used to predict which processes will be spontaneous at constant volume and temperature according to Table 6.1.2.

Table 6.1.2: Spontaneity Conditions for a Process under Constant Temperature and Volume

ΔU	ΔS	Spontaneous?
> 0	> 0	At high T
> 0	< 0	At no T
< 0	> 0	At all T
< 0	< 0	At low T

Calculating ΔG for Reactions

Much like in the case of enthalpy (and unlike entropy), free energy functions do not have an unambiguous zero to the energy scale. So, just like in the case of enthalpies of formation, by convention, the standard free energy of formation (ΔG_f^o) for elements in their standard states is defined as zero. This allows for two important things to happen. First, ΔG_f^o can be measured and tabulated for any substance (in principle, at least.) ΔG_f^o is determined to be ΔG_{rxn}^o for the reaction that forms one mole of a compound from elements in their standard states (similarly to how ΔH_f^o is defined.) Secondly, tabulated (ΔG_f^o) can be used to calculate standard reaction free energies (ΔG_{rxn}^o) in much the same way as ΔH_f^o is used for reaction enthalpies.

Example 5.1.1:

Given the following data at 298 K, calculate ΔG^o at 298 K for the following reaction:

$$C_2H_4(g) + H_2(g) \rightarrow C_2H_6(g)$$

Substance	ΔG_f^o (kJ/mol)
$C_2H_4(g)$	68.4
$C_2H_6(g)$	-32.0

Solution

The ΔG_f^o values can be used to calculate ΔG^o for the reaction in exactly the same method as ΔH_f^o can be used to calculate a reaction enthalpy.

$$\Delta G^o = (1\,mol)(-32.0\,kJ/mol) - (1\,mol)(68.4\,kJ/mol)$$

$$\Delta G^o = 100.4\,kJ$$

Note: $H_2(g)$ is not included in the calculation since ΔG_f^o for $H_2(g)$ is 0 since it is an element in its standard state.

This page titled 5.1: Free Energy Functions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 6.1: Free Energy Functions by Patrick Fleming is licensed CC BY-NC-SA 4.0.

5.2: ΔA, ΔG, and Maximum Work

The functions A and G are oftentimes referred to as **free energy** functions. The reason for this is that they are a measure of the maximum work (in the case of ΔA) or non p-V work (in the case of ΔG) that is available from a process. To show this, consider the total differentials.

First, consider the differential of A.

$$dA = dU - TdS - SdT$$

Substituting the combined first and second laws for dU, but expressing the work term as dw, yields

$$dA = TdS - dw - TdS - SdT$$

And cancelling the TdS terms gives

$$dA = dw - SdT$$

or at constant temperature ($dT = 0$)

$$dA = dw$$

Since the only assumption made here was that the change is reversible (allowing for the substitution of TdS for dq), and dw for a reversible change is the maximum amount of work, it follows that dA gives the maximum work that can be produced from a process at constant temperature.

Similarly, a simple expression can be derived for dG. Starting from the total differential of G.

$$dG = dU + pdV\text{–}pdV + Vdp\text{–}TdS\text{–}SdT$$

Using an expression for $dU = dq + dw$, where $dq = TdS$ and dw is split into two terms, one (dw_{pV}) describing the work of expansion and the other (dw_e) describing any other type of work (electrical, stretching, etc.)

$$dU - TdS + dW_{pV} + dW_e$$

dG can be expressed as

$$dG = \cancel{TdS} - \cancel{pdV} + dw_e + \cancel{pdV} + Vdp\text{–}\cancel{TdS}\text{–}SdT$$

Cancelling the TdS and pdV terms leaves

$$dG = +dw_e + Vdp\text{–}SdT$$

So at constant temperature ($dT = 0$) and pressure ($dp = 0$),

$$dG = dw_e$$

This implies that dG gives the maximum amount of non p-V work that can be extracted from a process.

This concept of dA and dG giving the maximum work (under the specified conditions) is where the term “free energy” comes from, as it is the energy that is *free* to do work in the surroundings. If a system is to be optimized to do work in the soundings (for example a steam engine that may do work by moving a locomotive) the functions A and G will be important to understand. It will, therefore, be useful to understand how these functions change with changing conditions, such as volume, temperature, and pressure.

Contributors

- Patrick E. Fleming (Department of Chemistry and Biochemistry; California State University, East Bay)

This page titled 5.2: ΔA, ΔG, and Maximum Work is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 6.3: ΔA, ΔG, and Maximum Work by Patrick Fleming is licensed CC BY-NC-SA 4.0.

5.3: Pressure Dependence of Gibbs Energy

The pressure and temperature dependence of G is also easy to describe. The best starting place is the definition of G.

$$G = U + pV - TS \tag{5.3.1}$$

Taking the total differential of G yields

$$dG = dU + pdV - pdV + Vdp - TdS - SdT$$

The differential can be simplified by substituting the combined first and second law statement for dU (consider a reversible process and pV work only).

$$dG = \cancel{TdS} \cancel{-pdV} + \cancel{pdV} + Vdp - \cancel{TdS} - SdT$$

Canceling the TdS and pdV terms leaves

$$dG = V\,dp - S\,dT \tag{5.3.2}$$

This suggests that the natural variables of G are p and T. So the total differential dG can also be expressed

$$dG = \left(\frac{\partial G}{\partial p}\right)_T dp + \left(\frac{\partial G}{\partial T}\right)_p dT \tag{5.3.3}$$

And by inspection of Equations 5.3.2 and 5.3.3, it is clear that

$$\left(\frac{\partial G}{\partial p}\right)_T = V$$

and

$$\left(\frac{\partial G}{\partial T}\right)_p = -S$$

It is also clear that the Maxwell relation on G is given by

$$\left(\frac{\partial V}{\partial T}\right)_p = \left(\frac{\partial S}{\partial p}\right)_T$$

which is an extraordinarily useful relationship, since one of the terms is expressible entirely in terms of measurable quantities!

$$\left(\frac{\partial V}{\partial T}\right)_p = V\alpha$$

The pressure dependence of G is given by the pressure derivative at constant temperature

$$\left(\frac{\partial G}{\partial p}\right)_T = V \tag{5.3.4}$$

which is simply the molar volume. For a fairly incompressible substance (such as a liquid or a solid) the molar volume will be essentially constant over a modest pressure range.

Example 5.3.1: Gold under Pressure

The density of gold is 19.32 g/cm^3. Calculate ΔG for a 1.00 g sample of gold when the pressure on it is increased from 1.00 atm to 2.00 atm.

Solution

The change in the Gibbs function due to an isothermal change in pressure can be expressed as

$$\Delta G = \int_{p_1}^{p_2} \left(\frac{\partial G}{\partial p}\right)_T dp$$

And since substituting Equation 5.3.4, results in

$$\Delta G = \int_{p_1}^{p_2} V dp$$

Assuming that the molar volume is independent or pressure over the stated pressure range, ΔG becomes

$$\Delta G = V(p_2 - p_1)$$

So, the molar change in the Gibbs function can be calculated by substituting the relevant values.

$$\Delta G = \left(\frac{197.0\,g}{mol} \times \frac{1}{19.32\,g} \times \frac{1\,L}{1000\,cm^3}\right)(2.00\,atm - 1.00\,atm)\underbrace{\left(\frac{8.315\,J}{0.08206\,atm\,L}\right)}_{\text{conversion unit}} \tag{5.3.5}$$

$$= 1.033\,J \tag{5.3.6}$$

Contributors

- Patrick E. Fleming (Department of Chemistry and Biochemistry; California State University, East Bay)

This page titled 5.3: Pressure Dependence of Gibbs Energy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 6.5: Pressure Dependence of Gibbs Energy by Patrick Fleming is licensed CC BY-NC-SA 4.0.

5.4: Temperature Dependence of A and G

In differential form, the free energy functions can be expressed as

$$dA = -pdV - SdT$$

and

$$dG = -Vdp - SdT$$

So by inspection, it is easy to see that

$$\left(\frac{\partial A}{\partial T}\right)_V = -S$$

and

$$\left(\frac{\partial G}{\partial T}\right)_p = -S$$

And so, it should be fairly straightforward to determine how each changes with changing temperature:

$$\Delta A = -\int_{T_1}^{T_2} \left(\frac{\partial A}{\partial T}\right)_V dT = -\int_{T_1}^{T_2} S\, dT$$

and

$$\Delta G = -\int_{T_1}^{T_2} \left(\frac{\partial G}{\partial T}\right)_p dT = -\int_{T_1}^{T_2} S\, dT$$

But the temperature dependence of the entropy needed to be known in order to evaluate the integral. A convenient work-around can be obtained starting from the definitions of the free energy functions.

$$A = U - TS$$

and

$$G = H - TS$$

Dividing by T yields

$$\frac{A}{T} = \frac{U}{T} - S$$

and

$$\frac{G}{T} = \frac{H}{T} - S$$

Now differentiating each expression with respect to T at constant V or p respectively yields

$$\left(\frac{\partial \left(\frac{A}{T}\right)}{\partial T}\right)_V = -\frac{U}{T^2}$$

and

$$\left(\frac{\partial \left(\frac{G}{T}\right)}{\partial T}\right)_p = -\frac{H}{T^2}$$

Or differentiating with respect to $1/T$ provides a simpler form that is mathematically equivalent:

$$\left(\frac{\partial\left(\frac{A}{T}\right)}{\partial\left(\frac{1}{T}\right)}\right)_V = U$$

and

$$\left(\frac{\partial\left(\frac{G}{T}\right)}{\partial\left(\frac{1}{T}\right)}\right)_p = H$$

Focusing on the second expression (since all of the arguments apply to the first as well), we see a system that can be integrated. Multiplying both sides by $d(1/T)$ yields:

$$d\left(\frac{G}{T}\right) = Hd\left(\frac{1}{T}\right)$$

Or for finite changes ΔG and ΔH:

$$d\left(\frac{\Delta G}{T}\right) = \Delta Hd\left(\frac{1}{T}\right)$$

and integration, assuming the *enthalpy change is constant over the temperature* interval yields

$$\int_{T_1}^{T_2} d\left(\frac{\Delta G}{T}\right) = \Delta H \int_{T_1}^{T_2} d\left(\frac{1}{T}\right)$$

$$\frac{\Delta G_{T_2}}{T_2} - \frac{\Delta G_{T_1}}{T_1} = \Delta H\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \tag{5.4.1}$$

Equation ??? is the **Gibbs-Helmholtz equation** and can be used to determine how ΔG changes with changing temperature. The equivalent equation for the Helmholtz function is

$$\frac{\Delta A_{T_2}}{T_2} - \frac{\Delta A_{T_1}}{T_1} = \Delta U\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \tag{5.4.2}$$

Example 5.4.1:

Given the following data at 298 K, calculate ΔG at 500 K for the following reaction:

$$CH_4(g) + 2O_2(g) \rightarrow CO_2(g) + H_2O(g)$$

Compound	ΔG_f^o (kJ/mol)	ΔH_f^o (kJ.mol)
CH_4(g)	-50.5	-74.6
CO_2(g)	-394.4	-393.5
H_2O(g)	-228.6	-241.8

Solution

ΔH and $\Delta G_{298\,K}$ and can be calculated fairly easily. It will be assumed that ΔH is constant over the temperature range of 298 K – 500 K.

$$\Delta H = (1\,mol)(-393.5\,kJ/mol) + (2\,mol)(-241.8\,kJ/mol) - (1\,mol)(-74.5\,kJ/mol) = -820.6\,kJ$$

$$\Delta G_{298} = (1\,mol)(-394.4\,kJ/mol) + (2\,mol)(-228,6\,kJ/mol) - (1\,mol)(-50.5\,kJ/mol) = -801.1\,kJ$$

So using Equation 5.4.1 with the data just calculated gives

$$\frac{\Delta G_{500\,K}}{500\,K} - \frac{-801.1\,kJ}{298\,K} = (-820.6\,kJ)\left(\frac{1}{500\,K} - \frac{1}{298\,K}\right)$$

$$\Delta G_{500\,K} = -787.9\,kJ$$

Note: ΔG became a little bit less negative at the higher temperature, which is to be expected for a reaction which is exothermic. An increase in temperature should tend to make the reaction less favorable to the formation of products, which is exactly what is seen in this case!

This page titled 5.4: Temperature Dependence of A and G is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 6.6: Temperature Dependence of A and G by Patrick Fleming is licensed CC BY-NC-SA 4.0.

5.5: When Two Variables Change at Once

So far, we have derived a number of expressions and developed methods for evaluating how thermodynamic variables change as one variable changes while holding the rest constant. But real systems are seldom this accommodating. For example, a piece of metal (such as a railroad rail) left in the sun will undergo both an increase in temperature and an expansion due to the absorption of energy from sunlight. So both T and V are changing at the same time! If the change in a thermodynamic variable (such as G) is needed, contributions from both changes are required to be taken into account. We've already seen how to express this in terms of a total differential.

$$dG = \left(\frac{\partial G}{\partial p}\right)_T dp + \left(\frac{\partial G}{\partial T}\right)_p dT \tag{5.5.1}$$

Fortunately, G (like the other thermodynamic functions U, H, S, and A) is kind enough to be a state variable. This means that we can consider the changes independently and then simply add the results. Another way to think of this is that the system may follow either of two pathways to get from the initial conditions to the final conditions:

- *Pathway I:*
    1. An isothermal expansion from V_1 to V_2 at T_1 followed by
    2. An isochoric temperature increase from T_1 to T_2 at V_2
- *Pathway 2:*
    1. An isochoric temperature increase from T_1 to T_2 at V_1 followed by
    2. And isothermal expansion from V_1 to V_2 at T_2

And since G has the good sense to be a state variable, the pathway connecting the initial and final states is unimportant. We are free to choose any path that is convenient to calculate the change.

Example 5.5.1: Non-Isothermal Gas Expansion

Calculate the entropy change for 1.00 mol of a monatomic ideal gas (C_V = 3/2 R) expanding from 10.0 L at 273 K to 22.0 L at 297 K.

Solution

If one considers entropy to be a function of temperature and volume, one can write the total differential of entropy as

$$dS = \left(\frac{\partial S}{\partial T}\right)_V dT + \left(\frac{\partial S}{\partial V}\right)_T dV$$

and thus

$$\Delta S = \int_{T_1}^{T_2} \left(\frac{\partial S}{\partial T}\right)_V dT + \int_{V_1}^{V_2} \left(\frac{\partial S}{\partial V}\right)_T dV$$

The first term is the contribution due to an *isochoric temperature* change:

$$\Delta S_{T_1 \to T_2} = \int_{T_1}^{T_2} \left(\frac{\partial S}{\partial T}\right)_V dT \tag{5.5.2}$$

$$= \int_{T_1}^{T_2} \frac{nC_V}{T} dT \tag{5.5.3}$$

$$= nC_V \ln\left(\frac{T_2}{T_1}\right) \tag{5.5.4}$$

$$= (1.00\, mol)\left(\frac{3}{2} \cdot 8.314 \frac{J}{mol\, K}\right) \ln\left(\frac{297\, K}{273\, K}\right) \tag{5.5.5}$$

$$= 13.57\, J/K \tag{5.5.6}$$

The second term is the contribution due to an isothermal expansion:

$$\Delta S_{V_1 \to V_2} = \int_{V_1}^{V_2} \left(\frac{\partial S}{\partial V}\right)_T dV \tag{5.5.7}$$

From the Maxwell relation on A

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V$$

So Equation 5.5.7 becomes

$$\Delta S_{V_1 \to V_2} = \int_{V_1}^{V_2} \left(\frac{\partial p}{\partial T}\right)_V dV \tag{5.5.8}$$

$$= \int_{V_1}^{V_2} \left(\frac{nR}{V}\right) dV \tag{5.5.9}$$

$$= nR \ln\left(\frac{V_2}{V_1}\right) \tag{5.5.10}$$

$$= (1.00\, mol)\left(8.314 \frac{J}{mol\, K}\right) \ln\left(\frac{22.0\, L}{10.0\, L}\right) \tag{5.5.11}$$

$$= 6.56\, J/K \tag{5.5.12}$$

And the total entropy change is

$$\Delta S_{tot} = \Delta S_{V_1 \to V_2} + \Delta S_{V_1 \to V_2} \tag{5.5.13}$$

$$= 13.57\, J/K + 6.56\, J/K \tag{5.5.14}$$

$$= 20.13\, J/K \tag{5.5.15}$$

Deriving an expression for a partial derivative (Type III)

Thermodynamics involves many variables. But for a single component sample of matter, only two state variables are needed to describe the system and fix all of the thermodynamic properties of the system. As such, it is conceivable that two functions can be specified as functions of the same two variables. In general terms: $z(x, y)$ and $w(x, y)$.

So an important question that can be answered is, “What happens to z if w is held constant, but x is changed?” To explore this, consider the total differential of z:

$$dz = \left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x dy \tag{5.5.16}$$

but z can also be considered a function of x and $w(x, y)$. This implies that the total differential can also be written as

$$dz = \left(\frac{\partial z}{\partial x}\right)_w dx + \left(\frac{\partial z}{\partial w}\right)_x dy \tag{5.5.17}$$

and these two total differentials must be equal to one another!

$$= \left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x dy = \left(\frac{\partial z}{\partial x}\right)_w dx + \left(\frac{\partial z}{\partial w}\right)_x dw$$

If we constrain the system to a change in which w remains constant, the last term will vanish since $dw = 0$.

$$\left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x dy = \left(\frac{\partial z}{\partial x}\right)_w dx \tag{5.5.18}$$

but also, since w is a function x and y, the total differential for w can be written

$$dw = \left(\frac{\partial w}{\partial x}\right)_y dx + \left(\frac{\partial w}{\partial y}\right)_x dy$$

And it too must be zero for a process in which w is held constant.

$$0 = \left(\frac{\partial w}{\partial x}\right)_y dx + \left(\frac{\partial w}{\partial y}\right)_x dy$$

From this expression, it can be seen that

$$dy = -\left(\frac{\partial w}{\partial x}\right)_y \left(\frac{\partial y}{\partial w}\right)_x dx$$

Substituting this into the Equation 5.5.18, yields

$$\left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x \left[-\left(\frac{\partial w}{\partial x}\right)_y \left(\frac{\partial y}{\partial w}\right)_x dx\right] = \left(\frac{\partial z}{\partial x}\right)_w dx \tag{5.5.19}$$

which simplifies to

$$\left(\frac{\partial z}{\partial x}\right)_y dx - \left(\frac{\partial z}{\partial w}\right)_x \left(\frac{\partial w}{\partial x}\right)_y dx = \left(\frac{\partial z}{\partial x}\right)_w dx$$

So for $dx \neq 0$, implies that

$$\left(\frac{\partial z}{\partial x}\right)_y - \left(\frac{\partial z}{\partial w}\right)_x \left(\frac{\partial w}{\partial x}\right)_y = \left(\frac{\partial z}{\partial x}\right)_w$$

or

$$\left(\frac{\partial z}{\partial x}\right)_y = \left(\frac{\partial z}{\partial x}\right)_w + \left(\frac{\partial z}{\partial w}\right)_x \left(\frac{\partial w}{\partial x}\right)_y \tag{5.5.20}$$

As with partial derivative transformation types I and II, this result can be achieved in a formal, albeit less mathematically rigorous method.

Consider $z(x, w)$. This allows us to write the total differential for z:

$$dz = \left(\frac{\partial z}{\partial x}\right)_w dx + \left(\frac{\partial z}{\partial w}\right)_x dw$$

Now, divide by dx and constrain to constant y.

$$\left.\frac{dz}{dx}\right|_y = \left(\frac{\partial z}{\partial x}\right)_w \left.\frac{dx}{dx}\right|_y + \left(\frac{\partial z}{\partial w}\right)_x \left.\frac{dw}{dx}\right|_y$$

noting that $dx/dx = 1$ and converting the other ratios to partial derivatives yields

$$\left(\frac{\partial z}{\partial x}\right)_y = \left(\frac{\partial z}{\partial x}\right)_w + \left(\frac{\partial z}{\partial w}\right)_x \left(\frac{\partial w}{\partial x}\right)_y \tag{5.5.21}$$

which agrees with the previous result (Equation 5.5.20)! Again, the method is not mathematically rigorous, but it works so long as w, x, y, and z are **state functions** and the total differentials dw, dx, dy, and dz are **exact**.

Contributors

- Patrick E. Fleming (Department of Chemistry and Biochemistry; California State University, East Bay)

This page titled 5.5: When Two Variables Change at Once is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 6.7: When Two Variables Change at Once by Patrick Fleming is licensed CC BY-NC-SA 4.0.

5.6: Thermodynamic Functions have Natural Variables

The fundamental thermodynamic equations follow from five primary thermodynamic definitions and describe internal energy, enthalpy, Helmholtz energy, and Gibbs energy in terms of their natural variables. Here they will be presented in their differential forms.

Introduction

The fundamental thermodynamic equations describe the thermodynamic quantities U, H, G, and A in terms of their natural variables. The term "natural variable" simply denotes a variable that is one of the convenient variables to describe U, H, G, or A. When considered as a whole, the four fundamental equations demonstrate how four important thermodynamic quantities depend on variables that can be controlled and measured experimentally. Thus, they are essentially equations of state, and using the fundamental equations, experimental data can be used to determine sought-after quantities like G or H.

First Law of Thermodynamics

The first law of thermodynamics is represented below in its differential form

$$dU = đq + đw$$

where

- U is the internal energy of the system,
- q is heat flow of the system, and
- w is the work of the system.

The "đ" symbol represent **inexact differentials** and indicates that both q and w are path functions. Recall that U is a state function. The first law states that internal energy changes occur only as a result of heat flow and work done.

It is assumed that w refers only to PV work, where

$$w = -\int pdV$$

The fundamental thermodynamic equation for internal energy follows directly from the first law and the principle of Clausius:

$$dU = đq + đw$$
$$dS = \frac{\delta q_{rev}}{T}$$

we have

$$dU = TdS + \delta w$$

Since only PV work is performed,

$$dU = TdS - pdV \tag{5.6.1}$$

The above equation is the fundamental equation for U with natural variables of entropy S and volume$V$.

Principle of Clausius

The *Principle of Clausius* states that the entropy change of a system is equal to the ratio of heat flow in a reversible process to the temperature at which the process occurs. Mathematically this is written as

$$dS = \frac{\delta q_{rev}}{T}$$

where

- S is the entropy of the system,
- q_{rev} is the heat flow of a reversible process, and
- T is the temperature in Kelvin.

Enthalpy

Mathematically, enthalpy is defined as

$$H = U + pV \tag{5.6.2}$$

where H is enthalpy of the system, p is pressure, and V is volume. The fundamental thermodynamic equation for enthalpy follows directly from it deffinition (Equation 5.6.2) and the fundamental equation for internal energy (Equation 5.6.1) :

$$\begin{aligned} dH &= dU + d(pV) \\ &= dU + pdV + VdP \\ dU &= TdS - pdV \\ dH &= TdS - pdV + pdV + Vdp \\ dH &= TdS + Vdp \end{aligned}$$

The above equation is the fundamental equation for H. The natural variables of enthalpy are S and p, entropy and pressure.

Gibbs Energy

The mathematical description of Gibbs energy is as follows

$$G = U + pV - TS = H - TS \tag{5.6.3}$$

where G is the Gibbs energy of the system. The fundamental thermodynamic equation for Gibbs Energy follows directly from its definition 5.6.3 and the fundamental equation for enthalpy 5.6.2:

$$\begin{aligned} dG &= dH - d(TS) \\ &= dH - TdS - SdT \end{aligned}$$

Since

$$dH = TdS + VdP$$

$$dG = TdS + VdP - TdS - SdT$$

$$dG = VdP - SdT \tag{5.6.4}$$

The above equation is the fundamental equation for G. The natural variables of Gibbs energy are P and T.

Helmholtz Energy

Mathematically, Helmholtz energy is defined as

$$A = U - TS \tag{5.6.5}$$

where A is the Helmholtz energy of the system, which is often written as the symbol F. The fundamental thermodynamic equation for Helmholtz energy follows directly from its definition (Equation 5.6.5) and the fundamental equation for internal energy (Equation 5.6.1):

$$dA = dU - d(TS)$$
$$= dU - TdS - SdT$$

Since

$$dU = TdS - pdV$$
$$dA = TdS - pdV - TdS - SdT$$
$$dA = -pdV - SdT \tag{5.6.6}$$

Equation 5.6.6 is the fundamental equation for A with natural variables of V and T. For the definitions to hold, it is assumed that **only** PV work is done and that **only reversible** processes are used. These assumptions are required for the first law and the principle of Clausius to remain valid. Also, these equations do not account include n, the number of moles, as a variable. When n is included, the equations appear different, but the essence of their meaning is captured without including the n-dependence.

Chemical Potential

The fundamental equations derived above were not dependent on changes in the amounts of species in the system. Below the n-dependent forms are presented[1,4].

$$dU = TdS - PdV + \sum_{i=1}^{N} \mu_i dn_i$$
$$dH = TdS + VdP + \sum_{i=1}^{N} \mu_i dn_i$$
$$dG = -SdT + Vdp + \sum_{i=1}^{N} \mu_i dn_i$$
$$dA = -SdT - PdV + \sum_{i=1}^{N} \mu_i dn_i$$

where μ_i is the chemical potential of species i and dn_i is the change in number of moles of substance i.

Importance/Relevance of Fundamental Equations

The differential fundamental equations describe U, H, G, and A in terms of their natural variables. The natural variables become useful in understanding not only how thermodynamic quantities are related to each other, but also in analyzing relationships between measurable quantities (i.e. P, V, T) in order to learn about the thermodynamics of a system. Below is a table summarizing the natural variables for U, H, G, and A:

Thermodynamic Quantity	Natural Variables
U (internal energy)	S, V
H (enthalpy)	S, P
G (Gibbs energy)	T, P
A (Helmholtz energy)	T, V

References

1. DOI: 10.1063/1.1749582

2. DOI: 10.1063/1.1749549
3. DOI:10.1103/PhysRev.3.273
4. *A Treatise on Physical Chemistry,* 3rd ed.; Taylor, H. S. and Glasstone, S., Eds.; D. Van Nostrand Company: New York, 1942; Vol. 1; p 454-485.

Problems

1. If the assumptions made in the derivations above were not made, what would effect would that have? Try to think of examples were these assumptions would be violated. Could the definitions, principles, and laws used to derive the fundamental equations still be used? Why or why not?
2. For what kind of system does the number of moles not change? This said, do the fundamental equations without n-dependence apply to a wide range of processes and systems?

Answers

1. If it was not assumed that PV-work was the only work done, then the work term in the second law of thermodynamics equation would include other terms (e.g. for electrical work, mechanical work). If reversible processes were not assumed, the Principle of Clausius could not be used. One example of such situations could the movement of charged particles towards a region of like charge (electrical work) or an irreversible process like combustion of hydrocarbons or friction.
2. In general, a closed system of non-reacting components would fit this description. For example, the number of moles would not change for a closed system in which a gas is sealed (to prevent leaks) in a container and allowed to expand/is contracted.

Contributors and Attributions

- Andreana Rosnik, Hope College

This page titled 5.6: Thermodynamic Functions have Natural Variables is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 22.5: Thermodynamic Functions have Natural Variables by Andreana Rosnik is licensed CC BY-NC-SA 4.0.

5.7: The Gibbs-Helmholtz Equation

Ideal gas

For a mole of ideal gas we can use the gas law to integrate volume over pressure and we get

$$\Delta G_{molar} = RT \ln\left(\frac{P_2}{P_1}\right)$$

It is customary to identify one of the pressures (P_1) with the *standard state* of **1 bar** and use the plimsoll to indicate the fact that we are referring to a standard state by writing:

$$G_{molar}(P) = G^o_{molar} + RT \ln\left(\frac{P}{1}\right) = G^o_{molar} + RT \ln[P]$$

The fact that we are making the function intensive (per mole) is usually indicated by putting a bar over the G symbol, although this is often omitted for $G^{\ominus}_{molar}$

The Gibbs-Helmholtz Expression

$$\frac{G}{T} = \frac{H}{T} - S$$

Take the derivative under constant pressure of each side to get

$$\left(\frac{\partial G/T}{\partial T}\right)_P = -\frac{H}{T^2} + \frac{1}{T}\left(\frac{\partial H}{\partial T}\right)_P - \left(\frac{\partial S}{\partial T}\right)_P$$

We make use of the relationship between C_p and H and C_p and S

$$\left(\frac{\partial G/T}{\partial T}\right)_P = -\frac{H}{T^2} + \frac{\cancel{C_p}}{\cancel{T}} - \frac{\cancel{C_p}}{\cancel{T}} \tag{5.7.1}$$

$$= -\frac{H}{T^2} \tag{5.7.2}$$

We said before that S is a first order derivative of G. As you can see from this derivation the enthalpy H is also a first order derivative, albeit not of G itself, but of G/T.

$$\left(\frac{\partial \Delta G/T}{\partial T}\right)_P = -\frac{\triangle H}{T^2}$$

The last step in the derivation simply takes the step before twice -say for the G and H at the begin and end of a *process*- and subtracts the two identical equations leading to a Δ symbol. In this differential form the Gibbs-Helmholtz equation can be applied to any process.

Gibbs Energy as a Function of Temperature

If heat capacities are know from 0 K we could determine both enthalpy and entropy by integration:

$$S(T) = S(0) + \int_0^T \frac{C_p}{T} dT$$

$$H(T) = H(0) + \int_0^T C_p \, dT$$

As we have seen we must be careful at phase transitions such as melting or vaporization. At these points the curves are

discontinuous and the derivative C_p is undefined.

$$H(T) = H(0) + \int_0^{T_{fus}} C_p(T)_{solid} dT + \Delta H_{fus} + \int_{T_{fus}}^{T_{boil}} C_p(T)_{liquid} dT + \Delta H_{vap} + etc. \tag{5.7.3}$$

$$S(T) = S(0) + \int_0^{T_{fus}} \frac{C_p(T)_{solid}}{T} dT + \Delta S_{fus} + \int_{T_{fus}}^{T_{boil}} \frac{C_p(T)_{liquid}}{T} dT + \Delta S_{vap} + etc. \tag{5.7.4}$$

$$= S(0) + \int_0^{T_{fus}} \frac{C_p(T)_{solid}}{T} dT + \frac{\Delta H_{fus}}{T_{fus}} + \int_{T_{fus}}^{T_{boil}} \frac{C_p(T)_{liquid}}{T} dT + \frac{\Delta H_{vap}}{T_{boil}} + etc. \tag{5.7.5}$$

with $H(T = 0) = \text{undefined}$ and $S(T = 0) = 0$ from the third law of thermodynamics.

> We also discussed the fact that the third law allows us to define $S(0)$ as zero in most cases. For the enthalpy we cannot do that so that our curve is with respect to an undefined zero point. We really should plot $H(T) - H(0)$ and leave $H(0)$ undefined.

Because the Gibbs free energy $G = H - TS$ we can also construct a curve for G as a function of temperature, simply by combining the H and the S curves (Equations 5.7.3 and 5.7.5):

$$G(T) = H(T) - TS(T)$$

Interestingly, if we do so, the **discontinuties** at the phase transition points will **drop out** for G because at these points $\Delta_{trs}H = T_{trs}\Delta_{trs}S$. Therefore, G is **always** continuous.

The $H(0)$ problem does not disappear so that once again our curve is subject to an arbitrary offset in the y-direction. The best thing we can do is plot the quantity $G(T) - H(0)$ and leave the offset $H(0)$ undefined.

We have seen above that the derivative of G with temperature is $-S$. As entropy is always positive, this means that the G-curve is always *descending* with temperature. It also means that although the curve is *continuous* even at the phase transitions, the slope of the G curve is not, because the derivative $-S$ makes a jump there. Fig. 22.7 in the book shows an example of such a curve for benzene. Note the kinks in the curve at the mp and the boiling point.

This page titled 5.7: The Gibbs-Helmholtz Equation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 22.7: The Gibbs-Helmholtz Equation is licensed CC BY-NC-SA 4.0.

5.E: Gibbs Energy (Exercises)

Q6.1

Using Table T1, calculate the standard reaction Gibbs functions (\(\Delta G^o\)) for the following reactions at 298 K.

a. $CH_3CH_2OH(l) + 3O_2(g) \rightarrow 2CO_2(g) + 3H_2O(l)$
b. $C_6H_{12}O_6(s) + 6O_2 \rightarrow 6CO_2(g) + 6H_2O(l)$
c. $2POCl_3(l) \rightarrow 2PCl_3(l) + O_2(g)$
d. $2KBr(s) + Cl_2(g) \rightarrow 2KCl(s) + Br_2(l)$
e. $SiH_4(g) + 2Cl(g) \rightarrow SiCl_4(l) + 2H_2(g)$

Q6.2

Estimate \(\Delta G\) at 1000 K from its value at 298 K for the reaction

$$C(s) + 2H_2(g) \rightarrow CH_4(g) \tag{5.E.1}$$

with $\Delta G = -50.75\ kJ\ at\ 298\ K$

This page titled 5.E: Gibbs Energy (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

6: CHEMICAL KINETICS I

Physical Chemistry: Thermodynamics

Patrick Fleming

I II III IV V VI VII VIII IX X XI XII • Homework Exercises

Chemical kinetics is the study of how fast chemical reactions proceed from reactants to products. This is an important topic because while thermodynamics will tell us about the direction of spontaneous change, it is silent as to how fast processes will occur. But additionally, the power of studying reaction rates is that it gives us insight into the actual pathways chemical processes follow to proceed from reactants to products.

6.1: Reaction Rate

6.2: Rate Laws

6.3: 0th order Rate Law

6.4: 1st order rate law

6.5: 2nd order Rate Laws

6.6: The Method of Initial Rates

6.7: The Method of Half-Lives

6.8: Temperature Dependence

6.E: Chemical Kinetics I (Exercises)

6.S: Chemical Kinetics I (Summary)

This page titled 6: Chemical Kinetics I is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11: Chemical Kinetics I by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.1: Reaction Rate

The rate of a chemical reaction (or the **reaction rate**) can be defined by the time needed for a change in concentration to occur. But there is a problem in that this allows for the definition to be made based on concentration changes for either the reactants or the products. Plus, due to stoichiometric concerns, the rates at which the concentrations are generally different! Toward this end, the following convention is used.

For a general reaction

$$aA + bB \rightarrow cC + dD$$

the reaction rate can be defined by any of the ratios

$$\text{rate} = -\frac{1}{a}\frac{\Delta[A]}{dt} = -\frac{1}{b}\frac{\Delta[B]}{dt} = +\frac{1}{c}\frac{\Delta[C]}{dt} = +\frac{1}{d}\frac{\Delta[D]}{dt}$$

Or for infinitesimal time intervals

$$\text{rate} = -\frac{1}{a}\frac{d[A]}{dt} = -\frac{1}{b}\frac{d[C]}{dt} = +\frac{1}{c}\frac{d[C]}{dt} = +\frac{1}{d}\frac{d[D]}{dt}$$

Example 6.1.1:

Under a certain set of conditions, the rate of the reaction

$$N_2 + 3H_2 \rightarrow 2NH_3$$

the reaction rate is $6.0 \times 10^{-4}\ M/s$. Calculate the time-rate of change for the concentrations of N_2, H_2, and NH_3.

Solution

Due to the stoichiometry of the reaction,

$$\text{rate} = -\frac{d[N_2]}{dt} = -\frac{1}{3}\frac{d[H_2]}{dt} = +\frac{1}{2}\frac{d[NH_3]}{dt}$$

so

$$\frac{d[N_2]}{dt} = -6.0 \times 10^{-4}\, M/s$$

$$\frac{d[H_2]}{dt} = -2.0 \times 10^{-4}\, M/s$$

$$\frac{d[NH_3]}{dt} = 3.0 \times 10^{-4}\, M/s$$

Note: The time derivatives for the reactants are negative because the reactant concentrations are decreasing, and those of products are positive since the concentrations of products increase as the reaction progresses.

This page titled 6.1: Reaction Rate is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.1: Reaction Rate by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.2: Rate Laws

A rate law is any mathematical relationship that relates the concentration of a reactant or product in a chemical reaction to time. Rate laws can be expressed in either derivative (or ratio, for finite time intervals) or integrated form. One of the more common general forms a rate law for the reaction

$$A + B \rightarrow products$$

may take is

$$\text{rate} = k[A]^{\alpha}[B]^{\beta}$$

where k, α, and β are experimentally determined rate order values. However, a rate law can take many different forms, some of which can be quite intricate and complex. The powers α and β need not be integers. For example

$$\text{rate} = k[A]^{\alpha}[B]^{1/2} \tag{6.2.1}$$

is a rate law that is observed for some reactions. Sometimes, the concentrations of products must be included.

$$\text{rate} = \frac{k[A]^{1/2}[B]}{[P]}$$

In some cases, the concentration for a catalyst or enzyme is important. For example, many enzyme mitigated reactions in biological systems follow the Michaelis-Menten rate law, which is of the form

$$\text{rate} = \frac{V_{max}[S]}{K_m + [S]}$$

where V_{max} and K_M are factors that are determined experimentally, and $[S]$ is the concentration of the substrate in the reaction.

Order

For those cases where the rate law can be expressed in the form

$$\text{rate} = k[A]^{\alpha}[B]^{\beta}[C]^{\gamma}$$

where A, B, and C are reactants (or products or catalysts, etc.) involved in the reaction, the reaction is said to be of α order in A, β order in B, and γ order in C. The reaction is said to be $\alpha + \beta + \gamma$ order overall. Some examples are shown in the following table:

Table 11.3.1: Example Rate Laws

Rate law	Order with respect to A	Order with respect to B	Order with respect to C	Overall order
$\text{rate} = k$	0	0	0	0
$\text{rate} = k[A]$	0	0	0	1
$\text{rate} = k[A]^2$	0	0	0	2
$\text{rate} = k[A][B]$	0	1	0	2
$\text{rate} = k[A]^2[B]$	0	1	0	3
$\text{rate} = k[A][B][C]$	0	1	1	3

Reaction orders can also be fractional such as for Equation 6.2.1 which is 1st order in A, and half order in B. The order can also be negative such as

$$\text{rate} = k\frac{[A]}{[B]}$$

which is 1st order in A, and -1 order in B. In this case, an build-up of the concentration of B will retard (slow) the reaction.

In all cases, the order of the reaction with respect to a specific reactant or product (or catalyst, or whatever) must be determined experimentally. As a general rule, the stoichiometry cannot be used to predict the form of the rate law. However, the rate law can be used to gain some insight into the possible pathways by which the reaction can proceed. That is the topic of Chapter 12. For now we will focus on three useful methods that are commonly used in chemistry to determine the rate law for a reaction from experimental data.

Empirical Methods

Perhaps the simplest of the methods to be used are the empirical methods, which rely on the qualitative interpretation of a graphical representation of the concentration vs time profile. In these methods, some function of concentration is plotted as a function of time, and the result is examined for a linear relationship. For the following examples, consider a reaction of the form

$$A + B \rightarrow products$$

in which A is one of the reactants. In order to employ these empirical methods, one must generate the forms of the integrated rate laws.

This page titled 6.2: Rate Laws is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.3: Rate Laws by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.3: 0th order Rate Law

If the reaction follows a zeroth order rate law, it can be expressed in terms of the time-rate of change of [A] (which will be negative since A is a reactant):

$$-\frac{d[A]}{dt} = k$$

In this case, it is straightforward to separate the variables. Placing time variables on the right and [A] on the left

$$d[A] = -k\,dt$$

In this form, it is easy to integrate. If the concentration of A is $[A]_0$ at time t = 0, and the concentration of A is [A] at some arbitrary time later, the form of the integral is

$$\int_{[A]_o}^{[A]} d[A] = -k\int_{t_o}^{t} dt$$

which yields

$$[A] - [A]_o = -kt$$

or

$$[A] = [A]_o - kt$$

This suggests that a plot of concentration as a function of time will produce a straight line, the slope of which is –k, and the intercept of which is $[A]_0$. If such a plot is linear, then the data are consistent with 0th order kinetics. If they are not, other possibilities must be considered.

This page titled 6.3: 0th order Rate Law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.4: 0th order Rate Law by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.4: 1st order rate law

A first order rate law would take the form

$$\frac{d[A]}{dt} = k[A]$$

Again, separating the variables by placing all of the concentration terms on the left and all of the time terms on the right yields

$$\frac{d[A]}{[A]} = -k\,dt$$

This expression is also easily integrated as before

$$\int_{[A]=0}^{[A]} \frac{d[A]}{[A]} = -k \int_{t=0}^{t=t} dt$$

Noting that

$$\frac{dx}{x} = d(\ln x)$$

The form of the integrated rate law becomes

$$\ln[A] - \ln[A]_o = kt$$

or

$$\ln[A] = \ln[A]_o - kt \tag{6.4.1}$$

This form implies that a plot of the natural logarithm of the concentration is a linear function of the time. And so a plot of ln[A] as a function of time should produce a linear plot, the slope of which is -k, and the intercept of which is $\ln[A]_0$.

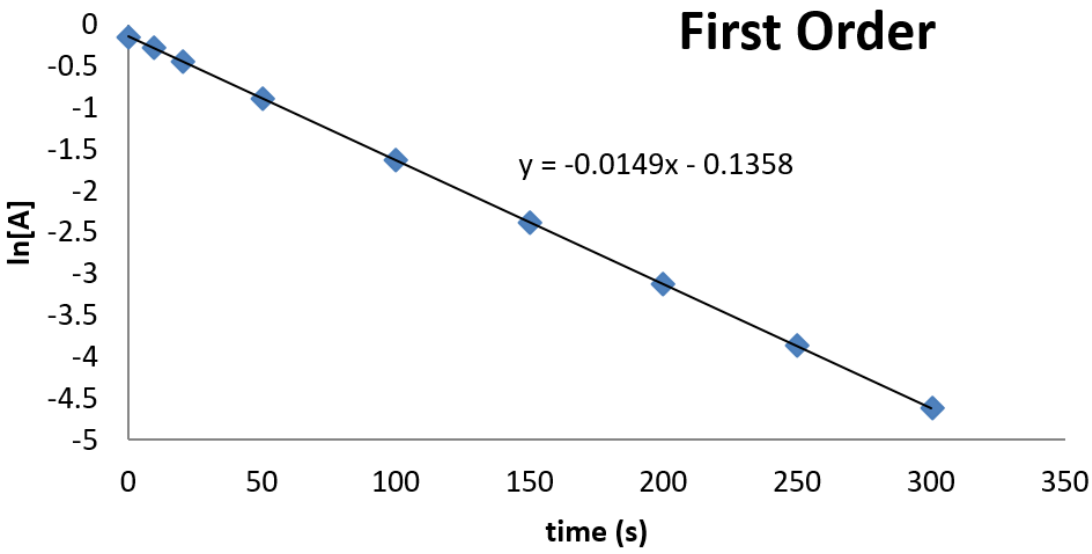
Example 6.4.1:

Consider the following kinetic data. Use a graph to demonstrate that the data are consistent with first order kinetics. Also, if the data are first order, determine the value of the rate constant for the reaction.

Time (s)	0	10	20	50	100	150	200	250	300
[A] (M)	0.873	0.752	0.648	0.414	0.196	0.093	0.044	0.021	0.010

Solution

The plot looks as follows:



From this plot, it can be seen that the rate constant is 0.0149 s^{-1}. The concentration at time $t = 0$ can also be inferred from the intercept.

It should also be noted that the integrated rate law (Equation 6.4.1) can be expressed in exponential form:

$$[A] = [A]_o e^{-kt}$$

Because of this functional form, 1st order kinetics are sometimes referred to as exponential decay kinetics. Many processes, including radioactive decay of nuclides follow this type of rate law.

This page titled 6.4: 1st order rate law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.5: 1st order rate law by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.5: 2nd order Rate Laws

If the reaction follows a second order rate law, the some methodology can be employed. The rate can be written as

$$-\frac{d[A]}{dt} = k[A]^2 \tag{6.5.1}$$

The separation of concentration and time terms (this time keeping the negative sign on the left for convenience) yields

$$-\frac{d[A]}{[A]^2} = k\,dt$$

The integration then becomes

$$-\int_{[A]_o}^{[A]} \frac{d[A]}{[A]^2} = \int_{t=0}^{t} k\,dt \tag{6.5.2}$$

And noting that

$$-\frac{dx}{x^2} = d\left(\frac{1}{x}\right)$$

the result of integration Equation 6.5.2 is

$$\frac{1}{[A]} - \frac{1}{[A]_o} = kt$$

or

$$\frac{1}{[A]} = \frac{1}{[A]_o} + kt$$

And so a plot of $1/[A]$ as a function of time should produce a linear plot, the slope of which is k, and the intercept of which is $1/[A]_0$.

Other 2nd order rate laws are a little bit trickier to integrate, as the integration depends on the actual stoichiometry of the reaction being investigated. For example, for a reaction of the type

$$A + B \rightarrow P$$

That has rate laws given by

$$-\frac{d[A]}{dt} = k[A][B]$$

and

$$-\frac{d[B]}{dt} = k[A][B]$$

the integration will depend on the decrease of [A] and [B] (which will be related by the stoichiometry) which can be expressed in terms the concentration of the product [P].

$$[A] = [A]_o - [P] \tag{6.5.3}$$

and

$$[B] = [B]_o - [P] \tag{6.5.4}$$

The concentration dependence on A and B can then be eliminated if the rate law is expressed in terms of the production of the

product.

$$\frac{d[P]}{dt} = k[A][B] \tag{6.5.5}$$

Substituting the relationships for $[A]$ and $[B]$ (Equations 6.5.3 and 6.5.4) into the rate law expression (Equation 6.5.5) yields

$$\frac{d[P]}{dt} = k([A]_o - [P])([B] = [B]_o - [P]) \tag{6.5.6}$$

Separation of concentration and time variables results in

$$\frac{d[P]}{([A]_o - [P])([B] = [B]_o - [P])} = k\,dt$$

Noting that at time $t = 0$, $[P] = 0$, the integrated form of the rate law can be generated by solving the integral

$$\int_{[A]_o}^{[A]} \frac{d[P]}{([A]_o - [P])([B]_o - [P])} = \int_{t=0}^{t} k\,dt$$

Consulting a table of integrals reveals that for $a \neq b$[1],

$$\int \frac{dx}{(a-x)(b-x)} = \frac{1}{b-a} \ln\left(\frac{b-x}{a-x}\right)$$

Applying the definite integral (as long as $[A]_0 \neq [B]_0$) results in

$$\frac{1}{[B]_0 - [A]_0} \ln\left(\frac{[B]_0 - [P]}{[A]_0 - [P]}\right)\Bigg|_0^{[A]} = k\,t|_0^t$$

$$\frac{1}{[B]_0 - [A]_0} \ln\left(\frac{[B]_0 - [P]}{[A]_0 - [P]}\right) - \frac{1}{[B]_0 - [A]_0} \ln\left(\frac{[B]_0}{[A]_0}\right) = k\,t \tag{6.5.7}$$

Substituting Equations 6.5.3 and 6.5.4 into Equation 6.5.7 and simplifying (combining the natural logarithm terms) yields

$$\frac{1}{[B]_0 - [A]_0} \ln\left(\frac{[B][A]_o}{[A][B]_o}\right) = kt$$

For this rate law, a plot of $\ln([B]/[A])$ as a function of time will produce a straight line, the slope of which is

$$m = ([B]_0 - [A]_0)k.$$

In the limit at $[A]_0 = [B]_0$, then $[A] = [B]$ at all times, due to the stoichiometry of the reaction. As such, the rate law becomes

$$\text{rate} = k[A]^2$$

and integrate direct like in Equation 6.5.1 and the integrated rate law is (as before)

$$\frac{1}{[A]} = \frac{1}{[A]_o} + kt$$

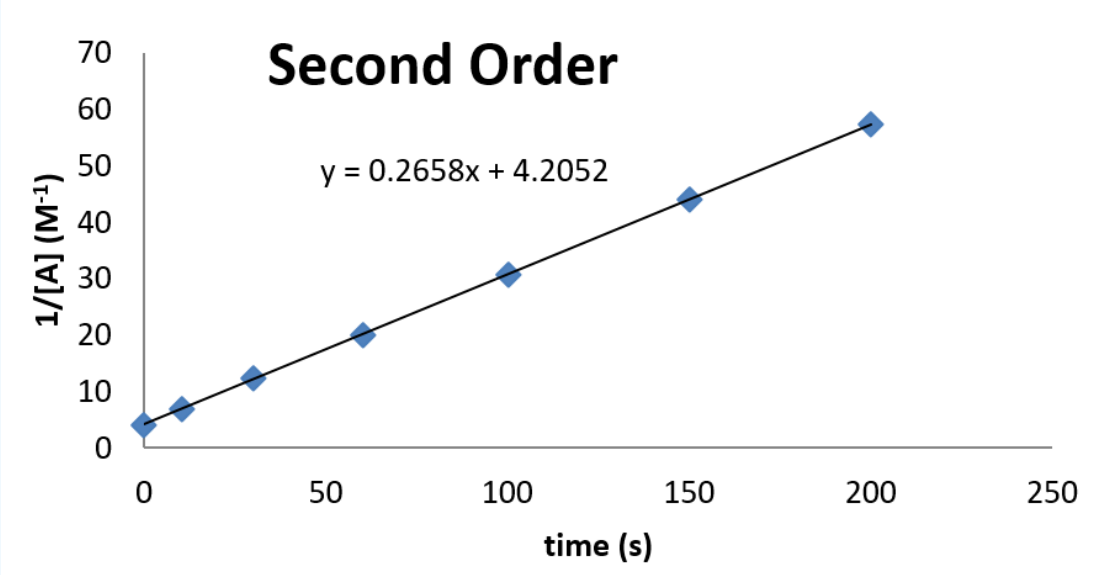
Example 6.5.2: Confirming Second Order Kinetics

Consider the following kinetic data. Use a graph to demonstrate that the data are consistent with second order kinetics. Also, if the data are second order, determine the value of the rate constant for the reaction.

time (s)	0	10	30	60	100	150	200
[A] (M)	0.238	0.161	0.098	0.062	0.041	0.029	0.023

Solution

The plot looks as follows:



From this plot, it can be seen that the rate constant is 0.2658 M^{-1} s^{-1}. The concentration at time $t = 0$ can also be inferred from the intercept.

---

[1] This integral form can be generated by using the method of partial fractions. See (House, 2007) for a full derivation.

This page titled 6.5: 2nd order Rate Laws is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.6: 2nd order Rate Laws by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.6: The Method of Initial Rates

The **method of initial rates** is a commonly used technique for deriving rate laws. As the name implies, the method involves measuring the initial rate of a reaction. The measurement is repeated for several sets of initial concentration conditions to see how the reaction rate varies. This might be accomplished by determining the time needed to exhaust a particular amount of a reactant (preferably one on which the reaction rate does not depend!). A typical set of data for a reaction:

$$A + B \rightarrow \text{products}$$

might appear as follows:

Run	[A] (M)	[B] (M)	Rate (M/s)
1	0.0100	0.0100	0.0347
2	0.0200	0.0100	0.0694
3	0.0200	0.0200	0.2776

The analysis of this data involves taking the ratios of rates measured where one of the concentrations does not change. For example, assuming a rate law of the form

$$\text{rate} = k[A]^{\alpha}[B]^{\beta} \tag{6.6.1}$$

The ratio of runs i and j generate the following relationship.

$$\frac{\text{rate}_i}{\text{rate}_j} = \frac{k[A]_i^{\alpha}[B]_i^{\beta}}{k[A]_j^{\alpha}[B]_j^{\beta}}$$

So using runs 1 and 2,

$$\frac{0.0347\,M/s}{0.0694\,M/s} = \frac{\cancel{k}\,(0.01\,M/s)^{\alpha}\,\cancel{(0.01\,M/s)^{\beta}}}{\cancel{k}\,(0.02\,M/s)^{\alpha}\,\cancel{(0.01\,M/s)^{\beta}}}$$

this simplifies to

$$\frac{1}{2} = \left(\frac{1}{2}\right)^{\alpha}$$

So clearly, $\alpha = 1$ and the reaction is 1st order in A. Taking the ratio using runs 2 and 3 yields

$$\frac{0.0694\,M/s}{0.2776\,M/s} = \frac{\cancel{k}\,\cancel{(0.02\,M)^{\alpha}}\,(0.01\,M)^{\beta}}{\cancel{k}\,\cancel{(0.02\,M)^{\alpha}}\,(0.02\,M)^{\beta}}$$

This simplifies to

$$\frac{1}{4} = \left(\frac{1}{2}\right)^{\beta} \tag{6.6.2}$$

By inspection, one can conclude that $\beta = 2$, and that the reaction is second order in B. But if it is not so clear (as it might not be if the concentration is not incremented by a factor of 2), the value of β can be determined by taking the natural logarithm of both sides of the Equation 6.6.2.

$$\begin{aligned}\ln\frac{1}{4} &= \ln\left(\frac{1}{2}\right)^{\beta} \\ &= \beta\ln\left(\frac{1}{2}\right)\end{aligned}$$

dividing both sides by $\ln(1/2)$

$$\frac{\ln\left(\frac{1}{4}\right)}{\ln\left(\frac{1}{2}\right)} = \beta\frac{\ln\left(\frac{1}{2}\right)}{\ln\left(\frac{1}{2}\right)}$$

or

$$\beta = \frac{-1.3863}{-0.69315} = 2$$

And so the rate law (Equation 6.6.1) can be expressed as

$$\text{rate} = k[A][B]^2$$

And is 1st order in A, 2nd order in B, and 3rd order overall. The rate constant can then be evaluated by substituting one of the runs into the rate law (or using all of the data and taking an average). Arbitrarily selecting the first run for this,

$$0.0347\,M/s = k(0.01\,M/s)(0.01\,M/s)^2$$

This results in a value of k

$$k = \frac{0.0347\,M/s}{(0.01\,M/s)(0.01\,M/s)^2} = 3.47 \times 10^5\,M^{-2}s^{-1}$$

It is useful to note that the units on k are consistent with a 3rd order rate law.

This page titled 6.6: The Method of Initial Rates is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.7: The Method of Initial Rates by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.7: The Method of Half-Lives

Another method for determining the order of a reaction is to examine the behavior of the **half-life** as the reaction progresses. The half-life can be defined as the time it takes for the concentration of a reactant to fall to half of its original value. The method of half-lives involved measuring the half-life's dependence on concentration. The expected behavior can be predicted using the integrated rate laws we derived earlier.

Using the definition of the half-life, at time $t_{1/2}$ the concentration $[A]$ drops to half of its original value, $[A]_0$.

$$[A] = \frac{1}{2}[A]_o$$

at $t = t_{1/2}$.

So if the reaction is 0th order in A, after one half-life

$$\frac{1}{2}[A]_o = [A]_o - kt_{1/2}$$

Solving for $t_{1/2}$ reveals the dependence of the half-life on the initial concentration.

$$\frac{[A]_o}{2k} = t_{1/2}$$

So as the original concentration is **decreased**, the half-life of a 0th order reaction will also **decrease**.

Similarly, for a first order reaction,

$$\frac{1}{2}[A]_o = [A]_o e^{-kt_{1/2}}$$

and solving for $t_{1/2}$ results in a concentration independent expression

$$\frac{\ln 2}{k} = t_{1/2}$$

It is because the half-life of a 1st order reaction is **independent** of concentration that it is oftentimes used to describe the rate of first order processes (such as radioactive decay.)

For a 2nd order reaction, the half-life can be expressed based on the integrated rate law.

$$\frac{1}{\frac{1}{2}[A]_o} = \frac{1}{[A]_o} + kt_{1/2}$$

solving for $t_{1/2}$ yields

$$\frac{1}{t_{1/2}} = t_{1/2}$$

In the case of a second order reaction, the half-life **increases** with **decreasing** initial concentration.

Table 6.7.1: Calculated half lives for Reactions following simple Rate Laws

Order	Half-life	Behavior
0th	$\frac{1}{2}[A]_o = [A]_o - kt_{1/2}$	Decreases as the reaction progresses (as [A] decreases
1st	$[\frac{\ln 2}{k} = t_{1/2}$	Remains constant as the reaction progresses (is independent of concentration)

Order	Half-life	Behavior
2nd	$\frac{1}{\frac{1}{2}[A]_o} = \frac{1}{[A]_o} + kt_{1/2}$	Increases with decreasing concentration.

For reactions in which the rate law depends on the concentration of more than one species, the half-life can take a much more complex form that may depend on the initial concentrations of multiple reactants, or for that matter, products!

Example 6.7.1: Radiocarbon Dating

Carbon-14 decays into nitrogen-14 with first order kinetics and with a half-life of 5730 years.

$$^{14}\mathrm{C} \rightarrow {}^{14}\mathrm{N}$$

What is the rate constant for the decay process? What percentage of carbon-14 will remain after a biological sample has stopped ingesting carbon-14 for 1482 years?

Solution

The rate constant is fairly easy to calculate:

$$t_{1/2} = \frac{\ln 2}{k} = \frac{\ln 2}{5730\,yr} = 1.21 \times 10^{-4}\,yr^{-1}$$

Now the integrated rate law can be used to solve the second part of the problem.

$$[^{14}\mathrm{C}] = [^{14}\mathrm{C}_o]e^{-kt}$$

this can be rewritten in term of relative loss of $[^{14}\mathrm{C}]$.

$$\frac{[^{14}\mathrm{C}]}{[^{14}\mathrm{C}_o]} = e^{-kt}$$

so

$$\frac{[^{14}\mathrm{C}]}{[^{14}\mathrm{C}_o]} = e^{-(1.21\times10^{-4}\,yr^{-1})(1482\,ys)} = 0.836$$

So after 1482 years, there is 83.6 % of $^{14}\mathrm{C}$ still left.

Example 6.7.2:

Based on the following concentration data as a function of time, determine the behavior of the half-life as the reaction progresses. Use this information to determine if the following reaction is 0th order, 1st order, or 2nd order in A. Also, use the data to estimate the rate constant for the reaction.

time (s)	[A] (M)
0	1.200
10	0.800
20	0.600
30	0.480
40	0.400

time (s)	[A] (M)
50	0.343
60	0.300
70	0.267
80	0.240
90	0.218
100	0.200

Solution

If the original concentration is taken as 1.200 M, half of the original concentration is 0.600 M. The reaction takes 20 seconds to reduce the concentration to half of its original value. If the original concentration is taken as 0.800 M, it clearly takes 30 seconds for the concentration to reach half of that value. Based on this methodology, the following table is easy to generate:

$[A]_o$ **(M)**	1.200	0.800	0.600	0.400
$t_{1/2}$ **(s)**	20	30	40	60

The rate constant can be calculated using any of these values:

$$k = \frac{1}{[A]t_{1/2}} \tag{6.7.1}$$

$$= \frac{1}{(0.8\,M)(30\,s)} \tag{6.7.2}$$

$$= 0.0417\,M^{-1}s^{-1} \tag{6.7.3}$$

This page titled 6.7: The Method of Half-Lives is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.8: The Method of Half-Lives by Patrick Fleming is licensed CC BY-NC-SA 4.0.

6.8: Temperature Dependence

This page titled 6.8: Temperature Dependence is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

6.E: Chemical Kinetics I (Exercises)

This page titled 6.E: Chemical Kinetics I (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

6.S: Chemical Kinetics I (Summary)

The results of the integration of these simple rate laws can be summarized in the following table.

Order	Elementary Reaction	Integrated rate law	Linear plot
0	-	$[A] = [A]_o - kt$	$[A]$ vs. t
1	$A \rightarrow P$	$\ln[A] = \ln[A]_o - kt$ $[[A] = [A]_o e^{-kt}$	$\ln[A]$ vs. t
2	$A + A \rightarrow P$	$\frac{1}{[A]} = \frac{1}{[A]_o} + kt$	$\frac{1}{[A]}$ vs. t
	$A + B \rightarrow P$	$\frac{1}{[B]_0 - [A]_0} \ln\left(\frac{[B][A]_o}{[A][B]_o}\right) = kt$	$\ln\left(\frac{[B]}{[A]}\right)$ vs. t

This page titled 6.S: Chemical Kinetics I (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 11.S: Chemical Kinetics I (Summary) by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7: CHEMICAL KINETICS II

Physical Chemistry: Thermodynamics

Patrick Fleming

In the previous chapter, we discussed the rates of chemical reactions. In this chapter, we will expand on the concepts of chemical reaction rates by exploring what the rate law implies about the mechanistic pathways that reactions actually follow to proceed from reactants to products. Typically, one determines a rate law that describes a chemical reaction, and then suggests a mechanism that can be (or might not be!) consistent with the observed kinetics. This chapter will be concerned with reconciling reaction mechanisms with predicted rate laws.

7.1: Reaction Mechanisms
7.2: Concentration Profiles for Some Simple Mechanisms
7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws
7.4: The Rate Determining Step Approximation
7.5: The Steady-State Approximation
7.6: The Equilibrium Approximation
7.7: The Lindemann Mechanism
7.8: The Michaelis-Menten Mechanism
7.9: Catalysis
7.10: Theories of Reaction Rates
7.E: Chemical Kinetics II (Exercises)
7.S: Chemical Kinetics II (Summary)

This page titled 7: Chemical Kinetics II is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12: Chemical Kinetics II by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.1: Reaction Mechanisms

A **reaction mechanism** is a set of **elementary reactions** steps, that when taken in aggregate define a chemical pathway that connects reactants to products. An elementary reaction is one that proceeds by a single process, such a molecular (or atomic) decomposition or a molecular collision. Typically, elementary reactions only come in **unimolecular**

$$A \rightarrow products$$

and **bimolecular**

$$A + B \rightarrow products$$

form. Occasionally, an elementary step that is **termolecular**

$$A + B + C \rightarrow products$$

(involved the simultaneous collision of three atoms or molecules) but it is generally a pair of bimolecular steps acting in rapid succession, the first forming an activated complex, and the second stabilizing that complex chemically or physically.

$$A + B \rightarrow AB^*$$

$$AB^* + C \rightarrow AB + C^*$$

The wonderful property of elementary reactions is that the **molecularity** defines the order of the rate law for the reaction step.

The Requirements of a Reaction Mechanism

A valid reaction mechanism must satisfy three important criteria:

1. The sum of the steps must yield the overall stoichiometry of the reaction.
2. The mechanism must be consistent with the observed kinetics for the overall reaction.
3. The mechanism must account for the possibility of any observed side products formed in the reaction.

Example 7.1.1:

For the reaction

$$A + B \rightarrow C$$

is the following proposed mechanism valid?

$$A + A \xrightarrow{k_1} A_2$$

$$A_2 + B \xrightarrow{k_1} C + A$$

Solution

Adding both proposed reactions gives

$$\cancel{2}\,A + \cancel{A_2} + B \rightarrow \cancel{A_2} + C + \cancel{A}$$

Canceling those species that appear on both sides of the arrow leaves

$$A + B \rightarrow C$$

which is the reaction, so the mechanism is at least stoichiometrically valid. However, it would still have to be consistent with the observed kinetics for the reaction and account for any side-products that are observed.

This page titled 7.1: Reaction Mechanisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.1: Reaction Mechanisms by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.2: Concentration Profiles for Some Simple Mechanisms

To illustrate how mechanisms may affect the concentration profile for a reaction, we can examine some simple mechanisms

$$A \rightarrow B$$

In this type of reaction, one substance is simply converting into another. An example of this type of reaction might be the isomerization of methylisocyanide to form acetonitrile (methylcyanide) (Redmon, Purvis, & Bartlett, 1978). If the reaction mechanism consists of a single unimolecular step, which is characterized by the rate constant k_1:

$$A \xrightarrow{k_1} B$$

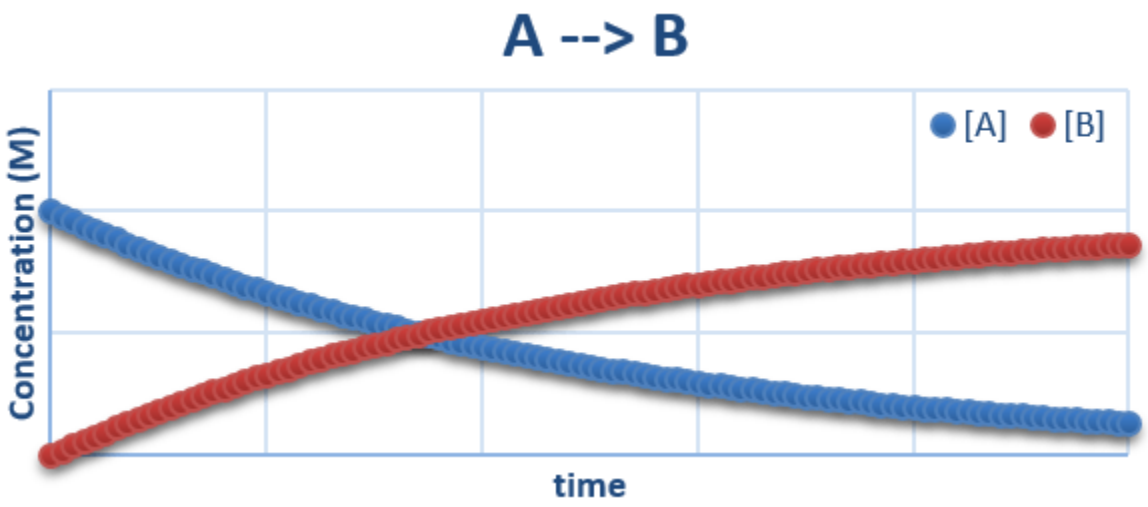
then rate of change of the concentrations of A and B may be written

$$\frac{d[A]}{dt} = -k_1[A]$$

and

$$\frac{d[B]}{dt} = +k_1[A]$$

A plot the concentrations as a function of time would look as follows:



It can be easily seen that the concentration of the reactant (A) decreases as time moves forward, and that of the product (B) increases. This will continue until reactant A is depleted.

$$A \rightleftharpoons B$$

When the system can establish equilibrium, the rate of change of the concentration of A and B will depend on both the forward and reverse reactions. If k_1 is the rate constant that characterizes the forward reaction

$$A \xrightarrow{k_1} B$$

and k_{-1} that which characterizes the reverse

$$B \xrightarrow{k_{-1}} A$$

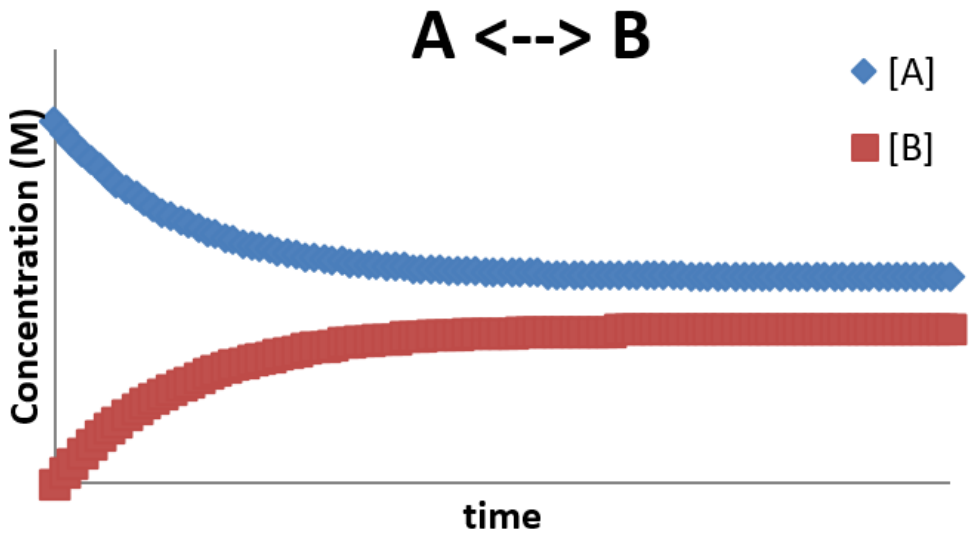
the the rate of change on concentrations of A and B can be expressed

$$\frac{d[A]}{dt} = -k_1[A] + k_{-1}[B]$$

and

$$\frac{d[B]}{dt} = +k_1[A] - k_{-1}[B]$$

The concentration profile for this situation looks as follows:



This profile is characterized by the fact that after a certain amount of time, the system achieves equilibrium and the concentrations stop changing (even though the forward and reverse reactions are still taking place. This is the nature of a **dynamic equilibrium** about which we speak off of the time in chemistry. The final concentrations of [A] and [B] once equilibrium is established will depend on the ratio of k_1 and k_{-}

Since the rate of formation of A (from the reverse step) is equal to the rate of consumption of A (from the forward step, the overall rate of change of the concentration of A is zero once equilibrium has been established. So it should be clear that

$$k_1[A] = k_{-1}[B]$$

or

$$\frac{k_1}{k_{-1}} = \frac{[B]}{[A]}$$

and the ratio o fk_1 to k_{-1} gives the value of the equilibrium constant!

$A + C \rightarrow B + C$

Some reactions require a **catalyst** to mediate the conversion of reactants in to products. The definition of a catalyst is a species that must be added (it is not formed as an intermediate) shows up in the mechanism (usually in a very early step) and this ends up as part of the rate law, but is reformed later on so that it does not appear in the overall stoichiometry. If the reaction

$$A \rightarrow B$$

is aided by a catalyst C, then one possible (single-step) reaction mechanism might be

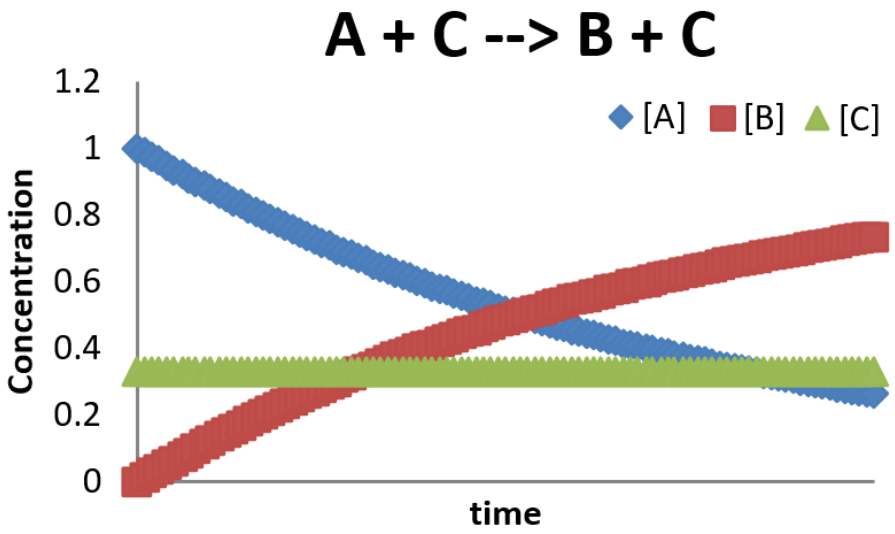
$$A + C \rightarrow B + C$$

In this case, C is acting as a catalyst to the reaction. The rate of change of the concentrations can be found by

$$\frac{[A]}{dt} = -k[A][C]$$

$$\frac{[B]}{dt} = k[A][C]$$

$$\frac{[C]}{dt} = -k[A][C] + k[A][C] = 0$$



This is a very simplified picture of a catalyzed reaction. Generally a catalyzed reaction will require at least two steps:

$$A + C \rightarrow AC$$

$$AC \rightarrow B + C$$

Later, we will see how the steady-state approximation actually predicts the above depicted concentration profile for the two-step mechanism when AC is a short-lived species that can be treated as having a constant and small concentration.

$A \rightarrow B \rightarrow C$

Another important (and very common) mechanistic feature is the formation of an intermediate. This is a species that is formed in at least one of the mechanism step, but does not appear in the overall stoichiometry for the reaction. This is different from a catalyst which must be added to speed the reaction. A simple example of a reaction mechanism involving the formation of a catalyst is

$$A \xrightarrow{k_1} B$$

$$B \xrightarrow{k_2} C$$

In this case, C cannot form until an appreciable concentration of the intermediate B has been created by the first step of the mechanism.

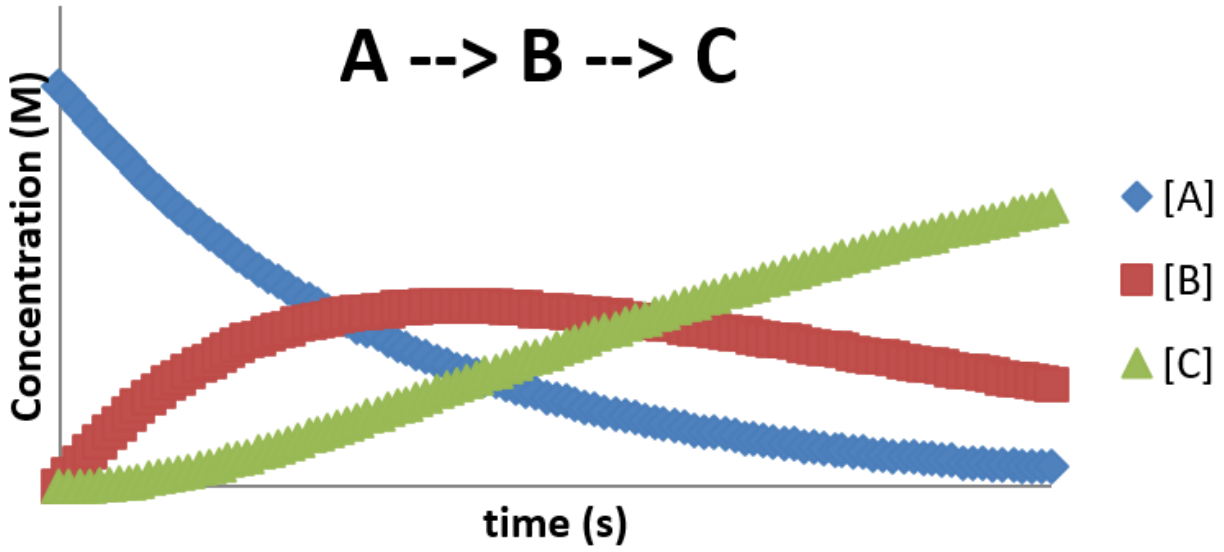
The rate of change of the concentrations of A, B, and C can be expressed

$$\frac{[A]}{dt} = -k_1[A]$$

$$\frac{[B]}{dt} = k_1[A] - k_2[B]$$

$$\frac{[C]}{dt} = k_2[B]$$

The concentration profile is then shown below. Notice the delay in the formation of C.



$A \rightleftharpoons B \rightarrow C$

In many cases, the formation of an intermediate involves a reversible step. This step is sometimes referred to as a **pre-equilibrium** step since it oftentimes will establish a near equilibrium while the reaction progresses. The result of combining a pre-equilibrium with an intermediate produces a profile that shows features of both of the simpler mechanisms. An example of such a mechanism is

$$A \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} B$$

$$B \xrightarrow{k_2} C$$

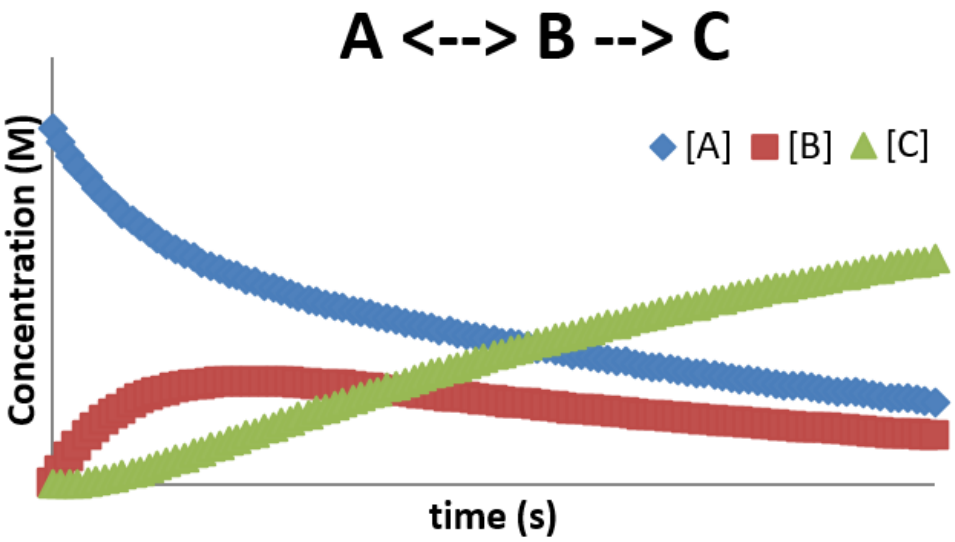
In this case, the rate of change for the concentrations of A, B, and C can be expressed by

$$\frac{[A]}{dt} = -k_1[A] + -k_{-1}[B]$$

$$\frac{[B]}{dt} = k_1[A] - k_{-1}[B] - k_2[B]$$

$$\frac{[C]}{dt} = k_2[B]$$

The concentration profile for this mechanism is shown below. Again, notice the delay in the production of the product C, due to the requirement that the concentration of B be sufficiently high to allow the second step to occur with an appreciable rate.



$A \rightarrow B$ and $A \rightarrow C$

There are many cases where a reactant can follow pathways to different products (or sometimes even the same products!), and those pathways compete with one another. An example is the following simple mechanism:

$$A \xrightarrow{k_{-1}} B$$

$$A \xrightarrow{k_2} C$$

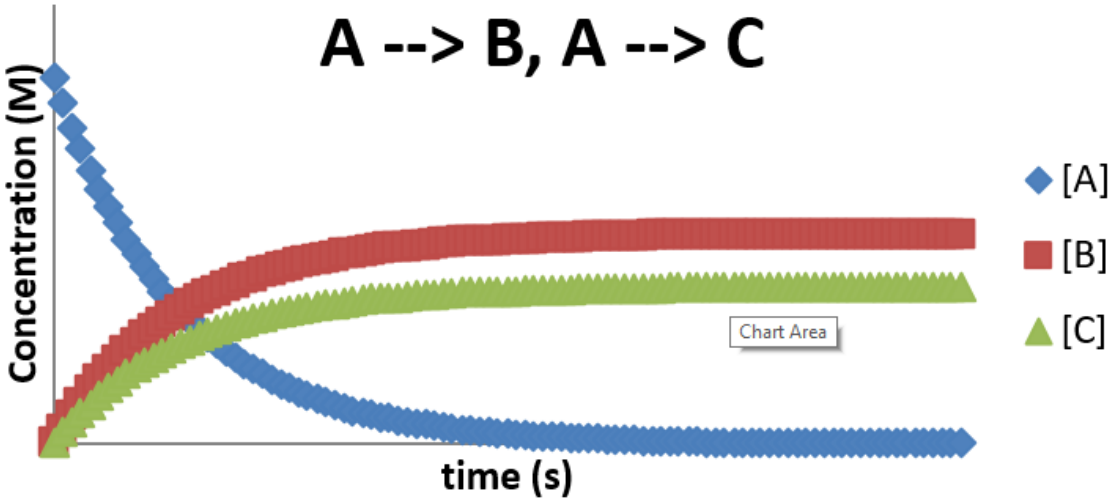
In this case, the rate of change on concentrations can be expressed as

$$\frac{[A]}{dt} = -k_1[A] + -k_2[A]$$

$$\frac{[B]}{dt} = +k_1[A]$$

$$\frac{[C]}{dt} = +k_2[A]$$

Overall, the profile looks like two first order decompositions occurring at the same time, with the final concentration of the product formed with the larger rate constant being favored.



One of the goals of studying chemical kinetics is to understand how to alter reaction condition to favor the production of desirable reaction products. This can be accomplished by a number of means, such as alteration of concentrations, temperature, addition of catalysts, etc. Understanding the basics will (hopefully) lead to a better understanding of how concentration profiles can be altered by changing conditions.

This page titled 7.2: Concentration Profiles for Some Simple Mechanisms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.2: Concentration Profiles for Some Simple Mechanisms by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws

The great value of chemical kinetics is that it can give us insights into the actual reaction pathways (mechanisms) that reactants take to form the products of reactions. Analyzing a reaction mechanism to determine the type of rate law that is consistent (or not consistent) with the specific mechanism can give us significant insight. For example, the reaction

$$A + B \rightarrow C$$

might be proposed to follow one of two mechanistic pathways:

$$\underbrace{A + A \xrightarrow{k_1} A_2}_{\text{step 1}}$$

$$\underbrace{A_2 + B \xrightarrow{k_2} C}_{\text{step 2}}$$

or

$$\underbrace{A \xrightarrow{k_1} A^*}_{\text{step 1}}$$

$$\underbrace{A^* + B \xrightarrow{k_2} C}_{\text{step 2}}$$

The first rate law will predict that the reaction should be second order in A, whereas the second mechanism predicts that it should be first order in A (in the limit that the steady state approximation, discussed in the following sections, can be applied to A_2 and A^*). Based on the observed rate law being first or second order in A, one can rule out one of the rate laws. Unfortunately, this kind of analysis cannot confirm a specific mechanism. Other evidence is needed to draw such conclusions, such as the spectroscopic observation of a particular reaction intermediate that can only be formed by a specific mechanism.

In order analyze mechanisms and predict rate laws, we need to build a toolbox of methods and techniques that are useful in certain limits. The next few sections will discuss this kind of analysis, specifically focusing on

- the **Rate Determining Step** approximation,
- the **Steady State** approximation, and
- the **Equilibrium** approximation.

Each type of approximation is important in certain limits, and they are oftentimes used in conjunction with one another to predict the final forms of rate laws.

This page titled 7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.3: The Connection between Reaction Mechanisms and Reaction Rate Laws by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.4: The Rate Determining Step Approximation

The **rate determining step** approximation is one of the simplest approximations one can make to analyze a proposed mechanism to deduce the rate law it predicts. Simply stated, the rate determining step approximation says that a mechanism can proceed no faster than its slowest step. So, for example, if the reaction

$$A + B \rightarrow C$$

is proposed to follow the mechanism

$$\underbrace{A + A \xrightarrow{k_1} A_2}_{\text{slow}}$$

$$\underbrace{A_2 \xrightarrow{k_2} C + A}_{\text{fast}}$$

the rate determining step approximation suggests that the rate (expressed in terms of the appearance of product C) should be determined by the slow initial step, and so the rate law will be

$$\frac{[C]}{dt} = k_1[A]^2$$

matching the order of the rate law to the molecularity of the slow step. Conversely, if the reaction mechanism is proposed as

$$\underbrace{A \xrightarrow{k_1} A^*}_{\text{slow}}$$

$$\underbrace{A^* + B \xrightarrow{k_2} C}_{\text{fast}}$$

the rate determining step approximation suggests that the rate of the reaction should be

$$\frac{[C]}{dt} = k_1[A]$$

again, with the order of the rate law matching the molecularity of the rate determining step.

This page titled 7.4: The Rate Determining Step Approximation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.4: The Rate Determining Step Approximation by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.5: The Steady-State Approximation

One of the most commonly used and most attractive approximations is the **steady state approximation**. This approximation can be applied to the rate of change of concentration of a highly reactive (short lived) intermediate that holds a constant value over a long period of time. The advantage here is that for such an intermediate (I),

$$\frac{d[I]}{dt} = 0$$

So long as one can write an expression for the rate of change of the concentration of the intermediate I, the steady state approximation allows one to solve for its constant concentration. For example, if the reaction

$$A + B \rightarrow C \tag{7.5.1}$$

is proposed to follow the mechanism

$$A + A \xrightarrow{k_1} A_2 \tag{7.5.2}$$

$$A_2 + B \xrightarrow{k_2} C + A \tag{7.5.3}$$

The time-rate of change of the concentration of the intermediate A_2 can be written as

$$\frac{d[A_2]}{dt} = k_1[A]^2 - k_2[A_2][B]$$

In the limit that the steady state approximation can be applied to A_2

$$\frac{d[A_2]}{dt} = k_1[A]^2 - k_2[A_2][B] \approx 0$$

or

$$[A_2] \approx \frac{k_1[A]^2}{k_2[B]}$$

So if the rate of the overall reaction is expressed as the rate of formation of the product C,

$$\frac{d[C]}{dt} = k_2[A_2][B]$$

the above expression for $[A_2]$ can be substituted

$$\frac{d[C]}{dt} = k_2\left(\frac{k_1[A]^2}{k_2[B]}\right)[B]$$

of

$$\frac{d[C]}{dt} = k_1[A]^2$$

and the reaction is predicted to be second order in $[A]$.

Alternatively, if the mechanism for Equation 7.5.1 is proposed to be

$$A \xrightarrow{k_1} A^* \tag{7.5.4}$$

$$A^* + B \xrightarrow{k_2} C \tag{7.5.5}$$

then the rate of change of the concentration of A^* is

$$\frac{[A^*]}{dt} = k_1[A] - k_2[A^*][B]$$

And if the steady state approximation holds, then

$$[A^*] \approx \frac{k_1[A]}{k_2[B]}$$

So the rate of production of C is

$$\frac{d[C]}{dt} = k_2[A^*][B] \tag{7.5.6}$$

$$= \cancel{k_2}\left(\frac{k_1[A]}{\cancel{k_2}\,\cancel{[B]}}\right)\cancel{[B]} \tag{7.5.7}$$

or

$$\frac{d[C]}{dt} = k_1[A]$$

and the rate law is predicted to be first order in A. In this manner, the plausibility of either of the two reaction mechanisms is easily deduced by comparing the predicted rate law to that which is observed. If the prediction cannot be reconciled with observation, then the scientific method eliminates that mechanism from consideration.

This page titled 7.5: The Steady-State Approximation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.5: The Steady-State Approximation by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.6: The Equilibrium Approximation

In many cases, the formation of a reactive intermediate (or even a longer lived intermediate) involves a reversible step. This is the case if the intermediate can decompose to reform reactants with a significant probability as well as moving on to form products. In many cases, this will lead to a pre-equilibrium condition in which the **equilibrium approximation** can be applied. An example of a reaction mechanism of this sort is

$$A + B \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} AB$$

$$AB \xrightarrow{k_2} C$$

Given this mechanism, the application of the steady state approximation is cumbersome. However, if the initial step is assumed to achieve equilibrium, an expression can be found for $[AB]$. In order to derive this expression, one assumes that the rate of the forward reaction is equal to the rate of the reverse reaction for the initial step in the mechanism.

$$k_1[A][B] = k_{-1}[AB]$$

or

$$\frac{k_1[A][B]}{k_{-1}} = [AB]$$

This expression can be substituted into an expression for the rate of formation of the product C:

$$\frac{d[C]}{dt} = k_2[AB]$$

or

$$\frac{d[C]}{dt} = \frac{k_2 k_1}{k_{-1}}[A][B]$$

Which predicts a reaction rate law that is first order in A, first order in B, and second order overall.

Example 7.6.1:

Given the following mechanism, apply the equilibrium approximation to the first step to predict the rate law suggested by the mechanism.

$$A + A \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} A_2$$

$$A_2 + B \xrightarrow{k_2} C + A$$

Solution

If the equilibrium approximation is valid for the first step,

$$k_1[A]^2 = k_{-1}[A_2]$$

or

$$\frac{k_1[A]^2}{k_{-1}} \approx [A_2]$$

Plugging this into the rate equation for the second step

$$\frac{d[C]}{dt} = k_2[A_2][B]$$

yields

$$\frac{d[C]}{dt} = \frac{k_2 k_1}{k_{-1}}[A]^2[B]$$

Thus, the rate law has the form

$$\text{rate} = k'[A]^2[B]$$

which is second order in A, first order in B and third order over all, and in which the effective rate constant (k' is

$$k' = \frac{k_2 k_1}{k_{-1}}.$$

Sometimes, the equilibrium approximation can suggest rate laws that have negative orders with respect to certain species. For example, consider the following reaction

$$A + 2B \rightarrow 2C$$

A proposed mechanism for which might be

$$A + B \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} I + C$$

$$I + B \xrightarrow{k_2} C$$

in which I is an intermediate. Applying the equilibrium approximation to the first step yields

$$k_1[A][B] = k_{-1}[I][C]$$

or

$$\frac{k_1[A][B]}{k_{-1}[C]} \approx [I]$$

Substituting this into an expression for the rate of formation of C, one sees

$$\frac{d[C]}{dt} = k_2[I][B]$$

or

$$\frac{d[C]}{dt} = \frac{k_1[A][B]}{k_{-1}[C]}[B] = \frac{k_2 k_1[A][B]}{k_{-1}[C]}$$

The rate law is then of the form

$$\text{rate} = k\frac{[A][B]^2}{[C]}$$

which is first order in A, second order in B, negative one order in C, and second order overall. Also,

$$k' = \frac{k_2 k_1}{k_{-1}}.$$

In this case, the negative order in C means that a buildup of compound C will cause the reaction to slow. These sort of rate laws are not uncommon for reactions with a reversible initial step that forms some of the eventual reaction product.

This page titled 7.6: The Equilibrium Approximation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.6: The Equilibrium Approximation by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.7: The Lindemann Mechanism

Introduction

The **Lindemann mechanism** (Lindemann, Arrhenius, Langmuir, Dhar, Perrin, & Lewis, 1922) is a useful one to demonstrate some of the techniques we use for relating chemical mechanisms to rate laws. In this mechanism, a reactant is collisionally activated to a highly energetic form that can then go on to react to form products.

$$A + A \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} A^* + A$$

$$A^* \xrightarrow{k_2} P$$

Steady-state analysis

If the steady state approximation is applied to the intermediate A^*

$$\frac{d[A^*]}{dt} = k_1[A]^2 - k_{-1}[A^*][A] - k_2[A^*] \approx 0$$

an expression can be derived for $[A^*]$.

$$[A^*] = \frac{k_1[A]^2}{k_{-1}[A] + k_2}$$

Substituting this into an expression for the rate of the production of the product P

$$\frac{d[P]}{dt} = k_2[A^*]$$

yields

$$\frac{d[P]}{dt} = \frac{k_2 k_1[A]^2}{k_{-1}[A] + k_2}$$

In the limit that $k_{-1}[A] \ll k_2$, the rate law becomes first order in $[A]$ since $k_{-1}[A] + k_2 \approx k_{-1}[A]$.

$$\frac{d[P]}{dt} = \frac{k_2 k_1}{k_{-1}}[A]$$

This will happen if the second step is very slow (and is the rate determining step), such that the reverse of the first step “wins” in the competition for [A*]. However, in the other limit, that $k_2 \gg k_{-1}[A]$, the reaction becomes second order in $[A]$ since $k_{-1}[A] + k_2 \approx k_2$.

$$\frac{d[P]}{dt} = k_1[A]^2$$

which is consistent with the forward reaction of the first step being the rate determining step, since A^* is removed from the reaction (through the formation of products) very quickly as soon as it is formed.

Third-body Collisions

Sometimes, the **third-body collision** is provided by an inert species M, perhaps by filling the reaction chamber with a heavy non-reactive species, such as Ar. In this case, the mechanism becomes

$$A + M \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} A^* + M$$

$$A^* \xrightarrow{k_2} P$$

In the limit that $[A^*]$ can be treated using the steady state approximation, where [M] = concentration of all colliding species (e.g. [A] + [Ar], assuming [A*] is negligible) we get:

$$[A^*] = \frac{k_1[A][M]}{k_{-1}[M] + k_2}$$

and the rate of production of the product becomes

$$\frac{d[P]}{dt} = \frac{k_2 k_1[A][M]}{k_{-1}[M] + k_2} = \frac{k_2 k_1[M]}{k_{-1}[M] + k_2}[A]$$

And if the concentration of the third body collider is constant, it is convenient to define an **effective rate constant**, k_{uni}.

$$k_{uni} = \frac{k_2 k_1[M]}{k_{-1}[M] + k_2}$$

The utility is that important information about the individual step rate constants can be extracted by plotting $1/k_{uni}$ as a function of $1/[M]$.

$$\frac{1}{k_{uni}} = \frac{k_{-1}}{k_2 k_1} + k_2 \left(\frac{1}{[M]}\right)$$

The plot should yield a straight line, the slope of which gives the value of k_2, and the intercept gives $(k_{-1}/k_2 k_1)$.

This page titled 7.7: The Lindemann Mechanism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.7: The Lindemann Mechanism by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.8: The Michaelis-Menten Mechanism

The **Michaelis-Menten** mechanism (Michaelis & Menten, 1913) is one which many enzyme mitigated reactions follow. The basic mechanism involves an enzyme (E, a biological catalyst) and a substrate (S) which must connect to form an enzyme-substrate complex (ES) in order for the substrate to be degraded (or augmented) to form a product (P). The overall reaction is

$$S \rightarrow P$$

And the simple two-step mechanism is given by

$$E + S \underset{\blacksquare}{\overset{k_1}{\rightleftharpoons}} ES \tag{7.8.1}$$

$$ES \xrightarrow{k_2} P \tag{7.8.2}$$

Notice that the enzyme is necessary for the reaction to proceed, but is not part of the overall stoichiometry (as is the case for any catalyst!).

Equilibrium Approximation Derivation

Applying the equilibrium approximation to the first step

$$k_1[E][S] \approx k_{-1}[ES] \tag{7.8.3}$$

And using a mass conservation relationship on the enzyme (noting that the enzyme must be either in its bare form (E) or complexed with a substrate (ES)):

$$[E]_o = [E] + [ES]$$

or

$$[E] = [E]_o - [ES]$$

Substituting this into the equilibrium expression (Equation 7.8.3) yields

$$k_1([E]_o - [ES])[S] = k_{-1}[ES]$$

Solving this expression for $[ES]$ stepwise reveals

$$k_1[E]_o[S] - k_1[ES][S] = k_{-1}[ES]$$

$$k_1[E]_o[S] = k_{-1}[ES] + k_1[ES][S]$$

$$= (k_{-1} + k_1)[ES]$$

$$\frac{k_1[E]_o[S]}{k_1[S] + k_{-1}} = [ES]$$

Substituting this into the expression for the rate of production of the product P

$$\frac{d[P]}{dt} = k_2[ES] \tag{7.8.4}$$

yields

$$\frac{d[P]}{dt} = \frac{k_2 k_1[E]_o[S]}{k_1[S] + k_{-1}}$$

Multiplying the top and bottom of the expression on the right hand side by 1/k_1 gives the result

$$\frac{d[P]}{dt} = \frac{k_2[E]_o[S]}{[S] + \frac{k_1}{k_{-1}}}$$

The ratio of k_{-1}/k_1 is the equilibrium constant that describes the dissociation of the enzyme-substrate complex, K_d in Equation 7.8.1. Noting that $k_2[E]_0$ gives the maximum rate (V_{max}), and that $\frac{d[P]}{dt}$ is the observed reaction rate, the rate law takes the form

$$\text{rate} = \frac{V_{max}[S]}{K_d + [S]}$$

> This is because the maximum reaction rate is achieved when [ES] is low. As [ES] increases, the likelihood of the complex decomposing to reform reactants is higher, slowing the conversion. [ES] will be low if the concentration of the enzyme is much larger than that of the substrate, so there is never a shortage of enzyme available to form the complex with the substrate. However, if the substrate concentration is higher, the lack of available enzyme active sites will slow the reaction and cause it to become 0th order.

In the limit that the substrate concentration is large compared to K_d (i.e., $K_d + [S] \approx [S]$), the reaction ends up zeroth order in substrate.

$$\text{rate} = \frac{V_{max}[S]}{K_d + [S]} \approx \frac{V_{max}\cancel{[S]}}{\cancel{[S]}} = V_{max}$$

Hence, adding more substrate to the system under this limiting condition will have no effect on the observed rate. This is characteristic of a bottleneck in the mechanism, which would happen if there is a shortage of enzyme sites to which the substrate can attach.

In the other extreme, in which K_d is very large compared to the substrate concentration (i.e., $K_d + [S] \approx K_d$), the reaction become first order in substrate.

$$\text{rate} = \frac{V_{max}[S]}{K_d + [S]} \approx \frac{V_{max}[S]}{K_d} = \frac{V_{max}}{K_d}[S]$$

Steady-State Approximation Derivation

In an alternate derivation (Briggs & Haldane, 1925) using the steady state approximation applied to the enzyme-substrate complex

$$\frac{[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES] \approx 0$$

Solving for $[ES]$ gives the result

$$ES] = \frac{k_1[E][S]}{k_{-1} + k_2}$$

or

$$[ES] = \frac{[E][S]}{K_m}$$

where

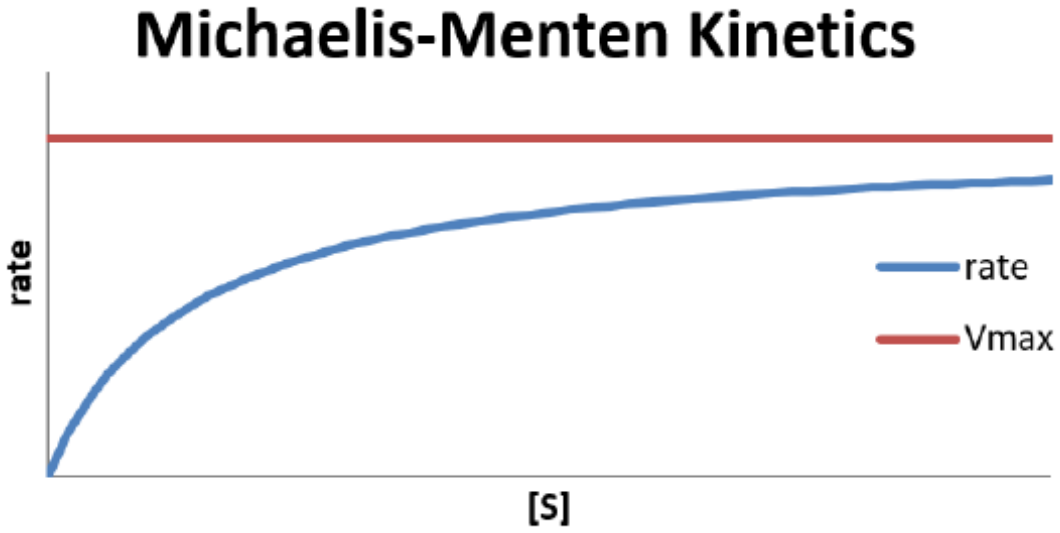
$$K_m = \frac{k_{-1} + K_2}{k_1}$$

K_M is the **Michaelis constant**, which is affected by a number of factors, including pH, temperature, and the nature of the substrate itself. Proceeding as before, though the conservation of mass relationship and substitution into the expression for rate

(Equation 7.8.4) results in

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]}$$

The advantage to this approach is that it accounts for the loss of ES complex due to the production of products as well as the decomposition to reform the reactants E and S. As before, in the limit that $[S] \gg K_M$, the reaction reaches its maximum rate (V_{max}) and becomes independent of any concentrations. However in the limit that $[S] \ll K_M$, the reaction becomes 1st order in $[S]$.



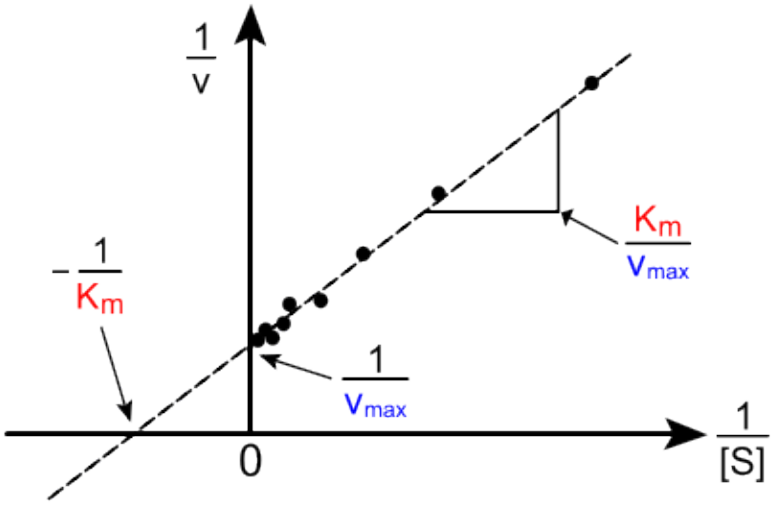
The Michalis constant and V_{max} parameters can be extracted in a number of ways. In the Lineweaver-Burk (Lineweaver & Burk, 1934) method, the reciprocal of the rate law is used to create a linear relationship.

$$\frac{1}{\text{rate}} = \frac{K_m + [S]}{V_{max}[S]}$$

or

$$\frac{1}{\text{rate}} = \frac{K_m}{V_{max}}\frac{1}{[S]} + \frac{1}{V_{max}}$$

So a plot of $1/rate$ as a function of $1/[S]$ results in a straight line, the slope of which is equal to K_M/V_{max} and the intercept is $1/V_{max}$. This is called a Lineweaver–Burk plot.



An example of a Lineweaver-Burk plot. (CC BY-SA 3.0; Diberri).

This page titled 7.8: The Michaelis-Menten Mechanism is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.8: The Michaelis-Menten Mechanism by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.9: Catalysis

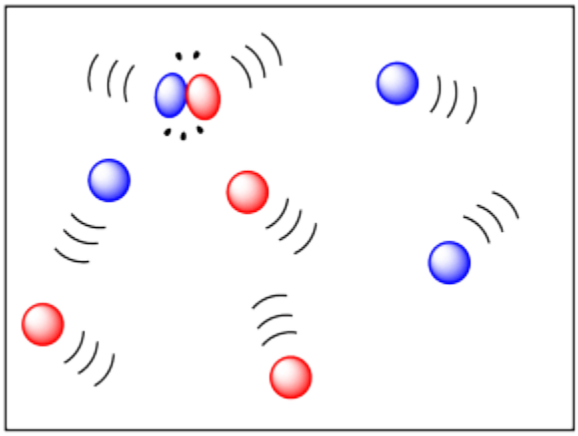
This page titled 7.9: Catalysis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

7.10: Theories of Reaction Rates

The macroscopic discussion of kinetics discussed in previous sections can be now expanded into a more microscopic picture in terms of molecular level properties (e..g, mass and velocities) involving two important theories: (1) collision theory and (2) transition-state theory.

Collision Theory

If two molecules need to collide in order for a reaction to take place, then factors that influence the ease of collisions will be important. The more energy there is available to the molecules, the faster they will move around, and the more likely they are to bump into each other. Higher temperatures ought to lead to more collisions and a greater frequency of reactions between molecules. In the drawing below, the cold, sluggish molecules on the left are not likely to collide, but the energetic molecules on the right are due to collide at any time.



The rate at which molecules collide which is the frequency of collisions is called the collision frequency, Z, which has units of collisions per unit of time. Given a container of molecules A and B, the collision frequency between A and B is defined by:

$$Z = N_A N_B \sigma_{AB} \sqrt{\frac{8k_B T}{\pi \mu_{AB}}} \tag{7.10.1}$$

where:

- N_A and N_B are the numbers of molecules A and B, and is directly related to the concentrations of A and B.
- The mean speed of molecules obtained from the Maxwell-Boltzmann distribution for thermalized gases

$$\sqrt{\frac{8k_B T}{\pi \mu_{AB}}} \tag{7.10.2}$$

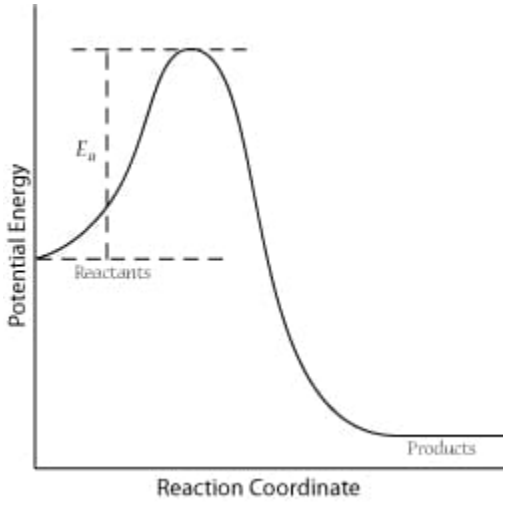
- σ_{AB} is the averaged sum of the collision cross-sections of molecules A and B. The collision cross section represents the collision region presented by one molecule to another.
- μ is the reduced mass and is given by

$$\mu = \frac{m_\mathrm{A} m_\mathrm{B}}{m_\mathrm{A} + m_\mathrm{B}} \tag{7.10.3}$$

The concepts of collision frequency can be applied in the laboratory: (1) The temperature of the environment affects the average speed of molecules. Thus, reactions are heated to increase the reaction rate. (2) The initial concentration of reactants is directly proportional to the collision frequency; increasing the initial concentration will speed up the reaction.

For a successful collision to occur, the reactant molecules must collide with enough kinetic energy to break original bonds and

form new bonds to become the product molecules. This energy is called the activation energy for the reaction; it is also often referred to as the energy barrier.



The fraction of collisions with enough energy to overcome the activation barrier is given by:

$$f = e^{\frac{-E_a}{RT}} \tag{7.10.4}$$

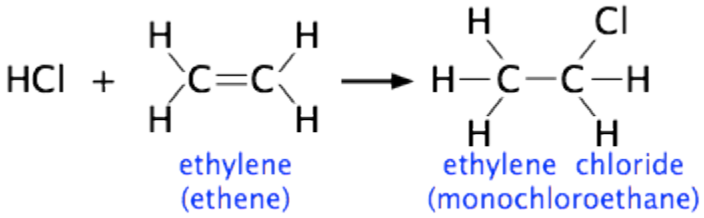
where:

- f is the fraction of collisions with enough energy to react
- E_a is the activation energy

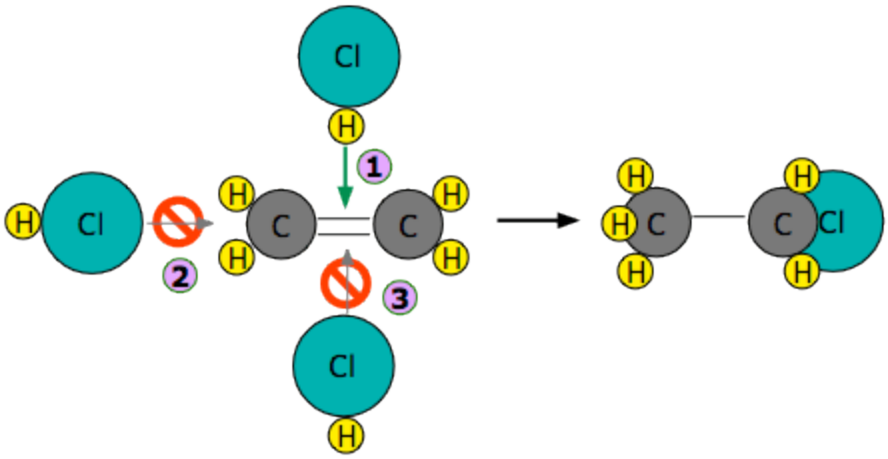
The fraction of successful collisions is directly proportional to the temperature and inversely proportional to the activation energy.

The fraction of successful collisions is directly proportional to the temperature and inversely proportional to the activation energy.

The more complicated the structures of the reactants, the more likely that the value of the rate constant will depend on the trajectories at which the reactants approach each other. This kind of *electrophilic addition reaction* is well-known to all students of organic chemistry. Consider the addition of a hydrogen halide such as HCl to the double bond of an alkene, converting it to a chloroalkane.



Experiments have shown that the reaction only takes place when the HCl molecule approaches the alkene with its hydrogen-end, and in a direction that is approximately perpendicular to the double bond, as shown at (1) below.



The reason for this becomes apparent when we recall that HCl is highly polar owing to the high electronegativity of chlorine, so that the hydrogen end of the molecule is slightly positive. The steric factor, ρ is then introduced to represent is the probability of the reactant molecules colliding with the right orientation and positioning to achieve a product with the desirable geometry and stereospecificity. Values of ρ are generally very difficult to assess and range from 0 to 1, but are sometime estimated by comparing the observed rate constant with the one in which the preexponential constant A is assumed to be the same as Z.

The lesson you should take from this example is that once you start combining a variety of chemical principles, you gradually develop what might be called "chemical intuition" which you can apply to a wide variety of problems. This is far more important than memorizing specific examples.

All Three Factors Combined

The rate constant of the gas-phase reaction is proportional to the product of the *collision frequency* and the *fraction of successful reactions*. As stated above, sufficient kinetic energy is required for a successful reaction; however, they must also collide properly. Compare the following equation to the Arrhenius equation:

$$k = Z\rho e^{\frac{-E_a}{RT}} \tag{7.10.5}$$

where

- k is the rate constant for the reaction
- ρ is the steric factor.
- Zρ is the pre-exponential factor, A, of the Arrhenius equation. It is the frequency of total collisions that collide with the right orientation. In practice, it is the pre-exponential factor that is directly determined by experiment and then used to calculate the steric factor.
- E_a is activation energy
- T is absolute temperature
- R is gas constant.

Although the collision theory deals with gas-phase reactions, its concepts can also be applied to reactions that take place in solvents; however, the properties of the solvents (for example: solvent cage) will affect the rate of reactions. Ultimately, collision theory illustrates how reactions occur; it can be used to approximate the rate constants of reactions, and its concepts can be directly applied in the laboratory. Read this for a more detailed discussion of Collision Theory.

Transition-State Theory

Transition state theory (TST) provides a more accurate alternative to the previously used Arrhenius equation and the collision theory. The transition state theory attempts to provide a greater understanding of activation energy, E_a, and the thermodynamic properties involving the transition state. Collision theory of reaction rate, although intuitive, lacks an accurate method to predict the probability factor for the reaction. The theory assumes that reactants are hard spheres rather than molecules with specific structures. In 1935, Henry Eyring helped develop a new theory called the transition state theory to provide a more accurate

alternative to the previously used Arrhenius equation and the collision theory. The Eyring equation involves the statistical frequency factory, v, which is fundamental to the theory.

According to TST, between the state where molecules are reactants and the state where molecules are products, there is a state known as the transition state. In the transition state, the reactants are combined in a species called the activated complex. The theory suggests that there are three major factors that determine whether a reaction will occur:

1. The concentration of the activated complex
2. The rate at which the activated complex breaks apart
3. The way in which the activated complex breaks apart: whether it breaks apart to reform the reactants or whether it breaks apart to form a new complex, the products.

Collision theory proposes that not all reactants that combine undergo a reaction. However, assuming the stipulations of the collision theory are met and a successful collision occurs between the molecules, transition state theory allows one of two outcomes: a return to the reactants, or a rearranging of bonds to form the products.

Consider a bimolecular reaction:

$$A + B \rightarrow C \tag{7.10.6}$$

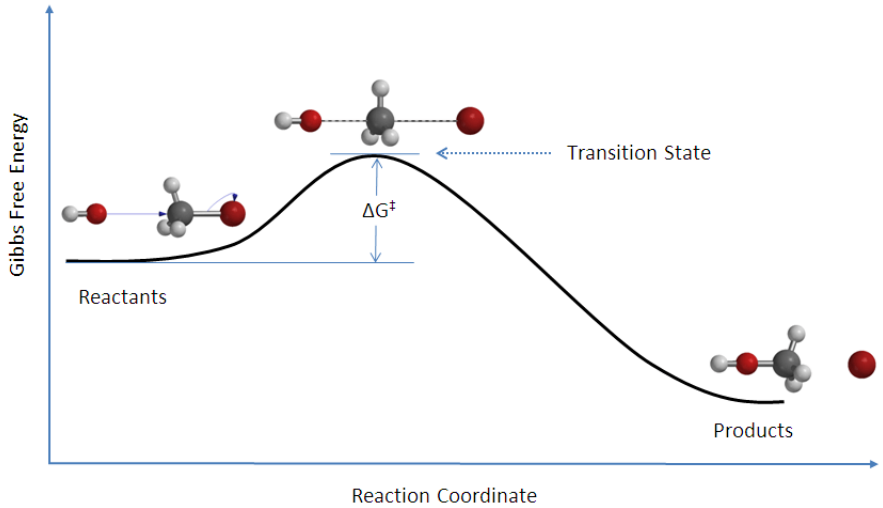
$$K = \frac{[C]}{[A][B]} \tag{7.10.7}$$

where K is the equilibrium constant. In the transition state model, the activated complex AB is formed:

$$A + B \rightleftharpoons AB^{\ddagger} \rightarrow C \tag{7.10.8}$$

$$K^{\ddagger} = \frac{[AB]^{\ddagger}}{[A][B]} \tag{7.10.9}$$

There is an energy barrier, called activation energy, in the reaction pathway. A certain amount of energy is required for the reaction to occur. The transition state, $AB^{\ddagger}$, is formed at maximum energy. This high-energy complex represents an unstable intermediate. Once the energy barrier is overcome, the reaction is able to proceed and product formation occurs.



Reaction: $HO^- + CH_3Br \rightarrow [HO{-}{-}{-}CH_3{-}{-}{-}Br]^{\ddagger} \rightarrow CH_3OH + Br^-$

Figure 7.10.1: Reaction coordinate diagram for the bimolecular nucleophilic substitution (S_N2) reaction between bromomethane and the hydroxide anion. form Wikipedia.

The rate of a reaction is equal to the number of activated complexes decomposing to form products. Hence, it is the concentration of the high-energy complex multiplied by the frequency of it surmounting the barrier.

$$rate = v[AB^{\ddagger}] \quad (7.10.10)$$
$$= v[A][B]K^{\ddagger} \quad (7.10.11)$$

The rate can be rewritten:

$$rate = k[A][B] \quad (7.10.12)$$

Combining Equations 7.10.12 and 7.10.11 gives:

$$k[A][B] = v[A][B]K^{\ddagger} \quad (7.10.13)$$
$$k = vK^{\ddagger} \quad (7.10.14)$$

where

- v is the frequency of vibration,
- k is the rate constant and
- $K^{\ddagger}$ is the thermodynamic equilibrium constant.

Statistical mechanics (not shown) provides that the frequency, v, is equivalent to the thermal energy, k_BT, divided by Planck's constant, h.

$$v = \frac{k_BT}{h} \quad (7.10.15)$$

where

- k_B is the Boltzmann's constant (1.381 x 10^{-23} J/K),
- T is the absolute temperature in Kelvin (K) and
- h is Planck's constant (6.626 x 10^{-34} Js).

Substituting Equation 7.10.15 into Equation 7.10.14 :

$$k = \frac{k_BT}{h}K^{\ddagger} \quad (7.10.16)$$

Equation ref is often tagged with another term (M^{1-m}) that makes the units equal with M is the molarity and m is the molecularly of the reaction.

$$k = \frac{k_BT}{h}K^{\ddagger}(M^{1-m}) \quad (7.10.17)$$

It is important to note here that the equilibrium constant $K^{\ddagger}$ can be calculated by absolute, fundamental properties such as bond length, atomic mass, and vibration frequency. This gives the transition rate theory the alternative name absolute rate theory, because the rate constant, k, can be calculated from fundamental properties.

Thermodynamics of Transition State Theory

To reveal the thermodynamics of the theory, $K^{\ddagger}$ must be expressed in terms of $\Delta G^{\ddagger}$. $\Delta G^{\ddagger}$ is simply,

$$\Delta G^{o\ddagger} = G^o(transitionstate) - G^o(reactants) \quad (7.10.18)$$

By definition, at equilibrium, $\Delta G^{\ddagger}$ can be expressed as:

$$\Delta G^{\ddagger} = -RT\ln K^{\ddagger} \quad (7.10.19)$$

Rearrangement gives:

$$[K]^{\ddagger} = e^{-\frac{\Delta G^{\ddagger}}{RT}} \quad (7.10.20)$$

From Equation 7.10.17

$$k = ve^{-\frac{\Delta G^{\ddagger}}{RT}}(M^{1-m}) \tag{7.10.21}$$

It is also possible to obtain terms for the change in enthalpy and entropy for the transition state. Because

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \tag{7.10.22}$$

it follows that the derived equation becomes,

$$k = \frac{k_B T}{h} e^{\Delta S^{\ddagger}/R} e^{-\Delta H^{\ddagger}/RT} M^{1-m} \tag{7.10.23}$$

Equation 7.10.23 is known as the Eyring Equation and was developed by Henry Eyring in 1935, is based on transition state theory and is used to describe the relationship between reaction rate and temperature. It is similar to the Arrhenius Equation, which also describes the temperature dependence of reaction rates.

The linear form of the Eyring Equation is given below:

$$\ln \frac{k}{T} = \frac{-\Delta H^{\dagger}}{R}\frac{1}{T} + \ln \frac{k_B}{h} + \frac{\Delta S^{\ddagger}}{R} \tag{7.10.24}$$

The values for $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ can be determined from kinetic data obtained from a $\ln \frac{k}{T}$ vs. $\frac{1}{T}$ plot. The Equation is a straight line with negative slope, $\frac{-\Delta H^{\ddagger}}{R}$, and a y-intercept, $\frac{\Delta S^{\ddagger}}{R} + \ln \frac{k_B}{h}$.

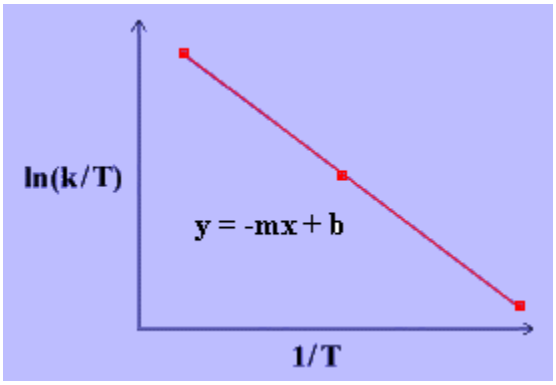


Figure 7.10.2: Linearized TST theory

Conclusion

In this article, the complete thermodynamic formulation of the transition state theory was derived. This equation is more reliable than either the Arrhenius equation and the equation for the Collision Theory. However, it has its limitations, especially when considering the concepts of quantum mechanics. Quantum mechanics implies that tunneling can occur, such that particles can bypass the energy barrier created by the transition state. This can especially occur with low activation energies, because the probability of tunneling increases when the barrier height is lowered.

In addition, transition state theory assumes that an equilibrium exists between the reactants and the transition state phase. However, in solution non-equilibrium situations can arise, upsetting the theory. Several more complex theories have been presented to correct for these and other discrepancies. This theory still remains largely useful in calculating the thermodynamic properties of the transition state from the overall reaction rate. This presents immense usefulness in medicinal chemistry, in which the study of transition state analogs is widely implemented.

References

1. Chang, Raymond. Physical Chemistry for the Biosciences. University Science Books, California 2005
2. Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J., Current Status of Transition-State Theory. The Journal of physical chemistry 1996, 100, (31), 12771-12800

3. Laidler, K.; King, C., Development of transition-state theory. The Journal of physical chemistry 1983, 87, (15), 2657
4. Lienhard, Gustav, Enzyme Catalysis and Transition -State Theory: Transition State Analogs. Cold Spring Harb Symp Quant Biol 1972. 36:45-51

Contributors and Attributions

- Phong Dao, Bilal Latif, Lu Zhao (UC Davis)

This page titled 7.10: Theories of Reaction Rates is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 9.7: Theories of Reaction Rates is licensed CC BY-NC-SA 4.0.

7.E: Chemical Kinetics II (Exercises)

In preparation

This page titled 7.E: Chemical Kinetics II (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.E: Chemical Kinetics II (Exercises) by Patrick Fleming is licensed CC BY-NC-SA 4.0.

7.S: Chemical Kinetics II (Summary)

Vocabulary and Concepts

- autocatalysis
- bimolecular
- catalyst
- chain reaction
- dynamic equilibrium
- effect rate constant
- elementary reaction
- equilibrium approximation
- equilibrium constant
- initiation step
- intermediate
- Lindemann mechanism
- Lotka-Voltera mechanism
- Michaelis constant
- Michaelis-Menten
- molecularity
- Montreal Protocol
- oscillating reaction
- predator-prey
- pre-equilibrium
- propagation step
- radical
- rate determining step
- reaction mechanism
- steady state approximation
- termination step
- termolecular
- third-body collision
- unimolecular

References

1. Briggs, G., & Haldane, J. (1925). A note on the kinematics of enzyme action. *Biochem J, 19*(2), 338-339.
2. Fahey, D. W. (2006). *Twenty Questions and Answers About the Ozone Layer: 2006 Update*. Earth System Research Laboratory; Chemical Sciences Division. Washington DC: U.S. Department of Commerce; National Oceanic and Atmospheric Administration.
3. K, J. (2015, May 14). *Ozone layer healing with 'hole' closing up says Nasa*. Retrieved May 3, 2016, from International Business Times: http://www.ibtimes.co.uk/ozone-layer...s-nasa-1501227
4. Lindemann, F. A., Arrhenius, S., Langmuir, I., Dhar, N. R., Perrin, J., & Lewis, W. C. (1922). Discussion on the "radiation theory of chemical reaction". *Trans. Faraday Soc., 17*, 598-606.
5. Lineweaver, H., & Burk, D. (1934). The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society, 56*(3), 658-666.
6. Michaelis, L., & Menten, M. (1913). Die Kinetik der Invertinwirkung. *Biochem Z, 49*, 333-369.
7. Redmon, L. T., Purvis, G. D., & Bartlett, R. J. (1978). The unimolecular isomerization of methyl isocyanide to methyl cyanide (acetonitrile). *Hournal of Chemical Physics, 69*, 5386.
8. Secretariat, O. (2015). *THE MONTREAL PROTOCOL ON SUBSTANCES THAT DEPLETE THE OZONE LAYER* . Retrieved May 3, 2016, from The Ozone Secretariat: ozone.unep.org/en/treaties-an...te-ozone-layer

This page titled 7.S: Chemical Kinetics II (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Patrick Fleming.

- 12.S: Chemical Kinetics II (Summary) by Patrick Fleming is licensed CC BY-NC-SA 4.0.

8: OPTIONAL- SPECIAL TOPICS

8.1: Chemistry of Cooking (Rodriguez-Velazquez)

8.1.1: Thickening and Concentrating Flavors

8.1.1.1: Viscosity

8.1.1.2: Thickening Agents

8.1.1.3: Types of Thickening Agents

8.1.1.4: Coagulation

8.1.1.5: Gelatinization

8.1.1.6: Crystallization

8.1.1.7: Non-traditional thickeners

8.1.1.8: Sauces

8.1.1.9: Low-temperature and sous-vide

8.1.2: Flour

8.1.2.1: Introduction - Understanding Ingredients

8.1.2.2: The History of Wheat Flour

8.1.2.3: Milling of Wheat

8.1.2.4: Flour Streams and Types of Wheat Flour

8.1.2.5: Flour Terms and Treatments

8.1.2.6: Flour Additives

8.1.2.7: Whole Grain and Artisan Milling

8.1.2.8: Flour in Baking

8.1.2.9: Rye Flour

8.1.2.10: Other Grains and Flours

8.1.3: Fat

8.1.3.1: Understanding Fats and Oils

8.1.3.2: Sources of Bakery Fats and Oils

8.1.3.3: Major Fats and Oils Used in Bakeries

8.1.3.4: Functions of Fat in Baking

8.1.4: Sugar

8.1.4.1: Sugar Chemistry (ADD US)

8.1.4.2: Sugar Refining

8.1.4.3: The Application of Sugar

8.1.4.4: Agave

8.1.4.5: Honey

8.1.4.6: Malt

8.1.4.7: Maple Syrup (ADD US)

8.1.4.8: Sugar Substitutes (ADD US)

8.1.4.9: Glucose/Dextrose

8.1.5: Leavening Agents
8.1.5.1: Introduction to Leavening Agents
8.1.5.2: Yeast
8.1.5.3: The Functions of Yeast
8.1.5.4: Using Yeast in Baking
8.1.5.5: Baking Powder
8.1.5.6: Sodium Bicarbonate
8.1.5.7: Ammonium Bicarbonate
8.1.5.8: Water Hardness and pH

8.1.6: Dairy Products
8.1.6.1: Introduction to Dairy Products
8.1.6.2: Milk
8.1.6.3: Milk Products ADD US
8.1.6.4: Milk in bread baking
8.1.6.5: Yogurt
8.1.6.6: Lactose
8.1.6.7: Cheese

8.1.7: Eggs
8.1.7.1: Eggs Grade
8.1.7.2: Composition and Nutrition
8.1.7.3: Egg Products
8.1.7.4: The Function of Eggs
8.1.7.5: Storing Eggs

8.1.8: Chocolate
8.1.8.1: From the Cocoa Bean to the Finished Chocolate
8.1.8.2: Chocolate Produced for the Baking Industry
8.1.8.3: Couverture
8.1.8.4: Definitions and Regulations (ADD US)

8.1.9: Spices
8.1.9.1: Elements of Taste
8.1.9.2: Introduction to Salt
8.1.9.3: Origins of Salt
8.1.9.4: Functions of Salt in Baking
8.1.9.5: Using Salt in Fermented Doughs
8.1.9.6: Storing Salt
8.1.9.7: Introduction to Spices and Other Flavorings
8.1.9.8: Seasoning and Flavoring
8.1.9.9: Herbs
8.1.9.10: Spices

8.1.9.11: Flavorings in Baking

8.2: Aquatic Chemistry

8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems

8.2.2: 2. Carbonate Equilibria in Natural Waters

8.2.3: 3. Redox Equilibria in Natural Waters

8.2.4: 4. Solids in Contact With Natural Waters

8.2.5: Fundamental Characteristics of Water

8.2.6: Natural Water

8.2.7: Water Biology

8.2.8: Water Chemistry

8.2.9: Water Physics

8.2.10: Water Treatment

8.3: Biochemical Cycles

8.3.1: Biochemical Cycles

8.3.2: Carbon Cycle

8.3.3: Nitrogen Cycle

8.4: Toxicology MSDT

8.4.1: Pathophysiology

8.4.1.1: What is Pathophysiology?

8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity

8.4.1.3: Cellular Response to Toxicant-Induced Injury

8.4.1.4: Repair and Adaptation

8.4.1.5: Patterns of Toxic Injury

8.4.1.6: Final Evaluation

8.4.2: Biochemistry and Molecular Genetics

8.4.2.1: Introduction to Biomolecules and Cell Components

8.4.2.2: Cell Structure and Subcellular Compartments

8.4.2.3: DNA and RNA Metabolism

8.4.2.4: Epigenetic Mechanisms

8.4.2.5: Section 2 Final Evaluation

8.4.3: Principles of Genetic Toxicology

8.4.3.1: Introduction to Genetic-toxicology Assay

8.4.3.2: Different Genetic Damages or Mutations

8.4.3.3: Different Genetic-Toxicology Assays

8.4.3.4: Different Cytotoxicity Assays

8.4.3.5: Epigenetics Assay

8.4.3.6: Section 3 Final Evaluation

8.4.4: Applied Systems Toxicology

8.4.4.1: Systems Toxicology

8.4.4.2: Dose Level and Applied Toxicology
8.4.4.3: Tools and Technologies in Systems Toxicology
8.4.4.4: Other Approaches for Predictive Toxicity Modeling
8.4.4.5: Technologies Used In Systems Biology/Toxicology
8.4.4.6: Takeaways Summary
8.4.4.7: Section 4 Final Evaluation

8.4.5: Regulatory Toxicology
8.4.5.1: Introduction to Regulatory Toxicology
8.4.5.2: Global Regulatory Toxicology
8.4.5.3: Topic 3- Regional Regulatory Toxicology
8.4.5.4: National Regulatory Toxicology
8.4.5.5: State Regulatory Toxicology
8.4.5.6: Non-Governmental Regulatory Toxicology
8.4.5.7: Section 5 Final Evaluation

8.4.6: Principles of Toxicology
8.4.6.1: Introduction to Toxicology
8.4.6.1.1: What is toxicology?
8.4.6.1.2: Basic Terminology

8.4.6.2: Dose and Dose Response
8.4.6.2.1: Dose and It's Impact on Toxicity
8.4.6.2.2: The Dose Response Relationship
8.4.6.2.3: Dose Estimates of Toxic Effects
8.4.6.2.4: Determining the Safety of a Drug
8.4.6.2.5: NOAEL and LOAEL

8.4.6.3: Toxic Effects
8.4.6.3.1: Types of Toxic Effects
8.4.6.3.2: Factors Affecting Toxicity
8.4.6.3.3: Systemic Toxic Effects
8.4.6.3.4: Organ Specific Toxic Effects

8.4.6.4: Interactions
8.4.6.4.1: Interactions

8.4.6.5: Toxicity Testing Methods
8.4.6.5.1: Testing and Assessing Toxicity
8.4.6.5.2: Clinical Investigations and Other Types of Human Data
8.4.6.5.3: Epidemiology Studies

8.4.6.6: Risk Assessment
8.4.6.6.1: Risk Assessment
8.4.6.6.2: Hazard Identification
8.4.6.6.3: Dose-Response Assessment
8.4.6.6.4: Exposure Assessment

8.4.6.6.5: Risk Characterization

8.4.6.7: Exposure Standards and Guidelines
8.4.6.7.1: Exposure Standards and Guidelines
8.4.6.7.2: Regulation of Consumer Products and Drug Safety
8.4.6.7.3: Environmental Exposure Standards/Guidelines
8.4.6.7.4: Occupational (Workplace) Exposure Standards/Guidelines/Approaches

8.4.6.8: Basic Physiology
8.4.6.8.1: Introduction to Basic Physiology
8.4.6.8.2: Homeostasis
8.4.6.8.3: Organs and Organ Systems
8.4.6.8.4: Tissues
8.4.6.8.5: Cells
8.4.6.8.6: Chemicals

8.4.6.9: Introduction to Toxicokinetics
8.4.6.9.1: What is Toxicokinetics

8.4.6.10: Absorption
8.4.6.10.1: Introduction to Absorption
8.4.6.10.2: Gastrointestinal Tract
8.4.6.10.3: Respiratory Tract
8.4.6.10.4: Dermal Route
8.4.6.10.5: Other Routes of Exposure

8.4.6.11: Distribution
8.4.6.11.1: Introduction to Distribution
8.4.6.11.2: Influence of Route of Exposure
8.4.6.11.3: Disposition Models
8.4.6.11.4: Structural Barriers to Distribution
8.4.6.11.5: Storage Sites

8.4.6.12: Biotransformation
8.4.6.12.1: Introduction to Biotransformation
8.4.6.12.2: Chemical Reactions
8.4.6.12.3: Biotransformation Sites
8.4.6.12.4: Modifiers of Biotransformation

8.4.6.13: Excretion
8.4.6.13.1: Introduction to Secretion
8.4.6.13.2: Urinary Excretion
8.4.6.13.3: Fecal Excretion
8.4.6.13.4: Exhaled Air
8.4.6.13.5: Other Routes

8.4.6.14: Cellular Toxicology
8.4.6.14.1: Adaptation

8.4.6.14.2: Cell Damage and Tissue Repair
8.4.6.14.3: Cancer
8.4.6.14.4: Neurotoxicity

8.4.6.15: Intuitive Toxicology and Risk Communication
8.4.6.15.1: Intuitive Toxicology
8.4.6.15.2: Risk Communication

8.4.6.16: Environmental Toxicology, Environmental Health, and One Health
8.4.6.16.1: Environmental Toxicology
8.4.6.16.2: Environmental Health
8.4.6.16.3: One Health

8.4.6.17: Conclusion

8.5: Food to energy metabolic pathways
8.5.1: Basics of metabolism
8.5.2: Digestion of food
8.5.3: Oxidation of glucose -the glycolysis
8.5.4: Citric acid cycle
8.5.5: Oxidative Phosphorylation
8.5.6: Oxidation of fatty acids
8.5.7: Degradation of amino acids

This page titled 8: Optional- Special topics is shared under a license and was authored, remixed, and/or curated by LibreTexts.

8.1: CHEMISTRY OF COOKING (RODRIGUEZ-VELAZQUEZ)

People around the world are fascinated about the preparation of food for eating. There are countless cooking books, TV shows, celebrity chefs and kitchen gadgets that make cooking an enjoyable activity for everyone. The chemistry of cooking course seeks to understand the science behind our most popular meals by studying the behavior of atoms and molecules present in food. This book is intended to give students a basic understanding of the chemistry involved in cooking such as caramelization, Maillard reaction, acid-base reactions, catalysis, and fermentation. Students will be able to use chemistry language to describe the process of cooking, apply chemistry knowledge to solve questions related to food, and ultimately create their own recipes.

8.1.1: Thickening and Concentrating Flavors

8.1.1.1: Viscosity

8.1.1.2: Thickening Agents

8.1.1.3: Types of Thickening Agents

8.1.1.4: Coagulation

8.1.1.5: Gelatinization

8.1.1.6: Crystallization

8.1.1.7: Non-traditional thickeners

8.1.1.8: Sauces

8.1.1.9: Low-temperature and sous-vide

8.1.2: Flour

8.1.2.1: Introduction - Understanding Ingredients

8.1.2.2: The History of Wheat Flour

8.1.2.3: Milling of Wheat

8.1.2.4: Flour Streams and Types of Wheat Flour

8.1.2.5: Flour Terms and Treatments

8.1.2.6: Flour Additives

8.1.2.7: Whole Grain and Artisan Milling

8.1.2.8: Flour in Baking

8.1.2.9: Rye Flour

8.1.2.10: Other Grains and Flours

8.1.3: Fat

8.1.3.1: Understanding Fats and Oils

8.1.3.2: Sources of Bakery Fats and Oils

8.1.3.3: Major Fats and Oils Used in Bakeries

8.1.3.4: Functions of Fat in Baking

8.1.4: Sugar

8.1.4.1: Sugar Chemistry (ADD US)

8.1.4.2: Sugar Refining

8.1.4.3: The Application of Sugar

8.1.4.4: Agave

8.1.4.5: Honey
8.1.4.6: Malt
8.1.4.7: Maple Syrup (ADD US)
8.1.4.8: Sugar Substitutes (ADD US)
8.1.4.9: Glucose/Dextrose

8.1.5: Leavening Agents

8.1.5.1: Introduction to Leavening Agents
8.1.5.2: Yeast
8.1.5.3: The Functions of Yeast
8.1.5.4: Using Yeast in Baking
8.1.5.5: Baking Powder
8.1.5.6: Sodium Bicarbonate
8.1.5.7: Ammonium Bicarbonate
8.1.5.8: Water Hardness and pH

8.1.6: Dairy Products

8.1.6.1: Introduction to Dairy Products
8.1.6.2: Milk
8.1.6.3: Milk Products ADD US
8.1.6.4: Milk in bread baking
8.1.6.5: Yogurt
8.1.6.6: Lactose
8.1.6.7: Cheese

8.1.7: Eggs

8.1.7.1: Eggs Grade
8.1.7.2: Composition and Nutrition
8.1.7.3: Egg Products
8.1.7.4: The Function of Eggs
8.1.7.5: Storing Eggs

8.1.8: Chocolate

8.1.8.1: From the Cocoa Bean to the Finished Chocolate
8.1.8.2: Chocolate Produced for the Baking Industry
8.1.8.3: Couverture
8.1.8.4: Definitions and Regulations (ADD US)

8.1.9: Spices

8.1.9.1: Elements of Taste
8.1.9.2: Introduction to Salt
8.1.9.3: Origins of Salt
8.1.9.4: Functions of Salt in Baking
8.1.9.5: Using Salt in Fermented Doughs

8.1.9.6: Storing Salt
8.1.9.7: Introduction to Spices and Other Flavorings
8.1.9.8: Seasoning and Flavoring
8.1.9.9: Herbs
8.1.9.10: Spices
8.1.9.11: Flavorings in Baking

This page titled 8.1: Chemistry of Cooking (Rodriguez-Velazquez) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- Chemistry of Cooking (Rodriguez-Velazquez) by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1: THICKENING AND CONCENTRATING FLAVORS

All sections in this chapter. So three more pages need to be made by you.

8.1.1.1: Viscosity

8.1.1.2: Thickening Agents

8.1.1.3: Types of Thickening Agents

8.1.1.4: Coagulation

8.1.1.5: Gelatinization

8.1.1.6: Crystallization

8.1.1.7: Non-traditional thickeners

8.1.1.8: Sauces

8.1.1.9: Low-temperature and sous-vide

Contributors and Attributions

Sorangel Rodriguez-Velazquez (American University).

Chemistry of Cooking by Sorangel Rodriguez-Velazquez is licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International License, except where otherwise noted

This page titled 8.1.1: Thickening and Concentrating Flavors is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1: Thickening and Concentrating Flavors by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.1: Viscosity

Viscous means “sticky”’ and the term viscosity refers to the way in which the chocolate flows. Chocolate comes in various viscosities, and the confectioner chooses the one that is most appropriate to his or her needs. The amount of cocoa butter in the chocolate is largely responsible for the viscosity level. Emulsifiers like lecithin can help thin out melted chocolate, so it flows evenly and smoothly. Because it is less expensive than cocoa butter at thinning chocolate, it can be used to help lower the cost of chocolate.

Molded pieces such as Easter eggs require a chocolate of less viscosity. That is, the chocolate should be somewhat runny so it is easier to flow into the moulds. This is also the case for coating cookies and most cakes, where a thin, attractive and protective coating is all that is needed. A somewhat thicker chocolate is advisable for things such as ganache and flavoring of creams and fillings. Where enrobers (machines to dip chocolate centers) are used, the chocolate may also be thinner to ensure that there is an adequate coat of couverture.

Viscosity varies between manufacturers, and a given type of chocolate made by one manufacturer may be available in more than one viscosity. Bakers sometimes alter the viscosity depending on the product. A vegetable oil is sometimes used to thin chocolate for coating certain squares. This makes it easier to cut afterwards.

Chips, Chunks, and Other Baking Products

Content and quality of chocolate chips and chunks vary from one manufacturer to another. This chocolate is developed to be more heat stable for use in cookies and other baking where you want the chips and chunks to stay whole. Ratios of chocolate liquor, sugar, and cocoa butter differ. All these variables affect the flavor. Chips and chunks may be pure chocolate or have another fat substituted for the cocoa butter. Some high quality chips have up to 65% chocolate liquor, but in practice, liquor content over 40% tends to smear in baking, so high ratios defeat the purpose.

Many manufacturers package their chips or chunks by count (ct) size. This refers to how many pieces there are in 1 kg of the product. As the count size number increases, the size of the chip gets smaller. With this information, you can choose the best size of chip for the product you are producing.

Other chocolate products available are chocolate sprinkles or “hail,” used as a decoration; chocolate curls, rolls, or decorative shapes for use on cakes and pastries; and chocolate sticks or “batons,” which are often baked inside croissants.

This page titled 8.1.1.1: Viscosity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.1: Viscosity by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.2: Thickening Agents

Learning Objectives

- Identify and describe thickening agents used in the food service industry
- Describe the production and properties of thickening agents Describe the function of thickening agents in baking

Two types of thickening agents are recognized: starches and gums. Most thickening agents are of vegetable origin; the only exception is gelatin. All the starches are products of the land; some of the gums are of marine origin.

Bakers use thickening agents primarily to:

- Make fillings easier to handle and bake
- Firm up products to enable them to be served easily
- Provide a glossy “skin” to improve finish and reduce drying

This page titled 8.1.1.2: Thickening Agents is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.2: Thickening Agents by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.3: Types of Thickening Agents

Cornstarch

Cornstarch is the most common thickening agent used in the industry. It is mixed with water or juice and boiled to make fillings and to give a glossy semi-clear finish to products. Commercial cornstarch is made by soaking maize in water containing sulphur dioxide. The soaking softens the corn and the sulphur dioxide prevents possible fermentation. It is then crushed and passed to water tanks where the germ floats off. The mass is then ground fine and, still in a semi-fluid state, passed through silk screens to remove the skin particles. After filtration, the product, which is almost 100% starch, is dried.

Cornstarch in cold water is insoluble, granular, and will settle out if left standing. However, when cornstarch is cooked in water, the starch granules absorb water, swell, and rupture, forming a translucent thickened mixture. This phenomenon is called *gelatinization*. Gelatinization usually begins at about 60°C (140°F), reaching completion at the boiling point.

The commonly used ingredients in a starch recipe affect the rate of gelatinization of the starch. Sugar, added in a high ratio to the starch, will inhibit the granular swelling. The starch gelatinization will not be completed even after prolonged cooking at normal temperature. The result is a filling of thin consistency, dull color, and a cereal taste. Withhold some of the sugar from the cooking step in such cases, and add it after gelatinization of the starch has been completed.

Other ingredients such as egg, fat, and dry milk solids have a similar effect. Fruits with high acidity such as rhubarb will also inhibit starch setting. Cook the starch paste first and add the fruit afterward.

In cooking a filling, about 1.5 kg (3 1/3 lb.) of sugar should be cooked with the water or juice for every 500 g (18 oz.) of starch used as a thickener. Approximately 100 g (4 oz.) of starch is used to thicken 1 L of water or fruit juice. The higher the acidity of the fruit juice, the more thickener required to hold the gel. Regular cornstarch thickens well but makes a cloudy solution. Another kind of cornstarch, waxy maize starch, makes a more fluid mix of great clarity.

Pre-gelatinized Starches

Pre-gelatinized starches are mixed with sugar and then added to the water or juice. They thicken the filling in the presence of sugar and water without heating. This is due to the starch being precooked and not requiring heat to enable it to absorb and gelatinize. There are several brands of these starches on the market (e.g., Clear Jel), and they all vary in absorption properties. For best results, follow the manufacturer's guidelines. Do not put pre-gelatinized starch directly into water, as it will form lumps immediately.

> Note
>
> If fruit fillings are made with these pre-cooked starches, there is a potential for breakdown if the fillings are kept. Enzymes in the uncooked fruit may "attack" the starch and destroy some of the gelatinized structure. For example, if you are making a week's supply of pie filling from fresh rhubarb, use a regular cooked formula.

Arrowroot

Arrowroot is a highly nutritious farinaceous starch obtained from the roots and tubers of various West Indian plants. It is used in the preparation of delicate soups, sauces, puddings, and custards.

Agar-Agar

Agar-agar is a jelly-like substance extracted from red seaweed found off the coasts of Japan, California, and Sri Lanka. It is available in strips or slabs and in powder form. Agar-agar only dissolves in hot water and is colorless. Use it at 1% to make a firm

gel. It has a melting point much higher than gelatin and its jellying power is eight times greater. It is used in pie fillings and to some extent in the stiffening of jams. It is a permitted ingredient in some dairy products, including ice cream at 0.5%. One of its largest uses is in the production of materials such as piping jelly and marshmallow.

Algin (Sodium Alginate)

Extracted from kelp, this gum dissolves in cold water and a 1% concentration to give a firm gel. It has the disadvantage of not working well in the presence of acidic fruits. It is popular in uncooked icings because it works well in the cold state and holds a lot of moisture. It reduces stickiness and prevents recrystallization.

Carrageenan or Irish Moss

Carrageenan is another marine gum extracted from red seaweed. It is used as a thickening agent in various products, from icing stabilizers to whipping cream, at an allowable rate of 0.1% to 0.5%.

Gelatin

Gelatin is a glutinous substance made from the bones, connective tissues, and skins of animals. The calcium is removed and the remaining substance is soaked in cold water. Then it is heated to 40°C to 60°C (105°F 140°F). The partially evaporated liquid is defatted and coagulated on glass plates and then poured into moulds. When solid, the blocks of gelatin are cut into thin layers and dried on wire netting.

Gelatin is available in sheets of leaf gelatin, powders, granules, or flakes. Use it at a 1% ratio. Like some of the other gelling agents, acidity adversely affects its gelling capacity.

The quality of gelatin often varies because of different methods of processing and manufacturing. For this reason, many bakers prefer leaf gelatin because of its reliable strength.

Gum Arabic or Acacin

This gum is obtained from various kinds of trees and is soluble in hot or cold water. Solutions of gum arabic are used in the bakery for glazing various kinds of goods, particularly marzipan fruits.

Gum Tragacanth

This gum is obtained from several species of Astragalus, low-growing shrubs found in Western Asia. It can be purchased in flakes or powdered form. Gum tragacanth was once used to make gum paste and gum paste wedding ornaments, but due to high labour costs and a prohibitive price for the product, its use nowadays is uncommon.

Pectin

Pectin is a mucilaginous substance (gummy substance extracted from plants), occurring naturally in pears, apples, quince, oranges, and other citrus fruits. It is used as the gelling agent in traditional jams and jellies.

This page titled 8.1.1.3: Types of Thickening Agents is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.3: Types of Thickening Agents by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.4: Coagulation

Coagulation is defined as the transformation of proteins from a liquid state to a solid form. Once proteins are coagulated, they cannot be returned to their liquid state. Coagulation often begins around 38°C (100°F), and the process is complete between 71°C and 82°C (160°F and 180°F). Within the baking process, the natural structures of the ingredients are altered irreversibly by a series of physical, chemical, and biochemical interactions. The three main types of protein that cause coagulation in the bakeshop are outlined below.

Egg proteins

Eggs contain many different proteins. The white, or albumen, contains approximately 40 different proteins, the most predominant being ovalbumin (54%) and ovotransferrin (12%). The yolk contains mostly lipids (fats), but also lipoproteins. These different proteins will all coagulate when heated, but do so at different temperatures. The separated white of an egg coagulates between 60°C and 65°C (140°F and 149°F) and the yolk between 62°C and 70°C (144°F and 158°F), which is why you can cook an egg and have a fully set white and a still runny yolk. These temperatures are raised when eggs are mixed into other liquids. For example, the coagulation and thickening of an egg, milk, and sugar mixture, as in custard, will take place between 80°C and 85°C (176°F and 185°F) and will start to curdle at 88°C to 90°C (190°F and 194°F).

Dairy and soy proteins

Casein, a semi-solid substance formed by the coagulation of milk, is obtained and used primarily in cheese. Rennet, derived from the stomach linings of cattle, sheep, and goats, is used to coagulate, or thicken, milk during the cheese-making process. Plant-based rennet is also available. Chymosin (also called rennin) is the enzyme used to produce rennet, and is responsible for curdling the milk, which will then separate into solids (curds) and liquid (whey).

Milk and milk products will also coagulate when treated with an acid, such as citric acid (lemon juice) or vinegar, used in the preparation of fresh ricotta, and tartaric acid, used in the preparation of mascarpone, or will naturally curdle when sour as lactic acid develops in the milk. In some cases, as in the production of yogurt or crème fraîche, acid-causing bacteria are added to the milk product to cause the coagulation. Similarly, tofu is made from soybean milk that has been coagulated with the use of either salt, acid, or enzyme-based coagulants.

Flour proteins (gluten)

Two main proteins are found in wheat flour: *glutenin* and *gliadin* (smaller quantities are also found in other grains). During mixing and in contact with liquid, these two form into a stretchable substance called gluten. The coagulation of gluten is what happens when bread bakes; that is, it is the firming or hardening of these gluten proteins, usually caused by heat, which solidify to form a firm structure.

This page titled 8.1.1.4: Coagulation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.4: Coagulation by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.5: Gelatinization

Hydrocolloids

A hydrocolloid is a substance that forms a gel in contact with water. There are two main categories: Thermo-reversible gel: A gel that melts upon reheating and sets upon cooling. Examples are gelatin and agar agar. Thermo-irreversible gel: A gel that does not melt upon reheating. Examples are cornstarch and pectin. Excessive heating, however, may cause evaporation of the water and shrinkage of the gel. Hydrocolloids do not hydrate (or dissolve) instantly, and that hydration is associated with swelling, which easily causes lumping. It is therefore necessary to disperse hydrocolloids in water. Classically, this has always been done with cornstarch, where a portion of the liquid from the recipe is mixed to form a “slurry” before being added to the cooking liquid. This can also be done with an immersion blender or a conventional blender, or by mixing the hydrocolloid with a helping agent such as sugar, oil, or alcohol prior to dispersion in water.

Starches

Starch **gelatinization** is the process where starch and water are subjected to heat, causing the starch granules to swell. As a result, the water is gradually absorbed in an irreversible manner. This gives the system a viscous and transparent texture. The result of the reaction is a gel, which is used in sauces, puddings, creams, and other food products, providing a pleasing texture. Starch-based gels are thermoirreversible, meaning that they do not melt upon heating (unlike gelatin, which we will discuss later). Excessive heating, however, may cause evaporation of the water and shrinkage of the gel.

The most common examples of starch gelatinization are found in sauce and pasta preparations and baked goods. In sauces, starches are added to liquids, usually while heating.

- The starch will absorb liquid and swell, resulting in the liquid becoming thicker. The type of starch determines the final product. Some starches will remain cloudy when cooked; others will remain clear.
- Pasta is made mostly of semolina wheat (durum wheat flour), which contains high amounts of starch. When pasta is cooked in boiling water, the starch in the pasta swells as it absorbs water, and as a result the texture of the pasta softens.

Starch molecules make up the majority of most baked goods, so starch is an important part of the structure. Although starches by themselves generally can’t support the shape of the baked items, they do give bulk to the structure. Starches develop a softer structure when baked than proteins do. The softness of the crumb of baked bread is due largely to the starch. The more protein structure there is, the chewier the bread.

Starches can be fairly straightforward extracts of plants, such as cornstarch, tapioca, or arrowroot, but there are also modified starches and pre-gelatinized starches available that have specific uses. See Table 1 for a list of different thickening and binding agents and their characteristics.

Table 1 – Common starches and gels used in the bakeshop

Starch or Gel	*Ratio*	*Preparation*	*Characteristics and Uses*
Cornstarch	20-40 g starch thickens 1 L liquid	A slurry (mixture of cornstarch and water) is mixed and added to a simmering liquid while whisking until it dissolves and the liquid thickens; or Cornstarch mixed with sugar, and cold liquid added Thickened mixture simmered until no starch taste remains	Used to thicken sauces when a clear glossy texture is desired, such as dessert sauces and in Asian-inspired dishes Translucent, thickens further as it cools; forms a “sliceable” gel Sensitive to extended heat exposure, so products become thin if held at heat for long periods of time

Starch or Gel	Ratio	Preparation	Characteristics and Uses
Agar agar	15-30g agar agar sets 1 L liquid	Powder dissolved in cold water Added to cold or simmering liquid Activates with heat, sets when cold	Extracted from seaweed Used in Asian desserts and molecular gastronomy cooking Used in place of gelatin in vegetarian dishes Clear firm texture Does not thin when reheated, thickens more when cold

Waxy maize, waxy rice
Dissolved in cold water 20-40 g starch thickens 1 L liquid
Added to hot liquid while whisking until it dissolves and the liquid thickens
Used in desserts and dessert sauces Clear, does not thicken further as it cools Does not gel at cool temperatures, good for cold sauces Quite stable at extreme temperatures (heat and freezing)
Modified starches
Dissolved in cold water 20-40 g starch thickens 1 L liquid
Added to hot liquid while whisking until it dissolves and the liquid thickens
Modified starches are often used in commercially processed foods and convenience products Modified to improve specific characteristics (e.g., stability or texture under extreme conditions; heat and freezing) Translucent, thickens further as it cools
Pre-gelatinized starches
Powder, dissolved in cold liquid 20-40 g starch thickens 1 L liquid
Added to liquid at any temperature
Used when thickening liquids that might lose color or flavor during cooking Become viscous without the need for additional cooking Translucent, fairly clear, shiny, does NOT gel when cold
Arrowroot
Powder, dissolved in cold liquid 20-40 g starch thickens 1 L liquid
Added to hot liquid while whisking until it dissolves and the liquid thickens
Derived from cassava root Used in Asian cuisines Very clear; possesses a gooey texture Translucent, shiny, very light gel when cold
Gelatin 15-30 g gelatin sets 1 L liquid
Powder or sheets (leaves) dissolved in cold water Added to cold or simmering liquid Activates with heat, sets when cold
Derived from collagens in bones and meats of animals Used in aspic, glazes, cold sauces, and desserts Clear, firm texture Dissolves when reheated, thickens when cold

Gelling agents

Gelatin is a water-soluble protein extracted from animal tissue and used as a gelling agent, a thickener, an emulsifier, a whipping agent, a stabilizer, and a substance that imparts a smooth mouth feel to foods. It is thermo-reversible, meaning the setting properties or action can be reversed by heating. Gelatin is available in two forms: powder and sheet (leaf). Gelatin is often used to stabilize whipped cream and mousses; confectionery, such as gummy bears and marshmallows; desserts including pannacotta; commercial products like Jell-O; “lite” or low-fat versions of foods including some margarines; and dairy products such as yogurt and ice cream. Gelatin is also used in hard and soft gel capsules for the pharmaceutical industry.

Agar agar is an extract from red algae and is often used to stabilize emulsions or foams and to thicken or gel liquids. It is thermo-reversible and heat resistant. It is typically hydrated in boiling liquids and is stable across a wide range of acidity levels. It begins to gel once it cools to around 40°C (100°F) and will not melt until it reaches 85°C (185°F).

Pectin

Pectin is taken from citrus and other tree fruits (apples, pears, etc.). Pectin is found in many different foods such as jam, milk-

based beverages, jellies, sweets, and fruit juices. Pectin is also used in molecular gastronomy mainly as a gelling agent, thickener, and stabilizer.

There are a variety of types of pectin that react differently according to the ingredients used. Low-methoxyl pectin (which is activated with the use of calcium for gelling) and high-methoxyl pectin that requires sugar for thickening are the two most common types used in cooking. High-methoxyl pectin is what is traditionally used to make jams and jellies. Low-methoxyl pectin is often used in modern cuisine due to the thermo-irreversible gel that it forms and its good reaction to calcium. Its natural capability to emulsify and gel creates stable preparations.

Increasingly, cooks, bakers, and pastry chefs are turning to many different gels, chemicals, and other substances used in commercial food processing as new ingredients to modify liquids or other foods. These will be outlined in detail in the section on molecular gastronomy.

This page titled 8.1.1.5: Gelatinization is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.5: Gelatinization by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.6: Crystallization

Many factors can influence crystallization in food. Controlling the crystallization process can affect whether a particular product is spreadable, or whether it will feel gritty or smooth in the mouth. In some cases, crystals are something you try to develop; in others, they are something you try to avoid. It is important to know the characteristics and quality of the crystals in different food. Butter, margarine, ice cream, sugar, and chocolate all contain different types of crystals, although they all contain fat crystals. For example, ice cream has fat crystals, ice crystals, and sometimes lactose crystals. The fact that sugar solidifies into crystals is extremely important in candy making. There are basically two categories of candies: crystalline (candies that contain crystals in their finished form, such as fudge and fondant); and non-crystalline (candies that do not contain crystals, such as lollipops, taffy, and caramels). Recipe ingredients and procedures for non-crystalline candies are specifically designed to prevent the formation of sugar crystals because they give the resulting candy a grainy texture. One way to prevent the crystallization of sucrose in candy is to make sure that there are other types of sugar—usually fructose and glucose—to get in the way and slow down or inhibit the process. Acids can also be added to “invert” the sugar, and to prevent or slow down crystallization. Fats added to certain confectionery items will have a similar effect.

When boiling sugar for any application, the formation of crystals is generally not desired. These are some of the things that can promote crystal growth:

- Pot and utensils that are not clean
- Sugar with impurities in it (A scoop used in the flour bin, and then used for sugar, may have enough particles on it to promote crystallization.)
- Water with a high mineral content (“hard water”)
- Too much stirring (agitation) during the boiling phase

Crystallization may be prevented by adding an interferent, such as acid (lemon, vinegar, tartaric, etc.) or glucose or corn syrup, during the boiling procedure. As mentioned above, ice cream can have ice and fat crystals that co-exist along with other structural elements (emulsion, air cells, and hydrocolloid stabilizers such as locust bean gum) that make up the “body” of the ice cream. Some of these components crystallize either partially or completely. The bottom line is that the nature of the crystalline phase in the food will determine the quality, appearance, texture, feel in the mouth, and stability of the product. The texture of ice cream is derived, in part, from the large number of small ice crystals. These small ice crystals provide a smooth texture with excellent melt-down and cooling properties. When these ice crystals grow larger during storage (recrystallization), the product becomes coarse and less enjoyable. Similar concerns apply to sugar crystals in fondant and frostings, and to fat crystals in chocolate, butter, and margarine.

Control of crystallization in fats is important in many food products, including chocolate, margarine, butter, and shortening. In these products, the aim is to produce the appropriate number, size, and distribution of crystals in the correct shape because the crystalline phase plays such a large role in appearance, texture, spreadability, and flavor release. Thus, understanding the processes that control crystallization is critical to controlling quality in these products.

To control crystallization in foods, certain factors must be controlled:

- Number and size of crystals
- Crystal distribution
- Proper polymorph (crystal shape)

Crystallization is important in working with chocolate. The tempering process, sometimes called precrystallization, is an important step that is used for decorative and moulding purposes, and is a major contributor to the mouth feel and enjoyment of chocolate. Tempering is a process that encourages the cocoa butter in the chocolate to harden into a specific crystalline pattern, which maintains the sheen and texture for a long time. When chocolate isn’t tempered properly it can have a number of problems. For example, it may not ever set up hard at room temperature; it may become hard, but look dull and blotchy; the internal texture may be spongy rather than crisp; and it can be susceptible to fat bloom, meaning the fats will migrate to the surface and make whitish streaks and blotches.

This page titled 8.1.1.6: Crystallization is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.6: Crystallization by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.7: Non-traditional thickeners

Non-traditional thickeners

In addition to traditional starches, there are new ways to thicken sauces and to change the texture of liquids. Some of these thickening agents work without heating and are simply blended with the cold liquid, such as modified starch or xanthan gum. These allow the creation of sauces and other liquids with a fresh, uncooked taste.

Foams, froths, and bubbles

Liquids can be stabilized with gelatin, lecithin, and other ingredients, and then used to create foams by whipping or using a special dispenser charged with nitrogen gas. A well-made foam adds an additional flavor dimension to the dish without adding bulk, and an interesting texture as the foamdissolves in the mouth (Figure 1).



Figure 1. "Dinner in the Dark 21-Dessert" by Esther Little is licensed under CC BY SA 2.0

Espuma

Espuma is the Spanish term for froth or foam, and it is created with the use of a siphon (ISO) bottle. This is a specific term, since culinary foams may be attained through other means.

Espuma from a siphon creates foam without the use of an emulsifying agent such as egg. As a result, it offers an unadulterated flavor of the ingredients used. It also introduces much more air into a preparation compared to other culinary aerating processes.

Espuma is created mainly with liquid that has air incorporated in it to create froth. But solid ingredients can be used too; these can be liquefied by cooking, puréeing, and extracting natural juices. It should be noted, though, that the best flavors to work with are those that are naturally diluted. Otherwise, the espuma tends to lose its flavor as air is introduced into it.

Stabilizers may be used alongside the liquids to help retain their shape longer; however, this is not always necessary. Prepared liquids can also be stored in a siphon bottle and kept for use. The pressure from the bottle will push out the aerated liquid, producing the espuma.

Foam

Foam is created by trapping air within a solid or liquid substance. Although culinary foams are most recently associated with

molecular gastronomy, they are part of many culinary preparations that date back to even earlier times. Mousse, soufflé, whipped cream, and froth in cappuccino are just some examples of common foams. Common examples of “set” foams are bread, pancakes, and muffins.

Foam does not rely on pressure to encase air bubbles into a substance. Like espuma, foam may also be created with the help of a surfactant and gelling or thickening agents to help it hold shape. The production of a culinary foam starts with a liquid or a solid that has been puréed. The thickening or gelling agent is then diluted into this to form a solution. Once dissolved, the solution is whipped to introduce air into it.

The process of whipping is done until the foam has reached the desired stiffness. Note that certain ingredients may break down if they are whipped for too long, especially without the presence of a stabilizing agent.

Gels Turning a liquid, such as a vegetable juice or raspberry purée, into a solid not only gives it a different texture but also allows the food to be cut into many shapes, enabling different visual presentations (Figure 2). Regular gelatin can be used as well as other gelling agents, such as agar agar, which is derived from red algae.



Figure 2. “Papayagelee” by hedonistin is licensed under CC BY NC 2.0

Brittle gels

Gelling agents are often associated with jelly-like textures, which may range from soft to firm. However, certain gels produced by specific agents may not fit this description.

Rather than forming an elastic or pliable substance, brittle gels may also be formed. These are gels that are firm in nature yet fragile at the same time. This characteristic is caused by the formation of a gel network that is weak and susceptible to breaking. This property allows brittle gels to crumble in the mouth and create a melt-in-the-mouth feeling. As a result, new sensations and textures are experienced while dining. At the same time, tastes within a dish are also enhanced due to the flavour release caused by the gel breakdown.Brittle gels are made by diluting the gelling agent into a liquid substance such as water, milk, or a stock. This mixture is left to set to attain a gelled end product. It should be noted that the concentration of gelling agents used, as well as the amount of liquid, both affect gelation.

Agar agar is a common agent used to create brittle gels. However, when combined with sugar it tends to create a more elastic substance. Low-acyl gellan gum, locust bean gum, and carrageenan also create brittle gels.

Fluid gels

A fluid gel is a cross between a sauce, gel, and purée. It is a controlled liquid that has properties of all three preparations. A fluid gel displays viscosity and fluidity at the same time, being thick yet still spreadable.

Fluid gels behave as solids when undisturbed, and flow when exposed to sufficient agitation. They are used in many culinary dishes where fluids need to be controlled, and they provide a rich, creamy texture.

A fluid gel is created using a base liquid that can come from many different sources. The base liquid is commonly extracted from fruits and vegetables, taken from stocks, or even puréed from certain ingredients. The longer the substance is exposed to stress, and the more intense the outside stress, the more fluidity is gained. More fluidity causes a finer consistency in the gel.

Fluid gels can be served either hot or cold, as many of the gelling agents used for such preparations are stable at high temperatures.

Drying and powdering

Drying a food intensifies its flavour and, of course, changes its texture. Eating a piece of apple that has been cooked and then dehydrated until crisp is very different from eating a fresh fruit slice. If the dehydrated food is powdered, it becomes yet another flavour and texture experience.

When maltodextrin (or tapioca maltodextrin) is mixed with fat, it changes to a powder. Because maltodextrin dissolves in water, peanut butter (or olive oil) that has been changed to a powder changes back to an oil in the mouth.

Freezing

In molecular gastronomy, liquid nitrogen is often used to freeze products or to create a frozen item without the use of a freezer.

Liquid nitrogen is the element nitrogen in a liquefied state. It is a clear, colourless liquid with a temperature of -196°C (-321°F). It is classified as a cryogenic fluid, which causes rapid freezing when it comes into contact with living tissues.

The extremely cold temperatures provided by this liquefied gas are most often used in modern cuisine to produce frozen foams and ice cream. After freezing food, nitrogen boils away, creating a thick nitrogen fog that may also add to the aesthetic features of a dish.

Given the extreme temperature of liquid nitrogen, it must be handled with care. Mishandling may cause serious burns to the skin. Nitrogen must be stored in special flasks and handled only by trained people. Aprons, gloves, and other specially designed safety gear should be used when handling liquid nitrogen.

Used mainly in the form of a coolant for molecular gastronomy, liquid nitrogen is not ingested. It is poured directly onto the food that needs to be cooled, causing it to freeze. Any remaining nitrogen evaporates, although sufficient time must be provided to allow the liquefied gas to be eliminated and for the dish to warm up to the point that it will not cause damage during consumption.

Spherification

Spherification is a modern cuisine technique that involves creating semi-solid spheres with thin membranes out of liquids. Spheres can be made in various sizes and of various firmnesses, such as the “caviar” shown in Figure 3. The result is a burst-in-the-mouth effect, achieved with the liquid. Both flavour and texture are enhanced with this culinary technique.

There are two versions of the spherification process: direct and reverse.

In direct spherification, a flavoured liquid (containing either sodium alginate, gellan gum, or carrageenan) is dripped into a water bath that is mixed with calcium (either calcium chloride or calcium lactate). The outer layer is induced by calcium to form a thin gel layer, leaving a liquid centre. In this version, the spheres are easily breakable and should be consumed immediately.

Calcium chloride and sodium alginate are the two basic components used for this technique. Calcium chloride is a type of salt used in cheese making, and sodium alginate is taken from seaweed. The sodium alginate is used to gel the chosen liquid by dissolving it directly into the fluid. This causes the liquid to become sticky, and proper dissolving must be done by mixing. The liquid is then left to set to eliminate any bubbles.

Once ready, a bath is prepared with calcium chloride and water. The liquid is then dripped into the bath using a spoon or syringe depending on the desired sphere size. The gel forms a membrane encasing the liquid when it comes into contact with the calcium

chloride. Once set, the spheres are then removed and rinsed with water to remove any excess calcium chloride.

In reverse spherification, a calcium-containing liquid (or ingredients mixed with a soluble calcium salt) is dripped into a setting bath containing sodium alginate. Surface tension causes the drop to become spherical. A skin of calcium alginate immediately forms around the top. Unlike in the direct version, the gelling stops and does not continue into the liquid orb. This results in thicker shells so the products do not have to be consumed immediately.



Figure 3. “White chocolate spaghetti with raspberry sauce and chocolate martini caviar” by ayngelina is licensed under CC BY NC-ND 2.0

Videos on spherification:

Direct: https://www.youtube.com/watch?v=BeRMBv95gLk

Reverse: https://www.youtube.com/watch?v=JPNo79U77yI

Specialty ingredients used in molecular gastronomy

There are a number of different ingredients used in molecular gastronomy as gelling, thickening, or emulsifying agents. Many of these are available in specialty food stores or can be ordered online.

Algin

Another name for sodium alginate, algin is a natural gelling agent taken from the cell walls of certain brown seaweed species.

Calcium chloride

Calcium chloride, also known as CaCl2, is a compound of chlorine and calcium that is a by-product of sodium bicarbonate (baking soda) manufacturing. At room temperature it is a solid salt, which is easily dissolved in water.

This is very salty and is often used for preservation, pickling, cheese production, and adding taste without increasing the amount of sodium. It is also used in molecular gastronomy in the spherification technique (see above) for the production of ravioli, spheres, pearls, and caviar (Figure 3).

Calcium lactate

Calcium lactate is a calcium salt resulting from the fermentation of lactic acid and calcium. It is a white crystalline power when solid and is highly soluble in cold liquids. It is commonly used as a calcium fortifier in various food products including beverages and supplements.

Calcium lactate is also used to regulate acidity levels in cheese and baking powder, as a food thickener, and as a preservative for fresh fruits. In molecular gastronomy, it is most commonly used for basic spherification and reverse spherification due to the

lack of bitterness in the finished products.

Like calcium chloride, calcium lactate is used alongside sodium alginate. In regular spherification, it is used in the bath. It is also used as a thickener in reverse spherification.

Carob bean gum

Carob bean gum is another name for locust bean gum. It is often used to stabilize, texturize, thicken, and gel liquids in the area of modern cuisine, although it has been a popular thickener and stabilizer for many years.

Carrageenan

Carrageenan refers to any linear sulfated polysaccharide taken from the extracts of red algae. This seaweed derivative is classified mainly as iota, kappa, and lambda. It is a common ingredient in many foods.

There are a number of purposes that it serves, including binding, thickening, stabilizing, gelling, and emulsifying. Carrageenan can be found in ice cream, salad dressings, cheese, puddings, and many more foods. It is often used with dairy products because of its good interaction with milk proteins. Carrageenan also works well with other common kitchen ingredients and offers a smooth texture and taste that blends well and does not affect flavour.

More often than not, carrageenan is found in powder form, which is hydrated in liquid before being used. For best results, carrageenan powder should be sprinkled in cold liquid and blended well to dissolve, although it may also be melted directly in hot liquids.

Citric acid

Classified as a weak organic acid, citric acid is a naturally occurring preservative that can be found in citrus fruits. Produced as a result of the fermentation of sugar, it has a tart to bitter taste and is usually in powder form when sold commercially. It is used mainly as a preservative and acidulent, and it is a common food additive in a wide range of foods such as candies and soda. Other than extending shelf life by adjusting the acidity or pH of food, it can also help enhance flavours. It works especially well with other fruits, providing a fresh taste.

In modern cooking, citric acid is often used as an emulsifier to keep fats and liquids from separating. It is also a common component in spherification, where it may be used as an acid buffer.

Gellan gum

Gellan gum is a water-soluble, high-molecular-weight polysaccharide gum that is produced through the fermentation of carbohydrates in algae by the bacterium Pseudomonas elodea. This fermented carbohydrate is purified with isopropyl alcohol, then dried and milled to produce a powder.

Gellan gum is used as a stabilizer, emulsifier, thickener, and gelling agent in cooking. Aspics and terrines are only some of the dishes that use gellan. It comes in both high-acyl and low-acyl forms. High-acyl gellan gum produces a flexible elastic gel, while low-acyl gellan gum will give way to a more brittle gel.

Like many other hydrocolloids, gellan gum is used with liquids. The powder is normally dispersed in the chosen liquid to dissolve it. Once dissolved, the solution is then heated to facilitate liquid absorption and gelling by the hydrocolloid. A temperature between 85°C and 95°C (185°F and 203°F) will start the
dissolution process. Gelling will begin upon cooling around 10°C and 80°C (50°F and 176°F).

Gellan gum creates a thermo-irreversible gel and can withstand high heat without reversing in form. This makes it ideal for the creation of warm gels.

Guar gum

Guar gum, or guaran, is a carbohydrate. This galactomannan is taken from the seeds of the guar plant by dehusking, milling, and screening. The end product is a pale, off-white, loose powder. It is most commonly used as a thickening agent and stabilizer for sauces and dressings in the food industry. Baked goods such as bread may also use guar gum to increase the amount of soluble fibre. At the same time, it also aids with moisture retention in bread and other baked items.

Being a derivative of a legume, guar gum is considered to be vegan and a good alternative to starches. In modern cuisine, guar gum is used for the creation of foams from acidic liquids, for fluid gels, and for stabilizing foams.

Guar gum must first be dissolved in cold liquid. The higher the percentage of guar gum used, the more viscous the liquid will become. Dosage may also vary according to the ingredients used as well as desired results and temperature.

Iota carrageenan

Iota carrageenan is a hydrocolloid taken from red seaweed (Eucheuma denticulatum). It is one of three varieties of carrageenan and is used mainly as a thickening or gelling agent.

Gels produced from iota carrageenan are soft and flexible, especially when used with calcium salts. It produces a clear gel that exhibits little syneresis. Iota is a fast-setting gel that is thermo-reversible and remains stable through freezing and thawing. In modern cuisine it is used to create hot foams as well as custards and jellies with a creamy texture.

Like most other hydrocolloids, iota carrageenan must first be dispersed and hydrated in liquid before use. Unlike lambda carrageenan, it is best dispersed in cold liquid. Once hydrated, the solution must be heated to about 70°C (158°F) with shear to facilitate dissolution. Gelling will happen between 40°C and 70°C (104°F and 158°F) depending on the number of calcium ions present.

Kappa carrageenan

Kappa carrageenan is another type of red seaweed extract taken specifically from Kappaphycus alvarezii. Like other types of carrageenan, it is used as a gelling, thickening, and stabilizing agent. When mixed with water, kappa carrageenan creates a strong and firm solid gel that may be brittle in texture.

This particular variety of carrageenan blends well with milk and other dairy products. Since it is taken from seaweed, it is considered to be vegan and is an alternative to traditional gelling agents such as gelatin.

Kappa carrageenan is used in various cooking preparations including hot and cold gels, jelly toppings, cakes, breads, and pastries. When used in molecular gastronomy preparations and other dishes, kappa carrageenan should be dissolved in cold liquid.

Once dispersed, the solution must be heated between 40°C and 70°C (104°F and 158°F). Gelling will begin between 30°C and 60°C (86°F and 140°F). Kappa carrageenan is a thermo-reversible gel and will stay stable up to 70°C (158°F). Temperatures beyond this will cause the gel to melt and become liquid once again.

Locust bean gum

Locust bean gum, also known as LBG and carob bean gum, is a vegetable gum derived from Mediterranean-region carob tree seeds. This hydrocolloid is used to stabilize, texturize, thicken, and gel liquids in modern cuisine, although it has been a popular thickener and stabilizer for many years.

It has a neutral taste that does not affect the flavour of food that it is combined with. It also provides a creamy mouth feel and has reduced syneresis when used alongside pectin or carrageenan for dairy and fruit applications. The neutral behaviour of this hydrocolloid makes it ideal for use with a wide range of ingredients.

To use locust bean gum, it must be dissolved in liquid. It is soluble with both hot and cold liquids.

Maltodextrin

Maltodextrin is a sweet polysaccharide that is produced from starch, corn, wheat, tapioca, or potato through partial hydrolysis and spray drying. This modified food starch is a white powder that has the capacity to absorb and hold water as well as oil. It is an ideal additive since it has fewer calories than sugar and is easily absorbed and digested by the body in the form of glucose.

Coming from a natural source, it ranges from nearly flavourless to fairly sweet without any odour. Maltodextrin is a common ingredient in processed foods such as soda and candies. In molecular gastronomy, it can be used both as a thickener and a stabilizer for sauces and dressings, for encapsulation, and as a sweetener. In many cases, it is also used as an aroma carrier due to its capacity to absorb oil. It is also often used to make powders or pastes out of fat.

Sodium alginate

Sodium alginate, which is also called algin, is a natural gelling agent taken from the cell walls of certain brown seaweed species. This salt is obtained by drying the seaweed, followed by cleaning, boiling, gelling, and pulverizing it. A light yellow powder is produced from the process. When dissolved in liquids, sodium alginate acts as a thickener, creating a viscous fluid. Conversely, when it is used with calcium it forms a gel through a cold process.

In molecular gastronomy, sodium alginate is most commonly used as a texturizing agent. Foams and sauces may be created with it. It is also used in spherification for the creation of pearls, raviolis, mock caviar, marbles, and spheres. Sodium alginate can be used directly by dissolving it into the liquid that needs to be gelled, as in the case of basic spherification. It may also be used inversely by adding it directly to a bath, as in the case of reverse spherification.

This versatile product is soluble in both hot and cold liquids, and gels made with it will set at any temperature.

Soy lecithin

Soy lecithin, also called just lecithin, is a natural emulsifier that comes from fatty substances found in plant tissues. It is derived from soybeans either mechanically or chemically, and is a by-product of soybean oil creation. The end product is a light brown powder that has low water solubility.

As an emulsifier, it works to blend immiscible ingredients together, such as oil and water, giving way to stable preparations. It can be whisked directly into the liquid of choice.

Soy lecithin is also used in creating foams, airs, mousses, and other aerated dishes that are long lasting and full of flavour. It is used in pastries, confections, and chocolate to enhance dough and increase moisture tolerance.

As with most ingredients, dosage and concentration for soy lecithin will depend on the ingredients used, specific properties desired in the resulting preparation, as well as other conditions.

Tapioca maltodextrin

Tapioca maltodextrin is a form of maltodextrin made from tapioca starch. It is a common ingredient in molecular gastronomy because it can be used both as a thickener and stabilizer for sauces and dressings, for encapsulation, and as a sweetener. In many cases it is also used as an aroma carrier due to its capacity to absorb oil. It is often used to make powders or pastes out of fat.

Xanthan gum

Xanthan gum is a food additive used as a thickening agent. It is produced through the fermentation of glucose. As a gluten-free additive it can be used as a substitute in cooking and baking.

As a thickener, when used in low dosages, xanthan gum produces a weak gel with high viscosity that is shear reversible with a high pourability. It also displays excellent stabilizing abilities that allow for particle suspension.

Moreover, xanthan gum mixes well with other flavours without masking them and provides an improved mouth feel to preparations. The presence of bubbles within the thickened liquids often makes way for light and creamy textures. It is used in the production of emulsions, suspensions, raviolis, and foams.

Being a hydrocolloid, xanthan gum must be hydrated before use. High versatility allows it to be dissolved over a wide range of temperatures, acid, and alcohol levels. Once set, xanthan gum may lose some of its effectiveness when exposed to heat.

This page titled 8.1.1.7: Non-traditional thickeners is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.7: Non-traditional thickeners by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.8: Sauces

Sauces enhance desserts by both their flavor and their appearance, just as savory sauces enhance meats, fish, and vegetables. Crème anglaise, chocolate sauce, caramel sauce, and the many fruit sauces and coulis are the most versatile. One or another of these sauces will complement nearly every dessert.

Examples of dessert sauces

Caramel sauce: A proper caramel flavor is a delicate balance between sweetness and bitterness. As sugar cooks and begins to change color, a flavor change will occur. The darker the sugar, the more bitter it will become. Depending on the application for the finished caramel, it can be made mild or strong. At this point, a liquid is added. This liquid will serve several roles: it will stop the cooking process, it can add richness and flavor, and it will soften the sauce. The fluidity of the finished sauce will depend on the amount of liquid added to it, and the temperature it is served at. Dairy products, such as cream, milk, or butter, will add richness; use water for a clear sauce; use fruit purées to add different flavor elements.

- **Chocolate sauce**: Sometimes called fudge sauce, chocolate sauce is generally made from cream (or milk), butter, and chocolate, and can be served hot or cold. The proportion of each of the ingredients will affect the thickness of the final product.
- **Compote**: French for “mixture,” a compote is cooked fruit served in its own cooking liquid, usually a sugar syrup. Compotes can be made with fresh, frozen, or dried fruits, and served hot or cold.
- **Coulis**: French for “strained liquid,” a coulis is most often an uncooked, strained purée. Flavors remain pure, and the colors bright. One of the drawbacks of using a coulis is that it may separate quickly when used as a plating sauce. It’s best to use à la minute.
- **Crème anglaise**: French for “English custard,” crème anglaise is a rich, pourable custard sauce that can be served hot or cold over cake, fruits, or other desserts. Made with eggs, sugar, and milk or cream, it is stirred over heat until it thickens into a light sauce. However, it’s a delicate operation: too much heat turns it into scrambled eggs! It should not get above 85°C (185°F) during the cooking process. Vanilla is the classic flavoring, but coffee, spices, chocolate, or liqueurs can be added. With additional yolks and heavy cream, it becomes the “custard” used for French ice cream. With additional yolks, gelatin, whipped cream, and flavoring, it becomes Bavarian cream.
- **Curd**: A curd is creamy and fruit based, with citrus and berry flavors being the most popular. Made from fruit juices, eggs, butter, and sugar cooked in a process similar to crème anglaise, curds can be thick, pourable sauces or spreads.
- **Fruit butter**: Fruit butter is a spread made from whole fruits, cooked, reduced, and puréed (if you don’t want any chunks in it) until very thick. It does not contain any butter; the term refers to the consistency.
- **Fruit sauce**: A fruit sauce is a fruit purée, cooked and thickened with a starch. It is normally served cold.
- **Hard sauce**: This traditional sauce for Christmas pudding, or any steamed pudding, is made by combining butter, sugar, and flavorings, often liqueurs. It is normally piped into shapes and chilled, then placed on the warm dessert just before serving.
- **Sabayon**: Sabayon is a mixture of egg yolks, flavoring, and sugar beaten over simmering water until thick, then beaten until cool. It is traditionally flavored with sweet white wine or liquor, then served over fresh fruit and grilled (when it is called a gratin). The Italian version of this is called azabaglione and isflavored with Madeira wine.
- **Whipped cream**: This very popular dessert topping can be served plain, sweetened, or flavoredC.rème chantilly, a classic version of this, is a combination of whipped cream, sugar, and vanilla.

Applying dessert sauces

Except in the case of some home-style or frozen desserts, sauces are usually not ladled over the dessert because doing so would mar the appearance. Instead, the sauce is applied in a decorative fashion to the plate rather than the dessert. Many different styles of plate saucing are available.

Pouring a pool of sauce onto the plate is knownfloaosding. Although plate flooding often looks old- fashioned today, it can still be a useful technique for many desserts. Flooded plates can be made more attractive by applying a contrasting sauce and then blending or feathering the two sauces decoratively with
a pick or the end of a knife. For this technique to work, the two sauces should be at about the same fluidity or consistency.

Rather than flooding the entire plate, it may be more appropriate for some desserts to apply a smaller pool of sauce to the plate,

as this avoids overwhelming the dessert with too much sauce.

A variation of the flooding technique is outlining, where a design is piped onto the plate with chocolate and allowed to set. The spaces can then be flooded with colorful sauces.

A squeeze bottle is useful for making dots, lines, curves, and streaks of sauce in many patterns. Or just a spoon is needed to drizzle random patterns of sauce onto a plate. Another technique for saucing is applying a small amount of sauce and streaking it with a brush, an offset spatula, or the back of a spoon.

Sauces are a great way to highlight flavors. Choose ones that will create balance on the plate, not just for color, but with all the components. A tart berry sauce will complement a rich cheesecake or chocolate dessert because sourness (acid) will cut through fat, making it taste lighter than it is. A sweet sauce served with a sweet dessert will have the overall effect of hiding flavors in both. Hold back on sweetness in order to intensify other flavors.

Many modern presentations may have a minimal amount of sauce. Sometimes this is done just for aesthetic reasons and not for how it will complement the dessert. Think of the dish and the balance of the components. This is the most important factor: flavor first, presentation second.

This page titled 8.1.1.8: Sauces is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.8: Sauces by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.1.9: Low-temperature and sous-vide

Sous-vide cooking is about immersing a food item in a precisely controlled water bath, where the temperature of the water is the same as the target temperature of the food being cooked. Food is placed in
a food-grade plastic bag and vacuum-sealed before going into the water bath. Temperatures will vary depending on desired end result. This allows the water in the bath to transfer heat into the food while preventing the water from coming into direct contact with it. This means the water does not chemically interact with the food: the flavors of the food remain stronger, because the water is unable to dissolve or carry away any compounds in the food (Figure 1).



Figure 1. “Img_0081” by Derek is licensed under CC BY- SA-ND 2.0

Sous-vide fruits and vegetables

Cooking vegetable and fruits sous-vide is a great way to tenderize them without losing as many of the vitamins and minerals that are normally lost through blanching or steaming. Fruits can also be infused with liquid when cooked at lower temperatures by adding liquid to the bag. Sous-vide helps preserve the nutrients present in fruits and vegetables by not cooking them above the temperatures that cause the cell walls to fully break down. This allows them to tenderize without losing all their structure. The bag also helps to catch any nutrients that do come out of the vegetable.

While time and temperature do not factor into safety for fruits and vegetables, they do have a unique effect on their structure. There are two components in fruits and vegetables that make them crisp: pectin and starch. Pectin, which is a gelling agent commonly used in jams and jellies for structure, breaks down at 83oC (183oF) at a slower rate than the starch cells do. In many cases this allows for more tender fruits and vegetables that have a unique texture to them.

Custards

The term custard spans so many possible ingredients and techniques that it is most useful to think of a custard as simply a particular texture and mouth feel. Custards have been made for centuries by lightly cooking a blend of eggs, milk, and heavy cream, but modernist chefs have invented myriad ways to make custards.

Using the sous-vide method to prepare crème anglaise, curds, ice cream bases, custard bases, sabayons, and dulce de leche is possible. The technique offers greater consistency and more control over the texture, which can range from airy, typical of a sabayon, to dense, as in a posset. For custards, eggs will be properly cooked at 82°C (180oF), so if the water bath is set to this temperature, no overcooking can happen. The one constant among custards is the use of plenty of fat, which not only provides that distinctive mouth feel but also makes custard an excellent carrier of fat-soluble flavors and aromas. Lighter varieties of custard, prepared sous-vide style and cooled, can be aerated in a whipping siphon into smooth, creamy foams.

Other applications for vacuum-seal processes

Fruit compression

Vacuum-compressing fruits and vegetables is a popular modern technique that can give many plant foods an attractive, translucent appearance (as shown in the watermelon in Figure 2) and a pleasant, surprising texture. This technique exploits the ability of a vacuum chamber to reduce surrounding pressure, which causes air and moisture within the plant tissue to rapidly expand and rupture the structures within the food. When the surrounding pressure is restored to a normal level, the labyrinth of air-filled spaces collapses. As a result, light tends to pass through the food rather than being scattered and diffused, which is why vacuum-compressed plant foods appear translucent. Causing the porous structure of a plant food to collapse also imparts a somewhat dense, toothsome texture that can give a familiar ingredient, such as watermelon, an entirely new appeal.



Figure 5. “WD-50 (7th Course)” by Peter Dillon is licensed under CC BY 2.0

Infusions

When adding liquids, the vacuum-seal process creates a rapid infusion—especially with more porous foods (such as adding spices to cream or herbs to melon). This can add flavor and texture in a shorter time than traditional infusions.

This page titled 8.1.1.9: Low-temperature and sous-vide is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 1.9: Low-temperature and sous-vide by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2: FLOUR

8.1.2.1: Introduction - Understanding Ingredients

8.1.2.2: The History of Wheat Flour

8.1.2.3: Milling of Wheat

8.1.2.4: Flour Streams and Types of Wheat Flour

8.1.2.5: Flour Terms and Treatments

8.1.2.6: Flour Additives

8.1.2.7: Whole Grain and Artisan Milling

8.1.2.8: Flour in Baking

8.1.2.9: Rye Flour

8.1.2.10: Other Grains and Flours

This page titled 8.1.2: Flour is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2: Flour by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.1: Introduction - Understanding Ingredients

Ingredients play an important role in baking. Not only do they provide the structure and flavour of all of the products produced in the bakery or pastry shop, their composition and how they react and behave in relation to each other are critical factors in understanding the science of baking. This is perhaps most evident when it comes to adapting formulas and recipes to accommodate additional or replacement ingredients while still seeking a similar outcome to the original recipe.

In this book, we look at each of the main categories of baking ingredients, listed below, and then explore their composition and role in the baking process. In addition to these categories, we will discuss the role that salt and water play in the baking process.

The main categories of baking ingredients are:

- Grains and flours Sweeteners
- Fats oils Leavening agents Eggs
- Dairy products
- Chocolate and other cocoa products Nuts and seeds
- Thickening agents
- Spices and other flavourings
- Fruit

Note: For most measurements used in the open textbook series, both S.I. (metric) and U.S./imperial values are given. The exception is nutritional information, which is always portrayed using metric values in both Canada and the United States.

This page titled 8.1.2.1: Introduction - Understanding Ingredients is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.1: Introduction - Understanding Ingredients by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.2: The History of Wheat Flour

Archaeologists who did excavations in the region of the lake dwellers of Switzerland found grains of wheat, millet, and rye 10,000 years old. The Romans perfected the rotary mill for turning wheat into flour. By the time of Christ, Rome had more than 300 bakeries, and Roman legions introduced wheat throughout their empire. Improved milling processes were needed because even when wheat was milled twice and bolted (sifted) through silk gauze, the result was still a yellowish flour of uneven texture and flecked with germ and bran.

In the second half of the 19th century, there were great changes in the flour milling process. An American inventor, Edmund LaCroix, improved the process with a purifier to separate the middlings (bran, germ,and other coarse particles) from the particles that form smooth-textured white flour. In recent years, the demand for whole grain milling has increased because whole grain food products have proved to be more nutritious than products made from white flour. (More information on whole grain and artisan milling is provided later in this section.)

In Canada, large-scale wheat growing didn't occur until after the Prairies were settled in the 1800s. Hard wheat, such as Red Fife, Marquis, and Selkirk, earned Canada a position as the granary for Britain and many other European countries. Today, most of the wheat grown in Western Canada is the hard Red Spring variety. Soft wheats, such as soft red and soft white, are primarily grown in Quebec and Ontario. Many of the original wheat growers have passed on their farms to the next generations, while others branched out to organic farming and milling. One of these farms, Nunweiler's, has a heritage that goes back to the early 1900s when the original wheat in Canada, Red Fife and Marquis, was grown on this farm.

Today, the major wheat growing areas of North America are in the central part of the continent, in the Great Plains of the United States and the Canadian Prairies. From Nebraska south, winter wheat can be grown, while to the north through Saskatchewan spring wheat dominates. Many American states and some Canadian provinces grow both kinds. In fact, there are very few states that don't grow some wheat. Kansas, the site of the American Institute of Baking, could be said to be at the heart of the U.S. wheat growing area, while Saskatchewan is the Canadian counterpart.

This page titled 8.1.2.2: The History of Wheat Flour is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.2: The History of Wheat Flour by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.3: Milling of Wheat

Milling of wheat is the process that turns whole grains into flours. The overall aims of the miller are to produce:

- A consistent product
- A range of flours suitable for a variety of functions
- Flours with predictable performance

The very first mill operation is analyzing the grain, which determines criteria such as thgeluten content and amylase activity. It is at this point that decisions about blending are made.

Following analysis, milling may be divided into three stages:

- Cleaning and conditioning – ridding the grain of all impurities and readying it for milling
- Crushing or breaking – breaking down the grain in successive stages to release its component parts Reduction – progressive rollings and siftings to refine the flour and separate it into various
- categories, called streams

Cleaning

Wheat received at the mill contains weeds, seeds, chaff, and other foreign material. Strong drafts of air from the aspirator remove lighter impurities. The **disc separator** removes barley, oats, and other foreign materials. From there, the wheat goes to the scourers in which it is driven vigorously against perforated steel casings by metal beaters. In this way, much of the dirt lodged in the crease of the wheat berry is removed and carried away by a strong blast of air. Then the magnetic separator removes any iron or steel.

At this point, the wheat is moistened. Machines known as whizzers take off the surface moisture. The wheat is then **tempered**, or allowed to lie in bins for a short time while still damp, to toughen the bran coat, thus making possible a complete separation of the bran from the flour-producing portion of the wheat berry. After tempering, the wheat is warmed to a uniform temperature before the crushing process starts.

Crushing or Breaking

The objectives at this stage are twofold:

- Separate as much bran and germ as possible from the endosperm
- Maximize the flour from the resulting endosperm

Household grain mills create flour in one step — grain in one end, flour out the other — but the commercial mill breaks the grain down in a succession of very gradual steps, ensuring that little bran and germ are mixed with any endosperm.

Although the process is referred to ascrushing, flour mills crack rather than crush the wheat with large steel rollers. The rollers at the beginning of the milling system are corrugated and break the wheat into coarse particles. The grain passes through screens of increasing fineness. Air currents draw off impurities from the middlings. Middlings is the name given to coarse fragments of endosperm, somewhere between the size of semolina and flour. Middlings occur after the “break” of the grain.

Bran and germ are sifted out, and the coarse particles are rolled, sifted, and purified again. This separation of germ and bran from the endosperm is an important goal of the miller. It is done to improve dough-making characteristics and colour. As well, the germ contains oil and can affect keeping qualities of the flour.

Reduction

In the reduction stage, the coarser particles go through a series of fine rollers and sieves. After the first crushing, the wheat is separated into five or six streams. This is accomplished by means of machines called **plansifters** that contain sieves, stacked

vertically, with meshes of various sizes. The finest mesh is as fine as the finished flour, and some flour is created at an early stage of reduction.

Next, each of the divisions or streams passes through cleaning machines, known apsurifiers, a series of sieves arranged horizontally and slightly angled. An upcurrent draught of air assists in eliminating dust. The product is crushed a little more, and each of the resulting streams is again divided into numerous portions by means of sifting. The final crushings are made by perfectly smooth steel rollers that reduce the middlings into flour. The flour is then bleached and put into bulk storage. From bulk storage, the flour is **enriched** (thiamine, niacin, riboflavin, and iron are added), and either bagged for home and bakery use or made ready for bulk delivery.

Extraction Rates

The extraction rate is a figure representing the percentage of flour produced from a given quantity of grain. For example, if 82 kg of flour is produced from 100 kg of grain, the extraction rate is 82% (82÷100×100). Extraction rates vary depending on the type of flour produced. A whole grain flour, which contains all of the germ, bran, and endosperm, can have an extraction rate of close to 100%, while white all-purpose flours generally have extraction rates of around 70%. Since many of the nutrients are found in the germ and bran, flours with a higher extraction rate have a higher nutritional value.

This page titled 8.1.2.3: Milling of Wheat is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.3: Milling of Wheat by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.4: Flour Streams and Types of Wheat Flour

Modern milling procedures produce many different **flour streams** (approximately 25) that vary in quality and chemical analysis. These are combined into four basic streams of edible flour, with four other streams going to feed.

- Top patent flour: This stream is composed of only the purest and most highly refined streams from the mill. It is low in ash and is approximately 50% of the flour extracted. The term ash indicates the mineral content (e.g., phosphorus) of the flour. When flour is burned, all that is left is the burned mineral elements that constitute ash.
- Second patent flour: This flour is composed of streams with an intermediate degree of refinement. It has an average ash content of approximately 0.45% and represents about 35% of the total flour. First clear flour: This stream contains the balance of the flour that possesses baking properties, and is high in ash and protein content. It is usually about 15% of the total flour.
- Second clear flour: This grade contains the poorest flour streams. It is very high in ash (approximately 0.75%), and has little or no baking quality. It is about 2% of the total flour.
- Feed streams: The balance of the streams from the mill are classed as feed. Feeds are marketed as bran, wheat shorts, flour middlings, and wheat germ.

Within the streams of edible flours, there are a number of different types of flour used in food preparation. Each has different characteristics, and with those come different uses, as described below.

All-Purpose Flour

General purpose or home use flours are usually a blend of hard spring wheats that are lower in protein (gluten) content than bread flours. They are top patent flours and contain sufficient protein to make good yeast breads, yet not too much for good quick breads, cakes, and cookies.

Note: A word about gluten quality as opposed to gluten quantity: The fact that a particular flour contains a high quantity of protein, say 13% to 15%, does not necessarily mean that it is of high quality. It may contain too much ash or too much damaged starch to warrant this classification. High quality is more important in many bread applications than high quantity. All-purpose flour is an example of a high-quality flour, with a protein content of about 12%.

Graham Flour

A U.S. patented flour, graham flour is a combination of whole wheat flour (slightly coarser), with added bran and other constituents of the wheat kernel.

Bread Flour

Bread flour is milled from blends of hard spring and hard winter wheats. They average about 13% protein and are slightly granular to the touch. This type of flour is sold chiefly to bakers because it makes excellent bread with bakery equipment, but has too much protein for home use. It is also called strong flour or hard flour and is second patent flour.

For example, the specification sheet on bread flour produced by a Canadian miller might include the following information:

- Ingredients: Wheat flour, amylase, ascorbic acid, niacin, iron, thiamine mononitrate, riboflavin, azodicarbonamide, folic acid.
- Moisture: 14.2%
- Ash: 0.54%
- Protein (5.7 x N) 13.00%

Along with this information there is microbiological data and an allergen declaration. (Note that the formula in parentheses beside “Protein” is simply the laboratory’s way of deriving the protein figure from the nitrogen content.)

Cake Flour

Cake flour is milled from soft winter wheats. The protein content is about 7% and the granulation is so uniform and fine that the flour feels satiny. An exception is a high-protein cake flour formulated especially for fruited pound cakes (to prevent the fruit from sinking).

Clear Flour

Clear flour comes from the part of the wheat berry just under the outer covering. Comparing it to first patent flour is like comparing cream to skim milk. It is dark in colour and has a very high gluten content. It
is used in rye and other breads requiring extra strength.

Gluten Flour

Gluten flour is made from wheat flour by removing a large part of the starch. It contains no more than 10% moisture and no more than 44% starch.

Pastry Flour

Pastry flour is made from either hard or soft wheat, but more often from soft. It is fairly low in protein and is finely milled, but not so fine as cake flour. It is unsuitable for yeast breads but ideal for cakes, pastries, cookies, and quick breads.

Self-Rising Flour

Self-rising flour has leavening and salt added to it in controlled amounts at the mill.

Wheat Germ Flour

Wheat germ flour consists entirely of the little germ or embryo part of the wheat separated from the rest of the kernel and flattened into flakes. This flour should be refrigerated.

Whole Wheat Flour

Whole wheat flour contains all the natural parts of the wheat kernel up to 95% of the total weight of the wheat. It contains more protein than all-purpose flour and produces heavier products because of the bran particles.

Whole Wheat Pastry Flour

Whole wheat pastry flour is milled from the entire kernel of soft wheat, is low in gluten, and is suitable for pastry, cakes, and cookies.

Hovis Flour

Most of the germ goes away with the shorts and only a small fraction of the total quantity can be recovered in a fairly pure form. At the mill, a special process developed in England to improve its keeping qualities and flavour cooks this fraction. It is then combined with white flour to make Hovis flour, which produces a loaf that, though small for its weight, has a rich, distinctive flavour.

Triticale Flour

The world's first new grain, triticale is a hybrid of wheat and rye. It combines the best qualities of both grains. It is now grown commercially in Manitoba.

Semolina

Semolina is the granular product consisting of small fragments of the endosperm of the durum wheat kernel. (The equivalent particles from other hard wheat are called farina.) The commonest form of semolina available commercially is the breakfast cereal Cream of Wheat.

No-Time Flour

The primary goal of all bakers has been to reduce production time and keep costs to a minimum without losing quality, flavour, or structure. After extensive research, millers have succeeded in eliminating bulk fermentation for both sponge and straight dough methods. No-time flour is flour with additives such as ascorbic acid, bromate, and cysteine. It saves the baker time and labour, and reduces floor spac requirements. The baker can use his or her own formulas with only minor adjustments.

Blending Flours

Blending of flours is done at the mill, and such is the sophistication of the analysis and testing of flours (test baking, etc.) that when problems occur it is generally the fault of the baker and not the product. Today the millers and their chemists ensure that bakers receive the high grade of flour that they need to produce marketable products for a quality-conscious consumer. Due to the vagaries of the weather and its effect on growing conditions, the quality of the grain that comes into the mill is hardly ever constant. For example, if damp weather occurs at harvest time, the grain may start to sprout and will cause what is known as damaged starch. Through analysis and adjustments in grain handling and blending, the miller is able to furnish a fairly constant product.

Bakers do blend flours, however. A portion of soft flour may be blended with the bread flour to reduce the toughness of a Danish pastry or sweet dough, for example. Gluten flour is commonly used in multigrain bread to boost the aeration.

This page titled 8.1.2.4: Flour Streams and Types of Wheat Flour is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.4: Flour Streams and Types of Wheat Flour by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.5: Flour Terms and Treatments

In addition to types of flour, you may come across various other terms when purchasing flour. These include some terms that refer to the processing and treatment of the flour, and others outlining some of the additives that may be added during the milling and refining process.

Bleached

Bleaching and maturing agents are added to whiten and improve the baking quality quickly, making it possible to market the freshest flour. Even fine wheat flours vary in colour from yellow to cream when freshly milled. At this stage, the flour produces doughs that are usually sticky and do not handle well. Flour improves with age under proper storage conditions up to one year, both in color and quality.

Because storing flour is expensive, toward the close of the 19th century, millers began to treat freshly milled flour with oxidizing agents to bleach it and give it the handling characteristics of naturally aged flour. Under the category of maturing agents are included materials such as chlorine dioxide, chlorine gas plus a small amount of nitrosyl chloride, ammonium persulfate, and ascorbic acid. No change occurs in the nutritional value of the flour when these agents are present.

There are two classes of material used to bleach flour. A common one, an organic peroxide, reacts with the yellow pigment only, and has no effect on gluten quality. Chlorine dioxide, the most widely used agent in North America, neutralizes the yellow pigment and improves the gluten quality. It does, however, destroy the tocopherols (vitamin E complex).

Enriched

Iron and three of the most necessary B vitamins (thiamin, riboflavin, and niacin), which are partially removed during milling, are returned to white flour by a process known as enrichment. No change occurs in taste, colour, texture, baking quality, or caloric value of the flour.

Pre-sifted

During the milling process, flour is sifted many times through micro-fine silk. This procedure is known as pre-sifting. The mesh size used for sifting varies from flour to flour. There are more holes per square inch for cake flour than, for example, bread flour, so that a cup of cake flour has significantly more minute particles than does a cup of bread flour, is liable to be denser, and weigh slightly more. Sifted flour yields more volume in baked bread than does unsifted flour, simply because of the increased volume of air.

This page titled 8.1.2.5: Flour Terms and Treatments is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.5: Flour Terms and Treatments by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.6: Flour Additives

A number of additives may be found in commercial flours, from agents used as dough conditioners, to others that aid in the fermentation process. Why use so many additives? Many of these products are complementary – that is, they work more effectively together and the end product is as close to “ideal” as possible. Nevertheless, in some countries the number of additives allowed in flour are limited. For instance, in Germany, ascorbic acid remains the only permitted additive. Some of the additives that are commonly added to flour include those described below.

Bromate

Until the early 1990s, bromate was added to flour because it greatly sped up the oxidation or aging of flour. Millers in Canada stopped using it after health concerns raised by the U.S. Food and Drug Administration (FDA). In the United States, bromate is allowed in some states but banned in others (e.g., California).

Azodicarbonamide (ADA)

Approved in the United States since 1962, but banned in Europe, ADA falls under the food additives permitted in Canada. ADA is a fast-acting flour treatment resulting in a cohesive, dry dough that tolerates high water absorption. It is not a bleach, but because it helps produce bread with a finer texture it gives an apparently whiter crumb. It does not destroy any vitamins in the dough. Bakers who want to know if their flours contain ADA or other chemical additives can request the information from their flour suppliers.

L-Cysteine

An amino acid, L-cysteine speeds up reactions within the dough, thus reducing or almost eliminating bulk fermentation time. In effect, it gives the baker a “no-time” dough. It improves dough elasticity and gas retention.

Ascorbic Acid

Ascorbic acid was first used as a bread improver in 1932, after it was noticed that old lemon juice added to dough gave better results because it improved gas retention and loaf volume. Essentially vitamin C (ascorbic acid) has the advantage of being safe even if too much is added to the dough, as the heat of baking destroys the vitamin component. The addition of ascorbic acid consistent with artisan bread requirements is now routine for certain flours milled in North America.

Calcium Peroxide

Calcium peroxide (not to be confused with the peroxide used for bleaching flour) is another dough-maturing agent.

Glycerides

Glycerides are multi-purpose additives used in both cake mixes and yeast doughs. They are also known as surfactants, which is a contraction for “surface-acting agents.” In bread doughs, the main function of glycerides is as a crumb-softening agent, thus retarding bread staling. Glycerides also have some dough strengthening properties.

Sodium Stearoyl Lactylate

Approved for use in the United States since 1961, this additive improves gas retention, shortens proofing time, increases loaf volume, and works as an anti-staling agent.

This page titled 8.1.2.6: Flour Additives is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.6: Flour Additives by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.7: Whole Grain and Artisan Milling

Whole grain and artisan milling is the type of milling that was practiced before the consumer market demanded smooth white flours that are refined and have chemical additives to expedite aging of flours. Artisan milling produces flours that are less refined and better suited to traditional breads, but also contain little to no additives and have higher nutritional content. For that reason, demand for these types of flour is on the rise.

Artisan millers (also known as micro millers) process many non-stream grains, including spelt, kamut, buckwheat, and other non-gluten grains and pulses. This offers bakers opportunities to work with different grains and expand their businesses. Artisan flours are readily available directly from millers or through a distributor. Knowing the origin of the grains and the quality of the ingredients in baking is important for artisan bakers.

Whole grain flours are on the increase as consumers become more aware of their benefits. Whole grain flour, as the name suggests, is made from whole grains.

Many artisan millers purchase their grains directly from growers. This method of purchasing establishes trustworthy working relationships with the grain growers and promotes transparency in grain growing and food safety practices. Grain growers that sell their grains to artisan millers apply conventional or organic growing practices. Grain growers and millers have to go through vigorous processes to obtain the **certified organic** certification for their grains or products, which guarantees that no chemical additives have been used.

How organic grain is processed varies. Stone milling and impact hammer milling methods are typical when minimal refined whole grain flour is preferred. Information on several American artisan millers that produce various whole grain flours can be found Faitrebid Mills; Hayden Flour Mills; and Baker Miller Chicago. Organic flours have gained popularity in the baking industry. As consumers become more aware of them, we see the demand swinging back toward whole grain and artisan milling as a preference.

This page titled 8.1.2.7: Whole Grain and Artisan Milling is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.7: Whole Grain and Artisan Milling by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.8: Flour in Baking

Flour forms the foundation for bread, cakes, and pastries. It may be described as the skeleton, which supports the other ingredients in a baked product. This applies to both yeast and chemically leavened products.

The strength of flour is represented in protein (gluten) quality and quantity. This varies greatly from flour to flour. The quality of the protein indicates the strength and stability of the flour, and the result in bread making depends on the method used to develop the gluten by proper handling during the fermentation. Gluten is a rubber-like substance that is formed by mixing flour with water. Before it is mixed it contains two proteins. In wheat, these two proteins are **gliadin** and **glutenin**. Although we use the terms protein and gluten interchangeably, gluten only develops once the flour is moistened and mixed. The protein in the flour becomes gluten.

Hard spring wheat flours are considered the best for bread making as they have a larger percentage of good quality gluten than soft wheat flours. It is not an uncommon practice for mills to blend hard spring wheat with hard winter wheat for the purpose of producing flour that combines the qualities of both. Good bread flour should have about 13% gluten.

Storing Flour

Flour should be kept in a dry, well-ventilated storeroom at a fairly uniform temperature. A temperature of about 21°C (70°F) with a relative humidity of 60% is considered ideal. Flour should never be stored in a damp place. Moist storerooms with temperatures greater than 23°C (74°F) are conducive to mould growth, bacterial development, and rapid deterioration of the flour. A well-ventilated storage room is necessary because flour absorbs and retains odors. For this reason, flour should not be stored in the same place as onions, garlic, coffee, or cheese, all of which give off strong odors.

Flour Tests

Wheat that is milled and blended with modern milling methods produce flours that have a fairly uniform quality all year round and, if purchased from a reliable mill, they should not require any testing for quality. The teacher, student, and professional baker, however, should be familiar with qualitative differences in flours and should know the most common testing methods.

Flours are mainly tested for:

- Color
- Absorption
- Gluten strength
- Baking quality

Other tests, done in a laboratory, are done for:

- Albumen
- Starch
- Sugar
- Dextrin
- Mineral and fat content

Color

The color of the flour has a direct bearing on baked bread, providing that fermentation has been carried out properly. The addition of other ingredients to the dough, such as **brown sugar**, malt, molasses, salt, and colored margarine, also affects the color of bread.

To test the color of the flour, place a small quantity on a smooth glass, and with a spatula, work until a firm smooth mass about 5 cm (2 in.) square is formed. The thickness should be about 2 cm (4/5 in.) at the back
of the plate to a thin film at the front. The test should be made in comparison with a flour of known grade and quality, both flours

being worked side by side on the same glass. A creamy white color indicates a hard flour of good gluten quality. A dark or greyish color indicates a poor grade of flour or the presence of dirt. Bran specks indicate a low grade of flour.

After making a color comparison of the dry samples, dip the glass on an angle into clean water and allow to partially dry. Variations in color and the presence of bran specks are more easily identified in the damp samples.

Absorption

Flours are tested for absorption because different flours absorb different amounts of water and therefore make doughs of different consistencies. The absorption ability of a flour is usually between 55% and 65%. To determine the absorption factor, place a small quantity of flour (100 g/4 oz.) in a bowl. Add water gradually from a beaker containing a known amount of water. As the water is added, mix with a spoon until the dough reaches the desired consistency. You can knead the dough by hand for final mixing and determination of consistency. Weigh the unused water. Divide the weight of the water used by the weight of the flour used. The result is the absorption ability in percentage. For example:

- Weight of flour used 100 g (4 oz.)
- Weight of water used 60 g (2.7 oz.) T
- therefore absorption = 6/10 or 60%

Prolonged storage in a dry place results in a natural moisture loss in flour and has a noticeable effect on the dough. For example, a sack of flour that originally weighed 40 kg (88 lb.) with a moisture content of 14% may be reduced to 39 kg (86 lb.) during storage. This means that 1 kg (2 lb.) of water is lost and must be made up when mixing. The moisture content of the wheat used to make the flour is also important from an economic standpoint.

Hard wheat flour absorbs more liquid than soft flour. Good hard wheat flour should feel somewhat granular when rubbed between the thumb and fingers. A soft, smooth feeling indicates a soft wheat flour or a blend of soft and hard wheat flour. Another indicator is that hard wheat flour retains its form when pressed in the hollow of the hand and falls apart readily when touched. Soft wheat flour tends to remain lumped together after pressure.

Gluten Strength

The gluten test is done to find the variation of gluten quality and quantity in different kinds of flour. Hard flour has more gluten of better quality than soft flour. The gluten strength and quality of two different kinds of hard flour may also vary with the weather conditions and the place where the wheat is grown. The difference may be measured exactly by laboratory tests, or roughly assessed by the variation of gluten balls made from different kinds of hard flours.

For example, to test the gluten in hard flour and all-purpose flour, mix 250 g (9 oz.) of each in separate mixing bowls with enough water to make each dough stiff. Mix and develop each dough until smooth. Let the dough rest for about 10 minutes. Wash each dough separately while kneading it under a stream of cold water until the water runs clean and all the starch is washed out. (Keep a flour sieve in the sink to prevent dough pieces from being washed down the drain.) What remains will be crude gluten. Shape the crude gluten into round balls, then place them on a paper-lined baking pan and bake at 215°C (420°F) for about one hour. The gluten ball made from the hard flour will be larger than the one made from all-purpose flour. This illustrates the ability of hard flour to produce a greater volume because of its higher gluten content.

Ash Content

Ash or mineral content of flour is used as another measurement of quality. Earlier in the chapter, we talked about extraction rates as an indicator of how much of the grain has been refined. Ash content refers to the amount of ash that would be left over if you were to burn 100 g of flour. A higher ash content indicates that the flour contains more of the germ, bran, and outer endosperm. Lower ash content means that the flour is more highly refined (i.e., a lower extraction rate).

Baking Quality

The final and conclusive test of any flour is the kind of bread that can be made from it. The baking test enables the baker to check on the completed loaf that can be expected from any given flour. Good volume is related to good quality gluten; poor volume to

young or green flour. Flour that lacks stability or power to hold during the entire fermentation may result in small, flat bread. Flour of this type may sometimes respond to an increase in the amount of yeast. More yeast shortens the fermentation time and keeps the dough in better condition during the pan fermentation period.

This page titled 8.1.2.8: Flour in Baking is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.8: Flour in Baking by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.9: Rye Flour

Rye is a hardy cereal grass cultivated for its grain. Its use by humans can be traced back over 2,000 years. Once a staple food in Scandinavia and Eastern Europe, rye declined in popularity as wheat became more available through world trade. A crop well suited to northern climates, rye is grown on the Canadian Prairies and in the northern states such as the Dakotas and Wisconsin.

Rye flour is the only flour other than wheat that can be used without blending (with wheat flour) to make yeast-raised breads. Nutritionally, it is a grain comparable in value to wheat. In some cases, for example, its lysine content (an amino acid), is even biologically superior.

The brown grain is cleaned, tempered, and milled much like wheat grain. One difference is that the rye endosperm is soft and breaks down into flour much quicker that wheat. As a result, it does not yield semolina, so purifiers are seldom used. The bran is separated from the flour by the break roller, and the flour is further rolled and sifted while being graded into chop, meal, light flour, medium flour, and dark flour:.

- Chop: This is the miller's name for the coarse stock after grinding in a break roller mill.
- Meal: Like chop, meal is made of 100% extraction obtained by grinding the entire rye kernel.
- Light rye flour: This is obtained from the centre of the rye kernel and is low in protein and high in starch content. It can be compared to white bread flour and is used to make light rye breads. Medium rye flour: This is straight flour and consists of all the kernels after the bran and shorts have been removed. It is light grey in colour, has an ash content of 1%, and is used for a variety of sourdough breads.
- Dark rye flour: This is comparable to first clear wheat flour. It has an ash content of 2% and a protein content of 16%. It is used primarily for heavier types of rye bread.

The lighter rye flours are generally bleached, usually with a chlorine treatment. The purpose of bleaching is to lighten the colour, since there is no improvement on the gluten capability of the flour.

Extraction of Rye Flour

The grade of extraction of rye flour is of great importance to the yield of the dough and the creation of a particular flavour in the baked bread. Table 1 shows the percentage of the dry substances of rye flour by grade of extraction.

page146image23744

Table 1 Table of extraction for rye flour

page146image24224 page146image24384
page146image25144

Substance

Grade of Extraction

page146image26704

70%

85%

page146image29544 page146image29704 page146image30464 page146image30
page146image31208

Ash

0.8%

1.4%

page146image33144 page146image33304 page146image34064 page146image34808

Fat

1.2%

1.7%

page146image36744 page146image36904 page146image37664 page146image38408

Protein

8.1%

9.6%

page146image40344 page146image41440 page146image41600 page146image41920

Sugar

6.5%

7.5%

page146image44352 page146image44512 page146image44672

Starch	72.5%	65.1%
Crude fibre	0.5%	1.3%
Pentosans	5.2%	7.6%
Undefinable	5.2%	5.8%

Note that ash, fibre, and **pentosans** are higher in the 85% extraction rate flour, and starch is lower. Pentosans are gummy carbohydrates that tend to swell when moistened and, in baking, help to give the rye loaf its cohesiveness and structure. The pentosan level in rye flour is greater than that of wheat flour and is of more significance for successful rye bread baking.

Rye flours differ from wheat flours in the type of gluten that they contain. Although some dark rye flours can have a gluten content as high as 16%, this is only gliadin. The glutenin, which forms the elasticity in dough is absent, and therefore doughs made only with rye flour will not hold the gas produced by the yeast during fermentation. This results in a small and compact loaf of bread.

Starch and pentosans are far more important to the quality of the dough yield than gluten. Starch is the chief component of the flour responsible for the structure of the loaf. Its bread-making ability hinges on the age of the flour and the acidity. While rye flour does not have to be aged as much as wheat flour, it has both a “best after” and a “best before” date. Three weeks after milling is considered to be good.

When the rye flour is freshly milled, the starch gelatinizes (sets) quickly at a temperature at which amylases are still very active. As a result, bread made from fresh flour may be sticky and very moist. At the other extreme, as the starch gets older, it gelatinizes less readily, the enzymes cannot do their work, and the loaf may split and crack. A certain amount of starch breakdown must occur for the dough to be able to swell.

The moisture content of rye flour should be between 13% and 14%. The less water in the flour, the better its storage ability. Rye should be stored under similar conditions to wheat flour.

Differences between Rye and Wheat

Here is a short list of the differences between rye and wheat:

- Rye is more easily pulverized.
- Rye does not yield semolina.
- Gluten content in rye is not a significant dough-making factor.
- Starch is more important for bread making in rye flour than in wheat flour.
- The pentosan level in rye flour is higher and more important for bread making.
- Rye flour has greater water binding capability than wheat flour, due to its starch and pentosan content.

In summary, both wheat and rye have a long history in providing the “staff of life.” They are both highly nutritious. North American mills have state-of-the-art technology that compensates for crop differences, thus ensuring that the baker has a reliable and predictable raw material. Flour comes in a great variety of types, specially formulated so that the baker can choose according to product and customer taste.

This page titled 8.1.2.9: Rye Flour is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.9: Rye Flour by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.2.10: Other Grains and Flours

Several other types of grains are commonly used in baking. In particular, corn and oats feature predominantly in certain types of baking (quick breads and cookies respectively, for instance) but increasingly rice flour is being used in baked goods, particularly for people with gluten sensitivities or intolerances. The trend to whole grains and the influence of different ethnic cultures has also meant the increase in the use of other grains and pulses for flours used in breads and baking in general.

Corn

Corn is one of the most widely used grains in the world, and not only for baking. Corn in used in breads and cereals, but also to produce sugars (such as dextrose and corn syrup), starch, plastics, adhesives, fuel (ethanol), and alcohol (bourbon and other whisky). It is produced from the maize plant (the preferred scientific and formal name of the plant that we callcorn in North America). There are different varieties of corn, some of which are soft and sweet (corn you use for eating fresh or for cooking) and some of which are starchy and are generally dried to use for baking, animal feed, and popcorn.

Varieties Used in Baking

- Cornmeal has a sandy texture and is ground to fine, medium, and coarse consistencies. It has most of the husk and germ removed, and is used is recipes from the American South (e.g., cornbread) and can be used to add texture to other types of breads and pastry.
- Stone-ground cornmeal has a texture not unlike whole wheat flour, as it contains some of the husk and germ. Stone ground cornmeal has more nutrients, but it is also more perishable. In baking, it acts more like cake flour due to the lack of gluten.
- Corn flour in North America is very finely ground cornmeal that has had the husk and germ removed. It has a very soft powdery texture. In the U.K. and Australia, corn flour refers to cornstarch. Cornstarch is the starch extracted from the maize kernel. It is primarily used as a thickener in baking and other cooking. Cornstarch has a very fine powdery consistency, and can be dissolved easily in water. As a thickening agent, it requires heat to set, and will produce products with a shiny, clear consistency.
- Blue cornmeal has a light blue or violet colour and is produced from whole kernels of blue corn. It is most similar to stone-ground cornmeal and has a slightly sweet flavour.

Rice

Rice is another of the world's most widely used cereal crops and forms the staple for much of the world's diet. Because rice is not grown in Canada, it is not regulated by the Canadian Grain Commission.

Varieties Used in Baking

- Rice flour is prepared from finely ground rice that has had the husks removed. It has a fine, slightly sandy texture, and provides crispness while remaining tender due to its lack of gluten. For this reason, many gluten-free breads are based on rice flours or blends that contain rice flour.
- Short grain or pearl rice is also used in the pastry shop to produce rice pudding and other desserts.

Oats

Oats are widely used for animal feed and food production, as well as for making breads, cookies, and dessert toppings. Oats add texture to baked goods and desserts.

Varieties Used in Baking

- Bakers will most often encounter rolled oats, which are produced by pressing the de-husked whole kernels through rollers.
- Oat bran and oat flour are produced by grinding the oat kernels and separating out the bran and endosperm.

- Whole grain oat flour is produced by grinding the whole kernel but leaving the ground flour intact. Steel-cut oats are more commonly used in cooking and making breakfast cereals, and are the chopped oat kernels.

Other Grains and Pulses

A wide range of additional flours and grains that are used in ethnic cooking and baking are becoming more and more widely available in Canada. These may be produced from grains (such as kamut, spelt, and quinoa), pulses (such as lentils and chickpeas), and other crops (such as buckwheat) that have a grain-like consistency when dried. Increasingly, with allergies and intolerances on the rise, these flours are being used in bakeshops as alternatives to wheat-based products for customers with special dietary needs.

This page titled 8.1.2.10: Other Grains and Flours is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 2.10: Other Grains and Flours by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.3: FAT

Add all sections in this chapter

8.1.3.1: Understanding Fats and Oils

8.1.3.2: Sources of Bakery Fats and Oils

8.1.3.3: Major Fats and Oils Used in Bakeries

8.1.3.4: Functions of Fat in Baking

This page titled 8.1.3: Fat is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 3: Fat by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.3.1: Understanding Fats and Oils

Fats and oils are organic compounds that, like carbohydrates, are composed of the elements carbon (C), hydrogen (H), and oxygen (O), arranged to form molecules. There are many types of fats and oils and a number of terms and concepts associated with them, which are detailed further here.

Lipids

In baking, lipids are generally a synonym for fats. Baking books may talk about the “lipid content of eggs,” for example.

Triglycerides

Triglycerides is another chemical name for the most common type of fats found in the body, indicating that they are usually made up of three (tri) fatty acids and one molecule of glycerol (glycerine is another name) as shown in Figure 3. (The mono and diglycerides that are used as emulsifiers have one and two fatty acids respectively.)

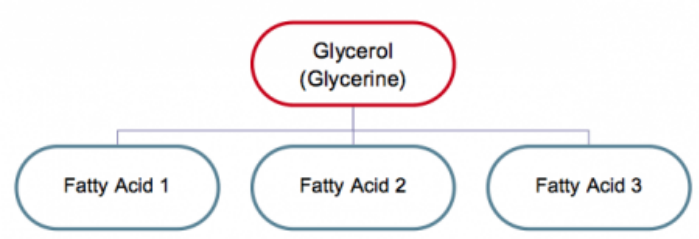


Figure 1 Composition of fats (triglycerides)

Fatty Acids

Each kind of fat or oil has a different combination of fatty acids. The nature of the fatty acid will determine the consistency of the fat or oil. For example, stearic acid is the major fatty acid in beef fat, and linoleic acid is dominant in seed oils. Fatty acids are defined as short, medium, or long chain, depending on the number of atoms in the molecule. The reason that some fat melts gradually is that as the temperature rises, each fatty acid will, in turn, soften, as its melting point is reached. Fats that melt all of a sudden mean that the fatty acids are of the same or similar type and have melting points within a narrow range. An example of such a fat is coconut fat: one second it is solid, the next, liquid.

Table 1 shows the characteristics of three fatty acids.

Table 1: Characteristics of Fatty Acids Type of Fatty Acid Melting Point Physical State (at room temperature) Stearic 69°C (157°F) Solid Oleic 16°C (61°F) Liquid Linoleic -12°C (9°F) Liquid

Rancid

Rancid is a term used to indicate that fat has spoiled. The fat takes on an unpleasant flavor when exposed to air and heat. Unsalted butter, for example, will go rancid quickly if left outside the refrigerator, especially in warm climates.

Oxidation/Antioxidants

Oxidation (exposure to air) causes rancidity in fats over time. This is made worse by combination with certain metals, such as copper. This is why doughnuts are never fried in copper pans!

Some oils contain natural antioxidants, such as tocopherols (vitamin E is one kind), but these are often destroyed during the

processing. As a result, manufacturers add synthetic antioxidants to retard rancidity. BHA and BHT are synthetic antioxidants commonly used by fat manufacturers.

Saturated/Unsaturated

Saturated and unsaturated refer to the extent to which the carbon atoms in the molecule of fatty acid are linked or bonded (saturated) to hydrogen atoms. One system of fatty acid classification is based on the number of double bonds.

- 0 double bonds: saturated fatty acids. Stearic acid is a typical long-chain saturated fatty acid (Figure 2).[1]

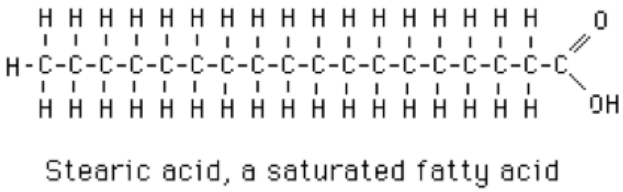


Figure 2 *Stearic Acid*

- 1 double bond: monounsaturated fatty acids. Oleic acid is a typical monounsaturated fatty acid (Figure 3).[2]

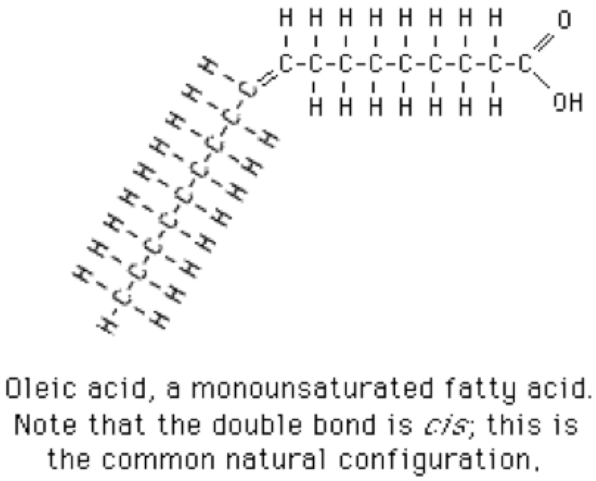
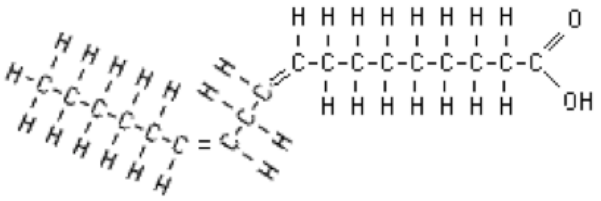


Figure 3 *Oleic Acid*

- 2 or more double bonds: polyunsaturated fatty acids. Linoleic acid is a typical polyunsaturated fatty acid (Figure 4).[3]



Linoleic acid, a polyunsaturated fatty acid.
Both double bonds are *cis*.

Figure 4 *Linoleic Acid*

Saturated fat is a type of fat found in food. For many years, there has been a concern that saturated fats may lead to an increased risk of heart disease; however, there have been studies to the contrary and the literature is far from conclusive. The general assumption is that the less saturated fat the better as far as health is concerned. For the fat manufacturer, however, low saturated fat levels make it difficult to produce oils that will stand up to the high temperatures necessary for processes such as deep-frying. Hydrogenation has been technology's solution. Hydrogenation will be discussed later in the chapter.

Saturated fat is found in many foods:

- Animal foods (like beef, chicken, lamb, pork, and veal)
- Coconut, palm, and palm kernel oils
- Dairy products (like butter, cheese, and whole milk)
- Lard
- Shortening

Unsaturated fat is also in the foods you eat. Replacing saturated and trans fats (see below) with unsaturated fats has been shown to help lower cholesterol levels and may reduce the risk of heart disease. Unsaturated fat is also a source of omega-3 and omega-6 fatty acids, which are generally referred to as "healthy" fats. Choose foods with unsaturated fat as part of a balanced diet using the U.S. Department of Health and Human Service's Dietary Guidelines.

Even though unsaturated fat is a "good fat," having too much in your diet may lead to having too many calories, which can increase your risk of developing obesity, type 2 diabetes, heart disease, and certain types of cancer.

There are two main types of unsaturated fats:

- Monounsaturated fat, which can be found in:

1. Avocados
2. Nuts and seeds (like cashews, pecans, almonds, and peanuts)
3. Vegetable oils (like canola, olive, peanut, safflower, sesame, and sunflower)

- Polyunsaturated fat, which can be found in:

1. Fatty fish (like herring, mackerel, salmon, trout and smelt)
2. Fish oils
3. Nuts and seeds (like cashews, pecans, almonds and peanuts)
4. Vegetable oils (like canola, corn, flaxseed, soybean and sunflower)

Hydrogenation

Simply put, hydrogenation is a process of adding hydrogen gas to alter the melting point of the oil or fat. The injected hydrogen bonds with the available carbon, which changes liquid oil into solid fat. This is practical, in that it makes fats versatile. Think of the different temperature conditions within a bakery during which fat must be workable; think of the different climatic conditions

encountered in bakeries.

Trans Fat Trans fat is made from a chemical process known as “partial hydrogenation.” This is when liquid oil is made into a solid fat. Like saturated fat, trans fat has been shown to raise LDL or “bad” cholesterol levels, which may in turn increase your risk for heart disease. Unlike saturated fat, trans fat also lowers HDL or “good” cholesterol. A low level of HDL-cholesterol is also a risk factor for heart disease.

Until recently, most of the trans fat found in a typical American diet came from:

- Fried foods (like doughnuts)
- baked goods including cakes, pie crusts, biscuits, frozen pizza, cookies, and crackers
- stick margarine and other spreads

The US Food and Drug Administration (FDA) specifically prescribe what information must be displayed on a label. The trans fat content of food is one piece of core nutrition information that is required to be declared in a nutrition facts table. More information on a nutrition facts table and labeling details can be found in www.fda.gov/food/ingredientsp.../ucm274590.htm

Emulsification (Emulsified Shortenings)

Emulsification is the process by which normally unmixable ingredients (such as oil and water) can be combined into a stable substance. Emulsifiers are substances that can aid in this process. There are natural emulsifiers such as lecithin, found in egg yolks. Emulsifiers are generally made up of monoglycerides and diglycerides and have been added to many hydrogenated fats, improving the fat’s ability to:

- Develop a uniformly fine structure
- Absorb a high percentage of sugar
- Hold in suspension a high percentage of liquid

Emulsified shortenings are ideal for cakes and icings, but they are not suitable for deep-frying.

Stability

Stability refers to the ability of a shortening to have an extended shelf life. It refers especially to deepfrying fats, where a smoke point (see below) of 220°C to 230°C (428°F to 446°F) indicates a fat of high stability.

Smoke Point

The smoke point is the temperature reached when fat first starts to smoke. The smoke point will decline over time as the fat breaks down (see below).

Fat Breakdown

The technical term for fat breakdown is hydrolysis, which is the chemical reaction of a substance with water. In this process, fatty acids are separated from their glycerol molecules and accumulate over time in the fat. When their concentration reaches a certain point, the fat takes on an unpleasant taste, and continued use of the fat will yield a nasty flavor. The moisture, which is at the root of this problem, comes from the product being fried. This is why it is a good reason to turn off the fryer or turn it to “standby” between batches of frying foods such as doughnuts. Another cause of fat breakdown is excessive flour on the product or particles breaking off the product. Attribution

Figure 2. Stearic Acide. Retrieved from http://library.med.utah.edu/NetBioch...Acids/3_3.html

Figure 3 Oleic Acid Retrieved from: http://library.med.utah.edu/NetBioch...Acids/3_3.html

Figure 4 Linoleic Acid Retrieved from: http://library.med.utah.edu/NetBioch...Acids/3_3.html

This page titled 8.1.3.1: Understanding Fats and Oils is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 3.1: Understanding Fats and Oils by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.3.2: Sources of Bakery Fats and Oils

Edible fats and oils are obtained from both animal and vegetable sources. Animal sources include: Beef, Pork, Sheep, and Fish. In North America, the first two are the prime sources. Vegetable sources include canola, coconut, corn, cotton, olive, palm fruit and palm kernel, peanut, soya bean, safflower, and sunflower.

Refining of Fats and Oils

The major steps in refining fats and oils are as follows:

- Free fatty acids are neutralized and treated with an alkali.
- Color is removed.
- The fat is hydrogenated.
- The fat is deodorized.
- The fat is chilled and beaten to make it softer and whiter. This is done by a votator (a machine that cools and kneads liquid margarine).
- Fat is stored to facilitate the correct crystallization (tempering).

This page titled 8.1.3.2: Sources of Bakery Fats and Oils is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 3.2: Sources of Bakery Fats and Oils by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.3.3: Major Fats and Oils Used in Bakeries

Lard

Lard is obtained from the fatty tissues of pigs, with a water content of 12% to 18%. Due to dietary concerns, lard has gradually lost much of its former popularity. It is still extensively used, however, for:

- Yeast dough additions
- Pie pastry
- Pan greasing

Lard has a good plastic range, which enables it to be worked in a pie dough at fairly low temperatures (try the same thing with butter!). It has a fibrous texture and does not cream well. It is therefore not suitable for cake making. Some grades of lard also have a distinctive flavor, which is another reason it is unsuitable for cake making.

Butter

Butter is made from sweet, neutralized, or ripened creams pasteurized and standardized to a fat content of 30% to 40%. When cream is churned or overwhipped, the fat particles separate from the watery liquid known as buttermilk. The separated fat is washed and kneaded in a water wheel to give it plasticity and consistency. Color is added during this process to make it look richer, and salt is added to improve its keeping quality.

In Canada, the following regulations apply to butter:

- Minimum 80% milk fat by weight
- Permitted ingredients: milk solids, salt, air or inert gas, permitted food color, permitted bacterial culture
- The grade and grade name for butter and butter products is Canada 1.

Sweet (or unsalted) butter is made from a cream that has a very low acid content and no salt is added to it. It is used in some baking products like French butter cream, where butter should be the only fat used in the recipe. Keep sweet butter in the refrigerator.

From the standpoint of flavor, butter is the most desirable fat used in baking. Its main drawback is its relatively high cost. It has moderate but satisfactory shortening and creaming qualities. When used in cake mixing, additional time, up to five minutes more, should be allowed in the creaming stage to give maximum volume. Adding an emulsifier (about 2% based on flour weight) will also help in cake success, as butter has a poor plastic range of 18°C to 20°C (64°F to 68°F).

Butter and butter products may also be designated as “whipped” where they have had air or inert gas uniformly incorporated into them as a result of whipping. Whipped butter may contain up to 1% added edible casein or edible caseinates.

Butter and butter products may also be designated as “cultured” where they have been produced from cream to which a permitted bacterial culture has been added.

Margarine

Margarines are made primarily from vegetable oils (to some extent hydrogenated) with a small fraction of milk powder and bacterial culture to give a butter-like flavor. Margarines are very versatile and include:

- General purpose margarine with a low melting point, suitable for blending in dough and general baking
- Cake margarine with excellent creaming qualities
- Roll-in margarine, which is plastic and suitable for Danish pastries
- Puff pastry roll-in, which is the most waxy and has the highest melting point

Margarine may be obtained white, but is generally colored. Margarine has a fat content ranging from 80% to 85%, with the

balance pretty much the same as butter.

Oil content claims on margarine

The claim that margarine contains a certain percentage of a specific oil in advertisements should always be based on the percentage of oil by weight of the total product. All the oils used in making the margarine should be named. For example, if a margarine is made from a mixture of corn oil, cottonseed oil, and soybean oil, it would be considered misleading to refer only to the corn oil content in an advertisement for the margarine. On the other hand, the mixture of oils could be correctly referred to as vegetable oils.

It used to be that you could only buy margarines in solid form full of saturated and trans fat. The majority of today's margarines come in tubs, are soft and spreadable, and are non-hydrogenated, which means they have low levels of saturated and trans fat. Great care must be taken when attempting to substitute spreadable margarine for solid margarine in recipes.

Shortenings

Since the invention of hydrogenated vegetable oil in the early 20th century, shortening has come almost exclusively to mean hydrogenated vegetable oil. Vegetable shortening shares many properties with lard: both are semi-solid fats with a higher smoke point than butter and margarine. They contain less water and are thus less prone to splattering, making them safer for frying. Lard and shortening have a higher fat content (close to 100%) compared to about 80% for butter and margarine. Cake margarines and shortenings tend to contain a bit higher percentage of monoglycerides that margarines. Such "high-ratio shortenings" blend better with hydrophilic (attracts water) ingredients such as starches and sugar.

Health concerns and reformulation

Early in this century, vegetable shortening became the subject of some health concerns due to its traditional formulation from partially hydrogenated vegetable oils that contain trans fats, which have been linked to a number of adverse health effects. Consequently, a low trans-fat variant of Crisco brand shortening was introduced in 2004. In January 2007, all Crisco products were reformulated to contain less than one gram of trans fat per serving, and the separately marketed trans-fat free version introduced in 2004 was consequently discontinued. Since 2006, many other brands of shortening have also been reformulated to remove trans fats. Non-hydrogenated vegetable shortening can be made from palm oil.

Hydrogenated vegetable shortenings

Hydrogenated shortenings are the biggest group of fats used in the commercial baking industry. They feature the following characteristics:

- They are made from much the same oils as margarine.
- They are versatile fats with good creaming ability.
- Their hydrogenation differs according to the specific use for which the fat is designed.
- They are 100% fat – no water.
- They keep well for six to nine months.

Variations on these shortenings are: emulsified vegetable shortenings, roll-in pastry shortenings, and deepfrying fats.

Emulsified vegetable shortenings

Emulsified vegetable shortenings are also termed high-ratio fats. The added emulsifiers (mono- and diglycerides) increase fat dispersion and give added fineness to the baked product. They are ideal for highratio cakes, where relatively large amounts of sugar and liquid are incorporated. The result is a cake:

- Fine in texture
- Light in weight and of excellent volume
- Superior in moisture retention (good shelf life)
- Tender to eat

This is also the fat of choice for many white cake icings.

Roll-in pastry shortenings

This type of shortening is also called special pastry shortening (SPS). These fats have a semi-waxy consistency and offer:

- Large plastic range
- Excellent extensibility
- Excellent lifting ability

They are primarily used in puff pastry and Danish pastry products where lamination is required. They come in various specialized forms, with varying qualities and melting points. It is all a matter of compromise between cost, palatability, and leavening power. A roll-in that does not have “palate cling” may have a melting point too low to guarantee maximum lift in a puff pastry product.

Deep-Frying Fats

Deep-frying fats are special hydrogenated fats that have the following features:

- High smoke point of up to 250°C (480°F)
- High heat stability and resistance to fat breakdown
- No undesirable flavor on finished products
- No greasiness when cold
- These fats contain an anti-foaming agent.

Vegetable Oils

Vegetable oil is an acceptable common name for an oil that contains more than one type of vegetable oil. Generally, when such a vegetable oil blend is used as an ingredient in another food, it may be listed in the ingredients as “vegetable oil.”

There are two exceptions: if the vegetable oils are ingredients of a cooking oil, salad oil, or table oil, the oils must be specifically named in the ingredient list (e.g., canola oil, corn oil, safflower oil), and using the general term vegetable oil is not acceptable. As well, if any of the oils are coconut oil, palm oil, palm kernel oil, peanut oil, or cocoa butter, the oils must be specifically named in the ingredient list.

When two or more vegetable oils are present and one or more of them has been modified or hydrogenated, the common name on the principal display panel and in the list of ingredients must include the word “modified” or “hydrogenated,” as appropriate (e.g., modified vegetable oil, hydrogenated vegetable oil, modified palm kernel oil).

Vegetable oils are used in:

- Chemically leavened batters (e.g., muffin mixes)
- Dough additives (to replace the fat)
- Short sponges (to replace the butter or fat)

Coconut Fat

Coconut fat is often used to stabilize butter creams as it has a very small plastic range. It has a quite low melting point and its hardness is due to other factors. It can be modified to melt at different temperatures, generally between 32°C and 36°C (90°F and 96°F).

The Importance of Melting Points

As mentioned above, all fats become oils and vice versa, depending on temperature. Physically, fats consist of minute solid fat particles enclosing a microscopic liquid oil fraction. The consistency of fat is very important to the baker. It is very difficult to work with butter (relatively low melting point) in hot weather, for example. At the other extreme, fats with a very high melting point are not very palatable, since they tend to stick to the palate. Fat manufacturers have therefore attempted to customize fats to

accommodate the various needs of the baker.

Fats with a melting range between 40°C and 44°C (104°F and 112°F) are considered to be a good compromise between convenience in handling and palatability. New techniques allow fats with quite high melting points without unpleasant palate-cling. Table 1 shows the melting points of some fats.

Table 1 Melting points of typical fats.

Type of Fat	Melting Point
Coconut fat	32.5°C-34.5°C (90.5°F-4.1°F)
Regular margarine	34°C (93°F)
Butter	38°C (100°F)
Regular shortenings	44°C-47°C (111°F-116°F)
Roll-in shortenings	40°C-50°C (104°F-122°F)
Roll-in margarine	44°C-54°C (111°F-130°F)

Blending

It is probably safe to say that most fats are combinations or blends of different oils and/or fats. They may be all vegetable sources. They may be combined vegetable and animal sources. A typical ratio is 90% vegetable source to 10% animal (this is not a hard and fast rule). Formerly, blends of vegetable and animal oils and fats were termed compound fats. Nowadays, this term, if used at all, may refer also to combinations of purely vegetable origin.

This page titled 8.1.3.3: Major Fats and Oils Used in Bakeries is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 3.3: Major Fats and Oils Used in Bakeries by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.3.4: Functions of Fat in Baking

The following summarize the various functions of fat in baking.

Tenderizing Agents

Used in sufficient quantity, fats tend to “shorten” the gluten strands in flour; hence their name: shortenings. Traditionally, the best example of such fat was lard.

Creaming Ability

This refers to the extent to which fat, when beaten with a paddle, will build up a structure of air pockets. This aeration, or creaming ability, is especially important for cake baking; the better the creaming ability, the lighter the cake. Plastic Range Plastic range relates to the temperature at which the fatty acid component melts and over which shortening will stay workable and will “stretch” without either cracking (too cold) or softening (too warm). A fat that stays “plastic” over a temperature range of 4°C to 32°C (39°F to 90°F) would be rated as excellent. A dough made with such a fat could be taken from the walk-in cooler to the bench in a hot bakeshop and handled interchangeably. Butter, on the other hand, does not have a good plastic range; it is almost too hard to work at 10°C (50°F) and too soft at 27°C (80°F).

Lubrication

In dough making, the fat portion makes it easier for the gluten network to expand. The dough is also easier to mix and to handle. This characteristic is known as lubrication.

Moistening Ability

Whether in dough or in a cake batter, fat retards drying out. For this purpose, a 100% fat shortening will be superior to either butter or margarine.

Nutrition

As one of the three major food categories, fats provide a very concentrated source of energy. They contain many of the fatty acids essential for health.

This page titled 8.1.3.4: Functions of Fat in Baking is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 3.4: Functions of Fat in Baking by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4: SUGAR

8.1.4.1: Sugar Chemistry (ADD US)

8.1.4.2: Sugar Refining

8.1.4.3: The Application of Sugar

8.1.4.4: Agave

8.1.4.5: Honey

8.1.4.6: Malt

8.1.4.7: Maple Syrup (ADD US)

8.1.4.8: Sugar Substitutes (ADD US)

8.1.4.9: Glucose/Dextrose

Contributors and Attributions

Sorangel Rodriguez-Velazquez (American University).

Chemistry of Cooking by Sorangel Rodriguez-Velazquez is licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International License, except where otherwise noted

This page titled 8.1.4: Sugar is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4: Sugar by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.1: Sugar Chemistry (ADD US)

Chemically, sugar consists of carbon (C), oxygen (O), and hydrogen (H) atoms, and is classified as a **carbohydrate**. There are three main groups of sugars, classified according to the way the atoms are arranged together in the molecular structure. These groups are the following:

- **Monosaccharides** or simple sugars. **Dextrose (glucose)** is the major monosaccharide. Others are l**evulose** or **fructose** (found in honey and many fruits), and galactose, which is a milk sugar. Such sugars do not readily crystallize. (Mono means one, indicating that the sugar consists of only one molecule.)
- **Disaccharides** or complex sugars. Sucrose (common sugar) is the primary example of a disaccharide. Maltose, found in cereals, and **lactose**, found in milk, are others.
- **Polysaccharides**. Examples are starches, dextrins, and cellulose.

Bakers are not concerned with polysaccharides but rather with the monosaccharides and disaccharides. The latter two both sweeten, but they cannot be used interchangeably because they have different effects on the end product. These differences are touched on later in the book.

Sugar Names

It is helpful to understand some of the conventions of the names of different sugars. Note that sugar names often end in “ose”: sucrose, dextrose, maltose, lactose, etc. **Sucrose** is the chemical name for sugar that comes from the cane and beet sugar plants.

Note that glucose is the chemical name for a particular type of sugar. What is sometimes confusing is that glucose occurs naturally, as a sugar molecule in substances such as honey, but it is also produced industrially from the maize plant (corn).

The Canadian Food and Drug Regulations (FDR) govern the following definitions:

- Sugars: All monosaccharides and disaccharides. Used for nutrition labelling purposes.
- Sweetening agent: Any food for which a standard is provided in Division 18 of the Food and Drug Regulation, or any combination of these. Includes sugar (sucrose), sugar syrups, and molasses derived from sugar cane or sugar beet, dextrose, glucose and syrups, honey and lactose. Excludes sweeteners considered to be food additives.
- Sweetening ingredient: Any sugar, invert sugar, honey, dextrose, glucose, or glucose solids, or any combination of these in dry or liquid form. Designed for sweetening fruits, vegetables, and their products and substitutes.
- Maple syrup: The syrup obtained by the concentration of maple sap or by the dilution or solution of a maple product, other than maple sap, in potable water.
- Sweetener: Any food additive listed as a sweetener. Includes both sugar alcohols and high intensity- sweeteners such as acesulfame-potassium, aspartame, and sucralose.
- Sugar alcohols: Food additives that may be used as sweeteners. Includes isomalt, lactitol, maltitol, maltitol syrup, mannitol, sorbitol, sorbitol syrup, xylitol, and erythritol.

This page titled 8.1.4.1: Sugar Chemistry (ADD US) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.1: Sugar Chemistry by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.2: Sugar Refining

While some refining usually occurs at source, most occurs in the recipient country. The raw sugar that arrives at the ports is not legally edible, being full of impurities.

At the refinery, the raw brown sugar goes through many stages:

- Washing and boiling
- Filtering to remove impurities
- Evaporation to the desired crystal size under vacuum to avoid caramelization Centrifuging, in which the fluid is spun off leaving the crystals
- Drying in a rotating drum with hot air
- Packaging in various sizes, depending on the intended market

Sugar beet undergoes identical steps after the initial processing, which involves:

- Slicing the beets and extracting the sugar with hot water
- Removing impurities
- Filtration
- Concentration in evaporators

From here, the process is identical to the final steps in cane processing. See Figure 2 which illustrates the process.

Some of the sugar passes through a machine that presses the moist sugar into cubes and wraps and packages them; still other sugar is made into icing sugar. The sugar refining process is completely mechanical, and machine operators' hands never touch the sugar.

Brown and yellow sugars are produced only in cane sugar refineries. When sugar syrup flows from the centrifuge machine, it passes through further filtration and purification stages and is re-boiled in vacuum pans such as the two illustrated in Figure 2. The sugar crystals are then centrifuged but not washed, so the sugar crystals still retain some of the syrup that gives the product its special flavour and colour.

During the whole refining process almost 100 scientific checks for quality control are made, while workers in research laboratories at the refineries constantly carry out experiments to improve the refining processand the final product. Sugar is carefully checked at the mills and is guaranteed to have a high purity. Government standards both in the United States and Canada require a purity of at least 99.5% sucrose.

Are animal ingredients included in white sugar?

Bone char — often referred to as natural carbon — is widely used by the sugar industry as a decolourizing
filter, which allows the sugar cane to achieve its desirable white colour. Other types of filters involve granular carbon or an ion-exchange system rather than bone char.

Bone char is made from the bones of cattle, and it is heavily regulated by the European Union and the USDA. Only countries that are deemed BSE-free can sell the bones of their cattle for this process.

Bone char is also used in other types of sugar. Brown sugar is created by adding molasses to refined sugar,
so companies that use bone char in the production of their regular sugar also use it in the production of their brown sugar. Confectioner's sugar — refined sugar mixed with cornstarch — made by these companies also involves the use of bone char. Fructose may, but does not typically, involve a bone-char filter.

Bone char is not used at the sugar beet factory in Taber, Alberta, or in Montreal's cane refinery. Bone char is used only at the Vancouver cane refinery. All products under the Lantic trademark are free of bone char. For the products under the Rogers trademark, all Taber sugar beet products are also free of bone char. In order to differentiate the Rogers Taber beet products from the Vancouver cane products, you can verify the inked-jet code printed on the product. Products with the code starting with the number "22" are from Taber, Alberta, while products with the code starting with the number "10" are from Vancouver.

If you want to avoid all refined sugars, there are alternatives such as sucanat and turbinado sugar, which are not filtered with bone char. Additionally, beet sugar — though normally refined — never involves the use of bone char.

This page titled 8.1.4.2: Sugar Refining is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.2: Sugar Refining by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.3: The Application of Sugar

Sugar is the third most used ingredient in the bakeshop. Sugar has several functions in baking. The most recognized purpose is, of course, to sweeten food, but there are many other reasons sugar is used in cooking and baking:

- It can be used for browning effect, both caramelization and the Maillard reaction, on everything from breads to cookies to cakes. Browning gives a pleasant colour and flavour to the finished product. Caramelization results from the action of heat on sugars. At high temperatures, the chemical changes associated with melting sugars result in a deep brown colour and new flavours. The Maillard reaction results from chemical interactions between sugars and proteins at high heat. An amino group from a protein combines with a reducing sugar to produce a brown colour in a variety of foods (e.g., brewed coffee, fried foods, and breads).
- It acts as the most important tenderizing agent in all baked goods, and one of the factors responsible for the spread in cookies. It helps delay the formation of gluten, which is essential for maintaining a soft or tender product.
- It makes an important contribution to the way we perceive the texture of food. For example, adding sugar to ice cream provides body and texture, which is perceived as smoothness. This addition helps prevent lactose crystallization and thus reduces sugar crystal formation that otherwise causes a grainy texture sometimes associated with frozen dairy products.
- It preserves food when used in sufficient quantity.
- In baking, it increases the effectiveness of yeast by providing an immediate and more usable source of nourishment for the yeast's growth. This hastens the leavening process by producing more carbon dioxide, which allows the dough to rise at a quicker and more consistent rate.

Just as there are many functions of sugar in the bakeshop, there are different uses for the various types of sugar as well:

- Fine granulated sugar is most used by bakers. It generally dissolves easily in mixes and is pure enough for sugar crafters to boil for "pulled" sugar decorations.
- Coarse granulated sugar may be used for a topping on sugar cookies, puff pastry, and Danish pastries as it doesn't liquefy or caramelize so readily. In some European countries, an extra coarse sugar (called hail — a literal translation) is used for this purpose.
- Icing or powdered sugar is used in icings and fillings and in sifted form as a top decoration on many baked goods.
- Brown or yellow sugars are used where their unique flavor is important, or in bakeries where an old- fashioned or rustic image is projected. Brown sugar can usually be substituted for white sugar without technical problems in sugar/batter mixes such as cakes and muffins, and in bread dough.

This page titled 8.1.4.3: The Application of Sugar is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.3: The Application of Sugar by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.4: Agave

Agave has gained popularity in the food industry due to some of its nutritional properties. The agave nectar is obtained from the sap of the heart of the agave plant, a desert succulent, which is also used to produce tequila. The syrup/sugar production process of agave is similar to that of sugar. See more about the nutritional properties and application of agave in the chapter Special Diets, Allergies, Intolerances, Emerging Issues, and Trends in the open textbook Nutrition and Labelling for the Canadian Baker.

A video on the production of agave syrup is available

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This page titled 8.1.4.4: Agave is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.4: Agave by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.5: Honey

Honey is a natural food, essentially an invert sugar. Bees gather nectar and, through the enzyme invertase, change it into honey. Honey varies in composition and flavor depending on the source of the nectar. The average composition of honey is about 40% levulose, 35% dextrose, and 15% water, with the remainder being ash, waxes, and gum.

Blended honey is a mixture of pure honey and manufactured invert sugar, or a blend of different types of honey mixed together to produce a good consistency, color, and aroma. Dehydrated honey is available in a granular form.

Store honey in a tightly covered container in a dry place and at room temperature because it is hygroscopic, meaning it absorbs and retains moisture. Refrigeration or freezing won't harm the color or flavor but it may hasten granulation. Liquid honey crystallizes during storage and is re-liquefied by warming in a double boiler not exceeding a temperature of 58°C (136°F).

Honey is used in baking:

- As a sweetener
- To add unique flavor
- In gingerbread and special cookies where a certain moistness is characteristic of the product
- To improve keeping qualities

There are several types of honey available:

- Comb honey is "packed by the bees" directly from the hive.
- Liquid honey is extracted from the comb and strained. It is the type used by most bakers.
- Creamed honey has a certain amount of crystallized honey added to liquid honey to give body to the final product.
- Chunk honey consists of pieces of comb honey as well as liquid.
- Granulated honey has been crystallized.

In the United States, honey categories are based on color, from white to dark amber. Honey from orange blossom is an example of white honey. Clover honey is an amber honey, and sage and buckwheat honeys are dark amber honeys.

This page titled 8.1.4.5: Honey is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.5: Honey by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.6: Malt

Malt is the name given to a sweetening agent made primarily from barley. The enzymes from the germ of the seeds become active, changing much of the starch into maltose, a complex sugar. Maltose has a distinct flavor and is used for making yeast products such as bread and rolls. Malt is considered to be relatively nutritious compared to other sweeteners.

Malt is available as:

- Flour
- Malt syrup
- Malt extract
- Dried malt

The flour is not recommended since it can lead to problems if not scaled precisely. Malt syrup is inconvenient to work with, as it is sticky, heavy, and bulky. Dried malt is the most practical, though it must be kept protected from humidity.

There are two distinct types of malt:

- Diastatic malt flour is dried at low temperature, thus retaining the activity of the diastatic enzymes.
- Non-diastatic malt flour is darker in color. It is treated at high temperature, which kills the enzymes, and the result is non-diastatic malt.

Crushing malted grain in water produces malt syrup. This dissolves the maltose and soluble enzymes. The liquid is concentrated, producing the syrup. If the process is continued, a dry crystallized product called dried malt syrup is obtained.

Malt syrup has a peculiar flavor, which many people find desirable. It is used in candy, malted milk, and many other products. The alcoholic beverage industry is the largest consumer of malt by far, but considerable quantities are used in syrup and dried malt syrup, both of which are divided into diastatic and non-diastatic malt.

Both diastatic and non-diastatic malts add sweetness, color, and flavor to baked products. Both are valuable since they contain malt sugar, which is fermented by the yeast in the later stages of fermentation. Other sugars such as glucose and levulose are used up rapidly by fermenting yeast in the early stages of fermentation. Diastatic malt is made with various levels of active enzymes. Malt with medium diastatic activity is recommended. Normally, bread bakers will find sufficient enzymes in well-balanced flour from a good mill, so it is unnecessary to use diastatic malt.

When using dry diastatic malt, about the same weight should be used as liquid regular diastatic malt. Adjustment is made at the factory insofar as the enzyme level is increased in the dry product to compensate. Since the dry type contains about 20% less moisture than the liquid type, add water to make up the difference if dry diastatic malt is substituted for malt syrup.

The main uses of malt in the bakery are to:

- Add nutritive value, as it is rich in vitamins and essential amino acids
- Lengthen shelf life through its ability to attract moisture
- Help fermentation by strengthening the gluten and feeding the yeast Make products more appealing through browning of the crust
- Add unique flavor to products when used in sufficient quantity

Table 1 shows the suggested use levels for malt.

Table 1 Recommended level of malt for various baked goods

Product	Percentage of Flour Weight
White pan bread	0.5-1.5
Sweet goods	1.5-3.0
French/Italian bread	0.5-2.0

Product	Percentage of Flour Weight
Whole wheat bread	5.0-9.0
Pretzels	1.5-6.0
Hard rolls	3.0-5.5

This page titled 8.1.4.6: Malt is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.6: Malt by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.7: Maple Syrup (ADD US)

Canada is responsible for 84% of the world's maple syrup production, with the United States being responsible for the remaining 16%. Maple syrup is made by boiling and evaporating the sap of the sugar maple tree. Because sap is only 2% or 3% sugar, it takes almost 40 liters of sap to make 1 liter of syrup. This makes maple syrup a very expensive sweetener. It is prized for its unique flavor and sweet aroma. Don't confuse maple-flavored pancake or table syrup with real maple syrup. Table syrup is made from inexpensive glucose or corn syrup, with added caramel coloring and maple flavoring.

Maple syrup in Canada has two categories:

1. Canada Grade A, which has four color/flavor classes
    - golden, delicate taste
    - amber, rich taste
    - dark, robust taste
    - very dark, strong taste
2. Canada Processing Grade, which has no color descriptors (any maple syrup that possesses minimal food quality defects but still meets all government regulatory standards for food quality and safety for human consumption)

This definition and grading system gives consumers more consistent and relevant information about the varieties, and helps them make informed choices when choosing maple syrup.

Darker maple syrups are better for baking as they have a more robust flavor. Using maple sugar is also a good way to impart flavor. Maple sugar is what remains after the sap of the sugar maple is boiled for longer than is needed to create maple syrup. Once almost all the water has been boiled off, all that is left is a solid sugar. It can be used to flavor some maple products and as an alternative to cane sugar.

For a video on maple syrup production, see: https://www.youtube.com/watch?v=OFIj4pMYpTQ

This page titled 8.1.4.7: Maple Syrup (ADD US) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.7: Maple Syrup (ADD US) by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.8: Sugar Substitutes (ADD US)

In Canada, food additives such as sugar substitutes, which cover both artificial sweeteners and intense sweeteners obtained from natural sources, are subject to rigorous controls under theFood and Drugs Act and Regulations. New food additives (or new uses of permitted food additives) are permitted only once a safety assessment has been conducted and regulatory amendments have been enacted.

Several sugar substitutes have been approved for use in Canada. These include acesulfame-potassium, aspartame, polydextrose, saccharin, stevia, sucralose, thaumatin, and sugar alcohols (polyols) like sorbitol, isomalt, lactitol, maltitol, mannitol, and xylitol. Please see the Health Canada website for more information on sugar substitutes.

Bakers must be careful when replacing sugar (sucrose) with these sugar substitutes in recipes. Even though the sweetness comparison levels may be similar (or less), it is generally not possible to do straight 1-for-1 substitution. Sugar (sucrose) plays many roles in a recipe:

- It is a bulking agent.
- It absorbs moisture.
- It is a tenderizer.
- It adds moisture and extends shelf life. It adds color (caramelization).

Sugar substitutes may not work in a recipe in the same way. More information on sugar substitutes and their relative sweetness can be found here: http://www.sugar-and-sweetener-guide...er-values.html

This page titled 8.1.4.8: Sugar Substitutes (ADD US) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.8: Sugar Substitutes (ADD US) by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.4.9: Glucose/Dextrose

The sugar known as glucose has two origins:

1. In a natural form in most fruits
2. In a processed form from corn (corn syrup)

In baking, we usually refer to industrially made glucose. It is made from corn and the resulting product, a thick syrup, is then adjusted to a uniform viscosity or consistency. The particular form of the syrup is defined by what is known as the *dextrose equivalent,* or DE for short. Corn syrup is the most familiar form of glucose.

In plant baking, high-fructose corn syrup (HFCS) is the major sweetening agent in bread and buns. It consists of roughly half fructose and half dextrose. Dextrose (chemically identical to glucose) is available in crystalline form and has certain advantages over sucrose:

- It is easily fermentable.
- It contributes to browning in bread and bun making.
- In crystalline form, it is often used in doughnut sugars as it is more inclined to stay dry and non- greasy.
- It is hygroscopic and valued as a moisture-retaining ingredient.
- It retards crystallization in syrups, candies, and fondant.

Corn syrup is made from the starch of maize (corn) and contains varying amounts of glucose and maltose, depending on the processing methods. Corn syrup is used in foods to soften texture, add volume, prevent crystallization of sugar, and enhance flavor.

Glucose/dextrose has a sweetening level of approximately three-quarters that of sugar. Table 1 shows the amount of corn syrup or HFCS needed to replace sugar in a formula.

p[illegible]4

Table 1:Replacement factor for Corn Syrup and High-Fructose Corn Syrup Type of Sugar

	Solids	Replacement Factor
Granulated sugar	100%	1.0
Regular corn syrup	80%	1.25
High-fructose corn syrup	71%	1.41

Glucose, HFCS, and corn syrup are not appropriate substitutions for sucrose in all bakery products. Certain types of cakes, such as white layer cakes, will brown too much if glucose or HFCS is used in place of sugar.

This page titled 8.1.4.9: Glucose/Dextrose is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 4.9: Glucose/Dextrose by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5: LEAVENING AGENTS

The word leavening in the baking trade is used to describe the source of gas that makes a dough or batter expand in the presence of moisture and heat. Leavening agents are available in different forms, from yeast (the organic leavener) to chemical, mechanical, and physical leaveners. Bakers choose the appropriate type of leavening based on the product they are making.

8.1.5.1: Introduction to Leavening Agents

8.1.5.2: Yeast

8.1.5.3: The Functions of Yeast

8.1.5.4: Using Yeast in Baking

8.1.5.5: Baking Powder

8.1.5.6: Sodium Bicarbonate

8.1.5.7: Ammonium Bicarbonate

8.1.5.8: Water Hardness and pH

This page titled 8.1.5: Leavening Agents is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5: Leavening Agents by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.1: Introduction to Leavening Agents

The word leavening in the baking trade is used to describe the source of gas that makes a dough or batter expand in the presence of moisture and heat. Leavening agents are available in different forms, from yeast (the organic leavener) to chemical, mechanical, and physical leaveners. Bakers choose the appropriate type of leavening based on the product they are making.

This page titled 8.1.5.1: Introduction to Leavening Agents is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.1: Introduction to Leavening Agents by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.2: Yeast

Yeast is a microscopic unicellular fungus that multiplies by budding, and under suitable conditions, causes fermentation. Cultivated yeast is widely used in the baking and distilling industries. History tells us that the early Chaldeans, Egyptians, Greeks, and Romans made leavened bread from fermented doughs. This kind of fermentation, however, was not always reliable and easy to control. It was Louis Pasteur, a French scientist who lived in the 19th century, who laid the foundation for the modern commercial production of yeast as we know it today through his research and discoveries regarding the cause and prevention of disease.

Types of Yeast

Wild yeast spores are found floating on dust particles in the air, in flour, on the outside of fruits, etc. Wild yeasts form spores faster than cultivated yeasts, but they are inconsistent and are not satisfactory for controlled fermentation purposes.

Compressed Yeast

Compressed yeast is made by cultivating a select variety, which is known by experiment to produce a yeast that is hardy, consistent, and produces a fermentation with strong enzymatic action. These plants are carefully isolated in a sterile environment free of any other type of yeast and cultivated on a plate containing nutrient agar or gelatin. Wort, a combination of sterilized and purified molasses or malt, nitrogenous matter, and mineral salts is used to supply the food that the growing yeast plants need to make up the bulk of compressed yeast.

After growing to maturity in the fermentation tank, the yeast is separated from the used food or wort by means of centrifugal machines. The yeast is then cooled, filtered, pressed, cut, wrapped, and refrigerated. It is marketed in 454 g (1 lb.) blocks, or in large 20 kg (45 lb.) bags for wholesale bakeries.

Figure 1 illustrates the process of cultivating compressed yeast, and Table 1 summarizes its composition.

page185image17960 page185image18120 page185image18280

page186image432

Figure 1 Cultivating compressed yeast

Table 1 Average composition of fresh (compressed) yeast

Water	68% to 73%
Protein	12% to 14%
Fat	0.6% to 0.8%
Carbohydrate	9% to 11%
Mineral Matter	1.7% to 2%

Active Dry Yeast

Active dry yeast is made from a different strain than compressed yeast. The manufacturing process is the same except that the cultivated yeast is mixed with starch or other absorbents and dehydrated. Its production began after World War II, and it was used mainly by the armed forces, homemakers, and in areas where fresh yeast was not readily available.

Even though it is a dry product, it is alive and should be refrigerated below 7°C (45°F) in a closed container for best results. It has a moisture content of about 7%. Storage without refrigeration is satisfactory only for a limited period of time. If no refrigeration is available, the yeast should be kept unopened in a cool, dry place. It should be allowed to warm up to room temperature slowly before being used.Dry yeast must be hydrated for about 15 minutes in water at least four times its weight at a temperature between 42°C and 44°C (108°F and112°F). The temperature should never be lower than 30°C (86°F), and dry yeast should never

be used before it is completely dissolved.

It takes about 550 g (20 oz.) of dry yeast to replace 1 kg (2.2 lb.) of compressed yeast, and for each kilogram of dry yeast used, an additional kilogram of water should be added to the mix. This product is hardly, if ever, used by bakers, having been superseded by instant yeast (see below).

Instant Dry Yeast

Unlike instant active dry yeast that must be dissolved in warm water for proper rehydration and activation, instant dry yeast can be added to the dough directly, either by:

Mixing it with the flour before the water is added

Adding it after all the ingredients have been mixed for one minute

This yeast can be reconstituted. Some manufacturers call for adding it to five times its weight of water at a temperature of 32°C to 38°C (90°F to 100°F). Most formulas suggest a 1:3 ratio when replacing compressed yeast with instant dry. Others vary slightly, with some having a 1:4 ratio. In rich Danish dough,
it takes about 400 g (14 oz.), and in bread dough about 250 g to 300 g (9 oz. to 11 oz.) of instant dry yeast

to replace 1 kg (2.2 lb.) of compressed yeast. As well, a little extra water is needed to make up for the moisture in compressed yeast. Precise instructions are included with the package; basically, it amounts to the difference between the weight of compressed yeast that would have been used and the amount of dry yeast used.

Instant dry yeast has a moisture content of about 5% and is packed in vacuum pouches. It has a shelf life of about one year at room temperature without any noticeable change in its gassing activity. After the seal is broken, the content turns into a granular powder, which should be refrigerated and used by its best-before date, as noted on the packaging.

Instant dry yeast is especially useful in areas where compressed yeast is not available. However, in any situation, it is practical to use and has the advantages of taking up less space and having a longer shelf life than compressed yeast.

Cream Yeast

Creamy yeast is a soft slurry-type yeast that is used only in large commercial bakeries and is pumped into the dough.

Yeast Food

Yeast food is used in bread production to condition the dough and speed up the fermentation process. It consists of a blend of mineral salts such as calcium salt or ammonium salt and potassium iodate. It has a tightening effect on the gluten and is especially beneficial in dough where soft water is used. The addition of yeast food improves the general appearance and tasting quality of bread. The retail baker does not use it much.

This page titled 8.1.5.2: Yeast is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.2: Yeast by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.3: The Functions of Yeast

Yeast has two primary functions in fermentation:

To convert sugar into carbon dioxide gas, which lifts and aerates the dough
To mellow and condition the gluten of the dough so that it will absorb the increasing gases evenly and hold them at the same time

In baked products, yeast increases the volume and improves the flavor, texture, grain, color, and eating quality. When yeast, water, and flour are mixed together under the right conditions, all the food required for fermentation is present as there is enough soluble protein to build new cells and enough sugar to feed them.

Activity within the yeast cells starts when enzymes in the yeast change complex sugar into invert sugar. The invert sugar is, in turn, absorbed within the yeast cell and converted into carbon dioxide gas and alcohol. Other enzymes in the yeast and flour convert soluble starch into malt sugar, which is converted again by other enzymes into fermentable sugar so that aeration goes on from this continuous production of carbon dioxide.

Proper Handling of Yeast

Compressed yeast ages and weakens gradually even when stored in the refrigerator. Fresh yeast feels moist and firm, and breaks evenly without crumbling. It has a fruity, fresh smell, which changes to a sticky mass with a cheesy odor. It is not always easy to recognize whether or not yeast has lost enough of its strength to affect the fermentation and the eventual outcome of the baked bread, but its working quality definitely depends on the storage conditions, temperature, humidity, and age.

The optimum storage temperature for yeast is -1°C (30°F). At this temperature it is still completely effective for up to two months. Yeast does not freeze at this temperature.

Other guidelines for storing yeast include:

Rotating it properly and using the older stock first
Avoiding overheating by spacing it on the shelves in the refrigerator

Yeast needs to breathe, since it is a living fungus. The process is continuous, proceeding slowly in the refrigerator and rapidly at the higher temperature in the shop. When respiration occurs without food, the yeast cells starve, weaken, and gradually die.

Yeast that has been frozen and thawed does not keep and should be used immediately. Freezing temperatures weaken yeast, and thawed yeast cannot be refrozen successfully.

This page titled 8.1.5.3: The Functions of Yeast is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.3: The Functions of Yeast by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.4: Using Yeast in Baking

Many bakers add compressed yeast directly to their dough. A more traditional way to use yeast is to dissolve it in lukewarm water before adding it to the dough. The water should never be higher than 50°C (122°F) because heat destroys yeast cells. In general, salt should not come into direct contact with yeast, as salt dehydrates the yeast. (Table 1 indicates the reaction of yeast at various temperatures.)

It is best to add the dissolved yeast to the flour when the dough is ready for mixing. In this way, the flour is used as a buffer. (Buffers are ingredients that separate or insulate ingredients, which if in too close contact, might start to react prematurely.) In sponges where little or no salt is used, yeast buds quickly and fermentation of the sponge is rapid.

Table 1 How yeast reacts at different temperatures

Temperature	Reaction
15°C -20°C (60°F -68°F)	slow reaction
26°C -29°C (80°F -85°F)	normal reaction
32°C -38°C (90°F -100°F)	fast reaction
59°C (138°F)	terminal death point

Never leave compressed yeast out for more than a few minutes. Remove only the amount needed from the refrigerator. Yeast lying around on workbenches at room temperature quickly deteriorates and gives poor results. One solution used by some bakeries to eliminate steps to the fridge is to have a small portable cooler in which to keep the yeast on the bench until it is needed. Yeast must be kept wrapped at all times because if it is exposed to air the edges and the corners will turn brown. This condition is known as air- burn.

This page titled 8.1.5.4: Using Yeast in Baking is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.4: Using Yeast in Baking by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.5: Baking Powder

Baking powder is a dependable, high-quality chemical leavener. To be effective, all baking powders rely on the reaction between one or more acids on sodium bicarbonate to produce carbon dioxide gas. Just as with yeast leavening, the presence of carbon dioxide gas creates air bubbles that cause the product to rise.

There are two main types of baking powders available on the market:

- Continuous or single-action baking powder
- Double- or multiple-action baking powder

The difference between continuous- and double-action baking powders is simply the rate of reaction:

- Continuous-action baking powder uses one acid, which continuously reacts with the soda to release gas steadily throughout the baking process until all the gassing power is spent.
- Double-action baking powder contains two different acids, which react with soda at different stages of the baking process. One acid reacts to give off a small amount of gas at low temperature, and the other major acid reacts at baking temperatures to give off the bulk of the gas.

The Leavening Mechanism of Baking Powder

Before baking, approximately 15% of the CO2 gas is released in the cold stage. Eighty-five percent of the CO2 gas is released in the oven starting at approximately 40°C (105°F). Some leavening power is apparently lost in the cold stage, but there is usually still adequate gassing power in the remaining portion.

When the baking powder is activated through moisture and heat, the gas works its way into the many cells created by the mixing or creaming of the batter and starts to expand them. This process comes to a halt when the starch gelatinizes and the cells become rigid. This starts at about 60°C (140°F) and is more or less complete at around 75°C (167°F). After this point, some gas may still be created, but it simply escapes through the porous structure of the product.

Using Baking Powder

For even distribution throughout the batter, baking powder should be sifted with the flour or other dry ingredients. For most cakes, about 5% baking powder to the weight of the flour produces an optimum result. Accurate scaling is important, since a little too much may cause the product to collapse. (Note this is unlike yeast, where an “overdose” will usually simply cause a more rapid rise.)

This page titled 8.1.5.5: Baking Powder is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.5: Baking Powder by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.6: Sodium Bicarbonate

When sodium bicarbonate (baking soda) is moistened and heated, it releases carbon dioxide gas. If it is moistened and heated in the presence of sufficient acid, it will release twice as much gas as if it is moistened and heated without the presence of an acid.

Slightly acidic ingredients provide the mix with some of the necessary acids for the release of carbon dioxide gas. Examples are:

Honey Molasses Ginger Cocoa Bran

For this reason, some of the mixes contain baking powder only while others contain a combination of baking powder and baking soda. If an excessive amount of baking soda is used in a cake batter without the presence of sufficient acid, the normally white cake crumb will have a yellowish-brown color and a strong undesirable smell of soda.

The gas evolves very fast at the beginning of baking when thepH level is still on the acidic side (pH of around 5 to 6). Once the soda neutralizes the acid, the dough or batter quickly becomes alkaline and the release of gas is reduced. Mixes and doughs leavened with baking soda must be handled without delay, or the release of the gas may be almost exhausted before the product reaches the oven.

The darker color of the crumb found on the bottom half of a cake or muffins is caused by the partial dehydration of the batter that is heated first during baking. In spiced honey cookies and gingerbread, baking soda is used alone to give them quick color during baking and yet keep the products soft.

In chocolate cakes, baking soda is used in conjunction with baking powder to keep the pH at a desirable level. However, it is important to know whether the cocoa powder you are using is natural or treated by the Dutch process. In the Dutch process, some of the acid in the cocoa is already neutralized, and there is less left for the release of gas in the mix. This means more baking powder and less baking soda is used.

Baking soda in a chocolate mix not only counteracts the acid content in the baked cake but also improves the grain and color of the cake. A darker and richer chocolate color is produced if the acid level is sufficient to release all the carbon dioxide gas. On the other hand, the reddish, coarse, open-grained crumb in devil's food cake is the result of using baking soda as the principal leavening agent.

The level of baking soda depends on the nature of the product and on the other ingredients in the formula. Cookies, for example, with high levels of fat and sugar, do not require much, if any, leavening.

Table 1 provides the recommended amounts of baking soda for different products. Note that the percentages appear small compared to the 5% level of baking powder suggested because baking powder
contains both an acid agent and a leavening agent.

p[illegible]8

Table 1 Recommended amounts of baking soda	
Product	Amount of Baking Soda (% of flour weight)
Cookies	0.4-0.6

page207image33224 page207image33384 page207image33544

Cakes	0.5-1.0
Cake doughnuts	0.7-1.0
Pancakes	1.4-2.0

This page titled 8.1.5.6: Sodium Bicarbonate is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.6: Sodium Bicarbonate by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.7: Ammonium Bicarbonate

Ammonium bicarbonate is a white crystalline powder used in flat, spiced cookies, such as gingerbreads, and in eclair paste. It must be dissolved in the cold liquid portion of the batter. At room temperature, decomposition of CO_2 in the batter is minimal. When heated to approximately 60°C (140°F) decomposition is more noticeable, and at oven temperature, decomposition takes place in a very short time. Ammonium bicarbonate should only be used in low moisture-containing products that are not dense. Providing that these conditions are met, there will be no taste and odor remaining from the ammonium.

This page titled 8.1.5.7: Ammonium Bicarbonate is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.7: Ammonium Bicarbonate by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.5.8: Water Hardness and pH

Effects on Baking

Most municipal supplies of water contain chlorine, which is used to ensure the purity of the water. Some cities add fluoride to their water supply to stop tooth decay. Neither chlorine nor fluoride is present in large enough quantities to affect dough in any way. In addition, most municipal water is treated to reduce excessive acidity, since this could be corrosive for the water lines. It is therefore unlikely that bakers using municipal water need to be concerned about extremely acidic water.

Soft water is another matter, as it can lead to sticky dough. An addition of yeast food, or a reduction in dough water, will help. Alkaline water tends to tighten the dough and retard fermentation, since enzymes work best in slightly acidic dough.

If there is a possibility of water problems, a sample should be forwarded to a laboratory for a complete analysis.

This page titled 8.1.5.8: Water Hardness and pH is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 5.8: Water Hardness and pH by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6: DAIRY PRODUCTS

Add all section in this chapter

8.1.6.1: Introduction to Dairy Products

8.1.6.2: Milk

8.1.6.3: Milk Products ADD US

8.1.6.4: Milk in bread baking

8.1.6.5: Yogurt

8.1.6.6: Lactose

8.1.6.7: Cheese

Contributors and Attributions

Sorangel Rodriguez-Velazquez (American University).

Chemistry of Cooking by Sorangel Rodriguez-Velazquez is licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International License, except where otherwise noted

This page titled 8.1.6: Dairy Products is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6: Dairy Products by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6.1: Introduction to Dairy Products

Milk and milk products are some of our oldest and best-known natural foods. In baking, milk is used fresh, condensed, powdered, skimmed, or whole. The great bulk, weight, and perishability of fresh milk plus the expense of refrigeration makes it a relatively high-cost ingredient, and for this reason, most modern bakeries use non-fat powdered milk or buttermilk powder.

Over the past 20 years, there has been a trend to lower fat content in dairy products. This reflects the high caloric value of milk fat, and also is compatible with the trend to leaner, healthier nutrition. These “low-fat” products often have the fat replaced with sugars, so care must be taking in substituting these ingredients
in a recipe. For bakers, this trend has not meant any great changes in formulas: a 35% milk fat or a 15% cream cheese product usually works equally well in a cheesecake. Some pastry chefs find lowering the richness in pastries and plated desserts can make them more enjoyable, especially after a large meal.

Table 1 provides the nutritional properties of milk products.

Table 1 Nutritional properties of milk products (per 100 g)

	Whole Milk (3.5% milk fat)	Skim Milk (0.1% milk fat)	Coffee Cream (18% milk fat)	Heavy or Whipping Cream (36% milk fat)
Protein	3.22 g	3.37 g	3g	2g
Fat	3.25 g	0.08 g	19 g	37 g
Cholesterol	10 mg	2 mg	66 mg	137 mg
Potassium	143 mg	156 mg	122 mg	75 mg
Calcium	113 mg	125 mg	96 mg	65 mg
Magnesium	10 mg	11 mg	9 mg	7 mg
Sodium	40 mg	42 mg	40 mg	40 mg
Vitamin A (IU)	102 IU	204 IU	656 IU	1470 IU

Note: Besides the elements shown in Table 1, all dairy products contain vitamin B-complex. IU = International Units, a term used in nutritional measurement

This page titled 8.1.6.1: Introduction to Dairy Products is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6.1: Introduction to Dairy Products by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6.2: Milk

Homogenized milkis fresh milk in which the fat particles are so finely divided and emulsified mechanically that the milk fat cannot separate on standing. The milk fat is forced into tiny droplets. As soon as the droplets form, milk proteins and emulsifiers form a protective film around each one, preventing the fat from reuniting. The tiny droplets stay suspended indefinitely, and milk fat no longer separates and rises

to the top as a cream layer. In other words, homogenized dairy products are stable emulsions of fat droplets suspended in milk. It is also said that homogenized milk is more readily digestible.

Pasteurization of milk was developed in 1859 by the French chemist Louis Pasteur. One method of pasteurization is to heat milk to above 71°C (160°F), maintain it at this temperature for a set time, then cool it immediately to 10°C (50°F) or lower. This kills all harmful bacteria that carry the potential threat of bovine tuberculosis and fever from cows to humans.

The two main types of pasteurization used today are high-temperature, short-time (HTST, also known as “flash”) and higher-heat, shorter time (HHST). Ultra-high-temperature (UHT) processing is also used.

High-temperature, short-time (HTST) pasteurization is done by heating milk to 72°C (161°F) for 15 seconds. Milk simply labelled “pasteurized” is usually treated with the HTST method.
Higher-heat, shorter time (HHST) milk and milk products are pasteurized by applying heat continuously, generally above 100°C (212°F) for such time to extend the shelf life of the product under refrigerated conditions. This type of heat process can be used to produce dairy products with extended shelf life (ESL).

Ultra-high-temperature (UHT) processing holds the milk at a temperature of 140°C (284°F) for four seconds. During UHT processing, milk is sterilized rather than pasteurized. This process allows milk or juice to be stored several months without refrigeration. The process is achieved by spraying the milk or juice through a nozzle into a chamber that is filled with high-temperature steam under pressure. After the temperature reaches 140°C (284°F) the fluid is cooled instantly in a vacuum chamber and packed in a pre-sterilized, airtight container. Milk labelled UHT has been treated in this way.

For more information on pasteurization, visit the International Dairy Foods Association. Attribution

This page titled 8.1.6.2: Milk is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6.2: Milk by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6.3: Milk Products ADD US

Cream

The usual minimum standard for cream is 10% fat content, though it ranges between 10% and 18%. Cream in this range may be sold as half and half, coffee cream, or table cream.

Whipping cream is about 32% to 36% in milk fat content. Cream with 36% or higher is called heavy cream. This percentage of fat is not a mandated standard; much less than this and the cream simply will not whip. For best whipping results, the cream should be 48 to 60 hours old and be cold. A stabilizer, some sugar, and flavour may be added during whipping. Before adding stabilizer, check the ingredients on the carton; some whipping creams nowadays have added agents such as carrageenan, in which case an additional stabilizer may not be necessary.

Canadian cream definitions are similar to those used in the United States, except for that of “light cream.” In Canada, what the U.S. calls light cream is referred to most commonly as half and half. In Canada, “light cream” is low-fat cream, usually with 5% to 6% fat. You can make your own light cream by blending milk with half-and-half.

In Quebec, country cream is sold, which contains 15% milk fat. If you are usinga recipe that calls for country cream, you may substitute 18% cream.

If you have recipes from the UK, you might see references to double cream. This is cream with about 48% milk fat, which is not readily available in Canada, except in some specialty stores. Use whipping cream or heavy cream instead.

Table 1 lists some of the common cream types and their uses.

Table 1 Cream types and fat content Name

	Minimum Milk Fat	Additional Definition	Main Uses
Whipping cream	32%	Heavy cream has at least 36% milk fat	Whips well, can be piped; custards, cream fillings, confectionary products
Table cream	18%	Coffee cream	Added to coffee, poured over puddings, used in sauces
Half-and-half	10%-12%	Cereal cream	Added to coffee; custards and ice cream mixes
Light cream	5%-10%		Added to coffee

Buttermilk

Inoculating milk with a specific culture to sour it
Churning milk and separating the liquid left over from the butter

There are two methods to produce buttermilk:

page214image47224 page214image47384 page214image47544

The second method is where buttermilk gets its name, but today, most of what is commonly called buttermilk is the first type. Buttermilk has a higher acid content than regular milk (pH of 4.6 compared with milk’s pH of 6.6).

The fermented dairy product known ascultured buttermilk is produced from cow’s milk and has a characteristically sour taste caused by lactic acid bacteria. This variant is made using one of two species of bacteria — either Lactococcus lactis or Lactobacillus bulgaricus, which creates more tartness in certain recipes.

The acid in buttermilk reacts with the sodium bicarbonate (baking soda) to produce carbon dioxide, which acts as the leavening agent.

Sour Cream

Sour cream is made from cream soured by adding lactic acids and thickened naturally or by processing. Milk fat content may vary from 5.5% to 14%. The lactic acid causes the proteins in sour cream to coagulate to a gelled consistency; gums and starches may be added to further thicken it. The added gums and starches also keep the liquid whey in sour cream from separating.

Use sour cream in cheesecakes, coffee cakes, and pastry doughs. Low-fat and fat-free sour cream are available. Low-fat sour cream, which is essentially cultured half-and-half or light cream (and usually contains 7% to 10% milk fat), is often satisfactory as a substitute for regular sour cream in baking. These products are higher in moisture and less rich in flavor than regular sour cream.

Crème Fraîche

Crème fraîche (fresh cream) is a soured cream containing 30% to 45% milk fat and having a pH of around 4.5. It is soured with bacterial culture. Traditionally it is made by setting unpasteurized milk into a pan at room temperature, allowing the cream to rise to the top. After about 12 hours, the cream is skimmed off. During that time, natural bacteria in the unpasteurized milk ripens the cream, turning it into a mildly sour, thickened product.

An effective substitute can be made by adding a small amount of cultured buttermilk or sour cream to whipping cream and allowing it to stand in a warm spot for 10 hours or more before refrigerating. As the cream ripens from the growth of the lactic acid bacteria, it thickens and develops a sour flavour. This product is similar to sour cream, but it has a higher milk fat content.

Milk Substitutes

Milk substitutes are becoming increasingly popular as replacements for straight skim milk powders. Innumerable replacement blends are available to the baker. Their protein contents range from 11% to 40%; some are wet, some are dry-blended. Product types vary from all dairy to mostly cereal. All-dairy blends range from mostly dry skim milk to mostly whey. A popular blend is whey mixed with 40% soy flour solids and a small quantity of sodium hydroxide to neutralize the whey acidity.

Dough consistency may be a little softer if the milk in the replacement blend exceeds 3%, and this could dictate the need to increase dough mixing by at least half a minute. However, absorption and formula changes are seldom necessary when switching from dry milk to a blend, or from a blend to a blend.

For nutritional labelling, or when using a blend in a non-standardized product that must carry an itemized ingredient label, all blend components must be listed in their proper order on the label.

The Canadian Food Inspection Agency defines modified milk ingredients as any of the following in liquid, concentrated, dry, frozen, or reconstituted form:

Calcium-reduced skim milk
Casein: This a protein in milk and is used as a binding agent. Caseins are also used in wax to shine fruits and vegetables, as an adhesive, and to fortify bread. Caseins contain common amino acids. Caseinate: This protein is derived from skim milk. Bodybuilders sometimes take powder enriched with calcium caseinate because it releases proteins at an even, measured pace.

page215image32696 page215image32856 page215image33016

Cultured milk products: These are milk products that have been altered through controlled fermentation, including yogurt, sour cream, and cultured buttermilk.
Milk serum proteins
Ultra-filtered milk: The Canadian Food and Drug Regulations define this type of milk as that which “has been subjected to a process in which it is passed over one or more semi-permeable membranes to partially remove water, lactose, minerals, and water-soluble vitamins without altering the whey protein-to-casein ratio and that results in a liquid product.”

Whey: This is serum by-product created in the manufacture of cheese.
Whey butter: Typically oily in composition, whey butter is made from cream separated from whey. Whey cream: This is cream skimmed from whey, sometimes used as a substitute for sweet cream and butter.
Any component of milk that has been altered from the form in which it is found in milk.

Milk Powder

Milk powder is available in several different forms: whole milk, skim milk (non-fat dry milk), buttermilk, or whey. They are all processed similarly: the product is first pasteurized, then concentrated with an evaporator, and finally dried (spray or roller dried) to produce powder.

Whole milk powder must contain no less than 95% milk solids and must not exceed 5% moisture. The milk fat content must be no less than 2.6%. Vitamins A and D may be added and the emulsifying agent lecithin may also be added in an amount not exceeding 0.5%.
Skim milk powder (non-fat dry milk) must contain no less than 95% milk solids and must not exceed 4% moisture or 1.5% fat.
Buttermilk powder must contain no less than 95% milk solids and must not exceed 3% moisture or 6% fat.
Whey powder consists primarily of carbohydrate (lactose), protein (several different whey proteins, mainly lactalbumins and globulins), various minerals, and vitamins. Whey powder is a valuable addition to the functional properties of various foods as well as a source of valuable nutrients because it contains approximately 50% of the nutrients in the original milk.

Table 2 compares the composition of milk and two powdered milk products.

page216image63504

Table 2 Comparison of fresh and powdered milk products (% by weights)

	Whole Milk	Skim Milk Powder (Non-fat dry milk)	Buttermilk Powder
Milk fat	3.25	0.7	5.0
Protein	3.5	36.0	34.0
Milk sugar (lactose)	4.9	51.0	48.0
Minerals	0.8	8.2	7.9
Water	87.0	3.0	3.0
Calcium	0.12	1.3	1.3

To make 10 L (22 lb.) of liquid skim milk from skim milk powder, 9.1 L (2.4 gal.) of water and 900 g (2 lb.) of skim milk powder are required.
To make 10 L (22 lb.) of whole milk from skim milk powder, 8.65 L (2.25 gal) of water, 900 g (2 lb.) of skim milk powder, and 450 g (1 lb.) of butter are needed.

When reconstituting dried milk, add it to the water and whisk in immediately. Delaying this, or adding water to the milk powder, will usually result in clogging. Water temperature should be around 21°C (70°F).

page216image63504 page216image63664

Evaporated Milk

Sometimes called concentrated milk, this includes evaporated whole, evaporated partly skimmed, and evaporated skim milks, depending on the type of milk used in its production. Canadian standards require 25% milk solids and 7.5% milk fat.

All types of evaporated milk have a darker color than the original milk because at high temperatures a browning reaction occurs between the milk protein and the lactose. After 60% of the water is removed by evaporation, the milk is homogenized, cooled, restandardized, and canned. It is then sterilized by heating for 10 to 15 minutes at 99°C to 120°C (210°F to 248°F). Controlled amounts of disodium phosphate and/or sodium citrate preserve the “salt balance” and prevent coagulation of the milk that might occur at high temperatures and during storage.

Sweetened Condensed Milk

Sweetened condensed milk is a viscous, sweet-colored milk made by condensing milk to one-third of its original volume, which then has sugar added. It contains about 40% sugar, a minimum of 8.5% milk fat, and not less than 28% total milk solids.
Attribution

page217image10488 page217image10648 page217image10808

This page titled 8.1.6.3: Milk Products ADD US is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6.3: Milk Products ADD US by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6.4: Milk in bread baking

In the dough stage, milk increases water absorption. Consequently, dough made with milk should come softer from the mixer than dough made with water. Other aspects of milk in yeast doughs include:

Dough may be mixed more intensively.
Milk yields dough with a higher pH compared to water dough, and the fermentation will be slower. Fermentation tolerance (the ability of the dough to work properly in a range of temperatures) will be slightly improved.
Bench time will be extended as the dough ferments more slowly at this stage. (Final proof times will be about the same, as by this time the yeast has adjusted to the condition of the dough.)

Bread made with milk will color faster in the oven and allowance should be made for this. If taken out too early after a superficial examination of crust color, it may collapse slightly and be hard to slice. The loaf should be expected to have a darker crust color than bread made without milk.

In the finished product, milk will make bread that has:

Greater volume (improved capacity to retain gas) Darker crust (due to the lactose in the milk) Longer shelf life (due partly to the milk fat)
Finer and more “cottony” grain

Better slicing due to the finer grain

If skim milk or skim milk powder is used, some of the above benefits will not be so evident (e.g., longer shelf life, which is a result of the fat in the milk).

The type of sugar found in milk, lactose, has little sweetening power and does not ferment, so in dough made with skim milk powder, sugar has to be added or the fermentation will be very slow. While lactose is not fermentable, it caramelizes readily in the oven and produces a healthy crust color. The recommended amount of skim milk powder used in fermented dough is 2% to 8% based on flour, and up 15% in cakes.

Buttermilk and sour milk are used to make variety breads. They have a lower pH and require a shorter fermentation for good results.

This page titled 8.1.6.4: Milk in bread baking is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6.4: Milk in bread baking by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6.5: Yogurt

Yogurt is a thick or semi-solid food made from pasteurized milk fermented by lactic bacteria. The milk coagulates when a sufficient quantity of lactic acid is produced. Yogurt is a rich, versatile food capable of enhancing the flavor and texture of many recipes. It is prepared sweetened or unsweetened, and is used in baking to make yogurt-flavored cream cakes, desserts, and frozen products.

This page titled 8.1.6.5: Yogurt is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6.5: Yogurt by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6.6: Lactose

Lactose is a "milk sugar" and is a complex sugar. It is available commercially spray-dried and in crystalline form. There are many advantages to using it in various baking applications:

Because of its low sweetening value compared to sucrose, it can lend texture and create browning while keeping the sweetness level at low values, which many consumers prefer. It can be used to replace sucrose up to a 50% level, or replace it entirely in products like pie pastry. Lactose improves dough handling properties and the color of the loaf.

In pie crusts, it gives good color to top and bottom crusts, more tender crusts, and retards sogginess. In machine-dropped cookies, lactose can help the dough release better from the die. In cakes and muffins, it gives body without excessive sweetening and improves volume. Lactose binds flavors that are normally volatile and thus intensifies or enhances flavor.

This page titled 8.1.6.6: Lactose is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6.6: Lactose by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.6.7: Cheese

Cheese is a concentrated dairy product made from fluid milk and is defined as the fresh or matured product obtained by draining the whey after coagulation of casein.

Cheese making consists of four steps:

1. Curdling of the milk, either by enzyme (rennet) or by lactic curdling (natural process)
2. Draining in which the whey (liquid part) is drained from the curd (firm part)
3. Pressing, which determines the shape
4. Ripening, in which the rind forms and the curd develops flavor

Cheese can be classified, with some exceptions, into five broad categories, as follows. Examples are given of specific cheeses that may be used in baking.

1. Fresh cheese: High moisture content and no ripening characterize these products. Examples: cottage cheese, baker's cheese, cream cheese, quark, and ricotta.
2. Soft cheeses: Usually some rind, but with a soft interior. Example: feta.
3. Semi-soft cheeses: Unripened cheeses of various moisture content. Example: mozzarella.
4. Firm cheeses: Well-ripened cheese with relatively low moisture content and fairly high fat content. Examples: Swiss, cheddar, brick.
5. Hard cheeses: Lengthy aging and very low moisture content. Example: Parmesan.

In baking, cheeses have different functions. Soft cheeses, mixed with other ingredients, are used in fillings
for pastries and coffeecakes. They are used for certain European deep-fried goods, such as cannoli. They may also be used, sometimes in combination with a richer cream cheese, for cheesecakes. All the cheeses itemized under fresh cheese (see above) are all more or less interchangeable for these functions. The coarser cheese may be strained first if necessary. The firmer cheeses are used in products like cheese bread, quiches, pizza, and cheese straws.

A brief description of the cheeses most likely to be used by bakers follows.

Dry Curd Cottage Cheese

This is a soft, unripened, acid cheese. Pasteurized skim milk is inoculated with lactic-acid-producing bacteria, and a milk-clotting enzyme (rennet) is added. Following incubation, the milk starts to clot, and it is then cut into cubes. After gentle cooking, the cubes or curds become quite firm. At this point, the whey is drained off, and the curd is washed and cooled with cold water.

Creamed Cottage Cheese

Creamed or dressed cottage cheese consists of dry curd cottage cheese combined with a cream dressing. The milk fat content of the dressing determines whether the final product is "regular" (4% milk fat) or low fat (1% to 2% milk fat).

Baker's Cheese

This is a soft, unripened, uncooked cheese. It is made following exactly the same process as for dry curd cottage cheese, up to and including the point when the milk clot is cut into cubes. This cheese is not cooked to remove the whey from the curd. Rather, the curd is drained through cloth bags or it may be pumped through a curd concentrator. The product is then ready to be packaged. The milk fat content is

page221image25480 page221image25640 page221image25800

Chemistry of Cooking 214

generally about 4%.

Quark

Quark (or quarg) is a fresh unripened cheese prepared in a fashion similar to cottage cheese. The mild flavor and smooth texture of quark make it excellent as a topping or filling for a variety of dishes. Quark is similar to baker's cheese, except acid is added to it (it is inoculated with lactic-acid-producing bacteria), and then it is blended with straight cream to produce a smooth spread containing approximately 7% milk fat. Today there are low-fat quarks with lower percentage, and high-fat versions with milk fat adjusted to 18%. Quark cheese can often be used in place of sour cream, cottage cheese, or ricotta cheese.

Cream Cheese

Cream cheese is a soft, unripened, acid cheese. A milk-and-cream mixture is homogenized and pasteurized, cooled to about 27°C (80°F), and inoculated with lactic-acid-producing bacteria. The resulting curd is not cut, but it is stirred until it is smooth, and then heated to about 50°C (122°F) for one hour. The curd is drained through cloth bags or run through a curd concentrator. Regular cream cheese is fairly high fat, but much lighter versions exist now.

Ricotta

Ricotta is a fresh cheese prepared from either milk or whey that has been heated with an acidulating agent added. Traditionally lemon juice or vinegar was used for acidulation, but in commercial production, a bacterial culture is used. The curds are then strained and the ricotta is used for both sweet and savory applications.

Mascarpone

Mascarpone is a rich, fresh cheese that is a relative of both cream cheese and ricotta cheese. Mascarpone is prepared in a similar fashion to ricotta, but using cream instead of whole milk. The cream is acidified (often by the direct addition of tartaric acid) and heated to a temperature of 85°C (185°F), which results in precipitation of the curd. The curd is then separated from the whey by filtration or mechanical means. The cheese is lightly salted and usually whipped. Note that starter culture and rennet are not used in the production of this type of cheese. The high-fat content and smooth texture of mascarpone cheese make it suitable as a substitute for cream or butter. Ingredient applications of mascarpone cheese tend to focus on desserts. The most famous application of mascarpone cheese is in the Italian dessert tiramisu.

Table 1 provides the composition of various types of cheeses.

Table 1 Composition of various cheeses (% by weight)

	Moisture %	Milk Fat %	Salt %
Dry curd cottage cheese	80	0.4	n/a
Regular creamed cottage cheese	79	4	1
Low fat (1% and 2%) creamed cottage cheese	79	1-2	1
Baker's cheese	79	4	1
Quark	72	5-7	n/a
Quark (high fat)	59	18	n/a
Cream cheese	54 (varies)	17- 37	1
Ricotta	72-75	8-13	n/a

page222image68672

Mascarpone	46	60-75	1

This page titled 8.1.6.7: Cheese is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 6.7: Cheese by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.7: EGGS

8.1.7.1: Eggs Grade

8.1.7.2: Composition and Nutrition

8.1.7.3: Egg Products

8.1.7.4: The Function of Eggs

8.1.7.5: Storing Eggs

This page titled 8.1.7: Eggs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 7: Eggs by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.7.1: Eggs Grade

Fresh hen eggs are sold by grade in all provinces. All shell eggs that are imported, exported, or shipped from one province to another for commercial sale must be graded. In Canada, it is mandatory to have all eggs graded by the standards set by Agriculture and Agri-Foods Canada (AAFC). The grade name appears on cartons. The grades Canada A and Canada B bear the maple leaf symbol with the grade name inside, and Canada C and Nest Run eggs will have the grade name printed in block text. The grades indicate the quality of the egg and should not be confused with size. Only Canada A are available in different sizes. The average large size egg weighs about 56 g (2 oz.) as indicated in Table 1.

Table 1: Canada Grade A egg sizes

Size	Weight (including shell)
Peewee	Less than 42 g (1.5 oz.)
Small	At least 42 g (1.5 oz.)
Medium	At least 49 g (1.75 oz.)
Large	At least 56 g (2 oz.)
Extra Large	At least 63 g (2.25 oz.)
Jumbo	70 g (2.5 oz.) or more

The Canada grade symbol does not guarantee that the eggs are of Canadian origin, but it does guarantee that the products meet Canadian government standards. Agriculture Canada inspects all egg-processing plants to ensure that the products are wholesome and processed according to sanitary standards. The pasteurization of “packaged” egg product is also monitored.

The criteria for grading eggs are:

- Weight
- Cleanliness
- Soundness and shape of shell
- Shape and relative position of yolk within the egg Size of air cell free of abnormalities
- Freedom from dissolved yolk and blood spots

Canada A

Canada A eggs are clean, normal in shape with sound shells, and have the finest interior quality. They are ideal for all uses. The yolks are round and compact and surrounded by very thick, firm albumen. Canada A eggs are a premium quality and in limited supply on the retail market. If eggs are not sold within a limited time, unsold stocks are returned to the supplier. Eggs graded as A must meet the minimum weight for the declared size (see Table 12.) The size designation for Canada A eggs must appear on the label.

Canada B

Canada B eggs have very slight abnormalities. This grade is fine for baking, where appearance is not important. These eggs must weigh at least 49 g (1.75 oz.). There are no size designations on the label for Canada B eggs.

Canada C

Canada C is considered a processing grade and provides a safe outlet for the disposition of cracked eggs. Canada C eggs must be shipped to a federally registered processed egg station and pasteurized as a means of controlling the higher risk of salmonella or other microbial contamination that may be found in cracked eggs.

These eggs are suitable for processing into commercially frozen, liquid, and dried egg products. Sizes are not specified.

Canada Nest Run

Since Canada Nest Run eggs are generally sent for further processing, they are usually not washed, candled (a process discussed later in this chapter), or sized. However, nest run eggs must meet the minimum quality requirements prescribed by the egg regulations. This grade, as with other Canada grades, can only be applied to eggs in a federally registered egg station.

This page titled 8.1.7.1: Eggs Grade is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 7.1: Eggs Grade by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.7.2: Composition and Nutrition

Table 1 Composition of eggs by percent of weight. Traces of sugar and ash are also present.

	Composition of Eggs (%) Whole Egg	Composition of Eggs (%) Yolk	Composition of Eggs (%) White
Moisture	73.0	49.0	86.0
Protein	13.3	16.7	11.6
Lipid	11.5	31.6	0.2

Table 2 Nutritional content of a large egg			
	Whole Egg	Yolk	White
Weight	50 g	17 g	33 g
Protein	6g	3g	3g
Fat	5g	5g	Trace
Cholesterol	216 mg	216 mg	0
Calcium	25 mg	2 mg	27 mg
Iron	1.0 mg	0.6 mg	Trace
Sodium	63 mg	7 mg	54 mg
Potassium	60 mg	16 mg	47 mg
Vitamin A	96 RE	99 RE	0 RE

Note: B-complex vitamins, not itemized, are well represented in eggs, as are amino acids. RE = retinol equivalent, a term used in nutritional measurement.

Worth noting is the concentration of certain food elements in different parts of the egg. Note for example that all the cholesterol is in the yolk. The yolk is relatively rich in iron and the white is high in calcium.

In practice, when separating large eggs, one estimates the weight of the white as 30 g (1 oz) and the yolk as 20 g (0.7 oz). The color of the shell, which is either a creamy white or brown, is relevant to the breed of the hen, and there is no other basic difference in the content of the egg or the shell.

The color of the yolk depends on the diet of the hens. Bakers have a preference for eggs with dark yolks. Certainly the appearance of cakes made with such eggs is richer. Tests have found that, although eggs with darker yolks tend to produce moister sponge cakes, the cakes are somewhat coarser and less tender.

This page titled 8.1.7.2: Composition and Nutrition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 7.2: Composition and Nutrition by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.7.3: Egg Products

A number of egg products besides whole shell eggs are used in the baking and food service industry. By law, all egg products other than shell eggs are pasteurized to protect them against salmonella, and the low temperature at which they are kept inhibits bacterial activity, although under certain conditions they may spoil very rapidly.

The chief categories of egg products available are:

Liquid eggs (whole eggs and whole eggs with additional yolks)
Frozen eggs (whole eggs, egg whites, and egg yolks)
Dried and powdered eggs (whole eggs, egg whites, and meringue powder)

Liquid and Frozen Eggs

Liquid and frozen whole eggs are preferred in large bakeries where cracking and emptying of shells is not economical. They are also one of the most economical ways of purchasing eggs. Liquid and frozen whole eggs are sometimes “fortified” by the addition of egg yolks. Some bakers feel that liquid or frozen eggs don’t yield the same volume in sponge cakes as fresh eggs, and there is a certain bias in favor of shell eggs.

If stored in the freezer at -18°C (0°F) or lower, liquid and frozen eggs will keep for long periods with minimum loss of quality. Thawing should take place in the refrigerator or under cold water without submerging the container. Leaving frozen eggs at room temperature to thaw is a bad practice because the outside layers of egg can reach a temperature favorable to bacteria while the centre is still frozen. Heat should never be used to defrost eggs. Unused portions must be refrigerated and used within 24 hours.

Frozen egg yolks consist of 90% egg yolks and 10% sugar to prevent the yolk from gelling and to avoid separation of the fat.

Spray-Dried Whole Eggs and Egg Whites

Dried eggs are used by some bakers as a convenience and cost saver. As with frozen eggs, some bakers doubt their performance in products such as sponge cakes. But dried eggs produce satisfactory results because of the addition of a carbohydrate to the egg before the drying process, usually corn syrup, which results in foaming comparable to fresh eggs.

Dried whole eggs should be stored unopened in a cool place not over 10°C (50°F), preferably in the refrigerator. They are reconstituted by blending 1 kg (2.2 lb.) of powdered whole egg with 3 kg (6.6 lb) of cold water. The water is added slowly while mixing. Once reconstituted, dried eggs should be used immediately or refrigerated promptly and used within an hour.

In mixes such as muffins and cake doughnuts, dried eggs can be mixed in with the other dry ingredients and do not have to be reconstituted. In layer cake formulas, dried eggs are blended with the other dry ingredients before the fat and some water are added, followed by the balance of liquid in two stages.

Spray-dried egg whites are reconstituted by mixing 1 kg (2.2 lb.) of powdered egg white with 1 kg (2.2 lb.) of cold water, letting it stand for 15 minutes, and then adding 9 kg (20 lb.) of cold water. When used in cake mixes, the powdered egg white is blended with the other dry ingredients, but only 7 L (7 qt.) of cold water is used for every 1 kg (2.2 lb.) of powdered egg white.

page228image28800 page228image28960 page228image29120

221 Chemistry of Cooking

Dry Egg Substitutes or Replacements

Egg substitutes are made from sweet cheese, whey, egg whites, dextrose, modified tapioca starch, sodium caseinate, and artificial color and flavor. They are cost-cutters and can be used alone or in combination with fresh or dried eggs in cakes, cookies, and fillings. One kg (2.2 lb.) of powder is mixed with 4 kg (9 lb.) of water to replace powdered eggs.

Meringue Powder

While it is not a pure dehydrated egg white, meringue powder is widely used by bakers to make baked Alaska, royal icing, and

toppings. It contains vegetable gums and starches to absorb moisture and make it whip better.

This page titled 8.1.7.3: Egg Products is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 7.3: Egg Products by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.7.4: The Function of Eggs

Eggs are a truly multifunctional ingredient and have many roles to play in the bakeshop. Their versatility means that product formulas may be adjusted once the properties of eggs are understood. For example, in French butter cream, egg whites may be substituted in the summer for whole eggs to give a more stable and bacteria-free product (egg white is alkaline, with pH 8.5). A yolk or two may be worked into a sweet short paste dough to improve its extensibility. Sponge cake formulas can be adjusted, for example, with the addition of egg yolks in jelly rolls to improve rolling up.

If a recipe is changed by replacing some or all of the eggs with water, two factors must be remembered:

Water replacement is about 75% of the egg content, since egg solids constitute about 25% of the egg.
Leavening ability is lessened and must be made up by the addition of chemical leavening.

Other uses of eggs are:

Leavening: They will support many times their own weight of other ingredients through their ability to form a cell structure either alone or in combination with flour. The egg white in particular is capable of forming a large mass of cells by building a fine protein network.
Moistening and binding: The fat in eggs provides a moistening effect, and the proteins present coagulate when heated, binding ingredients together.

Thickening: Eggs are valuable thickeners in the cooking of chiffon pie fillings and custard. Emulsifying: Lecithin, present in the yolk, is a natural emulsifier and assists in making smooth batters.
Customer appeal: Eggs enhance the appearance of products through their colour and flavour, and they improve texture and grain.

Structure: Eggs bind with other ingredients, primarily flour, creating the supporting structure for other ingredients.
Shelf life: The shelf life of eggs is extended through the fat content of the yolk.
Nutrition: Eggs are a valuable food in every respect. Note, however, that 4% of the lipid in egg yolk is cholesterol, which may be a concern to some people. Developments in poultry feed claim to have reduced or eliminated this cholesterol level.

Tenderizing: The fat in eggs acts like a shortening and improves the tenderness of the baked cake.

Keep these points in mind when using eggs:

Spots in eggs are due to blood fragments in the ovary. Such eggs are edible and may be used.
The albumen or egg white is soluble in cold water, congeals at 70°C (158°F), and remains insoluble from then on.
Cover leftover yolks or whites tightly and refrigerate. Add a little water on top of yolks, or mix in 10% sugar, to prevent crusting.
Do not return unused portions to the master container.
Use clean utensils to dip egg products from their containers.

This page titled 8.1.7.4: The Function of Eggs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 7.4: The Function of Eggs by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.7.5: Storing Eggs

Whole eggs are the perfect medium for the development of bacteria and mould. Eggs with an undesirable odor may be high in bacteria or mould. While some of these odors disappear in baking, some will remain and give an off-taste to the product if the odor is concentrated and strong.

Store fresh eggs in the refrigerator in cartons to prevent moisture loss and absorption of odours. If refrigerator space is at a premium, eggs are stable for up to three weeks if kept at a temperature of 13°C to 15°C (55°F to 60°F). Naturally, this must be in a location with invariable conditions.

Food poisoning can result from using eggs held too long before using. Liquid or cracked eggs should be kept under refrigeration at all times.

Whole eggs can be checked for freshness with the candling or salt water method:

Candling method: Hold the egg up to a light in a darkened room or positioned so that the content or condition of the egg may be seen. If the yolk is held firmly by the white when the egg is turned, and the egg is clean and not broken, then the egg is of good quality. Smell or odor is not readily revealed unless the shell is broken.

Salt water method: Add 100 g of salt to 1 L (3.5 oz. to 1 qt.) of water. Allow to dissolve completely. When an egg is placed in this mixture, its level of buoyancy determines the age of the egg. An old egg will float to the surface, while a fresher egg will sink to the bottom.

This page titled 8.1.7.5: Storing Eggs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 7.5: Storing Eggs by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.8: CHOCOLATE

Add all sections in this chapter

8.1.8.1: From the Cocoa Bean to the Finished Chocolate

8.1.8.2: Chocolate Produced for the Baking Industry

8.1.8.3: Couverture

8.1.8.4: Definitions and Regulations (ADD US)

Contributors and Attributions

Sorangel Rodriguez-Velazquez (American University).

Chemistry of Cooking by Sorangel Rodriguez-Velazquez is licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International License, except where otherwise noted

This page titled 8.1.8: Chocolate is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 8: Chocolate by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.8.1: From the Cocoa Bean to the Finished Chocolate

In North America, chocolate manufacturing started in Massachusetts in 1765. Today, in the factory, the beans get cleaned, and magnets take out metallic parts, and then sand, dust, and other impurities are removed. Some starch will be changed into dextrins in the roasting process to improve flavor. Machines break the beans and grind them fine until a flowing liquid is produced, called chocolate liquor. Through hydraulic pressure, cocoa butter is reduced from 55% to approximately 10% to 24% or less, and the residue forms a solid mass called press cake.

The press cake is then broken, pulverized, cooled, and sifted to produce commercial cocoa powder. The baking industry uses primarily cocoa powders with a low fat content.

At the factory, chocolate is also subject to an additional refining step calcleodnching. Conching has a smoothing effect. The temperature range in this process is between 55°C and 65°C (131°F and 149°F). Sugar interacts with protein to form amino sugars, and the paste losesacids and moisture and becomes smoother.

This video explains the chemical reactions related to heat, melting point, and formation of crystal structures in science360.gov/obj/video/27d9...stry-chocolate

This page titled 8.1.8.1: From the Cocoa Bean to the Finished Chocolate is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 8.1: From the Cocoa Bean to the Finished Chocolate by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.8.2: Chocolate Produced for the Baking Industry

True chocolate contains cocoa butter. The main types of chocolate, in decreasing order of cocoa liquor content, are:

Unsweetened (bitter) chocolate Dark chocolate
Milk chocolate
White chocolate

Unsweetened Chocolate

Unsweetened chocolate, also known as bitter chocolate, baking chocolate, or cooking chocolate, is pure cocoa liquor mixed with some form of fat to produce a solid substance. The pure ground, roasted cocoa beans impart a strong, deep chocolate flavor. With the addition of sugar in recipes, however, it is used as the base for cakes, brownies, confections, and cookies.

Dark (Sweet, Semi-Sweet, Bittersweet) Chocolate

Dark chocolate has an ideal balance of cocoa liquor, cocoa butter, and sugar. Thus it has the attractive, rich color and flavor so typical of chocolate, and is also sweet enough to be palatable. It does not contain any milk solids. It can be eaten as is or used in baking. Its flavor does not get lost or overwhelmed, as in many cases when milk chocolate is used. It can be used for fillings, for which more flavorful chocolates with high cocoa percentages ranging from 60% to 99% are often used.Dark is synonymous with semi-sweet, and extra dark with bittersweet, although the ratio of cocoa butter to solids may vary.

Sweet chocolate has more sugar, sometimes almost equal to cocoa liquor and butter amounts (45% to 55% range).
Semi-sweet chocolate is frequently used for cooking. It is a dark chocolate with less sugar than sweet chocolate.

Bittersweet chocolate has less sugar and more liquor than semi-sweet chocolate, but the two are often interchangeable when baking. Bittersweet and semi-sweet chocolates are sometimes referred to as couverture (see below). The higher the percentage of cocoa, the less sweet the chocolate is.

Milk Chocolate

Milk chocolate is solid chocolate made with milk, added in the form of milk powder. Milk chocolate contains a higher percentage of fat (the milk contributes to this) and the melting point is slightly lower. It is used mainly as a flavoring and in the production of candies and moulded pieces.

White Chocolate

The main ingredient in white chocolate is sugar, closely followed by cocoa butter and milk powder. It has no cocoa liquor. It is used mainly as a flavoring in desserts, in the production of candies and, in chunk form in cookies.

This page titled 8.1.8.2: Chocolate Produced for the Baking Industry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 8.2: Chocolate Produced for the Baking Industry by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.8.3: Couverture

The usual term for top quality chocolate is **couverture**. Couverture chocolate is a very high-quality chocolate that contains extra cocoa butter. The higher percentage of cocoa butter, combined with proper tempering, gives the chocolate more sheen, firmer “snap” when broken, and a creamy mellow flavor. Dark, milk, and white chocolate can all be made as couvertures.

The total percentage cited on many brands of chocolate is based on some combination of cocoa butter in relation to cocoa liquor. In order to be labelled as couverture by European Union regulations, the product must contain not less than 35% total dry cocoa solids, including not less than 31% cocoa butter and not less than 2.5% of dry non-fat cocoa solids. Couverture is used by professionals for dipping, coating, moulding, and garnishing.

What the percentages don’t tell you is the proportion of cocoa butter to cocoa solids. You can, however, refer to the nutrition label or company information to find the amounts of each. All things being equal, the chocolate with the higher fat content will be the one with more cocoa butter, which contributes to both flavor and mouthfeel. This will also typically be the more expensive chocolate, because cocoa butter is more valuable than cocoa liquor.

But keep in mind that just because two chocolates from different manufacturers have the same percentages, they are not necessarily equal. They could have dramatically differing amounts of cocoa butter and liquor, and dissimilar flavors, and substituting one for the other can have negative effects for
your recipe. Determining the amounts of cocoa butter and cocoa liquor will allow you to make informed decisions on chocolate choices.

This page titled 8.1.8.3: Couverture is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 8.3: Couverture by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.8.4: Definitions and Regulations (ADD US)

The legislation for cocoa and chocolate products in Canada is found in Division 4 of the Food and Drug Regulations (FDR), under the Food and Drugs Act (FDA). The Canadian Food Inspection Agency (CFIA) is responsible for administering and enforcing the FDR and FDA. Here are some of the regulations governing cocoa and chocolate:

- Cocoa butter must be the only fat source. Chocolate sold in Canada cannot contain vegetable fats or oils.
- Chocolate must contain chocolate liquor.

The only sweetening agents permitted in chocolate in Canada are listed in Division 18 of the Food and Drug Regulations.

- Artificial sweeteners such as aspartame, sucralose, acesulfame potassium, and sugar alcohols (sorbitol, maltitol, etc.) are not permitted.
- Milk and/or milk ingredients are permissible.
- Emulsifying agents are permissible, as are flavors such as vanilla.

Cocoa butter and sugar quantities are not defined in the regulations. Some semi-sweet chocolate may be sweeter than so-called sweet chocolate. And remember that bittersweet chocolate is not, as you might expect, sugarless. Only if the label states “unsweetened,” do you know that there is no sugar added.

Products manufactured or imported into Canada that contain non-permitted ingredients (vegetable fats or oils, artificial sweeteners) cannot legally be called chocolate when sold in Canada. A non-standardized name such as “candy” must be used.

Finally, lecithin, which is the most common emulsifying agent added to chocolate, is approved for use in chocolate in North America and Europe, but Canadian regulations state that no more than 1% can be added during the manufacturing process of chocolate. Emulsifiers like lecithin can help thin out melted chocolate so it flows evenly and smoothly. Because it is less expensive than cocoa butter at thinning chocolate, it can be used to help lower the cost. The lecithin used in chocolate is mainly derived from soy. Both GMO (genetically modified organism) and non-GMO soy lecithin are available. Check the manufacturer’s packaging and ingredient listing for the source of soy lecithin in your chocolate.

This page titled 8.1.8.4: Definitions and Regulations (ADD US) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 8.4: Definitions and Regulations (ADD US) by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9: SPICES

8.1.9.1: Elements of Taste

8.1.9.2: Introduction to Salt

8.1.9.3: Origins of Salt

8.1.9.4: Functions of Salt in Baking

8.1.9.5: Using Salt in Fermented Doughs

8.1.9.6: Storing Salt

8.1.9.7: Introduction to Spices and Other Flavorings

8.1.9.8: Seasoning and Flavoring

8.1.9.9: Herbs

8.1.9.10: Spices

8.1.9.11: Flavorings in Baking

This page titled 8.1.9: Spices is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9: Spices by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.1: Elements of Taste

Essentially there are a handful of elements that compose all of the taste profiles found in the foods we eat. Western definitions of taste conventionally define four major elements of taste:

- Salty
- Sweet
- Sour
- Bitter

Asian cultures have added the following to the list:

- Umami (literally “pleasant savory taste”)
- Spiciness
- Astringency

Foods and recipes that contain a number of these elements in balance are generally those that we think of as tasting good.

This page titled 8.1.9.1: Elements of Taste is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.1: Elements of Taste by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.2: Introduction to Salt

Historically, salt was a prestigious commodity. “The salt of the earth” describes an outstanding person. The word salary comes from the Latin salaria, which was the payment made to Roman soldiers for the purchase of salt. In Arabic, the phrase translated as “there is salt between us” expresses the covenant between humans and the divine. Though no longer a valuable commodity in the monetary sense, salt is still valuable in the sense of being crucial to human health. Common salt (sodium chloride) is 40% sodium and 60% chloride. An average adult consumes about 7 kg (15 lb.) per year.

Salt can be found deposited in Earth’s layers in rock salt deposits. These deposits formed when the water in the oceans that covered Earth many millions of years ago evaporated. The salt was then covered by various types of rocks.Today, we have three basic methods of obtaining salt from natural sources:

- Mining rock salt
- Extracting salt from salt brines created by pumping water into underground salt deposits
- Evaporating salt water from oceans, seas, and salt lakes

This page titled 8.1.9.2: Introduction to Salt is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.2: Introduction to Salt by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.3: Origins of Salt

Mined Rock Salt

In some countries, salt is mined from salt beds approximately 150 m to 300 m (490 ft. to 985 ft.) below Earth's surface. Sometimes, impurities such as clay make it impossible to use rock salt without purification. Purification makes it possible to get the desired flavor and color, thus making it edible. Edible salt is highly refined: pure and snow white.

Salt from Salt Brines

Salt can also be mined from natural salt beds by using water to extract the salt in the form of a brine, which saves having to construct a mine. Holes are drilled approximately 20 cm (8 in.) in diameter until the salt deposits are reached. A pipe is then driven into the salt beds and another pipe is driven inside the larger pipe further into the deposits. Pressurized water is forced through the outer pipe into the salt beds, and then pumped back out through the smaller pipe to the refineries. Through separation of the impurities, eventually all water in the brine will evaporate, leaving crystallized salt, which then can be dried, sifted, and graded in different sizes.

Ocean, Sea, and Lake Salt

In some countries, especially those with dry and warm climates, salt is recovered straight from the ocean or salt lakes. The salt water is collected in large shallow ponds (also calledsalt gardens) where, through the heat of the sun, the water slowly evaporates. Moving the salt solution from one pond to another until the salt crystals become clear and the water has evaporated eliminates impurities. The salt is then purified, dried completely, crushed, sifted, and graded.

This page titled 8.1.9.3: Origins of Salt is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.3: Origins of Salt by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.4: Functions of Salt in Baking

Salt has three major functions in baking. It affects: (1) Fermentation, (2) Dough, and (3) conditioning Flavor

Fermentation

Fermentation is salt's major function:

- Salt slows the rate of fermentation, acting as a healthy check on yeast development.
- Salt prevents the development of any objectionable bacterial action or wild types of fermentation.
- Salt assists in oven browning by controlling the fermentation and therefore lessening the destruction of sugar.
- Salt checks the development of any undesirable or excessive acidity in the dough. It thus protects against undesirable action in the dough and effects the necessary healthy fermentation required to secure a finished product of high quality.

Dough Conditioning

Salt has a binding or strengthening effect on gluten and thereby adds strength to any flour. The additional firmness imparted to the gluten by the salt enables it to hold the water and gas better, and allows the dough to expand without tearing. This influence becomes particularly important when soft water is used for dough mixing and where immature flour must be used. Under both conditions, incorporating a maximum amount of salt will help prevent soft and sticky dough. Although salt has no direct bleaching effect, its action results in a fine-grained loaf of superior texture. This combination of finer grain and thin cell walls gives the crumb of the loaf a whiter appearance.

Flavor

One of the important functions of salt is its ability to improve the taste and flavor of all the foods in which it is used. Salt is one ingredient that makes bread taste so good. Without salt in the dough batch, the resulting bread would be flat and insipid. The extra palatability brought about by the presence of salt is only partly due to the actual taste of the salt itself. Salt has the peculiar ability to intensify the flavor created in bread as a result of yeast action on the other ingredients in the loaf. It brings out the characteristic taste and flavor of bread and, indeed, of all foods. Improved palatability in turn promotes the digestibility of food, so it can be said that salt enhances the nutritive value of bakery products. The lack of salt or too much of it is the first thing noticed when tasting bread. In some bread 2% can produce a decidedly salty taste, while in others the same amount gives a good taste. The difference is often due to the mineralization of the water used in the dough.

This page titled 8.1.9.4: Functions of Salt in Baking is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.4: Functions of Salt in Baking by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.5: Using Salt in Fermented Doughs

The average amount of salt to use in dough is about 1.75% to 2.25% based on the flour used. Some authorities recommend that the amount of salt used should be based on the actual quantity of water used in making the dough, namely about 30 g per L (1 oz. per qt.) of water.

During the hot summer months, many bakers find it advantageous to use slightly more salt than in the winter as a safeguard against the development of any undesirable changes in the dough fermentation. Salt should never be dissolved in the same water in which yeast is dissolved. It is an antiseptic and dehydrates yeast cells and can even kill part of them, which means that less power is in the dough and a longer fermentation is needed. In bread made by the sponge dough method and in liquid fermentation systems, a small amount of salt included in the first stage strengthens the gluten.

This page titled 8.1.9.5: Using Salt in Fermented Doughs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.5: Using Salt in Fermented Doughs by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.6: Storing Salt

Salt is very stable and does not spoil under ordinary conditions. However, it may have a slight tendency to absorb moisture and become somewhat lumpy and hard. Therefore, it is advisable to store it in a clean, cool, and dry place. Inasmuch as salt can absorb odors, the storage room should be free from any odor that might be taken up and carried by the salt.

This page titled 8.1.9.6: Storing Salt is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.6: Storing Salt by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.7: Introduction to Spices and Other Flavorings

Food touches all of the senses. We taste, we smell, we see color and shape, we feel texture and temperature, and we hear sounds as we eat. All of these elements together create a palette with an infinite number of combinations, but the underlying principles that make food taste good are unchanged.

- Variety and diversity in textures and the elements of taste make for interesting food.
- Contrast is as important as harmony; but avoid extremes and imbalance.

This page titled 8.1.9.7: Introduction to Spices and Other Flavorings is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.7: Introduction to Spices and Other Flavorings by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.8: Seasoning and Flavoring

Many ingredients are used to enhance the taste of foods. These ingredients can be used to provide both seasoning and flavoring.

- Seasoning means to bring out or intensify the natural flavor of the food without changing it. Seasonings are usually added near the end of the cooking period. The most common seasonings are salt, pepper, and acids (such as lemon juice). When seasonings are used properly, they cannot be tasted; their job is to heighten the flavors of the original ingredients.
- Flavoring refers to something that changes or modifies the original flavor of the food. Flavoring can be used to contrast a taste such as adding liqueur to a dessert where both the added flavor and the original flavor are perceptible. Or flavorings can be used to create a unique flavor in which it is difficult to discern what the separate flavorings are. Spice blends used in pumpkin pies are a good example of this.

Knowing how to use seasonings and flavorings skillfully provides cooks and bakers with an arsenal with which they can create limitless flavor combinations. Flavoring and seasoning ingredients include wines, spirits, fruit zests, extracts, essences, and oils. However, the main seasoning and flavoring ingredients are classified as herbs and spices.

Knowing the difference between herbs and spices is not as important as knowing how to use seasonings and flavorings skillfully. In general, fresh seasonings are added late in the cooking process while dry ones tend to be added earlier. It is good practice to under-season during the cooking process and then add more seasonings (particularly if you are using fresh ones) just before presentation. This is sometimes referred to as “layering.” When baking, it is difficult to add more seasoning at the end, so testing recipes to ensure the proper amount of spice is included is a critical process.

This page titled 8.1.9.8: Seasoning and Flavoring is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.8: Seasoning and Flavoring by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.9: Herbs

Herbs tend to be the leaves of fragrant plants that do not have a woody stem. Herbs are available fresh or dried, with fresh herbs having a more subtle flavor than dried. You need to add a larger quantity of fresh herbs (up to 50% more) than dry herbs to get the same desired flavor. Conversely, if a recipe calls for a certain amount of fresh herb, you would use about one-half of that amount of dry herb.

The most common fresh herbs are basil, coriander, marjoram, oregano, parsley, rosemary, sage, tarragon, and thyme. Fresh herbs should have a clean, fresh fragrance and be free of wilted or brown leaves. They can be kept for about five days if sealed inside an airtight plastic bag. Fresh herbs are usually added near the completion of the cooking process so flavors are not lost due to heat exposure.

Dried herbs lose their power rather quickly if not properly stored in airtight containers. They can last up to six months if properly stored. Dried herbs are usually added at the start of the cooking process as their flavor takes longer to develop than fresh herbs.

This page titled 8.1.9.9: Herbs is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.9: Herbs by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.10: Spices

Spices are aromatic substances obtained from the dried parts of plants such as the roots, shoots, fruits, bark, and leaves. They are sold as seeds, blends of spices, whole or ground spices, and seasonings. The aromatic substances that give a spice its particular aroma and flavor are the essential oils. The flavor of the essential oil or flavoring compound will vary depending on the quality and freshness of the spice.

The aromas of ground spices are volatile. This means they lose their odor or flavoring when left exposed to the air for extended periods. They should be stored in sealed containers when not in use. Whole beans or unground seeds have a longer shelf life but should also be stored in sealed containers.

Allspice

Allspice is only one spice, yet it has a flavor resembling a blend of cloves, nutmeg, and cinnamon. At harvest time, the mature (but still green) berries from the allspice trees (a small tropical evergreen) are
dried in the sun. During drying they turn reddish-brown and become small berries. The berries are about
0.6 cm (1/4 in.) in diameter and contain dark brown seeds.

Allspice is grown principally in Jamaica and to a lesser degree in Mexico. Allspice is available whole or ground. Bakers usually use ground allspice in cakes, cookies, spices, and pies.

Anise

Anise is the small, green-grey fruit or seed of a plant of the parsley family. The plant grows to a height of 45 cm (18 in.) and has fine leaves with clusters of small white flowers. It is native to Mexico and Spain, with the latter being the principal producer. Anise seeds are added to pastries, breads, cookies, and candies.

Caraway

Caraway is the dried fruit or seed of a biennial plant of the parsley family, harvested every second year, primarily in the Netherlands. It is also produced in Poland and Russia. The many-branched, hollow-stemmed
herb grows up to 60 cm (24 in.) high and has small white flowers. Caraway is a small crescent-shaped brown seed with a pleasant aroma but somewhat sharp taste. Although it is most familiar in rye bread, caraway is also used in cookies and cakes.

Cardamom

Native to India, Sri Lanka, and Guatemala, cardamom is the fruit or seed of a plant of the ginger family. The three-sided, creamy-white, flavorless pod holds the tiny aromatic, dark brown seeds. It is available in whole and ground (pod removed). Cardamom in ground form flavors Danish pastries and coffee cakes, Christmas baking, and Easter baking such as hot cross buns.

Cinnamon

Cinnamon comes from the bark of an aromatic evergreen tree. It is native to China, Indonesia, and Indochina. Cinnamon may be purchased in ground form or as cinnamon sticks. Ground cinnamon is used in pastries, breads, puddings, cakes, candy, and cookies. Cinnamon sticks are used for preserved fruits and flavoring puddings. Cinnamon sugar is made with approximately 50 g (2 oz.) of cinnamon to 1 kg (2.2 lb.) of granulated sugar.

Cassia

Cassia, sometimes known as Chinese cinnamon, is native to Assam and Myanmar. It is similar to cinnamon but a little darker with a sharper taste. It is considered better for savory rather than sweet foods. It is prized in Germany and some other countries as a

flavor in chocolate.

Cloves

Cloves are the dried, unopened buds of a tropical evergreen tree, native to Indonesia. The flavor is characterized by a sweet, pungent spiciness. The nail-shaped whole cloves are mainly used in cooking, but
the ground version of this spice heightens the flavor of mincemeat, baked goods, fruit pies, and plum pudding.

Ginger

Ginger is one of the few spices that grow below the ground. It is native to southern Asia but is now imported from Jamaica, India, and Africa. The part of the ginger plant used is obtained from the root. Ground ginger is the most commonly used form in baking — in fruitcakes, cookies, fruit pies, and gingerbread. Candied ginger is used in pastries and confectionery.

Mace

Originating in the East and West Indies, mace is the fleshy growth between the nutmeg shell and outer husk, yellow-orange in color. It is usually sold ground, but sometimes whole mace (blades of mace) is available. Mace is used in pound cakes, breads, puddings, and pastries.

Nutmeg

Nutmeg is the kernel or seed of the nutmeg fruit. The fruit is similar to the peach. The fleshy husk, grooved on one side, splits, releasing the deep-brown aromatic nutmeg. It is available whole or ground. Ground nutmeg is used extensively in custards, cream puddings, spice cakes, gingerbread, and doughnuts.

Poppy Seed

Poppy seed comes from the Netherlands and Asia. The minute, blue-grey, kidney-shaped seeds are so small
they seem to be round. Poppy seeds are used in breads and rolls, cakes and cookies, and fillings for pastries.

Sesame or Benne Seed

Sesame or benne seeds are the seeds of the fruit of a tropical annual herb grown in India, China, and Turkey. The seeds are tiny, shiny, and creamy-white with a rich almond-like flavor and aroma. Bakers use sesame seeds in breads, buns, coffee cakes, and cookies.

Vanilla

The Spaniards named vanilla. The word derives from vaina, meaning pod. Vanilla is produced from an orchid-type plant native to Central America. The vanilla beans are cured by a complicated process, which helps explain the high cost of genuine vanilla. The cured pods should be black in color and packed in airtight boxes. Imitation vanilla extracts are made from a colorless crystalline synthetic compound called vanillin. Pure vanilla extract is superior to imitation vanilla. Artificial vanilla is more intense than real vanilla by a factor of 3 to 4 and must be used sparingly.

To use vanilla beans, split the pod down the middle to scrape out the seeds. The seeds are the flavoring agents. Alternatively, the split pod can be simmered in the milk or cream used in dessert preparation. Its flavoring power is not spent in one cooking and it can be drained, kept frozen, and reused. A vanilla bean kept in a container of icing sugar imparts the flavor to the sugar, all ready for use in cookies and cakes.

Vanilla extract is volatile at temperatures starting at 138°C (280°F) and is therefore not ideal for flat products such as cookies. It is suitable for cakes, where the interior temperature does not get so high.

Vanilla beans and vanilla extract are used extensively by bakers to flavor a wide range of desserts and other items.

This page titled 8.1.9.10: Spices is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.10: Spices by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.1.9.11: Flavorings in Baking

Flavors cannot be considered a truly basic ingredient in bakery products but are important in producing the most desirable products. Flavoring materials consist of:

- Extracts or essences
- Emulsions
- Aromas
- Spices

Note: Salt may also be classed as a flavoring material because it intensifies other flavors.

These and others (such as chocolate) enable the baker to produce a wide variety of attractively flavored pastries, cakes, and other bakery products. Flavor extracts, essences, emulsions, and aromas are all solutions of flavor mixed with a solvent, often ethyl alcohol.

The flavors used to make extracts and essences are the extracted essential oils from fruits, herbs, and vegetables, or an imitation of the same. Many fruit flavors are obtained from the natural parts (e.g., rind of lemons and oranges or the exterior fruit pulp of apricots and peaches). In some cases, artificial flavor is added to enhance the taste, and artificial coloring may be added for eye appeal. Both the Canadian and U.S. departments that regulate food restrict these and other additives. The flavors are sometimes encapsulated in corn syrup and emulsifiers. They may also be coated with gum to preserve the flavor compounds and give longer shelf life to the product. Some of the most popular essences are compounded from both natural and artificial sources. These essences have the true taste of the natural flavors.

Aromas are flavors that have an oil extract base. They are usually much more expensive than alcoholic extracts, but purer and finer in their aromatic composition. Aromas are used for flavoring delicate creams, sauces, and ice creams.

Emulsions are homogenized mixtures of aromatic oils and water plus a stabilizing agent (e.g., vegetable gum). Emulsions are more concentrated than extracts and are less susceptible to losing their flavor in the oven. They can therefore be used more sparingly.

This page titled 8.1.9.11: Flavorings in Baking is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Sorangel Rodriguez-Velazquez via source content that was edited to the style and standards of the LibreTexts platform.

- 9.11: Flavorings in Baking by Sorangel Rodriguez-Velazquez is licensed CC BY-NC-SA 4.0. Original source: https://chem.libretexts.org//@api/deki/files/122891/Chemistry-of-Cooking.pdf.

8.2: AQUATIC CHEMISTRY

Water, a natural occurring and abundant substance that exists in solid, liquid, and gas forms on the planet Earth, has attracted the attention of artists, engineers, poets, writers, philosophers, environmentalists, scientists, and politicians. Every aspect of life involves water as food, as a medium in which to live, or as the essential ingredient of life. The food-science aspects of water range from agriculture, aquaculture, biology, biochemistry, cookery, microbiology, nutrition, photosynthesis, power generation, to zoology. Even in the narrow sense of food technology, water is intimately involved in the production, washing, preparation, manufacture, cooling, drying, and hydration of food. Water is eaten, absorbed, transported, and utilized by cells. Facts and data about water are abundant and diverse. This article can only selectively present some fundamental characteristics of water molecules and their collective properties for readers when they ponder food science at the molecular level.

8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems

8.2.2: 2. Carbonate Equilibria in Natural Waters

8.2.3: 3. Redox Equilibria in Natural Waters

8.2.4: 4. Solids in Contact With Natural Waters

8.2.5: Fundamental Characteristics of Water

8.2.6: Natural Water

8.2.7: Water Biology

8.2.8: Water Chemistry

8.2.9: Water Physics

8.2.10: Water Treatment

This page titled 8.2: Aquatic Chemistry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems

Natural waters contain a wide variety of solutes that act together to determine the pH, which typically ranges from 6 to 9. Some of the major processes that affect the acid-base balance of natural systems are:

- Contact with atmospheric carbon dioxide
- Input of acidic gases from volcanic and industrial emissions
- Contact with minerals, rocks, and clays
- Presence of buffer systems such as carbonate, phosphate, silicate, and borate
- Presence of acidic anions, such as $Fe(H_2O)_6^{3+}$
- Input and removal of CO_2 through respiration and photosynthesis
- Other biological processes, such as oxidation ($O_2 + H^+ + e^- \rightarrow H_2O$), nitrification, denitrification, and sulfate reduction.

In this chapter and also in the next one which deals specifically with the carbonate system, we will consider acid-base equilibria as they apply to natural waters. We will assume that you are already familiar with such fundamentals as the Arrhenius and Brfinsted concepts of acids and bases and the pH scale. You should also have some familiarity with the concepts of free energy and activity. The treatment of equilibrium calculations will likely extend somewhat beyond what you encountered in your General Chemistry course, and considerable emphasis will be placed on graphical methods of estimating equilibrium concentrations of various species.

Contributors and Attributions

Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

This page titled 8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.2: 2. Carbonate Equilibria in Natural Waters

http://www.chem1.com/acad/webtext/pdf/c3carb.pdf

This page titled 8.2.2: 2. Carbonate Equilibria in Natural Waters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.3: 3. Redox Equilibria in Natural Waters

http://www.chem1.com/acad/webtext/pdf/c3redox.pdf

Back to Advanced Aquatic Chemistry

Back to chem1 (Stephen Lower/Simon Frasier University)

This page titled 8.2.3: 3. Redox Equilibria in Natural Waters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.4: 4. Solids in Contact With Natural Waters

http://www.chem1.com/acad/webtext/pdf/c3solids.pdf

This page titled 8.2.4: 4. Solids in Contact With Natural Waters is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.5: Fundamental Characteristics of Water

I. Introduction

Water, a natural occurring and abundant substance that exists in solid, liquid, and gas forms on the planet Earth, has attracted the attention of artists, engineers, poets, writers, philosophers, environmentalists, scientists, and politicians. Every aspect of life involves water as food, as a medium in which to live, or as the essential ingredient of life. The food-science aspects of water range from agriculture, aquaculture, biology, biochemistry, cookery, microbiology, nutrition, photosynthesis, power generation, to zoology. Even in the narrow sense of food technology, water is intimately involved in the production, washing, preparation, manufacture, cooling, drying, and hydration of food. Water is eaten, absorbed, transported, and utilized by cells. Facts and data about water are abundant and diverse. This article can only selectively present some fundamental characteristics of water molecules and their collective properties for readers when they ponder food science at the molecular level.

The physics and chemistry of water is the backbone of engineering and sciences. The basic data for the properties of pure water, which are found in the CRC Handbook of Chemistry and Physics (1), are useful for food scientists. However, water is a universal solvent, and natural waters contain dissolved substances present in the environment. All solutes in the dilute solutions modify the water properties. Lang's Handbook of Chemistry (2) gives solubilities of various gases and salts in water. Water usage in the food processing industry is briefly described in the Nalco Water Handbook (3). For water supplies and treatments, the Civil Engineering Handbook (4) provides practical guides. The Handbook of Drinking Water Quality (5) sets guidelines for waters used in food services and technologies. Wastewater from the food industry needs treatment, and the technology is usually dealt with in industrial chemistry (6). Most fresh food contains large amounts of water. Modifying the water content of foodstuffs to extend storage life and enhance quality is an important and widely used process (7).

A very broad view and deep insight on water can be found in "Water – A Matrix of Life" (8). Research leading to our present-day understanding of water has been reviewed in the series "Water – A Comprehensive Treatise" (9). The interaction of water with proteins (10, 11) is a topic in life science and food science. Water is the elixir of life and H_2O is a biomolecule.

II. Water and Food Technology

Water is an essential component of food (12). Philosophical conjectures abound as to how Earth evolved to provide the mantle, crust, atmosphere, hydrosphere, and life. Debates continue, but some scientists believe that primitive forms of life began to form in water (13). Complicated life forms developed, and their numbers grew. Evolution produced anaerobic, aerobic and photosynthetic organisms. The existence of abundant life forms enabled parasites to appear and utilize plants and other organisms. From water all life began (14). Homo sapiens are integral parts of the environment, and constant exchange of water unites our internal space with the external.

The proper amount of water is also the key for sustaining and maintaining a healthy life. Water transports nutrients and metabolic products throughout the body to balance cell contents and requirements. Water maintains biological activities of proteins, nucleotides, and carbohydrates, and participates in hydrolyses, condensations, and chemical reactions that are vital for life (15). On average, an adult consumes 2 to 3 L of water: 1-2 L as fluid, 1 L ingested with food, and 0.3 L from metabolism. Water is excreted via the kidney, skin, lung, and anus (16). The amount of water passing through us in our lifetimes is staggering.

Aside from minute amounts of minerals, food consists of plant and animal parts. Water is required for cultivating, processing, manufacturing, washing, cooking and digesting food. During or after eating, a drink, which consists of mostly water, is a must to hydrate or digest the food. Furthermore, water is required in the metabolic process.

Cells and living organisms require, contain and maintain a balance of water. An imbalance of water due to freezing, dehydration, exercise, overheating, etc. leads to the death of cells and eventually the whole body. Dehydration kills far more quickly than starvation. In the human body, water provides a medium for the transportation, digestion and metabolism of food in addition to many other physiological functions such as body temperature regulation (17).

Two-thirds of the body mass is water, and in most soft tissues, the contents can be as high as 99% (16). Water molecules interact with biomolecules intimately (9); it is part of us. Functions of water and biomolecules collectively manifest life. Water is also required for running households, making industrial goods, and generating electric power.

Water has shaped the landscape of Earth for trillions of years, and it covers 70% of the Earth's surface. Yet, for food production and technology it is a precious commodity. Problems with water supply can lead to disaster (5). Few brave souls accept the challenge to stay in areas with little rainfall. Yet, rainfall can be a blessing or a curse depending on the timing and amount. Praying for timely and bounty rainfall used to be performed by emperors and politicians, but water for food challenges scientists and engineers today.

III. Water Molecules and Their Microscopic Properties

Plato hypothesized four *primal substances*: water, fire, earth, and air. His doctrine suggested that a combination and permutation of various amounts of these four *primal substances* produced all the materials of the world. Scholars followed this doctrine for 2000 years, until it could not explain experimental results. The search of fundamental substances led to the discovery of hydrogen, oxygen, nitrogen, etc., as chemical elements. Water is made up of hydrogen (H) and oxygen (O). Chemists use H_2O as the universal symbol for water. The molecular formula, H_2O, implies that a water molecule consists of two H atoms and one O atom. However, many people are confused with its other chemical names such as hydrogen oxide, dihydrogen oxide, dihydrogen monoxide, etc.

A. Isotopic Composition of Water

The discoveries of electrons, radioactivity, protons, and neutrons implied the existence of isotopes. Natural isotopes for all elements have been identified. Three isotopes of hydrogen are protium (^1H), deuterium (D, ^2D or ^2H), and radioactive tritium (T, ^3T or ^3H), and the three stable oxygen isotopes are ^{16}O, ^{17}O, and ^{18}O. The masses and abundances of these isotopes are given in Table 1. For radioactive isotopes, the half-lives are given.

Table 1. Isotopes of hydrogen and oxygen; and isotopic water molecules molar mass (amu), relative abundance (%) or half life

Isotopes of hydrogen			**Stable isotopes of oxygen**		
^1H	**^2D**	**^3T**	**^{16}O**	**^{17}O**	**^{18}O**
1.007825	2.00141018	3.0160493	15.9949146	16.9991315	17.9991604
99.985%	0.015%	12.33 years	99.762%	0.038%	0.200%
Isotopic water molecules molar mass (amu) and relative abundance (%, ppm or trace)					
$H_2{}^{16}O$	**$H_2{}^{18}O$**	**$H_2{}^{17}O$**	**$HD^{16}O$**	**$D_2{}^{16}O$**	**$HT^{16}O$**
18.010564 99.78%	20.014810 0.20%	19.014781 0.03%	19.00415 0.0149%	19.997737 0.022 ppm	20.018789 trace

Random combination of these isotopes gives rise to the various isotopic water molecules, the most abundant one being $^1H_2{}^{16}O$ (99.78%, its mass is 18.010564 atomic mass units (amu)). Water molecules with molecular masses about 19 and 20 are present at some fractions of a percent. Although $HD^{16}O$ (0.0149%) is much more abundant than $D_2{}^{16}O$ (heavy water, 0.022 part per

million), D_2O can be concentrated and extracted from water. In the extraction process, HDO molecules are converted to D_2O due to isotopic exchange. Rather pure heavy water (D_2O) is produced on an industrial scale especially for its application in nuclear technology, which provides energy for the food industry.

A typical mass spectrum for water shows only mass-over-charge ratio of 18 and 17 respectively for H_2O^+ and OH^+ ions in the gas phase. Other species are too weak for detection, partly due to condensation of water in mass spectrometers.

The isotopic composition of water depends on its source and age. Its study is linked to other sciences (18). For the isotopic analysis of hydrogen in water, the hydrogen is reduced to a hydrogen gas and then the mass spectrum of the gas is analyzed. For isotopes of oxygen, usually the oxygen in H_2O is allowed to exchange with CO_2, and then the isotopes of the CO_2 are analyzed. These analyses are performed on archeological food remains and unusual food samples in order to learn their origin, age, and history.

B. Structure and Bonding of Water Molecules

Chemical bonding is a force that binds atoms into a molecule. Thus, chemists use H-O-H or HOH to represent the bonding in water. Furthermore, spectroscopic studies revealed the H–O–H bond angle to be 104.5^{o} and the H–O bond length to be 96 picometers (pm) for gas H_2O molecules (19). For solid and liquid, the values depend on the temperature and states of water. The bond length and bond angles are fundamental properties of a molecule. However, due to the vibration and rotational motions of the molecule, the measured values are average or equilibrium bond lengths and angles.

The mean van der Waals diameter of water has been reported as nearly identical with that of isoelectronic neon (282 pm). Some imaginary models of the water molecule are shown in Fig. 1

Text Box: Fig. 1. Some imaginative models of water molecules

An isolated water molecule is hardly static. It constantly undergoes a vibration motion that can be a combination of any or all of the three principle modes: symmetric stretching, asymmetric stretching, and bending (or deformation). These vibration modes are indicated in Fig. 2.

Text Box: Fig. 2. The three principle vibration modes of the water molecule, H2O: n1, symmetric stretching, n2, bending, and n3, asymmetric stretching.

Absorption of light (photons) excites water molecules to higher energy levels. Absorption of photons in the infrared (IR) region excites the vibration motion. Photons exciting the symmetric stretching, bending, and asymmetric stretching to the next higher energy levels have wave numbers 3656, 1594, and 3756 cm^{-1} respectively, for H_2O (20). These values and those for other water molecules involving only ^{16}O are given in Table 2

Table 2. Absorption frequencies of D2O, H2O, and HDO molecules for the excitation of fundamental modes to a higher energy level.
Vibration mode

Symmetric stretching	3656	2726	2671
Bending	1594	1420	1178
Asymmetric stretching	3756	3703	2788

The spectrum of water depends on temperature and density of the gaseous H_2O. A typical IR spectrum for the excitation of only the fundamental vibration modes consists of three peaks around 1594, 3656 and 3756 cm^{-1}. Additional peaks due to excitation to mixed modes appear at higher wave numbers.

Rotating the H_2O molecule around the line bisecting the HOH angle by 180° (360°/2) results in the same figure. Thus, the

molecules have a 2-fold rotation axis. There are two mirror planes of symmetry as well. The 2-fold rotation and mirror planes give the water molecules the symmetry point group C_{2v}.

Rutherford alpha scattering experiment in 1909 showed that almost all atomic mass is in a very small atomic nucleus. In a neutral atom the number of protons in the nucleus is the same as the number of electrons around the nucleus. A proton and an electron have the same amount, but different kind of charge. Electrons occupy nearly all of the atomic volume, because the radius of an atom is 100,000 times that of the nucleus.

Electrons, in quantum mechanical view, are waves confined in atoms, and they exist in several energy states called orbitals. Electrons in atoms and molecules do not have fixed locations or orbits. Electron states in an element are called electronic configurations, and their designation for H and O are $1s^1$, and $1s^2 2s^2 2p^4$, respectively. The superscripts indicate the number of electrons in the orbitals 1s, 2*s* or 2*p*. The electronic configuration for the inert helium (He) is $1s^2$, and $1s^2$ is a stable core of electrons. Bonding or valence electrons are $1s^1$ and $2s^2 2p^4$ for H and O respectively.

Text Box: Fig. 3. A tetrahedral arrangement of the Lewis dot-structure and charge distribution of the H2O molecule.

The valence bond approach blends one 2*s* and three 2*p* orbitals into four bonding orbitals, two of which accommodate two electron pairs. The other two orbitals have only one electron each, and they accommodate electrons of the H atoms bonded to O, thus forming the two H–O bonds. An electron pair around each H atom and four electron pairs around the O atom contribute stable electronic configurations for H and O, respectively. The Lewis dot-structure, Fig. 3, represents this simple view. The two bonding and two lone pairs are asymmetrically distributed with major portions pointing to the vertices of a slightly distorted tetrahedron in 3-dimensional space. The two lone pairs mark slightly negative sites and the two H atoms are slightly positive. This charge distribution around a water molecule is very important in terms of its microscopic and macroscopic, chemical and physical properties described later. Of course, the study of water continues and so does the evolution of bonding theories. Moreover, the distribution of electrons in a single water molecule is different from those of dimers, clusters, and bulk water.

The asymmetric distribution of H atoms and electrons around the O atom results in positive and negative sites in the water molecule. Thus, water consists of polar molecules.

Text Box: Fig. 4. Separation of a positive and a negative charge q at a distance d results in a dipole, m = d q

The **dipole moment, *m*,** is a measure of polarity and a useful concept. A pair of opposite charge, *q*, separated by a distance, *d*, has a dipole moment of $m = d\,q$ with the direction pointing towards the positive charge as shown in Fig. 4.

The dipole moment of individual water molecules is 6.187×10^{-30} C m (or 1.855 D) (21). This quantity is the vector resultant of two dipole moments due to the O–H bonds. The bond angle H–O–H of water is 104.5^{o}. Thus, the dipole moment of an O–H bond is 5.053×10^{-30} C m. The bond length between H and O is 0.10 nm, and the partial charge at the O and the H is therefore $q = 5.053 \times 10^{-20}$ C, 32 % of the charge of an electron (1.6022×10^{-19} C). Of course, the dipole moment may also be considered as separation of the electron and positive charge by a distance 0.031 nm.

It should be pointed out that the dipole moments of liquid and solid water appear to be higher due to the influence of neighboring molecules. For the liquid and solid, macroscopic properties need be considered.

C. Hydrogen Bonds

Text Box: Fig. 5. Two possible structures of the water dimer, (H2O)2

Attraction among water molecules is more than polar-polar in nature. The O atoms are small and very electronegative. As a result, the positive H atoms (protons) are very attractive to the negative O atoms of neighboring molecules. This O- - -H – O strong attraction is called a **hydrogen bond**, a concept popularized by L. Pauling (22). Furthermore, hydrogen atoms bonded to atoms of N and F, neighboring elements of O in the periodic table, are positive, and they form hydrogen bonds with atoms of N, O, or F. The strength of hydrogen bonds depends on the X–H - - - Y (X or Y are N, O, or F atoms) distances and angles; the shorter the distances, the stronger are the hydrogen bonds.

When two isolated water molecules approach each other, a dimer is formed due to hydrogen bonding. The dimer may have one or two hydrogen bonds. Dimers exist in gaseous and liquid water. When more water molecules are in close proximity, they form trimmers, tetramers and clusters. Hydrogen bonds are not static, they exchange protons and partners constantly. Hydrogen

bonding is a prominent feature in the structures of various solid phases of water usually called ice as we shall see later.

Text Box: Fig. 6. Hydrogen-bond and the van der Waal sphere of two water molecules, by permission from M.F. Martin (23), sbu.ac.uk/water/molecule.html

Water molecules not only form hydrogen bonds among themselves, they form hydrogen bonds with any molecule that contains N–H, O–H and F–H bonds. Foodstuffs such as starch, cellulose, sugars, proteins, DNA, and alkaloids contain N–H and O–H groups, and these are both H-donors and H-acceptors of hydrogen bonds of the type N- - H–O, O- -H–N, N- -H–N, etc. A dimer depicting the hydrogen bond and the van der Waal sphere of two molecules is shown in Fig. 6 (23).

Carbohydrates (starch, cellulose and sugars) contain H–C–O–H groups. The O–H groups are similar to those of water molecules, and they are H-acceptors and H-donors for hydrogen bonds. Proteins contain O–H, R–NH_2 or R_2>NH groups, and the O–H and N–H groups are both H-donors and H-acceptors for the formation of hydrogen bonds. Thus, water molecules have intimate interactions with carbohydrates and proteins.

IV. Macroscopic Properties of Water

Text Box: Table 3. Thermodynamic constants for phase transitions of H2O (molar mass, 18.015268 g mol–1) at 101.314 kPa pressure. Fusion - - - - Vaporization - - - - (melting) Boiling Sublimation Temperature K 273.15 373.15 273.16 DH kJ mol-1 6.01 40.66 51.06 DS J mol-1 K-1 22.00 108.95 186.92 DE kJ mol-1 6.01 37.61 48.97 DV L mol-1 -1.621 30.10 - Energy change of phase transition: DH, enthalpy; DS, entropy; DE, internal energy. DV: volume change of phase transition.

Collectively, water molecules exist as gas, liquid, or solid depending on the temperature and pressure. These phases of water exhibit collective or macroscopic properties such as phase transitions, crystal structures, liquid structures, vapor pressures, and volume-pressure relationships of vapor. In addition, energies or enthalpies for melting, vaporization and heating are also important for applications in food technology.

Thermodynamic constants for phase transitions given in Table 3 are those of pure water. Natural waters, of course, contain dissolved air, carbon dioxide, organic substances, microorganisms, and minerals. Water in food or used during food processing usually contains various organic and inorganic substances. These solutes modify the properties of water and caution should be taken to ensure proper values are applied in food technology.

The triple point and boiling points of water are defined as 273.16 and 373.16 K (kelvin) in the SI unit of temperature, respectively. Thus, the temperature differences can be in units of K or oC.

Water has many unusual properties due to its ability to form hydrogen bonds and its large dipole moment. As a result, the melting, boiling and critical points for water are very high compared to substances of similar molar masses. In general, the higher the molar masses, the higher are the melting and boiling points of the material. Associated with these properties are its very large heat of melting, heat capacity, heat of vaporization and heat of sublimation. Moreover, its surface tension and viscosity are also very large. Thermodynamic energies, and volume changes for phase transitions of H_2O are summarized in Table 3. These data are mostly taken from the Encyclopedia of Chemical Technology, Vol. 25 (1991) (24).

A. Crystal Structures and Properties of Ice

Text Box: Table 4. Properties of Ice at 273.15 K Heat of formation DH f 292.72 kJ mol-1 Density 0.9168 g cm-3 Heat capacity 2.06 J g-1 K-1 DHfusion 6.01 kJ mol-1 Dielectric constant at 3 kHz 79 Thermal expansion coefficient Volumetric 120´10-6 cm3 g-1 K-1 Linear 52.7´10-6 cm g-1 K-1

Hydrogen bonding is prominent in the crystal structures of various solid phases of H_2O. The triple point of water is at 273.16 K and 4.58 torr (611 Pa). The melting point at 1.00 atm (760 torr or 101.325 kPa) is used to define the Kelvin scale as 273.15 K. When water freezes at these temperatures and atmospheric pressure or lower, the solids are hexagonal ice crystals usually designated as I_h. Properties of I_h are given in Table 4. Snowflakes have many shapes because their growth habit depends on temperature and vapor pressure, but they all exhibit hexagonal symmetry, due to the hexagonal structure of ice (25).

However, from a geometric point of view, the same bonding may also be arranged to have cubic symmetry. The existence of cubic ice has been confirmed. When water vapor deposits onto a very cold, 130 – 150 K, surface or when small droplets are cold under low pressure at high altitude, the ice has a cubic symmetry usually designated as I_c. At still higher pressures, different crystal forms designated as ice II, III, IV, ... etc., up to 13 phases of cubic, hexagonal, tetragonal, monoclinic, and orthorhombic symmetries have been identified (26). The polymorphism of solid water is very complicated. Some of these ice

forms are made under very high pressures, and water crystallizes into solid at temperatures above the normal melting or even boiling temperatures. Ice VII is formed above 10 G Pa (gigapascal) at 700 K (26).

When liquid water is frozen rapidly, the molecules have little chance of arranging into crystalline ice. The frozen liquid is called **amorphous ice** or **glassy ice**.

Text Box: Fig. 7. Ordered crystal structure of ice, bond distance: O–H, 100 pm; H- - -O, 175 pm; and O–H- - -O, 275 pm.

The basic relationships between nearest neighboring water molecules are the same in both I_h and I_c. All O atoms are bonded to four other O atoms by hydrogen bonds, which extend from an oxygen atom towards the vertices of a tetrahedron. A sketch of the crystal structure of hexagonal I_h is shown in Fig. 7 (27). In I_h, hydrogen positions are somewhat random due to thermal motion, disorder and exchanges. For example, the hydrogen may shift between locations to form H_3O^+ and OH^- ions dynamically throughout the structure. In this structure, bond angles or hydrogen-bond angles around oxygen atoms are those of the idealized tetrahedral arrangement of 109.5° rather than 104.5° observed for isolated molecules. Formation of the hydrogen bond in ice lengthens the O–H bond distance, 100 pm compared to 96 pm in a single water molecule. The diagram illustrates a crystal structure that is completely hydrogen bonded, except for the molecules at the surface.

Each O atom of hexagonal ice I_h is surrounded by four almost linear O–H - - O hydrogen bonds of length 275 pm, in a tetrahedral fashion. Each C atom of cubic diamond is also surrounded by four C–C covalent bonds of length 154 pm. Thus, the tetrahedral coordination can either be cubic or hexagonal, from a geometrical viewpoint. Indeed, the uncommon cubic ice and hexagonal diamond have been observed, giving a close relationship in terms of spatial arrangement of atoms between ice and diamond (26). Strong hydrogen bonds make ice hard, but brittle. The structure is related to its physical properties, which vary with temperature.

The pressure of H_2O vapor in equilibrium with ice is called the **vapor pressure of ice,** which decreases as the temperature decreases. At the triple point or 0° C, the pressure is 611.15 Pa. When ice is slightly overheated to 0.01° C, the pressure increases to 611.657 Pa. However, at this temperature, the vapor pressure of liquid water is lower. The vapor pressures of ice between 0° C and – 40° C are listed in Table 5 at 1°C interval. Various models can be used to estimate the vapor pressure at other temperatures. One method uses the Clausius-Clapeyron differential equation

$$\frac{dp}{dT} = \frac{H}{T\,DV}$$

where p is the pressure, T is the temperature (K), H is the latent heat or **enthalpy** of phase transition, and DV is the difference in volume of the phases. The enthalpy of sublimation for ice depends on the temperature. At the freezing point, the enthalpy of sublimation for ice is 51 (51.06 in Table 3) kJ mol^{-1}, estimated from the vapor pressure at 0 and –1° C. The enthalpy of sublimation is required to overcome hydrogen bonding, dipole, and intermolecular attractions. The energy required in freeze-drying processes varies, depending on temperature and other conditions. Water in solutions and in food freezes below 0°C.

Text Box: Table 5. Aqueous vapor pressure (Pa) in equilibrium with ice between 0 and –40°C (at 0.01°C, p is 611.657 Pa) t (° C) – 0< – 1< –2< –3< – 4< – <0 611.15 259.90 103.26 38.01 12.84 – <1 562.67 237.74 93.77 34.24 – <2 517.72 217.32 85.10 30.82 – <3 476.06 198.52 77.16 27.71 – <4 437.47 181.22 69.91 24.90 – <5 401.76 165.30 63.29 22.35 7.20 – <6 368.73 150.68 57.25 20.04 – <7 338.19 137.25 51.74 17.96 – <8 309.98 124.92 46.73 16.07 – <9 283.94 113.62 42.16 14.37 Missing digits of t, <, in the row are found in the column, and vice versa. These values are taken from the CRC Handbook of Chemistry and Physics (1).

The number of hydrogen bonds is twice the number of water molecules, when surface water molecules are ignored. The energy required to separate water molecules from the solid is the enthalpy of sublimation (55.71 J mol^{-1}). Half of this value, 26 kJ mol^{-1}, is the energy to separate the H- -O linkages, and it translates into 0.26 eV, per H- -O bond. These values are close to those obtained by other means (25, 26, 28, 29, 30). Several factors contribute to this linkage, and the hydrogen-bond energy is less than 0.26 eV.

Text Box: Table 6 Properties of liquid water at 298 K Heat of formation DHf 285.89 kJ mol-1 Density at 3.98o C 1.000 g cm-3 Density at 25o C 0.9970480 g cm-3 Heat capacity 4.17856 J g-1 K-1 DHvaporization 55.71 kJ mol-1 Dielectric constant 80. Dipole moment 6.24´10-30 C m Viscosity 0.8949 mP? s–1 Velocity of sound 1496.3 m s–1 Volumetric thermal expansion coefficient 0.0035 cm3 g-1 K-1

B. Properties of Liquid Water

The macroscopic physical properties of this common but eccentric fluid at 298 K (25° C) are given in Table 6. Water has an unusually high melting and boiling points for a substance of molar mass of only 18 daltons. Strong hydrogen bonds and high polarity

accounts for this.

The heat of formation is the energy released when a mole of hydrogen and half a mole of oxygen at 298 K and 1.00 atm react to give one mole of water at 298 K. This value differs from that for ice in Table 4 due to both temperature and phase differences. As temperature increases, the average kinetic energy of molecules increases, and this affects water's physical properties. For example, surface tension of water decreases, whereas the thermal conductance increases as the temperature increases. Heat capacity at constant pressure (*C*p), vapor pressure, viscosity, thermal conductance, dielectric constant, and surface tension in the temperature range 273–373 K (0–100°C) are given in Table 7.

Table 7. Properties of liquid water in the range 273 – 373 K (0 – 100° C)

Temp.	Heat capacity	Viscosity	Thermal conductance	Dielectric constant	Surface tension
t	*C*p				
(°C)	(J g^{-1} K^{-1})	(mPa s)	(W K^{-1} m^{-1})		(mN m^{-1})
0	4.2176	1.793	561.0	87.90	75.64
10	4.1921	1.307	580.0	83.96	74.23
20	4.1818	1.002	598.4	80.20	72.75
30	4.1784	0.797	615.4	76.60	71.20
40	4.1785	0.653	630.5	73.17	69.60
50	4.1806	0.547	643.5	69.88	67.94
60	4.1843	0.466	654.3	66.73	66.24
70	4.1895	0.404	663.1	63.73	64.47
80	4.1963	0.354	670.0	60.86	62.67
90	4.2050	0.315	675.3	58.12	60.82
100	4.2159	0.282	679.1	55.51	58.91

More detailed data can be found in the CRC Handbook of Chemistry and Physics (1)

Liquid water has the largest heat capacity per unit mass of all substances. Large quantities of energy are absorbed or release when its temperatures changes. The large heat capacity makes water an excellent reservoir and transporter of energy. A large body of water moderates climate. The heat capacity C_p of water varies between 4.1 to 4.2 J g^{-1} K^{-1} (74 to 76 J mol^{-1} K^{-1}) even at temperature above 100° C and high pressure. The enthalpy of vaporization for water is also very large (55.71 kJ mol^{-1} at 298 K). Thus, energy consumption is high for food processing when water is involved.

Water and aqueous solutions containing only low molar-mass solutes are typical Newtonian fluids for which the **shear stress** is proportional to **shear strain rate**. Viscosity is the ratio of shear stress to shear strain rate. On the other hand, viscosity of solutions containing high molar-mass substances depends on shear strain rate. For pure water, the viscosity decreases from 1.793 to 0.282 mPa s (millipascal seconds; identical to centipoise (cp)) as temperature increases from 0 to 100° C. Thus, the flow rate through pipes increases as water or solution temperature increases.

The dielectric constant of water is very large, and this enables water to separate ions of electrolytes, because it reduces the electrostatic attraction between positive and negative ions. Many salts dissolve in water. When an electric field is applied to water,

its dipole molecules orient themselves to decrease the field strength. Thus, its dielectric constant is very large. The dielectric constant decreases as temperature increases, because the percentage of molecules involved in hydrogen bonding and the degree of order decrease (28, 29). The measured dielectric constant also depends on the frequency of the applied electric field used in the measurement, but the variation is small when the frequency of the electric field is less than 100 MHz. The dielectric behavior of water allows water vapor pressure to be sensed by capacitance changes when moisture is absorbed by a substance that lies between the plates of a capacitor. These sensors have been developed for water activity measurement (31).

The light absorption coefficients are high in the infrared and ultraviolet regions, but very low in the visible region. Thus, water is transparent to human vision.

Text Box: Table 8. Vapor pressure (kPa) of liquid H2O between triple and critical points at every 10° C t ° C 0<< 1<< 2<< 3<< <00 0.6113 101.32 1553.6 8583.8 <10 1.2281 143.24 1906.2 9860.5 <20 2.3388 198.48 2317.8 11279 <30 4.2455 270.02 2795.1 12852 <40 7.3814 361.19 3344.7 14594 <50 12.344 475.72 3973.6 16521 <60 19.932 617.66 4689.4 18665 <70 31.176 791.47 5499.9 21030 <80 47.375 1001.9 6413.2 22055* <90 70.117 1254.2 7438.0 * Critical pressure at 373.98° C Missing digits of t, <, in the row are found in the column, and vice versa. Values from the CRC Handbook of Physics and Chemistry (2003) (1), which lists vapor pressure at 1o interval.

The variation of vapor pressure as a function of temperature is the bases for defining **water activities** of food. Liquid water exists between the triple-point and the critical-point temperatures (0 – 373.98°C) at pressures above the vapor pressures in this range.

As with ice, the vapor pressure of liquid water increases as the temperature increases. Vapor pressures of water (in kPa instead of Pa for ice in Table 5) between the triple and critical points, at 10°C interval, are given in Table 8. When the vapor pressure is 1.00 atm (101.32 kPa) the temperature is the boiling point (100° C). At slightly below 221°C, the vapor pressure is 2.00 atm. The critical pressure at the critical temperature, 373.98° C, is 217.67 atm (22,055 kPa). Above this temperature, water cannot be liquefied, and the phase is called **supercritical water**.

The partial pressure of H_2O in the air at any temperature is the **absolute humidity**. When the air is saturated with water vapor, the **relative humidity** is 100%. The unsaturated vapor pressure divided by the vapor pressure of water as given in Table 8 at the temperature of the air is the **relative humidity**. The temperature at which the vapor pressure in the air becomes saturated is the **dew point**, at which dew begins to form. However when the dew point is below 273 K or 0°C, ice crystals (frost) begin to form. Thus, the relative humidity can be measured by finding the dew point. Dividing the vapor pressure at the dew point by the vapor pressure of water at the temperature of the air gives the relative humidity. The transformations between solid, liquid, and gaseous water play important roles in hydrology and in the transformation of the environment on Earth. Phase transitions of water combined with the energy from the sun make the weather.

Text Box: Table 9. The density of water (g / mL) as a function of temperature between 0 and 39o C (31). t ° C 0< 1< 2< 3< <0 0.9998426 0.9997021 0.9982063 0.9956511 <1 0.9999015 0.9996074 0.9979948 0.9953450 <2 0.9999429 0.9994996 0.9977730 0.9950302 <3 0.9999627 0.9993792 0.9975412 0.9947971 <4 0.9999750* 0.9992464 0.9972994 0.9943756 <5 0.9999668 0.9991016 0.9970480 0.9940359 <6 0.9999430 0.9989450 0.9967870 0.9936883 <7 0.9999043 0.9987769 0.9975166 0.9933328 <8 0.9998509 0.9985976 0.9962370 0.9929695 <9 0.9997834 0.9984073 0.9956511 0.9925987 Missing digits of t, <, in the row are found in the column, and vice versa.

Density is a collective property, and it varies with temperature, isotopic composition, purity, etc. The International Union of Pure and Applied Chemistry (IUPAC) has adopted the density of pure water from the ocean as the density standard. The isotopic composition of ordinary water is constant, and the density of pure water between 0 and 39° C extracted from (32) is given in Table 9.

The density of cage-like ice I_h, due to 100% of its molecules involved in hydrogen-bonded is only 9% lower than that of water. This indicates that water has a high percentage of molecules involved in the transient and dynamic hydrogen bonding. The percentage of hydrogen-bonded water molecules in water decreases as temperature increases, causing water density to increase. As temperature increases the thermal expansion causes its density to decreases. The two effects cause water density to increase from 0 to 3.98^{o}C, reaching its maximum of 1.0000 g mL^{-1} and then decrease as temperature increases.

Incidentally at 8° C, the density of water is about the same as that at 0° C. At 25° C, the density decreases 0.3 % with respect to its maximum density, whereas at 100° C, it decreases by 4 %. Dense water sinks, and convection takes place when temperature fluctuates at the surface of lakes and ponds, bringing dissolved air and nutrients to various depths of waters for the organism living in them. On the other hand, the pattern of density dependence on temperature of water makes temperatures at the bottoms of lakes and oceans vary little if the water is undisturbed. When water freezes, ice begins to form at the surface, leaving the water at some depth undisturbed. Water at the bottom remains at 4°C preserving various creatures living in water.

When hydrogen-bonded to tissues and cells or in food, water has a unique order and structure, and the vapor pressure and density differ from those of pure water. Yet, the collective behavior of water molecules sheds some light regarding their properties in food, cells, tissues, and solutions.

V. Chemical Properties of Water

Water is a chemical as is any substance, despite the confusion and distrust of the public regarding the term “chemical”. Thus, water has lots of interesting chemical properties. It interacts intimately with components of food particularly as a solvent, due to its dipole moment and its tendency to form hydrogen bond. These interactions affect the chemical properties of nutrients, including their tendency to undergo oxidation or reduction, to act as acids or bases, and to ionize.

A. Water as a Universal Solvent

Water is dubbed a universal solvent, because it dissolves many substances due to strong interactions between water molecules and those of other substances. Entropy is another driving force for a liquid to dissolve or mix with other substances. Mixing increases disorder or entropy.

(a) hydrophobic effect and hydrophilic effect

Because of its large dielectric constant, high dipole moment and ability to donate and accept protons for hydrogen bonding, water is an excellent solvent for polar substances and electrolytes, which consist of ions. Molecules strongly interact with or love water molecules are **hydrophilic**, due to hydrogen bonding, polar-ionic or polar-polar attractions. Nonpolar molecules that do not mix with water are **hydrophobic** or **lipophilic**, because they tend to dissolve in oil. Large molecules such as proteins and fatty acids that have hydrophilic and hydrophobic portions are **amphipathic** or **amphiphilic.** Water molecules strongly intermingle with hydrophilic portions by means of dipole-dipole interaction or hydrogen bonding.

The lack of strong interactions between water molecules and lipophilic molecules or the nonpolar portions of amphipathic molecules is called the **hydrophobic effect**, a term coined by Charles Tanford (33). Instead of a direct interaction with such solutes, water molecules tend to form hydrogen-bonded cages around small nonpolar molecules when the latter are dispersed into water. Hydrogen-bonded water molecules form cages, called **hydrates** or **clathrates**, For example, the clathrate of methane forms stable crystals at low temperatures (34).

Nonpolar chains in proteins prefer to stay together as they avoid contact with water molecules. Hydrophilic and hydrophobic effects play important roles for the stability and state of large molecules such as enzymes, proteins, and lipids. Hydrophobic portions of these molecules stay together forming pockets in globular proteins. Hydrophilic and hydrophobic effects cause nonpolar portions of phospholipids, proteins, and cholesterol to assemble into bilayers or biological membranes (34).

(b) Hydration of ions

Text Box: Fig. 8. The first hydration sphere of most cations M(H2O)6+, and anions X(H2O)6–1 . Small water molecules are below the plane containing the ions, and large water molecules are above the plane.

Due to its high dielectric constant, water reduces the attractions among positive and negative ions of electrolytes and dissolves them. The polar water molecules coordinate around ions forming hydrated ions such as $Na(H_2O)_6^{+}$, $Ca(H_2O)_8^{2+}$, $Al(H_2O)_6^{3+}$ etc. Six to eight water molecules form the first sphere of hydration around these ions. Fig. 8 is a sketch of the interactions of water molecules with ions. The water molecules point their negative ends of their dipoles towards positive ions, and their positive ends towards negative ions. Molecules in the hydration sphere constantly and dynamically exchange with those around them. The number of hydrated-water molecules and their lifetimes have been studied by various methods. These studies reveal that the hydration sphere is one-layer deep, and the life times of these hydrated-water molecules are in the order of picoseconds (10^{-12} s). The larger negative ions also interact with the polar water molecules, not as strong as those of cations. The presence of ions in the solution changes the ordering of molecules even if they are not in the first hydration sphere (9).

The hydration of ions releases energy, but breaking up ions from a solid requires energy. The amounts of energy depends on the substance, and for this reason, some are more soluble than others. Natural waters in the ocean, streams, rivers and lakes are in contact with minerals and salts. The concentrations of various ions depend on the solubility of salts (35) and the contact time.

Drinking water includes all waters used in growth, processing, and manufacturing of food. J. De Zuane divides ions in natural water into four types in The Handbook on Drinking Water Quality (5).

Type A includes arsenic, barium, cadmium, chromium, copper, fluoride, mercury, nitrate, nitrite, and selenium ions. They are highly toxic, yet abundant.

Type B includes aluminum, nickel, sodium, cyanide, silver, zinc, molybdenum, and sulfate ions. Their concentrations are also high, but they are not very toxic.

Type C consists of calcium, carbonate, chloride, iron, lithium, magnesium, manganese, oxygen, phosphate, potassium, silica, bromine, chlorine, and iodine and ozone. They are usually present at reasonable levels.

Type D ions are present usually at low levels: antimony, beryllium, cobalt, tin, thorium, vanadium and thallium.

Most metals are usually present in water as cations, with a few as anions. However, some chemical analyses may not distinguish their state in water. The most common anions are chloride, sulfate, carbonate, bicarbonate, phosphate, bromide, iodide, etc. Toxicity is a concern for ions in water, but some of these ions are essential for humans.

Pure water has a very low electric conductivity, but ions in solutions move in an electric field making electrolyte solutions highly conductive. The conductivity is related to total dissolved solids (TDS), salts of carbonate, bicarbonate, chloride, sulfate, and nitrate. Sodium, potassium, calcium and magnesium ions are often present in natural waters because their soluble salts are common minerals in the environment. The solubilities of clay (alumina), silicates, and most common minerals in the Earth crust, are low.

(c) Hard waters and their treatments

Waters containing plenty of dissolved CO_2 (H_2CO_3) are acidic and they dissolve $CaCO_3$ and $MgCO_3$. Waters with dissolved Ca^{2+}, Mg^{2+}, HCO_3^- and CO_3^{2-} are called **temporary hard waters** as the hardness can be removed by boiling, which reduces the solubility of CO_2. When CO_2 is driven off, the solution becomes less acidic due to the following equilibria (the double arrows, D, indicate reversible reactions):

$H^+ (aq) + HCO_3^- (aq)$ D $H_2CO_3 (aq)$ D $H_2O + CO_2 (g)$

$HCO_3^- (aq)$ D $H^+ (aq) + CO_3^{2-} (aq)$

Reducing the acidity increases the concentration of CO_3^{2-} and solids $CaCO_3$ and $MgCO_3$ precipitate:

$Ca^{2+} (aq) + CO_3^{2-} (aq)$ D $CaCO_3 (s)$

$Mg^{2+} (aq) + CO_3^{2-} (aq)$ D $MgCO_3 (s)$

Water containing less than 50 mg L^{-1} of these substances is considered soft; 50-150 mg L^{-1} moderately hard; 150 - 300 mg L^{-1} hard; and more than 300 mg L^{-1} very hard.

For the **lime treatment**, we determine the amount of dissolved Ca^{2+} and Mg^{2+} first; and then add an equal number of moles of lime, $Ca(OH)_2$, to remove them by these reactions:

$$Mg^{2+} + Ca(OH)_2(s) \text{ ® } Mg(OH)_2(s) + Ca^{2+}$$

$$Ca^{2+} + 2\ HCO_3^- + Ca(OH)_2(s) \text{ ® } 2\ CaCO_3(s) + 2\ H_2O$$

Permanent hard waters contain sulfate (SO_4^{2-}) ions with Ca^{2+} and Mg^{2+}. Calcium ions, Ca^{2+}, of the sulfate solution can be removed by adding sodium carbonate:

$$Ca^{2+} + Na_2CO_3 \text{ ® } CaCO_3 (s) + 2\ Na^+$$

Hard waters cause scales or deposits to build up in boilers and pipes, and they are usually softened by ion exchange with resins or zeolites. In these processes, the calcium and magnesium ions are taken up by the zeolite or resin that releases sodium or

hydrogen ions into the water. Reverse osmosis has also being used to soften hard water.

However, water softening replaces desirable calcium and other ions by sodium ions. Thus, soft waters are not suitable drinking waters. Incidentally, bakers use hard water because the calcium ions strengthen the gluten proteins in dough mixing. Some calcium salts are added to dough to enhance bread quality.

(d) Properties of aqueous solutions

Waters containing dissolved substances are aqueous solutions; their physical properties differ from those of pure water. For example, at the same temperature, the H_2O vapor pressures of solutions are lower than that of pure water, resulting in boiling point elevation (higher), freezing point depression (lower) and osmotic pressure.

Several ways can be used to express concentrations: part per million (ppm), percent, moles per liter, mole fraction, etc. The mole fraction of water is the fraction of water molecules among all molecules and ions in the system. The vapor pressure of an ideal solution, $P_{solution}$, is the vapor pressure of water (at a given temperature), P^{o}_{water}, modified by the mole fraction X_{water}.

$P_{solution} = X_{water} P^{o}_{water}$. ($X_{water} < 1$)
If the solute has a significant vapor pressure, P_{solute} is also modified by its mole fraction,

$P_{solute} = X_{solute} P^{o}_{solute}$.

For non-ideal solutions, in which water and solute strongly interact, the formulas require modifications. A practical method is to use an **effective mole fraction** X defined by:
$P_{solution} / P^{o}_{water.} = X$

In any case, the vapor pressures of solutions containing nonvolatile electrolytes are lower than those of pure water at their corresponding temperature.

Phase transitions take place when the vapor pressures of the two phases are the same. Because solutions' vapor pressures are lower, their melting points are lower but their boiling points are higher. The difference in temperature, ΔT, is proportional to the concentrations of all solutes, $m_{all\text{-}solute}$ (molality),

$\Delta T = K\, m_{all\text{-}solute}$
where K is either the molar boiling point elevation constant, K_b, or the molar freezing point depression constant K_f. For water, K_f = 1.86 K L kg^{-1}, and K_b = 0.52 K L kg^{-1}. Due to ionization of electrolytes, positive and negative ions should be treated as separate species and all species should be included in $m_{all\text{-}solute}$.

The tendency of water molecules from a dilute solution to diffuse into a more concentrated solution, through semipermeable membranes, has a measurable quantity called **osmotic pressure**, π, which is proportional to the concentration (mol per kg of water) of all dissolved species, $m_{all\text{-}solute}$ in mol kg^{-1} and temperature T in K,

$\pi = - m_{all\text{-}solute} R T$
where R is the gas constant 8.3145 J mol^{-1} K^{-1}. Water molecules diffuse from pure water (π = 0) into the solution, and the osmotic pressure is therefore given as a negative value here. Theoretically, a solution with $m_{all\text{-}solute}$ = 1.0 mol kg^{-1} of water, π = – 2477 J kg or
– 2.477 kJ kg at 298 K. Note $m_{all\text{-}solute} = m\, i$; ($i$ being the number of ions produced by the solute) in the van't Hoff equation, which is often used in other literature.

Solutions having the identical osmotic pressure are isotonic. Applying more pressure to a solution to compensate for the osmotic pressure causes water molecules to diffuse through membranes, generating pure or fresh waters. This process is called **reverse osmosis**, and it has been used to soften water or desalinate seawater, converting it to fresh water.

The lowering of vapor pressure and the osmotic pressure of solutions play important roles in hydration and dehydration of food and in living cells. Solutions containing proper concentrations of nutrients and electrolyte have been used to medically treat dehydrated patients. J.R. Cade and his coworkers applied these principles to formulate drinks for athletes; he and his coworkers were credited as the inventor of the sports drink Gaterade (36). The concept of a balanced solution for hydration became a great

business decades after its invention.

B. Acidity and Alkalinity of Water

Acidity and alkalinity are also important characteristics of water due to its dynamic self-ionization equilibrium,

H_2O (l) D $H^+ + OH^-$, $K_w = [H^+][OH^-] = 1 \times 10^{-14}$ at 298 K and 1 atm

where $[H^+]$ and $[OH^-]$ represent the molar concentrations of H^+ (or H_3O^+) and OH^- ions, respectively, and K_w is called the **ion product of water**, (see tables in ref. 1 and 36). Values of K_w under various conditions have been evaluated theoretically (37, 38). Solutions in which $[H^+] = [OH^-]$ are said to be neutral. At 298 K, for a neutral solution,

pH = – log [H+] = pOH = – log $[OH^-]$ = 7 (at 298 K)

The H^+ ions or protons dynamically exchange with protons in other water molecules. The self-ionization and equilibrium are present in all aqueous solutions, including acid and base solutions, as well as in pure water. Water is both an acid and a base.

Strong acids such as $HClO_4$, $HClO_3$, HCl, HNO_3, and H_2SO_4 ionize completely in their solutions to give H^+ (H_3O^+) ions and anions: ClO_4^-, ClO_3^-, Cl^-, NO_3^-, and HSO_4^-, respectively. Strong bases such as NaOH, KOH, and $Ca(OH)_2$ also ionize completely giving OH^- ions and Na^+, K^+, and Ca^{2+} ions respectively. In an acidic solution, $[H^+]$ is greater than $[OH^-]$. In a 1.0 mol L^{-1} HCl solution, $[H^+]$ = 1.0 mol L^{-1}, pH = 0.

Weak acids such as formic acid, HCOOH, acetic acid (CH_3COOH), ascorbic acid ($C_6H_8O_6$), oxalic acid ($H_2C_2O_4$), carbonic acid (H_2CO_3), benzoic acid (C_6H_5COOH), malic acid ($C_4H_6O_5$), lactic acid $H_3CCH(OH)COOH$, and phosphoric acid (H_3PO_4) also ionize in their aqueous solutions, but not completely. The ionization of acetic acid, is represented by the equilibrium,

CH_3COOH (aq) D H^+ (aq) + CH_3COO^- (aq), K_a = 1.75 x 10^{-5} at 298 K

where K_a is the acid dissociation constant.

The solubility of CO_2 in water increases with its (CO_2 partial) pressure, according to Henry's law, and the chemical equilibria for the dissolution is,

$H_2O + CO_2$ (g) D H_2CO_3 (aq)

Of course, H_2CO_3 dynamically exchanges H^+, and H_2O with other water molecules, and this weak diprotic acid ionizes in two stages with their acid constants, K_{a1} and K_{a2}.

$H_2O + CO_2$ (aq) D H^+ (aq) + HCO_3^- (aq), K_{a1} = 4.30 x 10^{-7} (at 298 K)

HCO_3^- (aq) D H^+ (aq) + CO_3^{2-} (aq), K_{a2} = 5.61 x 10^{-11}

Constants K_{a1} and K_{a2} increase with temperature. At 298 K, the pH of a solution containing 0.1 mol L^{-1} H_2CO_3 is 3.7, acidophilic organisms may grow, but most pathogenic organisms are neutrophiles and they cease growing. Soft drinks contain other acids – citric, malic, phosphoric, ascorbic acids etc; they lower the pH further.

Ammonia and many nitrogen-containing compounds are weak bases. The ionization equilibrium of NH_3 in water and the base dissociation constant K_b are,

$NH_3 + H_2O$ D NH_4^+ (aq) + OH^- K_b = 1.70 x 10^{-5} at 298 K.

Other weak bases react with H_2O similarly.

The ionization or dissociation constants of inorganic and organic acids and bases are extensive, and they have been tabulated in various books (39, 40, 41).

Amino acids and proteins contain acidic and basic groups. At some specific pH called the **isoelectric point**, they carry no charge, but exist as zwitterions. For example, the isoelectric point for glycine is pH = 6.00 and it exists as the zwitterion $H_2C(NH_3^+)COO^-$.

C. Oxidation-Reduction Reactions in Water

Oxidation of hydrogen by oxygen not only produces water, but also releases energy. At the standard conditions, the electrochemical half reaction equations are:

$H_2 = 2\ H^+ + 2\ e^-\ E^o = 0.000$ V (defined)

$O_2 + 4\ H^+ + 4\ e^- = 2\ H_2O\ E^o = 1.229$ V

The cell reaction and the cell potential at the standard condition for it are:

$2\ H_2 + O_2 = 2\ H_2O\ DE^o = 1.229$ V

Proper setups for harvesting this energy are the goals of hydrogen-fuel-cell technology. The cell potential DE for non-standard conditions depends on pH and temperature. Its value is related to the energy released in the reaction. A plot of DE versus pH yields a Pourbaix diagram, which is useful to evaluate the stability of various species in water. Water can be a reducing or oxidizing reagent, because it offers protons or electrons. Applying a voltage to pass electrons through a chemical cell decomposes water by electrolysis.

Waters containing dissolved oxygen cause additional reactions, for example:

$2\ H_2O + 2\ e^- = H_2 + 2\ OH^-\ E^o = -0.828$ V

$O_2 + 2\ H^+ + 2\ e^- = H_2O_2\ E^o = 0.682$ V

$O_2 + H_2O + 2\ e^- = HO_2^- + OH^-\ E^o = 0.076$ V

$O_2 + 2\ H_2O + 2\ e^- = 4\ OH^-\ E^o = 0.401$ V

At the proper conditions, a suitable chemical reaction driven by the potential takes place.

Oxidation-reduction reactions involving water usually are due to proton or electron transfer. These oxidation-reduction reactions occur for the growth, production, manufacture, digestion, and metabolism of food.

Water participates in oxidation-reduction reactions in many steps of photosynthesis, resulting in the fixation of CO_2 into biomolecules, releasing oxygen atoms of water as O_2. Engineering a new generation of plants with greater photosynthetic capacity facing lack of waters challenges geneticists (42) and botanists. We now understand photosynthesis to great details, from the studies by many scientists. Photosynthetic reactions are related to food production, but they are so complex that we can only mention them (43).

The oxidation-reduction reactions of water cause corrosion on metal surfaces. Not only deterioration of facilities is very costly for the food industry, corrosion of pipes results in having toxic metal ions Cu^{2+} and Pb^{2+} in drinking water. The concern of lead ions in drinking water led the Environmental Protection Agency to ban the use of high-lead solders for water pipes. These reactions are electrochemical processes. Galvanic effects, high acidity, high flow rate, high water temperature, and the presence of suspended solids accelerate corrosion, as do lack of Ca^{2+} and Mg^{2+} ions in purified waters. The formation of scales protects the metal surface. However, balancing the clogging against surface protection of pipes is a complicated problem, requiring scientific testing and engineering techniques for a satisfactory solution.

D. The Hydrogen Bond and Chemical Reactions

Enzymes are mostly large protein molecules, and they are selective and specific catalysts responsible for most of reactions in biological bodies. Folding of the long protein provides specific 3-dimensional selective pockets for their substrates. The pockets not only fix the substrates in position, they also weaken certain bonds to facilitate specific reactions. This is the mechanism by which enzymes select their substrates and facilitate their specific reactions.

Hydrogen bond strength is stronger in nonaqueous media than in aqueous solutions as the charge densities on the donor and

acceptor atoms increase (44). Hydrogen bonds between the enzyme and its substrate can be stronger than those in an aqueous environment, thus speeding up the reaction rate even further.

The hydrolysis of peptide linkage is the reaction of a protein with water,

R-C(=O)-NH-R' + H_2O ® R-C(=O)OH + H_2N-R'

This type of reaction can be catalyzed by acids, bases, and enzymes.

VI. Water Activity

Water is a nutrient and a component in food groups: grains, meat, dairy, fruits and vegetables. Furthermore, major nutrients such as carbohydrates, proteins, water-soluble vitamins, and minerals are hydrophilic. Even parts of fat or lipid molecules are hydrophilic, but the alkyl chains of fats and proteins experience the hydrophobic effect in an aqueous environment. (45).

Foodstuffs interact with water by means of polar, hydrogen-bonding and hydrophobic interactions. The results of these interactions change the chemical potential (properties) of water. Foodstuffs dissolve in or absorb water. Thus, water within food may be divided into bound water, affected water, and free water in the order of their interaction strength. The bound water molecules are similar to those in the first hydration sphere of ions, and those close to the first sphere are affected water molecules. Further away from the interface are free water molecules. The structure and properties of the first two types change. Interaction of water with dietary fiber is an example (46). Thus, properties of water in food are different from those of pure water.

Water molecules in both liquid and vapor phases can participate in hydration reactions. At equilibrium in a system with two or more phases, their vapor pressure or chemical potential, m, must be equal. The chemical potential, m, of a solution or water-containing foodstuff must be equal at a given temperature *T*, and

$\mathrm{m} = \mathrm{m_w} + R\,T \ln (p / p_w)$,

where *R* is the gas constant (8.3145 J mol^{-1} K^{-1}), and *p* is the vapor pressure of the solution or of water in foodstuff, and p_w is the vapor pressure of pure water at the same temperature. The ratio p / p_w is called the **water activity** $a_{w\,(=\,p\,/\,p_w)}$, which is related to the water chemical potential of water in solutions or in the foodstuff. For ideal solutions and for most moist foods, a_w is less than unity, $a_w < 1.0$ (31).

Both water activity and relative humidity are fractions of the vapor pressure of pure water. Methods for their measurements are the same. We have mentioned the measurement by changes in capacitance earlier. Water contents have a sigmoidal relationship with water activities, $a_w = 1.0$ for infinitely dilute solutions, $a_w > 0.7$ for dilute solutions and moist foods and $a_w < 0.6$ for dry foods. Of course, the precise relationship depends on the material in question. In general, if the water vapor of the atmosphere surrounding the food is greater than the activity of the food, water is absorbed. Otherwise, dehydration takes place. The water activity reflects the combined effects of water-solute, water-surface, capillary, hydrophilic and hydrophobic interactions. The water activity of a foodstuff is a vital parameter, because it affects its texture, taste, safety, shelf life, and appearance.

Furthermore, controlling water activity, rather than water content is important. When $a_w < 0.9$, most molds are inhibited. Growth of yeasts and bacteria also depends on a_w. Microorganisms cease growing if $a_w < 0.6$.

VII Water Potential

Similar to water activity in food, **water potential** is a term used in plant, soil, and crop sciences. Water potential, represented by Y (psi) or Yw, is a measure of the free energy of water in a system: soil, material, seeds, plants, roots, leaves, or an organism. Water potential is the difference between the chemical potential of pure water and water in the system at the same temperature. Pure water has the highest free energy: $\mathrm{Y} = 0$ for pure water by convention, and $\mathrm{Y} < 0$ for solutions. Water diffuses from high potential to low potential. Physiological processes decrease as the water potential decreases.

In general, water potential, Yw ,is a combined effect of osmotic (Ys), matrix (interface and water binding Ym), turgor (Yt) pressures, and gravity (Yg).

$\mathrm{Yw} = \mathrm{Ys} + \mathrm{Ym} + \mathrm{Yt} + \mathrm{Yg}$

Osmotic pressure, Ys, is always present due to solutes in the fluids. The metric pressure, Ym, is related to bound-, affected- and

free-waters in the system. The outwardly directed pressure extended by the swelling protoplast against the wall is called turgor pressure, Υt. Usually, this term is insignificant until the cell is full, and at such point, Υt increases rapidly and stops when $\Upsilon w = \Upsilon t$. Otherwise, the cell ruptures. The mechanical rigidity of succulent plant parts, the opening of stomata and the blossom are usually the results of turgor pressure. In systems such as tall plants and soil science, pressure due to the gravitational pull of water, Υg, is also included in the water potential.

For example, the water potential of potato tissues can be measured by incubating them in a serious of solutions of known osmotic pressures. The potato will neither lose nor gain water if the osmotic pressure of the solution equals the water potential of potato tissues. The osmotic pressure ($\pi = - m R T$) may be evaluated from a known concentration m, using the equation given earlier. Instead of energy units, water potential is often expressed in units of pressure (megapascal, MPa), which is derived by dividing the energy by the molar volume (0.018 L mol^{-1} for H_2O) of water (47).

There are many other methods for water potential measurements depending on the system: soil, leaf, stem, organism, etc. The soil water potential is related to the water available for the plants growing on the soil. Water potential of a plant or leaf indicates its health or state with respect to water. Thus, water potential is a better indicator for plant, agriculture, irrigation, and environmental managements than water content. Water moves through plants because
$\Upsilon_{water} = 0 > \Upsilon_{soil} > \Upsilon_{root} > \Upsilon_{stem} > \Upsilon_{leaf}$

Thus, the concept of water potential and water activity are very useful in growth, manufacture, handling, storage, and management of food.

VIII. Living Organisms in Water

The closer we look, the more we see. Living organisms on Earth are so complicated that their classification and phylogeny are still being studied and revised. New relationships are proposed to modify the five kingdoms proposed by Robert Whittaker in 1969. Nevertheless, most of the earliest unicellular living organisms in the Monera, and Protista kingdoms are still living in water. Both the numbers of species and individuals are staggering. For example, photosynthesis by algae in oceans consumes more CO_2 than that by all plants on land. Algae were probably present on Earth before other organisms. Many phyla (divisions) of fungi, plantae, and animalia kingdoms also make water their homes. Both numbers and species of organisms living in water are probably more than those on land. The subject on living organisms in water is fascinating, but we can only mention some fundamentals about their relationships to water here. Certainly, every aspect of living organisms in water is related to food, because Homo sapiens is part of the food chain, if not at the top of it.

All life requires energy or food. Some living organisms receive their energy from the sun whereas others get their energy from chemical reactions in the aquatic media. Chemical reactions are vital during their lives. For example, some bacteria derive energy by catalyzing the oxidation of iron sulfide, FeS_2, to produce iron ions $Fe(H_2O)_6^{2+}$ and elemental sulfur. Water is the oxidant, which in turn reduces oxygen (48). Chemical reactions provide energy for bacteria to sustain their lives and to reproduce. Factors affecting life in water are minerals, solubility of the mineral, electrochemical potentials of the material, acidity (pH), sunlight, dissolved oxygen level, presence of ions, chemical equilibria, etc. Properties of water influence life in general, and in the aquatic system in particular. As the population grows, aquaculture probably will be seen as a more efficient way of supplying protein for the ever-increasing population.

Regarding drinking water, we are concerned with aquatic organisms invisible to the naked eye. Pathogenic organisms present in drinking water cause intestinal infections, dysentery, hepatitis, typhoid fever, cholera, and other illnesses. Pathogens are usually present in waters contaminated with human and animal wastes that enter the water system via discharge, run offs, flood, and accidents at sewage treatment facilities. Insects, rodents, and animals can also bring bacteria to the water system (49, 50). Testing for all pathogenic organisms is impossible, but some organisms have common living conditions to some pathogenic bacteria. Thus, water testing can use these harmless bacteria as indicators for drinking water safety.

IX. Water Resource, Supply, Treatment, and Usage

About 70% of the Earth surface is covered with water, but only about 2% is covered by fresh water. Ocean waters are salty, and only the small percentages that is fresh water resources (lakes, rivers, and underground). Fresh water is needed for drinking, food, farming, washing, and manufacturing.

When salty water freezes, the ice so formed contains very little salt, if any. Thus, nearly all ice, including the massive ice at the

polar cap, is fresh water. In fact, the ice cap in the Antarctic contains a lot of fresh-water ice, but that cannot be considered a water resource.

Hydrologists, environmentalists, and scientists, engineers, sociologists, economists, and politicians are all concerned with problems associated with water resources. Solutions to these problems require experts and social consensus.

X. Subcritical and Supercritical Waters

Waters at temperatures between the normal boiling and critical points (0 to 373.98° C) are called **subcritical waters**, whereas the phase above the critical point is **supercritical water**. In the 17th century, Denis Papin (a physicist) generated high-pressure steam using a closed boiler, and thereafter pressure canners were used to preserve food. Pressure cookers were popular during the 20th century. Analytical chemists have used subcritical waters to extract chemicals from solids for analysis since 1994 (51).

Water vapor pressures up to its critical point are given in Table 8, but data on polarity, dielectric constant, surface tension, density and viscosity above 100°C are scarce. In general, these properties decrease as the temperature increases. In fact, some drop dramatically for supercritical water. On the other hand, some of them increase with pressure. Thus, properties of sub- and super-critical waters can be manipulated by adjusting temperature and pressure to attain desirable properties.

As the polarity and dielectric constant decrease, water becomes an excellent solvent for non-polar substances such as those for flavor and fragrance. However, foodstuffs may degrade at high temperatures. Applications of sub- and super-critical water are relatively recent events, but applications of supercritical CO_2 (critical temperature 32°C) for chemical analyses started in the 1980s, and investigations of supercritical water followed. However, research and development have been intensified in recent years (52). Scientists and engineers explore the usage of supercritical water for waste treatment, polymer degradation, pharmaceutical manufacturing, chromatographic analysis, nuclear reactor cooling, etc. Significant advances have also been made in material processing, ranging from fine particle manufacture to the creation of porous materials.

Water has been called a green solvent compared to the polluting organic solvents. Sub- and super-critical waters have been explored as replacement of organic solvents in many applications including the food industry (53). However, supercritical water is very reactive, and it is corrosive for stainless steels that are inert to ordinary water. Yet, the application of sub- and super-critical waters is a wide-open field.

XI Postscript

Water, ice, and vapor are collections of H_2O molecules, whose characteristics determine the properties of all phases of water. Together and in concert water molecules shape the landscape, nurture lives, fascinate poets, and captivate scientists. Human efforts in understanding water have accumulated a wealth of science applicable in almost all disciplines, while some people take it for granted.

Water molecules are everywhere, including outer space. They not only intertwine with our history and lives, they are parts of us. How blessed we are to be able to associate and correlate the phenomena we see or experience to the science of water.

An article has a beginning and an end, but in the science of water, no one has the last word, as research and exploration, including its presence in outer space, on water continue (54). Writing this article induced my fascination on this subject, and for this reason, I am grateful to Professors Wai-kit Nip, Lewis Brubacher, and Peter F. Bernath for their helpful discussions and encouragement.

Contributors and Attributions

- Chung (Peter) Chieh (Professor Emeritus, Chemistry @ University of Waterloo)

This page titled 8.2.5: Fundamental Characteristics of Water is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Chung (Peter) Chieh.

8.2.6: Natural Water

Discussion Questions

- What is hard water?
- What are the differences between temporary and permanent hard water?
- How can hard water be converted to soft water?
- How to produce deionized water?

Water is the most important resource. Without water life is not possible. From a chemical point of view, water, H_2O, is a pure compound, but in reality, you seldom drink, see, touch or use pure water. Water from various sources contains dissolved gases, minerals, organic and inorganic substances.

The Hydrosphere

The total water system surrounding the planet Earth is called the **hydrosphere**. It includes freshwater systems, oceans, atmosphere vapour, and biological waters. The Arctic, Atlantic, Indian, and Pacific oceans cover 71% of the Earth surface, and contain 97% of all water. Less than 1% is fresh water, and 2-3 % is ice caps and glaciers. The Antarctic Ice Sheet is almost the size of North America continent. These waters dominate our weather and climate, directly and indirectly affecting our daily lives. They cover 3.35×10^8 km^2. The four oceans have a total volume of 1.35×10^9 km^3.

- The sunlight dims by 1/10 for every 75 m in the ocean, and humans barely see light below 500 m. The temperature of almost all of the deep ocean is 4°C (277 K).
- The average ocean depth is 4 km, and the deepest point at the Mariana Trench is 10,912 m (35,802 ft), which compares to the height of 8.8 km for Mount Everest.

Hydrospheric processes are steps by which water cycles on the planet Earth. These processes include sublimation of ice, evaporation of liquid, transportation of moisture by air, rain, snow, river, lake, and ocean currents. All these processes are related to the physical and chemical properties of water, and many government agencies are set up to study and record phenomena related to them. The study of these processes is called **hydrology**

Among the planets, Earth is the only one in which there are solid, liquid and gaseous waters. These conditions are just right for life, for which water is a vital part. Water is the most abundant substance in the biosphere of Earth. Groundwater is an important part of the water system. When vapor is cooled, clouds and rain develop. Some of the rain percolate through the soil and into the underlying rocks. The water in the rocks is **groundwater**, which moves slowly.

A body of rock, which contains appreciable quantities of water, is called an **aquifier**. Below the **water table**, the aquifier is filled (or saturated) with water. Above the water table is the unsaturated zone. Some regions have two or more water tables. These zones are usually separated by water-impermeable material such as boulder and clay. Groundwater can be brought to the surface by drilling below the water table, and pumped out. The amount of water that can be pumped out depends on the structure of the aquifer. Little water is stored in tight granite layers, but large quantities of water are stored in limestone aquifier layers. In some areas, there are under ground rivers.

Table 8.2.6.1*: Ions in sea water*

Species	Cl^-	Na^+	SO_4^-	Mg^{2+}	Ca^{2+}	K^+	HCO_3^-	Br^-	Sr^{2+}	BO_4^{3-}	F^-	H_4SiO_4	H^+
mg/Kg	10,760	2,710	2,710	1,290	411	399	142	67	8	4.5	1.3	0.5-10	$10^{-8.35}$

Common Ions Present in Natural Water

Hydrology is also the study of how solids and solute interact in, and with, water. In this link, the compositions of seawater, composition of the atmosphere, compositions of rain and snow, and compositions of river waters and lake waters are given in details. Table 8.2.6.1 list the major ions present in seawater. The composition does vary, depending on region, depth, latitude,

and water temperature. Waters at the river mouths contain less salt. If the ions are utilized by living organism, its contents vary according to the populations of organisms.

Dust particles and ions present in the air are nucleation center of water drops. Thus, waters from rain and snow also contain such ions: Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, NH_4^{+}. These cations are balanced by anions, HCO_3^{-}, SO_4^{-}, NO_2^{-}, Cl^{-}, and NO_3^{-}. The pH of rain is between 5.5 and 5.6. Rain and snow waters eventually become river or lake waters. When the rain or snow waters fall, they interact with vegetation, top soil, bed rock, river bed and lake bed, dissolving whatever is soluble. Bacteria, algae, and water insects also thrive. Solubilities of inorganic salts are governed by the kinetics and equilibria of dissolution. The most common ions in lake and river waters are the same as those present in rainwater, but at higher concentrations. The pH of these waters depends on the river bed and lake bed. Natural waters contain dissolved minerals. Waters containing Ca^{2+} and Mg^{2+} ions are usually called **hard water**.

Hard Water

Minerals usually dissolve in natural water bodies such as lakes, rivers, springs, and underground waterways (ground waters). Calcium carbonate, $CaCO_3$, is one of the most common inorganic compounds in the Earth crust. It is the ingredient for both calcite and aragonite. These two minerals have different crystal structures and appearance. This photograph shows crystals of typical Calcite.

Calcium-carbonate minerals dissolve in water, with a solubility product as shown below.

$$CaCO_3 \rightleftharpoons Ca^{2+} + CO_3^{2-} \quad K_{sp} = 5 \times 10^{-9} \tag{8.2.6.1}$$

From the solubility product, we can (see example 1) evaluate the molar solubility to be 7.1×10^{-5} M or 7.1 mg/L (7.1 ppm of $CaCO_3$ in water). The solubility increases as the pH decrease (increase acidity). This is compounded when the water is saturated with carbon dioxide, CO_2. Saturated CO_2 solution contains carbonic acid, which help the dissolution due to the reaction:

$$H_2O + CO_2 \rightleftharpoons H_2CO_3 \tag{8.2.6.2}$$

$$CaCO_3 + H_2CO_3 \rightleftharpoons Ca^{2+} + 2HCO_3^{-} \tag{8.2.6.3}$$

Because of these reactions, some natural waters contain more than 300 ppm calcium carbonates or its equivalents.

The carbon dioxide in natural water creates an interesting phenomenon. Rainwater is saturated with CO_2, and it dissolves limestones. When CO_2 is lost due to temperature changes or escaping from water drops, the reverse reaction takes place. The solid formed, however, may be a less stable phase called aragonite, which has the same chemical formula as, but a different crystal structure than that of calcite.

The rain dissolves calcium carbonate by the two reactions shown above. The water carries the ions with it, sips through the crack of the rocks. When it reached the ceiling of a cave, the drop dangles there for a long time before fallen. During this time, the carbon dioxide escapes and the pH of the water increases. Calcium carbonate crystals begin to appear. Calcite, aragonite, stalactite, and stalagmite are four common solids found in the formation of caves.

Natural waters contain metal ions. Water containing calcium, magnesium and their counter anions are called **hard waters**. Hard waters need to be treated for the following applications.

- Heat transfer carrier in boilers and in cooling systems
- Solvents and reagents in industrial chemical applications
- Domestic water for washing and cleaning

Temporary vs. Permanent Hard Water

Due to the reversibility of the reaction,

$$CaCO_{3(s)} + H_2CO_3 \rightleftharpoons Ca^{2+} + 2HCO_3^{-} \tag{8.2.6.4}$$

water containing Ca^{2+}, Mg^{2+} and CO_3^{2-} ions is called **temporary hard water**, because the hardness can be removed by boiling.

Boiling drives the reverse reaction, causing deposit in pipes and scales in boilers. The deposits lower the efficiency of heat transfer in boilers, and diminish flow rates of water in pipes. Thus, temporary hard water has to be softened before it enters the boiler, hot-water tank, or a cooling system. The amount of metal ions that can be removed by boiling is called **temporary hardness**

After boiling, metal ions remain due to presence of chloride ions, sulfate ions, nitrate ions, and a rather high solubility of $MgCO_3$. Amount of metal ions that can not be removed by boiling is called **permanent hardness. Total hardness** is the sum of temporary hardness and permanent hardness. Hardness is often expressed as equivalence of amount of calcium ions in the solution. Thus, water conditioning is an important topic. The value of water treatment market has been estimated to be worth $30 billion.

Lime-soda Softening

Lime-soda softening is the removal of temporary hardness by adding a calculated amount of hydrated lime, $Ca(OH)_2$:

$$Ca^{2+} + 2HCO_3^- + Ca(OH)_{2(s)} \rightarrow 2CaCO_{3(s)} + 2H_2O \tag{8.2.6.5}$$

Adding more lime causes the pH of water to increase, and as a result, magnesium ions are removed by the reaction:

$$Mg^{2+} + Ca(OH)_{2(s)} \rightarrow Mg(OH)_{2(s)} + Ca^{2+} \tag{8.2.6.6}$$

The extra calcium ions can be removed by the addition of sodium carbonate.

$$Na_2CO_3 \rightarrow 2Na^+ + CO_3^{2-} \tag{8.2.6.7}$$

$$Ca^{2+} + CO_3^{2-} + \rightarrow CaCO_{3(s)} \tag{8.2.6.8}$$

In this treatment, the amount of $Ca(OH)_3$ required is equivalent to the temporary hardness plus the magnesium hardness. The amount of sodium carbonate required is equivalent to the permanent hardness. Thus, lime-soda softening is effective if both the temporary and total hardness have been determined. The sodium ion will remain in the water after the treatment. The pH of the water is also rather high depending on the amount of lime and sodium carbonates used.

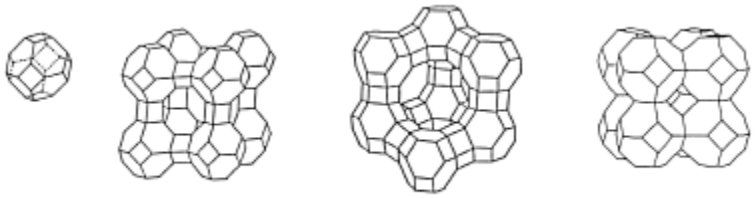
Complexation Treatment

Addition of complexing reagent to form soluble complexes with Ca^{2+} and Mg^{2+} prevents the formation of solid. One of the complexing agents is sodium triphosphate Na_3PO_4, which is marketed as Calgon, etc. The phosphate is the complexing agent. Other complexing agents such as Na_2H_2EDTA can also be used, but the complexing agent $EDTA^{4-}$ forms strong complexes with transition metals. This causes corrosion problem, unless the pipes of the system are made of stainless steel.

Ion Exchange

Today, most water softeners are using zeolites and employing ion exchange technique to soften hard water. Zeolites are a group of hydrated crystalline aluminosilicates found in certain volcanic rocks. The tetrahedrally coordinated aluminum and silicon atoms form AlO_4 and SiO_4 tetrahedral groups. They interconnect to each other sharing oxygen atoms forming cage-type structures as shown on the right. This diagram and the next structural diagram are taken from an introduction to zeolites There are many kinds of zeolites, some newly synthesized.

Whatever kind, the crystal structure of zeolites contains large cages. The cages are connected to each other forming a framework with many cavities and channels. Both positive and negative ions can be trapped in these cavities and channels as shown below.



For each oxygen that is not shared in the AlO_4 and SiO_4 tetrahedral groups, a negative charge is left on the group. These negative charges are balanced by trapping alkali metal and alkaline earth metal ions. When more cations are trapped, hydroxide and chloride ions will remain in the cavities and channels of the zeolites.

To prepare a zeolite for water treatment, they are soaked in concentrated NaCl solution. The cavities trap as many sodium ions as they can accommodate. After the treatment, the zeolite is designated as Na-zeolite. Then the salt solution is drained, and the zeolite is washed with water to eliminate the extra salt. When hard water flow through them, calcium and magnesium ions will be trapped by the Na-zeolite. For every Ca^{2+} or Mg^{2+} trapped, two Na^+ ions are released. The treated water contains a rather high concentration of Na^+ ions, but low concentrations of Mg^{2+} and Ca^{2+}. Thus, zeolite ion exchange convert hard water into soft water.

Pure Water by Ion Exchange

In most cases, the resins are polystyrene with functional $-SO_3H$ groups attached to the polymer chain for cation exchange resin, and with functional group $-N(CH_3)_3^+$ attached to the chain for anion exchange resin. To prepare the resin for making *pure or deionized water*, the cation resin is regenerated with HCl so that the groups are really $-SO_3H$. The anion resin is regenerated with NaOH, so that the functional groups are $-N(CH_3)_3(OH)$. When water containing any metal ion M^+ and anion A^- passes through the ion exchange resins in two stages, the following reactions take place,

$$\mathrm{M^+ + -SO_3H \ \rightarrow H^+ + -SO_3M} \tag{8.2.6.9}$$

$$\mathrm{A^- + -N(CH_3)_3(OH) \ \rightarrow OH^- + -N(CH_3)_3A} \tag{8.2.6.10}$$

$$\mathrm{H^+ + OH^- \rightleftharpoons H_2O} \tag{8.2.6.11}$$

Thus, ion exchange provides *pure water* to meet laboratory requirement.

Reverse Osmosis Water Filter System

This method can also be used to prepare water for domestic and laboratory applications. This method has been discussed in Wastewater treatment

Magnetic Water Treatment

The following is a list of companies selling magnetic devices for **magnetic water treatment**. All devices are based on results of some ressearch indicating that when water runs through a magnetic field, the calcium carbonate will precipitate as aragonite rather than the usual calcite. For example, K.J. Kronenberg has published an article in *IEEE Transactions on Magnetics*, (Vol. Mag-21, No. 5, September 1985, pages 2059-2061). and stated the following:

> *The crystallization mode of the water's mineral content was found to change from a dendritic, substrate-bound solidification habit to the form of separate disc-shaped crystals after the water had moved through a number of magnetic fields. The former scarcity of crystallization nucleii in the water had been turned into an abundance of nucleation centers in the water. The reduction of the number of the substrate-bound crystals has been used as a quantitative measure of the magnetic effect.*

Many companies have made various devices for magnetic conditioning of water, and they claim that their devices will clean up the pipes and boilers at little or no cost. I have yet to test one of these devices for its claim, but my preliminary tests shows that permanent magnet has little effect on the calcium carbonate deposit of temporary hard water. The cleaning effect they have claimed is probably much overstated.

Example 8.2.6.1

From the solubility product shown for the dissolution of calcium carbonate,

$$CaCO_3 \rightleftharpoons Ca^{2+} + CO_3^{2-} \quad K_{sp} = 5 \times 10^{-9} \tag{8.2.6.12}$$

Evaluate the molar solubility of Ca^{2+} in saturated solution.

Solution

From the definition of solubility product, we have

$$[Ca^{2+}][CO_3^{2-}] = 5 \times 10^{-9} \tag{8.2.6.13}$$

Thus,

$$[Ca^{2+}] = [CO_3^{2-}] = 7.1 \times 10^{-5}\ M \tag{8.2.6.14}$$

The concentration of 7.1×10^{-5} M is equivalent to 7.1 mg/L (7.1 ppm of $CaCO_3$ in water).

DISCUSSION

There may be other ions present in the system and other equilibria conditions in addition to the equilibrium mentioned here. Problems are more complexed in the real world.

Exercise 8.2.6.1

Boiling of 1.0 L of water produced 10 mg of CaCO3 solid. What is the temporary hardness of the water?

Answer available at the end of this section.

Contributors and Attributions

- Chung (Peter) Chieh (Professor Emeritus, Chemistry @ University of Waterloo)

Answers

Exercise 8.2.6.1

10 ppm

This page titled 8.2.6: Natural Water is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Chung (Peter) Chieh.

8.2.7: Water Biology

Discussion Questions

- Is water the source of life, why?
- What is life?
- Is life a symbiotic system of many lives?
- What properties of water make water the prime ingredient of life?
- What are the essential substances of life?
- Other than material, what else is essential for life?
- What is the role of water in supporting life?
- Is there a life form involving no water?
- Is there a method to determine the total water content?

Not finished yet, but the major ideas are here now!

Water Biology



Chemical and physical properties of water discussed in other pages are essential considerations for water biology. Natural water also contains biological matters as well as living creatures. In the discusssion of biology and water, you must feel pleasant, because they make each other grow more lovely as this picture from water lily cottage. The water pages of National Wildlife Federation, offers some interesting reading too.

All living things have a cycle of life. A cycle involves all or some of these process: birth, growth, mature, reproduction, matamorphosis, and death. There are millions of living organisms on Earth, ranging from single-cell amebas, bacteria, to the complicated homosapiens. There are also viruses which are fragment of DNA or RNA that depend on host cells for their reproduction. They are not cells.

Living things usually have cells that isolate their systems so that the cells contain unique materials to sustain the lives of cells. Cells regulate their contents (homeostatic), and carry out their metabolsims. They divide or making copies of themselves. Many reproduction process involve two individuals and future populations are subject to a greater diversity. Mutation is a fact of life, and many adopt to their changing environment.

How life started? Let the research and debate continue by not giving any clusive statements here. A physical geography course suggests that the marine invertebrates began their life 600 million years ago, and they are followed by fish, land plants, amphibians, reptiles, mommals, and then flowering plants, in this order. All these started more than a hundred million years ago,

and hominid (primate) line began its evolution 20 to 15 million years ago.

There are strong evidences that life on earth appeared in a body of water. Only the planet Earth has three states of water, and it offers a suitable environment for life to began, among all nine solar planets. Since all life forms involve water. Water is seen as the source, matrix, and mother of life. Water is important, because water is required for life, and some people even consider water as life blood.

Since water supports life, living organisms also modify their environment, changing the nature of the water in which they live. Biology of water pollution, lists the syllabus on a course including a laboratory section. Water and biology interweave into an entangled maze waiting for explorers and curious minds.

Water dissolves or emulsifies other life-supporting substances and transport them to intercellular and intracellular fluids. It is also a medium in which reactions take place. Reactions provide energy (non-matter) for living. Energy causes changes, and manifestation of changes is at least related to, if not the whole, life. An organized and systematized set of reactions is essential in each life.

Balancing Water in Bio-systems

Many living organisms live their lives entirely in water as shown here in this photo from a job center talking about work in marine biology. Aquatic living organisms extract neutrients from water, yet maintaining a balance of electrolyte and nurrishment concentration in their cells. For living things not living in water, they extract water from their environment by whatever mechanism they can. Cells in their body are surrounded by body fluid, and all cells maintain constant concentrations of electrolytes, neutrints, and metabolites. The process of maintaining constant concentrations is called homeostasis. Certainly, some active transport mechanisms are involved in this balance.

The rooting of every type of plants is unique. Generally speaking, plants having extensive roots are able to extract water under harsh conditions. On the other hand, some plants such as cactus, jade and juniper have little roots, but their leaves have a layer of wax that prevents water from evaporation. Water conserving plants tolerate draught, and they survive under harsh conditions. The picture shown here is a jade plant from the above link.

Lately, some pumpkin growers harvested squash weighing almost 500 kg. At the peak of the growing season, the squash grows almost 0.5 kg a day. That is equivalent to 25 moles of water collected by the roots, discounting the water evaporated through the leaves. The growth is particularly good during a hot and wet day, but during a hot sunny after noon, the temperature of the leaves and fruits get very hot.

Essential Electrolytes for Life Support

In addition to water, many inorganic substances or minerals are essential to life. These substances ionize in water to form ions and their solutions conduct electricity. Therefore, they are called **electrolytes**. Because most of these substances are already dissoved in natural water, we list ions instead of the mineral they come from.

When ions dissove, they form complexes with water molecules. For most metals, the first sphere of coordination usually involve 6 water molecules. For example, when sodium chloride dissolve, we have

$$\mathrm{NaCl} + 12\,\mathrm{H_2O} \rightleftharpoons \mathrm{Na(H_2O)_6}^{+} + \mathrm{Cl(H_2O)_6}^{-} \tag{8.2.7.1}$$

$$\mathrm{FeCl_2} + 18\,\mathrm{H_2O} \rightleftharpoons \mathrm{Fe(H_2O)_{62}}^{+} + 2\,\mathrm{Cl(H_2O)_6}^{-} \tag{8.2.7.2}$$

Formation of complexes are due to the high dipole moment of water, and the dissolution can be attribute to the high dielectric constant (80). However, in most publications, we ignore the water molecules in the complexes, and simply consider them as ions.

In the following, we describe some essential ions or salts as electrolytes.

- Sodium chloride, Na^+ and Cl^-
  NaCl is readily dissolved and absorbed in extracellular fluid. The two ions help to balance water, acid/base, osmotic pressure, carbon dioxide transport, and excreted in human urine and sweat. Lack of sodium chloride shows symptoms of dehydration.
- Potassium, K^+
  Good sources of potassium ions are vegetables, fruits, grains, meat, milk, and legumes. It is readily absorbed, and actively

transported into the intracellular fluid. Its function is similar to that of sodium ions, but cells prefer potassium ions over sodium. Lack of potassium leads to cardiac arrest.

- Calcium, Ca^{2+}
  Divalent calcium ions are usually poorly absorbed by human, but the are essential for the bones, teeth, blood clotting. Lack of calcium hinders growth and osteoporosis in old age.
- Phosphates, PO_4^{3-}
  Calcium phosphate is essential for bones, teeth, etc. However, phosphates are also responsible for many life reactions. ATP, NAD, FAD etc are metabolic intermediates, and they involve phosphate. Phospholipids and phosphoproteins are some other phosphate containing species.
- Magnesium, Mg^{2+}
  Magnesium ions are essential in chlorophyll. These ions are absorbed readily, and compete with calcium at times. Magnesium and calcium ions are present in hard water, and this link alerts the lack of magnesium leading to cardiovascular disease.
- Ferrous or ferric ions, Fe^{2+} or Fe^{3+}
  Usually known as iron, but iron are present either as divalent or as trivalent ions. Iron is absorbed according to body need; aided by HCl, ascorbic acid (vitamin C), and regulated by apoferritin. In mammals, iron is stored in liver as ferritin and hemosiderin. Iron deficiency leads to anemia. Good food sources of iron are liver, meats, egg yolk, green vegetables, and whole grains.
- Zinc ions, Zn^{2+}
  Zinc ions are important ingredients for many enzymes. They are present in insulin, carbonic anhydrase, carboxypeptidase, lactic dehydrogenase, alcohol dehydrogenase, alkaline phosphatase etc. Like iron, zinc deficiency leads to anemia and poor growth.
- Copper ions, Cu^{2+}
  Copper ions help iron utilization, and this metal is present in may enzymes.
- Cobalt ions, Co^{2+}
  Cobalt ions are centers of vitamin B12, and deficiency of which leads to anemia.
- Iondine ions, I^-
  Iondine is a constituent of thyroxin, which regulates cellular oxidation.
- Fluoride ions, F^-
  Fluoridation of drinking water is often a controversal issue. Childen's teeth are less seceptible to decay. Once they began to bruch their teeth, the fluoride in tooth paste is sufficient.

The eletrolytes listed above are present in significant amount in water, or fluids of organisms. There are some metals present in very minute quantities in biological systems, and these are not listed above.

Metal ions also interact with proteins. An **enzyme** is usually a very large protein molecule, and it folds into a kidney shape enclosing one or more metal ions forming a complex. The metal is usually responsible for the enzyme activity. Cobalt, copper, iron, molebdenium, nickel, and zinc have groups of enzymes each, and further discussion can be found in The Prosthetic groups and Metal Ions in Protein Active Sites Database (PROMISE). A general discussion is called **bioinorganic chemistry** and this site has an extensive General references on Bioinorganic Chemistry.

Balancing Electrolytes

Ions	Extracellular	Intracellular	Interstitial
Na^+	140	10	150
K^+	5	150	4
Ca^{2+}	10	4	6
Mg^{2+}	6	80	4
Total	**161**	**244**	**164**
Cl^-	103	2	120

Ions	Extracellular	Intracellular	Interstitial
$HCO3^-$	30	10	30
$HPO4^{2-}$	4	177	4
$SO4^{2-}$	2	10	2
Organic acid	6	5	6
Protein	16	40	2
Total	**161**	**244**	**164**

Electrolyte balance are maintained by **passive transport or diffusion and slective active transport mechanisms**. Diffusion process tends to make the concentration all the same throughout the entire fluid, but active or selective transport moves ions to special compartment. For example, the active transport of sodium and potasium by an enzyme called **sodium-potassium ATPase** is usually known as *sodium-potassium pump*. This process pumps potassium ions inside a cell while removing sodium ions from the cells. Thus, a high concentration of potassium is maintained inside cells. Energy is required in active transport, and cellular metabolism provides the energy and the necessary molecular motions to facilitate the process.

Hormons are produced by special cells, and they are responsible for the communication between various part of the body. Some complicate harmon actions regulates the rate of transport and balance the ion concentrations depending on the portion of the tissue and the need. This is generally called the **hormonal effects** following the suggestion of *human biochemistry*.

Gibbs-Donnan effect considers the equilibrium in compartments that are separated by memberances or cell walls. There will be no net change when the products of concentrations of say $[Na^+]_1$, $[Cl^-]_1$ are the same for compartments 1 and 2.

$$[Na+]1\ [Cl-]1 = [Na+]2\ [Cl-]2$$

the subscript 1 and 2 refer to the two compartments. When no other components are present, we have

$$[Na+]1 = [Cl-]1 = [Na+]2 = [Cl-]2$$

But if compartment 2 has a sodium salt with other anions, this salt ionize to give Na^+ too. The above condition will not be maintained, in this case. In other words, thermodynamic will be a force to adjust the concentrations.

In general, the cations should be balanced with anions. Otherwise, the solution will be charged.

Water in Human Biology

In human, water in the tissue and body fluid is mostly free, but some fraction may be bounded in pockets of hydrophilic compartments. Body fluids have many electrolytes and neutrients dissolve in them.

Intracellular fluid	70%
Interstitial fluid (lymph)	20%
Blood plasma	7%
Intestinal lumen etc.	3%

Human Biochemistry by J.M. Orten and O.W. Neuhaus (1982), 10th Ed. suggests that about 70% of human body weight is water, most found in three major compartments: 70% intracellular fluid, 20% interstitial fluid, and 7% blood plasma, and only 3 % in intestinal lumen, cerebrospinal fluid and other compartments.

However, *Human Biochemistry* also suggests that blood makes up about 8% of the total body weight.

Example 1

For a person weighing 50 kg (110 lb), what is the weight of the blood?

Solution
From the distribution given above,

$$\text{Amount of blood} = 50\ \text{kg} * 0.08$$
$$= 4\ \text{kg}.$$

This is a lot of blood, and donating 0.5 L of blood will not affect the normal function of the blood.

Input		Output	
Drinking	400 g	Skin	500 g
Beverage	580	Expired air	350
Preformed water in solid food	720	Urine	1100
Metabolic water	320	Feces	150
Total	2020	**Total**	2100
Balance	-80 g?		

Water in human comes from ingestion. Aside from drinking water, there is other beverages. Much of the food also contain water. When food is oxidized in the cells, all hydrogen in food converts to water, which is called metabolic water. Water is excreted via urine, feces, skin, and expiration. A typical daily water balance is shown in a table here. Water balance is maintained between cells and fluid, and the output depends on kidney functions and body insensible perspiration (Expired air from the lung is saturated with water vapor, and evaporation from the skin).

Drinking Water

Drinking water affects health. An Excite search using the phrase "drinking water" came up with 57890 documents. Drinking Water Resources gives annotated links to web sites that provide information about the drinking water.

A rather recent book *Chemistry of Water Treatment* by S.D. Faust and O.M. Aly, 2nd Ed. (1998) [TD433 F38 1998] addresses the standards for drinking water in the first Chapter. The standards have changed over the years, as we better understand the science.

Safe drinking water is a suitable combination of minerals and electrolytes. Usually, one should not drink water softened by water softeners. Using distilled water for beverages and cooking may not reach your set goals. Hard water with calcium and magnesium ions is good for drinking.

Usually, a government set up a non-profit organization to provide rules for safe drinking water. This organization has an infrastructure to monitor drinking water systems, and it shall also carry out research to improve the quality of drinking water.

Regarding making rules, reliable tests should be developed to determine the electrolyte we have mentioned, plus others such as lead ions, Pb^{2+}, mercury Hg^{2+}, methylmercury, arsenic, radioactivity, etc. Bacteria tests should be carried out regularly. This organization should also have a communication channel to release relavant message.

The Environmental Protection Agency of the U.S. gives a list of comtaminants. The list has suggested limits, and it divides the contaminants into

- Inorganic substnaces
- Organic substances
- Radioactivities (alpha and beta rays, radium)
- Micro-organisms

Among inorganic substances, limits are given to contents of antimony, arsenic, asbestos, barium, beryllum, cadmium, chromium, copper, mercury, nitrate, nitrite, selenium, and thallium.

More than 50 organic compounds are on the list, and some familar ones are: acrylamide, benzene, carbon tetrachloride, chlorobenzene, 2 4 D, dichlorobenzen, dioxin, polychlorinated biphenyls (PCBs), toluene, and vinyl chloride. Many of these have a zero limit.

In terms of microorganisms, *Giardia lamblia*, and *Legionella*, are checked. Furthermore, viruses, turbidity, total coliforms, and heterotrophic plate should be checked.

Contaminant	Standard
Aluminum	0.05 to 0.2 mg/L
Chloride	250 mg/L
Color	15 (color units)
Copper	1.0 mg/L
Corrosivity	noncorrosive
Fluoride	2.0 mg/L
Foaming Agents	0.5 mg/L
Iron	0.3 mg/L
Manganese	0.05 mg/L
Odor	3 threshold odor number
pH	6.5-8.5
Silver	0.10 mg/L
Sulfate	250 mg/L
Total Dissolved Solids	500 mg/L
Zinc	5 mg/L

The secondary standard lists most electrolytes as shown in this table on the right.

The Secondary Drinking Water Standards are non-enforceable guidelines regulating contaminants that may cause cosmetic effects (such as skin or tooth discoloration) or aesthetic effects (such as taste, odor, or color) in drinking water.

There are many brands of bottled drinking water, they have been very popular only in recent years. Do we know the bottle procedure? Is the industry regulated? Is the water quality reliable? Are all bottled water the same? Is there a brand bottled water for real good health? Do we know what should be in healthy drinking water? There is an opinion expressed in Ontario Clean Water Agency (OCWA). Check it out.

The magnesium web site gave the following news release Oct. 4, 1999. *According to the U.S. National Academy of Sciences (1977) there have been more than 50 studies, in nine countries, that have indicated an inverse relationship between water hardness and mortality from cardiovascular disease. That is, people who drink water that is deficient in magnesium and calcium generally appear more susceptible to this disease. The U.S. National Academy of Sciences has estimated that a nation-wide initiative to add calcium and magnesium to soft water might reduce the annual cardiovascular death rate by 150,000 in the United States.* This is a good summary from the report.

Sports Drinks

Sports talk gives information about Sports Drinks. This is science, art, testing, and myth. However, some fundamentals should be considered.

Our bodies are mostly water, about 70%. The body fluid has many different things dissolved in it, particularly salt. the salinity -

varies somewhat with where you take the sample of water to measure. Don't worry about that). I recall the concentration as being 0.5%, but I'm not sure. Doctors call this "normal saline".

Now if you put a human, or other animal, cell in saltwater that is the same concentration as the saltwater inside the cell, the cell pretty much just sits there. If you put it in distilled water, the cell absorbs the water through its cell membrane - called diffusion - until it eventually pops. If you put the cell in concentrated saltwater, the cell looses water. The water diffuses out of the cell through the membrane, leaving a small, shriveled up cell.

What does this have to do with sports drinks? If you give someone distilled water, it seems like they would absorb the water faster because of what I just described. On the other hand, during sweating, you're loosing sodium, potassium and small quantities of other electrolytes. If you're exercising particularly long or hard, you need to replace those electrolytes. Researchers found that adding some salt to water replaced the salt lost through sweating and helped the body to get water to the cells. If you look at a label on a Gatorade or other drink, you'll find that the main electrolyte is simple salt. But if you put too much of the electrolytes in the water, the cells shrivel up just like the way I described above.

I hope that helps you understand what happens. This stuff on how cells swell up or shrivel up is in high school biology books, and maybe even in something you can find in your school's library.

Taste and Order of Drinks

Taste and order are sensations, and thus hard to quantify and systematize. Often, the Weber Fechner law is used. This law expresses the taste or order sensations S as being proportional to the logarithm of the stimulus *R*, with the proportional constant *K*,

$$S = K \log R$$

For some common substances, the minimum amounts detected by *expert nose* or *expert taster* gives the sensation as a threshold, below which no taste or order was detected.

However, orderous sensation of water may be reported as threshold order number (TON). If *A* mL of ordorous sample is diluted with *B* mL of order-free water to be "just detectable" to the *expert nose*, the TON is defined as

$$TON = \frac{A + B}{A}$$

Similarly, a flavour threshold number (FTN) can be defined in the same manner.

$$FTN = \frac{A + B}{A}$$

except that *A* and *B* are volumes of samples and taste free water used.

These formulations show a method of defining some quantities that are, otherwise, very diffcult to quantify. There are other methods of reporting orders and taste, and some bottling companies have their standard methods of comparison.

Of course, the source of order and taste comes are organic, inorganic compounds as well as bacteria, algae. For example, mercaptans such as C_2H_2SH and ammonia offers disagreeable smell and taste.

Example 2

A sample of water was tested by 10 expert noses, and only 5 of them can detect an order. Thus, this is "just detected". What is the TON for this sample?

Solution

Since no dilution was used,

$$TON = A / A = 1.$$

DISCUSSION

If equal amount of orderless water is required to dilute it so that the order is "just detected", then the test order number is (1+1)/1 = 2.

Contributors and Attributions

- Chung (Peter) Chieh (Professor Emeritus, Chemistry @ University of Waterloo)

This page titled 8.2.7: Water Biology is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Chung (Peter) Chieh.

8.2.8: Water Chemistry

Water is an unusual compound with unique physical properties. As a result, its the compound of life. Yet, its the most abundant compound in the biosphere of Earth. These properties are related to its electronic structure, bonding, and chemistry. However, due to its affinity for a variety of substances, ordinary water contains other substances. Few of us has used, seen or tested pure water, based on which we discuss its chemistry.

The chemistry of water deals with the fundamental chemical property and information about water. Water chemistry is discussed in the following subtitles.

- Composition of water
- Structure and bonding of water
- Molecular Vibration of water
- Symmetry of water molecules
- Formation of hydrogen bonding in water
- Structure of ice
- Autoionization
- Leveling effect of water and acid-base characters
- Amphiprotic nature
- Reactivity of water towards alkali metals; alkaline earth metals; halogens; hydrides; methane; oxides; and oxygen ions.
- Electrolysis of water

Composition of Water

Water consists of only hydrogen and oxygen. Both elements have natural stable and radioactive isotopes. Due to these isotopes, water molecules of masses roughly 18 ($H_2{}^{16}O$) to 22 ($D_2{}^{18}O$) are expected to form. Isotopes and their abundances of H and O are given below. From these data, we can estimate the relative abundances of all isotopic water molecules.

Abundances (% or halflife) of hydrogen and oxygen isotopes

^{1}H	^{2}D	^{3}T		
99.985%	0.015%	12.33 y		
^{14}O	^{15}O	^{16}O	^{17}O	^{18}O
70.6 s	122 s	99.762%	0.038%	0.200%

The predominant water molecules $H_2{}^{16}O$ have a mass of 18 amu, but molecules with mass 19 and 20 occur significantly. Because the isotopic abundances are not always the same due to their astronomical origin, The isotopic distribution of water molecules depends on its source and age. Its study is linked to other sciences. (*See* Dojlido, J.R. & Best, G.A. (1993) *Chemistry of Water and Water Pollution*, Ellis Harwood for isotopic distribution of water.)

Relative abundance of isotopic water

$H_2{}^{16}O$	$H_2{}^{18}O$	$H_2{}^{17}O$	$HD^{16}O$	$D_2{}^{16}O$	$HT^{16}O$
99.78%	0.20%	0.03%	0.0149%	0.022 ppm	trace
18	20	19	19	20	20 amu

In particular, $D_2{}^{16}O$ is called heavy water, and it is produced by enrichment from natural water. Properties of heavy water are particularly interesting due to its application in nuclear technology.

Structure and Bonding of the Water Molecule

Pure water, H_2O, has a unique molecular structure. The O-H bondlengths are 0.096 nm and the H-O-H angle = 104.5°. This strange geometry can be explained by various methods. From carbon to neon, the numbers of valence electrons increase from 4 to 8. These elements require 4, 3, 2, 1, and 0 H atoms to share electrons in order to complete the octet requirement. Their **Lewis dot structures** are shown on the right, and note the trend in bondlengths.

There are six valance electrons on the oxygen, and one each from the hydrogen atom in the water molecule. The eight electrons form two H-O bonds, and left two lone pairs. The long pairs and bonds stay away from each other and they extend towards the corners of a tetrahedron. Such an ideal structure should give H-O-H bond angle of 109.5°, but the lone pairs repel each other more than they repel the O-H bonds. Thus, the O-H bonds are pushed closer, making the H-O-H angle less than 109°.

After the introduction of quantum mechanics, the electronic configuration for the valence electron of oxygen are $2s^2\ 2p^4$. Since the energy levels of 2*s* and 2*p* are close, valence electrons have characters of both *s* and *p*. The mixture is called ***sp*^{3} hybridization.** These hybridized orbitals are shown on the right. The structures of CH_4, NH_3, and H_2O can all explained by these hybrid orbitals of the central atoms. The above approach is the valence bond theory, and both the C-H bonds and lone electron pairs are counted as **VSPER pairs** in the Valence-shell Electron-Pair Repulsion (VSEPR) model, according to which, the four groups point to the corners of a tetrahedron.

For triatomic molecules such as water, molecular orbital (MO) approach can also be applied to discuss the bonding. The result however is similar to the valence bond approach, but the MO theory gives the energy levels of the electron for further exploration.

Molecular Vibration of Water

Atoms in a molecule are never at rest, and for each type of molecule, there are some normal vibration modes. For the water molecule, the three normal modes of vibrations are symmetric stretching, bending and asymmetric stretching. The vibrations are quantized, as do any microscopic system, and their quantum numbers are designated as v_1, v_2 and v_3. The observed transition bands of D_2O, H_2O, and HDO are given in the table below.

Transition bands of D2O, H2O, and HDO

Quantum numbers of upper state			**Absorption wavenumbers of bands /cm^{-1}**		
v_1	v_2	v_3	D_2O	H_2O	HDO
0	1	0	1178	1594	1402
1	0	0	2671	3656	2726
0	0	1	2788	3756	3703
0	1	1	3956	5332	5089
Data from Eisenberg, D. and Kauzmann, W. (1969) *Structure and properties of water*, Oxford University press.					

The ideal transition bands are centered in the given wavenumbers. However, these wavenumbers are calculated based on isolated molecules with no interaction with any neighbor. When molecules interact with each other, the energy levels are modified, and the bands shift.

Many more less intense absorption bands extend into the green part of the visible spectrum. The absorption spectrum of water may contribute to the blue color for lake, river and ocean waters.

Symmetry of Water Molecules

The water molecules are rather symmetric in that there are two mirror planes of symmetry, one containing all three atoms and one perpendicular to the plane passing through the bisector of the H-O-H angle. Furthermore, if the molecules are rotated 180° (360°/2) the shape of the molecule is unperturbed. This indicates that the molecules have a 2-fold rotation axis. The three symmetry elements are 2-fold rotation, and two mirror planes. Both mirror planes contain the rotation axis, and this type of

symmetry belongs to the **point group** C_{2v}. A **point group** has a definite number of symmetry elements arranged in certain fashion. Molecules can be classified according to their point groups. Molecules of the same point group have similar spectroscopic characters. Other molecules of C_{2v} point group are CH_2=O, CH_2Cl_2, the bent O_3 etc.

Formation of Hydrogen Bonding

Under certain conditions, an atom of hydrogen is attracted by rather strong forces to two atoms instead of only one, so that it may be considered to be acting as a bond between them. This is called ***hydrogen bond***. This statement is from Linus Pauling (1939) in his book *The Nature of the Chemical Bond*. He gave the ion $[F:H:F]^-$ as an example. At that time, the hydrogen bond was recognized as mainly ionic in nature. The energy associated with hydrogen bond is 8 to 40 kJ/mol.

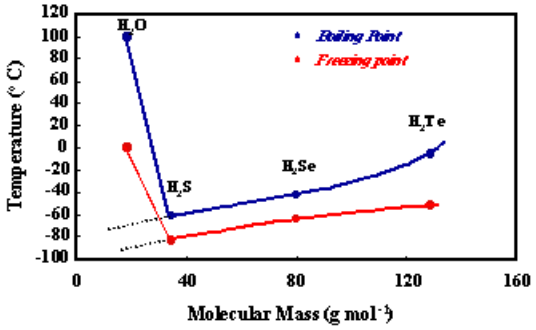
Comparison of melting and boiling points for a few substances

Molecule	Molar mass	m.p.	b.p. /° C
NH_3	17	-77.8	-33.5
H_2O	18	-0	100
H_2S	34	-85.6	-60
H_2Se	81	-60.4	-41.5
H_2Te	128.6	-51	-1.8
CH_3OH	32	?	65
C_2H_5OH	46	?	78
$C_2H_5OC_2H_5$	74	?	34

Normally, the melting point and boiling point of a substance increase with molecular mass. For example the melting points of inert gases are 0.95, 24.48, 83.8, and 116.6 K respectively for He, Ne, Ar, and Kr.

In this table, the melting and boiling points for water are particular high for its small molecular mass. This is usually attributed to the formation of hydrogen bonds. The small electronegative atoms F, O and N are somewhat negatively charged when they are bonded to hydrogen atoms. The negative charges on F, O and N attract the slightly positive hydrogen atoms, forming a strong interaction called hydrogen bond.

A graph showing the melting points and boiling points of group 16 provided by Prof. J. Boucher illustrates the same point.



Boiling and Freezing Points of Group 16 Hydrides

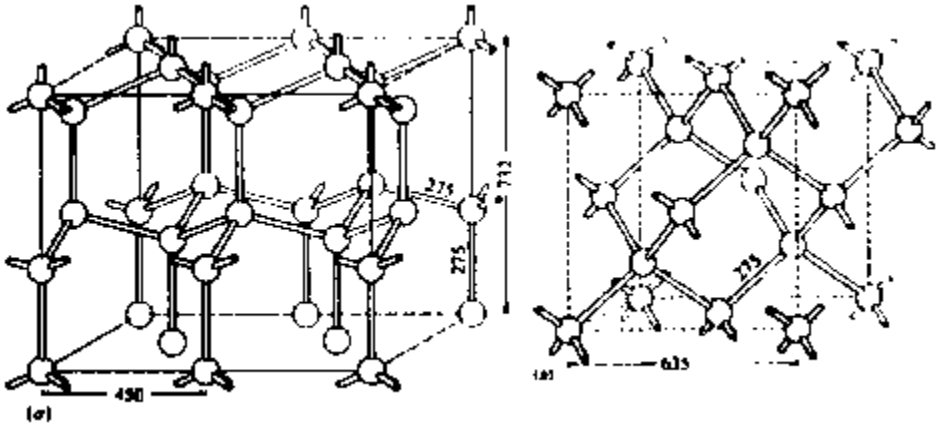
Based on the observed absorption at 3546 and 3691 cm^{-1}, Van Thiel, Becker, and Pinmentel (1957, *J. Chem. Phys.* **27** 386) suggested the formation of water dimer when trapped in a matrix of nitrogen.

Due to hydrogen bonding, water molecules form dimers, trimers, polymers, and clusters. The hydrogen bonds are not necessarily liner.

Structure of Ice

Ice occurs in many places, including the Antarctic. If all the ice melted, the water level of the oceans will rise about 70 m. The structure of ice and the caption are from this link.

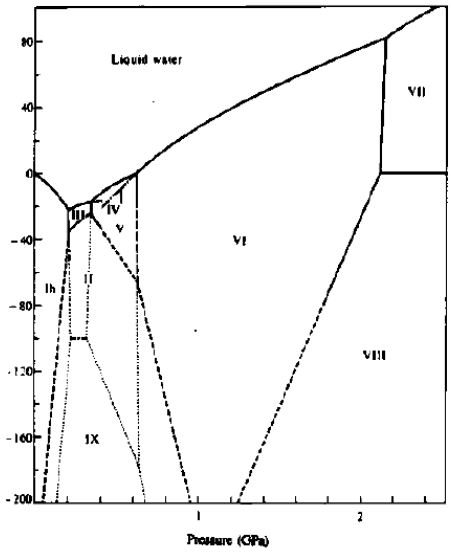
The density of ice is dramatically smaller than that of water, due to the regular arrangement of water molecule via hydrogen bonds. In an idealized structure of ice, every hydrogen atom is involved in hydrogen bond. Every oxygen atom is surrounded by four hydrogen bonds.



This diagram from caltech.edu, shows the structure of hexagonal ice in (a) and cubic ice in (b). A rod here represents a hydrogen bond. Since the hydrogen bonds are not linear, the real structure is a little more complicated.

The tetrahedral coordination opens up the space between molecules. On each hydrogen bond, shown by a rod joining the oxygen atoms, lies one proton in an asymmetric position (not shown). Bond lengths, 275 pm, are indicated. Ordinary ice is hexagonal. and the hexagonal *c* axis is labelled 732 pm, and one of the hexagonal *a* axes is labelled 450 pm. If water vapor condenses on very cold substrate at 143-193 K (-130 to -80°C) a cubic phase is formed. In (b) the cubic unit cell is outlined with dashed lines; dimensions are in pm determined at 110 K.

These diagrams can also be used to represent the two forms of diamond, and in this case, the rods joining the atoms represent C-C bonds. Each C-C bondlength is 154 pm. Silicon and germanium crystals have the same structure, but their bondlengths are longer. The two diamond types of structure are related to the packing of spheres. The hexagonal type has the ABABAB... sequence, whereas the cubic type has the ABCABC... sequence. In both cases, half of the tetrahedral sites are occupied by tetrahedrally bonded carbon atoms. Hexagonal diamonds have been observed in meteorites.



The four hydrogen bonds around an oxygen atom form a tetrahedron in a fashion found in the two types of diamonds. Thus, ice, diamond, and close packing of spheres are somewhat topologically related.

A phase diagram of water shows 9 different solid phases (ices). Ice Ih is the ordinary ice. In addition to ice Ic from vapor deposition, conditions for nine phases are shown. Aside from ice I, other phases are formed and observed under high pressure generated by machines built by scientists. So far, ten different forms of ice have been observed, and some ice forms exist at very high pressure. The pressure deep under the polar (Antarctic) ice cap is very high, but we are not able to make any direct observation or study.

There is a report of the 11th ice, and the ice phase diagram and drawings of ice structures given here is extremely interesting.

The Autoionization of Water

The Autoionization of Water in the formation of ions according to

$$HOH(l) + HOH(l) = H3O+ + OH-$$

This is an equilibrium process and is characterised by an equilibrium constant, K'_w:

$$K'w = \frac{[H3O+]\,[OH-]}{[H2O]}$$

Since $[H_2O] = 1000/18 = 55.56$ M, and remains rather constant under any circumstance, we usually write

Kw = [H3O+] [OH-]
= 10-14 (or 1e-14)
pKw = -log Kw (defined)
= 14 (at 298 K)

t°C	K_w
20	1.14e-15
25	1.00e-14
35	2.09e-14
40	2.92e-14
50	5.47e-14

For neutral water, $[H_3O^+] = [OH^-]$ = 1e-7 at this temperature. Furthermore, we define

pH = -log[H3O+]
pOH = -log[OH-]
pH = pOH = 7 at 298 K; (in neutral solutions)

It is important to realize that K_w depends on temperature as shown in the Table here.

Leveling Effect of Water and Acid-Base Characters

The strength of strong acids and bases is dominated by the autoionization of water. In aqueous solutions, the strongest acid and base are the hydronium ion, H_3O^+, and the hydroxide ion OH^- respectively. Acids HCl, HBr, HI, HNO_3, $HClO_3$, $HClO_4$, and H_2SO_4 completely ionize in water, making them as strong as H_3O^+ due to the **leveling effect** of water. Furthermore, strong acids, strong bases, and salts completely ionize in their aqueous solutions.

For example, HCl is a stronger acid than H_2O, and the reaction takes place as HCl dissolves in water.

HCl + H2O = Cl- + H3O+

A similar equation can be written for another strong acid.

On the other hand, a stong base also react with water to give the stong base species, OH^-.

H2O + B- = OH- + HB

For example, O^{2-}, CH_3O^-, and NH_3 are strong bases. The leveling effect also apply to bases.

Amphiprotic Species

Equilibria of acids and bases, are interesting chemistry. When an acid and a base differ by a proton, they are called a conjugate acid-base pair. A water molecule is a weak acid and base, due to its ability to accept or donate a proton. Such properties make water an **amphiprotic species**. In fact, H_3O^+, H_2O and OH^- are amphiprotic, as are some other conjugate acid-base pairs of weak acids and bases.

If several acids and bases are dissolved in water, all equilibria must be considered. To estimate the pH of these solutions requires the exact treatment of several equilibrium constants. For example, many species dissolve in rain water, and many equilibria must be considered. Detail consideration and examples are given in Acid-Base Reactions

Carbon dioxide in the air dissolve in rain water, lakes and rivers. A solution of CO_2 involves the following reaction:

Reaction	***K* formula**	***K* value**
$H_2O(l) + CO_2(g) = H_2CO_3(l)$	$1/P_{CO_2}$	?
$H_2CO_3 = HCO_3^- + H^+$	$[HCO_3^-]\,[H^+] / [H_2CO_3]$	5e-7
$HCO_3^- = CO_3^{-2} + H^+$	$[CO_3^{-2}]\,[H^+] / [HCO_3^-]$	5e-11
$HOH(l) + HOH(l) = H_3O^+ + OH^-$	$[H_3O^+]\,[OH^-]$	1e-14

These complicated equilibria make natural water a buffer.

Example 1

Assume that the partial pressure of carbon dioxide causes a total concentration of carbonic species to be 8e-4 M. Estimate the pH of this solution.

Solution
From the given data, we have the following five equations and five unknowns:

Equilibrium	Equations	No.
H_2CO_3 « HCO_3^- + H^+	$\frac{[HCO_3^-]\,[H^+]}{[H_2CO_3]}$ = 5e-7	(1)
HCO_3^- « CO_3^{2-} + H^+	$\frac{[CO_3^{2-}]\,[H^+]}{HCO_3^-}$ = 5e-11	(2)
2 H_2O « H_3O^+ + OH^-	$[H_3O^+]\,[OH^-]$ = 1e-14	(3)
Charge balance	$[H^+]$ = $[HCO_3^-]$ + $[OH^-]$ + 2 $[CO_3^{2-}]$	(4)
All species containing C	$[H_2CO_3]$ + $[HCO_3^-]$ + $[CO_3^{2-}]$ = 8.0e-4 M	(5)
Unknowns		
$[H^+]$, $[OH^-]$, $[H_2CO_3]$, $[HCO_3^-]$, $[CO_3^{2-}]$		

Solving these equations for the 5 unknowns can be done using Maple, Mathcad, spread sheet, or approximation. In any case, we are interested in the pH, and we can make the following approximations or assumptions

Assume H^+ mostly comes from (1)	$[H^+]$ = $[HCO_3^-]$	
H_2CO_3 is a weak acid most unionize	$[H_2CO_3]$ = 8.0e-4 M	(6)
Let x = $[HCO_3^-]$ = $[H^+]$	$\frac{[HCO_3^-]\,[H^+]}{[H_2CO_3]}$ = $\frac{x^2}{[H_2CO_3]}$ = 5.0e-7	(1)

Combining (1) and (6) gives $[H^+]^2 = x^2$ = 8.0e-4 * 5.0e-7 = 4.0e-10. Therefore,

[H+] = 2.0e-5
pH = -log(2.0e5) = 4.7

DISCUSSION
Generally speaking, rain water has a pH about 5, rather acidic. It dissolves limestone and marble readily. Due to the dissolved carbon dioxide, rain water is a buffer solution.

Increased carbon dioxide level forces an increase in dissolved carbon dioxide. Would this causes pH of rain water to decrease or increase? Justify your answer by giving the reasons.

Since $[H^+]$ = 2.0e-5, $[OH^-]$ = 5e-9, the amount of H^+ from ionization of water is also 5.0e-9, small with respect to 2.0e-5 from ionization of H_2CO_3. Similarly, the ionization from

HCO3- « CO32- + H+

is also small. Most of the C-containing species is H_2CO_3

H_2CO_3 is a weak acid, its ionization is small indeed.

Now, you may proceed to evaluate other concentrations: [OH^-], [HCO_3^-], and [CO_3^{2-}]

Reactivity of Water Towards Metals

Alkali metals react with water readily. Contact of cesium metal with water causes immediate explosion, and the reactions become slower for potassium, sodium and lithium. Reaction with barium, strontium, calcium are less well known, but they do react readily. Warm water may be needed to react with calcium metal, however. Many metals displace H^+ ions in acidic solutions. This is often seen as a property of acids.

Electrolysis of Water

The *enthalpy of formation* for liquid water, $H_2O(l)$, is -285.830 and that of water vapour is -241.826 kJ/mol. The difference is the heat of vaporization at 298 K. Liquid water and vapor entropies (S) are 69.95 and 188.835 kJ K^{-1} mol^{-1} respectively, (*see* Thermodynamic Data. These are entropies, not standard entropies of formation. The entropy of formation for water is obtained by,

DSof water = Sowater - SoH2 - 0.5 SoO2
= 69.95 - 130.68 - 0.5*205.14 (data from Thermodynamic Data)
= - 163.3 J K-1 mol-1
DGowater = DH - T DS (note H in kJ/mol and S in J/mol)
DGowater = -285.83 - 298.15 * 163.3/1000 = -237.13 kJ

The equilibrium constant and Gibb's energy are related,

$$DGo = - R T \ln K$$
$$K = \exp(- DGo / R T)$$
$$= 3.5e41 \text{ atm-3/2}$$

This is a very large value for the formation of water,

H2 + 0.5 O2 = 0.5 H2O(l).

In other words, the reaction is complete, and the possibility of water dissociated into hydrogen and oxygen is very small. A negative value for DG^o indicates an exothermic reaction.

The Gibb's energy is the energy released other than pressure-volume work. This redox reaction to form water can be engineered to proceed in a Daniel cell. In this case, the energy is converted into electric energy according to this equation.

$$DGowater = - n F E = -237.13 \text{ kJ}$$

where *n* is the number of electrons (= 2) in the redox equation, *F* is the Faraday constant (= 96485 C), and *E* is the potential of the Daniel cell. Thus,

$$E = - \frac{-237130 \text{ J}}{2*96485 \text{ C}}$$

$$= 1.23 \text{ V}$$

Ideally, a reverse voltage of 1.23 V is required for the electrolysis of water. But in reality, a little over voltage is required to carry out the electrolysis to decompose water. Furthermore, pure water does not conduct electricity, and acid, base or salt is often

added for the electrolysis of water. This link has a demonstration.

Example 2

In order to carry out the electrolysis of water, 1.50 V is applied. Assume the energy not converted to chemical energy is converted to heat. How much heat is generated for the electrolysis of 1 mole water?

Solution

Ideally, 1.23 V will be used for the electrolysis. Energy due to the over voltage of 1.50 - 1.23 = 0.27 V is converted to heat.

Heat = 0.27 V * 2 * 96485 C
= 52102 J
= 52 kJ

DISCUSSION
The excess energy can also be evaluated using

Heat = n F *1.50 - 237130

This problem also illustrates the principle of conservation of energy.

Questions

1. **For the reaction**

   H2O(l) -> H2(g) + 0.5 OH2(g)

   the equilibrium constant as shown earlier is 1/3.5e41 = 2.9e-43 $atm^{3/2}$. What is the partial pressure of $H_2(g)$?

Solutions

1. Hint: 2.6e-64

   Skill -
   Evaluate this value please!

Contributors and Attributions

- Chung (Peter) Chieh (Professor Emeritus, Chemistry @ University of Waterloo)

This page titled 8.2.8: Water Chemistry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Chung (Peter) Chieh.

8.2.9: Water Physics

Physical Properties of Water

Chemical and physical properties of water are often discussed together. These properties are fundamentals of many disciplines such as hydrology, environmental studies, chemical engineering, environmental engineering, civil engineering etc. They are of interest to chemists and physicists of course.

Here are some highlights of the physical properties of water. Pure liquid water has a high heat capacity of 4.182 J K^{-1} g^{-1}; it is a good heat conductor, but a poor electric conductor. It is a solvent for dissolving ionic and polar substances, but interact with non-polar substance weakly. Surface tension of water is rather high, and little quantities aggregate into drops rather than spread out as thin layers.

Hydrogen bonding contributes to many of the physical and chemical properties, such as the unusual but normally known freezing point of 273.15 K, and 373.15 K respectively. The critical temperature and pressure are 647.3 K and 220.5 bar (22050 kPa) and critical volume = 0.056 m^3 $kmol^{-1}$.

Because of the many applications of water, some more details on the properties of water are desirable. Thus, keeping track of water properties is of national interest. For example, the American Society of Mechanical Engineers (ASME) is such an organization. International co-operation on research and information exchange are more economical, and for thermal properties, The International Association for the Properties of Water and Steam (IAPWS) is set up. Canada is a member of this organization.

From the application point of view, the variations of the following properties as functions of temperature and pressure are required.

- Compressibility of steam and water as a function of pressure at various temperature
- Density of water as a function of temperature
- Viscosity of water as a function of temperature
- Enthalpy of water for various thermodynamic evaluations
- Molar volumes and expansion coefficients of water and vapor as functions of pressure and temperature
- Speed of sound in water and vapor and speed of sound in air-vapor mixture as functions of temperature and pressure
- Entropy of water as a function of temperature and pressure
- Thermal conductivity of water and steam
- Viscosity of water at any temperature for pipe and pump design
- Dielectric constant as functions of temperature and pressure
- Surface tension as functions of temperature and pressure
- Gibbs energy at various temperature and pressure
- Properties such as dielectric constants and ion products of supercritical water (fluid)
- Applications of supercritical water: Plastic waste recycle, recovery of toluenediamine, hydrolysis of PET (polyethylene terephthalate) etc.

We will discuss some of these to illustrate the point, but not all of them.

Density of Water

Density of Water		
T /K	**Density g/mL**	**D_2O**
273	0.999841	1.10469
274	0.999900	
275	0.999941	
276	0.999965	

Density of Water		
277	0.999973	1.1057
278	0.999965	1.10562
279	0.999941	
280	0.999902	
281	0.999849	
282	0.999781	
281	0.999700	

Density is the mass per unit volume. The density of water is usually taken as 1.0 g/mL or 1.000e3 kg m^{-3} at 277 K. This suggests that the density varies with temperature and water density is the highest at 277 K, and the density between 273 and 281 K from the *CRC Handbook of Chemistry and Physics* are given in the Table here. These data are calculated from experimental data for pure water based on the standard at 276.98 K. The same source gives the density of ordinary water as 1.000000 g/mL at 277 K.

The volume occupied by one mole of substance is called the **molar volume**. The molar volume of liquid water is 18.016/*density*. At 277 K, the molar volume is 18.016 mL. For liquid water, the molar volumes of liquid water increase to 18.03 mL at both 269 K and 285 K.

The density of ice is 0.917 at 273 K, and the molar volume is 19.65 mL, 9% more than the molar of liquid. Thus, 9% of an ice cube containing no air bubble float above the surface, and 91% of it is below the waterline. The density makes behavior of icebergs interesting. Ice bergs are major tourist attractions in Newfoundland and Labrador, Canada.

Electric Dipole Moment and Dielectric Constant

Charged ions interact with each other due to electrostatic attraction or repulsion. The force *F* between two charge particles with charges q_1, and q_2 separated by a distance *r* is

$$F = \frac{q_1 q_2}{4 \pi \varepsilon_o r^2}$$

Cl d-
|
Na d+

Uncharged molecules still interact with each other, not due to electrostatic interaction, but due to electric dipole interaction. The electric dipole moment is a vector due to uneven distribution of unlike charges. In diatomic systems, the magnitude of the electric dipole moment can be estimated as the difference between the Pauling electronegativities of the two atoms. For convenience, let us assume that centres of positive and negatives of Na-Cl are separated by a distance l, then the electric dipole moment, μ is

$$m = q\, l$$

Traditionally, the dipole moments of molecules have been tabulated in electrostatic units, in which case the charge of an electron is 4.80e-10 esu (= (1.6e-19 C) (3e9 esu/C)). In NaCl crystals, the distance between Na and Cl ions is 240 pm. If the NaCl molecule (in a gas) has the same distance between the ideallized ions, then the dipole moment is calculated below:

$$m = q\, l$$
$$= (1.60e\text{-}19)\ (240e\text{-}12\ m)$$
$$= 3.84e\text{-}29\ C\ m.$$

or in cgs-esu units

$$m = 4.8e\text{-}10\ esu * 2.40e\text{-}8\ cm$$
$$= 11.5e\text{-}18\ esu\ cm$$

In the cgs-esu unit, 1e-18 esu cm is define a Debye (symbol D). Thus, we have

1 D = 1e-18 esu cm
= 3.355e-30 C m (from the calculation above)
m = 11.5 D for ideallized NaCl gas molecule
but
mobserved = 9 D in NaCl gas =3e-29 C m.

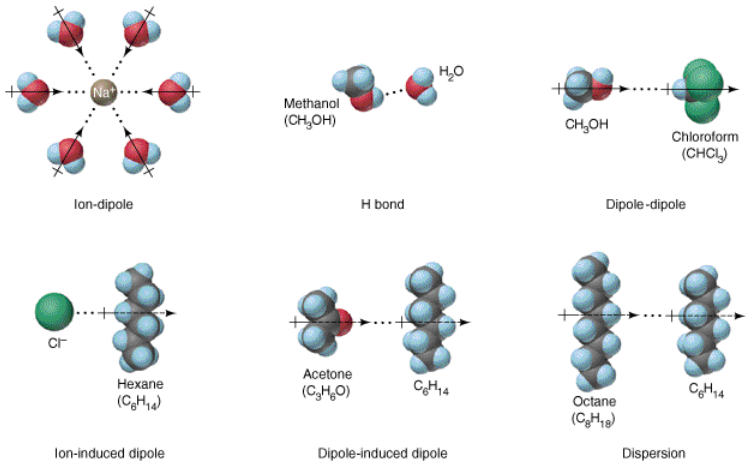
Electric Dipole Moment of Some Gas Molecules	
Molecule	**m /D**
NaCl	9.0
KCl	10.3
CO	0.1
HF	1.8
HCl	1.1
HBr	0.8
H_2O	**1.8**
SO_2	1.6
N_2O	0.2
NH_3	1.5

However, experimental dipole moment = 9 D for NaCl gas. Thus, the model for the calculation has to be modified to account for the partial delocalization of the charges or by including some covalent character in the Na-Cl bond. In any case, the model shows a physical method (model) for the evaluation of dipole moment. Dipole moments of some gas molecules are given in the table here.

The dipole moment is a vector from the negative to the positive charge along the bond. For triatomic moecules such as those of water, the total electric dipole moment is the sum of al dipoles for each bond. The experimental dipole moment for water is 1.8 D, which is the same as that of H-F. Water is a very polar compound. Ammonia with three N-H bonds has a dipole moment of 1.5 D.

Homonuclear diatomic molecules have zero dipole moment, of course. So are linear CO_2 and CS_2 molecules.

The high dipole moment makes water a very special substance. Water has a very high dielectric constant, 80. Due to dipole-charge interaction, water is the universal solvent for ionic substances, especially mono-valent ions. The diagram below shows some typical ion-dipole and dipole-dipole interaction in solutions. The dissolution in water is called hydration.



The dipole moment has something to do with its interaction with microwave. Application of Ground Penetrating Radar in Glaciology is a web site that gives the theory of radar and discusses the interaction of radar with water. It further illustrates the application in glaciology.

In contrast, supercritical water has a low dielectric constant, making it a good solvent for non-polar substances.

Example 1

Let us assume the H-F molecule as composed of two ions, a positive and a negative ion. What is the distance that separating these two ions in order to give a dipole moment of 1.8 D?

Solution
Recall the first formula from the discussion of dipole moment above, and proceed with the calculation as shown:

$$m = q\, l$$

$$q\, l = 4.8e\text{-}10 \text{ esu} * l$$
$$= 1.8e\text{-}18 \text{ esu cm.}$$

$$l = \frac{1.8e\text{-}18 \text{ esu cm}}{4.8e\text{-}10 \text{ esu}} = 0.375e\text{-}8 \text{ cm (37.5 pm)}$$

DISCUSSION
What is the normal H-F bond distance? (Ans. 92 pm).

Dividing the apparent bondlength of 38 pm by the observed bondlength 92 pm gives the ionic character 38/92 = 41 % ionic character.

Questions

1. **In a fire or nuclear power generator, steam at 600°C under pressure is often used to transfer the heat from the boiler to the turbine. What phase is this steam?**

Solutions

1. Hint: supercritical fluid of water.

Contributors and Attributions

- Chung (Peter) Chieh (Professor Emeritus, Chemistry @ University of Waterloo)

This page titled 8.2.9: Water Physics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.2.10: Water Treatment

Water treatment is` a process of making water suitable for its application or returning its natural state. Thus, water treatment required before and after its application. The required treatment depends on the application. For example, treatment of greywater (from bath, dish and wash water) differs from the black water (from flush toilets). Composting toilet is not allowed in urban dwelling. Yet, composting toilets are used in a 30,000-square-foot office complex at the Institute of Asian Research, University of British Columbia.

Water treatment involves science, engineering, business, and art. The treatment may include mechanical, physical, biological, and chemical methods. As with any technology, science is the foundation, and engineering makes sure that the technology works as designed. The appearance and application of water is an art.

In terms of business, RGF Environmental, Water Energy Technologies, Aquasana Store, Vitech, Recalyx Industrial SDN BHD and PACE Chemicals ltd are some of many companies that offer various processes for water treatment. Millipore, a Fisher Scientific partner, offers many lines of products to produce *ultrapure water*, using a combination of active charcoal membranes, and reverse osmosis filter. Internet sites of these companies offer useful information regarding water.

An environmental scientist or consultant matches the service provider, modify if necessary, with the requirement.

- Natural Water includes some discussion on hard and soft water. Softening hard water for boiler, cooler, and domestic application is discussed therein. These treatments prepare water so that it is suitable for the applications.
- Water Biology deals with water and biology. Drinking water is part of making water suitable for living. Thus, this link gives some considerations to drinking water problems.
- There are many different industry types, and waters from various sources are usually treated before and after their applications. Pre-application treatment and wastewater treatment offer a special opportunity or challenge. Only a general consideration will be given to some industrial processes.
- General municipal and domestic wastewater treatment converts used water (waste) into environmentally acceptable water or even drinking water. Every urban centre requires such a facility.

General Wastewater Treatment

Water is a renewable resource. All water treatments involve the removal of solids, bacteria, algae, plants, inorganic compounds, and organic compounds. Removal of solids is usually done by filtration and sediment. Bacteria digestion is an important process to remove harmful *pollutants*. Converting used water into environmentally acceptable water or even drinking water is **wastewater treatment**

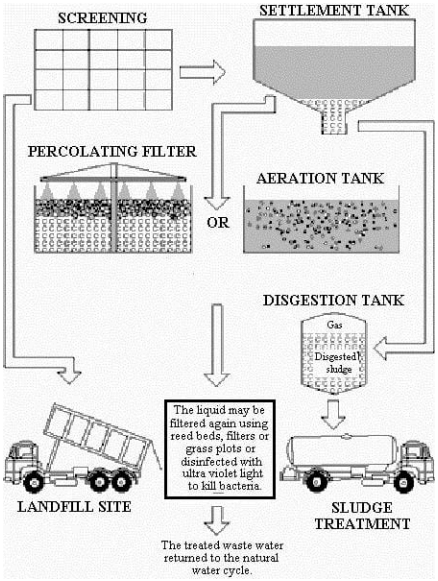
Water in the Great Lakes Region is an organization dealing with the water resources. Ontario Clean Water Agency (OCWA) is a provincial Crown corporation in business to provide environmentally responsible and cost-efficient water and wastewater services. It currently operates more than 400 facilities for 200 municipalities. This web site provides information on water and water treatment.

In April 1993, 403,000 people in Milwaukee were ill as a result of *cryptosporidium* contaimination of water due to spring run off. This outbreak caused the more stringent regulations to be implemented in the public dringking water system. The measures were aimed at removing cryptosporidium.

In May 2000, due to torrential downpour surface water got into shallow wells in a small town Walkerton, Ontario, Canada. On May 17, some residents complained of fever, bloody diarrhea and vomiting. This was know as the Walkerton *E. Coli* Outbreak. Nearly half of the population of the town fell ill, and several people died due to the E. Coli O157:H7 infection. A public inquiry recommended many measures to prevent similar outbreaks. These measures were aimed at eliminating E. Coli.

Sewage Treatment

As a general discussion, let us look at a typical process in sewage treatment. A flow diagram for a general sewage treatment plant from Water Education, Department of Computer Science, University of Exeter, U.K., is shown below:



Sewage is SCREENED to remove large solid chunks, which are disposed in LANDFILL SITE. It flows over to the SETTLEMENT TANK to let the fine particles to settle. The settlement is called the activated SLUDGE. The supernatant is then PERCOLATING FILTERED and/or AERATED. The water can be filtered again, and then disinfected (chlorinated in most cases). When there is no other complication, the water is returned to nature back to the ecological cycle.

The SLUDGE removed from the settlement is composed of living biological material. A portion of it may be returned to the AERATION TANK, but the raw SLUDGE is digested by both microorganism. Anaerobic (without oxygen) and aerobic (with air) bacteria digestions are used. At the digestion stage, carbon dioxide, ammonia, and methane gases are evolved. Volume of the digested sludge is reduced, and it is acceptable as a fertilizer supplement in farming.

Wastewater Treatment

Although the sewage water may be discharged back to the ecological system after AERATED DIGESTION and PERCOLATING FILTRATION, but in some cases, further treatment is required. Some general consideration of water treatment is given below.

A rather recent book, *Chemistry of Water Treatment* by S.D. Faust and O.M. Aly, 2nd Ed. (1998) [TD433 F38 1998], addresses the problem of quality natural and treated water.

The first three chapters discuss the criteria and standards for drinking water quality, organic compounds in waters, taste and order of water. Understandably, the standards change over the years. So are the standards of treated waters. Guidelines are available from government agencies such as Environment Canada which is equivalent to U.S. Public Health Service and the Environment Protection Agency (EPA). We have talked about drinking water in Water Biology.

Next seven chapters deal with the removal of the following:

- organics and inorganics by activated carbon
- particulate matter by coagulation
- particulate matter by filtration and sedimentation
- hardness and other scale-forming substances
- inorganic contaminants
- corrosive substances
- pathogenic (disease producing) bacteria, viruses, and protozoans (microorganisms).

There is a chapter dealing with **aeration**

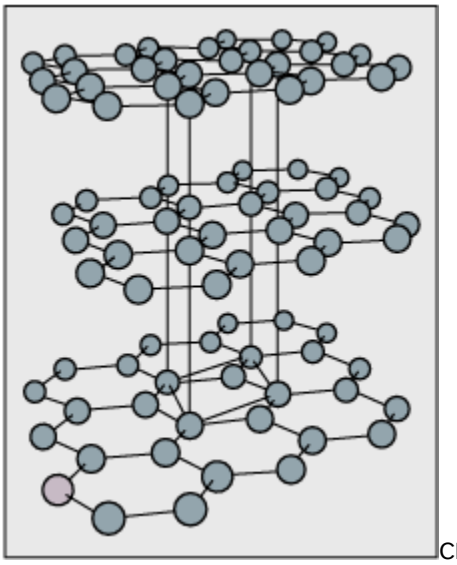
These items cover the chemistry, biology, and physics involved in the treatment of water. Some of these topics have been discussed in chemistry of water, physical properties of water, biology of water, and natural water. Introductions are going to be given to some selected topics below.

Treatment by Activated Carbon

Treatment by activated carbon is mostly due to adsorption or absorption. When a chemical species is adhered to the surface of a solid, it is an **adsorption**. When partial chemical bonds are formed between adsorbed species or when the absorbate got into the channels of the solids, we call it **absorption**. However, these two terms are often used to mean the same, because to distinguish one from type from the other is very difficult.

Application of activated charcoal for the removal of undesirable order and taste in drinking water has been recognized at the dawn of civilization. Using bone char and charred vegetation, gravel, and sand for the filtration of water for domestic application has been practised for thousands of years. Active research and production of activated charcoal was accelerated during the two world wars. The use of poison gas prompted the development of masks. They are still in use today.

Charcoal absorbs many substances, ranging from colored organic particulates to inorganic metal ions. Charcoal has been used to remove the colour of raw sugar from various sources.



Charcoal consists of microcrystallites of graphite. The particles are so small in charcoal that they were considered amorphous. The crystal structure of graphite consists of layers of hexagonal networks, stacked on top of each other. Today, making activated carbon is a new and widely varied industry. Other molecules attach themselves to the porous surface and dangling carbons in these microcrystallites.

Carbon containing substances are charred at less than 900 K to produce carbon in the manufacture of **activated carbon**. However, the carbon is activated at 1200 K using oxidizing agent to selectively oxidize portions of the char to produce pores in the material. Because of the special process to produce used, these materials with high surface to mass ratio, they are called **activated carbon** rather than activated charcoal. Factors affecting the absorption are particle size, surface area, pore structure, acidity (pH), temperature, and the nature of the material to be absorbed. Usually, adsorption (absorption) equilibria and rate of adsorption must be considered for effective applications.

Coagulation, Flucculation and Sedimentation

Natural and wastewater containing small particulates. They are suspended in water forming a **colloid**. These particles carry the same charges, and repulsion prevents them from combining into larger particulates to settle. Thus, some chemical and physical techniques are applied to help them settle. The phenomenon is known as **coagulation**. A well known method is the addition of electrolyte. Charged particulates combine with ions neutralizing the charges. The neutral particulates combine to form larger particles, and finally settle down.

Another method is to use high-molecular-weight material to attract or trap the particulates and settle down together. Such a process is called **flocculation**. Starch and multiply charged ions are often used.

Historically, dirty water is cleaned by treating with alum, $Al_2(SO_4)_3 \cdot 12\ H_2O$, and lime, $Ca(OH)_2$. These electrolytes cause the pH of the water to change due to the following reactions:

Al2(SO4)3.12 H2O, -> Al(aq)3+ + 3 SO4(aq)2- + 12 H2O
SO4(aq)2- + H2O -> HSO4(aq)- + OH- (causing pH change)
Ca(OH)2 -> Ca(aq)2+ + 2 OH- (causing pH change)

The slightly basic water causes $Al(OH)_3$, $Fe(OH)_3$ and $Fe(OH)_2$ to precipitate, bringing the small particulates with them and the water becomes clear. Some records have been found that Egyptians and Romans used these techniques as early as 2000 BC.

Suspension of iron oxide particulates and humic organic matter in water gives water the yellow muddy appearance. Both iron oxide particulates and organic matter can be removed from coagulation and flocculation. The description given here is oversimplified, and many more techniques have been applied in the treatment of water. Coagulation is a major application of lime in the treatment of wastewater.

Other salts such as iron sulfates $Fe_2(SO_4)_3$ and $FeSO_4$, chromium sulfate $Cr_2(SO_4)_3$, and some special polymers are also useful. Other ions such as sodium, chloride, calcium, magnesium, and potassium also affect the coagulation process. So do temperature, pH, and concentration.

Disposal of coagulation sludge is a concern, however.

Sedimentation let the water sit around to let the floculated or coagulated particles to settle out. It works best with relatively dense particles (e.g. silt and minerals), while flotation works better for lighter particles (e.g. algae, color). A settling tank should be big enough so that it takes a long time (ideally 4 hours +) to get through. Inlets and outlets are designed so the water moves slowly in the tank. Long and narrow channels are installed to let the water to snake its way through the tank. The settled particles, sludge, must occasionally be removed from the tanks. The water is next ready to be filtered. Sedimentation is used in pre-treatment and wastewater treatment.

Filtration

Filtration is the process of removing solids from a fluid by passing it through a porous medium. Coarse, medium, and fine porous media have been used depending on the requirement. The filter media are artificial membranes, nets, sand filter, and high technological filter systems. The choice of filters depends on the required filtering speed and the *cleanness* requirement. The flow required for filtration can be achieved using gravity or pressure. In pressure filtration, one side of the filter medium is at higher pressure than that of the other so that the filter plane has a pressure drop. Some portion of this filter type must be enclosed in a container.

The process of removing the clogged portion of the filter bed by reversing the flow through the bed and washing out the solid is called **back washing**. During this process, the solid must be removed out of the system, but otherwise the filters must be either replaced or taken out of service to be cleaned.



Aqua-Rain manufacture water filters as shown here. This unit consists of four filters. Regarding the filtering system, its technical info gave the following statement.

> *At the heart of the AquaRain™ Water Filtration System are Marathonr State-of-the-art ceramic elements utilizing a long-proven filtration process that is over 100 years old which will safely remove dangerous waterborne pathogens such as cysts (Cryptospordium, Giardia lamblia) and bacteria (E. coli, Samonelli typhus, etc...). These innovative Marathonr ceramic elements are also filled with a high grade silvered granulated activated carbon (GAC). The GAC reduces pesticides, chemicals, chlorine, tastes & odors, while leaving the naturally occurring minerals found in the water unaffected.*

The units are designed for emergency and perhaps undeveloped countries.

AquaSelect of Mississauga has a pitcher water filter system, and *its cartridge contains hundreds of high efficiency activated carbon and ion exchange beads*, its web site claims. Brita filters is very popular.

Aeration

Bringing air into intimate contact with water for the purpose of exchanging certain components between the two phases is called **aeration**. Oxygenation is one of the purposes of aeration. Others are removal of volatile organic substances, hydrogen sulfide, ammonia, and volatile organic compounds.

A gas or substance dissolved in water may further react with water. Such a reaction is called **hydration**. Ionic substance dissolve due to hydration, for example:

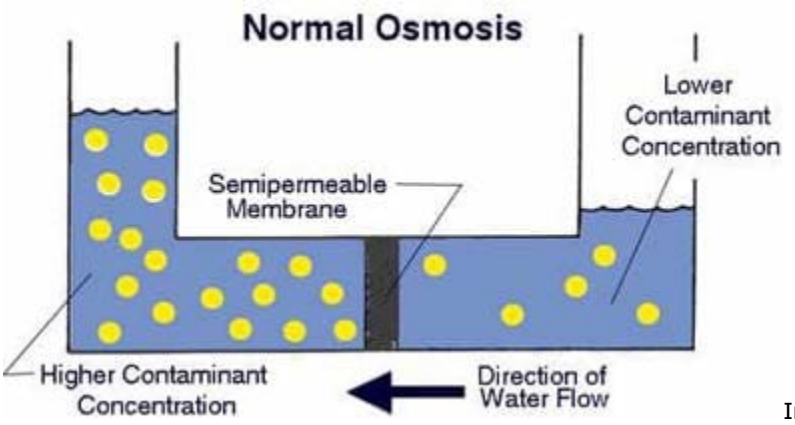
$$\mathrm{HCl\,(g)} + x\,\mathrm{H_2O} \rightleftharpoons \mathrm{H(H_2O)}_x{}^{+} + \mathrm{Cl(aq)}^{-} \tag{8.2.10.1}$$

$$\mathrm{H_2S} \rightleftharpoons \mathrm{H^{+}\,(aq)} + \mathrm{HS^{-}\,(aq)} \tag{8.2.10.2}$$

These reactions are reversible, and aeration may also causes dehydration resulting in releasing the gas from water. Henry's law is applicable to this type of equilibrium for consideration. Methods of aeration are

- Diffused aeration - Air bubbles through water.
- Spray aeration - Water is sprayed through air.
- Multiple-tray aeration - Water flows through several trays to mix with air.
- Cascade aeration - Water flows downwards over many steps in the form of thin water falls.
- Air stripping - A combination of multiple tray and cascade technique plus random packed blocks causing water to mix thoroughly with air.

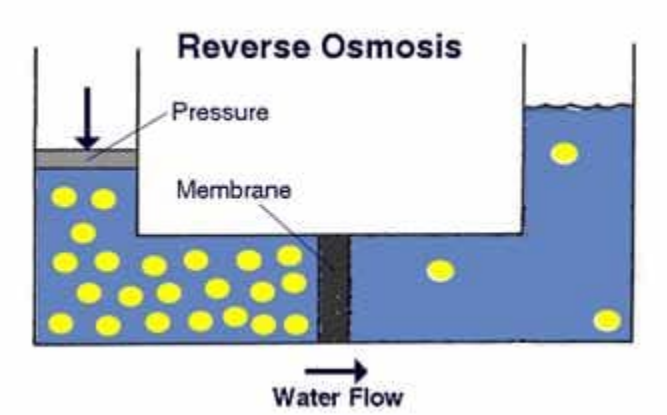
Reverse Osmosis Water Filter System



In the following discussion, a dilute solution and a concentrated solution are considered. The dilute solution can be a clean water whereas the concentrated solution contains undesirable solute (electrolyte or others).

When a compartment containing a dilute solution is connected to another compartment containing a concentrated solution by a semipermeable membrane, water molecules move from the dilute solution to concentrated solution. This phenomenon is called osmosis. Pig bladders are natural semipermeable membranes. As the water molecules migrate through the semipermeable membrane, water level in the solution will increase until the (osmotic) pressure prevents a net migration of water molecules in one direction. A pressure equivalent to the height difference is called the osmotic pressure. The illustration given on the right is from

the PurePro, one of the many companies that manufacture reverse osmosis water filter devices. Millipore also use this technique.



By applying pressure in the higher concentration solution, water molecules migrate from a high concentration solution to a low concentration solution. This method is called **reverse osmosis water filter system**. The concept of reverse osmosis is illustrated in the diagram here from PurePro.

In this technique, the membrane must be able to tolerate the high pressure, and prevent solute molecules to pass through. Regarding membranes, PurePro made the following statement:

> *Semipermeable membranes have come a long way from the natural pig bladders used in the earlier osmosis experiments. Before the 1960's, these membranes were too inefficient, expensive, and unreliable for practical applications outside the laboratory. Modern advances in synthetic materials have generally solved these problems, allowing membranes to become highly efficient at rejecting contaminants, and making them tough enough to withstand the greater pressures necessary for efficient operation.*

This technology certainly works, and it has been used to convert salt (ocean or sea) water into fresh water. With this technique, the water with higher concentration is discharged. Thus, this technology is costly in regions where the water cost is high. Free Drinking Water also uses reverse osmosis filter system for domestic applications.

Industrial Wastewater Treatment

The Environment Canada's (Atlantic Region) Waste Management and Remediation Section contains links to detailed information on the programs and activities relating to Petroleum and Allied Petroleum Storage Tank Systems, Ocean Disposal, Contaminated Sites, and Hazardous Waste Disposal Advice for the provinces of Newfoundland and Labrador, Nova Scotia, New Brunswick and Prince Edward Island. It also contains links to sites within and outside of Environment Canada with related information.

Contributors and Attributions

- Chung (Peter) Chieh (Professor Emeritus, Chemistry @ University of Waterloo)

This page titled 8.2.10: Water Treatment is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Chung (Peter) Chieh.

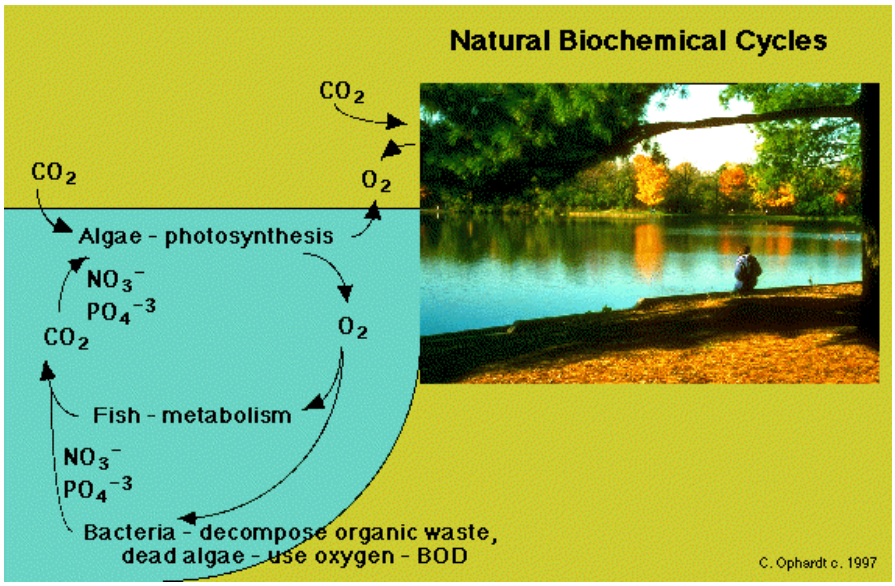
8.3: BIOCHEMICAL CYCLES

8.3.1: Biochemical Cycles
8.3.2: Carbon Cycle
8.3.3: Nitrogen Cycle

This page titled 8.3: Biochemical Cycles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.3.1: Biochemical Cycles

Plants such as trees and algae undergo the photosynthesis reaction where carbon dioxide and water in the presence of sunlight are converted to organic materials and oxygen. An important reverse reaction occurs in the water: Fish use metabolism where oxygen and organic materials - other small fish or algae - as food is converted to carbon dioxide, water, and energy. Bacteria in water, as well as land, also undergo metabolism and use oxygen and decompose organic wastes as food to convert to carbon dioxide, water, and energy. By products in the decomposition of organic waste are nitrates and phosphates. The major natural biochemical cycles include the carbon, nitrogen, and phosphate cycles. They are presented in brief in this graphic.



The overall health of a body of water depends upon whether these factors are in balance. Municipal sewage systems are now doing a better job of removing most of the organic waste products in the discharge water, but some organic waste still enters the streams and lakes. If an excess amount of organic waste is present in the water, the bacteria use all of the available oxygen in the water in an attempt to decompose the organic waste.

The amount of organic waste in water is represent by a chemical test called BOD - Biological Oxygen Demand. The concentration of oxygen is measured in a water sample at the beginning of the test and again after five days. The difference between the oxygen concentrations represents the amount of oxygen consumed by the bacteria in the metabolism of the waste organics present.

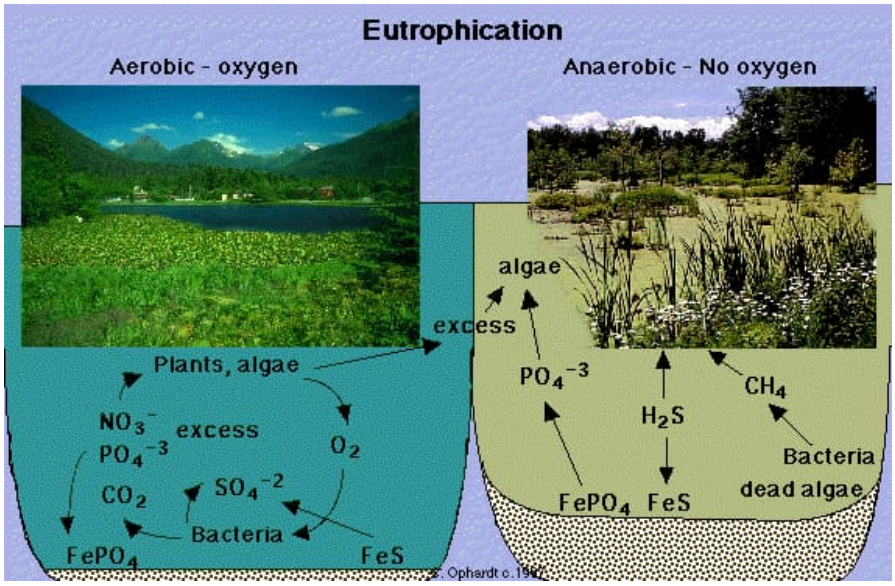
Eutrophication

In situations where eutrophication occurs, the natural cycles are overwhelmed by an excess of one or more of the following: nutrients such as nitrate or phosphate, or organic waste.

In the first case under aerobic conditions (presence of oxygen), the natural cycles may be more or less in balance until an excess of nitrate and/or phosphate enters the water. At this time the water plants and algae begin to grow more rapidly than normal. As this happens there is also an excess die off of the plants and algae as sunlight is blocked at lower levels. Bacteria try to decompose the organic waste, consuming the oxygen, and releasing more phosphate and nitrate to begin the cycle anew. Some of the phosphate may be precipitated as iron phosphate to remove the soluble form from the water solution.

In the second case under anaerobic conditions (absence of oxygen), as conditions worsen as more phosphates and nitrates may be added to the water, all of the oxygen may be used up by bacteria in trying to decompose all of the waste. Different bacteria continue to carry on decomposition reactions, however the products are drastically different. The carbon is converted to methane gas instead of carbon dioxide, sulfur is converted to hydrogen sulfide gas. Some of the sulfide may be precipitated as iron sulfide. Under anaerobic conditions the iron phosphate in the sediments may be solubilized into solution to make it available as a nutrient for the algae which would start the growth and decay cycle over again. The pond may gradually fill with undecayed plant materials

to make a swamp.

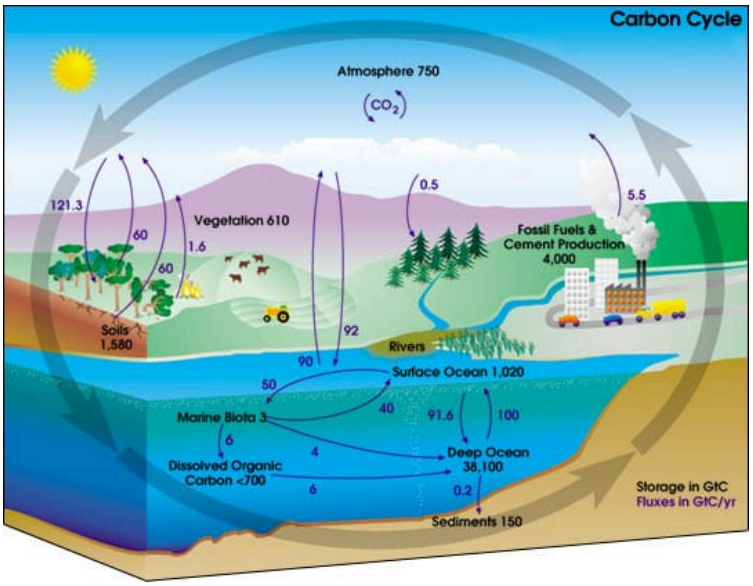


Contributors

- Charles Ophardt, Professor Emeritus, Elmhurst College; Virtual Chembook

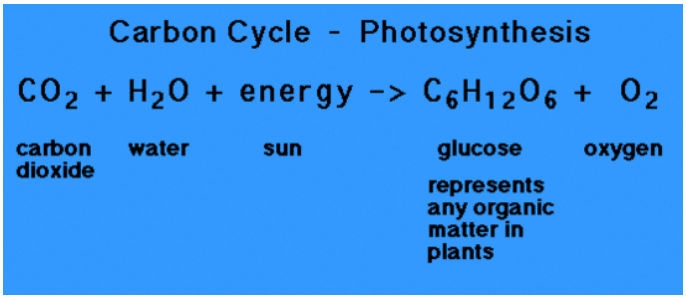
This page titled 8.3.1: Biochemical Cycles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.3.2: Carbon Cycle



Photosynthesis

Photosynthesis is a complex series of reactions carried out by algae, phytoplankton, and the leaves in plants, which utilize the energy from the sun. The simplified version of this chemical reaction is to utilize carbon dioxide molecules from the air and water molecules and the energy from the sun to produce a simple sugar such as glucose and oxygen molecules as a by product. The simple sugars are then converted into other molecules such as starch, fats, proteins, enzymes, and DNA/RNA i.e. all of the other molecules in living plants. All of the "matter/stuff" of a plant ultimately is produced as a result of this photosynthesis reaction. An important summary statement is that during photosynthesis plants use carbon dioxide and produce oxygen.

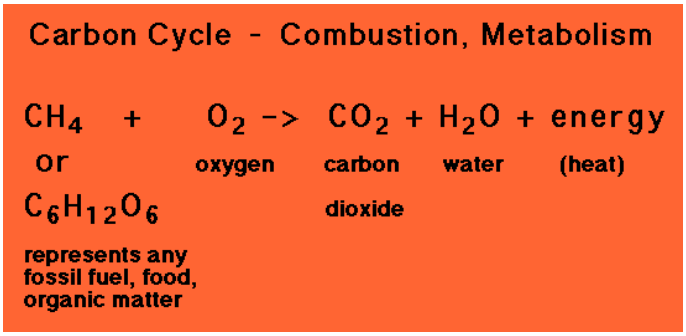


Combustion/Metabolism Reaction

Combustion occurs when any organic material is reacted (burned) in the presence of oxygen to give off the products of carbon dioxide and water and ENERGY. The organic material can be any fossil fuel such as natural gas (methane), oil, or coal. Other organic materials that combust are wood, paper, plastics, and cloth. Organic materials contain at least carbon and hydrogen and may include oxygen. If other elements are present they also ultimately combine with oxygen to form a variety of pollutant molecules such as sulfur oxides and nitrogen oxides.

Metabolism occurs in animals and humans after the ingestion of organic plant or animal foods. In the cells a series of complex reactions occurs with oxygen to convert for example glucose sugar into the products of carbon dioxide and water and ENERGY. This reaction is also carried out by bacteria in the decomposition/decay of waste materials on land and in the water.

An important summary statement is that during combustion/metabolism oxygen is used and carbon dioxide is a product. The whole purpose of both processes is to convert chemical energy into other forms of energy such as heat.

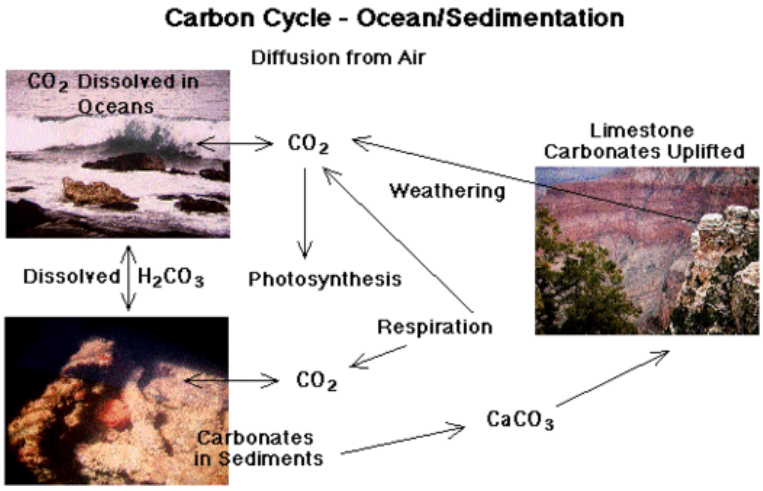


Sedimentation

Carbon dioxide is slightly soluble and is absorbed into bodies of water such as the ocean and lakes. It is not overly soluble as evidenced by what happens when a can of carbonated soda such as Coke is opened. Some of the dissolved carbon dioxide remains in the water, the warmer the water the less carbon dioxide remains in the water.

Some carbon dioxide is used by algae and phytoplankton through the process of photosynthesis.

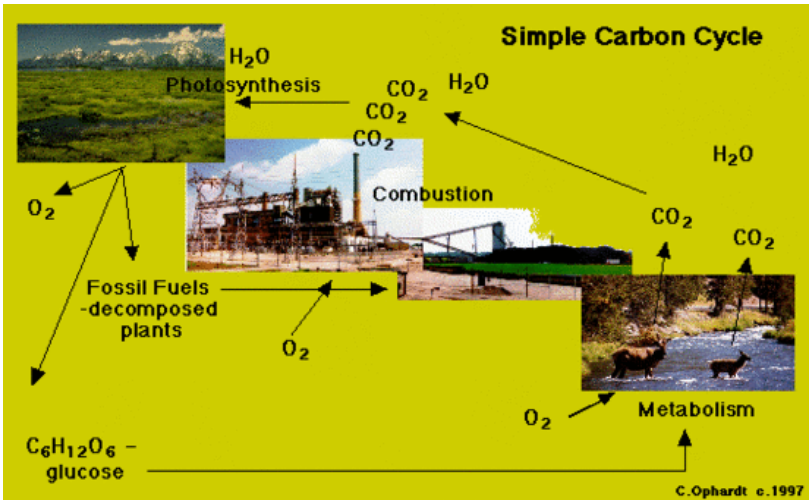
In other marine ecosystems, some organisms such as coral and those with shells take up carbon dioxide from the water and convert it into calcium carbonate. As the shelled organisms die, bits and pieces of the shells fall to the bottom of the oceans and accumulate as sediments. The carbonate sediments are constantly being formed and redissolved in the depths of the oceans. Over long periods of time, the sediments may be raised up as dry land or into mountains. This type of sedimentary rock is called limestone. The carbonates can redissolve releasing carbon dioxide back to the air or water.



Human Impacts: Fossil Fuels

In the natural carbon cycle, there are two main processes which occur: photosynthesis and metabolism. During photosynthesis, plants use carbon dioxide and produce oxyge and during metabolism oxygen is used and carbon dioxide is a product. Humans impact the carbon cycle during the combustion of any type of fossil fuel, which may include oil, coal, or natural gas. Fossil Fuels were formed very long ago from plant or animal remains that were buried, compressed, and transformed into oil, coal, or natural gas. The carbon is said to be "fixed" in place and is essentially locked out of the natural carbon cycle. Humans intervene during by burning the fossil fuels. During combustion in the presence of air (oxygen), carbon dioxide and water molecules are released into the atmosphere. The question becomes as to what happens to this extra carbon dioxide that is released into the atmosphere. This is the subject of considerable debate and about it possible effect in enhancing the greenhouse ffect which may than result

in global warming.



Contributors

- Charles Ophardt, Professor Emeritus, Elmhurst College; Virtual Chembook

This page titled 8.3.2: Carbon Cycle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

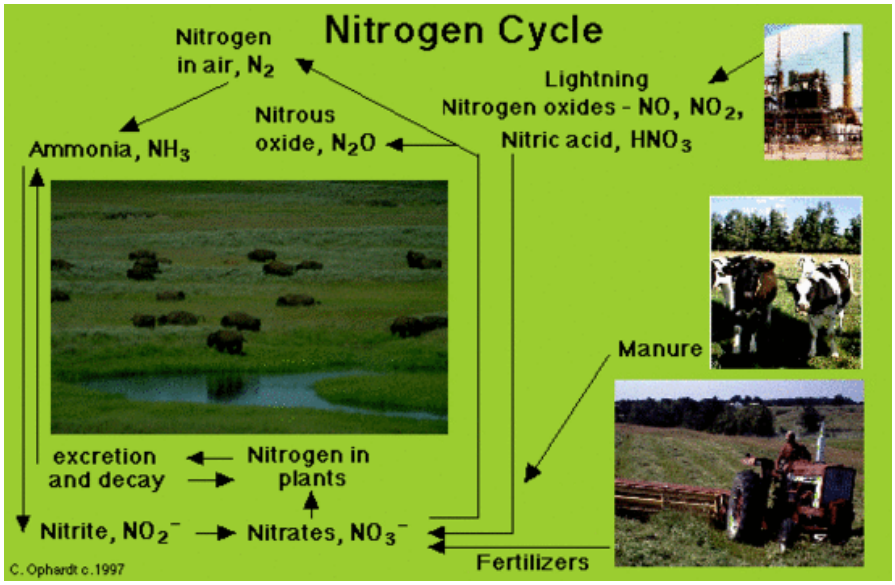
8.3.3: Nitrogen Cycle

The main component of the nitrogen cycle starts with the element nitrogen in the air. Two nitrogen oxides are found in the air as a result of interactions with oxygen. Nitrogen will only react with oxygen in the presence of high temperatures and pressures found near lightning bolts and in combustion reactions in power plants or internal combustion engines. Nitric oxide, NO, and nitrogen dioxide, NO_2, are formed under these conditions. Eventually nitrogen dioxide may react with water in rain to form nitric acid, HNO_3. The nitrates thus formed may be utilized by plants as a nutrient.

Nitrogen in the air becomes a part of biological matter mostly through the actions of bacteria and algae in a process known as nitrogen fixation. Legume plants such as clover, alfalfa, and soybeans form nodules on the roots where nitrogen fixing bacteria take nitrogen from the air and convert it into ammonia, NH_3. The ammonia is further converted by other bacteria first into nitrite ions, NO^{2-}, and then into nitrate ions, NO^{3-}. Plants utilize the nitrate ions as a nutrient or fertilizer for growth. Nitrogen is incorporate in many amino acids which are further reacted to make proteins.

Ammonia is also made through a synthetic process called the Haber Process. Nitrogen and hydrogen are reacted under great pressure and temperature in the presence of a catalyst to make ammonia. Ammonia may be directly applied to farm fields as fertilizer. Ammonia may be further processed with oxygen to make nitric acid. The reaction of ammonia and nitric acid produces ammonium nitrate which may then be used as a fertilizer. Animal wastes when decomposed also return to the earth as nitrates.

To complete the cycle other bacteria in the soil carry out a process known as denitrification which converts nitrates back to nitrogen gas. A side product of this reaction is the production of a gas known as nitrous oxide, N_2O. Nitrous oxide, also known as "laughing gas" - mild anesthetic, is also a greenhouse gas which contributes to global warming



Contributors

- Charles Ophardt, Professor Emeritus, Elmhurst College; Virtual Chembook

This page titled 8.3.3: Nitrogen Cycle is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.4: TOXICOLOGY MSDT

One of the challenges of recruiting undergraduates into the toxicology graduate research programs is a lack toxicology training programs and courses at the undergraduate level. Course content developed through this program are accessible to the general public. The intent is to provide content where other undergraduate students, K-12 educators, and the interested public encounter content providing an introduction to the discipline of toxicology.

8.4.1: Pathophysiology

8.4.1.1: What is Pathophysiology?

8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity

8.4.1.3: Cellular Response to Toxicant-Induced Injury

8.4.1.4: Repair and Adaptation

8.4.1.5: Patterns of Toxic Injury

8.4.1.6: Final Evaluation

8.4.2: Biochemistry and Molecular Genetics

8.4.2.1: Introduction to Biomolecules and Cell Components

8.4.2.2: Cell Structure and Subcellular Compartments

8.4.2.3: DNA and RNA Metabolism

8.4.2.4: Epigenetic Mechanisms

8.4.2.5: Section 2 Final Evaluation

8.4.3: Principles of Genetic Toxicology

8.4.3.1: Introduction to Genetic-toxicology Assay

8.4.3.2: Different Genetic Damages or Mutations

8.4.3.3: Different Genetic-Toxicology Assays

8.4.3.4: Different Cytotoxicity Assays

8.4.3.5: Epigenetics Assay

8.4.3.6: Section 3 Final Evaluation

8.4.4: Applied Systems Toxicology

8.4.4.1: Systems Toxicology

8.4.4.2: Dose Level and Applied Toxicology

8.4.4.3: Tools and Technologies in Systems Toxicology

8.4.4.4: Other Approaches for Predictive Toxicity Modeling

8.4.4.5: Technologies Used In Systems Biology/Toxicology

8.4.4.6: Takeaways Summary

8.4.4.7: Section 4 Final Evaluation

8.4.5: Regulatory Toxicology

8.4.5.1: Introduction to Regulatory Toxicology

8.4.5.2: Global Regulatory Toxicology

8.4.5.3: Topic 3- Regional Regulatory Toxicology

8.4.5.4: National Regulatory Toxicology

8.4.5.5: State Regulatory Toxicology
8.4.5.6: Non-Governmental Regulatory Toxicology
8.4.5.7: Section 5 Final Evaluation

8.4.6: Principles of Toxicology

8.4.6.1: Introduction to Toxicology
8.4.6.1.1: What is toxicology?
8.4.6.1.2: Basic Terminology

8.4.6.2: Dose and Dose Response
8.4.6.2.1: Dose and It's Impact on Toxicity
8.4.6.2.2: The Dose Response Relationship
8.4.6.2.3: Dose Estimates of Toxic Effects
8.4.6.2.4: Determining the Safety of a Drug
8.4.6.2.5: NOAEL and LOAEL

8.4.6.3: Toxic Effects
8.4.6.3.1: Types of Toxic Effects
8.4.6.3.2: Factors Affecting Toxicity
8.4.6.3.3: Systemic Toxic Effects
8.4.6.3.4: Organ Specific Toxic Effects

8.4.6.4: Interactions
8.4.6.4.1: Interactions

8.4.6.5: Toxicity Testing Methods
8.4.6.5.1: Testing and Assessing Toxicity
8.4.6.5.2: Clinical Investigations and Other Types of Human Data
8.4.6.5.3: Epidemiology Studies

8.4.6.6: Risk Assessment
8.4.6.6.1: Risk Assessment
8.4.6.6.2: Hazard Identification
8.4.6.6.3: Dose-Response Assessment
8.4.6.6.4: Exposure Assessment
8.4.6.6.5: Risk Characterization

8.4.6.7: Exposure Standards and Guidelines
8.4.6.7.1: Exposure Standards and Guidelines
8.4.6.7.2: Regulation of Consumer Products and Drug Safety
8.4.6.7.3: Environmental Exposure Standards/Guidelines
8.4.6.7.4: Occupational (Workplace) Exposure Standards/Guidelines/Approaches

8.4.6.8: Basic Physiology
8.4.6.8.1: Introduction to Basic Physiology
8.4.6.8.2: Homeostasis
8.4.6.8.3: Organs and Organ Systems

8.4.6.8.4: Tissues
8.4.6.8.5: Cells
8.4.6.8.6: Chemicals

8.4.6.9: Introduction to Toxicokinetics
8.4.6.9.1: What is Toxicokinetics

8.4.6.10: Absorption
8.4.6.10.1: Introduction to Absorption
8.4.6.10.2: Gastrointestinal Tract
8.4.6.10.3: Respiratory Tract
8.4.6.10.4: Dermal Route
8.4.6.10.5: Other Routes of Exposure

8.4.6.11: Distribution
8.4.6.11.1: Introduction to Distribution
8.4.6.11.2: Influence of Route of Exposure
8.4.6.11.3: Disposition Models
8.4.6.11.4: Structural Barriers to Distribution
8.4.6.11.5: Storage Sites

8.4.6.12: Biotransformation
8.4.6.12.1: Introduction to Biotransformation
8.4.6.12.2: Chemical Reactions
8.4.6.12.3: Biotransformation Sites
8.4.6.12.4: Modifiers of Biotransformation

8.4.6.13: Excretion
8.4.6.13.1: Introduction to Secretion
8.4.6.13.2: Urinary Excretion
8.4.6.13.3: Fecal Excretion
8.4.6.13.4: Exhaled Air
8.4.6.13.5: Other Routes

8.4.6.14: Cellular Toxicology
8.4.6.14.1: Adaptation
8.4.6.14.2: Cell Damage and Tissue Repair
8.4.6.14.3: Cancer
8.4.6.14.4: Neurotoxicity

8.4.6.15: Intuitive Toxicology and Risk Communication
8.4.6.15.1: Intuitive Toxicology
8.4.6.15.2: Risk Communication

8.4.6.16: Environmental Toxicology, Environmental Health, and One Health
8.4.6.16.1: Environmental Toxicology
8.4.6.16.2: Environmental Health

8.4.6.16.3: One Health

8.4.6.17: Conclusion

This page titled 8.4: Toxicology MSDT is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Toxicology MSDT via source content that was edited to the style and standards of the LibreTexts platform.

- Toxicology MSDT by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.1: PATHOPHYSIOLOGY

Pathophysiology is the study of the physical and functional changes that occur during a disease process. In this e-module you will learn about the concept of pathophysiology, types of toxicity, repair and adaptation, and patterns of toxic injury.

8.4.1.1: What is Pathophysiology?

8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity

8.4.1.3: Cellular Response to Toxicant-Induced Injury

8.4.1.4: Repair and Adaptation

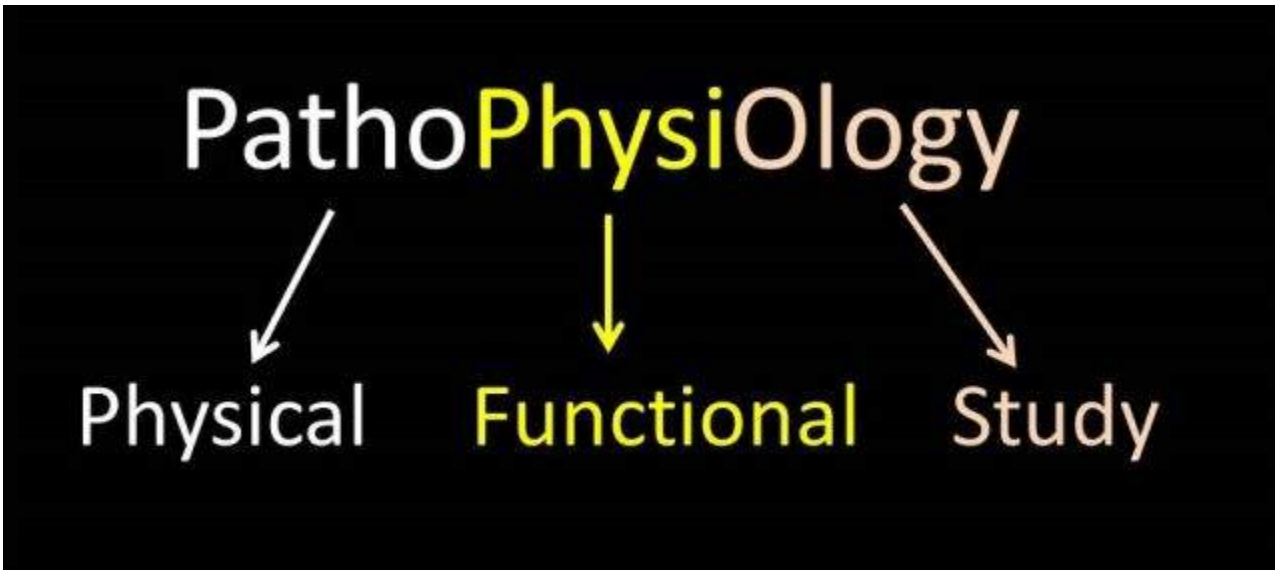
8.4.1.5: Patterns of Toxic Injury

8.4.1.6: Final Evaluation

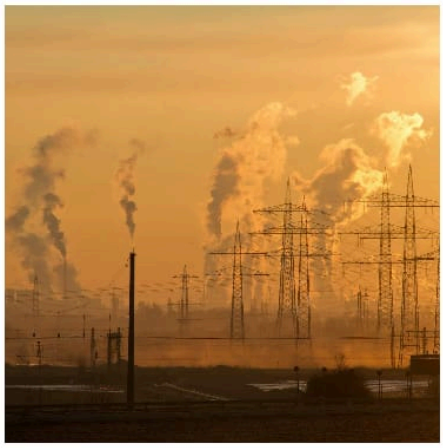
This page titled 8.4.1: Pathophysiology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 1: Pathophysiology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.1.1: What is Pathophysiology?



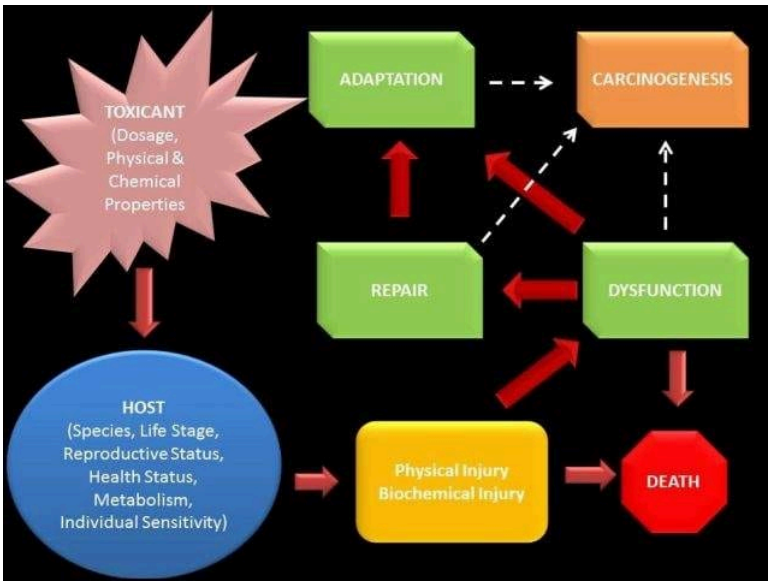
Pathophysiology describes the changes that occur during a disease process, with “patho-“ referring to the physical changes that are observed and “physio-“ referring to the functional processes or mechanisms that occur during a disease process. In toxicology, pathophysiology encompasses the biochemical and physical alterations that occur upon exposure of an individual (generally termed the “host”) to harmful amounts of a toxicant.







In toxicology, pathophysiology takes into account how the characteristics of the toxicant (for example, dosage, physical properties and chemical properties) and the characteristics of the host (including species, life stage, health/reproductive status, metabolism, and individual sensitivity) interact to produce physical and/or biochemical changes in the host. Pathophysiology also encompasses the host response to the effects of a toxicant. With acutely lethal intoxications, the physical and chemical injury may be sufficient to cause rapid death of the organism. In non-lethal toxic exposures, toxicant-induced injury results in dysfunction of cells, tissues and/or organs that may persist or that may progress to death. Persistent toxic injury that does not result in death generally leads to attempts at repair of toxicant-induced damage. With some toxicants, the host is able to develop strategies to adapt to continued exposure to toxicants. Dysfunction, repair and adaptive processes that occur in response to exposure to certain toxicants may trigger development of unregulated cell growth leading to tumor formation in a process termed carcinogenesis.



In many cases the toxicant is the unchanged xenobiotic to which the host was exposed, but in some cases the xenobiotic itself may be relatively innocuous and requires bioactivation to more toxic metabolites before toxic effects occur. A variety of endogenous systems have evolved to mitigate the effects of many toxicants and/or their metabolites. However, when these systems fail or when the dose of toxicant exceeds the capacity of the system to neutralize the toxic effects, poisoning occurs. The clinical syndrome associated with a poisoning is referred to as toxicosis.

Most toxicants exert their effects on specific molecules, tissues or organs based on the physical and chemical makeup of the toxicant as well as the absorption, distribution and metabolism of the toxicant within the body.

Toxicants such as some strong acids or alkalis (e.g. concentrated hydrochloric acid) are not systemically absorbed, so are limited to causing local injury upon contact with skin, eyes or mucous membranes. Toxicants that are ingested and absorbed from the gastrointestinal tract are shuttled via the portal vein to the liver, where they may cause direct injury, where they may be bioactivated to toxic metabolites, or where they may be detoxified before they reach the general circulation (a process termed “first pass effect”). Inhaled toxicants such as smoke may cause local tissue injury due to irritants or corrosive components as well as systemic intoxication from toxic gases such as carbon monoxide or cyanide.

Topic 1: Key Points
In this section, we explored the following main points:

- 1: Pathophysiology is the study of the physical and functional changes that occur during a disease process.
- 2: Toxic insults can result in physical and biochemical alterations that may lead to cellular dysfunction, repair, adaptation, carcinogenesis and/or death.

This page titled 8.4.1.1: What is Pathophysiology? is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 1.1: What is Pathophysiology? by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity

Learning Objectives

After completing this lesson, you will be able to:

- 1: Discuss the various effects that toxicants have on target molecules and how these effects result in injury to the host.
- 2: Discuss the pathogenesis of the rash that occurs following exposure to poison ivy.

The interaction of toxicants with host molecules may lead to the dysfunction or destruction of the target molecule, or it may result in formation of adducts that the immune system identifies as “foreign,” triggering immune responses against these neoantigens.

Target Molecule Dysfunction

Target molecule dysfunction is a common mechanism by which xenobiotics, particularly drugs, exert their effects; remember it is the dose of xenobiotic that determines whether the effect will be therapeutic (pharmacologic) or harmful (toxicologic). Target molecule dysfunction may occur through activation of cellular membrane receptors, resulting in over-stimulation of some cellular function. For instance, the toxic effects of methomyl, a carbamate insecticide, include over-stimulation of cells of excretory glands, resulting in excessive salivation, excessive tear formation and excessive secretion of mucus by goblet cells within the respiratory tract.

Conversely, other toxicants may inhibit or impede the action of cellular receptors; the resulting clinical effects will depend on the type of receptor affected and what action is impeded. For example, there are channels in nerve cell membranes that allow sodium to pass into and out of the cell; when exposed to pyrethrins, insecticides extracted from chrysanthemum flowers, these channels are unable to close, which results in excessive stimulation seen as muscle twitches, tremors and convulsions. However, when these channels are exposed to tetrodotoxin, the infamous puffer fish toxin, these channels are unable to open, which prevents stimulation of muscle and results in paralysis.



Figure 8.4.1.2.1: Pufferfish (fugu) is a delicacy in some countries, but if not prepared correctly it can cause paralysis and even death.

Toxicants may induce target molecule dysfunction by altering protein structure such that the protein is no longer functional, resulting in disruption of membrane protein channels, interference with transmembrane signaling or loss of enzyme function. Many of these types of effects involve the toxicant or its metabolite binding to reactive moieties on the protein molecule; the sulfhydryl or thiol (S-H) moiety is particularly susceptible to binding with other reactive compounds. Toxicant-induced alteration of DNA structure can lead to mispairing of nucleotides during mitosis, with potential effects ranging from altered protein synthesis to initiation of carcinogenesis.



Figure 8.4.1.2.2: DNA strand breaks may result in cell death, genetic mutation or development of cancers

Neoantigen formation results when a xenobiotic or its metabolite **binds** to a larger protein to form a novel molecule that elicits an immune response. Molecules that trigger an immune response upon binding to carrier proteins are termed **haptens,** and the process of neoantigen formation in this manner is termed **haptenization**. Neoantigens can trigger humoral immune responses resulting in the development of antibodies that can trigger acute **allergic** reactions such as hives or anaphylaxis. Neoantigens that trigger cellular-mediated immune responses cause injury to specific tissues or organs such as skin, liver or blood vessels in a process termed **autoimmunity.**

DID YOU KNOW?

The rash caused by poison ivy (Toxicodendron spp.) is caused by urushiol, an oily mixture of chemicals called catechols, in the sap of the plant. Upon exposure to human skin, urushiols bind to membranes on skin cells and serve as haptens, changing the shape of proteins in the membranes. The body's immune cells no longer recognize these skin cells as normal parts of the body and mount an immune response, resulting in inflammation, itching, blisters, swelling and redness at the site of urushiol contact. In addition to local irritation, serious, systemic reactions to urushiol can occur if the leaves are ingested, or if smoke from burning poison ivy is inhaled.



Figure 8.4.1.2.3: Skin contact with leaves of poison ivy can result in a blistering rash.

Topic 3: Key Points
In this section, we explored the following main points:

- 1: Toxicant-induced target molecule dysfunction can occur through activation or inhibition of cellular receptors, denaturing of membrane proteins, and destruction of target molecules.
- 2: Haptenization results in the formation of neoantigens that can trigger immune responses against cells and tissues of the body, resulting in allergic or autoimmune reactions.
    - The rash caused by poison ivy is an autoimmune reaction against skin cells whose membranes have bound to the toxicant urushiol from the plant.

Knowledge Check

1. On a protein molecule highly reactive toxicant molecules have a predilection for____.

    sulfhydryl (S-H) moiety

    Haptenization

dose

Answer available at the end of this section.

2. The therapeutic or toxic effect of a xenobiotic is entirely dependent on its____.

sulfhydryl (S-H) moiety

Haptenization

dose

Answer available at the end of this section.

3. A xennobiotic binding a larger protein molecule, resulting in an immune response.

sulfhydryl (S-H) moiety

Haptenization

dose

Answer available at the end of this section.

Answers

Knowledge Check
sulfhydryl (S-H) moiety

Knowledge Check
dose

Knowledge Check
Haptenization

This page titled 8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

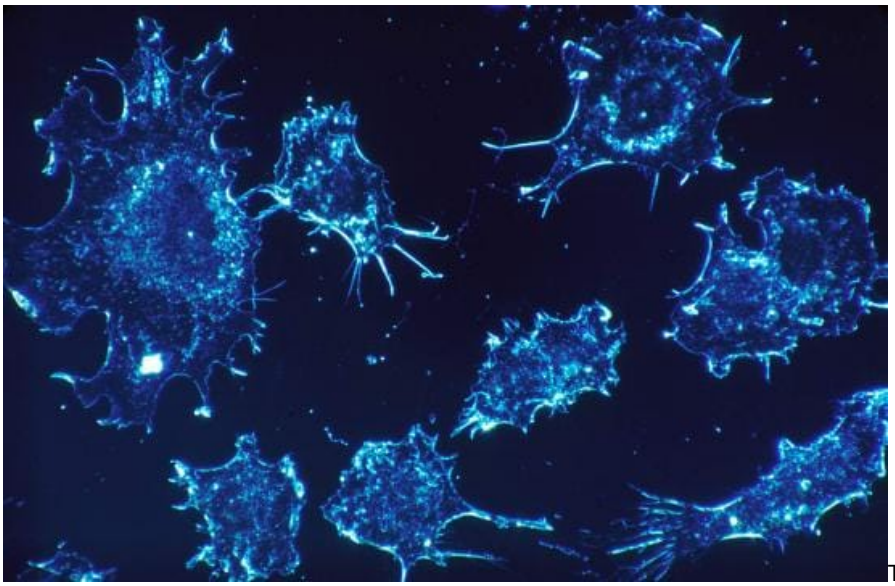
- 1.3: Outcomes of Targeted and Non-Targeted Toxicity by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.1.3: Cellular Response to Toxicant-Induced Injury

Learning Objectives

After completing this lesson, you will be able to:

- 1: Describe the range of cellular injury that may occur following exposure to a toxicant.
- 2: Discuss the major mechanisms of toxicant-induced cell death.
- 3: Compare the processes of necrosis and apoptosis in terms of inciting causes and the cellular changes that occur with each.



Toxicants can exert a variety of effects at the molecular level that have significant repercussions at the cellular, tissue and organ levels. The effects of a toxic exposure can range from reversible cellular dysfunction to irreversible cellular injury to cell death, all of which can alter normal organ function and have significant impact on the health and well-being of the body as a whole. The ability of cells, tissues and organs to overcome the effect of a toxicant through repair and/or adaptation will dictate the ultimate outcome of a toxic exposure.

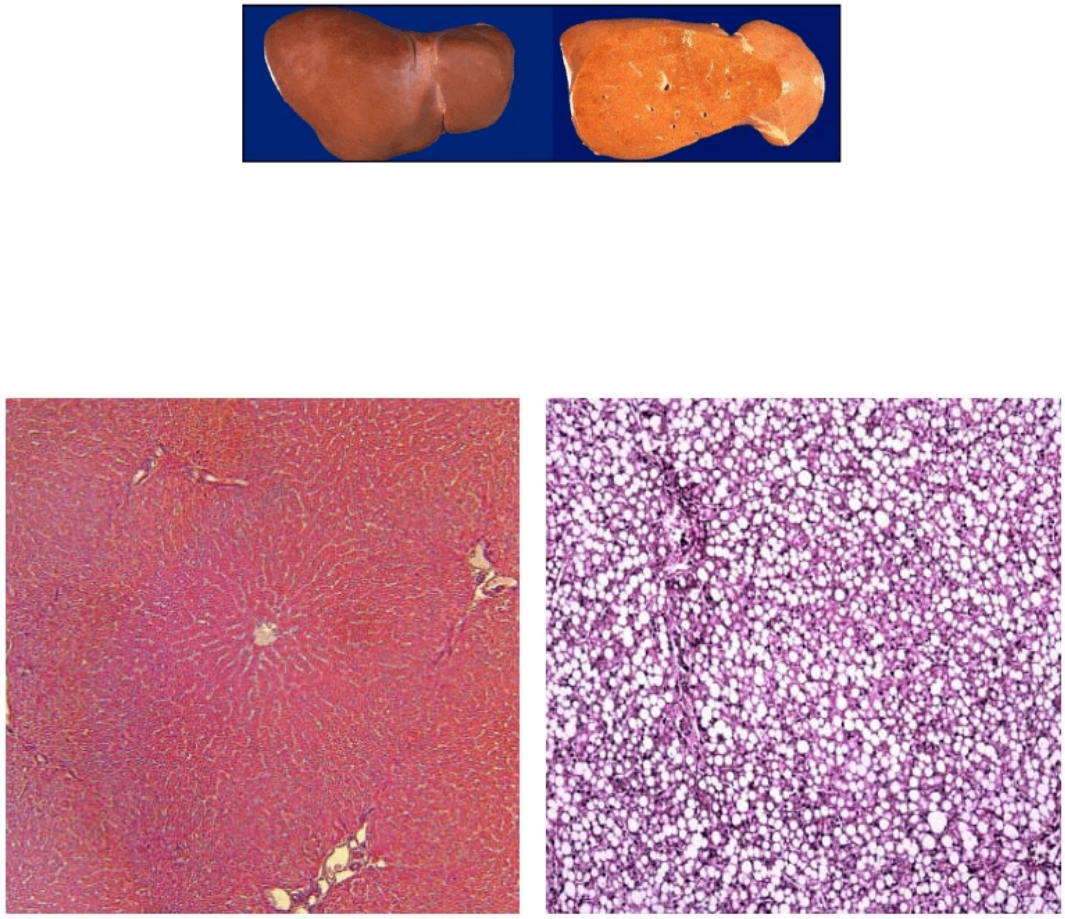


Figure 8.4.1.3.1: Lipid accumulation in the liver comparing normal liver (left) to liver with steatosis (center). In the photomicrographs at the right, the white spaces are areas of lipid accumulation.

Following toxic insult, cells have a limited repertoire of responses. Nonlethal cell injury may lead to cellular degeneration, seen microscopically as swelling of the cells. Toxicant-injured cells may accumulate water, lipid, pigments, glycogen or metabolic waste products due to impairment of normal maintenance functions. Accumulation of lipids in hepatocytes, termed **steatosis** or **hepatic lipidosis**, is a common toxic effect seen in cases of alcohol-related liver disease. Degenerative changes in cells are often reversible if the inciting cause is removed. When cell injury proceeds beyond the self-repair capability of the cell, cell death ensues.

Mechanisms of Cell Death

Major mechanisms of toxicant-induced cell death include disruption of cell membrane structure and/or function, loss of cellular maintenance functions, and impairment of cellular energy production. Loss of membrane structural or functional integrity can result in uncontrolled passage of water, ions and other compounds into or out of the cell. The subsequent loss of normal cytosolic environment interferes with normal biochemical processes necessary for cell function and/or survival. Loss of ability to synthesize proteins and other macromolecules impedes maintenance of organelles and enzymatic pathways vital to cellular survival. Impairment of cellular energy production generally occurs when toxic effects alter mitochondrial function and/or structure and can lead to cell death due to failure to produce sufficient ATP to power essential cellular functions.

Necrosis

Necrosis is the term used to describe cell death due to irreversible injury. Necrotic cells undergo degenerative processes including swelling of organelles, loss of organelle function, oxidative and hydrolytic degradation of intracellular membranes and macromolecules by electophiles and free radicals, and, ultimately, **lysis** (loss of cellular constituents to surrounding tissues due to cell membrane rupture). Necrosis generally results in the generation of an inflammatory response as cellular components and free radicals that are released to the extracellular matrix attract inflammatory cells.

Apoptosis

In contrast, **apoptosis** (sometimes nicknamed “cell suicide” or “programmed cell death”) is a more orderly form of cell death. Apoptosis is an active process involving activation of specific enzymes which triggers the systematic fragmentation of cell constituents into blebs of cell membrane that pinch off of the main cell to form **apoptotic bodies**. During this fragmentation, the cell continues to produce energy and proteins, unlike necrosis where organelle and energy production cease prior to cellular fragmentation. The end result of apoptosis is numerous apoptotic bodies, each composed of a cellular membrane surrounding intact and functional cellular components. Apoptosis can be triggered by various forms of oxidative stress, particularly the presence of excessive oxygen-derived free radicals, due to excessive free radical generation and/or to lack or exhaustion of endogenous antioxidants. Because intracellular components are not spilled into the extracellular matrix, apoptosis generally does not incite an inflammatory response; instead the apoptotic bodies are removed by local phagocytes.

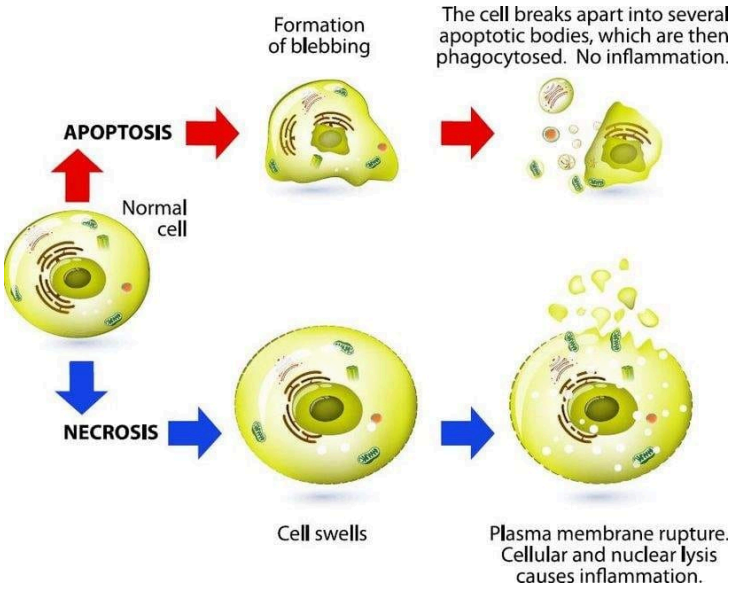


Figure 8.4.1.3.2: Apoptosis and Necrosis are two very different forms of cell death.

Topic 4: Key Points
In this section, we explored the following main points:

- 1: Toxicant-induced cellular injury can range from reversible cellular dysfunction to irreversible cellular injury to cell death.
- 2: Cellular responses to toxic injury may include cellular degeneration and accumulation of substrates within the cell.
  - Steatosis is lipid accumulation within hepatocytes and is a common toxic effect secondary to alcohol exposure.
- 3: Major mechanisms of cell death include disruption of cell membrane structure and/or function, loss of cellular maintenance functions, and impairment of cellular energy production.
- 4: Necrosis is cell death resulting from cessation of organelle function, degradation of intracellular structures and culminating in lysis of the cell and attraction of inflammatory cells.
- 5: Apoptosis is an active form of cell death whereby cellular function is maintained as the cell components are compartmentalized and packaged into apoptotic bodies that pinch off of the main cell.
  - Apoptosis can be triggered by oxidative stresses caused by oxygen-derived free radicals.
  - Apoptosis is not generally associated with an inflammatory response.

Knowledge Check

1. The accumulation of excess lipid within liver cells:

Answer available at the end of this section.

2. Cellular degeneration secondary to a toxic insult is seen microscopically as:

Answer available at the end of this section.

3. Loss of organelle function, hydrolytic degradation of intracellular membranes and lysis of cells are characteristics of_____.

Answer available at the end of this section.

4. The orderly decommissioning of cellular organelles without loss of energy and protein synthesis leading to fragmentation into membrane-bound packets is characteristic of____.

Answer available at the end of this section.

Answers

Knowledge Check
Steatosis

Knowledge Check
Cellular Swelling

Knowledge Check
necrosis

Knowledge Check
apoptosis

This page titled 8.4.1.3: Cellular Response to Toxicant-Induced Injury is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 1.4: Cellular Response to Toxicant-Induced Injury by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

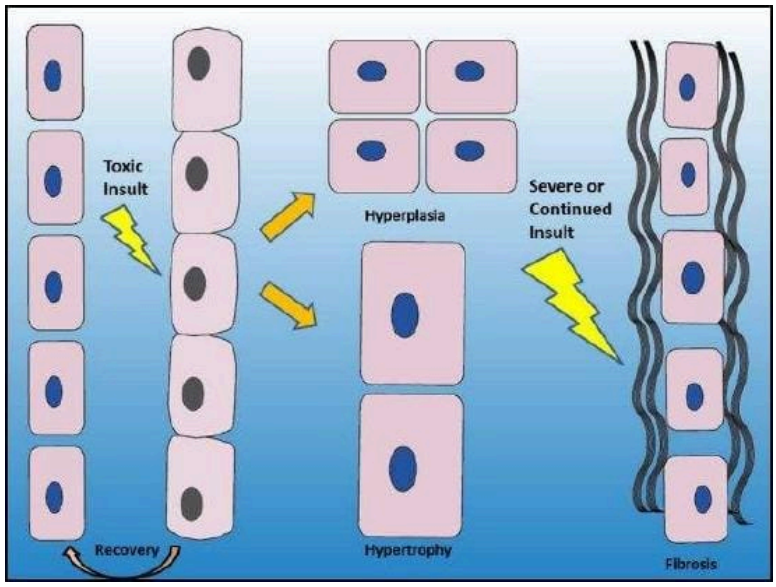
8.4.1.4: Repair and Adaptation

Learning Objectives

After completing this lesson, you will be able to:

1. Explain cell and tissue repair mechanisms and how toxicants may alter these processes.
2. Discuss adaptation mechanisms that may occur with exposures to toxicants.

Repair



The ability of cells to repair toxicant-induced damage plays a large role in the outcome of exposure to toxicants. Repair mechanisms include molecular repair such as reversal of thiol oxidation, removal of damaged units with replacement by newly synthesized units, and degradation followed by resynthesis of damaged structures. Cellular repair mechanisms include autophagy of damaged organelles and resynthesis/regeneration of damaged structures. Tissue repair mechanisms include active removal of damaged cells via apoptosis, regeneration of cells by **hyperplasia** (increase in cell numbers) or **hypertrophy** (increase in cell size), and resynthesis of extracellular matrix. When cellular or tissue repair is unable to restore original tissue architecture, production of extracellular connective tissue results in **fibrosis**.

Some toxicants can interfere with the normal cellular repair mechanisms, resulting in early cellular senescence and death. Inability to repair damage to DNA can result in disruption of normal protein synthesis or initiation of carcinogenesis. Other toxicants trigger exaggerated repair mechanisms which themselves pose a hazard to the survival of the individual. For instance, the herbicide paraquat causes lung injury mediated by oxygen-derived free radicals; because the lung is a highly oxygenated organ, this damage can be quite extensive as the ready availability of oxygen provides plenty of fuel for the snowballing generation of free radicals such as superoxide anions. The immediate effect of paraquat on the lung is damage to alveolar walls, resulting in stimulation repair mechanisms that cause intense, progressive fibrosis of the lung over several weeks, which ultimately leads to the death of the patient from asphyxia. Intense fibrosis is also seen in alcoholic cirrhosis of the liver, with normal liver parenchyma being replaced by fibrous connective tissue that impedes normal liver function.

ADAPTATION



Adaptation mechanisms have evolved at the cellular, tissue and organ levels that allow the individual to survive in the face of exposure to levels of toxicants that might otherwise result in serious or lethal injury. Adaptation requires time to develop, so generally occurs upon multiple exposures to levels of toxicants that are too low to cause severe acute injury. Adaptation mechanisms include alteration of toxicant delivery to target sites, decreasing reactivity of the target site to the toxicant, increasing local repair mechanisms, and development of compensatory mechanisms to mitigate toxicant-induced injury.

Decreasing toxicant delivery to target site may include decreasing toxicant absorption, detoxification of the toxicant before it can reach the target site, or binding of the toxicant with a neutral molecule. The opioid fentanyl is a powerful narcotic when injected, but it is quickly metabolized in the liver if ingested, greatly reducing the amount that enters the systemic circulation and reaches the central nervous system; this “first pass effect” occurs with many substances and is often the reason why some drugs must be administered via injection rather than the oral route. Similarly, chronic ingestion of ethyl alcohol can result in the **induction** (increased expression) of the enzyme alcohol dehydrogenase in the gastrointestinal tract and liver, resulting in more alcohol being metabolized before it can reach the systemic circulation.

Some toxic heavy metals that reach the bloodstream become bound to metal-binding proteins called **metallothioneins**; because only unbound metals are free to react, metallothioneins prevent the toxic metals from interacting with their target sites. Many sea-dwelling mammals such as killer whales (Orca spp.) bind organic mercury to selenium-containing metallothioneins which allows them to accumulate large amounts of organic mercury that would be lethal to terrestrial mammals.

Adaptation by decreased reactivity to the target site is classically illustrated by the **tolerance** that can develop in people addicted to opioids such as heroin. Constant stimulation of opioid receptors results in **downregulation** (decreased expression) of those receptors and increased amounts of opioids required to stimulate remaining receptors.

Adaptive mechanisms may include increased function or number of cells to compensate for the loss of function of similar cells due to toxic insult. A variety of toxicants can cause injury to kidney tubules, and as a result, other kidney tubules may undergo hypertrophy and/or hyperplasia in an attempt to provide adequate kidney function.

Adaptive and repair mechanisms frequently occur simultaneously, resulting in characteristic lesions in some organs or tissues, For example, toxic hepatocellular injury from chronic alcohol exposure may result in nodular areas of regenerating hepatocytes (adaptation) surrounded by fibrous connective tissue formed to replace lost hepatocytes (repair), resulting in the characteristic “bubbly” look of hepatic cirrhosis.

Topic 5: Key Points
In this section, we explored the following main points:

- 1: Tissue repair mechanisms include removal of damaged cells via apoptosis, regeneration of cells by hyperplasia (increase in cell numbers) or hypertrophy (increase in cell size), and regeneration of extracellular matrix.

    - Fibrosis may result when tissue repair is incomplete.
- 2: Adaptation mechanisms may allow the host to survive in the face of continuous exposure to toxicants and include:
    - alteration of toxicant delivery to target sites.
    - decreasing reactivity of target site to toxicant
    - increasing local repair processes
    - compensatory mechanisms to mitigate toxicant-induced injury

Knowledge Check

1. An example of decreasing the reactivity of a target site to a toxicant

Answer available at the end of this section.

2. A regenerative response that results in an increased number of cells being produced

Answer available at the end of this section.

Answers

Knowledge Check

Downregulation of toxicant receptors

Knowledge Check

Hyperplasia

This page titled 8.4.1.4: Repair and Adaptation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 1.5: Repair & Adaptation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.1.5: Patterns of Toxic Injury

Learning Objectives

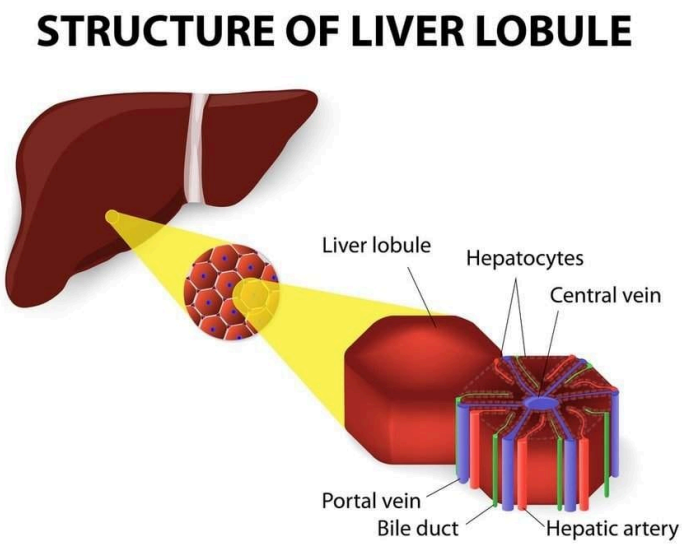
- 1: Explain how the structure and function of the liver relates to its susceptibility and responses to toxicants.
- 2: Discuss how the pattern of toxic injury seen in an organ may provide clues to the toxicant that is involved.



Toxicants can exert their effects systemically or by altering target tissues or organs, and sometimes the patterns of injury left behind can aid in the identification of an unknown toxicant, or at least allow the list of suspects to be narrowed down to a smaller number than the thousands of potential toxicants that exist. Because most absorbed toxicants are carried in the bloodstream, blood flow patterns often dictate which tissues may be affected. Within the tissues, variations in cell type and function can also influence the pattern of injury that can be seen with a given toxicant. Because they receive high volumes of blood flow, and because of their roles in metabolism and excretion of xenobiotics, the liver and kidney are particularly vulnerable to effect from systemically absorbed toxicants.

Patterns of Hipatic Injury

The microscopic appearance of the liver is that of a hexagonal lobule. Blood enters the parenchyma from the hepatic artery and portal vein situated, along with a bile duct, at an area termed the **portal triad** situated at each of the six vertices of the lobule (**periportal** regions). From the triad region, the blood passes through endothelial-lined **sinusoids** that separate anastomosing sheets of hepatocytes before exiting via the terminal hepatic, or **central** vein (**centrilobular** region). Physiologically, the liver has highly oxygenated blood entering at **zone 1**, which is roughly equivalent to the periportal areas, surrendering its oxygen as it filters through the sinusoids, and reaching **zone 3** (centrilobular region) as poorly oxygenated blood that exits via the central vein. **Zone 2**, also called the **midzonal** region, is the intermediate area between zones 1 and 3.



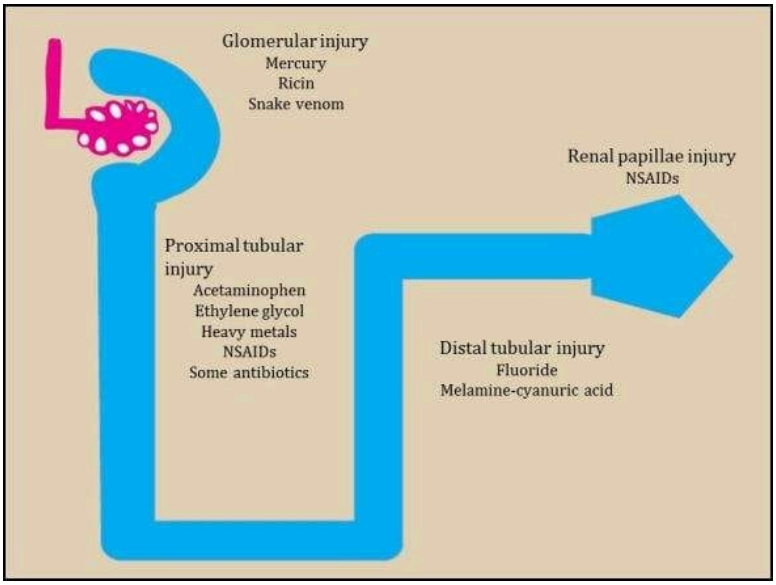
Within the different zones of the liver, hepatocytes vary in their physiologic functions, with those in zone 1 being more efficient in oxidative metabolism and zone 3 hepatocytes being efficient at xenobiotic biotransformation. Hepatocytes in zone 1 are the first to be exposed to toxicants that enter the liver; if those toxicants are directly injurious to hepatocytes (e.g. white phosphorus), the pattern of cell injury will be periportal. Toxicants requiring bioactivation to cause injury (e.g. acetaminophen) will generally cause zone 3 (centrilobular) hepatic injury, since this area contains higher levels of biotransforming enzymes. This area is also at risk for hypoxic injury due to toxicants that alter oxygen delivery to cells (e.g. carbon monoxide). Because of these vulnerabilities, centrilobular injury is the most common form of toxicant-induced hepatic injury. **Massive** liver necrosis affects entire hepatic lobules, and has been associated with exposure to a variety of toxicants including acetaminophen, aflatoxin, blue-green algae, and hepatotoxic mushrooms. **Piecemeal** necrosis is a less commonly seen form of liver injury wherein scattered individual hepatocyte necrosis or apoptosis occurs along the limiting plate between portal triads; this form of hepatic injury has been associated with immune-mediated processes such as is seen with non-steroidal anti-inflammatory drug-induced hepatopathy.

Zone 1 (Periportal)	Zone 2 (Midzonal)	Zone 3 (Centrilobular)
Iron Salts	Furans	Acetaminophen
White Phosphorus	Ngaione (*Myoporum* spp.)	Aflatoxin
	Beryllium	Microcystin (Blue-Green Algae)
		Ricin (Castor Bean)

Patterns of Renal Injury

The nephron is the functional unit of the kidney, consisting of the renal corpuscle (**Bowman's capsule** and **glomerulus**), **proximal tubule**, **loop of Henle**, and **distal tubules**. Toxicants that are directly injurious may cause damage to the glomerular structures or to the anterior portion of the proximal tubule. Toxicants that require bioactivation generally cause injury to the more distal section of the proximal renal tubules, as this is where the majority of biotransforming enzymes occur. As the glomerular filtrate passes through the loop of Henle and distal tubules and becomes more concentrated, toxicants that were too dilute to affect earlier renal structures may cause injury in these more distal sections of the kidney. Toxicants that decrease renal blood flow (e.g.

non-steroidal anti-inflammatory drugs) can also cause injury to more distal tubules, collecting ducts and renal papillae as these regions receive less blood flow than more proximal structures.



Topic 6: Key Points
In this section, we explored the following main points:

- 1: Toxic injury to a tissue or organ will depend largely on the pattern of blood flow to that tissue or organ.
  - Cells in areas that normally have low oxygenation are at increased risk of injury from toxicants that reduce blood flow or oxygen delivery.
- 2: Variations in cell type and function can affect the pattern of toxicant-induced injury.
  - Cells with large capacity for metabolism of xenobiotics will be at higher risk for injury by toxicants requiring bioactivation.
- 3: Identification of the pattern of toxicant-induced injury can provide clues to the types of toxicants that may have caused the injury.

Knowledge Check

1. Which of the following patterns of injury would be consistent with a toxicant that is directly injurious to liver and kidney cells?

Centrilobular hepatic injury, distal tubular renal injury

Centrilobular hepatic injury, proximal tubular renal injury

Periportal hepatic injury, distal tubular renal injury

Periportal hepatic injury, proximal tubular renal injury

Answer available at the end of this section.

2. Which features make the centrilobular region of the liver lobule more susceptible to toxic insult?

Relatively low levels of biotransformation enzymes and relatively low oxygenation

Relatively high levels of biotransformation enzymes and relatively low oxygenation

Relatively low levels of biotransformation enzymes and relatively high oxygenation

Relatively high levels of biotransformation enzymes and relatively high oxygenation

Answer available at the end of this section.

Answers

Knowledge Check

Periportal hepatic injury, proximal tubular renal injury

Knowledge Check

Relatively high levels of biotransformation enzymes and relatively low oxygenation

This page titled 8.4.1.5: Patterns of Toxic Injury is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 1.6: Patterns of Toxic Injury by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.1.6: Final Evaluation

1. General rule of thumb for apoptosis is that it is more commonly seen at higher levels of toxicant exposure while necrosis occurs more frequently at relatively lower levels of toxicant exposure. *True or False?*

True

False

Answer

False

2. Neoantigen formation results when a xenobiotic or its metabolite binds to a larger protein to form a novel molecule that elicits an immune response. Molecules that trigger this immune response are called:

Haptens

Free radicals

Electrophils

Poisons

Answer

Haptens

3. Lipid peroxidation occurs when an _____ compound steals an electron from a membrane phospholipids resulting in the production of a fatty acid radical.

Electrophobic

Hydroxyl radical

Electrophilic

Hydrophilic

Answer

Electrophilic

4. Electron transfer can result in oxidation of some endogenous macromolecules and formation of _____ such as superoxide ion (•O2-) and hydroxyl radical (HO•).

Enzymes

Hormones

Free radicals

Proteins

Answer

Free radicals

5. Upon exposure of human skin with leaves of poison ivy, the toxin found in this plant binds to membranes on skin cells to stimulate an immune response resulting in blistering rash. This toxin is known as:

Urushiol

Ricin

Tetrodotoxin

Pyrethrins

Answer

Urushiol

6. Which of the following covalently binds to the acetaminophen metabolite N-Acetyl-P-Benzoquinone Imine (NAPQI) to detoxify it?

N-acetylcysteine

Superoxide dismutase

Catalase

Amylase

Answer

N-acetylcysteine

7. Differences between Necrosis and Apoptosis include the following EXCEPT:

Necrosis is a degenerative process while Apoptosis is an active process

Necrosis triggers inflammatory response while Apoptosis does not incite inflammatory response

Necrosis results in loss of energy while Apoptosis does not result in loss of energy

Necrosis is an active process while Apoptosis is a degenerative process

Answer

Necrosis is an active process while Apoptosis is a degenerative process

8. A regenerative response that results in an increase in cell size is termed _____.

Hypertrophy

Hypoplasia

Hyperplasia

Atrophy

Answer

Hypertrophy

9. _____ helps to terminate lipid peroxidation when present in sufficient quantities.

Vitamin E

Hydroxyl radical

Vitamin A

Oxygen

Answer

Vitamin E

10. An "adduct" is the product of an irreversible bond between the toxicant and target molecule; this type of binding is known as:

Noncovalent binding

Covalent binding

Hydrogen abstraction

Ionic binding

Answer

Covalent binding

11. An example of decreasing the reactivity of a target site to a toxicant is describe as:

Binding of toxicant to target receptors

Downregulation of toxicant target receptors

Increased metabolism of alcohol within the stomach and liver

Stimulation of membrane receptors

Answer

Downregulation of toxicant target receptors

12. 10% of acetaminophen bioactivated by Cytochrome P450 in the liver to a toxic metabolite is called:

Superoxide dismutase

N-acetylcysteine

N-acetyl-P-Benzoquinone Imine (NAPQ1)

Superoxide ion

Answer

N-acetyl-P-Benzoquinone Imine (NAPQ1)

This page titled 8.4.1.6: Final Evaluation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Final Evaluation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.2: BIOCHEMISTRY AND MOLECULAR GENETICS

Biochemistry is the study of chemical processes within and related to living organisms. In this e-module you will learn about biomolecules and cell components, cell structure and subcellular compartments, DNA and RNA metabolism, and epigenetic mechanisms.

8.4.2.1: Introduction to Biomolecules and Cell Components

8.4.2.2: Cell Structure and Subcellular Compartments

8.4.2.3: DNA and RNA Metabolism

8.4.2.4: Epigenetic Mechanisms

8.4.2.5: Section 2 Final Evaluation

This page titled 8.4.2: Biochemistry and Molecular Genetics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2: Biochemistry and Molecular Genetics by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.2.1: Introduction to Biomolecules and Cell Components

Learning Objectives

- 1: Define the basic structure of biomolecules, such as: amino acids and proteins, carbohydrates, fatty acids, triacylglycerol, phospholipids, steroids and nucleic acids.
- 2: Define the meaning and significance of essential and non-essential amino acids.
- 3: Understand the function of enzymes.
- 4: Define the basic structure of ribonucleic acid (RNA) and deoxyribonucleic acid (DNA).

1.1

Amino Acids and Proteins

Amino acids are the basic units of proteins. All amino acids present in proteins carry a carboxyl- and an amino group, hydrogen and variable side chains (R) at a single α – carbon atom.

Amino Acid Basic Structure:

Every amino acid has four components linked together with a central carbon atom α – carbon:

- **Amino group (NH2)**
- **Carboxylic acid group (COOH)**
- **Hydrogen atom (H)**
- **R-group, which varies with each amino acid (R)**

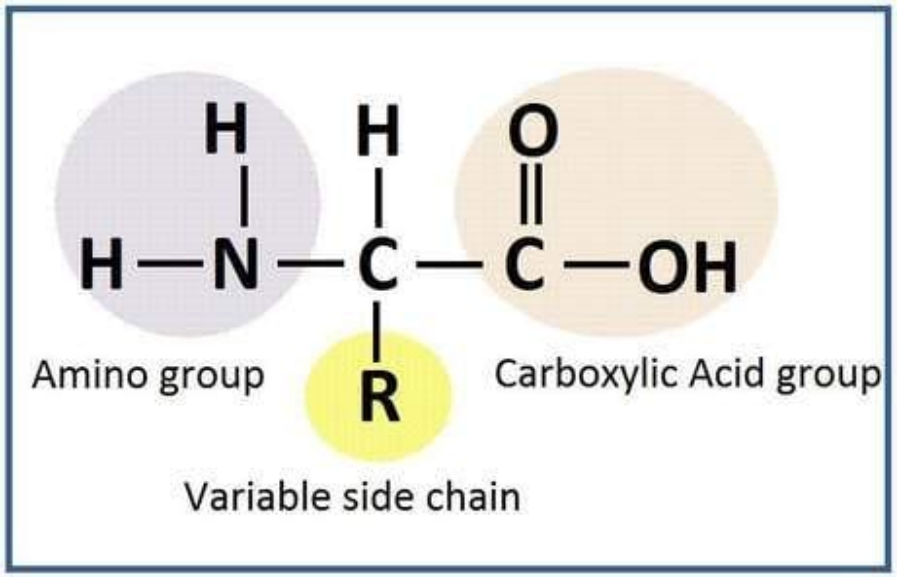
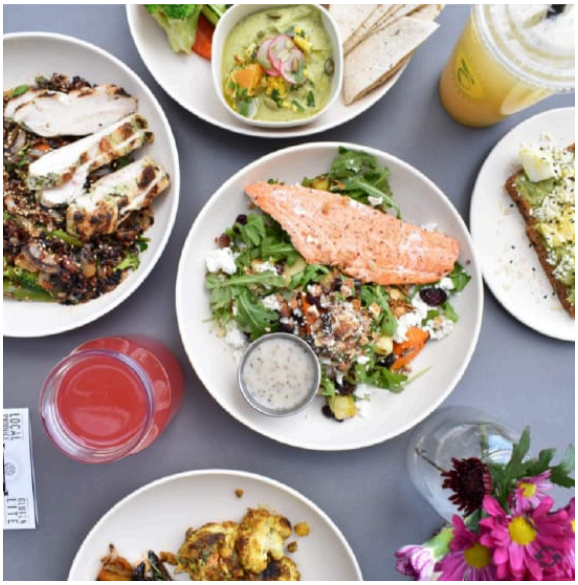


Figure 8.4.2.1.1: Basic structure of an amino acid. Credit: Aline de Conti

R groups may be:

- Hydrophobic
- Hydrophilic
- Charged R-groups: *positive or negative charged*
- Special R-groups: *conjugated with other molecules*

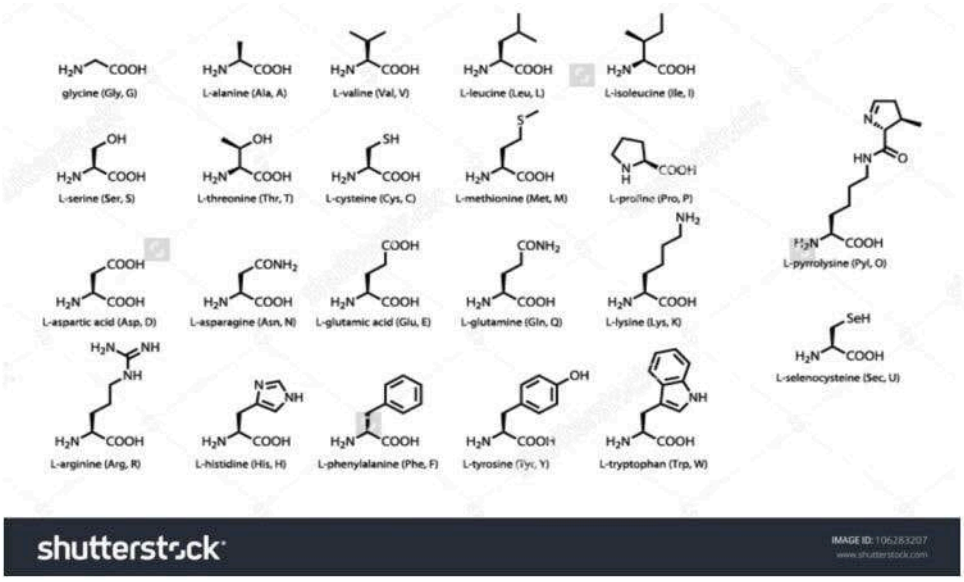
Amino Acids are classified into two groups.



Essential: Humans cannot synthesize them and must be obtained directly from food (phenylalanine, valine, threonine, tryptophan, methionine, leucine, isoleucine, lysine, histidine, cysteine, and arginine).

Non-essential: The human body is able to produce them (glycine, alanine, serine, asparagine, glutanine, tyrosine, aspartic acid, glutanic acid, and proline).

Below, you can see different structures of the most common amino acids in humans. Amino acids link together, in a reaction known as **peptide bond**, to form proteins.



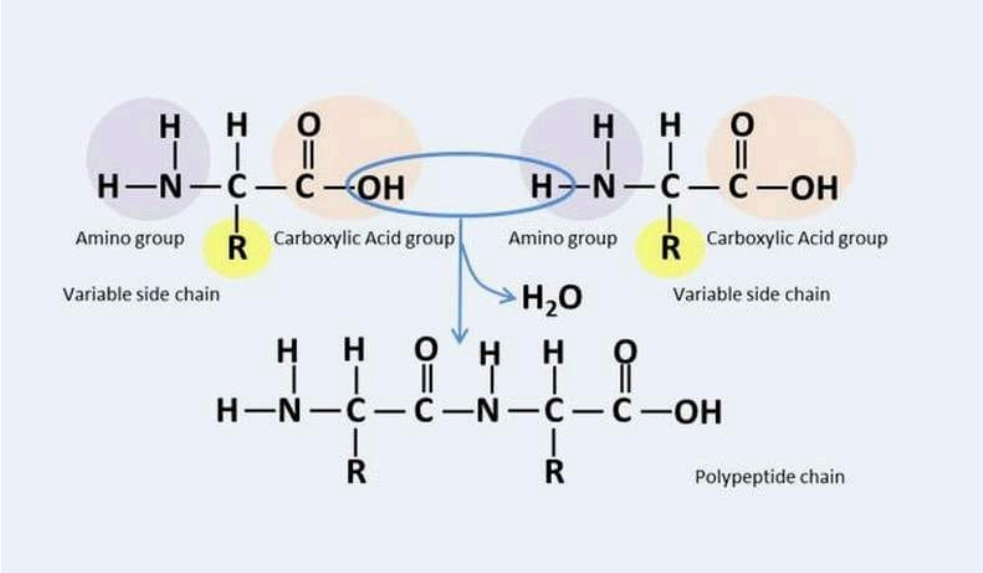


Figure 8.4.2.1.2: Peptide bond. Credit: Aline de Conti

Levels of Protein Structure

Primary (1°) Structure

The sequence of amino acids in a protein is named as **primary structure.** The amino acids are linked via peptide bonds formed with the carboxylic acid group of one amino acid and the amino group of a other amino acid.

Secondary (2°) Structure

The secondary structure is the way a polypeptide folds to form **α-helix**, **β-strand**, or **β-turn**.

Tertiary (3°) Structure

The tertiary structure is the way the polypeptide chain coils and turns to form a complex molecular shape. Additionally, tertiary structure starts to develop an active sites of proteins where critical actions and interactions will take place.

Quaternary (4°) Structure

The quaternary structure is the combination of the multiple protein subunits that interact to form a single, larger, biologically active protein.

Protein Functions

Proteins have several functions in the human body including hormonal, enzymes, structural proteins in cell membranes, proteins also receive signals from outside the cell and mobilize intracellular response, and they are part of the immune system.

Enzymes are specialized proteins that accelerate a chemical reaction by serving as a biological catalyst. By catalyzing these reactions, enzymes cause them to take place one million or more times faster than in their absence. Several biochemical reactions important for cellular maintenance occur due to enzymes activity. For example: environmental response and metabolic pathways.

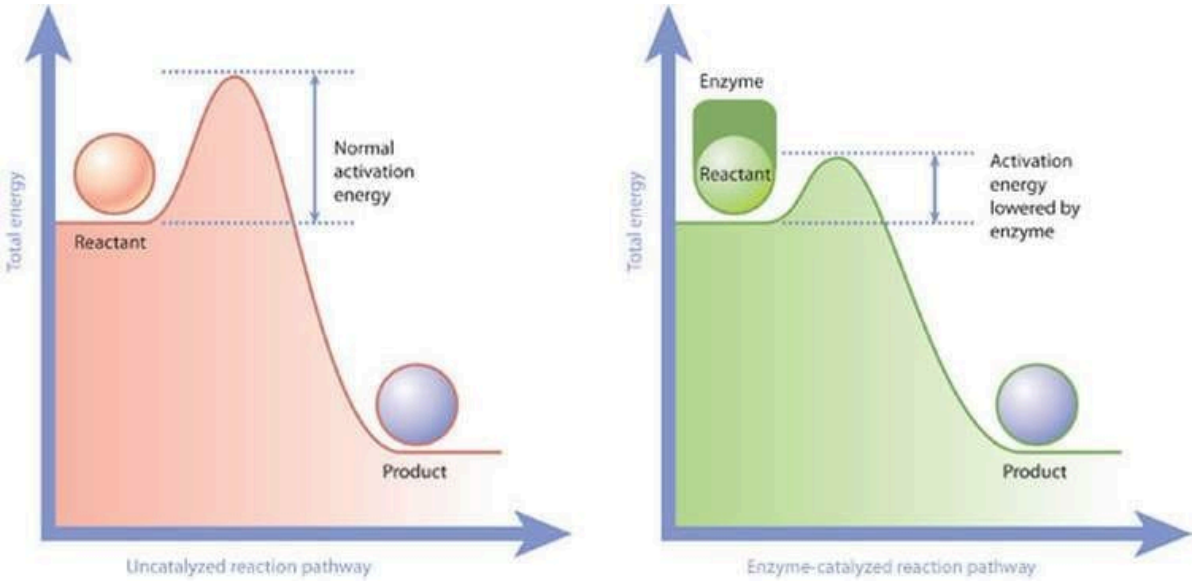
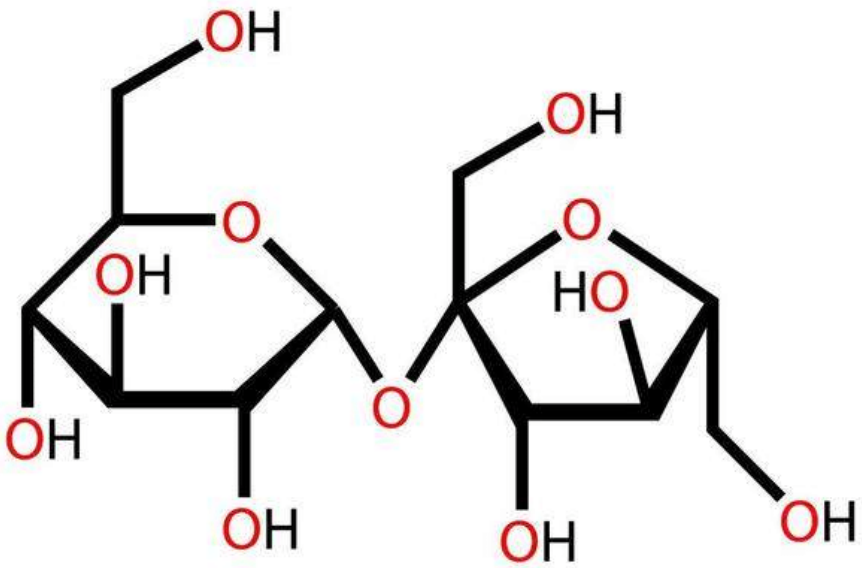


Figure 8.4.2.1.3: Enzymes and Activation Energy - © 2010 Nature Education

1.2

Carbohydrates



Carbohydrates are made of molecules of carbon (C), Hydrogen (H) and Oxygen (O), and are composed of recurring monomers called monosaccharides (which typically form ring structures). A common name of monomers and dimers is ‘sugar’.

Carbohydrates are classified into three subtypes.



1. Monosaccharides: 1 unit of monomer. Examples: fructose, glucose, galactose.

Present in fruits, etc.

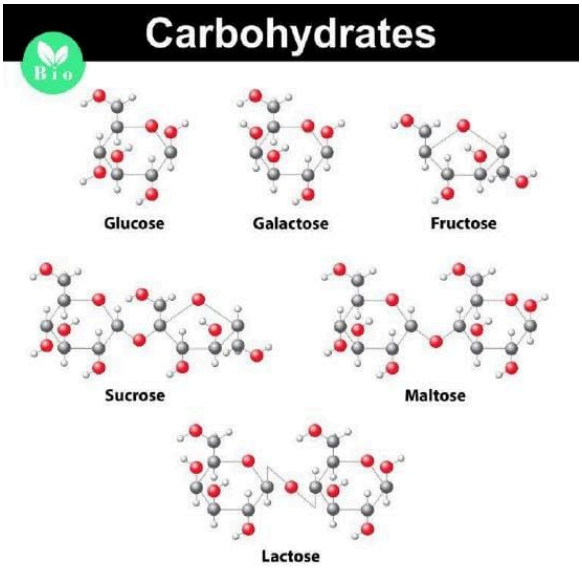
2. Disaccharides: 2 units of monosaccharides. Examples: lactose, maltose, sucrose.

Present in milk, etc.

3. Polysaccharides: Many monosaccharide units. Examples: cellulose, glycogen, starch.

Present in breads, grass, etc.

Carbohydrates are a group of macromolecules that are important energy source required for various metabolic activities. Carbohydrates may bind to proteins and lipids that play important roles in cell interactions e.g. receptor molecules and immune system e.g. antigens.



1.3

Lipids

Lipid molecules are mainly hydrophobic molecules i.e. found in areas away from water molecules, but also present smaller hydrophilic parts that are important for its biological function. The major roles of lipid molecules are to serve as storage of biological energy (Example: **triacylglycerols**) and provide the building blocks for biological membranes (Example: **phospholipids** and **cholesterol**). Although there are other types of lipids, in this topic we will discuss the structure and function of these main groups of lipids.

Triacylglycerols

Triacylglycerols are composed of fatty acids and glycerol.

- **Glycerol** is a simple three-carbon molecule with hydroxyl groups at each carbon.
- **Fatty acids** are chains of carbon molecules with a carboxylic acid (COOH) in the first carbon and a CH3 (methyl) group at the end of the chain.



Fatty acids can be...

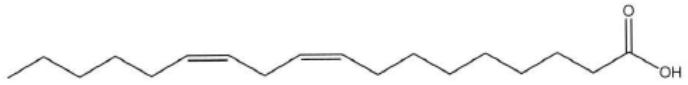
Saturated: Fatty acids contain only single carbon-carbon bonds, and all of the carbon molecules are bonded to the maximum number of hydrogen molecules.

Unsaturated: Fatty acids have at least one double carbon-carbon bond with the potential for additional hydrogen atom bonding still existing for some of the carbon atoms in the backbone chain. If more than one double bond is present, the term polyunsaturated is used.

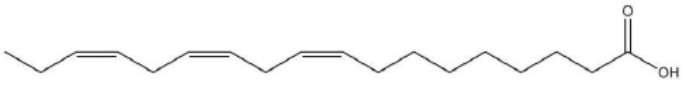
Essential Fatty Acids

Examples of two essential fatty acids, linoleic acid (known as **omega-6;ω-6**) and linolenic acid (known as **omega-3; ω-3**). These fatty acids present double bonds at the sixth and third carbon atoms, respectively, counting from the methyl end of their chains. They are considered essential because humans do not have the ability to produce double bonds at these locations and, therefore, must obtain these two fatty acid from vegetable oils.

linoleic acid (omega-6 fatty acid)



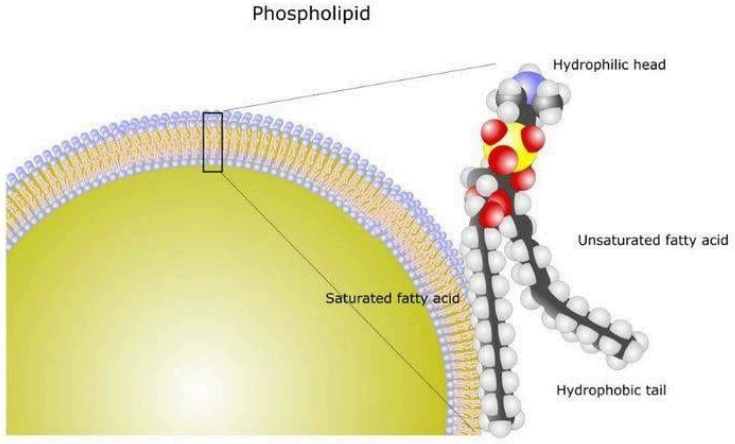
alpha-linolenic acid (omega-3 fatty acid)



Phospholipids

Phospholipids are the major component of cell membranes. They form lipid bilayers because of their amphiphilic characteristic.

The structure of the **phospholipid** molecule generally consists of two hydrophobic fatty acid "tails" and a hydrophilic "head" consisting of a phosphate group (PO4−3) attached to the third glycerol carbon. This head group is usually charged, creating a part of the lipid that is hydrophilic, and wants to be near water, a quality that is essential for the formation of biological membranes and many lipid functions.



Steroids

Steroids are lipids that have four rings made of carbon atoms—three rings have six sides and one has five sides—with a six-carbon ring tail. Examples: bile salts, cholesterol, the sexual hormones estrogen, progesterone and testosterone, corticosteroids and pro-vitamin D.

Cholesterol

Cholesterol is an important molecule found only in eukaryotic organisms with a variety of functions. Cholesterol is also a component of biological membranes and its main function is to control the fluidity of membranes. Cholesterol does not like to be exposed to water environments, preferring to be shielded by other hydrophobic molecules such as lipids or hydrophobic parts of proteins.

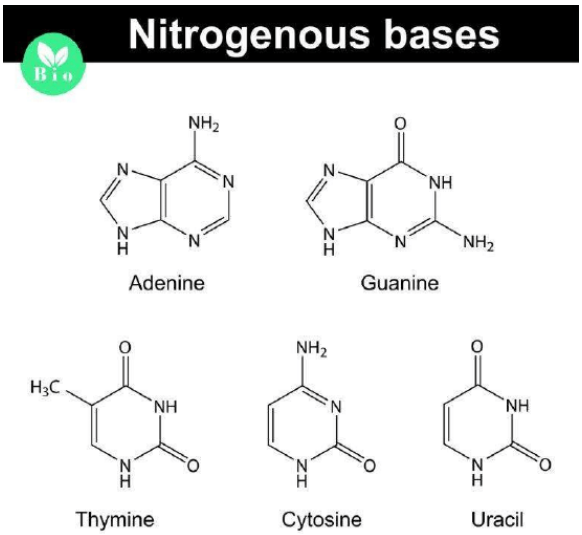
Cholesterol also serves as the primary source for the production of steroid hormones, bile salts, and vitamin D.



1.4

Nucleosides and nucleotides are involved in the preservation and transmission of the genetic information of all living creatures. In addition, they play roles in biological energy storage and transmission, signaling and regulation of various aspects of metabolism.

These molecules can be divided into two major families.



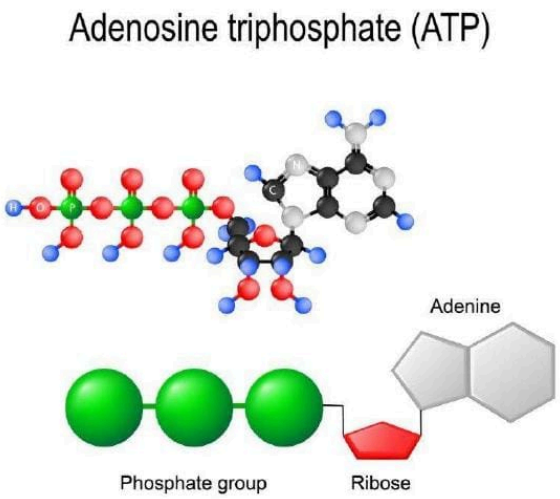
Purines: They are two-ring structures: adenine and guanine.

Pyrimidines: They are one-ring structures: thymine, cytosine, and uracil.

The unique structure and interaction of these molecules serve as the fundamental building block of RNA and DNA molecules and allow fundamental processes of DNA replication and protein synthesis to occur.

Components of Nucleotides

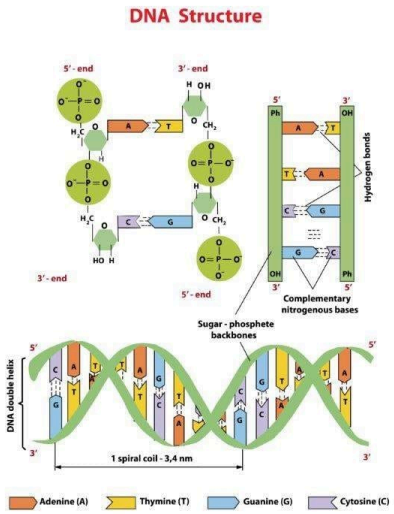
- **Nitrogenous base:** The nitrogenous base of a nucleoside or nucleotide may be either a purine or a pyrimidine.
- **Carbohydrate:** The carbohydrate component of nucleosides and nucleotides is usually the sugar ribose for RNA molecules and deoxyribose for DNA molecule
- **Phosphate Group:** One or more phosphate groups (PO4−3) may be attached to the carbon 5 of the carbohydrate molecule.



1.5

DNA

DNA stands for deoxyribonucleic acid.
It is an extremely long molecule that forms a double-helix.



DNA components:

- Sugars - Deoxyribose

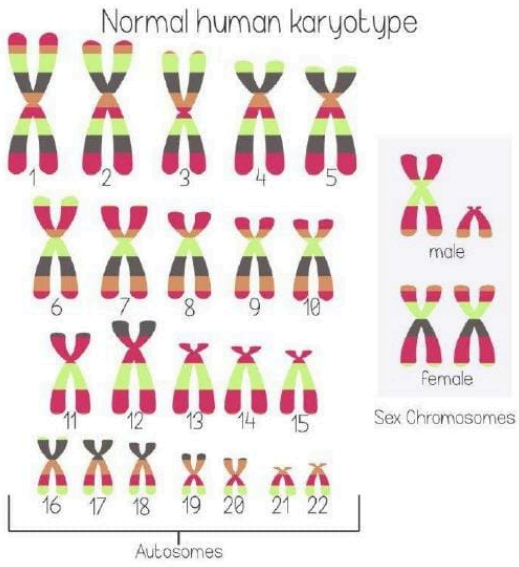
- Phosphates - (PO4−3)

- Base - cytosine (C), guanine (G), adenine (A) and thymine (T).

The DNA consists of two strands attached to each other by hydrogen bond created by nucleotide pairing (A-T and C-G).

The **double-helix** structure of DNA is important for its function because these two bonded strands can temporarily separate to allow for DNA replication.

The sequences of nucleotides (A, C, T, G) in the DNA molecule will make up the **genes** and, subsequently, proteins are referred to as “expressed sequences” or “exons.” Sequences that do not code for a protein are called “intervening sequences” or “introns.”



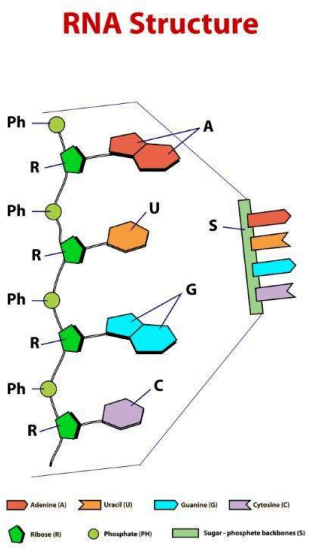
Human Genome

The genome of humans is estimated to contain approximately 20,000–25,000 different genes arranged on multiple chromosomes.

Twenty three pairs of chromosomes:
Twenty two pairs (autosomes).
One pair (sex chromosome) (xx) (female) or (xy) (male).

Humans have 23 pairs of chromosome in every cell (except mature red blood cells); Gametes or sex cells (sperm and eggs) have half the normal complement of chromosomes.

1.6



RNA

RNA stands for ribonucleic acid.
RNA molecules are single strands.

RNA components:

- Sugars - Ribose
- Phosphates - (PO4−3)
- Base: cytosine (C), guanine (G), adenine (A) and uracil (U)

RNA molecules often form secondary (2°) structures and may interact with DNA, other RNA molecules, and proteins. These interactions help to define the particular function of each type of RNA.

Types of RNA molecules and functions:

Messenger RNA (mRNA):

Molecules which function as the transmitter of genetic information from the DNA genetic code to the resulting protein.

Transfer RNA (tRNA)

Molecules that carry amino acids and match them with a specific mRNA sequence during protein synthesis.

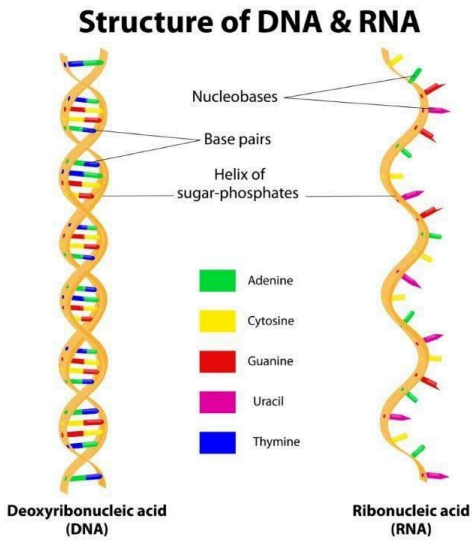
Ribosomal RNA (rRNA)

Molecules associated with proteins and are responsible for the synthesis of protein molecules.

Regulatory RNA

Molecules involved in regulation of DNA expression, posttranscriptional mRNA processing, and the activity of the transcribed mRNA message.

The basic structure of DNA and RNA are similar, however with 3 main differences:



Nitrogenous Bases: Three of the nitrogenous bases are the same in the DNA and RNA: adenine, cytosine, and guanine. The fourth base for DNA is thymine while for RNA it is uracil. Thymine and uracil both bind to adenine.

Number of Strands: The DNA molecule is usually double-stranded and most cellular RNA molecules are single-stranded.

Type of Sugar: In the DNA molecule the sugar is **deoxyribose** and in the RNA molecule the sugar is **ribose**.

Topic 1: Key Points

In this section, we explored the following main points:

- 1: Amino acids link together, in a reaction known as peptide bond, to form proteins.
- 2: One important function of protein is to act as enzymes to accelerate chemical reactions.
- 3: Carbohydrates are important energy source required for various metabolic activities and may bind to proteins and lipids that play important roles in cell interactions
- 4: Lipid molecules serve as storage of biological energy and provide the building blocks for biological membranes
- 5: DNA and RNA structures have 3 main differences .The nitrogenous bases (DNA has thymine and RNA has uracil). The DNA molecule is usually double stranded and most of the RNA molecules are single stranded. In the DNA molecule the sugar is deoxyribose and in the RNA molecule the sugar is ribose.

Knowledge Check

1. What type of nucleic acid does thymine belong to?

Answer available at the end of this section.

2. Uracil?

Answer available at the end of this section.

3. Enzymes are...

specialized proteins that accelerate a chemical reaction by serving as a biological catalyst.

specialized proteins that stops a chemical reaction.

Answer available at the end of this section.

4. **A nucleotide consists of...**

Check all that apply.

- A sugar (either deoxyribose or ribose)
- Uracil as the nitrogen base
- A phosphate group
- One of the four nitrogen bases

Answer available at the end of this section.

5. **Lipid molecules are mainly hydrophilic molecules.** *True or False?*

Answer available at the end of this section.

Answers

Knowledge Check
DNA

Knowledge Check
RNA

Knowledge Check
specialized proteins that accelerate a chemical reaction by serving as a biological catalyst.

Knowledge Check
A sugar, A phosphate group, and One of the four nitrogen bases

Knowledge Check
false

This page titled 8.4.2.1: Introduction to Biomolecules and Cell Components is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.1: Introduction to Biomolecules and Cell Components by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.2.2: Cell Structure and Subcellular Compartments

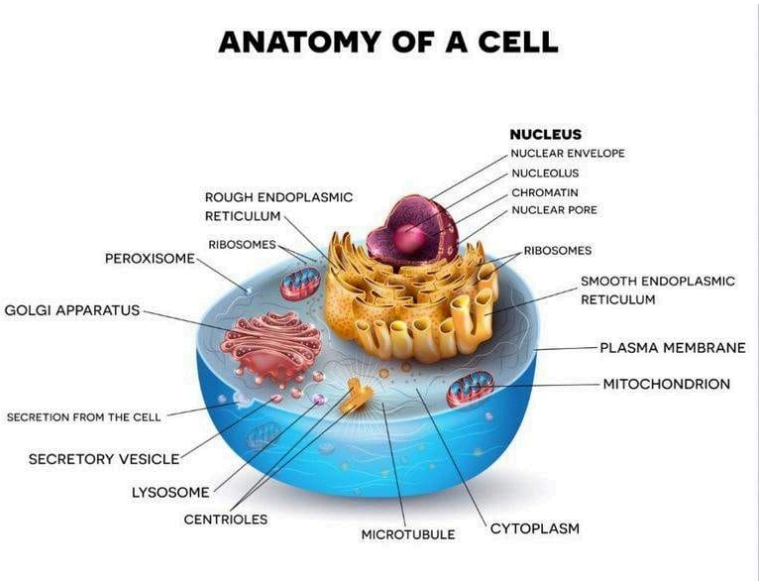
Learning Objectives

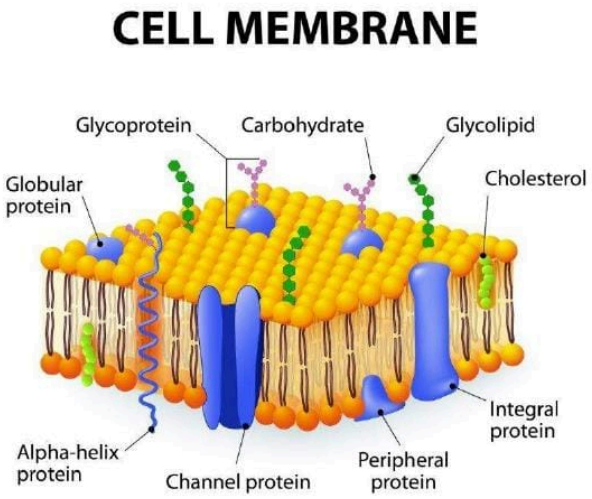
- 1: Describe the basic components of cell structure and organelles.
- 2: Understand the chromatin organization into nucleus.
- 3: Understand basic mechanisms of cell signaling to respond to changes in the environment.

2.1

Structure of a cell

Cells are the smallest unit of life. A typical eukaryotic cell and it components are illustrated in the next figure.





Membrane:

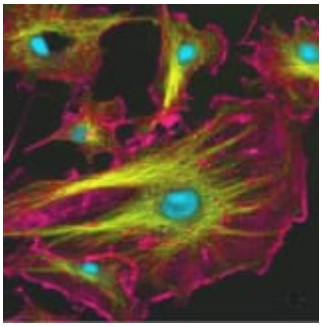
Plasma membrane, nuclear envelope, and membranes provide separation of different environments to permit a variety of biological functions. Membranes are dynamic and fluid structures that allow selective movement of ions, energy sources, vitamins and cofactors, and waste. Membrane components include lipid, carbohydrates and protein molecules. Membrane fluidity is very important for its functions and, as a consequence, is important in disease processes, as well as, treatments.

Figure 8.4.2.2.1: Structure of cell membrane

Membranes are composed of lipids arranged in a **lipid bilayer,** with the hydrophilic glycerol and phosphate “head” groups of the lipid molecules forming the two outside layers and the hydrophobic “tail” groups arranged inside. Proteins are the second major part of biological membranes and makeup approximately 20%–80% of both the structural and functional components of these membranes. Many of these proteins are embedded into the membrane and stick out on both sides; these are called **transmembrane proteins.**

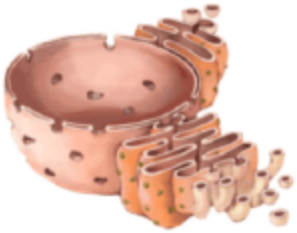
2.2

Subcellular compartments



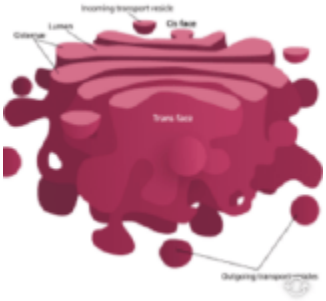
CYTOSKELETON

is a structure that helps cells maintain their shape and internal organization, and it also provides mechanical support that enables cells to carry out essential functions like division and movement. The cytoskeleton is made up of microtubules, actin filaments, and intermediate filaments.



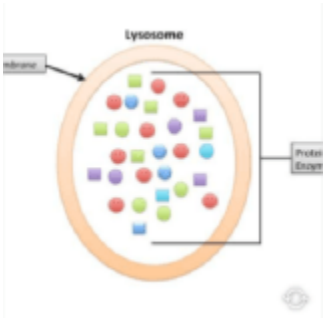
ENDOPLASMIC RETICULUM (ER)

It consists of a system of sac- and tube-like structures, which locally expand into cisterns. Its internal lumen is connected with the intermembrane space of the nuclear membrane. Part of the ER is studded on the outside with ribosomes (rough ER), which take part in protein synthesis. The other part of the ER is free of ribosomes (smooth ER). Enzymes of the smooth ER are involved in the synthesis of fatty acids. The smooth ER also plays a role in detoxification by hydroxylation reactions.



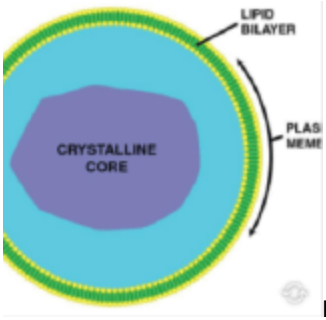
GOLGI APPARATUS

This organelle consists of stacks of flattened membrane sacs. Their main function is the further processing and sorting of proteins and their export to the final targets. In most cases, these are secreted as membrane proteins. In addition, the Golgi apparatus also produces polysaccharides.



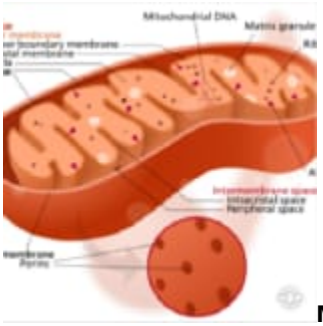
LYSOSOMES

Lysosomes are vesicles enclosed by a lipid bilayer. These organelles are filled with many enzymes for polysaccharide, lipid, protein and nucleic acid degradation. They act also on intracellular material to be removed and even contribute to the apoptosis of their own cell. Lysosomes of special cells (e.g., macrophages) destroy bacteria or viruses as a defense mechanism.



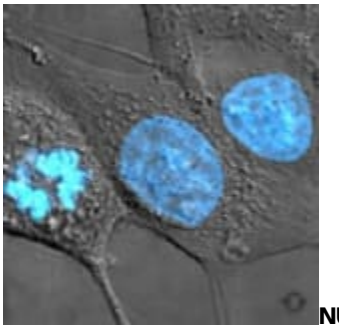
PEROXISOMES

Peroxisomes are surrounded by a single membrane. They are generated from components of the cytosol and do not bud from other membranes. The main task of these organelles is the performance of monooxygenase (hydroxylase) or oxidase reactions, which produce hydrogen peroxide (H2O2).



MITOCHONDRIA

In a typical eukaryotic cell, there are in the order of 2000 of these organelles, which are often of ellipsoidal shape. They have a smooth outer membrane and a highly folded inner membrane with numerous invaginations (cristae), which contain most of the membrane-bound enzymes of mitochondrial metabolism. Mitochondria are the site of respiration and ATP synthesis, but also of many other central reactions of metabolism, e.g., citrate cycle, fatty acid oxidation, glutamine formation, and part of the pathway leading to steroid hormones. Mitochondria are the only organelles which are equipped with their own (circular) DNA, RNA and ribosomes and thus can perform their own protein synthesis.



NUCLEUS

All eukaryotic cells show the presence of a separate nucleus, which contains the major portion of the genetic material of the cell (DNA). The nuclear DNA is organized in a number of chromosomes. The nucleus is surrounded by a double membrane of lipid bilayers with integrated proteins, called the nuclear membrane (also known as the nuclear envelope). Nuclear pores span the nuclear membrane and enable the transport of proteins, rRNA etc.

Image credits

2. DataBase Center for Life Science (DBCLS) licensed under CC BY 4.0

3. Kelvin Song licensed under CC BY 3.0

4. Lumoreno licensed under CC BY-SA 3.0

5. Agateller licensed under CC BY-SA 3.0

2.3

Structure of Chromatin

Chromosomal DNA is packaged inside microscopic nuclei by its association of histones H2A, H2B, H3, and H4. These are positively-charged proteins that strongly adhere to negatively-charged DNA and form complexes called nucleosomes. (11-nm "beads on a string" structure). Each nucleosome is composed of DNA wound 1.65 times around eight histone proteins. Nucleosomes fold up to form a 30-nanometer chromatin fiber, which forms loops averaging 300 nanometers in length. The 300 nm fibers are compressed and folded to produce a 250 nm-wide fiber, which is tightly coiled into the chromatid of a chromosome. The level of structure varies depending on the cell cycle stage and, as a result, the requirement for DNA transcription or replication.

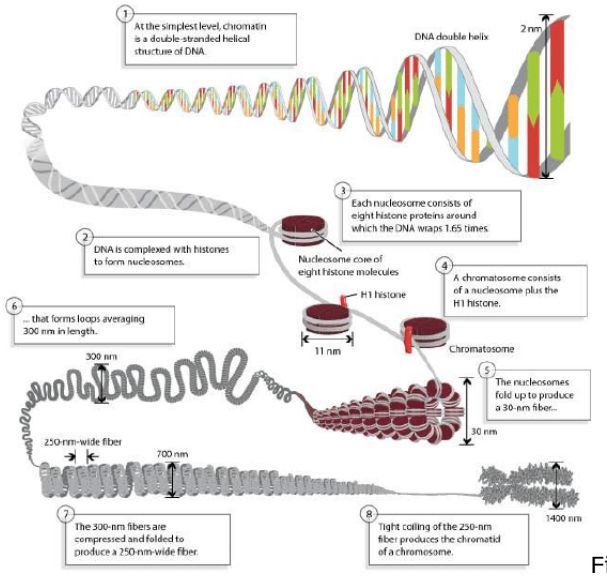
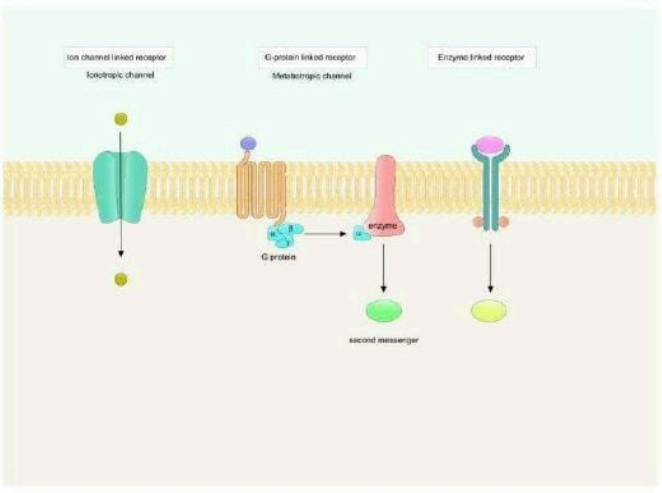


Figure 8.4.2.2.2: Chromosomes are composed of DNA tightly-wound around histones - © 2010 Nature Education

Chromatin: The most important structure inside the nucleus is **chromatin**, consisting, in humans, of the 46 chromosomes. Chromatin is the combination or complex of DNA and proteins that make up the contents of the nucleus of a cell. The most abundant protein in the nucleus are histones. Histones are rich in basic amino acids (positively charged), which interact with negative charges of the DNA.

2.4

Cell Signaling



Cell signaling consists in the ability of cells to respond to environment changes through signals received outside their borders. Cells may receive many signals simultaneously and also send out messages to other cells.
Cells have proteins called receptors, that are generally transmembrane proteins, which bind to signaling molecules outside the cell and subsequently transmit the signal through a sequence of molecular switches to internal signaling pathways and initiate a physiological response. Different receptors are specific for different molecules. In fact, there are hundreds of receptor types found in cells, and varying cell types have different populations of receptors.

Figure 8.4.2.2.3: Main types of transmembrane receptors

Examples of receptors membrane:

- G-protein-coupled receptors
- Ion channel receptors
- Enzyme-linked receptors

Receptor may be located in the cellular membrane (important to receive extracellular signals) but also may be present inside the cell or inside the nucleus.

Topic 2: Key Points

In this section, we explored the following main points:

- 1: The characteristics of the cell membrane, such as lipid bilayer and transmembrane proteins are very important for the cell function.
- 2: The chromatin is the combination of complex DNA and proteins that make up the contents of the nucleus of a cell.
- 3: Cells have proteins called receptors that are important to receive extracellular signals and initiate signaling pathways in the cell.

Knowledge Check

1. Histones are proteins present in the nucleus of the cells with the function of:

interacting with the DNA to form nucleosomes.

regulating the fluidity of cell membranes.

Answer available at the end of this section.

2. Membranes are composed of lipids arranged in a ...

Check all that apply.

- lipid monolayer, with...
- lipid bilayer, with...
- ...the hydrophilic glycerol and phosphate “head” groups of the lipid molecules forming the two outside layers and the hydrophobic “tail” groups arranged inside.
- ...the hydrophilic and hydrophobic “tail” groups arranged inside.

Answer available at the end of this section.

Answers

Knowledge Check

interacting with the DNA to form nucleosomes.

Knowledge Check

lipid bilayer, with...the hydrophilic glycerol and phosphate “head” groups of the lipid molecules forming the two outside layers and the hydrophobic “tail” groups arranged inside.

This page titled 8.4.2.2: Cell Structure and Subcellular Compartments is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.2: Cell Structure and Subcellular Compartments by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.2.3: DNA and RNA Metabolism

Learning Objectives

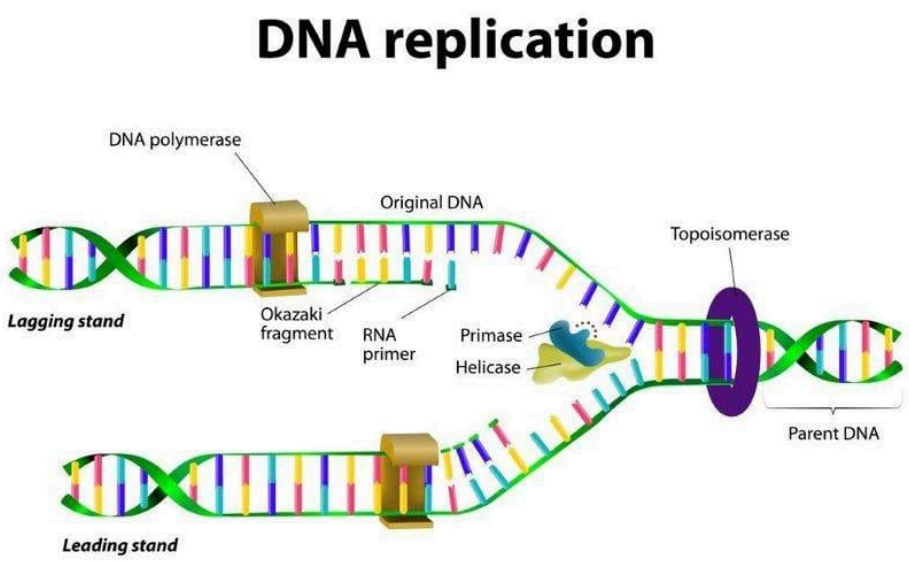
After completing this lesson, you will be able to:

- Understand the mechanism of DNA replication.
- Understand the process of gene expression.
- List the main types of DNA mutation and mechanism of DNA repair.

DNA Replication

The process of DNA replication consists of uncoiling double-stranded DNA, copying each DNA strand and then separating the two, new, double-stranded copies. The process starts at an **origin of replication (ori)**, a nucleic acids sequence where the replication can start. There are around 100,000 origins of replication in each human cell. This means that the DNA replication may start simultaneously in different positions at the same time. The **replication fork** is the point where two DNA strands, one termed the **leading strand** and the other the **lagging strand**, are separated and DNA copying occurs. The coiled-coil, double-helical DNA structure is initially unwound by the enzyme **DNA helicase** by breaking the hydrogen bonds between complementary nucleic acids. **Single-stranded binding proteins** attach to the new DNA strands to keep them separated.

An enzyme termed **primase** then produces a short strand of RNA to serve as a primer for the remainder of the process. The enzyme **DNA polymerase** replicates each DNA strand in the 5′ to 3′ direction by adding the correct, matching nucleotide triphosphate to the 3′-hydroxyl end of the primer strand. As each new nucleic acid is added, a new phosphodiester bond is formed, utilizing the energy contained in the remaining diphosphate group.



RNA Transcription

RNA Transcription is the process whereby a particular segment of DNA is transcripted into an equivalent RNA sequence.

- mRNA: For genes that codes for a protein.
- tRNA: For a transfer RNA.
- rRNA: For assembly of a ribosome.
- miRNA (micro RNA): Which binds to mRNA and inhibits its translation.

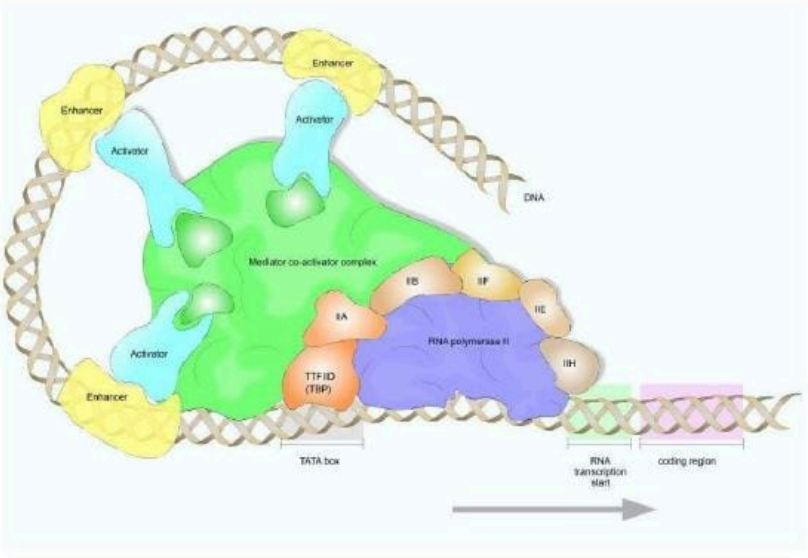
- siRNA (small-interfering RNA): Which binds to mRNA and aids in its degradation.
- snRNA (small nuclear RNA): Which participates in RNA processing as part of the spliceosomes.
- snoRNA (small nucleolar RNA): Which participate in nucleolar RNA processing.

The transcription starts with binding of the enzyme **RNA polymerase** to a **promoter** sequence on the DNA, a regulatory region that dictates where the transcription should start.

The DNA is transcribed from 3′ to 5′ and occurs only on one of the DNA strands, the **template strand.**

As in DNA replication, energy for the formation of the phosphodiester bond is derived from hydrolysis of the two terminal phosphate bonds of the nucleoside triphosphate.

Multiple RNA polymerases can transcribe on a single DNA gene sequence, allowing rapid production of the RNA product.



Enhancer: This is a short region of DNA that can be bound by transcription factors to increase or facilitate the transcription of a particular gene. They can be located up away from the gene, upstream, or downstream from the start site.

Transcription Factors (TFs): These include a wide number of proteins, excluding RNA polymerase, that promotes (as an activator), or blocks (as a repressor) the recruitment of RNA polymerase to DNA. TFs bind to promoter regions of DNA.

Promoter Region: These are specific DNA sequences, usually located upstream and near the transcription start sites of genes and serves as a binding site for proteins called transcription factors that recruit RNA polemerase. Example: TATA bos, CpG island.

Figure 8.4.2.3.1: Machinery of RNA transcription

RNA Processing

The newly synthesized RNA transcripts are processed prior to their use in the cell as mature RNA.

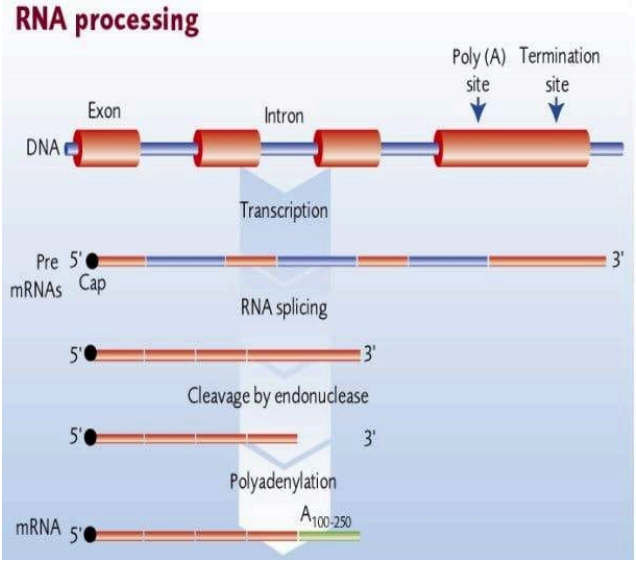


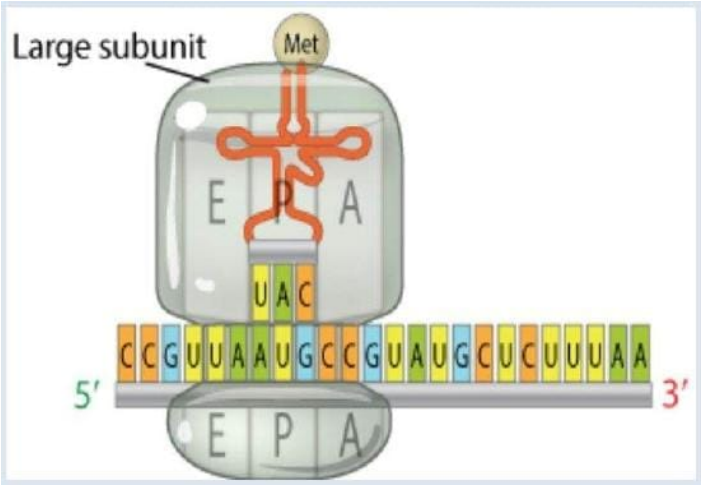
Figure 8.4.2.3.2: RNA processing

A 7-methyl guanosine nucleic acid is added to the 5′-end (known as a **5′ cap**) of the pre-mRNA as it emerges from RNA polymerase II (Pol II). The cap protects the RNA from being degraded by enzymes and serves as an assembly point for the proteins to begin translation to protein.

Removal of **introns** present in the pre-mRNA and splicing of the remaining **exons**, in a process called **RNA splicing**. The continuous series of DNA bases coding for a protein are interrupted by base sequences that are not translated. The translated sequences are referred to as exons (expressed sequences) and the nontranslated sequences as introns (intervening sequences). This completes the mRNA molecule, which is now ready for export to the cytosol. (The remainder of the transcript is degraded, and the RNA polymerase leaves the DNA.)

3.4 Protein Translation

Protein synthesis requires the interaction of **mRNA, tRNA**, several accessory proteins, called **initiation factor (IF)** and **elongation (EF)** factor, and **ribosomes.**

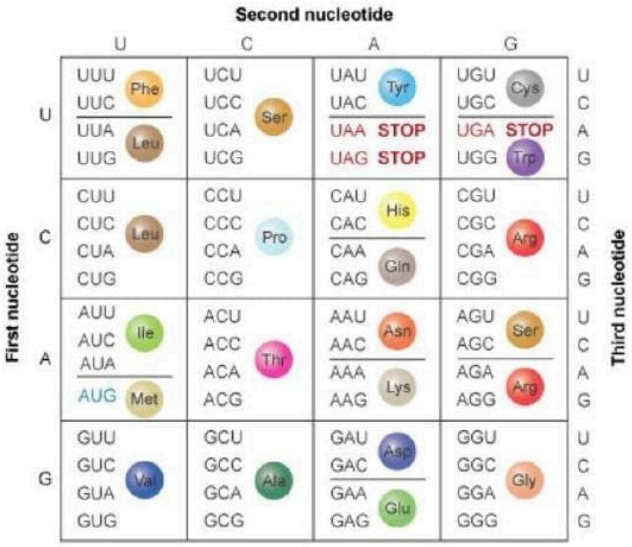


The large ribosomal subunit binds to the small ribosomal subunit to complete the initiation complex. The tRNA molecule bind to one amino acids at the top of the structure. In the base if the tRNA there is the three base sequence known as the

anticodon

. This anticodon binds, through base pairing, to a three base codon on mRNA. It is this interaction between the mRNA and the amino acid-tRNA that provides the high degree of fidelity observed in the transfer of genetic information from DNA to proteins.

Figure 8.4.2.3.3: Transfer RNA (tRNA) - © 2013 Nature Education



The Triplet Code (Genetic Code)

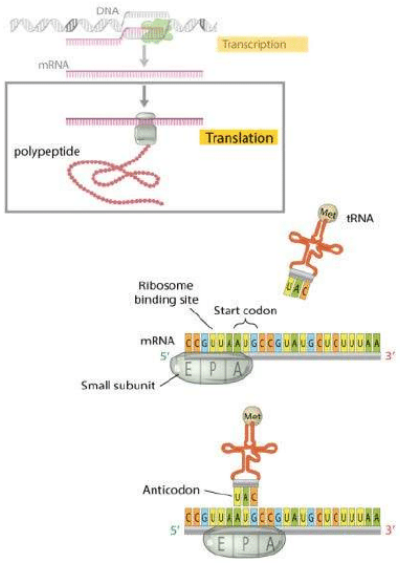
A sequence of three bases in DNA identifies each of the 20 amino acids that are to be incorporated into the newly synthesized protein. This information is incorporated into mRNA which is synthesized using DNA as the template. Considering that three bases are required at a minimum, and we have 4 nucleotides (43 = 64 code words are possible). This is more than the 20 amino acids. In fact, many triplets are used to define one same amino acid. In addition, some “extra” triplet sequences are used as stop codons to terminate protein synthesis. AUG is used as the start codon for the N-terminal amino acid in eukaryotes.

Using the same binding rules as DNA double strands (e.g., a tRNA that binds the starting mRNA codon AUG has an anticodon sequence of UAC), insuring the specific order of AAs required for proper production of the protein.

Figure 8.4.2.3.4: Genetic code - © 2014 Nature Education

RNA plays three distinct and important roles:

- mRNA: The intermediary between gene and protein – provides the message.
- tRNA: The key or adaptor – reads the genetic code, brings amino acids to the growing polypeptide chain
- rRNA: In ribosome, provides a scaffold for protein synthesis, catalyzes peptide bond formation



The IF and EF accessory proteins serve a number of roles, including enabling binding of the mRNA molecule to the ribosome,

movement of the mRNA along the ribosome to the start point of the synthesis, docking of the tRNA–amino acid, and movement of the mRNA and growing peptide chain, as well as accuracy assurance.

The protein biosynthesis, can be divided into three phases: initiation, elongation, and termination.

Initiation

Initiation of protein synthesis begins when the protein initiator factor IF-3 binds to the small subunit of the ribosome and causes its dissociation. The small ribosomal subunit then binds to the 5' side of mRNA which carries information in a triplet code from DNA. The small subunit is then translocated where it meets the large ribosomal subunit, other protein initiator factors, and initiator tRNA. The tRNA is bound to methionine at a site in the ribosome known as the P site.

Elongation

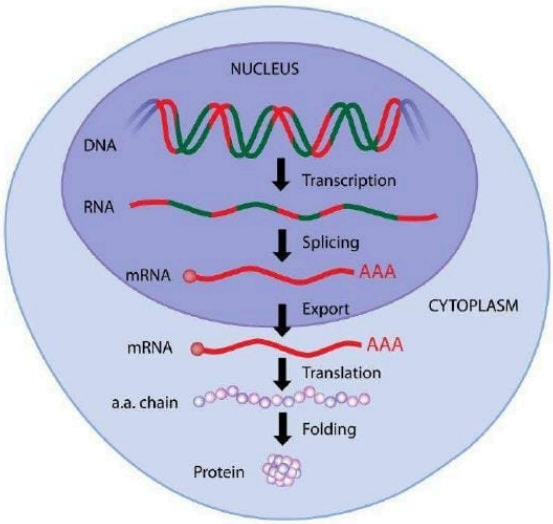
In the elongation phase of protein synthesis, a specific aminoacyl-tRNA, directed by hydrogen bonding interactions between the anticodon region of the aminoacyl-tRNA and the codon region of mRNA, adds to a site distinct from the P site, the A site. The A and P sites are in close proximity, allowing the peptide bond formation between the amino acids. The newly synthesized tRNA bound dipeptide then moves from the A site to the P site. After a translocation of the ribosome in the 5' - 3' direction along the mRNA occurs to expose a new codon. Then, another amino acid-tRNA identity and binds to the mRNA at the A site and the peptidyl transferase reaction is again initiated. As the polypeptide chain grows through subsequent cycles of amino acid residue incorporation, it emerges from the ribosome and undergoes folding into its native secondary and tertiary conformations.

Termination

In termination step, the peptide bond synthesis ceases when a stop codon on the mRNA is reached. This termination site will not bind aminoacyl-tRNA and peptide synthesis stops. Release factors allow the newly synthesized protein to dissociate from the ribosome.

3.5 Gene Expression

Gene expression is a two-step process in which DNA is converted into a protein.



Step 1: The first step is DNA transcription to RNA. In this step, the information from the archival copy of DNA is imprinted into mRNA. The structure of RNA is a little different, it contains ribose instead of deoxyribose, and the four bases that bind to it are cytosine (C), guanine (G), adenine (A), and uracil (U). During transcription, DNA unfolds, and mRNA is created by pairing mRNA bases with the bases of DNA. In this process C in DNA translates to G, G to C, **A to U**, and T to A. After mRNA is transcripted it is transported to the ribosome.

Step 2: The second step, protein translation occurs at the ribosome. During translation, the sequence of codons (triplet of bases) of mRNA is, with the help of tRNA, translated into a sequence of amino acids

Figure 8.4.2.3.5: Gene expression seems to be a straightforward process, the mechanism that control the gene expression that causes most phenotypic differences in organisms.

3.6 Mutation



Mutations are changes in the genetic sequence (DNA or RNA sequence) , and they are a main cause of diversity among organisms. Although some of mutations are beneficial, offering resistance to disease or improved structure and/or function, some other specific mutations can lead to disease and/or death of the cell or organism.

Mutations can occur due to assaults from the environment or spontaneous mutation may occur during the DNA replication. Mutations are estimated to occur at an approximate rate of 1000–1,000,000 per cell per day in the human genome, and every new cell is believed to contain approximately 120 new mutations.

TYPES OF MUTATION

Point mutations when only a single base pair is changed into another base pair. They can be classified as the following:

- **Transition:** When a purine nucleotide is changed to a different purine (A ↔ G) or a pyrimidine nucleotide is changed to a different pyrimidine nucleotide [C ↔ T(U)].
- **Transversion:** When the orientation of a single purine and pyrimidine nucleotide is reversed [A/G ↔ C/T(U)].
- **Silent:** When the same AA is coded.
- **Missense:** When a different AA is coded.
- **Neutral:** When an AA change occurs but does not affect the protein's structure or function.
- **Nonsense:** When a stop codon results, terminating translation and shortening the resulting protein.

Insertion and ***deletion*** mutations, which are together known as ***indels.*** Indels can have a wide variety of lengths. At the short end of the spectrum, indels of one or two base pairs within coding sequences have the greatest effect, because they will inevitably cause a frameshift, i.e. change the entire reading of the mRNA sequence. At the intermediate level, indels can affect parts of a gene or whole groups of genes. At the largest level, whole chromosomes or even whole copies of the genome can be affected by insertions or deletions. At this high level, it is also possible to invert or translocate entire sections of a chromosome, and chromosomes can even fuse or break apart.

If a large number of genes are lost as a result of one of these processes, then the consequences are usually very harmful.

3.7 DNA Repair

The human body have mechanisms to detect and repair the various types of damage that can occur to DNA, no matter whether this damage is caused by the environment or by errors in replication.

Because DNA is a molecule that plays an active and critical role in cell division, during the cell cycle, checkpoint mechanisms ensure that the DNA is intact before permitting DNA replication and cell division to occur. Failures in these checkpoints can lead

to an accumulation of damage, which in turn leads to mutations.

UV radiation causes DNA lesions that may distort DNA's structure, introducing bends or kinks and thereby impeding transcription and replication. These lesions may be repaired through a process known as **nucleotide excision repair (NER)**, a mechanism where an enzyme catalyze the removal of damaged nucleotides, and replacement of the correct sequence, guided by the intact complementary DNA strand. Defects in this mechanism is related to human diseases like skin cancer.

Another repair mechanism that handles the spontaneous DNA damage caused oxidation or hydroxylation generated by metabolism is the **base excision repair (BER)**. In this mechanism, enzymes known as DNA glycosylases remove damaged bases by literally cutting them out of the DNA strand through cleavage of the covalent bonds between the bases and the sugar-phosphate backbone. The resulting gap is then filled by a specialized repair polymerase and sealed by ligase.

DNA damage also may occur in form of double-strand breaks, which are caused by ionizing radiation, including gamma rays and X-rays. Double-strand breaks may be repaired through one of two mechanisms: **nonhomologous end joining (NHEJ)**, where an enzyme called DNA ligase IV uses overhanging pieces of DNA adjacent to the break to join and fill in the ends; or **homologous recombination repair (HRR)** where the homologous chromosome itself is used as a template for repair.

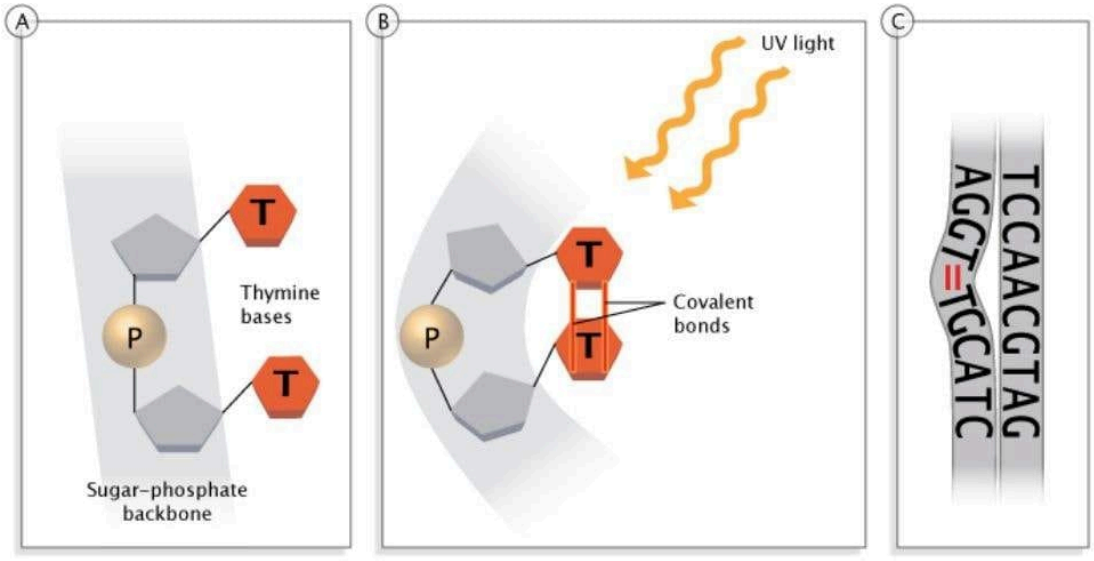


Figure 8.4.2.3.6: UV Radiation can distort DNA, impeding transcription and replication. A region of DNA is shown before (A) and after (B) its structure is distorted by UV radiation. The distortion in panel B is due to a convex bend in one strand (C). - Source: © Nature Education

Topic 3: Key Points

In this section, we explored the following main points:

1. The process of DNA replication is controlled by several proteins that act together to assure the correct base pairing for creation of the new DNA strand.
2. Transcription is the first step of gene expression, in which a particular segment of DNA is copied into RNA (mRNA) by the enzyme RNA polymerase.
3. In translation, messenger RNA (mRNA)—produced by transcription from DNA—is decoded by a ribosome and tRNA to produce a specific amino acid chain, or protein.
4. Mutations are changes in the genetic sequence (DNA or RNA sequence), that can be beneficial or may result in damage, if not repaired.

Knowledge Check

1. RNA splicing is the process which involves the removal of introns present in the pre-mRNA and splicing of the remaining exons. True or False?

True

False

Answer available at the end of this section.

2. Which of the following does not belong to the process of DNA replication?

Primase

DNA Polymerase

RNA Polymerase

Answer available at the end of this section.

3. During protein translation, the sequence of codons (triplets of bases) of mRNA is important to:

Maintain the structure of the mRNA

Translated the correct a sequence of amino acids.

DNA replication

Answer available at the end of this section.

Answers

Knowledge Check

true

Knowledge Check

RNA Polymerase

Knowledge Check

Translated the correct a sequence of amino acids.

This page titled 8.4.2.3: DNA and RNA Metabolism is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.3: DNA and RNA Metabolism by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

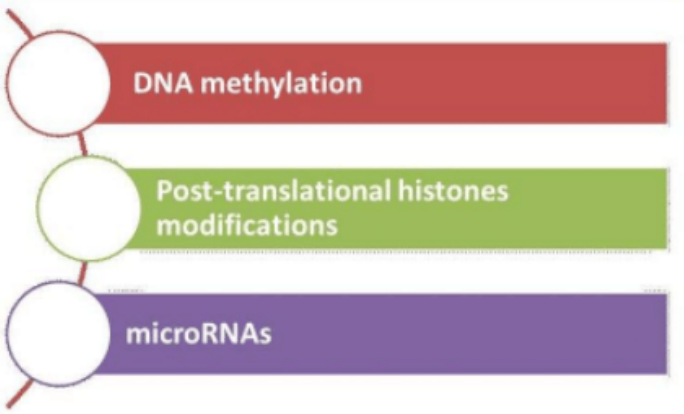
8.4.2.4: Epigenetic Mechanisms

Learning Objectives

- 1: Identify the main epigenetic mechanisms related to control of gene expression.

4.1

Introduction

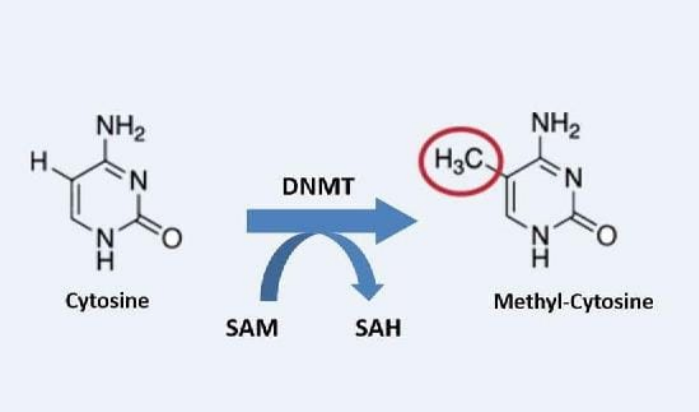


Epigenetics is defined as potentially heritable and reversible changes in gene expression mediated by methylation of DNA, modifications of histone proteins or by non-coding RNAs that are not due to any alteration in the DNA sequence. These processes singularly or jointly affect transcript stability, DNA folding, nucleosome positioning, chromatin compaction, and ultimately nuclear organization. They determine whether a gene is silenced or activated and when and where this occurs.

Epigenetic change is a regular and natural occurrence, essential for normal cell development, but can also be influenced by several factors including age, the environment/lifestyle, and disease state.

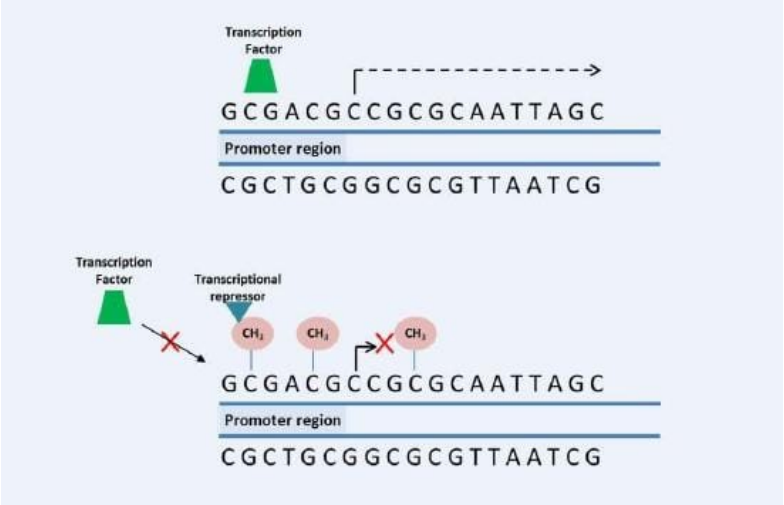
4.2

DNA Methylation



DNA methylation is a covalent modification of DNA, in which a methyl group is transferred from S-adenosylmethionine (SAM), that is converted to S-adenosylhomocisteine (SAH), to the 5 position of cytosine by a family of enzymes known as DNA methyltransferases (DNMT).

Figure 8.4.2.4.1: Disaccharide structure



DNA methylation occurs predominantly in cytosines located in proximity of guanines, known as CpG dinucleotides (CpGs) or CpG island. These CpG island are found in promoter region of genes. The methyl group inhibits the binding of transcription factors to their recognition site, resulting in inhibition of gene transcription. Furthermore, the methyl group may attract other proteins know as transcriptional repressors that contributes for inhibition of gene transcription.

Figure 8.4.2.4.2: DNA methylation in the promoter region – Credit Aline de Conti

4.3

Histones Post-Translational Modifications

The nucleosome is composed of five histone proteins (H1, H2A, B, H3, and H4). The N-terminus of these histone proteins are subject to covalent modifications such as methylation, phosphorylation, acetylation, ubiquitination or sumoylation by a group of histone-modifying enzymes . Alterations in these proteins contribute to the accessibility and compactness of the chromatin, and result in activation or suppression of particular genes.

EXAMPLES OF TYPES AND ROLES OF HISTONE MODIFICATIONS

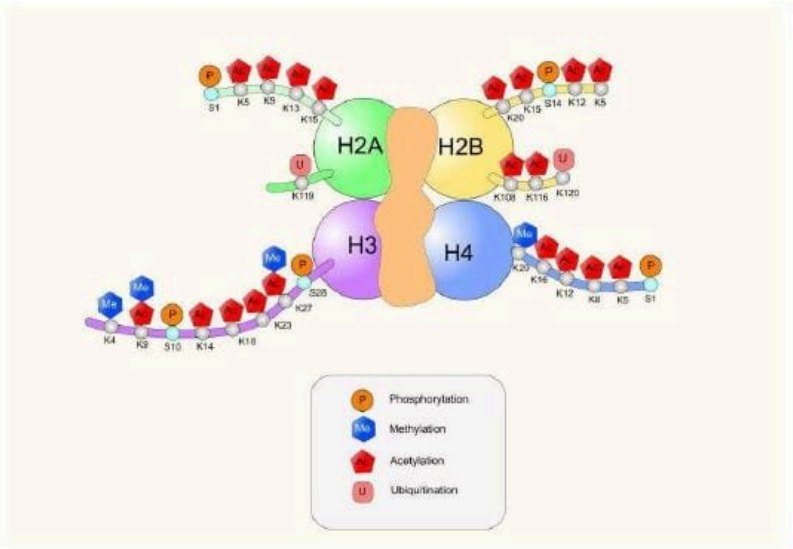


Figure 8.4.2.4.3: Histone sites of post-translational modifications.

Activation of gene Transcription:

Acelylation in the lysine 9(K9) of histone 3(H3) - H3K9ac

Acetylation in the lysine 27(K27) of histone 3(H3) - H3K27ac

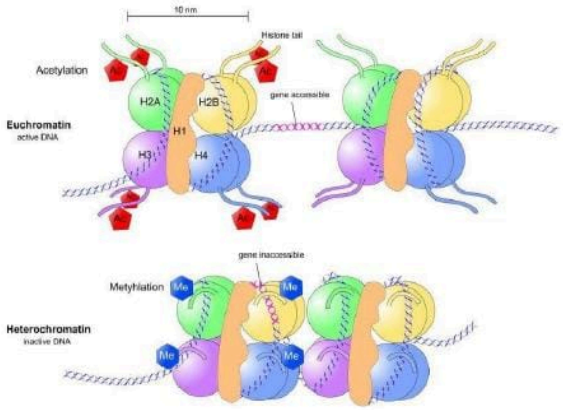
Trimethylation in the lysine 4(K4) of histone 3(H3) - H3K4me3

Repression of gene transcription:

Trimethylation in the lysine 4(K3) of histone 3(H3) - H3K9me3

Trimethylation in the lysine 27(K27) of histone 3(H3) - H3K27me3

Trimethylation in the lysine 20(K20) of histone 3(H3) - H3K20me3



DNA methylation and histone post-translational modifications play important role in the establishment of chromatin structure and in consequence in the gene expression modulation. These epigenetic mechanisms can be independent or can happen together to control the gene expression. They also may influence each other, for example the histone methylation can help to direct DNA methylation patterns, and DNA methylation might serve as a template for some histone modifications after DNA replication.

Figure 8.4.2.4.4: Histone acetylation and methylation in gene regulation.

4.4

microRNAs

miRNAs have an important role in gene regulation and they can influence biological functions including cell differentiation and proliferation during normal development and pathological responses. They are small non-coding RNA molecules (containing about 22 nucleotides), derived from regions of RNA transcripts that fold back on themselves to form short hairpins. They regulate gene expression at the post-transcriptional level. A number of miRNAs may bind to specific regions of the messenger RNA (mRNA) and block its translation to proteins. Alteration of the expression of miRNAs is believed to contribute to the progression of tumorigenesis and other diseases.

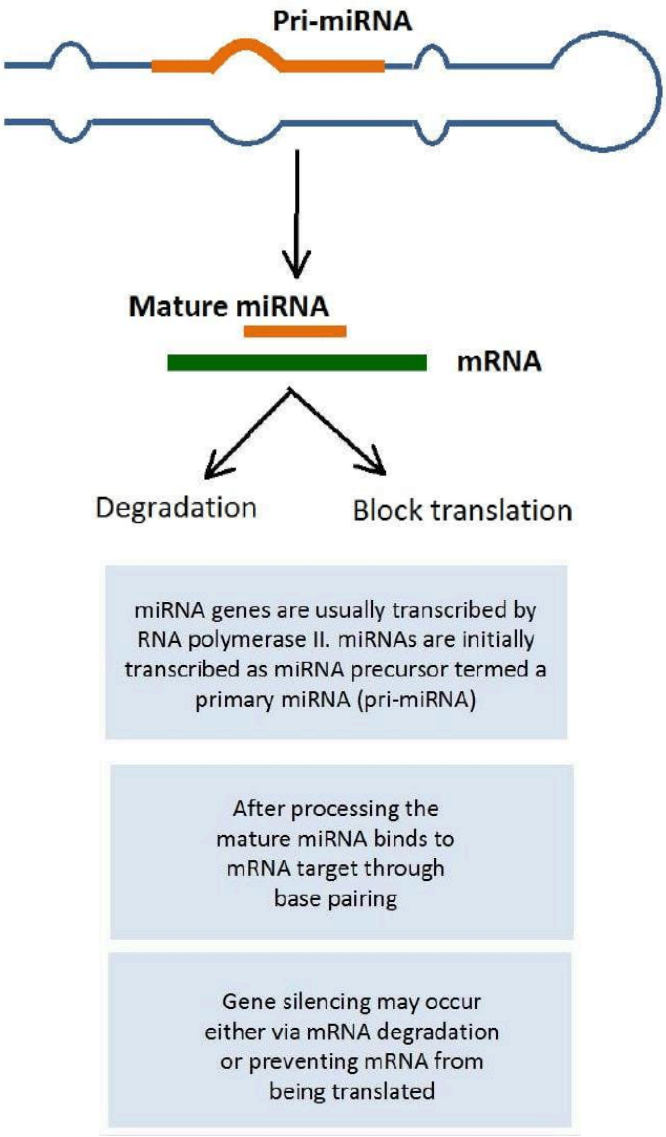


Figure 8.4.2.4.5: Schematic representation of gene silencing by microRNA. Credit: Aline de Conti

Topic 4: Key Points

In this section, we explored the following main points:

- 1: Epigenetic mechanisms may influence gene expression without alteration in the DNA sequence.
- 2: Main epigenetic mechanisms are DNA methylation, histones post-translational modification and alteration in the

expression of microRNAs.
- 3: Epigenetic change is a regular and natural occurrence but can also be influenced by several factors including age, the environment/lifestyle, and disease state.
- 4: Different from genetic alterations, epigenetic alterations are considered reversible.

Knowledge Check

1. Epigenetic mechanisms are considered:

    Reversible

    Irreversible

Answer available at the end of this section.

2. DNA methylation and histone post-translational modifications play important role in the establishment of chromatin structure and in consequence in the gene expression modulation. *True or False?*

    False

    True

Answer available at the end of this section.

3. DNA methylation may inhibit the binding of transcription factors to their recognition site, at promoter regions resulting in:

    Inhibition of gene transcription

    Activation of gene transcription

Answer available at the end of this section.

Answers

Knowledge Check
Reversible

Knowledge Check
True

Knowledge Check
Inhibition of gene transcription

This page titled 8.4.2.4: Epigenetic Mechanisms is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.4: Epigenetic Mechanisms by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.2.5: Section 2 Final Evaluation

1. The carbohydrate component of a nucleoside and nucleotide is usually:

The sugar ribose for RNA molecule and deoxyribose for DNA molecule

Carboxylic group for DNA molecule and amino group for RNA molecule

The deoxyribose for RNA molecule and sugar ribose for DNA molecule

The deoxyribose for RNA molecule and deoxyribose for DNA molecule

Answer

The sugar ribose for RNA molecule and deoxyribose for DNA molecule

2. During protein translation, the sequence of codons (triplets of bases) of mRNA is important for:

Maintaining the structure of the mRNA

Translating the correct sequence of an amino acid

DNA replication

DNA synthesis

Answer

Translating the correct sequence of an amino acid

3. In DNA replication, the lagging strand short chains of nucleic acids, called Okazaki fragments, are generated and the enzyme DNA _____ joins the Okazaki fragments together with lagging strand for replication to proceeds.

topoisomerase

polymerase

helicase

ligase

Answer

ligase

4. Point mutation when the orientation of a single base pair purine and pyrimidine nucleotide is reversed is termed:

Transition

Missense

Transversion

Nonsense

Answer

Transversion

5. The main function of _____ is to control the fluidity of biological membranes.

Glycerol

Cholesterol

Glucose

Phospholipid

Answer

Cholesterol

6. The smooth endoplasmic reticulum (ER) is involved in the synthesis of fatty acids and

Detoxification by hydroxylation reactions

Provides mechanical support for cells

Monooxygenase reaction to produce hydrogen peroxide (H2O2)

Production of energy

Answer

Detoxification by hydroxylation reactions

7. The only organelles which are equipped with their own (circular) DNA, RNA and ribosomes and can perform their own protein synthesis are:

Nucleus

Golgi apparatus

Peroxisomes

Mitochondria

Answer

Mitochondria

8. Phospholipids are the major component of cell membranes forming lipid bilayers because of their amphiphilic property whose structure consist of:

Two hydrophilic fatty acid “tails” and a hydrophobic head consisting of a phosphate group (PO4-3) attached to the third glycerol carbon.

Two hydrophobic fatty acid “tails” and a hydrophobic head consisting of a phosphate group (PO4-3) attached to the third glycerol carbon.

Two hydrophobic fatty acid “tails” and a hydrophilic head consisting of a phosphate group (PO4-3) attached to the third glycerol carbon.

Four hydrophilic fatty acid “tails” and a hydrophobic head consisting of a phosphate group (PO4-3) attached to the third glycerol carbon.

Answer

Two hydrophobic fatty acid “tails” and a hydrophilic head consisting of a phosphate group (PO4-3) attached to the

third glycerol carbon.

9. The most important structural component of a nucleus consisting of positively charged protein and negatively charged DNA is called:

Lysosome

Chromatin

Cytoskeleton

Nucleolus

Answer

chromatin

10. The difference between amino acids and fatty acids in terms of structure is that amino acids carry a carboxyl and an amino group while fatty acids carry a carboxyl and a methyl group.

True

False

Answer

true

11. An example of a nucleotide with a two ring structures is:

Cytosine (C)

Adenine (A)

Uracil (U)

Thymine (T)

Answer

Adenine (A) [and also Guanine(G)]

12. DNA methylation occurs predominantly in _____ to inhibit the binding of transcription factors to their recognition site which results in the inhibition of gene transcription.

Guanine

Cytosine

Adenine

Thymine

Answer

Cytosine (C)

This page titled 8.4.2.5: Section 2 Final Evaluation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 2 Final Evaluation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.3: PRINCIPLES OF GENETIC TOXICOLOGY

The genetic toxicology methodology or assay technique helps to test or evaluate the level of damages of the genetic information caused by toxicants or agents within the cells. In this e-module you will learn about different genetic toxicology assays, different genetic damages, and cytotoxicity and epigenetics assays.

8.4.3.1: Introduction to Genetic-toxicology Assay

8.4.3.2: Different Genetic Damages or Mutations

8.4.3.3: Different Genetic-Toxicology Assays

8.4.3.4: Different Cytotoxicity Assays

8.4.3.5: Epigenetics Assay

8.4.3.6: Section 3 Final Evaluation

This page titled 8.4.3: Principles of Genetic Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3: Principles of Genetic Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.3.1: Introduction to Genetic-toxicology Assay

Learning Objectives

- Know about the definition of genetic toxicology assay.

1.1. What is Genetic-Toxicology Assay?

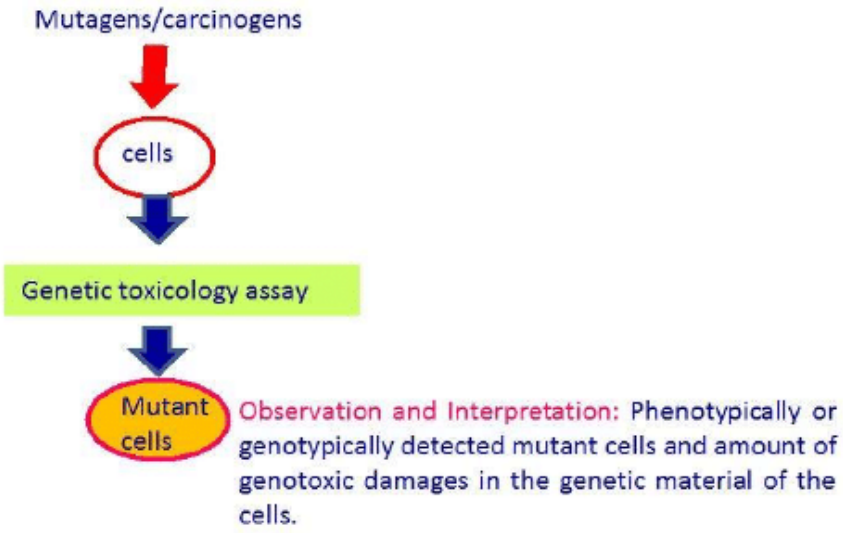


Figure 8.4.3.1.1: Diagrammatic steps in genetic toxicology assay.

The genetic-toxicology methodology or assay technique helps to test or evaluate the level of damage to the genetic information caused by toxicants or agents within the cells (Figure 1). Damages resulted as induced **mutations**, which may lead to different diseases including cancer. The causative toxic agents are know as **mutagens**. Mutagens caused for cancer disease, is known as **carcinogens.**

Topic 1: Key Points

In this section, we explored the following main points:

- 1: Definition of genetic toxicology methodology
- 2: Causative toxic agents as mutagens.
- 3: Genotoxic damages result in the phenotypically or genotypically mutant cells.

Knowledge Check

Genetic-toxicology methodology or assay technique helps to test...

the level of damages caused by toxicants within the cells.

the level of reactive oxygen species (ROS) within the cells.

the stages of cancer disease.

Answer available at the end of this section.

Answers

Knowledge Check

the level of damages caused by toxicants within the cells.

This page titled 8.4.3.1: Introduction to Genetic-toxicology Assay is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.1: Introduction to Genetic-toxicology Assay by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.3.2: Different Genetic Damages or Mutations

Learning Objectives

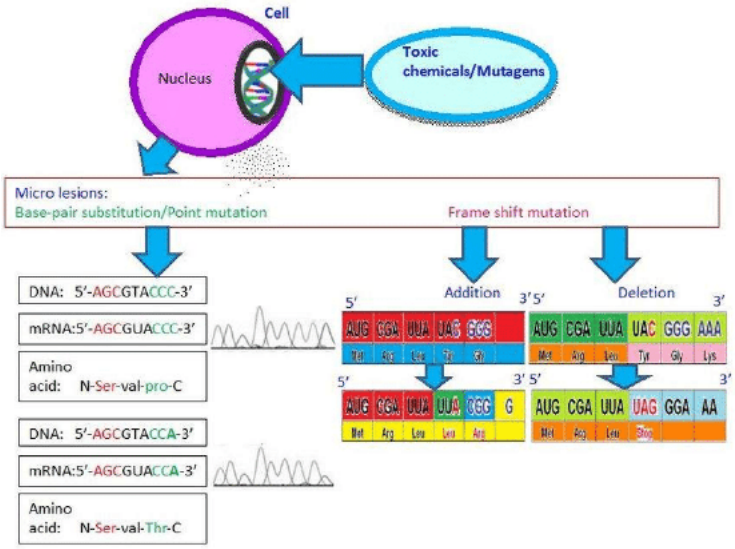
- 1: Know different types of genetic damages or mutations.
- 2: Know how different mutations resulted in the cells by mutagens.

Do you know how many mutations or damages are possible by mutagens in the cells?

Mutations can be:

- Microlesions (gene mutation)
- Macro lesions (Chromosomal mutation)

2.1: Microlesions (Gene Mutation)



Microlesions are the damages or mutations in DNA bases. These mutations are with invisible phenotypic changes. The types of microlesions are illustrated in the figure on the right.

Base-pair substitution mutation (qualitative change in nucleotide pairs)

- In this mutation, single base nucleotide is replaced by another nucleotide.

Frame shift (quantitative change in nucleotide pairs)

- In this mutation, addition or deletion of nucleotide in the DNA sequence resulted to shift or change the entire DNA or amino acid sequence.

Figure 8.4.3.2.1: Micro lesions mutation by mutagens.

2.2 Macro Lesions (Chromosomal Mutation)

Macro lesions are chromosomal mutations with mutagens and are with distinct morphological changes in the phenotype. These

morphological changes of chromosomes can be cytologically visible under microscope. Macro lesions are following types:

Numerical changes in chromosomes

- Polyploidy: Duplication of entire set of chromosome to triploid or tetraploid.
- Aneuploidy: Changes of single missing chromosome to **monosomy** or three copies of a single chromosome to **trisomy**.

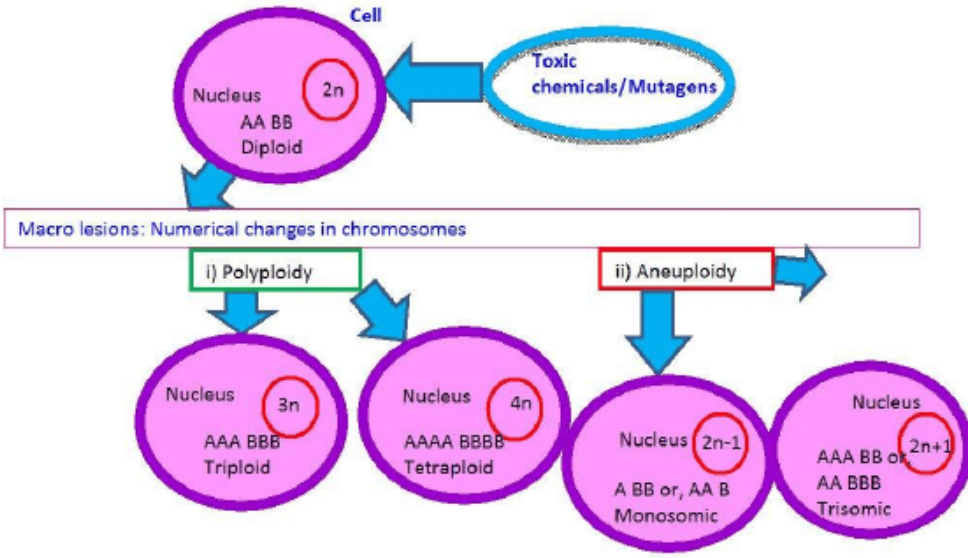


Figure 8.4.3.2.2: Macro lesions mutation by mutagens.

Structural changes in chromosomes

- Deletion: loss of chromosome segment
- Translocation: A segment of one chromosome becomes attached to a non homologous chromosome. It can be one way transfer as simple translocation and two way transfer as reciprocal translocation.
- Inversion: A change in the direction of material along a single chromosome.
- Duplication: Repetition of chromosome segment

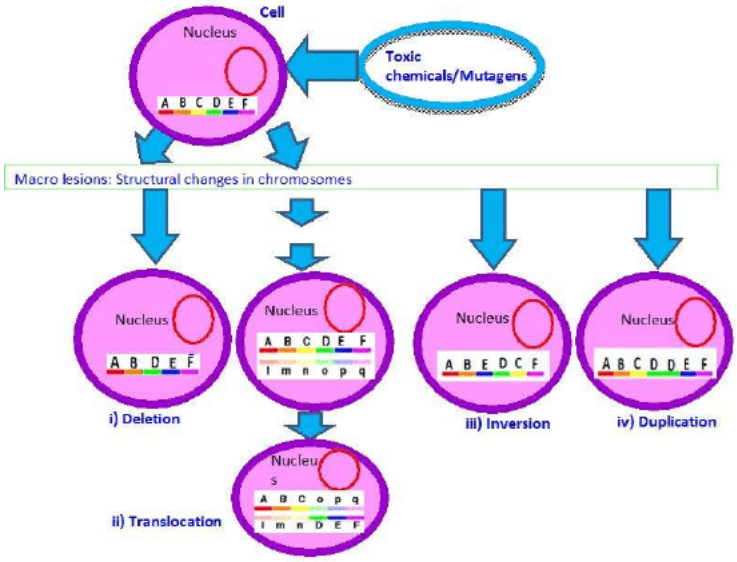


Figure 8.4.3.2.3: Structural changes in chromosomes resulted to macro lesions mutation.

Micronuclei changes

- Micronuclei (MN) are the damaged chromosome fragments or whole chromosomes that were not incorporated into the cell nucleus and stayed as the extra-nuclear bodies after the cell division.
- MN can be resulted by the defects of the cell repair machinery and by the accumulation of damaged DNA and chromosomal aberrations.genetic.html

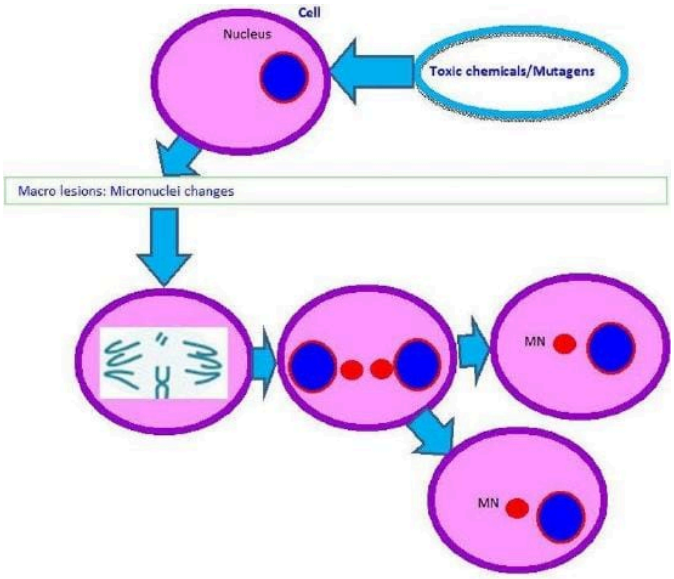


Figure 8.4.3.2.4: Micronuclei changes resulted to macro lesions mutation.

Topic 2: Key Points

In this section, we explored the following main points:

- 1: Different types of Microlesions (gene mutation) and Macro-lesions (Chromosomal mutation).
- 2: How toxic agents or mutagens modulate the two different microlesions namely base pair substitution and frame shift mutation.
- 3: Genotoxic mutagens involve also in different macro-lesions namely numerical or structural changes in chromosomes or changes of micronuclei in cell.

Knowledge Check

1. What is monosomy?

    Duplication of entire set of chromosome .

    Single missing chromosome from diploid set.

    Three copies of a single chromosome.

Answer available at the end of this section.

2. In base-pair substitution mutation, single base nucleotide is replaced by another nucleotide.

    True

    False

Answer available at the end of this section.

3. What are the structural changes in chromosomes caused by toxicants?

Deletion

Translocation

Inversion

Duplication

All of the above

Answer available at the end of this section.

4. Micronuclei (MN) changes are the damaged chromosome fragments or whole chromosomes that were not incorporated into the cell nucleus and stayed as the extra-nuclear bodies after the cell division.

True

False

Answer available at the end of this section.

5. Frame shift mutation, resulted to shift or change of entire DNA or amino acid sequence by:

addition of nucleotide in the DNA.

deletion of nucleotide in the DNA.

addition or deletion of nucleotide in the DNA

Answer available at the end of this section.

Answers

Knowledge Check
Single missing chromosome from diploid set.

Knowledge Check
True

Knowledge Check
All of the above

Knowledge Check
true

Knowledge Check
addition or deletion of nucleotide in the DNA

This page titled 8.4.3.2: Different Genetic Damages or Mutations is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.2: Different Genetic Damages or Mutations by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.3.3: Different Genetic-Toxicology Assays

Learning Objectives

- 1: Know different types of genetic-toxicology assays
- 2: Know how different genetic-toxicology assays are used in toxicology when cells are exposed to mutagens.

The goal of genetic toxicology assay is to determine whether any chemical or mutagen will do any adverse effect on genetic material or may cause different diseases including cancer. The assays can be performed using bacterial, yeast, or mammalian cells. One can early control and save vulnerable organisms from genotoxic chemicals by preforming genetic toxicology assay.

The following different types of genetic toxicology assays are used now a days:

- Bacterial Reverse Mutation Assay (Ames Assay)
- Genetic mutation assay
  - Allele-Specific PCR
  - Sanger Dideoxy Sequencing
- Chromosome aberration study
- Micronucleus assay

3.1: Bacterial Reverse Mutation Assay (Ames Assay)

This assay was discovered by Bruce Ames in 1970. This assay is widely used to test for gene mutation. The technique uses several strains of the bacterium Salmonella typhimurium which carry mutations in genes involved in histidine synthesis. These strains are auxotrophic mutants and they require histidine for growth and they cannot produce it. This assay examine the ability of the chemical or mutagen in creating mutations or a "prototrophic" state of strains, when the strains can grow on a histidine-free medium.

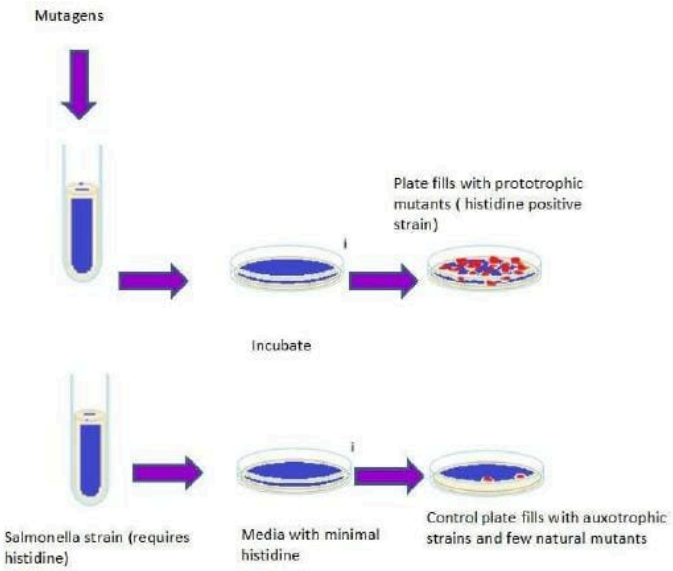


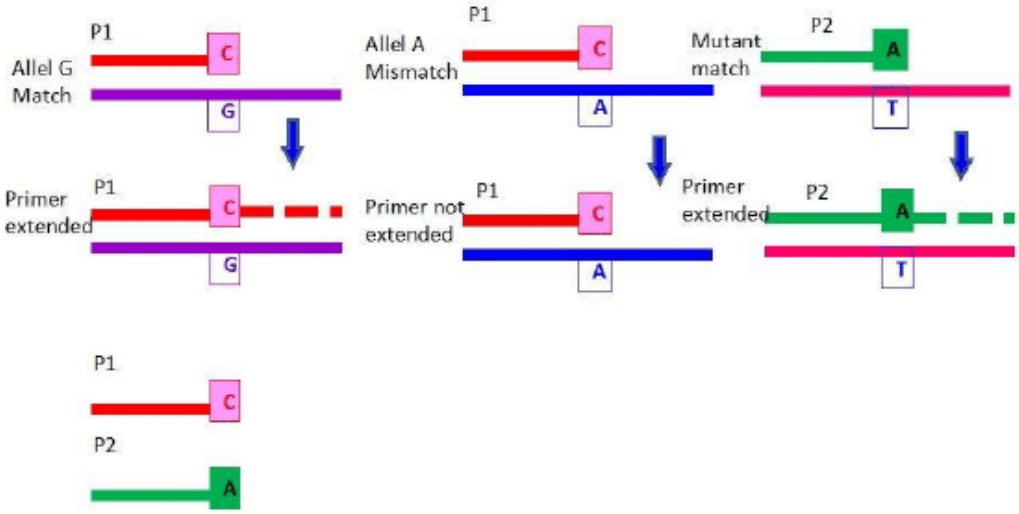
Figure 8.4.3.3.1: Ames test

3.2: Genetic Mutation Assay

Following are the different molecular assay to study nucleotide variants or alternation of genetic material caused by mutagens:

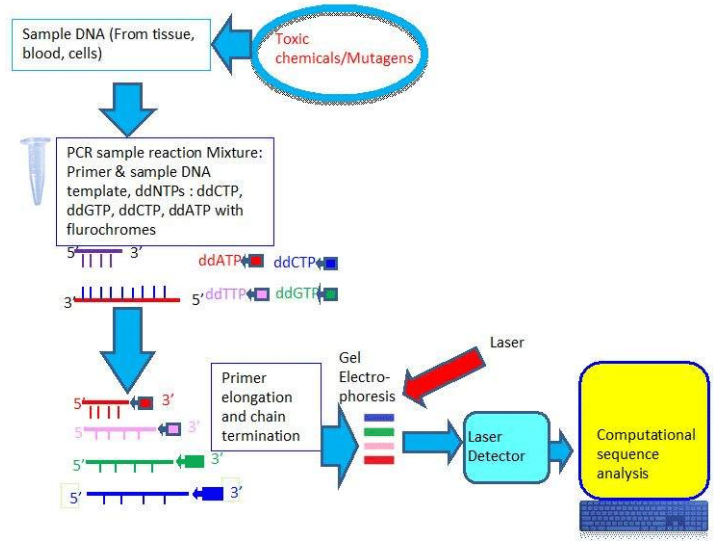
Allele-Specific PCR

Single nucleotide polymorphism (SNP) resulted from base substitution mutation can be analyzed by this method. In this real-time PCR, fluorescent reporter probes are added to the reaction mixture and one fluorescent reporter probe is selected for wild type and other fluorescent probe is used for mutant. The PCR primers with fluorescent probe will match or mismatch one of the alleles at 3' end of the primer. DNA polymerase extends the probes in a complementary fashion and releasing the reporter fluorescent molecules for detection. The PCR cycles with the reporter probes show the amplified signals and allow for precise measurement of one or both alleles of interest. Similarly, the 3' end of the mutant-specific primer is extended only in the presence of DNA with that mutation.



Sanger Dideoxy Sequencing

The goal of this method is to detect unknown mutations including single nucleotide variants (SNVs) and small duplications, insertions, deletions, and indels of interest caused by mutagens. In this method, sequencing primers hybridized to the PCR product and are extended using the four deoxynucleotides (dNTPs), a mixture of fluorescently labeled dideoxynucleotides (ddNTPs) and DNA polymerase. Four ddNTPs are marked with a different fluorescent dye. Random incorporation of the marked ddNTPs shows in termination of strands at each location along the sequence. The gel electrophoresis separates the strands by size. Fluorescence spectroscopy measured the terminating nucleotides.



3.3: Chromosome Aberration Study

Cytogenetic assays of mammalian cells are performed to detect different types of structural and numerical chromosomal aberrations caused by genotoxic chemicals. The clastogenic or aneugenic effects from the genotoxic chemicals will result in an increase in frequency of structural (premature centric separation, chromosome breaks, dicentric chromosomes, ring) complex rearrangements (Figure 4) or numerical aberrations of the genetic material in mammalian cells.

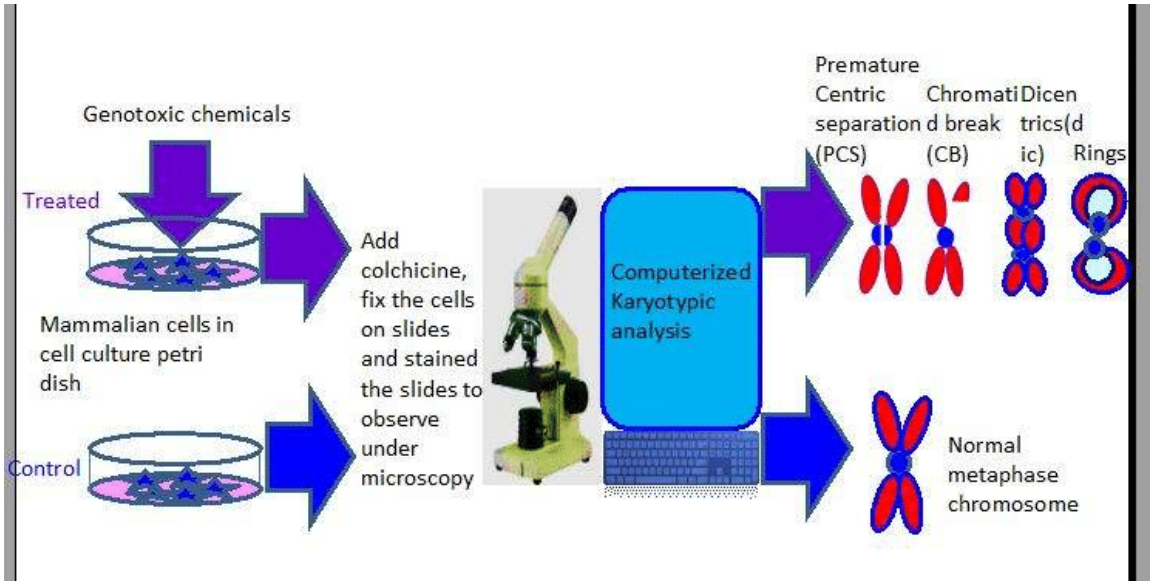


Figure 8.4.3.3.2: Chromosome aberration study

3.4: Micronucleus Assay

Micronucleus assay is used as a tool to evaluate genetic damage caused by genotoxic chemicals. The number of micronuclei (Figure 5) generated directly relates to the amount of DNA damage in the cells.

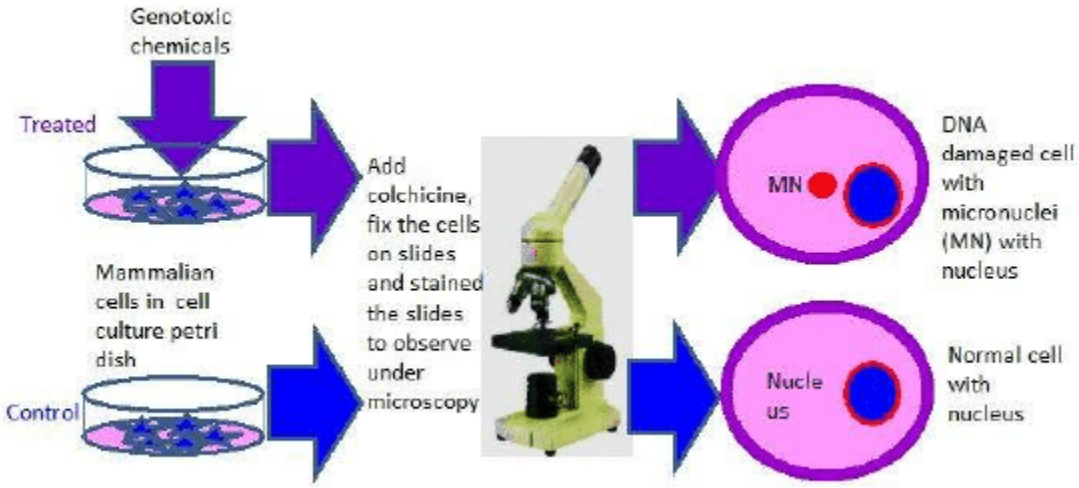


Figure 8.4.3.3.3: Micronucleus assay

Topic 3: Key Points

In this section, we explored the following main points:

- 1: Definition of genetic toxicology assay
- 2: Different types of genetic-toxicology assays.
- 3: How genetic mutation assays are performed by Allele-Specific PCR and Sanger Dideoxy Sequencing techniques.
- 4: What are the different types of chromosomal aberrations observed under microscope by chromosome aberration study?
- 5: Changes of micronuclei observed under microscope by Micronucleus assay.

Knowledge Check

1. Cytogenetic assays of mammalian cells are performed to detect different types of structural and numerical chromosomal aberrations caused by a genotoxic chemicals. The structural chromosomal aberrations are:

premature centric separation

ring

chromosome breaks

dicentric chromosomes

All of the above

Answer available at the end of this section.

2. In Allele-Specific PCR, fluorescent reporter probes are added to the reaction mixture and one fluorescent reporter probe is selected for wild type and other fluorescent probe is used for mutant.

True

False

Answer available at the end of this section.

3. Which instrument is used to measure the terminating nucleotides in Sanger Dideoxy Sequencing?

Fluorescence spectroscopy

Spectrophotometer

Fluorescence microscopy

None of the above

Answer available at the end of this section.

4. The Ames technique uses several strains of the bacterium Salmonella typhimurium which carry mutations in genes involved in:

Arginine Synthesis.

Histidine synthesis.

Lysine Synthesis.

None of the above.

Answer available at the end of this section.

Answers

Knowledge Check

All of the above

Knowledge Check

True

Knowledge Check

Fluorescence spectroscopy

Knowledge Check

Histidine synthesis.

This page titled 8.4.3.3: Different Genetic-Toxicology Assays is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.3: Different Genetic-Toxicology Assays by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.3.4: Different Cytotoxicity Assays

Learning Objectives

- 1: Know different types of cytotoxicity assays.
- 2: Know how different cytotoxicity assays are used when cells are exposed to toxicants or mutagens.

The goal of cytotoxicity assay is to determine whether any chemicals or drugs will do any toxic effect or load on milieu or genetic material of the cells caused for lethality of the cells or caused for different diseases.

The following are the different types of cytotoxicity assays:

- DNA fragmentation/ladder assay
- Comet assay
- Necrosis assay
- Enzyme assay
- Proteomics assay
- Expression array assay

4.1: DNA Fragmentation/Ladder Assay

DNA fragmentation or ladder assay are used to know the fragmented DNA of the cells caused by chemicals or drugs. Fragmented DNA can be separated by agarose gel electrophoresis and can be visualized as “ladder” by ethidium bromide staining. The evaluation of cytotoxicity through cell death is an acceptable common assessment.

Ladder assays are performed for the following reasons:

1. to simply characterize the toxicity of the chemicals or drugs in cells, or
2. to determine the maximum doses of the test chemicals or drugs that can be used for cells without causing too much cell death.

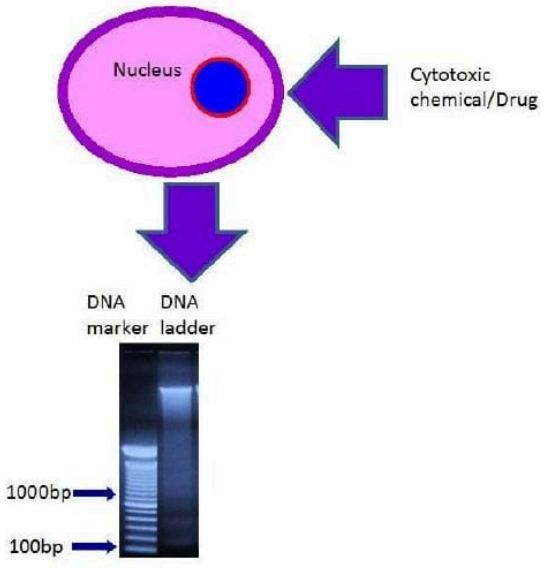


Figure 8.4.3.4.1: DNA ladder. Nucleotide base pair mentioned as,“bp”.

4.2: Comet Assay

The Comet Assay (single cell gel electrophoresis /SCGE) is used to detect DNA damage by using a micro gel electrophoresis. The image of the damaged DNA shows a comet with head and tail. The analysis of image for comet assay is calculated for the “tail length” of the comet which is the measurement from the point of highest intensity within the comet head as well as the “tail moment” which is the product of the tail length and the fraction of total DNA present within the tail.

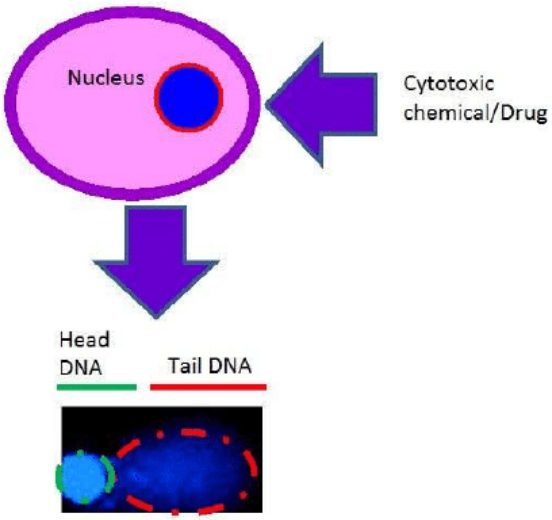


Figure 8.4.3.4.2: Comet assay showed amount of damaged DNA by comet tail moment.

4.3: Necrosis Assay

The necrosis assay is performed by flow cytometry analysis with staining of Annexin V and propidium iodide (PI) in the cells. The cells are considered in the stage of necrosis if the cells lose membrane integrity and die promptly due to cell lysis when exposed to chemicals, drugs, toxins or foreign antigens. In necrosis, cells show swelling, loss of membrane integrity and disruption of metabolism. Cells with necrosis do not go to stage of apoptosis, apoptotic cell may undergo secondary necrosis. These necrotic cells will shut down metabolism, lose membrane integrity, lyse and formed cell injury autolysis.

The flow cytometric analysis for necrosis assay showed that the cells stained positive for both FITC Annexin V and PI are in the end stage of apoptosis and are undergoing to the stage of necrosis as dead cells stained PI positive. Cells that stain negative for both FITC Annexin V and PI are alive and not undergoing apoptosis or necrosis.

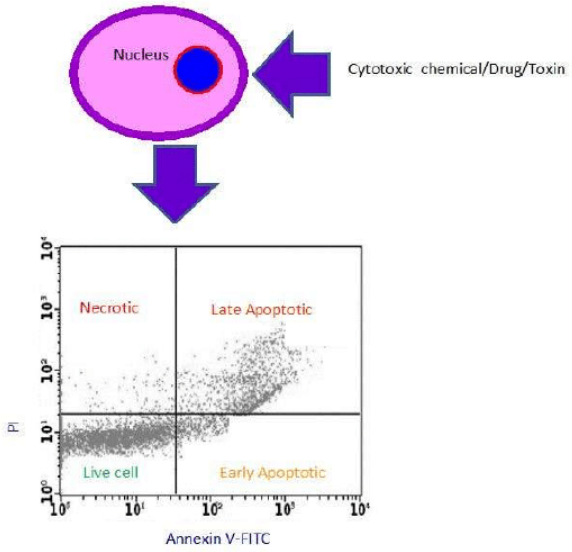


Figure 8.4.3.4.3: Flow cytometry analysis of cells showing necrosis stage.

4.4: Enzyme Assay

The enzyme assay is used to monitor passaging of lactate dehydrogenase (LDH), due to loss of cell membrane integrity when cells are exposed to cytotoxic compounds. LDH reduces NAD to NADH which generates a color change by interaction with a specific probe (Figure 4a). In other enzyme assay, Adenosine triphosphate (ATP)-based assay combined with bioluminescent assay are used to measure cytotoxicity of the cells in which ATP is the reagent for the luciferase reaction.

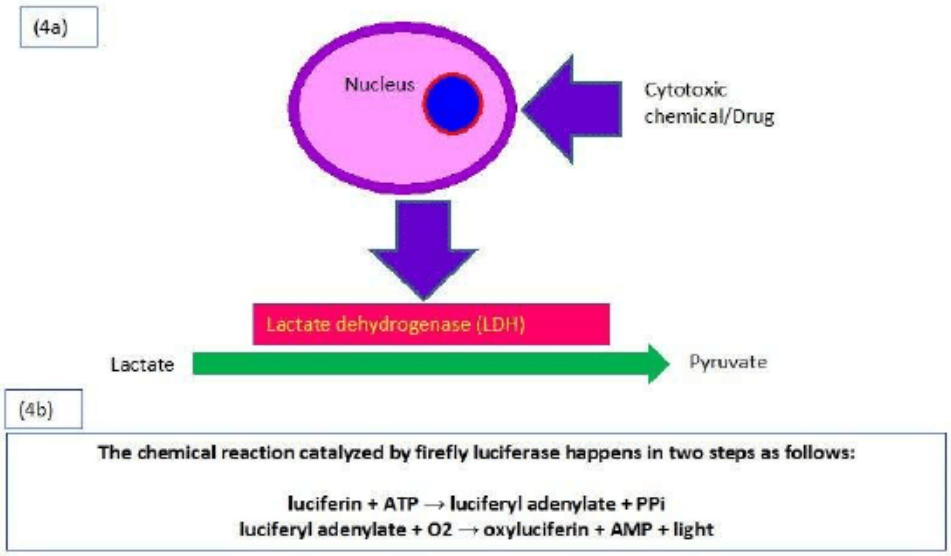


Figure 8.4.3.4.4: Schematic presentation of enzyme assay. (a) LDH assay (b) ATP based assay

4.5: Proteomics Assay

The proteomic assay is performed to know the mechanism of cellular toxicity by measuring expression of a specific protein which may consider as a biomarker for particular toxic mechanism or cellular toxicity signaling pathway. Immunofluorescence, immunoprecipitation and immunoblot assay are mainly used to know the effect of toxicants in cellular toxicity signaling pathway or mechanism.

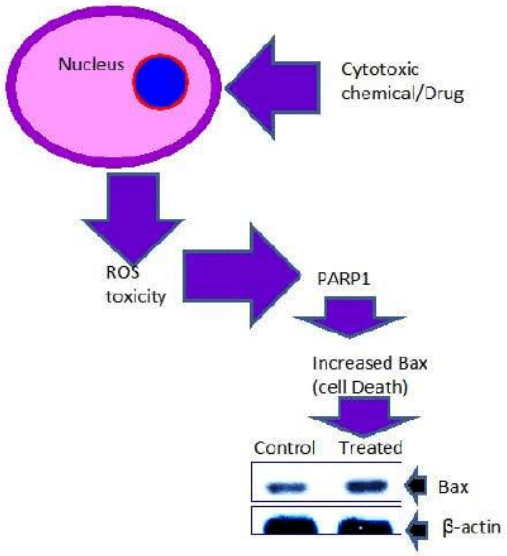


Figure 8.4.3.4.5: Schematic presentation of proteomic assay.

4.6: Expression Array Assay

The expression array is the chip based microarray of more gene expressions (finger print of genes) by the effect of cellular toxicants. This is a rapid and sensitive detection method which allows to detect all toxicological end points at wide range of molecular level changes in the cell at single assay. The microarray process can be divided into two main parts. First is the printing of known gene sequences onto glass slides or other solid support followed by hybridization of fluorescently labeled cDNA (containing the unknown sequences to be interrogated) to the known genes immobilized on the glass slide. After hybridization, arrays are scanned using a fluorescent microarray scanner. Analyzing the relative fluorescent intensity of different genes provides a measure of the differences in gene expression.

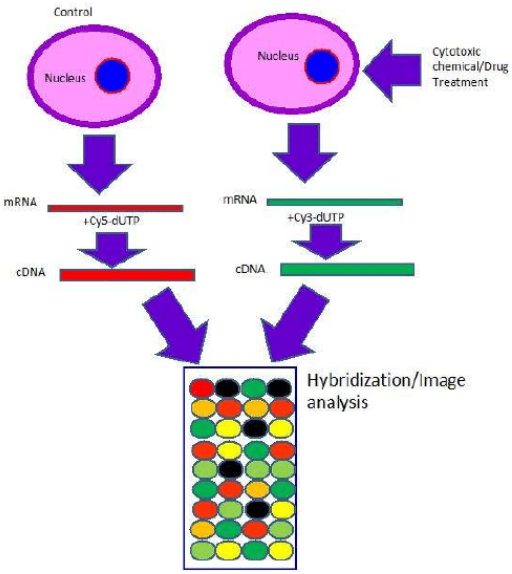


Figure 8.4.3.4.6: Schematic representation of gene expression array showed that Cy5(red) and Cy3 (green) labeled cDNA hybridized to a DNA microarray. Yellow spots indicate the genes are expressed in both samples. The intensity and different types of color at each spot indicate the level and presence of genes in samples. Black spots show low level of expression or do not show any expression of genes.

Topic 4: Key Points

In this section, we explored the following main points:

- 1: Different types of cytotoxicity assays.
- 2: How different cytotoxicity assay namely DNA fragmentation/ladder assay, Comet assay, Necrosis assay, Enzyme assay, Proteomics assay, and Expression array assay are used when cells are exposed to cytotoxic agents.

Knowledge Check

1. The cells are considered in the stage of necrosis, if the cells lose membrane integrity and die promptly due to cell lysis when exposed to chemicals, drugs, toxins or foreign antigens.

True

False

Answer available at the end of this section.

2. The Comet Assay is used to detect DNA damage by using a micro gel electrophoresis:

True

False

Answer available at the end of this section.

3. The expression array is the chip based microarray of more gene expressions (finger print of genes) by the effect of cellular toxicants.

True

False

Answer available at the end of this section.

4. Ladder assay are performed for the following reasons: 1) to simply characterize the toxicity of the chemicals or drugs in cells, or 2) to determine the maximum doses of the test chemicals or drugs that can be used for cells without causing too much cell death.

True

False

Answer available at the end of this section.

5. What are the proteomic assays to know the effect of toxicants in cellular toxicity signaling pathway or mechanism?

Immunofluorescence assay

ring

Immunoprecipitation assay

Immunoblot assay

All of the above

Answer available at the end of this section.

Answers

Knowledge Check

True

Knowledge Check

True

Knowledge Check

True

Knowledge Check

True

Knowledge Check

All of the above

This page titled 8.4.3.4: Different Cytotoxicity Assays is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.4: Different Cytotoxicity Assays by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.3.5: Epigenetics Assay

Learning Objectives

- 1: Know different types of epigenetic assays.
- 2: Know how different epigenetic assays are used when cells are exposed to toxicants or mutagens.

5.1: DNA Methylation Assay

DNA methylation assays are important to know the epigenetic modification which is a heritable, enzyme-induced modification without alteration of the nucleotide base pairs. The transfer of a methyl-group to the 5-carbon on the cytosine in a CpG dinucleotide happens in the DNA methylation by DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B). The high level of promoter CpG island methylation results in gene silencing. The methylated DNA immunoprecipitation (MeDIP)-chip technique is used for DNA methylation assay.

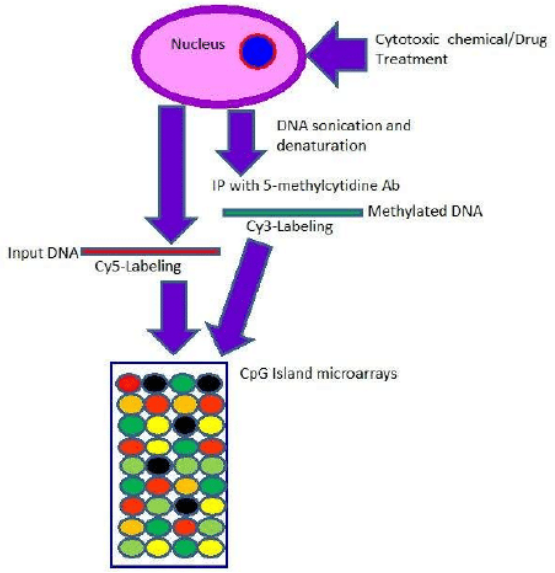


Figure 8.4.3.5.1: Schematic diagram of MeDIP protocol for DNA methylation assay.

In brief, the MeDIP-chip procedure is mentioned as follows. The genomic DNA is sheared to low molecular weight fragments (approximately 400 bp) by sonication. Then, the methylated DNAs are immunoprecipitated with the anti-methyl-cytosine antibody, and are amplified with PCR, if source material is less. Input and methylated DNA are labeled with fluorescent dyes Cy3 (green) and Cy5 (red), pooled, denatured, and are hybridized to a microarray slide containing all the annotated human CpG islands or other whole genome or promoter microarray designs. Then the slide is scanned using a scanner and each image is analyzed with the image analysis software (Figure 1).

5.2: Histone Modification Assay

Histone modification assays are useful to find the modification of histone proteins (e.g. lysine acetylation, lysine and arginine methylation, serine and threonine phosphorylation, and lysine ubiquitination and sumoylation) which have important roles in epigenetic inheritance. The chromatin immunoprecipitation (ChIP) assay followed by hybridization to microarrays (ChIP-chip) (left) or by high-throughput sequencing (ChIP-seq) (right) are both powerful techniques to find histone modification.

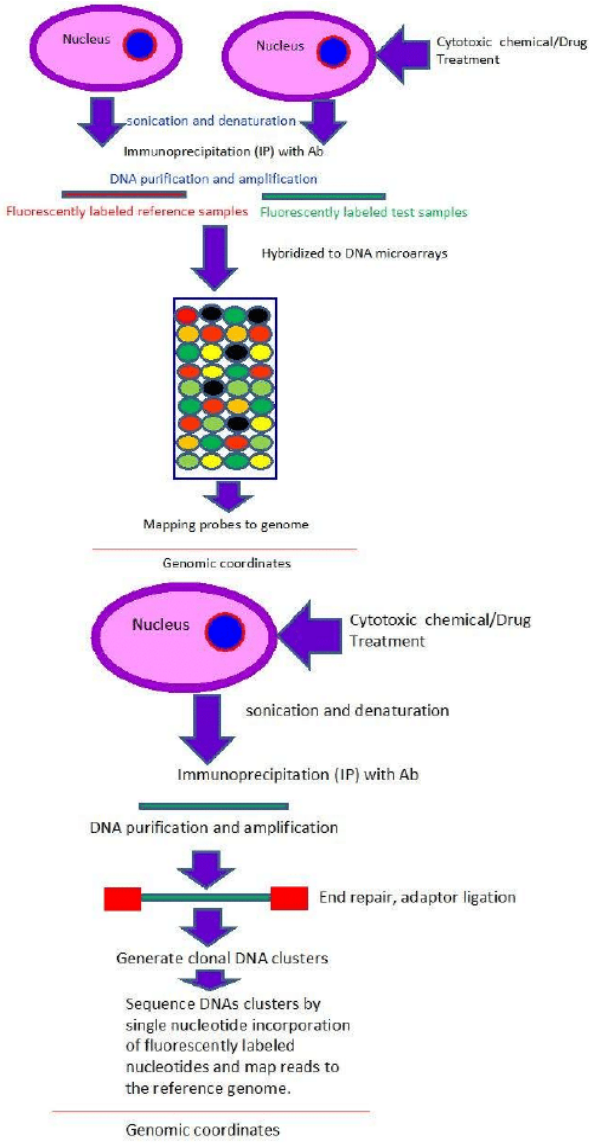


Figure 8.4.3.5.2: Schematic presentation of ChIP – chip microarray and sep of histone modification assay

5.3: MicroRNAs Assay

MicroRNAs assays are used to know the non-coding RNAs (17-25 nucleotides) which target messenger RNAs (mRNAs) and decayed the mRNAs or downregulated at the level of translation into protein. Almost, 60% of human protein coding genes are controlled by miRNAs and these miRNAs are epigenetically regulated. About 50% of miRNA genes are related with CpG islands, which may be repressed by epigenetic methylation. Other miRNAs are epigenetically controlled by either histone modifications or by DNA methylation. The expression of microRNAs are quantified by RT-PCR followed by quantitative PCR (qPCR). Then, miRNAs are hybridized to microarrays, slides or chips with probes to hundreds or thousands of miRNA targets. The microRNAs can be both invented and profiled by sequencing methods (microRNA sequencing).

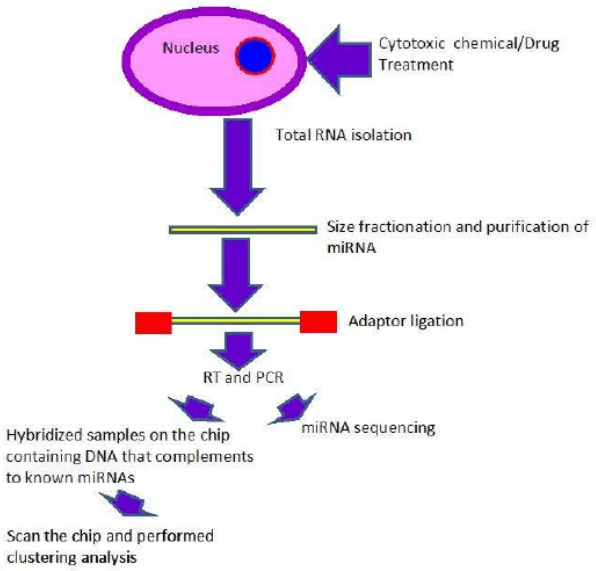


Figure 8.4.3.5.3: Schematic presentation of miRNA assay in epigenetic toxicology

Topic 5: Key Points

In this section, we explored the following main points:

- 1: Different types of Epigenetic assays.
- 2: How different Epigenetic assays namely DNA methylation assay, histone modification assay and MicroRNAs assay are used when cells are exposed to toxic chemicals or agents.

Knowledge Check

1. DNA methylation assays are important to know the non-epigenetic modification.

True

False

Answer available at the end of this section.

2. MicroRNAs assays are used to know the non-coding RNAs. These non-coding RNAs are :

17 to 25 nucleotides.

50 to 100 nucleotides.

200 to 400 nucleotides.

None of the above.

Answer available at the end of this section.

3. Histone modification assays are useful to find the modification of histone proteins which have important roles in epigenetic inheritance:

True

False

> *Answer available at the end of this section.*

Answers

Knowledge Check

false

Knowledge Check

17 to 25 nucleotides.

Knowledge Check

true

This page titled 8.4.3.5: Epigenetics Assay is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.5: Epigenetics Assay by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.3.6: Section 3 Final Evaluation

1. _____ is a method used in analyzing base substitution mutation resulting from single nucleotide polymorphism.

Micronucleus assay

Chromosome aberration study

Allele-Specific PCR

Chromatography

Answer

Allele-Specific PCR

2. Proteomic assay analyzes the effect of toxicants in cellular toxicity signaling pathways or mechanisms through:

Immunofluorescence

Immunoblot

Immunoprecipitation

All of the above

Answer

All of the above

3. In Ladder Assay, fragmented DNA can be separated by agarose gel electrophoresis and can be visualized as “ladder” by _____ staining.

Ethidium bromide

Eosin

Gram

Wright’s

Answer

Ethidium bromide

4. Trisomy is a form of aneuploidy interpreted as:

Single missing chromosome from diploid set.

Three copies of a single chromosome from a diploid set

Three copies of a single chromosome from a triploid set

Single missing chromosome from a triploid set

Answer

Three copies of a single chromosome from a diploid set

5. Gene mutations in which a single base nucleotide is replaced by another nucleotide are known as:

Frame shift

Quantitative change in nucleotide

Qualitative change in nucleotide

Left shift

Answer

Qualitative change in nucleotide

6. Some structural chromosomal aberrations caused by genotoxic chemicals in cytogenetic assays of mammalian cells include the following EXCEPT:

Dicentric Chromosomes

Ring Chromosomes

Spiral Chromosomes

Chromosome breaks

Answer

Spiral Chromosomes

7. Structural changes in chromosomes include the following EXCEPT:

Aneuploidy

Inversion

Translocation

Deletion

Answer

Aneuploidy

8. Addition or deletion of nucleotides in the DNA sequence results in the change of the entire DNA or amino acid sequence. This process is known as:

Frame shift

Base-pair substitution mutation

Qualitative change in nucleotide

Right shift

Answer

Frame shift

9. In Allele-Specific PCR, fluorescent reporter probes are added to the reaction mixture, one fluorescent reporter probe is selected for the wild type and the other fluorescent probe is used for the mutant.

True

False

Answer

true

10. The analysis of image to determine DNA damage in comet assay is calculated for the _____ and _____.

“tail length” and “head length”

“head length” and “head moment”

“head length” and “tail moment”

“tail length” and “tail moment”

Answer

“tail length” and “tail moment”

11. Cells which stain negative for both fluorescein isothiocyanate Annexin V and propidium iodide in flow cytometric analysis for necrosis assay are:

Dead and undergoing apoptosis and necrosis

Alive and not undergoing apoptosis or necrosis

Dead and undergoing necrosis only

Alive and undergoing apoptosis only

Answer

Alive and not undergoing apoptosis or necrosis

12. Histone Modification Assay uses Chromatin immunoprecipitation assay (ChIP) followed by:

Hybridization to microarrays (ChIP-chip)

Immunofluorescence

Fluorescence spectroscopy

Fluorescence microscope

Answer

Hybridization to microarrays (ChIP-chip)

This page titled 8.4.3.6: Section 3 Final Evaluation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 3 Final Evaluation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4: APPLIED SYSTEMS TOXICOLOGY

Systems toxicology is a branch of science that utilizes data from different branches of toxicology and integrates them to provide a holistic approach for safety assessment. In this e-module you will learn about the concept of systems toxicology, dose level in toxicology, and different approaches to traditional and new toxicology.

8.4.4.1: Systems Toxicology

8.4.4.2: Dose Level and Applied Toxicology

8.4.4.3: Tools and Technologies in Systems Toxicology

8.4.4.4: Other Approaches for Predictive Toxicity Modeling

8.4.4.5: Technologies Used In Systems Biology/Toxicology

8.4.4.6: Takeaways Summary

8.4.4.7: Section 4 Final Evaluation

This page titled 8.4.4: Applied Systems Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

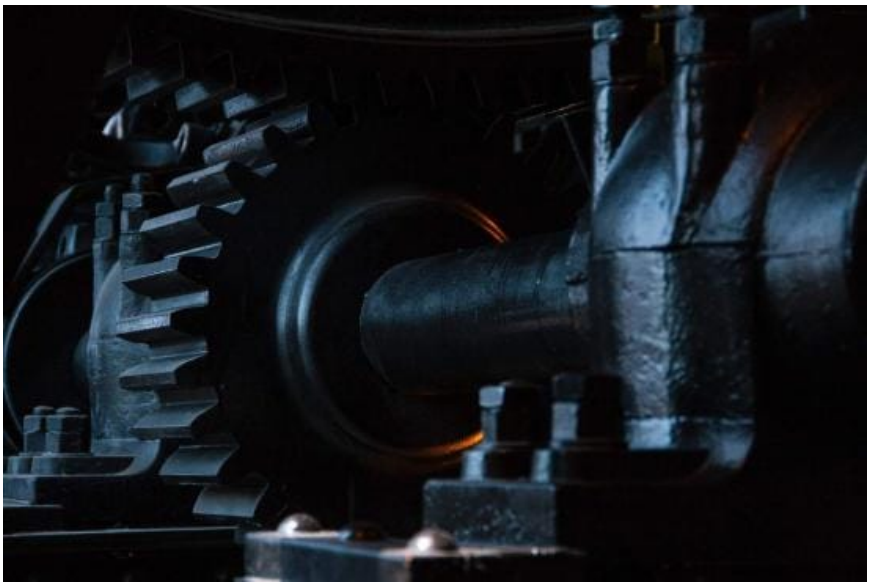
- 4: Applied Systems Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4.1: Systems Toxicology

Learning Objectives

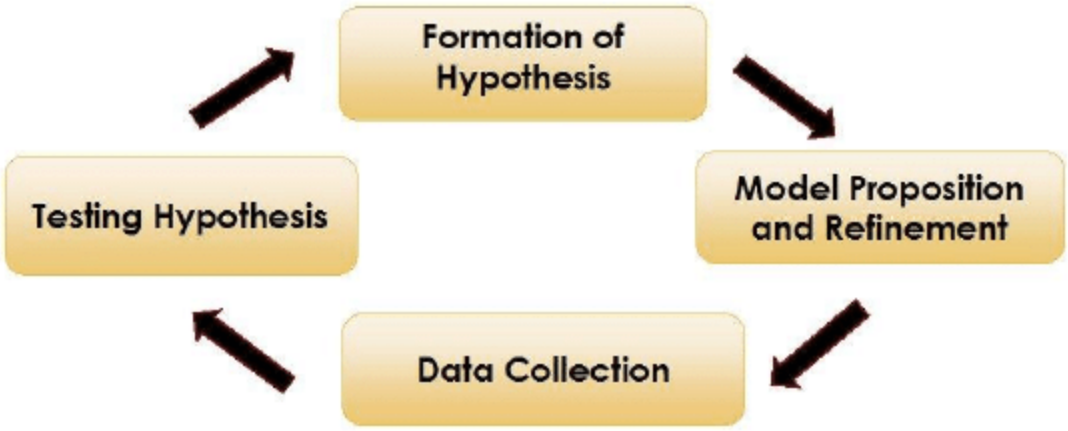
- 1: Understand the concept of systems toxicology.
- 2: Understand the approaches of traditional toxicology approaches vs. the new toxicity testing paradigm.
- 3: Recognize the driving force behind the growth of this field.
- 4: Applications of this field.

The word “systems” originates from the Latin word “systema” which means a complete concept that has several parts. Similarly systems toxicology is a branch of science that utilizes data from different branches of toxicology and integrates them to provide a holistic approach for safety assessment.



Toxicology is the science of understanding the adverse effects of xenobiotics (drugs, chemicals, etc.) on biological systems. Biological systems are extremely complex. Due to the vast number of toxicology research approaches over the years, and a lot of data have been generated in different systems- in vivo, in vitro, in silico (especially due to “omics” approaches). However, there is currently a lack of interpreting/utilizing that data for efficient safety assessment of xenobiotics.

Systems toxicology aims to fill this gap and utilize these data from different systems and integrate them into meaningful assessment for safety. It relies heavily on mathematical and computational models to link the data from various systems. So, in order to have fully validated systems toxicology approaches it is important to have “real” (in life) data from animal models to validate the hypothesis.



The Driving Force

The main driving force behind the development of systems toxicology approaches is the fact that the whole “safety assessment” process is a very lengthy, time consuming and expensive process in case of chemicals as well as the pharmaceutical industry. In order to make this process more efficient, it is important for early pharmaceutical/chemical (especially pharmaceuticals)candidate selection/screening. Screening thousands of compounds is a lengthy process and current high-throughput screening approaches together with large volume data analysis techniques have helped in more efficient selection of target molecules.



Schematic for the drug development process

Topic 1: Key Points
In this section, we explored the following main points:

- 1: The concept of systems toxicology and the different approaches and the driving force behind the development of this field.

Knowledge Check

1. Systems toxicology is usually applicable in the ...

Early discovery phases of drug development

During marketing of the drug

During regulatory safety testing phases

None of the above

Answer available at the end of this section.

2. Systems toxicology approach involves...

Traditional animal experiments

Alternative in vitro methods

Computational and mathematical models

All of the above

Answer available at the end of this section.

Answers

Knowledge Check

Early discovery phases of drug development

Knowledge Check

All of the above

This page titled 8.4.4.1: Systems Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 4.1: Systems Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4.2: Dose Level and Applied Toxicology

Learning Objectives

After completing this lesson, you will be able to:

- 1: Assess toxicity in response to dose levels.

As students of toxicology it is critically important to understand that everything including water and oxygen has the potential to act as poison. It is only the dose that determines the toxic/beneficial effect.

Dose level and Applied Toxicology

While traditional toxicology approaches in many academic laboratories and also in industry utilized very high dose levels of xenobiotics in order to study various mechanisms of toxicity, it has been identified recently that this is not a very ideal approach especially for industrial applications.

Step 1 Title

- Using very high dose levels may saturate the physiological processes in a living system and hence may cause “forced toxicity” which is not very relevant in terms of real life exposures (could be valid in case of overdose/accidental overexposure).

In industry, a lot of preliminary research is conducted in order to determine dose levels for toxicology experiments. Typically, in case of pharmaceutical compounds a lot of pharmacokinetic modeling and simulations is performed to come up with exposure levels that are multiples of the real drug exposure level in humans. In the chemical industry relevant exposure levels that human beings may be exposed to based on the use of the chemical is determined.

Dose Level Selection

- 1: Traditional toxicity testing involved using a large number of animals and using very high dose levels. Study designs such as LD 50 (Lethal Dose for 50% animals in the study) are no longer used. It is not feasible from a scientific or animal welfare point of view
- 2: Nowadays, dose level selection often involves mathematical simulation and modeling utilizing various forms of in vitro data that are analyzed with the help of medium and high-throughput modeling tools. This is a systems toxicology approach where data from different platforms is utilized to come up with relevant dose levels that are scientifically justifiable and also to refine and minimize animal experiments.

Schematic Approach: Systems Toxicology

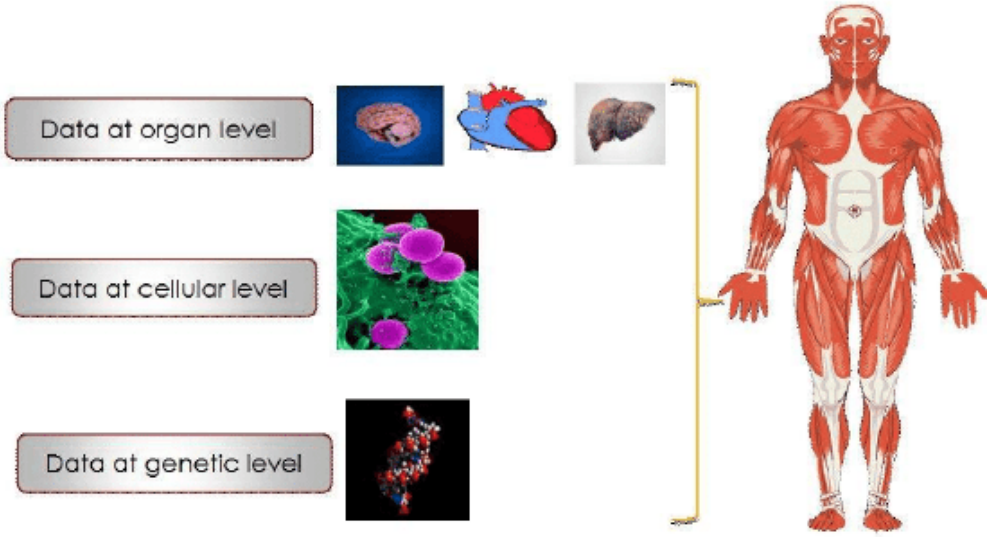


Figure 8.4.4.2.1: Schematic Approach: Systems Toxicology. Credit: pixabay.com

Key Points

In this section, we explored the following main points:

- 1: The understanding of how toxicity is driven strongly by the dose response relationship.

Knowledge Check

1. Utilizing very high dose levels in animal models are not useful because

It causes forced toxicity due to saturation of the absorption and elimination processes

It is too toxic and no meaningful information is available from such data

It is necessary for planning acute studies

It causes a lot of wastage of test article

Answer available at the end of this section.

2. Traditional LD 50 studies evaluated...

Adverse effects in 50 animals per dose group

Death in 50 % of the population

Effects in 50% of the animal population

Doses to be used in subsequent chronic studies

Answer available at the end of this section.

Answers

Knowledge Check

It causes forced toxicity due to saturation of the absorption and elimination processes

Knowledge Check

Death in 50% of the population

This page titled 8.4.4.2: Dose Level and Applied Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 4.2: Dose Level and Applied Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4.3: Tools and Technologies in Systems Toxicology

Learning Objectives

- 1: Understand tools and technologies used in systems toxicology, the concept of adverse effects and Adverse Outcome Pathways.

Tools and Technologies



1. Tools at the molecular level: Omics technologies such as genomics, protemics, metabolimics.

2. Tools at the cellular level include highly refined high-throughput *in vitro* model systems.

3. Modeling softwares that can utilize in vitro data and mathematically translate that into relevant in vivo information utilizing pharmacokinetic (PK) or physiologically based pharmacokinetic models (PBPK). This is known as in vitro to in vivo extrapolation (IVIVE).

4. Modeling softwares that can perform sophisticated species scaling (prediction of human parameters from nonclinical species such as rat, dog, monkey, etc.)

5. Risk assessment and exposure modeling tools that enable calculation of hazard and risk in populations on a whole and specific subpopulations.

Toxicity Testing in Chemicals vs. Pharmaceuticals

Regulatory based toxicity testing is different for pharmaceutical based compounds versus chemical and agrochemical compounds. Non clinical safety assessment for phamaceutical products is governed by the different stages of drug development while the toxicological data requirement for chemicals/agrochemicals is based on the amount (tonnage) produced. While most pharmaceutical compounds undergo extensive animal toxicity testing, there are thousands of chemicals in commerce today that have undergone very limited or non toxicity testing. In order to address this several governmental mandates are being put into action.

Government Initiatives & Mandates for 21st Century Toxicity Testing

- In Europe, the Registration Evaluation Authorization and Restriction of Chemicals (REACH) was initially implemented in 2007.
- This substantially altered the safety testing performed on new as well as existing chemicals.
- In The United States too several initiatives to increase safety testing on more and more chemicals are underway which would increase the cost of safety assessments astronomically.
- The regulations under REACH have been directly estimated to cost the industry more than 4.2 billion dollars (Brown, 2003).



Figure 8.4.4.3.1: Brown, V. J. (2003). REACHing for chemical safety. Environ. Health Perspect. 111, A766–A769.

21st Century Toxicity Testing

Traditional toxicity testing involves the use of a lot of animals and is an extremely expensive and time consuming process. In order to address the large number of untested chemicals the US Environmental Protection Agency (EPA) initiated the **ToxCast program.** The ToxCast Program is a high-throughput screening program that would enable the prioritization of chemicals so that resources can be channelized towards those chemicals that possess the greatest risk to human safety.

THE TOXCAST PROGRAM



The ToxCast program developed and utilized automated in vitro assays (to test effects of chemicals on various biological processes using living cells, isolated proteins etc.). The assay designs included endpoints such as cytotoxicity, enzyme activity, endocrine endpoints, gene expression etc. A total of approximately 600 endpoints were evaluated.

The biggest question to answer was how relevant was this data to human safety and how to utilize this *in vitro* information to predict human safety.

Figure 8.4.4.3.2: National Institute of Environmental Health Sciences – NIH, 50th Anniversary

The Overall Perspective

In vitro assays may be relevant in the light of the fact that overall disesase/toxicology processes are actually mediated by molecular and cellular perturbations. However, the overall picture at the whole organism level includes several other complex

factors such as pharmacokinectics of the compound, metabolism, clearance etc. This gap could be filled by utilizing computational modeling tools that would utilize the in vitro data and integrate them into human physiology with the help of pharmacokinetic (PK) or physiologically based pharmacokinetic models (PBPK).

Understanding the concept of effects Vs. adverse Effects

In order to conduct safety assessments it is important to understand the concept of “adverse effects” versus simply “effects”. It is also important to understand the significance of biological relevance of isolated *in vitro*, molecular assays as it pertains to the whole organism.

Adverse Outcome Pathways (AOPs) have been developed to try and link the causal molecular initiating event to a host of intermediate processes at the cellular level that finally leads to adverse outcome (AO) in the whole organism that can be used for safety assessment purposes.

Adverse Outcome Pathways

The AOP programme was launched by the Organization for Economic Co-operation and Development (OECD) in 2012. The objective was to link the main molecular initiating event with the phenotypic/functional toxicity/adverse effect at the organism level.



Figure 8.4.4.3.3: Adverse Outcome Pathways

Using Systems Toxicology Approach For Toxicity Predictions

The molecular initiating events could be used in selecting in vitro assays that could have possible potential for predicting toxicity at the whole organism level.

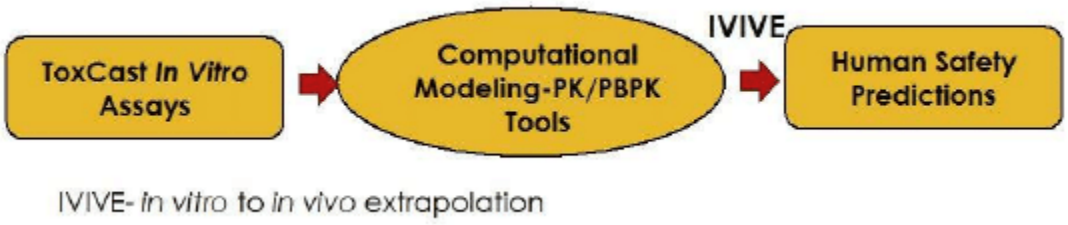


Figure 8.4.4.3.4: IVIVE- in vitro to in vivo extrapolation

Topic 3: Key Points

In this section, we explored the following main points:

- 1: Tools and Technologies, used in the field of systems toxicology.
- 2: Concept of Adverse effects and AOPs .

Knowledge Check

1. The ToxCast program utilizes which kind of assays?

In vivo

In silico

In vitro

None of the above

Answer available at the end of this section.

2. Adverse Outcome Pathway (AOP) Programme was launched by...

National Toxicology Program (NTP)

National Institute of Health (NIH)

Environmental protection Agency (EPA)

Organization for Economic Co-Operation and Development

Answer available at the end of this section.

Answers

Knowledge Check

in vitro

Knowledge Check

Organization for Economic Co-Operation and Development

This page titled 8.4.4.3: Tools and Technologies in Systems Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 4.3: Tools and Technologies in Systems Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4.4: Other Approaches for Predictive Toxicity Modeling

Learning Objectives

- 1: Understand what Quantitative Structure Activity Relationships are and how they are used in the field of systems toxicology.

Quantitative Structure Activity Relationships (QSAR)

QSAR models are classification models that are used to link particular structures in some molecules to causative/adverse effects at cellular/organism level. This is based on the fact that compounds that are structurally similar would have similar mechanisms of mediating toxicity. For example compounds belonging to the structural class of triazoles have been reported to cause developmental malformations such as cleft palate in rats.

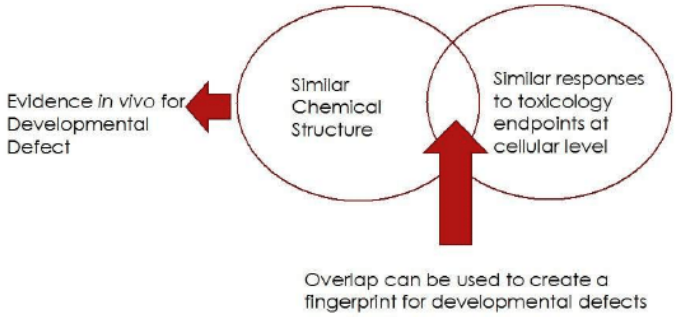


Figure 8.4.4.4.1: QSAR Approaches for Predictive Toxicity Testing

Mechanisms at the cellular level include events such as positive response for the TGFb1 signaling pathway. This information can be used to build a molecular fingerprint where compound having similar chemical structures as triazoles and having a positive response to TGFb1 signaling pathway could be flagged for potential developmental toxicity (teratogens).

QSAR Approaches for Predictive Toxicity Testing



This method however requires a lot of data in order to build robust databases. This also requires validation of QSAR relationships with large sets of *in vitro* and *in vivo* datasets. This requires extensive collaboration between scientists across different disciplines of toxicology.

Currently, several softwares are available commercially that are used for predictive toxicology assessments. These softwares have been validated against compounds across different classes and various toxicological end points. Utilizing such tools requires detailed research and information on the extent of evaluation and validation conducted in order to develop these tools. For example softwares that have been validated with large data sets for pharmaceutical compounds may not be appropriate for using in a chemical space and vice versa.

Names of certain commercially available Computational Tools Used In Industry

- QSAR tools: Derek, GastroPlus, ADMET Predictor, OECD QSAR Toolbox, ACD ToxSuite
- PK/PBPK Modeling Tools: GastroPlus, Simcyp, WinNonlin, Berkeley Madonna

Topic 4: Key Points

In this section, we explored the following main points:

- 1: The application of QSARS in systems toxicology.

Knowledge Check

1. QSAR databases link:

Relationship between chemical structures of compounds and their activity/toxicity

Relationship between different responses at different organizational levels

Physiologically relevant modeling tools

In vitro to in vivo extrapolation

Answer available at the end of this section.

2. Development of efficient QSAR tools depend on...

The user's familiarity with the software/tool

Knowledge of multiple QSAR models

A large database of validation training sets

All of the above

Answer available at the end of this section.

Answers

Knowledge Check

Relationship between chemical structures of compounds and their activity/toxicity

Knowledge Check

A large database of validation training sets

This page titled 8.4.4.4: Other Approaches for Predictive Toxicity Modeling is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 4.4: Other Approaches for Predictive Toxicity Modeling by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4.5: Technologies Used In Systems Biology/Toxicology

Learning Objectives

- 1: Understand the different omics technologies and how they are used.

"Systems biology/toxicology uses very powerful high-throughput platforms/tools such as the “omics” technologies. The human genome was first sequenced in 2003. It took 13 years and was extremely expensive. Advancement in technology has now made it possible to have genome sequencing completed in less than a day. It has also made handling and interpretation of large volumes of data at different levels (gene, transcriptome, protein, metabolic) possible. Scientists across different disciplines are now using these technologies to provide a more holistic approach to disease and toxicity."

Systems Biology in Toxicology and Environmental Health; First Edition, Chapter 1

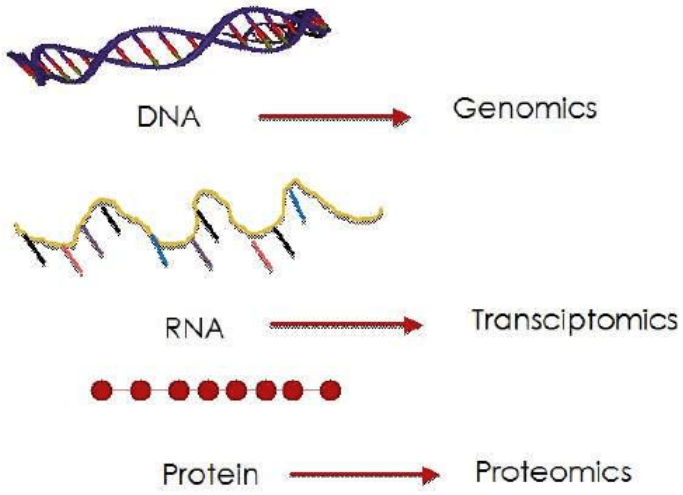


Figure 8.4.4.5.1: A schematic approach for “omics” technologies. Parts of illustration from www.pixabay.com

Technologies Used In Systems Biology/Toxicology

Genomics: This refers to the technology which allows us to study the complete genetic material of an organism. This involves DNA sequencing and analysis. Some of the most common methods used for high-throughput genotyping for genome wide association studies (GWAS) include Illumina Omni Arrays which can simultaneously analyze upto 5 million markers per sample. Other examples are the HumanOmni 5 Quad (Omni 5) and the Affymetrix platforms.

GWAS: A GWAS involves analysis of genetic variants in different individuals to study particular traits associated with genetic variations. This enables scientists to study the underlying mechanisms of different diseases at the genomic level. Single base pair changes are the most common form of variants of the genome and are known as single nucleotide polymorphisms (SNP). While most of them are functionally harmless, rare ones can lead to changes at the protein level leading to functional impairment and diseases.

Genomics: Example Application

SNPs that results in functional changes at the protein level thereby causing a difference in phenotype such as a disease state is known as a mutation. These are considered rare genetic variants. An example of a disease state is cystic fibrosis which is cause by multiple mutation in the CFTR gene. This was made possible by genotyping families that were affected by this disease and

identifying markers (genetic variants) that could be linked to this disease. This is known as linkage analysis. This type of analysis was also successfully used to identify unique mutations that lead to rare diseases like Huntington's disease. While this approach has been successful for rare diseases it has been a challenge to utilize this for more common disease states such as cancer, heart and liver disease etc.

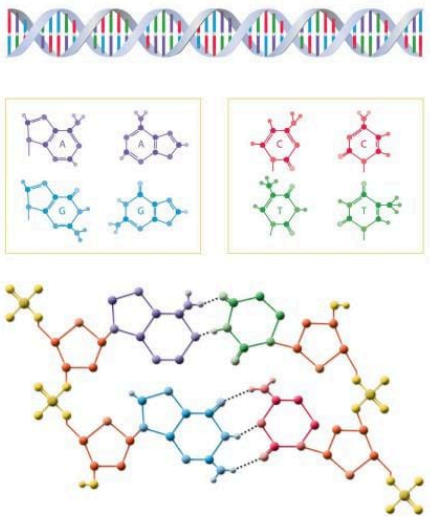


Figure \PageInde2: PLoS Comput Biol. 2012 Dec; 8(12)

Transcriptomics

"Transcriptomics is the study of all transcriptomes (RNA) in a genome. In other words, it is the study that enables us to study large volumes (> 200,000) of data related to gene expression at the RNA level.

Traditionally RNA expression was studied through a technique known as Northern blots. But this technique could not handle large volumes of data. Current techniques used for high-throughput data analysis are different kinds of microarrays and biochips. The most commonly used are the DNA based microarrays.

Quantitative gene expression analysis helps us understand the difference in expression of genetic product between different cells,tissues, species etc. Such microarrays are commercially made available from companies like Affymetrix, Agilent, Applied Microarrays, Illumina etc."

Systems Biology in Toxicology and Environmental Health; First Edition, Chapter 4

Different kinds of microarrays

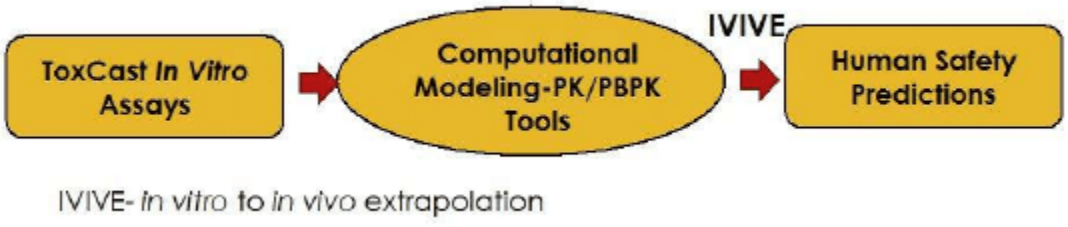
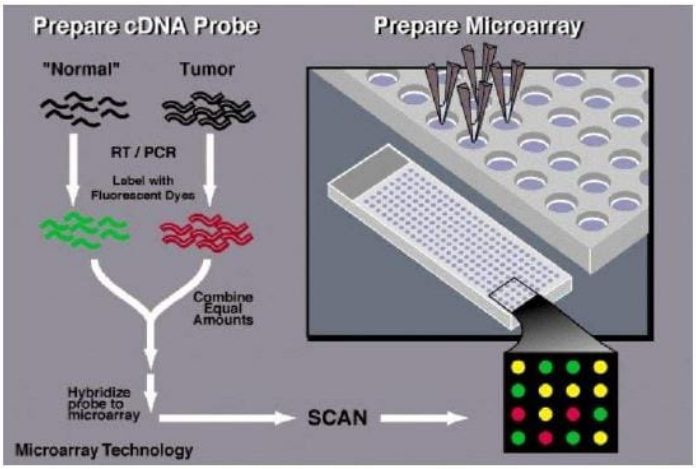


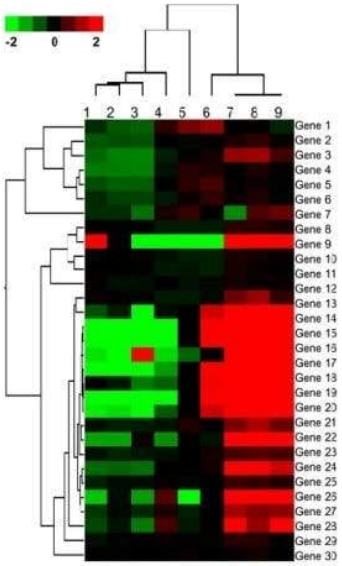
Figure 8.4.4.5.3: Different kinds of microarrays. Rebecca C. Fry. Systems Biology in Toxicology and Environmental Health, Chapter 4 (Kindle Location 1954). Elsevier Inc.

Analysis of microarray data



Microarray data provides relative expression levels of various genes. Quantitaive analysis of microarray data require sophisticated statistical tools and is expressed in the form of a “heat map” which provides expression patterns (upregulation/downregulation of genes) in the form of various colors. Significance analysis of microarrays (SAM is common technique) is a statistical tools that allows for quantitative determination of expression patterns.

Figure 8.4.4.5.4: Sample processing for microarrays. https://www.genome.gov/10000533/dna-...ay-technology/



The output which is the heatmap has different colors signifying differential expression patterns. In the example given below red color signifies upregulated gene expression while green signifies downregulation.

Figure 8.4.4.5.5: Heatmap generated from microarray. Rebecca C. Fry. Systems Biology in Toxicology and Environmental Health, Chapter 4 (Kindle Location 1954). Elsevier Inc.

Proteomics

"Proteomics is the field that studies total proteins in a system. This involves high-throughput profiling of proteins. This kind of expression pattern is especially useful since it allows us to study post translational effects, since all transcription products are not always converted to proteins; however, all functional aspects at the phenotypic level are almost always driven by proteins.

Proteomics is a powerful tool that can help us in identification of protein biomarkers specific to toxicity due to particular exposures or specific diseases states. Mass spectrometers are commercially sold by several companies such as Waters, Thermofisher, Agilent, Shimadzu, Perkin Elmer etc.

A mass spectrometer is used for analysis of proteomics data. Interpretation requires sophisticated bioinformatics tools."

Rebecca C. Fry. Systems Biology in Toxicology and Environmental Health, Chapter 4 (Kindle Location 1954). Elsevier Inc.

Sample processing and workflow for proteomics studies:

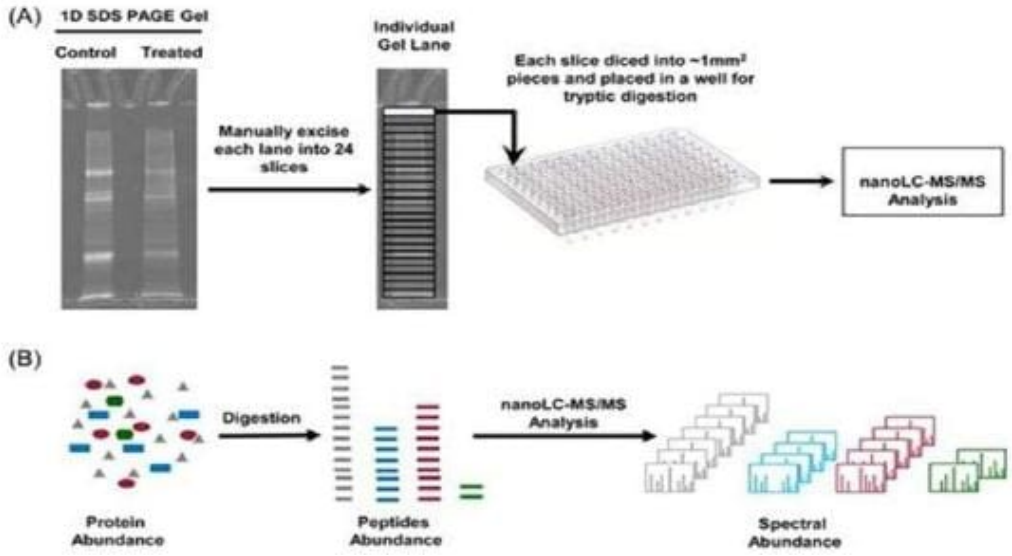


Figure 8.4.4.5.6: Sample processing and workflow for proteomics studies. Rebecca C. Fry. Systems Biology in Toxicology and Environmental Health, Chapter 4 (Kindle Location 1954). Elsevier Inc.

Metabolomics

"Metabolomics refers to the study of metabolites (low molecular weight products of cellular /biological processes that are found in cells, tissues, biological fluids etc.). Metabolomics provides an understanding of the differences in the biochemistry between different variants (such as diseased and healthy patient populations, control group versus treatment group, high dose group versus low dose group etc.)

Metabolic profiling can be easily performed in biological matrices such as urine, blood, plasma, serum and also a wide variety of tissues. Hence this can be used very efficiently in human health and safety assessment for evaluating biomarkers for exposure to toxicity to various agents. It can be used in drug discovery in the pharmaceutical industry for comparing metabolic profile between different dose level and control groups. In the clinical setting metabolomics can provide an understanding of differences in metabolite products in diseased versus healthy patient populations.

Platforms/tools used in metabolomics is similar to proteomics. Liquid chromatography mass spectrometers (MS) are broadly used for metabolite profiling of different biological matrices. Additionally, nuclear magnetic resonance (NMS) imaging is also widely used for metabolomics analysis. A major difference between the two tools is that NMR is a non destructive process and samples used for NMR can be reused for other purposes or returned to the biorepository whereas this is not possible with the LCMS method.

Metabolomic studies can be conducted using a “targeted approach” where a few selected analytes (metabolites) are analyzed based on hypothetical research. Such studies are mainly performed in early discovery phases. Alternatively, a more “broad spectrum” approach may be used in order to develop biomarkers for certain treatments or specific disease states.

As with other “omics” approaches, state-of-the art bioinformatics and statistical tools are used for quantitative interpretation of the data generated from metabolomics platforms (LCMS, NMR)."

Rebecca C. Fry. Systems Biology in Toxicology and Environmental Health, Chapter 4 (Kindle Location 1954). Elsevier Inc.

Metabolomics Workflow

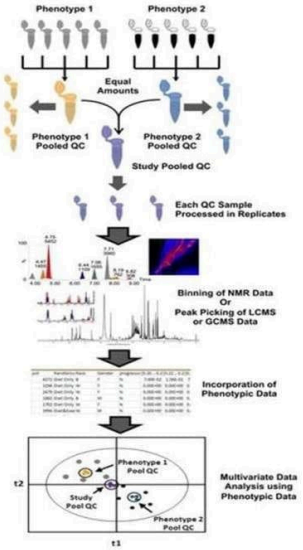


Figure 8.4.4.5.7: Metabolomics Workflow. Rebecca C. Fry. Systems Biology in Toxicology and Environmental Health, Chapter 4 (Kindle Location 1954). Elsevier Inc.

Topic 5: Key Points

In this section, we explored the following main points:

- 1: The application of the different “omics” technologies in systems toxicology.

Knowledge Check

1. Which of the following processes is a non destructive process in which samples can reused or returned back to the biorepository?

Microarray

DNA sequencing

NMR

Protein Isolation

Answer available at the end of this section.

2. Which of the following technologies is used to study complete RNA profiles

Nuclear Magnetic Resonance

Microarrays

DNA sequencing

Physiologically based pharmacokinetic models

Answer available at the end of this section.

3. The central dogma of molecular biology states is...

DNA to RNA and RNA to protein

RNA to protein to DNA

Protein to DNA to RNA

None of the above

Answer available at the end of this section.

4. Liquid Chromatography mass spectroscopy is used for...

Genomics

Proteomics

Metabolomics

Proteomics and metabolomics

Answer available at the end of this section.

Answers

Knowledge Check
NMR

Knowledge Check
Microarrays

Knowledge Check
DNA to RNA and RNA to protein

Knowledge Check
Proteomics and metabolomics

This page titled 8.4.4.5: Technologies Used In Systems Biology/Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 4.5: Technologies Used In Systems Biology/Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4.6: Takeaways Summary

The tremendous advancement at the technological level has made it possible to generate data at “high-throughput” levels. It has also enabled scientists to study toxicological processes from a more holistic approach. Instead of answering single questions at a time, a more comprehensive approach is now being applied to understand toxicological responses. “Systems” approach is being effectively used in industry, government and clinical settings.

In biopharmaceutical/chemical industries, thousands of molecules are screened in the early discovery phase to select target compounds with efficacy. Similarly, compounds can also be screened based on their toxicity fingerprints. Molecular fingerprints generated for each class of compounds/ specific chemistries can be used to screen compounds for different indications for future uses. This has also significantly reduced the time and increased efficiency in discovery programs in industry settings.

The government has also been using “the systems” approach successfully for safety assessment programs. Efforts such as the ToxCast and Tox 21 are examples where “systems” toxicology is being efficiently used to prioritize animal testing towards chemicals that pose the greatest risk to human health and safety.

In the clinical setting more and more information is becoming available via the “systems” approach for specific disease states that enable physicians and researchers to develop personalized medicines based on specific needs of patients.

While, all these efforts mark the beginning of a very promising future, there is still a lot of research necessary to utilize these tools and technologies to their full potential.

REFERENCES

- Systems Biology in Toxicology and Environmantal Health. First Edition. Editor Rebecca Fry. Chapters 1 and 4
- Incorporating Human Dosimetry and Exposure into High-Throughput In Vitro Toxicity Screening. *Rotroff et al., Toxicological Sciences;* 117(2), 348–358 (2010)

This page titled 8.4.4.6: Takeaways Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 4.6: Takeaways Summary by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.4.7: Section 4 Final Evaluation

1. *In vitro* to *in vivo* extrapolation (IVIVE) is modeling software that can utilize *in vitro* data and mathematically translate that into relevant *in vivo* information utilizing pharmacokinetic (PK).

True

False

Answer

True

2. Which of the following is a common tool used for both Quantitative Structure Activity Relationships (QSAR) and Physiologically based pharmacokinetic (PBPK)?

GastroPlus

ADMET Predictor

ACD Toxsuite

Derek

Answer

GastroPlus

3. HumanOmni 5 Quad (Omni 5), illumina Omni Arrays and the Affymetrix platforms are common methods used in:

Transcriptomics

Genomics

Proteomics

Metabolomics

Answer

Genomics

4. Quantitative Structure Activity Relationships (QSAR) approaches for predictive toxicity testing requires all the following EXCEPT:

Lots of data in order to build robust databases.

Validation of QSAR relationships with large sets of in vitro and in vivo datasets.

Extensive collaboration between scientists across different disciplines of toxicology.

A mass spectrophotometer.

Answer

A mass spectrophotometer.

5. Computational modeling tools utilizes the in vitro data and integrate them into human physiology with the help of...

Physiologically based pharmacokinetic model.

High-throughput in vitro model.

Hypothesis and diagnostic model.

Mathematical and computational model.

Answer

Physiologically based pharmacokinetic model.

6. Systems toxicology aims to fill this gap and utilize these data from different systems and integrate them into meaningful assessment for safety. This gap is:

Lack of modern equipment and technology for data collection.

Lack of interpretation and utilization of collected data.

Lack of human resources and management.

Lack of organ and system models.

Answer

Lack of interpretation and utilization of collected data.

7. Adverse Output Pathways (AOP) can be assessed in the following stages:

Molecular initiating event

Intermediate response at cellular level

Toxic response at organic level

All of the above

Answer

All of the above

8. In Metabolomics, the major difference between liquid chromatography mass spectrometers (LCMS) and nuclear magnetic resonance (NMS) is...

NMR samples cannot be reused for other purposes while LCMS samples can be reused for other purposes.

NMR is a destructive process whereas LCMS is a non-destructive process.

NMR is a non-destructive process whereas LCMS is a destructive process.

NMR samples cannot be returned to the biorepository whereas LCMS samples can be returned to the biorepository.

Answer

NMR is a non-destructive process whereas LCMS is a destructive process.

9. In the study of the underlying mechanisms of different diseases at the genomic level, the most common form of variants at this level is:

Polyploidy

Aneuploidy

Single Nucleotide Polymorphism (SNP)

Micronuclei variants

Answer

Single Nucleotide Polymorphism (SNP)

10. ___ is used for analysis of proteomics data:

A mass spectrophotometer

DNA micro array

Illumni Omni Array

DNA sequencing

Answer

A mass spectrophotometer

11. The ToxCast program utilizes which kind of assays?

In vivo

In vitro

In silico

All of the above

Answer

In vitro

12. The following agencies were established to address the challenge of large number of chemicals that go to market with very limited or no toxicity testing:

Environmental Protection Agency (EPA) in the US.

Registration Evaluation Authorization and Restriction of Chemicals (REACH) in Europe.

Both A & B are correct.

Occupational Safety and Health Administration (OSHA).

Answer

Both A & B are correct.

This page titled 8.4.4.7: Section 4 Final Evaluation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 4 Final Evaluation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5: REGULATORY TOXICOLOGY

Regulatory toxicology is where the science of toxicology meets the regulations, policies and guidelines that protect human health and the environment from chemicals. In this e-module you will learn about global, regional, national, state, and non-governmental regulatory toxicology.

8.4.5.1: Introduction to Regulatory Toxicology

8.4.5.2: Global Regulatory Toxicology

8.4.5.3: Topic 3- Regional Regulatory Toxicology

8.4.5.4: National Regulatory Toxicology

8.4.5.5: State Regulatory Toxicology

8.4.5.6: Non-Governmental Regulatory Toxicology

8.4.5.7: Section 5 Final Evaluation

This page titled 8.4.5: Regulatory Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5: Regulatory Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5.1: Introduction to Regulatory Toxicology

Learning Objectives

- 1: Identify regulatory toxicology inside and outside of government.
- 2: Give examples of regulatory toxicology at various scales and locales.
- 3: Explain the difference between regulation and guidance.

What is Regulatory Toxicology?

Regulatory Toxicology is where the science of toxicology meets the regulations, policies and guidelines that protect human health and the environment from chemicals. Regulatory toxicology commonly is associated with government agencies. These agencies may vary dramatically in their size and scope. For example, the United Nations covers the entire globe, while agencies within a city are limited to the area covered by the municipality.

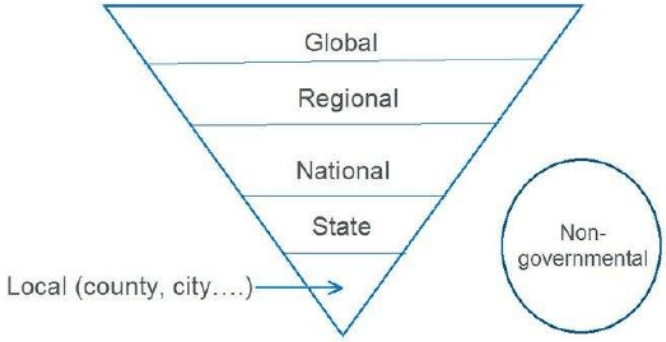


Figure 8.4.5.1.1: Introduction to regulatory toxicology

Regulatory agencies generally have specific focus areas that they address. For example, the U.S. Occupational Safety and Health Administration (OSHA) covers hazardous chemicals in the workplace, while the U.S. Consumer Product Safety Commission (CPSC) addresses chemical hazards in consumer products. Lastly, regulatory toxicology also occurs in non-governmental agencies such as professional societies, private industry and various advocacy groups. Further discussion of these is provided later in the module. More information on OSHA and CPS.

What is the difference between a regulation, policy, and guideline?

A regulation is a rule or order issued by a governmental authority that has the force of law. Often regulations are developed by experts in a governmental authority to enforce legislation. An example of a regulation is the Food Quality Protection Act (FQPA) passed by the U.S. Congress and signed into law by the President in 1996.

Policies and guidelines are principles and approaches that clarify and interpret regulations. As such, policies and guidelines do not carry the force of law but provide important direction.

The Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) is a U.S. regulation that regulates the broad class of chemicals used as pesticides (i.e., substances used to combat “pests”). The U.S. Environmental Protection Agency (EPA) has authority over FIFRA, and has in turn established many policies and guidelines concerning pesticides.

One area that EPA has established multiple policies and guidelines is in the registration of pesticides (i.e., EPA review and approval).

Topic 1: Key Points

In this section, we explored the following main points:

- 1: What is Regulatory Toxicology?
- 2: What is the difference between a regulation, policy and guidance

Knowledge Check

1. What is regulatory toxicology?

The forces that govern the regulation of homeostasis in the body

The branch of toxicology that studies regulations

The intersection of the science of toxicology with the regulatory world protecting human health and the environment

Regulations that dictate the science underlying toxicology

Answer available at the end of this section.

2. Which of the following have the force of law?

A regulation

A policy

A guideline

All of the above

Answer available at the end of this section.

Answers

Knowledge Check

The intersection of the science of toxicology with the regulatory world protecting human health and the environment

Knowledge Check

A regulation

This page titled 8.4.5.1: Introduction to Regulatory Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.1: Introduction to Regulatory Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5.2: Global Regulatory Toxicology

Learning Objectives

- 1: Define what is meant by “global regulatory toxicology”
- 2: Give an example of a global regulatory guideline

What is Global Regulatory Toxicology?

Global regulatory toxicology is exactly as it sounds, it deals with regulatory toxicology on a global scale (i.e., the entire planet). Relative to other jurisdictions (e.g., nations, states, cities), there are few regulatory toxicology initiatives that are global in nature. Some initiatives (e.g., clean drinking water, clean air) with similar intent may span many parts of the globe and appear global, but they lack a global consensus.

Global regulatory toxicology initiatives often originate from activities by the United Nations (UN), a global organization bringing together member countries to confront common challenges.

Example of Global Regulatory Toxicology: Globally Harmonized System (GHS)

Full name is the Globally Harmonized System of Classification and Labelling of Chemicals (GHS). It was created by the UN. Work on GHS began in 1992 and the first edition was released in 2003. GHS is updated every two years. The goal is to harmonize the criteria by which chemicals are classified in terms of their hazards. Hazards include physical (e.g., flammability), environmental (e.g., toxicity to fish), and human heath (e.g., acute toxicity people).

Example of GHS classification – Hazards to the Eye

Category	Hazard	Hazard statement	Requirement for eye protection?	Signal Word	Symbol
1	Irreversible effects	Causes serious eye damage.	Yes	DANGER!	
2A	Reversible effects	Causes serious eye irritation.	Yes	WARNING!	
2B	Reversible effects	Causes eye irritation.	No	WARNING!	None
Not classified	None	None	No	None	None

Figure 8.4.5.2.1: The Globally Harmonized System (GHS)

The Globally Harmonized System (GHS)

Prior to Globally Harmonized System (GHS), each country had its own criteria for hazard classification, and some countries had multiple criteria. This provided challenges and confusion to the general public and other stakeholders. GHS is a guideline, not a regulation. However, once adopted by a country, GHS generally becomes a regulation.

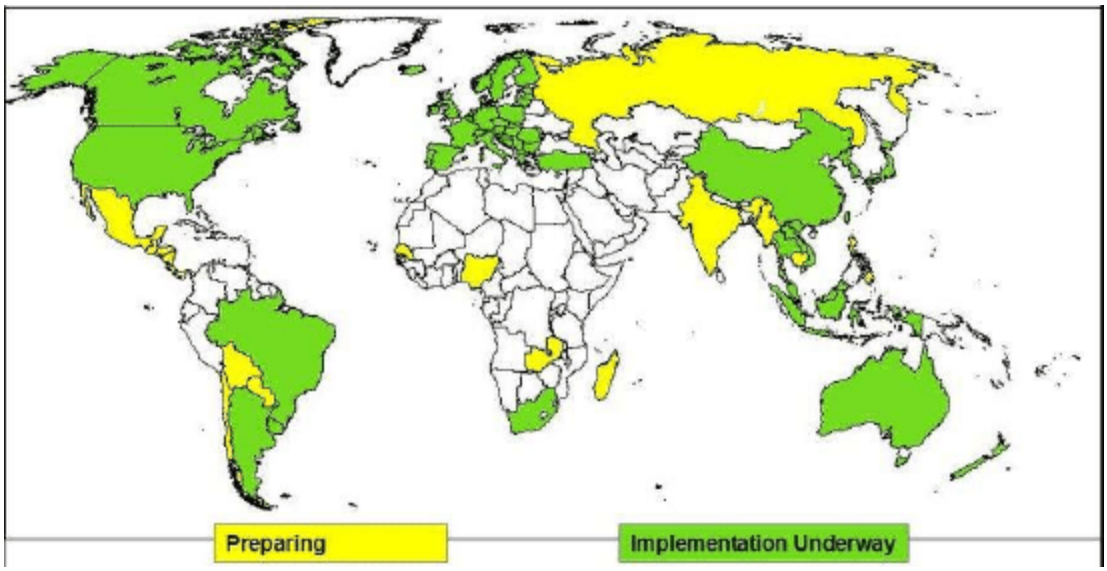


Figure 8.4.5.2.2: GHS Implementation Status – May, 2017

Topic 2: Key Points

In this section, we explored the following main points:

- 1: What is Global Regulatory Toxicology?
- 2: The Globally Harmonized System (GHS)

Knowledge Check

1. Which of the following generally is involved with establishing a global regulatory toxicology initiative?

Environmental Protection Agency (EPA)

United Nations (UN)

Food and Drug Administration (FDA)

All of the above

Answer available at the end of this section.

2. Which of the following is an example of a global regulatory toxicology initiative?

Clean Air Act

Clean Water Act

Globally Harmonized System

Safe Drinking Water Act

Answer available at the end of this section.

Answers

Knowledge Check

United Nations (UN)

Knowledge Check

Globally Harmonized System

This page titled 8.4.5.2: Global Regulatory Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.2: Global Regulatory Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5.3: Topic 3- Regional Regulatory Toxicology

Learning Objectives

- 1: Define what is meant by a “regional regulatory toxicology”
- 2: Give an example of a regional regulation

What is Regional Regulatory Toxicology?

Regional regulatory toxicology exists between global and national regulatory toxicology. It deals with regulatory toxicology that includes multiple countries. Often these countries are adjacent or in close proximity (e.g., United States and Canada) but that is not a requirement.

A group of nations that frequently exert regulations on a regional scale is the European Union (EU). The EU started in 1951 with six European countries agreeing to cooperate and has expanded over the ensuing decades to the current list of 28 member countries. -One current member of the EU, the United Kingdom has decided through a popular vote to eventually exit the EU (i.e., Brexit).

It started as a paper in 2001 and entered into force on June 1, 2007. It's a new approach for the regulation of chemicals in the EU. Required key hazard information to be obtained for existing chemicals to remain in commerce, or for a new chemicals to enter into commerce.

Information falls into three broad hazard categories: physical-chemical, human health and environmental health. The types and amounts of information varies depending on the volume of chemical used in commerce. The more the chemical is used, the more the data.



In addition to hazard information, REACH requires assessments to determine if chemicals in commerce pose unreasonable risk(s). REACH has resulted in the assessment and/or re-assessment of 1000s of chemicals. A large database of toxicity information was created. Some countries are using REACH as a model to develop new, or update existing, chemical regulations, for example, South Korea with K-REACH.

Topic 3: Key Points

In this section, we explored the following main points:

- 1: What is Regional Regulatory Toxicology?
- 2: The Registration, Evaluation, Authorization and restriction of CHemicals (REACH) regulation

Knowledge Check

1. Which of the following generally is a trait commonly found in a regional regulatory toxicology initiative?

Involvement of multiple regulatory agencies within a country

Involvement of all/most countries around the globe

Involvement of multiple states within the United States

Involvement of two or more countries in a regulatory initiative

Answer available at the end of this section.

2. Which of the following is an example of a regional regulatory toxicology initiative?

GHS

REACH

K-REACH

FQPA

Answer available at the end of this section.

Answers

Knowledge Check

Involvement of two or more countries in a regulatory initiative

Knowledge Check

REACH

This page titled 8.4.5.3: Topic 3- Regional Regulatory Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.3: Topic 3: Regional Regulatory Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5.4: National Regulatory Toxicology

Learning Objectives

After completing this lesson, you will be able to:

- Define what is meant by “National Regulatory Toxicology”.
- Give an example of a national regulatory regulation.

What is National Regulatory Toxicology?

As the name suggests, it focuses on regulatory toxicology at the national level (e.g., United States, Canada, China). Regulations are set by authorities that apply for the entire country. The focus on the United States (US) deals with regulations established at the federal level, and all states within the US must comply with federal regulations.

Agency	Population Served	Mainly Regulates Substances in:
EPA	General population (humans); Environment (e.g., aquatic species)	• Environment (e.g., air, soil, water)
FDA	General population; Medical Patients	• Food • Drugs and Medical Devices
OSHA	Workers	• Occupational environment (e.g., air in a factory)
CPSC	Consumers	• Consumer products (e.g., shampoos, toys, clothing)

Figure 8.4.5.4.1: Example agencies that set national toxicology regulations in the U.S.

Example agencies that set national toxicology regulations in the U.S.:

- Environmental Protection Agency (EPA)
- Food and Drug Administration (FDA)
- Occupational Safety and Health Administration (OSHA)
- Consumer Product Safety Commission (CPSC)

All agencies aim to set regulations to protect against adverse health effects from exposure to substances. Different agencies regulate on different types of substances, that are used for different purposes and affect different populations.

Goal of **EPA** is to protect human and environmental health; deals with exposures to substances in the environment (e.g., air, water, soil). It typically pertains to exposures that the general population generally does not have control over (e.g., air pollution) as well as modernizing chemical regulation in the US, e.g. Toxic Substances Control Act (TSCA) was recently revised.

TSCA was first passed in 1976. TSCA was revised with passage of the **Frank R. Lautenberg Chemical Safety for the 21st Century Act** in 2016. It regulates new or already existing chemicals manufactured or in use in the US.

Topic 4: Key Points

In this section, we explored the following main points:

- 1: What is National Regulatory Toxicology?

- 2: Examples of National Agencies in the U.S.
- 3: What do these Agencies regulate?
- 4: Highlight: EPA and TSCA Reform.

Knowledge Check

1. EPA regulates chemicals that are in:

Food

Environmental media (e.g., water, air)

Drugs and Medical Devices

Shampoo

Answer available at the end of this section.

2.US States can choose not to comply with regulations set by the FDA

True

False

Answer available at the end of this section.

Answers

Knowledge Check

Environmental media (e.g., water, air)

Knowledge Check

False

This page titled 8.4.5.4: National Regulatory Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.4: National Regulatory Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5.5: State Regulatory Toxicology

Learning Objectives

- 1: Define what is meant by “State Regulatory Toxicology”.
- 2: Give an example of a state regulatory regulation.

What is State Regulatory Toxicology?

As the name suggests, it deals with regulatory toxicology for the US on a state level (e.g. Texas, Delaware, Arizona). It was established by authorities applicable to a given state. State regulatory toxicology only applies with the state’s boundaries; however, they may influence adjacent states or even national regulatory toxicology.

State regulatory toxicology is under the layer of the US national government. It sits under a complex web of state and local laws and policies, in addition to regulatory authorities. The make-up of state and local governments varies widely across the US; while they have mutual specific features, their organizations differ. Whatever their design, state and local governments can sometimes have a much greater impact on people's lives than the federal government.

The Federal-State Toxicology and Risk Analysis Committee (FSTRAC) is made up of representatives from U.S. state health and environmental agencies and U.S. EPA personnel.

FSTRAC is an integral part of EPA’s communication strategy with states and tribes for human health risks associated with water contamination. It fosters cooperation, consistency, and an understanding of EPA’s and different states’ goals and problems in human health risk assessment. Additionally, it allows states and the federal government to work together on issues related to the development and implementation of regulations and criteria under the Safe Drinking Water Act and Clean Water Act.

FSTRAC members have supported development of Human Health Benchmarks for Pesticides (HHBP).

- 1: Represent levels of pesticides in drinking water that are not anticipated to cause health effects.
- 2: Used to help assess drinking water quality for pesticides that do not have other regulatory toxicology standards.

Examples of agencies that set national toxicology regulations in the U.S.:

-
-
-
-
-
-

The goal of state agencies is the same as federal agencies, protect people and the environment from health effects associated with chemical exposures.

States have differing regulatory toxicology requirements and focuses. Reasons for the differences could be due to: state history, geography, culture, population size and diversity, major industries, etc...

Examples of state regulatory toxicology as it concerns chemicals in air:

TEXAS

Effects Screening Levels (ESLs) – Air concentrations generally applicable to a specific chemical set by the Texas Commission on Environmental Quality (TCEQ). Play an important role in the regulation of air emissions from companies located in the state. List of ESLs here.

CALIFORNIA

Reference Exposure Levels (RELs) – Air concentrations generally applicable to a specific chemical set by the California Environmental Protection Agency (CalEPA). Represent an air concentration that does not pose a health risk to people. List of RELs here.

Agency	State Specific Rule or Regulation	Requirement
CalEPA	Proposition 65	• List of chemicals known to the state to cause cancer or reproductive toxicity
TCEQ	Texas Risk Reduction Program (TRRP)	• Established cleanup standards for contaminated chemical sites

Figure 8.4.5.5.1: Example of state toxicology regulations and the requirement that must be followed.

Example agencies that set state toxicology regulations

-
-
-

Agency	Regulation
Office of Environmental Health Hazard Assessment	Proposition 65 (Prop 65)
State of Washington Department of Ecology	Children's Safe Products Act (CSPA)
Minnesota Department of Health – Environmental Health	Minnesota Wellhead Protection Program

Figure 8.4.5.5.2: Example agencies that set state toxicology regulations.

An example of a State Regulatory Agency is the **CalEPA Office of Environmental Health Hazard Assessment (OEHHA)**: Provides requirements and guidance for the California Proposition 65 Safe Drinking Water and Toxic Enforcement Act of 1986 (Proposition 65).

Proposition 65

Purpose

Enable consumers to make informed decisions regarding chemical exposures.

Reason

Established to protect California citizens from chemicals known to the state to cause cancer, birth defects, or other reproductive harms.

Scope

Addresses chemical exposures to the citizens of California that may occur through consumer products, workplace exposures, and exposures occurring via the environment.

Basics

OEHHA publishes a list of chemicals known to cause cancer, birth defects or other reproductive harm. The list is updated regularly and currently contains approximately 900 chemicals. Once a chemical is listed, companies have 12 months to comply with warning requirements under the regulation.

Proposition 65 is referred to as a "risk-based" regulation in that the warning requirements only apply if the risk from chemical exposure are too high as defined by the regulation.

Exposure examples:

- Oral
- Inhalation
- Skin contact

Proposition 65 upcoming changes...

The regulation has undergone revisions: New Proposition 65 Warnings, that will now require companies to add a symbol and change the phrasing of the warning. For example: "**WARNING:** This product can expose you to chemicals including arsenic, which is known to the State of California to cause cancer. For more information, visit here.

Topic 5: Key Points

In this section, we explored the following main points:

- 1: What is State Regulatory Toxicology?
- 2: An example of how federal and state agencies work together
  - What is the Federal-State Toxicology and Risk Analysis Committee (FSTRAC)
  - Two example outcomes of the FSTRAC's workgroup
- 3: Example of State Agencies
- 4: Highlight two state specific guidance and rules as it concerns chemicals in air
  - Guidance
    - Effects Screening Levels (ESLs)
    - Reference Exposure Levels (RELs)
- 5: Rules
  - California Proposition 65
  - Texas Risk Reduction Program rule

Knowledge Check

1. Proposition 65 is a regulation established in:

Florida

Texas

California

New York

Answer available at the end of this section.

2. Which of the following could explain some of the variability observed between states in terms of regulatory toxicology?

Degree of urbanization

Type of natural resources

Degree of industrialization

All of the above

Answer available at the end of this section.

Answers

Knowledge Check

California

Knowledge Check

All of the above

This page titled 8.4.5.5: State Regulatory Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.5: State Regulatory Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5.6: Non-Governmental Regulatory Toxicology

Learning Objectives

- 1: Define what is meant by a “Non-Governmental Regulatory Toxicology”.
- 2: Give several examples of non-governmental entities that conduct and influence regulatory toxicology.

What is Non-Governmental Regulatory Toxicology?

They are groups that are not officially part of a government agency that conduct and/or influence regulatory toxicology. Examples of such groups may include not-for-profit organizations, advocacy groups, professional societies, industry trade associations and individual companies. The term “NGO” (short for non-governmental organization(s)) is sometimes used; however, the term “NGO” often is limited to advocacy groups.

These groups often can conduct or influence regulatory toxicology on a much faster timeframe than government agencies. Government agencies often take many months or years to enact regulations, policies, and guidance; while non-governmental entities often proceed at a fraction of this time. Lastly, they often lack the transparency found in government agencies in terms of decision making, as well as the ability to provide input.

Examples of a non-governmental entities that conduct and influence regulatory toxicology

-

- Advocacy groups – Often have a particular focus such as the proper treatment of animals, environmental sustainability, women’s health…
- Individual companies – For-profit businesses that decide to address a particular topic related to regulatory toxicology. Examples include eliminating use of certain chemicals, reduce use of water, increasing recycling…
- Trade Associations – represent business that have a common or similar commercial interest (e.g., chemical manufacture, consumer products). Often create industry best practices or guidance for issues related to regulatory toxicology.

Topic 6: Key Points

In this section, we explored the following main points:

- 1: What is Non-Governmental Regulatory Toxicology?
- 2: Examples of different types of non-governmental entities that conduct and influence regulatory toxicology.

Knowledge Check

1. Which of the following is a trait found in a non-governmental entities?

Must engage with the general public prior to taking actions.

Have lengthy processes that often take decades to complete.

Is not a government agency or formally associated with a government agency.

A narrowly defined entity that is not common in society.

Answer available at the end of this section.

2. Which of the following is an example of non-governmental regulatory toxicology?

A not-for profit company developing standards and conducting certifications regarding the sustainability of products.

An advocacy group that provides input to government and industry related to reducing lead in consumer products.

An organization representing chemists develops guidance on workplace safety concerning corrosive chemicals.

All of the above

Answer available at the end of this section.

Answers

Knowledge Check

Is not a government agency or formally associated with a government agency.

Knowledge Check

All of the above

This page titled 8.4.5.6: Non-Governmental Regulatory Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.6: Non-Governmental Regulatory Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.5.7: Section 5 Final Evaluation

1. _____ was created by UN to harmonize the criteria by which chemicals are classified in terms of their hazards.

Federal Insecticide, Fungicide and Rodenticide Act (FIFRA).

Consumer Product Safety Commission (CPSC).

Globally Harmonized System of Classification and Labelling of Chemicals (GHS).

The US Food and Drug Administration (FDA).

Answer

Globally Harmonized System of Classification and Labelling of Chemicals (GHS).

2. Regulatory agencies generally have specific focus areas that they address, examples include the following EXCEPT:

Food Quality Protection Act (FQPA) addresses the quality of food and chemicals used in an environment.

Consumer Product Safety Commission (CPSC) addresses chemical hazards in consumer products.

Occupational Safety and Health Administration (OSHA) addresses hazardous chemicals in the work place.

Food and Drug Administration (FDA) regulates pharmaceutical drugs used in humans and animals.

Answer

Food Quality Protection Act (FQPA) addresses the quality of food and chemicals used in an environment.

3. Food and Drug Administration (FDA) is a national regulatory agency that serves general population and medical patients while Consumer Product Safety Commission (CPSC) serves general population and environment.

True

False

Answer

False

4. The purpose of Proposition 65 is to enable consumers make informed decisions regarding chemical exposures while its scope is to address chemical exposures to citizens of California that may occur through consumer products, workplace and environmental exposures.

True

False

Answer

True

5. Registration, Evaluation, Authorization and restriction of CHemicals (REACH) is an example of:

Global regulatory toxicology.

Regional regulatory toxicology.

National regulatory toxicology.

State regulatory toxicology.

Answer

Regional regulatory toxicology.

6. One of the advantages of a Non-Governmental Organization (NGO) over government agencies is:

NGOs often conduct regulatory toxicology on a much faster timeframe than government agencies.

NGOs are transparent in terms of decision making while government agencies are not transparent.

NGOs are not only limited to advocacy group but it influences global regulations.

NGOs are officially part of government agency that influence regulatory toxicology.

Answer

NGOs often conduct regulatory toxicology on a much faster timeframe than government agencies.

7. The following are examples of national toxicology regulatory agencies in the U.S. EXCEPT:

Consumer Product Safety Commission (CPSC).

Environmental Protection Agency (EPA).

Texas Commission on Environmental Quality (TCEQ).

Food and Drug Administration (FDA).

Answer

Texas Commission on Environmental Quality (TCEQ).

8. State of Washington Department of Ecology is an agency that sets state toxicology regulations for:

Proposition 65.

Wellhead Protection Program.

Children's Safe Products Act (CSPA).

Risk reduction Program.

Answer

Children's Safe Products Act (CSPA).

9.Global regulatory toxicology brings together member countries to confront common challenges on a global scale, an example is:

United Nations (UN).

Environmental Protection Agency (EPA).

Consumer Product Safety Commission (CPSC).

Occupational Safety and Health Administration (OSHA).

Answer

United Nations (UN).

10. Which of the following statements is true?

CPSC mainly regulates substance in consumer products e.g. shampoos, clothing etc.

EPA mainly regulates substance in occupational environment e.g. air in a factory.

CPSC mainly regulates pharmaceutical drugs and medical devices.

OSHA mainly regulates toxic substances in consumer products.

Answer

CPSC mainly regulates substance in consumer products e.g. shampoos, clothing etc.

11. Which of the following statements is true?

Policies are principles that clarify regulation and carry the force of law.

Guidelines are approaches that interpret regulations and carry the force of law.

Regulations are rules issued by governmental authority and carry the force of law.

Regulations are orders issued by government authority and do not carry the force of law.

Answer

Regulations are rules issued by governmental authority and carry the force of law.

12. Key hazard information to be obtained for existing chemicals to remain in commerce or for new chemicals to enter into commerce falls into three broad hazard categories:

Physical-chemical, animal health and environmental health.

Animal health, human health and environmental health.

Environmental health, physical-chemical and biosafety.

Physical-chemical, human health and environmental health

Answer

Physical-chemical, human health and evironmental health

This page titled 8.4.5.7: Section 5 Final Evaluation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 5 Final Evaluation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6: PRINCIPLES OF TOXICOLOGY

Introduction

ToxTutor is a self-paced tutorial covering key principles of toxicology and was adopted from the National Library of Medicine (NLM) chemical and toxicology databases. While a knowledge of anatomy and physiology is not required for viewing ToxTutor, the **Introduction to the Human Body** from the National Cancer Institute provides a good introduction to the topic.

Topics Covered in this Course

ToxTutor is divided into the following sections:



Each section of ToxTutor contains one or more related content pages. *Next* and *Back* buttons are provided to allow you to navigate through these pages. For more information, see the "Getting Around" section below.
The basic principles of toxicology described in ToxTutor are similar to those taught in university programs and are well described in toxicology literature. A list of the textbooks used as the primary resources for the tutorials is found in the **Bibliography**.

Using ToxTutor

Getting Around

It will take approximately three hours to complete this self-paced tutorial.
There are a variety of ways you can navigate ToxTutor. You can:

- Use the **Glossary** button in the upper right corner of each page to access the **glossary**.
- Click the links in the list above (or any underlined words in any module which are linked) to access sections directly.
- Click the Previous and *Next* links at the bottom of each page to move through the material.
- Use a combination of the above methods to explore the course contents.

Progress

At the top of each page in ToxTutor are links indicating where the current page falls within the overall ToxTutor program. You can click these links to return to this homepage or to the section that contains the page.

Links

Throughout the course, you will encounter links. Any link to a resource outside of ToxTutor, will typically open in a new tab or window. All other links are to other areas within ToxTutor. Links are bold and underlined as seen **here** and **here**.

In addition, clicking on some images may open another external website for more information.

Credits

ToxTutor was adopted from the U.S National Library of Medicine in 2021. More Information can be seen in the ToxTutor Bibliography.

8.4.6.1: Introduction to Toxicology

8.4.6.1.1: What is toxicology?

8.4.6.1.2: Basic Terminology

8.4.6.2: Dose and Dose Response

8.4.6.2.1: Dose and It's Impact on Toxicity

8.4.6.2.2: The Dose Response Relationship

8.4.6.2.3: Dose Estimates of Toxic Effects

8.4.6.2.4: Determining the Safety of a Drug

8.4.6.2.5: NOAEL and LOAEL

8.4.6.3: Toxic Effects

8.4.6.3.1: Types of Toxic Effects

8.4.6.3.2: Factors Affecting Toxicity

8.4.6.3.3: Systemic Toxic Effects

8.4.6.3.4: Organ Specific Toxic Effects

8.4.6.4: Interactions

8.4.6.4.1: Interactions

8.4.6.5: Toxicity Testing Methods

8.4.6.5.1: Testing and Assessing Toxicity

8.4.6.5.2: Clinical Investigations and Other Types of Human Data

8.4.6.5.3: Epidemiology Studies

8.4.6.6: Risk Assessment

8.4.6.6.1: Risk Assessment

8.4.6.6.2: Hazard Identification

8.4.6.6.3: Dose-Response Assessment

8.4.6.6.4: Exposure Assessment

8.4.6.6.5: Risk Characterization

8.4.6.7: Exposure Standards and Guidelines

8.4.6.7.1: Exposure Standards and Guidelines

8.4.6.7.2: Regulation of Consumer Products and Drug Safety

8.4.6.7.3: Environmental Exposure Standards/Guidelines

8.4.6.7.4: Occupational (Workplace) Exposure Standards/Guidelines/Approaches

8.4.6.8: Basic Physiology

8.4.6.8.1: Introduction to Basic Physiology

8.4.6.8.2: Homeostasis
8.4.6.8.3: Organs and Organ Systems
8.4.6.8.4: Tissues
8.4.6.8.5: Cells
8.4.6.8.6: Chemicals

8.4.6.9: Introduction to Toxicokinetics

8.4.6.9.1: What is Toxicokinetics

8.4.6.10: Absorption

8.4.6.10.1: Introduction to Absorption
8.4.6.10.2: Gastrointestinal Tract
8.4.6.10.3: Respiratory Tract
8.4.6.10.4: Dermal Route
8.4.6.10.5: Other Routes of Exposure

8.4.6.11: Distribution

8.4.6.11.1: Introduction to Distribution
8.4.6.11.2: Influence of Route of Exposure
8.4.6.11.3: Disposition Models
8.4.6.11.4: Structural Barriers to Distribution
8.4.6.11.5: Storage Sites

8.4.6.12: Biotransformation

8.4.6.12.1: Introduction to Biotransformation
8.4.6.12.2: Chemical Reactions
8.4.6.12.3: Biotransformation Sites
8.4.6.12.4: Modifiers of Biotransformation

8.4.6.13: Excretion

8.4.6.13.1: Introduction to Secretion
8.4.6.13.2: Urinary Excretion
8.4.6.13.3: Fecal Excretion
8.4.6.13.4: Exhaled Air
8.4.6.13.5: Other Routes

8.4.6.14: Cellular Toxicology

8.4.6.14.1: Adaptation
8.4.6.14.2: Cell Damage and Tissue Repair
8.4.6.14.3: Cancer
8.4.6.14.4: Neurotoxicity

8.4.6.15: Intuitive Toxicology and Risk Communication

8.4.6.15.1: Intuitive Toxicology
8.4.6.15.2: Risk Communication

8.4.6.16: Environmental Toxicology, Environmental Health, and One Health

8.4.6.16.1: Environmental Toxicology

8.4.6.16.2: Environmental Health

8.4.6.16.3: One Health

8.4.6.17: Conclusion

This page titled 8.4.6: Principles of Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 6: Principles of Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.1: INTRODUCTION TO TOXICOLOGY

Learning Objectives

After completing this lesson, you will be able to:

- Define toxicology and identify adverse effects.
- Recognize the history of toxicology.
- Explain how dose determines whether a substance is a remedy or a poison.
- Differentiate between toxic agents and toxic substances.

In this section...

Topics include:

8.4.6.1.1: What is toxicology?

8.4.6.1.2: Basic Terminology

What We've Covered

In this section, we covered several important concepts:

- Toxicology is the study of adverse effects of chemicals and physical agents on living organisms.
- A xenobiotic is a foreign substance taken into the body.
- A toxic agent is any chemical, physical, or biological agent that can produce an adverse biological effect.
- Toxic substances can be systemic toxicants, which affect the entire body or multiple organs, or organ toxicants, which affect a specific organ or tissues.
- The dose of a substance is the most important determinant of toxicity.

Coming Up...

In the next section, we will explore the concept of dose and its importance to toxicology in greater detail.

This page titled 8.4.6.1: Introduction to Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 1: Introduction to Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.1.1: What is toxicology?

What is Toxicology?

Toxicology is traditionally defined as "the science of poisons." Over time, our understanding of how various agents can cause harm to humans and other organisms has increased, resulting in a more descriptive definition of toxicology as "the *study of the adverse effects of chemical, physical, or biological agents on living organisms and the ecosystem, including the prevention and amelioration of such adverse effects.*"



Figure 8.4.6.1.1.1: Photos of a Gila monster anda Bothrops jararaca snake. (CC BY-SA 4.0; Leandro Avelar via source)

These adverse effects can take many forms, ranging from immediate death to subtle changes not appreciated until months or years later. They may occur at various levels within the body, such as an organ, a type of cell, or a specific biochemical. Our understanding of how toxic agents damage the body has progressed along with medical knowledge. We now know that various observable changes in anatomic or bodily functions actually result from previously unrecognized changes in specific biochemicals in the body.

Did you know?

The study of toxicology may appear to focus only on poisonings or disasters, but some toxic chemicals can have positive effects. Animal venoms, whether from bees, wasps, snakes, or Gila monsters, are composed of hundreds of chemicals that are being studied as treatments for human diseases.

For example, exantide, a drug derived from Gila monster saliva, has been approved for use in Type 2 diabetes. Captopril, which is used to treat hypertension and heart failure, was developed from studies on the chemical bradykinin-potentiating factor (BPF) in the venom of a South American snake Bothrops jararaca. Melitten, which comes from honeybee venom, is being investigated for its anticancer and antifungal properties.

Figure 8.4.6.1.1.1: Figure 1. Gila monster (top); Bothrops jararaca (bottom)
*(Image Source: Wikipedia, adapted under GNU Free Documentation License - **Original Images**)*

History of Toxicology

Prehistory

Poisonous plants and animals were recognized and their extracts used for hunting or in warfare.

1500 BC

Written records indicate that hemlock, opium, arrow poisons, and certain metals were used to poison enemies or for state executions.

c. 1198

With time, people began to make the connection between exposure to a specific substance and illness or death.

In 1198, Moses Maimonides wrote what may be the first collection of writings on toxicology, *The Treatise on Poisons and Their Antidotes*.

Renaissance and Age of Enlightenment

Certain fundamental toxicology concepts began to take shape. Noteworthy studies include those by Paracelsus in the 16th century and Orfila in the 19th century.

Paracelsus (16th Century)

Determined that specific chemicals were actually responsible for the toxicity of a plant or animal poison.

Documented that the body's response to those chemicals depended on the dose received.

Studies revealed that small doses of a substance might be harmless or beneficial, whereas larger doses could be toxic. This is now known as the **dose-response relationship**, a major concept in toxicology.

"*All substances are poisons; there is none which is not a poison. The right dose differentiates a poison and a remedy.*"
- Paracelsus

Orfila, the founder of toxicology (19th Century)

A Spanish physician, Orfila is often referred to as the **founder of toxicology**.

Orfila was the first to describe a systematic correlation between the chemical and biological properties of poisons of the time.

Orfila demonstrated the effects of poisons on specific organs by analyzing autopsy materials for poisons and tissue damage associated with them.

20th and 21st Centuries

Marked by great advancements in the level of understanding of toxicology.

DNA and various biochemicals that maintain body functions have been discovered.

Our level of knowledge of toxic effects on organs and cells has expanded to the molecular level.

Virtually all toxic effects are recognized as being caused by changes in specific cellular molecules and biochemicals.

Remedy or Poison?

Xenobiotic is the general term that is used for a ***foreign*** substance taken into the body. It is derived from the Greek term *xeno* which means "foreigner." Xenobiotics may produce beneficial effects (such as pharmaceuticals) or they may be toxic (such as lead).

As Paracelsus proposed centuries ago, dose differentiates whether a substance will be a remedy or a poison. A xenobiotic in small amounts may be nontoxic and even beneficial, but when the dose is increased, toxic and lethal effects may result.

The following image provides some examples that illustrate this concept.

Alcohol is non-toxic at 0.05%, toxic at 0.1%, and lethal at 0.5% ethanol blood levels. Carbon monoxide is non-toxic at less than 10%, toxic at 20 to 30%, and lethal at 60% or more % hemoglobin bound. Secobarbital, a sleep aid, is non-toxic or benefical at 0.1 mg/dL, toxic at 0.7 mg/dL, and lethal at over 1 mg/dL blood levels. Aspirin is non-toxic or beneficial at 0.65 g (2 tablets), toxic at 9.75 g (30 tablets), and lethal at 34 g (105 tablets). Ibuprofen is non-toxic or beneficial at 400 mg (2 tablets), toxic at 1,400 mg (7 tablets), and lethal at 12,000 mg (60 tablets).

Figure 8.4.6.1.1.2: Figure 3. Examples of varying *doses of the same substance as non-toxic or beneficial, toxic, and lethal (Image Source: Adapted from T. Gossel and J. Bricker, eds)*

This page titled 8.4.6.1.1: What is toxicology? is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 1.1: What is toxicology? by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.1.2: Basic Terminology

Toxicology Defined

Toxicology is an evolving medical science and toxicology terminology is evolving with it. Most terms are very specific and will be defined as they appear in the tutorial. However, some terms are more general and used throughout the various sections. The most commonly used terms are introduced in this section.

- **Toxicology** is the study of the adverse effects of chemicals or physical agents on living organisms.
- A **toxicologist** is a scientist who determines the harmful effects of agents and the cellular, biochemical, and molecular mechanisms responsible for the effects.
- **Toxinology**, a specialized area of study, looks at microbial, plant and animal venoms, poisons, and toxins.

Terminology and definitions for materials that cause toxic effects are not always consistently used in the literature. The most common terms are toxicant, toxin, poison, toxic agent, toxic substance, and toxic chemical.

Toxicant, toxin, and poison are often used interchangeably in the literature but there are subtle differences as shown below:

Toxicants:

- Substances producing adverse biological effects of any kind.
- May be chemical or physical in nature.
- Effects may be acute or chronic.

Person spraying a chemical onto a field of grass

Figure 8.4.6.1.2.1: *Pesticide chemicals are toxicants*
(Image Source: iStock Photos, ©)

Toxins:

- Peptides or proteins produced by living organisms.
- Venoms are toxins injected by a bite or sting.

Mushrooms

Figure 8.4.6.1.2.2: Amanita muscaria mushroom contains a *neurotoxin*
(Image Source: iStock Photos, ©)

Poisons:

- Toxins produced by organisms.

Black widow spider with a distinctive red hourglass shape on its abdomen

Figure 8.4.6.1.2.3: Black Widow spiders produce a *poison that is a toxin*
(Image Source: Texas Parks & Wildlife, ©)

A **toxic agent** is anything that can produce an adverse biological effect. It may be chemical, physical, or biological in form. For example, toxic agents may be:

Chemical (such as cyanide)

Physical (such as radiation)

Biological (such as snake venom)

The toxicity of the agent is dependent on the dose.

A distinction is made for diseases people get from *living* organisms. Organisms that invade and multiply within another organism and produce their effects by biological activity are not classified as toxic agents but as **biological agents**. An example of this is a virus that damages cell membranes resulting in cell death.

If the invading organisms excrete chemicals which are the basis for their toxicity, the excreted substances are known as **biological toxins**. In that case, the organisms are called **toxic organisms**. A specific example is tetanus. Tetanus is caused by a bacterium, *Clostridium tetani*. The bacteria *C. tetani* itself does not cause disease by invading and destroying cells. Rather, a toxin (neurotoxin) that the bacteria excrete travels to the nervous system and produces the disease (Figure 8).

Toxic Substances

A **toxic substance** is simply a material that has toxic properties. It may be a discrete toxic chemical or a mixture of toxic chemicals. For example, lead chromate, asbestos, and gasoline are all toxic substances. More specifically:

- Lead chromate is a discrete **toxic chemical**.
- Asbestos is a **toxic material** that does not have an exact chemical composition but comprises a variety of fibers and minerals.
- Gasoline s a **toxic substance** rather than a toxic chemical in that it contains a mixture of many chemicals. Toxic substances may not always have a constant composition. The composition of gasoline varies with octane level, manufacturer, time of season, and other factors.

Lead chromate, asbestos, and gasolene

Figure 8.4.6.1.2.4: Examples of *toxic substances: lead chromate (left), asbestos (center), and gasoline (right)*
(Image Source: iStock Photos, ©)

Systemic Toxicants and Organ Toxicants

Toxic substances may be **systemic toxicants** or **organ toxicants**.

A **systemic toxicant** affects the entire body or many organs rather than a specific site. For example, potassium cyanide is a systemic toxicant in that it affects virtually every cell and organ in the body by interfering with the cells' ability to use oxygen.

Toxicants may also affect only specific tissues or organs while not producing damage to the body as a whole. These specific sites are known as the **target organs** or **target tissues**.

- Benzene is a **specific organ toxicant** in that it is primarily toxic to the blood-forming tissues.
- Lead is also a specific organ toxicant; however, it has three **target organs:** the central nervous system, the kidneys, and the hematopoietic system.

A toxicant may affect a specific type of tissue (such as connective tissue) that is present in several organs. The toxic site is then considered the **target tissue**.

A systemic toxicant affects the entire body or several organs (left); an organ toxicant affects a specific organ (right)

Figure 8.4.6.1.2.5: Systemic toxicant *and organ toxicant*
(Image Source: iStock Photos, ©)

Types of Cells

The body is composed of many types of cells, which can be classified in several ways. Table 1 shows examples of one classification of one type of cells.

Cell Types	Examples

Basic structure	cuboidal cells
Tissue type	hepatocytes of the liver
Germ cells	ova and sperm
Somatic cells	non-reproductive cells of the body

Germ cells are involved in reproduction and can give rise to a new organism. They have only a single set of chromosomes peculiar to a specific sex. Male germ cells give rise to sperm and female germ cells develop into ova. Toxicity to germ cells can cause effects in a developing fetus that lead to outcomes such as birth defects or miscarriage.

Somatic cells are all body cells except the reproductive germ cells. (Somatic cells include the "basic structure" and "tissue type" cells listed in Table 1). They have two sets (or pairs) of chromosomes. In an exposed individual, toxicity to somatic cells causes a variety of toxic effects, such as dermatitis, death, and cancer.

Figure 6 illustrates the differences between germ cells and somatic cells.

Germ cells (sperm and egg cells) have a single set of chromosomes. Somatic cells (all other body cells) have two sets, or pairs, of chromosomes.

Figure 8.4.6.1.2.6: Germ cells and somatic cells (Image Source: National Human Genome Research Institute, **http://www.genome.gov**)

Natural and Man-Made Chemicals

Often, people mistakenly assume that all man-made chemicals are harmful and natural chemicals are beneficial. In reality, natural chemicals can be just as harmful to human health as man-made chemicals, and in many cases, more harmful. Figure 12 compares the toxicity of several natural and man-made chemicals.

Infographic titled 'Natural and Man-made chemicals.' A common misconception is that all man-made chemicals are harmful and all natural chemicals are good for us. However, many natural chemicals are just as harmful to human health, if not more so, than man-made chemicals. Examples of chemicals that result in toxic effects at 1000 mg/kg of body weight include natural chemicals such as muscimol (found in fly agaric mushrooms), solanine (found in green potatoes), and amygdalin (found in apple seeds), and man-made chemicals such as ethylene glycol (used in anti-freeze), aspirin, and sodium thiopental (formerly used for lethal injections). Examples of checmials with no toxic effects seen at 1000 mg/kg of body weight include natural chemicals such as citric acid (found in lemons and limes), sucrose (table sugar), and water; and man-made chemicals such as MSG (a food additive), teflon (used in non-stick pans), and aspartame (an artificial sweetener). Any substance, if given in large enough amounts, can cause death. Some are lethal after only a few nanograms, whilst others require kilograms to achieve a lethal dose. Chemical toxicity is a sliding scale, not black and white - and whether a chemical is naturally occuring or man-made tells us nothing about its toxicity.

Figure 8.4.6.1.2.7: Natural and man-made *chemicals (Image Source: **http://pbs.twimg.com/media/CTn6HlDWwAEB1Jc.png:large**, Creative Commons license)*

This page titled 8.4.6.1.2: Basic Terminology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 1.2: Basic Terminology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.2: DOSE AND DOSE RESPONSE

Learning Objectives

After completing this lesson, you will be able to:

- Define *dose* and explain its importance in determining toxicity.
- Analyze and compare dose-response curves.
- Describe the safety or toxicity of substances based on specific dose levels.

In this section...

Topics include:

8.4.6.2.1: Dose and It's Impact on Toxicity

8.4.6.2.2: The Dose Response Relationship

8.4.6.2.3: Dose Estimates of Toxic Effects

8.4.6.2.4: Determining the Safety of a Drug

8.4.6.2.5: NOAEL and LOAEL

What We've Covered

In this section, we explored the following main points:

- Dose is the amount of a substance administered; however, several parameters are required to characterize exposure to xenobiotics, including the:
  - Number of doses
  - Frequency of doses
  - Total time period of exposure
- The dose-response relationship helps establish causality, or that the chemical induced the observed effects; the threshold effect, or the lowest dose that induced effects; and the slope, or the rate at which effects increase with dose increases.
- Estimating doses for toxic effects involves:
  - Lethal Doses/Concentrations, such as LD0, LD10, and LC50, which denote doses or concentrations that are expected to lead to death in specific percentages of a population.
  - Effective Doses, such as ED50 and ED90, which denote doses that are effective in achieving a desired endpoint in specific percentages of a population.
  - Toxic Doses, such as TD0 and TD50, which denote doses that cause adverse toxic effects in specific percentages of a population.
- The Therapeutic Index (TI) compares the effective dose to the toxic dose of a drug.
- The Margin of Safety (MOS) compares the toxic dose to 1% of the population to the effective dose to 99% of the population.
- NOAEL is the **highest** dose at which there **is no** observed toxic effect.
- LOAEL is the **lowest** dose at which there **is** an observed toxic effect.

Coming Up...

In the next section, we will take a closer look at various types of toxic effects.

This page titled 8.4.6.2: Dose and Dose Response is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 2: Dose and Dose Response by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.2.1: Dose and It's Impact on Toxicity

Dose Defined

Dose by definition is the amount of a substance administered at one time. However, other parameters are needed to characterize the exposure to xenobiotics. The most important are the number of doses, frequency, and total time period of the treatment.

For example:

- 650 -3 g)." tabindex="0">mg acetaminophen (Tylenol® products) as a single dose.
- 500 -3 g)." tabindex="0">mg penicillin every 8 hours for 10 days.
- 10 -3 g)." tabindex="0">mg DDT per day for 90 days.

Substances can enter the body from either:

1. Encountering them in the environment (**exposure**).
2. Intentionally consuming or administering a certain quantity of a substance.

Environments in which xenobiotics are present include outdoor air, indoor air, and water. Xenobiotics can travel into the body through the skin, eyes, lungs, and digestive tract. **Exposure to a xenobiotic** can occur in any environment where a substance can enter the:

- Skin through dermal absorption (air and water).
- Respiratory tract through inhalation.
- Digestive tract through ingestion.

Exposure to a xenobiotic through dermal absorption

Figure 8.4.6.2.1.1: Dermal
(Image Source: ORAU, ©)

A xenobiotic breathed into the lungs

Figure 8.4.6.2.1.2: *Inhalation*
(Image Source: ORAU, ©)

Ingestion of a xenobiotic

Figure 8.4.6.2.1.3: *Ingestion*
(Image Source: ORAU, ©)

A dose can be considered either:

1. A measurement of environmental exposures.
2. The amount of a substance administered over a period of time.

Types of doses include:

- **Absorbed dose** — the amount of a substance that entered the body through the skin, eyes, lungs, or digestive tract and was taken up by organs or particular tissues. Absorbed dose can also be called **internal dose**.
- **Administered dose** — the quantity administered usually orally or by injection (note that an administered dose taken orally may not necessarily be absorbed).
- **Total dose** — the sum of all individual doses.

Not all substances that enter the body are necessarily absorbed by it. This concept applies to water intake. When a person drinks a large quantity of water at one time, some of it is absorbed while the rest of the water is eliminated.

If an individual drinks 1 liter of water every hour for 3 hours, each *administered dose* would be 1 liter. The *total dose* would be the amount the person drank over the time period that the water was consumed. The *absorbed dose, however,* would likely be less than the total dose because it would depend on how much of the water the individual's body absorbed which can be affected by various factors. The water intake example is represented in Table 1.

Doses	Administered Dose	Total Dose	Absorbed Dose
Dose 1 (8:00 a.m.): 1 L water	1 L	1 L	Less than 1 L
Dose 2 (9:00 a.m.): 1 L water	1 L	2 L	Less than 2 L
Dose 3 (10:00 a.m.): 1 L water	1 L	3 L	Less than 3 L

The terms for types of doses help account for the amount of a substance that entered the body by different means, but the amount *absorbed* is what is most important.

In a later section, we will review specifics about how the body handles substances after they enter the body.

Fractionating Doses

Fractionating a total dose usually decreases the probability that the total dose will cause toxicity. The reason is that the body often can repair the effect of each **subtoxic** dose if sufficient time elapses before the next dose is received. In that case, a total dose that would be harmful if received all at once is non-toxic when administered over a period of time. For example, 30 -3 g)." tabindex="0">mg of strychnine swallowed at one time could be fatal to an adult whereas 3 -3 g)." tabindex="0">mg of strychnine swallowed each day for 10 days is not considered a fatal dose.

The **units** used in toxicology are basically the same as those used in medicine. The **gram (g)** is the standard unit. Because most exposures are in smaller quantities, the **-3 g)." tabindex="0">milligram (-3 g)." tabindex="0">mg)** is commonly used. For example, the common adult dose of acetaminophen is 650 -3 g)." tabindex="0">mg.

Importance of Age, Body Size, and Time

A person's age and body size affect the clinical and toxic effects of a given dose. Age and body size usually are connected, particularly in children. This relationship is important because a person's body size can affect the burden that a substance has on it. For example, a 650--3 g)." tabindex="0">mg dose of acetaminophen is typical for adults but it would be toxic to young children. Therefore, a tablet of an acetaminophen product designed for children (Children's Tylenol®) contains only 80 -3 g)." tabindex="0">mg of the drug.

Stick figure humans of various ages and body sizes gathered around a clock

Figure 8.4.6.2.1.4: Age, body size, and time are key factors when considering the clinical and toxic effects of a dose (Image Source: iStock Photos, ©)

One way to compare the effectiveness of a dose and its toxicity is to assess the amount of a substance administered with respect to **body weight**. A common dose measurement is **mg/kg** which stands for -3 g)." tabindex="0">mg of substance per kg of body weight. Another method used to compare doses among different species is to use **body surface area**, rather than simply body weight.

Units

Because some xenobiotics are toxic in quantities much smaller than the -3 g)." tabindex="0">milligram, smaller fractions of the gram, such as **-6) of a gram." tabindex="0">microgram (µg)**. Table 2 shows other units.

Unit	Gram Equivalents	Exp. Form
Kilogram (kg)	1000.0 g	10^{3}g
Gram (g)	1.0 g	1 g
-3 g)." tabindex="0">Milligram (-3 g)." tabindex="0">mg)	0.001 g	10^{-3} g
-6) of a gram." tabindex="0">Microgram (µg)	0.000,001 g	10^{-6} g
-9 g)." tabindex="0">Nanogram (-9 g)." tabindex="0">ng)	0.000,000,001 g	10^{-9} g
-12 g)." tabindex="0">Picogram (-12 g)." tabindex="0">pg)	0.000,000,000,001 g	10^{-12} g
Femtogram (fg)	0.000,000,000,000,001 g	10^{-15} g

Table 8.4.6.2.1.1: *Various units, their gram equivalents, and their exponential form*

Concentration

Environmental exposure units are expressed as the amount of a xenobiotic in a unit of the media, which could be liquid, solid, or air. **Concentration** is the amount of a substance found in a certain amount of another substance, such as water, air, soil, food, blood, hair, urine, or breath. For example, the weight of a toxic substance found in a certain weight of food is indicated as a measure of concentration rather than the total amount. Knowing how concentrated the toxic substance is in a sample of food that weighs 100 g allows for easy comparison when testing for that toxic substance in other samples of food that weigh more or less than 100 g.

Figure 7 illustrates this concept. The two glasses contain samples of juice that are being tested for contamination with lead. The volume of juice in Glass A is 100 mL and the volume of juice in Glass B is 50 mL. The concentration of lead is the same in both samples of juice: 20 parts per billion (ppb). The total amount of lead would be higher in Glass A but the concentration of lead per unit volume is the same in both glasses.

Two glasses of juice, one containing 100 mL and the other containing
50 mL. Both glasses contain a concentration of 20 ppb lead.

Figure 8.4.6.2.1.5: The concentration of lead is the same in samples with different volumes, but the total amount of lead in each is not
(Image Source: Adapted from iStock Photos, ©)

Assessing Exposure

An individual's exposure to a substance can be assessed based on the relationship between the person's *body weight* and these factors:

- *Concentration* of the substance in the environmental media (for example, in µg/ml).
- Amount of the substance taken into the body.
- Duration and frequency of individual events during which the body was in contact with the environmental media.

Environmental exposure units used in toxicology include:

- -3 g)." tabindex="0">mg/liter (**-3 g)." tabindex="0">mg/L**) for liquids.
- -3 g)." tabindex="0">mg/gram (**-3 g)." tabindex="0">mg/g**) for solids.
- -3 g)." tabindex="0">mg/cubic meter (**-3 g)." tabindex="0">mg/m^3**) for air.

Smaller units are used as needed; for example, **-6) of a gram." tabindex="0">µg/mL**. Other commonly used dose units for substances in media are parts per million **(ppm)**, parts per billion **(ppb)**, and parts per trillion **(ppt)**. When smaller units are used to quantify exposure, the mg/kg/day unit can be adapted to the smaller unit. For example, parts per billion per kg per day (ppb/kg/day) could be used.

An important thing to remember is that the use of a small dose unit is *not* related to the burden a substance has on the body. An exposure unit describes only the *quantity* of the substance.

This page titled 8.4.6.2.1: Dose and It's Impact on Toxicity is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.1: Dose and It's Impact on Toxicity by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.2.2: The Dose Response Relationship

The Dose-Response Relationship

The dose-response relationship is an essential concept in toxicology. It correlates exposures with changes in body functions or health.

In general, the higher the dose, the more severe the response. The dose-response relationship is based on observed data from experimental animal, human clinical, or cell studies.

Knowledge of the dose-response relationship establishes:

- **Causality** — that the chemical has induced the observed effects.
- The **threshold effect** — the lowest dose where an induced effect occurs.
- The **slope** for the dose response — the *rate* at which injury builds up.

Within a population, the majority of responses to a toxicant are similar; however, there are differences in how responses may be encountered – some individuals are susceptible and others resistant. As demonstrated in Animation 1, a graph of the individual responses can be depicted as a bell-shaped standard distribution curve. There is a wide variance in responses as demonstrated by the mild reaction in resistant individuals, the typical response in the majority of individuals, and the severe reaction in sensitive individuals.

*Animation 1: A graph of individual responses to a substance, which generally take the form of a bell-shaped curve (**view full-text, PDF version**)*

The **dose-response curve** is a visual representation of the response rates of a population to a range of doses of a substance, as demonstrated in Animation 2 (available at ToxTutor). *The graph of a dose-response relationship typically has an* "s" *shape. (**view full-text, PDF version**)*

Knowledge Check

1. The quantity of a substance administered to an individual over a period of time or in several individual doses is known as the:

- administered dose
- absorbed dose
- total dose

Answer available at the end of this section.

2. Fractionation of a total dose so that the total amount administered is given over a period of time usually results in:

- decreased toxicity
- increased toxicity

Answer available at the end of this section.

3. The usual dosage unit that incorporates the amount of material administered or absorbed in accordance with the size of the individual over a period of time is:

- PPM/hour

- ◯ mg/kg/day
- ◯ kg/100 lb/week

Answer available at the end of this section.

4. The dose at which a toxic effect is first encountered is called the:

- ◯ median toxic dose
- ◯ absorbed dose
- ◯ threshold dose

Answer available at the end of this section.

5. The dose-response relationship helps a toxicologist determine:

- ◯ whether exposure has caused an effect
- ◯ the threshold dose
- ◯ the rate of increasing effect with increasing dose levels
- ◯ all of the above

Answer available at the end of this section.

Answers

Knowledge Check

total dose

It is the quantity of a substance administered to an individual over a period of time or in several individual doses. It is particularly important when evaluating cumulative poisons.

Knowledge Check

decreased toxicityFractionation of a total dose so that the total amount administered is given over a period of time usually results in **decreased toxicity**. This applies to most forms of toxicity but not necessarily to carcinogenicity or mutagenicity.

Knowledge Check

mg/kg/day

The usual dosage unit that incorporates the amount of material administered or absorbed in accordance with the size of the individual over a period of time is **mg/kg/day**. In some cases, much smaller dosage units, such as µg/kg/day, are used.

Knowledge Check

threshold dose

Knowledge Check

all of the above

The dose-response relationship demonstrates whether any effect has occurred, the threshold dose, and the rate at which the effect increases with increasing dose levels.

This page titled 8.4.6.2.2: The Dose Response Relationship is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.2: The Dose Response Relationship by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.2.3: Dose Estimates of Toxic Effects

Dose Estimates

Dose-response curves are used to derive dose estimates of chemical substances.

Historically, **LD50** (Lethal Dose 50%) has been a common dose estimate for acute toxicity. It is a **statistically derived maximum dose** at which 50% of the group of organisms (rat, mouse, or other species) would be expected to die. LD50 testing is no longer the recommended method for assessing toxicity because of the ethics of using large numbers of animals, the variability of responses in animals and humans, and the use of mortality as the only endpoint. Regulatory agencies use LD50 only if it is justified by scientific necessity and ethical considerations.

The Three Rs

The current practice for estimating acute toxicity emphasizes the following approaches, known as the Three Rs:

1. **Replacing** animals in science by *in vitro*, ***in silico***, and other approaches.
2. **Reducing** the number of animals used. For example, the oral LD50 approach has been replaced in some circumstances by an **up-and-down method** in which animals are dosed one at a time.
3. **Refining** care and procedures to minimize pain and distress.

Other dose estimates also may be used.

Lethal Doses/Concentrations

- **Lethal Dose 0% (LD0)** — represents the dose at which no individuals are expected to die. This is just below the threshold for lethality.
- **Lethal Dose 10% (LD10)** — refers to the dose at which 10% of the individuals will die.
- **Lethal Concentration 50% (LC50)** — for inhalation toxicity, air concentrations are used for exposure values. The LC50 refers to the calculated concentration of a gas lethal to 50% of a group. Occasionally LC0 and LC10 are also used.

Effective Doses (EDs)

Effective Doses (**EDs**) are used to indicate the effectiveness of a substance. Normally, effective dose refers to a beneficial effect such as relief of pain. It may also stand for a harmful effect such as paralysis. Thus, the specific endpoint must be indicated. The usual terms are:

Term	Effective for this percentage of the population
ED0	0%
ED10	10%
ED50	50%
ED90	90%

Table 8.4.6.2.3.1. Typical terms for describing effective doses

Toxic Doses (TDs)

Toxic Doses (**TDs**) are used to indicate doses that cause adverse toxic effects. The usual dose estimates include:

Term	Toxic to this percentage of the population
TD0	0%
TD10	10%
TD50	50%
TD90	90%

Table 8.4.6.2.3.2. Typical terms for describing toxic doses

Determining the Relative Safety of Pharmaceuticals

Toxicologists, pharmacologists, and others use **effective** and **toxic dose** levels to determine the relative safety of pharmaceuticals. As shown in Figure 1, two dose-response curves are presented for the same drug, one for effectiveness and the other for toxicity. In this case, a dose that is 50% to 75% effective does not cause toxicity. However, a 90% effective dose may result in a small amount of toxicity.

Graph of two dose-response curves for the same drug - one for effectiveness and one for toxicity. ED10, ED50, ED90, TD10, TD50, TD90 are marked on the graph. A 50-75% effective dose does not cause toxicity in this example, but a 90% effective dose could result in a small amount of toxicity.

Figure 8.4.6.2.3.1. Dose-response curves representing effective dose and toxic dose for the same drug (Image Source: NLM)

It should be noted that a desired effect in a drug is often an undesired effect with an environmental chemical.

This page titled 8.4.6.2.3: Dose Estimates of Toxic Effects is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.3: Dose Estimates of Toxic Effects by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.2.4: Determining the Safety of a Drug

Do you know?

What are measures for describing the safety of a drug? This section describes the:

- Therapeutic index
- Margin of safety

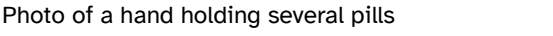


Figure 8.4.6.2.4.1: (Image Source: iStock Photos, ©)

Therapeutic Index

The **Therapeutic Index (TI)** is used to compare the therapeutically effective dose to the toxic dose of a pharmaceutical agent. The TI is a statement of relative safety of a drug. It is the ratio of the dose that produces toxicity to the dose needed to produce the desired therapeutic response. The common method used to derive the TI is to use the 50% dose-response points, including TD50 (toxic dose) and ED50 (effective dose).

Therapeutic Index, or TI, equals the toxic dose divided by the dose for therapeutic response. In other words, TI equals TD50 divided by ED50.

For example, if the TD50 is 200 and the ED50 is 20 -3 g)." tabindex="0">mg, the TI would be 10.

TI equals TD50 divided by ED50. In this example, this is 200 divided by 20, which means TI equals 10.

Figure 1).

Illustration of two different drugs on a single axis graph. The x-axis represents increasing therapeutic index from left to right. The drug with TI of 3 is more toxic and less safe than the drug with TI of 10. As the TI increases, toxicity decreases.

Figure 8.4.6.2.4.2. A higher value on the Therapeutic Index indicates a more favorable safety profile (Image Source: ORAU, ©)

However, the use of the **ED50** and **TD50** doses to derive the **TI** may be misleading about a drug's safety, depending on the slope of the dose-response curves for therapeutic and toxic effects. To overcome this deficiency, toxicologists often use another term to denote the safety of a drug: the Margin of Safety.

Margin of Safety (MOS)

The **Margin of Safety (MOS)** is usually calculated as the ratio of the toxic dose to 1% of the population (TD01) to the dose that is 99% effective to the population (ED99).

Margin of Safety, or MOS, equals TD01 divded by ED99.

Figure 2 shows the relationship between effective dose response and toxic dose response. The shaded area represents the doses at which the substance produces an effective dose response while the toxic dose response remains below the TD50. The slope of a curve shows how dose increases result in responses to the effective or toxic dose.

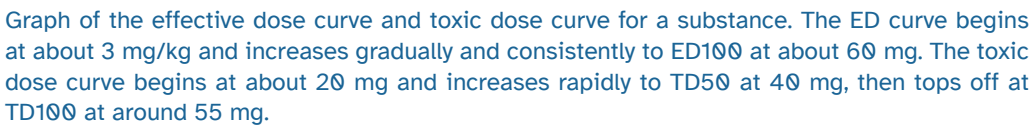


Figure 8.4.6.2.4.3. Relationship between effective dose response and toxic dose response
(Image Source: NLM)

Because of differences in slopes and threshold doses, low doses may be effective without producing toxicity. Although more patients may benefit from higher doses, that is offset by the probability that toxicity will occur.

The toxicity of various substances can be compared using the slopes for each curve (Figure 3).



Figure 8.4.6.2.4.4. Comparison of the toxicity of two substances
(Image Source: NLM)

For some substances, a small increase in dose causes a large increase in response, which is seen in Toxicant A's steep slope. For other substances, a much larger increase in dose is required to cause the same increase in response, as indicated in Toxicant B's shallow slope.

This page titled 8.4.6.2.4: Determining the Safety of a Drug is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.4: Determining the Safety of a Drug by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.2.5: NOAEL and LOAEL

NOAEL and LOAEL

Results from research studies establish the highest doses at which no toxic effects were identified and the lowest doses at which toxic or adverse effects were observed. The terms often used to describe these outcomes are:

- **No Observed Adverse Effect Level (NOAEL)**
- **Lowest Observed Adverse Effect Level (LOAEL)**

These terms refer to the actual doses used in human clinical or experimental animal studies. They are defined as follows:

- NOAEL — *Highest dose* at which there *was not* an observed toxic or adverse effect.
- LOAEL — *Lowest dose* at which there *was* an observed toxic or adverse effect.

Figure 1 shows a dose-response curve where the NOAEL occurs at 10 mg and the LOAEL occurs at 18 mg.

A dose response curve of a hypothetical example where NOAEL occurs at 10 mg and LOAEL is at about 18 mg.

Figure 8.4.6.2.5.1. A dose-response curve showing doses where the NOAEL and LOAEL occur for a substance (Image Source: NLM)

Sometimes the terms **No Observed Effect Level (NOEL)** and **Lowest Observed Effect Level (LOEL)** are also used. NOELs and LOELs do not necessarily imply toxic or harmful effects and can be used to describe beneficial effects of substances.

The NOAEL, LOAEL, NOEL, and LOEL are commonly used in risk assessments and research. For example, **this U.S. Food and Drug Administration (FDA) publication** for industry describes a process for estimating the maximum safe starting dose of drugs tested in clinical trials. It provides extensive information about these concepts and their utility when developing new drugs.

NOEALs and LOAELs are also included in the **Noncarcinogenic Risk Assessment section** where they are applied using the benchmark dose (BMD) method.

Knowledge Check

1. Which of the following is **not** one of the Three Rs of estimating acute toxicity?

- Replace animals in science by *in vitro*, *in silico*, and other approaches
- Reduce the number of animals used
- Refine care and procedures to minimize pain and distress
- Randomize the selected test animals

Answer available at the end of this section.

2.

The Therapeutic Index (TI) is used to:

- Compare the therapeutically effective dose to the toxic dose of a pharmaceutical agent
- Calculate the lethal dose level for different pharmaceuticals

◯Compare the lethal dose to the therapeutically effective dose of a pharmaceutical agent

Answer available at the end of this section.

3. The Margin of Safety (MOS) of a drug is the:

◯Amount of a pharmaceutical that can be given before toxicity first appears

◯Difference between the Effective Dose to 50% of the population (ED50) and the Toxic Dose to 50% of the population (TD50)

◯Ratio of the Toxic Dose to 1% of the population (TD01) to the Effective Dose to 99% of the population (ED99)

Answer available at the end of this section.

4. The No Observed Adverse Effect Level (NOAEL) is the:

◯*Lowest dose* at which there *was no* observed toxic or adverse affect

◯*Highest dose* at which there *was no* observed toxic or adverse effect

◯*Highest dose* at which there *was* an observed adverse effect

Answer available at the end of this section.

5. The Lowest Observed Adverse Effect Level (LOAEL) is the:

◯*Lowest dose* at which there *was* an observed toxic or adverse effect

◯*Lowest dose* at which there *was no* observed toxic or adverse effect

◯*Highest dose* at which there *was* an observed adverse effect

Answer available at the end of this section.

Answers

Knowledge Check

Randomize the selected test animalsThe Three Rs involve replacing animals in science by *in vitro*, *in silico*, and other approaches; reducing the number of animals used in testing; and refining care and procedures to minimize pain and distress.

Knowledge Check

Compare the therapeutically effective dose to the toxic dose of a pharmaceutical agent

Knowledge Check

◯Ratio of the Toxic Dose to 1% of the population (TD01) to the Effective Dose to 99% of the population (ED99)

Knowledge Check

Highest dose at which there *was no* observed toxic or adverse effect

Knowledge Check

Lowest dose at which there *was* an observed toxic or adverse effect

This page titled 8.4.6.2.5: NOAEL and LOAEL is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 2.5: NOAEL and LOAEL by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.3: TOXIC EFFECTS

Learning Objectives

After completing this lesson, you will be able to:

- Explain factors that influence the toxicity of a substance.
- Define types of systemic and organ-specific toxic effects.

In this section...

Topics include:

8.4.6.3.1: Types of Toxic Effects

8.4.6.3.2: Factors Affecting Toxicity

8.4.6.3.3: Systemic Toxic Effects

8.4.6.3.4: Organ Specific Toxic Effects

Did you know?

In December 1984, the world's worst industrial accident occurred in Bhopal, India. More than 40 tons of **methyl isocyanate** leaked from a pesticide plant, killing thousands of people and injuring hundreds of thousands. Follow-up studies have shown that the incident caused increased mortality and continued effects on health, including airway disease, eye diseases, and pregnancy losses.

The company involved in the leak tried to distance itself from the accident and prevent those affected from learning the true nature of the accident. The legal case went on for years. Eventually, families of the dead received an average of about $2,200. While the company ceased operation at its Bhopal plant after the disaster, it did not clean up the site completely. The plant continues to leak several toxic chemicals and heavy metals into local aquifers.

Survivors of the deadly industrial incident in Bhopal, India in 1984, protesting over the mishandling of the disaster

Figure 1. Survivors of the Bhopal disaster of 1984 protest over the mishandling of the disaster
Bhopal disaster protestors. [Photo]. In Encyclopædia Britannica. Retrieved from **http://www.britannica.com/event/Bhopal-disaster/images-videos/Survivors-of-the-1984-deadly-industrial-accident-in-Bhopal-India/192038**

What We've Covered

This section made the following main points:

- Toxicity can result from adverse cellular, biochemical, or macromolecular changes.
- Some chemicals affect only specific target organs; others can damage any cell or tissue they contact.
- Chemicals can affect organisms by multiple mechanisms and at the molecular level, leading to modern approaches such as Adverse Outcome Pathways (AOPs) and Mechanism of Actions (MOAs).
- Several factors influence toxicity, including form and innate chemical activity, dosage, exposure route, species, life stage, gender, absorption ability, metabolism, distribution, excretion, health and nutritional status, the presence of other chemicals, and circadian rhythms.
- Systemic toxic effects, which can occur at multiple sites, include:
    - Acute toxicity, which occurs almost immediately (seconds/minutes) after a single dose or series of doses within 24

hours.
    - Subchronic toxicity, which results from repeated exposure for several weeks or months.
    - Chronic toxicity, which damages specific organ systems over the course of many months or years.
    - Carcinogenicity, or abnormal cell growth and differentiation that can lead to cancer.
    - Developmental toxicity, which adversely affects the developing embryo or fetus.
    - Genetic toxicity, caused by damage to DNA and altered genetic expressions.
- Organ specific toxic effects include:
    - Blood/cardiovascular toxicity, affecting the blood, bone marrow, or heart.
    - Dermal toxicity, impacting the skin.
    - Epigenetic alterations, changing genetic programming.
    - Optical toxicity, adversely affecting the eyes.
    - Hepatotoxicity, impacting the liver, bile duct, or gall bladder.
    - Immunotoxicity, affecting the immune system.
    - Nephrotoxicity, affecting the kidneys.
    - Neurotoxicity, impacting the central nervous system.
    - Reproductive toxicity, damaging the reproductive system.
    - Respiratory toxicity, affecting the respiratory system.

Coming Up...

In the next section, we will look at effects that can occur when two or more chemicals interact.

This page titled 8.4.6.3: Toxic Effects is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 3: Toxic Effects by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.3.1: Types of Toxic Effects

Types of Toxic Effects

Many factors play a potential role in toxicity. The dosage (or amount of exposure) is the most important factor. A well-known saying, **"the dose makes the poison"** speaks to this principle.

Toxicity can result from adverse cellular, biochemical, or macromolecular changes. Some examples are noted below.

Many chemicals distribute in the body and often affect only specific **target organs**. However, other chemicals can damage any cell or tissue that they contact. The target organs that are affected may vary depending on dosage and route of exposure. For example, the central nervous system may be the target organ for toxicity from a chemical after acute exposure whereas the liver may be affected after chronic exposures.

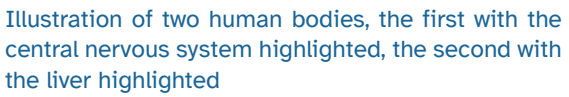


Figure 8.4.6.3.1.1. Central nervous system (left); Liver (right)
(Image Source: iStock Photos, ©)

Chemicals can cause many types of toxicity by a variety of mechanisms. Some act locally such as when direct exposure triggers skin or eye irritation, whereas other chemical cause systemic effects in the body in sites remote from where the actual exposure occurred. Toxicity can act directly affect subcellular components, such as cell receptors, or it can cause problems at the cellular level, such as with exposures to caustic or corrosive substances.

For example, chemicals might:

- Themselves be toxic or require metabolism (chemical change within the body) before they cause toxicity.
- Cause damage leading to fibrosis as the body attempts to repair the toxicity.
- Damage or disrupt an enzyme system or protein synthesis.
- Produce reactive chemicals in cells.
- Cause changes in hormone signaling or other effects.
- Produce DNA damage or epigenetic changes.

Some chemicals may also act indirectly by:

- Modifying an essential biochemical function.
- Interfering with nutrition.
- Altering a physiological mechanism.

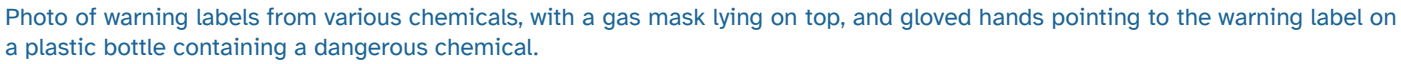


Figure 8.4.6.3.1.2. Chemicals can have a wide range of toxic effects
(Image Source: iStock Photos, ©)

Did you know?

Mercury is a naturally occurring heavy metal. **Methylmercury** , the most common organic mercury compound, can be formed in water and soil by bacteria. It builds up in the tissues of fish. Exposure to high levels of mercury and mercury compounds can cause death or permanently damage the brain and kidneys.

In the late 1950s, people living around Japan's Minamata Bay developed symptoms of severe methylmercury poisoning, some of whom died. Children exposed *in utero* were born with disabilities. Investigations showed that heavily contaminated sludge from

a factory had been released into the bay, contaminating fish and shellfish. People who ate the fish and shellfish became ill. The events led to a better understanding of industrial pollution and how heavy metals can accumulate in systems.

In January 2013 the **Minamata Convention on Mercury** global treaty was agreed to by an intergovernmental committee. It seeks to protect human health and the environment from the adverse effects of mercury.

Photo of raw fish

Figure 8.4.6.3.1.3. Fish and shellfish became contaminated with mercury, causing severe methylmercury poisoning in people and animals who consumed them
(Image Source: iStock Photos, ©)

Because chemicals can affect organisms by different mechanisms and at the molecular level, there are new ways to conduct toxicity testing.

An emerging approach is to use **Adverse Outcome Pathways (AOPs)**, which evaluate changes in normal cellular pathways. AOPs reflect the move away from high-dose studies in laboratory animals for toxicity testing to *in vitro* methods that evaluate changes in normal cellular pathways using human-relevant cells or tissues.

Other terms that describe changes resulting from the exposure of a living organism to a substance include mode of action (MoA) and mechanism of action (MOA).

- **Mode of action (MoA)** (older term) — describes a functional or anatomical change at the *cellular* level.
- **Mechanism of action (MOA)** — describes such changes at the *molecular* level.

More information about toxicity testing can be found in the **Hazard Identification** section.

This page titled 8.4.6.3.1: Types of Toxic Effects is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.1: Types of Toxic Effects by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.3.2: Factors Affecting Toxicity

Factors Influencing Toxicity

In some instances, individuals can have unpredictable reactions, or **idiosyncratic responses**, to a drug or other substance. An idiosyncratic response is uncommon, and it is sometimes impossible to understand whether it is the result of a genetic predisposition or has some other cause such as the status of the immune system. It could result in an abnormally small or short, or abnormally large or long response to the drug or other substance. Or, the response could be qualitatively different than what has been observed in most other individuals.

The toxicity of a substance usually depends on the following factors:

- Form and innate chemical activity
- Dosage, especially dose-time relationship
- Exposure route
- Species
- Life stage, such as infant, young adult, or elderly adult
- Gender
- Ability to be absorbed
- Metabolism
- Distribution within the body
- Excretion
- Health of the individual, including organ function and pregnancy, which involves physiological changes that could influence toxicity
- Nutritional status
- Presence of other chemicals
- Circadian rhythms (the time of day a drug or other substance is administered)

Factors Related to the Substance

Form and Innate Chemical Activity

The **form** of a substance may have a profound impact on its toxicity especially for metallic elements, also termed heavy metals. For example, the toxicity of mercury vapor differs greatly from methyl mercury. Another example is chromium. Cr^{3+} is relatively nontoxic whereas Cr^{6+} causes skin or nasal corrosion and lung cancer.

The **innate chemical activity** of substances also varies greatly. Some can quickly damage cells causing immediate cell death. Others slowly interfere only with a cell's function. For example:

- Hydrogen cyanide binds to the enzyme cytochrome oxidase resulting in cellular hypoxia and rapid death.
- Nicotine binds to cholinergic receptors in the central nervous system (CNS) altering nerve conduction and inducing gradual onset of paralysis.

Dosage

The **dosage** is the most important and critical factor in determining if a substance will be an *acute* or a *chronic* toxicant. Virtually all chemicals can be acute toxicants if sufficiently large doses are administered. Often the toxic mechanisms and target organs are different for acute and chronic toxicity. Examples are:

Toxicant	Acute Toxicity	Chronic Toxic Effects
Ethanol	CNS depression	Liver cirrhosis

Arsenic	Gastrointestinal damage	Skin/liver cancer

Table 8.4.6.3.2.1. Examples of acute and chronic toxicity

Exposure Route

The way an individual comes in contact with a toxic substance, or **exposure route**, is important in determining toxicity. Some chemicals may be highly toxic by one route but not by others. Two major reasons are differences in absorption and distribution within the body. For example:

- Ingested chemicals, when absorbed from the intestine, distribute first to the liver and may be immediately detoxified.
- Inhaled toxicants immediately enter the general blood circulation and can distribute throughout the body prior to being detoxified by the liver.

Different target organs often are affected by different routes of exposure.

Diagram of toxic substances being ingested

Figure 8.4.6.3.2.1. Ingestion
(Image Source: ORAU, ©)

Diagram of toxic substances being breathed into the lungs

Figure 8.4.6.3.2.2. Inhalation
(Image Source: ORAU, ©)

Absorption

The **ability to be absorbed** is essential to systemic toxicity. Some chemicals are readily absorbed and others are poorly absorbed. For example, nearly all alcohols are readily absorbed when ingested, whereas there is virtually no absorption for most polymers. The rates and extent of absorption may vary greatly depending on the form of a chemical and the route of exposure to it. For example:

- Ethanol is readily absorbed from the gastrointestinal tract but poorly absorbed through the skin.
- Organic mercury is readily absorbed from the gastrointestinal tract; inorganic lead sulfate is not.

Factors Related to the Organism

Species

Toxic responses can vary substantially depending on the species. Most differences between species are attributable to differences in metabolism. Others may be due to anatomical or physiological differences. For example, rats cannot vomit and expel toxicants before they are absorbed or cause severe irritation, whereas humans and dogs are capable of vomiting.

Selective toxicity refers to species differences in toxicity between two species simultaneously exposed. This is the basis for the effectiveness of pesticides and drugs. For example:

- An insecticide is lethal to insects but relatively nontoxic to animals.
- Antibiotics are selectively toxic to microorganisms while virtually nontoxic to humans.

Life Stage

An individual's age or **life stage** may be important in determining his or her response to toxicants. Some chemicals are more toxic to infants or the elderly than to young adults. For example:

- Parathion is more toxic to young animals.

- Nitrosamines are more carcinogenic to newborn or young animals.

Illustrated silhouette people of varying ages and sizes standing outdoors under a tree

Figure 8.4.6.3.2.3. An individual's life stage can impact that person's response to toxicants
(Image Source: iStock Photos, ©)

Gender

Gender can play a big role in influencing toxicity. Physiologic differences between men and women, including differences in pharmacokinetics and pharmacodynamics, can affect drug activity.

In comparison with men, pharmacokinetics in women generally can be impacted by their lower body weight, slower gastrointestinal motility, reduced intestinal enzymatic activity, and slower kidney function (glomerular filtration rate). Delayed gastric emptying in women may result in a need for them to extend the interval between eating and taking medications that require absorption on an empty stomach. Other physiologic differences between men and women also exist. Slower renal clearance in women, for example, may result in a need for dosage adjustment for drugs such as digoxin that are excreted via the kidneys.

In general, pharmacodynamic differences between women and men include greater sensitivity to and enhanced effectiveness, in women, of some drugs, such as beta blockers, opioids, and some antipsychotics.

Studies in animals also have identified gender-related differences. For example:

- Male rats are 10 times more sensitive than females to liver damage from DDT.
- Female rats are twice as sensitive to parathion as are male rats.

Male and female gender symbols

Figure 8.4.6.3.2.4. *Gender symbols for female (left) and male (right)*
(Image Source: iStock Photos, ©)

Metabolism

Metabolism, also known as biotransformation, is the conversion of a chemical from one form to another by a biological organism. Metabolism is a major factor in determining toxicity. The products of metabolism are known as metabolites. There are two types of metabolism:

1. Detoxification
2. Bioactivation

In **detoxification**, a xenobiotic is converted to a less toxic form. This is a natural defense mechanism of the organism. Generally, detoxification converts lipid-soluble compounds to polar compounds.

In **bioactivation**, a xenobiotic may be converted to more reactive or toxic forms. Cytochrome P-450 (CYP450) is an example of an enzyme pathway used to metabolize drugs. In the elderly, CYP450 metabolism of drugs such as phenytoin and carbamazepine may be decreased. Therefore, the effect of those drugs may be less pronounced. CYP450 metabolism also can be inhibited by many drugs. Risk of toxicity may be increased if a CYP450 enzyme-inhibiting drug is given with one that depends on that pathway for metabolism.

There is awareness that the gut microbiota can impact the toxicity of drugs and other chemicals. For example, gut microbes can metabolize some environmental chemicals and bacteria-dependent metabolism of some chemicals can modulate their toxicity. Also, environmental chemicals can alter the composition and/or the metabolic activity of the gastrointestinal bacteria, thus contributing in a meaningful way to shape an individual's microbiome. The study of the consequences of these changes is an emerging area of toxicology.

Learn more about human exposure to pollutants and their interaction with the GI microbiota.

Learn more about the microbiome and toxicology.

Distribution Within the Body

The **distribution** of toxicants and toxic metabolites throughout the body ultimately determines the sites where toxicity occurs. A major determinant of whether a toxicant will damage cells is its lipid solubility. If a toxicant is lipid-soluble, it readily penetrates cell membranes. Many toxicants are stored in the body. Fat tissue, liver, kidney, and bone are the most common storage sites. Blood serves as the main avenue for distribution. Lymph also distributes some materials.

Excretion

The site and rate of **excretion** is another major factor affecting the toxicity of a xenobiotic. The kidney is the primary excretory organ, followed by the gastrointestinal tract, and the lungs (for gases). Xenobiotics may also be excreted in sweat, tears, and milk.

A large volume of blood serum is filtered through the kidney. Lipid-soluble toxicants are reabsorbed and concentrated in kidney cells. Impaired kidney function causes slower elimination of toxicants and increases their toxic potential.

Health Status

The health of an individual or organism can play a major role in determining the levels and types of potential toxicity. For example, an individual may have pre-existing kidney or liver disease. Certain conditions, such as pregnancy, also are associated with physiological changes in kidney function that could influence toxicity.

Nutritional Status

Diet (nutritional status) can be a major factor in determining who does or does not develop toxicity. For example:

- Consumption of fish that have absorbed mercury from contaminated water can result in mercury toxicity; an antagonist for mercury toxicity is the nutrient selenium.
- Some vegetables can accumulate cadmium from contaminated soil; an antagonist for cadmium toxicity is the nutrient zinc.
- Grapefruit contains a substance that inhibits the P450 drug detoxification pathway, making some drugs more toxic.

Find out more about nutrition and chemical toxicity **here**.

Circadian Rhythms

Circadian rhythms can play a role in toxicity. For example, rats administered an immunosuppressive drug had severe toxicity in their intestines 7 hours after light onset compared to controls and to other times in the day. The rats had changes in their digestive enzyme activity and other physiological indicators at this dosing time.

Find out more about circadian rhythm and gut toxicity **here**.

Other Factors

Presence of Other Chemicals

The presence of other chemicals, at the same time, earlier, or later may:

- **Decrease toxicity** (antagonism)
- **Add to toxicity** (additivity)
- **Increase toxicity** (synergism or potentiation)

For example:

- Antidotes used to counteract the effects of poisons function through antagonism (atropine counteracts poisoning by

organophosphate insecticides).
- Alcohol may enhance the effect of many antihistamines and sedatives.
- A synergistic interaction between the antioxidant butylated hydroxytoluene (BHT) and a certain concentration of oxygen results in lung damage in the form of interstitial pulmonary fibrosis.

Information on additional examples of lung damage from chemical interactions can be found **here**.

Knowledge Check

1. A target organ is an organ that:

- Absorbs a toxic substance
- Stores an absorbed substance or its metabolite
- Is damaged by a toxic substance

Answer available at the end of this section.

2. What are the important factors that influence the degree of toxicity of a substance?

- Innate chemical activity, form, dosage, and exposure route
- The species, life stage, gender, health status, nutritional status, and circadian rhythms of the organism
- Absorption, metabolism, distribution within the body, excretion, and presence of other chemicals
- All of the above

Answer available at the end of this section.

3. Metabolism, or biotransformation, of a xenobiotic:

- Always results in reduced toxicity of the xenobiotic
- May result in detoxification or bioactivation
- Has no influence on the toxicity of the xenobiotic

Answer available at the end of this section.

4. An antibiotic administered to humans kills bacteria in the body but does not harm human tissues. This is an example of:

- Selective toxicity
- Acute toxicity
- Varying absorption of the antibiotic

Answer available at the end of this section.

5. A major determinant of whether a toxicant will damage cells is its:

- Acidity
- Biotransformation

- Lipid solubility

Answer available at the end of this section.

Answers

Knowledge Check

Is damaged by a toxic substanceA target organ is an organ in which a substance exerts a toxic effect.

Knowledge Check

All of the above

Knowledge Check

May result in detoxification or bioactivation
Metabolism of a xenobiotic results in either detoxification, which converts the xenobiotic to a less toxic form, or bioactivation, which converts the xenobiotic to more reactive or toxic forms. For example, a xenobiotic itself might not be carcinogenic, but a metabolite of the xenobiotic might be.

Knowledge Check

Selective Toxicity

Selective toxicity refers to differences in toxicity between two species simultaneously exposed, much like the antibiotic in this example.

Knowledge Check

Lipid solubilityA major determinant of whether or not a toxicant will damage cells is its lipid solubility. If a toxicant is lipid-soluble, it readily penetrates cell membranes.

This page titled 8.4.6.3.2: Factors Affecting Toxicity is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.2: Factors Affecting Toxicity by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.3.3: Systemic Toxic Effects

Types of Systemic Toxic Effects

Toxic effects are generally categorized according to the site of the toxic effect. In some cases, the effect may occur at only one site. This site is termed the **specific target organ**.

In other cases, toxic effects may occur at multiple sites. This is known as **systemic toxicity**. Types of systemic toxicity include:

- Acute Toxicity
- Subchronic Toxicity
- Chronic Toxicity
- Carcinogenicity
- Developmental Toxicity
- Genetic Toxicity (somatic cells)

Acute Toxicity

Acute toxicity occurs almost immediately (seconds/minutes/hours/days) after an exposure. An **acute exposure** is usually a single dose or a series of doses received within a 24-hour period. Death can be a major concern in cases of acute exposures. For example:

- In 1989, 5,000 people died and 30,000 were permanently disabled due to exposure to methyl isocyanate from an industrial accident in India.
- Many people die each year from inhaling carbon monoxide from faulty heaters.

Close-up photo of the burner on a gas heater

Figure 8.4.6.3.3.1. Faulty gas heaters can emit toxic carbon monoxide
(Image Source: iStock Photos, ©)

Subchronic Toxicity

Subchronic toxicity results from repeated exposure for several weeks or months. This is a common human exposure pattern for some pharmaceuticals and environmental agents. For example:

- Ingestion of warfarin (Coumadin®) tablets (blood thinners) for several weeks as a treatment for venous thrombosis can cause internal bleeding.
- Workplace exposure to lead over a period of several weeks can result in anemia.

Photo of an opened box of Warfarin Tablets (left); photo of old, rusty pipes containing lead or lead soldering (right)

Figure 8.4.6.3.3.2. Warfarin Tablets (left); old lead pipes (right)
(Image Source: iStock Photos, ©)

Chronic Toxicity

Chronic toxicity represents cumulative damage to specific organ systems and takes many months or years to become a recognizable clinical disease. Damage due to subclinical individual exposures may go unnoticed. With repeated exposures or long-term continual exposure, the damage from this type of exposure slowly builds up (cumulative damage) until the damage exceeds the threshold for chronic toxicity. Ultimately, the damage becomes so severe that the organ can no longer function normally and a variety of chronic toxic effects may result.

Chronic toxic effects include:

- Cirrhosis in alcoholics who have ingested ethanol for several years.
- Chronic kidney disease in workmen with several years of exposure to lead.
- Chronic bronchitis in long-term cigarette smokers.
- Pulmonary fibrosis in coal miners (black lung disease).

Photo of cigarettes and liquor

Figure 8.4.6.3.3.3. Smoking cigarettes and/or drinking alcohol over a long period of time can lead to chronic toxicity (Image Source: iStock Photos, ©)

Carcinogenicity

Carcinogenicity is a **complex multistage process** of abnormal cell growth and differentiation that can lead to cancer. The two stages of carcinogenicity are:

1. **Initiation** — a normal cell undergoes irreversible changes.
2. **Promotion** — initiated cells are stimulated to progress to cancer.

Chemicals can act as **initiators** or **promoters**.

The initial transformation that causes normal cells to undergo irreversible changes results from the mutation of the cellular genes that control normal cell functions. The mutation may lead to abnormal cell growth. It may involve a loss of suppresser genes that usually restrict abnormal cell growth. Many other factors are involved, such as growth factors, immune suppression, and hormones.

A **tumor (neoplasm)** is simply an uncontrolled growth of cells:

- **Benign tumors** grow at the site of origin; do not invade adjacent tissues or metastasize; and generally are treatable.
- **Malignant tumors (cancer)** invade adjacent tissues or migrate to distant sites **(metastasis)**. They are more difficult to treat and often cause death.

Developmental Toxicity

Developmental toxicity pertains to adverse toxic effects to the developing embryo or fetus. It can result from toxicant exposure to either parent before conception or to the mother and her developing embryo or fetus. The three basic types of developmental toxicity are:

1. **Embryolethality** — failure to conceive, spontaneous abortion, or stillbirth.
2. **Embryotoxicity** — growth retardation or delayed growth of specific organ systems.
3. **Teratogenicity** — irreversible conditions that leave permanent birth defects in live offspring, such as cleft palette or missing limbs.

Chemicals cause developmental toxicity in two ways:

1. They act directly on cells of the embryo, causing cell death or cell damage, leading to abnormal organ development.
2. They induce a mutation in a parent's germ cell, which is transmitted to the fertilized ovum. Some mutated fertilized ova develop into abnormal embryos.

Ultrasound images of a developing fetus with a background image of heart rate graphs

Figure 8.4.6.3.3.4. Ultrasound images of a developing fetus (Image Source: iStock Photos, ©)

Genetic Toxicity

Genetic toxicity results from damage to DNA and altered genetic expression. This process is known as **mutagenesis**. The genetic change is referred to as a **mutation** and the agent causing the change is called a **mutagen**. There are three types of genetic changes:

1. **Gene mutation** — change in DNA sequence within a gene.
2. **Chromosome aberration** — changes in the chromosome structure.
3. **Aneuploidy** or **polyploidy** — increase or decrease in number of chromosomes.

If the mutation occurs in a germ cell, the effect is **heritable**. This means there is no effect on the exposed person; rather, the effect is passed on to future generations.

If the mutation occurs in a **somatic** cell, it can cause altered cell growth (for example, cancer) or cell death (for example, teratogenesis) in the exposed person.

Photo of a lab technician holding a vile, with a computer monitor displaying various genetic sequences in the background

Figure 8.4.6.3.3.5. Genetic toxicity results from damage to DNA and altered genetic expression
(Image Source: iStock Photos, ©)

This page titled 8.4.6.3.3: Systemic Toxic Effects is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.3: Systemic Toxic Effects by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.3.4: Organ Specific Toxic Effects

Organ Specific Toxic Effects

Toxic effects that pertain to specific organs and organ systems include:

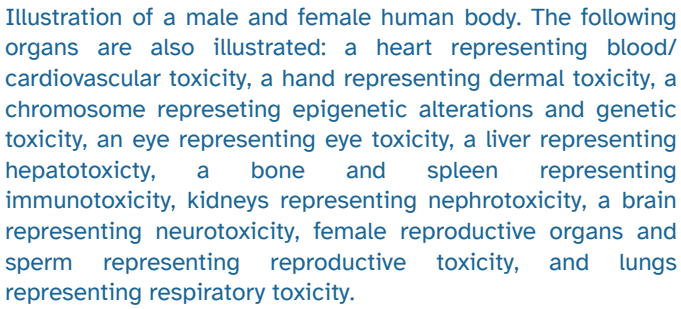


Figure 8.4.6.3.4.1. Organ-specific toxic effects pertain to specific organs and organ systems (Image Source: Adapted from iStock Photos, ©)

Blood and Cardiovascular/Cardiac Toxicity

Blood and Cardiovascular/Cardiac Toxicity results from xenobiotics acting directly on cells in circulating blood, bone marrow, and the heart. Examples of blood and cardiovascular/cardiac toxicity are:

- Hypoxia due to carbon monoxide binding of hemoglobin preventing transport of oxygen.
- Decrease in circulating leukocytes due to chloramphenicol damage to bone marrow cells.
- Leukemia due to benzene damage of bone marrow cells.
- Arteriosclerosis due to cholesterol accumulation in arteries and veins.
- Death of normal cells in and around the heart as a result of exposure to **drugs used to treat cancer**.

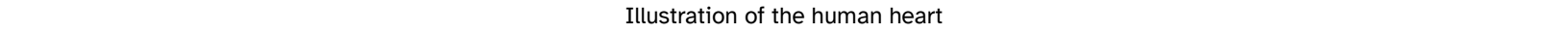


Figure 8.4.6.3.4.2. Heart cells can be damaged by exposure to certain drugs (Image Source: iStock Photos, ©)

Dermal Toxicity

Dermal Toxicity can occur when a toxicant comes into direct contact with the skin or is distributed to it internally. Effects range from mild irritation to severe changes, such as irreversible damage, hypersensitivity, and skin cancer. Examples of dermal toxicity include:

- Dermal irritation from skin exposure to gasoline.
- Dermal corrosion from skin exposure to sodium hydroxide (lye).
- Dermal itching, irritation, and sometimes painful **rash** from poison ivy, caused by **urushiol**.
- Skin cancer due to ingestion of arsenic or skin exposure to UV light.

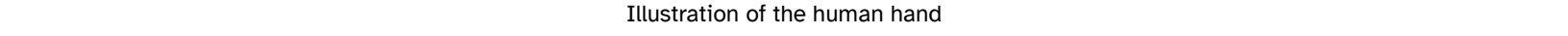


Figure 8.4.6.3.4.3. Hand (Image Source: iStock Photos, ©)

Epigenetic Alterations

Epigenetics is an emerging area in toxicology. In the field of genetics, epigenetics involves studying how external or environmental factors can switch genes on and off and change the programming of cells.

More specifically, epigenetics refers to stable changes in the programming of gene expression which can alter the phenotype without changing the DNA sequence (genotype). Epigenetic modifications include DNA methylation, covalent modifications of histone tails, and regulation by non-coding RNAs, among others.

Toxicants are examples of factors that can alter genetic programming.

Illustration of a human chromosome

Figure 8.4.6.3.4.4. Chromosome, which contains DNA (Image Source: iStock Photos, ©)

In the past, toxicology studies have assessed toxicity without measuring its impact at the level where gene expression occurs. Exogenous agents could cause long-term toxicity that continues after the initial exposure has disappeared, and such toxicities remain undetected by current screening methods. Thus, a current challenge in toxicology is to **develop screening methods that would detect epigenetic alterations** caused by toxicants.

Research is being done to assess epigenetic changes caused by toxicants. For example, the **National Institutes of Health (NIH) National Institute of Environmental Health Sciences (NIEHS) Environmental Epigenetics program** provides funding for a variety of research projects that use state-of-the-art technologies to analyze epigenetic changes caused by environmental exposures. NIEHS-supported researchers use animals, cell cultures, and human tissue samples to pinpoint how epigenetic changes can lead to harmful health effects and can potentially be passed down to the next generation.

Eye Toxicity

Eye Toxicity results from direct contact with or internal distribution to the eye. Because the cornea and conjunctiva are directly exposed to toxicants, conjunctivitis and corneal erosion may be observed following occupational exposure to chemicals. Many household items can cause conjunctivitis. Chemicals in the circulatory system can distribute to the eye and cause corneal opacity, cataracts, and retinal and optic nerve damage. For example:

- Acids and strong alkalis may cause severe corneal corrosion.
- Corticosteroids may cause cataracts.
- Methanol (wood alcohol) may damage the optic nerve.

Illustration of a human eye

Figure 8.4.6.3.4.5. *Eye (Image Source: iStock Photos, ©)*

Hepatotoxicity

Hepatotoxicity is toxicity to the liver, bile duct, and gall bladder. Because of its extensive blood supply and significant role in metabolism, the liver is particularly susceptible to xenobiotics Thus, it is exposed to high doses of the toxicant or its toxic metabolites. The primary forms of hepatotoxicity are:

- **Steatosis** — lipid accumulation in the hepatocytes.
- **Chemical hepatitis** — inflammation of the liver.
- **Hepatic necrosis** — death of the hepatocytes.
- **Intrahepatic cholestasis** — backup of bile salts into the liver cells.
- **Hepatic cancer** — cancer of the liver.
- **Cirrhosis** — chronic fibrosis, often due to alcohol.
- **Hypersensitivity** — immune reaction resulting in hepatic necrosis.

Illustration of a human liver

Figure 8.4.6.3.4.6. *Liver (Image Source: iStock Photos, ©)*

Related Resource: LiverTox®

Immunotoxicity

Immunotoxicity is toxicity of the immune system. It can take several forms:

- Hypersensitivity (allergy and autoimmunity)
- Immunodeficiency
- Uncontrolled proliferation (leukemia and lymphoma)

The normal function of the immune system is to recognize and defend against foreign invaders. This is accomplished by production of cells that engulf and destroy the invaders or by antibodies that inactivate foreign material. Examples include:

- Contact dermatitis due to exposure to poison ivy.
- Systemic lupus erythematosus ("lupus") in workers exposed to hydrazine.
- Immunosuppression by cocaine.
- Leukemia induced by benzene.

Illustration of a human chromosome

Figure 8.4.6.3.4.7. *Bone (which contains bone marrow) and spleen, both components of the immune system, which recognizes and defends against foreign invaders*
(Image Source: iStock Photos, ©)

Nephrotoxicity

The kidney is highly susceptible to toxicants because a high volume of blood flows through the organ and it filters large amounts of toxins which can concentrate in the kidney tubules.

Nephrotoxicity is toxicity to the kidneys. It can result in systemic toxicity causing:

- Decreased ability to excrete body wastes.
- Inability to maintain body fluid and electrolyte balance.
- Decreased synthesis of essential hormones (for example, erythropoietin, which increases the rate of blood cell production).

Illustration of human kidneys

Figure 8.4.6.3.4.8. *Kidneys (Image Source: iStock Photos, ©)*

Neurotoxicity

Neurotoxicity represents toxicant damage to cells of the central nervous system (brain and spinal cord) and the peripheral nervous system (nerves outside the CNS). The primary types of neurotoxicity are:

- Neuronopathies (neuron injury)
- Axonopathies (axon injury)
- Demyelination (loss of axon insulation)
- Interference with neurotransmission

Illustration of a human brain and synapse

Figure 8.4.6.3.4.9. *Brain and synapse are susceptible to toxicant damage (Image Source: iStock Photos, ©)*

Reproductive Toxicity

Reproductive Toxicity involves toxicant damage to either the male or female reproductive system. Toxic effects may cause:

- Decreased libido and impotence.
- Infertility.
- Interrupted pregnancy (abortion, fetal death, or premature delivery).
- Infant death or childhood morbidity.
- Altered sex ratio and multiple births.
- Chromosome abnormalities and birth defects.
- Childhood cancer.

Illustration of a human chromosome

Figure 8.4.6.3.4.10. *Female reproductive organs (left); male and female germ cells (right) (Image Source: iStock Photos, ©)*

Respiratory Toxicity

Respiratory Toxicity relates to effects on the upper respiratory system (nose, pharynx, larynx, and trachea) and the lower respiratory system (bronchi, bronchioles, and lung alveoli). The primary types of respiratory toxicity are:

- Pulmonary irritation
- Asthma/bronchitis
- Reactive airway disease
- Emphysema
- Allergic alveolitis
- Fibrotic lung disease
- Pneumoconiosis
- Lung cancer

Illustration of a human chromosome

Figure 8.4.6.3.4.11. *Lungs (Image Source: iStock Photos, ©)*

Knowledge Check

Knowledge Check

1. Toxic effects are primarily categorized into two general types:

◯ Systemic or organ-specific effects

◯ Carcinogenic or teratogenic effects

◯ Hepatic or nephrotoxic effects

Answer available at the end of this section.

2. What is the main difference between acute and chronic toxicity?

◯ Different organs are involved

◯ Acute toxicity occurs only after a single dose, whereas chronic toxicity occurs with multiple doses

◯ Acute toxicity appears within hours or days of an exposure, whereas chronic toxicity takes many months or years to become a recognizable clinical disease

◯ Acute toxicity is less likely to lead to death than is chronic toxicity

Answer available at the end of this section.

3. Police respond to a 911 call in which two people are found dead in an enclosed bedroom heated by an unvented kerosene stove. There was no sign of trauma or violence. A likely cause of death is:

◯ Excess oxygen generated by the combustion of kerosene

◯ Acute toxicity due to uncombusted kerosene fumes

◯ Acute toxicity due to carbon monoxide poisoning

Answer available at the end of this section.

4. Genetic toxicity can result in:

- ◯ Gene mutation
- ◯ Changes in the structure and/or number of chromosomes
- ◯ Epigenetic alterations
- ◯ All of the above

Answer available at the end of this section.

Answers

Knowledge Check

Systemic or organ-specific effects - **This is the correct answer.**
Toxic effects are broadly categorized as either systemic or organ- specific effects.

Knowledge Check

Acute toxicity appears within hours or days of an exposure, whereas chronic toxicity takes many months or years to become a recognizable clinical disease - **This is the correct answer.**

Knowledge Check

Acute toxicity due to carbon monoxide poisoning - **This is the correct answer.**
The victims most likely died as a result of acute toxicity from exposure to carbon monoxide.

Knowledge Check

All of the above - **This is the correct answer.**
Genetic toxicity can cause gene mutations, changes in chromosome structure (aberration), increases or decreases in the number of chromosomes (aneuploidy or polyploidy), and changes to genetic programming (epigenetic alterations).

This page titled 8.4.6.3.4: Organ Specific Toxic Effects is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 3.4: Organ Specific Toxic Effects by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.4: INTERACTIONS

Learning Objectives

After completing this lesson, you will be able to:

- Explain the impacts that can be experienced when two or more chemicals interact.
- List realistic examples of chemical interactions.

In this section...

Topics include:

8.4.6.4.1: Interactions

Did you know?

Gasoline is a volatile, complex mixture of hydrocarbon compounds. The mixture is easily vaporized during handling in normal conditions. People are exposed to this complex substance during refueling at service stations. **More information** is available on consumer exposure to gasoline.

In this section, we will look into the effects of interactions among such chemicals.

Figure 1. Refueling car
(Image Source: iStock Photos, ©)

What We've Covered

In this section, we explored the following key points:

- Interactions between multiple chemicals can:
  - Decrease toxicity (antagonism).
  - Add to toxicity (additivity).
  - Increase toxicity (synergism or potentiation).
- Interactions can occur by simultaneous exposure or if exposure to the agents is separated by time.
- People are normally exposed to many chemicals and combinations of chemicals every day.
- Emerging approaches in assessing interactions include:
  - Adverse outcome pathways (AOPs).
  - *In vitro* methods.
  - "Omics" techniques.
  - *In silico* approaches.

Coming Up...

In the next section, we will explore various methods for testing toxicity.

This page titled 8.4.6.4: Interactions is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 4: Interactions by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.4.1: Interactions

Interactions

As noted in "**Factors Influencing Toxicity**," the presence of other chemicals, at the same time, earlier, or later may:

- Decrease toxicity (**antagonism**).
- Add to toxicity (**additivity**).
- Increase toxicity (**synergism** or **potentiation**) of some chemicals.

For example, interactions between two or more toxic agents can produce lung damage by interactions:

- Between chemicals.
- Between chemicals and receptors.
- In which a first agent modifies the cell and tissue response to a second agent.

Interactions may occur by:

- Simultaneous exposure to two or more agents.
- Exposure to two or more agents at different times.

Lab worker pouring a red chemical into a yellow chemical

Figure 8.4.6.4.1.1. Interactions between two or more substances often impact toxicity
(Image Source: iStock Photos, ©)

Sources of Interactions

Humans are normally exposed to many chemicals at one time. For example, the use of consumer products, medical treatments, and exposures from the diet and environment (such as from soil, air, and water) can consist of exposures to hundreds, if not thousands, of chemicals. Other examples include:

- Hospital patients receive an average of six drugs daily.
- Consumers may use five or more consumer products before breakfast (for example, soap, shampoo, conditioner, toothpaste, and deodorant).
- Home influenza treatment consists of aspirin, antihistamines, and cough syrup taken simultaneously.
- Drinking water may contain small amounts of pesticides, heavy metals, solvents, and other organic chemicals.
- Air often contains mixtures of hundreds of chemicals such as automobile exhaust and cigarette smoke.
- Gasoline vapor at service stations is a mixture of 40-50 chemicals.

Photo of various personal care products, unlabeled; including hand soap, mouthwash, shampoo, body wash, toothpaste, cotton swabs, cotton balls, and a towel

Figure 8.4.6.4.1.2. The use of personal care products can result in expsoures to hundreds of chemicals
(Image Source: iStock Photos, ©)

Photo of cold and flu over the counter remedies; includes a cap full of cough syrup, cough drops, and a spilled bottle of pills

Figure 8.4.6.4.1.3. *Cold and flu remedies are another source of chemical exposure*
(Image Source: iStock Photos, ©)

Toxicology studies and human health risk assessments have traditionally focused primarily on a single chemical. However, as noted above, people are exposed to many chemicals every day. They are also exposed to non-chemical stressors every day and throughout a lifetime.

In addition, non-chemical stressors include infectious agents, diet, and psychosocial stress, all of which have potential roles in contributing to the health effects associated with chemical exposures.

Approaches for Assessing Interactions

Development of methods to assess the health effects associated with complex exposures is underway at **various organizations**.

Non-animal tools and approaches are demonstrating high potential for use in assessing combined effects of chemicals on humans and the environment. These tools and approaches may help uncover information about new mixture components or entire mixtures, which can promote understanding of the underlying mechanisms of their combined effects. The strategies for assessing interactions rely less on *in vivo* testing and more on non-animal studies and computational tools and incorporate emerging approaches such as:

- The adverse outcome pathway (AOP) concept.
- *In vitro* methods.
- "Omics" techniques.
- *In silico* approaches such as quantitative structure activity relationships (QSARs).
- Read-across.
- Toxicokinetic modeling.
- Integrated approaches to testing and assessment (IATA).

The goals include the ability to development more effective and comprehensive regulatory assessments while reducing the reliance on animal testing.

Photo of a lab scientist typing on a computer keyboard, with a microscope and chemical flasks in the background

Figure 8.4.6.4.1.4. *Modern testing methods rely heavily on computational toxicology*
(Image Source: iStock Photos, ©)

Knowledge Check

The presence of one chemical decreasing toxicity of another chemical is called:

- ○ Additivity
- ○ Antagonism
- ○ Synergism

Answer available at the end of this section.

Additivity is when the combined toxic effect of two chemicals when given together is less than the sum of their individually measured toxic effects.

- ○ True
- ○ False

Answer available at the end of this section.

Piperonyl butoxide is not an insecticide; however, it can greatly increase the effects of a pyrethrum insecticide. Thus, piperonyl butoxide can be called a synergist and this interaction can be called synergism.

- ○ True
- ○ False

Answer available at the end of this section.

Answers

Knowledge Check

Antagonism - **This is the correct answer.**

When the presence of another chemical decreases toxicity of a chemical, this is called antagonism.

Knowledge Check

False - **This is the correct answer.**

An additive effect occurs when the combined effects of two or more chemicals is equal to the sum of the effects of each chemical given alone.

Knowledge Check

True - **This is the correct answer.**

The interaction of this combination is synergism. Synergists are used to enhance the toxicity of several commonly used insecticides.

This page titled 8.4.6.4.1: Interactions is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 4.1: Interactions by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.5: TOXICITY TESTING METHODS

Learning Objectives

After completing this lesson, you will be able to:

- Explain modern approaches to testing for and assessing toxicity.
- Identify sources of information related to alternatives to using animals to assess toxicity.
- Explain how clinical investigations and epidemiology studies are used to evaluate toxicity to humans.

In this section...

Topics include:

8.4.6.5.1: Testing and Assessing Toxicity

8.4.6.5.2: Clinical Investigations and Other Types of Human Data

8.4.6.5.3: Epidemiology Studies

What We've Covered

This section included the following key points:

- The 3Rs concept of using test methods replace the use of animals with other types of studies and approaches, reduce the number of animals used in studies, and refine study procedures to cause less pain or stress to animals.
- ALTBIB is a comprehensive starting point provided by NLM to find information related to alternatives to animal testing.
- Animal tests for toxicity have been conducted prior to and in parallel with human clinical investigations.
- Standardized animal tests have been developed for testing:
  - Acute toxicity
  - Subchronic toxicity
  - Chronic toxicity
  - Carcinogenicity
  - Reproductive toxicity
  - Developmental toxicity
  - Dermal toxicity
  - Ocular toxicity
  - Neurotoxicity
  - Genetic toxicity
- Modern approaches to toxicity testing are preferred over animal testing and include:
  - *In vitro* methods, which are performed outside living organisms.
  - *In silico* methods, which are performed using computers and computer simulation.
  - Chip models, which include human cell cultures placed on computer chips for study.
- Approaches used for testing pharmaceuticals include:
  - Clinical investigations, in which human subjects are studied with clinical observations and laboratory measurements.
  - Epidemiological studies, involving observation of humans exposed to xenobiotics in their regular life or occupation.
  - Reports of adverse reactions to drugs.
- Consumer products and the chemicals they contain are tested through:
  - *In silico* data from computer models.
  - *In vitro* data from tests performed as alternatives to animal testing.
  - Animal study data.
  - Human data from premarketing and postmarketing studies.

Coming Up...

In the next section, we will explore the concept of risk assessment.

This page titled 8.4.6.5: Toxicity Testing Methods is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 5: Toxicity Testing Methods by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.5.1: Testing and Assessing Toxicity

Testing and Assessing Toxicity

Alternatives to animal testing have emerged in recent years.

Since about 1990, numerous attempts have been made around the world to reduce the use of and replace laboratory animals in toxicology and other studies. These efforts have involved finding alternatives to animal testing and incorporating the "3Rs" concept (Replace, Reduce and Refine), which means using test methods that:

- **Replace** the use of animals with other types of studies and approaches.
- **Reduce** the number of animals in studies.
- **Refine** the procedures to make studies less painful or stressful to the animals.

Regulatory authorities, companies, and others have endorsed the principle of the 3Rs, and alternative testing methods have been and are being developed. An international group that has played a key role is the International Cooperation on Alternative Test Methods (ICATM). Established in 2009, ICATM includes representatives of **organizations from various countries**.

Finding Information about Alternatives to Animal Testing

Many countries including the United States, Canada, and the European Union member states, require that a comprehensive search for possible alternatives be completed before some or all research involving animals is begun. Because numerous Web resources are now available to provide guidance and other information on *in vitro* and other alternatives to animal testing, completing such searches and keeping current with information associated with alternatives to animal testing is much easier than it used to be.

ALTBIB, from NLM

The NLM ALTBIB ("Resources for Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing") portal is a comprehensive starting point for finding information related to alternatives to animal testing. ALTBIB is available at **http://toxnet.nlm.nih.gov/altbib.html**.

It provides access to **PubMed®/MEDLINE®** citations relevant to alternatives to use of live vertebrates in biomedical research and testing.

ALTBIB's topics and subtopics are aligned with current approaches. For example, information is provided on *in silico*, *in vitro*, and improved (refined) animal testing methods and on testing strategies that incorporate these methods and other approaches.

Screenshot of the ALTBIB homepage

Figure 8.4.6.5.1.1. NLM ALTBIB homepage

ALTBIB also provides access to news and additional resources, including information on the status of the evaluation and acceptance of alternative methods. Main categories include:

- Animal Alternatives News
- Additional Resources
- Evaluation/Acceptance of Test Methods

Links to Specific Resources (Sources Providing Animal Alternatives News, Key Organizations Providing Resources, and the Regulatory Acceptance of Specific Alternative Methods and Milestones in Non-animal Toxicity Testing)

Animal Tests

NOTE: This information is provided for historical and other reasons, especially since animal testing is still being done in some cases, and because toxicologists, risk assessors, and others are faced with interpreting the results of new and old studies that used animals.

Animal tests for toxicity have been conducted prior to and in parallel with human clinical investigations as part of the non-clinical laboratory tests of pharmaceuticals. For pesticides and industrial chemicals, human testing is rarely conducted. Years ago, results from animal tests were often the only way to effectively predict toxicity in humans.

Animal tests were developed and used because:

- Chemical exposure can be precisely controlled.
- Environmental conditions can be well-controlled.
- Virtually any type of toxic effect can be evaluated.
- The mechanism by which toxicity occurs can be studied.



Figure 8.4.6.5.1.2. *Rats have traditionally been used in toxicity studies using animals (Image Source: iStock Photos, ©)*

Standardized Animal Toxicity Tests

Animal methods to evaluate toxicity have been developed for a wide variety of toxic effects. Some procedures for routine safety testing have been standardized. **Standardized animal toxicity** tests have been highly effective in detecting toxicity that may occur in humans. As noted above, concern for animal welfare has resulted in tests that use humane procedures and only as many animals as are needed for statistical reliability.

To be standardized, a test procedure must have scientific acceptance as the most meaningful assay for the toxic effect. Toxicity testing can be very **specific** for a particular effect, such as dermal irritation, or it may be **general**, such as testing for unknown chronic effects.

Standardized tests have been developed for the following effects:

- Acute Toxicity
- Subchronic Toxicity
- Chronic Toxicity
- Carcinogenicity
- Reproductive Toxicity
- Developmental Toxicity
- Dermal Toxicity
- Ocular Toxicity
- Neurotoxicity
- Genetic Toxicity

Species Selection

Species selection varies with the toxicity test to be performed. There is no single species of animal that can be used for all toxicity tests. Different species may be needed to assess different types of toxicity. The published literature (such as via PubMed) and online databases (such as TOXNET) should be searched for information from non-animal and animal studies, as well as for possible best approaches, most applicable species, and strains and gender of a species. Here are two examples:

- It would have been invaluable years ago for toxicologists and risk assessors to have known that carcinogenic effects in male rats are considered irrelevant for humans if the α(2u)-globulin protein is involved because humans lack that protein. See **another example**
- Many physiological, pharmacological, and toxicological findings related to organic anion and cation transport and

transporters in rodents and rabbits do not apply to humans. **Learn more**

In some cases, it may not be possible to use the most desirable animal for testing because of animal welfare or cost considerations.

- For example, use of dogs and non-human primates is now restricted to special cases or banned by some organizations, even though they represent the species that may respond the closest to humans in terms of chemical and other exposures (however, note the examples above).

Rodents and rabbits are the most commonly used laboratory species because they are readily available, inexpensive to breed and house, and they have a history of producing reliable results in experiments.

The toxicologist attempts to design an experiment to duplicate the potential exposure of humans as closely as possible. For example:

- The **route of exposure** should simulate that of human exposure. Most standard tests use inhalation, oral, or dermal routes of exposure.
- The **age of test animals** should relate to that of humans. Testing is normally conducted with young adults, although in some cases, newborn or pregnant animals may be used.
- For most routine tests, **both sexes** are used. Sex differences in toxic response are usually minimal, except for toxic substances with hormonal properties.
- **Dose levels** are normally selected so as to determine the threshold as well as a dose-response relationship. Usually, a minimum of three dose levels are used.

Rat in a cage on a table inside a laboratory

Figure 8.4.6.5.1.3. *Rodents have commonly been used in animal testing*
(Image Source: iStock Photos, ©)

Acute Toxicity

Historically, **acute toxicity tests** were the first tests conducted. They provide data on the relative toxicity likely to arise from a single or brief exposure, or sometimes multiple doses over a brief period of time. Standardized tests are available for oral, dermal, and inhalation exposures, and many regulatory agencies still require the use of all or some of these tests. Table 8.4.6.5.1.1 lists basic parameters historically used in acute toxicity testing.

Category	Parameter
Species	Rats preferred for oral and inhalation tests; rabbits preferred for dermal tests
Age	Young adults
Number of animals	5 of each sex per dose level
Dosage	Three dose levels recommended; exposures are single doses or fractionated doses up to 24 hours for oral and dermal studies and 4-hour exposure for inhalation studies
Observation period	14 days

Table 8.4.6.5.1.1. *Acute toxicity test parameter*

Subchronic Toxicity

Subchronic toxicity tests are employed to determine toxicity likely to arise from repeated exposures of several weeks to several months. Standardized tests are available for oral, dermal, and inhalation exposures. Detailed information is obtained during and after the study, ranging from body weight, food and water consumption measurements, effects on eyes and behavior, composition of blood, and microscopic examination of selected tissues and organs.

Table 8.4.6.5.1.2 lists basic parameters previously used in subchronic toxicity testing.

Category	Parameter
Species	Rodents (usually rats) preferred for oral and inhalation studies; rabbits for dermal studies; non-rodents (usually dogs) recommended as a second species for oral tests
Age	Young adults
Number of animals	10 of each sex for rodents; 4 of each sex for non-rodents per dose level
Dosage	Three dose levels plus a control group; includes a toxic dose level plus NOAEL; exposures are 90 days
Observation period	90 days (same as treatment period)

Table 8.4.6.5.1.2. Subchronic toxicity test parameter

Chronic Toxicity

Chronic toxicity tests determine toxicity from exposure for a substantial portion of a subject's life. They are similar to the subchronic tests except that they extend over a longer period of time and involve larger groups of animals.

Table 8.4.6.5.1.3 includes basic parameters previously used in chronic toxicity testing.

Category	Parameter
Species	Two species recommended; rodent and non-rodent (rat and dog)
Age	Young adults
Number of animals	20 of each sex for rodents, 4 of each sex for non-rodents per dose level
Dosage	Three dose levels recommended; includes a toxic dose level plus NOAEL. The recommended maximum chronic testing durations for pharmaceuticals are now 6 and 9 months in rodents and non-rodents, respectively. (Historically exposures were for 12 months, 24 months for food chemicals.)
Observation period	12-24 months

Table 8.4.6.5.1.3. Chronic toxicity test parameter

Carcinogenicity

Carcinogenicity tests are similar to chronic toxicity tests. However, they extend over a longer period of time and require larger groups of animals in order to assess the potential for cancer.

Table 8.4.6.5.1.4 lists basic parameters used in the past in carcinogenicity testing.

Category	Parameter
Species	Testing in two rodents species—rats and mice—is preferable given their relatively short life spans.
Age	Young adults
Number of animals	50 of each sex per dose level
Dosage	Three dose levels recommended; highest should produce minimal toxicity; exposure periods are at least 18 months for mice and 24 months for rats

Observation period	18-24 months for mice and 24-30 months for rats

Table 4. Carcinogenicity test parameter

Reproductive Toxicity

Reproductive toxicity testing is intended to determine the effects of substances on gonadal function, conception, birth, and the growth and development of offspring. The oral route of administration is preferable.

Table 8.4.6.5.1.5 lists basic parameters historically used in reproductive toxicity testing.

Category	Parameter
Species	Rat is recommended
Age	Young adults
Number of animals	20 of each sex per dose level
Dosage	Three dose levels recommended; highest dose should produce toxicity but not mortality in parents; lowest dose should not produce toxicity
Observation period	Test substance given to parental animals (P1) prior to mating, during pregnancy, and through weaning of first generation (F1) offspring; substance then given to selected F1 offspring during their growth into adulthood, mating, and production of second generation (F2) until the F2 generation is 21 days old.

Table 8.4.6.5.1.5. Reproductive toxicity test parameter

Developmental Toxicity

Developmental toxicity testing detects the potential for substances to produce embryotoxicity and birth defects.

Table 8.4.6.5.1.6 lists basic parameters previously used in developmental toxicity tests.

Category	Parameter
Species	Two species recommended; rat, mouse, hamster, and rabbit are used most commonly.
Age	Young adult females
Number of animals	20 pregnant females per dose level
Dosage	At least three dose levels are used; includes a toxic dose level plus NOAEL; occurs throughout organ development in the fetus for teratogenic effects; starts with parents prior to breeding, continues during pregnancy for all developmental effects
Observation period	Offspring sacrificed and examined the day prior to expected birth for teratogenic effects; offspring observed for growth retardation and abnormal function through infancy and examined for teratogenic effects

Table 8.4.6.5.1.6. Developmental toxicity test parameter

Dermal Toxicity

Dermal toxicity tests determine the potential for an agent to cause irritation and inflammation of the skin. Those reactions may be a result of direct damage to the skin cells by a substance or an indirect response due to sensitization from prior exposure. *In vitro* approaches to dermal toxicity testing are being developed, in part because this type of testing has received so much publicity.

Table 8.4.6.5.1.7 lists basic parameters historically used in dermal toxicity testing.

Category	Parameter
Primary dermal irritation	Determines direct toxicity. The substance is applied to the skin of 6 albino rabbits for 4 hours and the rabbits are observed for 72 hours for irritation.
Dermal sensitization	Assays for immune hypersensitivity of the skin, consisting of two phases: 1) Application of the test substance to the skin of guinea pigs for 4 hours in the sensitization phase; and 2) a challenge phase at least 1 week later in which the substance is reapplied to the skin. An inflammatory reaction indicates that the skin has been sensitized to the substance.

Table 8.4.6.5.1.7. Dermal toxicity test parameter

Ocular Toxicity

Ocular toxicity was at one time determined by applying a test substance for 1 second to the eyes of 6 test animals, usually rabbits. The eyes were then carefully examined for 72 hours, using a magnifying instrument to detect minor effects. An ocular reaction can occur on the cornea, conjunctiva, or iris. It may be simple irritation that is reversible and quickly disappears.

This eye irritation test was commonly known as the **"Draize Test."** This test has received much attention, such as the development of a **"low volume" variation and *in vitro* approaches**.

Neurotoxicity

A battery of **standardized neurotoxicity tests** were developed to supplement the **delayed neurotoxicity test** in domestic chickens (hens). The hen assay determines delayed neurotoxicity resulting from exposure to anticholinergic substances, such as certain pesticides. The hens are protected from immediate neurological effects of the test substance and observed for 21 days for delayed neurotoxicity.

Table 8.4.6.5.1.8 lists measurements included in other neurotoxicity tests.

Category	Parameter
Motor activity	Tests for decreased motor activity, such as cage movement. Rats or mice are used.
Peripheral nerve conduction	Tests for electrical conduction in motor and sensory nerves. Rodents are exposed to the test substance for 90 days.
Neuropathy	Tests for nerve damage by microscopic examination. This is one aspect of other standardized toxicity tests.

Table 8.4.6.5.1.8. Neurotoxicity test parameter

Genetic Toxicity

Genetic toxicity is determined using a wide range of test species including whole animals and plants *(for example, rodents, insects, and corn)*, microorganisms, and mammalian cells. A large variety of tests have been developed to measure gene mutations, chromosome changes, and DNA activity.

Table 8.4.6.5.1.9 lists parameters used for common gene mutation tests.

Category	Parameter
Microorganisms	*Salmonella typhimurium* and *Escherichia coli* are commonly used bacterial tests. The S. *typhimurium* assay is known as the Ames Test. Yeasts are also used to detect gene mutations in culture systems.

Mammalian cells	The two main cell lines are mouse lymphoma and Chinese hamster ovary (CHO) cells.
Fruit Flies	*Drosophila melanogaster* is used to detect sex-linked recessive lethal mutations.
Mice	The Mouse Specific-Locus Test is the major gene mutation test that employs whole animals. Exposed mice are bred and observed for hereditary changes.

Table 8.4.6.5.1.9. *Genetic toxicity test parameter*

Chromosomal effects can be detected with a variety of tests, some of which utilize entire animals *(in vivo) and* some which use cell systems *(in vitro)*. Several assays are available to test for chemically induced chromosome aberrations in whole animals. Table 8.4.6.5.1.10 lists common *in vivo means of* testing chromosomal effects.

Category	Parameter
Rodent chromosomal assay	Involves exposure of mice or rats to a single dose of a substance. Their bone marrow is analyzed for chromosome aberrations over a 48-hour period.
Dominant lethal assay	Exposed male mice or rats are mated with untreated females. The presence of dead implants or fetuses is the result of the fertilized ovum receiving damaged DNA from the sperm. This leads to the death of the embryo or fetus. The genetic defect in the sperm is thus a heritable dominant lethal mutation.
Micronucleus test	Mice are exposed once and their bone marrow or peripheral blood cells are examined for 72 hours for the presence of micronuclei, such as broken pieces of chromosomes surrounded by a nuclear membrane.
Heritable translocation assay	Exposed male Drosophila or mice are bred to non-exposed females. The offspring males (F1 generation) are then bred to detect chromosomal translocations.
Sister chromatid exchange assay (SCE)	Mice are exposed to a substance and their bone marrow cells or lymphocytes are examined microscopically for complete chromosomal damage. This is indicated by chromatid fragments joining sister chromatids rather than their own.

Table 8.4.6.5.1.10. *Chromosomal effects (In vivo) test parameter*

In Vitro Testing

In vitro tests for chromosomal effects involve exposure of cell cultures and followed by microscopic examination of them for chromosome damage.

The most commonly used cell lines are Chinese Hamster Ovary (CHO) cells and human lymphocyte cells. The CHO cells are easy to culture, grow rapidly, and have a low chromosome number (22), which makes for easier identification of chromosome damage.

Human lymphocytes are more difficult to culture. They are obtained from healthy human donors with known medical histories. The results of these assays are potentially more relevant to determine effects of xenobiotics that induce mutations in humans.

Two widely used genotoxicity tests measure DNA damage and repair that is not mutagenicity. DNA damage is considered the first step in the process of mutagenesis. Common assays for detecting DNA damage include:

1. **Unscheduled DNA synthesis (UDS)** — involves exposure of mammalian cells in culture to a test substance. UDS is measured by the uptake of tritium-labeled thymidine into the DNA of the cells. Rat hepatocytes or human fibroblasts are the mammalian cell lines most commonly used.
2. **Exposure of repair-deficient *E. coli* or *B. subtilis*** — DNA damage cannot be repaired so the cells die or their growth may be inhibited.

Emerging Approaches and Methods

In the future, there will likely be additional and refined *in vitro* methods, and the emergence of *in silico* and "chip" approaches.

Many current efforts are underway to refine, develop, and validate *in vitro* methods.

Did you know?

The Human Toxicology Project Consortium provides a video series called **Pathways to a Better Future**. These videos discuss the future of toxicology, if you would like to know more about where the field is headed.

In Silico Methods

Also emerging are *in silico* methods, meaning "performed on computer or via computer simulation." This term was developed as an analogy to the Latin phrases *in vivo* and *in vitro*.

Advanced computer models called "Virtual Tissue Models" are being developed by the U.S. **EPA's National Center for Computational Toxicology (NCCT)**. The EPA's Virtual Tissue Models are described as using "new computational methods to construct advanced computer models capable of simulating how chemicals may affect human development. Virtual tissue models are some of the most advanced methods being developed today. The models will help reduce dependence on animal study data and provide much faster chemical risk assessments" (**source**).

One example is the Virtual Embryo (**v-Embryo™**) research effort, aimed at developing prediction models to increase our understanding of how chemical exposure may affect unborn children. Researchers are integrating new types of *in vitro*, *in vivo*, and *in silico* models that simulate critical steps in fetal development. Virtual Embryo models simulate biological interactions observed during development and predict chemical disruption of key biological events in pathways that is believed to lead to adverse effects.

v-Embryo logo from EPA

Figure 8.4.6.5.1.4. *v-Embryo logo*

"Chip" Models

Also emerging are microphysiological systems (MPS) that are used in "tissue chip" and "organs-on-chips" models. Chip models include human cell cultures that are placed on a computer chip and studied there. The **Wyss (pronounced "Veese") Institute for Biologically Inspired Engineering** is a helpful resource for more information.

For example, the "Lung-on-a-chip" is described as "combining microfabrication techniques with modern tissue engineering, lung-on-a-chip offers a new *in vitro* approach to drug screening by mimicking the complicated mechanical and biochemical behaviors of a human lung." To learn more, watch a **video** from the Wyss Institute that shows a human lung-on-a-chip. **Another Wyss Institute video** illustrates how researches have used long-on-a-chip to mimic pulmonary edema.

A human hand is shown holding the computer chip containing lung data.

Figure 8.4.6.5.1.5. *Lung-on-a-chip used to mimic pulmonary edema*
(Image Source: The Wyss Institute for Biologically Inspired Engineering)

Using a connected series of tissue chips as an integrated multi-organ system can allow for the creation of a "**human-on-a-chip**," to be used to model the metabolism and effects of drugs and other substances moving through a human. For example, a liver chip could provide fluids and metabolites to a kidney chip, allowing for the assessment of the nephrotoxic (kidney damage) potential of a substance metabolized in the liver.

Induced Pluripotent Stem Cells (iPSCs)

Induced pluripotent stem cells (iPSCs) are an emerging approach using *in vitro* cultures of cells. The cells of mammals and plants can be reprogrammed via "cellular reprogramming" to generate iPSCs. Like human embryonic stem cells, iPSCs are

pluripotent (capable of giving rise to several different cell types) and these cells can renew themselves. As examples, iPSC-derived hepatocytes, cardiomyocytes, and neural cells can serve as tools for the screening of drugs and other substances for potential toxicity, and also can be used to study disease mechanisms and pathways. Further, iPSCs have been studied in immunotherapy and regenerative cellular therapies.

Schematic representation of how somatic cells taken from a patient can be reprogrammed into induced pluripotent stem cells (iPSCs) using the 'Yamanaka' factors, OCT4, KLF4, c-MYC and SOX2. Subsequent differentiation of human iPSCs (hiPSCs) into neurons of define lineage allow for investigations into disease pathophysiology and identification of potential drug targets. In addition, hiPSC derived neurons may function as a cellular platform in which drug screens can be carried out using disease relevant neurons.

Figure 8.4.6.5.1.6. *Promise of hiPSCs. Schematic representation of how somatic cells taken from a patient can be reprogrammed into induced pluripotent stem cells (iPSCs) using the 'Yamanaka' factors, OCT4, KLF4, c-MYC and SOX2. Subsequent differentiation of human iPSCs (hiPSCs) into neurons of define lineage allow for investigations into disease pathophysiology and identification of potential drug targets. In addition, hiPSC derived neurons may function as a cellular platform in which drug screens can be carried out using disease relevant neurons.*
(Image Source: ***Adapted under Creative Commons Attribution License (CC BY)****.*
doi: ***10.1016/j.yhbeh.2015.06.014***
Original image: ***https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4579404/figure/f0005/****)*

Did you know?

Professor Shinya Yamanaka (Kyoto University, Japan) received the 2012 Nobel Prize in Physiology or Medicine for discovering that mature cells can be reprogrammed to iPSCs that can differentiate into any type of cell. Key to this discovery was his use of four "reprogramming factors" referred to as c-Myc, Klf4, Oct3/4, and Sox2.

Learn more

Combining "Chips" and iPSCs

The emerging approaches of "chips" and iPSCs are being combined. One example is for the evaluation of drugs as potential countermeasures for biological and chemical threats that can be a substitute for human clinical trials. The "chips" and "humans on a chip" can be used as complex *in vitro* human models to simulate the biology and function of an organ.

This page titled 8.4.6.5.1: Testing and Assessing Toxicity is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.1: Testing and Assessing Toxicity by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.5.2: Clinical Investigations and Other Types of Human Data

For Drugs

The focus of this section is on the U.S. Food and Drug Administration (FDA), but regulatory agencies worldwide have very similar approaches. The main methods of determining the toxicity of drugs to humans are:

- **Clinical investigations** — administration of chemicals to human subjects with careful clinical observations and laboratory measurements.
- **Epidemiological studies** — observation of humans who have been exposed to xenobiotics in the normal course of their life or occupation.
- **Adverse reactions to drug reports** — reports voluntarily submitted by physicians to the FDA after a drug has been approved and is in widespread use.

Closeup photo of various unlabeled pills and capsules

Figure 8.4.6.5.2.1. Drugs can be toxic to humans
(Image Source: iStock Photos, ©)

Clinical Investigations

Clinical investigations are a component of **Investigational New Drug Applications (INDs)** submitted to the FDA. Clinical investigations are conducted only after a minimal battery of nonclinical laboratory studies has been completed.

Toxicity studies using human subjects require strict ethical considerations. They are primarily conducted for new pharmaceutical applications submitted to the FDA for approval.

Generally, toxicity found in animal studies occurs with similar incidence and severity in humans. Differences sometimes occur, thus clinical tests with humans are needed to confirm the results of nonclinical laboratory studies.

FDA clinical investigations are conducted in three phases, as outlined below.

Screenshot of the first page of the Investigational
New Drug Application (IND)

Figure 8.4.6.5.2.2. *Portion of the Investigational New Drug Application (IND)*
(Image Source: FDA)

Phase 1 consists of testing the drug in a small group of 20 to 80 healthy volunteers. Information obtained in Phase 1 studies is used to design Phase 2 studies, in particular, to determine the drug's:

- Initial tolerability in human subjects.
- Pharmacokinetics and pharmacological effects.

Phase 1 - study of a small group of 20-80 healthy volunteers

Phase 2 studies are more extensive, involving several hundred patients and are used to:

- Determine the short-term side effects of the drug.
- Determine the risks associated with the drug.
- Evaluate the effectiveness of the drug for treatment of a particular disease or condition.
- Elucidate the drug's metabolism.

Phsae 2 - more extensive studies of several hundred patients

Phase 3 studies are **controlled** and uncontrolled trials conducted with several hundred to several thousand patients. They are designed to:

- Gather additional information about effectiveness and safety.
- Evaluate overall risk:benefit profile of the drug.
- Provide the basis for the precautionary information that accompanies the drug.

Phase 3 - controlled and uncontrolled trials of several hundred to several thousand patients

For Consumer Products

Health-related data for a chemical in a consumer product (and for the consumer product itself for the human studies) can come from the following types of studies:

- *In silico* data — from computer programs that estimate toxic properties based on data for similar chemicals, and/or from the physical chemical properties.
- *In vitro* data — from the results of alternatives to animal tests, such as from cell cultures used to assess the potential for eye or skin irritation.
- Animal (toxicological) study data — for example, from studies that assessed eye or skin irritation potential.
- Human data – from studies conducted before (**premarketing**) and after (**postmarketing**) a product had been sold to consumers. More specifically, from:
  - Premarketing **clinical studies**, such as from patch tests to assess skin irritation potential.
  - Premarketing **"controlled use" studies** that are designed to assess the skin effects from using a new type of personal care product.
  - Postmarketing **studies conducted by physicians or dermatologists**, such as testing a diagnostic patch with their patients.
  - Postmarketing **epidemiological studies**, including studies developed by Poison Control Centers, companies, and academia that look at the "real world" health reports of effects associated with consumer use of a product.

Photo of two scientists in a lab with chemicals, looking at a laptop screen for research

Figure 8.4.6.5.2.3. *Toxicologists in a lab, using a computer for research*
(Image Source: iStock Photos, ©)

Did you know?

Bisphenol A (BPA) and **phthalates** are chemicals that have been widely found in consumer products. BPA has been used in some food can linings, polycarbonate food and beverage containers, tooth sealants applied to dentists, and even in cash register receipts! Examples of potential exposures to BPA include eating or drinking foods or liquids from those containers, and skin exposures from handing the cash register receipts. Workers involved in making products with BPA can be exposed during production.

Often called plasticizers, phthalates are used to make plastics more flexible. Some phthalates are used as solvents. They can be found in vinyl flooring and shower curtains, children's toys, personal care products, and as contaminants in the food supply. As with BPA, exposures can come from many sources.

Toxicologists and others are still assessing the full extent of the potential impacts on health. Studies suggest that BPA and phthalates affect the reproductive system, impacting how hormones such as estrogen and testosterone work in the body. The impact of fetal or early childhood exposures is still being assessed. Because of the ubiquity of the possible products containing these chemicals, thorough assessments of potential exposures, toxicities, and potential substitutes are essential.

Plastic products often contain BPA
and phthalates

Figure 8.4.6.5.2.4. *Plastic food and beverage containers are common sources of BPA and phthalates in consumer products*
(Image Source: iStock Photos, ©)

This page titled 8.4.6.5.2: Clinical Investigations and Other Types of Human Data is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.2: Clinical Investigations and Other Types of Human Data by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.5.3: Epidemiology Studies

What Are Epidemiology Studies?

Epidemiology studies are conducted using human populations to evaluate whether there is a **correlation or causal relationship** between exposure to a substance and adverse health effects.

These studies differ from clinical investigations in that individuals have already been administered the drug during medical treatment or have been exposed to it in the workplace or environment.

Epidemiological studies measure the risk of illness or death in an exposed population compared to that risk in an identical, unexposed population (for example, a population the same age, sex, race and social status as the exposed population).

Figure 8.4.6.5.3.1. Epidemiology studies tend to produce graphs and charts for data analysis and presentation
(Image Source: Adapted from iStock Photos, ©)

Types of Studies

There are four primary types of epidemiology studies. They are:

1. **Cohort studies** — A cohort (group) of individuals with exposure to a chemical and a cohort without exposure are followed over time to compare disease occurrence.
2. **Case control studies** — Individuals with a disease (such as cancer) are compared with similar individuals without the disease to determine if there is an association of the disease with prior exposure to an agent.
3. **Cross-sectional studies** — The prevalence of a disease or clinical parameter among one or more exposed groups is studied, such as:
    - The prevalence of respiratory conditions among furniture makers.
4. **Ecological studies** – The incidence of a disease in one geographical area is compared to that of another area, such as:
    - Cancer mortality in areas with hazardous waste sites as compared to similar areas without waste sites.

Cohort Studies

Cohort studies are the most commonly conducted epidemiology studies and they frequently involve occupational exposures. Exposed persons are easy to identify and their exposure levels are usually higher than in the general public. There are two types of cohort studies:

1. **Prospective**, in which cohorts are identified based on current exposures and followed into the future.
2. **Retrospective,** in which cohorts are identified based on past exposure conditions and study "follow-up" proceeds forward in time; data come from past records.

Common Statistical Measures

Standard, quantitative measures are used to determine if epidemiological data are meaningful. The most commonly used measures are:

- **Odds Ratio (O/R)** — The ratio of risk of disease in a case-control study for an exposed group to an unexposed group. An odds ratio equal to 2 (O/R = 2) means that the exposed group has twice the risk as the non-exposed group.
- **Standard Mortality Ratio (SMR)** — The relative risk of death based on a comparison of an exposed group to non-exposed group. A standard mortality ratio equal to 150 (SMR = 150) indicates that there is a 50% greater risk.
- **Relative Risk (RR)** — The ratio expressing the occurrence of disease in an exposed population to that of an unexposed

population. A relative risk of 175 (RR = 175) indicates a 75% increase in risk.

Study Design

When designing an epidemiology study, the most critical aspects include:

- An appropriate control group.
- An adequate time span.
- The statistical ability to detect an effect.

More specifically, the control population used as a comparison group must be as similar as possible to that of the test group, for example, same age, sex, race, social status, geographical area, and environmental and lifestyle influences.

Many epidemiology studies evaluate the potential for an agent to cause cancer. Because most cancers require long latency periods, the study must cover that period of time.

The statistical ability to detect an effect is referred to as the **power** of the study. To gain precision, the study and control populations should be as large as possible.

Bias Errors

Epidemiologists attempt to control errors that can occur in the collection of data, which are known as bias errors. The three main types of bias errors are:

1. **Selection bias**, which occurs when the study group is not representative of the population from which it came.
2. **Information bias**, which occurs when study subjects are misclassified as to disease or exposure status. Recall bias occurs when individuals are asked to remember exposures or conditions that existed years before.
3. **Confounding factors**, which occur when the study and control populations differ with respect to factors which might influence the occurrence of the disease. For example, smoking might be a confounding factor and should be considered when designing studies.

Postmarketing Studies

Finally, for consumer products, **postmarketing** epidemiological studies can be performed. Examples include studies developed by Poison Control Centers, companies, academia, and other sources to look at the "real world" health reports of effects associated with consumer use of a product or article under reasonably foreseeable conditions.

Knowledge Check

While animal testing was historically the primary method used in testing for toxicity, modern testing methods prefer:

- ◯ *In silico*
- ◯ *In vitro*
- ◯ Refined animal testing
- ◯ All of the above

Answer available at the end of this section.

In testing a pharmaceutical to comply with FDA requirements, the initial testing consists of:

- ◯ Clinical investigations

- ◯ Non-clinical laboratory studies
- ◯ Epidemiology studies
- ◯ All of the above

Answer available at the end of this section.

The primary goal of a Phase 1 clinical investigation is to:

- ◯ Obtain information to design a more definitive Phase 2 clinical investigation
- ◯ Evaluate the effectiveness of a drug for treating disease
- ◯ Provide the basis for pharmaceutical labeling

Answer available at the end of this section.

Determining the overall risk versus the benefit of a new pharmaceutical is part of:

- ◯ Phase 2 clinical investigation
- ◯ Phase 3 clinical investigation
- ◯ Epidemiology study

Answer available at the end of this section.

The type of epidemiology study in which individuals are identified according to exposure and followed to determine subsequent disease risk is known as:

- ◯ Cohort study
- ◯ Case control study
- ◯ Cross-sectional study
- ◯ Ecological study

Answer available at the end of this section.

An epidemiological study in which the individuals that make up the test cohort are identified according to past exposures is known as:

- ◯ Case control study
- ◯ Prospective cohort study
- ◯ Retrospective cohort study

Answer available at the end of this section.

Answers

Knowledge Check

All of the above - **This is the correct answer.**
Modern approaches to testing for toxicity include *in silico*, *in vitro*, and improved (refined) animal testing.

Knowledge Check

Non-clinical laboratory studies - **This is the correct answer.**
Investigational New Drug Applications (IND) require clinical investigations. Before clinical investigations begin, a minimal battery of non-clinical laboratory studies must be completed.

Knowledge Check

Obtain information to design a more definitive Phase 2 clinical investigation - **This is the correct answer.**
The primary goal of a Phase 1 clinical investigation is to obtain information that is used to design more extensive, Phase 2 studies.

Knowledge Check

Phase 3 clinical investigation - **This is the correct answer.**
Determining the overall risk versus the benefit of a new pharmaceutical is part of Phase 3 clinical study. The risk versus benefit is one of the last steps in the drug evaluation process.

Knowledge Check

Cohort study - **This is the correct answer.**
The type of epidemiology study in which individuals are identified according to exposure and followed to determine subsequent disease risk is known as a cohort study. In a cohort study individuals are selected to be part of the group based on their exposure to a particular substance.

Knowledge Check

Retrospective cohort study - **This is the correct answer.**
This is known as a retrospective cohort study. As the name implies, retrospective cohorts are identified according to past exposure conditions and the follow-up study proceeds forward in time.

This page titled 8.4.6.5.3: Epidemiology Studies is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 5.3: Epidemiology Studies by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.6: RISK ASSESSMENT

Learning Objectives

After completing this lesson, you will be able to:

- Identify the basic steps in the risk assessment process.
- Explain the framework for risk-based decision-making.
- Describe methods for identifying hazards.
- Explain methods for toxicity assessment, including dose-response and exposure.

In this section...

Topics include:

8.4.6.6.1: Risk Assessment

8.4.6.6.2: Hazard Identification

8.4.6.6.3: Dose-Response Assessment

8.4.6.6.4: Exposure Assessment

8.4.6.6.5: Risk Characterization

What We've Covered

This section made the following main points:

- A **hazard** is the capability of a substance to cause an adverse effect.
  A **risk** is the probability that the hazard will occur under specific conditions.
- **Risk assessment** is the process of determining hazard, exposure, and risk.
  Risk management is the process of weighing policy alternatives and deciding on the most appropriate regulatory action.
- There are **four basic steps to risk assessment**:
  1. **Hazard Identification**
     - Identify or develop information suggesting or confirming whether a chemical poses a potential hazard to humans.
     - (Quantitative) Structure Activity, or (Q)SAR methods, including computer models, help consider closely related chemicals as a group or category.
     - Read-across involves estimating what a chemical may be like, including the presence or absence of certain properties or activities, based on one or more other chemicals.
     - Adverse Outcome Pathways (AOPs) involve *in vitro* methods that evaluate changes in normal cellular signaling pathways.
     - Other emerging methods include (Quantitative) *in vitro* to *in vivo* extrapolation, or (Q)IVIVE, Integrated Testing Strategies, and Integrated Approaches to Testing and Assessment (IATA).
  2. **Dose-Response Assessment**
     - Carcinogenic (cancer) risk assessment involves two steps:
       1. Perform qualitative evaluation of all epidemiology studies, animal bioassay data, and biological activity.
       2. Quantitation of the risk for substances classified as definite or probably human carcinogens.
     - Non-carcinogenic risk assessment includes:
       - Acceptable Daily Intake (ADI), which divides the NOAEL by uncertainty/safety factors.
       - Reference Dose (RfD), which divides the NOAEL or LOAEL by uncertainty/safety factors.
       - Benchmark Dose Method (BMD), which extrapolates data to determine a point of departure (POD) that accounts for study quality.
       - Assessments for noncancer toxicity effects, acute or short-term exposures, and occupational exposures.

3. **Exposure Assessment**
    - People are exposed to mixtures of hundreds of chemicals in everyday life.
    - An **exposure pathway** describes the:
        - Route a substance takes from its source to its endpoint.
        - How people can be exposed to the substance.
    - The three steps of exposure assessment are to:
        1. Characterize the point of exposure setting and exposure scenario.
        2. Identify exposure pathways.
        3. Quantify the exposure.
    - Exposure models are commonly used because actual exposure measurements are often not available.
4. **Risk Characterization**
    - This final phase predicts the frequency and severity of effects in exposed populations.
    - Biological and statistical uncertainties are described.
    - For carcinogenic risks, the probability of a person developing cancer over a lifetime is estimated by multiplying the cancer slope factor for the substance by the chronic, 70-year average daily intake.
    - For noncarcinogenic effects, the exposure level is compared with an ADI, RfD, or MRL derived for similar exposure periods.

This page titled 8.4.6.6: Risk Assessment is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 6: Risk Assessment by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.6.1: Risk Assessment

Did you know?

For many years, the terminology and methods used in human risk or hazard assessment were inconsistent, which led to confusion among scientists, the public, and others.

Illustration of a magnifying glass over the word 'Risk'

(Image Source: iStock Photos, ©)

"Red Book" for Risk Assessment (1983)

In 1983, the (U.S.) National Academy of Sciences (NAS) published **Risk Assessment in the Federal Government: Managing the Process**. Often called the "Red Book" by toxicologists and others, it addressed the standard terminology and concepts for risk assessments.

Cover of the publication, 'Risk Assessment in the Federal Government: Managing the Progress,' commonly known as the 'Red Book'

Figure 8.4.6.6.1.1. Toxicology-based approaches to hazard identification, dose-response assessment, exposure analysis, and characterization of risks were described in the 1983 Red Book
(Image Source: National Academies Press)

Key Terms

The following terms are routinely used in risk assessments:

- **Hazard** — capability of a substance to cause an adverse effect.
- **Risk** — probability that the hazard will occur under specific exposure conditions.
- **Risk assessment** — the process by which hazard, exposure, and risk are determined.
- **Risk management** — the process of weighing policy alternatives and selecting the most appropriate regulatory action based on the results of risk assessment and social, economic, and political concerns.

Risk Assessment Steps

The four basic steps in the risk assessment process as defined by the NAS are:

1. **Hazard identification** — characterization of innate adverse toxic effects of agents.
2. **Dose-response assessment** — characterization of the relation between doses and incidences of adverse effects in exposed populations.
3. **Exposure assessment** — measurement or estimation of the intensity, frequency, and duration of human exposures to agents.
4. **Risk characterization** — estimation of the incidence of health effects under the various conditions of human exposure.

Once risks are characterized in step 4, the process of risk management begins (Figure 2).

Illustration of the relationship between risk assessment and risk management, each represented by a circle that overlaps. The circle on the left represents risk assessment and begins with hazard identification, followed by dose-response assessment and exposure assesment, each leading to the overlapping portion of risk characterization. The circle on the right represents risk management. Control alternatives, determination of the acceptable risk level, and outcomes of risk characterization lead to control decision. A feedback loop also connects from risk management to risk assessment.

Figure 8.4.6.6.1.2. *Interaction between processes of risk assessment and risk management*
(Image Source: ORAU, ©)

"Silver Book" for Advancing Risk Assessment (2009)

A newer book by the NAS, **Science and Decisions: Advancing Risk Assessment** (2009), often called the “Silver Book” by toxicologists and others, emphasizes uncertainty and variability and cumulative risk, and notes that risk assessment "is at a crossroads."

Cover of the publication, 'Science and Decisions: Advancing Risk Assessment,' commonly called the 'Silver Book'

Figure 8.4.6.6.1.3. *The 2009 Silver Book includes approaches for improving risk analysis and a framework for risk-based decision-making*
(Image Source: National Academies Press)

Risk-Based Decision Making

The co-authors of this Silver Book proposed a framework for risk-based decision-making (Figure 4). The framework consists of three phases:

Enhanced problem formulation and scoping — available risk-management options are identified.

Phase 1 - Enhanced problem formulation and scoping

Planning and assessment — risk-assessment tools are used to determine risks under existing conditions and under potential risk-management options.

Phase 2 - Planning and assessment

Risk management — risk and non-risk information is integrated to inform choices among options.

Phase 3 - Risk management

The core of the framework, as noted in the Silver Book, includes the risk assessment paradigm of the Red Book, but differs primarily in its initial and final steps:

- "The framework systematically identifies problems and options that risk assessors should evaluate at the earliest stages of decision-making."
- "It expands the array of impacts assessed beyond individual effects (for example, cancer, respiratory problems, and individual species) to include broader questions of health status and ecosystem protection."
- "It provides a formal process for stakeholder involvement throughout all stages but has time constraints to ensure that decisions are made."
- "It increases understanding of the strengths and limitations of risk assessment by decision-makers at all levels, for example, by making uncertainties and choices more transparent."

The framework for risk-based decision-making includes three phases: problem formulation and scoping, planning and conduct of risk assessment, and risk management. Stage 1 involves planning, stage 2 involves risk assessment, and stage 3 is confirmation of utility. There are formal provisions for internal and external stakeholder involvement at all stages.

Figure 8.4.6.6.1.4. *Framework for risk-based decision-making (Source: "Silver Book," chapter 8)*

Latest Approaches

Other parts of ToxTutor highlight the latest approaches used in risk assessment. For example:

- **Hazard Identification** describes the emerging approach to hazard identification called **Adverse Outcome Pathways (AOPs)**.
- **Testing for, and Assessing Toxicity** describes the numerous global efforts made, since approximately 1990, to reduce and replace use of laboratory animals in toxicology studies. Included are efforts to develop and validate *in vitro* methods and the emerging use of *in silico* methods.

Knowledge Check

Risk is the:

- ◯ Capability of a substance to cause an adverse effect
- ◯ Weighing of possible alternatives and selecting the most appropriate regulatory actions
- ◯ Probability that a hazard will occur under specific exposure conditions

Answer available at the end of this section.

In the risk assessment process, what happens during the hazard identification step?

- ◯ Characterization of the relation between doses and incidences of adverse effects
- ◯ Characterization of innate adverse toxic effects of agents
- ◯ Measurement or estimation of intensity, frequency, and duration of human exposures to agents
- ◯ Estimation of the incidence of health effects under the various conditions of human exposure

Answer available at the end of this section.

What are the phases of the risk-based decision-making framework proposed by the co-authors of the "Silver Book?"

- ◯ Enhanced problem formulation and scoping; planning and assessment; and risk management
- ◯ Hazard identification; dose-response assessment; exposure assessment; risk characterization
- ◯ Hazard identification; risk assessment; action planning

Answer available at the end of this section.

Answers

Knowledge Check

Probability that a hazard will occur under specific exposure conditions - **This is the correct answer.**
Risk is the *probability* that a hazard will occur.

Knowledge Check

Characterization of innate adverse toxic effects of agents - **This is the correct answer.**
Hazard identification is the first step in the risk assessment process as defined by the National Academy of Sciences.

Knowledge Check

Enhanced problem formulation and scoping; planning and assessment; and risk management - **This is the correct answer.**
The framework proposed by the co-authors of the "Silver Book" involves enhanced problem formulation and scoping; planning and assessment; and risk management.

This page titled 8.4.6.6.1: Risk Assessment is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 6.1: Risk Assessment by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.6.2: Hazard Identification

The goal of hazard identification in toxicology is to identify or develop information suggesting or confirming that a chemical (or, for example, a consumer product) poses or does not pose a potential hazard to humans.

During earlier years of toxicology, this process relied primarily on human epidemiology data and on various types of animal testing data, supplemented in more recent years with the development of *in vitro* methods such as those focused on assessing the potential for mutations and DNA damage. The future of hazard identification is promising and toxicologists now have various types of *in vitro* methods to explore for hazard identification, along with the emergence of "*chip*" approaches.

Hazard identification is the first step in Risk Assessment

Figure 8.4.6.6.2.1. Hazard identification is the first component of risk assessment
(Image Source: ORAU, ©)

These emerging methods are based, in part, on **(Quantitative) Structure Activity, or (Q)SAR methods**. Q(SAR) methods, such as computer models, help toxicologists and others to consider closely related chemicals as a group, or chemical category, rather than as individual chemicals. Not every chemical needs to be tested for every toxicity endpoint, and the data for chemicals and endpoints that have been tested are used to estimate the corresponding properties for other chemicals and endpoints of interest. Data from a chemical category must be judged as adequate to support at least a "screening-level" hazard identification.

One approach involves using endpoint information for one chemical to predict the same endpoint for another chemical that is considered "similar" in some way (such as having structural similarity and similar properties and/or activities).

Read-Across

Another approach for hazard identification used since about 2000 is **read-across**. Read-across can be qualitative or quantitative:

- In **qualitative read-across**, the presence (or absence) of a property/activity such as a particular type of toxic effect for the chemical of interest is inferred from the presence (or absence) of the same property/activity for one or more other chemicals. This qualitative approach provides a "yes/no" answer.
- **Quantitative read-across** uses information for one or more chemicals to estimate what the chemical of interest will be like. Thus, quantitative read-across can be used to obtain a quantitative value for an endpoint, such as a dose-response relationship.

Adverse Outcome Pathways (AOPs)

An emerging approach to hazard identification is the use of **Adverse Outcome Pathways (AOPs)**. AOPs reflect the move in toxicity testing from high-dose studies in laboratory animals to *in vitro* methods that evaluate changes in normal cellular signaling pathways using human-relevant cells or tissues. The AOP concept has emerged as a framework for connecting high throughput toxicity testing (HTT, or high throughput toxicity screening, HTS) and other results.

AOP Learning Channel

The **Human Toxicology Project Consortium** provides a collection of informational videos about AOPs for your further exploration. These videos are available on the **AOP Learning Channel**.

Other Computer Models

Another emerging term is **(quantitative) *in vitro* to *in vivo* extrapolation, or (Q)IVIVE**, used together with what are being called **Integrated Testing Strategies** and Integrated Approaches to Testing and Assessment (IATA).

Toxicology Testing in the 21st Century - A New Strategy

The High Throughput Screening (HTS) Initiative is part of the new toxicology testing strategy developed from the 2004 National Toxicology Program (NTP) **Vision and Roadmap for the 21st Century**.

Traditional toxicological testing is based largely on the use of laboratory animals. However, this approach suffers from low throughput, high cost, and difficulties inherent to inter-species extrapolation – making it of limited use in evaluating the very large number of chemicals with inadequate toxicological data.

NTP recognized that the dramatic technological advances in molecular biology and computer science offered an opportunity to use *in vitro* biochemical- and cell-based assays and non-rodent animal models for toxicological testing. These assays allow for much higher throughput at a much reduced cost. In some assays, many thousands of chemicals can be tested simultaneously in days.

The goal is to move toxicology from a predominantly observational science at the level of disease-specific models to a predominantly predictive science focused upon broad inclusion of target-specific, mechanism-based, biological observations.

The High Throughput Screening program represents a new paradigm in toxicological testing. The HTS program approach to toxicological testing screens for mechanistic targets active within cellular pathways considered critical to adverse health effects such as carcinogenicity, reproductive and developmental toxicity, genotoxicity, neurotoxicity, and immunotoxicity in humans.

National Toxicology Program, A National Toxicology Program for the 21st Century

Figure 8.4.6.6.2.2. *National Toxicology Program vision and roadmap*
(Image Source: National Toxicology Program)

Goals of the HTS Program

- To prioritize substances for further in-depth toxicological evaluation.
- To identify mechanisms of action for further investigation (for example, disease-associated pathways).
- To develop predictive models for *in vivo* biological response (predictive toxicology).

Reference:
National Toxicology Program. (2016, March 21). *Tox 21*. U.S. Department of Health and Human Services. Retrieved from **http://ntp.niehs.nih.gov/results/tox21/index.html**

As described in the ***Testing for, and Assessing Toxicity*** section, the EPA is developing "Virtual Tissue Models" such as the Virtual Embryo (v-Embryo™). These types of advanced computer models are being designed to be capable of simulating how chemicals may affect human development and will help reduce dependence on animal study data. They will also provide faster ways of developing chemical risk assessments.

Finally, also noted in the *Testing for and Assessing Toxicity* section, emerging in the toxicologist's tool box are "chip" models (for example, an "organ on a chip"). One example is the "Lung-on-a-chip" that "...offers a new *in vitro* approach to drug screening by mimicking the complicated mechanical and biochemical behaviors of a human lung."

A human hand is shown holding the computer chip containing lung data.

Figure 8.4.6.6.2.3. *Lung-on-a-chip used to mimic pulmonary edema*
(Image Source: The Wyss Institute for Biologically Inspired Engineering)

Knowledge Check

- ○ Extensive animal testing for toxicity
- ○ Computer models like (Q)SAR

◯Risk assessment strategies

Answer available at the end of this section.

Adverse Outcome Pathways (AOPs) are methods of hazard identification that:

◯Evaluate changes in normal cellular signaling pathways using human-relevant cells or tissues

◯Identify adverse outcomes in test subjects administered increasing doses of potential toxicants

◯Measure the intensity, frequency, and duration of human exposures to agents

◯Evaluate the health effects under various conditions of human exposure

Answer available at the end of this section.

Can quantitative read-across be used to determine the value of an endpoint, such as dose-response relationship?

◯Yes

◯No

Answer available at the end of this section.

Answers

Knowledge Check

Computer models like (Q)SAR - **This is the correct answer.**
Part of the basis for emerging approaches to hazard identification, such as assessing for potential mutations and DNA damage, relies on (Quantitative) Structure Activity (Q)SAR methods.

Knowledge Check

Evaluate changes in normal cellular signaling pathways using human-relevant cells or tissues - **This is the correct answer.**
Adverse Outcome Pathways (AOPs) are *in vitro* methods that evaluate changes in normal cellular signaling pathways using human-relevant cells or tissues.

Knowledge Check

Yes - **This is the correct answer.**
Quantitative read-across can lead to a measurable value for an endpoint, such as a dose-response relationship.

This page titled 8.4.6.6.2: Hazard Identification is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 6.2: Hazard Identification by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.6.3: Dose-Response Assessment

Dose-Response Assessment

The dose-response assessment step of the risk assessment process quantitates the hazards that were identified in the previous step. It determines the relationship between dose and incidence of effects in humans. There are normally two major extrapolations required:

1. From high experimental doses to low environmental doses.
2. From animal doses to human doses.

The procedures used to extrapolate from high to low doses are different for assessing carcinogenic effects and noncarcinogenic effects:

- **Carcinogenic effects** in general are not considered to have a threshold and mathematical models are generally used to provide estimates of carcinogenic risk at very low dose levels.
- **Noncarcinogenic effects** *(for example neurotoxicity)* are considered to have dose thresholds below which the effect does not occur. The lowest dose with an effect in animal or human studies is divided by safety factors to provide a margin of safety.

Risk assessment and risk management process, with the dose-response assessment phase highlighted

Figure 8.4.6.6.3.1. Dose-response assessment is a step in the risk assessment process
(Image Source: ORAU, ©)

Carcinogen (Cancer) Risk Assessment

Cancer risk assessment involves two steps:

1. **Perform qualitative evaluation of all epidemiology studies, animal bioassay data, and biological activity *(for example, mutagenicity)*.**The substance is classified as to its carcinogenic risk to humans based on the weight of evidence. If the evidence is sufficient, the substance may be classified as a definite, probable or possible human carcinogen.
2. **Quantitate the risk for those substances classified as definite or probable human carcinogens.** Mathematical models are used to extrapolate from the high experimental doses to the lower environmental doses.

The two primary cancer classification schemes are those of the **Environmental Protection Agency** (EPA) and the **International Agency for Research on Cancer** (IARC). The EPA and IARC classification systems are quite similar.

1. Qualitative Evaluation of Cancer Risk

The EPA's cancer assessment procedures have been used by several Federal and State agencies. The Agency for Toxic Substances and Disease Registry (ATSDR) relies on EPA's carcinogen assessments. A substance is assigned to one of five descriptors shown below in Table 8.4.6.6.3.1.

Descriptor	Definition
Carcinogenic to Humans	Strong evidence of human carcinogenicity
Likely to Be Carcinogenic to Humans	Evidence is adequate to demonstrate carcinogenic potential to humans but does not reach the weight of evidence for the "Carcinogenic to Humans" descriptor.
Suggestive Evidence of Carcinogenic Potential	The weight of evidence is suggestive of carcinogenicity; a concern for potential carcinogenic effects in humans is raised, but the data are judged insufficient for a stronger conclusion.

Inadequate Information to Assess Carcinogenic Potential	Available data are judged inadequate for applying one of the other descriptors. Additional studies generally would be expected to provide further insights.
Not Likely to Be Carcinogenic to Humans	Available data are considered robust for deciding that there is no basis for a substance to be considered a human carcinogen.

Table 8.4.6.6.3.1. *Hazard Descriptors from the EPA's* ***Guidelines for Carcinogen Risk Assessment*** *(March 2005)*

Cancer Data for Humans

The basis for **sufficient human evidence** is an epidemiology study that clearly demonstrates a causal relationship between exposure to the substance and cancer in humans.

The data are determined to be **limited evidence in humans** if there are alternative explanations for the observed effect.

The data are considered to be **inadequate evidence in humans** if no satisfactory epidemiology studies exist.

Cancer Data for Animals

An increase in cancer in more than one species or strain of laboratory animals or in more than one experiment is considered **sufficient evidence in animals**. Data from a single experiment can also be considered sufficient animal evidence if there is a high incidence or unusual type of tumor induced. Normally, however, a carcinogenic response in only one species, strain, or study is considered as only **limited evidence in animals**.

2. Quantitative Evaluation of Cancer Risk

When an agent is classified as a Human or Probable Human Carcinogen, it is then subjected to a **quantitative risk assessment**. For those designated as a Possible Human Carcinogen, the risk assessor can determine on a case-by-case basis whether a quantitative risk assessment is warranted.

The key risk assessment parameter derived from the EPA carcinogen risk assessment is the **cancer slope factor**. This is a toxicity value that quantitatively defines the relationship between dose and response. The cancer slope factor is a plausible upper-bound estimate of the probability that an individual will develop cancer if exposed to a chemical for a lifetime of 70 years. The cancer slope factor is expressed as mg/kg/day.

Linearized Multistage Model (LMS)

Mathematical models are used to extrapolate from animal bioassay or epidemiology data to predict low-dose risk. Most assume linearity with a zero threshold dose.

A dose-response curve is shown for a substance based on actual test results. The lowest dose that caused cancer is marked. Linear extrapolation is then used through zero threshold dose from upper confidence level of lowest dose that caused cancer.

Figure 8.4.6.6.3.2. *The Linearized Multistage Model is used to extrapolate cancer risk from a dose-response curve using the cancer slope factor*
(Image Source: NLM)

EPA uses the **Linearized Multistage Model (LMS)** illustrated in Figure 2 to conduct its cancer risk assessments. It yields a cancer slope factor, known as the **q1*** (pronounced "Q1-star"), which can be used to predict cancer risk at a specific dose. It assumes linear extrapolation with a zero dose threshold from the upper confidence level of the lowest dose that produced cancer in an animal test or in a human epidemiology study.

Other Models

Other models that have been used for cancer assessments include:

- **One-hit model**, which assumes there is a single stage for cancer and that one molecular event induces a cell transformation. This is a very conservative model.
- **Multi-hit model**, which assumes several interactions are needed before a cell can be transformed. This is one of the least conservative models.
- **Probit model**, which assumes log normal distribution (Probit) for tolerances of exposed population. This model is sometimes used, but generally considered inappropriate for assessing cancer risk.
- **Physiologically Based Pharmacokinetic (PBPK) Models**, which incorporate pharmacokinetic and mechanistic data into the extrapolation process. This model requires extensive data and is becoming commonly used.

Application of Models to Estimate Chemical Concentrations in Drinking Water

The chemical chlordane has been found to cause a lifetime risk of one cancer death in a million persons. Different cancer risk assessment models vary in their estimates of drinking water concentrations for chlordane as illustrated in Table 8.4.6.6.3.2:

Model	Concentration (µg/L)
Probit	50
Multi-hit	2
Linearized multistage	0.07
One-hit	0.03

Table 8.4.6.6.3.2. Estimates of drinking water chlordane concentrations by various cancer assessment models

PBPK models are relatively new and are being employed when biological data are available. They quantitate the absorption of a foreign substance, its distribution, metabolism, tissue compartments, and elimination. Some compartments store the chemical (such as bone and adipose tissue) whereas others biotransform or eliminate it (such as liver or kidney). All these biological parameters are used to derive the target dose and comparable human doses.

Noncarcinogenic Risk Assessment

Historically, the **Acceptable Daily Intake (ADI)** procedure has been used to calculate permissible chronic exposure levels for humans based on noncarcinogenic effects. The ADI is the amount of a chemical to which a person can be exposed each day for a long time (usually lifetime) without suffering harmful effects. It is determined by applying safety factors (to account for the uncertainty in the data) to the highest dose in human or animal studies that has been demonstrated not to cause toxicity (NOAEL).

The EPA has slightly modified the ADI approach and calculates a **Reference Dose (RfD)** as the acceptable safety level for chronic noncarcinogenic and developmental effects. Similarly, the ATSDR calculates **Minimal Risk Levels** (MRLs) for noncancer endpoints.

The **critical toxic effect** used in the calculation of an ADI, RfD, or MRL is the serious adverse effect that occurs at the lowest exposure level. It may range from lethality to minor toxic effects. It is assumed that humans are as sensitive as the animal species unless evidence indicates otherwise.

Assessment of Chronic Exposures

In determining the ADIs, RfDs or MRLs, the **NOAEL** is divided by safety factors (uncertainty factors) in order to provide a margin of safety for allowable human exposure.

ADI (human dose) equals NOAEL (experimental dose) divded by safety factors

When a NOAEL is not available, a **LOAEL** can be used to calculate the RfD.

An additional safety factor is included if a LOAEL is used. A Modifying Factor of 0.1–10 allows risk assessors to use scientific judgment in upgrading or downgrading the total uncertainty factor based on the reliability and quality of the data. For example,

if a particularly good study is the basis for the risk assessment, a modifying factor of <1 may be used. If a poor study is used, a factor of >1 can be incorporated to compensate for the uncertainty associated with the quality of the study.

Dose-response graph. Two data points are shown near the x-axis, one of which is marked the NOAEL. The threshold dose is then marked, from which a diagonal line is drawn at an approximate 45 degree angle. The LOAEL is also marked.

Figure 8.4.6.6.3.3. *Dose-response curve for noncarcinogenic effects*
(Image Source: NLM)

Figure 3 above shows a dose-response curve for noncarcinogenic effects which also identifies the NOAEL and LOAEL. Any toxic effect might be used for the NOAEL/LOAEL so long as it is the most sensitive toxic effect and considered likely to occur in humans.

The **Uncertainty Factors** or **Safety Factors** used to derive an ADI or RfD are listed in Table 8.4.6.6.3.3.

Situation	Uncertainty/Safety Factor
Human variability	10x
Extrapolation from animals to humans	10x
Use of less than chronic data	10x
Use of LOAEL instead of NOAEL	10x
Modifying factor	0.1—10x

Table 8.4.6.6.3.3. Uncertainty/Safety factors used to derive an Acceptable Daily Intake (ADI) or Reference Dose (RfD)

The modifying factor is used only in deriving EPA Reference Doses. The number of factors included in calculating the ADI or RfD depends upon the study used to provide the appropriate NOAEL or LOAEL.

The general formula for deriving the RfD is:

RfD equals the NOAEL or LOAEL divided by the
product of the uncertainty factors

The more uncertain or unreliable the data become, the higher the total uncertainty factor that is applied. An example of an RfD calculation is provided below. A subchronic animal study with a LOAEL of 50 mg/kg/day was used in the numerator. Uncertainty factors used in the denominator are 10 for human variability, 10 for an animal study, 10 for less than chronic exposure, and 10 for use of an LOAEL instead of a NOAEL.

RfD equals 50 mg/kg/day divided by the product of the uncertainty factors,
which in this case is ten times ten times ten times ten. The result is 0.005
mg/kg/day.

In addition to chronic effects, RfDs can also be derived for other long-term toxic effects, including developmental toxicity.

Traditionally, the **NOAEL method** has been used to determine the **point of departure (POD)** from animal toxicology data for use in risk assessments. However, this approach has limitations such as a strict dependence on the dose selection, dose spacing, and sample size of the study from which the critical effect has been identified. Also, using the NOAEL does not take into consideration the shape of the dose-response curve and other related information.

Benchmark Dose Method

The **benchmark dose (BMD) method**, first proposed as an alternative in the 1980s, addresses many limitations of the NOAEL method. It is less dependent on dose selection and spacing and takes into account the shape of the dose-response curve (Figure 4). In addition, the estimation of a BMD 95% lower bound confidence limit (BMDL) results in a POD that appropriately accounts for study quality (i.e., sample size). With the availability of user-friendly BMD software programs, including the EPA's Benchmark

Dose Software (BMDS), the BMD has become the method of choice for many health organizations worldwide.

A dose-response graph is shown, with the NOAEL and LOAEL indicated. Above and to the left of the dose-response curve is another dose-response curve, in a dotted line, which represents the confidence limit on dose. The BMD and BMDL are marked as the response level as percent or standard deviation units. These values are used to extrapolate dose-response values.

Figure 8.4.6.6.3.4. *Extrapolated values using the benchmark dose method reflect the shape of a dose-response curve (Image Source: EPA)*

Assessment of Noncancer Toxicity Effects

While the Agency for Toxic Substances and Disease Registry (ATSDR) does not conduct cancer risk assessments, it does derive **Minimal Risk Levels (MRLs)** for noncancer toxicity effects (such as birth defects or liver damage). The MRL is defined as an estimate of daily human exposure to a substance that is likely to be without an appreciable risk of adverse effects over a specified duration of exposure. For inhalation or oral routes, MRLs are derived for acute (14 days or less), intermediate (15–364 days), and chronic (365 days or more) durations of exposures.

The method used to derive MRLs is a modification of the EPA's RfD methodology. The primary modification is that the uncertainty factors of 10 may be lower, either 1 or 3, based on scientific judgment. These uncertainty factors are applied for human variability, interspecies variability (extrapolation from animals to humans), and use of a LOAEL instead of NOAEL. As in the case of RfDs, the product of uncertainty factors multiplied together is divided into the NOAEL or LOAEL to derive the MRL.

Assessment of Acute or Short-Term Exposures

Risk assessments are also conducted to derive permissible exposure levels for acute or short-term exposures to chemicals. Health Advisories (HAs) are determined for chemicals in drinking water. HAs are the allowable human exposures for 1 day, 10 days, longer-term, and lifetime durations. The method used to calculate HAs is similar to that for the RfDs using uncertainty factors. Data from toxicity studies with durations of length appropriate to the HA are being developed.

Assessment of Occupational Exposures

For **occupational exposures**, Permissible Exposure Levels (PELs), Threshold Limit Values (TLVs), and National Institute for Occupational Safety and Health (NIOSH) Recommended Exposure Levels (RELs) are developed. They represent dose levels that will not produce adverse health effects from repeated daily exposures in the workplace. The method used to derive is conceptually the same. Safety factors are used to derive the PELs, TLVs, and RELs.

Conversion of Animal Doses to Human Dose Equivalents

Animal doses must be converted to human dose equivalents. The **human dose equivalent** is based on the assumption that different species are equally sensitive to the effects of a substance per unit of body weight or body surface area.

Historically, the FDA used a ratio of body weights of humans to animals to calculate the human dose equivalent. The EPA has used a ratio of surface areas of humans to animals to calculate the human dose equivalent. Some current approaches include multiplying the animal dose by the ratio of human to animal body weight raised to either the 2/3rd or 3/4th power (to convert from body weight to surface area). Toxicologists and risk assessors should check to make sure that the approach they are using is the one mandated or recommended by the regulatory agency of most relevance to their efforts.

Allowable Exposures to Contamination Sources

The last step in risk assessment is to express the risk in terms of allowable exposure to a contaminated source. Risk is expressed in terms of the concentration of the substance in the environment where human contact occurs. For example, the unit for assessing risk in air is risk per -3 g)." tabindex="0">mg/m^3 whereas the unit for assessing risk in drinking water is risk per -3 g)." tabindex="0">mg/L.

For carcinogens, the media risk estimates are calculated by dividing cancer slope factors by 70 kg (average weight of a man) and multiplying by 20 m^3/day (average inhalation rate of an adult) or 2 liters/day (average water consumption rate of an adult).

Knowledge Check

The procedures used to extrapolate from high to low doses primarily depend upon the:

- ◯ Threshold dose of the substance
- ◯ Rate of lethality in laboratory animals
- ◯ Carcinogenicity of the substance

Answer available at the end of this section.

According to EPA, a substance is classified as likely to be carcinogenic to humans when:

- ◯ There is strong evidence of human carcinogenicity
- ◯ Evidence is adequate to demonstrate potential carcinogenicity to humans, but not strongly enough to definitively classify as carcinogenic
- ◯ The weight of evidence suggests human carcinogenicity, but the data are determined not to be sufficient for a stronger conclusion
- ◯ Robust data lead to the conclusion that a substance is clearly carcinogenic to humans

Answer available at the end of this section.

The primary cancer risk assessment model used by the EPA is known as the:

- ◯ Linearized Multistage Model (LMS)
- ◯ Probit Model
- ◯ Physiologically Based Pharmacokinetic Model (PB-PK)

Answer available at the end of this section.

The Acceptable Daily Intake (ADI) is calculated by:

- ◯ Dividing the NOAEL by safety factors
- ◯ Dividing the NOAEL by the LOAEL
- ◯ Multiplying the RfD by a modifying factor
- ◯ Linear extrapolation from the LOAEL to the zero intercept

Answer available at the end of this section.

Animal doses must be converted to human dose equivalents for risk assessment. When doing this, toxicologists and risk assessors must:

- ◯ Multiply the animal dose by the ratio of human to animal body weight raised to the $^2/_3$ power

- ◯ Ensure they use the conversion method mandated or recommended by the regulatory agency most relevant to their efforts
- ◯ Multiply the animal dose by the ratio of human to animal body weight raised to the $^{3}/_{4}$ power

Answer available at the end of this section.

Minimal Risk Levels (MRLs) are derived:

- ◯ Similarly to deriving the RfD, but with a potentially lower uncertainty factor
- ◯ By multiplying the cancer slope factor by the lowest exposure dose
- ◯ By multiplying the LOAEL by safety factors

Answer available at the end of this section.

Answers

Knowledge Check

Genotoxic carcinogenicity of the substance - **This is the correct answer.**
The procedure for extrapolation from high to low doses depend on whether or not the effects are carcinogenic. Carcinogenic effects are not considered to have a threshold dose and mathematical models are used to estimate the risk of carcinogenicity at very low doses. Noncarcinogenic effects are considered to have threshold doses and the margin of safety (MOS) is calculated.

Knowledge Check

Evidence is adequate to demonstrate potential carcinogenicity to humans, but not strongly enough to definitively classify as carcinogenic - **This is the correct answer.**
A substance is classified as likely to be carcinogenic to humans when evidence is adequate to demonstrate carcinogenic potential to humans but does not reach the weight of evidence for the descriptor Carcinogenic to Humans.

Knowledge Check

Linearized Multistage Model (LMS) - **This is the correct answer.**
EPA uses the Linearized Multistage Model (LMS) to conduct its cancer risk assessments, producing the q1* that is used to predict cancer risk at a specific dose.

Knowledge Check

Dividing the NOAEL by safety factors - **This is the correct answer.**
The ADI is calculated by dividing the NOAEL by safety factors.

Knowledge Check

Ensure they use the conversion method mandated or recommended by the regulatory agency most relevant to their efforts - **This is the correct answer.**
Toxicologists and risk assessors should check to ensure they use the approach mandated or recommended by the regulatory agency most relevant to their efforts.

Knowledge Check

Similarly to deriving the RfD, but with a potentially lower uncertainty factor - **This is the correct answer.**
The MRL is calculated much like the RfD, except that the uncertainty factors of 10 may be lower (1 or 3), based on scientific judgment.

This page titled 8.4.6.6.3: Dose-Response Assessment is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 6.3: Dose-Response Assessment by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.6.4: Exposure Assessment

Exposure Assessment

No Exposure = No Risk

An expression used in toxicology is "no exposure = no risk." Exposure assessment is a key step in the risk assessment process because without an exposure, even the most toxic chemical does not present a threat. Our understanding of potential exposures to chemicals has grown significantly since approximately 1980. For example, research has identified previously "missing" sources and pathways of potential indoor air exposures such as chemicals from consumer products or elsewhere that end up in household dust.

Environmental contaminants are analyzed according to their releases, movement and fate in the environment, and the exposed populations. Consumer products and pharmaceuticals are analyzed in terms of reasonably foreseeable potential exposures.

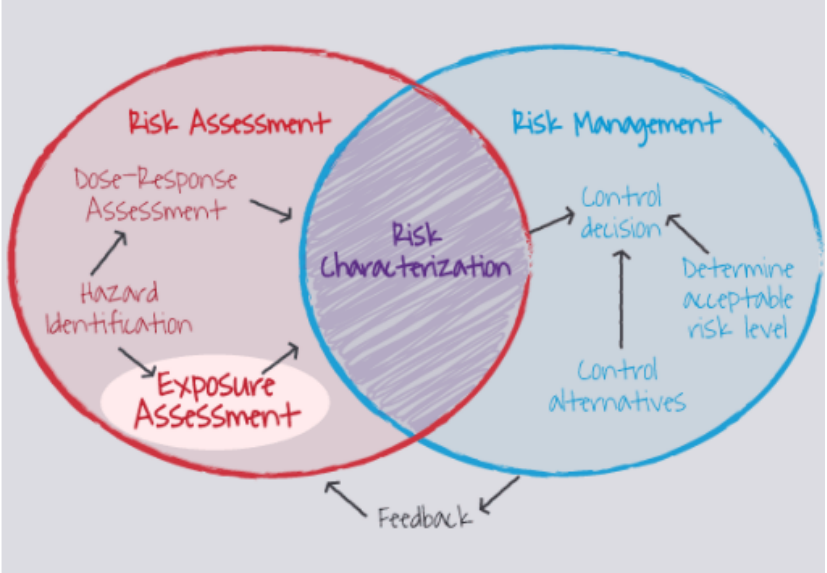


Figure 8.4.6.6.4.1. Exposure assessment as a component of risk assessment (Image Source: ORAU, ©)

Sources of Exposure

Everyday life can involve a person being exposed to mixtures of hundreds of chemicals. Examples of documented non-occupational sources include:

- Consumer products (perhaps 20 or more during a day).
- Clothing.
- Residential and other water.
- Indoor and outdoor air.
- Food and food packaging.
- Beverages.
- Toys.
- Furniture.
- Carpeting, paint, and other building materials.
- Tobacco smoke.
- Household dust containing chemicals from consumer products.
- Outdoor soil and soil tracked indoors.

Figure 8.4.6.6.4.2 highlights potential exposures from a chemical used in furniture fabric.

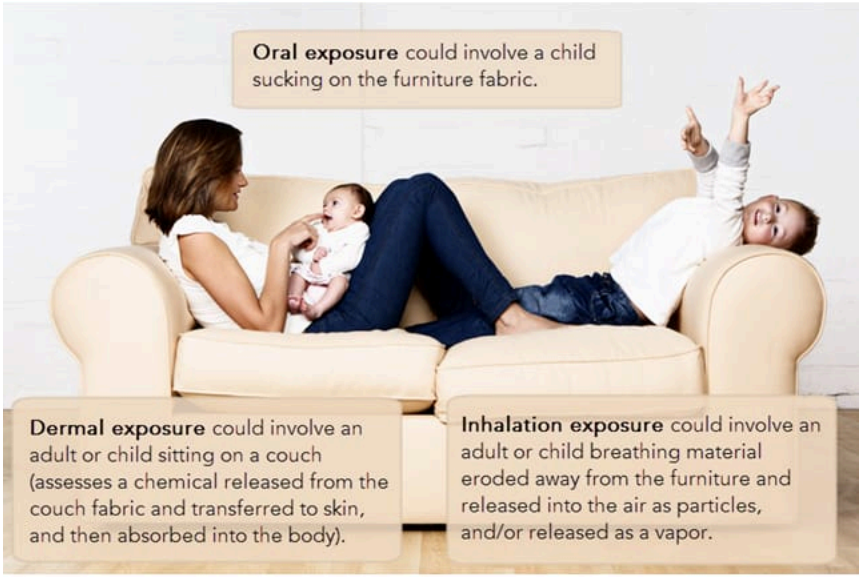


Figure 8.4.6.6.4.2. Exposure to a chemical used in furniture textiles
(Image Source: Adapted from U.S. National Academy Press **publication** and iStock Photos, ©)

Figure 8.4.6.6.4.3 illustrates some of the types of possible consumer exposures from a hand dishwashing product.



Figure 8.4.6.6.4.3. The many uses and types of exposures for a dishwashing product
(Image Source: Adapted from iStock Photos, ©)

Other residential sources of chemicals can come from household air and water. For example, chemicals in air can deposit on, absorb into, or **adsorb** (Adsorption occurs when molecules of a gas, liquid, or dissolved solid adhere to a surface and create a thin film around it) onto household materials (such as carpets and foods and food packaging), which can lead to dermal and oral exposures. When chemicals are confined to indoor spaces and not diluted in outdoor air, there can be large differences in indoor versus outdoor levels of a chemical.

Figure 8.4.6.6.4.4 illustrates some other residential exposures to chemicals via water.

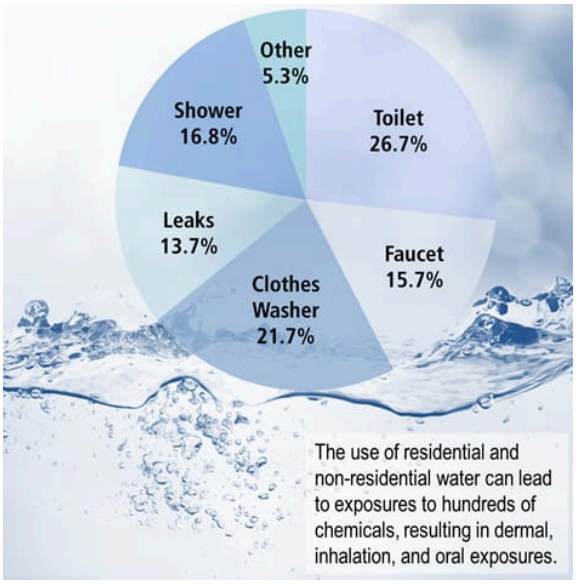


Figure 8.4.6.6.4.4. Water is a household source of chemicals
(Data Source: American Water Works Association Research Foundation, "Residential End Users of Water," 1999.
Learn more at **EPA WaterSense**.
Image Source: Adapted from iStock Photos, ©)

Exposure Pathways

The route a substance takes from its source (where it began) to its endpoint (where it ends), and how people can come into contact with (or be exposed to) it is defined as an **exposure pathway**.
An exposure pathway has five parts:

1. A **source of exposure**, such as using a consumer product for a household task or a chemical spilled from a truck onto a highway.
2. An **environmental media and transport mechanism**, such as movement through the indoor or outdoor air or groundwater.
3. A **point of exposure**, such as a person's house or a private well.
4. A **route of exposure** — eating, drinking, breathing, or touching.
5. A **receptor population** — a person or group of people potentially or actually exposed.

When all five parts are present, the exposure pathway is termed a **completed exposure pathway**.
(Source: **ATSDR Glossary**)

Process of Exposure Assessment

Exposure assessment is a three-step process:



The main variables in the exposure assessment are:

- Exposed populations.
- Types of substances.
- Single substance or mixture of substances.
- Frequency and duration of exposure.
- Pathways and types of exposure.

All possible types of reasonably foreseeable exposures are considered in order to assess the toxicity and risk that might occur.

Considerations for Environmental Exposure

For an environmental exposure, the risk assessor would look at the physical environment and the potentially exposed populations. The physical environment may include considerations about climate, vegetation, soil type, groundwater and surface water. Populations that may be exposed as the result of chemicals that migrate from the site of pollution are also considered.

Subpopulations may be at greater risk due to a higher level of exposure or because they have increased sensitivity. Examples include infants, the elderly, pregnant women, and those with chronic illness.

Pollutants may be transported away from the source and may be physically, chemically, or biologically transformed. They may also accumulate in various materials. Assessment of the chemical fate requires knowledge of many factors, including:

- Organic carbon and water partitioning at equilibrium (Koc).
- Chemical partitioning between soil and water (Kd).
- Partitioning between air and water (Henry's Law Constant).
- Solubility constants.
- Vapor pressures.
- Partitioning between water and octanol (Kow).
- Bioconcentration factors.

These factors are integrated with the data on sources, releases, and routes of the pollutants to determine the exposure pathways of importance, such as groundwater, surface water, air, soil, food, and/or breastmilk.

Use of Exposure Models

Because actual measurements of exposures are often not available, exposure models may be used. For example:

- In air quality studies, chemical emission and air dispersion models are used to predict the air concentrations to downwind residents.
- Residential wells downstream from a site may not currently show signs of contamination. They may become contaminated in the future as chemicals in the groundwater migrate to the well site.

In these situations, groundwater transport models can be used to estimate when chemicals of potential concern will reach the wells.

Information Sources on Chemical Exposures and Health

The future of exposure assessment promises to involve more information and more approaches.
Find out more information about:

- **Chemicals in personal care products** from the National Institute of Environmental Health Sciences (NIEHS).
- **Data connections between the environment and diseases** listed in the Comparative Toxicogenomics Database (CTD).
- **Chemical safety tools and databases** from the Environmental Protection Agency (EPA).
- **Health effects of chemical exposures fact sheet** and **Web site information** from the Agency for Toxic Substances & Disease Registry (ATSDR).
- Publications like "**Exploring Global Exposure Factors Resources for Use in Consumer Exposure Assessments**."

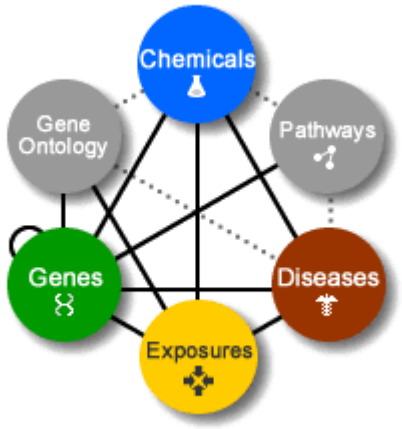


Figure 8.4.6.6.4.5. Topics covered by CTD (Image Source: CTD)

Considerations When Reading About a New Exposure Study

When reading about a new exposure study, there are several questions you can consider to critically evaluate studies about everyday chemical exposures (**source**):

1. Is the study published in a peer-reviewed journal?
2. Are there other publications that lend or do not lend support to the current research, or is this study possibly the first of its kind that suggests an emerging issue for regulators, chemical companies, consumer product manufacturers, and others to consider?
3. In addition to scientific journal publications, is there information available from government or other Web sites that provides perspectives about the exposures and potential risks?
4. How was the study conducted? For example, is it a preliminary or "pilot" study involving a small number of people who were studied, and did the participants represent a narrow or broad, wide range of the types of potentially affected consumers?
5. Did the study use household products/materials or food to which some consumers are likely to be exposed? If yes, are there geographical or other limitations that should be noted by the authors such as the products or food being likely to be sold only in one country or region of the world?

6. Did the study try to approach "reasonably foreseeable" consumer exposure conditions?
7. Is there a known or reasonably foreseeable association between these types of exposures and human adverse effects?

This page titled 8.4.6.6.4: Exposure Assessment is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 6.4: Exposure Assessment by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.6.5: Risk Characterization

Risk Characterization

Approaches to risk characterization continue to evolve. The final stage in the risk assessment process involves predicting the frequency and severity of effects in exposed populations. The conclusions reached from the stages of hazard identification and exposure assessment are integrated to determine the probability of effects likely to occur in humans exposed under similar conditions.

Because most risk assessments include major uncertainties, it is important to describe biological and statistical uncertainties in risk characterization. The assessment should identify which components of the risk assessment process involve the greatest degree of uncertainty.

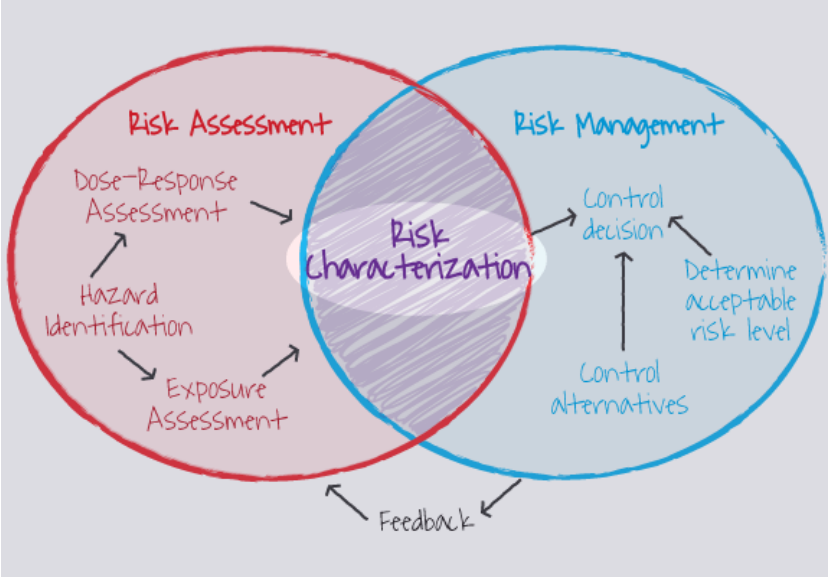


Figure 1. Risk characterization is the final phase of risk assessment (Image Source: ORAU, ©)

For Carcinogenic Risks

Potential human carcinogenic risks associated with chemical exposure are expressed in terms of an increased probability of developing cancer during a person's lifetime. For example, a 10-6 increased cancer risk represents an increased lifetime risk of 1 in 1,000,000 for developing cancer. For carcinogenicity, the probability of an individual developing cancer over a lifetime has historically been estimated by multiplying the cancer slope factor *(mg/kg/day)* for the substance by the chronic *(70-year average)* daily intake *(mg/kg/day)*.

For Noncarcinogenic Effects

For noncarcinogenic effects, the exposure level has historically been compared with an ADI, RfD, or MRL derived for similar exposure periods. Three exposure durations are considered: acute, intermediate, or chronic. For humans:

- **Acute effects** — arise within days to a few weeks.
- **Intermediate effects** — evident in weeks to a year.
- **Chronic effects** — manifest in a year or more.

For Multiple Exposures

In some complex risk assessments, such as for hazardous waste sites, the risk characterization must consider multiple chemical

exposures and **multiple exposure pathways** (described in Exposure Assessments).
Simultaneous exposures to several chemicals, each at a subthreshold level, can often cause adverse effects by "adding" the multiple exposures together, called **dose additivity**.
The assumption of dose additivity is most acceptable when substances induce the same toxic effect by the same mechanism. When available, information on mechanisms of action and chemical interactions are considered and are useful in deriving more scientific risk assessments.
Individuals are often exposed to a substance by more than one exposure pathway *(for example, drinking contaminated water and inhaling contaminated dust)*.

Knowledge Check

1) A major component of exposure assessment involves:

a) Identifying the exposure pathways
b) Measuring the amount of a substance that is metabolized in the body
c) Determining the amount of exposure that must be reduced to comply with the acceptable risk level

Answer available at the end of this section.

2) The movement of substances in environmental media is primarily predicted by:

a) Tagging substances with radioactive tracers and measuring radioactivity at various times and locations within the environmental media
b) Using exposure models to derive scientific estimates
c) Performing actual measurements of exposure pathways

Answer available at the end of this section.

3) Which of the following is not true about risk characterization?

a) It involves predicting the frequency and severity of effects in exposed populations
b) It determines the amount of exposure that must be reduced to comply with the acceptable risk level
c) It integrates conclusions reached in hazard identification and exposure assessment
d) It yields probabilities of effects likely to occur in humans exposed under similar conditions

Answer available at the end of this section.

4) An increased cancer risk of 2.0 X 10^-6 (^ indicates to the -6 power) means that:

a) It is likely that two people out of one million will develop the specific type of cancer in their lifetime due to exposure to the chemical
b) The xeniobiotic for which the cancer risk assessment was performed is likely to cause cancer in two people on a yearly basis
c) It is likely that two people out of one thousand will develop the specific type of cancer in their lifetime due to exposure to the chemical
d) It is probable that two million people will develop cancer if they are continuously exposed to the chemical for life

Answer available at the end of this section.

Answers

Knowledge Check

Identifying the exposure pathways - **This is the correct answer.**
Exposure pathways are key to exposure assessment because they identify the route a substance takes from its source to its end point, as well as how people can be exposed to the substance.

Knowledge Check

Using exposure models to derive scientific estimates - **This is the correct answer.**
Since actual measurements of exposures are often unavailable, exposure models may be used.

Knowledge Check

It determines the amount of exposure that must be reduced to comply with the acceptable risk level - **This is the correct answer.**
Risk characterization involves predicting the frequency and severity of effects in exposed populations. It integrates conclusions reached in hazard identification and exposure to yield probabilities of effects likely to occur in humans exposed under similar conditions.

Knowledge Check

It is likely that two people out of one million will develop the specific type of cancer in their lifetime due to exposure to the chemical - **This is the correct answer.**
An increased cancer risk of 2 times 10^-6 means two in a million people will likely develop the specific type of cancer in their lifetime due to exposure to the chemical.

This page titled 8.4.6.6.5: Risk Characterization is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 6.5: Risk Characterization by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.7: EXPOSURE STANDARDS AND GUIDELINES

Learning Objectives

After completing this lesson, you will be able to:

- Explain the difference between exposure standards and guidelines.
- Identify approaches to regulating consumer products and drug safety.
- Describe standards and guidelines for environmental and occupational exposure.

In this section...

Topics include:

8.4.6.7.1: Exposure Standards and Guidelines

8.4.6.7.2: Regulation of Consumer Products and Drug Safety

8.4.6.7.3: Environmental Exposure Standards/Guidelines

8.4.6.7.4: Occupational (Workplace) Exposure Standards/Guidelines/Approaches

What We've Covered

This section made the following main points:

- **Standards** are legally acceptable exposure levels or controls set by Congressional or Executive mandate.
- **Guidelines** are recommended maximum exposure levels and are voluntary and not legally enforceable.
- Consumer products
  - The U.S. Consumer Product Safety Commission (CPSC) protects the public from unreasonable risks of harm connected with consumer products.
  - The CPSC establishes consumer exposure standards for hazardous substances and articles.
  - The CPSC requires warning labels on containers of household products that are toxic, corrosive, irritating, or sensitizing.
- Drugs
  - FDA approval is required before pharmaceuticals can be marketed.
  - Animal studies and human clinical trials are required to determine toxic dose levels.
  - The New Drug Application (NDA) contains guidance for drug usage and warnings regarding side effects and interactions.
  - Information about a drug's harmful side effects must be provided through labeling and package inserts, publication in the Physicians' Desk Reference (PDR), and direct-to-consumer marketing.
- Food additives
  - The FDA is responsible for approving food additives.
  - **Direct additives** are intentionally added to foods for functional purposes and include processing aids, flavors, appearance agents, and nutritional supplements.
  - **Indirect additives** are not intentionally added to foods and are not natural constituents of foods, but become constituents during production, processing, packaging, and storage.
  - FDA scientists must review new direct food additives before they can be used in foods.
  - Generally Recognized as Safe (GRAS) additives are generally accepted as safe for an intended use and can be introduced into the food supply without prior FDA approval.
- Environment
  - The EPA establishes exposure standards for pesticides, water pollutants, air pollutants, and hazardous wastes.
  - Pesticides must be registered with EPA after undergoing extensive analyses.
  - The EPA prepares health advisories (HAs) as voluntary exposure guidelines for drinking water contamination.
  - Ambient water quality criteria help control pollution sources at the point of release into the environment.

- National Ambient Air Quality Standards (NAAQS) protect public health and welfare from air pollution.
- Hazardous wastes are regulated under the Resource Conservation and Recovery Act (RCRA) and Superfund.
- RCRA regulates hazardous and non-hazardous solid waste.

- Occupational Safety
  - The Occupational Safety and Health Administration (OSHA) establishes legal standards for worker exposure in the United States.
  - Permissible Exposure Limits (PELs) list air concentration limits for chemicals, but not skin absorption or sensitization.
  - Short Term Exposure Limit (STELs) PELs are concentration limits of substances in the air that workers may be exposed to for 15 minutes without adverse effects.
  - Ceiling limits are concentration limits for airborne substances that must not be exceeded.
  - Immediately dangerous to life or health (IDLH) designates an airborne exposure or atmosphere that could lead to death or immediate or delayed permanent adverse health effects.
  - Control banding (CB) determines a control measure based on a band of hazards, such as skin irritation or carcinogenic potential, and exposures.

This page titled 8.4.6.7: Exposure Standards and Guidelines is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 7: Exposure Standards and Guidelines by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.7.1: Exposure Standards and Guidelines

Exposure Standards and Guidelines

Exposure standards and guidelines are developed by governments to protect the public from harmful substances and activities that can cause serious health problems. This section describes standards and guidelines relating to protection from the toxic effects of chemicals only.

Exposure standards and guidelines are determined by risk management decisions. Risk assessments provide regulatory agencies with estimates of numbers of persons potentially harmed under specific exposure conditions. Regulatory agencies then propose exposure standards and guidelines designed to protect the public from "unacceptable risk" levels.

Exposure standards and guidelines usually provide numerical exposure levels for various media (such as food, consumer products, water, and air) that should not be exceeded. Alternatively, these standards may be preventive measures to reduce exposure (such as labeling, special ventilation, protective clothing and equipment, and medical monitoring).



Figure 8.4.6.7.1.1. Standards and guidelines protect the public from harmful substances
(Image Source: iStock Photos, ©)

Standards and Guidelines

More specifically, standards and guidelines for chemical exposure levels consist of the following:

- **Standards** — legally acceptable exposure levels or controls issued as the result of Congressional or Executive mandate. They result from formal rulemaking and are legally enforceable. Violators are subject to punishment, including fines and imprisonment.
- **Guidelines** — recommended maximum exposure levels which are voluntary and not legally enforceable. Guidelines may be developed by regulatory and non-regulatory agencies, or by some professional societies.

Federal and state regulatory agencies have the authority to issue permissible exposure standards and guidelines in the following categories:

- Consumer Product Exposure Standards and Guidelines
- Environmental Exposure Standards and Guidelines
- Occupational Exposure Standards and Guidelines

Knowledge Check

1) Exposure standards are:

Developed by chemical manufacturers
Recommended maximum exposure levels which are voluntary and not legally enforceable
Legally enforceable acceptable exposure levels or controls

Answer available at the end of this section.

2) Exposure guidelines are:

Developed by chemical manufacturers
Recommended maximum exposure levels which are voluntary and not legally enforceable
Legally enforceable acceptable exposure levels or controls

Answer available at the end of this section.

Answers

Knowledge Check

Legally enforceable acceptable exposure levels or controls - **This is the correct answer.**
Exposure standards are legally enforceable acceptable exposure levels or controls resulting from Congressional or Executive mandate.

Knowledge Check

Recommended maximum exposure levels which are voluntary and not legally enforceable - **This is the correct answer.**
Exposure guidelines are recommended maximum exposure levels that are voluntary and not legally enforceable.

This page titled 8.4.6.7.1: Exposure Standards and Guidelines is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 7.1: Exposure Standards and Guidelines by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.7.2: Regulation of Consumer Products and Drug Safety

Regulation of Consumer Products and Drug Safety

Consumer products are often called household products. It is important to know what a category of product is called in the area of the world of interest.

For example, in the United States, cosmetic products are defined by the **FDA** as those products "intended to be applied to the human body for cleansing, beautifying, promoting attractiveness, or altering the appearance without affecting the body's structure or functions." While cosmetics are often thought of as being make-up products such as eye-liner, lipstick, or nail polish, the FDA definition includes skin-care creams, lotions, powders, and perfumes. However, soap products are excluded from FDA's definition of cosmetics.

In comparison, in the European Union (EU), the European Commission's definition of cosmetic products includes soap, shampoo, deodorant, toothpaste, perfumes, and makeup.

It is also important to keep in mind that some products are known by different names. For example, what is known as a cloth or disposable diaper in the United States is called a nappy in several other countries.

Figure 8.4.6.7.2.1. Shopping for consumer products
(Image Source: iStock Photos, ©)

U.S. Consumer Product Safety Commission (CPSC)

The U.S. **Consumer Product Safety Commission** (CPSC) is charged with "protecting the public from unreasonable risks of injury or death associated with the use of the consumer products" under its jurisdiction such as toys, cribs, power tools, cigarette lighters, and household chemical-containing products. The CPSC considers if a product could pose a fire, electrical, chemical, or mechanical (such as choking) hazard. CPSC's work, including research, product recalls, education, and administration of regulations, laws, and standards, has resulted in a decline in the rate of deaths and injuries associated with consumer products over the past several decades.

Consumer exposure standards are developed for hazardous substances and articles by the CPSC. The authority under the **Federal Hazardous Substance Act** (FHSA) pertains to substances other than pesticides, drugs, foods, cosmetics, fuels, and radioactive materials. The CPSC-required warning labels on containers of household products that are toxic, corrosive, irritants, or sensitizers help consumers to safely store and use those products and to give them information about immediate first aid steps to take if an accident happens. Highly toxic substances are labeled with DANGER; less toxic substances are labeled with WARNING or CAUTION.

Figure 8.4.6.7.2.2. CPSC danger label for a gas-powered generator
(Image Source: CPSC)

The CPSC's basis for a determination of highly toxic has been death in laboratory rats at an oral dose of 50 mgs, or an inhaled dose in rats of 200 ppm for one hour, or a 24-hour dermal dose in rabbits of 200 mg/kg. A substance is corrosive if it causes visible destruction or irreversible damage to the skin or eye. If it causes damage that is reversible within 24 hours, it is designated an irritant. An immune response from a standard sensitization test in animals is sufficient for designating the substance a sensitizer.

CPSC is a member of the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (**ICCVAM**), a permanent committee of the National Institutes of Environmental Health Science (NIEHS) under the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (**NICEATM**). ICCVAM was created "to establish, wherever feasible, guidelines, recommendations, and regulations that promote the regulatory acceptance of new or revised scientifically valid toxicological tests that protect human and animal health and the environment while reducing, refining, or

replacing animal tests and ensuring human safety and product effectiveness."

European Commission General Product Safety Directive (GPSD)

In the EU, the European Commission's **General Product Safety Directive** (GPSD) requires producers and distributors to place only safe consumer products on the market and to take all necessary measures to prevent risks to consumers. The GPSD excludes certain product categories covered by specific European safety regulations. It provides an alert system (Rapid Alert System for non-food dangerous products – "RAPEX") between the Commission and EU Member States, Norway, Iceland and Liechtenstein.

U.S. Food and Drug Administration (FDA)

The FDA oversees food safety, tobacco products, dietary supplements, prescription and non-prescription drugs, vaccines, blood products, medical devices, electromagnetic radiation emitting devices, cosmetics, animal foods, and veterinary products.

Drugs

Manufacturers of new pharmaceuticals must obtain formal FDA approval before their products can be marketed. Drugs intended for use in humans must undergo animal studies and human clinical trials to determine toxic dose levels prior to filing a **New Drug Application** (NDA).

Did you know?

During the late 1950s the drug **Thalidomide** was used to control nausea and vomiting in pregnancy in Canada, Europe, Australia, and parts of Asia. After the drug came on the market, reports appeared of children born with missing limbs, with upper limbs usually more affected than lower limbs. In addition to damage to arms and legs, faces, eyes, ears, and genitalia, internal organs including the heart, kidney, and gastrointestinal tract were damaged.

Thalidomide did not reach the US market due to the efforts of an astute FDA drug reviewer, Dr. Frances Oldham Kelsey, who insisted that thalidomide be fully tested before approval. In response to the episode, the Kefauver Harris Amendment to the Federal Food, Drug, and Cosmetic Act became law in 1962. The amendment requires drug manufacturers to give proof of the effectiveness and safety of a drug before it can be approved.

Thalidomide continues to be prescribed under strict supervision as a treatment for multiple myeloma, leprosy, certain complications of human immunodeficiency virus (HIV), and autoimmune disorders. In the 2000s, children with thalidomide-related birth defects were noted in Brazil due to the use of thalidomide to treat leprosy. Pregnant women were being exposed when family members took the drug.

Figure 8.4.6.7.2.3. Thalidomide (Image Source: **NLM Circulating Now**)

An NDA covers all aspects of a drug's effectiveness and safety, including:

- Pharmacokinetics and pharmacological effects.
- Metabolism and postulated mechanism of action.
- Associated risks of the drug.
- Intended uses and therapeutic efficacy.
- Risk:benefit relationship.
- Basis for package inserts supplied to physicians.

The FDA does not issue exposure standards for drugs. Instead, it approves an NDA that contains guidance for usage and warnings concerning potential side effects of a drug. The manufacturer is required to provide this information to physicians prescribing the drug as well as to others that may purchase or use the drug. Information on a drug's harmful side effects is provided in three main ways:

1. Labeling and package inserts that accompany a drug and explain approved uses, recommended dosages, and effects of overexposure.
2. Publication of information in the Physicians' Desk Reference (PDR) and other publically available databases.
3. Direct-to-consumer advertisements.

Figure 8.4.6.7.2.4. Drugs must include labeling and package inserts
(Image Source: iStock Photos, ©)

The package insert labels and the PDR contain the following information:

- Description
- Clinical pharmacology
- Indications and usage
- Contraindications
- Warnings
- Precautions
- Adverse reactions
- Interactions
- Overdosage
- Available forms
- Dosage and administration

Figure 8.4.6.7.2.5. Sample package insert for a fictional prescription drug
(Image Source: **FDA**)

Food Additives

The FDA is responsible for the approval of food additives. Standards are different depending on whether they are direct food additives or indirect food additives.
Direct food additives are intentionally added to foods for functional purposes. Examples of direct food additives include:

- Processing aids
- Texturing agents
- Preservatives
- Flavoring and appearance agents
- Nutritional supplements

Approval usually designates the maximum allowable concentrations in a food product.

Figure 8.4.6.7.2.6. A worker in a food production plant adds a mix of dry flavoring and processing agents to meat
(Image Source: iStock Photos, ©)

Indirect food additives are not intentionally added to foods and they are not natural constituents of foods. They become

a constituent of the food product from environmental contamination during production, processing, packaging, and storage. Examples of indirect food additives are:

- Antibiotics administered to cattle.
- Pesticide residues remaining after production or processing of foods.
- Chemicals that migrate from packaging materials into foods.

Exposure standards indicate the maximum allowable concentration of these substances in food.
New direct food additives must undergo stringent review by FDA scientists before they can be approved for use in foods. The manufacturer of a direct food additive must provide evidence of the safety of the food additive in accordance with specified uses. The safety evaluation is conducted by the toxicity testing and risk assessment procedures previously discussed with derivation of the ADI. In contrast to pharmaceutical testing, virtually all toxicity evaluations are conducted with experimental laboratory animals.
FDA approval of all new food additives has been requires since the Food, Drug and Cosmetic Act (FDCA) was amended in 1958. At that time, all existing food additives were **Generally Recognized as Safe** *(*GRAS*)* and no exposure standard was developed. Many of GRAS substances have since been reevaluated and maximum acceptable levels have been established. However, under the law, a substance may be determined to be GRAS for an intended use and introduced into the food supply as such without prior approval by FDA. FDA maintains a searchable database of **GRAS substances**.

Figure 8.4.6.7.2.7. Select Committee on GRAS Substances searchable database

The FDA reevaluation of GRAS substances requires that specific toxicity tests be conducted based on the level of the GRAS substance in a food product. For example, the lowest level of concern is for an additive used at 0.05 ppm in the food product. Only short-term tests (a few weeks) are required for those compounds. In contrast, a food additive used at levels higher than 1.0 ppm must be tested for carcinogenicity, chronic toxicity, reproductive toxicity, developmental toxicity, and mutagenicity.
The 1958 amendment to the Food, Drug and Cosmetic Act law is known as the Delaney Clause. This clause prohibited the addition of any substance to food that has been shown to induce cancer in man or animals. The implication was that any positive result in an animal test, regardless of dose level or mechanism, is sufficient to prohibit use of the substance. In this case, the allowable exposure level is zero. In 1958, chemical levels could only be measured in parts per thousand whereas analytical methods today allow some chemical levels to be measured down to parts per trillion or quintillion. Such levels might represent negligible cancer risks and Congress has repeatedly amended the Delaney Clause to create more and more exceptions. In 1996, the Delaney Clause was repealed, and the "zero–risk" standard was changed to one of "reasonable certainty."

Food Safety in the European Union (EU)

General Food Law

The basic principles for the EU's food safety policy are defined in the EU's General Food Law (**Regulation (EC) No 178/2002**), adopted in 2002. This regulation:

- Lays down general principles, requirements and procedures that underpin decision making in matters of food and feed safety, covering all stages of food and feed production and distribution.
- Sets up an independent agency responsible for scientific advice and support, the **European Food Safety Authority (EFSA)** - see below for more information.
- Creates the main procedures and tools for the management of emergencies and crises as well as the **Rapid Alert System for Food and Feed (RASFF)**.

Figure 8. Flags of the European Union and various member nations *(Image Source: iStock Photos, ©*

European Food Safety Authority (EFSA)

The **European Food Safety Authority (EFSA)** was set up in 2002 and is based in Parma, in Italy. It carries out risk assessments before certain foods are allowed to be sold in the EU. The agency was legally established by the EU under the **General Food Law - Regulation 178/2002**. The General Food Law created a European food safety system in which responsibility for risk assessment (science) and for risk management (policy) are kept separate. **EFSA** is responsible for the former area, and also has a duty to communicate its scientific findings to the public.

EFSA provides scientific advice to the European Commission and EU countries, to help them take effective decisions to protect consumers. It also plays an essential role in helping the EU respond swiftly to food safety crises. EFSA's remit covers:

- Food and feed safety
- Nutrition
- Animal health and welfare
- Plant protection
- Plant health

It also considers, through environmental risk assessments, the possible impact of the food chain on the biodiversity of plant and animal habitats (discover **EFSA topics**, **EFSA scientific work areas**, and **application resources by food sector area**). EFSA publishes all its scientific outputs, including its scientific opinions, in the **EFSA Journal**. It also issues a range of **supporting publications**.
Most of **EFSA's work** is undertaken in response to requests for scientific advice from the European Commission, the European Parliament and EU Member States. EFSA also carry out scientific work on own initiative, in particular to examine emerging issues and new hazards and to update our assessment methods and approaches. This is known as "self-tasking." EFSA's quality management system (QMS) has been awarded an ISO 9001:2015 certificate, the international standard for quality management.
EFSA's Scientific Panels of experts are responsible for the bulk of EFSA's scientific assessment work. Each of the 10 Panels is dedicated to a different area of the food and feed chain. The Scientific Committee has the task of supporting the work of the Panels on cross-cutting scientific issues. It focuses on developing harmonized risk assessment methodologies in fields where EU-wide approaches are not yet defined.
EFSA staff support the Scientific Panels and Scientific Committee in carrying out most of the Authority's **scientific work**. The membership of EFSA's Scientific Committee and Panels is renewed every three years (see also **Working practices** and **EFSA Strategy 2020 - Trusted science for safe food**).

The EU General Food Law deals with a wide range of issues related to food in general and food safety in particular, including food information and animal welfare. It covers all parts of the food chain from animal feed and food production to processing, storage, transport, import and export, as well as retail sales. It also establishes the principles for risk analysis. These stipulate how, when, and by whom scientific and technical assessments should be carried out in order to ensure that humans, animals, and the environment are properly protected.

EU Food Safety Policy

The EU's food safety policy covers food from farm to fork. The EU food policy comprises:

- Comprehensive legislation on food and animal feed safety and food hygiene.
- Sound scientific advice on which to base decisions.
- Enforcement and checks.

Where specific consumer protection is justified, there may be special rules on:

- Use of pesticides, food supplements, colorings, antibiotics, or hormones.
- Food additives, such as preservatives and flavorings
- Substances in contact with foodstuffs, for example, plastic packaging.
- Labeling of ingredients that may cause allergies.
- Health claims such as "low-fat" or "high-fiber."

The EU's **Rapid Alert System for Food and Feed (RASFF)** was launched in 1979 and allows information on food and feed to be shared quickly and efficiently between all the relevant bodies at national and EU-level. In a similar vein, the **EU Notification System for Plant Health Interceptions (EUROPHYT)** is the EU's notification and rapid alert system for plant products entering and being traded within the EU. It helps to prevent the introduction and spread of plant disease and plant pests.
The EU's **Trade Control and Expert System (TRACES)** is a system for tracking live animals and food and feed of animal origin as they enter the EU and are traded within the EU. It links veterinary authorities across and outside the EU, and enables veterinary services and businesses to react swiftly when a health threat is discovered.

Regulatory Science in the European Union

The following tables describe regulatory science in the EU, including links to the relevant agencies. The list is a work in progress.

Regulatory Science (the table below is best viewed on a computer browser)

Area	Regulatory Science (click links for more information)
Animal health and welfare	European Food Safety Authority
Animal feed	European Food Safety Authority
Biological hazards	European Food Safety Authority
Biocidal Active Substances	European Chemicals Agency
Biocidal Products	European Chemicals Agency
Botanicals	European Food Safety Authority
Consumer products	European Commission General Product Safety Directive (GPSD)
Communicable diseases	European Centre for Disease Prevention and Control
Food additives	European Food Safety Authority
Food ingredients	European Food Safety Authority
Food Colors	European Food Safety Authority
Food Contact Materials	European Food Safety Authority
Food enzymes	European Food Safety Authority
Food Supplements	European Food Safety Authority
Non-plastic food contact materials	European Food Safety Authority
Plastics and plastic recycling	European Food Safety Authority
Smoke Flavorings	European Food Safety Authority
Sweeteners	European Food Safety Authority
Food packaging	European Food Safety Authority
GMO	European Food Safety Authority
Plant Health	European Food Safety Authority
Pesticides	European Food Safety Authority
Nutrition	European Food Safety Authority
Nutrient sources	European Food Safety Authority
Vitamins and minerals	European Food Safety Authority

List above shows regulatory authorities over specific areas of science in the European Union

Regulatory Authorities (the table below is best viewed on a computer browser)

Agency **(click links for more information)**	Area of Authority
European Food Safety Authority (EFSA)	• Animal health and welfare • Biological hazards • Chemical contaminants • Cross-cutting science • Food ingredients and food packaging • GMO • Nutrition • Pesticides • Plant health
European Chemicals Agency (ECHA)	• Chemicals • Biocides
European Medicines Agency (EMEA)	Scientific evaluation, supervision and safety monitoring of medicines
European Centre for Disease Prevention and Control (ECDC)	Communicable diseases

List above shows regulatory authorities and their specific areas of science in the European Union

Cross-Cutting Science (the table below is best viewed on a computer browser)

Area	Regulatory Agency **(click links for more information)**
Chemical mixtures	European Food Safety Authority
Cloning	European Food Safety Authority
Emerging risks	European Food Safety Authority
Endocrine active substances	European Food Safety Authority
Margin of Exposure	European Food Safety Authority
Nanotechnology	European Food Safety Authority
Threshold of Toxicological Concern	European Food Safety Authority
Bee health	European Food Safety Authority
Machine learning	European Food Safety Authority

List above shows regulatory authorities over cross-cutting science in the European Union.

Knowledge Check

1) Consumer exposure standards are developed for hazardous substances and articles by the:
a) U.S. Food and Drug Administration (FDA)
b) U.S. Consumer Product Safety Commission (CPSC)
c) General Product Safety Directive (GPSD)

Answer available at the end of this section.

2) Exposure standards for pharmaceuticals are:
a) Issued by the U.S. Food and Drug Administration (FDA)
b) Developed by the Environmental Protection Agency
c) Recommended guidance developed by the U.S. Food and Drug Administration (FDA)

Answer available at the end of this section.

3) The FDA develops exposure standards for both direct and indirect food additives. Which of the following is an example of an indirect food additive?
a) A preservative added to food products
b) An antibiotic administered to cattle
c) Natural and artificial flavorings
d) A nutritional supplement, such as Vitamin A

Answer available at the end of this section.

4) Under the Delaney Clause of 1958, the FDA:

a) Required physicians to strictly adhere to exposure standards for pharmaceuticals
b) Prohibited the addition of any substance to food that has been shown to induce cancer in humans or animals
c) Authorized the addition of potentially carcinogenic substances to food as long as the concentration is at 0.005 ppm or less

Answer available at the end of this section.

5) In the European Union, what regulatory authority is responsible for chemicals and biocides?

a) European Centre for Disease Prevention and Control (ECDC)
b) European Food Safety Authority (EFSA)
c) European Chemicals Agency (ECHA)

Answer available at the end of this section.

Answers

Knowledge Check

U.S. Consumer Product Safety Commission (CPSC)

Knowledge Check

Recommended guidance developed by the U.S. Food and Drug Administration (FDA)

Knowledge Check

An antibiotic administered to cattle

Knowledge Check

Prohibited the addition of any substance to food that has been shown to induce cancer in humans or animals

Knowledge Check

European Food Safety Authority (EFSA)

This page titled 8.4.6.7.2: Regulation of Consumer Products and Drug Safety is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 7.2: Regulation of Consumer Products and Drug Safety by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.7.3: Environmental Exposure Standards/Guidelines

Environmental Exposure Standards/ Guidelines

The Environmental Protection Agency (EPA) is responsible for US-wide **laws** that require determination and enforcement of environmental exposure standards. In addition, EPA has the authority to prepare recommended exposure guidelines for selected environmental pollutants. Similar organizations exist in US states, and in other countries and groups of countries (such as the European Commission for European Union member countries).
The EPA is responsible for developing exposure standards for:

- Pesticides
- Water pollutants
- Air pollutants
- Hazardous wastes

Pesticides

Pesticides cannot be marketed until they have been registered by the EPA in accordance with the **Federal Insecticide, Fungicide, and Rodenticide Act** (FIFRA). In order to obtain registration, a pesticide must undergo an extensive battery of toxicity tests, chemistry analyses, and environmental fate tests.

The **Food Quality Protection Act** of 1996 (FQPA) changed the way that pesticides were regulated by requiring investigation of:

- Nonoccupational exposure to pesticides.
- The cumulative effects of pesticides having a common toxicity mechanism.
- Any increased susceptibility in infants, children and other sensitive groups.
- Whether the pesticide has endocrine-disrupting effects.

Certain pesticides have been determined by the EPA to pose minimal risk to health or to the environment and are **exempt** from FIFRA registration.
A primary exposure standard for pesticides is the **pesticide tolerance for food use** This standard specifies the pesticide residue allowed to remain in or on each treated food product. The EPA residues risk assessment covers:

- Toxicity.
- Amount used and how often.
- How much residue remains by the time a product reaches the market.
- Other ways of being exposed to the pesticide, if any.

Figure 8.4.6.7.3.1. Dusting crops with pesticides
(Image Source: iStock Photos, ©)

Figure 8.4.6.7.3.2. Aerosol spray cans of pesticide
(Image Source: iStock Photos, ©)

Water Pollutants

Access to safe drinking water and control of water pollution are regulated by the **Safe Drinking Water Act** (SDWA) and the **Clean Water Act** (CWA). Under the SDWA, the EPA conducts risk assessments and issues **Maximum Contaminant Levels** **(MCLs)** for naturally-occurring and man-made contaminants in drinking water. The MCL is the acceptable exposure level which, if exceeded,

requires immediate water treatment to reduce the contaminant level.

In addition to establishing MCLs, the EPA can propose recommended exposure guidance for drinking water contamination. As an interim procedure, **maximum contaminant level goals (MCLGs)** may be recommended for long-term exposures to contaminants in drinking water. Generally, no allowable exposure can be recommended for a carcinogenic chemical. When the MCL is issued, an acceptable exposure level based on a cancer risk assessment may be proposed for the MCL.

EPA prepares **health advisories (HAs)** as voluntary exposure guidelines for drinking water contamination. The HAs provide exposure limits for 1-day, 10-day, longer-term, or lifetime exposure periods. They pertain to both cancer and noncancer risks. The formula used to derive a health advisory differs from that for the ADI or RfD in that the HAs pertain to short-term as well as long-term exposures. In addition, human body weight and drinking water consumption are included in the formula. Data from toxicology studies such as the duration of exposure and the exposure route (such as oral) must be represented in the HAs. For example, the 10-day HA must be based on a NOAEL or LOAEL derivation that was obtained from an animal toxicology study of approximately 10 days duration (routinely 7- to 14-day toxicity studies). A longer-term HA applies to humans drinking contaminated water for up to 7 years (which could represent 10% of a human's 70-year lifespan). Because 90 days is about 10% of a rat's expected lifespan, the 90-day subchronic study with rats is considered appropriate for providing the basis for the longer-term HA assessment. A life-time HA (representing lifetime exposure to a toxicant in drinking water) is also determined for noncarcinogens. The procedure uses the RfD risk assessment with adjustments for body weight of an adult human (70 kg) and drinking water consumption of 2 L/day.	Figure 8.4.6.7.3.3. EPA issues health advisories as voluntary exposure guidelines for drinking water contamination (Image Source: Adapted from EPA)

In addition to drinking water standards, the EPA is authorized under the Clean Water Act (CWA) to issue exposure guidance for controlling pollution in ground water. The intent is to provide clean water for fishing and swimming rather than for drinking purposes. It provides a scheme for controlling the introduction of pollutants into navigable surface water. The recommendations for ground water protection are known as ambient water quality criteria.

The **ambient water quality criteria** are intended to control pollution sources at the point of release into the environment. While these criteria may be less restrictive than the drinking water standards, they usually are the same numeric value. For example, the MCL for drinking water and the ambient water quality criteria for ground water for lead are the same: 0.05 mg/L of water.

Figure 8.4.6.7.3.4. A scientist tests the water for toxicity
(Image Source: iStock Photos, ©)

Air Pollutants

Air emission standards are issued by EPA under the **Clean Air Act** (CAA). The CAA authorizes the issuance of **National Ambient Air Quality Standards (NAAQS)** for air pollution. There are two types of NAAQS:

1. **Primary NAAQS standards** set limits to protect public health, including people at increased risk (preexisting heart or lung disease, children, older adults).
2. **Secondary NAAQS standards** relate to public welfare, such as crops, animals, and structures.

NAAQS have been established for the following major atmospheric pollutants:

- Carbon monoxide
- Sulfur oxide
- Oxides of nitrogen
- Ozone
- Hydrocarbons

- Particulates
- Lead

When air emissions exceed the NAAQS levels, the polluting industry must take action to reduce emissions to acceptable levels.

Figure 8.4.6.7.3.5. A factory releases pollutants into the air
(Image Source: iStock Photos, ©)

Hazardous Wastes

Hazardous wastes are regulated under the **Resource Conservation and Recovery Act** (RCRA) and the **Comprehensive Environmental Response, Compensation and Liability Act** (CERCLA), commonly known as **Superfund**. RCRA regulates hazardous and nonhazardous solid waste, including chemical waste produced by industrial processes, medical waste, and underground storage tanks.

The main purpose of CERCLA is to clean up hazardous waste disposal sites. EPA has established standards known as **Reportable Quantities (RQs)**. Companies must report to EPA any chemical release that exceeds the RQ. RQs evaluate physical, chemical, and toxicological properties of a substance. These are called **primary criteria** and include aquatic toxicity, acute mammalian toxicity (oral, dermal, and inhalation), ignitability, reactivity, chronic toxicity, and potential carcinogenicity. **Secondary criteria** evaluate how a substance degrades in the environment.

Minimal Risk Levels (MRLs) for noncancer toxic effects are derived by the **Agency for Toxic Substances and Disease Registry** (ATSDR), which has a congressional mandate to investigate the health effects of hazardous substances in the environment. MRLs are estimates of daily human exposures that are likely to be without an appreciable risk of adverse effects over a specified duration of exposure. MRLs are derived for acute (14 days or less), intermediate (15–364 days), and chronic (365 days or more) exposures for inhalation or oral routes.

Knowledge Check

1) The exposure standard established by the EPA for the pesticide residue allowed to remain in or on each treated food product is called the:

a) Pesticide tolerance for food use
b) Reportable quantity of pesticides
c) Maximum Contaminant Level (MCL)

Answer available at the end of this section.

2) The EPA establishes exposure standards for natural and man-made contaminants in drinking water. These standards are called:

a) Ambient Water Quality Criteria
b) Maximum Contaminant Level Goals (MCLGs)
c) Maximum Contaminant Levels (MCLs)

Answer available at the end of this section.

3) What is the difference between primary National Ambient Air Quality Standards (NAAQS) and secondary NAAQS?

a) Primary NAAQS relate to public welfare (for example, crops, animals, and structures); secondary NAAQS set limits to protect public health
b) Primary NAAQS set limits to protect public health; secondary NAAQS relate to public welfare
c) Primary NAAQS are legally enforceable; secondary NAAQS are not

Answer available at the end of this section.

4) The Agency for Toxic Substances and Disease Registry (ATSDR) estimates levels for daily human exposure to chemicals that are likely to be without an appreciable risk of adverse effects for specified periods of exposure. These are known as:

a) Minimal Risk Levels (MRLs)
b) Maximum Contaminant Levels (MCLs)
c) Reportable Quantities (RQs)

Answer available at the end of this section.

Answers

Knowledge Check

Pesticide tolerance for food use - **This is the correct answer.**
The pesticide tolerance for food use standard established by the EPA specifies the pesticide residue allowed to remain in or on treated food products.

Knowledge Check

Maximum Contaminant Levels (MCLs)- **This is the correct answer.**
The EPA conducts risk assessments and issues Maximum Contaminant Levels (MCLs) for naturally-occurring and man-made contaminants in drinking water.

Knowledge Check

Primary NAAQS set limits to protect public health; secondary NAAQS relate to public welfare - **This is the correct answer.**
Primary NAAQS set limits to protect public health. Secondary NAAQS relate to public welfare, such as crops, animals, and structures.

Knowledge Check

Minimal Risk Levels (MRLs) - **This is the correct answer.**
Minimal Risk Levels (MRLs) for noncancer toxic effects estimate the daily human exposures that are likely to be without an appreciable risk of adverse effects over a specific duration of exposure.

This page titled 8.4.6.7.3: Environmental Exposure Standards/Guidelines is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 7.3: Environmental Exposure Standards/Guidelines by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.7.4: Occupational (Workplace) Exposure Standards/ Guidelines/Approaches

Occupational Safety and Health Administration (OSHA) Standards

Recommended or mandatory **occupational exposure limits (OELs)** for chemicals exist in many countries. For example, legal standards in the United States are established by the Occupational Safety and Health Administration (OSHA). These standards are known as **Permissible Exposure Limits** (PELs). The majority of PELs were issued after the 1970 **Occupational Safety and Health (OSH) Act**.

OSHA maintains the "Permissible Exposure Limits – Annotated Tables" that contain comparative information taken from federal, state, and professional organizations such as the:

- California Division of Occupational Safety and Health (Cal/OSHA) PELs.
- ACGIH® (formerly known as the **American Conference of Governmental Industrial Hygienists**) Threshold Limit Values (TLVs®).
- **National Institute for Occupational Safety and Health** (NIOSH) Recommended Exposure Limits (RELs).

These tables list air concentration limits for chemicals but do not include notations for skin absorption or sensitization. The PEL Annotated Tables include the following: **Table Z-1**, **Table Z-2**, and **Table Z-3**.

Figure 8.4.6.7.4.1. Screenshot of a portion of OSHA PEL Annotated Table Z-2

Most OSHA PELs are for airborne substances with allowable exposure limits averaged over an 8-hour day in a 40-hour week. This is known as the **Time-Weighted-Average (TWA) PEL**. While adverse effects are not expected to be encountered with repeated exposures at the PEL, OSHA recommends that employers consider using the **alternative occupational exposure limits** because it believes levels above some of the alternative occupational exposure limits may be hazardous to workers even when the exposure levels are in compliance with the relevant PELs.

Short Term Exposure Limits, Ceiling Limits, and Skin Designations

OSHA also issues Short Term Exposure Limit (STELs) PELs, Ceiling Limits, and PELs that carry a skin designation.

- **Short Term Exposure Limit** **(STELs)** -- PEL STELs are concentration limits of substances in the air that a worker may be exposed to for 15 minutes without suffering adverse effects. The 15-minute STEL is usually considerably higher than the 8-hour TWA exposure level.
- **Ceiling Limits** are concentration limits for airborne substances that should never be exceeded.
- A **skin designation** indicates that the substance can be readily absorbed through the skin, eye or mucous membranes, and substantially contribute to the dose that a worker receives from inhalation of the substance. OSHA standards do not include surface contamination criteria or quantifications for skin absorption.

Theoretically, an occupational substance could have PELs as TWA, STEL, and Ceiling Value, and with a skin designation, but that is rare. Usually, an OSHA-regulated substance will have only a PEL as a time-weighted average.

Immediately Dangerous to Life or Health (IDLH)

The **Immediately Dangerous to Life or Health (IDLH)** occupational exposure guideline was developed jointly by the OSHA and NIOSH **Standards Completion Program** in 1974. IDLH represents:

- An airborne exposure "likely to cause death or immediate or delayed permanent adverse health effects or prevent escape from such an environment" (NIOSH).
- "An atmosphere that poses an immediate threat to life, would cause irreversible adverse health effects, or would impair an individual's ability to escape from a dangerous atmosphere" (OSHA).

IDLH values can be used in assigning respiratory protection equipment.

Recommended Exposure Limits and Biological Exposure Indices

The NIOSH **Recommended Exposure Limits** (RELs) are also designated as time-weighted average, short-term exposure limits, and ceiling limits. NIOSH also uses immediately dangerous to life or health (IDLH) values.
ACGIH® also has developed **Biological Exposure Indices (BEIs®)** as guidance values for assessing biological monitoring results (concentration of a chemical in biological media such as blood or urine). The OSHA **Policy Statement** on the Uses of TLVs® and BEIs® provides an overview on using these guidelines.

Control Banding (CB)

Control banding **(CB)** is an emerging area internationally for guiding the assessment and management of workplace chemical risks. CB is a technique that determines a control measure such as dilution by air ventilation or engineering controls based on a range or “band” of hazards such as skin irritation or carcinogenic potential and exposures such as an assessment of a small, medium, or large exposure. It is based on the fact that there are a limited number of control approaches, and a history of having many problems solved in the past. CB is used with other health and safety practices such as chemical substitution. It is not a replacement for the use of experts in occupational safety and health and it does not eliminate the need to perform exposure monitoring.

Knowledge Check

1) The Occupational Safety and Health Administration (OSHA) develops legal standards for workplace exposure. These standards are called:
a) Threshold Limit Values (TLVs)
b) Recommended Exposure Limits (RELs)
c) Permissable Exposure Limits (PELs)

Answer available at the end of this section.

2) What exposure standards can be used to assign respiratory protection equipment?
a) Immediately Dangerous to Life or Health (IDLH)
b) Short Term Exposure Limits (STELs)
c) Biological Exposure Indices (BEIs)

Answer available at the end of this section.

Answers

Knowledge Check

Permissable Exposure Limits (PELs)

OSHA establishes Permissable Exposure Limits, or PELs, which are legal standards for workplace exposure.

Knowledge Check

Immediately Dangerous to Life or Health (IDLH)

Immediately Dangerous to Life or Health (IDLH) values can be used in assigning respiratory protection equipment.

This page titled 8.4.6.7.4: Occupational (Workplace) Exposure Standards/Guidelines/Approaches is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 7.4: Occupational (Workplace) Exposure Standards/Guidelines/Approaches by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.8: BASIC PHYSIOLOGY

Learning Objectives

After completing this lesson, you will be able to:

- Explain the basic functional and structural organization of the human body.
- Describe homeostasis.
- Explain the role of the four types of tissues in the human body.
- Explain the role of physiological chemicals in normal body functions.

In this section...

Topics include:

8.4.6.8.1: Introduction to Basic Physiology

8.4.6.8.2: Homeostasis

8.4.6.8.3: Organs and Organ Systems

8.4.6.8.4: Tissues

8.4.6.8.5: Cells

8.4.6.8.6: Chemicals

Section 8: Key Points

What We've Covered

This section made the following main points:

- In most cases, toxic substances exert their harmful effects directly on specific cells or biochemicals within the affected organ (specific toxic effects).
- Homeostasis is the ability of the body to maintain relative stability and function despite drastic changes in the external environment or one portion of the body. The primary components of homeostasis include:
  - Stimulus — a change in the environment.
  - Receptor — the site within the body that detects or receives the stimulus.
  - Control center — the operational point at which the signals are received, analyzed, and an appropriate response is determined.
  - Effector — the body site where a response is generated.
  - Feedback mechanisms — methods by which the body regulates the response
- The basic structure and functional organization of the human body is:
  Chemicals → Cells → Tissues → Organs → Organ Systems → Organism
- The human body consists of eleven organ systems.
- Tissues in organs are precisely arranged to work in harmony to perform organ functions.
- There are four types of tissues in the body:
  1. Epithelial tissue protects, absorbs, and secretes substances, and detects sensations.
  2. Connective tissue provides support and holds body tissues together.

3. Muscle tissue has the ability to contract.
4. Nerve tissue conducts electrical impulses and conveys information from one area of the body to another.

- The cell is the smallest component of the body that can perform all of the basic life functions.
- Cell components that are most susceptible to cellular damage include the cell membrane, nucleus, ribosomes, peroxisomes, lysosomes, and mitochondria.
- The three categories of physiological chemicals normally functioning in the body are:
  1. Elements — made up of only one atom (examples include hydrogen and oxygen).
  2. Inorganic compounds — simple molecules made up of one or two different elements (examples include water and carbon dioxide).
  3. Organic compounds — contain covalently-bonded carbon and hydrogen and often other elements (examples include DNA, RNA, ATP, and proteins).

This page titled 8.4.6.8: Basic Physiology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 8: Basic Physiology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.8.1: Introduction to Basic Physiology

Introduction to Basic Physiology

In order to understand how toxic substances cause harmful changes in organs, tissues, or cells, knowledge of normal physiology and anatomy is needed. This section is an overview of normal physiology, especially as related to the normal body components and how they function. While we show how some xenobiotics can damage the different body components, detailed examples of toxic cellular and biochemical reactions will be covered in later sections.

Complexity of the Body

The body is immensely complex with numerous components, all of which perform precise functions necessary for the body to maintain health and well-being. Malfunction of any component can result in disease or a breakdown of a portion of the body. Toxic substances can damage an organ or organ system so that it cannot function properly, leading to death or sickness of the organism (for example, liver or kidney failure). However, in nearly all cases, the toxic substance actually exerts its harmful effect directly on specific cells or biochemicals within the affected organ. These cell and chemical changes in turn cause the tissue or organ to malfunction.

Specific Toxic Effects

Most toxic substances are usually specific in their toxic damage to particular tissues or organs, referred to as the "target tissues" or "target organs." Toxic effects may affect only a specific type of cell or biochemical reaction. For example:

- The toxic effect of carbon monoxide is due to its binding to a specific molecule (hemoglobin) of a specific cell (red blood cell).
- Organophosphate toxic substances, which inhibit an enzyme (acetylcholinesterase) responsible for modulating neurotransmission at nerve endings.

Systemic Toxic Effects

On the other hand, the effect of some toxic substances may be generalized and potentially damage all cells and thus all tissues and all organs.

- An example is the production of free radicals by whole body radiation. Radiation interacts with cellular water to produce highly reactive free radicals that can damage cellular components. The result can be a range of effects from the death of the cell, to cell malfunction, and to the failure of normal cell division (for example, cancer).
- An example of a multi-organ chemical toxic substance is lead, which damages several types of cells, including kidney cells, nerve cells, and red blood cells.

The body is a remarkable complex living "machine" consisting of trillions of cells and multitudes of biochemical reactions. Each cell has a specific function and cells work together to promote the health and vitality of the organism. The number and types of toxic reactions are likewise very large. While this tutorial cannot possibly present all these types of cellular and biochemical toxic reactions, it is our goal to provide an overview of the primary toxic mechanisms with a few examples that illustrate these mechanisms. It is important to understand that changes at one level in the body can affect homeostasis at several other levels.

This page titled 8.4.6.8.1: Introduction to Basic Physiology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 8.1: Introduction to Basic Physiology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.8.2: Homeostasis

Homeostasis

Homeostasis is the ability of the body to maintain relative stability and function even though drastic changes may take place in the external environment or in one portion of the body. A series of control mechanisms, some functioning at the organ or tissue level and other centrally controlled, maintain homeostasis. The major central homeostatic controls are the nervous and endocrine systems.
Physical and mental stresses, injury, and disease continually challenge us and any of them can interfere with homeostasis. When the body loses its homeostasis, it may plunge out of control, into dysfunction, illness, and even death. Homeostasis at the tissue, organ, organ system, and organism levels reflects the combined and coordinated actions of many cells. Each cell contributes to maintaining homeostasis.

Maintaining Homeostasis

To maintain homeostasis, the body reacts to an abnormal change (induced by a toxic substance, biological organism, or other stress) and makes certain adjustments to counter the change (a defense mechanism). The primary components responsible for the maintenance of homeostasis include:

- **Stimulus** — a change in the environment, such as an irritant, loss of blood, or presence of a foreign chemical.
- **Receptor** — the site within the body that detects or receives the stimulus, senses the change from normal, and sends signals to the control center.
- **Control center** — the operational point at which the signals are received, analyzed, and an appropriate response is determined. This is sometimes referred to as the integration center since it integrates the signals with other information to determine if a response is needed and the nature of a response.
- **Effector** — the body site where a response is generated, which counters the initial stimulus and thus attempts to maintain homeostasis.
- **Feedback mechanisms** — methods by which the body regulates the degree of response that has been elicited. A negative feedback depresses the stimulus to shut off or reduce the effector response, whereas a positive feedback has the effect of increasing the effector response.

Example: Reaction to a Toxin

An example of a homeostatic mechanism can be illustrated by the body's reaction to a toxin that causes anemia and hypoxia (low tissue oxygen) (Figure 1). The production of red blood cells (**erythropoiesis**) is controlled primarily by the hormone, erythropoietin. When the body goes into a state of hypoxia (the stimulus), it prompts the heme protein (the receptor) that signals the kidney to produce erythropoietin (the effector). This, in turn, stimulates the bone marrow to increase red blood cells and hemoglobin, raising the ability of the blood to transport oxygen and thus raises the tissue oxygen levels in the blood and other tissues. This rise in tissue oxygen levels serves to suppress further erythropoietin synthesis (feedback mechanism). In this example, cells and chemicals interact to produce changes that can either disturb homeostasis or restore homeostasis. Toxic substances that damage the kidney can interfere with the production of erythropoietin or toxic substances that damage the bone marrow can prevent the production of red blood cells. This interferes with the homeostatic mechanism described resulting in anemia.

Figure 8.4.6.8.2.1. Homeostatic mechanism to restore levels of red blood cells
(Image Source: NLM)

Knowledge Check

1) The ability of the body to maintain relative stability and function even though drastic changes may take place in the external environment or in one portion of the body is known as:
a) Physiology
b) Homeostasis
c) Toxicity

Answer available at the end of this section.

2) To maintain homeostasis, the body reacts to an abnormal change *(induced by a toxin, biological organism, or other stress)* and makes certain adjustments to counter the change *(a defense mechanism)*. The component of the homeostasis process which detects the change in the environment is known as the:
a) Effector
b) Stimulus
c) Receptor

Answer available at the end of this section.

Answers

Knowledge Check

Legally enforceable acceptable exposure levels or controls

Knowledge Check

Recommended maximum exposure levels which are voluntary and not legally enforceable

This page titled 8.4.6.8.2: Homeostasis is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 8.2: Homeostasis by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.8.3: Organs and Organ Systems

Organ Systems and Organs

Before one can understand how xenobiotics affect these different body components, it's important to understand normal body components and how they function. For this reason, this section provides a basic overview of anatomy and physiology as it relates to toxicity mechanisms.

Basic Body Structure and Organization

We can think of the basic structure and functional organization of the human body as a pyramid or hierarchical arrangement in which the lowest level of organization (the foundation) consists of cells and chemicals. Organs and organ systems represent the highest levels of the body's organization (Figure 1).

Figure 8.4.6.8.3.1. Pyramid represents a hierarchical organization of human body components
(Image Source: NLM)

Simplified definitions of the various levels of organization within the body are:

- **Organ system** — a group of organs that contribute to specific functions within the body. Examples include:
  - Gastrointestinal system
  - Nervous system
- **Organ** — a group of tissues precisely arranged so that they can work together to perform specific functions. Examples include:
  - Liver
  - Brain
- **Tissue** — a group of cells with similar structure and function. There are only four types of tissues:
  1. Epithelial
  2. Connective
  3. Muscle
  4. Nerve
- **Cell** — the smallest living units in the body. Examples include:
  - Hepatocyte
  - Neuron
- **Chemicals** — atoms or molecules that are the building blocks of all matter. Examples include:
  - Oxygen
  - Protein

Organ Systems of the Human Body

The human body consists of eleven organ systems, each of which contains several specific organs. An organ is a unique anatomic structure consisting of groups of tissues that work in concert to perform specific functions. Table 1 includes the structures and functions of these eleven organ systems.

Organ Systems of the Human Body

The human body consists of eleven organ systems, each of which contains several specific organs. An organ is a unique anatomic structure consisting of groups of tissues that work in concert to perform specific functions. Table 1 includes the structures and functions of these eleven organ systems.

Organ System	Functions	Organs
Integumentary	• Barrier to invading organisms and chemicals • Temperature control	• Skin • Hair • Subcutaneous tissue
Skeletal	• Supports and moves body • Protects internal organs • Mineral storage • Blood formation	• Bones • Cartilage • Ligaments • Bone marrow
Muscular	• Locomotion • Heat production	• Muscles • Tendons
Nervous	• Coordinates activities of other organ systems • Responds to sensations	• Brain • Spinal cord • Nerves • Eyes • Ears
Endocrine	• Regulates body functions by chemicals (hormones)	• Pituitary gland • Parathyroid gland • Thyroid gland • Adrenal gland • Thymus • Pancreas • Gonads
Cardiovascular	• Transports oxygen and nutrients to tissues • Removes waste products	• Heart • Blood • Blood vessels
Lymphatic	• Returns tissue fluid to blood • Defends against foreign organisms	• Spleen • Lymph nodes • Thymus • Lymphatic vessels
Respiratory	• Oxygen/carbon dioxide exchange	• Lungs • Trachea • Larynx • Nasal cavities • Pharynx
Digestive	• Processes foods • Absorption of nutrients into body	• Stomach • Intestinal tract • Liver • Pancreas • Esophagus • Salivary glands
Urinary	• Elimination of wastes • Regulates pH and volume of blood	• Kidneys • Urinary bladder • Urethra
Reproductive	• Produces germ cells (eggs and sperm) • Environment for growth of fetus (female)	• Ovaries • Uterus • Mammary glands • Testes • Prostate gland • External genitalia

Organ Systems of the Human Body

The human body consists of eleven organ systems, each of which contains several specific organs. An organ is a unique anatomic structure consisting of groups of tissues that work in concert to perform specific functions. Table 1 includes the structures and functions of these eleven organ systems.

Organ System	Functions	Organs
Integumentary	• Barrier to invading organisms and chemicals • Temperature control	• Skin • Hair • Subcutaneous tissue
Skeletal	• Supports and moves body • Protects internal organs • Mineral storage • Blood formation	• Bones • Cartilage • Ligaments • Bone marrow
Muscular	• Locomotion • Heat production	• Muscles • Tendons
Nervous	• Coordinates activities of other organ systems • Responds to sensations	• Brain • Spinal cord • Nerves • Eyes • Ears
Endocrine	• Regulates body functions by chemicals (hormones)	• Pituitary gland • Parathyroid gland • Thyroid gland • Adrenal gland • Thymus • Pancreas • Gonads
Cardiovascular	• Transports oxygen and nutrients to tissues • Removes waste products	• Heart • Blood • Blood vessels
Lymphatic	• Returns tissue fluid to blood • Defends against foreign organisms	• Spleen • Lymph nodes • Thymus • Lymphatic vessels
Respiratory	• Oxygen/carbon dioxide exchange	• Lungs • Trachea • Larynx • Nasal cavities • Pharynx
Digestive	• Processes foods • Absorption of nutrients into body	• Stomach • Intestinal tract • Liver • Pancreas • Esophagus • Salivary glands
Urinary	• Elimination of wastes • Regulates pH and volume of blood	• Kidneys • Urinary bladder • Urethra
Reproductive	• Produces germ cells (eggs and sperm) • Environment for growth of fetus (female)	• Ovaries • Uterus • Mammary glands • Testes • Prostate gland • External genitalia

Table 8.4.6.8.3.1. Organ systems of the human body

Knowledge Check

1) Groups of cells with similar structure and function are known as:
a) Tissues
b) Organs
c) Organ systems

Answer available at the end of this section.

2) The organ system that transports oxygen and nutrients to tissues and removes waste products is the:
a) Urinary system
b) Integumentary system
c) Cardiovascular system

Answer available at the end of this section.

3) The organ system that regulates body functions by chemicals (hormones) is known as the:
a) Nervous system
b) Reproductive system
c) Endocrine system

Answer available at the end of this section.

Answers

Knowledge Check

Tissues - **This is the correct answer.**

Tissues are groups of cells with similar structure and function. There are only four types of tissues: epithelial tissue, connective tissue, muscle tissue, and nerve tissue.

Knowledge Check

Cardiovascular system - **This is the correct answer.**

The cardiovascular system functions to transport oxygen and nutrients to tissues and removes waste products. The primary organs are the heart, blood, and blood vessels.

Knowledge Check

Endocrine system - **This is the correct answer.**
The endocrine system functions to regulate body functions by chemicals (hormones). It contains several organs including the pituitary gland, parathyroid gland, thyroid gland, adrenal gland, thymus, pancreas, and gonads.

This page titled 8.4.6.8.3: Organs and Organ Systems is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 8.3: Organs and Organ Systems by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.8.4: Tissues

Tissues

There are four types of tissues dispersed throughout the body, as described below. A type of tissue is not unique for a particular organ and all types of tissue are present in most organs, just as certain types of cells are found in many organs. For example, nerve cells and circulating blood cells are present in virtually all organs.

An "Orchestra" of Tissues

Tissues in organs are precisely arranged so that they can work in harmony in performing organ functions. This is similar to an orchestra that contains various musical instruments, each of which is located in a precise place and contributes exactly at the right time to create harmony. Like musical instruments that are mixed and matched in various types of musical groups, tissues and cells also are present in several different organs and contribute their part to the function of the organ and the maintenance of homeostasis.

Kinds of Tissues in the Body

The four types of tissues are:

1. Epithelial tissue
2. Connective tissue
3. Muscle tissue
4. Nerve tissue

The four types of tissues are similar in that each consists of cells and extracellular materials. However, the types of tissues have different types of cells and differ in the percentage composition of cells and the extracellular materials. Figure 1 illustrates how tissues fit into the hierarchy of body components.

Figure 8.4.6.8.4.1. Hierarchy of body components
(Image Source: NLM)

Epithelial Tissue

Epithelial tissue is specialized to protect, absorb, and secrete substances, as well as detect sensations. It covers every exposed body surface, forms a barrier to the outside world, and controls absorption. Epithelium forms most of the surface of the skin, and the lining of the intestinal, respiratory, and urogenital tracts. Epithelium also lines internal cavities and passageways such as the chest, brain, eye, inner surfaces of blood vessels, the heart, and the inner ear.
Functions of epithelium include:

- Providing physical protection from abrasion, dehydration, and damage by xenobiotics.
- Controlling the permeability of substances in entering or leaving the body.
  - Some epithelia are relatively impermeable; others are readily crossed.
  - Various toxins can damage this epithelial barrier.
- Detecting sensation (sight, smell, taste, equilibrium, and hearing) and conveying this information to the nervous system.
  - For example, touch receptors in the skin respond to pressure by stimulating adjacent sensory nerves.

The epithelium also contains glands and secretes substances such as sweat or digestive enzymes. Others secrete substances into the blood (hormones), such as the pancreas, thyroid, and pituitary gland.
The epithelial cells are classified according to the shape of the cell and the number of cell layers. Three primary cell shapes exist: squamous (flat), cuboidal, and columnar. There are two types of layering: 1) simple and 2) stratified. Figure 2 illustrates these types of epithelial cells.

Figure 8.4.6.8.4.2. Classification of epithelial tissues
(Source: Adapted from iStock Photos, ©)

Connective Tissue

Connective tissues provide support and hold the body tissues together. They contain more intercellular substances than the other tissues. Connective tissues include blood; bone; cartilage; adipose (fat); and the fibrous and areolar (loose) connective tissues that give support to most organs. The blood and lymph vessels are immersed in the connective tissue media of the body. The blood-vascular system is a component of connective tissue.

In addition to connecting, the connective tissue plays a major role in protecting the body from outside invaders. The hematopoietic tissue is a form of connective tissue responsible for the manufacture of all the blood cells and immunological capability. Phagocytes are connective tissue cells and produce antibodies. If invading organisms or xenobiotics get through the epithelial protective barrier, the connective tissue acts to defend against them.

Figure 8.4.6.8.4.3. Connective tissues
(Image Source: Adapted from Wikimedia Commons, Public Domain)

Muscle Tissue

Muscular tissue is specialized for an ability to contract. Muscle cells are elongated and called muscle fibers. When one end of a muscle cell receives a stimulus, a wave of excitation is conducted through the entire cell so that all parts contract in harmony.
There are three types of muscle cells:

1. **Skeletal muscle** — attached to bone and contracts causing the bones to move.
2. **Cardiac muscle** — contracts to force blood out of the heart and around the body.
3. **Smooth muscle** — can be found in several organs, including the digestive tract, reproductive organs, respiratory tract, and the lining of the bladder. Examples of smooth muscle activity are the:
    - Contraction of the bladder to force urine out.
    - Peristaltic movement to move feces down the digestive system.
    - Contraction of smooth muscle in the trachea and bronchi which decreases the size of the air passageway.

Figure 8.4.6.8.4.4. Muscle tissues
(Image Source: Adapted from iStock Photos, ©)

Nerve Tissue

Nervous tissue is specialized with a capability to conduct electrical impulses and convey information from one area of the body to another. Most of the nervous tissue (98%) is located in the central nervous system, the brain, and spinal cord.

There are two types of nervous tissue: 1) neurons and 2) neuroglia. Neurons (Figure 5) actually transmit the impulses. Neuroglia (Figure 6) provide physical support for the neural tissue, control tissue fluids around the neurons, and help defend the neurons from invading organisms and xenobiotics. Receptor nerve endings of neurons react to various kinds of stimuli (for example, light, sound, touch, and pressure) and can transmit waves of excitation from the farthest point in the body to the central nervous system.

Figure 8.4.6.8.4.5. Human neuron anatomy
(Image Source: Adapted from iStock Photos, ©)

Figure 8.4.6.8.4.6. Types of neuroglia
(Image Source: Adapted from Wikimedia Commons. Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". *WikiJournal of Medicine* 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. View original image.)

Knowledge Check

1) There are only four types of tissues in the body. The type of tissue that is specialized to protect, absorb and secrete substances, detect sensations, covers every exposed body surface, and forms a barrier to the outside world is:
a) Nerve tissue
b) Epithelial tissue
c) Connective tissue
d) Muscle tissue

Answer available at the end of this section.

Answers

Knowledge Check

Epithelial tissue

Epithelial tissue is specialized to protect, absorb and secrete substances, and detect sensations. It covers every exposed body surface, forms a barrier to the outside world and controls absorption. Epithelium forms most of the surface of the skin, and the lining of the intestinal, respiratory, and urogenital tracts. Epithelium also lines internal cavities and passageways such as the chest, brain, eye, inner surfaces of blood vessels, and heart and inner ear.

This page titled 8.4.6.8.4: Tissues is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 8.4: Tissues by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.8.5: Cells

Cells

Cells are the smallest component of the body that can perform all of the basic life functions. Each cell performs specialized functions and plays a role in the maintenance of homeostasis. While each cell is an independent entity, it is highly affected by damage to neighboring cells. These various cell types combine to form tissues, which are collections of specialized cells that perform a relatively limited number of functions specific to that type of tissue. Several trillion cells make up the human body. These cells are of various types, which can differ greatly in size, appearance, and function.

Primary Cell Components

While there are approximately 200 types of cells, they all have similar features: cell membrane, cytoplasm, organelles, and nucleus. The only exception is that the mature red blood cell does not contain a nucleus. Toxins can injure any of the components of the cell causing cell death or damage and malfunction.
Figure 1 shows the various components of a composite cell.

Figure 8.4.6.8.5.1. Basic cell structure
(Image Source: adapted from National Cancer Institute - **SEER Training**)

The primary components of a typical cell include the following:

- **Cell membrane** — a phospholipid bilayer which also contains cholesterol and proteins; its functions are to provide support and to control entrance and exit of all materials. We will discuss the structure of the cell membrane and the mechanisms by which chemicals can penetrate or be absorbed into or out of the cell in the **Introduction to Absorption** section later in ToxTutor.
- **Cytoplasm** — a watery solution of minerals, organic molecules, and gases found between the cell membrane and nucleus.
- **Nucleus** — a membrane-bound part of a cell that contains nucleotides, enzymes, and nucleoproteins; the nucleus controls metabolism, protein synthesis, and the storage and processing of genetic information.
- **Cytosol** — the liquid part of the cytoplasm which distributes materials by diffusion throughout the cell.
- **Nucleolus** — a dense region of the nucleus which contains the RNA and DNA; It is the site for rRNA synthesis and assembly of the ribosome components.
- **Endoplasmic reticulum** — an extensive network of membrane-like channels that extends throughout the cytoplasm; it synthesizes secretory products and is responsible for intracellular storage and transport.
- **Ribosomes** — very small structures that consist of RNA and proteins and perform protein synthesis; some ribosomes are fixed (bound to the endoplasmic reticulum) while other ribosomes are free and scattered within the cytoplasm.
- **Mitochondria** — oval organelles bound by a double membrane with inner folds enclosing important metabolic enzymes; they produce nearly all (95%) of the ATP and energy required by the cell.
- **Lysosomes** — vesicles that contain strong digestive enzymes; lysosomes are responsible for the intracellular removal of damaged organelles or pathogens.
- **Peroxisomes** — very small, membrane-bound organelles which contain a large variety of enzymes that perform a diverse set of metabolic functions.
- **Golgi apparatus** — stacks of flattened membranes containing chambers; they synthesize, store, alter, and package secretory products.
- **Centrioles** — there are two centrioles, aligned at right angles, each composed of 9 microtubule triplets; they organize specific fibers of chromosomes during cell division, which move the chromosomes.
- **Cilia** — thread-like projections of the outer layer of the cell membrane, which serve to move substances over the cell surface.

Cell Components Most Susceptible to Xenobiotics

While all components of the cell can be damaged by xenobiotics or body products produced in reaction to the xenobiotics, the components most likely to be involved in cellular damage are the cell membrane, nucleus, ribosomes, peroxisomes, lysosomes, and mitochondria.

Agents that can lead to changes in the permeability of the membrane and the structural integrity of a cell can damage **cell membranes**. The movement of substances through cell membranes is precisely controlled to maintain homeostasis of the cell. Changes in toxin-induced cell membrane permeability may directly cause cell death or make it more susceptible to the entrance of the toxin or to other toxins that follow. The effects in this case may be cell death, altered cell function, or uncontrolled cell division (neoplasia).

Nuclei contain the genetic material of the cell (chromosomes or DNA). Xenobiotics can damage the **nucleus**, which in many cases lead to cell death, by preventing its ability to divide. In other cases, the genetic makeup of the cell may be altered so that the cell loses normal controls that regulate division. That is, it continues to divide and become a neoplasm. How this happens is described in the **Cancer section** of ToxTutor.

Ribosomes use information provided by the nuclear DNA to manufacture proteins. Cells differ in the type of protein they manufacture. For example, the ribosomes of liver cells manufacture blood proteins whereas the ribosomes of fat cells manufacture triglycerides. Ribosomes contain RNA, structurally similar to DNA. Agents capable of damaging DNA may also damage RNA. Thus, toxic damage to ribosomes can interfere with protein synthesis. In the case of damage to liver cell ribosomes, a decrease in blood albumin may result with impairment in the immune system and blood transport.

Lysosomes contain digestive enzymes that normally function in the defense against disease. They can break down bacteria and other materials to produce sugars and amino acids. When xenobiotics damage lysosomes, the enzymes can be released into the cytoplasm where they can rapidly destroy the proteins in the other organelles, a process known as **autolysis**. In some hereditary diseases, the lysosomes of an individual may lack a specific lysosomal enzyme. This can cause a buildup of cellular debris and waste products that is normally disposed of by the lysosomes. In such diseases, known as lysosomal storage diseases, vital cells (such as in heart and brain) may not function normally resulting in the death of the diseased person.

Peroxisomes, which are smaller than lysosomes, also contain enzymes. Peroxisomes normally absorb and neutralize certain toxins such as hydrogen peroxide (H2O2) and alcohol. Liver cells contain considerable peroxisomes that remove and neutralize toxins absorbed from the intestinal tract. Some xenobiotics can stimulate certain cells (especially liver) to increase the number and activity of peroxisomes. This, in turn, can stimulate the cell to divide. The xenobiotics that induce the increase in peroxisomes are known as "peroxisome proliferators." Their role in cancer causation are discussed in the **Cancer section** of ToxTutor.

Mitochondria provide the energy for a cell (required for survival), by a process involving ATP synthesis. If a xenobiotic interferes with this process, the death of the cell will rapidly ensue. Many xenobiotics are mitochondrial poisons.

- Examples of poisons that cause cell death by interfering with mitochondria include cyanide, hydrogen sulfide, cocaine, DDT, and carbon tetrachloride.

Knowledge Check

1) There are several different types of cellular organelles. The very small structures *(fixed to the endoplasmic reticulum or free within the cytoplasm)* that consist of RNA and proteins, and function in protein synthesis, are:

a) Nucleus
b) Peroxisomes
c) Lysosomes
d) Ribosomes

Answer available at the end of this section.

2) The organelle that produces nearly all (95%) of the energy required by the cell is the:

a) Nucleolus
b) Golgi apparatus
c) Mitochondria
d) Centioles

Answer available at the end of this section.

Answers

Knowledge Check

Legally enforceable acceptable exposure levels or controls
Exposure standards are legally enforceable acceptable exposure levels or controls resulting from Congressional or Executive mandate.

Knowledge Check

Recommended maximum exposure levels which are voluntary and not legally enforceable
Exposure guidelines are recommended maximum exposure levels that are voluntary and not legally enforceable.

This page titled 8.4.6.8.5: Cells is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 8.5: Cells by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.8.6: Chemicals

Chemicals

Most toxic effects are initiated by chemical interactions in which a foreign chemical or physical agent interferes with or damages normal chemicals of the body. This interaction results in a body chemical being unable to carry out its function in maintaining homeostasis.
There are many ways this can happen; for example:

- Interference with absorption or disposition of an essential nutrient.
- Interference with nerve transmission (see **Neurotoxicity**).
- Damage to a cell organelle preventing its functioning (see **Cell Damage and Tissue Repair**).

Types of Physiological Chemicals

There are three categories of chemicals normally functioning in the body:

1. **Elements** — substances made up of only one atom; for example:
   - Hydrogen
   - Calcium
   - Singlet oxygen (O)
2. **Inorganic compounds** — simple molecules that usually consist of one or two different elements; for example:
   - Water (H_2O)
   - Carbon dioxide (CO_2)
   - Bimolecular oxygen (O_2)
   - Sodium chloride ($NaCl$)
3. **Organic compounds** — substances that contain covalently-bonded carbon and hydrogen and often other elements; for example:
   - Sugars
   - Lipids
   - Amino acids
   - Proteins

Elements

Elements are components of all chemical compounds. Of the 92 naturally occurring elements, only 20 are normally found in the body. Seven of these, carbon, oxygen, hydrogen, calcium, nitrogen, phosphorous, and sulfur make up approximately 99% of the human body weight. In most cases, the elements are components of inorganic or organic compounds. However, in a few cases the actual elements may enter into chemical reactions in the body, such as oxygen during cell respiration, sodium in neurotransmission, and arsenic and lead in impaired mitochondrial metabolism.

Inorganic Compounds

Inorganic compounds are important in the body and responsible for many simple functions. The major inorganic compounds are water (H_2O), bimolecular oxygen (O_2), carbon dioxide (CO_2), and some acids, bases, and salts. The body is composed of 60–75%

water. Oxygen is required by all cells for cellular metabolism and circulating blood must be well oxygenated for maintenance of life. Carbon dioxide is a waste product of cells and must be eliminated or a serious change in pH can occur, known as acidosis. A balance in acids, bases, and salts must be maintained to assure homeostasis of blood pH and electrolyte balance.

Organic Compounds

Organic compounds are involved in nearly all biochemical activities related to normal cellular metabolism and function. The mechanisms by which xenobiotics cause cellular and biochemical toxicity are predominantly related to changes to organic compounds. The main feature that differentiates organic compounds from inorganic compounds is that organic compounds always contain carbon. Most organic compounds are also relatively large molecules. There are five major categories of organic compounds involved in normal physiology of the body:

1. Carbohydrates
2. Lipids
3. Proteins
4. Nucleic acids
5. High-energy compounds

Figure 8.4.6.8.6.1. Organic compounds are involved in numerous structures and functions of biochemical processes in the body (Image Source: NLM)

Carbohydrates

Most **carbohydrates** serve as sources of energy for the body. They are converted to glucose, which in turn is used by the cells in cellular respiration. Other carbohydrates become incorporated as structural components of genetic macromolecules.

- For example, deoxyribose is part of DNA, the genetic material of chromosomes, and ribose is part of RNA, which regulates protein synthesis.

Lipids

Lipids are essential substances of all cells and serve as a major energy reserve. They may be stored as fatty acids or as triglycerides. Other types of lipids are the steroids and phospholipids.

- Cholesterol is a lipid that is a component of cell membranes and is used to produce sex hormones such as testosterone and estrogen.
- Phospholipids serve as the main components of the phospholipid bilayer cell membrane.

Proteins

The most diverse and abundant of organic compounds in the body is the group of **proteins**. There are about 100,000 different kinds of proteins, which account for about 20% of the body weight. The building blocks for proteins are the 20 amino acids, which contain carbon, hydrogen, oxygen, nitrogen, and sometimes sulfur. Most protein molecules are large and consist of 50–1000 amino acids bonded together in a very precise structural arrangement. Even the slightest change in the protein molecule alters its function.
Proteins perform a large variety of important functions. Some proteins have a structural function such as the protein pores in cell membranes, keratin in skin and hair, collagen in ligaments and tendons, and myosin in muscles.

- Hemoglobin and albumin are proteins that carry oxygen and nutrients in the circulating blood.
- Antibodies and hormones are proteins.

A particularly important group of proteins are the enzymes.
Enzymes, which are catalysts, are compounds that accelerate chemical reactions, without themselves being permanently changed. Each enzyme is specific in that it will catalyze only one type of reaction. Enzymes are vulnerable to damage by xenobiotics and many toxic reactions occur by changing the shape of the enzyme ("denaturation") or by inhibiting the enzyme ("inhibition").

Nucleic Acids

Nucleic acids are large organic compounds that store and process information at the molecular level inside virtually all body cells. Three types of nucleic acids are present:

- Deoxyribonucleic acid (DNA)
- Ribonucleic acid (RNA)
- Adenosine triphosphate (ATP)

Nucleic acids are very large molecules composed of smaller units known as nucleotides. A nucleotide consists of a pentose sugar, a phosphate group, and four nitrogenous bases. The sugar in DNA is deoxyribose while the bases are adenine, guanine, cytosine, and thymine. RNA consists of the sugar, ribose, plus the four bases adenine, guanine, cytosine, and uracil. These two types of molecules are known as the molecules of life. For without them, cells could not reproduce and animal reproduction would not occur.
DNA is in the nucleus and makes up the chromosomes of cells. It is the genetic code for hereditary characteristics. RNA is located in the cytoplasm of cells and regulates protein synthesis, using information provided by the DNA. Some toxic agents can damage the DNA causing a mutation, which can lead to the death of the cell, cancer, birth defects, and hereditary changes in offspring. Damage to the RNA causes impaired protein synthesis, responsible for many types of diseases. Figure 2 shows the structure of DNA and RNA. Note that DNA is double-stranded and known as the double helix. RNA is a single strand of nucleotides.

Figure 8.4.6.8.6.2. Structures of DNA and RNA
(Image Source: Adapted from iStock Photos, ©)

High-Energy Compounds

Adenosine triphosphate (ATP) is the most important high-energy compound. It is a specialized nucleotide located in the cytoplasm of cells that serves as a source of cellular energy. ATP contains adenine (amino acid base), ribose (sugar), and three phosphate groups. ATP is created from adenine diphosphate using the energy released during glucose metabolism. One of the phosphates in ATP can later be released along with energy from the broken bond induced by a cellular enzyme.

Knowledge Check

1) A substance in the body that contains covalently-bonded carbon and hydrogen is an:
a) Organic compound
b) Inorganic compound
c) Element

Answer available at the end of this section.

2) The nucleic acid located in the nucleus, which makes up the chromosomes of cells, is:
a) ATP
b) RNA
c) DNA

Answer available at the end of this section.

Answers

Knowledge Check

Organic compound
Organic compounds contain covalently-bonded carbon and hydrogen and often other elements. For example, sugars, lipids, amino acids, and proteins are organic compounds.

Knowledge Check

DNA
DNA is in the nucleus and makes up the chromosomes of cells. It is the genetic code for hereditary characteristics.

This page titled 8.4.6.8.6: Chemicals is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 8.6: Chemicals by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.9: INTRODUCTION TO TOXICOKINETICS

Learning Objectives

After completing this lesson, you will be able to:

- Define toxicokinetics.
- Summarize the four inter-related processes of toxicokinetics.
- Identify examples of transporter proteins and their role in toxicokinetics.

In this section...

Topics include:

8.4.6.9.1: What is Toxicokinetics

What We've Covered

This section made the following main points:

- Toxicokinetics is essentially the study of how a substance enters the body and what happens to it inside the body.
  - The term "disposition" is often used in place of toxicokinetics to describe how the body disposes of a xenobiotic over time.
- The four inter-related processes of toxicokinetics are:
  1. Absorption — the substance enters the body.
  2. Distribution — the substance moves from the site of entry to other areas of the body.
  3. Biotransformation — the substance is transformed into new chemicals (metabolites).
  4. Excretion — the substance or its metabolites leave the body.
- The disposition of a toxicant and its biological reactivity are the factors that determine the severity of toxicity when a xenobiotic enters the body.

This page titled 8.4.6.9: Introduction to Toxicokinetics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 9: Introduction to Toxicokinetics by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.9.1: What is Toxicokinetics

What is Toxicokinetics?

Toxicokinetics Defined

Toxicokinetics is essentially the study of "how a substance gets into the body and what happens to it in the body." Before this term was used, the study of the kinetics (movement) of chemicals was originally conducted with pharmaceuticals and the term pharmacokinetics became commonly used. Similarly, toxicology studies were initially conducted with drugs. Toxicokinetics deals with what the body does with a drug when given a relatively high dose relative to the therapeutic dose. Read more about **differences between pharmacokinetics and toxicokinetics**.

Processes

Four processes are involved in toxicokinetics: 1. **Absorption** — the substance enters the body. 2. **Distribution** — the substance moves from the site of entry to other areas of the body. 3. **Biotransformation** — the body changes (transforms) the substance into new chemicals (metabolites). 4. **Excretion** — the substance or its metabolites leave the body. The science of toxicology has evolved to include environmental and occupational chemicals as well as drugs. Toxicokinetics is thus the appropriate term for the study of the kinetics of all substances at toxic dose/ exposure levels. Frequently the terms toxicokinetics, pharmacokinetics, or disposition have the same meaning. Disposition is often used in place of toxicokinetics to describe the movement of chemicals through the body over the course of time, that is, how the body disposes of a xenobiotic.	Figure 8.4.6.9.1.1. Processes of toxicokinetics (Image Source: Adapted from iStock Photos, ©)

Factors Determining the Severity of Toxicity

The disposition of a toxicant and its biological reactivity are the factors that determine the severity of toxicity that results when a xenobiotic enters the body. The most important aspects of disposition include:

- **Duration and concentration** of a substance at the portal of entry.
- **Rate and amount** of the substance that can be absorbed.
- **Distribution** in the body and **concentration** of the substance at specific body sites.
- **Efficiency** of biotransformation and nature of the metabolites.
- **Ability** of the substance or its metabolites **to pass through cell membranes** and come into contact with specific cell components (for example, DNA).
- **Amount and duration of storage** of the substance (or its metabolites) in body tissues.
- **Rate and sites of excretion** of the substance.
- **Age and health status** of the person exposed.

Here are some examples of how toxicokinetics of a substance can influence its toxicity:

- **Absorption** — A highly toxic substance that is poorly absorbed may be no more hazardous than a substance of low toxicity that is highly absorbed.
- **Biotransformation** — Two substances with equal toxicity and absorption may differ in how hazardous they are depending on the nature of their biotransformation. A substance that is biotransformed into a more toxic metabolite (bioactivated) is a

greater hazard than a substance that is biotransformed into a less toxic metabolite (detoxified).

Inter-Related Processes of Absorption, Distribution, Biotransformation, and Elimination

Absorption, distribution, biotransformation, and elimination are inter-related processes as illustrated in Figure 2 below. After the substance is absorbed, it is distributed through the blood, lymph circulation, and extracellular fluids into organs or other storage sites and may be metabolized. Then, the substance or its metabolites are eliminated through the body's waste products.

Figure ({8{.}4{.}6{.}9{.}1{.}\eatSpaces{2} \endSpaces}\). Absorption, Distribution, Metabolism, and Elimination
(Image Source: NLM)

What are Transporters?

Transporters, also called transporter proteins, play an important role in the processes of absorption, distribution, metabolism, and elimination (ADME). They are important to pharmacological, toxicological, clinical, and physiological applications. For example:

- In the **liver** — transmembrane transporters, together with drug metabolizing enzymes, are important in drug metabolism and drug clearance by the liver. Xenobiotics, endogenous metabolites, bile salts, and cytokines affect the levels (or "expression") of these transporters in the liver. Adverse reactions in the liver to a xenobiotic such as a drug could be caused by genetic or disease-induced variations of transporter expression or drug-drug interactions at the level of these transporters.
- In the **kidneys** — renal proximal tubules are targets for toxicity partly because of the expression of transporters that mediate the secretion and reabsorption of xenobiotics. Changes in transporter expression and/or function could enhance the accumulation of toxicants and make the kidneys more susceptible to injury, for example, when xenobiotic uptake by carrier proteins is increased or the efflux of toxicants and their metabolites is reduced. The list of nephrotoxic chemicals is a long one and includes:
  - Environmental contaminants such as some hydrocarbon solvents, some heavy metals, and the fungal toxin ochratoxin.
  - Some antibiotics.
  - Some antiviral drugs.
  - Some chemotherapeutic drugs.

The competition of xenobiotics for transporter-related excretion and genetic polymorphisms affecting transporter function affect the likelihood of nephrotoxicity.
Because of concerns that such changes to transporter expression and function can adversely affect clinical outcomes and physiological regulation, increased drug transporter activity is important to study and understand. There is clinical and laboratory research including *in vitro*, *ex vivo*, and *in vivo* studies that shows how powerful drug-drug interactions can be.
For example, drugs might compete with each other for binding to a transporter, which can lead to changes in serum and tissue drug levels and possible side effects.

- This is one possible explanation for the rare occurrence of potentially severe toxicity when the drug methotrexate and nonsteroidal anti-inflammatory drugs are given at the same time.
- The drug probenecid, which competitively inhibits some transporters, has been used to increase the half-life of antibiotics such as penicillin and antiviral drugs and improve their therapeutic value.

Pharmacokinetics and Toxicokinetics: Now and in the Future

Current research priorities suggest that we can anticipate important strides in the following areas of pharmacokinetics and

toxicokinetics:

- An increased understanding of human variability of pharmacokinetics and pharmacodynamics in the population.
- Further exploration of mode of action hypotheses (**MoA**).
  - Is a MoA the same as a MOA? No. A **mode of action (MoA)** describes a functional or anatomical change, at the cellular level, resulting from exposure to a substance. A **mechanism of action (MOA)** describes changes at the molecular level.
- Further application of biological modeling in the risk assessment of individual chemicals and chemical mixtures.
- Further identification and discussion of uncertainties in the modeling process.
- Further use of "Reverse Toxicokinetics," also called **"IVIVE" (*In vitro* to *in vivo* extrapolation)**. IVIVE *in vitro* data to estimate exposures that could be associated with adverse effects *in vivo*.

This page titled 8.4.6.9.1: What is Toxicokinetics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 9.1: What is Toxicokinetics by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.10: ABSORPTION

Learning Objectives

After completing this lesson, you will be able to:

- Explain absorption and its role in toxicokinetics.
- Describe the primary routes of exposure.
- Explain the role of cell membranes in absorption.
- Identify ways in which xenobiotics pass across cell membranes.

In this section...

Topics include:

8.4.6.10.1: Introduction to Absorption

8.4.6.10.2: Gastrointestinal Tract

8.4.6.10.3: Respiratory Tract

8.4.6.10.4: Dermal Route

8.4.6.10.5: Other Routes of Exposure

Section 10: Key Points

What We've Covered

This section made the following main points:

- Absorption is the process by which toxicants gain entrance into the body.
- Ingested and inhaled materials are considered outside the body until they cross the cellular barriers of the gastrointestinal tract or respiratory system.
- The likelihood of absorption depends on the:
  - Route of exposure.
  - Concentration of the substance at the site of contact.
  - Chemical and physical properties of the substance.
- Exposure routes include:
  - Primary routes:
    - Gastrointestinal (GI) tract
      - Mouth and esophagus — poorly absorbed under normal conditions due to short exposure time (nicotine and nitroglycerin are notable exceptions).
      - Stomach — significant site for absorption of weak organic acids, but weak bases are poorly absorbed.
      - Intestine — greatest absorption of both weak bases and weak acids, particularly in the small intestine.
      - Colon and rectum — very little absorption, unless administered via suppository.
    - Respiratory tract
      - Mucociliary escalator — movements of the cilia push mucus and anything contained within up and out into the

throat to be swallowed or removed through the mouth.
        - Pulmonary region — most important site for absorption with about 50 times the surface area of the skin and very thin membranes.
    - Skin
        - Epidermis and stratum corneum — the only layer important in regulating the penetration of a skin contaminant.
        - Toxicants move across the stratum corneum by passive diffusion.
        - If a toxicant penetrates through the stratum corneum, it enters lower layers of the epidermis, dermis, and subcutaneous tissue, which are far less resistant to further diffusion.
  - Other exposure routes:
    - Injections
    - Implants
    - Conjunctival instillations (eye drops)
    - Suppositories
- Cell membranes surround all body cells and are made up of a phospholipid bilayer in which each molecule contains a:
  - Polar (hydrophilic, or attracted to water) phosphate head
  - Lipophilic (attracted to lipid-soluble substances) lipid tail
- Xenobiotics must pass across cell membranes to enter, move within, and leave the body. This movement can be either:
  - Passive transfer (most common) — simple diffusion or osmotic filtration with no cellular energy or assistance required.
  - Facilitated transport — similar to passive transport, but a carrier-mediated transport mechanism and thus faster and capable of moving larger molecules.
  - Active transport — movement against the concentration gradient (from lower to higher concentrations), requiring cellular energy from ATP.
  - Endocytosis — the cell surrounds the substance with a section of its cell wall, separating from the membrane and moving into the interior of the cell.

This page titled 8.4.6.10: Absorption is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 10: Absorption by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.10.1: Introduction to Absorption

Introduction to Absorption

Toxicants gain entrance into the body by absorption. The body considers ingested and inhaled materials as being outside of it until those materials cross the cellular barriers of the gastrointestinal tract or respiratory system. A substance must be absorbed to exert an effect on internal organs, although local toxicity, such as irritation, may occur.

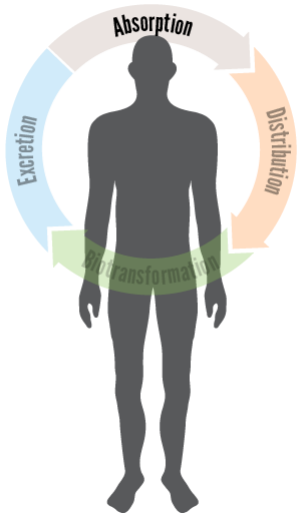


Figure 8.4.6.10.1.1. Processes of toxicokinetics
(Image Source: Adapted from iStock Photos, ©)

Absorption Variability

Absorption varies greatly by specific chemicals and the route of exposure.

- For skin, oral or respiratory exposure, the absorbed dose is only a fraction of the exposure dose (external dose).
- For substances injected or implanted directly into the body, exposure dose is the *same* as the absorbed or internal dose.

Several factors affect the likelihood that a xenobiotic will be absorbed. The most important factors are the:

- Route of exposure.
- Concentration of the substance at the site of contact.
- Chemical and physical properties of the substance.

The route of exposure influences how the concentration and properties of the substance vary. In some cases, a high percentage of a substance may not be absorbed from one route whereas a low amount may be absorbed via another route.

- For example, very little **DDT powder** will penetrate the skin whereas a high percentage will be absorbed when it is swallowed.

Due to such route-specific differences in absorption, xenobiotics are often ranked for hazard in accordance with the route of exposure. A substance may be categorized as relatively non-toxic by one route and highly toxic via another route.

Routes of Exposure

The **primary routes of exposure** by which xenobiotics can gain entry into the body are:

- **Gastrointestinal (GI) tract** — important for environmental exposure to contaminants from food and water; the main route for many pharmaceuticals.
- **Respiratory tract** — important for environmental and occupational exposure to air contaminants; some pharmaceuticals (such as nasal or oral aerosol inhalers) use this route.
- **Skin** — important environmental and occupational exposure route; many consumer and pharmaceutical products are applied directly to the skin.

Other routes of exposure – used primarily for specific medical purposes:

- **Injections** — primarily used for pharmaceuticals.
- **Implants** — pharmaceuticals may be implanted to permit slow, time-release (for example, hormones). Many medical devices are implanted for which minimal absorption is desired (such as artificial lens or tendons). Some materials enter the body via skin penetration as the result of accidents or weapons.
- **Conjunctival instillations (eye drops)** — primarily for treating ocular conditions; however, in some cases, considerable absorption can occur and cause systemic toxicity.
- **Suppositories** — used for medicines that may not be adequately absorbed after oral administration or that are intended for local therapy; usual locations for suppositories are the rectum and vagina.

Cell Membranes

Cell membranes (often referred to as plasma membranes) surround all body cells and are similar in structure. They consist of two layers of phospholipid molecules arranged like a sandwich, referred to as a "phospholipid bilayer." Each phospholipid molecule consists of a phosphate head and a lipid tail. The phosphate head is polar, meaning it is hydrophilic *(attracted to water)*. In contrast, the lipid tail is lipophilic *(attracted to lipid-soluble substances)*.
The two phospholipid layers are oriented on opposing sides of the membrane so that they are approximate mirror images of each other. The polar heads face outward and the lipid tails face inward in the membrane sandwich (Figure 8.4.6.10.1.2).

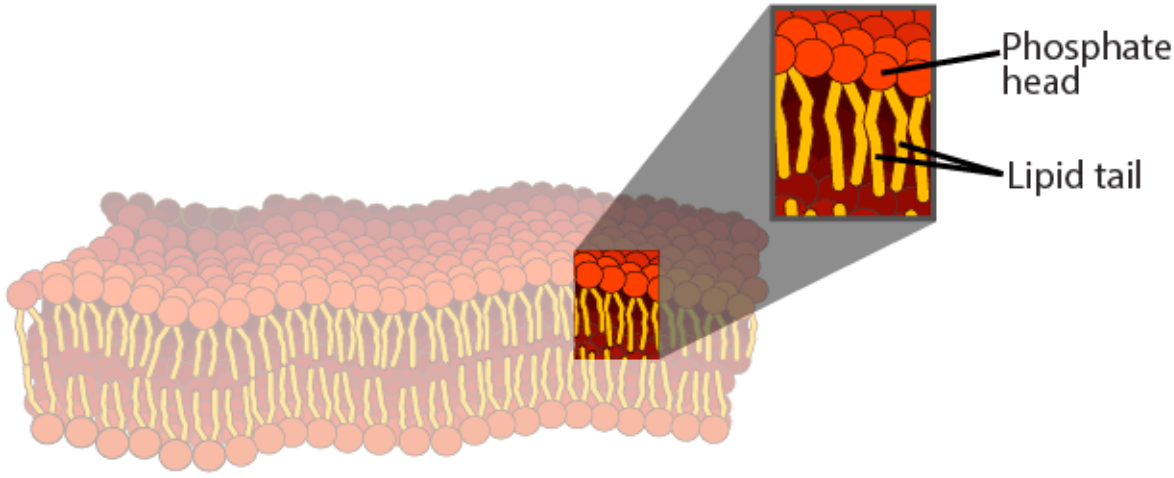


Figure 8.4.6.10.1.2. Each phospholipid molecule consists of a phosphate head and lipid tail
(Image Source: Adapted from Wikimedia Commons, obtained under Public Domain, original image)

The cell membrane is tightly packed with these phospholipid molecules interspersed with various proteins and cholesterol molecules. Some proteins span across the entire membrane which can create openings for aqueous channels or pores.

A typical cell membrane structure is illustrated in Figure 8.4.6.10.1.3.

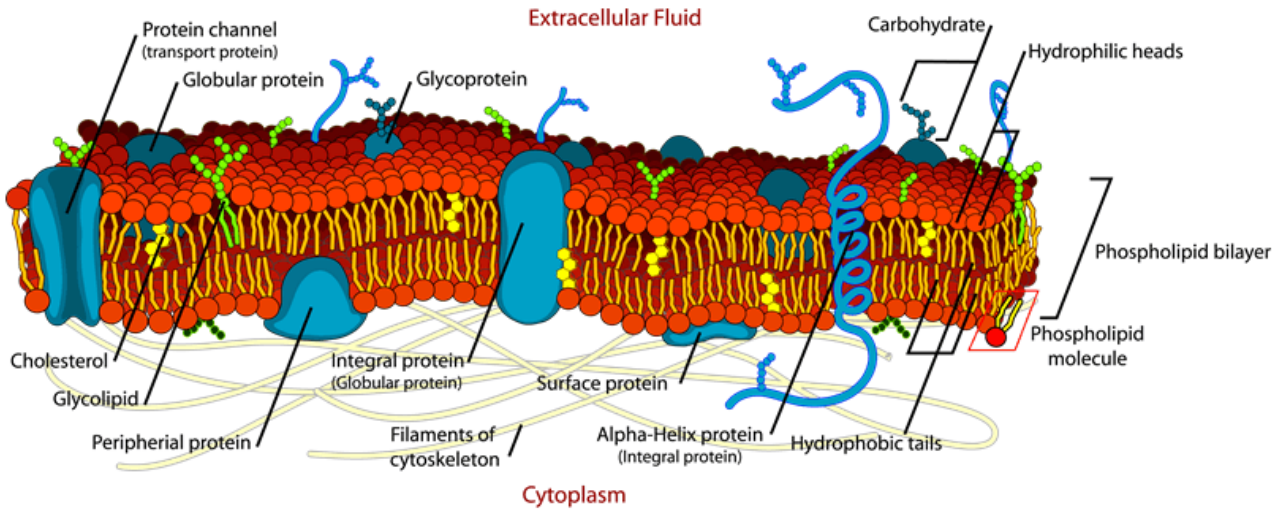


Figure 8.4.6.10.1.3. Typical cell membrane structure (Image Source: Adapted from Wikimedia Commons, obtained under Public Domain, original image)

Role of Cell Membranes in Absorption

For a xenobiotic to enter the body (as well as move within, and leave the body) it must pass across cell membranes (cell walls). Cell membranes are formidable barriers and a major body defense that prevents foreign invaders or substances from gaining entry into body tissues. Normally, cells in solid tissues (such as skin, or mucous membranes of the lung or intestine) are so tightly compacted that substances cannot pass between them. This requires that the xenobiotic have the ability to penetrate cell membranes. It must cross several membranes to go from one area of the body to another.

For a substance to move through one cell requires that it first move across the cell membrane into the cell, pass across the cell, and then cross the cell membrane again to leave the cell. This is true whether the cells are in the skin, the lining of a blood vessel, or an internal organ such as the liver. In many cases, in order for a substance to reach the site where it exerts toxic effects, it must pass through several membrane barriers.

Animation 1 depicts how a chemical from a theoretical consumer product called a "Shower Gel" might get to the surface of the skin during showering and then pass through several membranes before coming in contact with the inside of a liver cell

http://www.toxmsdt.com/uploads/b/886...a_cell_446.mp4

Animation 8.4.6.10.1.1. From a Gel to a Cell: Following the journey of a chemical from a theoretical shower gel product through several membranes and ultimately into a cell
(Image Source: iStock Photos, ©)

Movement of Toxicants Across Cell Membranes

Some toxicants move across a membrane barrier with relative ease while others find it difficult or impossible. Those that can cross the membrane use one of two general methods: 1) **passive transfer** or 2) **facilitated transport**.
Passive transfer consists of **simple diffusion** (or osmotic filtration) and is "passive" because no cellular energy or assistance is required.

Some toxicants cannot simply diffuse across the membrane but require assistance by specialized transport mechanisms. The primary types of specialized transport mechanisms are:

- Facilitated diffusion
- Active transport
- Endocytosis (phagocytosis and pinocytosis)

Passive Transfer

Passive transfer is the most common way that xenobiotics cross cell membranes. Two factors determine the rate of passive transfer:

1. The difference in concentrations of the substance on opposite sides of the membrane (this occurs when a substance moves from a region of high concentration to one having a lower concentration. Diffusion will continue until the concentration is equal on both sides of the membrane).
2. The ability of the substance to move either through the small pores in the membrane or the lipophilic interior of the membrane.

Properties affecting a chemical substance's ability for passive transfer are:

- Lipid solubility
- Molecular size
- The degree of ionization

Substances with high lipid solubility readily diffuse through the phospholipid membrane. Small water-soluble molecules can pass across a membrane through the aqueous pores, along with normal intracellular water flow.
Large water-soluble molecules usually cannot make it through the small pores, although some may diffuse through the lipid portion of the membrane, but at a slow rate.

- Most aqueous pores are about 4 Angstrom (Å) in size and allow chemicals of molecular weight 100-200 to pass through. Exceptions are membranes of capillaries and kidney glomeruli which have relatively large pores (about 40 Angstrom [Å]) that allow molecules up to a molecular weight of about 50,000 (molecules slightly smaller than albumin which has a molecular weight of 60,000) to pass through.

In general, highly ionized chemicals have low lipid solubility and pass with difficulty through the lipid membrane.
Figure 4 demonstrates the passive diffusion and filtration of xenobiotics through a typical cell membrane.

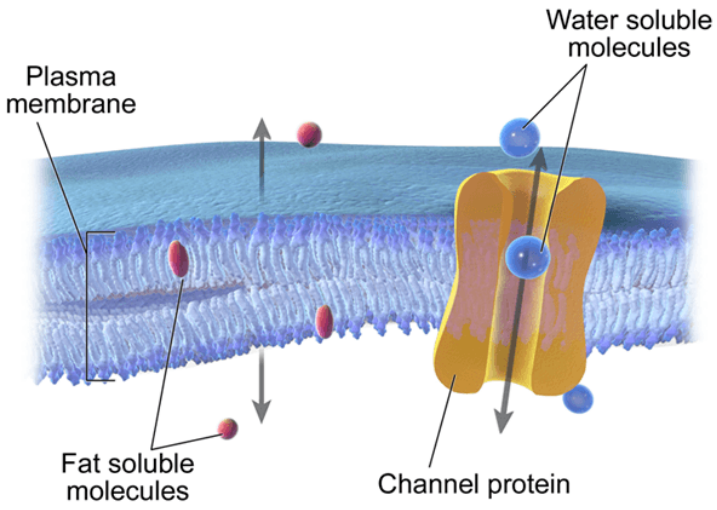


Figure 8.4.6.10.1.4. Cellular diffusion
(Image Source: Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. Obtained under Creative Commons license.)

Facilitated Diffusion

Facilitated diffusion is similar to simple diffusion in that it does not require energy and follows a concentration gradient. The difference is that it is a carrier-mediated transport mechanism (Figure 8.4.6.10.1.5)—that is, special transport proteins, which are embedded within the cell membrane, facilitate movement of molecules across the membrane. The results are similar to passive transport but faster and capable of moving larger molecules that have difficulty diffusing through the membrane without a carrier.

- Examples are the transport of sugar and amino acids into RBCs and the CNS.

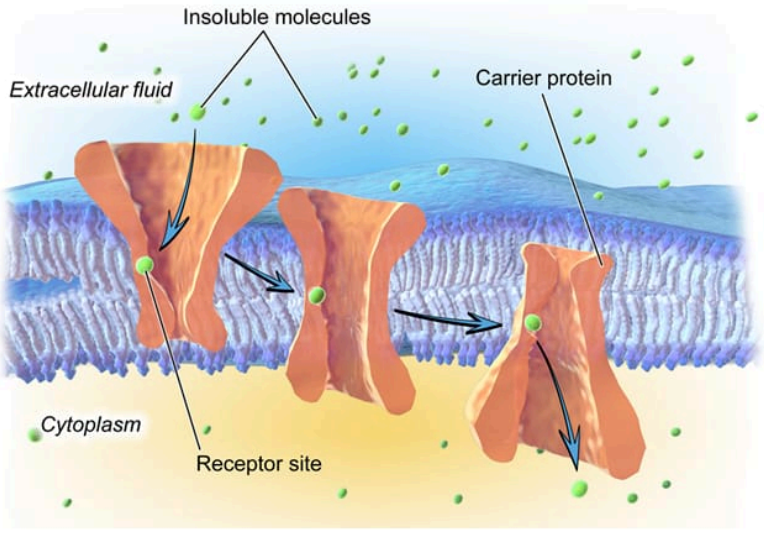


Figure 8.4.6.10.1.5. Facilitated Diffusion
(Image Source: Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. Obtained under Creative Commons license.)

Active Transport

Some substances are unable to move with diffusion, unable to dissolve in the lipid layer, and are too large to pass through the aqueous channels. For some of these substances, **active transport** processes exist in which movement through the membrane may be *against* the concentration gradient, that is, from low to higher concentrations. Cellular energy from adenosine triphosphate (ATP) is required in order to accomplish this. The transported substance can move from one side of the membrane to the other side by this energy process. Active transport is important in the transport of xenobiotics into the liver, kidney, and central nervous system and for maintenance of electrolyte and nutrient balance.

Figure 8.4.6.10.1.6 shows sodium and potassium moving against concentration gradient with the help of the ATP sodium-potassium exchange pump.

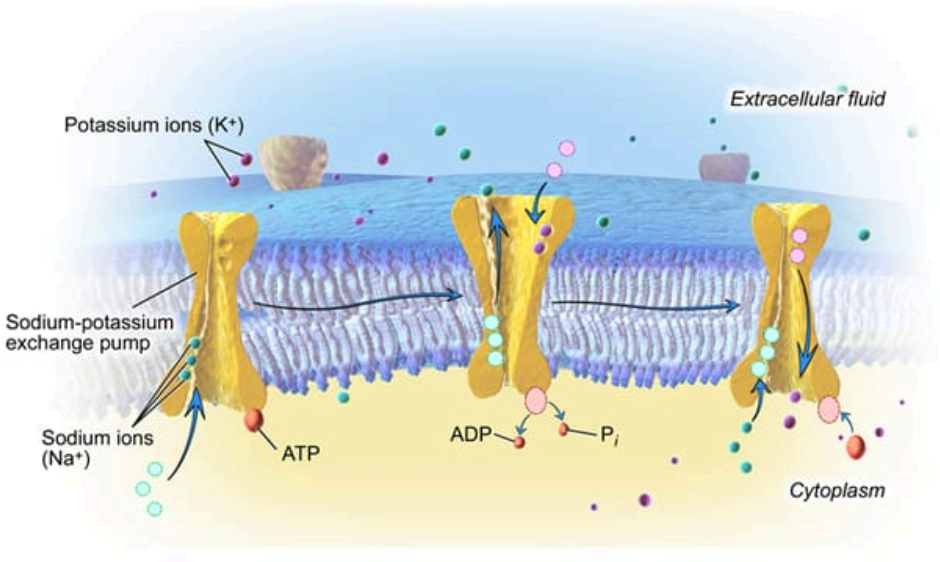


Figure 8.4.6.10.1.6. Sodium-Potassium Exchange Pump
(Image Source: Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. Obtained under Creative Commons license.)

Endocytosis (Phagocytosis and Pinocytosis)

Many large molecules and particles cannot enter cells via passive or active mechanisms. However, some may still enter by a process known as **endocytosis**.
In endocytosis, the cell surrounds the substance with a section of its cell wall. This engulfed substance and section of membrane then separates from the membrane and moves into the interior of the cell. The two main forms of endocytosis are 1) phagocytosis and 2) pinocytosis.
In **phagocytosis** (cell eating), large particles suspended in the extracellular fluid are engulfed and either transported into cells or are destroyed within the cell. This is a very important process for lung phagocytes and certain liver and spleen cells.
Pinocytosis (cell drinking) is a similar process but involves the engulfing of liquids or very small particles that are in suspension in the extracellular fluid.

Figure 8.4.6.10.1.7 demonstrates the types of membrane transport by endocytosis.

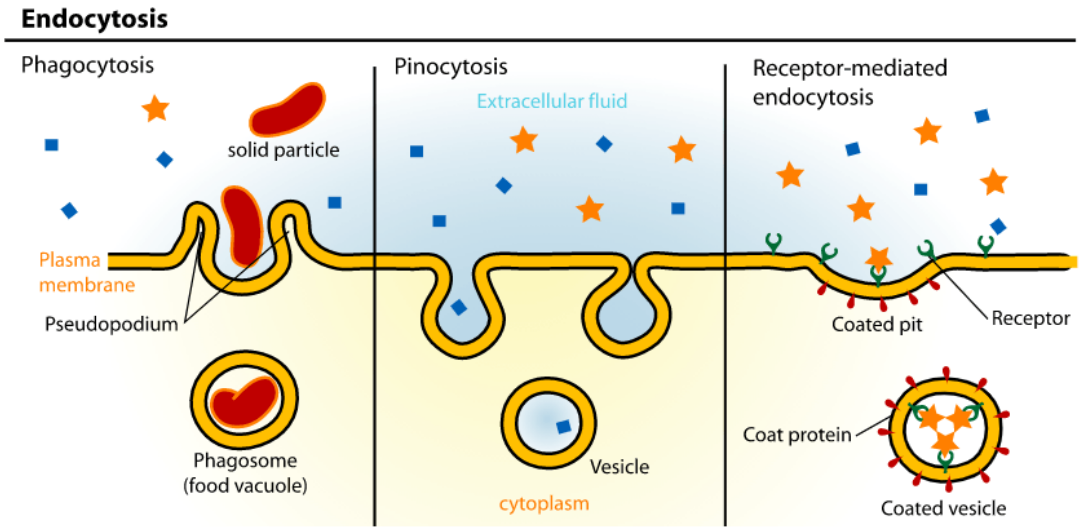


Figure 8.4.6.10.1.7. Types of Endocytosis
(Image Source: Wikimedia Commons, obtained by Public Domain License, original image)

Knowledge Check

1) The process whereby a substance moves from outside the body into the body is known as:
a) Distribution
b) Biotransformation
c) Absorption

Answer available at the end of this section.

2) For a xenobiotic to move from outside the body to a site of toxic action requires that it:
a) Possess hydrophilic (water-soluble) properties
b) Possess hydrophobic (lipophilic) properties
c) Pass through several cell membranes

Answer available at the end of this section.

3) The basic structure of the cell membrane consists of:
a) A thick protein layer containing phospholipid channels
b) A bilayer of phospholipids with scattered proteins within the layers
c) Cholesterol outer layer with a phospholipid inner layer

Answer available at the end of this section.

4) The membrane transport process by which large hydrophobic molecules cross membranes via the lipid portion of the membrane, follow the concentration gradient, and do not require energy or carrier molecules is known as:

a) Simple diffusion
b) Active transport
c) Facilitated diffusion

Answer available at the end of this section.

5) Endocytosis is a form of specialized membrane transport in which the cell surrounds the substance with a section of its cell membrane. The specific endocytosis process by which liquids or very small particles are engulfed and transported across the membrane is known as:
a) Phagocytosis
b) Pinocytosis
c) Exocytosis

Answer available at the end of this section.

Answers

Knowledge Check

Absorption - **This is the correct answer.**
Absorption is the first and crucial step in the toxicokinetics of a xenobiotic. Without absorption, a toxic substance does not represent a human health hazard.

Knowledge Check

Pass through several cell membranes - **This is the correct answer.**
In order for a xenobiotic to move from outside the body to an internal site of toxic action (target cells), a xenobiotic must pass through several membrane barriers. The first membranes are those at the portal of entry, for example, lung or intestinal tract.

Knowledge Check

A bilayer of phospholipids with scattered proteins within the layers - **This is the correct answer.**
The typical cell membrane consists of two layers of phospholipids with polar head groups consisting of phospholipid molecules and the lipid inner portion consisting primarily of cholesterol molecules. The phospholipid layers are oriented on opposing sides of the membrane so that they are approximate mirror images of each other. Various proteins are scattered throughout the lipid bilayers of the membrane.

Knowledge Check

Simple diffusion - **This is the correct answer.**
Large hydrophobic molecules must diffuse through the lipid portion of the membrane, with the rate of transport correlating with its lipid solubility. In general, highly ionized chemicals have low lipid solubility and do not readily pass through the lipid membrane.

Knowledge Check

Pinocytosis - **This is the correct answer.**
Pinocytosis (cell drinking) involves the engulfing of liquids or very small particles that are in suspension within the extracellular fluid.

This page titled 8.4.6.10.1: Introduction to Absorption is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 10.1: Introduction to Absorption by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.10.2: Gastrointestinal Tract

Gastrointestinal Tract

The gastrointestinal (GI) tract can be viewed as a tube going through the body (Figure 8.4.6.10.2.1). Its contents are considered exterior to the body until absorbed. Salivary glands, liver, and the pancreas are considered accessory glands of the GI tract as they have ducts entering the GI tract and secrete enzymes and other substances. For foreign substances to enter the body, they must pass through the gastrointestinal mucosa, crossing several membranes before entering the bloodstream.

Substances must be absorbed from the gastrointestinal tract in order to exert a toxic effect throughout the whole body, although local gastrointestinal damage may occur from direct exposures to toxicants. Absorption can occur at any place along the entire gastrointestinal tract. However, the degree of absorption depends on the site.

Three main factors affect absorption within the various sites of the gastrointestinal tract:

1. Type of cells at the specific site.
2. Period of time that the substance remains at the site.
3. pH of stomach or intestinal contents at the site.

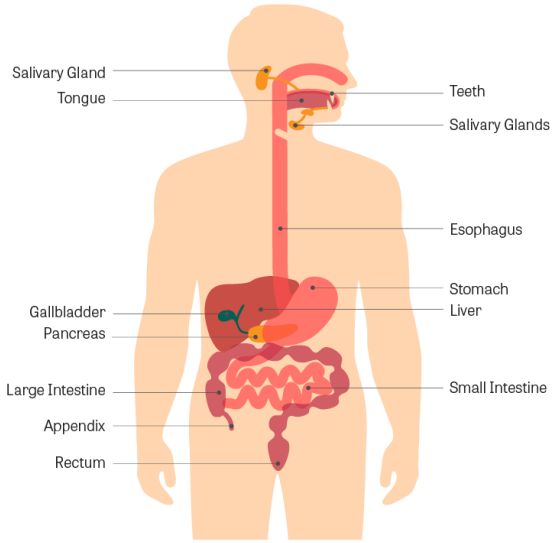


Figure 8.4.6.10.2.1. Anatomy of gastrointestinal tract
(Image Source: adapted from iStock Photos, ©)

Mouth and Esophagus

Under normal conditions, xenobiotics are poorly absorbed within the **mouth** and **esophagus**, due mainly to the very short time that a substance resides within these portions of the gastrointestinal tract. There are some notable exceptions. For example:

- Nicotine readily penetrates the mouth mucosa.
- Nitroglycerin is placed under the tongue (sublingual) for immediate absorption and treatment of heart conditions.

The sublingual mucosa under the tongue and in some areas of the mouth is thin and highly vascularized and allows some substances to be rapidly absorbed.

Stomach

The **stomach**, with its high acidity (pH 1-3), is a significant site for the absorption of weak organic acids, which exist in a diffusible, nonionized and lipid-soluble form. In contrast, weak bases will be highly ionized and therefore poorly absorbed. The acidic stomach may chemically break down some substances. For this reason, those substances must be administered in gelatin capsules or coated tablets, which can pass through the stomach into the intestine before they dissolve and release their contents.

Another determinant that affects the amount of a substance that will be absorbed in the stomach is the presence of food in the stomach. Food ingested at the same time as the xenobiotic may result in a considerable difference in absorption of the xenobiotic.

Intestine

The greatest absorption of chemicals, as with nutrients, takes place in the **intestine**, particularly in the small intestine. The intestine has a large surface area consisting of outward projections of the thin (one-cell thick) mucosa into the lumen of the intestine (the villi) (Figure 8.4.6.10.2.2). This large surface area facilitates diffusion of substances across the cell membranes of the intestinal mucosa.

Since the pH is near neutral (pH 5-8), both weak bases and weak acids are nonionized and are usually readily absorbed by passive diffusion. Lipid soluble, small molecules effectively enter the body from the intestine by passive diffusion.

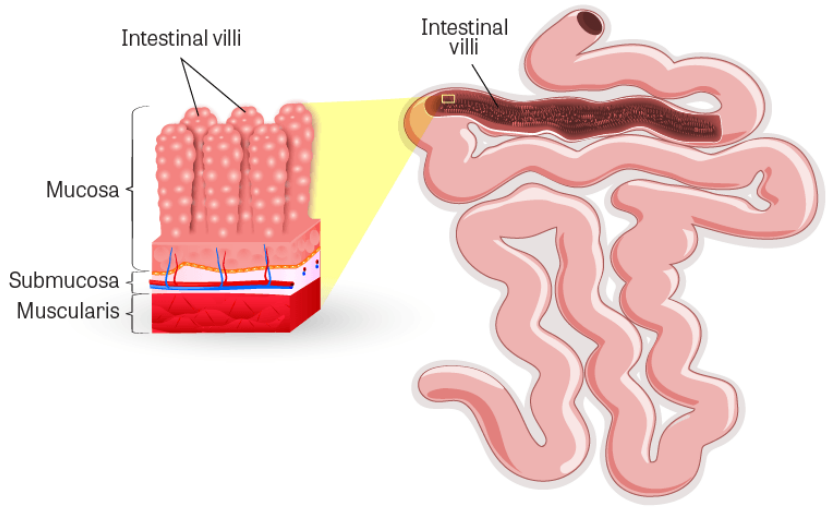


Figure 8.4.6.10.2.2. Anatomy of structures in small intestine used for absorption
(Image Source: adapted from iStock Photos, ©)

In addition to passive diffusion, facilitated and active transport mechanisms move certain substances across the intestinal cells into the body, including essential nutrients such as glucose, amino acids, and calcium. These mechanisms also transport strong acids, strong bases, large molecules, and metals, including some important toxins.

- For example, lead, thallium, and paraquat (herbicide) are toxins that active transport systems move across the intestinal wall.

The slow movement of ingested substances through the intestinal tract can influence their absorption. This slow passage increases the length of time that a compound is available for absorption at the intestinal membrane barrier.

Intestinal microflora and gastrointestinal enzymes can affect the toxicity of ingested substances. Some ingested substances may be only poorly absorbed but they may be biotransformed within the gastrointestinal tract. In some cases, their biotransformed

products may be absorbed and be more toxic than the ingested substance.

- An important example is the formation of carcinogenic nitrosamines from non-carcinogenic amines by intestinal flora.

Colon and Rectum

Very little absorption takes place in the **colon** and **rectum**. As a rule, if a xenobiotic has not been absorbed after passing through the stomach or small intestine, very little further absorption will occur. However, there are some exceptions, as some medicines may be administered as rectal suppositories with significant absorption.

- An example is Anusol (hydrocortisone preparation) used for the treatment of local inflammation which is partially absorbed (about 25%).

Knowledge Check

1) The most important factor that determines whether a substance will be absorbed within the stomach is the:
a) Physical form as a solid or liquid
b) Molecular size
c) pH

Answer available at the end of this section.

2) The primary routes for absorption of environmental agents are:
a) Gastrointestinal tract, respiratory tract, and skin
b) Conjunctival exposures and skin wounds

Answer available at the end of this section.

3) The site of the gastrointestinal tract where most absorption takes place is:
a) Stomach
b) Small intestine
c) Colon and rectum

Answer available at the end of this section.

Answers

Knowledge Check

pH - **This is the correct answer.**
The most important factor that determines absorption within the stomach is pH. Weak organic acids, which exist in a diffusible, nonionized and lipid-soluble form are readily absorbed in the high acidity of the stomach (pH 1-3). In contrast, weak bases will be highly ionized and therefore poorly absorbed.

Knowledge Check

Gastrointestinal tract, respiratory tract, and skin - **This is the correct answer.**
Environmental agents may be found in contaminated food, water, or air. As such, they may be ingested, inhaled, or present on the skin.

Knowledge Check

Small intestine - **This is the correct answer.**
By far, the greatest absorption takes place in the intestine. This is due to the neutral pH and the large, thin, surface area that allows easy penetrable by passive diffusion. Weak bases, weak acids, lipid soluble substances and small molecules effectively enter the body from the intestine. In addition, special carrier-mediated and active transport systems exist.

This page titled 8.4.6.10.2: Gastrointestinal Tract is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 10.2: Gastrointestinal Tract by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.10.3: Respiratory Tract

Respiratory Tract

Many environmental and occupational agents as well as some pharmaceuticals enter the respiratory tract through inhalation. Absorption can occur at any place within the upper respiratory tract. However, the amount of a particular xenobiotic that can be absorbed at a specific location depends highly on its physical form and solubility.
There are three basic regions to the respiratory tract:

1. Nasopharyngeal region
2. Tracheobronchial region
3. Pulmonary region

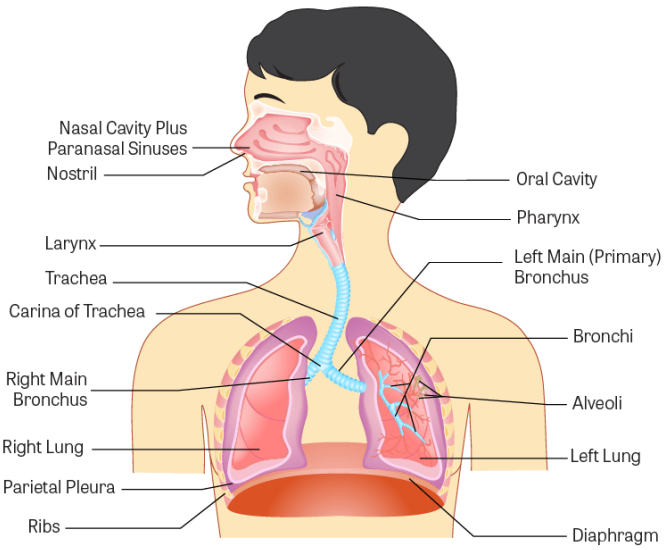
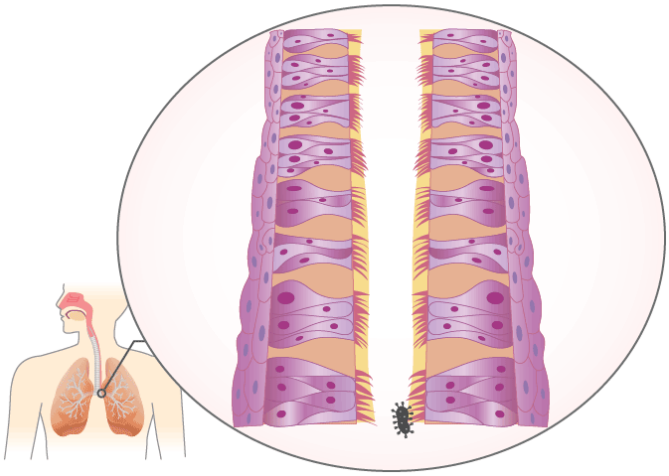


Figure 8.4.6.10.3.1. Anatomy of the respiratory tract
(Image Source: adapted from iStock Photos, ©)

Mucociliary Escalator

The mucociliary escalator covers most of the bronchi, bronchioles, and nose. It contains mucus-producing goblet cells and ciliated epithelium. The movements of the cilia push it and anything in it such as inhaled particles or microorganisms up and out into the throat, which can either get swallowed or removed through the mouth.



Animation 8.4.6.10.3.1. The mucociliary escalator provides a barrier against infection

Pulmonary Region

By far, the most important site for absorption is the pulmonary region consisting of the very small airways called bronchioles and the alveolar sacs of the lung.
The alveolar region has a very large surface area, about 50 times that of the skin. In addition, the alveoli consist of only a single layer of cells with very thin membranes that separate the inhaled air from the blood stream. Oxygen, carbon dioxide, and other gases readily pass through this membrane. Gases and particles, which are water-soluble (and thus blood-soluble), are absorbed more efficiently from the lung alveoli compared to their absorption via the gastrointestinal tract or through the skin. Water-soluble gases and liquid aerosols can pass through the alveolar cell membrane by simple passive diffusion.

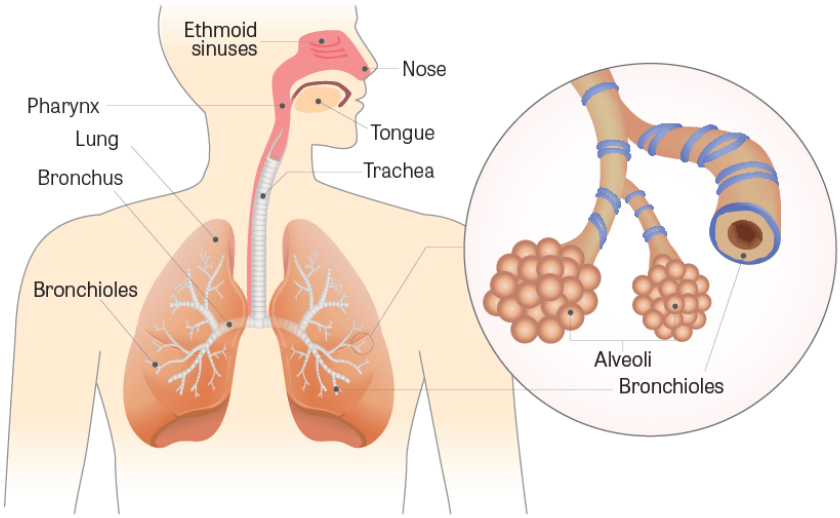


Figure 8.4.6.10.3.2. Detailed view of alveoli and bronchioles
(Image Source: adapted from iStock Photos, ©)

Impact of Physical Form on Absorption

In addition to solubility, the ability to be absorbed depends highly on the physical form of the agent (that is, whether the agent is a gas/vapor or a particle). The physical form determines the extent of its penetration into the deep lung.

Gases and Vapors

A **gas** or **vapor** can be inhaled deep into the lung and if it has high solubility in the blood, it is almost completely absorbed in one respiration (a single breath). Absorption through the alveolar membrane is by passive diffusion, following the concentration gradient. As the agent dissolves in the circulating blood, it leaves the lung and a large amount of gas or vapor can be absorbed and enter the body.
Blood-soluble gases or vapors can often be exhaled instead of absorbed right away. For blood-soluble gases, the equilibrium between the concentration of the agent in the inhaled air and that in the blood is difficult to achieve. Inhaled gases or vapors, which have poor solubility in the blood, have a limited capacity for absorption. The main reason for this is that the blood can become quickly saturated. Once saturated, blood will not be able to accept the gas and it will remain in the inhaled air and then get exhaled. One way that the amount of gas absorbed could increase is if the rate and depth of breathing were increased (this concept is known as ventilation limitation). More specifically, the amount of gas absorbed could be increased if the rate of blood supply to the lung were increased by **flow limitation**.

In contrast, *insoluble* gases or vapors can be absorbed into the body by the lungs before getting exhaled (for example, nitrogen dioxide and carbon monoxide). This is because the equilibrium between the inhaled air and the blood is reached more quickly for relatively insoluble gases than for soluble gases.

Airborne Particles

The absorption of airborne particles is usually quite different from that of gases or vapors. The absorption of solid particles, regardless of solubility, depends upon particle size:

- Large particles (>5 μM) are generally deposited in the nasopharyngeal (head airways region) region with little absorption.
- Particles 2-5 μM can penetrate into the tracheobronchial region.
- Very small particles (<1 μM) are able to penetrate deep into the alveolar sacs where they can deposit and be absorbed.

Differences in Absorption Among Regions of the Respiratory Tract

Nasopharyngeal Region

Minimal absorption takes place in the nasopharyngeal region due to the cell thickness of the mucosa and the rapid movement of gases and particles through the region.

Tracheobronchial Region

Relatively soluble gases can quickly enter the blood stream. Most deposited particles are moved back up to the mouth where they are swallowed.

Pulmonary Region

Absorption in the **alveoli** of the pulmonary region is quite efficient compared to other areas of the respiratory tract. Relatively soluble materials (gases or particles) are quickly absorbed into systemic circulation. Pulmonary macrophages exist on the surface of the alveoli. They are not fixed and not a part of the alveolar wall. They can engulf particles just as they engulf and kill

microorganisms. These alveolar macrophages can scavenge and clear some insoluble particle into the lymphatic system.

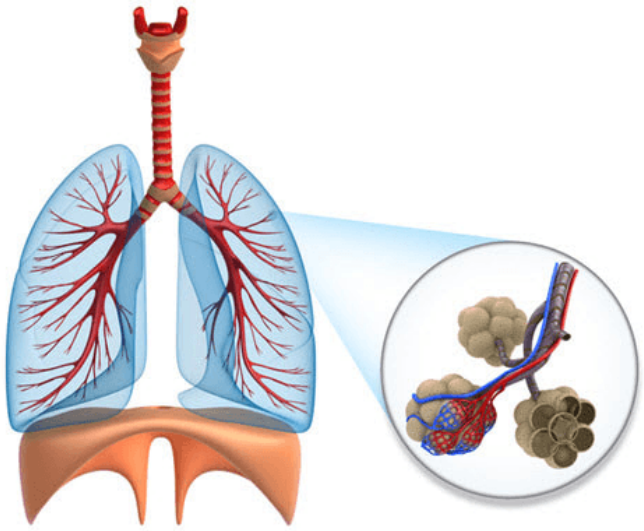


Figure 8.4.6.10.3.3. Alveoli
(Image Source: iStock Photos, ©

Some other particles may remain in the alveoli indefinitely. For example:

- Coal dust and asbestos fibers may lead to black lung or asbestosis, respectively.
- Carbon nanotubes (CNT), tiny tube-shaped structures smaller than a human hair, have been found in the lungs long after exposure. CNT are used in materials like polymers and anti-static packaging for their electric, magnetic, and mechanical properties. Studies of what happens to different forms of single-walled and multi-walled carbon nanotubes (CNT) found that pristine CNT could remain in the lung for months or even years after pulmonary deposition. However, some CNT can move to the gastrointestinal (GI) tract via the mucocilary escalator where, if swallowed, there appears to be no uptake of CNT from the GI tract (with a possible exception of the smallest functionalized single-walled CNT). In addition, under some experimental conditions in animals, some carbon nanotubes moved from the alveolar space to the nearby pulmonary region including lymph nodes, subpleura and pleura, and smaller amounts went to distal organs including the liver, spleen, and bone marrow.

Factors Affecting the Toxicity of Inhaled Materials

The nature of toxicity of inhaled materials depends on whether the material is absorbed or remains within the alveoli and small bronchioles. If the agent is absorbed and is lipid soluble, it can rapidly distribute throughout the body, passing through the cell membranes of various organs or into fat depots. Lipid-soluble substances take a longer time to reach equilibrium than water soluble substances. Chloroform and ether are examples of lipid-soluble substances with high blood solubility.

Non-absorbed foreign material can also cause severe toxic reactions within the respiratory system. These reactions may take the form of chronic bronchitis, alveolar breakdown (emphysema), fibrotic lung disease, and even lung cancer. In some cases, the toxic particles can kill the alveolar macrophages, which results in lowering the body's respiratory defense mechanism.

Pharmaceuticals Targeted to the Respiratory Tract

Inhaled drug delivery devices can be very effective and safe for getting active agents directly to their site of action. Inhalation

is used to deliver locally acting drugs to treat respiratory conditions, including asthma, chronic obstructive pulmonary disease (COPD), and airway infections. Advantages of targeted delivery to the lungs include a more rapid onset of action and an increased therapeutic effect. Depending on the drug inhaled, there can be reduced systemic side effects since a lower dose can deliver the required local concentration.

Toxicogenomics and Toxicity Testing

Toxicogenomics applies genomics concepts and technologies to study gene and protein activities within a type of tissue or type of cell in response to a chemical exposure. Toxicogenomics helps in the understanding of what genes and proteins interact with a chemical, and what diseases are associated with various genes, proteins, and chemicals. This example used toxicogenomics to evaluate the response of the respiratory tract to one type of inhaled material:
For example, titanium dioxide nanoparticles (TiO2NPs) induce lung inflammation in experimental animals and **one study** included a comprehensive toxicogenomic analysis of lung responses in mice exposed to six individual TiO2NPs exhibiting different sizes, crystalline structure, and surface modifications. The goal was to investigate whether the mechanisms leading to TiO2NP-induced lung inflammation are property specific. The results suggest that the severity of lung inflammation is property specific; however, the underlying mechanisms (genes and pathways perturbed) leading to inflammation were the same for all particle types. While the particle size clearly influenced the overall acute lung responses, a combination of small size, crystalline structure, and hydrophilic surface contributed to the long-term pathological effects observed at the highest dose.

Knowledge Check

1) An inhaled material will most likely be absorbed into the body if it has the following characteristics:

a) High lipid solubility and poorly ionized
b) Large particle size and low water solubility
c) High water solubility and small particle size

Answer available at the end of this section.

2) Particles of size 2-5 µM are most likely to settle out in which location of the respiratory tract?

a) Nasopharyngeal region
b) Tracheobronchial region
c) Pulmonary region

Answer available at the end of this section.

Answers

Knowledge Check

High water solubility and small particle size - **This is the correct answer.**
In contrast to absorption via the gastrointestinal tract or through the skin, gases and particles, which are water-soluble (and thus blood soluble), will be absorbed more efficiently from the lung alveoli. Very small particles (<1 µM) are able to penetrate deep into the alveolar sacs where they can deposit and be absorbed.

Knowledge Check

Tracheobronchial region - **This is the correct answer.**
Particles 2-5 µM can penetrate into the tracheobronchial region. Very small particles (<1 µM) are able to penetrate deep into the alveolar sacs where they can deposit and be absorbed.

This page titled 8.4.6.10.3: Respiratory Tract is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 10.3: Respiratory Tract by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.10.4: Dermal Route

Dermal Route

In contrast to the thin membranes of the respiratory alveoli and the gastrointestinal villi, the skin is a complex, multilayer tissue. It is relatively impermeable to most ions and aqueous solutions, and serves as a barrier to most xenobiotics.

Did you know?

Dimethyl sulfoxide (DMSO) has been used in research, human and veterinary medicine, and as a solvent. After applying to the skin, some people can quickly detect a garlic taste as the DMSO is absorbed and enters the body. DMSO also increases the rate of absorption of some other compounds through the skin.

For transdermal drug delivery (TDD), the big challenge is the barrier property of skin, especially the stratum corneum (SC). Different methods have been developed to enhance the penetration of drugs through the skin, with the most popular approach being the use of penetration enhancers (PEs), including natural terpenes. Terpenes, a large and diverse class of organic compounds produced by a variety of plants, are a very safe and effective class of PEs. **Limonene** is one example of a terpene used as a penetration enhancer. The main mechanism for the penetration enhancing action of terpenes is the interaction with SC intercellular lipids. The key factor affecting the enhancement is the lipophilicity of the terpenes and the drug molecules.



Entry of Toxicants via Skin

Some notable toxicants can gain entry into the body following skin contamination. For example:

- Certain commonly used organophosphate pesticides have poisoned agricultural workers following dermal exposure.
- The neurological warfare agent, sarin, readily passes through the skin and can produce quick death to exposed persons.
- Several industrial solvents can cause systemic toxicity by penetrating the skin. For example:
  - Carbon tetrachloride enters the skin and causes liver injury.
  - Hexane can pass through the skin and cause nerve damage.

The skin consists of three main layers of cells as illustrated in Figure 8.4.6.10.4.1:

1. Epidermis
2. Dermis
3. Subcutaneous tissue

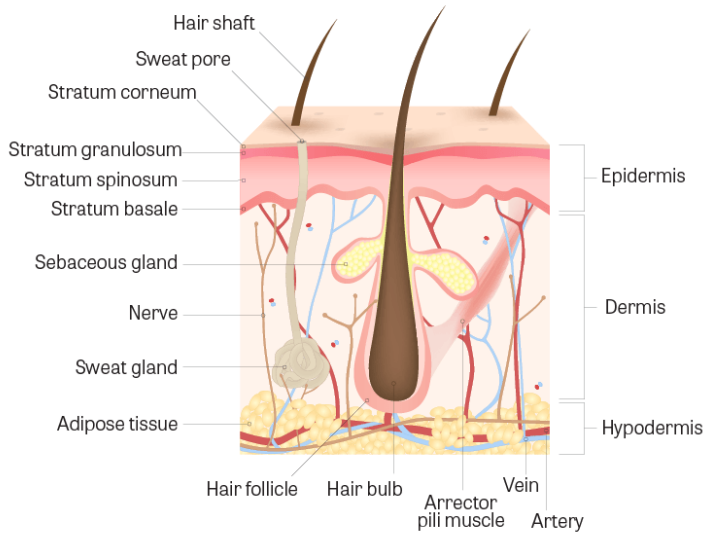


Figure 8.4.6.10.4.1. Layers of the skin
(Image Source: Adapted from iStock Photos, ©)

Epidermis and Stratum Corneum

The epidermis (and particularly the stratum corneum) is the only layer that is important in regulating the penetration of a skin contaminant. It consists of an outer layer of cells, packed with keratin, known as the **stratum corneum** layer. The stratum corneum is devoid of blood vessels. The cell walls of the keratinized cells are apparently double in thickness due to the presence of the keratin, which is chemically resistant and an impenetrable material. The blood vessels are usually about 100 μM from the skin surface. To enter a blood vessel, an agent must pass through several layers of cells that are generally resistant to penetration by chemicals.

Factors Influencing Penetration of the Stratum Corneum

Thickness

The thickness of the stratum corneum varies greatly with regions of the body. The stratum corneum of the palms and soles is very thick (400-600 μM) whereas that of the arms, back, legs, and abdomen is much thinner (8-15 μM). The stratum corneum of the axillary (underarm) and inguinal (groin) regions is the thinnest with the scrotum especially thin. As expected, the ability of toxicants to penetrate that stratum corneum inversely relates to the thickness of the epidermis.

Damage

Any process that removes or damages the stratum corneum can enhance penetration of a xenobiotic. Abrasion, scratching, or cuts to the skin will make it more penetrable. Some acids, alkalis, and corrosives can injure the stratum corneum and make it easier for agents to penetrate this layer. The most prevalent skin conditions that enhance dermal absorption are skin burns and dermatitis.

Passive Diffusion

Toxicants move across the stratum corneum by passive diffusion. There are no known active transport mechanisms functioning within the epidermis. Polar and nonpolar toxicants diffuse through the stratum corneum by different mechanisms:

- **Polar compounds,** which are water soluble, appear to diffuse through the outer surface of the hydrated keratinized layer.
- **Nonpolar compounds,** which are lipid soluble, dissolve in and diffuse through the lipid material between the keratin

filaments.

Water

Water plays an important role in dermal absorption. Normally, the stratum corneum is partially hydrated (approximately 7% by weight). Penetration of polar substances is about 10 times more effective than when the skin is completely dry. Additional hydration on the skin's surface increases penetration by 3–5 times, which further increases the ability of a polar compound to penetrate the epidermis.

Species

Skin penetration can vary by species which can influence the selection of species used for safety testing. Penetration of chemicals through the skin of the monkey, pig, and guinea pig is often similar to that of humans. The skin of the rat and rabbit is generally more permeable whereas the skin of the cat is generally less permeable. For practical reasons and to assure adequate safety, the rat and rabbit have been used for dermal toxicity safety tests.

Other Sites of Dermal Absorption

In addition to the stratum corneum, small amounts of chemicals may be absorbed through the sweat glands, sebaceous glands, and hair follicles. However, since these structures represent only a very small percentage of the skin's total surface area, they are not ordinarily viewed as important contributors to dermal absorption.

Dermis and Subcutaneous Tissue

Once a substance penetrates through the stratum corneum, it enters lower layers of the epidermis, the dermis, and subcutaneous tissue. These layers are far less resistant to further diffusion. They contain a porous, nonselective aqueous diffusion medium which can be penetrated by simple diffusion. Most toxicants that have passed through the stratum corneum can now readily move through the remainder of the skin and enter the circulatory system via the large numbers of venous and lymphatic capillaries in the dermis.

Semivolatile Organic Compounds (SVOCs)

Exposure to **semivolatile organic compounds (SVOCs)** via the dermal route can occur. The amount of SVOCs absorbed via air-to-skin uptake has been estimated to be comparable to or larger than the amount taken in via inhalation for many SVOCs encountered indoors, including:

- Butylated hydroxytoluene (BHT)
- Chlordane
- Chlorpyrifos
- Diethyl phthalate
- Nicotine (in free-base form)
- Other chemicals

The influence of particles and dust on dermal exposure, the role of clothing and bedding as transport vectors, and the potential significance of hair follicles as transport shunts through the epidermis are all areas of research interest.
Human exposure to indoor SVOCs through the dermal pathway has often been underestimated and not considered in exposure assessments. However, exposure scientists, risk assessors, and public health officials are increasingly aware of and interested in the health impacts of dermal exposure. Further, experts seek to understand how health consequences can vary by the exposure pathway. For example, an SVOC that enters the blood through the skin does not encounter the same detoxification pathways that it would encounter when ingested and processed by the stomach, intestines, and liver before entering the blood; its direct entry into the blood can make it potentially more toxic.



Figure 8.4.6.10.4.2. Examples of SVOCs from consumer goods
(Image Source: Adapted from iStock Photos, ©)

Knowledge Check

1) The main barrier to dermal absorption is the:
a) Stratum corneum
b) Dermis
c) Subcutaneous tissue

Answer available at the end of this section.

2) The two primary factors that can increase dermal penetration are:
a) Neutralizing pH and aerosolizing
b) Increasing hydration and disruption of the stratum corneum
c) Dehydrating a substance and increasing particle size

Answer available at the end of this section.

Answers

Knowledge Check

Stratum corneum - **This is the correct answer.**
The epidermis (and particularly the stratum corneum) is the only layer that is important in regulating penetration of a skin contaminant.

Knowledge Check

Increasing hydration and disruption of the stratum corneum - **This is the correct answer.**
Water plays an important role in dermal absorption. Normally, the stratum corneum is partially hydrated (~7% by weight). Penetration of polar substances is about 10 times as effective as when the skin is completely dry. Additional hydration can increase penetration by 3-5 times which further increases the ability of a polar compound to penetrate the epidermis. Any process that removes or damages the stratum corneum can enhance penetration of a xenobiotic.

This page titled 8.4.6.10.4: Dermal Route is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 10.4: Dermal Route by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.10.5: Other Routes of Exposure

Other Routes of Exposure

In addition to the common routes of environmental, occupational, and medical exposure (oral, respiratory, and dermal), other routes of exposure may be used for medical purposes. Many pharmaceuticals are given by parenteral routes via injection into the body using a syringe and hollow needle.

Other Exposure Routes

Intradermal injections are made directly into the skin, just under the stratum corneum. Tissue reactions are minimal and absorption is usually slow. A **subcutaneous injection** is beneath the skin. Since the subcutaneous tissue is quite vascular (consisting of vessels especially those carrying blood), absorption into the systemic circulation is generally rapid. Tissue sensitivity is also high and thus irritating substances may induce pain and an inflammatory reaction.

The **intramuscular** route is used to inject many pharmaceuticals, especially antibiotics and vaccines, directly into muscle tissue. It is an easy procedure and the muscle tissue is less likely to become inflamed compared to subcutaneous tissue. Absorption from muscle is about the same as from subcutaneous tissue.

The **intravenous** (vein) or **intra-arterial** (artery) routes are used to inject substances directly into large blood vessels when they are irritating or when an immediate action is desired, such as anesthesia.

Parenteral injections may also be made directly into body cavities, rarely in humans but frequently in laboratory animal studies. An **intraperitoneal injection** goes directly into the abdominal cavity and an **intrapleural injection** directly into the chest cavity. Since the pleura and peritoneum have minimal blood vessels, irritation is usually minimal and absorption is relatively slow.



Figure 8.4.6.10.5.1. Injection using a syringe
(Image Source: iStock Photos, ©)

Implantation is another route of exposure of increasing concern. A large number of pharmaceuticals and medical devices are

now implanted in various areas of the body. Implants may be used to allow slow, time-release of a substance such as hormones. Implanted medical devices and materials like artificial lenses, tendons, and joints, and cosmetic reconstruction do not involve absorption.

Some materials enter the body via **skin penetration** as the result of accidents or violence (weapons, etc.). The absorption in these cases depends highly on the nature of the substance. Metallic objects (such as bullets) may be poorly absorbed whereas more soluble materials that thrust through the skin and into the body from accidents may be absorbed rapidly into the circulation.

Novel methods of introducing substances into specific areas of the body are often used in medicine. For example, **conjunctival instillations** (eye drops) treat ocular conditions where high concentrations are needed on the outer surface of the eye, not possible by other routes.

Therapies for certain conditions require that a substance is deposited in body openings where high concentrations and slow release may be needed while keeping systemic absorption to a minimum. For these substances, the pharmaceutical agent is suspended in a poorly absorbed material such as beeswax with the material known as a **suppository**. The usual locations for use of suppositories are the rectum and vagina.

Did you know?

Cinnamic aldehyde, also called **cinnamaldehyde**, gives cinnamon its flavor and odor. It occurs naturally in the bark of cinnamon trees and other species. Cinnamic aldehyde and cinnamic alcohol are well known in the scientific literature as being associated with skin allergy in humans. Skin allergy is also called skin sensitization or allergic contact dermatitis. Cinnamic aldehyde is a more potent sensitizer than cinnamic alcohol. The skin absorption and metabolism of cinnamic aldehyde and cinnamic alcohol play an important role in the development of skin sensitization following skin exposures. Cinnamic alcohol applied to human skin is converted to cinnamic aldehyde and cinnamic acid.

Cinnamic aldehyde is a good example of how an assessment of the risk of skin sensitization can be conducted prior to the introduction of new ingredients and products into the marketplace. A **published quantitative risk assessment** for cinnamic aldehyde used the understanding of its chemical, cellular, and molecular properties. By estimating the exposure to cinnamic aldehyde and knowing its allergenic potency, it was possible to assess the sensitization risk of cinnamic aldehyde in different types of consumer products. This publication applied exposure-based risk assessment tools to two hypothetical products containing cinnamic aldehyde. The risk assessment predicted that an eau de toilette leave-on product containing 1000 ppm or more of cinnamic aldehyde would pose an *unacceptable risk* of skin sensitization getting induced. However, a shampoo containing the same level of cinnamic aldehyde would pose an *acceptable risk* of skin sensitization getting induced, based on there being limited exposure to the ingredient from a rinse-off product application.



Figure 8.4.6.10.5.2. Cinnamon sticks and powder
(Image Source: iStock Photos, ©)

Knowledge Check

1) If an immediate therapeutic effect is needed, the route of exposure that would most likely be used is the:

a) Intradermal route
b) Intramuscular injection
c) Intravenous injection

Answer available at the end of this section.

2) A pharmaceutical may be implanted in the body to:

a) Allow slow-release over a long period of time
b) Assure that the substance is distributed equally throughout the body
c) Reduce irritation from the substance

Answer available at the end of this section.

Answers

Knowledge Check

Intravenous injection - **This is the correct answer.**
Substances injected into the circulatory system go directly to the target tissue where immediate reactions can occur.

Knowledge Check

Allow slow-release over a long period of time - **This is the correct answer.**
Treatment with pharmaceuticals in time-release implants is a relatively new therapeutic technique that has gained popularity for long-term chronic chemotherapy.

This page titled 8.4.6.10.5: Other Routes of Exposure is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 10.5: Other Routes of Exposure by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.11: DISTRIBUTION

Learning Objectives

After completing this lesson, you will be able to:

- Explain distribution and its role in toxicokinetics.
- Describe the impact of exposure route on distribution.
- Describe three models of disposition.
- Identify structural barriers to distribution.

In this section...

Topics include:

8.4.6.11.1: Introduction to Distribution

8.4.6.11.2: Influence of Route of Exposure

8.4.6.11.3: Disposition Models

8.4.6.11.4: Structural Barriers to Distribution

8.4.6.11.5: Storage Sites

Section 11: Key Points

What We've Covered

This section made the following main points:

- Distribution is the process in which an absorbed chemical moves away from the site of absorption to other areas of the body.
- An absorbed chemical passes through cell linings of the absorbing organ (skin, lung, or gastrointestinal tract) into the interstitial fluid of that organ.
- The toxicant can leave the interstitial fluid by entering local tissue cells, blood capillaries and the blood circulatory system, or the lymphatic system.
- If the toxicant gains entrance into the blood plasma, it:
  - Travels bound or unbound along with the blood.
  - May be excreted, stored, or biotransformed, or may interact or bind with cellular components.
- The volume of distribution (VD) is the total volume (in liters) of body fluids in which a toxicant is distributed.
- The route of exposure is an important factor affecting the concentration of the toxicant or its metabolites at any specific location within the blood or lymph.
  - Toxicants entering from the GI tract or peritoneum are immediately subject to biotransformation or excretion by the liver and elimination by the lung (this is often called the "first-pass effect").
  - Toxicants absorbed through the lung or skin enter the blood and go directly to the heart and systemic circulation, thus being distributed to various organs before going to the liver (not subject to the first-pass effect).
  - Toxicants that enter the lymph will not go to the liver first, but will slowly enter systemic circulation.

    - The blood level of a toxicant depends on the site of absorption and the rate of biotransformation and excretion.
- Disposition is the combined processes of distirbution, biotransformation, and elimination. Disposition models can be:
    - One-Compartment Open Model — disposition of a substance introduced and distributed instantaneously and evenly in the body and eliminated proportionally to the amount left in the body ("first-order" rate).
    - Two-Compartment Open Model — the chemical enters and distributes in the first compartment (usually blood), then distributed to another compartment where it can be eliminated or may return to the first compartment.
        - The biological half-life, the most commonly used measure of the kinetic behavior of a xenobiotic, is the half-life for a chemical in a two-compartment model.
    - Multiple Compartment Model — the chemical involves several peripheral body compartments, including long-term storage, or biotransformation and elimination at varying rates as blood levels change.
- Organs or tissues differ in the amount of a chemical they may receive, depending on:
    - Volume of blood — organs that receive larger blood volumes potentially accumulate more of a given toxicant.
    - Tissue affinity — some tissues have a higher affinity for specific chemicals, accumulating a toxicant in great concentrations despite a rather low flow of blood.
- Structural barriers to distribution include the blood-brain barrier and the placental barrier.
- Toxicants can also be stored:
    - When bound to plasma proteins in the blood
    - In adipose tissues
    - In bone
    - In the liver
    - In the kidneys

This page titled 8.4.6.11: Distribution is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 11: Distribution by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.11.1: Introduction to Distribution

Introduction to Distribution

So far, we have described the absorption of substances into the body. Now we will focus on what happens next to substances in the body after they are absorbed.

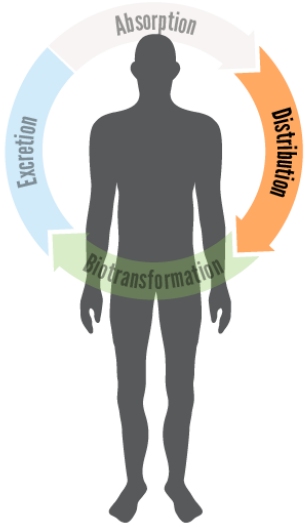


Figure 8.4.6.11.1.1. The processes of toxicokinetics
(Image Source: Adapted from iStock Photos, ©)

Distribution Defined

Distribution is the process in which an absorbed chemical moves away from the site of absorption to other areas of the body. In this section, we will answer the following questions:

- How do chemicals move through the body?
- Does distribution vary with the route of exposure?
- Is a chemical distributed evenly to all organs or tissues?
- How fast is a chemical distributed?
- Why do some chemicals stay in the body for a long time while others are eliminated quickly?

Body Fluids

When a chemical is absorbed, it passes through cell linings of the absorbing organ (skin, lung, or gastrointestinal tract) into the **interstitial fluid** (fluid surrounding cells) of that organ.
The other body fluids are **intracellular fluid** (fluid inside cells) and **blood plasma**. However, the body fluids are not isolated but represent one large pool. The interstitial and intracellular fluids, in contrast to fast moving blood, remain in place with certain components (for example, water and electrolytes) moving slowly into and out of cells. A chemical, while immersed in the interstitial fluid, is not mechanically transported the way it is in blood. Table 1 lists the approximate percentage of body weight each of these body fluids comprise.

Body fluid	Approximate % of total body weight
Interstitial fluid	15%
Intracellular fluid	40%
Blood plasma	8%

Table 1. Approximate percentages that various body fluids contribute to the total body weight

A toxicant can leave the interstitial fluid by entering:

- Local tissue cells.
- Blood capillaries and the blood circulatory system.
- The lymphatic system.

Blood Plasma

If the toxicant gains entrance into the blood plasma, it travels along with the blood, either in a bound or unbound form. Blood moves rapidly through the body via the cardiovascular circulatory system. In contrast, lymph (fluid) moves slowly through the lymphatic system. The major distribution of an absorbed chemical is by blood with only minor distribution by lymph. Since virtually all tissues have a blood supply, all organs and tissues of the body are potentially exposed to the absorbed chemical.

Distribution of a chemical to body cells and tissues requires that the toxicant penetrate a series of cell membranes. It must first penetrate the cells of the capillaries (small blood vessels) and later the cells of the target organs. The **factors pertaining to passage across membranes** apply to these other cell membranes as well. For example, important factors include the concentration gradient; molecular weight; lipid solubility; and polarity, with the smaller, nonpolar toxicants in high concentrations being most likely to gain entrance.

The distribution of a xenobiotic can be affected by whether it binds to plasma protein. Some toxicants may bind to these plasma proteins (especially albumin), which removes the toxicant from potential cell interaction. Within the circulating blood, the non-bound (free) portion is in equilibrium with the bound portion. However, only the free substance is available to pass through the capillary membranes. Those substances that are extensively bound are limited in terms of equilibrium and distribution throughout the body. Protein binding in the plasma greatly affects distribution, prolongs the half-life within the body, and affects the dose threshold for toxicity.

The plasma level of a xenobiotic is important since it generally reflects the concentration of the toxicant at the site of action. The passive diffusion of the toxicant into or out of these body fluids will be determined mainly by the toxicant's concentration gradient.

Volume of Distribution (written as V subscript D - see example in formula below)

The **apparent volume of distribution (VD)** is the total volume of body fluids in which a toxicant is distributed. The VD is expressed in liters.

If a toxicant is distributed only in the plasma fluid, the plasma concentration will remain high and a low VD results; however, if a toxicant is distributed in all sites (blood plasma, interstitial, and intracellular fluids) there is greater dilution in plasma

concentration and a higher VD will result. Binding in effect reduces the concentration of free toxicants in the plasma or VD. Toxicants that undergo rapid storage, biotransformation, or elimination further affect the VD. Toxicologists determine the VD of a toxicant in order to know how extensively a toxicant is distributed in the body fluids. The volume of distribution can be calculated by the formula:

$$V_D = \frac{\text{dose } (mg)}{\text{plasma conc } (mg/L)}$$

The volume of distribution may provide useful estimates as to how extensive the toxicant is distributed in the body. For example, a very high apparent VD may indicate that the toxicant has distributed to a particular tissue or storage area such as adipose tissue. In addition, the body burden for a toxicant can be estimated from knowledge of the VD by using the formula:

$$\text{body burden } (mg) = \text{plasma conc } (mg/L) \times V_D\ (L)$$

Once a chemical is in the blood stream:

- It may be excreted.
- It may be stored.
- It may be biotransformed into different chemicals (metabolites).
- Its metabolites may be excreted or stored.
- The chemical or its metabolites may interact or bind with cellular components.

Most chemicals undergo some biotransformation. The degree with which various chemicals are biotransformed and the degree with which the parent chemical and its metabolites are stored or excreted vary with the nature of the exposure (dose level, frequency, and route of exposure).

Knowledge Check

1) When an ingested toxicant is absorbed, it passes through the cells lining the GI tract into the:

a) Intracellular fluid
b) Gastric fluid
c) Interstitial fluid

Answer available at the end of this section.

2) The apparent volume of distribution (VD) represents the:

a) Total volume of body fluids in which a toxicant is distributed
b) Amount of blood plasma in which a toxicant is dissolved
c) Amount of interstitial fluid that contains a toxicant

Answer available at the end of this section.

Answers

Knowledge Check

Interstitial fluid - **This is the correct answer.**
When a chemical is absorbed it passes through cell linings of the absorbing organ (in this case, the gastrointestinal tract) into the interstitial fluid (fluid surrounding cells) within that organ.

Knowledge Check

Total volume of body fluids in which a toxicant is distributed - **This is the correct answer.**
The apparent volume of distribution (VD) represents the total volume of body fluids in which a toxicant is distributed. It consists of the interstitial fluid, intracellular fluid, and the blood plasma. Soon after absorption, a toxicant may be distributed to all three types of fluids, although the concentrations may be quite different. Rarely will a toxicant be distributed to only one type of fluid.

This page titled 8.4.6.11.1: Introduction to Distribution is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 11.1: Introduction to Distribution by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.11.2: Influence of Route of Exposure

Influence of Exposure Route

The **route of exposure** is an important factor that can affect the concentration of the toxicant (or its metabolites) at any specific location within the blood or lymph. This can be important since the time and path taken by the chemical as it moves through the body influences the degree of biotransformation, storage, and elimination (and thus toxicity).

For example, if a chemical goes to the liver before going to other parts of the body, much of it may be biotransformed quickly. In this case, the blood levels of the toxicant "downstream" may be diminished or eliminated. This way of processing the chemical right away can dramatically affect its potential toxicity.

Gastrointestinal Tract and Peritoneum

When toxicants are absorbed through the **gastrointestinal (GI) tract**, a similar biotransformation process occurs. Blood carries absorbed toxicants entering the vascular system of the GI tract directly to the liver via the portal system. This is also true for those drugs administered by intraperitoneal injection. Blood from most of the peritoneum also enters the portal system and goes immediately to the liver. Blood from the liver then flows to the heart and then on to the lung, before going to other organs.

Thus, toxicants entering from the GI tract or peritoneum are immediately subject to biotransformation or excretion by the liver and elimination by the lung. This is often referred to as the "first-pass effect."

For example, first-pass biotransformation of the drug propranolol (cardiac depressant) is about 70% when given orally. This means the blood level of this medication is only about 30% of a comparable dose administered intravenously.

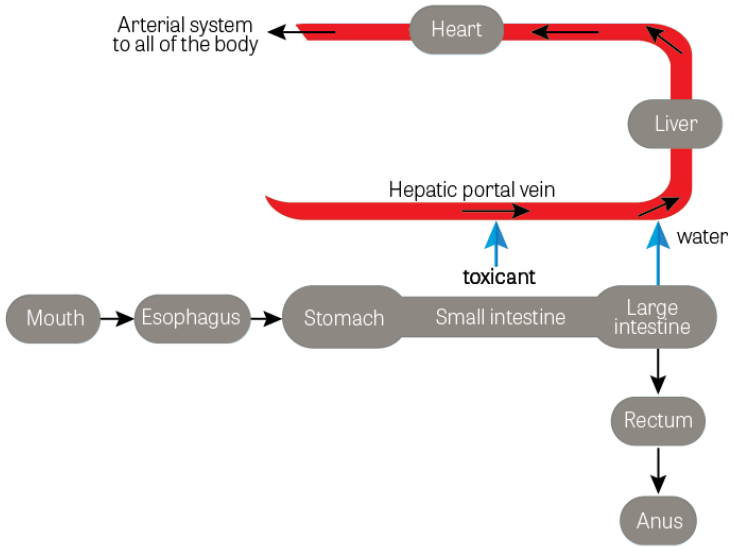


Figure 8.4.6.11.2.1. Movement of a toxicant through the portal system
(Image Source: Adapted from Kimball's Biology Pages. Original author: John W. Kimball, obtained under Creative Commons Attribution 3.0 Unported license, ©. View original image.)

Lung and Skin

Drugs and other substances that are absorbed through the **lungs** or **skin** enter the bloodstream to be carried throughout the body. Thus, they avoid the liver (hepatic) first-pass effect that would have occurred if they had been absorbed from the gastrointestinal tract. These substances can have local effects in the lungs or skin in addition to having systemic effects, and some cells in the lungs and skin may metabolize the drug or other substance. Examples of a "local first-pass effect" in the skin due to metabolism are when nitroglycerin and cortisol applied to the skin. Drugs administered intravenously or intramuscularly also enter the bloodstream to be carried throughout the body and avoid the liver (hepatic) first-pass effect.

Did you know?

Some advantages of transdermal drug delivery (skin patches):

- They are a better way to deliver substances that are broken down by the stomach acids, not well absorbed from the gut, or extensively broken down by the liver.
- They are a substitute for oral route.
- They permit constant dosing rather than the peaks and valley in medication level associated with orally administered medication.
- They can minimize undesirable side effects.
- They can be used to prescribe drugs that have short biological half-lives or a narrow therapeutic window.
- They can be removed, thereby terminating therapy easily.
- They are noninvasive, avoiding the inconvenience of IV therapy or injections.
- They can be used with patients who are nauseated or unconscious.
- They are cost-effective.



Figure 8.4.6.11.2.2. Nicotine patch (Image Source: iStock Photos, ©)

Lymph

The delivery of drugs and bioactive compounds via the lymphatic system avoids first-pass metabolism by the liver and increases oral bioavailability. It is also a way to deliver drugs for diseases that spread through the lymphatic system such as certain types of cancer and the human immunodeficiency virus (HIV). For example, **liposomes** composed of phosphatidylethanol can enhance the oral bioavailability of poorly absorbed hydrophilic drugs such as cefotaxime.

Blood

The blood levels of a drug or other substance depend on the site of absorption, whether being absorbed after subcutaneous injection or more quickly from intramuscular injection. These blood levels also depend on the individual's rate of local and systemic biotransformation, and the rate of excretion. Uptake and release can occur in areas of the body away from the first site of absorption. Some anesthetics can be taken up by the lungs and later released, impacting blood levels. Lidocaine, given intravenously, is one example of this later release. Further, as noted elsewhere in ToxTutor, the metabolism of a substance can vary widely from person-to-person due to factors such as genetic differences, age, diet, and diseases that affect metabolism.

Some advantages of intramuscular injections:

- They are absorbed faster than subcutaneous injection, partly because muscle tissue has a larger blood supply than tissue just under the skin.

- They can hold a greater injected volume of drug (or vaccine) than a subcutaneous tissue injection can.
- They can be used instead of intravenous injection if a drug is irritating to veins or if a suitable vein cannot be located.
- They may be used instead of oral delivery if a drug is known to be degraded by stomach acids.

Knowledge Check

1) The main difference in distribution of a toxicant absorbed from the gastrointestinal tract from toxicants absorbed through the skin or from inhalation is:

a) The toxicant is distributed to more organs
b) A greater amount of the toxicant that is absorbed will be distributed to distant parts of the body
c) The toxicant enters the systemic circulatory system after first passing through the liver

Answer available at the end of this section.

Answers

Knowledge Check

The toxicant enters the systemic circulatory system after first passing through the liver - **This is the correct answer.**
Toxicants that enter the vascular system of the gastrointestinal tract are carried directly to the liver by the portal system. Thus, toxicants are immediately subject to biotransformation or excretion by the liver. This is often referred to as the "first pass effect."

This page titled 8.4.6.11.2: Influence of Route of Exposure is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 11.2: Influence of Route of Exposure by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.11.3: Disposition Models

Disposition Models

Disposition is the term often used to describe the combined processes of distribution, biotransformation, and elimination. The most used disposition models are compartmental models, which are categorized as one-compartment, two-compartment, and multicompartment models. Compartmental models can be used to predict the time course of drug concentrations in the body. These compartments could represent a group of similar tissues or fluids.
For example:

- Blood is a compartment.
- Fat (adipose) tissue, bone, liver, kidneys, and brain are other major compartments.

Kinetic models may be a 1) one-compartment open model; 2) two-compartment open model; or 3) multiple compartment model.

One-Compartment Model

A **one-compartment open model** may be used for drugs like aminoglycosides which rapidly distribute (equilibrate) to tissues and fluids within the body. In other words, the entire body acts like one uniform compartment. This model is also called a one-compartment open model, with "open" being the assumption that the drug can enter and leave the body via excretion. The figure below shows the disposition of a drug or other substance that distributes instantaneously and evenly in the body, and is eliminated at a rate and amount that is proportional to the amount left in the body. This is known as a "first-order" rate and represented as the logarithm of concentration in blood as a linear function of time (Figure 8.4.6.11.3.1).

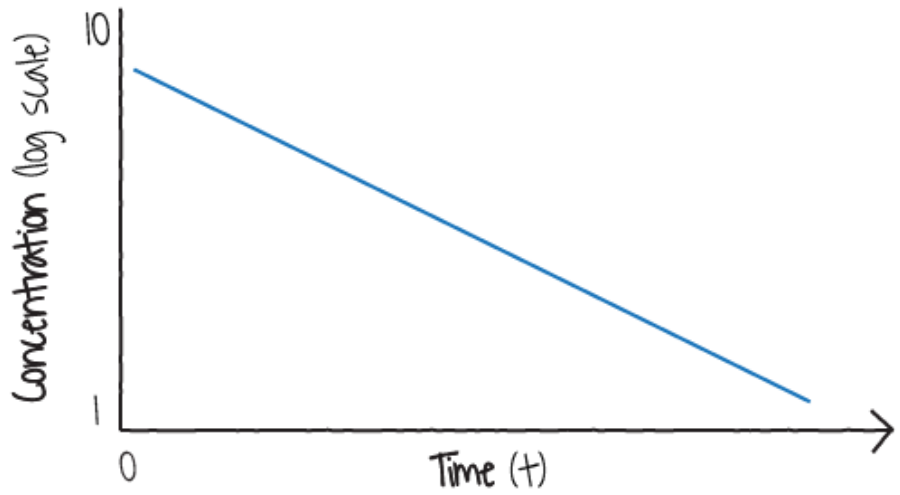


Figure 8.4.6.11.3.1. One-compartment open model
(Image Source: NLM)

The **half-life** of the chemical that follows a one-compartment model is simply the time required for half of the chemical to be lost from the plasma. Only a few chemicals actually follow the simple, first-order, one compartment model.

Two-Compartment Model

For most chemicals, it is necessary to describe the kinetics in terms of at least a two-compartment model. A two-compartment model is used for drugs which distribute slowly within the body. This model is also called a two-compartment open model, with "open" being the assumption that the drug can enter and leave the body.
For example, a one-time (bolus) intravenous administration over a short time period could lead to a drug distributing rapidly in the blood and also to highly perfused (by blood) organs like the liver and kidneys. This would be one compartment of the two-compartment model. There would be a slower distribution to other parts of the body as the second compartment.

- Two examples are vancomycin and digoxin. As shown in Figure 8.4.6.11.3.2, the drug or other substance enters and distributes in the first compartment. It is then distributed to another compartment. The concentration in the first compartment declines with time while the concentration in the second compartment rises, peaks, and declines as the chemical is eliminated from the body.

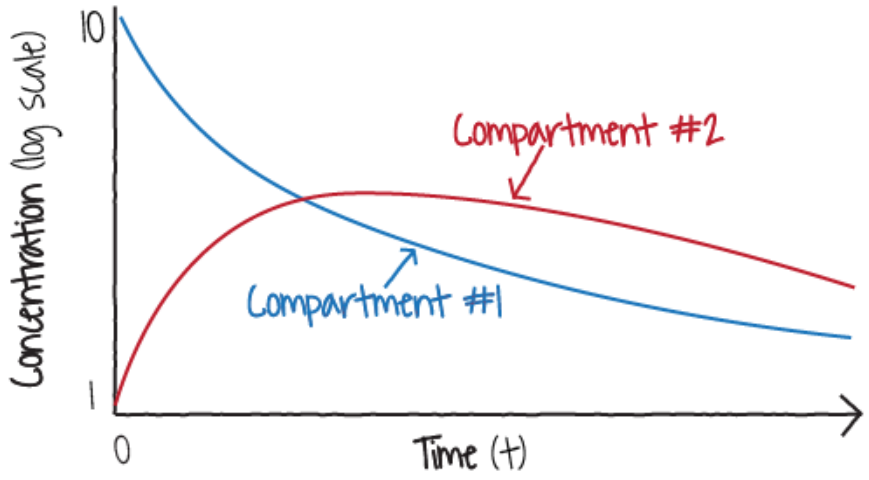


Figure 8.4.6.11.3.1. Two-compartment open model
(Image Source: NLM)

A half-life for a chemical whose kinetic behavior fits a two-compartment model is often referred to as the "biological half-life." This is the most commonly used measure of the kinetic behavior of a xenobiotic.

Multiple Compartment Model

Frequently the one- and two-compartment models cannot adequately describe the kinetics of a chemical within the body since there may be several peripheral body compartments to which the chemical may go, including long-term storage. In addition, biotransformation and elimination of a chemical may not be simple processes but subject to different rates as blood levels change.

Knowledge Check

1) Disposition models describe:

a) How a toxicant moves within the body over time
b) How the body eliminates the toxicant
c) The pathway for biotransformation of the toxicant within the liver

Answer available at the end of this section.

Answers

Knowledge Check

How a toxicant moves within the body over time - **This is the correct answer.**
Disposition models, also known as kinetic models, describe how a toxicant moves within the body compartments with time.

This page titled 8.4.6.11.3: Disposition Models is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 11.3: Disposition Models by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.11.4: Structural Barriers to Distribution

Structural Barriers to Distribution

Organs or tissues differ in the amount of a chemical that they receive or to which they are exposed. This is primarily due to two factors: the 1) **volume of blood** flowing through a specific tissue and the 2) **presence of special barriers** to slow down a toxicant's entrance.

Volume of Blood and Tissue Affinity

Organs that receive larger **blood volumes** can potentially accumulate more of a given toxicant. Body regions that receive a large percentage of the total cardiac output include the liver (28%), kidneys (23%), heart muscle, and brain. Bone and adipose tissues have relatively low blood flow, even though they serve as primary storage sites for many toxicants. This is especially true for toxicants that are fat-soluble and those that readily associate (or form complexes) with minerals commonly found in bone.
Tissue affinity determines the degree of concentration of a toxicant. In fact, some tissues have a higher affinity for specific chemicals and accumulate a toxicant in great concentrations despite a rather low flow of blood.
For example, adipose tissue, which has a meager blood supply, concentrates lipid-soluble toxicants. Once deposited in these storage tissues, toxicants may remain for long periods, due to their solubility in the tissue and the relatively low blood flow.

Structural Barriers

During distribution, the passage of toxicants from capillaries into various tissues or organs is not uniform. **Structural barriers** exist that restrict the entrance of toxicants into certain organs or tissues. The primary barriers are those of the brain, placenta, and testes.

Blood-Brain Barrier

The **blood-brain barrier** protects the brain from most toxicants. Specialized cells called astrocytes possess many small branches, which form a barrier between the capillary endothelium and the neurons of the brain. Lipids in the astrocyte cell walls and very tight junctions between adjacent endothelial cells limit the passage of water-soluble molecules. The blood-brain barrier is not completely impenetrable and its penetrability can vary with health status/disease state, but it does slow down the rate at which toxicants cross into brain tissue while allowing essential nutrients, including oxygen, to pass through.

Placental Barrier

The **placental barrier** protects the sensitive, developing fetus from most toxicants distributed in the maternal circulation. This barrier consists of several cell layers between the maternal and fetal circulatory vessels in the placenta. Lipids in the cell membranes limit the diffusion of water-soluble toxicants. However, nutrients, gases, and wastes of the developing fetus can pass through the placental barrier. As in the case of the blood-brain barrier, the placental barrier is not completely impenetrable but effectively slows down the diffusion of most toxicants from the mother into the fetus.

Knowledge Check

1) Organs may differ greatly in the concentration of a toxicant in them, due primarily to the:

a) Rate of elimination of the toxicant by the kidneys
b) Distance of the organ from the heart since the toxicant disintegrates quickly in the blood plasma
c) Volume of blood flow and the presence of special barriers

Answer available at the end of this section.

2) The placental barrier protects the fetus from toxicants in the maternal blood because:

a) Substances in the maternal blood must move through several layers of cells in order to gain entrance to placental blood
b) The placenta does not contain circulating fetal blood that can absorb toxicants from the maternal blood
c) Toxicants in maternal blood are usually lipid soluble and must be water-soluble in order to penetrate through the placental cell layers

Answer available at the end of this section.

Answers

Knowledge Check

Volume of blood flow and the presence of special barriers - **This is the correct answer.**
Organs or tissues differ in the amount of a chemical that they receive or to which they are exposed. This is primarily due to two factors, the **volume of blood** flowing through a specific tissue and the **presence of special "barriers"** to slow down toxicant entrance. Organs that receive larger blood volumes can potentially accumulate more of a given toxicant.

Knowledge Check

Substances in the maternal blood must move through several layers of cells in order to gain entrance to placental blood - **This is the correct answer.**
The placental barrier protects the developing and sensitive fetus from most toxicants distributed in the maternal circulation. This barrier consists of several cell layers between the maternal and fetal circulatory vessels in the placenta. Lipids in the cell membranes limit the diffusion of water-soluble toxicants.

This page titled 8.4.6.11.4: Structural Barriers to Distribution is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 11.4: Structural Barriers to Distribution by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.11.5: Storage Sites

Storage Sites

Storage of toxicants in body tissues sometimes occurs. Initially, when a toxicant enters the blood plasma, it may be bound to plasma proteins. Toxicants attached to proteins are considered a form of storage because they do not contribute to the chemical's toxic potential. Albumin is the most abundant plasma protein that binds toxicants. Normally, the toxicant is only bound to the albumin for a relatively short time.

The primary sites for toxicant storage are adipose tissue, bone, liver, and kidneys.

Adipose Tissue

Lipid-soluble toxicants are often stored in **adipose tissues**. Adipose tissue is located in several areas of the body but mainly in subcutaneous tissue. Lipid-soluble toxicants can be deposited along with triglycerides in adipose tissues. The lipids are in a continual exchange with blood and thus the toxicant may be mobilized into the blood for further distribution and elimination, or redeposited in other adipose tissue cells.

Bone

Bone is another major site for storage. Bone is composed of proteins and the mineral salt hydroxyapatite. Bone contains a sparse blood supply but is a live organ. During the normal processes that form bone, calcium and hydroxyl ions are incorporated into the hydroxyapatite-calcium matrix. Several chemicals, primarily elements, follow the same kinetics as calcium and hydroxyl ions and therefore can be substituted for them in the bone matrix.

For example, strontium (Sr) or lead (Pb) may be substituted for calcium (Ca), and fluoride (F-) may be substituted for hydroxyl (OH-) ions. Bone is continually being remodeled under normal conditions. Calcium and other minerals are continually being resorbed and replaced, on the average about every 10 years. Thus, any toxicants stored in the matrix will eventually be released to re-enter the circulatory system.

Liver and Kidneys

The **liver** is a storage site for some toxicants. It has a large blood flow and its hepatocytes (that is, liver cells) contain proteins that bind to some chemicals, including toxicants.
As with the liver, the **kidneys** have a high blood flow, which preferentially exposes these organs to toxicants in high concentrations. Storage in the kidneys is associated primarily with the cells of the nephron (the functional unit for urine formation).

Knowledge Check

1) The areas of the body which most frequently store toxicants are:

a) Adrenal gland, thyroid gland, and pancreas
b) Adipose tissue, bone, liver, and kidney
c) Skeletal muscle, tendons, and leg joints

Answer available at the end of this section.

Answers

Knowledge Check

1) Adipose tissue, bone, liver, and kidney - **This is the correct answer.**
The primary sites for toxicant storage are adipose tissue, bone, liver and kidneys. Lipid-soluble toxicants store in adipose tissues; chemicals that follow calcium or hydroxyl ion kinetics store in bone; and the liver and kidney cells are subjected to high concentrations of toxicants.

This page titled 8.4.6.11.5: Storage Sites is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 11.5: Storage Sites by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.12: BIOTRANSFORMATION

Learning Objectives

After completing this lesson, you will be able to:

- Explain biotransformation, including its importance to survival and the body sites it involves.
- Define enzymes and the three types of enzyme specificity.
- Explain the two phases of biotransformation.
- Identify factors that influence the effectiveness of biotransformation.

In this section...

Topics include:

8.4.6.12.1: Introduction to Biotransformation

8.4.6.12.2: Chemical Reactions

8.4.6.12.3: Biotransformation Sites

8.4.6.12.4: Modifiers of Biotransformation

Section 12: Key Points

What We've Covered

This section made the following main points:

- Biotransformation is the process by which a substance changes from one chemical to another (transformed) by a chemical reaction within the body.
- Biotransformation is vital to survival because it transforms absorbed nutrients into substances required for normal body functions.
- Potential complications of biotransformation include:
  - Detoxification — biotransformation results in metabolites of lower toxicity than the parent substance.
  - Bioactivation — biotransformation results in metabolites of greater toxicity than the parent substance.
- Chemical reactions continually occur in the body to build up new tissue, tear down old tissue, convert food to energy, dispose of waste materials, and eliminate toxic xenobiotics.
- Enzymes are catalysts for nearly all biochemical reactions in the body; essential biotransformation reactions would be slowed or prevented without these enzymes, causing major health problems.
- There are generally three types of enzyme specificity:
  1. Enzymes with absolute specificity catalyze only one reaction.
  2. Enzymes with group specificity act only on molecules that have specific functional groups.
  3. Enzymes with linkage specificity act on a particular type of chemical bond regardless of the rest of the molecular structure.
- There are two biotransformation reaction phases:
  1. Phase I reactions modify the chemical by adding a functional structure, allowing the substance to "fit" into a second (Phase II) enzyme:
     - Oxidation — the substrate loses electrons.
     - Reduction — the substrate gains electrons.

    - Hydrolysis — the addition of water splits the toxicant into two fragments or smaller molecules.
2. Phase II reactions conjugate (join together) the modified xenobiotic with another substance. The most important Phase II reactions are:
    - Glucuronide conjugation, a high-capacity pathway — glucuronic acid is added directly to the toxicant or its Phase I metabolite, generally resulting in hydrophilic conjugates excreted by the kidney or bile.
    - Sulfate conjugation, a low-capacity pathway — decreases the toxicity of xenobiotics, resulting in highly polar sulfate conjugates readily secreted in the urine.

- Biotransformation sites are the:
    - Liver (primary site, which also makes it the most susceptible to damage by ingested toxicants).
    - Kidneys (about 10-30% of the liver's capacity).
    - Skin, intestines, testes, and placenta (low capacity).
- Biotransformation effectiveness depends on factors that can inhibit or induce enzymes and dose levels, including species, age, gender, genetic variability, nutrition, disease, exposure to other chemicals, and the dose level.

This page titled 8.4.6.12: Biotransformation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 12: Biotransformation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.12.1: Introduction to Biotransformation

Biotransformation is the process by which a substance changes from one chemical to another (transformed) by a chemical reaction within the body. **Metabolism** or **metabolic transformations** are terms frequently used for the biotransformation process. However, metabolism is sometimes not specific for the transformation process but may include other phases of toxicokinetics.

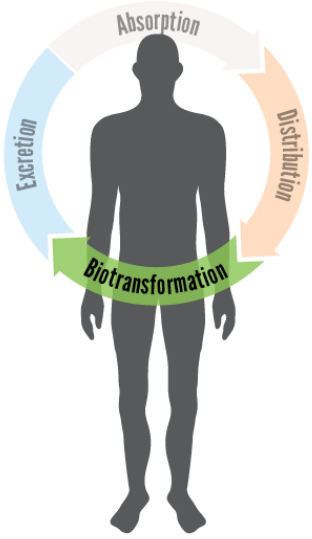


Figure 8.4.6.12.1.1. Processes of toxicokinetics (Image Source: Adapted from iStock Photos, ©)

Importance of Biotransformation

Biotransformation is vital to survival because it transforms absorbed nutrients (food, oxygen, etc.) into substances required for normal body functions. For some pharmaceuticals, it is a metabolite that is therapeutic and not the absorbed drug. For example, phenoxybenzamine, a drug given to relieve hypertension caused by pheochromocytoma, a kind of tumor, is biotransformed into a metabolite, which is the active agent. Biotransformation also serves as an important defense mechanism since toxic xenobiotics and body wastes are converted into less harmful substances and substances that can be excreted from the body.

Toxicants that are lipophilic, non-polar, and of low molecular weight are readily absorbed through the cell membranes of the skin, GI tract, and lung. These same chemical and physical properties control the distribution of a chemical throughout the body and its penetration into tissue cells. Lipophilic toxicants are hard for the body to eliminate and can accumulate to hazardous levels. However, most lipophilic toxicants can be transformed into hydrophilic metabolites that are less likely to pass through membranes of critical cells. Hydrophilic chemicals are easier for the body to eliminate than lipophilic substances. Biotransformation is thus a key body defense mechanism.

Fortunately, the human body has a well-developed capacity to biotransform most xenobiotics as well as body wastes.

Did you know?

Hemoglobin, the oxygen-carrying iron-protein complex in red blood cells, is an example of a body waste that must be eliminated. The normal destruction of aged red blood cells releases hemoglobin. Bilirubin is one of several hemoglobin metabolites. If the body cannot eliminate bilirubin via the liver because of disease, medicine, or infection, bilirubin builds up

in the body and the whites of the eyes and the skin may look yellow. Bilirubin is toxic to the brain of newborns and, if present in high concentrations, may cause irreversible brain injury. Biotransformation of the lipophilic bilirubin molecule in the liver results in the production of water-soluble (hydrophilic) metabolites excreted into bile and eliminated via the feces.

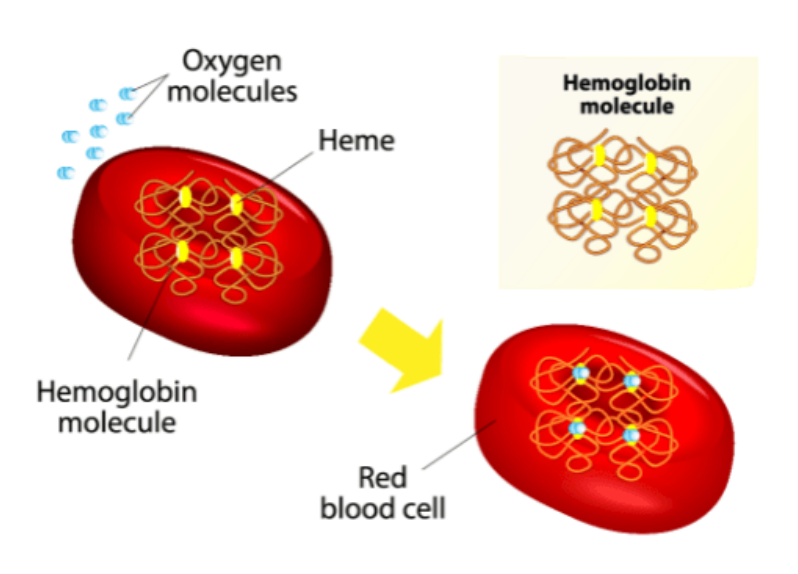


Figure 8.4.6.12.1.2. Human hemoglobin (Image Source: Adapted from iStock Photos, ©)

Potential Complications

The biotransformation process is not perfect. **Detoxification** occurs when biotransformation results in metabolites of lower toxicity. In many cases, however, the metabolites are more toxic than the parent substance, a process called **bioactivation**. Occasionally, biotransformation can produce an unusually reactive metabolite that may interact with cellular macromolecules like DNA. This can lead to very serious health effects such as cancer or birth defects.
An example is the biotransformation of vinyl chloride into vinyl chloride epoxide, which covalently binds to DNA and RNA, a step leading to cancer of the liver.

Knowledge Check

The term "biotransformation" refers to:

a. An increase in electrical charge in tissues produced by a biological transformer
b. Chemical reactions in the body that create a new chemical from another chemical
c. The transformation of one type of cell in a tissue to another type of cell

Answer available at the end of this section.

Detoxification is a biotransformation process in which:

a. Metabolites of lower toxicity are produced
b. Metabolites of higher toxicity are produced

Answer available at the end of this section.

Bioactivation is a biotransformation process in which:

a. Metabolites of lower toxicity are produced
b. Metabolites of higher toxicity are produced

Answer available at the end of this section.

Answers

Knowledge Check

Chemical reactions in the body that create a new chemical from another chemical - **This is the correct answer.**
Biotransformation is the process whereby a substance is changed from one chemical to another (transformed) by a chemical reaction within the body.

Knowledge Check

Metabolites of lower toxicity are produced - **This is the correct answer.**
When biotransformation results in metabolites of lower toxicity, the process is known as detoxification.

Knowledge Check

Metabolites of higher toxicity are produced - **This is the correct answer.**
When biotransformation results in metabolites of higher toxicity, this is known as bioactivation.

This page titled 8.4.6.12.1: Introduction to Biotransformation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 12.1: Introduction to Biotransformation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.12.2: Chemical Reactions

Chemical Reactions

Chemical reactions are continually taking place in the body. They are a normal aspect of life, participating in the:

- Building up of new tissue.
- Tearing down of old tissue.
- Conversion of food to energy.
- Disposal of waste materials.
- Elimination of toxic xenobiotics.

Within the body is a magnificent assembly of chemical reactions, which is well orchestrated and called upon as needed. Most of these chemical reactions occur at significant rates only because specific proteins, known as enzymes, are present to catalyze them, that is, accelerate the reaction. A catalyst is a substance that can accelerate a chemical reaction of another substance without itself undergoing a permanent chemical change.

Enzymes

Enzymes are the catalysts for nearly all biochemical reactions in the body. Without these enzymes, essential biotransformation reactions would take place slowly or not at all, causing major health problems.

Did you know?

Phenylketonuria (PKU) is the genetic condition in which the enzyme that biotransforms phenylalanine to tyrosine (another amino acid) is defective. As the result, phenylalanine can build up in the body and cause severe mental retardation. Babies are routinely checked at birth for PKU. If they have PKU, they need to follow a special diet to restrict the intake of phenylalanine in infancy and childhood.



Figure 8.4.6.12.2.1. Phenylketonuia (PKU) testing in an infant
(Image Source: Wikimedia Commons, obtained under Public Domain license. Author: U.S. Air Force Photographic Archives)

These enzymatic reactions are not always simple biochemical reactions. Some enzymes require the presence of cofactors or coenzymes in addition to the **substrate** (the substance to be catalyzed) before their catalytic activity can be exerted. These co-factors exist as a normal component in most cells and are frequently involved in common reactions to convert nutrients into energy (vitamins are an example of co-factors). It is the drug or chemical transforming enzymes that hold the key to xenobiotic transformation. The relationship of substrate, enzyme, coenzyme, and transformed product can be shown as:

$$substrate * \frac{enzyme}{co-enzyme} = transformedproduct$$

Most biotransforming enzymes are high molecular weight proteins, composed of chains of amino acids linked together by peptide bonds. A wide variety of biotransforming enzymes exist. Most enzymes will catalyze the reaction of only a few substrates, meaning that they have high specificity. Specificity is a function of the enzyme's structure and its catalytic sites. While an enzyme may encounter many different chemicals, only those chemicals (substrates) that fit within the enzyme's convoluted structure and spatial arrangement will be locked on and affected. This is sometimes referred to as the "lock and key" relationship.
As shown in Figure 8.4.6.12.2.2, when a substrate fits into the enzyme's structure, an enzyme-substrate complex can be formed. This allows the enzyme to react with the substrate with the result that two different products are formed. If the substrate does not fit into the enzyme ("incompatible"), no complex will be formed and thus no reaction can occur.

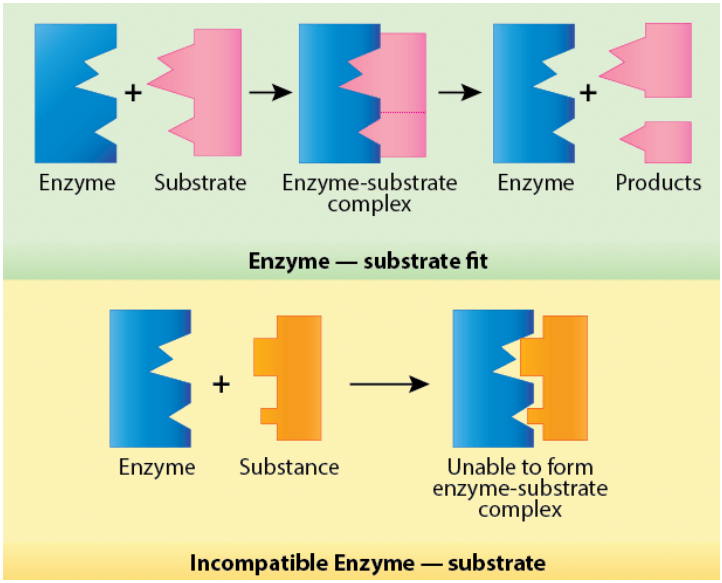


Figure 8.4.6.12.2.2. If the substrate does not fit into the enzyme, no complex will be formed and no reaction will occur.
(Image Source: NLM)

Enzyme Specificity

Enzymes range from having absolute specificity to broad and overlapping specificity. In general, there are three main types of specificity:

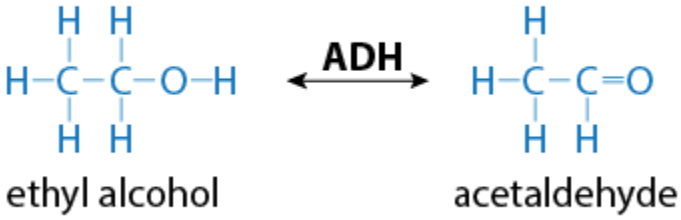
1. **Absolute** — the enzyme will catalyze only one reaction. Examples:
   - Formaldehyde dehydrogenase catalyzes only the reaction for formaldehyde.
   - Acetylcholinesterase biotransforms the neurotransmitting chemical, acetylcholine.
2. **Group** — the enzyme will act only on molecules that have specific functional groups, such as amino, phosphate, or methyl groups.
   - For example, alcohol dehydrogenase can biotransform several different alcohols, including methanol and ethanol.
3. **Linkage** — the enzyme will act on a particular type of chemical bond regardless of the rest of the molecular structure.
   - For example, N-oxidation can catalyze a reaction of a nitrogen bond, replacing the nitrogen with oxygen.

Enzyme Naming Convention

The names assigned to enzymes may seem confusing at first. However, except for some of the originally studied enzymes (such as pepsin and trypsin), a convention has been adopted to name enzymes. Enzyme names end in "ase" and usually combine the substrate acted on and the type of reaction catalyzed.

For example, alcohol dehydrogenase is an enzyme that biotransforms alcohols by the removal of a hydrogen. The result is a completely different chemical, an aldehyde or ketone.

The biotransformation of ethyl alcohol to acetaldehyde is depicted in Figure 8.4.6.12.2.3.



ADH = alcohol dehydrogenase, a specific catalyzing enzyme

Figure 8.4.6.12.2.3. Biotransformation of ethyl alcohol
(Image Source: NLM)

Beneficial or Harmful?

At this point in ToxTutor you likely see that the transformation of a specific xenobiotic can be either beneficial or harmful, and perhaps both depending on the dose and circumstances.
A good example is the biotransformation of acetaminophen (Tylenol®). When the prescribed doses are taken, the desired therapeutic response is observed with little or no toxicity. However, when excessive doses of acetaminophen are taken, hepatotoxicity can occur. This is because acetaminophen normally undergoes rapid biotransformation with the metabolites quickly eliminated in the urine and feces.
At high doses, the normal level of enzymes may be depleted and the acetaminophen is available to undergo the reaction by an additional biosynthetic pathway, which produces a reactive metabolite that is toxic to the liver. For this reason, a user of Tylenol® is warned not to take the prescribed dose more frequently than every 4–6 hours and not to consume more than four doses within a 24-hour period.
Biotransforming enzymes, like most other biochemicals, are available in a normal amount and in some situations can be "used up" at a rate that exceeds the body's ability to replenish them. This illustrates the frequently used phrase, the "dose makes the poison."



Figure 8.4.6.12.2.4. Generic acetaminophen tablets
(Image Source: iStock Photos, ©)

Biotransformation Reaction Phases

Biotransformation reactions are categorized not only by the nature of their reactions, for example, oxidation, but also by the

normal sequence with which they tend to react with a xenobiotic. They are usually classified as Phase I and Phase II reactions.

Phase I reactions are generally reactions which modify the chemical by adding a functional structure. This allows the substance to "fit" into a second, or Phase II enzyme, so that it can become conjugated (joined together) with another substance.

Phase II reactions consist of those enzymatic reactions that conjugate the modified xenobiotic with another substance. The conjugated products are larger molecules than the substrate and generally polar in nature (water soluble). Thus, they can be readily excreted from the body. Conjugated compounds also have poor ability to cross cell membranes.

In some cases, the xenobiotic already has a functional group that can be conjugated and the xenobiotic can be biotransformed by a Phase II reaction without going through a Phase I reaction.

For example, phenol can be directly conjugated into a metabolite that can then be excreted. The biotransformation of benzene requires both Phase I and Phase II reactions. As illustrated in Figure 8.4.6.12.2.5, benzene is biotransformed initially to phenol by a Phase I reaction (oxidation). Phenol has a structure including a functional hydroxyl group that is then conjugated by a Phase II reaction (sulfation) to phenyl sulfate.

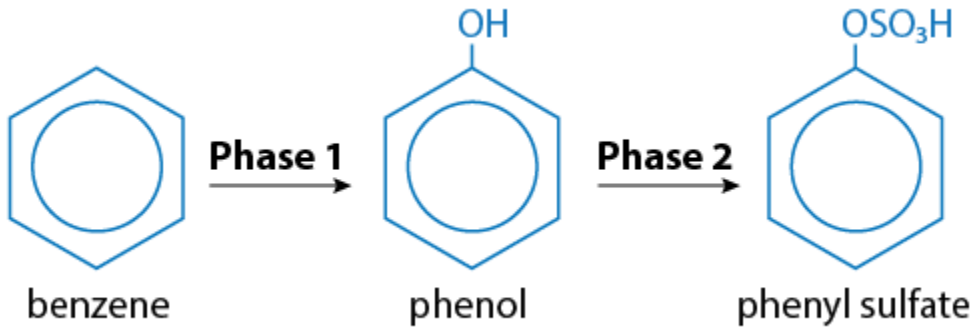


Figure 8.4.6.12.2.5. Biotransformation of benzene into phenol in Phase 1 (oxidation), which is then conjugated by a Phase 2 reaction (sulfation) to phenyl sulfate
(Image Source: NLM)

Table 1 lists the major transformation reactions for xenobiotics broken into Phase I and Phase II reactions. These reactions are discussed in more detailed below.

Phase I	Phase II
Oxidation	Sulfate conjugation
Reduction	Glucuronide conjugation
Hydrolysis	Glutathione conjugation
Acetylation	Amino acid conjugation

Table 1. Phase I and II biotransformation reactions for xenobiotics

Phase I Reactions

Phase I biotransformation reactions are simple reactions compared to Phase II reactions. In Phase I reactions, a small polar group (containing both positive and negative charges) is either exposed on the toxicant or added to the toxicant. The three main Phase I reactions are 1) oxidation; 2) reduction; and 3) hydrolysis.

Oxidation

Oxidation is a chemical reaction in which a *substrate loses electrons*. There are a number of reactions that can achieve the removal of electrons from the substrate.

- The **addition of oxygen**, or **oxygenation**, was the first of these reactions discovered and thus the reaction was named oxidation. However, many of the oxidizing reactions do not involve oxygen.
- The simplest type of oxidation reaction is **dehydrogenation**, which is the removal of hydrogen from the molecule.
- Another example of oxidation is **electron transfer** that consists simply of the transfer of an electron from the substrate.

Figure 8.4.6.12.2.6 shows these types of oxidizing reactions.

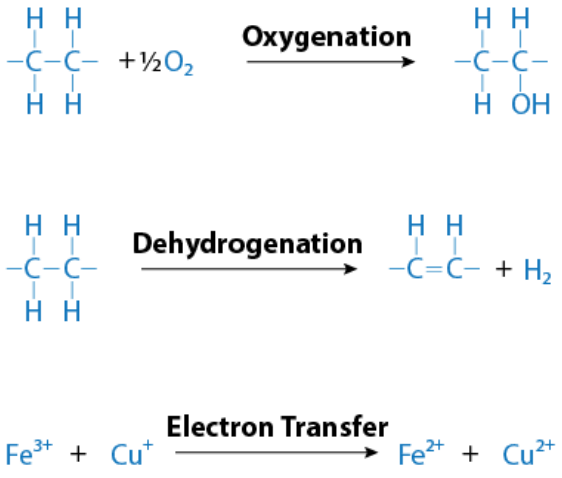


Figure 8.4.6.12.2.6. Three types of oxidation reactions
(Image Source: NLM)

The specific oxidizing reactions and oxidizing enzymes are numerous and several textbooks are devoted to this subject. Most of the reactions are described by the name of the reaction or enzyme involved. Some of these oxidizing reactions include:

- Alcohol dehydrogenation
- Aldehyde dehydrogenation
- Alkyl/acyclic hydroxylation
- Aromatic hydroxylation
- Deamination
- Desulfuration
- N-dealkylation
- N-hydroxylation
- N-oxidation
- O-dealkylation
- Sulphoxidation

Reduction

Reduction is a chemical reaction in which the *substrate gains electrons*. Reductions are most likely to occur with xenobiotics in which oxygen content is low. Reductions can occur across nitrogen-nitrogen double bonds (azo reduction) or on nitro groups (NO2). Frequently, the resulting amino compounds are oxidized which forms toxic metabolites. Some chemicals such as carbon tetrachloride can be reduced to free radicals, which are quite reactive with biological tissues. Thus, reduction reactions frequently result in activation of a xenobiotic rather than detoxification. An example of a reduction reaction in which the nitro group is reduced is illustrated in Figure 8.4.6.12.2.7.

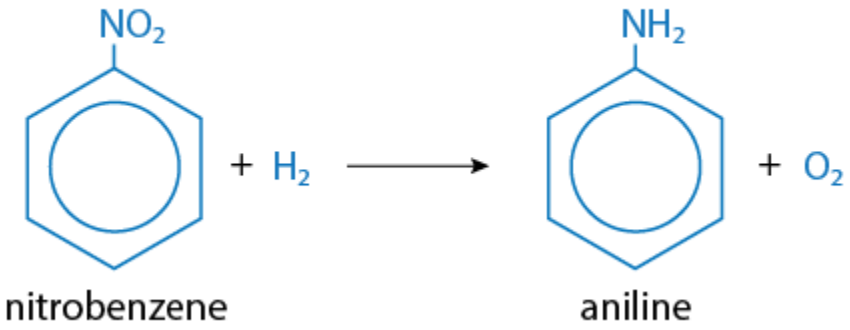


Figure above

There are fewer specific reduction reactions than oxidizing reactions. The nature of these reactions is also described by their name. Some reducing reactions include:

- Azo reduction
- Dehalogenation
- Disulfide reduction
- Nitro reduction
- N-oxide reduction
- Sulfoxide reduction

Hydrolysis

Hydrolysis is a chemical reaction in which the *addition of water* splits the toxicant into two fragments or smaller molecules. The hydroxyl group (OH-) is incorporated into one fragment and the hydrogen atom is incorporated into the other. Larger chemicals such as esters, amines, hydrazines, and carbamates are generally biotransformed by hydrolysis.
An example of hydrolysis is illustrated in the biotransformation of procaine (local anesthetic) which is hydrolyzed to two smaller chemicals (Figure 8.4.6.12.2.8).

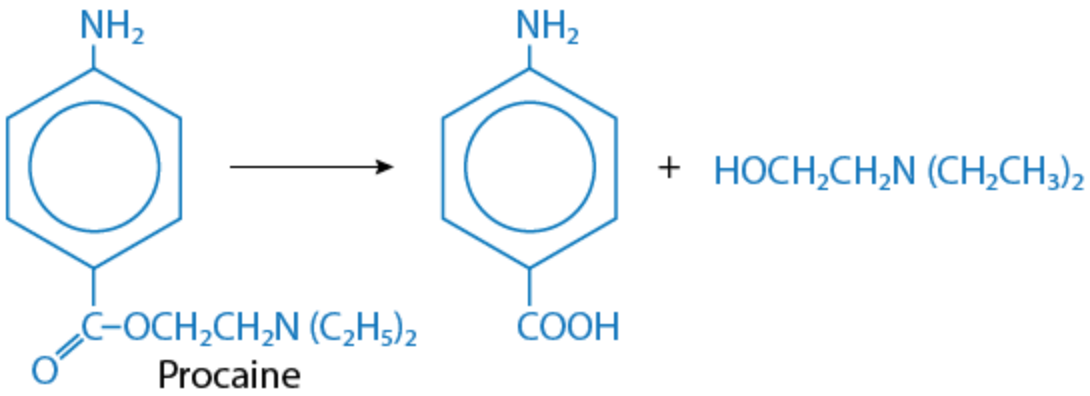


Figure 8.4.6.12.2.8. Hydrolysis of procaine
(Image Source: Adapted from Humboldt State University, Department of Chemistry. Author: Richard A. Paselk, Professor Emeritus. View original image.)

Toxicants that have undergone Phase I biotransformation are converted to metabolites that are sufficiently ionized, or hydrophilic, to be either eliminated from the body without further biotransformation or converted to an intermediate metabolite that is ready for Phase II biotransformation. The intermediates from Phase I transformations may be pharmacologically more effective and in many cases more toxic than the parent xenobiotic.

Phase II Reactions

A xenobiotic that has undergone a Phase I reaction is now a new intermediate metabolite that contains a reactive chemical group such as hydroxyl (-OH), amino (-NH2), and carboxyl (-COOH). Many of these intermediate metabolites do not possess sufficient hydrophilicity to permit elimination from the body. These metabolites must undergo additional biotransformation as a Phase II reaction.
Phase II reactions are conjugation reactions where a molecule normally present in the body is added to the reactive site of the Phase I metabolite. The result is a conjugated metabolite that is more water soluble than the original xenobiotic or Phase I metabolite. Usually, the Phase II metabolite is quite hydrophilic and can be readily eliminated from the body. The primary Phase II reactions are:

- Glucuronide conjugation – most important reaction (detailed below)
- Sulfate conjugation – important reaction (detailed below)
- Acetylation
- Amino acid conjugation
- Glutathione conjugation
- Methylation

Glucuronide Conjugation

Glucuronide conjugation is one of the most important and common Phase II reactions. The glucuronic acid molecule is used in this reaction. It is derived from glucose, a common carbohydrate (sugar) that is the primary source of energy for cells. In this reaction, glucuronic acid is added directly to the toxicant or its phase I metabolite. The sites of glucuronidation reactions are substrates having an oxygen, nitrogen, or sulfur bond, which apply to a wide array of xenobiotics as well as endogenous substances, such as bilirubin, steroid hormones, and thyroid hormones.

Glucuronidation is a pathway that conjugates xenobiotics at a high capacity ("high-capacity pathway"). Glucuronide conjugation usually decreases toxicity although there are some notable exceptions, for example, where it can result in producing carcinogenic substances. The glucuronide conjugates are generally quite hydrophilic and are excreted by the kidney or bile, depending on the size of the conjugate. The glucuronide conjugation of aniline is illustrated in Figure 8.4.6.12.2.9.

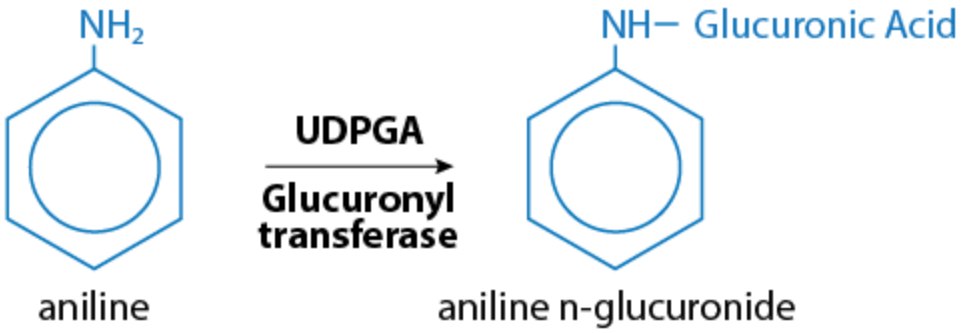


Figure 8.4.6.12.2.9. Glucuronide conjugation of aniline (which is used to make polyurethane, pharmaceuticals, and industrial chemicals)
(Image Source: NLM)

Sulfate Conjugation

Sulfate conjugation is another important Phase II reaction that occurs with many xenobiotics. In general, sulfation decreases the toxicity of xenobiotics. Unlike glucuronic acid conjugates that are often eliminated in the bile, the highly polar sulfate conjugates are readily secreted in the urine. In general, sulfation is a low-capacity pathway for xenobiotic conjugation. Often glucuronidation or sulfation can conjugate the same xenobiotics.

Knowledge Check

1) The substances in the body that accelerate chemical reactions are known as:

a) Amino acids
b) Enzymes
c) Substrates

Answer available at the end of this section.

2) The convention used to name specific enzymes consists of combining:

a) The substrate name with the type of chemical reaction
b) The target organ and the type of chemical reaction
c) The substrate name with the form of toxicity

Answer available at the end of this section.

3) Biotransformation reactions are classified as Phase I and Phase II. The basic difference is:

a) Phase I reactions conjugate a substrate whereas Phase II reactions oxidize the substance
b) Phase I reactions generally add a functional structure whereas Phase II reactions conjugate the substance
c) A Phase I reaction generally makes a substance more hydrophilic than a Phase II reaction

Answer available at the end of this section.

4) The difference between oxidation and reduction reactions is:

a) A substrate gains electrons from an oxidation reaction whereas it loses electrons by a reduction reaction
b) Oxygen is removed from a substrate in oxidation and added in the reduction reaction
c) A substrate losses electrons from an oxidation reaction whereas it gains electrons by a reduction reaction

Answer available at the end of this section.

5) Which conjugation reaction is the most common in the biotransformation of xenobiotics?

a) Amino acid conjugation
b) Glucuronide conjugation
c) Methylation

Answer available at the end of this section.

Answers

Knowledge Check

Enzymes - **This is the correct answer.**
Enzymes are proteins that catalyze nearly all biochemical reactions in the body.

Knowledge Check

The substrate name with the type of chemical reaction - **This is the correct answer.**
Enzyme names end in "ase" and usually combine the substrate acted on and the type of reaction catalyzed.

Knowledge Check

Phase I reactions generally add a functional structure whereas Phase II reactions conjugate the substance - **This is the correct answer.**
Phase I reactions are generally reactions which modify the chemical by adding a functional structure. This allows the substance to "fit" into the Phase II enzyme so that it can become conjugated (joined together) with another substance. Phase II reactions consist of those enzymatic reactions that conjugate the modified xenobiotic with another substance.

Knowledge Check

A substrate losses electrons from an oxidation reaction whereas it gains electrons by a reduction reaction - **This is the correct answer.**
Oxidation is a chemical reaction in which a substrate loses electrons. Reduction is a chemical reaction in which the substrate gains electrons.

Knowledge Check

Glucuronide conjugation - **This is the correct answer.**
Glucuronide conjugation is one of the most important and common Phase II reactions. Glucuronidation is a high-capacity pathway for xenobiotic conjugation.

This page titled 8.4.6.12.2: Chemical Reactions is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 12.2: Chemical Reactions by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.12.3: Biotransformation Sites

Biotransformation Sites

Biotransforming enzymes are widely distributed throughout the body.

- The liver is the primary biotransforming organ due to its large size and high concentration of biotransforming enzymes.
- The kidneys and lungs are next with 10-30% of the liver's capacity.
- A low capacity exists in the skin, intestines, testes, and placenta.

Primary Biotransformation Site: The Liver

Since the liver is the primary site for biotransformation, it is also potentially vulnerable to the toxic action of a xenobiotic that is activated to a more toxic compound.
Within the liver cell, the primary subcellular components containing the transforming enzymes are the **microsomes (small vesicles) of the endoplasmic reticulum** and the soluble fraction of the **cytoplasm (cytosol)**. The mitochondria, nuclei, and lysosomes contain a small level of transforming activity.

- Microsomal enzymes are associated with most Phase I reactions. Glucuronidation enzymes are also contained in microsomes.
- Cytosolic enzymes are non-membrane-bound and occur free within the cytoplasm. They are generally associated with Phase II reactions, although some oxidation and reduction enzymes are contained in the cytosol.
- The most important enzyme system involved in Phase I reactions are the **cytochromes P450**, also called the cytochrome P-450 system or the mixed function oxidase (MFO) system, but now mostly called CYP450 or CYPs by scientists and in research publications. It is found in microsomes and is responsible for oxidation reactions of a wide array of chemicals.

Susceptibility of the Liver

The liver is particularly susceptible to damage by ingested toxicants because it biotransforms most xenobiotics and receives blood directly from the gastrointestinal tract. Blood leaving the gastrointestinal tract does not flow directly into the general circulatory system. Instead, it flows into the liver first via the portal vein. This process is known as the "first pass." Blood leaving the liver is eventually distributed to all other areas of the body; however, much of the absorbed xenobiotic has undergone detoxification or bioactivation. The liver may have removed most of the potentially toxic chemical. On the other hand, some toxic metabolites are highly concentrated in the liver.

Knowledge Check

1) The organ that has the greatest ability to biotransform xenobiotics is the:

a) Liver
b) Pancreas
c) Skin

Answer available at the end of this section.

2) The "first pass" phenomenon pertains to:

a) The situation where xenobiotics that are absorbed from the GI tract first enter the circulating blood before going to the liver

b) A condition where the liver first biotransforms a xenobiotic by Phase II reaction before it is biotransformed by a Phase I reaction
c) An anatomical arrangement in which xenobiotics absorbed from the intestine go to the liver first rather than into the systemic circulation

Answer available at the end of this section.

Answers

Knowledge Check

Liver - **This is the correct answer.**
Biotransforming enzymes are widely distributed throughout the body. However, the liver has the largest concentration of all organs and thus has a very high capacity for biotransformation.

Knowledge Check

An anatomical arrangement in which xenobiotics absorbed from the intestine go to the liver first rather than into the systemic circulation - **This is the correct answer.**
Blood leaving the gastrointestinal tract does not directly flow into the general circulatory system. Instead, it flows into the liver first via the portal vein. This is known as the "first pass" phenomena.

This page titled 8.4.6.12.3: Biotransformation Sites is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 12.3: Biotransformation Sites by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.12.4: Modifiers of Biotransformation

Modifiers of Biotransformation

The relative effectiveness of biotransformation depends on several factors that can inhibit or induce enzymes and dose levels. Factors include:

- Species
- Age
- Gender
- Genetic variability
- Nutrition
- Disease
- Exposure to other chemicals

Species

It is well known that the capability to biotransform specific chemicals varies by **species**. These differences are termed selective toxicity, which refers to differences in toxicity between species similarly exposed. Research uses what is known about selective toxicity to develop chemicals that are effective but relatively safe in humans.

- For example, the pesticide malathion in mammals is biotransformed by hydrolysis to relatively safe metabolites, but in insects, it is oxidized to malaoxon, which is lethal to insects.

Age and Gender

Age may affect the efficiency of biotransformation. In general, human fetuses and newborns have limited abilities to carry out xenobiotic biotransformations. This limitation is due to inherent deficiencies in many of the enzymes responsible for catalyzing Phase I and Phase II biotransformations. While the capacity for biotransformation fluctuates with age in adolescents, by early adulthood the enzyme activities have essentially stabilized. The aged also have decreased biotransformation capability.
Gender may influence the efficiency of biotransformation for specific xenobiotics. This is usually limited to hormone-related differences in the oxidizing cytochrome P-450 enzymes.

Genetic Variability

Genetic variability in biotransforming capability accounts for most of the large variation among humans. In particular, human genetic differences influence the Phase II acetylation reaction. Some persons have rapid acetylation ("rapid acetylator") while others have a slow ability to carry out this reaction ("slow acetylator"). The most serious drug-related toxicity occurs in those who have slow acetylators, often referred to as "slow metabolizers." With slow acetylators, acetylation is so slow that blood or tissue levels of certain drugs (or Phase I metabolites) exceed their toxic threshold.

Table 8.4.6.12.4.1 includes examples of drugs that build up to toxic levels in slow metabolizers who have specific genetic-related defects in biotransforming enzymes.

Drug	Drug Category	Metabolic Defect	Toxic Effect
Isoniazid	antituberculosis drug	slow acetylation	nerve and liver damage
Hydralazine	antihypertensive drug	defect in mono-oxygenase enzyme	excessive fall in blood pressure
Dapsone	antibacterial agent	slow acetylation	systemic lupus erythematosus
Primaquine	antimalarial agent	defective G6PD	acute hemolytic anemia

Table 1. Examples of drugs that build to toxic levels in slow metabolizers with specific genetic-related defacts in biotransforming enzymes

Nutrition

Poor **nutrition** can have a detrimental effect on biotransforming ability. Poor nutrition relates to inadequate levels of protein, vitamins, and essential minerals. These deficiencies can decrease a person's ability to synthesize biotransforming enzymes. Many diseases can impair an individual's capacity to biotransform xenobiotics.
For example, hepatitis (a liver disease) is well known to reduce hepatic biotransformation to less than half of its normal capacity.
Prior or Simultaneous ExposurePrior or simultaneous exposure to xenobiotics can cause enzyme inhibition and enzyme induction. In some situations, exposure to a substance will inhibit the biotransformation capacity for another chemical due to **inhibition of specific enzymes**. A major mechanism for the inhibition is competition between the two substances for the available oxidizing or conjugating enzymes. The presence of one substance uses up the enzyme needed to metabolize the second substance.
Exposure to Other Environmental Chemicals and DrugsEnzyme induction is a situation where prior exposure to certain environmental chemicals and drugs results in an enhanced capability for biotransforming a xenobiotic. The prior exposures stimulate the body to increase the production of some enzymes. This increased level of enzyme activity results in increased biotransformation of a chemical subsequently absorbed.
Examples of enzyme inducers include:

- Alcohol
- Isoniazid
- Polycyclic halogenated aromatic hydrocarbons (for example, dioxin)
- Phenobarbital
- Cigarette smoke

The most commonly induced enzyme reactions involve the cytochrome P450 enzymes.
Dose level can affect the nature of the biotransformation. In certain situations, the biotransformation may be quite different at high doses compared to low dose levels. This difference in biotransformation contributes to a dose threshold for toxicity. The existence of different biotransformation pathways can usually explain what causes this dose-related difference in biotransformation. At low doses, a xenobiotic may follow a biotransformation pathway that detoxifies the substance. However, if the amount of xenobiotic exceeds the specific enzyme capacity, the biotransformation pathway is saturated. In that case, it is possible that the level of parent toxin builds up. In other cases, the xenobiotic may enter a different biotransformation pathway that may end up producing a toxic metabolite.
An example of a dose-related difference in biotransformation occurs with acetaminophen (Tylenol®):

- At normal doses:
  - About 96% of acetaminophen is biotransformed to non-toxic metabolites by sulfate and glucuronide conjugation.
  - About 4% of the acetaminophen oxidizes to a toxic metabolite.
  - That toxic metabolite is conjugated with glutathione and excreted.
- At 7-10 times the recommended therapeutic level:
  - The sulfate and glucuronide conjugation pathways become saturated and more of the toxic metabolite is formed.
  - The glutathione in the liver may also be depleted so that the toxic metabolite is not detoxified and eliminated.

- It can react with liver proteins and cause fatal liver damage.

Knowledge Check

1) Selective toxicity refers to a difference in the toxicity of a xenobiotic to different species. This selective toxicity can usually be attributed to differences in:

a) The ability to absorb the xenobiotic
b) Organ systems between species
c) Capability to biotransform the xenobiotic

Answer available at the end of this section.

Answers

Knowledge Check

Capability to biotransform the xenobiotic - **This is the correct answer.**
A difference between species in their capability to biotransform a specific chemical is normally the basis for a chemical's selective toxicity.

This page titled 8.4.6.12.4: Modifiers of Biotransformation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 12.4: Modifiers of Biotransformation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.13: EXCRETION

Learning Objectives

After completing this lesson, you will be able to:

- Define excretion.
- Identify the primary organ systems involved in excretion.
- Describe the basic mechanisms of excretion within each primary organ system involved.

In this section...

Topics include:

8.4.6.13.1: Introduction to Secretion

8.4.6.13.2: Urinary Excretion

8.4.6.13.3: Fecal Excretion

8.4.6.13.4: Exhaled Air

8.4.6.13.5: Other Routes

Section 13: Key Points

What We've Covered

This section made the following main points:

- Excretion, as used in ToxTutor, pertains to the elimination of a xenobiotic and its metabolites by specific excretory organs.
- The primary organ systems involved in excretion are the:
  - Urinary system, which involves:
    1. Filtration in the glomerulus.
    2. Secretion in the proximal tubule section of the nephron to transport certain molecules out of the blood and into the urine.
    3. Reabsorption in the proximal convoluted tubule of the nephron to reenter nearly all of the water, glucose, potassium, and amino acids lost during filtration back into the blood.
  - Gastrointestinal system, which occurs from two processes:
    1. Biliary excretion — generally active secretion by the liver into the bile and then into the intestinal tract, where it can be eliminated in the feces or reabsorbed.
    2. Intestinal excretion — an important elimination route only for xenobiotics that have slow biotransformation or slow urinary or biliary excretion.
  - Respiratory system, which is important for xenobiotics and metabolites that exist in a gaseous phase in the blood:
    - Excreted by passive diffusion from the blood into the alveolus.
- Minor routes of excretion occur including breast milk, sweat, saliva, tears, and semen.

This page titled 8.4.6.13: Excretion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 13: Excretion by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.13.1: Introduction to Secretion

Introduction to Excretion

Elimination from the body is very important in determining the potential toxicity of a xenobiotic. When the body rapidly eliminates a toxic xenobiotic (or its metabolites), it is less likely that they will be able to concentrate in and damage critical cells. The terms excretion and elimination are frequently used to describe the same process in which a substance leaves the body. **Elimination** is sometimes used in a broader sense and includes the removal of the absorbed xenobiotic through metabolic pathways as well as through excretion. **Excretion**, as used here, pertains to the elimination of the xenobiotic and its metabolites by specific excretory organs.

Except for the lung, polar (hydrophilic) substances are more likely than lipid-soluble toxicants to be eliminated from the body. Chemicals must again pass through membranes in order to leave the body, and the same chemical and physical properties that governed passage across other membranes apply to excretory organs as well.

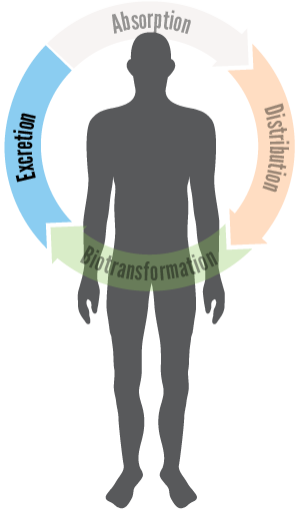


Figure 8.4.6.13.1.1. Processes of toxicokinetics (Image Source: Adapted from iStock Photos, ©)

Primary Routes of Excretion

The body uses several routes to eliminate toxicants or their metabolites. The main routes of excretion are via urine, feces, and exhaled air. Thus, the primary organ systems involved in excretion are the:

- Urinary system
- Gastrointestinal system
- Respiratory system

A few other avenues for elimination exist but they are relatively unimportant, except in exceptional circumstances.

Knowledge Check

1) The three major routes of excretion are:

a) Gastrointestinal tract, sweat, and saliva
b) Mother's milk, tears, and semen
c) Urinary excretion, fecal excretion, and exhaled air

Answer available at the end of this section.

Answers

Knowledge Check

Urinary excretion, fecal excretion, and exhaled air - **This is the correct answer.**
The main routes of excretion are via urine, feces, and exhaled air.

This page titled 8.4.6.13.1: Introduction to Secretion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 13.1: Introduction to Secretion by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.13.2: Urinary Excretion

Urinary Excretion

The primary route in which the body eliminates substances is through the kidneys. The main function of the kidney is the excretion of body wastes and harmful chemicals into the urine. The functional unit of the kidney responsible for excretion is the nephron. Each kidney contains about one million nephrons. The nephron has three primary regions that function in the renal excretion process: the glomerulus, proximal tubule, and the distal tubule (Figure 8.4.6.13.2.2).

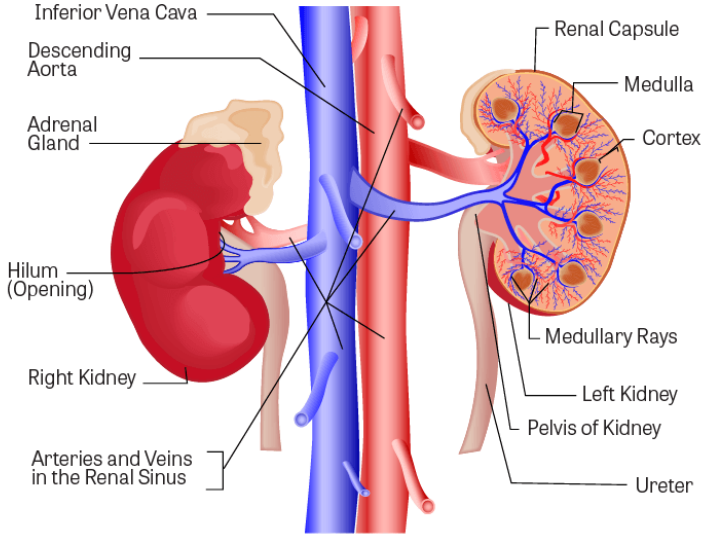
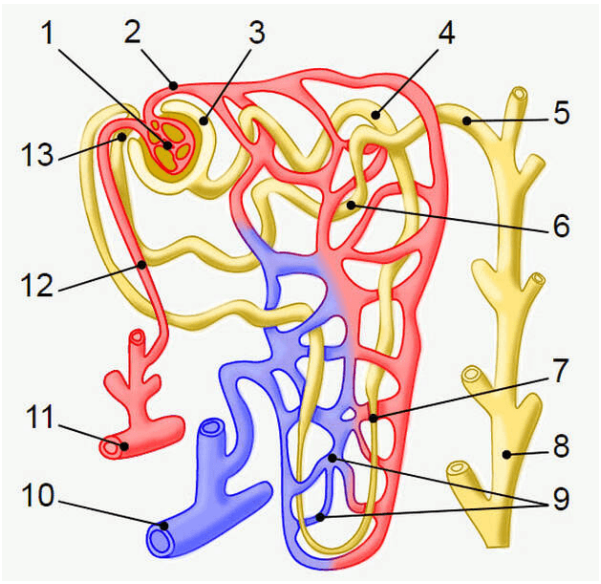


Figure 8.4.6.13.2.1. Components of the urinary system
(Image Source: Adapted from iStock Photos, ©)

Three processes are involved in urinary excretion:

1. Filtration
2. Secretion
3. Reabsorption



Legend:

1. Glomerulus
2. Efferent arteriole
3. Bowman's capsule
4. Proximal convoluted tubule
5. Cortical collecting duct
6. Distal convoluted tubule
7. Loop of Henle
8. Papillary duct
9. Peritubular capillaries
10. Arcuate vein
11. Arcuate artery
12. Afferent arteriole
13. Juxtaglomerular apparatus

Figure 8.4.6.13.2.2. Nephron of the kidney
(Image Source: Adapted from Wikimedia Commons, obtained by Public Domain, Creative Commons CC0 1.0 Universal Public Domain Dedication.
View original image.)

Filtration

Filtration takes place in the glomerulus, which is the vascular beginning of the nephron. Approximately one-fourth of the blood flow from cardiac output circulates through the kidney, the greatest rate of blood flow for any organ. A considerable amount of the blood plasma filters through the glomerulus into the nephron tubule. This results from the large amount of blood flow through the glomerulus, the large pores (40 Angstrom [Å]) in the glomerular capillaries, and the hydrostatic pressure of the blood. Small molecules, including water, readily pass through the sieve-like filter into the nephron tubule. Both lipid soluble and polar substances will pass through the glomerulus into the tubule filtrate. The amount of filtrate is very large, about 45 gallons per day in an adult human. About 99% of the water-like filtrate, small molecules, and lipid-soluble substances, are reabsorbed downstream in the nephron tubule. This means that the amount of urine eliminated is only about one percent of the amount of fluid filtrated through the glomeruli into the renal tubules.
Molecules with molecular weights greater than 60,000 (which include large protein molecules and blood cells) cannot pass through the capillary pores and remain in the blood. If urine contains albumin or blood cells, it indicates that the glomeruli have been damaged. Binding to plasma proteins will influence urinary excretion. Polar substances usually do not bind with the plasma proteins and thus can be filtered out of the blood into the tubule filtrate. In contrast, substances extensively bound to plasma proteins remain in the blood.

Secretion

Secretion, which occurs in the proximal tubule section of the nephron, is responsible for the transport of certain molecules out of the blood and into the urine. Secreted substances include potassium ions, hydrogen ions, and some xenobiotics. Secretion occurs by active transport mechanisms that are capable of differentiating among compounds based on polarity. Two systems exist, one that transports **weak acids** (such as many conjugated drugs and penicillins) and the other that transports **basic substances** (such as histamine and choline).

Reabsorption

Reabsorption takes place mainly in the proximal convoluted tubule of the nephron. Nearly all of the water, glucose, potassium,

and amino acids lost during glomerular filtration reenter the blood from the renal tubules. Reabsorption occurs primarily by passive transfer based on a concentration gradient, moving from a high concentration in the proximal tubule to the lower concentration in the capillaries surrounding the tubule (Figures 4-6).

A factor that greatly affects reabsorption and urinary excretion is the pH of the urine. This is especially the case with weak electrolytes. If the urine is alkaline, weak acids are more ionized and excretion is increased. Weak acids (such as glucuronide and sulfate conjugates) are less ionized if the urine is acidic and undergo reabsorption and renal excretion is reduced. Since the urinary pH varies in humans, the urinary excretion rates of weak electrolytes also vary.

- Examples are phenobarbital (an acidic drug) which is ionized in alkaline urine and amphetamine (a basic drug) which is ionized in acidic urine. Treatment of barbiturate poisoning (such as an overdose of phenobarbital) may include changing the pH of the urine to facilitate excretion.
- Diet may have an influence on urinary pH and thus the elimination of some toxicants. For example, a high-protein diet results in acidic urine.

The physical properties (primarily molecular size) and polarity of a substance in the urinary filtrate greatly affect its ultimate elimination by the kidney. Small toxicants (both polar and lipid-soluble) are filtered with ease by the glomerulus. In some cases, large molecules (including some that are protein-bound) may be secreted (by passive transfer) from the blood across capillary endothelial cells and nephron tubule membranes to enter the urine. The major difference in ultimate fate is governed by a substance's polarity. Those substances that are ionized remain in the urine and leave the body. Lipid-soluble toxicants can be reabsorbed and re-enter the blood circulation, which lengthens their half-life in the body and potential for toxicity.

Kidneys, which have been damaged by toxins, infectious diseases, or because of age, have diminished ability to eliminate toxicants thus making those individuals more susceptible to toxins that enter the body. The presence of albumin in the urine indicates that the glomerulus filtering system is damaged, letting large molecules pass through. The presence of glucose in the urine is an indication that tubular reabsorption has been impaired.

Knowledge Check

1) The reason that much of the blood plasma filters into the renal tubule is due to:

a) The large amount of blood, under relatively high pressure, that flows through kidney glomerulae whose capillaries have large pores
b) Its high lipid content
c) The high binding content of plasma

Answer available at the end of this section.

2) In which area of the nephron does active secretion take place?

a) The collecting duct of the nephron
b) The proximal tubule of the nephron
c) The glomerulus of the nephron

Answer available at the end of this section.

3) Most of the material filtered through the glomerulus is reabsorbed in the proximal convoluted tubule of the nephron. The primary property of a xenobiotic that determines whether it will be reabsorbed is:

a) Protein binding
b) Molecular size
c) Polarity

Answer available at the end of this section.

Answers

Knowledge Check

The large amount of blood, under relatively high pressure, that flows through kidney glomerulae whose capillaries have large pores - **This is the correct answer.**
A considerable amount of the blood plasma filters through the glomerulus into the nephron tubule. This results from the large amount of blood flow through the glomerulus, the large pores (40 Angstrom [Å]) in the glomerular capillaries, and the hydrostatic pressure of the blood.

Knowledge Check

The proximal tubule of the nephron - **This is the correct answer.**
Secretion occurs in the proximal tubule section of the nephron and is responsible for the transport of certain molecules out of the blood and into the urine.

Knowledge Check

Polarity - **This is the correct answer.**
The ultimate fate of a substance filtered into the renal tubule is governed by its polarity. Those substances that are ionized remain in the urine and leave the body.

This page titled 8.4.6.13.2: Urinary Excretion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 13.2: Urinary Excretion by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.13.3: Fecal Excretion

Fecal Excretion

Elimination of toxicants in the feces occurs from two processes:

1. Excretion in bile, which then enters the intestine ("biliary excretion").
2. Direct excretion into the lumen of the gastrointestinal tract ("intestinal excretion").

Biliary Excretion

The **biliary route** is an important mechanism for fecal excretion of xenobiotics and is even more important for the excretion of their metabolites. This route generally involves active secretion rather than passive diffusion. Specific transport systems appear to exist for certain types of substances, for example, organic bases, organic acids, and neutral substances. Some heavy metals are excreted in the bile, for example, arsenic, lead, and mercury. However, the most likely substances to be excreted via the bile are comparatively large, ionized molecules, such as those having a large molecular weight (conjugates greater than 300).

Once a substance has been excreted by the liver into the bile, and then into the intestinal tract, it can be eliminated from the body in the feces, or it may be reabsorbed. Since most of the substances excreted in the bile are water soluble, they are not likely to be reabsorbed as such. However, enzymes in the intestinal flora are capable of hydrolyzing some glucuronide and sulfate conjugates, which can release the less polar compounds that may then be reabsorbed. This process of excretion into the intestinal tract via the bile and reabsorption and return to the liver by the portal circulation is known as the **enterohepatic circulation** (Figure 1).

Enterohepatic circulation prolongs the life of the xenobiotic in the body. In some cases, the metabolite is more toxic than the excreted conjugate. Continuous enterohepatic recycling can occur and lead to very long half-lives of some substances. For this reason, drugs may be given orally to bind substances excreted in the bile.

- For example, a resin can be taken orally to bind with dimethylmercury, which had been secreted in the bile. The binding of the resin to dimethylmercury prevents its reabsorption and further toxicity.

Changes in the production and flow of bile into the liver affect the efficiency of biliary excretion.

- Liver disease usually causes a decrease in bile flow.
- Some drugs such as phenobarbital can produce an increase in bile flow rate. Administration of phenobarbital has been shown to enhance the excretion of methylmercury by this mechanism.

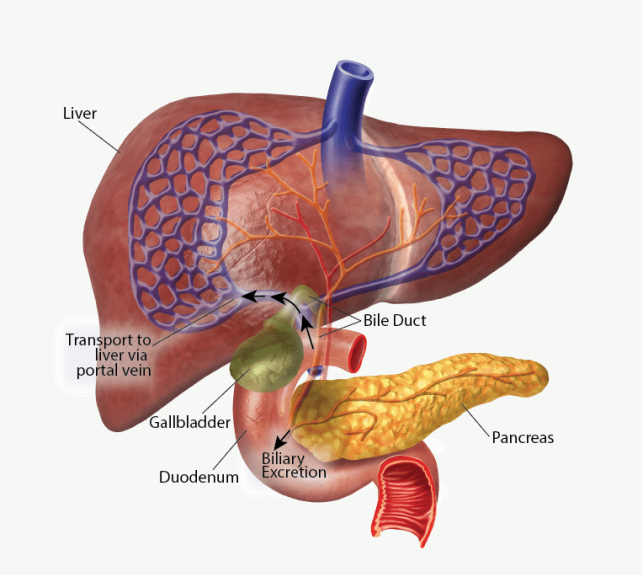


Figure 8.4.6.13.3.1. Biliary excretion and enterohepatic circulation
(Image Source: Adapted from iStock Photos, ©)

Intestinal Excretion

Another way that xenobiotics can be eliminated via the feces is by direct **intestinal excretion**. While this is not a major route of elimination, a large number of substances can be excreted into the intestinal tract and eliminated via feces. Some substances, especially those that are poorly ionized in plasma (such as weak bases), may passively diffuse through the walls of the capillaries, through the intestinal submucosa, and into the intestinal lumen to be eliminated in feces.

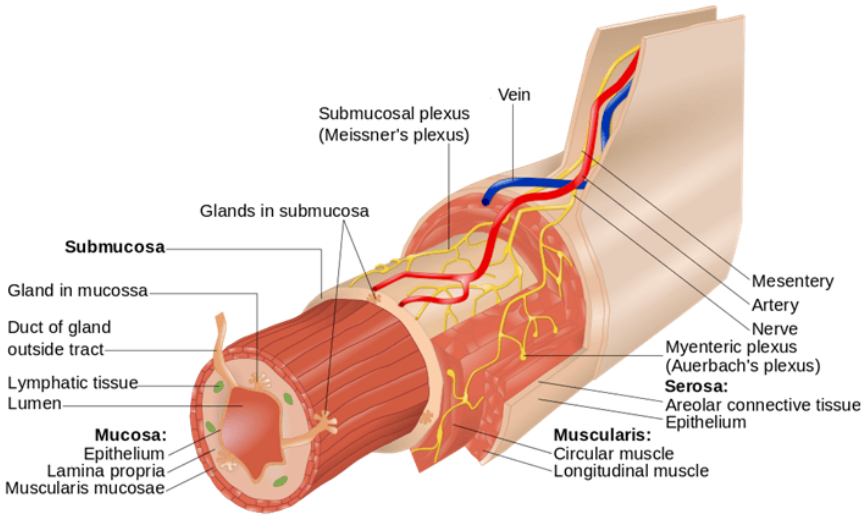


Figure 8.4.6.13.3.2. Layers of the Alimentary Canal
(Image Source: Wikimedia Commons, obtained under Creative Commons Attribution-Share Alike 3.0 Unported license. Author: Goran tek-en.
View original image.)

Intestinal excretion is a relatively slow process and therefore, it is an important elimination route only for those xenobiotics that have slow biotransformation, or slow urinary or biliary excretion. Increasing the lipid content of the intestinal tract can enhance intestinal excretion of some lipophilic substances. For this reason, mineral oil (liquid paraffin, derived from petroleum) is sometimes added to the diet to help eliminate toxic substances, which are known to be excreted directly into the intestinal tract.

Knowledge Check

1) Substances excreted in the bile are primarily:

a) Small, lipid soluble molecules
b) Comparatively large, ionized molecules
c) Large, lipid soluble molecules

Answer available at the end of this section.

2) Many substances excreted in bile undergo enterohepatic circulation, which involves:

a) Excretion of substances into the circulating system rather than into the intestine
b) Excretion into the intestinal tract and reabsorption and return to the liver by the portal circulation
c) The recycling of xenobiotics between the liver and gall bladder

Answer available at the end of this section.

Answers

Knowledge Check

Comparatively large, ionized molecules - **This is the correct answer.**
The most likely substances to be excreted via the bile are comparatively large, ionized molecules, such as large molecular weight (greater than 300) conjugates.

Knowledge Check

Excretion into the intestinal tract and reabsorption and return to the liver by the portal circulation - **This is the correct answer.**
The process of excretion into the intestinal tract via the bile and reabsorption and return to the liver by the portal circulation is known as the enterohepatic circulation. The effect of this enterohepatic circulation is to prolong the life of the xenobiotic in the body.

This page titled 8.4.6.13.3: Fecal Excretion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 13.3: Fecal Excretion by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.13.4: Exhaled Air

Exhaled Air

The lungs are an important route of excretion for xenobiotics (and metabolites) that exist in a gaseous phase in the blood.

Passive Diffusion

Blood gases are excreted by passive diffusion from the blood into the alveolus, following a concentration gradient. This type of excretion occurs when the concentration of the xenobiotic dissolved in capillary blood is greater than the concentration of the substance in the alveolar air. Gases with a low solubility in blood are more rapidly eliminated than those gases with a high solubility. Volatile liquids dissolved in the blood are also readily excreted via the expired air.

For example, breathalyzer devices can measure blood alcohol concentration because as alcohol in the blood moves across the alveoli the alcohol in the blood evaporates and is exhaled. The concentration of alcohol in the exhaled air relates to the level of alcohol in the blood.

Impact of Vapor Pressure

The amount of a liquid excreted by the lungs is proportional to its vapor pressure. Exhalation is an exception to most other routes of excretion in that it can be a very efficient route of excretion for lipid soluble substances. This is due to the very close proximity of capillary and alveolar membranes, which are thin and allow for the normal gaseous exchange that occurs in breathing.

Knowledge Check

1) Xenobiotics are eliminated in exhaled air by:

a) Passive diffusion
b) Active transport
c) Facilitated transport

Answer available at the end of this section.

Answers

Knowledge Check

Passive diffusion - **This is the correct answer.**
Blood gases are excreted by passive diffusion from the blood into the alveolus, following a concentration gradient. This occurs when the concentration of the xenobiotic dissolved in capillary blood is greater than the concentration of the substance in the alveolar air.

This page titled 8.4.6.13.4: Exhaled Air is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 13.4: Exhaled Air by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.13.5: Other Routes

Other Routes of Excretion

Several minor routes of excretion occur including mother's milk, sweat, saliva, tears, and semen.

Excretion into Breast Milk

Excretion into milk can be important since toxicants can be passed with milk to the nursing offspring. In addition, toxic substances can pass from cow's milk to people. Toxic substances are excreted into milk by simple diffusion. Both basic substances and lipid soluble compounds can be **excreted into milk (**The National Library of Medicine's **LactMed** is a resource for information on drugs, dietary supplements, and herbs that pass into breast milk.).

Basic substances can be concentrated in milk since milk is more acidic (pH approximately 6.5) than blood plasma. Since milk contains 3–4% lipids, lipid soluble xenobiotics can diffuse along with fats from plasma into the mammary gland and thus can be present in mother's milk. Substances such as lead, mercury, Bisphenol A (BPA), and phthalates that are chemically similar to calcium can also be excreted into milk along with calcium.

Did you know?

Volatile organic compounds (VOCs) found in indoor air can also be found in breast milk.

Examples include MTBE (methyl tert-butyl ether), chloroform, benzene, and toluene. For benzene, toluene, and MTBE, the levels in breast milk followed the indoor air concentrations. However, the infant average daily dose by inhalation exceeded ingestion rates by 25-to-135 fold. Thus, the amount of VOC exposure from indoor air in nonsmoking households is much greater than the VOC exposure from breast milk. Strategies to lessen infant VOC exposure should focus on improving indoor air quality.

Excretion into All Other Body Secretions or Tissues

Excretion of xenobiotics in **all other body secretions or tissues** (including the saliva, sweat, tears, hair, and skin) are of only minor importance. Under conditions of great sweat production, excretion in sweat may reach a significant degree. Some metals, including cadmium, copper, iron, lead, nickel, and zinc, may be eliminated in sweat to some extent. Xenobiotics that passively diffuse into saliva may be swallowed and absorbed by the gastrointestinal system. The excretion of some substances into saliva is responsible for the unpleasant taste that sometimes occurs with time after exposure to a substance.

Knowledge Check

1) The following are minor routes of excretion:

a) Sweat and saliva
b) Urinary excretion, fecal excretion, and exhaled air

Answer available at the end of this section.

Answers

Knowledge Check

Sweat and saliva - **This is the correct answer.**
Several minor routes of excretion exist, primarily via mother's milk, sweat, saliva, tears, and semen. The main routes of excretion are via urine, feces, and exhaled air.

This page titled 8.4.6.13.5: Other Routes is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 13.5: Other Routes by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.14: CELLULAR TOXICOLOGY

Learning Objectives

After completing this lesson, you will be able to:

- Explain cellular adaptation.
- Identify four possible endpoints to toxic damage to cells and tissues.
- Define commonly used cancer terms.
- Describe the phases of and genetic activity associated with carcinogenesis.
- Identify mechanisms and potential outcomes of neurotoxicity.

In this section...

Topics include:

8.4.6.14.1: Adaptation

8.4.6.14.2: Cell Damage and Tissue Repair

8.4.6.14.3: Cancer

8.4.6.14.4: Neurotoxicity

Section 14: Key Points

What We've Covered

This section made the following main points:

- To maintain homeostasis, cells and tissues undergo:
  - Physiological adaptation, which is beneficial in nature — for example, increased skeletal muscle cells in athletes.
  - Pathological adaptation, which is detrimental — for example, cellular changes in people who smoke cigarettes.
- Specific types of adaptation include:
  - Atrophy — a decrease in the size of cells.
  - Hypertrophy — an increase in the size of individual cells.
  - Hyperplasia — an increase in the number of cells in a tissue.
  - Metaplasia — the conversion from one type of mature cell to another type.
  - Dysplasia — abnormal cell changes or deranged cell growth.
  - Anaplasia — cells that are undifferentiated.
  - Neoplasia — new growth of tissue.
- Most toxic effects, especially due to xenobiotics, are due to specific biochemical interactions without causing recognizable damage to a cell or its organelles. Cellular or biochemical toxicity leads to:
  - The tissue being completely repaired and returned to normal.
  - The tissue being incompletely repaired but capable of functioning with reduced capacity.
  - Death of the organism or complete loss of a tissue or organ.
  - Neoplasm or cancers.

- Tumors are either:
  - Benign — similar to the cell of origin, slow-growing, and usually without systemic effects.
  - Malignant — dissimilar from the cell of origin, rapid-growing, and commonly with systemic effects and life-threatening. Most malignant tumors are either:
    - Carcinomas — arising in epithelium, the most common form of cancer, usually spread in the lymphatic system.
    - Sarcomas — arising in connective or muscle tissue, usually spread by the blood stream.
- Carcinogenesis is a multi-step, multi-factorial genetic disease consisting of at least three main phases:
  1. Initiation — irreversible alteration of the DNA (mutation) of a normal cell.
  2. Promotion/Conversion — promoters enhance further development of the initiated cells, often influencing further expression of the mutated DNA such that the initiated cell proliferates and progresses further.
  3. Progression — development of the initiated cell into a biologically malignant cell population, often with metastasis to other areas of the body.
- Regulatory genes control the activity of structural genes and direct the proliferation process of the cell. Regulatory genes that play roles in carcinogenesis include:
  - Proto-oncogenes — normal cellular genes that encode and instruct the production of regulatory proteins and growth factors within a cell or its membrane.
  - Oncogenes — altered or misdirected proto-oncogenes with the ability to direct the production of proteins within the cell that change or transform the normal cell into a neoplastic cell.
  - Tumor suppressor genes (anti-oncogenes) — present in normal cells and counteract and change the proto-oncogenes and altered proteins, preventing a cell with damaged DNA from proliferating and evolving into an uncontrolled growth.
    - The p53 gene normally halts cell division, stimulates repair enzymes, and if necessary, commands the mutated cell to self-destruct
    - p53 is the most frequently altered in human tumors and is incapable of its defense mechanisms
- Toxic damage to the nervous system is divided into three categories:
  1. Damage to sensory receptors and sensory neurons impacting the sensory functions.
  2. Damage to motor neurons causing muscular weakness and paralysis.
  3. Interneuronal damage causing learning deficiencies, memory loss, incoordination, and emotional conditions.

This page titled 8.4.6.14: Cellular Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 14: Cellular Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.14.1: Adaptation

This section discusses cellular effects yet cell and chemical effects cannot be conveniently separated because cells are constructed of a variety of chemicals of diverse types. Specific intracellular chemical changes may occur as changes in the cell and may affect either its appearance or function. The actual mechanisms leading to cell damage are usually biochemical in nature.

Adaptation Explained

To maintain homeostasis, cells and tissues:

- "Cope" with new demands placed on them by constantly adapting to changes in the tissue environment.
- Are usually capable of an amazing degree of cellular adaptability.
- Adapt in a way that may be beneficial in nature (physiological) or detrimental (pathological).

Examples of **physiological adaptation** are:

- An increase in skeletal muscle cells in athletes due to exercise and increased metabolic demand.
- The increase in number and size of epithelial cells in breasts of women resulting from endocrine stimulation during pregnancy.

When these cells or tissues are damaged, the body attempts to adapt and repair or limit the harmful effects. Often the adaptive changes result in cells or organs that cannot function normally. This imperfect adaptation is a pathological change.
Examples of **pathological adaptations** are:

- Cellular changes in people who smoke cigarettes: The ciliated columnar epithelium changes to non-ciliated squamous epithelium in the trachea and bronchi of cigarette smokers. The replacement of squamous epithelium can better withstand the irritation of the cigarette smoke. However, the loss of cilia and mucous secretions of columnar epithelium diminish the tracheobronchial defense mechanisms.
- Replacement of normal liver cells by fibrotic cells in chronic alcoholics (known as cirrhosis of the liver): A severely cirrhotic liver is incapable of normal metabolism, maintenance of nutrition, and detoxification of xenobiotics.

If the change is minor, cellular adaptation may result and the cells return to normal. When damage is very severe, the result may be cell death or permanent functional incapacitation.
Cellular adaptation to toxic agents includes three basic types:

1. Increase in cell activity.
2. Decrease in cell activity.
3. Alteration in cell morphology (structure and appearance) or cell function.

Specific Types of Cellular Adaptations

Atrophy

Atrophy is a **decrease in the size of cells**. If a sufficient number of cells are involved, the tissue or organ may also decrease in size. When cells atrophy, they have:

- Reduced oxygen needs.
- Reduced protein synthesis.
- Decreased number and size of the organelles.

The most common causes of atrophy are reduced use of the cells, lack of hormonal or nerve stimulation, decrease in nutrition,

reduced blood flow to the tissue, and natural aging.

- An example of atrophy is the decrease in the size of muscles and muscle cells in persons whose legs are paralyzed, in a cast, or infrequently used as when a patient is on bedrest.

Hypertrophy

Hypertrophy is an **increase in the size of individual cells**. This frequently results in an increase in the size of a tissue or organ. When cells hypertrophy, components of the cell increase in numbers with increased functional capacity to meeting increased cell needs. Hypertrophy generally occurs in situations where the organ or tissue cannot adapt to an increased demand by formation of more cells. This is commonly seen in cardiac and skeletal muscle cells, which do not divide to form more cells. Common causes for hypertrophy are increased work or stress placed on an organ or hormonal stimulation.

- An example of hypertrophy is the compensatory increase in the size of cells in one kidney after the other kidney has been removed or is in a diseased state.

Hyperplasia

Hyperplasia is an **increase in the number of cells in a tissue**. This generally results in an enlargement of tissue mass and organ size. It occurs *only* in tissues capable of mitosis such as the epithelium of skin, intestine, and glands. Some cells do not divide and thus cannot undergo hyperplasia, for example, nerve and muscle cells. Hyperplasia is often a compensatory measure to meet an increase in body demands. Hyperplasia is a frequent response to toxic agents and damage to tissues such as wounds or trauma. In wound healing, hyperplasia of connective tissue (for example, fibroblasts and blood vessels) contributes to the wound repair. In many cases, when the toxic stress is removed, the tissue returns to normal. Hyperplasia may result from hormonal stimulation, for example, breast and uterine enlargement due to increased estrogen production during pregnancy.

Metaplasia

Metaplasia is the **conversion** from one type of mature cell to another type of mature cell. It is a cellular replacement process. A metaplastic response often occurs with chronic irritation and inflammation. This results in a tissue more resistant to the external stress since the replacement cells are capable of survival under circumstances in which the original cell type could not survive. However, the cellular changes usually result in a loss of function, which was performed by the original cells that were lost and replaced.
Examples of metaplasia are:

- The common condition in which a person suffers from chronic reflux of acid from the stomach into the esophagus (Gastroesophageal Reflux Disease). The normal esophageal cells (squamous epithelium) are sensitive to the refluxed acid and die. They are replaced with the columnar cells of the stomach that are resistant to the stomach's acidity. This pathological condition is known as "Barrett's Esophagus."
- The change in the cells of the trachea and bronchi of chronic cigarette smokers from ciliated columnar epithelium to non-ciliated stratified squamous epithelium. The sites of metaplasia frequently are also sites for neoplastic transformations. The replacement cells lack the defense mechanism performed by the cilia in moving particles up and out of the trachea.
- With cirrhosis of the liver, which is a common condition of chronic alcoholics, the normal functional hepatic cells are replaced by nonfunctional fibrous tissue.

Dysplasia

Dysplasia is a condition of **abnormal cell changes or deranged cell growth** in which the cells are structurally changed in size, shape, and appearance from the original cell type. Cellular organelles also become abnormal. A common feature of dysplastic cells is that the nuclei are larger than normal and the dysplastic cells have a mitotic rate higher than the predecessor normal cells. Causes of dysplasia include chronic irritation and infection. In many cases, the dysplasia can be reversed if the stress is removed and normal cells return. In other cases, dysplasia may be permanent or represent a precancerous change.

- An example of dysplasia is the atypical cervical cells that precede cervical cancer. Routine examination of cervical cells is a routine screening test for dysplasia and possible early stage cervical cancer (Papanicolaou test).

- Cancer occurs at the site of Barrett's syndrome and in the bronchi of chronic smokers (bronchogenic squamous cell carcinoma).

Anaplasia

Anaplasia refers to **cells that are undifferentiated**. They have irregular nuclei and cell structure with numerous mitotic figures. Anaplasia is frequently associated with malignancies and serves as one criterion for grading the aggressiveness of a cancer. For example, an anaplastic carcinoma is one in which the cell appearance has changed from the highly differentiated cell of origin to a cell type lacking the normal characteristics of the original cell. In general, anaplastic cells have lost the normal cellular controls, which regulate division and differentiation.

Neoplasia

Neoplasia is a **new growth of tissue** and is commonly referred to as a tumor. There are two types of neoplasia: benign and malignant. Malignant neoplasia are cancers. Since cancer is such an important and complex medical problem, a **separate section** is devoted to cancer.

Interactions

Interactions between two or more toxic agents can produce damage by chemical-chemical interactions, chemical-receptor interactions, or by modification, by a first agent, of the cell and tissue response to a second agent. Interactions may occur by simultaneous exposure and if exposure to the two agents is separated in time.

Chemical-chemical interactions have been mostly studied in the toxicology of air pollutants, where it was shown that the untoward effect of certain oxidants may be enhanced in the presence of other aerosols.

Interactions at the *receptor site* have been found in isolated perfused lung experiments. Oxygen tolerance may be an example, when pre-exposure to one concentration of oxygen mitigates later exposure to 100% oxygen by modifying cellular and enzymatic composition of the lung.

Damage of the alveolar zone by the antioxidant butylated hydroxytoluene (BHT) in mice can be greatly enhanced by *subsequent exposure* to oxygen concentration which, otherwise, would have little if any demonstrable effect.

The *synergistic interaction* between BHT and oxygen in mice results in interstitial pulmonary fibrosis. Acute or chronic lung disease may then be caused not only by one agent, but also in many instances by the interaction of several agents.

Knowledge Check

1) An increase in skeletal muscle cells in athletes due to exercise and increased metabolic demand is an example of:

a) Pathological adaptation
b) Physiological adaptation

Answer available at the end of this section.

2) A cellular response in which there is an increase in the number of cells in a tissue is known as:

a) Atrophy
b) Hypertrophy
c) Hyperplasia
d) Metaplasia

Answer available at the end of this section.

3) A condition of abnormal cell changes or deranged cell growth in which the cells are structurally changed in size, shape, and appearance from the original cell type is known as:

a) Dysplasia
b) Anaplasia
c) Neoplasia

Answer available at the end of this section.

Answers

Knowledge Check

Physiological adaptation - **This is the correct answer.**
The increase in skeletal muscle cells in athletes due to exercise and increased metabolic demand is an example of physiological adaptation since the increased muscle is beneficial rather than harmful.

Knowledge Check

Hyperplasia - **This is the correct answer.**
Hyperplasia is an increase in the number of cells in a tissue.

Knowledge Check

Dysplasia - **This is the correct answer.**
Dysplasia is a condition of abnormal cell changes or deranged cell growth in which the cells are structurally changed in size, shape, and appearance from the original cell type.

This page titled 8.4.6.14.1: Adaptation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 14.1: Adaptation by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.14.2: Cell Damage and Tissue Repair

Cell Damage and Tissue Repair

Toxic damage to cells can cause individual cell death and if sufficient cells are lost, the result can be tissue or organ failure, ultimately leading to death of the organism. It is nearly impossible to separate a discussion of cellular toxicity and biochemical toxicity. Most observable cellular changes and cell death are due to specific biochemical changes within the cell or in the surrounding tissue. However, there are a few situations where a toxic chemical or physical agent can cause cell damage without actually affecting a specific chemical in the cell or its membrane. Physical agents such as heat and radiation may damage a cell by coagulating their contents (similar to cooking). In this case, there are no specific chemical interactions. Impaired nutrient supply (such as glucose and oxygen) may deprive the cell of essential materials needed for survival.

Toxic Effects

The majority of toxic effects, especially due to xenobiotics, are due to specific biochemical interactions without causing recognizable damage to a cell or its organelles.
Examples of these toxic effects include:

- Interference with a chemical that transmits a message across a neural synapse such as the inhibition of the enzyme acetylcholinesterase by organophosphate pesticides.
- When one toxic chemical inhibits or replaces another essential chemical such as the replacement of oxygen on the hemoglobin molecule with carbon monoxide.

The human body is extremely complex. In addition to over 200 different cell types and about as many types of tissues, there are literally thousands of different biochemicals, which may act alone or in concert to keep the body functions operating correctly. To illustrate the cell's structures and functions and the chemical toxicity of all tissues and organs would be impossible in this brief tutorial. This section presents only a general overview of toxic effects along with some specific types of toxicity that include cancer and neurotoxicity.

Capacity for Repair

Some tissues have a great capacity for repair, such as most epithelial tissues. Others have limited or no capacity to regenerate and repair, such as nervous tissue. Most organs have a functional reserve capacity so that they can continue to perform their body function although perhaps in somewhat diminished ability. For example:

- Half of a person's liver can be damaged, and the body can regenerate sufficient new liver or repair the damaged section by fibrous replacement to maintain most of the capacity of the original liver.
- The hypertrophy of one kidney to assume the capacity lost when the other kidney has been lost or surgically removed.

Toxic Damage to Cells and Tissues

Toxic damage to cells and tissues can be transient and non-lethal or, in severe situations, the damage may cause death of the cells or tissues. The following diagram illustrates the various effects that can occur with damage to cells. There are four main final endpoints to the cellular or biochemical toxicity:

1. The tissue may be completely repaired and return to normal.
2. The tissue may be incompletely repaired but is capable of sustaining its function with reduced capacity.
3. Death of the organism or the complete loss of a tissue or organ. In some instances, the organism can continue to live with the aid of medical treatment, for example, replacement of insulin or by organ transplantations.
4. Neoplasm or cancers may result, many of which will result in death of the organism and some of which may be cured by medical treatment.

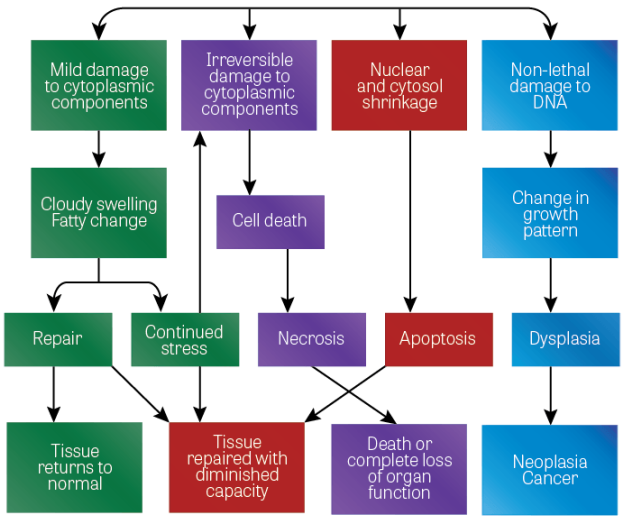


Figure 8.4.6.14.2.1. Toxic damage to cells
(Image Source: NLM)

Reversible Cell Damage

The response of cells to toxic injury may be transient and reversible once the stress has been removed or the compensatory cellular changes are made. In some cases, the full capability of the damaged cells returns. In other cases, a degree of permanent injury remains with a diminished cellular or tissue capacity. In addition to the adaptive cell changes discussed previously, two commonly encountered specific cell changes are associated with toxic exposures, cellular swelling, and fatty change.

Cellular swelling, which is associated with hypertrophy, is due to cellular hypoxia, which damages the sodium-potassium membrane pump. This in turn changes the intracellular electrolyte balance with an influx of fluids into the cell, causing it to swell. Cell swelling is reversible when the cause is eliminated.

Fatty change is more serious and occurs with severe cellular injury. In this situation, the cell has become damaged and is unable to adequately metabolize fat. The result is that small vacuoles of fat accumulate and become dispersed within the cytoplasm. While fatty change can occur in several organs, it is usually observed in the liver. This is because most fat is synthesized and metabolized in liver cells. Fatty change can be reversed but it is a much slower process than the reversal of cellular swelling.

Lethal Injury (Cell Death)

In many situations, the damage to a cell may be so severe that the cell cannot survive. Cell death occurs mainly by two methods: necrosis and apoptosis.

Necrosis is a progressive failure of essential metabolic and structural cell components usually in the cytoplasm. Necrosis generally involves a group of contiguous cells or occurs at the tissue level. Such progressive deterioration in structure and function rapidly leads to cell death or "necrotic cells." Necrosis begins as a reduced production of cellular proteins, changes in electrolyte gradient, or loss of membrane integrity (especially increased membrane permeability). Cytoplasmic organelles (such as mitochondria and endoplasmic reticulum) swell while others (especially ribosomes) disappear. This early phase progresses to fluid accumulation in the cells making them pale-staining or showing vacuoles, which pathologists call "cloudy swelling" or "hydropic degeneration." In some cells, they no longer can metabolize fatty acids so that lipids accumulate in the cytoplasmic vacuoles, referred to as "fatty accumulation" or "fatty degeneration." In the final stages of "cell dying," the nucleus becomes shrunken (pyknosis) or fragmented (karyorrhexis).

Apoptosis or "programmed cell death" is a process of self-destruction of the cell nucleus. Apoptosis is an individual or single cell death in that dying cells are not contiguous but are scattered throughout a tissue. Apoptosis is a normal process in cell turnover

in that cells have a finite lifespan and spontaneously die. During embryonic development, certain cells are programmed to die and are not replaced, such as the cells between each developing finger. If the programmed cells do not die, the fetus ends up with incomplete or fingers joined together in a web fashion.

In apoptosis, the cells shrink from a decrease of cytosol and the nucleus. The organelles (other than the nucleus) appear normal in apoptosis. The cell disintegrates into fragments referred to as "apoptotic bodies." These apoptotic bodies and the organelles are phagocytized by adjacent cells and local macrophages without initiation of an inflammatory response as is seen in necrosis. The cells undergo apoptosis and just appear to "fade away." Some toxicants induce apoptosis or, in other cases, they inhibit normal physiological apoptosis.

Following necrosis, the tissue attempts to regenerate with the same type of cells that have died. When the injury is minimal, the tissue may effectively replace the damaged or lost cells. In severely damaged tissues or long-term chronic situations, the ability of the tissue to regenerate the same cell types and tissue structure may be exceeded, so that a different and imperfect repair occurs.

- An example of this is with chronic alcoholic damage to liver tissue in which the body can no longer replace hepatocytes with hepatocytes but rather connective tissue replacement occurs. Fibrocytes with collagen replace the hepatocytes and normal liver structure with scar tissue. The fibrotic scar tissue shores up the damage but it cannot replace the function of the lost hepatic tissue. With constant fibrotic change, the liver function is continually diminished so that eventually the liver can no longer maintain homeostasis. This fibrotic replacement of the liver is known as cirrhosis (Figure 8.4.6.14.2.2). The normal dark-red, glistening smooth appearance of the liver has been replaced with light, irregular fibrous scar tissue that permeates the entire liver.

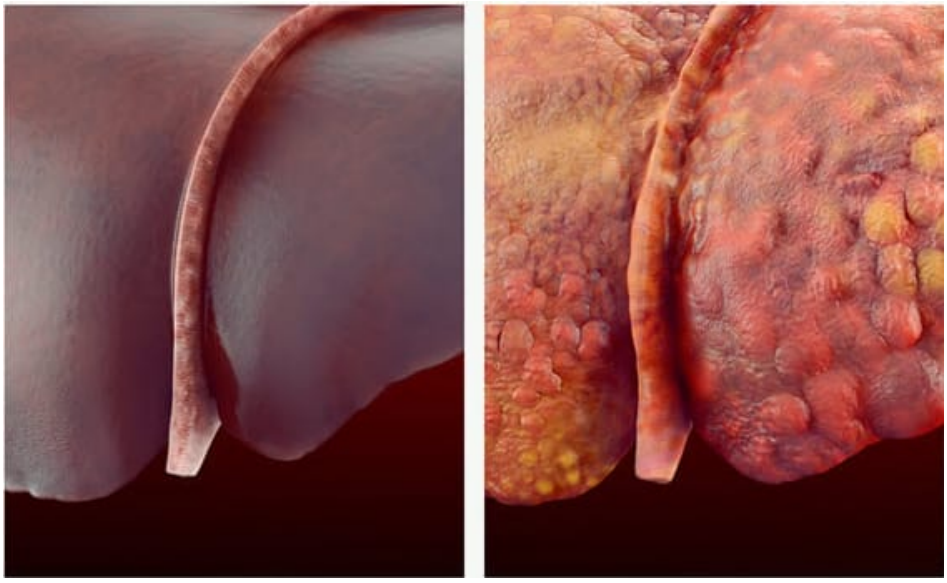


Figure 8.4.6.14.2.2. A healthy liver (left) and a liver with cirrhosis (right)
(Image Source: iStock Photos, ©)

We have so far discussed primarily changes to individual cells. However, a tissue and an organ consist of different types of cells that work together to achieve a particular function. As with a football team, when one member falters, the others rally to compensate. It is the same with a tissue. Damage to one cell type prompts reactions within the tissue to compensate for the injury. Within organs, there are two basic types of tissues: the parenchymal and stromal tissues. The **parenchymal tissues** contain the functional cells (for example, squamous dermal cells, liver hepatocytes, and pulmonary alveolar cells). The **stromal cells** are the supporting connective tissues (for example, blood vessels and elastic fibers).

Cell Repair

Repair of injured cells can be accomplished by either:

1. Regeneration of the parenchymal cells.

2. Repair and replacement by the stromal connective tissue.

The goal of the repair process is to fill the gap that results from the tissue damage and restore the structural continuity of the injured tissue. Normally a tissue attempts to regenerate the same cells that are damaged; however, in many cases, this cannot be achieved so that replacement with a stromal connective tissue is the best means for achieving the structural continuity.
The ability to **regenerate** varies greatly with the type of parenchymal cell. The regenerating cells come from the proliferation of nearby parenchymal cells, which serve to replace the lost cells. Based on regenerating ability, there are three types of cells:

1. **Labile cells** — cells that routinely divide and replace cells that have a limited lifespan (for example, skin epithelial cells, and hematopoietic stem cells).
2. **Stable cells** — cells that usually have a long lifespan with normally a low rate of division; they can rapidly divide upon demand.
3. **Permanent cells** — cells that never divide and do not have the ability for replication even when stressed or when some cells die.

Table 8.4.6.14.2.1 shows examples of cell types.

Cell Type	Examples
Labile cells	• Squamous epithelium of skin, mouth, vagina, and cervix • Columnar epithelium of intestinal tract • Transitional epithelium of urinary tract • Bone marrow cells
Stable cells	• Liver hepatocytes • Alveolar cells of lung • Epithelium of kidney tubules
Permanent cells	• Neurons • Skeletal and cardiac muscle

Table 1. Examples of three cell types of parenchymal cells

Cell Type	Examples
Labile cells	• Squamous epithelium of skin, mouth, vagina, and cervix • Columnar epithelium of intestinal tract • Transitional epithelium of urinary tract • Bone marrow cells
Stable cells	• Liver hepatocytes • Alveolar cells of lung • Epithelium of kidney tubules
Permanent cells	• Neurons • Skeletal and cardiac muscle

The labile cells have a great potential for regeneration by replication and repopulation with the same cell type so long as the supporting structure remains intact. Stable cells can also respond and regenerate but to a lesser degree and are quite dependent on the supporting stromal framework. When the stromal framework is damaged, the regenerated parenchymal cells may be irregularly dispersed in the organ resulting in diminished organ function. The tissue response for the labile and stable cells is initially hyperplasia until the organ function becomes normal again. When permanent cells die they are not replaced in kind but

instead connective tissue (usually fibrous tissue) moves in to occupy the damaged area. This is a form of metaplasia. Examples of replacement by metaplasia are:

- **Cirrhosis of the liver** — liver cells (hepatocytes) are replaced by bands of fibrous tissue, which cannot carry out the metabolic functions of the liver.
- **Cardiac infarcts** — cardiac muscle cells do not regenerate and thus are replaced by fibrous connective tissue (scar). The scar cannot transmit electrical impulses or participate in contraction of the heart.
- **Pulmonary fibrosis** — damaged or dead epithelial cells lining the pulmonary alveoli are replaced by fibrous tissue. Gases cannot diffuse across the fibrous cells and thus gas exchange is drastically reduced in the lungs.

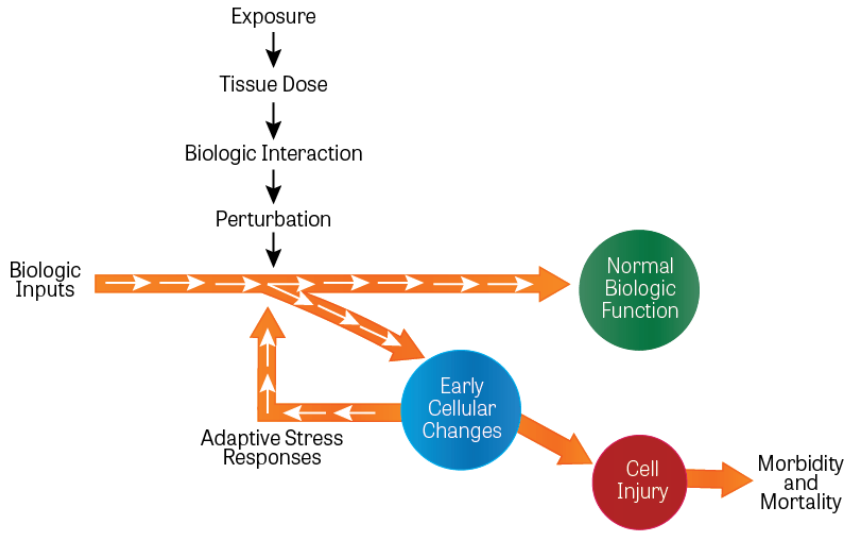


Figure 8.4.6.14.2.3. Activation of Toxicity Pathways
(Image Source: Adapted from Dr. Andrew Maier, adapted from National Research Council (NRC) 2007a.)

Knowledge Check

1) The process of self-destruction of the cell nucleus *(often referred to as "programmed cell death")* is known as:

a) Necrosis
b) Apoptosis
c) Cellular swelling
d) Fatty change

Answer available at the end of this section.

2) The category of cells that routinely divide and replace cells that have a limited lifespan is known as:

a) Labile cells
b) Stable cells
c) Permanent cells

Answer available at the end of this section.

Answers

Knowledge Check

Apoptosis - **This is the correct answer.**
Apoptosis *(referred to as "programmed cell death")* is a process of self-destruction of the cell nucleus.

Knowledge Check

Labile cells - **This is the correct answer.**
Labile cells are cells that routinely divide and replace cells that have a limited lifespan (e.g., skin epithelial cells and hematopoietic stem cells).

This page titled 8.4.6.14.2: Cell Damage and Tissue Repair is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 14.2: Cell Damage and Tissue Repair by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.14.3: Cancer

Cancer

Cancer has long been considered a cellular disease since cancers are composed of cells that grow without restraint in various areas of the body. Such growths of cancerous cells can replace normal cells or tissues causing severe malformations (such as with skin and bone cancers) and failure of internal organs which frequently leads to death. How do cells become cancerous? The development of cancer is an enormously complex process. For once a cell has started on the cancer path, it progresses through a series of steps, which continue long after the initial cause has disappeared.

Overview

There are about as many types of cancers as there are different types of cells in the body (over 100 types). Some cell types constantly divide and are replaced (such as skin and blood cells). Other types of cells rarely or never divide (such as bone cells and neurons). Sophisticated mechanisms exist in cells to control when, if, and how cells replicate. Cancer occurs when these mechanisms are lost and replication takes place in an uncontrolled and disorderly manner. It can arise when one cell or a small group of cells multiplies too many times because of damage to its DNA.

Recent research has begun to unravel the extremely complex pathogenesis of cancer. There is an intricate array of biochemical changes that take place within cells and between cells underlying the progression of cancer that transforms normal cells into cancerous cells. These biochemical changes lead a cell through a series of steps, changing it gradually from a normal to a cancer cell. The altered cell is no longer bound by the regulatory controls that govern the life and behavior of normal cells.

Cancer is not a single disease but a large group of diseases. The common aspect is that all cancers have the same basic property: they are composed of cells that no longer conform to the usual constraints on cell proliferation. In other words, they are uncontrolled growths of cells.

Terminology

The terminology associated with cancer can be confusing and may be used differently among the public and medical communities.
Here are definitions of the most frequently used cancer terms:

- **Cancer** — a malignant tumor that has the ability to metastasize or invade into surrounding tissues.
- **Tumor** — a general term for an uncontrolled growth of cells that becomes progressively worse with time. Tumors may be benign or malignant.
- **Neoplasm** — same as a tumor.
- **Neoplasia** — the growth of new tissue with abnormal and unregulated cellular proliferation.
- **Benign Tumor** — a tumor that does not metastasize or invade surrounding tissue.
- **Malignant Tumor** — a tumor that has the ability to metastasize or invade into surrounding tissues (same as cancer).
- **Metastasis** — ability to establish secondary tumor growth at a new location away from the original site.
- **Carcinogenesis** — the production of a carcinoma (epithelial cancer). Sometimes carcinogenesis is used as a general term for production of any type of tumor.

How are Cancers Named?

While most tumors are generally named in accordance with an **internationally agreed-upon classification scheme**, there are exceptions. Tumors are generally named and classified based on:

- The cell or tissue of origin
- Whether benign or malignant

Most tumor names end with the suffix "oma" which indicates a swelling or tissue enlargement. [Note: some terms ending with -oma are not cancers; for example, a hematoma is merely a swelling consisting of blood].

In naming tumors, qualifiers may be added in addition to the tissue of origin and structural features. For example, a "poorly-differentiated bronchogenic squamous cell carcinoma" is a malignant tumor (carcinoma) of squamous cell type (original cell type), which arose in the bronchi of the lung (site where the cancer started), and in which the cancer cells are poorly differentiated, meaning they have lost much of the normal appearance of squamous cells.

There are several historical exceptions to the standard nomenclature system, often based on their early and accepted use in the literature.

Examples include:

- Some tumors are named after the person that first described the tumor, for example, Wilms tumor (kidney tumor) and Hodgkin lymphoma (a specific form of lymphoid cancer).
- A few cancers are named for their physical characteristics such as pheochromocytomas (dark-colored tumors of the adrenal gland).
- A few cancers are composed of mixtures of cells, for example, fibrosarcoma and carcinosarcoma.

Most malignant tumors fall into one of two categories: carcinomas or sarcomas. The major differences between carcinomas and sarcomas are listed in Table 8.4.6.14.3.1:

Carcinoma	Sarcoma
• Malignant tumor arising in epithelium • The most common form of cancer • Usually spread in lymphatic system	• Malignant tumor arising in connective or muscle tissue • Usually spread by blood stream • Frequently metastasizes to lung

Differences between Benign and Malignant Tumors

The biological and medical consequences of a tumor depend on whether it is benign or malignant.
Table 8.4.6.14.3.2 provides a comparison of the primary differences between benign and malignant tumors:

Characteristics	Benign Tumors	Malignant Tumors
Cell characteristics	Similar to cell of origin (well differentiated)	Dissimilar from cell of origin (poorly differentiated)
Growth characteristics	Tumor edges move outward in a smooth manner (encapsulated), grows by expansion, and compresses and displaces surrounding tissues Tumor cells stay attached to the clone or mass of cells and do not break away and start new growths elsewhere in the body	Tumor edges move outward in an irregular fashion (usually no capsule) and can infiltrate, invade, and destroy surrounding tissues Tumor cells can break away from the cloned mass, live independently, move to other areas of the body, and start new clones or growths
Rate of growth	Slow	Rapid
Degree of vascularity	Slight	Moderate-marked
Recurrence after surgical removal	Seldom	Frequently
Degree of necrosis and ulceration	Unusual	Common
Likelihood of causing systemic effects	Unusual unless the tumor is a secreting endocrine neoplasm	Common and usually life-threatening

Common Sites for Cancer

Cancer can occur in almost any tissue or organ. Some cells and tissues are more likely to become cancerous than others are, particularly those **cells that normally undergo proliferation** to replace cells that have been lost due to injury or cell death. Those cells that don't proliferate (for example, neurons and heart muscle cells) rarely give rise to cancers. Figure 8.4.6.14.3.1 illustrates the most frequent occurrence of cancers in various body sites.

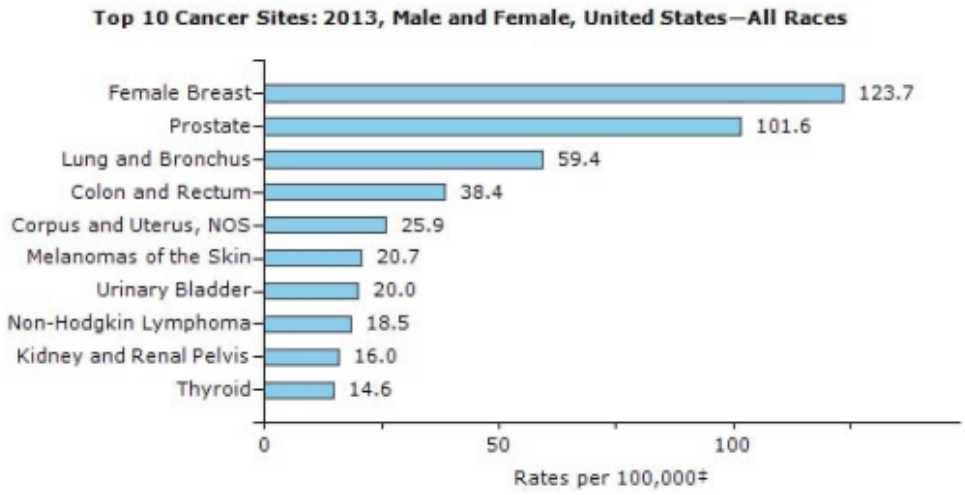


Figure 8.4.6.14.3.1. Top 10 Cancer Sites for Males and Females from All Races in the United States in 2013
(Image Source: CDC, **https://nccd.cdc.gov/uscs/toptencancers.aspx**)
‡ Rates are age-adjusted to the 2000 U.S. standard population (19 age groups – Census P25–1130)

While the prostate is the most common type of cancer that occurs in men, most survive with treatment. In contrast, other types of cancer are more often fatal. For example, the most common cancer, which causes death in men, even with treatment, is lung cancer. With women, a similar situation exists in that the breast is the most common site for cancer but more women die as a result of lung cancer.

What Do Cancers Look Like?

Cancer is a general term for more than 100 different cellular diseases, all with the same characteristic – the uncontrolled abnormal growth of cells in different parts of the body. Cancers appear in many forms. A few types are visible to the unaided eye but others grow inside the body and slowly destroy or replace internal tissues.

Skin Cancer

An example of a cancer that can be easily seen by the unaided eye is skin cancer. Skin cancers appear as raised, usually dark-colored, irregularly-shaped growths on the skin. As the cancer grows, it spreads to nearby areas of the skin. In advanced cases, the cancer metastasizes to lymph nodes and organs far away from the original site. The skin cancer illustrated in Figure 8.4.6.14.3.2 is known as a basal cell carcinoma. Melanomas and squamous cell carcinomas are other common skin cancers. Melanomas are usually the most malignant of the skin cancers.

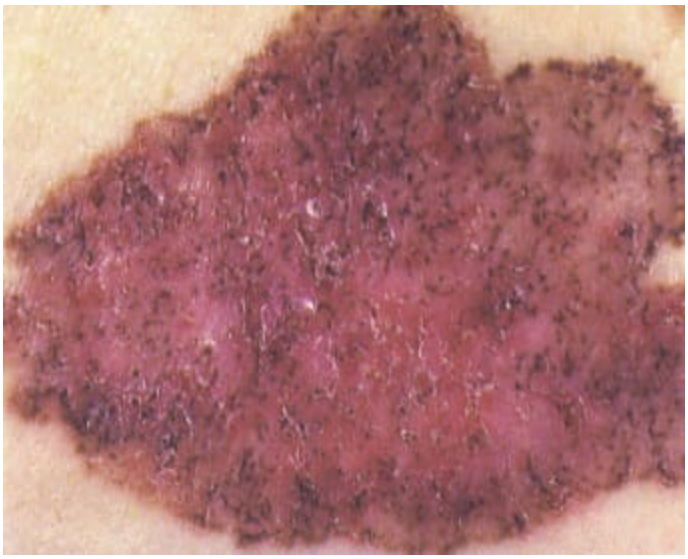


Figure 8.4.6.14.3.2. Photograph of basal cell carcinoma of the skin (Image Source: NLM)

Other Cancers

Most cancers involve internal organs and require elaborate diagnostic tests to diagnose. Some large internal tumors can be felt or will push the skin outward and can be detected by noting abnormal bulges or an abnormal feel (for example, a hard area) to the body. Thyroid tumors, bone tumors, breast tumors, and testicular tumors are cancers that might be felt or observed by the patient. Other internal tumors may only be suspected based on diminished organ function (such as difficulty breathing with lung cancers), pain, bleeding (for example, blood in feces with colon cancer), weakness, or other unusual symptoms. To confirm the actual existence of a cancer may require diagnostic tests. This is especially the case where the cancer is not growing as a single large lump, but rather as a series of small tumors (metastatic foci) or when widely dispersed throughout the body (such as leukemia).
A few examples of internal cancers are presented in the following figures.

Liver Cancer

Numerous cancer nodules can be seen showing that much of the liver has been destroyed (Figure 8.4.6.14.3.3).

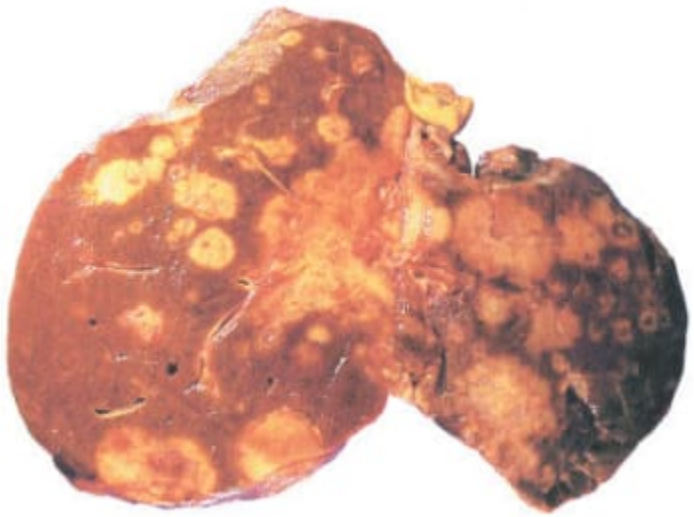


Figure 8.4.6.14.3.3. A liver with numerous cancer nodules
(Image Source: NLM)

Lung Cancer

An early developing squamous cell carcinoma can be seen growing in the middle of the lung (Figure 8.4.6.14.3.4). As the cancer develops, it will consume more of the lung and metastasize to other areas of the body.



Figure 8.4.6.14.3.4. A cancerous lung
(Image Source: NLM)

Kidney Cancer

The photograph in Figure 8.4.6.14.3.5 shows the cancer has consumed much of the upper portion of the kidney.

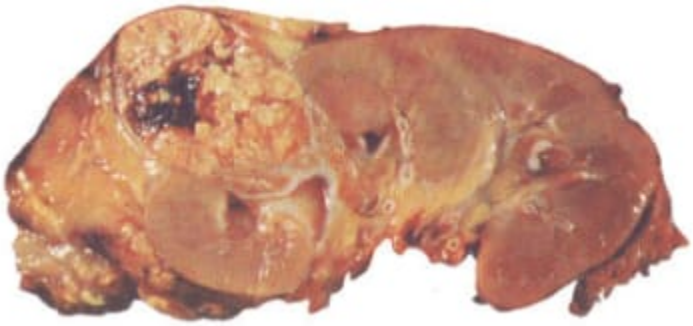


Figure 8.4.6.14.3.5. A kidney with cancer
(Image Source: NLM)

Historical Changes in Incidence of Cancer

Cancer has been recognized in humans for centuries. However, the incidence of various types of cancer has changed since the mid-1900s. This is especially true for lung and stomach cancer. Deaths from lung cancer hit a peak in the early 1990s and have been slowly declining since 2001. During that same period, deaths from stomach cancer decreased substantially. Breast cancer caused more deaths than any other type cancer in women for many decades. However, when women began smoking cigarettes, deaths from lung cancer outpaced deaths from breast cancer. These changes in types and incidences of cancer reflect the increased longevity of people as well as personal habits and environmental changes.

Latency Period for Cancer Development

Cancer is a chronic condition, which develops gradually over a period of time, and may become a clinical concern many years following the initial exposure to a carcinogen. This period of time is referred to as the **latency period**. The latency period varies with the type of cancers and may be as short as a few years to over 30 years. For example, the latency period for leukemia after benzene or radiation exposure may only be five years. In contrast, the latency period may be 20–30 years for skin cancer after arsenic exposure and mesothelioma (cancer of the pleura around the lungs) after asbestos exposure.

Survival Time

Success in treating cancer varies greatly with the type of cancer with some cancers responding to treatment whereas others do not. For example, medical treatment of cancers of the pancreas, liver, esophagus, and lung are largely unsuccessful. In contrast, cancers of the thyroid, testes, and skin respond quite well to treatment. Table 8.4.6.14.3.3 shows the 5-year survival rate by cancer location.

Organ	Male 5-year Survival	Female 5-year Survival
Pancreas	7.6%	7.8%
Liver	17.1%	18.5%
Esophagus	18.3%	18.5%
Lung and Bronchus	14.9%	20.8%
Stomach	28.5%	33.2%
Brain and Nervous System	32.8%	35.1%
Leukemia	60.7%	58.3%
Ovary	----------	46.2%
Non-Hodgkin Lymphoma	69.2%	72.4%
Oral Cavity and Pharynx	65.0%	67.0%
Kidney	73.0%	75.0%
Colon/Rectum	64.9%	65.2%
Cervix	----------	67.5%
Hodgkin Lymphoma	85.4%	87.7%
Breast	83.6%	89.7%
Prostate	98.9%	----------
Urinary Bladder	78.9%	73.0%
Melanoma of the Skin	89.5%	94.0%
Testes	95.4%	----------
Thyroid	95.7%	98.7%
All Sites	67.0%	66.9%

Table 8.4.6.14.3.3. Five year survival rate by primary cancer site (2006-2012)
(Source: Table 1. Surveillance, Epidemiology, and End Results Program (SEER), National Cancer Institute, https://seer.cancer.gov/csr/1975_2013/results_merged/topic_survival.pdf)

What Causes Cancer?

A large number of industrial, pharmaceutical, and environmental chemicals have been identified as potential carcinogens by animal tests. Human epidemiology studies have confirmed that many are human carcinogens as well. However, while it is apparent that chemicals and radiation play a substantial role, it appears that lifestyle factors (such as diet, obesity, and smoking), and infections (such as hepatitis B, hepatitis C, and Human Papillomaviruses) are also major factors leading to the likelihood that a person will develop cancer. Additional factors that are involved in the development of cancer include aging and heredity.

Pathogenesis of Cancer

Carcinogenesis is a multi-step, multi-factorial genetic disease. All known tumors are composed of cells with genetic alterations that make them perform differently from their progenitor (parent) cells. The carcinogenesis process is very complex and unpredictable consisting of several phases and involving multiple genetic events (mutations) that take place over a very long period of time, at least 10 years for most types of cancer.

Cancer cells do not necessarily proliferate faster than their normal progenitors. In contrast to normal proliferating tissues where there is a strict and controlled balance between cell death and replacement, cancers grow and expand since more cancer cells are produced than die in a given time period. For a tumor to be detected it must attain a size of at least one cubic centimeter (about the size of a pea). This small tumor contains 100 million to a billion cells at that time. The development from a single cell to that size also means that the mass has doubled at least 30 times. During the long and active period of cell proliferation, the cancerous cells may have become aggressive in growth and have reverted to a less differentiated type cell that is not similar to the original cell type.

While knowledge of carcinogenesis continues to evolve, it is clear that there are at least three main phases in cancer development:

1. Initiation
2. Promotion/Conversion
3. Progression

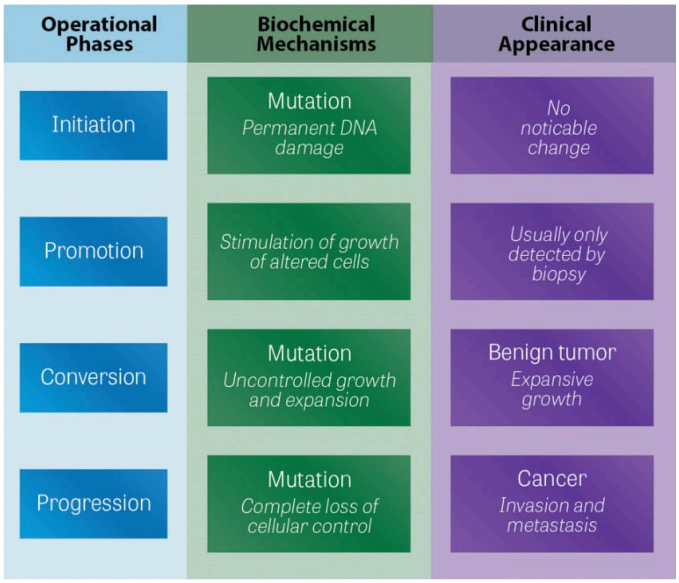


Figure 8.4.6.14.3.6. Phases of carcinogenesis
(Image Source: NLM)

1. Initiation

The **initiation** phase consists of the alteration of the DNA (mutation) of a normal cell, which is an irreversible change. The initiated cell has developed a capacity for individual growth. At this time, the initiated cell is indistinguishable from other similar cells in the tissue. The initiating event can consist of a single exposure to a carcinogenic agent or, in some cases, it may be an inherited genetic defect.

- An example is retinoblastoma in which some children who develop the disease may have inherited an altered copy of the gene involved and are at risk of passing the altered gene to successive generations.

The initiated cell, whether inherited or newly mutated, may remain dormant for months to years and, unless a promoting event occurs, may never develop into a clinical cancer case.

2. Promotion/Conversion

The **promotion/conversion** phase is the second major step in the carcinogenesis process in which specific agents (referred to as promoters) enhance the further development of the initiated cells. Promoters often, but not always, interact with the cell's DNA and influence the further expression of the mutated DNA so that the initiated cell proliferates and progresses further through the

carcinogenesis process. The clone of proliferating cells in this stage takes a form consistent with a benign tumor. The mass of cells remains as a cohesive group and physically keeps in contact with each other.

3. Progression

Progression is the third recognized step and is associated with the development of the initiated cell into a biologically malignant cell population. In this stage, a portion of the benign tumor cells may be converted into malignant forms so that a true cancer has evolved. Individual cells in this final stage can break away and start new clones of growth distant from the original site of development of the tumor. This is known as **metastasis**.

Genetic Activity

While the three-stage pathogenesis scheme describes the basic sequence of events in the carcinogenesis process, the actual events that take place in these various steps are due to activities of specific genes within the DNA of the cells. Cellular DNA contains two types of genes:

1. **Structural genes** direct the production of specific proteins within the cell.
2. **Regulatory genes** control the activity of the structural genes and direct the proliferation process of the cell.

The three classes of regulatory genes considered to have major roles in the carcinogenesis process are known as:

1. Proto-oncogenes
2. Oncogenes
3. Suppressor genes

Proto-oncogenes are normal cellular genes that encode and instruct the production of the regulatory proteins and growth factors within the cell or its membrane. The proteins encoded by proto-oncogenes are necessary for normal cellular cell growth and differentiation. Activation of a proto-oncogene can cause the alteration in the normal growth and differentiation of cells, which leads to neoplasia. Several agents can activate proto-oncogenes. This is the result of **point mutations** or by DNA rearrangements of the proto-oncogenes. The product of this proto-oncogene activation is an oncogene. Many proto-oncogenes have been identified and have usually been named after the source of their discovery, for example, the KRAS proto-oncogene was named for the discovery using the Kirsten rat sarcoma virus. HRAS, MYC, MYB, and SRC are other examples of proto-oncogenes. The proto-oncogenes are not specific for the original species but have been found in many other species, including humans. These proto-oncogenes are present in many cells but remain dormant until activated. Either a point mutation or chromosomal damage of various types can induce activation. Once activated they become an oncogene.

Oncogenes are altered or misdirected proto-oncogenes which now have the ability to direct the production of proteins within the cell that can change or transform the normal cell into a neoplastic cell. Most oncogenes differ from their proto-oncogenes by a single point mutation located at a specific codon (a group of three DNA bases that encodes for a specific amino acid) of a chromosome. The altered DNA in the oncogene results in the production of an abnormal protein that can alter cell growth and differentiation. It appears that a single activated oncogene is not sufficient for the growth and progression of a cell and its offspring to form a cancerous growth. However, it is a major step in the carcinogenesis process.

Tumor suppressor genes, sometimes referred to as anti-oncogenes, are present in normal cells and serve to counteract and change the proto-oncogenes and altered proteins that they are responsible for. The tumor suppressor genes serve to prevent a cell with damaged DNA from proliferating and evolving into an uncontrolled growth. They actively function to effectively oppose the action of an oncogene. If a tumor suppressor gene is inactivated (usually by a point mutation), its control over the oncogene and transformed cell may be lost. Thus the tumor-potential cell can now grow without restraint and is free of the normal cellular regulatory control. The suppressor gene most frequently altered in human tumors is the p53 gene. Damaged p53 genes have been identified in over 50% of human cancers.

The **p53 gene** normally halts cell division and stimulates repair enzymes to rebuild and restore the damaged regions of the DNA. If the damage is too extensive, the p53 commands the cell to self-destruct. An altered p53 is incapable of these defensive actions and cannot prevent the cell with damaged DNA from dividing and proliferating in an erratic and uncontrolled manner. This is the essence of cancer.

This section represents only a brief overview of an enormously complex process for which knowledge is continuously evolving with the tools of molecular biology. New factors are continuously being identified; however, many pieces of this cancer puzzle remain elusive at this time.

Knowledge Check

1) A body growth with the ability to metastasize or invade into surrounding tissues is known as a:

a) Benign tumor
b) Malignant tumor
c) Hyperplasia

Answer available at the end of this section.

2) Most cancers are thought to be due to the following:

a) Infections
b) Food additives
c) Lifestyle factors
d) Pollution

Answer available at the end of this section.

3) The initial stage in carcinogenesis in which there is an alteration of the DNA (mutation) is referred to as the:

a) Progression stage
b) Promotion stage
c) Initiation stage

Answer available at the end of this section.

4) The cellular gene which is present in most normal cells and serves as a balance to the genes for tumor expression is known as a:

a) Tumor suppressor gene
b) Oncogene
c) Proto-oncogene

Answer available at the end of this section.

Answers

Knowledge Check

Malignant tumor - **This is the correct answer.**
A malignant tumor that has the ability to metastasize or invade into surrounding tissues. It is the same as cancer.

Knowledge Check

Lifestyle factors - **This is the correct answer.**
Lifestyle (including diet, tobacco use, reproductive and sexual behavior, and alcohol consumption) is considered to cause about 75% of all cancers.

Knowledge Check

Initiation stage - **This is the correct answer.**
The initiation phase consists of the alteration of the DNA (mutation) of a normal cell, which is irreversible change.

Knowledge Check

Tumor suppressor gene - **This is the correct answer.**
Tumor suppressor genes, sometimes referred to as anti-oncogenes, are present in normal cells and serve as a balance to the genes for expression or proto-oncogenes.

This page titled 8.4.6.14.3: Cancer is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 14.3: Cancer by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.14.4: Neurotoxicity

Neurotoxicity

The nervous system is very complex and toxins can act at many different points in this complex system. The focus of this section is to provide a basic overview of how the nervous system works and how neurotoxins affect it. Due to the complexity of these topics, this section does not include extensive details related to the anatomy and physiology of the nervous system or the many neurotoxins in our environment and the subtle ways they can damage the nervous system or interfere with its functions.

Since the nervous system innervates all areas of the body, some toxic effects may be quite specific and others generalized depending upon where in the nervous system the toxin exerts its effect. Before discussing how neurotoxins cause damage, we will look at the basic anatomy and physiology of the nervous system.

Anatomy and Physiology of the Nervous System

The nervous system has three basic functions:

1. Specialized cells detect sensory information from the environment and relay that information to other parts of the nervous system.
2. It directs motor functions of the body usually in response to sensory input.
3. It integrates the thought processes, learning, and memory.

All of these functions are potentially vulnerable to the actions of toxicants.
The nervous system consists of two fundamental anatomical divisions:

1. Central nervous system (CNS)
2. Peripheral nervous system (PNS)

Central Nervous System

The CNS includes the brain and spinal cord. The CNS serves as the control center and processes and analyzes information received from sensory receptors and in response issues motor commands to control body functions. The brain, which is the most complex organ of the body, structurally consists of six primary areas (Figure 8.4.6.14.4.1):

1. **Cerebrum** — controls thought processes, intelligence, memory, sensations, and complex motor functions.
2. **Diencephalon *(thalamus, hypothalamus, pituitary gland)*** — relays and processes sensory information; controlls emotions, autonomic functions, and hormone production.
3. **Midbrain** — processes auditory and visual data; generates involuntary motor responses.
4. **Pons** — a tract and relay center which also assists in somatic and visceral motor control.
5. **Cerebellum** — voluntary and involuntary motor activities based on memory and sensory input.
6. **Medulla oblongata** — relays sensory information to the rest of the brain; regulates autonomic function, including heart rate and respiration.

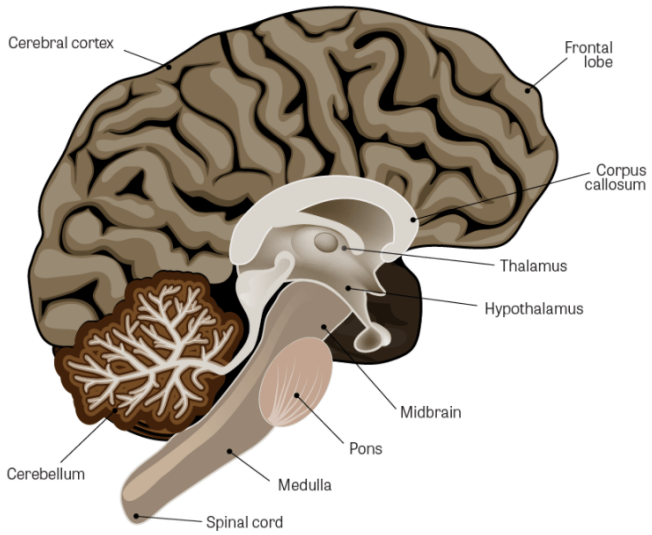


Figure 8.4.6.14.4.1. Internal anatomy of the brain
(Image Source: Adapted from iStock Photos, ©)

Peripheral Nervous System

The PNS consists of all nervous tissue outside the CNS (Figure 8.4.6.14.4.2). The PNS contains two forms of nerves:

1. Afferent nerves, which relay sensory information to the CNS.
2. Efferent nerves, which relay motor commands from the CNS to various muscles and glands.

Efferent nerves are organized into two systems. One is the **somatic nervous system** that is also known as the voluntary system and which carries motor information to skeletal muscles. The second efferent system is the **autonomic nervous system**, which carries motor information to smooth muscles, cardiac muscle, and various glands. The major difference between these two systems pertains to conscious control.

- The somatic system is under our voluntary control such as moving our arms by consciously telling our muscles to contract.
- In contrast, we cannot consciously control the smooth muscles of the intestine, heart muscle, or secretion of hormones. Those functions are automatic and involuntary as controlled by the autonomic nervous system.

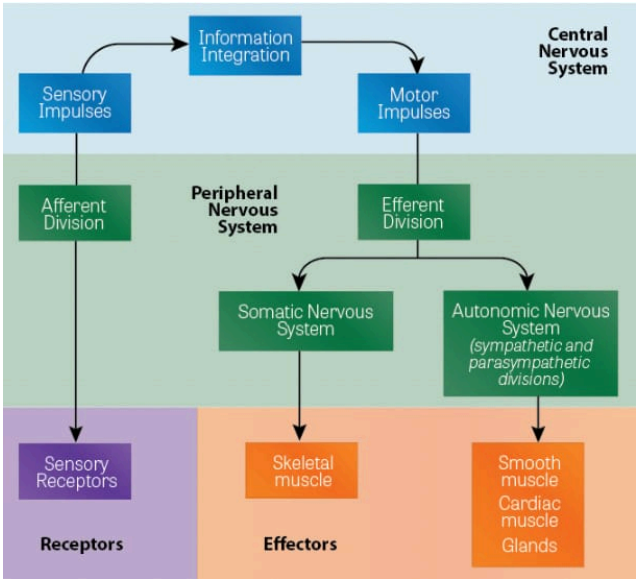


Figure 8.4.6.14.4.2. Structures of the central nervous system and peripheral nervous system (Image Source: NLM)

Cells of the Nervous System

There are two categories of cells found in the nervous system: neurons and glial cells. **Neurons** are the functional nerve cells directly responsible for transmission of information to and from the CNS to other areas of the body. **Glial cells** (also known as neuroglia) provide support to the neural tissue, regulate the environment around the neurons, and protect against foreign invaders.

Neurons communicate with all areas of the body and are present within both the CNS and PNS. They serve to transmit rapid impulses to and from the brain and spinal cord to virtually all tissues and organs of the body. As such, they are an essential cell and their damage or death can have critical effects on body function and survival. When neurons die, they are not replaced. As neurons are lost, so are certain neural functions such as memory, ability to think, quick reactions, coordination, muscular strength, and our various senses such as sight, hearing, and taste. If the neuron loss or impairment is substantial, severe and permanent disorders can occur, such as blindness, paralysis, and death.

A neuron consists of a cell body and two types of extensions, numerous dendrites, and a single axon (Figure 8.4.6.14.4.3). **Dendrites** are specialized in receiving incoming information and sending it to the neuron cell body with transmission (electrical charge) on down the axon to one or more junctions with other neurons or muscle cells (known as synapses). The **axon** may extend long distances, over a meter in some cases, to transmit information from one part of the body to another. The **myelin sheath** is a multi-layer coating that wraps some axons and helps insulate the axon from surrounding tissues and fluids, and prevents the electrical charge from escaping from the axon.

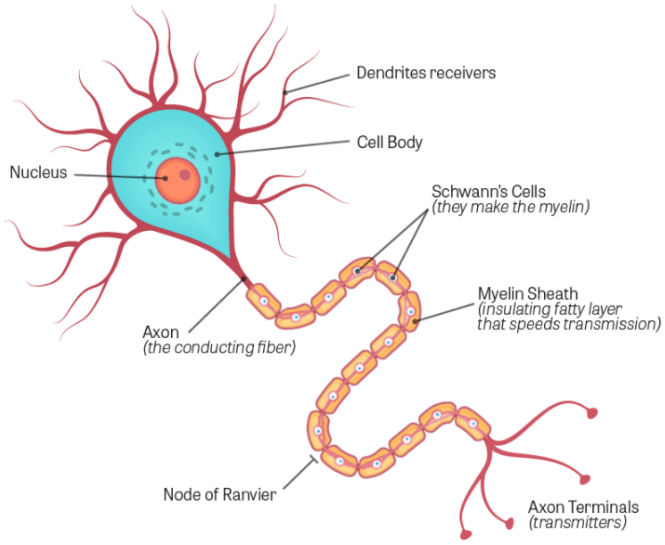


Figure 8.4.6.14.4.3. Neuron structure
(Image Source: Adapted from iStock Photos, ©)

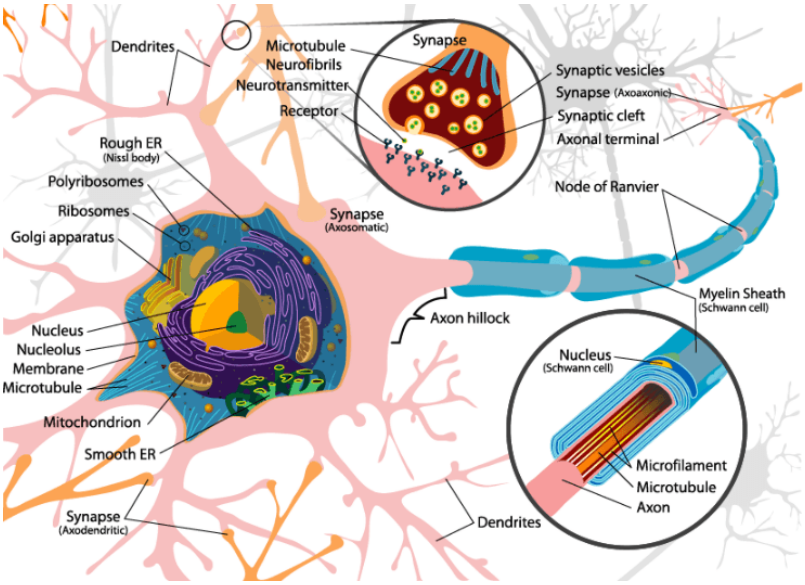


Figure 8.4.6.14.4.4. Complete neuron cell diagram
(Image Source: Adapted from Wikimedia Commons, obtained under Public Domain. Author: LadyofHats.)

Information passes along the network of neurons between the CNS and the sensory receptors and the effectors by a combination of electrical pulses and chemical neurotransmitters. The information (electrical charge) moves from the dendrites through the cell body and down the axon. The mechanism by which an electrical impulse moves down the neuron is quite complex. When the neuron is at rest, it has a negative internal electrical potential. This changes when a neurotransmitter binds to a dendrite receptor. Protein channels of the dendrite membrane open allowing the movement of charged chemicals across the membrane, which creates an electrical charge. The propagation of an electrical impulse (known as action potential) proceeds down the axon by a continuous series of openings and closings of sodium-potassium channels and pumps. The action potential moves like a wave from one end (dendritic end) to the terminal end of the axon.

However, the electrical charge cannot cross the gap (synapse) between the axon of one neuron and the dendrite of another neuron or an axon and a connection with a muscle cell (neuromuscular junction). Chemicals called neurotransmitters move the information across the synapse.

Neurons do not make actual contact with one another but have a gap, known as a **synapse**. As the electrical pulse proceeds up or down an axon, it encounters at least one junction or synapse. An electrical pulse cannot pass across the synapse. At the terminal end of an axon is a synaptic knob, which contains the neurotransmitters.

Neurotransmitters

Vesicles release **neurotransmitters** upon stimulus by an impulse moving down the presynaptic neuron. The neurotransmitters diffuse across the synaptic junction and bind to receptors on the postsynaptic membrane. The neurotransmitter-receptor complex then initiates the generation of an impulse on the next neuron or the effector cell, for example, a muscle cell or secretory cell.

After the impulse has again been initiated, the neurotransmitter-complex must be inactivated or continuous impulses (beyond the original impulse) will be generated. Enzymes perform this inactivation, which serves to break down the complex at precisely the right time and after the exact impulse has been generated. There are several types of neurotransmitters and corresponding inactivating enzymes. One of the major neurotransmitters is acetylcholine with acetylcholinesterase as the specific inactivator.

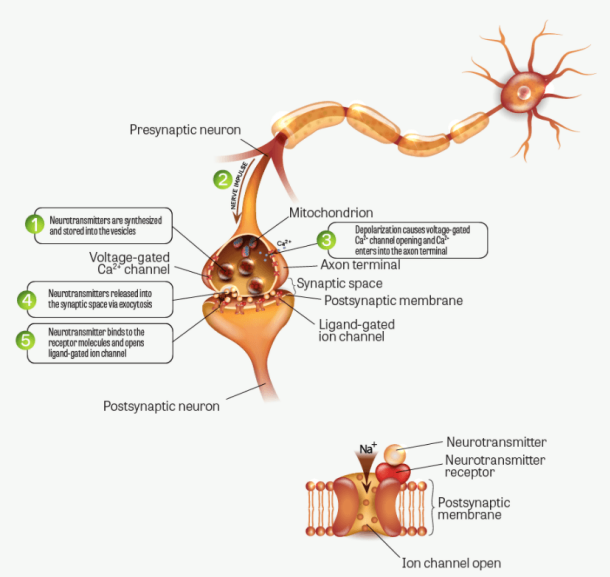


Figure 8.4.6.14.4.5. Impulse transmission across synapse
(Image Source: Adapted from iStock Photos, ©)

There are over 100 known neurotransmitters. Among the most well-known are:

- Acetylcholine
- Dopamine
- Serotonin
- Norepinephrine
- GABA (gamma-aminobutyric acid)

Types of Neurons

Neurons are categorized by their function and consist of three types:

1. **Sensory neurons (afferent neurons)** carry information from sensory receptors (usually processes of the neuron) to the CNS. Some sensory receptors detect external changes such as temperature, pressure, and the senses of touching and vision. Others monitor internal changes such as balance, muscle position, taste, deep pressure, and pain.
2. **Motor neurons (effector neurons)** relay information from the CNS to other organs terminating at the effectors. Motor neurons make up the efferent neurons of both the somatic and autonomic nervous systems.

3. **Interneurons (association neurons)** are located only in the CNS and provide connections between sensory and motor neurons. They can carry either sensory or motor impulses. They are involved in spinal reflexes, analysis of sensory input, and coordination of motor impulses. They also play a major role in memory and the ability to think and learn.

Glial Cells

Glial cells are important as they provide a structure for the neurons by protecting them from outside invading organisms, and maintaining a favorable environment (nutrients, oxygen supply, etc.). The neurons are highly specialized and do not have all the usual cellular organelles to provide them with the same life-support capability. They are highly dependent on the glial cells for their survival and function. For example, neurons have such a limited storage capacity for oxygen that they are extremely sensitive to decreases in oxygen (anoxia) and will die within a few minutes. The list below describes the types of glial cells:

- **Astrocytes** are big cells, only in the CNS, and maintain the blood-brain barrier that controls the entry of fluid and substances from the circulatory system into the CNS. They also provide rigidity to the brain structure.
- **Schwann cells and oligodendrocytes** wrap themselves around some axons to form **myelin**, which serves like insulation. Myelinated neurons usually transmit impulses at high speed, such as needed in motor neurons. Loss of myelination causes a dysfunction of these cells.
- **Microglia** are small, mobile, phagocytic cells.
- **Ependymal cells** produce the cerebrospinal fluid (CSF) which surrounds and cushions the central nervous system.

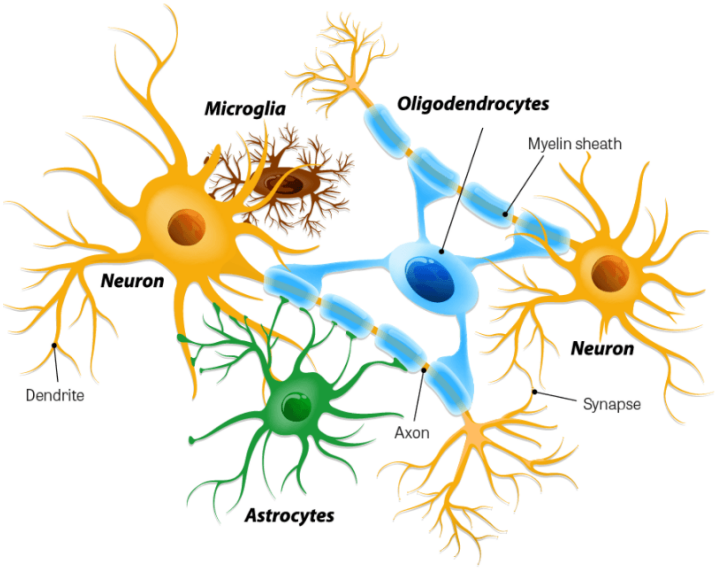


Figure 8.4.6.14.4.6. Neurons and neuroglial cells
(Image Source: Adapted from iStock Photos, ©)

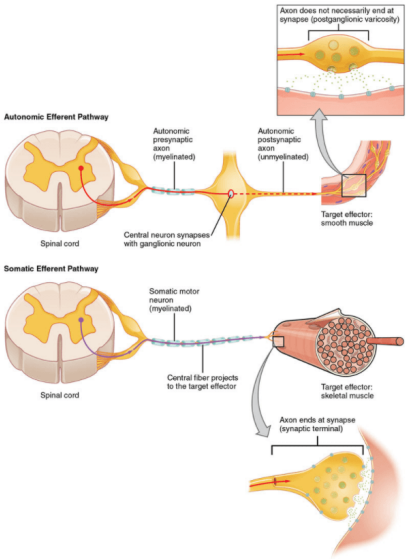


Figure 8.4.6.14.4.7. Comparison of somatic and visceral reflects
(Image Source: Wikimedia Commons, obtained under Creative Commons Attribution 3.0 Unported License. Author: OpenStax College. View original image. Source: Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, Jun 19, 2013.)

Toxic Damage to Nervous System

The nervous system is quite vulnerable to toxins since chemicals interacting with neurons can change the critical voltages, which must be carefully maintained. However, the nervous system has defense mechanisms that can protect it from toxins.

Most of the CNS is protected by an anatomical barrier between the neurons and blood vessels, known as the **blood-brain barrier**. It is protected from some toxin exposures by tightening junctions between endothelial cells of the blood vessels in the CNS and having astrocytes surround the blood vessels. This prevents the diffusion of chemicals out of the blood vessels and into the intracellular fluid except for small, lipid-soluble, non-polar molecules. Specific transport mechanisms exist to transport essential nutrients (such as glucose and amino acids and ions) into the brain. Another defense mechanism within the brain to counter chemicals that pass through the vascular barrier is the presence of metabolizing enzymes. Certain detoxifying enzymes, such as monoamine oxidase, can biotransform many chemicals to less toxic forms as soon as they enter the intercellular fluid.

The basic types of changes due to toxins can be divided into three categories – 1) sensory; 2) motor; and 3) interneuronal – depending on the type of damage sustained.

1. Damage can occur to sensory receptors and **sensory neurons**, which can affect the basic senses of pressure, temperature, vision, hearing, taste, smell, touch, and pain.
    - For example, heavy metal poisoning (especially lead and mercury) can cause deafness and loss of vision.
    - Several chemicals including inorganic salts and organophosphorus compounds can cause a loss of sensory functions.
2. Damage to **motor neurons** can cause muscular weakness and paralysis.
    - Isonicotinic hydrazide (used to treat tuberculosis) can cause such damage.
3. **Interneuronal** damage can cause learning deficiencies, loss of memory, incoordination, and emotional conditions.
    - Low levels of inorganic mercury and carbon monoxide can cause depression and loss of memory.

Mechanisms for Toxic Damage to the Nervous System

Toxic damage to the nervous system occurs by the following basic mechanisms:

1. Direct damage and death of neurons and glial cells.

2. Interference with electrical transmission.
3. Interference with chemical neurotransmission.

A. Death of Neurons and Glial Cells

The most common cause of death of neurons and glial cells is **anoxia**, an inadequate oxygen supply to the cells or their inability to utilize oxygen. Anoxia may result from the blood's decreased ability to provide oxygen to the tissues (impaired hemoglobin or decreased circulation) or from the cells unable to utilize oxygen.

- For example, carbon monoxide and sodium nitrite can bind to hemoglobin preventing the blood from being able to transport oxygen to the tissues.
- Hydrogen cyanide and hydrogen sulfide can penetrate the blood-brain barrier and is rapidly taken up by neurons and glial cells.
- Another example is sodium fluoroacetate (commonly known as Compound 1080, a rodent pesticide) which inhibits a cellular enzyme.

Those chemicals interfere with cellular metabolism and prevent nerve cells from being able to utilize oxygen. This is called **histoxic anoxia**.
Neurons are among the most sensitive cells in the body to inadequate oxygenation. Lowered oxygen for only a few minutes is sufficient to cause irreparable changes leading to the death of neurons.
Several other neurotoxins directly damage or kill neurons, including:

- Lead
- Mercury
- Some halogenated industrial solvents including methanol (wood alcohol)
- Toluene
- Trimethyltin polybrominated diphenyl ethers (PBDEs)

While some neurotoxic agents affect neurons throughout the body, others are quite selective.

- For example, methanol specifically affects the optic nerve, retina, and related ganglion cells while trimethyltin kills neurons in the hippocampus, a region of the cerebrum.

Other agents can degrade neuronal cell function by diminishing its ability to synthesize protein, which is required for the normal function of the neuron.

- Organomercury compounds exert their toxic effect in this manner.

With some toxins, only a portion of the neuron is affected. If the cell body is killed, the entire neuron will die. Some toxins can cause death or loss of only a portion of the dendrites or axon while the cell itself survives but with diminished or total loss of function. Commonly axons begin to die at the very distal end of the axon with necrosis slowly progressing toward the cell body. This is referred to as "dying-back neuropathy."

- Some organophosphate chemicals (including some pesticides) cause this distal axonopathy. The mechanism for the dying back is not clear but may be related to the inhibition of an enzyme (neurotoxic esterase) within the axon.
- Other well-known chemicals can cause distal axonopathy include ethanol, carbon disulfide, arsenic, ethylene glycol (in antifreeze), and acrylamide.

B. Interference with Electrical Transmission

There are two basic ways that a foreign chemical can interrupt or interfere with the propagation of the electrical potential (impulse) down the axon to the synaptic junction:

1. To interfere with the movement of the action potential down the intact axon.
2. To cause structural damage to the axon or its myelin coating. Without an intact axon, transmission of the electrical potential is not possible.

Agents that can block or interfere with the sodium and potassium channels and sodium-potassium pump cause interruption of the propagation of the electrical potential. This will weaken, slow, or completely interrupt the movement of the electrical potential. Many potent neurotoxins exert their toxicity by this mechanism.

- Tetrodotoxin (a toxin in frogs, pufferfish, and other invertebrates) and saxitoxin (a cause of shellfish poisoning) blocks sodium channels. Batrachotoxin (a toxin in South American frogs used as arrow poison) and some pesticides (DDT and pyrethroids) increases the permeability of the neuron membrane preventing closure of sodium channels which leads to repetitive firing of the electrical charge and an exaggerated impulse.

A number of chemicals can cause **demyelination**. Many axons (especially in the PNS) are wrapped with a protective myelin sheath that acts as insulation and restricts the electrical impulse within the axon. Agents that selectively damage these coverings disrupt or interrupt the conduction of high-speed neuronal impulses. Loss of a portion of the myelin can allow the electrical impulse to leak out into the tissue surrounding the neuron so that the pulse does not reach the synapse with the intended intensity.

- In some diseases, such as Multiple Sclerosis (MS) and Amyotrophic Lateral Sclerosis (ALS), the myelin is lost, causing paralysis and loss of sensory and motor function.

A number of chemicals can cause demyelination:

- Diphtheria toxin causes loss of myelin by interfering with the production of protein by the Schwann cells that produce and maintain myelin in the PNS.
- Triethyltin (used as a biocide, preservative, and polymer stabilizer) interrupts the myelin sheath around peripheral nerves.
- Lead causes loss of myelin primarily around peripheral motor axons.

C. Interference with Chemical Neurotransmission

Synaptic dysfunction is a common mechanism for the toxicity of a wide variety of chemicals. There are two types of synapses: *those between two neurons* (axon of one neuron and dendrites of another) and those *between a neuron and a muscle cell or gland*. The basic mechanism for the chemical transmission is the same. The major difference is that the neurotransmitting chemical between a neuron and muscle cell is acetylcholine whereas there are several other types of neurotransmitting chemicals involved between neurons, depending on where in the nervous system the synapse is located.
There are four basic steps involved in neurotransmission at the synapse:

1. Synthesis and storage of neurotransmitter (synaptic knob of axon).
2. Release of the neurotransmitter (synaptic knob with movement across synaptic cleft).
3. Receptor activation (effector membrane).
4. Inactivation of the transmitter (enzyme breaks down neurotransmitter stopping induction of action potential).

The arrival of the action potential at the synaptic knob initiates a series of events culminating in the release of the chemical neurotransmitter from its storage depots in vesicles. After the neurotransmitter diffuses across the synaptic cleft, it complexes with a receptor (membrane-bound macromolecule) on the post-synaptic side. This binding causes an ion channel to open, changing the membrane potential of the post-synaptic neuron or muscle or gland. This starts the process of impulse formation or action potential in the next neuron or receptor cell. However, unless this receptor-transmitter complex is inactivated, the channel remains open with continued pulsing. Thus, the transmitter action must be terminated. Specific enzymes that can break the bond and return the receptor-membrane to its resting state do this.

Drugs and environmental chemicals can interact at specific points in this process to change the neurotransmission. Depending on where and how the xenobiotics act, the result may be either an increase or a decrease in neurotransmission. Many drugs (such as tranquilizers, sedatives, stimulants, beta-blockers) are used to correct imbalances to neurotransmissions (such as occurs in depression, anxiety, and cardiac muscular weakness). The mode of action of some analgesics is to block receptors, which prevent transmission of pain sensations to the brain.

Exposure to environmental chemicals that can perturb neurotransmission is a very important area of toxicology. Generally, neurotoxins affecting neurotransmission act to:

1. Increase or decrease the release of a neurotransmitter at the presynaptic membrane.
2. Block receptors at the postsynaptic membrane.
3. Modify the inactivation of the neurotransmitter.

This is a list of only a few examples of neurotoxins to show the range of mechanisms:

- α-Bungarotoxin (a potent venom of elapid snakes) prevents the release of neurotransmitters.
- Scorpion venom potentiates the release of a neurotransmitter (acetylcholine).
- Black widow spider venom causes an explosive release of neurotransmitters.
- Botulinum toxin blocks the release of acetylcholine at neuromuscular junctions.
- Atropine blocks acetylcholine receptors.
- Strychnine inhibits the neurotransmitter glycine at postsynaptic sites resulting in an increased level of neuronal excitability in the CNS.
- Nicotine binds to certain cholinergic receptors.

A particularly important type of neurotoxicity is the inhibition of acetylcholinesterase. The specific function of acetylcholinesterase is to stop the action of acetylcholine once it has bound to a receptor and initiated the action potential in the second nerve or at the neuro-muscular or glandular junction. If the acetylcholine-receptor complex is not inactivated, continual stimulation will result leading to paralysis and death.

- Many commonly used chemicals, especially organophosphate and carbamate pesticides, poison mammals by this mechanism.
- The major military nerve gases are also cholinesterase inhibitors.

Acetylcholine is a common neurotransmitter. It is responsible for transmission at all neuromuscular and glandular junctions as well as many synapses within the CNS.

Events Involved in a Typical Cholinergic SynapseThe complexity of the sequence of events that takes place at a typical cholinergic synapse is indicated below:

Step	Events
1	• Electrical impulse arrives at synaptic knob and depolarizes the bulb and presynaptic membrane. • Synaptic vesicles release acetylcholine *(ACh)*.
2	• Calcium ions enter the synaptic knob's cytoplasm. • Synaptic vesicles release ACh.
3	• ACh release stops as calcium ions are removed from the synaptic knob cytoplasm. • The released ACh diffuses across synaptic cleft and binds to receptors on the post-synaptic membrane. • The chemically-regulated receptors cause a graded depolarization on the postsynaptic surface, which then is transmitted on down the axon or into the effector cell.
4	• ACh is broken down by acetylcholinesterase into choline and acetate at the receptor site on the postsynaptic membrane. • Choline is then reabsorbed from the synaptic cleft and is available for resynthesis into more ACh and stored by the synaptic vesicles for future use.

Table 1. Events that take place at a typical cholinergic synapse

The nervous system is the most complex system of the body. There are still many gaps in understanding how many neurotoxins act, yet research is discovering their possible effects on the body's structures and functions. It is important to understand that the most potent toxins (on a weight basis) are neurotoxins with extremely minute amounts sufficient to cause death.

Knowledge Check

1) The two fundamental anatomical divisions of the nervous system are the:
a) Cerebrum and cerebellum
b) Central nervous system and peripheral nervous system
c) Brain and spinal cord

Answer available at the end of this section.

2) The two major categories of cells found in the nervous system are:

a) Neurons and glial cells
b) Astrocytes and microglia
c) Schwann cells and oligodendrocytes

Answer available at the end of this section.

3) The propagation of an electrical impulse (action potential) down an axon consists of:

a) The transmission of the action potential by chemical neurotransmitters
b) The movement of sodium ions from the dendrite to the axon
c) A continuous series of opening and closing of sodium-potassium channels and pumps

Answer available at the end of this section.

4) The type of neuron that relays information from the CNS to other organs is a:

a) Motor neuron
b) Sensory neuron
c) Interneuron

Answer available at the end of this section.

5) The primary cause of death to neurons and glial cells is:

a) Interference with chemical transmission
b) Interference with electrical transmission
c) Anoxia

Answer available at the end of this section.

6) A major mechanism that prevents the action potential (impulse) from moving down an axon is:

a) Blockage or interference with movement of sodium and potassium ions in and out of neuron membrane, changing the action potential
b) Excessive release of chemical neurotransmitters
c) Blocking receptors at the post-synaptic membrane

Answer available at the end of this section.

7) What are the two basic types of synapses?

a) Neuro-muscular and neuro-glandular
b) CNS and PNS synapses
c) Those between two neurons and a neuron and effector

Answer available at the end of this section.

Answers

Knowledge Check

Central nervous system and peripheral nervous system - **This is the correct answer.**
The two fundamental anatomical divisions of the nervous system are the Central Nervous System (brain and spinal cord) and the Peripheral Nervous System, which consists of all nerves outside the brain and spinal cord.

Knowledge Check

Neurons and glial cells - **This is the correct answer.**
The two major categories of cells found in the nervous system are neurons and glial cells. Neurons are the functional nerve cells directly responsible for transmission of information to and from the CNS to other areas of the body. Glial cells (also known as neuroglia) provide support to the neural tissue, regulate the environment around the neurons, and protect against foreign invaders.

Knowledge Check

A continuous series of opening and closing of sodium-potassium channels and pumps - **This is the correct answer.**
The propagation of an electrical impulse (action potential) down an axon consists of a continuous series of opening and closing of sodium-potassium channels and pumps. The action potential moves like a wave from one end (dendritic end) to the terminal end of the axon.

Knowledge Check

Motor neuron - **This is the correct answer.**
Motor Neurons (effector neurons) relay information from the CNS to other organs terminating at the effectors.

Knowledge Check

Anoxia - **This is the correct answer.**
The most common cause of death of neurons and glial cells is anoxia, an inadequate oxygen supply to the cells or their inability to utilize oxygen.

Knowledge Check

Blockage or interference with movement of sodium and potassium ions in and out of neuron membrane, changing the action potential - **This is the correct answer.**
Interruption of the propagation of the electrical potential is caused by agents that can block or interfere with the sodium and potassium channels and sodium-potassium pump. This will weaken, slow, or completely interrupt the movement of the electrical potential.

Knowledge Check

Those between two neurons and a neuron and effector - **This is the correct answer.**
The two basic types of synapses are those between two neurons and those between a neuron and effectors, such as muscle cell or gland. The major difference in the two basic types is that the neurotransmitting chemical between a neuron and muscle cell is acetylcholine whereas there are several other types of neurotransmitting chemicals involved between neurons, depending on where in the nervous system the synapse is located.

This page titled 8.4.6.14.4: Neurotoxicity is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 14.4: Neurotoxicity by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.15: INTUITIVE TOXICOLOGY AND RISK COMMUNICATION

Learning Objectives

After completing this lesson, you will be able to:

- Describe aspects of Intuitive Toxicology.
- Describe aspects of Risk Communication.

In this section...

Topics include:

8.4.6.15.1: Intuitive Toxicology

8.4.6.15.2: Risk Communication

Section 15: Key Points

What We've Covered

This section made the following main points:

- Intuitive toxicology:
  - Combines the intuitive elements of expert and public risk judgments involved with exposure assessment, toxicology, and risk assessment.
  - Asks if toxicologists agree or disagree that an animal's reactions to a chemical reliably predict human reactions.
  - Asks if a chemical found in a scientific study to cause cancer in animals can be reasonably assumed to cause cancer in humans.
- When communicating risk, one needs to:
  - Accept and involve the public as a legitimate partner.
  - Listen to the public's specific concerns.
  - Be honest, frank, and open.
  - Coordinate and collaborate with other credible sources.
  - Meet the needs of the media.
  - Speak clearly and with compassion.
  - Plan carefully and evaluate your efforts.
  - Acknowledge and describe uncertainty, such as any data gaps or issues relating to methodology.

This page titled 8.4.6.15: Intuitive Toxicology and Risk Communication is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 15: Intuitive Toxicology and Risk Communication by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.15.1: Intuitive Toxicology

Intuitive Toxicology

Intuitive can be defined as "using or based on what one feels to be true even without conscious reasoning." Humans have always been intuitive toxicologists via the use of the senses of sight, taste, and smell to try to detect harmful or unsafe food, water, and air.

Intuition and Professional Judgment in Toxicology

Even well-established scientific approaches used in human risk assessment depend on extrapolations and judgments when assessing human, animal and other toxicology data. This led to the study of **intuitive toxicology**—the intuitive elements of expert and public risk judgments involved with exposure assessment, toxicology, and risk assessment.

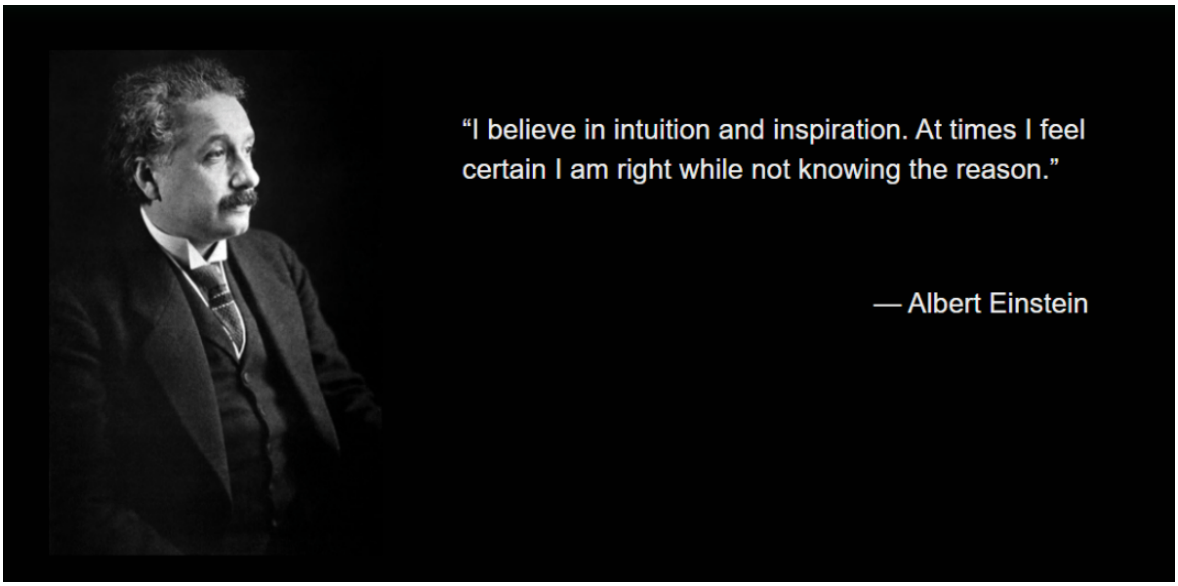


Figure 8.4.6.15.1.1. Albert Einstein understood the importance of intuition along with knowledge and experience (Image Source: Wikimedia Commons, Public Domain - **original image**)

Studies of Intuitive Toxicology

The studies of intuitive toxicology have surveyed toxicologists (for example, members of the Society of Toxicology) and others about a wide range of attitudes, beliefs, and perceptions regarding risks from chemicals. These have included basic concepts, assumptions, and interpretations related to the effects of chemical concentration, dose, and exposure on risk, and the value of animal studies for predicting the effects of chemicals on humans.

Two questions that have been studied repeatedly in intuitive toxicology are:

1. Would you agree or disagree that the way an animal reacts to a chemical is a reliable predictor of how a human would react to it?
2. If a scientific study produces evidence that a chemical causes cancer in animals, can we can then be reasonably sure that the chemical will cause cancer in humans?

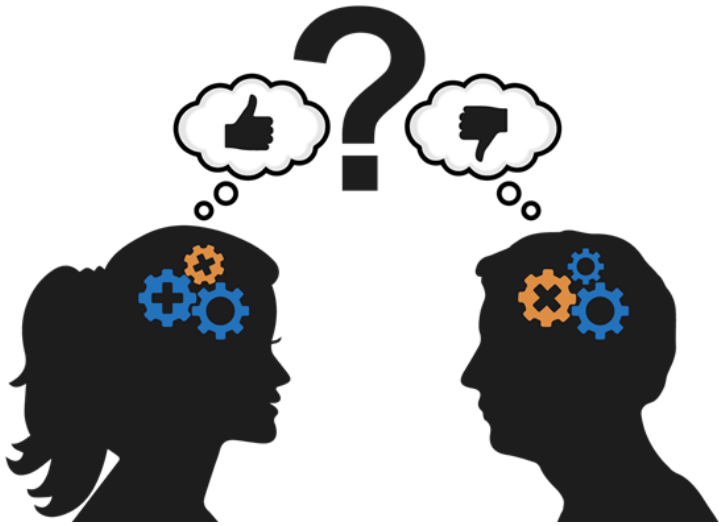


Figure 8.4.6.15.1.2. Understanding and application of toxicology sometimes involves elements of expert judgment and intuition (Image Source: Adapted from iStock Photos, ©)

Examples of Findings from Intuitive Toxicology

Examples of the findings from studies of intuitive toxicology in the United States, Canada, and the United Kingdom include:

- The public is more likely than toxicologists to think chemicals pose greater risks.
- The public finds it difficult to understand the concept of dose–response relationships.
- Much disagreement between toxicologists about how to interpret various results.
- Technical judgments of toxicologists were also found to be associated with factors such as their type of employment (for example, academia, government, or industry), gender, and age.

These types of studies have identified misconceptions that experts should try to clarify in interactions with the public. The results also suggest that disagreement among experts, especially as perceived by the news media and the public, can play a key role in controversies over toxicology-related risks.

Learn more about Intuitive Toxicology

- **Perception of risk.**
  Slovic P.
  Science. 1987 Apr 17;236(4799):280-5.
- **Risk perception: it's personal.**
  Brown VJ. Environ Health Perspect. 2014 Oct;122(10):A276-9. doi: 10.1289/ehp.122-A276
- **Intuitive toxicology: expert and lay judgments of chemical risks.** Neil N, Malmfors T, Slovic P. Toxicol Pathol. 1994 Mar-Apr;22(2):198-201.
- **Judgments of chemical risks: comparisons among senior managers, toxicologists, and the public.** Mertz CK, Slovic P, Purchase IF. Risk Anal. 1998 Aug;18(4):391-404.

Knowledge Check

1) Intuitive toxicology studies show that:

a) All toxicologists think the same
b) Members of the public usually think like toxicologists
c) There can be meaningful differences among toxicologists in how they look at the same set of toxicology study results
d) Intuitive toxicology is not important to consider in communication efforts

Answer available at the end of this section.

2) Which of the following statements is correct?

a) The concept of dose-response relationships is easily understood by the public
b) The public and toxicologists tend to agree about the risks of chemicals
c) Technical judgments of toxicologists have been found to not be associated with factors such as their type of employment (for example, academia, government of industry, gender, and age)
d) Technical judgments of toxicologists have been found to be associated with factors such as their type of employment (for example, academia, government of industry, gender, and age)

Answer available at the end of this section.

Answers

Knowledge Check

There can be meaningful differences among toxicologists in how they look at the same set of toxicology study results - **This is the correct answer.**
Intuitive toxicology studies show that there can be meaningful differences among toxicologists in how they look at the same set of toxicology study results.

Knowledge Check

Technical judgments of toxicologists have been found to be associated with factors such as their type of employment (for example, academia, government of industry, gender, and age) - **This is the correct answer.**
Technical judgments of toxicologists have been found to be associated with factors such as their type of employment (for example, academia, government of industry, gender, and age).

This page titled 8.4.6.15.1: Intuitive Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 15.1: Intuitive Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.15.2: Risk Communication

Risk Communication

Risk communication is the exchange of information about risks.

Rules for Communicating Risk

Much information about how risks could be communicated is available. Some key points about risk communication are identified in the "**Seven Cardinal Rules for Communicating Risk**" from the work of Dr. Vincent Covello and used by U.S. EPA and others:

- Accept and involve the public as a legitimate partner.
- Listen to the public's specific concerns.
- Be honest, frank, and open.
- Coordinate and collaborate with other credible sources.
- Meet the needs of the media.
- Speak clearly and with compassion.
- Plan carefully and evaluate your efforts.

Lessons Learned About Communicating Risk

Some of the lessons that organizations have learned about communicating exposure and health effects information to study subjects, the community, and the public include:

- Communication is not a "cheap add-on" to a study. It must be planned and budgeted at the start. The researcher must know the community and establish relationships early in the project. Communications should be tailored to the project and should contain what people really need to know. The study results that are most significant for the community should be emphasized. Moreover, results should be communicated in a format and a manner that subjects can readily understand. Researchers should evaluate and learn from each study.
- Ignoring communication may lead to legal problems.
- Communicating risk is part of societal accountability.
- Principles and guidelines, including proper terminology, are needed.
- Guidelines should be enforceable.
- Communication requires resources.
- It should be determined early in the project who has control of the release of results, and whether results will be presented in stages or all at once.
- A professional's credibility is at risk when decisions about communication of study results are being made.
- Mechanisms may be needed to proactively consider communication.
- The role of Institutional Review Boards (IRBs) must be considered in developing communication.

Learn more about communicating risk

Lessons Learned from a Crisis and Emergency

Six principles of effective crisis and risk communication are:

1. Be first
2. Be right
3. Be credible
4. Express empathy
5. Promote action
6. Show respect

"The CDC acknowledges that less-than-clear communication about what was known and not known about the possible health effects of the Elk River spill may have affected communities' trust in government." **Learn more**



Figure 8.4.6.15.2.1. Charleston, West Virginia viewed from across the Kanawha River, of which the Elk River is a tributary
(Image Source: iStock Photos, ©)

Uncertainty

Uncertainty is defined as "imperfect knowledge concerning the present or future state of an organism, system, or (sub)population under consideration." In other sources (**EFSA, 2018**), "uncertainty is defined as referring to all types of limitations in the knowledge available to assessors at the time an assessment is conducted and within the time and resources available for the assessment." There are different types of uncertainty, some quantifiable and others not, some reducible and others not.

Due to lack of knowledge, variability adds to the overall uncertainty. Ignoring uncertainty may lead to incomplete risk assessments, poor decision-making, and poor risk communication (**European Commission, 2015**). The degree to which characterization of uncertainty (and variability) is needed will depend on the risk assessment and risk management contexts as determined in the questions asked (problem formulation).



Figure 8.4.6.15.2.1. Uncertainty
(Image Source: Adapted from iStock Photos, ©)

Uncertainty should be acknowledged and described, such as outlining any data gaps or issues relating to methodology. What is being done to address the areas of uncertainty is also important. In its guideline **When Food Is Cooking Up a Storm**, the European Food Safety Authority provides a framework to assist decision-making about appropriate communications approaches in a wide variety of situations that can occur when assessing and communicating on risks related to food safety in Europe. It is directed towards governmental agencies that regulate the food sector.

EFSA has developed a harmonized approach to assessing and taking account of uncertainties in food safety, and animal and plant health. In 2018, EFSA published its **Guidance on Uncertainty Analysis in Scientific Assessment** which offers a diverse toolbox of scientific methods and technical tools for uncertainty analysis. It is sufficiently flexible to be implemented in such diverse areas as plant pests, microbiological hazards and chemical substances. Further, in a separate document EFSA (2018) **describes the principles and methods behind its guidance**. It provides a flexible framework within which different methods may be selected, according to the needs of each risk assessment. It is recommended that assessors should systematically identify sources of uncertainty, checking each part of their assessment to minimize the risk of overlooking important uncertainties.

Communicating Uncertainty in Risk Assessments and in Risk Management



Figure 8.4.6.15.2.3. Uncertainty
(Image Source: Adapted from iStock Photos, ©)

By late 2018, EFSA is expected to have a **practical guidance for communication specialists** on how to communicate the results of uncertainty analysis to different target audiences, including the public. The document aims to help EFSA to communicate scientific uncertainties to its different audiences by using more accessible language tailored to their needs.

Learn more about uncertainty and communicating about it

- U.S. FDA: "Elemental Analysis Manual for Food and Related Products" - **https://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm2006954.htm**
- U.S. FDA: "Best Practices in Risk Communication and Communicating Uncertainty: Applications to FDA-Regulated Products" - **https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/RiskCommunicationAdvisoryCommittee/UCM486965.pdf**
- EFSA: "**Risk, Precautionary Principle and Uncertainty**. European Commission **Uncertainty assessment**"
- EFSA: Workshop on the trial of the EFSA guidance document on uncertainty analysis in scientific assessments, 22 - 23 June 2017, Parma, Italy. **https://www.efsa.europa.eu/en/supporting/pub/en-1313**
- EFSA: Member State multilingual online survey on communicating uncertainty to different target audiences. European Food Safety Authority - **http://www.efsa.europa.eu/en/supporting/pub/en-1413**
- EFSA: "**Communicating uncertainty: some applications of indicators and their validity**" [PDF]

- EFSA: "When Food Is Cooking Up a Storm – Proven Recipes for Risk Communications"

Knowledge Check

1) Which of the following is not true about communicating risk to a community about exposures and health effects?

a) Results should be communicated in a format and a manner that subjects can readily understand
b) Results do not need to be communicated in a format and a manner that subjects can readily understand
c) Communications should be tailored to the project and should contain what people really need to know
d) Researchers and others can learn from studying good and bad risk communication efforts

Answer available at the end of this section.

2) According to the European Commission, ignoring uncertainty may lead to:

a) Great decision-making
b) Poor decision-making
c) Effective risk communication
d) Use of the most accurate knowledge available

Answer available at the end of this section.

Answers

Knowledge Check

Results do not need to be communicated in a format and a manner that subjects can readily understand - **This is the correct answer.**

Knowledge Check

Poor decision-making - **This is the correct answer.**
According to the European Commission, ignoring uncertainty may lead to poor decision-making.

This page titled 8.4.6.15.2: Risk Communication is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 15.2: Risk Communication by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.16: ENVIRONMENTAL TOXICOLOGY, ENVIRONMENTAL HEALTH, AND ONE HEALTH

Learning Objectives

After completing this lesson, you will be able to:

- Define Environmental Toxicology.
- Describe the differences between environmental toxicology and environmental health.
- Define One Health.

In this section...

Topics include:

8.4.6.16.1: Environmental Toxicology

8.4.6.16.2: Environmental Health

8.4.6.16.3: One Health

Section 16: Key Points

What We've Covered

This section made the following main points:

- Environmental Toxicology:
  - Is the multidisciplinary study of the effects of manmade and natural chemicals on health and the environment.
  - Includes work in academia, companies, government agencies, and elsewhere.
- Environmental Health:
  - Is a branch of public health focused on the relationships between people and their environment.
  - Seeks to advance policies and programs to reduce chemical and other environmental exposures in air, water, soil, and food.
  - Saves lives, saves money, and protects your future.
- One Health:
  - Is a global concept and strategy recognizing the links between the health of people, animals, and the environment.
  - Examines changes in the interactions between people, animals, and the environment and how these changes impact human and animal health.

This page titled 8.4.6.16: Environmental Toxicology, Environmental Health, and One Health is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 16: Environmental Toxicology, Environmental Health, and One Health by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.16.1: Environmental Toxicology

Environmental Toxicology

Environmental Toxicology is the multidisciplinary study of the effects of manmade and natural chemicals on health and the environment. This includes the study of the effects of chemicals on organisms in their natural environments and in the ecosystems to which they belong.

Branches of Environmental Toxicology

Environmental Toxicology covers a wide range of interdisciplinary studies, as illustrated in Figure 1:

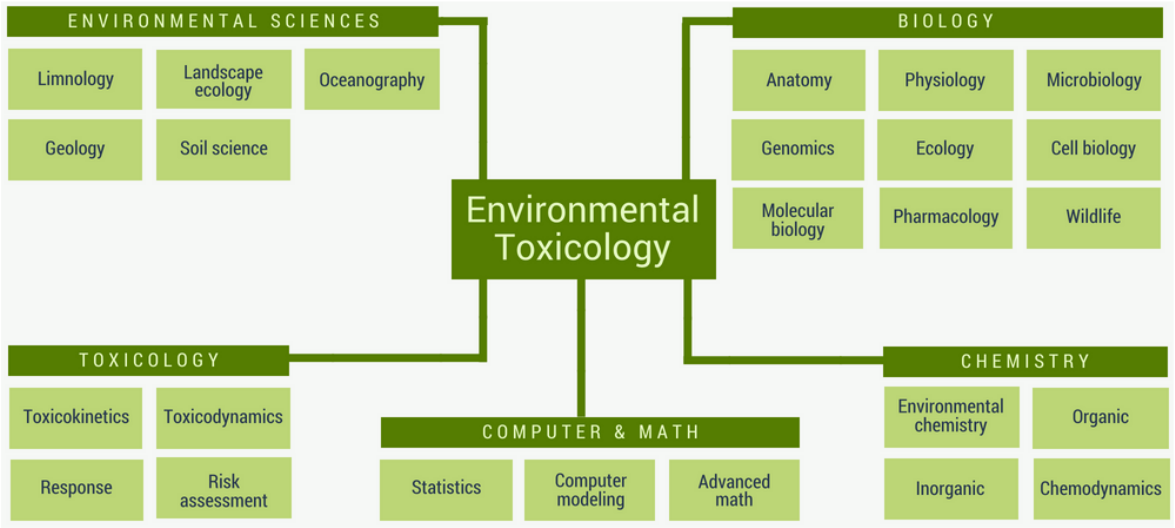


Figure 8.4.6.16.1.1. Environmental Toxicology interdisciplinary core (not comprehensive)
(Image Source: Adapted from Wikipedia under the Creative Commons Attribution-ShareAlike 3.0 License)

Scope of Work and Study

Environmental toxicologists work in academia, companies, government agencies, and elsewhere. The work can include laboratory studies, computer modeling, and work "in the field." It is not unusual for an environmental toxicologist to also have training in other areas—for example, public health, environmental chemistry, and pharmacology.
Some examples of what Environmental Toxicologists study include:

- The effects of a chemical or other substance at various concentrations on various species.
- Whether a chemical or other substance can bioaccumulate (increase over time) in animals or other organisms. This is important for human exposures if the bioaccumulation occurs in animals that are part of the human food chain, such as fish.
- Emerging issues such as the study of the sources and effects of microplastics that could become part of the human food chain.

Learn more about microplastics

- U.S. EPA "Toxicological Threats of Plastic" - https://www.epa.gov/trash-free-waters/toxicological-threats-plastic

- GreenFacts "Marine Litter" - https://www.greenfacts.org/en/marine-litter/index.htm#1
- European Chemicals Agency "Microplastics" - https://echa.europa.eu/hot-topics/microplastics



Figure 8.4.6.16.1.2. Microplastics are plastic debris less than five millimeters in length
(Image Source: National Ocean Service, National Oceanic and Atmospheric Administration. Original image)

Another emerging global issue is the health of bees. In the news in recent years are terms like the Colony Collapse Disorder (CCD), pesticides like the neonicotinic pesticides (also called neonicotinoids), and parasites that only reproduce in bee colonies. About 75% of all flowering plants rely on animal pollinators and about one-third of our food production is dependent on animal pollinators. The general declining health of honeybees and other bees is thought to be related to complex interactions among multiple stressors, including pesticides and parasites, and stressors like poor nutrition due to declining foraging habitats, bee management practices, and a lack of genetic diversity.

Neonicotinoids have been a focus of international attention and their mode of action is on the central nervous system of insects. Neonicotinoids are highly toxic to honeybees and also native bees like bumble bees and blue orchard bees, and sub-lethal levels can affect foraging and the ability to reproduce. Further, neonicotinoids can be persistent in the environment, and can be absorbed by plants and found in pollen and nectar.

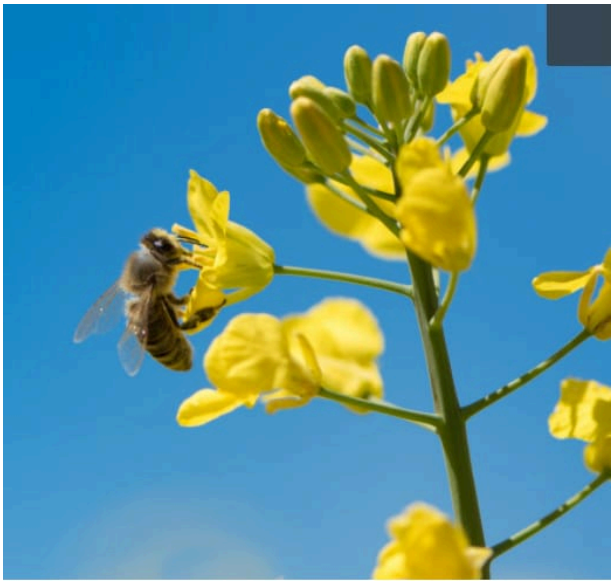


Figure 8.4.6.16.1.3. A honeybee gathers pollen from a flower
(Image Source: iStock Photos, ©)

Learn more about pollinators

- U.S. Fish and Wildlife Service "Pollinators" - https://www.fws.gov/Pollinators/
- U.S. Department of Agriculture, Forest Service "Pollinators" - https://www.fs.fed.us/wildflowers/pollinators/index.shtml
- U.S. EPA "Pollinator Health Concerns" - https://www.epa.gov/pollinator-protection/pollinator-health-concerns
- European Commission "The EU Approach to Tackle Pollinator Decline" - http://ec.europa.eu/environment/nature/

conservation/species/pollinators/index_en.htm

- European Food Safety Authority "Bees Under Siege: Making Sense of Multiple Stressors" - https://www.youtube.com/watch?v=ZVKJNc0tBDM&feature=youtu.be&list=PLGDvgn1aAEEbWUxOzO8FjrtBQpLiC5DOB
- Xerces Society for Invertebrate Conservation "Neonicotinoids and Bees" - http://xerces.org/neonicotinoids-and-bees/

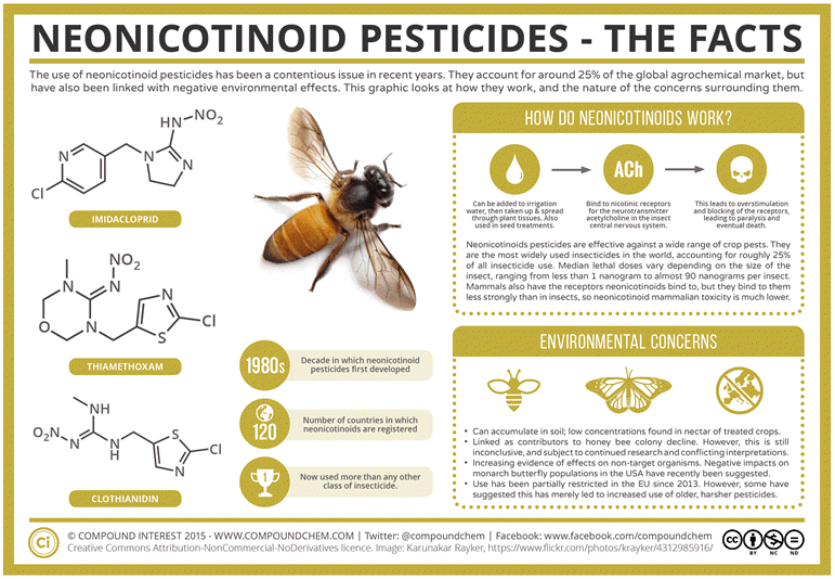


Figure 8.4.6.16.1.4. Neonicotinoid Pesticides
*(Image Source: Compound Interest, © 2015 - used under Creative Commons Attribution-NonCommercial-NoDerivatives license. Bee image captured by Karunakar Rayker. **Original Image**. Original Infographic)*

Knowledge Check

1) The interdisciplinary core ("branches") of environmental toxicology includes:

a) Environmental sciences; physics; toxicology; chemistry; biology
b) Environmental sciences; engineering; toxicology; biology; computer and math
c) Environmental sciences; computer and math; toxicology; biology; law
d) Environmental sciences; biology; toxicology; chemistry; computer and math

Answer available at the end of this section.

2) Which issues in environmental toxicology relate to the human food supply?

a) Bioaccumulation of substances in fish
b) Effects of neonicotinic pesticides on bees
c) Both the bioaccumulation of substances in fish and effects of neonicotinic pesticides on bees

Answer available at the end of this section.

Answers

Knowledge Check

Environmental sciences; biology; toxicology; chemistry; computer and math - **This is the correct answer.**
The interdisciplinary core ("branches") of environmental toxicology includes environmental sciences; biology; toxicology; chemistry; computer and math.

Knowledge Check

Both the bioaccumulation of substances in fish and effects of neonicotinic pesticides on bees - **This is the correct answer.**

Both the bioaccumulation of substances in fish and effects of neonicotinic pesticides on bees relate to the human food supply.

This page titled 8.4.6.16.1: Environmental Toxicology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 16.1: Environmental Toxicology by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.16.2: Environmental Health

Environmental Health

Environmental health is a branch of public health. It focuses on the relationships between people and their environment, and promotes human health and well-being. Further, it fosters healthy and safe communities and is a key part of any comprehensive public health system. The field works to advance policies and programs to reduce chemical and other environmental exposures in air, water, soil, and food to protect people and provide communities with an healthier environments.

Toxicology vs. Environmental Health

The two fields are closely connected, with a large intersection.
The terminology is loose. For example, is environmental health about the health of the environment and/or the effect of the environment on health?

Scope of Work and Study

Environmental health specialists identify and evaluate environmental hazards and their sources. They also limit exposures to physical, chemical, and biological agents in air, water, soil, food, and other environmental media or settings that may adversely affect human health (**Source**).



Figure 8.4.6.16.2.1. Environmental health saves lives, saves money, and protects your future
(Image Source: National Environmental Health Association (NEHA), © 2016. **Original image**)

National Center for Environmental Health (NCEH)

The CDC National Center for Environmental Health (NCEH) plans, directs, and coordinates a program to protect the American people from environmental hazards. The NCEH seeks to prevent premature death, avoidable illness, and disability caused by non-infectious, non-occupational environmental and related factors. One focus is on safeguarding the health of venerable populations, such as children, the elderly, and people with disabilities, from certain environmental hazards.

Learn more about environmental health

- U.S. CDC National Center for Environmental Health - **https://www.cdc.gov/nceh/**
- American Public Health Association (APHA): "Environmental Health" - **https://www.apha.org/topics-and-issues/environmental-health**
- National Environmental Health Association (NEHA): "Definitions of Environmental Health" - **https://www.neha.org/about-neha/definitions-environmental-health**

Knowledge Check

1) Which of the following are important environmental health issues?

a) Radon in air
b) Lead contamination in drinking water
c) Foodborne illness from contaminated food
d) All of these are important for human health

Answer available at the end of this section.

Answers

Knowledge Check

All of these are important for human health

This page titled 8.4.6.16.2: Environmental Health is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 16.2: Environmental Health by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.16.3: One Health

One Health

One Health is a worldwide concept and strategy recognizing that the health of people, animals, and the environment are all connected. For example, the Centers for Disease Control and Prevention (CDC) works with physicians, veterinarians, ecologists, and many others to monitor and control public health threats and to learn about how diseases spread among people, animals, and the environment.

Link Between Human, Animal, and Environmental Health

One Health is important at the local, regional, national, and global levels, and there are many examples of its importance. One example of how human, animal, and environmental health are linked involves bacteria, cows, farms, food, lettuce, and humans:



Cows graze next to a lettuce field. Cows can carry *E. coli* but still look healthy.



E. coli from cow manure in the nearby farm can contaminate the lettuce field.



People eat conta
infected with *E. c*
death can result.

Figure 8.4.6.16.3.1. Human, animal, and environmental health are linked
(Image Source: CDC - **original image**)

Another example of One Health involving animals and humans is the shared susceptibility to some diseases and environmental hazards. Animals can serve as early warning signs of potential human illness. An example is birds dying from West Nile virus before people in the same areas get sick from exposures to this virus.

Factors that Affect Human and Animal Health

Some interactions between people, animals, and the environment have changed in recent years and these changes have impacted animal and human health.

Factor (Cause)	Change (Effect)
Human populations are growing and expanding into new geographic areas.	More people live in close contact with wild and domestic animals. Close contact provides more opportunities for diseases to pass between animals and people.
The earth has experienced changes in climate and land use, such as deforestation and intensive farming practices.	Disruptions in environmental conditions and habitats provide new opportunities for diseases to pass to animals.
International travel and trade have increased.	Diseases can spread quickly across the globe.

Table 8.4.6.16.3.1. Factors that affect human and animal health
(Source: **CDC: One Health Basics**)

Learn more about One Health

- U.S. CDC: "One Health: From Concept to Action" - **https://youtu.be/TG0pduAYESA**
- U.S. CDC: "One Health Basics" - **https://www.cdc.gov/onehealth/basics/index.html**
- World Organization for Animal Health (OIE): "One Health: By Protecting Animals, We Preserve Our Future [PDF] - **http://www.oie.int/fileadmin/Home/eng/Media_Center/img/Infographies/A4-EN-WEB.pdf**
- U.S. FDA: "One Health: It's for All of Us" - **https://www.fda.gov/AnimalVeterinary/ResourcesforYou/AnimalHealthLiteracy/ucm278139.htm**

Knowledge Check

1) One Health is a concept and strategy recognizing that the health of _____, _____, and _____ are all connected.

a) People, plants, the environment
b) People, animals, the environment
c) People, animals, microbes

Answer available at the end of this section.

2) Which of the following factors that affect human and animal health is not correct?

a) Fewer people in recent years live in close contact with wild and domestic animals
b) Disruptions in environmental conditions and habitats provide new opportunities for diseases to pass to animals
c) International travel and trade have increased, and diseases can spread quickly across the globe

Answer available at the end of this section.

Answers

Knowledge Check

People, animals, the environment - **This is the correct answer.**
One Health is a concept and strategy recognizing that the health of people, animals, and the environment are all connected.

Knowledge Check

Fewer people in recent years live in close contact with wild and domestic animals - **This is the correct answer.**
The statement "Fewer people in recent years live in close contact with wild and domestic animals" is incorrect.

This page titled 8.4.6.16.3: One Health is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- 16.3: One Health by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.4.6.17: Conclusion

Thank you for completing ToxTutor! We trust it has given you a strong foundation of understanding in this important area of science.
In conclusion, we have now covered the following topics throughout this course:

1. Introduction to Toxicology
2. Dose and Dose Response
3. Toxic Effects
4. Interactions
5. Toxicity Testing Methods
6. Risk Assessment
7. Exposure Standards and Guidelines
8. Basic Physiology
9. Introduction to Toxicokinetics
10. Absorption
11. Distribution
12. Biotransformation
13. Excretion
14. Cellular Toxicology
15. Intuitive Toxicology and Risk Communication
16. Environmental Toxicology, Environmental Health, and One Health

Contact Information

If you would like additional information about toxicology and environmental health, or if you have any questions or comments, please contact us at **toxmsdt@gmail.com**

Credits

ToxTutor was adopted from the **U.S National Library of Medicine** in 2021. More Information can be seen in the ToxTutor **Bibliography**.



This page titled 8.4.6.17: Conclusion is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by ToxMSDT Online component via source content that was edited to the style and standards of the LibreTexts platform.

- Section 17: Conclusion by ToxMSDT Online component is licensed CC BY-NC 4.0. Original source: http://www.toxmsdt.com.

8.5: FOOD TO ENERGY METABOLIC PATHWAYS

8.5.1: Basics of metabolism

8.5.2: Digestion of food

8.5.3: Oxidation of glucose -the glycolysis

8.5.4: Citric acid cycle

8.5.5: Oxidative Phosphorylation

8.5.6: Oxidation of fatty acids

8.5.7: Degradation of amino acids

This page titled 8.5: Food to energy metabolic pathways is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9: Food to energy metabolic pathways by Muhammad Arif Malik is licensed Public Domain.

8.5.1: Basics of metabolism

Learning Objectives

- Define metabolism, its subclasses: catabolism and anabolism, and stages of food catabolism.
- Understand the structure of a typical cell and mitochondrion -the sites of catabolism.
- Understand the basic structural features and functions of some common compounds involved in the catabolism of food, including ATP/ADP, NAD^+/NADH, FAD/$FADH_2$ pairs.
- Understand the basic structural features of enzymes that are involved in their catalytic activity.

What is metabolism

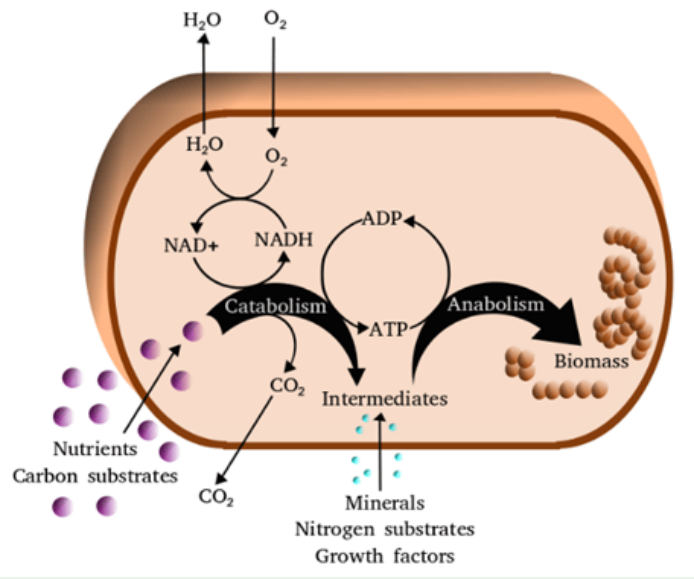
A large number of chemical reactions take place in living things almost all the time. Some of these reactions synthesize the substances living things need from the raw materials available. Usually, these reactions convert simple molecules or compounds to more complex molecules or compounds at the expense of energy. For example, photosynthesis converts carbon dioxide (CO_2) and water (H_2O) into glucose ($C_6H_{12}O_6$) by utilizing energy from sunlight.

$$6\,CO_2 + 6\,H_2O + \text{Energy}\ (2808\,\text{kJ/mol}) \longrightarrow C_6H_{12}O_2$$

Other reactions break down the complex molecules to release the energy needed as heat, to do work, to provide energy for synthetic reactions, to dispose of waste byproducts, etc. For example, the reverse of photosynthesis happens in the digestion of glucose.

$$6\,CO_2 + 6\,H_2O + \text{Energy}\ (2808\,\text{kJ/mol})C_6H_{12}O_2 \longrightarrow 6\,CO_2 + 6\,H_2O + \text{Energy}\ (2808\,\text{kJ/mol})$$

Metabolism



Metabolism is chemical reactions taking place in living things needed to sustain life.

Metabolism can be subdivided into two categories of reactions.

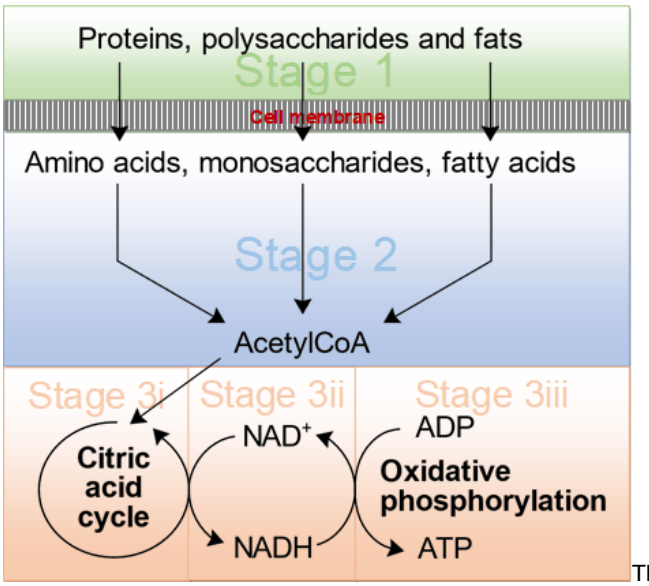
Catabolism and anabolism

- **Catabolism** is a set of chemical reactions in living things that breaks down molecules to release energy or obtain intermediates needed in other reactions.
- **Anabolism**, called biosynthesis, is a set of reactions in living things constructing molecules from smaller units.

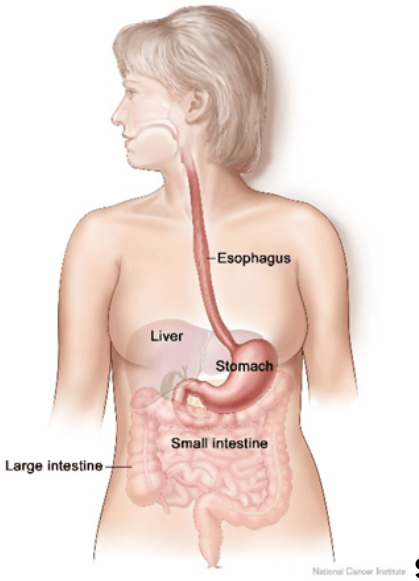
The figure on the right summarizes metabolism and its sub-categories: catabolism and anabolism (*Copyright: Linares-Pastén, J. A. (2018), CCBYSA 4, via Wikimedia Commons*)

Metabolism usually happens through a series of interconnected chemical reactions called **metabolic pathways**. Metabolic pathways have a lot of similarities in different species. It indicates their common origin in the early stages of the evolution of species and their retention due to their efficiency. Some diseases, such as type II diabetes and cancer, disrupt normal metabolism. The difference in metabolic pathways from the normal allows scientists to find therapeutic interventions. Some of those metabolic pathways, particularly the catabolic pathways related to the digestion of food and its conversions to obtain energy, will be described in this chapter.

Stages of catabolism of food



The catabolism of food starts from the digestion of food. It can be divided into three major stages, as illustrated in the figure on the right (*Copyright: modified from Tim Vickers, vectorized by Fvasconcellos, Public domain, via Wikimedia Commons*).



Stage 1

It is mainly the hydrolysis of food molecules in the digestive system, shown in the figure on the left (*Copyright: National Cancer Institute, Public domain, via Wikvia Wikimedians*).

Polysaccharides are hydrolyzed to monosaccharides, fats are hydrolyzed to glycerol and fatty acids, and proteins are hydrolyzed to amino acids. The products of the stage 1 reactions diffuse into the bloodstream and are transported to cells, entering the next phase of food catabolism.

Stage 2

It happens in the cells, which is the degradation of monosaccharides, fatty acids, and amino acids. It yields smaller groups, usually a two-carbon acetyl group or, in the case of some amino acids, four-carbon carboxylate groups, which enter the next stage.

Stage 3

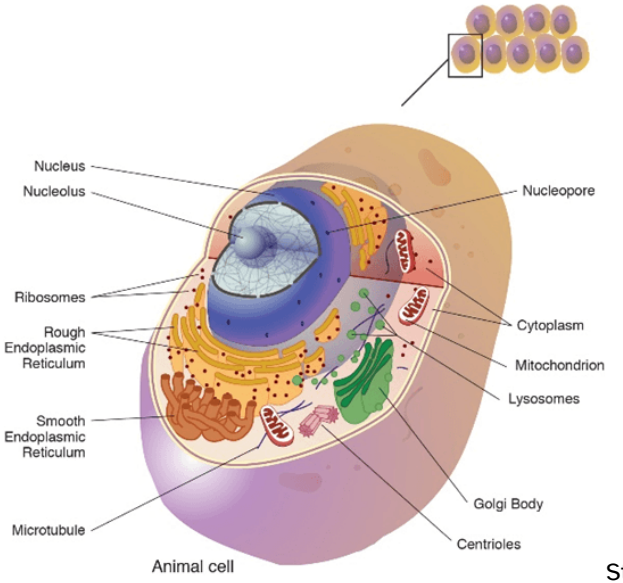
It happens in mitochondria in the cells. Stage 3 can be divided into three sub-stages described below.

Stage 3i: It is the **citric acid cycle**. In this stage, the two-carbon acetyl or four-carbon carboxylate groups are oxidized to carbon dioxide ($\mathrm{CO_2}$) at the expense of the reduction of coenzymes $\mathrm{NAD^+}$ and FAD to NADH and $\mathrm{FADH_2}$.

Stage 3ii: It is **electron transport** where coenzymes NADH and $\mathrm{FADH_2}$ are reduced back to their oxidized forms $\mathrm{NAD^+}$ and FAD at the expense of reduction of oxygen ($\mathrm{O_2}$) into water ($\mathrm{H_2O}$) via electron transport process with release of energy.

Stage 3iii: It is an **oxidative phosphorylation** process where the energy released during the electron transport stage is used to synthesize adenosine triphosphate (ATP), which is a high-energy molecule from adenosine diphosphate (ADP), and phosphate (Pi), which are lower-energy molecules. The energy is temporarily stored in the form of ATP and released wherever it may be needed by the reverse reaction, i.e., conversion of ATP into ADP and Pi.

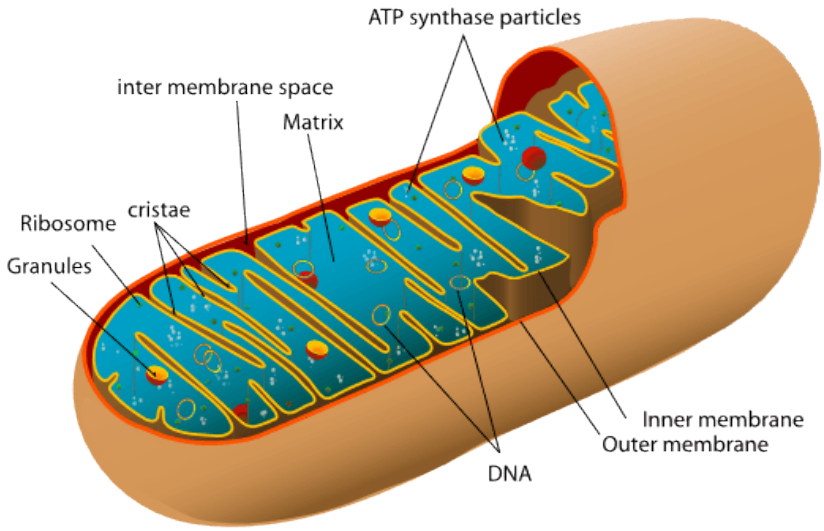
Cell structure related to food catabolism



Stages 2 and 3 of food catabolism happen in the cells. A cell membrane surrounds a typical animal cell. **Organelles** are organized or specialized structures within the cells. A typical animal cell is illustrated in the right figure with some organelles labeled (*Copyright; National Human Genome Research Institute, Public domain*).

The **nucleus** in the cell contains hereditary material, i.e., DNA. The space between the cell membrane and the nucleus is called the **cytoplasm**. The **cytosol** is the fluid part of the cytoplasm containing an aqueous solution of electrolytes and enzymes that catalyze many of the cell's chemical reactions. Within the cytoplasm are organelles that perform specialized functions. For example, **ribosomes** are the sites for protein synthesis, and **mitochondria** are the cell's energy factory where stage 3 of food catabolism occurs.

Mitochondrion



The mitochondrion (plural mitochondria) has an **outer membrane** and **inner membrane** and an **inter-membrane space** between the two, as shown in the figure on the left (*Copyright; LadyofHats, Public domain, via Wikimedia Commons*). The fluid section surrounded by the inner membrane is called the **matrix**. The enzymes that catalyze the chemical reactions of stage 3 of food catabolism are located in the matrix and along the inner membrane. These reactions ultimately convert the food molecules into CO_2, H_2O, and energy. The enzymes that catalyze stage 4 reactions that use this energy to produce high-energy ATP also occur in the matrix along the inner membrane.

The principal compounds involved in the common metabolic pathways

The principal compounds involved in the common metabolic pathways are adenosine triphosphate (ATP), which is the agent for the temporary storage of energy and transfer of phosphate (PO_4^{3-} or Pi) group; nicotinamide adenine dinucleotide (NAD^+) and flavin adenine dinucleotide (FAD) which are agents for the transfer of electrons during the biological oxidation-reduction sections; and coenzyme A (CoA) which is the agent for the transfer of acetyl (CH_3CO-) group. These compounds are described here.

ATP -the energy currency and agent for the transfer of phosphate groups

Adenosine triphosphate (ATP) is a nucleotide composed of adenine base bonded to ribose sugar by an N-glycosidic bond and triphosphate bonded to ribose by an ester bond, as shown in Figure 8.5.1.1. The N-glycoside of adenine with ribose is a nucleoside, i.e., adenosine. When adenosine is bonded to diphosphate, it makes adenosine diphosphate (ADP).

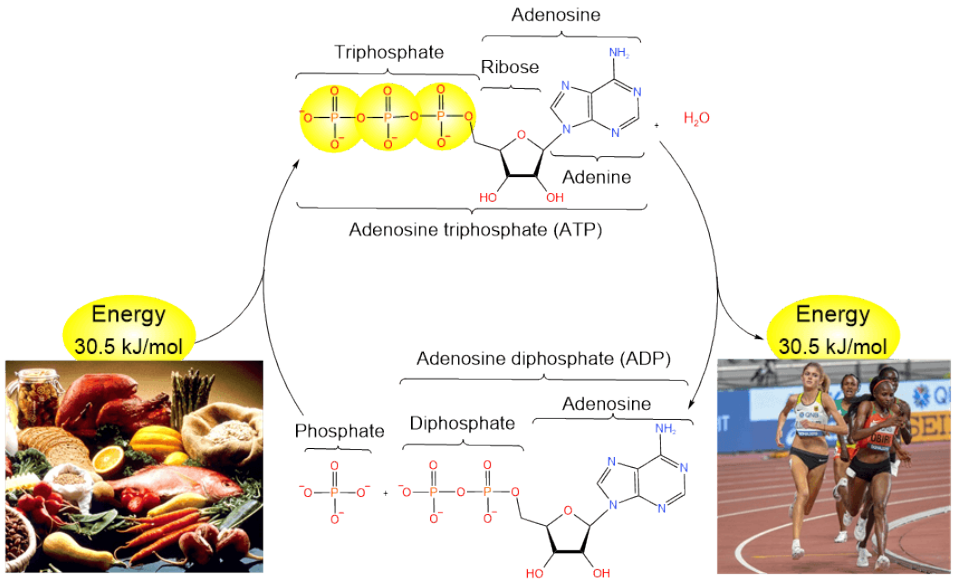


Figure 8.5.1.1: ATP $\rightleftharpoons$ ADP link the energy between energy-releasing relations, such as the catabolism of food, to the energy-consuming process. (Copyright; Public domain)

Hydrolysis of ATP splits one phosphate (PO_4^{3-} or Pi) and convert it to ADP with the release of 30.5 kJ/mol energy.

$$ATP + H_2O \longrightarrow ADP + Pi \ \delta H = -30.5 \text{ kJ/mol}$$

This energy is used in the processes that require energy, such as muscle contraction, nerve signal conduction, and biosynthesis. Catabolism of food releases energy temporarily stored in the form of ATP by reversing the above reaction.

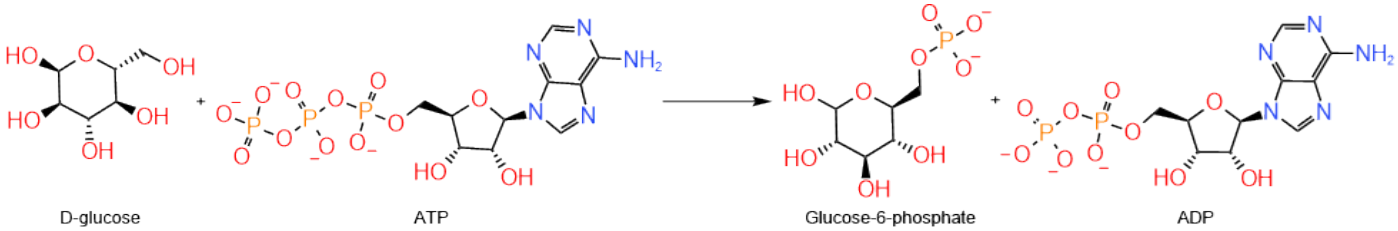
$$ADP + Pi \longrightarrow ATP + H_2O \ \delta H = +30.5 \text{ kJ/mol}$$

The cycle of reactions between ATP and ADP shown in Figure 8.5.1.1 happen rapidly, producing one to two million ATP's per second. An average human body produces ATP about equal to the human body mass per day, but it contains about 1 g of ATP at a time.

ADP can also hydrolyze with water releasing energy and converting to adenosine monophosphate (AMP

$$ADP + H_2O \longrightarrow AMP + Pi \ \delta H = -30.5 \text{ kJ/mol}$$

ATP is also a phosphorylation agent in metabolic reactions. For example, D-glucose is phosphorylated by the reaction shown below.



In summary the main roles of $\mathrm{ATP/ADP}$ pair are:

1. **energy transfer**, i.e., releases energy when ATP converts not ADP + $\mathrm{P_i}$ and absorbs energy in the reverse reaction,
2. **transfer of phosphate** ($\mathrm{P_i}$), i.e., releases phospahte when ATP converts not ADP + $\mathrm{P_i}$ and consumes phosphate in the reverse reaction.

$\mathrm{NAD^+}$ and FAD -agents for the transform of electrons

Oxidation and reduction

Oxidation is:

- loss of electron ($\mathrm{e^-}$),
- gain of oxygen (O), or
- loss of hydrogen (H).

Reduction is:

- gain of electron ($\mathrm{e^-}$),
- loss of oxygen (O), or
- gain of hydrogen (H).

Generally, oxidation reactions release energy, and reduction reactions gain energy.

Clarifications on oxidation-reduction processes

Oxidation and reductions are coupled, i.e., if one reagent is oxidized, another is reduced simultaneously. For example, in the reaction shown below, pyruvate is oxidized by losing hydrogen coenzyme $\mathrm{NAD^+}$ is reduced by gaining hydrogen. So, oxidation-reduction couples are collectively called **redox reactions**.

$$\underbrace{\mathrm{CH_3-\overset{\overset{\mathrm{O}}{\|}}{C}-\overset{\overset{\mathrm{O}}{\|}}{C}-O^-}}_{\text{Pyruvate}} + \mathrm{NADH} + \mathrm{H^+} \xleftrightarrow{\text{Lactate dehydrogenase}} \underbrace{\mathrm{CH_3-\overset{\overset{\mathrm{OH}}{|}}{CH}-\overset{\overset{\mathrm{O}}{\|}}{C}-O^-}}_{\text{Lactate}} + \mathrm{NAD^+}$$

Oxygen may be added to the compound, i.e., make a single or double bond with the substrate, or it may change from a single bond to a double bond in the substrate, e.g., $-\mathrm{C-O}$ to $-\mathrm{C=O}$, both ways it is oxidation. That is, an increase in the bond with oxygen is oxidation.

In metabolic reactions, H is represented as a proton $\mathrm{H^+}$ and an electron $\mathrm{e^-}$. So, the addition of $\mathrm{H^+}$ + $\mathrm{e^-}$ is reduction, and their removal is oxidation. Usually, $\mathrm{2\,H^+}$ + $\mathrm{2\,e^-}$ are transferred from or to a coenzyme in metabolic reactions.

If bonds with hydrogen added or removed are counterbalanced by bonds with oxygen removed or added, the overall reaction is not a redox. For example, an alkene's hydration, shown below, is not a redox reaction.

$$CH_3CH_2CH{=}CHCH_2CH_3 + H_2O \xrightarrow{H_2SO_4} CH_3CH_2CH_2{-}\overset{\overset{OH}{|}}{C}H{-}CH_2CH_3$$

Broader definition of oxidation and reduction

Oxidation and reduction are not limited to making or breaking a bond with oxygen or hydrogen; broader definitions are the following.

Adding a bond with a more electronegative atom is oxidation, and its removal is reduction. The opposite is true for a less electronegative atom, i.e., The addition of a bond with a less electronegative atom is reduction, and its removal is oxidation. For example, conversion of $R{-}\overset{\overset{OH}{|}}{C}H\text{-}R'$ into $R{-}\overset{\overset{O}{||}}{C}{-}R'$ replaces bond on C from less electronegative H with more electronegative O is oxidation. Similarly, conversion of $R{-}\overset{\overset{O}{||}}{C}{-}CH_2R'$ into $R{-}\overset{\overset{O}{||}}{C}{-}SCH_2R'$ replaces bond on C from a C with more electronegative S is oxidation. The reverse of these is reductions.

Nicotinamide adenine dinucleotide (NAD^+)

Nicotinamide adenine dinucleotide is a coenzyme composed of two nucleotides, an adenosine diphosphate (ADP) and a second nucleotide in which the nitrogen base is nicotinamide provided by vitamin niacin. The two nucleotides are linked by diphosphate linkage, as shown in Figure 8.5.1.2. The oxidized form is represented as NAD^+ and its reduced form is represented as NADH. The NAD^+ is reduced to NADH by reacting with two hydrogen ($2H^+ + 2e^-$) leaving one H^+ in the products, as shown in Figure 8.5.1.2 . Oxidation of NADH to NAD^+ is the reverse reaction.

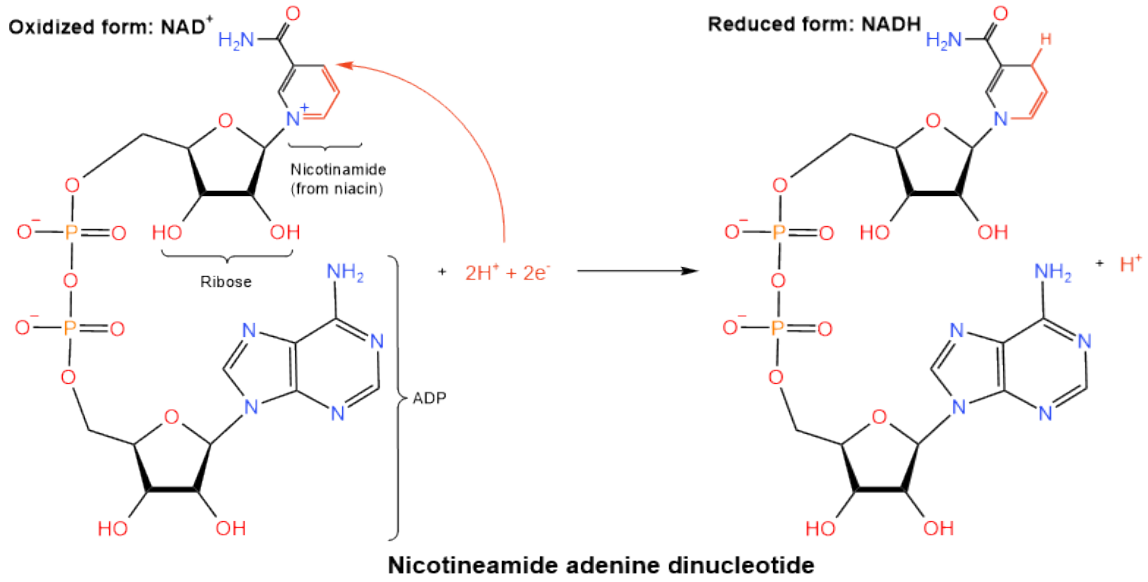
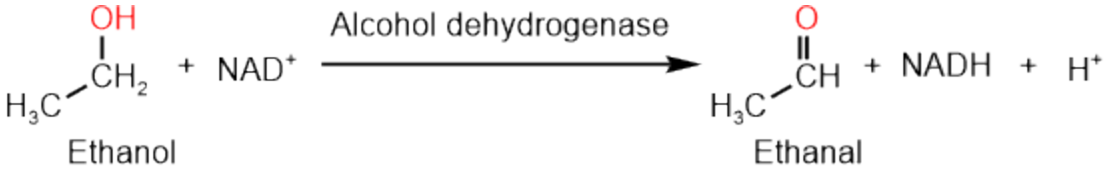


Figure 8.5.1.2: Illustration of the structure of the coenzyme nicotinamide adenine dinucleotide in its oxidized form NAD^+ and its conversion to the reduced form NADH by reacting with two H, i.e., $2H^+ + 2e^-$). (Copyright; Public domain)

The $NADH/NAD^+$ redox is coupled with $\underbrace{R{-}\overset{\overset{OH}{|}}{C}H\text{-}R'}_{\text{Alcohol}}$ / $\underbrace{R{-}\overset{\overset{O}{||}}{C}\text{-}R'}_{\text{Carbonyl}}$, redox reactions or it involves carbonyl (C=O) group in food catabolism .

An example of an oxidation reaction in metabolism is the oxidation of the alcohols (−OH) group to carbonyl (C=O) group. For example, ethanol is oxidized to ethanal in the liver at the expense of reduction of NAD^+ to NADH as shown below.



Flavin adenine dinucleotide (FAD)

Flavin adenine dinucleotide is a coenzyme composed of two nucleotides, an adenosine diphosphate (ADP) and a second nucleotide which is riboflavin (vitamin B_2). Riboflavin is composed of flavin and ribitol, which is a sugar alcohol. The two nucleotides are linked by an ester linkage between ribitol and diphosphate, as shown in Figure 8.5.1.3. The oxidized form is represented as FAD, and its reduced form is represented as $FADH_2$. The oxidation-reduction happens in the N containing rings of the flavin part. The FAD is reduced to $FADH_2$ by reacting with two hydrogen ($2\,H^+ + 2\,e^-$), as shown in Figure 8.5.1.2. Oxidation of $FADH_2$ to FAD is the reverse reaction.

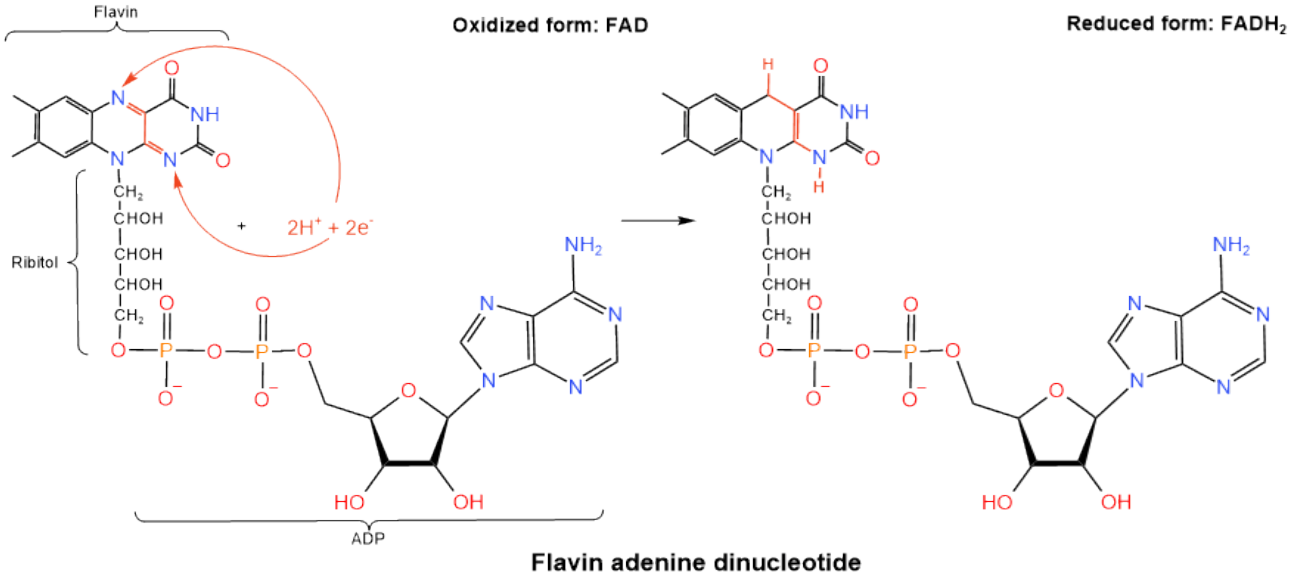
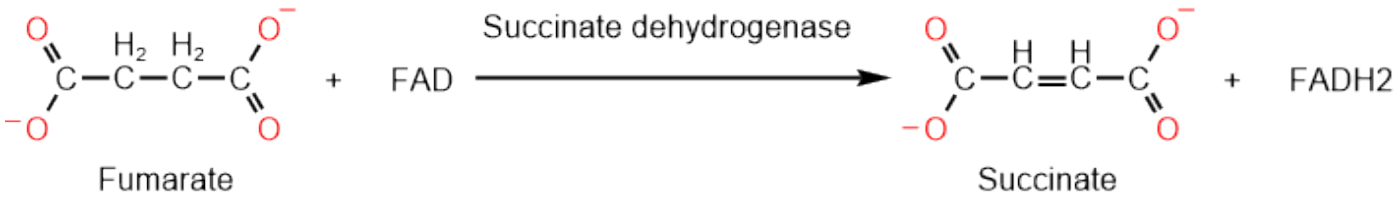


Figure 8.5.1.3: Illustration of the structure of the coenzyme flavin adenine dinucleotide in its oxidized form FAD and its conversion to the reduced form $FADH_2$ by reacting with two H, i.e., $2\,H^+ + 2\,e^-$. (Copyright; Public domain)

The $FADH_2$/FAD redox is coupled with $\underbrace{RR'\overset{H}{\overset{|}{C}}=\overset{H}{\overset{|}{C}}RR'}_{\text{Double bond}}$ / $\underbrace{RR'C=CRR'}_{\text{Double bond}}$, redox reactions, i.e., it involves C=C bond in food catabolism.

An example is the oxidation of a C−C bond to C=C bond of fumarate at the expense of reduction of FAD to $FADH_2$ as shown below.



Another example is the following step in the β-oxidation of fatty acids catalyzed by acyl-CoA dehydrogenase.

Coenzyme A (CoA) -the agent for transfer of acetyl ($\mathrm{CH_3CO}$-) group

Coenzyme A (CoA) is composed of several components, as illustrated in Figure $8.5.1.4$.

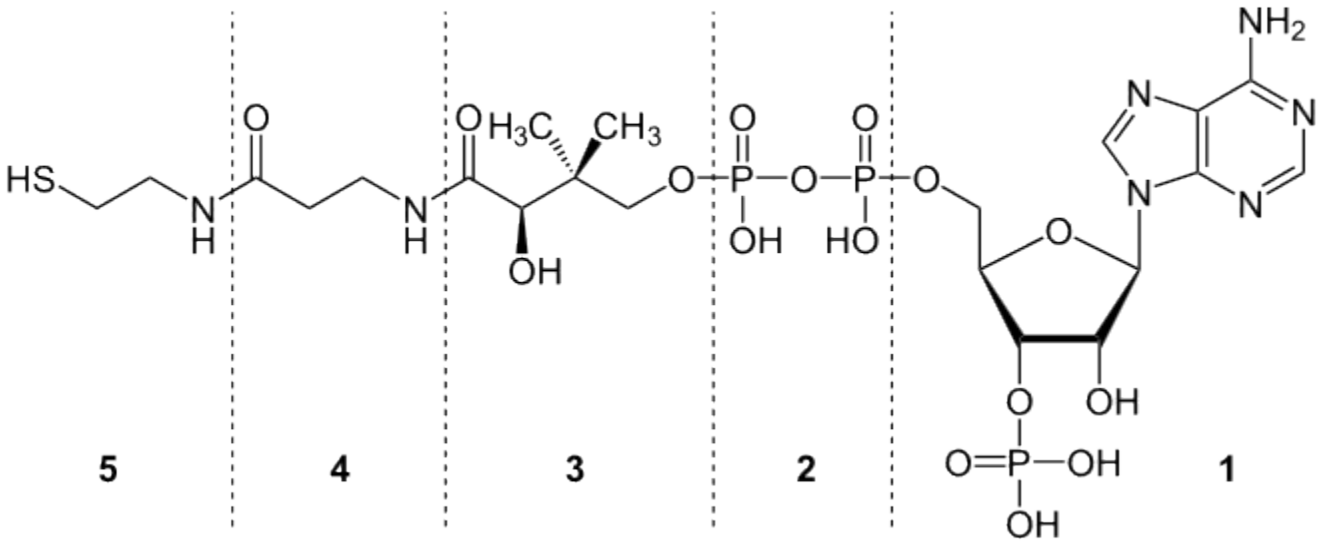
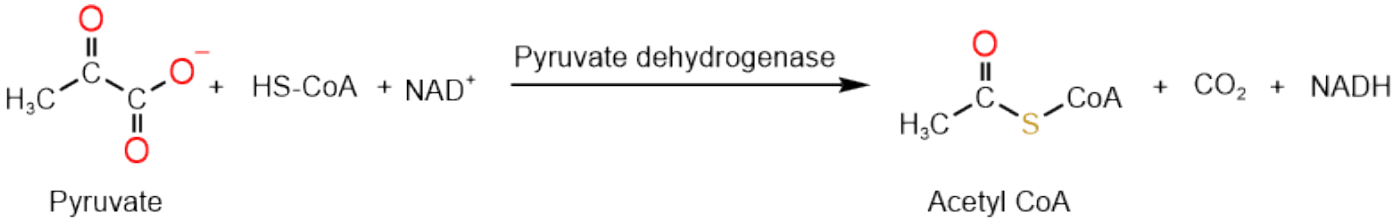
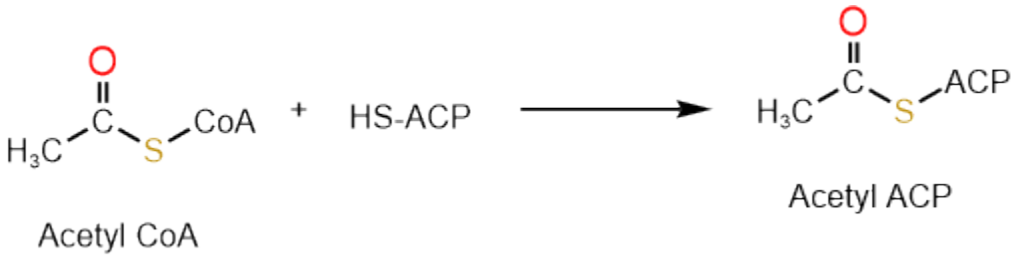


Figure $8.5.1.4$: Structure of coenzyme A: 1 is 3′-phosphoadenosine; 2 is diphosphate, organophosphate anhydride; 3 is pantoic acid; 4 is β -alanine; 5 is cysteamine. Pantoic acid and β-alanine together are pantothenic acid, also called vitamin $\mathrm{B_5}$. (Copyright; NEUROtiker, Public domain, via Wikimedia Commons)

The reactive part of CoA is the thiol ($-\mathrm{SH}$) group which forms a high-energy thioester bond with acyl groups. Free coenzyme is usually represented as $\mathrm{HS-CoA}$, and when it is as a thioester of the acyl group, it is represented as $\mathrm{Ac-S-CoA}$, where Ac- is an acyl group. For example, pyruvate transfers its acetyl ($\mathrm{CH_3CO}$-) groups to $\mathrm{HS-CoA}$ in the following reaction, which is a step in the catabolism of carbohydrates.



The $-\mathrm{S-CoA}$ group in $\mathrm{Ac-S-CoA}$ is a good leaving group that easily transfers the acyl group to other compounds during biosynthesis. For example, the acetyl group is transferred from $\mathrm{Ac-S-CoA}$ to an acyl carrier protein (ACP) which is a step in the synthesis of fatty acids.



Role of enzymes in metabolism

Nearly all metabolic reactions in living things are catalyzed by enzymes. Unlike chemical reactions in laboratories where different solvents, extreme temperature, and pressure conditions can be applied, and strong acids or bases can be used to catalyze the reactions, biomedical reactions in living things must occur under physiological conditions. Enzymes are specialized catalysts that catalyze the reactions under physiological conditions. The major factors responsible for the catalytic activity of the enzymes are the following.

1. Enzymes have functional groups in their active sites that act as handles to bind with and lock the substrate in a proper orientation such that the reagent can easily approach the reactive site on the substrate. This way, any steric hindrance to the reaction is minimized.
2. Enzymes and their co-factors also usually neutralize the ionic groups on the substrate that otherwise repel the charged or partially charged electrophile or nucleophile from approaching the substrate. This way, the electrostatic factors hindering the reaction are minimized.
3. Enzymes usually have functional groups in their active sites that catalyze the reaction of the substrate by activating it as a nucleophile, electrophile, acid, or base catalyst.

The following two examples from glucose catabolism illustrate the factors mentioned above.

The enzyme malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate in the citric acid cycle. The substrate is bound by the enzyme in such a way that i) the charge of its two carboxylate ($-\mathrm{COO}^-$) groups are neutralized by arginine side chains of the enzyme, ii) histidine side acts as a base catalyst and removes the proton from the $-\mathrm{OH}$ group of the substrate, and at the same time, iii) the H on the C carrying the $-\mathrm{OH}$ group is exposed to nicotine amide ring of NAD^+ that removes the proton, as illustrated in Figure \(\PageIndex{1}\),

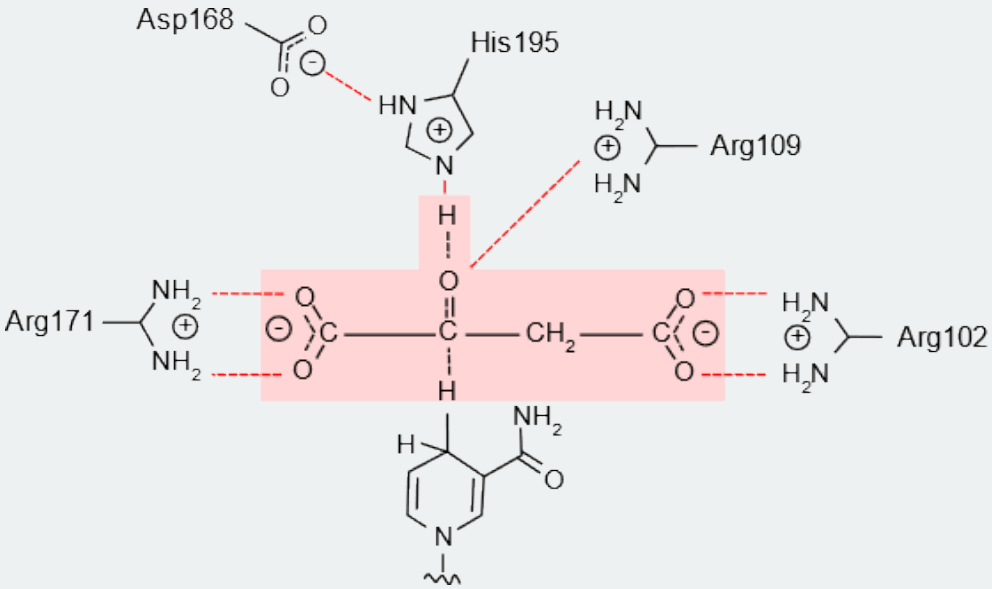


Figure 8.5.1.1: Active site of malate dehydrogenase with bound malate (highlighted in pink). (Copyright; Malate Dehydrogenase Active Site.jpg: Pssangderivative work:Miguelferig, CC0, via Wikimedia Commons)

Another example is the formation of citrate from oxaloacetate and acetyl-CoA in the citric acid cycle, catalyzed by the enzyme citrate synthase as illustrated in Figure 8.5.1.2.

1. The substrate is bonded by the enzyme in such a way that the two substrates are at a bonding distance. The $-\mathrm{CH_3}$ hydrogens of acetyl-CoA are acidic protons due to being α to the carbonyl group. The asparagine side chain acts as a base catalyst and removes the acidic proton from $-\mathrm{CH_3}$ group of acetyl-CoA. The proton transfer to asparagine is facilitated by another proton transfer from histidine (His^{274}) to the $\mathrm{C{=}O}$ group converting it into its enol form -a nucleophile.

2. The enol returns its proton to the histidine side chain in the second step activating the $\mathrm{C{=}C}$ bond of the enol for a nucleophilic attack on the electrophilic ketone ($\mathrm{C{=}O}$ group of oxaloacetate substrate, in the second step. This step is facilitated by the transfer of another proton from another histidine ((His^{320}), the ($\mathrm{C{=}O}$) group involved in the reaction. The result is the conversion of the $\mathrm{C{=}O}$ group into an ($\mathrm{-O-H}$ group of citroyl-CoA ubterneduate.
3. Citroyl-CoA intermediate is hydrogen-bonded to a histidine (His^{274}). This hydrogen bond activates the thioester group for a hydrolysis reaction. Water molecule hydrolyzes the thioester group releasing the citrate product from the enzyme in this reaction step.

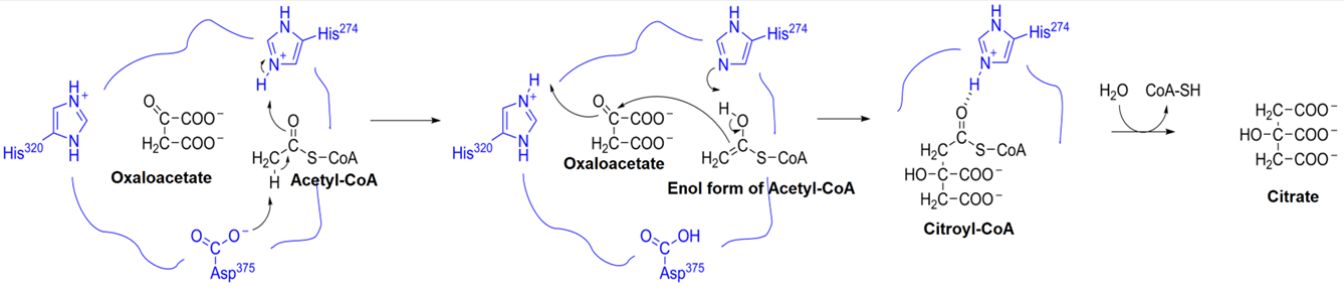


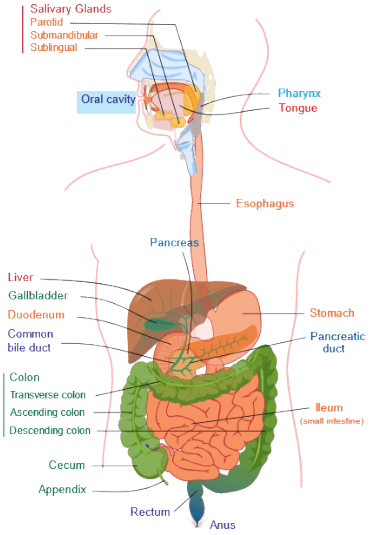
Figure 8.5.1.2: Illustration of enzyme action in the formation of citrate from oxaloacetate and acetyl-CoA in the citric acid cycle (Copyright; Modified from Maksim Fomich, via Wikimediaimedia Commons)

Nearly every metabolic reaction is catalyzed by an enzyme in a similar way as described above.

This page titled 8.5.1: Basics of metabolism is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9.1: Basics of metabolism by Muhammad Arif Malik is licensed Public Domain.

8.5.2: Digestion of food



Learning Objectives

- Understand what happens in stage 1 of food catabolism.

Food digestion, stage 1 of food catabolism, happens in the digestive system shown in the figure on the right (*Copyright: public domain, via Wikimedia Common*). It is primarily the hydrolysis of biopolymers or larger molecules in food into smaller molecules that can transfer from the digestive system to the bloodstream and transport to cells. Carbohydrates are hydrolyzed into monosaccharides, fats into fatty acids and mono- or di-glycerides, proteins into amino acids, and DNA and RNA into mononucleotides by different enzymes.

Digestion of carbohydrates

The carbohydrates in human food are starch, sucrose, and lactose. Their digestion begins in the mouth through enzymes in the saliva and continues in the small intestine by pancreatic amylase. The enzymes degrade the starch into smaller and smaller fragments that ultimately result in glucose and maltose, which the small intestine can absorb.

Sucrose (table sugar) is hydrolyzed into glucose and fructose by the enzyme sucrase. Lactose found in milk is hydrolyzed into glucose and galactose by the enzyme lactase. The majority of the adult population have problems digesting unfermented dairy because they can not produce sufficient amounts of the enzyme lactase -a condition called lactose intolerance.

Digestion of fats

Digestion of fats begins in the mouth through lingual lipase, but it mainly happens in the small intestine. Bile acids emulsify the fats, and the pancreatic enzyme lipase hydrolysis them into free fatty acids and mono- and diacyl-glycerides.

Digestion of proteins

Digestion of proteins happens in the stomach and the duodenum, i.e., the first portion of the small intestine. Three enzymes: pepsin, secreted by the stomach, and trypsin and chymotrypsin, secreted by the pancreas, hydrolyze proteins into smaller peptides which are then hydrolyzed into amino acids by various exopeptidases and dipeptidases.

This page titled 8.5.2: Digestion of food is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9.2: Digestion of food by Muhammad Arif Malik is licensed Public Domain.

8.5.3: Oxidation of glucose -the glycolysis

Learning Objectives

- Understand glycolysis and the essential reactions in this metabolic pathway responsible for glucose oxidation to two pyruvate molecules.
- Understand the reactions that pyruvate undergoes in the absence and presence of oxygen before entering the third stage of catabolism.

Glycolysis

Oxidation of glucose is the 2nd stage of the catabolism of carbohydrates. It happens in the cytoplasm of the cell. It is a metabolic pathway consisting of ten glycolysis reactions, as illustrated in Figure 8.5.3.1.

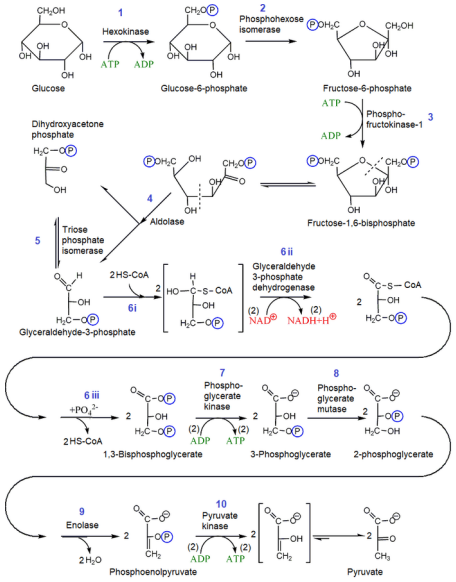
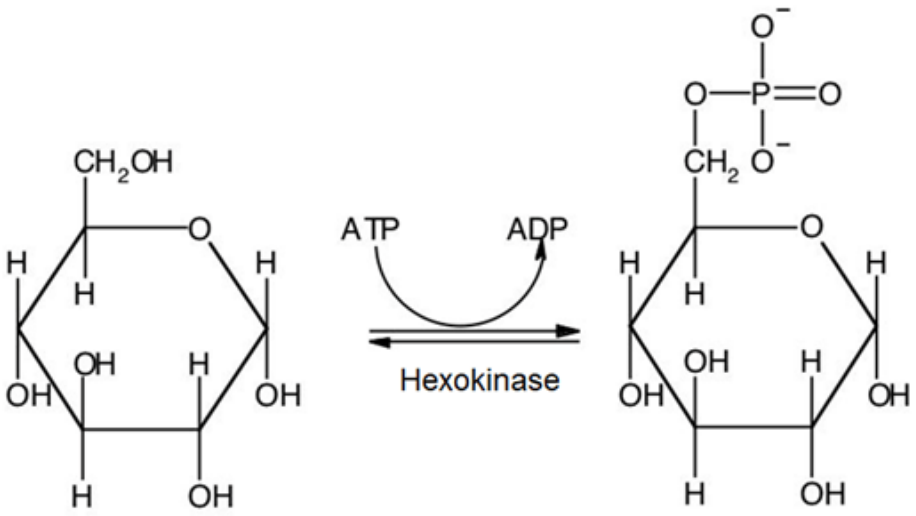


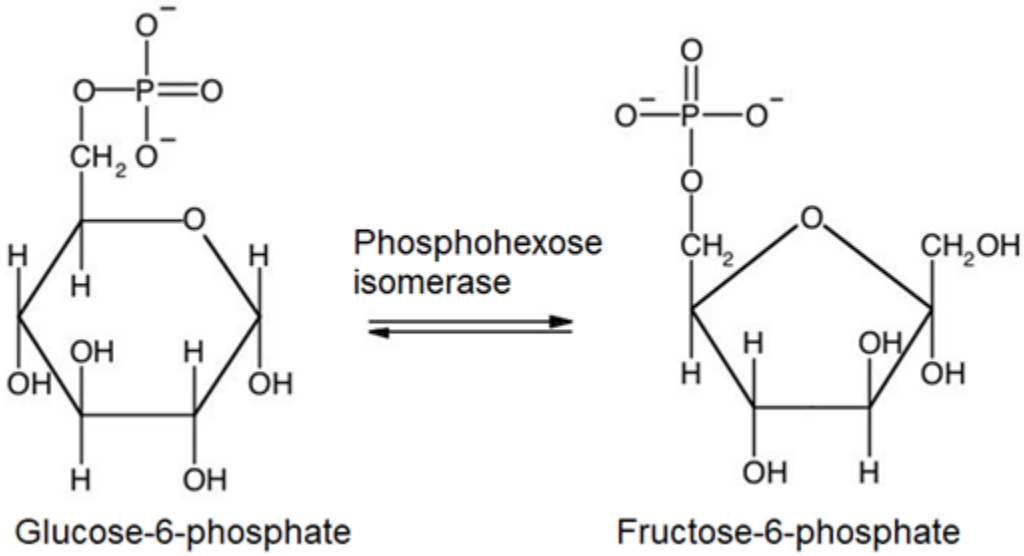
Figure 8.5.3.1: Illustration of ten steps of glycolysis with enzymes needed in each step. Phosphate (PO_4^{2-}) is shown as a letter P in a circle. (Copyright; Modified from created by Morglin, translated by Cryptex, Public domain, via Wikimedia Commons)

The ten reactions of glycolysis are the following (*copyright: modified from a Public Domain resource at Wikipedia.*)

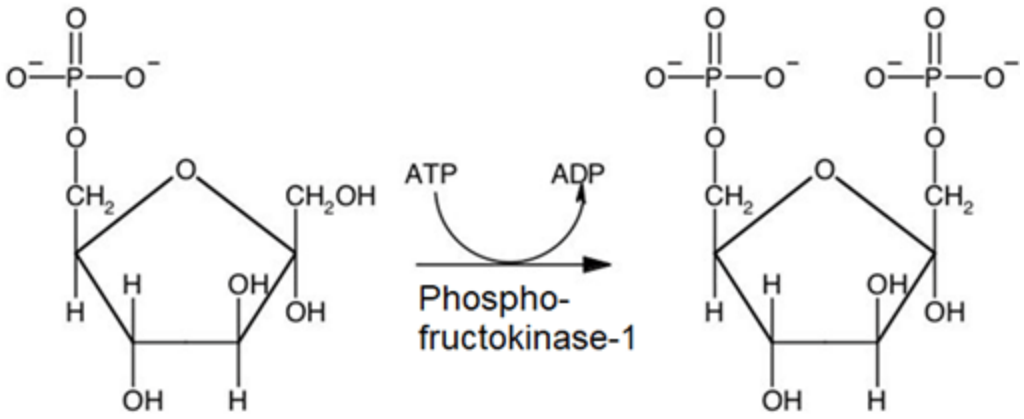
1. Glucose is converted to glucose-6-phosphate by an SN_2 reaction mechanism between a primary alcohol acting as a nucleophile and a P of ATP as an electrophile, where ADP acts as a good leaving group, as shown below.



2. Glucose-6-phosphate, which is an aldohexose, isomerizes to fructose-6-phosphate, which is a ketohexose.

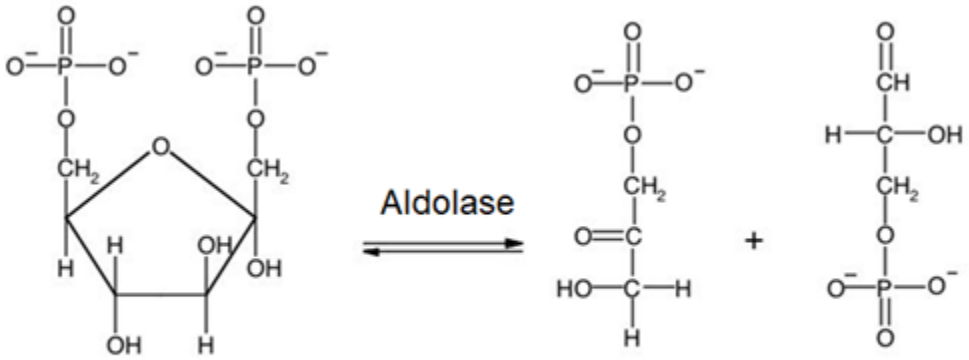


3. The same mechanism phosphorylates the primary alcohol group of fructose-6-phosphate as in the first step producing fructose-1,6-bisphosphate.



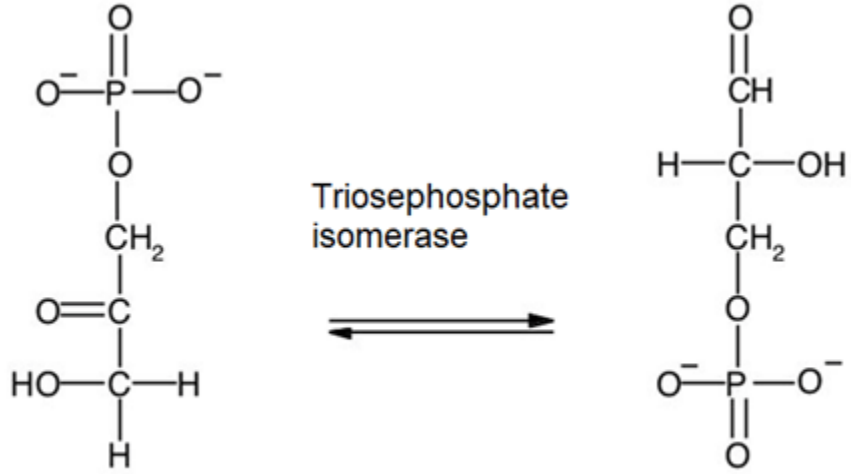
Fructose-6-phosphate Fructose-1,6-bisphosphate

4. The open-chain form of fructose-1,6-bisphosphate, which is in equilibrium with the cyclic form, splits into two compounds, glyceraldehyde-3-phosphate and dihydroxyacetone phosphate.



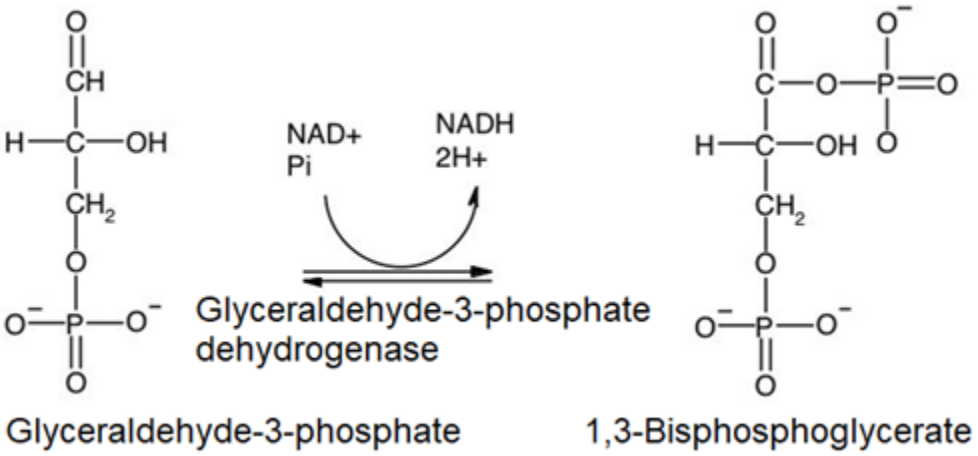
Fructose-1,6-bisphosphate Dihydroxyacetone phosphate Glyceraldehyde-3-phosphate

5. This reaction is an isomerization reaction between glyceraldehyde-3-phosphate and its isomer dihydroxyacetone phosphate. The aldehyde or ketone in these two isomers tautomerizes to their enol forms, which either revert to the initial compound or to its isomers when the enol form changes back to an aldo or keto form.



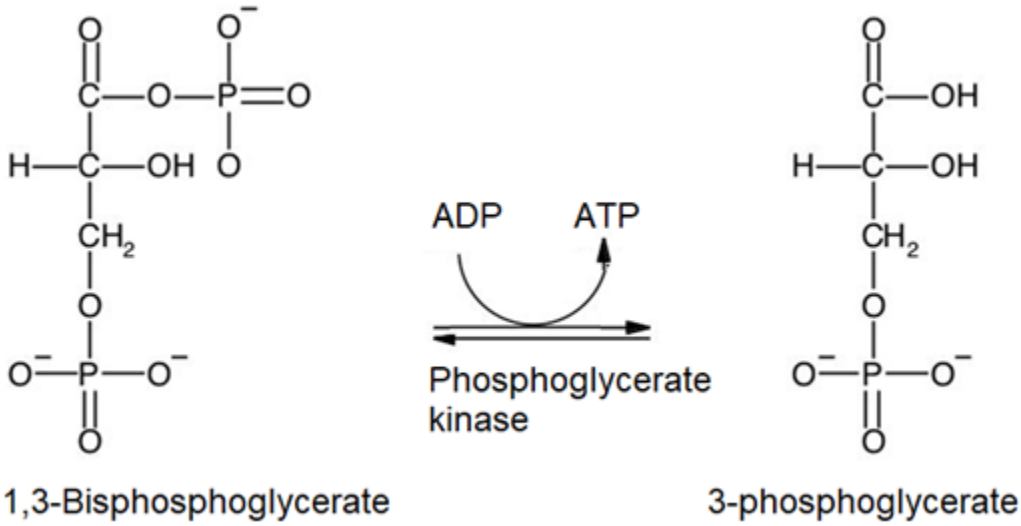
Dihydroxyacetone phosphate Glyceraldehyde-3-phosphate

6. The aldehyde ($-HC=O$) group of glyceraldehyde-3-phosphate is oxidized in three steps: activated by the addition of $HS-CoA$ (step 6i), followed by oxidation of $-OH$ to a carbonyl ($C=O$ group at the expense of reduction of NAD^+ to $NADH$ (step 6ii)); and finally displacement of $HS-CoA$ by a phosphate ($-PO_4^{2-}$) group (step 6iii). The overall reaction is the following.



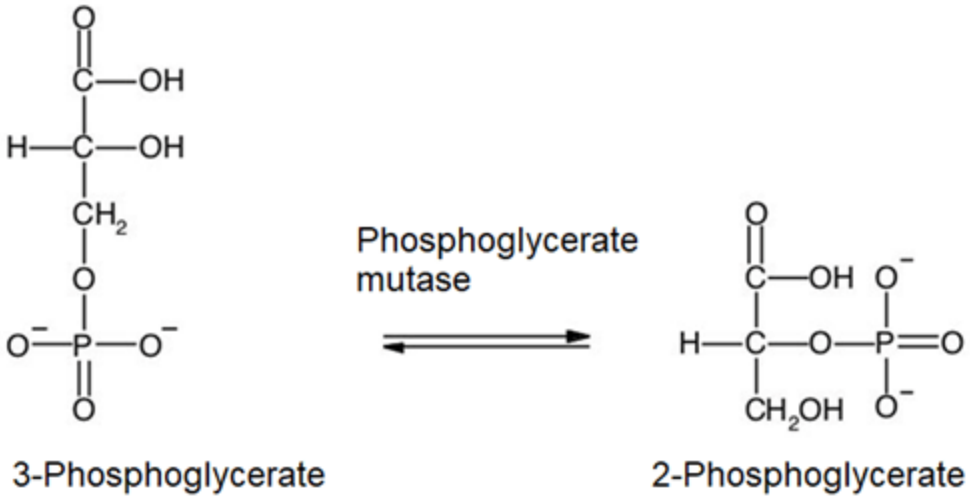
The ketone group in dihydroxyacetone phosphate does not oxidize directly but converts to glyceraldehyde-3-phosphate due to the equilibrium creation#5. So, reaction#6 and reactions after this happen twice for each glucose molecule.

7. A phosphate group is transferred from 1,3-bisphosphoglycerate to ADP producing an ATP and a 3-phosphoglycerate.



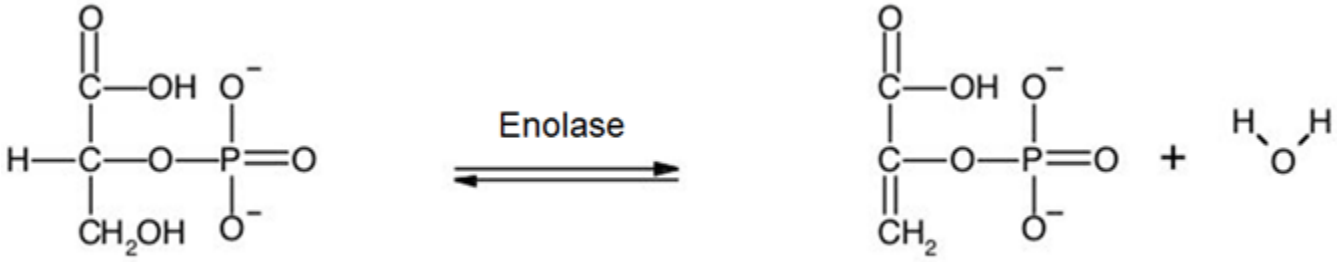
Since this step happens twice for each glucose molecule processed, it compensates for the two ATP consumed, one in step#1 and the other in step#3.

8. 3-Phosphoglycerate is isomerized to 2-phosphoglycerate. In this step, the enzyme phosphoglycerate mutase first transfers its phosphate group to the $-OH$ at position#2 and then receives the phosphate from position#3 of the intermediate, resulting in the isomerization of 3-Phosphoglycerate to 2-phosphoglycerate.



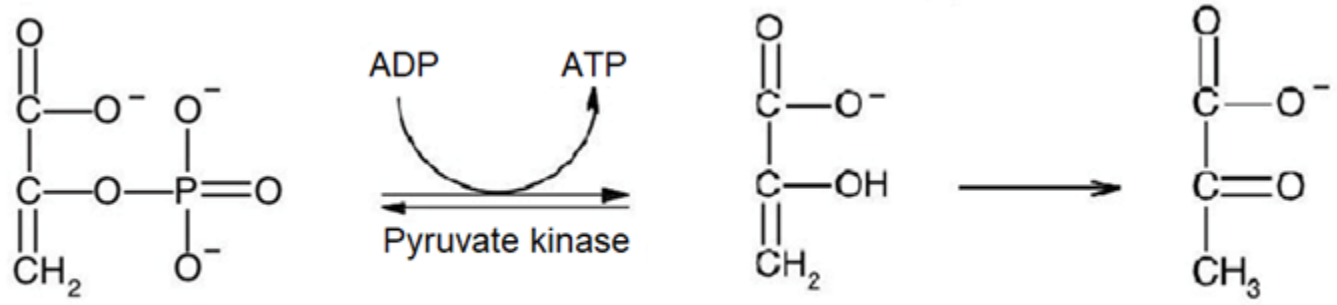
3-Phosphoglycerate 2-Phosphoglycerate

9. 2-Phosphoglycerate is dehydrated, i.e., H_2O is eliminated form it, producing phosphoenolpyruvate.



2-Phosphoglycerate Phosphoenolpyruvate

10. Phosphoenolpyruvate transfers its phosphate group to ADP producing pyruvate and ATP.



Phosphoenolpyruvate Enolpyruvate Pyruvate

Since the reaction happens two times for each glucose, two ATP are produced in this step.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In summary, glycolysis is the oxidation of glucose without the need for oxygen, i.e., **anaerobic oxidation**, consisting of two phases. The first phase comprises reactions#1 to 5. It is the preparatory phase, which is also energy consuming phase in which a six $C's$ glucose ($C_6H_{12}O_6$) is phosphorylated twice at the expense of two ATP's and converts to two glyceraldehyde-3-phosphate molecules having three $C's$ each. The second phase, comprising reactions#6 to 10, is the payoff phase in which four ATP's are produced, i.e., two more than the ATP consumed in the first phase. Two NAD^+ are reduced to two $NADH$ along with the production of two pyruvates CH_3COCOO^- in the second phase, as shown in the following reaction.

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 2\,NAD^+ + 2\,ADP + 2\,Pi \longrightarrow 2\,\underbrace{CH_3{-}\overset{O}{\overset{||}{C}}{-}\overset{O}{\overset{||}{C}}{-}O^-}_{\text{Pyruvate}} + 2\,NADH + 2\,ATP + 2\,H^+ + 2\,H_2O$$

Fate of pyruvate

Two NAD^+ are reduced to two $NADH$ which need to be oxidized back to NAD^+ for the process to continue. It happens in different ways depending on whether sufficient oxygen is present, i.e., **aerobic condition or aerobic respiration,** or if there is no oxygen, i.e., **anaerobic condition**. The fate of pyruvate also depends on whether the condition is aerobic or anaerobic, as described below.

Aerobic condition

In aerobic conditions, $NADH$ is not oxidized at this stage, it is oxidized to NAD^+ at the expense of oxygen in mitochondria. The energy released by the oxidation of $NADH$ is used to produce more $ATP's$ in the third stage of catabolism. Pyruvate is also transferred to mitochondria and undergoes oxidation at the cost of NAD^+ reduced to $NADH$ through a series of reactions catalyzed by a complex of three enzymes and five coenzymes, collectively called the pyruvate dehydrogenase complex. The overall reaction is the decarboxylation of pyruvate, i.e., carbon dioxide (CO_2) eliminated. The acetyl ($CH_3{-}CO^-$) group is transferred to $HS{-}CoA$ producing acetyl-CoA, as shown in the following reaction.

$$\underbrace{CH_3{-}\overset{O}{\overset{||}{C}}{-}\overset{O}{\overset{||}{C}}{-}O^-}_{\text{Pyruvate}} + HS{-}CoA + NAD^+ \xrightarrow{\text{Pyruvate dehydrogenase complex}} \underbrace{CH_3{-}\overset{O}{\overset{||}{C}}{-}S{-}CoA}_{\text{Acetyl CoA}} + CO_2 + NADH + H^+$$

Acetyl-CoA enters into the citric acid cycle, also called the Krebs cycle, in mitochondria, the third stage of catabolism.

The processing of pyruvate under aerobic conditions is also called **oxidative decarboxylation** or **link reaction,** which links glycolysis to the citric acid cycle, as explained in the video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Two pyruvates are produced per glycolysis of one glucose molecule. The overall reaction of one glucose molecule under aerobic conditions before entry into the citric acid cycle is the following.

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 2\,HS{-}CoA + 4\,NAD^+ + 2\,ADP + 2\,Pi \longrightarrow 2\,\underbrace{CH_3{-}\overset{O}{\overset{||}{C}}{-}S{-}CoA}_{\text{Acetyl CoA}} + 2\,CO_2 + 4\,NADH + 2\,ATP + 4\,H^+ + 2\,H_2O$$

There is a net gain of six energetic molecules: $2\,ATP$ and $4\,NADH$ from glycolysis and link reaction under aerobic conditions.

Anaerobic condition



During vigorous exercises, as shown in the figure on the right, or strenuous physical work, oxygen depletes in muscles creating anaerobic (oxygen-free) conditions. Under anaerobic conditions, NADH is oxidized to $\mathrm{NAD^+}$ in the cytoplasm at the expense of the reduction of pyruvate to lactate, as shown in the reaction below.

$$\underbrace{\mathrm{CH_3{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}O^-}}_{\text{Pyruvate}} + \mathrm{NADH} + \mathrm{H^+} \xrightarrow{\text{Lactate dehydrogenase}} \underbrace{\mathrm{CH_3{-}\overset{\overset{\mathrm{OH}}{|}}{C}H{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}O^-}}_{\text{Lactate}} + \mathrm{NAD^+}$$

Overall, glycolysis of glucose under anaerobic conditions is the following reaction.

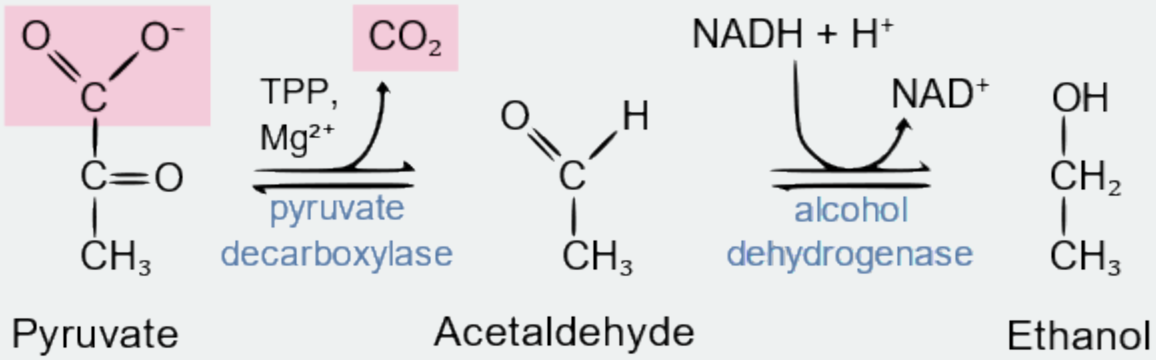
$$\underbrace{\mathrm{C_6H_{12}O_6}}_{\text{Glucose}} + 2\,\mathrm{ADP} + 2\,\mathrm{Pi} \longrightarrow 2\,\underbrace{\mathrm{CH_3{-}\overset{\overset{\mathrm{OH}}{|}}{C}{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}O^-}}_{\text{Lactate}} + 2\,\mathrm{ATP}$$

The two NADH produced in the glycolysis of one glucose are consumed in the anaerobic conversion of two pyruvates into two lactates.

- There is a net gain of two energetic molecules, i.e., $2\,\mathrm{ATP}$ in glycolysis of one glucose followed by anaerobic conversion of two pyruvates into two lactates.
- The accumulation of lactate makes muscles tire and sour. The person keeps beating heavily after the exercise to pay the oxygen debt. Most of the lactate is transported to the liver, where it is re-oxidized to pyruvate.

Fermentation

Instead of converting pyruvate to lactate, as in humans and animals, yeast has an enzyme called pyruvate decarboxylase that decarboxylates pyruvate to acetaldehyde. Then NADH is oxidized to $\mathrm{NAD^+}$ at the expense of the reduction of acetaldehyde to ethanol in a process called fermentation, as shown in the reaction below.



Copyright; User: Cwernert at en.Wikipedia, vectorized by User: GKFX., CC BY-SA 3.0, via Wikimedia Commons.

The overall reaction of glycolysis of glucose in the fermentation process by yeast.

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 2\,ADP + 2\,Pi \longrightarrow 2\,\underbrace{CH_3{-}CH_2{-}OH}_{\text{Ethanol}} + 2\,CO_2 + 2\,ATP$$

This page titled 8.5.3: Oxidation of glucose -the glycolysis is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9.3: Oxidation of glucose -the glycolysis by Muhammad Arif Malik is licensed Public Domain.

8.5.4: Citric acid cycle

Learning Objectives

- Understand the citric acid cycle, its reactions, and the yield of energetic molecules.

What is the citric acid cycle?

The **citric acid cycle** is the first part of the third stage of food catabolism, including carbohydrates, fats, and proteins. It is called the citric acid cycle because a $2\,\mathrm{C's}$ acetyl-CoA produced in the second stage of catabolism of foods reacts with a $4\,\mathrm{C's}$ oxaloacetate and produces $6\,\mathrm{C's}$-citrate in the first reaction. The citrate is oxidized through a series of eight reactions producing two carbon dioxide ($\mathrm{CO_2}$) and re-generates the oxaloacetate to repeat the next round of the cycle, as shown in Figure $8.5.4.1$. It is also called the **tricarboxylic acid cycle** because citrate has three carboxylates ($-\mathrm{COO^-}$) groups, i.e., triacid. Another name for it is the **Krebs cycle** in honor of Hans Krebs, who discovered this metabolic pathway.

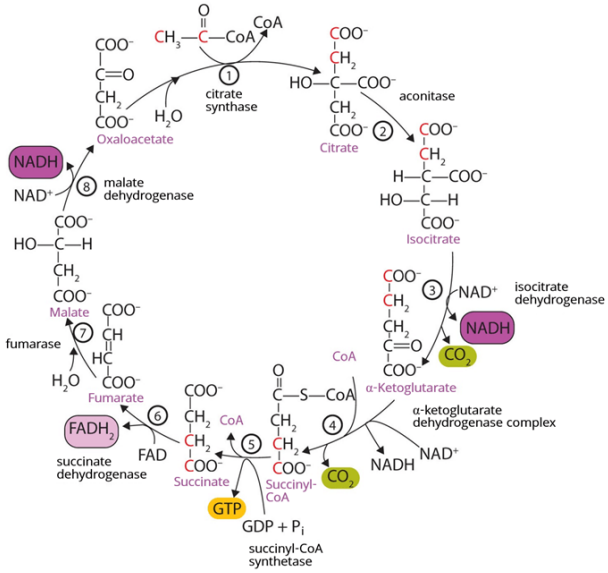


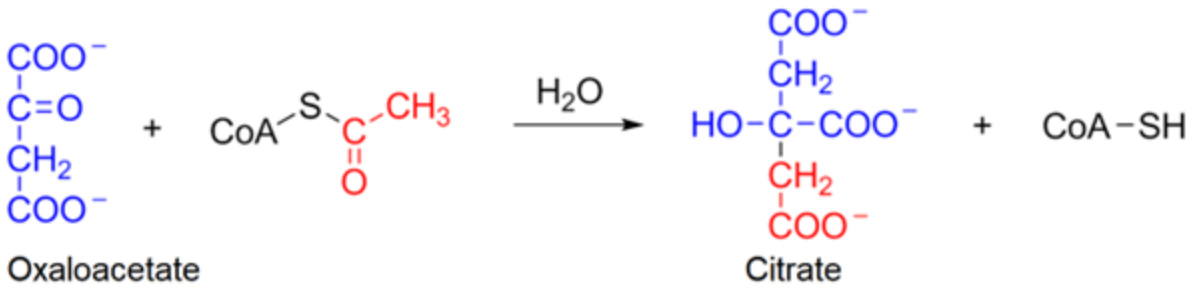
Figure $8.5.4.1$: Citric acid cycle. (Copyright; Theislikerice, CC BY-SA 4.0, via Wikimedia Commons)

The oxidation reactions in the citric acid cycle release energy, which is coupled to the reduction of $\mathrm{NAD^+}$ to NADH, FAD to FADH , or conversions of ADP to high-energy ATP. Guanosine triphosphate (GTP) -another high-energy molecule is also produced from guanosine diphosphate (GDP), but this reaction ultimately reverses to transfer its energy to make ATP.

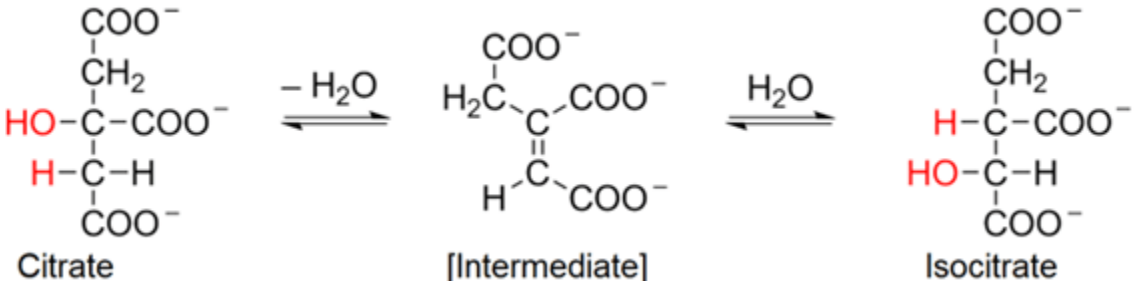
Reactions of the citric acid cycle

The following equations of the reactions are modified from a public domain at Wikipedia.

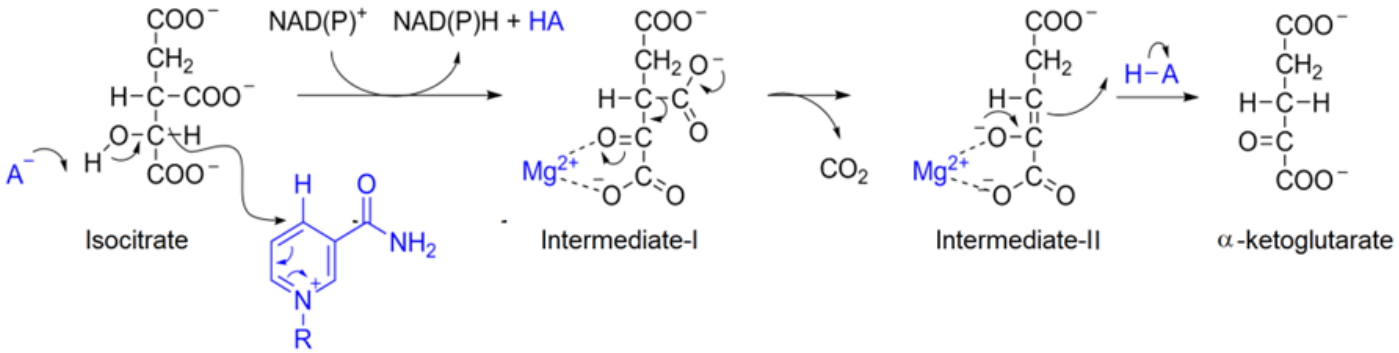
1. To understand the first reaction, recall that the $-\mathrm{CH_3}$ group of acetyl-CoA has relatively acidic protons as it is α to a carbonyl ($\mathrm{C{=}O}$) group. An enzyme, acting as a base, removes a proton from the $-\mathrm{CH_3}$ group making it a carbanion which, being a strong nucleophile, attacks the ketone-C of oxaloacetate, which is an electrophile. This nucleophilic addition reaction converts the $\mathrm{C{=}O}$ into an $-\mathrm{OH}$ group. It is followed by the hydrolysis of the thioester by $\mathrm{H_2O}$ through nucleophilic acyl substitution mechanism producing citrate and $\mathrm{HS{-}CoA}$, as shown in the following overall reaction.



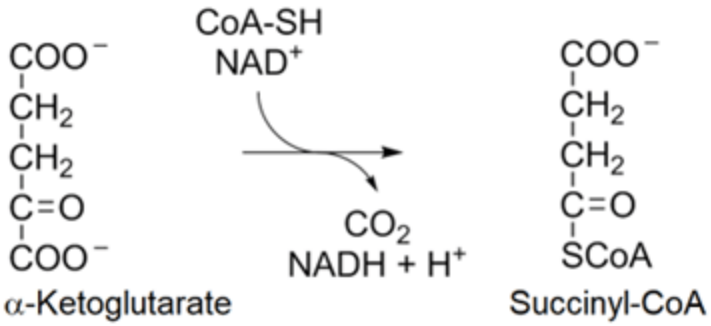
2. To understand the second reaction, recall that tertiary $-\mathrm{OH}$ do not oxidize but are easily eliminated through E2-dehydration mechanics. The tertiary $-\mathrm{OH}$ of citrate is eliminated, and then $\mathrm{H_2O}$ adds to the alkene intermediate, but it installs a secondary $-\mathrm{OH}$ group in isocitrate product, as shown below.



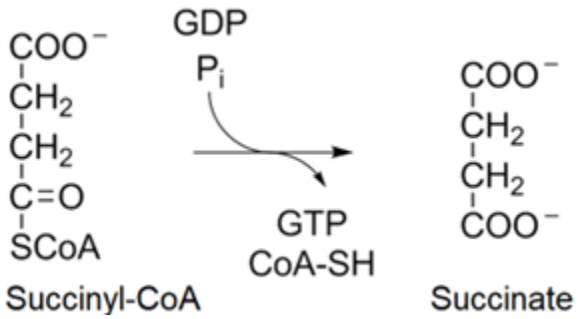
3. The third reaction has three steps. In the first step, the secondary $-\mathrm{OH}$ in isocitrate is oxidized to a ketone group at the expense of reduction of a $\mathrm{NAD^+}$ into NADH. In this step, one $-\mathrm{H}$ is picked up by an enzyme ($\mathrm{A^-}$) and the second by the nicotinamide ring (shown in blue color) of the coenzyme $\mathrm{NAD^+}$. The $\mathrm{COO^-}$ group in the intermediate-I, being β to the $\mathrm{C{=}O}$ group, eliminates (decarboxylate) easily as $\mathrm{CO_2}$. The decarboxylation is facilitated by cofactor $\mathrm{Mg^{2+}}$ that binds with and draws electrons from the $\mathrm{O's}$. The enzyme $\mathrm{H{-}A}$ protonates the enole $\mathrm{C{=}C}$ bond of the intermediate-II that produces α-ketoglutarate, as shown below.



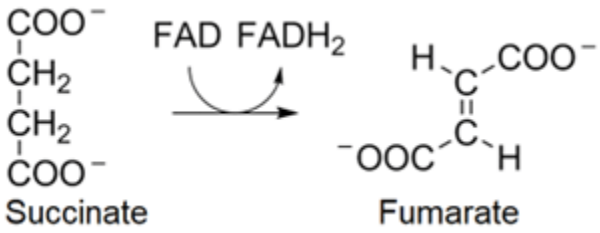
4. The fourth reaction resembles the previous link reaction's oxidative decarboxylation of pyruvate to acetyl-CoA. In this reaction, $\mathrm{NAD^+}$ is reduced into NADH at the expense of oxidative decarboxylation of α-ketoglutarate and the resulting acyl-group is transferred to $\mathrm{HS{-}CoA}$ resulting in succinyl-CoA product as shown in the reaction below.



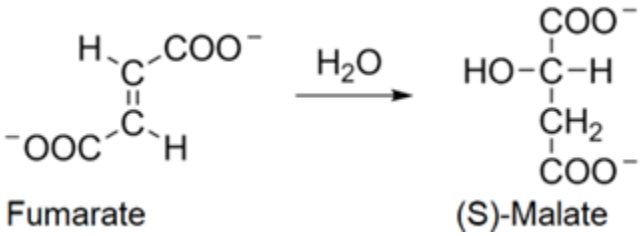
5. The fifth reaction has two steps. In the first step, a phosphate (Pi) group substitutes $-\mathrm{S}-\mathrm{CoA}$ group from succinyl-CoA. In the second step, the phosphate group is transferred to GDP that converts to a GTP by the same mechanism as in steps #7 and step#10 of glycolysis. GTP reverts back to GDP and transfers its phosphate group to ADP that converts to one ATP. So, production of one GTP is equal to production of one ATP. The overall reaction five is shown below.



6. In the sixth reaction, one FAD is reduced into $\mathrm{FADH_2}$ at the expense of oxidation of succinate to fumarate by the following overall reaction.

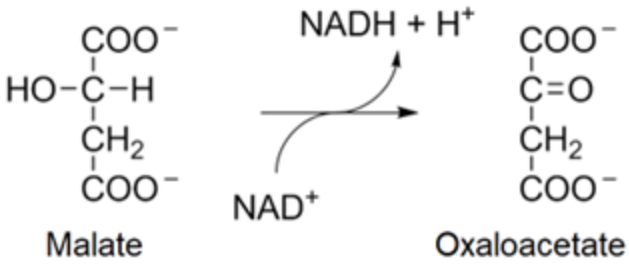


7. In the seventh reaction $\mathrm{H_2O}$ hydrates the $\mathrm{C}=\mathrm{C}$ of fumarate producing malate, as shown below.



Recall that the enzymes are stereospecific. In this case, only fumarate (not its cis-isomer) reacts with the enzyme, and only (S)-malate (not its enantiomers (R)-malate) is produced.

8. In the eighth step, one more $\mathrm{NAD^+}$ is reduced into NADH at the expense of oxidation of malate into oxaloacetate, as shown in the reaction below.



The oxaloacetate is the reactant of step#1 and begins the cycle again by reacting with another acetyl-. One citric acid cycle has the following overall reaction with one acetyl-CoA.

$$\underbrace{\mathrm{CH_3{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}S{-}CoA}}_{\text{Acetyl CoA}} + 3\,\mathrm{NAD^+} + \mathrm{FAD} + \mathrm{GDP} + \mathrm{P_i} + 2\,\mathrm{H_2O} \longrightarrow \mathrm{HS{-}CoA} + 3\,\mathrm{NADH} + 3\,\mathrm{H^+} + \mathrm{FADH_2} + \mathrm{GTP} + 2\,\mathrm{CO_2}$$

Recall that the catabolism of one glucose molecule produces two pyruvates that, intern, produce two acetyl-CoA in the oxidative decarboxylation of pyruvate under aerobic conditions. So, the citric acid cycle runs twice for the catabolism of one glucose molecule. One GTP ultimately converts into one ATP.

Two citric acids from one glucose molecule yield ten energetic molecules from two turns citric acid cycles: $6\,\mathrm{NADH}$, $2\,\mathrm{FADH_2}$, and $2\,\mathrm{ATP}$

Predicting the production or consumption of energetic molecules in food catabolism

The information in this note is meant to help predict which energetic molecule will be produced or consumed during food catabolism reactions. Recall that oxidation generally release energy (exothermic), and reduction consumes energy (endothermic). The energy released from oxidation reactions in food catabolism is utilized to produce one of the two energetic coenzymes, i.e., either NADH from the reduction of $\mathrm{NAD^+}$ or $\mathrm{FADH_2}$ from the reduction of FAD. Recall that $\mathrm{C{=}O}$ is stronger bond (~800 kJ/mol bond dissociation energy) compared to $\mathrm{C{=}C}$ bond (~600 kJ/mole bond dissociation energy). NADH is more energetic than $\mathrm{FADH_2}$. Therefore, any oxidation reaction involving carbonyl ($\mathrm{C{=}O}$) group produces NADH and any oxidation involving $\mathrm{C{=}C}$ bond produces $\mathrm{FADH_2}$. There is only one reaction described in food catabolism in this book, i.e., reaction#6 in the citric acid cycle that involved oxidation of $\mathrm{C{-}C}$ to $\mathrm{C{=}C}$ bond and produced $\mathrm{FADH_2}$. All other oxidation reactions in the citric acid cycle and glycolysis involve $\mathrm{C{=}O}$ group, i.e., either alcohol ($\mathrm{CH{-}OH}$) to carbonyl ($\mathrm{C{=}O}$) conversion, or ketone ($\mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}R'}$ to thioester ($\mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}S{-}R''}$) conversions produce NAD. Conversionerisons of pyruvate to lactate and ethanol under anaerobic conditions are reduction reactions involving $\mathrm{C{=}O}$ group and consume NADH.

Reactions that are not redox reactions, e.g., acyl ($\mathrm{R{-}\overset{\overset{\mathrm{O}}{\|}}{C}{-}}$) transfer, hydration (addition of $\mathrm{H_2O}$) , dehydration (elimination of $\mathrm{H_2O}$), and isomerization, involve a small amount of energy. They are usually not coupled with the formation or consumption of energetic molecules like NADH, $\mathrm{FADH_2}$, or ATP.

Exceptions include phosphate ester or thioester group, which are high-energy groups. The addition or removal of phosphate ester or thioester groups without redox reaction is usually coupled with ATP/ ADP conversion. Two reactions, i.e., reaction#1 and reaction#3 in glycolysis (Figure 9.3.1) add phosphate ester bond and consume ATP, and two reactions, i.e., reaction#& and reaction#10, remove phosphate ester and produce ATP. One reaction, i.e., reaction#5 in the citric acid cycle (Figure 9.4.1), involves the replacing a thioester group with a carboxylate group without redox. It is coupled with the production of GTP, which, in turn, is associated with the output of ATP.

In summary:

- redox reactions involving $\mathrm{C{=}O}$ group relate to NADH,
- redox reactions involving $\mathrm{C{=}C}$ group relate to $\mathrm{FADH_2}$, and

- all other reactions that involve energetic groups like phosphate esters or thioesters relate to ATP in food catabolism.

NADH and FADH are high-energy molecules that are oxidized by $\mathrm{O_2}$ in part two of the third stage of catabolism of food. The energy released from the oxidation of NADH and FADH is used to produce more ATPs, which is described in the next section.

The following video summarizes the citric acid cycle, also called the Krebs cycle.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This page titled 8.5.4: Citric acid cycle is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9.4: Citric acid cycle by Muhammad Arif Malik is licensed Public Domain.

8.5.5: Oxidative Phosphorylation

Learning Objectives

- Understand parts 2 and 3 of the third stage of food catabolism, i.e., oxidative phosphorylation, and the basics of the redox chemistry of the molecules involved, including flavin mononucleotide, ubiquinone, and cytochrome c.
- Understand the electron transport process basics, including basic redox reactions and proton pumping in complexes I through IV.
- Understand the ATP synthesis process, accounting for the ATP yield, body heating, and reactive oxygen species produced.

What is oxidative phosphorylation?

In food catabolism up to this point, the energy is transferred from food to high-energy molecules, such as energetic electron carriers NADH and $\mathrm{FADH_2}$. Recall that these electron carrier molecules do not deliver elections alone; they deliver protons and electrons, i.e., $2\,\mathrm{H^+} + 2\,\mathrm{e^-}$. The electrons in NADH and FADH are transported through a series of enzymes located in the inner membrane of mitochondria, and ultimately, the $2\,\mathrm{H^+} + 2\,\mathrm{e^-}$ reduce O into $\mathrm{H_2O}$ by the reaction shown below.

$$4\,\mathrm{H^+} + 4\,\mathrm{e^-} + \mathrm{O_2} \longrightarrow 2\,\mathrm{H_2O}$$

In each step, the electrons are transferred from a substance with a higher reduction potential to one with a lower reduction potential. Energy is released in each step proportional to the difference in the reduction potential. The energy is converted to electrochemical potential energy by pumping protons ($\mathrm{H^+}$) from the matrix to intermembrane spaces through sets of enzymes called complexes I, III, & IV. The electrochemical energy is harnessed by ATP synthase, which allows protons to flow back from the intermembrane space to the matrix and couples the energy released to synthesize higher energy ATP by phosphorylating lower energy ADP with inorganic phosphate ($\mathrm{P_i}$). These interconnected processes of electron transport and ATP synthesis are collectively called **oxidative phosphorylation**. It happens in the mitochondria of cells as illustrated in Figure $8.5.5.1$ and is explained in the following sections.

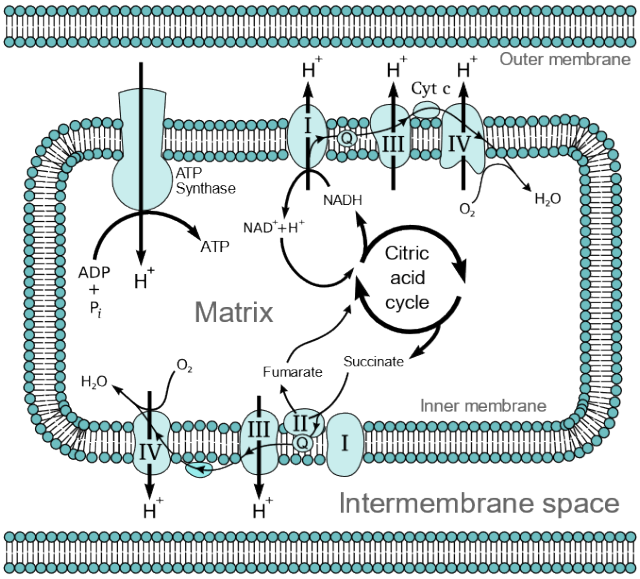
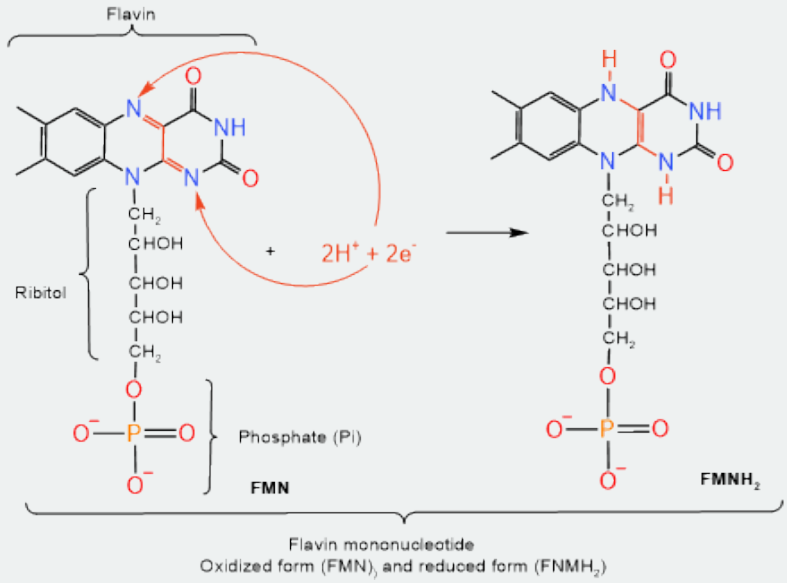
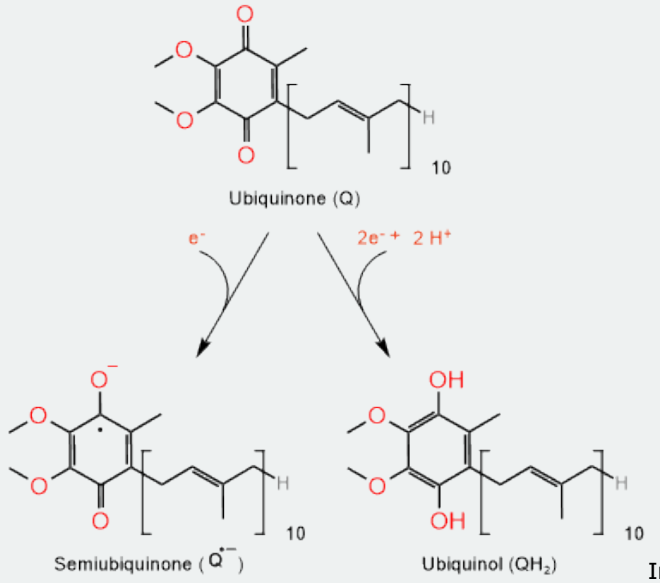


Figure $8.5.5.1$: Schematic diagram of the mitochondrial electron transport chain comprising of complexes I through IV and production of ATP by ATP-synthase. (Copyright; Fvasconcellos 22:35, 9 September 2007 (UTC), Public domain, via Wikimedia Commons)

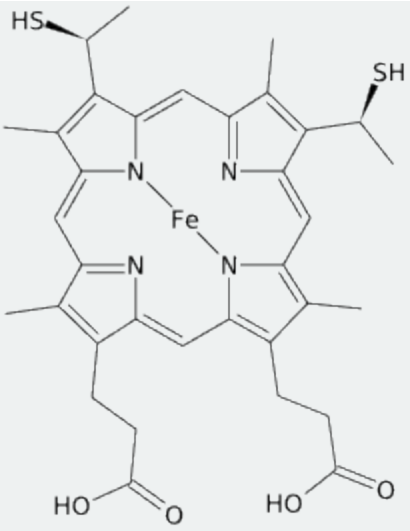
Electron and proton transport molecules





In addition to NAD^+ and FAD, the following molecules are involved in the transport of electrons and protons in the electron transport chain. The first electron acceptor from NADH or FADH_2 in the electron transport chain is **flavin mononucleotide** which changes from its oxidized form represented as FMN to its reduced form represented as FMNH_2 as illustrated in the figure on the right. Structure of FMN comprises three parts, phosphate, ribitol, and flavin. It is the same as one of the two nucleotides in flavin adenine dinucleotide. Nitrogen atoms in flavin part of FMN take part in the redo process, as illustrated in the figure on the right. FMNH_2 pass on the electrons to iron-sulfur clusters, which include $[2\,\mathrm{Fe}\text{-}2\,\mathrm{S}]$ and $[4\,\mathrm{Fe}\text{-}4\,\mathrm{S}]$ clusters. The redox couple in iron-sulfur clusters is $\mathrm{Fe}^{3+}/\mathrm{Fe}^{2+}$, i.e., the following redox reaction:

$$\mathrm{Fe}^{3+} + \mathrm{e}^- \underset{\text{Oxidation}}{\overset{\text{Reduction}}{\rightleftharpoons}} \mathrm{Fe}^{2+}.$$



Finally, the electrons are transferred to coenzyme-Q, which in oxidized form is ubiquinone represented as Q and in its reduced form is ubiquinol represented as $\mathrm{QH_2}$, as illustrated in the figure above on the left. Coenzyme-Q is hydrophobic, i.e., lipid-soluble. Coenzyme-Q moves freely in the hydrophobic environment within the biliary inner membrane to pass on electrons from complex I and II to complex III. Coenzyme-Q not only can accept two electrons and two protons, but it can also accept one electron to become semiubiquinone, a resonance-stabilized radical anion.

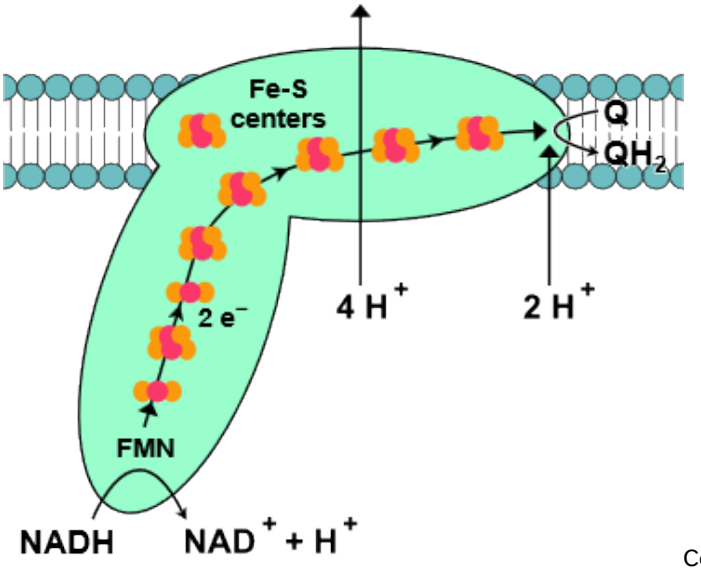
Cytochrome c (Cyt c) is another electron carrier protein with a heme group attached. Heme group is illustrated in the figure on the right. Iron in the heme group of cytochrome c is the redox couple ($\mathrm{Fe^{3+}/Fe^{2+}}$) as in the case of iron-sulfur clusters. $\mathrm{Fe^{3+}}$ changes to $\mathrm{Fe^{2+}}$ when an electron is added to it, and the reverse happens when an electron is removed. Cytochrome c is water soluble and travels between complex III and complex IV in the watery environment along the inner membrane's surface facing intermembrane space.

Electron transport chain

In this part 2 of stage 3, electrons are removed from NADH and $\mathrm{FADH_2}$ and passed on to reduce $\mathrm{O_2}$ into $\mathrm{H_2O}$ through a series of enzyme complexes, i.e., complexes I through IV and mobile electron carriers like ubiquinone (Q) and cytochrome c ($\mathrm{Cyt\ \ c}$) by a process called **electron transport chain**. At each step, electrons pass from a substance with a higher reduction potential to one with a lower one. Energy is released proportional to the difference in the reduction potentials.

The complex I, III, and IV extend from the matrix to the intermembrane space through the inner membrane, as illustrated in Figure $8.5.5.1$. These three complexes utilize the energy released to pump protons ($\mathrm{H^+}$) from the matrix to the intermembrane space. Complex II extends to the matrix from the inner membrane but does not cross the membrane to intermembrane space. Therefore complex II transports electrons but does not pump protons. Coenzyme-Q (Q) is a hydrophobic electron carrier molecule that moves in a hydrophobic environment inside the lipid bilayer of the inner membrane from complexes I and II to complex III. Cytochrome c ($\mathrm{Cyt\ \ c}$) is a hydrophilic electron carrier molecule that moves along the surface of the inner membrane facing the intermembrane space from complex III to complex IV. The enzyme complexes and the mobile electron carriers involved in the electron transport chain and ATP synthesis are described in the following sub-sections, where the chemical equations and figures are taken from the public domain source via Wikipedia.

Complex I



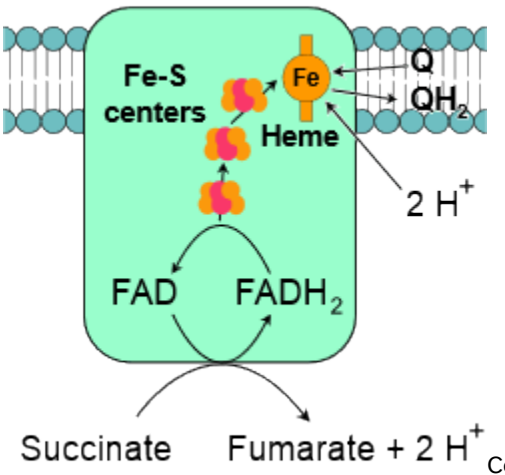
Complex I, also known as NADH-coenzyme Q oxidoreductase or NADH dehydrogenase, is the first enzyme complex in the electron transport chain. Complex I is illustrated in the figure on the right. The electron transport begins when an NADH binds with and donates two electrons to complex I. The electrons are received by flavin mononucleotide (FMN) attached to the complex, which is reduced to $\mathrm{FMNH_2}$ form. Then the electrons are transferred through a series of iron-sulfur clusters within complex I to coenzyme-Q (Q) that, in turn, reduces to its reduced form $\mathrm{QH_2}$. $\mathrm{QH_2}$ travels through the lipid bilayer to and delivers the electrons to complex III.

As the electrons move through complex I and release part of their energy, four protons ($4\,\mathrm{H^+}$) are pumped from the matrix to intermembrane space by utilizing the energy released. It generated polarity across the inner membrane, with the positive side (P-side) facing the intermembrane area and the negative side (N-side) facing the matrix. The exact mechanics of proton pumping is not yet clearly understood. Still, the research indicates that it happens through conformational changes in complex I such that the protein bind protons on the N-side and releases them on the P-side of the membrane. The overall redox reaction in complex I is the following.

$$\mathrm{NADH + Q + 5\,H^+_{matrix} \longrightarrow NAD^+ + QH_2 + 4\,H^+_{intermembrane}}$$

$\mathrm{NAD^+}$ liberated in this reaction becomes available to oxidize more substrates in the metabolic pathways.

Complex II



Complex II, also known as succinate-Q oxidoreductase or succinate dehydrogenase, is a second entry point to the electron transport chain. It is part of the citric acid cycle and the electron transport chain. It contains protein subunits bonded with flavin adenine dinucleotide FAD, iron-sulfur clusters, and a heme group, as illustrated in

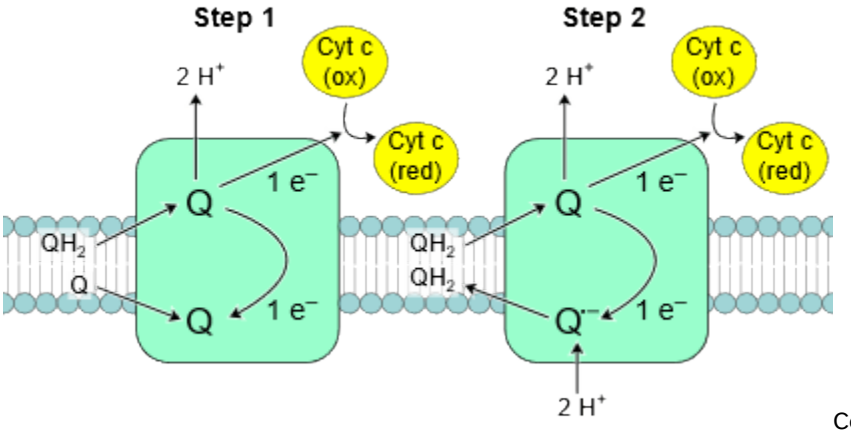
the figure on the right. Succinate is oxidized to fumarate in the citric acid cycle at the expense of oxidation of FAD to $\mathrm{FADH_2}$ in complex II. The electrons and protons transfer from $\mathrm{FADH_2}$ through iron-sulfur clusters to coenzyme-Q in complex II by the following overall reaction.

$$\mathrm{Succinate + Q \longrightarrow Fumarate + QH_2}$$

Coenzyme-Q passes on the electrons from complex II to complex III as from complex I to complex III.

Recall that (\ce{FADH2}\) is less energetic than (\ce{NADH}\). Since oxidation $\mathrm{FADH_2}$ to FAD at complex II releases less energy than oxidation of NADH to $\mathrm{NAD^+}$ at complex I, complex II do not pump protons from matrix to intermembrane space. The energy released by electron transport is utilized to pump protons in complex III and IV, as described in the next section.

Complex III



Complex III, also known as cytochrome C reductase or cytochrome $\mathrm{bc_1}$ complex, contains protein subunits, an iron-sulfur cluster, and cytochromes. A cytochrome is a protein with at least one heme group that transfers one electron by accepting and releasing an electron based on the following reaction:

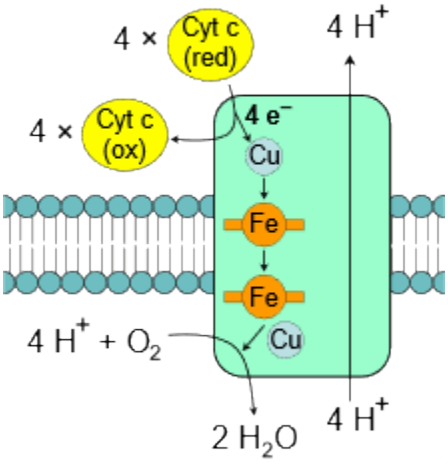
$$\mathrm{Fe^{3+} + e^- \rightleftharpoons Fe^{2+}}$$

Complex III passes on electrons from coenzyme-Q ($\mathrm{QH_2}$) to cytochrome c ($\mathrm{Cyt\ \ c}$). Since $\mathrm{QH_2}$ delivers two electrons ($\mathrm{e^-}$) while $\mathrm{Cyt\ \ c}$ can accept only one, the process happens in two steps, as illustrated in the figure on the right. In the first step, $\mathrm{QH_2}$ transfers one $\mathrm{e^-}$ to $\mathrm{Cyt\ \ c}$ and one to another coenzyme-Q in quinone (Q) form and releases $\mathrm{2\,H^+}$. Q converts to a semiubiquinone ($\mathrm{Q^{\bullet -}}$). In the second step, another $\mathrm{QH_2}$ transfers one electron to a second $\mathrm{Cyt\ \ c}$ and one to $\mathrm{Q^{\bullet -}}$ converting it to Q which is released from the complex and repeats the above two steps. This way, two electrons are transferred from a Q to two $\mathrm{Cyt\ \ c}$ and $\mathrm{4\,H^+}$ are pumped out into intermembrane space by complex III as shown in the following reaction.

$$\mathrm{QH_2 + 2\,Cyt\ \ c_{ox} + 2\,H^+_{matrix} \longrightarrow Q + 2\,Cyt\ \ c_{red} + 4\,H^+_{intermembrane}}$$

$\mathrm{Cyt\ \ c}$ carries the electron to complex IV.

Complex IV

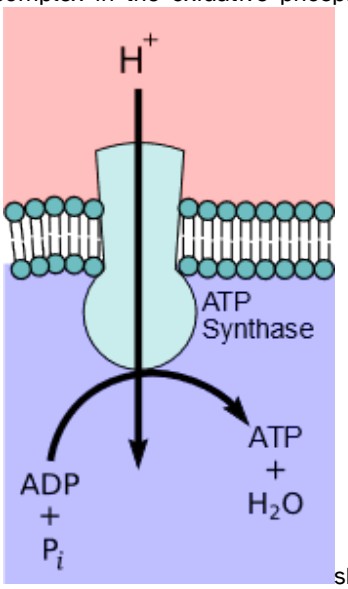


Complex IV, also known as cytochrome c oxidase, is the final complex in the electron transport chain. It contains several subunits, two heme groups, and several metal ion cofactors, including three atoms of copper, one magnesium, and one zinc. Complex IV reduces oxygen into water at the expense of oxidation of $\mathrm{Cyt\ c}$ and pumps $4\,\mathrm{H^+}$ from matrix to intermembrane space as illustrated in the figure on the right and shown in the following overall reaction.

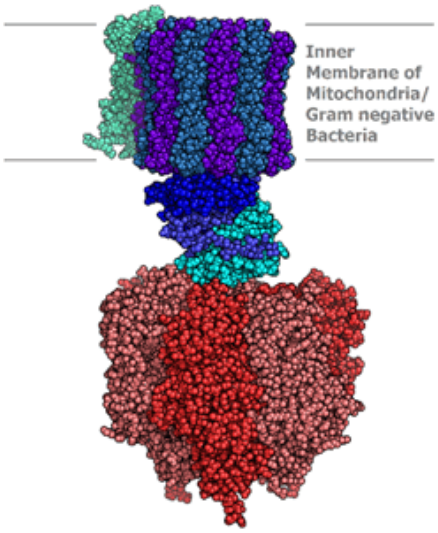
$$4\,\mathrm{Cyt\quad c_{red}} + \mathrm{O_2} + 8\,\mathrm{H^+_{matrix}} \longrightarrow 4\,\mathrm{Cyt\quad c_{ox}} + 2\,\mathrm{H_2O} + 4\,\mathrm{H^+_{intermembrane}}$$

ATP-synthase

ATP-synthase, also known as complex V is the final complex in the oxidative phosphorylation process, i.e., part 3 of stage



3 of food catabolism. It comprises two parts in the shape of a mushroom, as illustrated in the figure on the left. The first part, F_O, is the hydrophobic part embedded in the inner membrane (a blue



and purple region in the model image on the right (*Copyright; Alex. X, CC BY-SA 3.0, via Wikimedia Commons*). The F_O acts as a pore for the flow of protons (H^+) across the membrane. The second part is F_1 which is hydrophilic and spheroidal and protrudes into the matrix (red area in the image on the right). ATP synthesis takes place in the F_1 part of ATP-synthase. It operates on the principle of chemiosmosis, which is described below.

Chemiosmosis

Water movement from the direction of higher to lower concentration across a semipermeable membrane is called **osmosis.** When protons (H^+) are pumped from the matrix to intermembrane space by complex I, III, & IV, a proton concentration gradient is developed across the membrane. Since protons are positive charge ions (H^+), there is also an electrical gradient; collectively, it is called an electrochemical gradient. **Chemiosmosis** is the flow of ions (H^+) across a semipermeable membrane down their electrochemical gradient.

An electrochemical gradient developed due to protons (H^+) is a form of electrochemical potential energy. ATP-Synthase harnesses the electrochemical potential energy to synthesize ATP as the protons flow through it, just as electricity is generated as water from a dam flows through a turbine. The higher energy ATP is synthesized by condensing lower energy ADP and phosphate P_i in a process called **oxidative phosphorylation**. It takes three to four H^+ to synthesize one ATP by the following overall reaction.

$$ADP + P_i + 4\,H^+_{\text{intermembrane}} \rightleftharpoons ATP + H_2O + 4\,H^+_{\text{matrix}}$$

This is an equilibrium reaction, i.e., when the electrochemical gradient is low, ATP-synthase consumes ATP and returns protons from the matrix to the intermembrane space.

Structure and mechanisms of operation of ATP-synthase

ATP-synthase comprises several protein sub-units. It is in the shape of a mushroom with two major parts: a stem-like part which is a hydrophobic portion embedded in the inner membrane, is called **F_O,** and a mushroom head-like part that protrudes into the matrix, is called **F_1,** as illustrated in Figure $8.5.5.2$ a. Note: the subscript in F_O is the letter O, not the number zero.

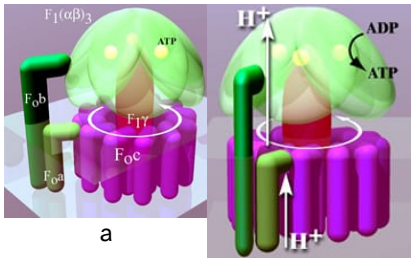


Figure 8.5.5.2: **a)** Simplified model of ATP-synthase (Copyright; Asw-hamburg, CC BY-SA 2.5, via Wikimedia commons), **b)** illustration proton flow in ATP-synthase (Copyright; Asw-hamburg at German Wikipedia., CC BY-SA 3.0, via Wikimedia Commons), **c)** Mechainsm of rotation of F_0c ring of ATP-synthase (Copyright: Asw-hamburg at German Wikipedia., CC BY-SA 3.0, via Wikimedia Commons), and **d)** Mechanism of ATP production in $F_1\beta$ subunits of ATP synthase. ATP is shown in red, ADP and phosphate in pink and the rotating γ subunit in black (Copyright; By US gov - US gov, Public Domain, https://commons.wikimedia.org/w/inde...?curid=3588598)

F_O portion of \(\ce{ATP}-synthase

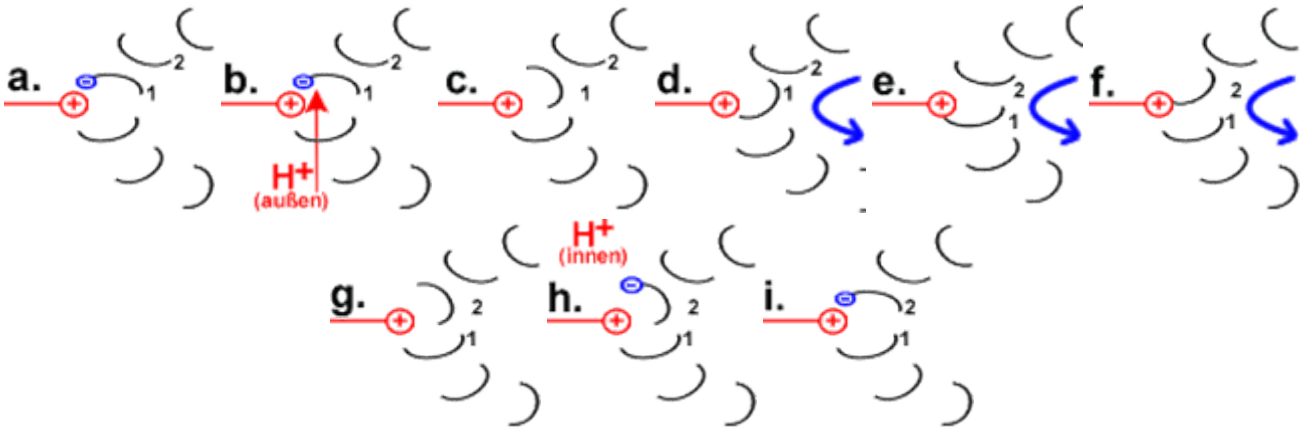
F_O part consists of six c subunits (F_Oc) arranged in a ring shape with proton channels between them that make the rotor part of this molecular machinery, as illustrated in Figure 8.5.5.2 a & b. The b subunit (F_Ob) connects to the F_1 portion and prevents it from rotating. The subunit F_Oa connects F_Ob to the F_Oc ring. F_Oc part is embedded in the membrane and couples the energy released as the protons flow through the channels to kinetic energy in the form of rotation of the F_Oc ring, like a turbine moves by water in a hydroelectric dam. Other subunits are not described here.

F_1 portion of \(\ce{ATP}-synthase

F_1 portion has a set of three α and thee β subunits arranged alternately like carpels of an orange ($F_1(\alpha\beta)_3$) with $F_1\gamma$ in the middle. The $F_1\gamma$ is like an axil connected with the rotary F_Oc, as shown in Figure 8.5.5.2 a. The $F_1(\alpha\beta)_3$ is prevented from rotating by the F_Ob subunit, but $F_1\gamma$ rotates along with the rotor F_Oc ring. The rotation of $F_1\gamma$ within $F_1(\alpha\beta)_3$ causes conformation changes in the β subunits, as illustrated in the video that will play by clicking this link. The conformational changes in the β subunits are linked to the mechanism of ATP synthesis. Other subunits of the F_1 portion are not described here.

Mechanism of F_Oc ring rotation in \(\ce{ATP}-synthase

The steps of the rotation of the F_Oc ring are illustrated below.



Steps in the rotation mechanism of F_Oc ring in ATP-synthase (Copyright; Asw-hamburg at German Wikipedia., CC BY-SA 3.0, via Wikimedia Commons)

Step a: F_Oc subunits are arranged in a circle. All of them have an amino acid asp residue with an acidic side chain but in neutral ($\text{peptide}-\text{COOH}$) form except one (labeled #1 in step a), which, being near a cation group ($\text{peptide}-\text{NH}_3^+$ of arg residue of F_Ob

subunit, is an anion, i.e., $\mathrm{peptide{-}COO^-}$. Electrostatic interaction between opposite charges builds mechanical tension in the F_Oc ring like a spiral spring.

1. Step b: A $\mathrm{H^+}$ from outside (intermembrance space) neutralizes the anion ($\mathrm{peptide{-}COO^- + H^+ \longrightarrow peptide{-}COOH}$).
2. Step c: The strained structure relaxes as the electrostatic tension is removed, causing the peptide to rotate.
3. Step d: Twisting motion bring F_Oc#1 back to the other side of the stator F_Ob $\mathrm{peptide{-}NH_3^+}$ group causing the rotor F_Oc to rotate by 30°.
4. Step e: F_Oc#1 subunit is neutral like other subunits in the ring.
5. Step f: The rotor F_Oc is rotated by 30°, and F_Oc#2 is under the spell of the positive charge of F_Ob $\mathrm{peptide{-}NH_3^+}$ group.
6. Step g: The asp residue of F_Oc#2 ionizes the \(\ce{-COOH) ($\mathrm{peptide{-}COOH \longrightarrow peptide{-}COO^- + H^+}$) and the spring comes under tension again.
7. Step h: The asp residue releases the $\mathrm{H^+}$ that moves inside (the matrix).
8. Step i: The initial position of step a is restored and the following has happed in the process:
    1. F_Oc#2 has taken the place of F_Oc#1,
    2. electrochemical energy of one $\mathrm{H^+}$ movement transferred to kinetic energy through 30° rotation of the rotor, and
    3. one $\mathrm{H^+}$ is smuggled from the intermembrane space to the matrix.

The simulation of the movement described above is illustrated in Figure $8.5.5.2$ c.

Mechanism of catalysis of ATP synthesis in $F_1(\alpha\beta)_3$ complex

The $F_1(\alpha\beta)_3$ complex catalyzes ATP synthesis. It has three α and three β subunits arranged alternately as three $\alpha\beta$ dimers in the shape of carpels in an orange. It is the β subunits that catalyze the ATP synthesis. Rotation of $F_1\gamma$ subunit inside the $F_1(\alpha\beta)_3$ complex causes conformational changes in the $F_1\beta$ subunits that are linked with the mechanism of ATP synthesis.

$F_1\beta$ switched between three states with each 360° rotation of $F_1\gamma$ subunit inside the $F_1(\alpha\beta)_3$ complex, as illustrated in Figure $8.5.5.2$ d.

1. First is the **open-state**, shown in brown, where ADP and $\mathrm{P_i}$ enter the active state.
2. Second is the **loose-state** in which the $F_1\beta$ closes up around the substrate ADP and $\mathrm{P_i}$ binding them loosely, shown in red.
3. Third, is the **tight-state** in which $F_1\beta$ tightens around the substrate ADP and $\mathrm{P_i}$ forcing them to condense into an ATP product.
4. Finally, $F_1\beta$ reverts to an open-state that releases the ATP and binds ADP and $\mathrm{P_i}$ to repeat the process.

Since one $\mathrm{H^+}$ cause a 30° turn, a complete (360°) turn of $F_1\gamma$ subunit inside $F_1(\alpha\beta)_3$ complex needs $12\,\mathrm{H^+}$ and produces three $\mathrm{ATP's}$ from three $F_1\beta$ subunits in it, that gives the following overall reaction.

$$\mathrm{ADP + P_i + 4\,H^+_{intermembrane} \rightleftharpoons ATP + H_2O + 4\,H^+_{matrix}}$$

The next video presents the oxidative phosphorylation of glucose in a summary form.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Heating the body

NADH is reduced in the electron transport chain at the expense of reduction of $\mathrm{O_2}$ by the following overall reaction.

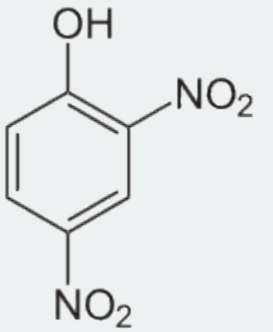
$$\frac{1}{2}\,O_2 + NADH + H^+ \longrightarrow H_2O + NAD^+$$

The potential difference between these redox pairs is 1.14 volts, equivalent to -218 kJ/mol. Reduction of one $NADH$ can produce three ATP. Production of ATP costs 30.5 kJ/mole, which is equivalent to 30.5 kJ/mole x 3 = 91.5 kJ/mole of $NADH$. So the percentage of energy conserved as ATP, i.e., the **energy efficiency** is: $\frac{91.5}{218} \times 100 = 42$. The remaining 58% of energy ends up heating the body.

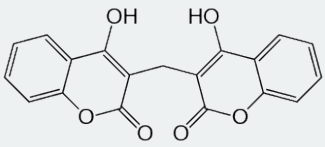


Some compounds, known as **uncouplers**, uncouple electron transport from ATP synthase, i.e., the electron transport pumps H^+ as usual but H^+ return to the matrix without producing $ATP's$. So, all energy released during the electron transport process becomes heat energy. Examples of uncouplers are 2,4-dinitrophenol and dicumarol that combine with H^+ and, being hydrophobic, carry them through the inner membrane. Some compounds, like oligomycin A prevent ATP synthesis by blocking H^+-channels in ATP synthase. Salicylic acid (aspirin), if taken in extreme excess, also blocks H^+-channels in ATP synthase. When the electron transport chain operates without ATP synthesis, all of the electron transport energy is used as heat.

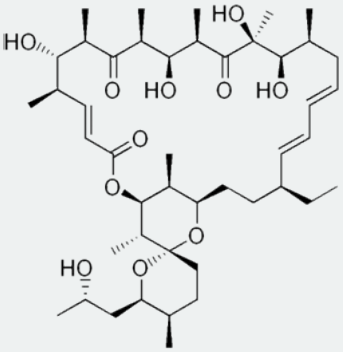
Certain animals adapted to the cold environment have developed uncoupling systems to generate more heat for heating their body. They contain a large amount of **brown fat,** a tissue with many mitochondria, and are brown due to iron in the cytochrome in the mitochondria. The electron transport works in the brown fat. Still, the mitochondria have certain proteins embedded in the inner membrane that allow H^+ to return to the matrix without ATP production. The body of newborn babies loses more heat per unit mass because of the larger surface area to mass ratio. Newborn babies have brown fat deposits on arteries that heat the blood circulating in their bodies. Adults usually do not have brown fat except those who live and work in a cold climate.



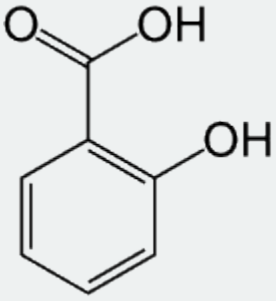
2,4-Dinitrophoenol



Dicumarol



Oligo
mycin A



Salicylic acid

Reactive oxygen species

The electron transport chain uses four electrons and four protons to reduce oxygen by the following overall reaction,

$$O_2 + 4\,e^- + 4\,H^+ \longrightarrow 2\,H_2O.$$

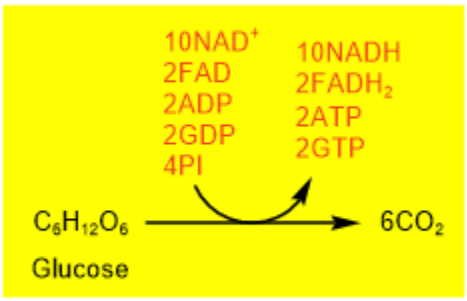
However, some electrons, particularly during the reduction of coenzyme-Q in the complex-III, leak and cause the following reaction.



$$O_2 \xrightarrow{e^-} \underset{\text{Superoxide}}{O_2^{\bullet -}} \xrightarrow{e^-} \underset{\text{Peroxide}}{O_2^{2-}}$$

Species like superoxide ($O_2^{\bullet -}$), peroxide (O_2^{2-}) and their product hydroxyl radical ($HO^{\bullet}$) are called **reactive oxygen species** which are harmful. They damage proteins, cause mutations by reacting with DNA, and are responsible for aging, as illustrated in the figure on the right (*taken from https://www.hiclipart.com/free-trans...-clipart-npvtr*). The body has mechanisms to suppress the production of reactive species and destroy them if formed. Production of reactive oxygen species increases with an increase in the inter-member potential. The reactive oxygen species, oxidants, activate uncoupling proteins that reduce the membrane potential. Some substances in the body, e.g., vitamins C, E and antioxidant enzymes, react with and destroy the reactive oxygen species.

Summary of glucose catabolism



Aerobic catabolism of a glucose molecule converts a $6\,C's$ glucose molecule to six carbon dioxide along with $ATP's$, $GTP's$, and reduced coenzyme $NADH's$, and $FADH_2$'s, as shown in the figure on the right.

Glucose first goes through glycolysis that produces two pyruvate and overall two ATP and two $NADH$. Two pyruvates go through oxidative decarboxylation producing two acetyl-CoA and two $NADH$. Each acetyl-CoA goes through one turn of the citric acid cycle. So, the two turns of the citric acid cycles convert two acetyl-CoA into four CO_2, six $NADH$, two $FADH_2$, and two GDP. These steps are summarized in Figure $8.5.5.3$ and Figure $8.5.5.4$ shown below.

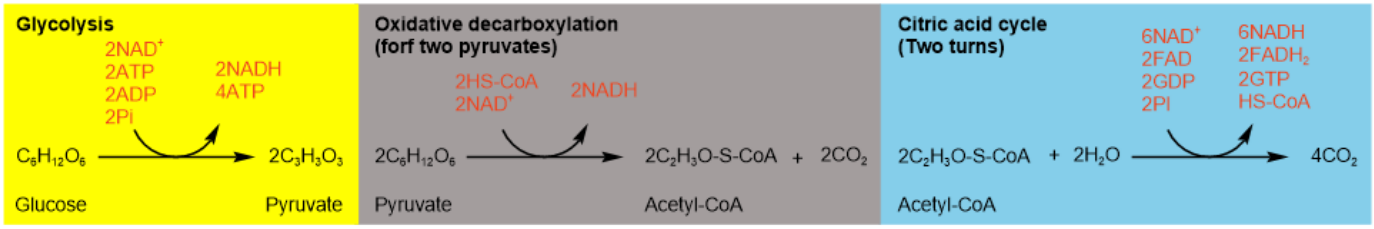


Figure $8.5.5.3$ illustrates the summary of aerobic catabolism of glucose and production of $ATP's$, $GTP's$, and reduced coenzyme $NADH's$, and $FADH_2$'s at different stages.

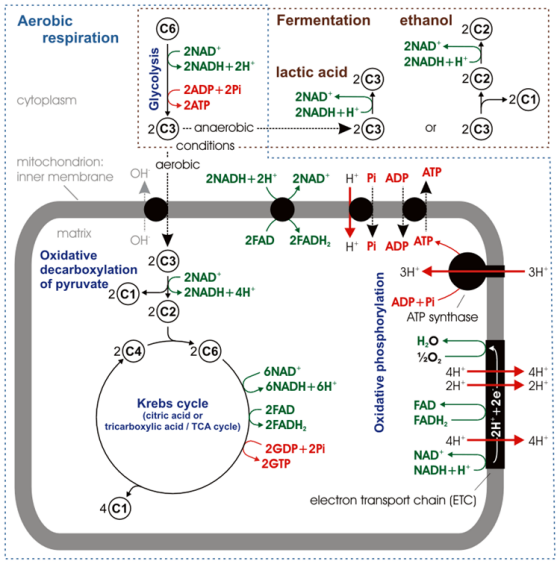


Figure 8.5.5.4: Stoichiometry of aerobic respiration and most known fermentation types in the eucaryotic cell. Numbers in circles indicate counts of carbon atoms in molecules; C6 is glucose $C_6H_{12}O_6$, and C1 carbon dioxide CO_2. The mitochondrial outer membrane is omitted. (Copyright; Darekk2, CC BY-SA 3.0, via Wikimedia Commons)

Insulin -the blood glucose-regulating hormone

Insulin is a hormone that regulates blood glucose levels. The glucose level increases after a meal, and the insulin promotes its absorption into the liver, fat, and muscle cells, converting it into glycogen or fats. High levels of insulin also inhibit the production of glucose by the liver. Low insulin levels have the opposite effects, which happen when a person has diabetes.

Comparison of ATP yield in aerobic and anaerobic catabolism of glucose

Anaerobic catabolism, i.e., the two forms of fermentation shown in Figure $8.5.5.3$ yields a net two $\mathrm{ATP's}$ from a glucose molecule.

Aerobic catabolism yields about thirty $\mathrm{ATP's}$. Two $\mathrm{ATP's}$ are produced before the citric acid cycle and two $\mathrm{GTP's}$, ten $\mathrm{NADH's}$, and two $\mathrm{FADH_2's}$ are produced during citric acid cycle. $\mathrm{GTP's}$ later convert to $\mathrm{ATP's}$. the reduced coenzymes, i.e., $\mathrm{NADH's}$ and $\mathrm{FADH_2's}$ enter the electron transport chain and produce heat and $\mathrm{ATP's}$ as explained in the next section.

Account of ATP production from complete oxidation of glucose

Two $\mathrm{ATP's}$ are produced before the citric acid cycle along with four NADH. The citric acid cycle yields $\mathrm{GTP's}$, six $\mathrm{NADH's}$, and two $\mathrm{FADH_2's}$. $\mathrm{GTP's}$ convert to $\mathrm{ATP's}$, i.e., one GTP is equivalent to one ATP. $\mathrm{NADH's}$ and $\mathrm{FADH_2's}$ enter the electron transport chain and produce heat and $\mathrm{ATP's}$. Recall that NADH is more energetic and pumps more proton through complex I, III, and IV than $\mathrm{FADH_2}$ which is less energetic and pumps less protons through complex III and IV only. Therefore, NADH yields more $\mathrm{ATP's}$ than $\mathrm{FADH_2}$. Earlier books state that one $\mathrm{NADH's}$ is equivalent to three $\mathrm{ATP's}$, but currently accepted average values are the following.

$$\mathrm{GTP} \approx 1\mathrm{ATP}$$

$$\mathrm{NADH} \approx 2.5\mathrm{ATP}$$

$$\mathrm{FADH_2} \approx 1.5\mathrm{ATP}$$

Based on these conversions, complete aerobic respiration of glucose produces approximately $32\,\mathrm{ATP's}$, as shown in the following equation and detailed in Table 1.

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 6\,O_2 + 10\,NAD^+ + 2\,FAD + \ + 2\,GTP + 2\,ADP + 4\,Pi \longrightarrow 6\,CO_2 + 6\,H_2O + 10\,NADH + 2\,FADH_2 + 2\,GTP + 2\,ATP$$

$$\underbrace{C_6H_{12}O_6}_{\text{Glucose}} + 6\,O_2 + 32\,ADP + 32\,Pi \longrightarrow 6\,CO_2 + 6\,H_2O + 32\,ATP$$

Table 1: ATP *Production from complete aerobic catabolism of one glucose molecule.*

Reaction	ATP or reduced coenzymes	Total ATP's output
Glycolysis		
glucose → glucose-6-phosphate		-1
fructose 6-phosphate → fructose 1,6-bisphosphate		-1
2 glyceraldehyde-3-phosphate → 2 1,3-bisphosphoglycerat	2 NADH	5
2 1,3-bisphosphoglycerate → 2 3-phosphoglycerate	2 ATP	2
2 phosphoenolpyruvate → 2 pyruvate	2 ATP	2
Oxidative phosphorylation of two pyruvate		
2 pyruvate → 2 acetyl-CoA	2 NADH	5
Two turns citric acid cycles to consume to 2 acetyl-CoA		
2 isocitrate → 2 α-ketoglutarate	2 NADH	5
2 α-ketoglutarate → 2 succinyl-CoA	2 NADH	5
2 succinyl-CoA → 2 succinate	2 GTP	2
2 succinate → 2 fumarate	2 $FADH_2$	3
2 malate → 2 oxaloacetate	2 NADH	5
Total		**32**

This page titled 8.5.5: Oxidative Phosphorylation is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9.5: Oxidative Phosphorylation by Muhammad Arif Malik is licensed Public Domain.

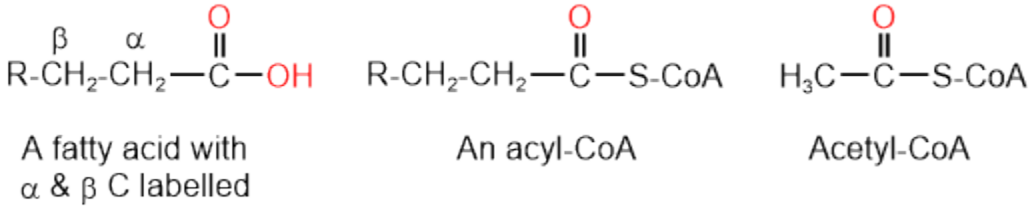
8.5.6: Oxidation of fatty acids

Learning Objectives

- Understand the catabolism of fatty acids, i.e., β-oxidation of fatty acids, the reactions involved in the process, the number of cycles needed, and the calculation of ATP yield per fatty acid.
- Understand obesity, ketogenesis, ketosis, ketoacidosis, and catabolism of glycerol, associated with β-oxidation of fatty acids.

β-Oxidation of fatty acids

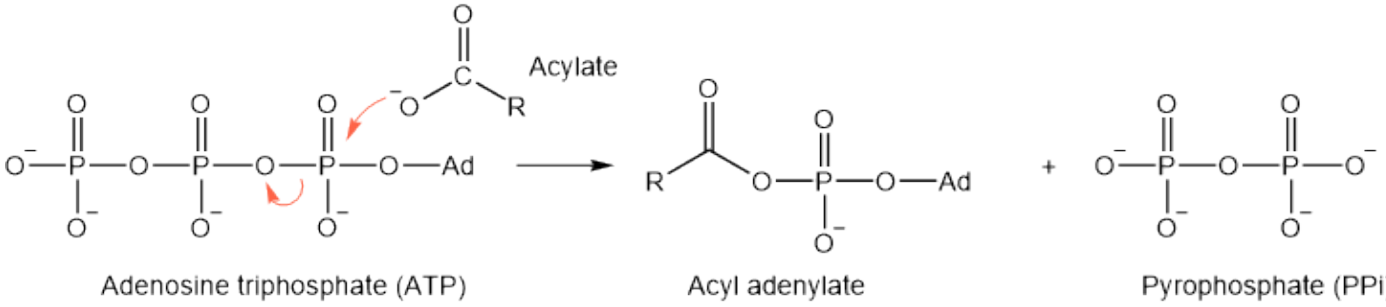
Although glucose is a quick energy source that animals need, fatty acids have higher energy per unit mass. Therefore, fats are the main energy storage compounds in animals. Fatty acids ($\mathrm{R{-}COOH}$) in fats are a major source of energy for animals. Hydrolysis of fats, i.e., the 1st stage of catabolism of fats, begins in the digestive tract and completes in the cytosol of the cells. The second stage of fatty acid catabolism starts with the conversion of ($\mathrm{R{-}COOH}$) to the thioester group of coenzyme A, ($\mathrm{R{-}CO{-}S{-}CoA}$), i.e., an acyl-CoA, at the cost of energy of ATP to AMP conversion. The thioester is called an activated fatty acid. Some of the important terms used in this section are illustrated in the figure below.



Acyl-CoA is transported from the cytosol into mitochondria for the second stage of its catabolism. In the second stage, the Acyl-CoA is fragmented into two $\mathrm{C's}$ fragments in the form of acetyl-CoA. The activated fatty acid goes through a set of four reactions in which the $\beta\mathrm{C}$ is oxidized to a carbonyl $\mathrm{C{=}O}$ group. Another coenzyme A makes thioester group with the β-$\mathrm{C{=}O}$, and an acetyl-CoA group is split off, leaving behind an acyl-CoA with two $\mathrm{C's}$ less than the initial acyl-CoA. This process is called **β-Oxidation of fatty acids**. The process repeats on the fragment acyl-CoA again and again until the last acyl-CoA fragment left is acetyl-CoA. The acetyl-CoA enters the citric acid cycle, which is the third stage of the catabolism, as in the case of the catabolism of glucose. These reactions are explained next.

Activation of fatty acids

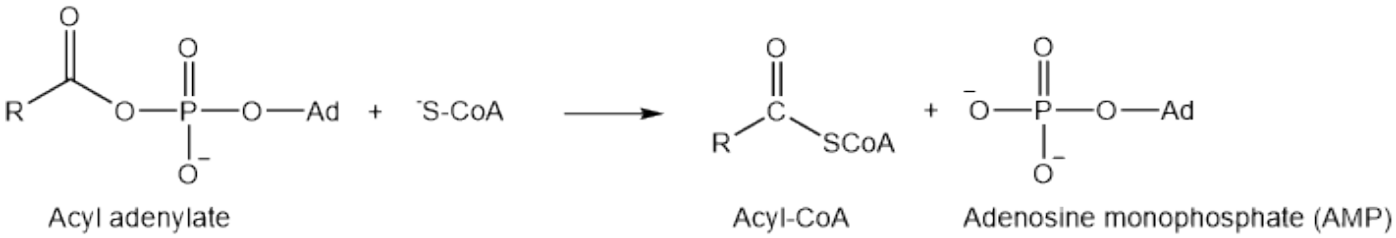
The activation of a fatty acid begins with an $\mathrm{SN_2}$ reaction of acylate anion as a nucleophile with ATP. It produces acyl-adenylate and pyrophosphate, as shown below.



Pyrophosphate is removed by hydrolysis, making the reaction irreversible: $\mathrm{PP_i + H_2O \longrightarrow 2\,P_i + 2\,H^+}$.

Acyl-adenylate goes through a nucleophilic acyl substitution reaction with coenzyme A ($\mathrm{HS{-}CoA}$) attacking as a nucleophile and

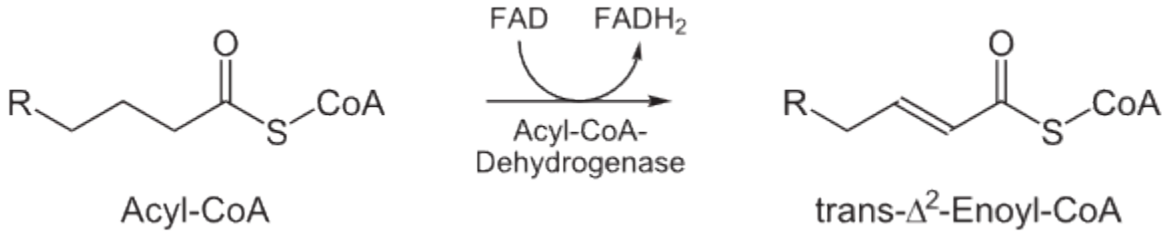
AMP as leaving group.



As described next, acyl-CoA product is transported from the cytosol into mitochondria for the β-oxidation process.

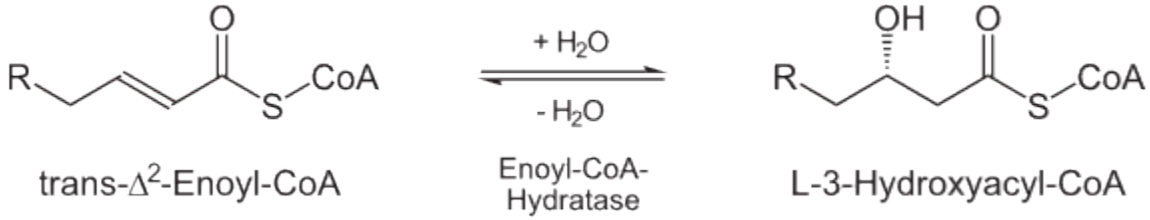
Reaction 1 -oxidation

Reaction 1 is the dehydrogenation of of acyl-CoA by enzyme acyl-CoA-dehydrogenase that removes H from α and β $\mathrm{C's}$ generating a trans $\mathrm{C{=}C}$ bond. An FAD is reduced to $\mathrm{FADH_2}$ as the acyl-CoA is oxidized in this reaction. Resonance stabilization of the $\mathrm{C{=}C}$ bond by conjugated $\mathrm{C{=}O}$ group makes the product stable, lowering the activation energy for the reaction.



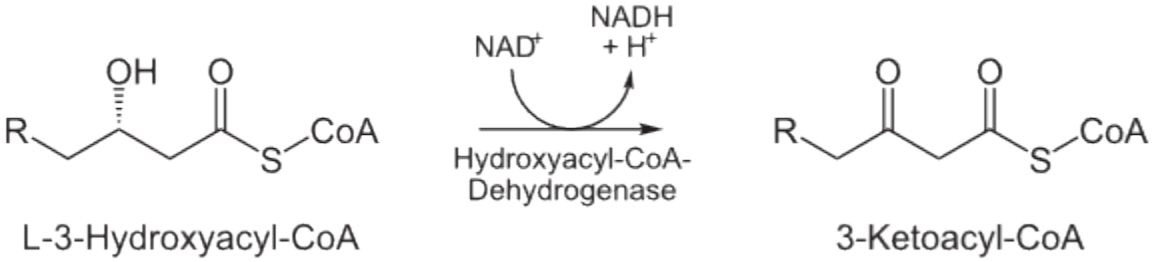
Reaction 2 -hydration

The $\mathrm{C{=}C}$ is hydrated by addition of $\mathrm{H_2O}$ that selectively installs an $-\mathrm{OH}$ group at $\beta\mathrm{C}$, creating a new chiral center in L-configuration. This is an electrophilic addition reaction. The $\mathrm{C{=}O}$ makes the $\beta\mathrm{C}$ more nucleophilic by withdrawing electrons from it by resonance. That is why the $\beta\mathrm{C}$ selectively reacts with the incoming $\mathrm{H_2O}$ electrophile.



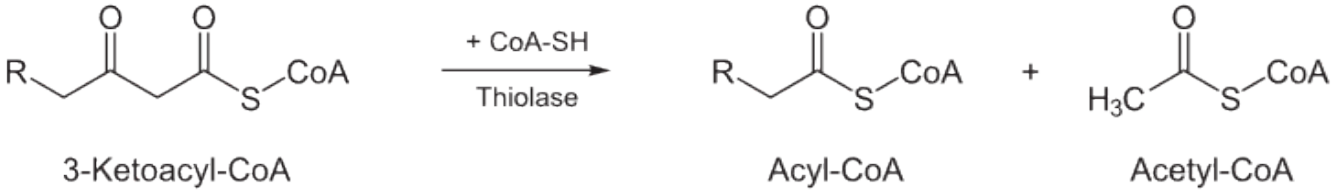
Reaction 3 -oxidation

The secondary $-\mathrm{OH}$ group is oxidized to a ketone ($\mathrm{C{=}O}$) group at the expense of reduction of $\mathrm{NAD^+}$ to NADH by the following reaction.



Reaction 4 -cleavage by thiolysis

The fourth reaction is a nucleophilic acyl substitution reaction in which an anion of coenzyme A ($^{-}\mathrm{S}-\mathrm{CoA}$) acts as a nucleophile, the ketone ($\mathrm{C}=\mathrm{O}$) as electrophile and $^{-}\mathrm{CH_2}-\mathrm{CO}-\mathrm{S}-\mathrm{CoA}$ as the leaving group. $^{-}\mathrm{CH_2}-\mathrm{CO}-\mathrm{S}-\mathrm{CoA}$ is a good leaving group because the negative charge is resonance stabilized and mainly resides on O of carbonyl ($\mathrm{C}=\mathrm{O}$) group. A proton later neutralizes the negative charge.



The second product of the above reaction is an acyl-CoA with two $2\,\mathrm{C's}$ less than the initial acyl-CoA.

Number of β-oxidation cycles

The product acyl-CoA of the first β-oxidation cycle goes through the process of four reactions again and again until the last acyl-CoA product is acetyl-CoA, i.e., $\mathrm{CH_3CO}-\mathrm{S}-\mathrm{COA}$. The process is illustrated in Figure $8.5.6.1$, with the help of an example of β -oxidation of stearic acid which is a typical $18\,\mathrm{C's}$ saturated fatty acid.

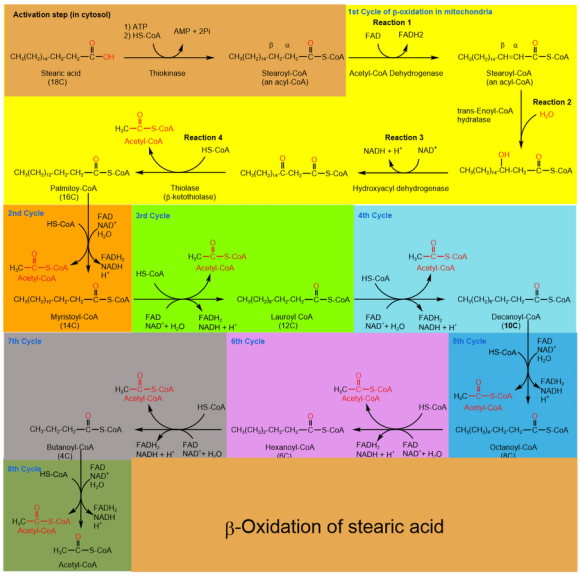


Figure $8.5.6.1$: β-Oxidation of fatty acids illustrated with the help of β-oxidation of stearic acid. (Copyright; Public domain)

If the starting fatty acid has $n\,\mathrm{C's}$, there are: $\frac{n}{2}$ acetyl-CoA produced. Since the last cycle produces two acetyl-CoA, the number of cycles for a fatty acid containing $n\,\mathrm{C's}$ is: $\frac{n}{2}-1$. For example, stearic acid has $18\,\mathrm{C's}$ and it goes through: $\frac{18}{2}-1=8$ β-oxidation cycles.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

ATP Yield from fatty acid oxidation

Each β-oxidation cycle of a fatty acid yields one $\mathrm{FADH_2}$, one NADH and one acetyl-CoA. These molecules produce $\mathrm{ATP's}$ when they enter the citric acid cycle and oxidative phosphorylation, as described earlier in the catabolism of glucose. One $\mathrm{FADH_2} \approx 1.5$ $\mathrm{ATP's}$, one $\mathrm{NADH} \approx 2.5$ $\mathrm{ATP's}$, and one acetyl-$\mathrm{CoA} \approx 10$ $\mathrm{ATP's}$ that makes 14 $\mathrm{ATP's}$ per β-oxidation cycle. Since there are $(\frac{n}{2} - 1)$ or $(0.5 \times n - 1)$ cycles for a fatty acid containing $n\,\mathrm{C's}$, there are: $((0.5 \times n - 1) \times 14))$ $\mathrm{ATP's}$ produced. The last cycle yields two acetyl-CoA, so there are addition $10\,\mathrm{ATP's}$ from the the last acetyl-CoA. The activation step converts one ATP to one AMP which is equivalent to conversion of $2\,\mathrm{ATP}$ to $2\,\mathrm{ADP}$. So, after adding ten from the last acetyl-CoA to the formula ($(0.5 \times n - 1)$ $\times 14)$ $\mathrm{ATP's}$ and subtracting 2 for the $2\,\mathrm{ATP's}$ consumed in the activation step, the formula for the number of $\mathrm{ATP's}$ produced per fatty acid containing $n\,\mathrm{C's}$ become the following:

$$\mathrm{ATP's\ produced\ per\ fatty\ acid\ containing}\ n\,\mathrm{C's} = ((0.5 \times n - 1) \times 14 + 10 - 2)\ \mathrm{ATP's}.$$

It simplifies to:

$$\mathrm{ATP's\ produced\ per\ fatty\ acid\ containing}\ n\,\mathrm{C's} = (7 \times n - 6)\ \mathrm{ATP's}.$$

For example, stearic acid contains $18\,\mathrm{C's}$ and produces $(7 \times 18 - 6) = 120$ $\mathrm{ATP's}$. So, one mole of stearic acid (284.48 g/mol) produces 120 mol $\mathrm{ATP's}$, which is significantly higher than 32 mol $\mathrm{ATP's}$ produced by one mole of glucose (180.156 g/mol). When converted to moles of $\mathrm{ATP's}$ produced per unit mass, stearic acid produces 0.42 mol $\mathrm{ATP's}$/g, which is more than two times higher than 0.18 mol $\mathrm{ATP's}$/g of glucose. This is because the average oxidation state of glucose $\mathrm{C's}$ is higher than those of fatty acid $\mathrm{C's}$. Glucose is a quick energy source, but animals use fats as relatively longer-term energy storage molecules due to their higher energy density.

Unsaturated fatty acids

Unsaturated fatty acids have $\mathrm{C{=}C}$ bond in their alkyl chain. β-Oxidation of unsaturated fatty acids happens the same way as for saturated fatty acids except for the following changes. If $\mathrm{C{=}C}$ bond is a *trans*-$\mathrm{C{=}C}$ bond between α and β $\mathrm{C'}$, reaction 1 in the β-oxidation is not needed and $\mathrm{FADH_2}$ from reaction 1 is not produced. If $\mathrm{C{=}C}$ bond is a cis-$\mathrm{C{=}C}$ bond or it is between β and γ $\mathrm{C'}$, It goes through isomerization reactions catalyzed by isomerase enzymes to convert the $\mathrm{C{=}C}$ bond into a *trans* $\mathrm{C{=}C}$ bond between α and β $\mathrm{C'}$ before reaction 3 happens on it. Unsaturated fatty acids yield a little less $\mathrm{ATP's}$ due to one $\mathrm{FADH_2}$ less produced pre $\mathrm{C{=}C}$ bond, but the difference is small and the formula: $(7 \times n - 6)$ $\mathrm{ATP's}$, still gives a reasonably accurate estimate of ATP yield.

Fat and obesity









Although carbohydrates and glucose are quick energy sources, fat has more energy per mass or volume. Storage of fats for long-term energy supply is an important survival feature for several animals. For example, hibernating animals like polar bear converts food into fats during summer when food is plenty and utilize stored fats to survive for months without food during hibernation in winter. Whales and penguins are kept warm by a layer of body fat called blubber and use it as an energy source to survive when food is unavailable. Camel store fat in their hump and can survive for months without food and water by utilizing the fat reserves. Migrator birds also store fat and use it as an energy source during migration.

Humans can store fats. In earlier times, major human food was vegetables, and fats in food accounted for about 20% of food calories, but these days more than 60% of food calories are from fats. The cumulation of too much fat is associated with overweight and obesity. **Obesity** is an excess accumulation of body fat that negatively affects health. It is measured in terms of **body mass index (BMI)**, which is the ratio of a person's mass to the square of a person's height. BMI 18.5 kg/m^2 to 24.9 kg/m^2 is normal, less is underweight, more is overweight, and over 30 kg/m^2 is obese, as shown in Figure 8.5.6.2. Obesity is associated with medical conditions like diabetes, high blood pressure, stroke, heart disease, gallstones, arthritis, and some cancers. According to NIH, nearly 3 out of 4 adults aged 20 or above in the US are overweight or obese. It is a preventable condition, as described in the following video message from NIH.

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

US national research authority video message.

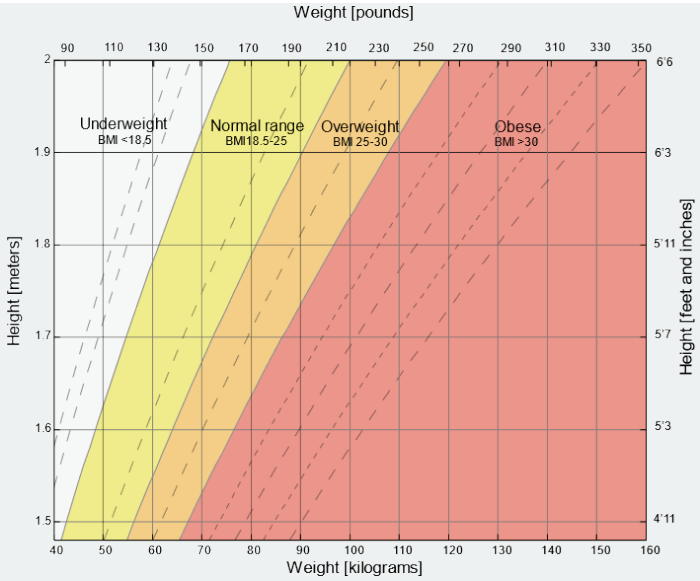
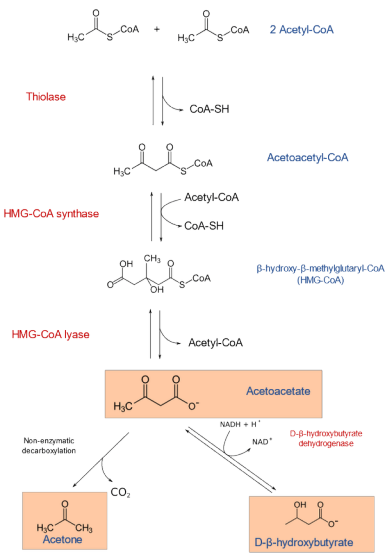


Figure 8.5.6.2: Body mass versus height chart showing ranges of underweight, normal weight, overweight, and obese regions based on body mass index (BMI) that can be read from the region marked by dotted lines in the chart at the cross-point of vertical lines starting from weight and horizontal lines starting from person's height. (Copyright; InvictaHOG, Public domain, via Wikimedia Commons)

Like insulin regulates blood glucose, **leptin** is a hormone adipose tissues produce. Leptin governs the long-term balance between food intake and energy expenditure. When there is more adipose tissue (fatty tissues), more leptin is released, which decreases hunger by giving a feeling of fullness, and less fatty tissue means less fat and less leptin, which has the opposite effect. Like diabetes disturbs the function of insulin, obesity may be related to the malfunctioning of leptin. This is an active area of research these days for finding obesity treatment.

Ketone bodies



Acetyl-CoA produced during the β-oxidation of fatty acids enter the citric acid cycle. When many fatty acids degrade, the citric acid cycle can not take all of the acetyl-CoA. The excess acetyl-CoA accumulates in the liver and converts into ketone bodies by a pathway known as **ketogenesis**, as shown in the figure on the right (*Copyright; Sav vas, CC0, via Wikimedia Commons*). In this figure, the enzymes are colored red, and the substrates/products are colored blue. The three **ketone bodies**, i.e., acetoacetate, acetone, and β-hydroxybutyrate, are marked within orange boxes.

Ketone bodies are transported from the liver to other tissues where acetoacetate and β-hydroxybutyrate are converted back to acetyl-CoA and enter the citric acid cycle to produce energy. Normally there is constant production and consumption of acetone bodies maintaining ~1 mg/dL concentration in blood.

Acetone cannot be converted back into acetyl-CoA except in the liver, where it can be converted into lactate and then to pyruvate.

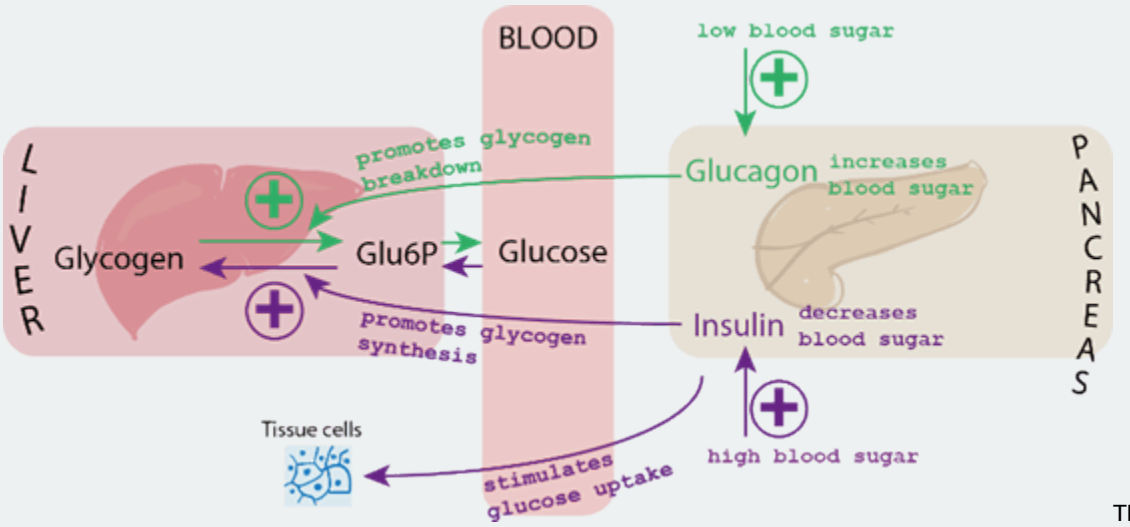
Ketosis



When the synthesis rate of ketone bodies exceeds their consumption rate, they start accumulating and may start excreting in urine and via breathing -a condition called **ketosis**. Since two of the three ketone bodies are acids, their accumulation in the blood lowers the blood pH -a condition called **ketoacidosis**. The smell of acetone, fruity or like nail polish remover smaller, is detectable in the breath of persons suffering from ketosis. Ketone bodies can be tested in blood, urine, or breath, e.g., by using test strips with a color chart to read the results, as shown in the figure on the left. In ketosis, the ketone bodies in the blood are between 0.5 to 3.0 millimole/L (mmol/L). In ketoacidosis, the concentration may go above 10 mmol/L.

Ketosis can happen during fasting, starvation, prolonged low carbohydrate diets, prolonged intense exercises, alcoholism, or due to uncontrolled diabetes.

Diabetes and ketone bodies

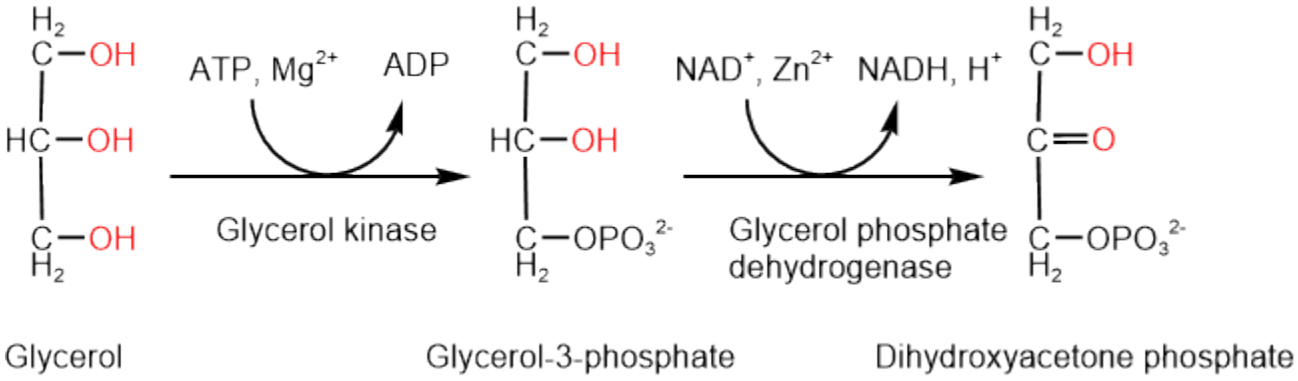


The liver and pancreas play important roles in maintaining blood glucose levels, illustrated in the figure on the right (*copyright: C. Muessig, CC BY-SA 3.0, via Wikimedia Commons*) and described next. The normal glucose level in the blood is 4.5 to 5.5 mM. When glucose level is above normal, usually after a meal, the pancreas senses it and secrets **insulin** that increases the flow of glucose into muscles and fatty tissues, where it is converted into glycogen, lowering blood glucose. When the blood glucose level is low, the pancreas secretes another hormone, **glucagon**, into the bloodstream. Glucagon triggers glycogen breakdown, particularly in the liver, releasing glucose into the bloodstream.

In **diabetes**, there is either insufficient insulin or not functioning properly. Less glucose is sent to muscles and, as a result, muscles cause β-oxidation of fatty acids to meet the energy needs. Higher levels of acetyl-CoA from the β-oxidation produce more ketone bodies than their consumption, accumulating ketone bodies. It causes **ketosis** or **ketoacidosis**. Low pH of blood plasma due to ketoacidosis triggers kidneys to excrete urine with high acid levels. Glucose and ketone bodies also spill into the urine. High glucose concentration in the blood causes higher osmotic pressure, increasing dehydration. The symptoms include frequent urination and excessive thirst. These conditions can be treated with low carbohydrate diets, insulin therapy, or other medications to control diabetes.

Catabolism of glycerol

Glycerol, a part of triglycerides (fats), accounts for about 5% of their energy. Glycerol is first converted into glycerol-3-phosphate by an SN_2 reaction of a primary alcohol group of glycerol acting as a nucleophile, phosphorus in the phosphate of ATP as a nucleophile, and ADP as a leaving group. Then, the secondary alcohol of glycerol is oxidized to a ketone group at the expense of reduction of a NAD^+ coenzyme into NADH as shown below.



Dihydroxyacetone phosphate produced by the second reaction is an intermediate in the glycolysis of glucose and enters the same

catabolic pathway.

This page titled 8.5.6: Oxidation of fatty acids is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9.6: Oxidation of fatty acids by Muhammad Arif Malik is licensed Public Domain.

8.5.7: Degradation of amino acids

Learning Objectives

- Understand the catabolism of amino acids, including transamination, oxidative amination, and urea cycle that takes care of the N and processing of the C skeleton of the amino acid to intermediates that enter into citric acid cycle for energy production.

Amino acids are the products of stage 1 of protein catabolism. Usually, amino acids are used to synthesize proteins and other substances that need nitrogen, e.g., nucleotides. Only about 10% of our energy is usually derived from amino acids. Still, when there is a shortage of carbohydrates and fats, e.g., during fasting or starvation, amino acids are used as an alternate energy source. Prolonged use of amino acids as an energy source may lead to the destruction of essential tissues.

Degradation of amino acids has two sets of catabolic pathways: one that deals with the processing of N and the other that deals with the processing of the remaining C skeleton of the amino acids.

Processing of N of amino acids

Processing of N usually happens in the liver and has three major stages: i) transamination, ii) oxidative deamination, and iii) urea cycle, as illustrated with some examples in Figure $8.5.7.1$.

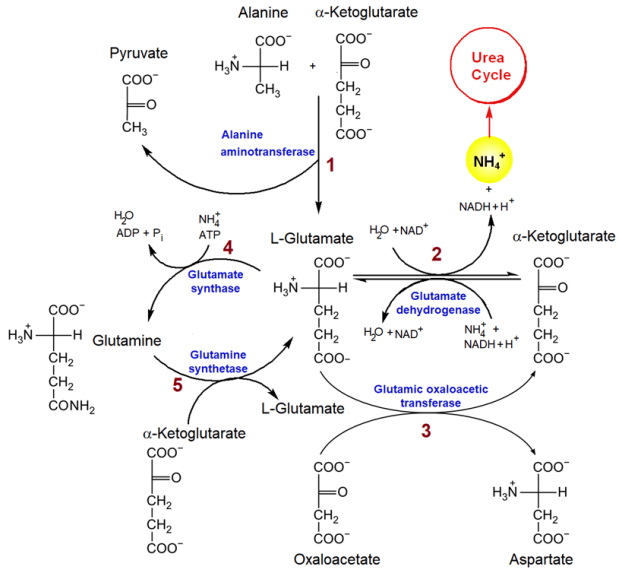


Figure $8.5.7.1$: Processing of N of amino acids Illustrated, including examples of transamination (Reactions 1, 3), oxidative deamination (reaction 2), and some other ways of $-\mathrm{NH_2}$ transfer (reactions 4 and 5) and production of $\mathrm{NH_4}^+$ (highlighted yellow) that enters the urea cycle. (Copyright; Public domain)

Transamination

Transamination is the process of exchange of an ammonium ($-\mathrm{NH_3}^+$) group of amino acids with ketone ($\mathrm{C{=}O}$) group of an α -ketoacid. Usually α-amino acids exchange $-\mathrm{NH_3}^+$ group with ketone ($\mathrm{C{=}O}$) group of an α-ketoglutarate. For example, alanine exchanges its $-\mathrm{NH_3}^+$ group with a $\mathrm{C{=}O}$ of an α-ketoglutarate producing a new α-keto acid (pyruvate in this example) and a new α-amino acids (L-glutamate in this example), as shown in reaction 1 in Figure $8.5.7.1$. Other α-keto acids can also receive $-\mathrm{NH_3}^+$ of amino acids also. For example, L-glutamate can transfer its $-\mathrm{NH_3}^+$ to oxaloacetate that regenerates α-ketoglutarate and produces aspartate, as shown in reaction 3 in Figure $8.5.7.1$.

Oxidative deamination

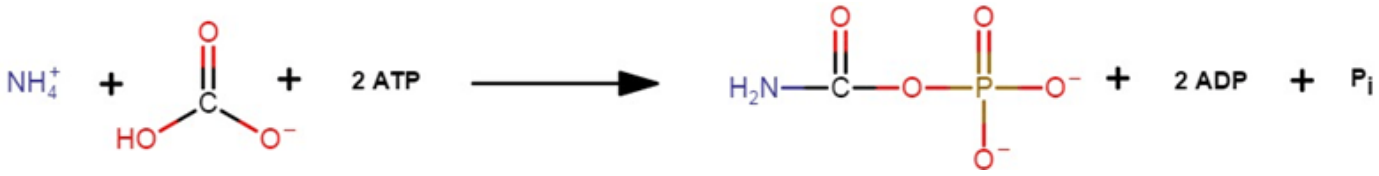
Oxidative deamination reaction replaces $-NH_3^+$ with a $C{=}O$, producing an α-keto acid and an ammonium (NH_4^+) ion, as shown in reaction 2 in Figure 8.5.7.1. This is an oxidation reaction that is coupled with the reduction of NAD^+ to $NADH$. It happens in mitochondria where $NADH$ enters oxidative phosphorylation pathway to produce $ATP's$. The other product, i.e., NH_4^+, is a toxic substance that needs to be disposed off. Some of NH_4^+ may be used in anabolic reactions, e.g., reactions 4 and 5 in Figure 8.5.7.1, but most of it enters the urea cycle.

Urea cycle

Ammonium ion NH_4^+ is a toxic substance that is converted to less toxic urea in the urea cycle by the following overall reaction.

$$2\,NH_4^+ + CO_2 \longrightarrow \underbrace{H_2N{-}\overset{\overset{O}{\|}}{C}{-}NH_2}_{\text{Urea}} + 2\,H^+ + H_2O.$$

Before entry into the urea cycle, NH_4^+ reacts with two $ATP's$ and bicarbonate (HCO_3^-), producing carbamoyl phosphate, two $ADP's$ and phosphate (P_i, as shown below.



This reaction is catalyzed by an enzyme carbamoyl phosphate synthetase I. Blood contains HCO_3^- which is the product of dissolution of CO_2. The product of the above reaction, i.e., Carbamoyl phosphate, goes through four reactions called the **urea cycle** illustrated in Figure 8.5.7.2.

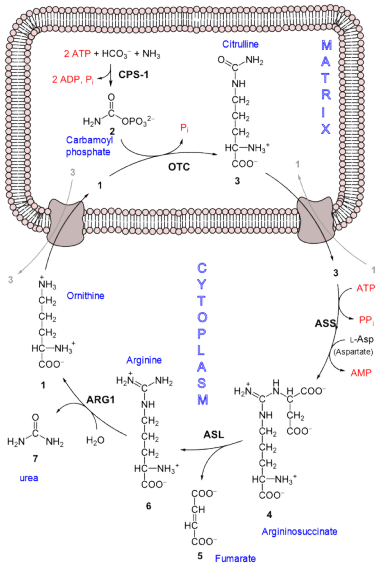
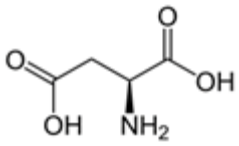


Figure 8.5.7.2: Urea cycle: CPS-1 is carbamoyl phosphate synthetase-1, OTC is ornithine transcarbamoylase, ASS is argininosuccinate synthetase, ASL is argininosuccinate lyase, and ARG1 is Arginase. Note: Inner membrane is shown, but the outer membrane of the mitochondrion is ignored in this diagram. (Copyright; Pk0001, CC BY-A 4.0, via Wikimedia Commons)

- The first reaction of the urea cycle is catalyzed by ornithine transcarbamylase. The $-COO^-$ group of ornithine substitutes phosphate (P_i from carbamoyl phosphate producing citrulline in mitochondria. Citrulline moves out from the matrix into the cytoplasm.



L-Aspartate

- The second reaction is a condensation reaction between the carbonyl ($C{=}O$) group of citrulline and $-NH_3^+$ group of aspartate that produces argininosuccinate. This reaction takes place in the cytosol and is catalyzed by argininosuccinate synthetase.
- In the third reaction, argininosuccinate is cleaved by the enzyme argininosuccinase to form arginine, which stays in the cycle, and fumarate, which leaves the cycle.
- In the fourth reaction, arginase cleaves arginine to form urea and ornithine. Ornithine is transported back to the matrix in mitochondria to start the next urea cycle, and urea is released into the blood.

Since HCO_3^- is produced by the dissolution of CO_2 in water, the NH_4^+ + HCO_3^- are equivalent to NH_3 + CO_2 + H_2O. So, the overall equation of the urea cycle becomes:

$$NH_3 + CO_2 + H_2O + \text{aspartate} + 3\,ATP + 3\,H_2O \longrightarrow \text{urea} + \text{fumarate} + 2\,ADP + 2\,P_i + AMP + PP_i + H_2O$$

Since NH_3 is removed to convert aspartate into fumarate along with $PP_i + H_2O \longrightarrow 2\,P_i$, substituting these for aspartate and fumarate simplifies the above equation into:

$$2\,NH_3 + CO_2 + 3\,ATP + 3\,H_2O \longrightarrow \text{urea} + 2\,ADP + 4\,P_i + AMP$$

One $NADH$ is produced during oxidative deamination of L-glutamate (reaction 2 in Figure $8.5.7.1$). Another $NADH$ is produced when fumarate from reaction 3 of the urea cycle is processed in the citric acid cycle, i.e., fumarate is converted into malate by enzyme fumarase and then malate is oxidized to oxaloacetate at the expense of reduction of NAD^+ to $NADH$ by enzyme malate dehydrogenase. Adding these reactions to the above reaction of the urea cycle results in the following overall reaction.

$$CO_2 + \text{glutamate} + \text{aspartate} + 3\,ATP + 2\,NAD^+ + 3\,H_2O \longrightarrow \text{urea} + \alpha\text{-ketoglutarate} + \text{oxaloacetate} + 2\,ADP + 2\,P_i + AMP + 2\,PP_i + 2\,NADH$$

Three ATP are consumed, one of them to AMP that makes total equivalent to four $ATP's$ consumed but two $NADH$ enter oxidative phosphorylation and yield ~5ATP, i.e., there is net production of an ATP in the process of N of an amino acid.

The urea released into the blood is later filtered out by the kidneys and excreted with urine. An adult passes about 25 to 30 g of urea in urine per day. If urea is not eliminated properly, it builds to toxic levels and needs medical treatment, like dialysis. Reducing protein intake also lowers urea output.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Urea cycle

The link between the urea cycle and the citric acid cycle

The urea cycle and citric acid cycle are two separate cycles but are linked with each other. Aspartate that provides one N in the urea cycle is produced by oxaloacetate transamination, which is an intermediate in the citric acid cycle. Fumarate is a product of the urea cycle that is an intermediate of the citric acid cycle and returns to it.

processing of C skeleton of amino acids

α-Ketoacids are left as C skeleton of amino acid after transamination or oxidative deamination. For example, pyruvate is left after transamination of alanine and α-ketoglutarate is left after oxidative deamination of glutamate, as shown in reactions 1 and 2, respectively, in Figure $8.5.7.1$. α-Ketoacids are either used as precursors for the synthesis of other compounds, or they enter the citric acid cycle to produce $\mathrm{CO_2}$, $\mathrm{H_2O}$, and energy. α-Ketoglutarate and oxaloacetate are α-keto acids that are intermediates in the citric acid cycle and directly enter the cycle. Other α-keto acids go through a series of reactions to convert into one of the intermediates in the citric acid cycle, or they convert to pyruvate or acetyl-CoA that enter the citric acid cycle, as illustrated in Figure $8.5.7.3$. Details of the conversation of α-keto acids into intermediates of the citric acid cycle or pyruvate or acetyl-CoA are not described here.

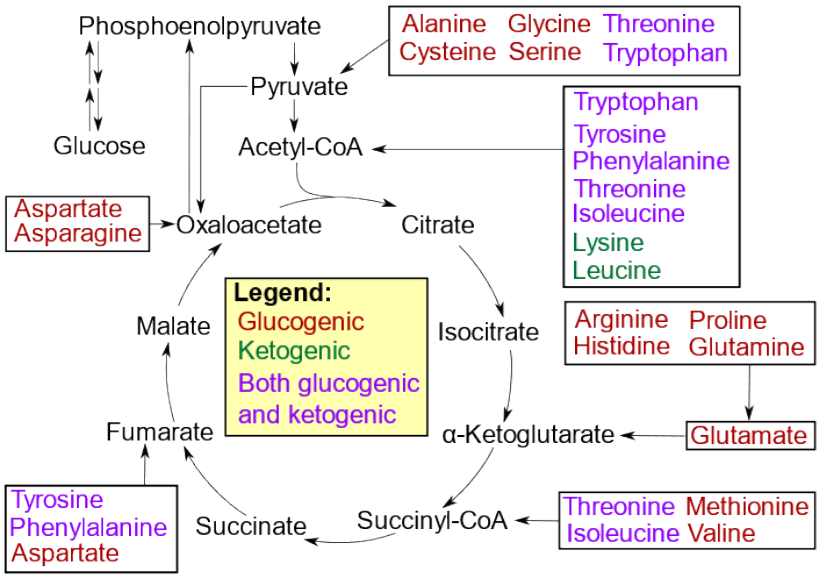


Figure $8.5.7.3$: Illustration of the feeding point of amino acid degradation products into the citric acid cycle and classification of amino acids as glucogenic, ketogenic, or both. (Copyright; Mikael Häggström, CC0, via Wikimedia Commons)

Amino acids that can degrade to pyruvate or oxaloacetate are called **glucogenic** because these products can form glucose through the glucogenesis pathway. For example, alanine yields pyruvate, and aspartate yield oxaloacetate, as shown in reactions 1 and 3 in Figure $8.5.7.1$. Amino acids that degrade to acetyl-CoA or acetoacetic acid, which can not form glucose but can be converted into ketone bodies, are called **ketogenic**. Lysine and leucine are ketogenic amino acids. Several amino acids can catabolize to produce both glucogenic and ketogenic intermediates, and they fall into both classes, as shown in Figure $8.5.7.3$.

This page titled 8.5.7: Degradation of amino acids is shared under a Public Domain license and was authored, remixed, and/or curated by Muhammad Arif Malik.

- 9.7: Degradation of amino acids by Muhammad Arif Malik is licensed Public Domain.

9: QUANTUM BASICS

9.1: Atoms and Photons- Origin of the Quantum Theory

9.2: Waves and Particles

9.3: The Free Particles

9.4: The Particle in a Box

9.5: Quantum Mechanics of Some Simple Systems

9.6: Principles of Quantum Mechanics

9.7: Harmonic Oscillator

9.8: Angular Momentum

9.9: Hydrogen Atom

9.10: Helium Atom

9.11: Atomic Structure and The Periodic Law

9.12: The Chemical Bond

9.13: Molecular Orbital Theory

9.14: Molecular Symmetry

9.15: Molecular Spectroscopy

9.16: Nuclear Magnetic Resonance

This page titled 9: Quantum Basics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

9.1: Atoms and Photons- Origin of the Quantum Theory

Atomic and Subatomic Particles

The notion that the building blocks of matter are invisibly tiny particles called atoms is usually traced back to the Greek philosophers Leucippus of Miletus and Democritus of Abdera in the 5th Century BC. The English chemist John Dalton developed the atomic philosophy of the Greeks into a true scientific theory in the early years of the 19th Century. His treatise *New System of Chemical Philosophy* gave cogent phenomenological evidence for the existence of atoms and applied the atomic theory to chemistry, providing a physical picture of how elements combine to form compounds consistent with the laws of definite and multiple proportions. Table 9.1.1 summarizes some very early measurements (by Sir Humphrey Davy) on the relative proportions of nitrogen and oxygen in three gaseous compounds.

Table 9.1.1: Oxides of Nitrogen

Compound	Percent N	Percent O	Ratio
I	29.50	70.50	0.418
II	44.05	55.95	0.787
III	63.30	36.70	1.725

We would now identify these compounds as NO_2, NO and N_2O, respectively. We see in data such as these a confirmation of Dalton's atomic theory: that compounds consist of atoms of their constituent elements combined in small whole number ratios. The mass ratios in Table 9.1.1 are, with modern accuracy, 0.438, 0.875 and 1.750.

After over 2000 years of speculation and reasoning from indirect evidence, it is now possible in a sense to actually *see* individual atoms, as shown for example in Figure 9.1.1. The word "atom" comes from the Greek *atomos*, meaning literally "indivisible." It became evident in the late 19th Century, however, that the atom was *not* truly the ultimate particle of matter. Michael Faraday's work had suggested the electrical nature of matter and the existence of subatomic particles. This became manifest with the discovery of radioactive decay by Henri Becquerel in 1896 the emission of alpha, beta and gamma particles from atoms. In 1897, J. J. Thompson identified the electron as a universal constituent of all atoms and showed that it carried a negative electrical charge, now designated *-e*.

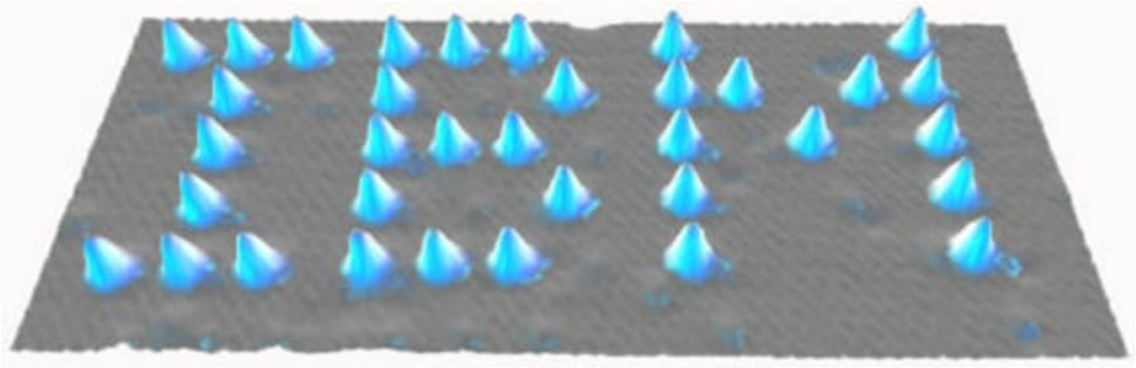


Figure 9.1.1: Image showing electron clouds of individual xenon atoms on a nickel(110) surface produced by a scanning tunneling microscope at IBM Laboratories.

To probe the interior of the atom, Ernest Rutherford in 1911 bombarded a thin sheet of gold with a stream of positively-charged alpha particles emitted by a radioactive source. Most of the high-energy alpha particles passed right through the gold foil, but a small number were strongly detected in a way that indicated the presence a small but massive positive charge in the center of the atom (Figure 9.1.2). Rutherford proposed the nuclear model of the atom. As we now understand it, an electrically-neutral atom of atomic number *Z* consists of a nucleus of positive charge *+Ze*, containing almost the entire the mass of the atom, surrounded by *Z* electrons of very small mass, each carrying a charge *-e*. The simplest atom is hydrogen, with *Z = 1*, consisting of a single electron outside a single proton of charge *+e*.

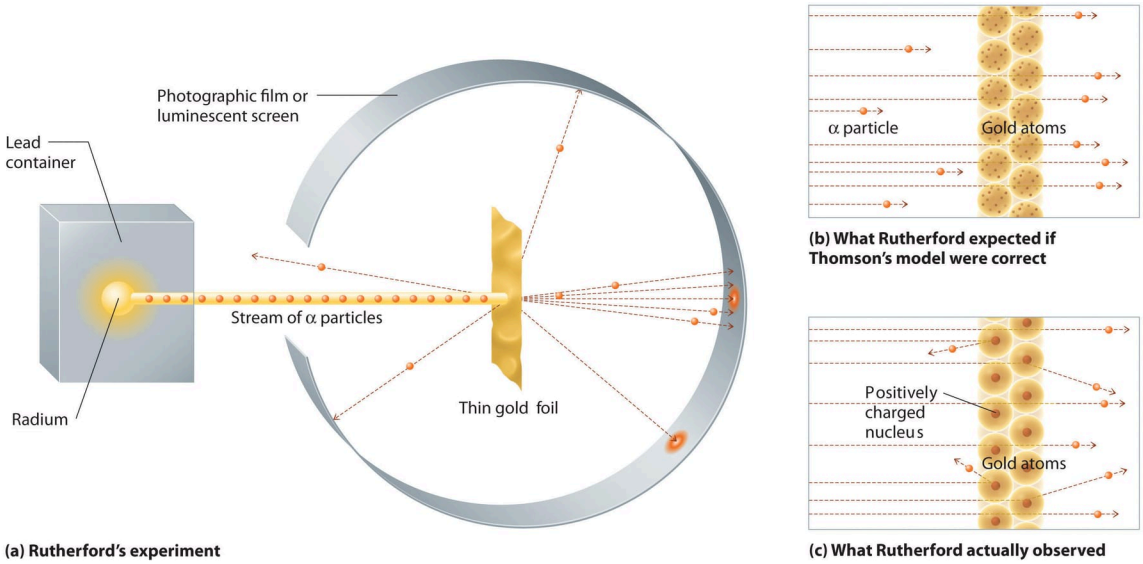


Figure 9.1.2: A Summary of Rutherford's Experiments*(a) A representation of the apparatus Rutherford used to detect deflections in a stream of α particles aimed at a thin gold foil target. The particles were produced by a sample of radium. (b) If Thomson's model of the atom were correct, the α particles should have passed straight through the gold foil. (c) But a small number of α particles were deflected in various directions, including right back at the source. This could be true only if the positive charge were much more massive than the α particle. It suggested that the mass of the gold atom is concentrated in a very small region of space, which he called the nucleus.*

With the discovery of the neutron by Chadwick in 1932, the structure of the atomic nucleus was clarified. A nucleus of atomic number *Z* and mass number *A* was composed of *Z* protons and *A*-*Z* neutrons. Nuclei diameters are of the order of several times 10^{-15}m. From the perspective of an atom, which is 10^{5} times larger, a nucleus behaves, for most purposes, like a point charge +*Ze*.

During the 1960's, compelling evidence began to emerge that protons and neutrons themselves had composite structures, with major contributions by Murray Gell-Mann. According to the currently accepted "Standard Model," the protons and neutron are each made of three *quarks*, with compositions *uud* and *udd*, respectively. The up quark *u* has a charge of $+\frac{2}{3}e$, while the down quark *d* has a charge of $-\frac{1}{3}e$. Despite heroic experimental efforts, individual quarks have never been isolated, evidently placing them in the same category with magnetic monopoles. By contrast, the electron maintains its status as an indivisible elementary particle.

Electromagnetic Waves

Perhaps the greatest achievement of physics in the 19th century was James Clerk Maxwell's unification in 1864 of the phenomena of electricity, magnetism and optics. An (optional) summary of Maxwell's equations is given in Supplement 1A. Heinrich Hertz in 1887 was the first to demonstrate experimentally the production and detection of the electromagnetic waves predicted by Maxwell-specifically radio waves-by acceleration of electrical charges. As shown in Figure 9.1.3, electromagnetic waves consist of mutually perpendicular electric and magnetic fields, **E** and **B** respectively, oscillating in synchrony at high frequency and propagating in the direction of **E** x **B**.

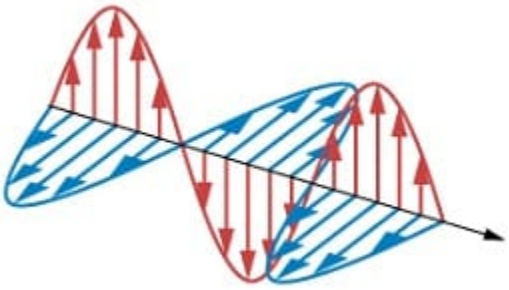


Figure 9.1.3: Schematic representation of monochromatic linearly-polarized electromagnetic wave.

The wavelength λ is the distance between successive maxima of the electric (or magnetic) field. The frequency ν represents the number of oscillations per second observed at a fixed point in space. The reciprocal of frequency $\tau = \frac{1}{\nu}$ represents period of oscillation-the time it takes for one wavelength to pass a fixed point. The speed of propagation of the wave is therefore determined by $\lambda = c\tau$ or in more familiar form

$$\lambda\nu = c \tag{9.1.1}$$

where $c = 2.9979 \times 10^8 m/sec$, usually called the *speed of light*, applies to all electromagnetic waves in vacuum. Frequencies are expressed in hertz (Hz), defined as the number of oscillations per second.

Electromagnetic radiation is now known to exist in an immense range of wavelengths including gamma rays, X-rays, ultraviolet, visible light, infrared, microwaves and radio waves, as shown in Figure $9.1.4$.

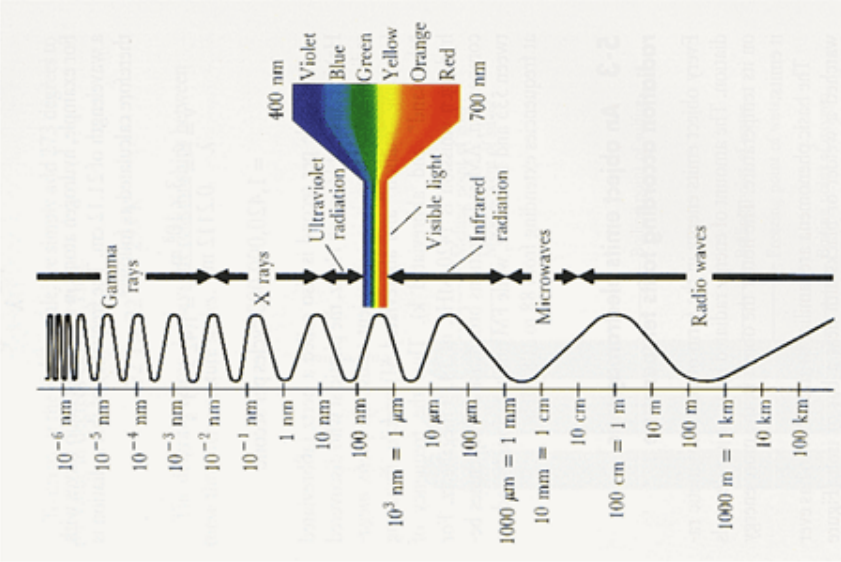


Figure $9.1.4$. The electromagnetic spectrum, showing wavelengths of different types of radiation. Adapted from R. A. Freedman and W. J. Kaufmann III, Universe (Freeman, New York, 2001).

Three Failures of Classical Physics

Isaac Newton's masterwork, *Pincipia*, published in 1687, can be considered to mark the beginning of modern physical science. Not only did Newton delineate the fundamental laws governing motion and gravitation but he established a general philosophical worldview which pervaded all scientific theories for two centuries afterwards. This system of thinking about the physical world is known as "Classical Physics." Its most notable feature is the primacy of cause and effect relationships. Given sufficient information about the present state of part of the Universe, it should be possible, at least in principle, to predict its future behavior (as well as its complete history.) This capability is known as *determinism*. For example, solar and lunar eclipses can be predicted centuries ahead, within an accuracy of several seconds. (But interestingly, we can't predict even a couple of days in advance if the weather will be clear enough to view the eclipse!) The other great pillar of classical physics is Maxwell's theory of electromagnetism.

The origin of quantum theory can be marked by three diverse phenomena involving electromagnetic radiation, which could *not* be adequately explained by the methods of classical physics. First among these was blackbody radiation, which led to the contribution of Max Planck in 1900. Next was the photoelectric effect, treated by Albert Einstein in 1905. Third was the origin of line spectra, the hero being Neils Bohr in 1913. A coherent formulation of quantum mechanics was eventually developed in 1925 and 1926, principally the work of Schrödinger, Heisenberg and Dirac. The remainder of this Chapter will describe the early contributions to the quantum theory by Planck, Einstein and Bohr.

Blackbody Radiation

It is a matter of experience that a hot object can emit radiation. A piece of metal stuck into a flame can become "red hot." At higher temperatures, its glow can be described as "white hot." Under even more extreme thermal excitation it can emit

predominantly blue light (completing a very patriotic sequence of colors!). Josiah Wedgwood, the famous pottery designer, noted as far back as 1782 that different materials become red hot at the same temperature. The quantitative relation between color and temperature is described by the *blackbody radiation law*. A blackbody is an idealized perfect absorber and emitter of all possible wavelengths λ of the radiation. Figure $9.1.5$ shows experimental wavelength distributions of thermal radiation at several temperatures. Consistent with our experience, the maximum in the distribution, which determines the predominant color, increases with temperature. This relation is given by Wien's displacement law, which can be expressed

$$T\lambda_{max} = 2.898 \times 10^6\, nmK \tag{9.1.2}$$

where the wavelength is expressed in nanometers (nm). At room temperature (300 K), the maximum occurs around $10\mu m$, in the infrared region. In Figure $9.1.5$, the approximate values of λ_{max} are 2900 nm at 1000 K, 1450 nm at 2000 K and 500n m at 5800 K, the approximate surface temperature of the Sun. The Sun's λ_{max} is near the middle of the visible range (380-750nm) and is perceived by our eyes as white light.

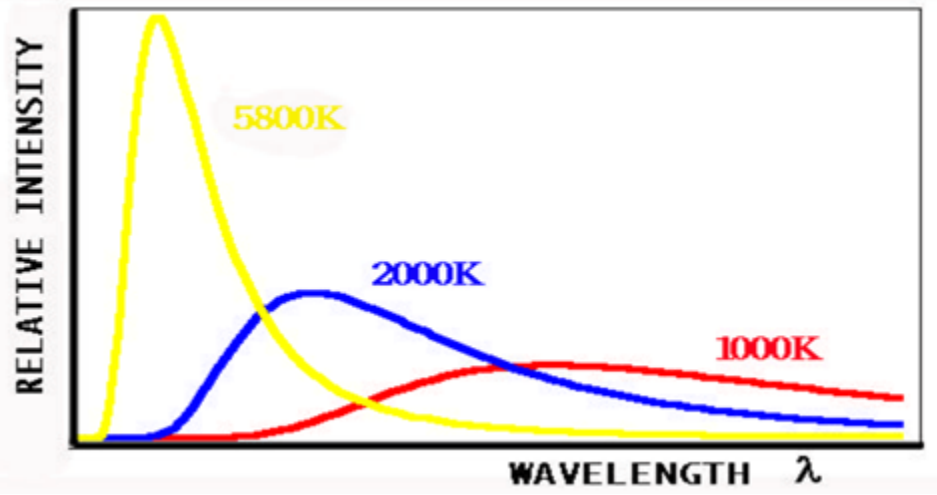


Figure $9.1.5$. Intensity distributions of blackbody radiation at three different temperatures. The total radiation intensity varies as T^4 (Stefan-Boltzmann law) so the total radiation at 2000K is actually 2^4 = 16 times that at 1000K.

The origin of blackbody radiation was a major challenge to 19th Century physics. Lord Rayleigh proposed that the electromagnetic field could be represented by a collection of oscillators of all possible frequencies. By simple geometry, the higher-frequency (lower wavelength) modes of oscillation are increasingly numerous since it it possible to fit their waves into an enclosure in a larger number of arrangements. In fact, the number of oscillators increases very rapidly as λ^{-4}. Rayleigh assumed that every oscillator contributed equally to the radiation (the equipartition principle). This agrees fairly well with experiment at low frequencies. But if ultraviolet rays and higher frequencies were really produced in increasing number, we would get roasted like marshmallows by sitting in front of a fireplace! Fortunately, this doesn't happen, and the incorrect theory is said to suffer from an "ultraviolet catastrophe."

Max Planck in 1900 derived the correct form of the blackbody radiation law by introducing a bold postulate. He proposed that energies involved in absorption and emission of electromagnetic radiation did not belong to a continuum, as implied by Maxwell's theory, but were actually made up of discrete bundles which he called "quanta." Planck's idea is traditionally regarded as marking the birth of the quantum theory. A quantum associated with radiation of frequency ν has the energy

$$E = h\nu \tag{9.1.3}$$

where the proportionality factor h = 6.626 x 10^{-34} J sec is known as Planck's constant. For our development of the quantum theory of atoms and molecules, we need only this simple result and do not have to follow the remainder of Planck's derivation. If you insist, however, the details are given in Supplement 1B.

The Photoelectric Effect

A familiar device in modern technology is the photocell or "electric eye," which runs a variety of useful gadgets, including automatic door openers. The principle involved in these devices is the photoelectric effect, which was first observed by Heinrich

Hertz in the same laboratory in which he discovered electromagnetic waves. Visible or ultraviolet radiation impinging on clean metal surfaces can cause electrons to be ejected from the metal. Such an effect is not, in itself, inconsistent with classical theory since electromagnetic waves are known to carry energy and momentum. But the detailed behavior as a function of radiation frequency and intensity can *not* be explained classically.

The energy required to eject an electron from a metal is determined by its *work function* Φ. For example, sodium has $\Phi = 1.82eV$. The electron-volt is a convenient unit of energy on the atomic scale: 1 eV = 1.602 x 10^{-19}J. This corresponds to the energy which an electron picks up when accelerated across a potential difference of 1 volt. The classical expectation would be that radiation of sufficient intensity should cause ejection of electrons from a metal surface, with their kinetic energies increasing with the radiation intensity. Moreover, a time delay would be expected between the absorption of radiation and the ejection of electrons. The experimental facts are quite different. It is found that no electrons are ejected, no matter how high the radiation intensity, unless the radiation *frequency* exceeds some threshold value ν_0 for each metal. For sodium ν_0 = 4.39 x 10^{14} Hz (corresponding to a wavelength of 683 nm), as shown in Figure $9.1.6$. For frequencies ν above the threshold, the ejected electrons acquire a kinetic energy given by

$$\frac{1}{2}mv^2 = h(\nu - \nu_0) = h\nu - \Phi \tag{9.1.4}$$

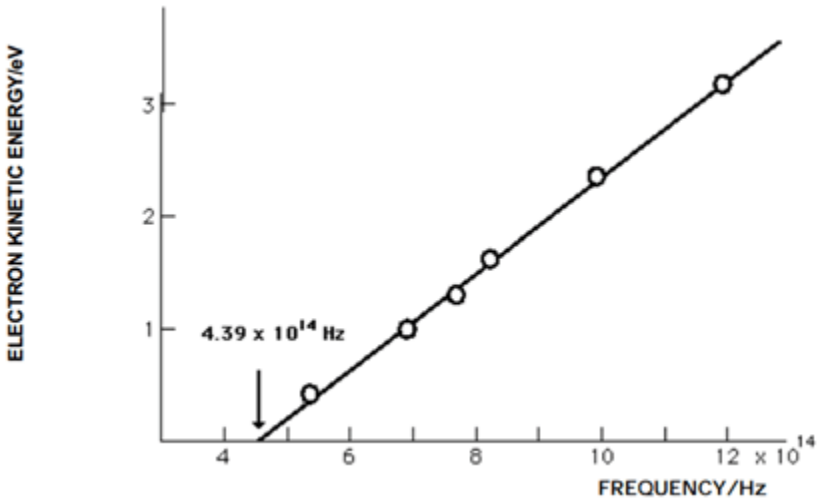


Figure $9.1.6$. Photoelectric data for sodium (Millikan, 1916). The threshold frequency ν_0, found by extrapolation, equals 4.39 x 10^{14}Hz.

Evidently, the work function Φ can be identified with $h\nu_0$, equal to 3.65 x 10^{-19}J=1.82 eV for sodium. The kinetic energy increases *linearly* with frequency above the threshold but is independent of the radiation intensity. Increased intensity does, however, increase the *number* of photoelectrons.

Einstein's explanation of the photoelectric effect in 1905 appears trivially simple once stated. He accepted Planck's hypothesis that a quantum of radiation carries an energy $h\nu$. Thus, if an electron is bound in a metal with an energy Φ, a quantum of energy $h\nu_0 = \Phi$ will be sufficient to dislodge it. And any excess energy $h(\nu - \nu_0)$ will appear as kinetic energy of the ejected electron. Einstein believed that the radiation field actually did consist of quantized particles, which he named *photons*. Although Planck himself never believed that quanta were real, Einstein's success with the photoelectric effect greatly advanced the concept of energy quantization.

Line Spectra

Most of what is known about atomic (and molecular) structure and mechanics has been deduced from spectroscopy. Figure $9.1.7$ shows two different types of spectra. A continuous spectrum can be produced by an incandescent solid or gas at high pressure. Blackbody radiation, for example, is a continuum. An emission spectrum can be produced by a gas at low pressure excited by heat or by collisions with electrons. An absorption spectrum results when light from a continuous source passes through a cooler gas, consisting of a series of dark lines characteristic of the composition of the gas. Frauenhofer between 1814 and 1823 discovered nearly 600 dark lines in the solar spectrum viewed at high resolution. It is now understood that these lines are caused by absorption by the outer layers of the Sun.

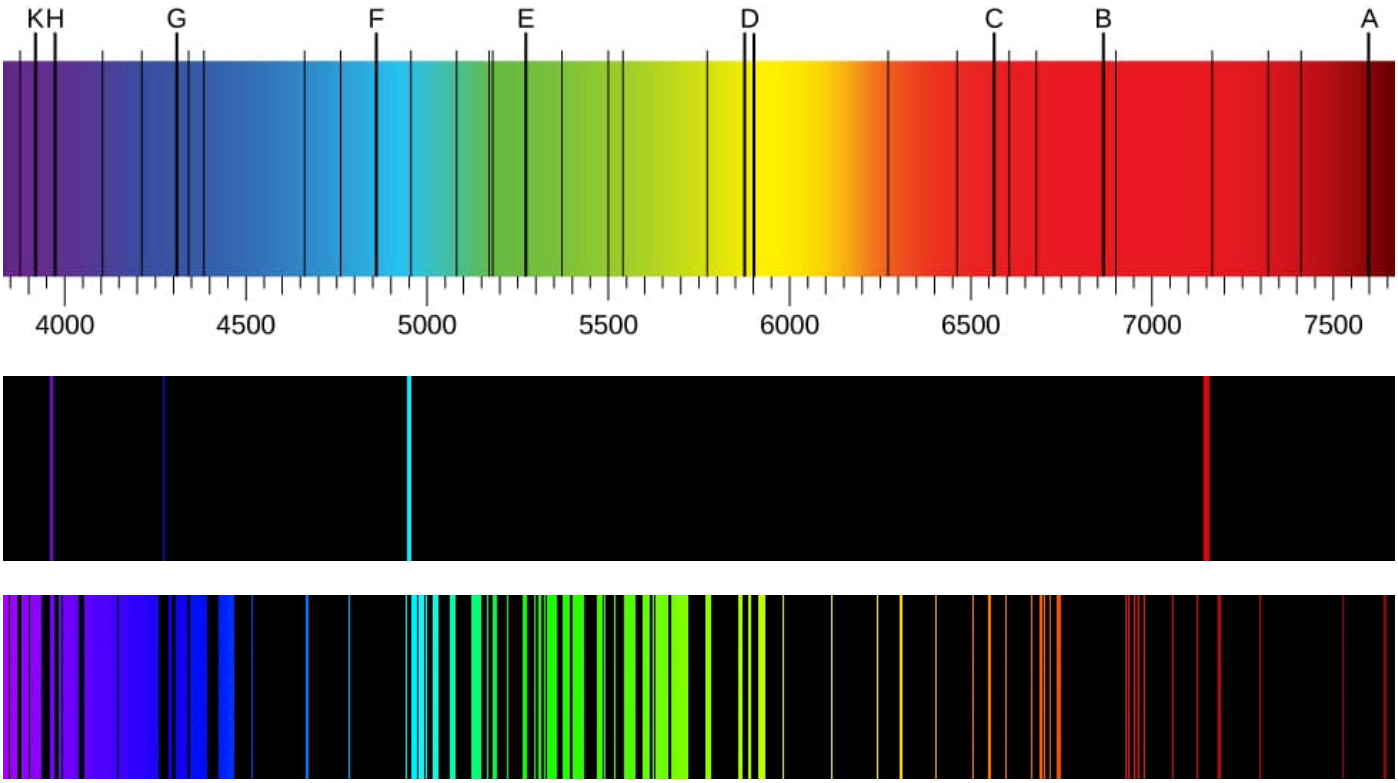


Figure 6.4.1. In the solar emission spectrum in the visible range from 380 nm to 710 nm, Fraunhofer lines are observed as vertical black lines at specific spectral positions in the continuous spectrum. Highly sensitive modern instruments observe thousands of such lines. (middle)The emission spectrum of atomic hydrogen: The spectral positions of emission lines are characteristic for hydrogen atoms. (credit: "Merikanto"/Wikimedia Commons). (bottom The emission spectrum of atomic iron: The spectral positions of emission lines are characteristic for iron atoms. Images used with permission (CC SA-BY; Openstax).

Gases heated to incandescence were found by Bunsen, Kirkhoff and others to emit light with a series of sharp wavelengths. The emitted light analyzed by a spectrometer (or even a simple prism) appears as a multitude of narrow bands of color. These so called *line spectra* are characteristic of the atomic composition of the gas. The line spectra of several elements are shown in Figure 9.1.8.

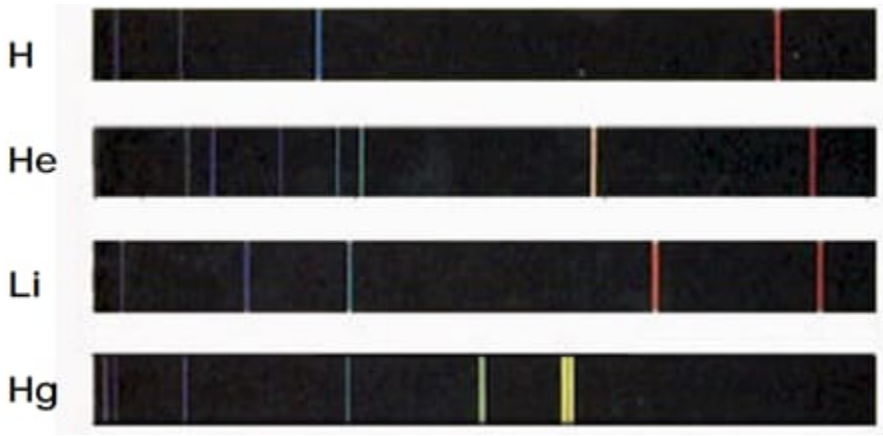


Figure 9.1.8. Emission spectra of several elements.

It is consistent with classical electromagnetic theory that motions of electrical charges within atoms can be associated with the absorption and emission of radiation. What is completely mysterious is how such radiation can occur for discrete frequencies, rather than as a continuum. The breakthrough that explained line spectra is credited to Neils Bohr in 1913. Building on the ideas of Planck and Einstein, Bohr postulated that the energy levels of atoms belong to a discrete set of values E_n, rather than a continuum as in classical mechanics. When an atom makes a downward energy transition from a higher energy level E_m to a lower energy level E_n, it caused the emission of a photon of energy

$$h\nu = E_m - E_n \tag{9.1.5}$$

This is what accounts for the discrete values of frequency ν in emission spectra of atoms. Absorption spectra are correspondingly associated with the annihilation of a photon of the same energy and concomitant excitation of the atom from E_n to E_m. Figure $9.1.9$ is a schematic representation of the processes of absorption and emission of photons by atoms. Absorption and emission processes occur at the same set frequencies, as is shown by the two line spectra in Figure $9.1.7$.

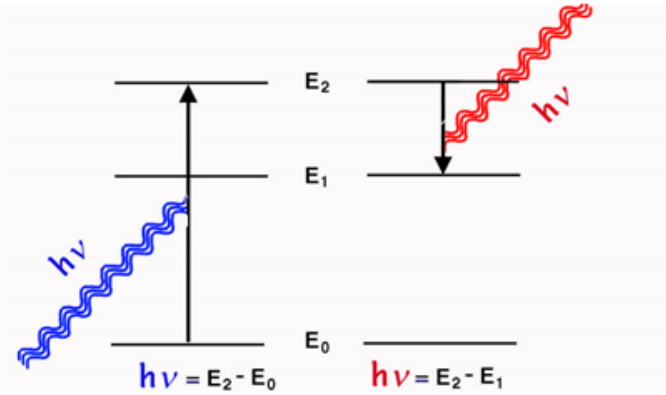


Figure $9.1.9$: Origin of line spectra. Absorption of the photon shown in blue causes atomic transition from E_0 to E_2. Transition from E_2 to E_1 causes emission of the photon shown in red.

Rydberg (1890) found that all the lines of the atomic hydrogen spectrum could be fitted to a simple empirical formula

$$\frac{1}{\lambda} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right), n = 1, 2, 3..., n_2 > n_1 \tag{9.1.6}$$

where R, known as the Rydberg constant, has the value 109,677 cm^{-1}. This formula was found to be valid for hydrogen spectral lines in the infrared and ultraviolet regions, in addition to the four lines in the visible region. No analogously simple formula has been found for any atom other than hydrogen. Bohr proposed a model for the energy levels of a hydrogen atom which agreed with Rydberg's formula for radiative transition frequencies. Inspired by Rutherford's nuclear atom, Bohr suggested a planetary model for the hydrogen atom in which the electron goes around the proton in one of a set of allowed circular orbits, as shown in Fig 8. A more fundamental understanding of the discrete nature of orbits and energy levels had to await the discoveries of 1925-26, but Bohr's model provided an invaluable stepping-stone to the development of quantum mechanics. We will consider the hydrogen atom in greater detail in Chap. 7.

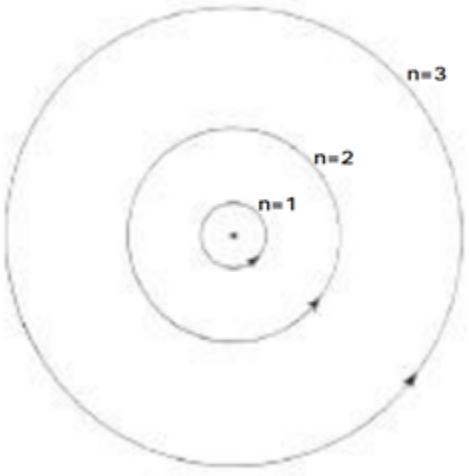


Figure $9.1.10$. Bohr model of the hydrogen atom showing three lowest-energy orbits.



Figure 9.1.11. A stylized representation of the Bohr model for a multielectron atom. From the logo of the International Atomic Energy Agency.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)
- Cameron Holder, Hope College. Holland, MI.

This page titled 9.1: Atoms and Photons- Origin of the Quantum Theory is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Chapter 1: Atoms and Photons: Origin of Quantum Theory is licensed CC BY-NC-SA 4.0.

9.2: Waves and Particles

Quantum mechanics is the theoretical framework which describes the behavior of matter on the atomic scale. It is the most successful quantitative theory in the history of science, having withstood thousands of experimental tests without a single verifiable exception. It has correctly predicted or explained phenomena in fields as diverse as chemistry, elementary-particle physics, solid-state electronics, molecular biology and cosmology. A host of modern technological marvels, including transistors, lasers, computers and nuclear reactors are offspring of the quantum theory. Possibly 30% of the US gross national product involves technology which is based on quantum mechanics. For all its relevance, the quantum world differs quite dramatically from the world of everyday experience. To understand the modern theory of matter, conceptual hurdles of both psychological and mathematical variety must be overcome. A paradox which stimulated the early development of the quantum theory concerned the indeterminate nature of light. Light usually behaves as a wave phenomenon but occasionally it betrays a particle-like aspect, a schizoid tendency known as the wave-particle duality. We consider first the wave theory of light.

The Double-Slit Experiment

Figure $9.2.1$ shows a modernized version of the famous double-slit diffraction experiment first performed by Thomas Young in 1801. Light from a monochromatic (single wavelength) source passes through two narrow slits and is projected onto a screen. Each slit by itself would allow just a narrow band of light to illuminate the screen. But with both slits open, a beautiful interference pattern of alternating light and dark bands appears, with maximum intensity in the center. To understand what is happening, we review some key results about electromagnetic waves.

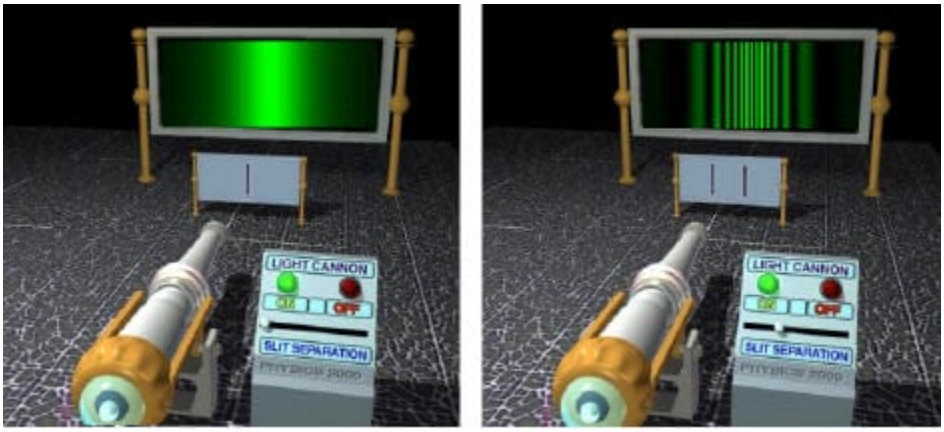


Figure $9.2.1$: Modern version of Young's interference experiment using a laser gun. Single slit (left) produces an intense band of light. Double slit (right) gives a diffraction pattern. See animated applet at www.colorado.edu/physics/2000...twoslitsa.html

Maxwell's theory of electromagnetism was an elegant unification of the diverse phenomena of electricity, magnetism and radiation, including light. Electromagnetic radiation is carried by transverse waves of electric and magnetic fields, propagating in vacuum at a speed $c \approx 3 \times 10^8 m/sec$, known as the "speed of light." As shown in Figure 2, the **E** and **B** fields oscillate sinusoidally, in synchrony with one another. The magnitudes of **E** and **B** are proportional ($B = E/c$ in SI units). The distance between successive maxima (or minima) at a given instant of time is called the wavelength λ. At every point in space, the fields also oscillate sinusoidally as functions of time. The number of oscillations per unit time is called the frequency ν. Since the field moves one wavelength in the time λ/c, the wavelength, frequency and speed for any wave phenomenon are related by

$$\lambda\nu = c \tag{9.2.1}$$

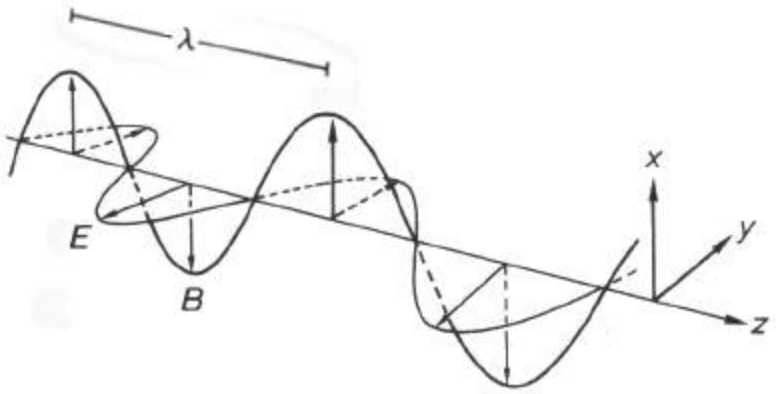


Figure 9.2.2: Schematic representation of electromagnetic wave. In electromagnetic theory, the intensity of radiation, energy flux incident on a unit area per unit time, is represented by the Poynting vector

$$\vec{S} = \mu_0 \vec{E} \times \vec{B} \tag{9.2.2}$$

The energy density contained in an electromagnetic field, even a static one, is given by

$$\rho = \frac{1}{2}\left(\epsilon_0 E^2 + \frac{B^2}{\mu_0}\right) \tag{9.2.3}$$

Note that both of the above energy quantities depend *quadratically* on the fields **E** and **B**. To discuss the diffraction experiments described above, it is useful to define the *amplitude* of an electromagnetic wave at each point in space and time **r**, *t* by the function

$$\Psi\left(\vec{r},t\right) = \sqrt{\epsilon_0} E\left(\vec{r},t\right) = \frac{B\left(\vec{r},t\right)}{\sqrt{\mu_0}} \tag{9.2.4}$$

such that the intensity is given by

$$\rho\left(\vec{r},t\right) = \left[\Psi\left(\vec{r},t\right)\right]^2 \tag{9.2.5}$$

The function $\Psi(\vec{r},t)$ will, in some later applications, have complex values. In such cases we generalize the definition of intensity to

$$\rho\left(\vec{r},t\right) = \left|\Psi\left(\vec{r},t\right)\right|^2 = \Psi\left(\vec{r},t\right)^* \Psi\left(\vec{r},t\right) \tag{9.2.6}$$

where $\Psi\left(\vec{r},t\right)^*$ represents the complex conjugate of $\Psi\left(\vec{r},t\right)$. In quantum mechanical applications, the function Ψ is known as the *wavefunction*.

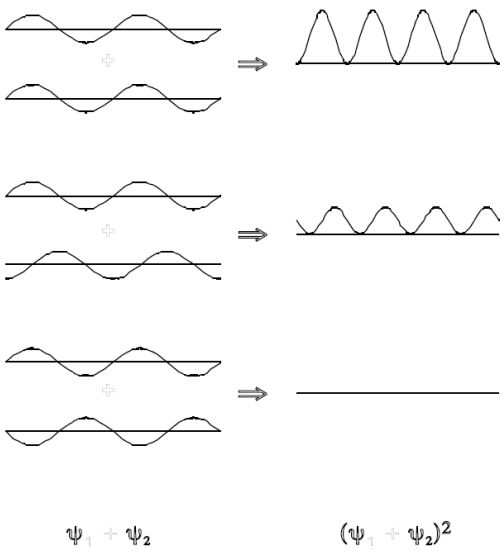


Figure $9.2.1$***:*** *Interference of two equal sinusoidal waves. Top: constructive interference. Bottom: destructive interference. Center: intermediate case. The resulting intensities* $\rho = \Psi^2$ *is shown on the right.*

The electric and magnetic fields, hence the amplitude Ψ, can have either positive and negative values at different points in space. In fact constructive and destructive interference arises from the superposition of waves, as illustrated in Figure $9.2.3$. By Equation $9.2.5$, the intensity $\rho \geq 0$ everywhere. The light and dark bands on the screen are explained by constructive and destructive interference, respectively. The wavelike nature of light is convincingly demonstrated by the fact that the intensity with both slits open is *not* the sum of the individual intensities, ie, $\rho \neq \rho_1 + \rho_2$. Rather it is the wave amplitudes which add:

$$\Psi = \Psi_1 + \Psi_2 \tag{9.2.7}$$

with the intensity given by the square of the amplitude:

$$\rho = \Psi^2 = \Psi_1^2 + \Psi_2^2 + 2\Psi_1\Psi_2 \tag{9.2.8}$$

The cross term $2\Psi_1\Psi_2$ is responsible for the constructive and destructive interference. Where Ψ_1 and Ψ_2 have the same sign, constructive interference makes the total intensity greater than the the sum of ρ_1 and ρ_2. Where Ψ_1 and Ψ_2 have opposite signs, there is destructive interference. If, in fact, $\Psi_1 = -\Psi_2$ then the two waves cancel exactly, giving a dark fringe on the screen.

Wave-Particle Duality

The interference phenomena demonstrated by the work of Young, Fresnel and others in the early 19th Century, apparently settled the matter that light was a wave phenomenon, contrary to the views of Newton a century earlier--case closed! But nearly a century later, phenomena were discovered which could *not* be satisfactorily accounted for by the wave theory, specifically blackbody radiation and the photoelectric effect.

Deviating from the historical development, we will illustrate these effects by a modification of the double slit experiment. Let us equip the laser source with a dimmer switch capable of reducing the light intensity by several orders of magnitude, as shown in Figure $9.2.4$. With each successive filter the diffraction pattern becomes dimmer and dimmer. Eventually we will begin to see localized scintillations at random positions on an otherwise dark screen. It is an almost inescapable conclusion that these scintillations are caused by *photons*, the bundles of light postulated by Planck and Einstein to explain blackbody radiation and the photoelectric effect.

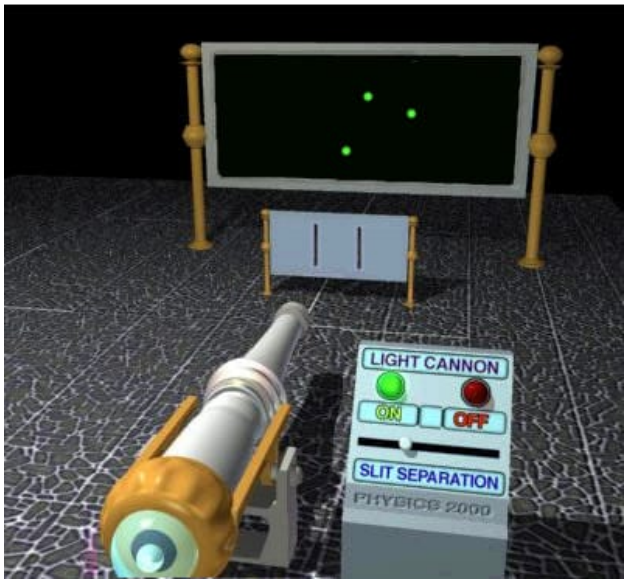


Figure 9.2.1: Scintillations observed after dimming laser intensity by several orders of magnitude. These are evidently caused by individual photons!

But wonders do not cease even here. Even though the individual scintillations appear at random positions on the screen, their statistical behavior reproduces the original high-intensity diffraction pattern. Evidently the statistical behavior of the photons follows a predictable pattern, even though the behavior of individual photons is unpredictable. This implies that each individual photon, even though it behaves mostly like a particle, somehow carry with it a "knowledge" of the entire wavelike diffraction pattern. In some sense, a single photon must be able to go through *both* slits at the same time. This is what is known as the *wave-particle duality* for light: under appropriate circumstances light can behave as a wave or as a particle.

Planck's resolution of the problem of blackbody radiation and Einstein's explanation of the photoelectric effect can be summarized by a relation between the energy of a photon to its frequency:

$$E = h\nu \tag{9.2.9}$$

where $h = 6.626 \times 10^{-34} Jsec$, known as Planck's constant. Much later, the Compton effect was discovered, wherein an x-ray or gamma ray photon ejects an electron from an atom, as shown in Figure 9.2.5. Assuming conservation of momentum in a photon-electron collision, the photon is found to carry a momentum *p*, given by

$$p = \frac{h}{\lambda} \tag{9.2.10}$$

Equation 9.2.9 and 9.2.10 constitute quantitative realizations of the wave-particle duality, each relating a particle-like property--energy or momentum--to a wavelike property--frequency or wavelength.

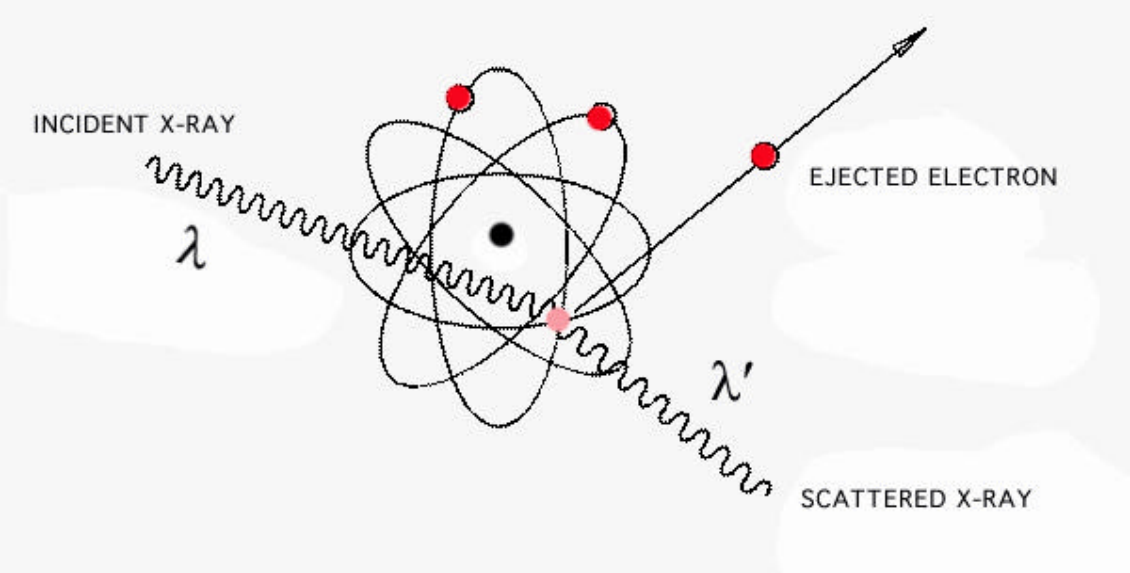


Figure 9.2.1: Compton effect. The momentum and energy carried by the incident x-ray photon are transferred to the ejected electron and the scattered photon.

According to the special theory of relativity, the last two formulas are actually different facets of the same fundamental relationship. By Einstein's famous formula, the equivalence of mass and energy is given by

$$E = mc^2 \tag{9.2.11}$$

The photon's rest mass is zero, but in travelling at speed c, it acquires a finite mass. Equating Equation 9.2.8) and 9.2.11 for the photon energy and taking the photon momentum to be $p = mc$, we obtain

$$p = \frac{E}{c} = \frac{h\nu}{c} = \frac{h}{\lambda} \tag{9.2.12}$$

Thus, the wavelength-frequency relation (Equation 9.2.1), implies the Compton-effect formula (Equation 9.2.10).

The best we can do is to *describe* the phenomena constituting the wave-particle duality. There is no widely accepted explanation in terms of everyday experience and common sense. Feynman referred to the "experiment with two holes" as the "central mystery of quantum mechanics." It should be mentioned that a number of models have been proposed over the years to rationalize these quantum mysteries. Bohm proposed that there might exist *hidden variables* which would make the behavior of each photon deterministic, ie, particle-like. Everett and Wheeler proposed the "many worlds interpretation of quantum mechanics" in which each random event causes the splitting of the entire universe into disconnected parallel universes in which each possibility becomes the reality. Needless to say, not many people are willing to accept such a metaphysically unwieldy view of reality. Most scientists are content to apply the highly successful computational mechanisms of quantum theory to their work, without worrying unduly about its philosophical underpinnings. Sort of like people who enjoy eating roast beef but would rather not think about where it comes from.

There was never any drawn-out controversy about whether electrons or any other constituents of matter were other than particle-like. Yet a variant of the double-slit experiment using electrons instead of light proves otherwise. The experiment is technically difficult but has been done. An electron gun, instead of a light source, produces a beam of electrons at a selected velocity, which is focused and guided by electric and magnetic fields. Then, everything that happens for photons has its analog for electrons. Individual electrons produce scintillations on a phosphor screen-this is how TV works. But electrons also exhibit diffraction effects, which indicates that they too have wavelike attributes. Diffraction experiments have been more recently carried out for particles as large as atoms and molecules, even for the C_{60} fullerene molecule.

De Broglie in 1924 first conjectured that matter might also exhibit a wave-particle duality. A wavelike aspect of the electron might, for example, be responsible for the discrete nature of Bohr orbits in the hydrogen atom. According to de Broglie's hypothesis, the "matter waves" associated with a particle have a wavelength given by

$$\lambda = h/p \tag{9.2.13}$$

which is identical in form to Compton's result (Equation 9.2.10) (which, in fact, was discovered later). The correctness of de Broglie's conjecture was most dramatically confirmed by the observations of Davisson and Germer in 1927 of diffraction of monoenergetic beams of electrons by metal crystals, much like the diffraction of x-rays. And measurements showed that de Broglie's formula (Equation 9.2.13) did indeed give the correct wavelength (see Figure 9.2.6).

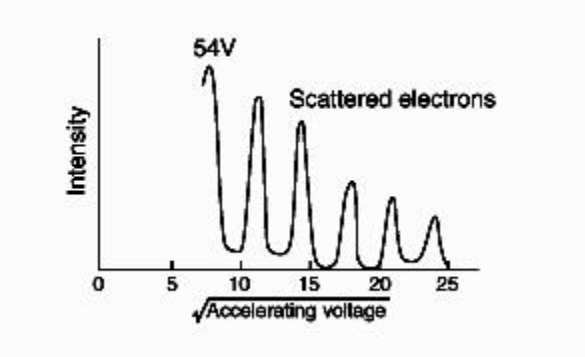


Figure 9.2.16: Intensity of electron scattered at a fixed angle off a nickel crystal, as function of incident electron energy. From C. J. Davisson "Are Electrons Waves?" *Franklin Institute Journal* **205**, 597 (1928).

The Schrödinger Equation

Schrödinger in 1926 first proposed an equation for de Broglie's matter waves. This equation cannot be derived from some other principle since it constitutes a fundamental law of nature. Its correctness can be judged only by its subsequent agreement with observed phenomena (*a posteriori* proof). Nonetheless, we will attempt a heuristic argument to make the result at least plausible. In classical electromagnetic theory, it follows from Maxwell's equations that each component of the electric and magnetic fields in vacuum is a solution of the wave equation

$$\nabla^2\Psi - \frac{1}{c^2}\frac{\partial^2\Psi}{\partial t^2} = 0 \tag{9.2.14}$$

where the Laplacian or "del-squared" operator is defined by

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{9.2.15}$$

We will attempt now to create an analogous equation for de Broglie's matter waves. Accordingly, let us consider a very general instance of wave motion propagating in the *x*-direction. At a given instant of time, the form of a wave might be represented by a function such as

$$\psi(x) = f\left(\frac{2\pi x}{\lambda}\right) \tag{9.2.16}$$

where $f(\theta)$ represents a sinusoidal function such as $\sin\theta$, $\cos\theta$, $e^{i\theta}$, $e^{-i\theta}$ or some linear combination of these. The most suggestive form will turn out to be the complex exponential, which is related to the sine and cosine by Euler's formula

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{9.2.17}$$

Each of the above is a periodic function, its value repeating every time its argument increases by 2π. This happens whenever *x* increases by one wavelength λ. At a fixed point in space, the time-dependence of the wave has an analogous structure:

$$T(t) = f(2\pi\nu t) \tag{9.2.18}$$

where ν gives the number of cycles of the wave per unit time. Taking into account both *x*- and *t*-dependence, we consider a wavefunction of the form

$$\Psi(x,t) = exp\left[2\pi i\left(\frac{x}{\lambda} - \nu t\right)\right] \tag{9.2.19}$$

representing waves travelling from left to right. Now we make use of the Planck and de Broglie formulas (Equation 9.2.8 and 9.2.13

, respectively) to replace ν and λ by their particle analogs. This gives

$$\Psi(x,t) = exp[i(px - Et)/\hbar] \tag{9.2.20}$$

where

$$\hbar \equiv \frac{h}{2\pi} \tag{9.2.21}$$

Since Planck's constant occurs in most formulas with the denominator 2π, this symbol was introduced by Dirac. Now Equation 9.2.18 represents in some way the wavelike nature of a particle with energy E and momentum *p*. The time derivative of Equation 9.2.20 gives

$$\frac{\partial \Psi}{\partial t} = -(iE/\hbar) \times \exp\left[\frac{i(px - Et)}{\hbar}\right] \tag{9.2.22}$$

Thus

$$i\hbar\frac{\partial \Psi}{\partial t} = E\Psi \tag{9.2.23}$$

Analogously

$$-i\hbar\frac{\partial \Psi}{\partial x} = p\Psi \tag{9.2.24}$$

and

$$-\hbar^2\frac{\partial^2 \Psi}{\partial x^2} = p^2\Psi \tag{9.2.25}$$

The energy and momentum for a nonrelativistic free particle are related by

$$E = \frac{1}{2}mv^2 = \frac{p^2}{2m} \tag{9.2.26}$$

Thus $\Psi(x,t)$ satisfies the partial differential equation

$$i\hbar\frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi}{\partial x^2} \tag{9.2.27}$$

For a particle with a potential energy $V(x)$,

$$E = \frac{p^2}{2m} + V(x) \tag{9.2.28}$$

we postulate that the equation for matter waves generalizes to

$$i\hbar\frac{\partial \Psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]\Psi \tag{9.2.29}$$

For waves in three dimensions should then have

$$i\hbar\frac{\partial}{\partial t}\Psi(\vec{r},t) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r})\right]\Psi(\vec{r},t) \tag{9.2.30}$$

Here the potential energy and the wavefunction depend on the three space coordinates *x, y, z*, which we write for brevity as **r**. This is the time-dependent Schrödinger equation for the amplitude $\Psi(\vec{r},t)$ of the matter waves associated with the particle. Its formulation in 1926 represents the starting point of modern quantum mechanics. (Heisenberg in 1925 proposed another version known as matrix mechanics.)

For *conservative* systems, in which the energy is a constant, we can separate out the time-dependent factor from (19) and write

$$\Psi(\vec{r},t) = \psi(\vec{r})e^{-iEt\div\hbar} \tag{9.2.31}$$

where $\psi(\vec{r})$ is a wavefunction dependent only on space coordinates. Putting Equation 9.2.31 into Equation 9.2.30 and cancelling the exponential factors, we obtain the time-independent Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r})\right]\psi(\vec{r}) = E\psi(\vec{r}) \tag{9.2.32}$$

Most of our applications of quantum mechanics to chemistry will be based on this equation.

The bracketed object in Equation 9.2.32 is called an *operator*. An operator is a generalization of the concept of a function. Whereas a function is a rule for turning one number into another, an operator is a rule for turning one function into another. The Laplacian (∇) is an example of an operator. We usually indicate that an object is an operator by placing a `hat' over it, eg, $\hat{A}$. The action of an operator that turns the function *f* into the function *g* is represented by

$$\hat{A}f = g \tag{9.2.33}$$

Equation 9.2.24 implies that the operator for the *x*-component of momentum can be written

$$\hat{p}_x = -i\hbar\frac{\partial}{\partial x} \tag{9.2.34}$$

and by analogy, we must have

$$\hat{p}_y = -i\hbar\frac{\partial}{\partial y} \tag{9.2.35}$$

and

$$\hat{p}_z = -i\hbar\frac{\partial}{\partial z} \tag{9.2.36}$$

The energy, as in Equation 9.2.28, expressed as a function of position and momentum is known in classical mechanics as the Hamiltonian. Generalizing to three dimensions,

$$\hat{H} = \frac{p^2}{2m} + V(\vec{r}) = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) + V(x,y,z) \tag{9.2.37}$$

We construct thus the corresponding quantum-mechanical operator

$$\hat{H} = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + V(x,y,z) = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}) \tag{9.2.38}$$

The time-independent Schrödinger equation (Equation 9.2.32) can then be written symbolically as

$$\hat{H}\Psi = E\Psi \tag{9.2.39}$$

This form is actually more generally to *any* quantum-mechanical problem, given the appropriate Hamiltonian and wavefunction. Most applications to chemistry involve systems containing many particles--electrons and nuclei. An operator equation of the form

$$\hat{A}\psi = const\psi \tag{9.2.40}$$

is called an eigenvalue equation. Recall that, in general, an operator acting on a function gives another function (e.g., Equation 9.2.33). The special case (Equation 9.2.40) occurs when the second function is a multiple of the first. In this case, ψ is known as an eigenfunction and the constant is called an eigenvalue. (These terms are hybrids with German, the purely English equivalents being `characteristic function' and `characteristic value.') To every dynamical variable A in quantum mechanics, there corresponds an eigenvalue equation, usually written

$$\hat{A}\psi = a\psi \tag{9.2.41}$$

The eigenvalues a represent the possible measured values of the variable A. The Schrödinger Equation (9.2.39) is the best known instance of an eigenvalue equation, with its eigenvalues corresponding to the allowed energy levels of the quantum system.

The Wavefunction

For a single-particle system, the wavefunction $\Psi(\vec{r},t)$, or $\psi(\vec{r})$ for the time-independent case, represents the amplitude of the still vaguely defined matter waves. The relationship between amplitude and intensity of electromagnetic waves we developed for Equation 9.2.6 can be extended to matter waves.

The most commonly accepted interpretation of the wavefunction is due to Max Born (1926), according to which $\rho(r)$, the square of the absolute value of $\psi(r)$ is proportional to the probability density (probability per unit volume) that the particle will be found at the position **r**. Probability density is the three-dimensional analog of the diffraction pattern that appears on the two-dimensional screen in the double-slit diffraction experiment for electrons described in the preceding Section. In the latter case we had the relative probability a scintillation would appear at a given point on the screen. The function $\rho(r)$ becomes equal, rather than just proportional to, the probability density when the wavefunction is normalized, that is,

$$\int |\psi(\vec{r})|^2 d\tau = 1 \tag{9.2.42}$$

This simply accounts for the fact that the total probability of finding the particle *somewhere* adds up to unity. The integration in Equation 9.2.42 extends over all space and the symbol $d\tau$ designates the appropriate volume element. For example, the volume differential in Cartesian coordinates, $d\tau = dx\ dy\ dz$ is changed in spherical coordinates to $d\tau = r^2 sin\theta\ dr\ d\theta\ d\phi$. The physical significance of the wavefunctions makes certain demands on its mathematical behavior. The wavefunction must be a single-valued function of all its coordinates, since the probability density ought to be uniquely determined at each point in space.

Moreover, the wavefunction should be finite and continuous everywhere, since a physically-meaningful probability density must have the same attributes. The conditions that the wavefunction be single-valued, finite and continuous--in short, "well behaved"--lead to restrictions on solutions of the Schrödinger equation such that only certain values of the energy and other dynamical variables are allowed. This is called *quantization* and is in the feature that gives *quantum* mechanics its name.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)
- Kevin Olson, Hope College. Holland, MI.

This page titled 9.2: Waves and Particles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Chapter 2: Waves and Particles is licensed CC BY-NC-SA 4.0.

9.3: The Free Particles

By definition, the particle does not feel any external force, therefore $V(x) = 0$, and the TISEq is written simply:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = E\psi(x). \tag{9.3.1}$$

This equation can be rearranged to:

$$\frac{d^2\psi}{dx^2} = -\frac{2mE}{\hbar^2}\psi(x), \tag{9.3.2}$$

which corresponds to a mathematical problem where the second derivative of a function should be equal to a constant, $-\frac{2mE}{\hbar^2}$ multiplied by the function itself. Such a problem is easily solved by the function:

$$\psi(x) = A\exp(\pm ikx). \tag{9.3.3}$$

The first and second derivatives of this function are:

$$\begin{aligned}\frac{d\psi(x)}{dx} &= \pm ikA\exp(\pm ikx) = \pm ik\psi(x)\\ \frac{d^2\psi(x)}{dx^2} &= \mp k^2 A\exp(\pm ikx) = -(\pm k^2)\psi(x).\end{aligned} \tag{9.3.4}$$

Comparing the second derivative in Equation ??? with Equation 9.3.2, we immediately see that if we set:

$$k^2 = \frac{2mE}{\hbar^2}, \tag{9.3.5}$$

we solve the original differential equation. Considering de Broglie's equation, Equation 17.5.4, we can replace $k = \frac{p}{\hbar}$, to obtain:

$$E = \frac{k^2\hbar^2}{2m} = \frac{p^2}{2m}, \tag{9.3.6}$$

which is exactly the classical value of the kinetic energy of a free particle moving in one direction of space. Since the function in Equation 9.3.3 solves the Schrödinger equation for the free particle, it is called an *eigenfunction* (or *eigenstate*) of the TISEq. The energy result of Equation 9.3.6 is called *eigenvalue* of the TISEq. Notice that, since k is continuous in the eigenfunction, the energy eigenvalue is also continuous (i.e., all values of E are acceptable).

This page titled 9.3: The Free Particles is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Roberto Peverati.

- 19.1: The Free Particles by Roberto Peverati is licensed CC BY-SA 4.0.

9.4: The Particle in a Box

Now we can consider a particle constrained to move in a single dimension, under the influence of a potential $V(x)$ which is zero for $0 \leq x \leq a$ and infinite elsewhere. Since the wavefunction is not allowed to become infinite, it must have a value of zero where $V(x)$ is infinite, so $\psi(x)$ is nonzero only within $[0, a]$. The Schrödinger equation is thus:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = E\psi(x) \qquad 0 \leq x \leq a. \tag{9.4.1}$$

In other words, inside the box $\psi(x)$ describes a free particle, but outside the box $\psi(x) = 0$. Since the Schrödinger equation involves derivatives, the function that solves it, $\psi(x)$, must be everywhere continuous and everywhere continuously differentiable. This fact means that the value of the wave function at the two extremes must be equal to zero:

$$\psi(0) = \psi(a) = 0. \tag{9.4.2}$$

Inside the box we can use Euler's formula to write the wave function as a linear combination of the positive and negative solutions:

$$\psi(x) = A\exp(\pm ix) = A\sin kx + B\cos kx, \tag{9.4.3}$$

where A and B are constants that we need to determine using the two constraints in Equation 9.4.2. For B it is straightforward to see that:

$$\psi(0) = 0 + B = 0 \implies B = 0. \tag{9.4.4}$$

For A we have:

$$\psi(a) = A\sin ka = 0, \tag{9.4.5}$$

which is trivially solved by $A = 0$, or by the more interesting condition of $ka = n\pi$. The trivial solution corresponds to a wave function uniformly equal to zero everywhere. This wave function is uninteresting, since it describes no particles in no boxes. The second set of solutions, however, is very interesting, since we can write it as:

$$\psi_n(x) = A\sin\left(\frac{n\pi x}{a}\right) \quad n = 1, 2, \ldots, \infty, \tag{9.4.6}$$

which represents an infinite set of functions, $\psi_n(x)$, determined by a positive integer number n, called *quantum number*. Since these functions solve the TISEq, they are also called *eigenfunctions*, but they are not a continuous set, unlike in the previous case. To calculate the energy eigenvalues, we can replace 9.4.6 into Equation 9.4.1, to obtain:

$$E_n = \frac{h^2 n^2}{8ma^2} \quad n = 1, 2, \ldots, \infty. \tag{9.4.7}$$

A few interesting considerations can be made from the results of Equation 9.4.7. First, although there is an infinite number of acceptable values of the energy (eigenvalues), these values are not continuous. Second, the lowest value of the energy is not zero, and it depends on the size of the box, a, since:

$$E_1 = \frac{h^2}{8ma^2} \neq 0. \tag{9.4.8}$$

This value is called *zero-point energy (ZPE)*, and is a purely quantum mechanical effect. Notice that we did not solve for the constant A. This task is not straightforward, and it can be achieved by requiring the wave function to describe one particle exclusively (we will come back to this task after chapter 23). Extending the problem to three dimensions is relatively straightforward, resulting in a set of three separate quantum numbers (one for each of the 3-dimensional coordinate n_x, n_y, n_z).

This page titled 9.4: The Particle in a Box is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Roberto Peverati.

- 19.2: The Particle in a Box by Roberto Peverati is licensed CC BY-SA 4.0.

9.5: Quantum Mechanics of Some Simple Systems

The simple quantum-mechanical problem we have just solved can provide an instructive application to chemistry: the *free-electron model* (FEM) for delocalized π-electrons. The simplest case is the 1,3-butadiene molecule

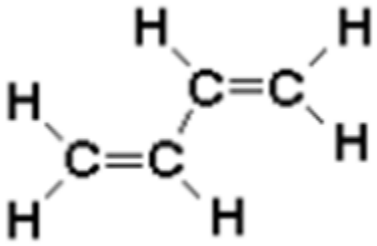


Figure 4) by

$$\rho = 2\psi_1^2 + 2\psi_2^2 \tag{9.5.1}$$

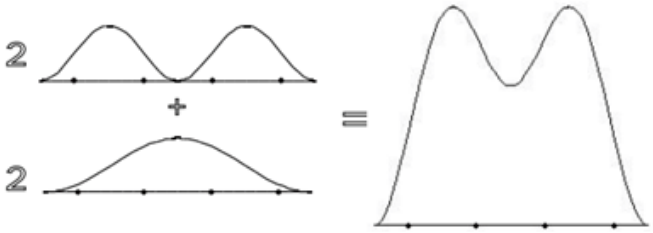
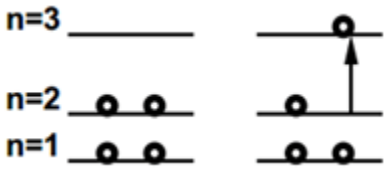


Figure 4. Pi-electron density in butadiene.

A chemical interpretation of this picture might be that, since the π-electron density is concentrated between carbon atoms 1 and 2, and between 3 and 4, the predominant structure of butadiene has double bonds between these two pairs of atoms. Each double bond consists of a π-bond, in addition to the underlying σ-bond. However, this is not the complete story, because we must also take account of the residual π-electron density between carbons 2 and 3. In the terminology of valence-bond theory, butadiene would be described as a *resonance hybrid* with the contributing structures CH_2=CH-CH=CH_2 (the predominant structure) and °CH_2-CH=CH-CH_2° (a secondary contribution). The reality of the latter structure is suggested by the ability of butadiene to undergo 1,4-addition reactions.

The free-electron model can also be applied to the electronic spectrum of butadiene and other linear polyenes. The lowest unoccupied molecular orbital (LUMO) in butadiene corresponds to the n=3 particle-in-a-box state. Neglecting electron-electron interaction, the longest-wavelength (lowest-energy) electronic transition should occur from n=2, the highest occupied molecular orbital (HOMO).



The energy difference is given by

$$\Delta E = E_3 - E_2 = (3^2 - 2^2)\frac{h^2}{8mL^2} \tag{9.5.2}$$

Here *m* represents the mass of an electron (not a butadiene molecule!), 9.1x10^{-31} Kg, and L is the effective length of the box,

4x1.40x10^{-10} m. By the Bohr frequency condition

$$\Delta E = h\nu = \frac{hc}{\lambda} \tag{9.5.3}$$

The wavelength is predicted to be 207 nm. This compares well with the experimental maximum of the first electronic absorption band, $\lambda_{max} \approx 210$ nm, in the ultraviolet region.

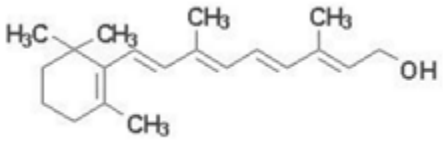
We might therefore be emboldened to apply the model to predict absorption spectra in higher polyenes CH_2=(CH-CH=)$_{n-1}CH_2$. For the molecule with 2*n* carbon atoms (*n* double bonds), the HOMO → LUMO transition corresponds to n → n + 1, thus

$$\frac{hc}{\lambda} \approx [(n+1)^2 - n^2] \frac{h^2}{8m(2nL_{CC})^2} \tag{9.5.4}$$

A useful constant in this computation is the **Compton wavelength**

$$\frac{h}{mc} = 2.426 \times 10^{-12} m. \tag{9.5.5}$$

For *n*=3, hexatriene, the predicted wavelength is 332 nm, while experiment gives $\lambda_{max} \approx 250$ nm. For *n*=4, octatetraene, FEM predicts 460 nm, while $\lambda_{max} \approx 300$ nm. Clearly the model has been pushed beyond range of quantitate validity, although the trend of increasing absorption band wavelength with increasing *n* is correctly predicted. Incidentally, a compound should be colored if its absorption includes any part of the visible range 400-700 nm. Retinol (vitamin A), which contains a polyene chain with *n*=5, has a pale yellow color. This is its structure:



Contributors and Attributions

Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)

This page titled 9.5: Quantum Mechanics of Some Simple Systems is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Free-Electron Model is licensed CC BY-NC-SA 4.0.

9.6: Principles of Quantum Mechanics

Here we will continue to develop the mathematical formalism of quantum mechanics, using heuristic arguments as necessary. This will lead to a system of postulates which will be the basis of our subsequent applications of quantum mechanics.

Hermitian Operators

An important property of operators is suggested by considering the Hamiltonian for the particle in a box:

$$\hat{H} = -\frac{h^2}{2m}\frac{d^2}{dx^2} \tag{9.6.1}$$

Let $f(x)$ and $g(x)$ be arbitrary functions which obey the same boundary values as the eigenfunctions of $\hat{H}$, namely that they vanish at $x = 0$ and $x = a$. Consider the integral

$$\int_0^a f(x)\,\hat{H}\,g(x)\,\mathrm{d}x = -\frac{\hbar^2}{2m}\int_0^a f(x)\,g''(x)\,\mathrm{d}x \tag{9.6.2}$$

Now, using integration by parts,

$$\int_0^a f(x)\,g''(x)\,\mathrm{d}x = -\int_0^a f'(x)\,g'(x)\,\mathrm{d}x + \Big[f(x)\,g'(x)\Big]_0^a \tag{9.6.3}$$

The boundary terms vanish by the assumed conditions on f and g. A second integration by parts transforms Equation 9.6.3 to

$$\int_0^a f''(x)\,g(x)\,\mathrm{d}x\; -\; \Big[f'(x)\,g(x)\Big]_0^a \tag{9.6.4}$$

It follows therefore that

$$\int_0^a f(x)\,\hat{H}\,g(x)\,\mathrm{d}x = \int_0^a g(x)\,\hat{H}\,f(x)\,\mathrm{d}x \tag{9.6.5}$$

An obvious generalization for complex functions will read

$$\int_0^a f^*(x)\,\hat{H}\,g(x)\,\mathrm{d}x = \left(\int_0^a g^*(x)\,\hat{H}\,f(x)\,\mathrm{d}x\right)^* \tag{9.6.6}$$

In mathematical terminology, an operator $\hat{A}$ for which

$$\int f^*\,\hat{A}\,g\,\mathrm{d}\tau = \left(\int g^*\,\hat{A}\,f\,\mathrm{d}\tau\right)^* \tag{9.6.7}$$

for all functions f and g which obey specified boundary conditions is classified as *hermitian* or *self-adjoint*. Evidently, the Hamiltonian is a hermitian operator. It is postulated that *all* quantum-mechanical operators that represent dynamical variables are hermitian.

Properties of Eigenvalues and Eigenfunctions

The sets of energies and wavefunctions obtained by solving any quantum-mechanical problem can be summarized symbolically as solutions of the eigenvalue equation

$$\hat{H}\,\psi_n = E_n\,\psi_n \tag{9.6.8}$$

For another value of the quantum number, we can write

$$\hat{H}\,\psi_m = E_m\,\psi_m \tag{9.6.9}$$

Let us multiply Equation 9.6.8 by ψ_m^* and the complex conjugate of Equation 9.6.9 by ψ_n. Then we subtract the two expressions and integrate over $\mathrm{d}\tau$. The result is

$$\int \psi_m^*\,\hat{H}\,\psi_n\,\mathrm{d}\tau - \left(\int \psi_n^*\,\hat{H}\,\psi_m\,\mathrm{d}\tau\right)^* = (E_n - E_m^*)\int \psi_m^*\,\psi_n\,\mathrm{d}\tau \tag{9.6.10}$$

But by the hermitian property (Equation 9.6.6), the left-hand side of Equation 9.6.10 equals zero. Thus

$$(E_n - E_m^*)\int \psi_m^*\,\psi_n\,\mathrm{d}\tau = 0 \tag{9.6.11}$$

Consider first the case $m = n$. The second factor in Equation 9.6.11 then becomes the normalization integral $\int \psi_n^*\ \psi_n\ \mathrm{d}\tau$, which equals 1 (or at least a nonzero constant). Therefore the first factor in Equation 9.6.11 must equal zero, so that

$$E_n^* = E_n \tag{9.6.12}$$

implying that the energy eigenvalues must be real numbers. This is quite reasonable from a physical point of view since eigenvalues represent possible results of measurement. Consider next the case when $E_m \neq E_n$. Then it is the second factor in Equation 9.6.11 that must vanish and

$$\int \psi_m^*\,\psi_n\,\mathrm{d}\tau = 0 \quad when\ E_m \neq E_n \tag{9.6.13}$$

Thus eigenfunctions belonging to different eigenvalues are orthogonal. In the case that ψ_m and ψ_n are degenerate eigenfunctions, so $m \neq n$ but $E_m = E_n$, the above proof of orthogonality does not apply. But it is always possible to construct degenerate functions that are mutually orthogonal. A general result is therefore the orthonormalization condition

$$\int \psi_m^*\,\psi_n\,\mathrm{d}\tau = \delta_{mn} \tag{9.6.14}$$

It is easy to prove that a linear combination of degenerate eigenfunctions is itself an eigenfunction of the same energy. Let

$$\hat{H}\,\psi_{nk} = E_n\,\psi_{nk}, \qquad k = 1, 2, \ldots d \tag{9.6.15}$$

where the ψ_{nk} represent a *d*-fold degenerate set of eigenfunctions with the same eigenvalue E_n. Consider now the linear combination

$$\psi = c_1\psi_{n,1} + c_2\psi_{n,2} + \ldots + c_d\psi_{n,d} \tag{9.6.16}$$

Operating on ψ with the Hamiltonian and using (14), we find

$$\hat{H}\,\psi = c_1\hat{H}\,\psi_{n,1} + c_2\hat{H}\,\psi_{n,2} + \ldots = E_n(c_1\psi_{n,1} + c_2\psi_{n,2} + \ldots) = E_n\,\psi \tag{9.6.17}$$

which shows that the linear combination ψ is also an eigenfunction of the same energy. There is evidently a limitless number of possible eigenfunctions for a degenerate eigenvalue. However, only *d* of these will be linearly independent.

Dirac Notation

The term *orthogonal* has been used both for perpendicular vectors and for functions whose product integrates to zero. This actually connotes a deep connection between vectors and functions. Consider two orthogonal vectors **a** and **b**. Then, in terms of their *x, y, z* components, labeled by 1, 2, 3, respectively, the scalar product can be written

$$\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3 = 0 \tag{9.6.18}$$

Suppose now that we consider an analogous relationship involving vectors in *n*-dimensional space (which you need not visualize!). We could then write

$$a \cdot b = \sum_{k=0}^{n} a_k b_k = 0 \tag{9.6.19}$$

Finally let the dimension of the space become non-denumerably infinite, turning into a continuum. The sum in Equation 9.6.19 would then be replaced by an integral such as

$$\int a(x)\, b(x) dx = 0 \tag{9.6.20}$$

But this is just the relation for orthogonal functions. A function can therefore be regarded as an abstract vector in a higher-dimensional continuum, known as *Hilbert space*. This is true for eigenfunctions as well. Dirac denoted the vector in Hilbert space corresponding to the eigenfunction ψ_n by the symbol $|n\rangle$. Correspondingly, the complex conjugate ψ_m^* is denoted by $\langle m|$. The integral over the product of the two functions is then analogous to a scalar product (or inner product in linear algebra) of the abstract vectors, written

$$\int \psi_m^* \, \psi_n \, \mathrm{d}\tau = \langle m| \cdot |n\rangle \equiv \langle m|n\rangle \tag{9.6.21}$$

The last quantity is known as a *bracket*, which led Dirac to designate the vectors $\langle m|$ and $|n\rangle$ as a "bra" and a "ket," respectively. The orthonormality conditions (Equation 9.6.14) can be written

$$\langle m|n\rangle = \delta_{mn} \tag{9.6.22}$$

The integral of a "sandwich" containing an operator $\hat{A}$ can be written very compactly in the form

$$\int \psi_m^* \, \hat{A} \, \psi_n \, \mathrm{d}\tau = \langle m|A|n\rangle \tag{9.6.23}$$

The hermitian condition on $\hat{A}$ [cf. Eq (6)] is therefore expressed as

$$\langle m|A|n\rangle = \langle n|A|m\rangle^* \tag{9.6.24}$$

Expectation Values

One of the extraordinary features of quantum mechanics is the possibility for superpositions of states. The state of a system can sometimes exist as a linear combination of other states, for example,

$$\psi = c_1\psi_1 + c_2\psi_2 \tag{9.6.25}$$

Assuming that all three functions are normalized and that ψ_1 and ψ_2 are orthogonal, we find

$$\int \psi^* \, \psi \, \mathrm{d}\tau = |c_1|^2 + |c_2|^2 = 1 \tag{9.6.26}$$

We can interpret $|c_1|^2$ and $|c_2|^2$ as the probabilities that a system in a state described by ψ can have the attributes of the states ψ_1 and ψ_2, respectively. Suppose ψ_1 and ψ_2 represent eigenstates of an observable A, satisfying the respective eigenvalue equations

$$\hat{A}\psi_1 = a_1\psi_1 \quad and \quad \hat{A}\psi_2 = a_2\psi_2 \tag{9.6.27}$$

Then a large number of measurements of the variable A in the state ψ will register the value a_1 with a probability $|c_1|^2$ and the value a_2 with a probability $|c_2|^2$. The average value or *expectation value* of A will be given by

$$\langle A\rangle = |c_1|^2 a_1 + |c_2|^2 a_2 \tag{9.6.28}$$

This can be obtained directly from ψ by the "sandwich construction"

$$\langle A\rangle = \int \psi^* \hat{A} \, \psi \, \mathrm{d}\tau \tag{9.6.29}$$

or, if ψ is not normalized,

$$\langle A \rangle = \frac{\int \psi^* \hat{A}\, \psi \, \mathrm{d}\tau}{\int \psi^* \, \psi \, \mathrm{d}\tau} \tag{9.6.30}$$

Note that the expectation value need not itself be a possible result of a single measurement (like the centroid of a donut, which is located in the hole!). When the operator $\hat{A}$ is a simple function, not containing differential operators or the like, then Equation 9.6.29 reduces to the classical formula for an average value:

$$\langle A \rangle = \int A \, \rho \, \mathrm{d}\tau \tag{9.6.31}$$

More on Operators

An operator represents a prescription for turning one function into another: in symbols, $\hat{A}\psi = \phi$. From a physical point of view, the action of an operator on a wavefunction can be pictured as the process of measuring the observable A on the state ψ. The transformed wavefunction ϕ then represents the state of the system *after* the measurement is performed. In general, ϕ is different from ψ, consistent with the fact that the process of measurement on a quantum system produces an irreducible perturbation of its state. Only in the special case that ψ is an eigenstate of A, does a measurement preserve the original state. The function ϕ is then equal to an eigenvalue a times ψ.

The product of two operators, say $\hat{A}\hat{B}$, represents the successive action of the operators, reading from *right to left*---i.e., first $\hat{B}$ then $\hat{A}$. In general, the action of two operators in the reversed order, say $\hat{B}\hat{A}$, gives a different result, which can be written

$$\hat{A}\hat{B} \neq \hat{B}\hat{A}. \tag{9.6.32}$$

We say that the operators do not *commute*. This can be attributed to the perturbing effect one measurement on a quantum system can have on subsequent measurements. An example of non-commuting operators from everyday life. In our usual routine each morning, we shower and we get dressed. But the result of carrying out these operations in reversed order will be dramatically different!

The *commutator* of two operators is defined by

$$\left[\hat{A},\, \hat{B}\right] \equiv \hat{A}\hat{B} - \hat{B}\hat{A} \tag{9.6.33}$$

When $\left[\hat{A},\, \hat{B}\right] = 0$, the two operators are said to *commute*. This means their combined effect will be the same whatever order they are applied (like brushing your teeth and showering).

The uncertainty principle for simultaneous measurement of two observables A and B is closely related to their commutator. The uncertainty Δa in the observable A is defined in terms of the mean square deviation from the average:

$$(\Delta a)^2 = \langle (\hat{A} - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2 \tag{9.6.34}$$

It corresponds to the *standard deviation* in statistics. The following inequality can be proven for the product of two uncertainties:

$$\Delta a \Delta b \geq \frac{1}{2} |\langle \left[\hat{A},\, \hat{B}\right] \rangle| \tag{9.6.35}$$

The best known application of Equation 9.6.35 is to the position and momentum operators, say $\hat{x}$ and $\hat{p_x}$. Their commutator is given by

$$[\hat{x},\, \hat{p_x}\,] = i\hbar \tag{9.6.36}$$

so that

$$\Delta x \Delta p \geq \frac{\hbar}{2} \tag{9.6.37}$$

which is known as the *Heisenberg uncertainty principle*. This fundamental consequence of quantum theory implies that the position and momentum of a particle cannot be determined with arbitrary precision--the more accurately one is known, the more uncertain is the other. For example, if the momentum is known exactly, as in a momentum eigenstate, then the position is

completely undetermined.

If two operators commute, there is no restriction on the accuracy of their simultaneous measurement. For example, the x and y coordinates of a particle can be known at the same time. An important theorem states that two commuting observables can have simultaneous eigenfunctions. To prove this, write the eigenvalue equation for an operator $\hat{A}$

$$\hat{A}\,\psi_n = a_n\,\psi_n \tag{9.6.38}$$

then operate with $\hat{B}$ and use the commutativity of $\hat{A}$ and $\hat{B}$ to obtain

$$\hat{B}\,\hat{A}\,\psi_n = \hat{A}\,\hat{B}\,\psi_n = a_n\,\hat{B}\,\psi_n \tag{9.6.39}$$

This shows that $\hat{B}\ \psi_n$ is also an eigenfunction of $\hat{A}$ with the same eigenvalue a_n. This implies that

$$\hat{B}\,\psi_n = const\,\psi_n = b_n\,\psi_n \tag{9.6.40}$$

showing that ψ_n is a simultaneous eigenfunction of $\hat{A}$ and $\hat{B}$ with eigenvalues a_n and b_n, respectively. The derivation becomes slightly more complicated in the case of degenerate eigenfunctions, but the same conclusion follows.

After the Hamiltonian, the operators for angular momenta are probably the most important in quantum mechanics. The definition of angular momentum in classical mechanics is $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. In terms of its Cartesian components,

$$L_x = yp_z - zp_y \tag{9.6.41}$$

$$L_y = zp_x - xp_z \tag{9.6.42}$$

$$L_z = xp_y - yp_x \tag{9.6.43}$$

In future, we will write such sets of equation as "$L_x = yp_z - zp_y,\ \ et\ \ cyc$," meaning that we add to one explicitly stated relation, the versions formed by successive cyclic permutation $x \to y \to z \to x$. The general prescription for turning a classical dynamical variable into a quantum-mechanical operator was developed in Chap 2. The key relations were the momentum components

$$\hat{p_x} = -i\hbar\frac{\partial}{\partial x},\ \ \hat{p_y} = -i\hbar\frac{\partial}{\partial y},\ \ \hat{p_z} = -i\hbar\frac{\partial}{\partial z} \tag{9.6.44}$$

with the coordinates x, y, z simply carried over into multiplicative operators. Applying Equation 9.6.44 to Equation 9.6.43, we construct the three angular momentum operators

$$\hat{L_x} = -i\hbar\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right) \quad et\,cyc \tag{9.6.45}$$

while the total angular momentum is given by

$$\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 \tag{9.6.46}$$

The angular momentum operators obey the following commutation relations:

$$\left[\hat{L_x},\,\hat{L_y}\right] = i\hbar\hat{L_z}\ \ et\,cyc \tag{9.6.47}$$

but

$$\left[\hat{L}^2,\,\hat{L_z}\right] = 0 \tag{9.6.48}$$

and analogously for $\hat{L_x}$ and $\hat{L_y}$. This is consistent with the existence of simultaneous eigenfunctions of $\hat{L}^2$ and any one component, conventionally designated $\hat{L_z}$. But then these states *cannot* be eigenfunctions of either $\hat{L_x}$ or $\hat{L_y}$.

Postulates of Quantum Mechanics

Our development of quantum mechanics is now sufficiently complete that we can reduce the theory to a set of five postulates.

Postulate 1: Wavefunctions

The state of a quantum-mechanical system is completely specified by a wavefunction Ψ that depends on the coordinates and time. The square of this function $\Psi^*\Psi$ gives the probability density for finding the system with a specified set of coordinate values.

The wavefunction must fulfill certain mathematical requirements because of its physical interpretation. It must be single-valued, finite and continuous. It must also satisfy a normalization condition

$$\int \Psi^* \, \Psi \, \mathrm{d}\tau = 1 \tag{9.6.49}$$

Postulate 2: Observables

Every observable in quantum mechanics is represented by a linear, hermitian operator.

The hermitian property was defined in Equation 9.6.7. A linear operator is one which satisfies the identity

$$\hat{A}(c_1\psi_1 + c_2\psi_2) = c_1 \, \hat{A}\psi_1 + c_2 \, \hat{A}\psi_2 \tag{9.6.50}$$

which is required in order to have a superposition property for quantum states. The form of an operator which has an analog in classical mechanics is derived by the prescriptions

$$\hat{\mathbf{r}} = \mathbf{r}, \quad \hat{\mathbf{p}} = -i\hbar\nabla \tag{9.6.51}$$

which we have previously expressed in terms of Cartesian components [cf. Equation 9.6.44].

Postulate 3: Eigenstates

In any measurement of an observable A, associated with an operator $\hat{A}$, the only possible results are the eigenvalues a_n, which satisfy an eigenvalue equation

$$\hat{A}\psi_n = a_n \, \psi_n \tag{9.6.52}$$

This postulate captures the essence of quantum mechanics--the quantization of dynamical variables. A continuum of eigenvalues is not forbidden, however, as in the case of an unbound particle.

Every measurement of A invariably gives one of the eigenvalues. For an arbitrary state (not an eigenstate of A), these measurements will be individually unpredictable but follow a definite statistical law, which is the subject of the fourth postulate:

Postulate 4: Expectation Values

For a system in a state described by a normalized wave function Ψ, the average or expectation value of the observable corresponding to A is given by

$$\langle A \rangle = \int \Psi^* \, \hat{A} \, \Psi \, \mathrm{d}\tau \tag{9.6.53}$$

Finally,

Postulate 5: Time-dependent Evolution

The wavefunction of a system evolves in time in accordance with the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\,\Psi \tag{9.6.54}$$

For time-independent problems this reduces to the time-independent Schrödinger equation

$$\hat{H}\,\psi = E\,\psi \tag{9.6.55}$$

which is the eigenvalue equation for the Hamiltonian operator.

The Variational Principle

Except for a small number of intensively-studied examples, the Schrödinger equation for most problems of chemical interest *cannot* be solved exactly. The variational principle provides a guide for constructing the best possible approximate solutions of a specified functional form. Suppose that we seek an approximate solution for the ground state of a quantum system described by a Hamiltonian $\hat{H}$. We presume that the Schrödinger equation

$$\hat{H}\,\psi_0 = E\,\psi_0 \tag{9.6.56}$$

is too difficult to solve exactly. Suppose, however, that we have a function $\tilde{\psi}$ which we think is an approximation to the true ground-state wavefunction. According to the variational principle (or variational theorem), the following formula provides an *upper bound* to the exact ground-state energy E_0:

$$\tilde{E} \equiv \frac{\int \tilde{\psi}^*\hat{H}\,\tilde{\psi}\,\mathrm{d}\tau}{\int \tilde{\psi}^*\,\tilde{\psi}\,\mathrm{d}\tau} \geq E_0 \tag{9.6.57}$$

Note that this ratio of integrals has the same form as the expectation value $\langle H\rangle$ defined by Equation 9.6.30. The better the approximation $\tilde{\psi}$, the lower will be the computed energy $\tilde{E}$, though it will still be greater than the exact value. To prove Equation 9.6.57, we suppose that the approximate function can, in concept, be represented as a superposition of the actual eigenstates of the Hamiltonian, analogous to Equation 9.6.25,

$$\tilde{\psi} = c_0\psi_0 + c_1\psi_1 + \ldots \tag{9.6.58}$$

This means that $\tilde{\psi}$, the approximate ground state, might be close to the actual ground state ψ_0 but is "contaminated" by contributions from excited states ψ_1, ... Of course, none of the states or coefficients on the right-hand side is actually known, otherwise there would be no need to worry about approximate computations. By Equation 9.6.26, the expectation value of the Hamiltonian in the state Equation 9.6.58 is given by

$$\tilde{E} = |c_0|^2E_0 + |c_1|^2E_1 + \ldots \tag{9.6.59}$$

Since all the excited states have *higher* energy than the ground state, $E_1,\ E_2 \ldots \geq E_0$, we find

$$\tilde{E} \geq \left(|c_0|^2 + |c_1|^2 + \ldots\right) E_0 = E_0 \tag{9.6.60}$$

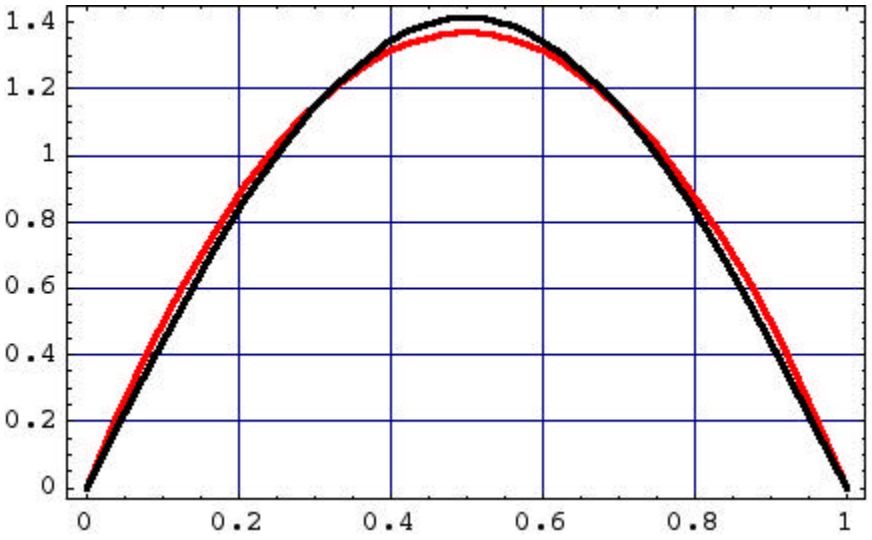
assuming $\tilde{\psi}$ has been normalized. Thus $\tilde{E}$ must be greater than the true ground-state energy E_0, as implied by Equation 9.6.57.

As a very simple, although artificial, illustration of the variational principle, consider the ground state of the particle in a box. Suppose we had never studied trigonometry and knew nothing about sines or cosines. Then a reasonable approximation to the ground state might be an inverted parabola such as the normalized function

$$\tilde{\psi}(x) = \left(\frac{30}{a^5}\right)^{\frac{1}{2}} x(a-x) \tag{9.6.61}$$

Fig. 1 shows this function along with the exact ground-state eigenfunction

$$\psi_1(x) = \left(\frac{2}{a}\right)^{\frac{1}{2}} sin\frac{\pi x}{a} \tag{9.6.62}$$



$$\frac{x}{a} \tag{9.6.63}$$

Figure 9.6.1***:*** *Variational approximation for particle in a box. Red line represents* $\tilde{\psi}$ *and black line represents* ψ_1

A variational calculation gives

$$\tilde{E} = \int_0^a \tilde{\psi}(x) \left(-\frac{\hbar^2}{2m}\right) \tilde{\psi}''(x)\, dx = \frac{5}{4\pi^2}\frac{h^2}{ma^2} = \frac{10}{\pi^2} E_1 \approx 1.01321 E_1 \tag{9.6.64}$$

in terms of the exact ground state energy $E_1 = \frac{h^2}{8ma^2}$. In accord with the variational theorem, $\tilde{E} > E_1$. The computation is in error by about 1%.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)
- Hannah Jaroh, Hope College, Holland, MI

This page titled 9.6: Principles of Quantum Mechanics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Chapter 4. Principles of Quantum Mechanics is licensed CC BY-NC-SA 4.0.

9.7: Harmonic Oscillator

The harmonic oscillator is a model which has several important applications in both classical and quantum mechanics. It serves as a prototype in the mathematical treatment of such diverse phenomena as elasticity, acoustics, AC circuits, molecular and crystal vibrations, electromagnetic fields and optical properties of matter.

Classical Oscillator

A simple realization of the harmonic oscillator in classical mechanics is a particle which is acted upon by a restoring force proportional to its displacement from its equilibrium position. Considering motion in one dimension, this means

$$F = -kx \tag{9.7.1}$$

Such a force might originate from a spring which obeys Hooke's law, as shown in Figure 9.7.1. According to Hooke's law, which applies to real springs for sufficiently small displacements, the restoring force is proportional to the displacement—either stretching or compression—from the equilibrium position.

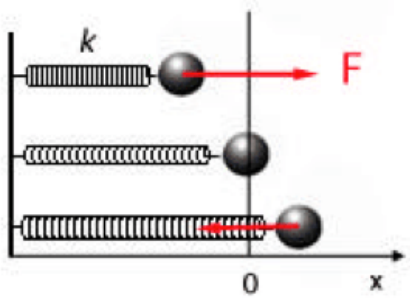


Figure 9.7.1: Spring obeying Hooke's law.

The *force constant* k is a measure of the stiffness of the spring. The variable x is chosen equal to zero at the equilibrium position, positive for stretching, negative for compression. The negative sign in Equation 9.7.1 reflects the fact that F is a *restoring* force, always in the opposite sense to the displacement x.

Applying Newton's second law to the force from Equation 9.7.1, we find x

$$F = m\frac{d^2x}{dx^2} = -kx \tag{9.7.2}$$

where m is the mass of the body attached to the spring, which is itself assumed massless. This leads to a differential equation of familiar form, although with different variables:

$$\ddot{x}(t) + \omega^2 x(t) = 0 \tag{9.7.3}$$

with

$$\omega^2 \equiv \frac{k}{m} \tag{9.7.4}$$

The dot notation (introduced by Newton himself) is used in place of primes when the independent variable is time. The general solution to Equation 9.7.3 is

$$x(t) = A\sin\omega t + B\cos\omega t \tag{9.7.5}$$

which represents periodic motion with a sinusoidal time dependence. This is known as *simple harmonic motion* and the corresponding system is known as a *harmonic oscillator*. The oscillation occurs with a constant angular frequency

$$\omega = \sqrt{\frac{k}{m}} \text{ radians per second} \tag{9.7.6}$$

This is called the *natural frequency* of the oscillator. The corresponding *circular (or angular) frequency* in Hertz (cycles per second) is

$$\nu = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \text{ Hz} \tag{9.7.7}$$

The general relation between force and potential energy in a conservative system in one dimension is

$$F = \frac{-dV}{dx} \tag{9.7.8}$$

Thus the potential energy of a harmonic oscillator is given by

$$V(x) = \frac{1}{2}kx^2 \tag{9.7.9}$$

which has the shape of a parabola, as drawn in Figure 9.7.2. A simple computation shows that the oscillator moves between positive and negative turning points $\pm x_{max}$ where the total energy E equals the potential energy $\frac{1}{2}kx_{max}^2$ while the kinetic energy is momentarily zero. In contrast, when the oscillator moves past $x = 0$, the kinetic energy reaches its maximum value while the potential energy equals zero.

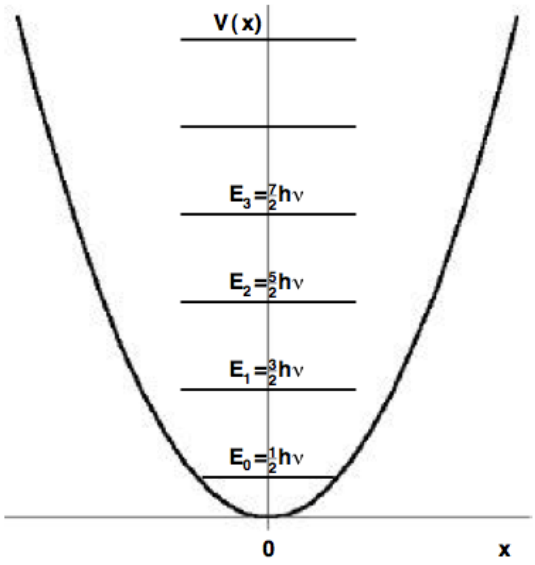


Figure 9.7.2: Potential energy function and first few energy levels for harmonic oscillator.

Harmonic Oscillator in Quantum Mechanics

Given the potential energy in Equation 9.7.9, we can write down the Schrödinger equation for the one-dimensional harmonic oscillator:

$$-\frac{\hbar^2}{2m}\psi''(x) + \frac{1}{2}kx^2\psi(x) = E\psi(x) \tag{9.7.10}$$

For the first time we encounter a differential equation with *non-constant* coefficients, which is a much greater challenge to solve. We can combine the constants in Equation 9.7.10 to two parameters

$$\alpha^2 = \frac{mk}{\hbar^2} \tag{9.7.11}$$

and

$$\lambda = \frac{2mE}{\hbar^2\alpha} \tag{9.7.12}$$

and redefine the independent variable as

$$\xi = \alpha^{1/2}x \tag{9.7.13}$$

This reduces the Schrödinger equation to

$$\psi''(\xi) + (\lambda - \xi^2)\psi(\xi) = 0 \tag{9.7.14}$$

The range of the variable x (also ξ) must be taken from $-\infty$ to $+\infty$, there being no finite cutoff as in the case of the particle in a box. A useful first step is to determine the asymptotic solution to Equation 9.7.13, that is, the form of $\psi(\xi)$ as $\xi \to \pm\infty$. For sufficiently large values of $|\xi|$, $\xi^2 \gg \lambda$ and the differential equation is approximated by

$$\psi''(\xi) - \xi^2\psi(\xi) \approx 0 \tag{9.7.15}$$

This suggests the following manipulation:

$$\left(\frac{d^2}{d\xi^2} - \xi^2\right)\psi(\xi) \approx \left(\frac{d}{d\xi} - \xi\right)\left(\frac{d}{d\xi} + \xi\right)\psi(\xi) \approx 0 \tag{9.7.16}$$

The first-order differential equation

$$\psi'(\xi) + \xi\psi(\xi) = 0 \tag{9.7.17}$$

can be solved exactly to give

$$\psi(\xi) = \text{const.}\, e^{-\xi^2/2} \tag{9.7.18}$$

Remarkably, this turns out to be an *exact* solution of the Schrödinger equation (Equation 9.7.14) with $\lambda = 1$. Using Equation 9.7.12 , this corresponds to an energy

$$E = \frac{\lambda\hbar^2\alpha}{2m} = \frac{1}{2}\hbar\sqrt{\frac{k}{m}} = \frac{1}{2}\hbar\omega \tag{9.7.19}$$

where ω is the natural frequency of the oscillator according to classical mechanics. The function in Equation 9.7.18 has the form of a Gaussian, the bell-shaped curve so beloved in the social sciences. The function has no nodes, which leads us to conclude that this represents the ground state of the system.The ground state is usually designated with the quantum number $n = 0$ (the particle in a box is a exception, with $n = 1$ labeling the ground state). Reverting to the original variable x, we write

$$\psi_0(x) = \text{const}e^{-\alpha x^2/2} \tag{9.7.20}$$

with

$$\alpha = (mk/\hbar^2)^{1/2} \tag{9.7.21}$$

With help of the well-known definite integral (Laplace 1778)

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}dx = \sqrt{\frac{\pi}{\alpha}} \tag{9.7.22}$$

we find the normalized eigenfunction

$$\psi_0(x) = (\frac{\alpha}{\pi})^{1/4}e^{-\alpha x^2/2} \tag{9.7.23}$$

with the corresponding eigenvalue

$$E_0 = \frac{1}{2}\hbar\omega \tag{9.7.24}$$

Drawing from our experience with the particle in a box, we might surmise that the first excited state of the harmonic oscillator

would be a function similar to Equation 9.7.23, but with a node at $x = 0$, say,

$$\psi_1(x) = const x e^{-\alpha x^2/2} \tag{9.7.25}$$

This is orthogonal to $\psi_0(x)$ by symmetry and is indeed an eigenfunction with the eigenvalue

$$E_1 = \frac{3}{2}\hbar\omega \tag{9.7.26}$$

Continuing the process, we try a function with two nodes

$$\psi_2 = const(x^2 - a)e^{-\alpha x^2/2} \tag{9.7.27}$$

Using the integrals tabulated in the Supplement 5, on Gaussian Integrals, we determine that with $a = \frac{1}{2}$ makes $\psi_2(x)$ orthogonal to $\psi_0(x)$ and $\psi_1(x)$. We verify that this is another eigenfunction, corresponding to

$$E_2 = \frac{5}{2}\hbar\omega \tag{9.7.28}$$

The general result, which follows from a more advanced mathematical analysis, gives the following formula for the normalized eigenfunctions:

$$\psi_n(x) = (\frac{\sqrt{\alpha}}{2^n n!\sqrt{\pi}})^{1/2} H_n(\sqrt{\alpha}x)e^{-\alpha x^2/2} \tag{9.7.29}$$

where $H_n(\xi)$ represents the Hermite polynomial of degree n. The first few Hermite polynomials are

$$H_0(\xi) = 1 \tag{9.7.30}$$

$$H_1(\xi) = 2\xi \tag{9.7.31}$$

$$H_2(\xi) = 4\xi^2 - 2 \tag{9.7.32}$$

$$H_3(\xi) = 8\xi^3 - 12\xi \tag{9.7.33}$$

The four lowest harmonic-oscillator eigenfunctions are plotted in Figure 9.7.3. Note the topological resemblance to the corresponding particle-in-a-box eigenfunctions.

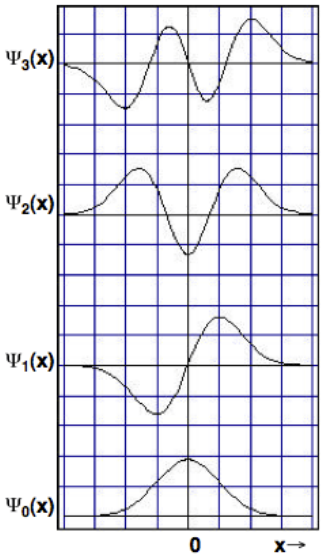


Figure 9.7.3: Harmonic oscillator eigenfunctions for n=0, 1, 2, 3.

The eigenvalues are given by the simple formula

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \tag{9.7.34}$$

These are drawn in Figure $9.7.2$, on the same scale as the potential energy. The ground-state energy $E_0 = \frac{1}{2}\hbar\omega$ is greater than the classical value of zero, again a consequence of the uncertainty principle. This means that the oscillator is always oscillating.

It is remarkable that the difference between successive energy eigenvalues has a constant value

$$\Delta E = E_{n+1} - E_n = \hbar\omega = h\nu \tag{9.7.35}$$

This is reminiscent of Planck's formula for the energy of a photon. It comes as no surprise then that the quantum theory of radiation has the structure of an assembly of oscillators, with each oscillator representing a mode of electromagnetic waves of a specified frequency.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)

This page titled 9.7: Harmonic Oscillator is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Harmonic Oscillator is licensed CC BY-NC-SA 4.0.

9.8: Angular Momentum

Particle in a Ring

Consider a variant of the one-dimensional particle in a box problem in which the x-axis is bent into a ring of radius R. We can write the same Schrödinger equation

$$\frac{-\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x) \tag{9.8.1}$$

There are no boundary conditions in this case since the x-axis closes upon itself. A more appropriate independent variable for this problem is the angular position on the ring given by, $\phi = x/R$. The Schrödinger equation would then read

$$-\frac{\hbar^2}{2mR^2}\frac{d^2\psi(\phi)}{d(\phi)^2} = E\psi(\phi) \tag{9.8.2}$$

The kinetic energy of a body rotating in the xy-plane can be expressed as

$$E = \frac{L_z^2}{2I} \tag{9.8.3}$$

where

$$I = mR^2 \tag{9.8.4}$$

is the moment of inertia and L_z, the *z*-component of angular momentum. (Since $L = r \times p$, if ***r*** and ***p*** lie in the *xy*-plane, ***L*** points in the *z*-direction.) The structure of Equation 9.8.2 suggests that this angular-momentum operator is given by

$$\hat{L_z} = -i\hbar\frac{\partial}{\partial\phi} \tag{9.8.5}$$

This result will follow from a more general derivation in the following Section. The Schrödinger equation (Equation 9.8.2) can now be written more compactly as

$$\psi'' (\phi) + m^2\psi(\phi) = 0 \tag{9.8.6}$$

where

$$m^2 \equiv 2IE/\hbar^2 \tag{9.8.7}$$

(Please do not confuse this variable m with the mass of the particle!) Possible solutions to (Equation 9.8.6) are

$$\psi(\phi) = \text{const}\, e^{\pm im\phi} \tag{9.8.8}$$

For this wavefunction to be physically acceptable, it must be *single-valued*. Since ϕ increased by any multiple of 2π represents the same point on the ring, we must have

$$\psi(\phi + 2\pi) = \psi(\phi) \tag{9.8.9}$$

and therefore

$$e^{im(\phi+2\pi)} = e^{im\phi} \tag{9.8.10}$$

This requires that

$$e^{2\pi im} = 1 \tag{9.8.11}$$

which is true only if m is an integer:

$$m = 0, \pm 1, \pm 2... \tag{9.8.12}$$

Using Equation 9.8.7, this gives the quantized energy values

$$E_m = \frac{\hbar^2}{2I} m^2 \tag{9.8.13}$$

In contrast to the particle in a box, the eigenfunctions corresponding to $+m$ and $-m$ (Equation 9.8.8) are linearly independent, so both must be accepted. Therefore all eigenvalues, except E_0, are two-fold (or doubly) degenerate. The eigenfunctions can all be written in the form const $e^{im\phi}$, with *m* allowed to take either positive and negative values (or 0), as in Equation 9.8.11. The normalized eigenfunctions are

$$\psi_m(\phi) = \frac{1}{\sqrt{2\pi}} e^{im\phi} \tag{9.8.14}$$

and can be verified to satisfy the normalization condition containing the complex conjugate

$$\int_0^{2\pi} \psi_m^*(\phi)\psi_m(\phi)d\phi = 1 \tag{9.8.15}$$

where we have noted that $\psi_m^*(\phi) = (2\pi)^{-1/2} e^{-im\phi}$. The mutual orthogonality of the functions (Equation 9.8.14) also follows easily, for

$$\int_0^{2\pi} \psi_{m'}^* \psi_m(\phi)d\phi = \frac{1}{2\pi} \int_0^{2\pi} e^{i(m-m')\phi} d\phi \tag{9.8.16}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} [cos(m-m')\phi + isin(m-m')\phi]d\phi = 0 \tag{9.8.17}$$

for $m' \neq m$.

The solutions in Equation 9.8.13 are also eigenfunctions of the angular momentum operator (Equation 9.8.5), with

$$\hat{L}_z \psi_m(\phi) = m\hbar\psi_m(\phi), m = 0, \pm 1, \pm 2... \tag{9.8.18}$$

This is a instance of a fundamental result in quantum mechanics, that any measured component of orbital angular momentum is restricted to integral multiples of $\hbar$. The Bohr theory of the hydrogen atom, to be discussed in the next Chapter, can be derived from this principle alone.

Free Electron Model for Aromatic Molecules

The benzene molecule consists of a ring of six carbon atoms around which six delocalized *pi*-electrons can circulate. A variant of the FEM for rings predicts the ground-state electron configuration which we can write as $1\pi^2 2\pi^4$, as shown here:

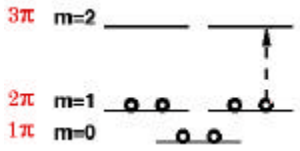


Figure taken from: S. M. Blinder "Introduction to Quantum Mechanics: In Chemistry, Materials Science, and Biology," pg 79, 2004.

The enhanced stability the benzene molecule can be attributed to the complete shells of π-electron orbitals, analogous to the way that noble gas electron configurations achieve their stability. Naphthalene, apart from the central C-C bond, can be modeled as a ring containing 10 electrons in the next closed-shell configuration$1\pi^2 2\pi^4 3\pi^4$. These molecules fulfill Hückel's "4*N*+2 rule" for aromatic stability. The molecules cyclobutadiene $(1\pi^2 2\pi^2)$ and cyclooctatetraene$(1\pi^2 2\pi^4 3\pi^2)$, even though they consist of rings with alternating single and double bonds, do *not* exhibit aromatic stability since they contain partially-filled orbitals.

The longest wavelength absorption in the benzene spectrum can be estimated according to this model as

$$\frac{hc}{\lambda} = E_2 - E_1 = \frac{\hbar^2}{2mR^2}(2^2 - 1^2) \tag{9.8.19}$$

The ring radius *R* can be approximated by the C-C distance in benzene, 1.39 Å. We predict $\lambda \approx 210$ nm, whereas the experimental absorption has $\lambda_{max} \approx 268$ nm.

Spherical Polar Coordinates

The motion of a free particle on the surface of a sphere will involve components of angular momentum in three-dimensional space. Spherical polar coordinates provide the most convenient description for this and related problems with spherical symmetry. The position of an arbitrary point ***r*** is described by three coordinates r, θ, ϕ as shown in Figure 9.8.2.

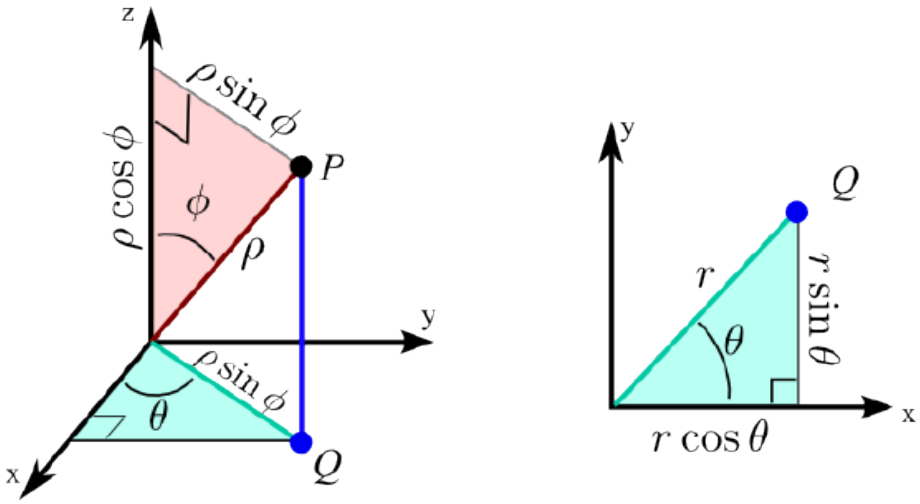


Figure 9.8.2: Spherical coordinates.

These are connected to Cartesian coordinates by the relations

$$x = r \sin\theta \cos\phi \tag{9.8.20}$$

$$y = r \sin\theta \sin\phi \tag{9.8.21}$$

$$z = r \cos\theta \tag{9.8.22}$$

The radial variable *r* represents the distance from **r** to the origin, or the length of the vector **r**:

$$r = \sqrt{x^2 + y^2 + z^2} \tag{9.8.23}$$

The coordinate θ is the angle between the vector **r** and the *z*-axis, similar to latitude in geography, but with \(\theta= 0 \) and $\theta = \pi$ corresponding to the North and South Poles, respectively. The angle ϕ describes the rotation of **r** about the *z*-axis, running from 0 to 2π, similar to geographic longitude. The volume element in spherical polar coordinates is given by

$$d\tau = r^2 \sin\theta dr d\theta d\phi, \tag{9.8.24}$$

$$r^{\text{int}}\{0, \infty\}, \theta^{\text{int}}\{0, \pi\}, \phi^{\text{int}}\{0, 2\pi\} \tag{9.8.25}$$

and represented graphically by the distorted cube in Figure 9.8.1.

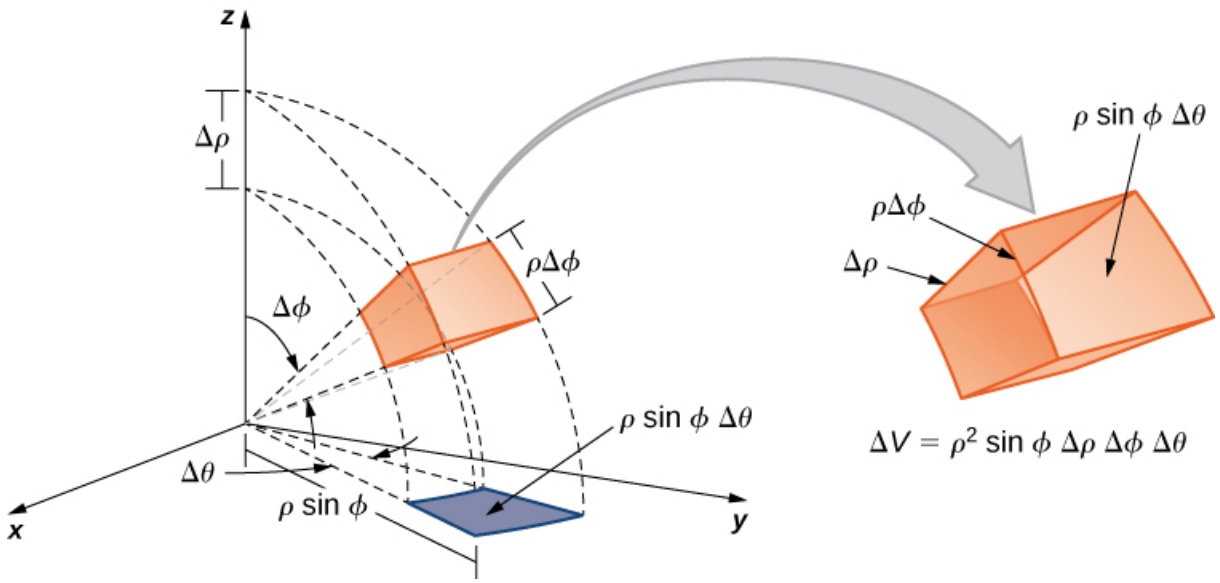


Figure 9.8.3***:*** *Volume element in spherical polar coordinates. (CC BY; OpenStax).*

We also require the Laplacian operator

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{r^2 sin\theta}\frac{\partial}{\partial\theta}sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2 sin^2\theta}\frac{\partial^2}{\partial\phi^2} \tag{9.8.26}$$

A detailed derivation is given in Supplement 6.

Rotation in Three Dimensions

A particle of mass *M*, free to move on the surface of a sphere of radius *R*, can be located by the two angular variables θ, ϕ. The Schrödinger equation therefore has the form

$$-\frac{\hbar^2}{2M}\nabla^2 Y(\theta,\phi) = EY(\theta,\phi) \tag{9.8.27}$$

with the wavefunction conventionally written as $Y(\theta,\phi)$. These functions are known as *spherical harmonics* and have been used in applied mathematics long before quantum mechanics. Since \(*r* = *R*\), a constant, the first term in the Laplacian does not contribute. The Schrödinger equation reduces to

$$\left\{\frac{1}{sin\theta}\frac{\partial}{\partial\theta}sin\theta\frac{\partial}{\partial\theta} + \frac{1}{sin^2\theta}\frac{\partial^2}{\partial\phi^2} + \lambda\right\}Y(\theta,\phi) = 0 \tag{9.8.28}$$

where

$$\lambda = \frac{2MR^2E}{\hbar^2} = \frac{2IE}{\hbar^2} \tag{9.8.29}$$

again introducing the moment of inertia $I = MR^2$. The variables θ and ϕ can be separated in Equation ??? after multiplying through by $sin^2\theta$. If we write

$$Y(\theta,\phi) = \Theta(\theta)\Phi(\phi) \tag{9.8.30}$$

and follow the procedure used for the three-dimensional box, we find that dependence on ϕ alone occurs in the term

$$\frac{\Phi''(\phi)}{\Phi(\phi)} = const \tag{9.8.31}$$

This is identical in form to Equation 9.8.6, with the constant equal to $-m^2$, and we can write down the analogous solutions

$$\Phi_m(\phi) = \sqrt{\frac{1}{2\pi}} e^{im\phi}, m = 0, \pm 1, \pm 2... \tag{9.8.32}$$

Substituting Equation ??? into Equation ??? and cancelling the functions $\Phi(\phi)$, we obtain an ordinary differential equation for $\Theta(\theta)$

$$\left\{ \frac{1}{sin\theta} \frac{d}{d\theta} sin\theta \frac{d}{d\theta} - \frac{m^2}{sin^2\theta} + \lambda \right\} \Theta(\theta) = 0 \tag{9.8.33}$$

Consulting our friendly neighborhood mathematician, we learn that the single-valued, finite solutions to (Equation ???) are known as *associated Legendre functions*. The parameters λ and m are restricted to the values

$$\lambda = \ell(\ell + 1), \ell = 0, 1, 2... \tag{9.8.34}$$

while

$$m = 0, \pm 1, \pm 2... \pm \ell(2\ell + 1 values) \tag{9.8.35}$$

Putting Equation ??? into Equation ???, the allowed energy levels for a particle on a sphere are found to be

$$E_\ell = \frac{\hbar^2}{2I} \ell(\ell + 1) \tag{9.8.36}$$

Since the energy is independent of the second quantum number *m*, the levels (Equation ???) are $(2\ell + 1)$-fold degenerate. The spherical harmonics constitute an orthonormal set satisfying the integral relations

$$\int_0^\pi \int_0^{2\pi} Y^*_{\ell' m'}(\theta, \phi) Y_{\ell m}(\theta, \phi) sin\theta d\theta d\phi = \delta_{\ell\ell'} \delta_{mm'} \tag{9.8.37}$$

The following table lists the spherical harmonics through ℓ = 2, which will be sufficient for our purposes.

$$Spherical Harmonics Y_{\ell m}(\theta, \phi) \tag{9.8.38}$$

$$Y_{00} = \left(\frac{1}{4\pi} \right)^{1/2} \tag{9.8.39}$$

$$Y_{10} = \left(\frac{3}{4\pi} \right)^{1/2} cos\theta \tag{9.8.40}$$

$$Y_{1\pm1} = \mp \left(\frac{3}{4\pi} \right)^{1/2} sin\theta e^{\pm i\phi} \tag{9.8.41}$$

$$Y_{20} = \left(\frac{5}{16\pi} \right)^{1/2} (3cos^2\theta - 1) \tag{9.8.42}$$

$$Y_{2\pm1} = \mp \left(\frac{15}{8\pi} \right)^{1/2} cos\theta sin\theta e^{\pm i\phi} \tag{9.8.43}$$

$$Y_{2\pm2} = \left(\frac{15}{32\pi} \right)^{1/2} sin^2\theta e^{\pm 2i\phi} \tag{9.8.44}$$

A graphical representation of these functions is given in Figure 9.8.4. Surfaces of constant absolute value are drawn, positive where green and negative where red.

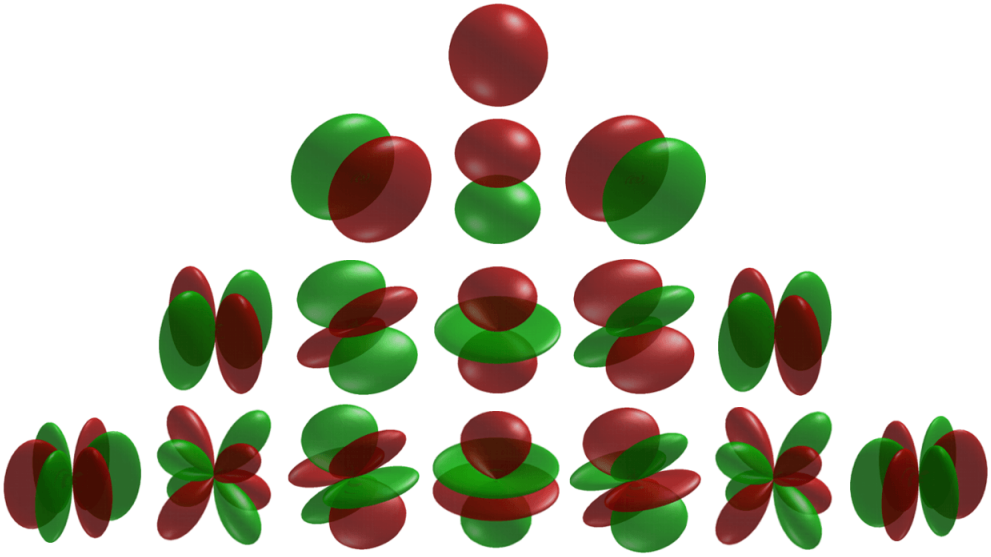


Figure used with permission (CC-SA-BY- 3.0; Sarxos)

Theory of Angular Momentum

Generalization of the energy-angular momentum relation in Equation 9.8.3 to three dimensions gives

$$E = \frac{L^2}{2I} \tag{9.8.45}$$

Thus from Equation ???-??? we can identify the operator for the square of total angular momentum

$$\hat{L}^2 = -\hbar^2 \left\{ \frac{1}{sin\theta} \frac{\partial}{\partial\theta} sin\theta \frac{\partial}{\partial\theta} + \frac{1}{sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right\} \tag{9.8.46}$$

By Equations ??? and ???, the functions $Y(\theta, \phi)$ are simultaneous eigenfunctions of $\hat{L}^2$ and $\hat{L}_z$ such that

$$\hat{L}^2 Y_{\ell m}(\theta, \phi) = \ell(\ell + 1)\hbar^2 Y_{\ell m}(\theta, \phi) \tag{9.8.47}$$

and

$$\hat{L}_z Y_{\ell m}(\theta, \phi) = m\hbar Y_{\ell m}(\theta, \phi) \tag{9.8.48}$$

But the $Y_{\ell m}(\theta, \phi)$ are not eigenfunctions of either L_x and L_y (unless ℓ = 0). Note that the magnitude of the total angular momentum $\sqrt{\ell(\ell + 1)}\hbar$ is greater than its maximum observable component in any direction, namely $\ell\hbar$. The quantum-mechanical behavior of the angular momentum and its components can be represented by a vector model, illustrated in Figure 5. The angular momentum vector **L**, with magnitude $\sqrt{\ell(\ell + 1)}\hbar$, can be pictured as precessing about the z-axis, with its *z*-component L_z constant. The components L_x and L_y fluctuate in the course of precession, corresponding to the fact that the system is not in an eigenstate of either. There are 2ℓ + 1 different allowed values for L_z, with eigenvalues $m\hbar(m = 0, \pm1, \pm2... \pm \ell)$ equally spaced between $+\ell\hbar$ and $-\ell\hbar$.

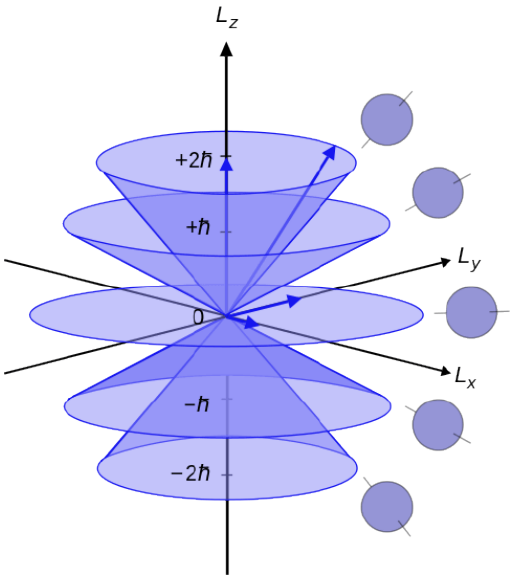


Figure $9.8.5$***:*** *Vector model for angular momentum, showing the case* ℓ*= 2. (Public Domain; Maschen).*

This discreteness in the allowed directions of the angular momentum vector is called *space quantization*. The existence of simultaneous eigenstates of $\hat{L}^2$ and any one component, conventionally $\hat{L}_z$, is consistent with the commutation relations derived in Chap. 4:

$$\left[\hat{L}_x, \hat{L}_y\right] = i\hbar\hat{L}_z etcyc \tag{9.8.49}$$

and

$$\left[\hat{L}^2, \hat{L}_z\right] = 0 \tag{9.8.50}$$

Electron Spin

The electron, as well as certain other fundamental particles, possesses an intrinsic angular momentum or *spin*, in addition to its orbital angular momentum. These two types of angular momentum are analogous to the daily and annual motions, respectively, of the Earth around the Sun. To distinguish the spin angular momentum from the orbital, we designate the quantum numbers as s and m_s, in place of ℓ and *m*. For the electron, the quantum number *s* always has the value $\frac{1}{2}$, while m_s can have one of two values,$\pm\frac{1}{2}$. The electron is said to be an elementary particle of spin $\frac{1}{2}$. The proton and neutron also have spin $\frac{1}{2}$ and belong to the classification of particles called *fermions*, which are governed by the Pauli exclusion principle. Other particles, including the photon, have integer values of spin and are classified as *bosons*. These do *not* obey the Pauli principle, so that an arbitrary number can occupy the same quantum state. A complete theory of spin requires relativistic quantum mechanics. For our purposes, it is sufficient to recognize the two possible internal states of the electron, which can be called `spin up' and `spin down.' These are designated, respectively, by α and β as factors in the electron wavefunction. Spins play an essential role in determining the possible electronic states of atoms and molecules.

This page titled 9.8: Angular Momentum is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- 6: Angular Momentum by Chris Geiersbach is licensed CC BY-NC-SA 4.0.

9.9: Hydrogen Atom

Atomic Spectra

When gaseous hydrogen in a glass tube is excited by a 5000-volt electrical discharge, four lines are observed in the visible part of the emission spectrum: red at 656.3 nm, blue-green at 486.1 nm, blue violet at 434.1 nm and violet at 410.2 nm:



Figure $9.9.1$: Visible spectrum of atomic hydrogen.

Other series of lines have been observed in the ultraviolet and infrared regions. Rydberg (1890) found that all the lines of the atomic hydrogen spectrum could be fitted to a single formula

$$\frac{1}{\lambda} = \mathcal{R}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right), \quad n_1 = 1,\ 2,\ 3..., \ n_2 > n_1 \tag{9.9.1}$$

where $\mathcal{R}$, known as the Rydberg constant, has the value $109,677$ cm^{-1} for hydrogen. The reciprocal of wavelength, in units of cm^{-1}, is in general use by spectroscopists. This unit is also designated *wavenumbers*, since it represents the number of wavelengths per cm. The Balmer series of spectral lines in the visible region, shown in Figure $9.9.1$, correspond to the values $n_1 = 2,\ n_2 = 3,\ 4,\ 5$ and 6. The lines with $n_1 = 1$ in the ultraviolet make up the Lyman series. The line with $n_2 = 2$, designated the Lyman alpha, has the longest wavelength (lowest wavenumber) in this series, with $1/\lambda = 82.258$ cm^{-1} or $\lambda = 121.57$ nm.

Other atomic species have line spectra, which can be used as a "fingerprint" to identify the element. However, no atom other than hydrogen has a simple relation analogous to Equation $9.9.1$ for its spectral frequencies. Bohr in 1913 proposed that all atomic spectral lines arise from transitions between discrete energy levels, giving a photon such that

$$\Delta E = h\nu = \frac{hc}{\lambda} \tag{9.9.2}$$

This is called the *Bohr frequency condition*. We now understand that the atomic transition energy ΔE is equal to the energy of a photon, as proposed earlier by Planck and Einstein.

The Bohr Atom

The nuclear model proposed by Rutherford in 1911 pictures the atom as a heavy, positively-charged nucleus, around which much lighter, negatively-charged electrons circulate, much like planets in the Solar system. This model is however completely untenable from the standpoint of classical electromagnetic theory, for an accelerating electron (circular motion represents an acceleration) should radiate away its energy. In fact, a hydrogen atom should exist for no longer than 5×10^{-11} sec, time enough for the electron's death spiral into the nucleus. This is one of the worst quantitative predictions in the history of physics. It has been called the Hindenberg disaster on an atomic level. (Recall that the Hindenberg, a hydrogen-filled dirigible, crashed and burned in a famous disaster in 1937.)

Bohr sought to avoid an atomic catastrophe by proposing that certain orbits of the electron around the nucleus could be exempted from classical electrodynamics and remain stable. The Bohr model was quantitatively successful for the hydrogen atom, as we shall now show.

We recall that the attraction between two opposite charges, such as the electron and proton, is given by Coulomb's law

$$F = \begin{cases} -\dfrac{e^2}{r^2} & \text{(gaussian units)} \\ -\dfrac{e^2}{4\pi\epsilon_0 r^2} & \text{(SI units)} \end{cases} \tag{9.9.3}$$

We prefer to use the Gaussian system in applications to atomic phenomena. Since the Coulomb attraction is a central force (dependent only on r), the potential energy is related by

$$F = -\frac{dV(r)}{dr} \tag{9.9.4}$$

We find therefore, for the mutual potential energy of a proton and electron,

$$V(r) = -\frac{e^2}{r} \tag{9.9.5}$$

Bohr considered an electron in a circular orbit of radius r around the proton. To remain in this orbit, the electron must be experiencing a centripetal acceleration

$$a = -\frac{v^2}{r} \tag{9.9.6}$$

where v is the speed of the electron. Using Equations 9.9.4 and 9.9.6 in Newton's second law, we find

$$\frac{e^2}{r^2} = \frac{mv^2}{r} \tag{9.9.7}$$

where m is the mass of the electron. For simplicity, we assume that the proton mass is infinite (actually $m_p \approx 1836 m_e$) so that the proton's position remains fixed. We will later correct for this approximation by introducing reduced mass. The energy of the hydrogen atom is the sum of the kinetic and potential energies:

$$E = T + V = \frac{1}{2}mv^2 - \frac{e^2}{r} \tag{9.9.8}$$

Using Equation 9.9.7, we see that

$$T = -\frac{1}{2}V \qquad \text{and} \qquad E = \frac{1}{2}V = -T \tag{9.9.9}$$

This is the form of the virial theorem for a force law varying as r^{-2}. Note that the energy of a bound atom is *negative*, since it is lower than the energy of the separated electron and proton, which is taken to be zero.

For further progress, we need some restriction on the possible values of r or v. This is where we can introduce the quantization of angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. Since $\mathbf{p}$ is perpendicular to $\mathbf{r}$, we can write simply

$$L = rp = mvr \tag{9.9.10}$$

Using Equation 9.9.9, we find also that

$$r = \frac{L^2}{me^2} \tag{9.9.11}$$

We introduce angular momentum quantization, writing

$$L = n\hbar, \qquad n = 1,\ 2... \tag{9.9.12}$$

excluding $n = 0$, since the electron would then not be in a circular orbit. The allowed orbital radii are then given by

$$r_n = n^2 a_0 \tag{9.9.13}$$

where

$$a_0 \equiv \frac{\hbar^2}{me^2} = 5.29 \times 10^{-11}\ \text{m} = 0.529 \mathring{A} \tag{9.9.14}$$

which is known as the *Bohr radius*. The corresponding energy is

$$E_n = -\frac{e^2}{2a_0 n^2} = -\frac{me^4}{2\hbar^2 n^2}, \qquad n = 1,\ 2... \tag{9.9.15}$$

Rydberg's formula (Equation 9.9.1) can now be deduced from the Bohr model. We have

$$\frac{hc}{\lambda} = E_{n_2} - E_{n_1} = \frac{2\pi^2 me^4}{h^2}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) \tag{9.9.16}$$

and the Rydbeg constant can be identified as

$$\mathcal{R} = \frac{2\pi^2 me^4}{h^3 c} \approx 109,737\ \text{cm}^{-1} \tag{9.9.17}$$

The slight discrepency with the experimental value for hydrogen $(109,677)$ is due to the finite proton mass. This will be corrected later.

The Bohr model can be readily extended to hydrogenlike ions, systems in which a single electron orbits a nucleus of arbitrary atomic number Z. Thus $Z = 1$ for hydrogen, $Z = 2$ for He^+, $Z = 3$ for Li^{++}, and so on. The Coulomb potential 9.9.5 generalizes to

$$V(r) = -\frac{Ze^2}{r}, \tag{9.9.18}$$

the radius of the orbit (Equation 9.9.13) becomes

$$r_n = \frac{n^2 a_0}{Z} \tag{9.9.19}$$

and the energy Equation 9.9.15 becomes

$$E_n = -\frac{Z^2 e^2}{2a_0 n^2} \tag{9.9.20}$$

De Broglie's proposal that electrons can have wavelike properties was actually inspired by the Bohr atomic model. Since

$$L = rp = n\hbar = \frac{nh}{2\pi} \tag{9.9.21}$$

we find

$$2\pi r = \frac{nh}{p} = n\lambda \tag{9.9.22}$$

Therefore, each allowed orbit traces out an integral number of de Broglie wavelengths.

Wilson (1915) and Sommerfeld (1916) generalized Bohr's formula for the allowed orbits to

$$\oint p\, dr = nh, \qquad n = 1,\ 2... \tag{9.9.23}$$

The Sommerfeld-Wilson quantum conditions Equation 9.9.23 reduce to Bohr's results for circular orbits, but allow, in addition, elliptical orbits along which the momentum p is variable. According to Kepler's first law of planetary motion, the orbits of planets are ellipses with the Sun at one focus. Figure 9.9.2 shows the generalization of the Bohr theory for hydrogen, including the elliptical orbits. The lowest energy state $n = 1$ is still a circular orbit. But $n = 2$ allows an elliptical orbit in addition to the circular one; $n = 3$ has three possible orbits, and so on. The energy still depends on n alone, so that the elliptical orbits represent degenerate states. Atomic spectroscopy shows in fact that energy levels with $n > 1$ consist of multiple states, as implied by the splitting of atomic lines by an electric field (Stark effect) or a magnetic field (Zeeman effect). Some of these generalized orbits are drawn schematically in Figure 9.9.2.

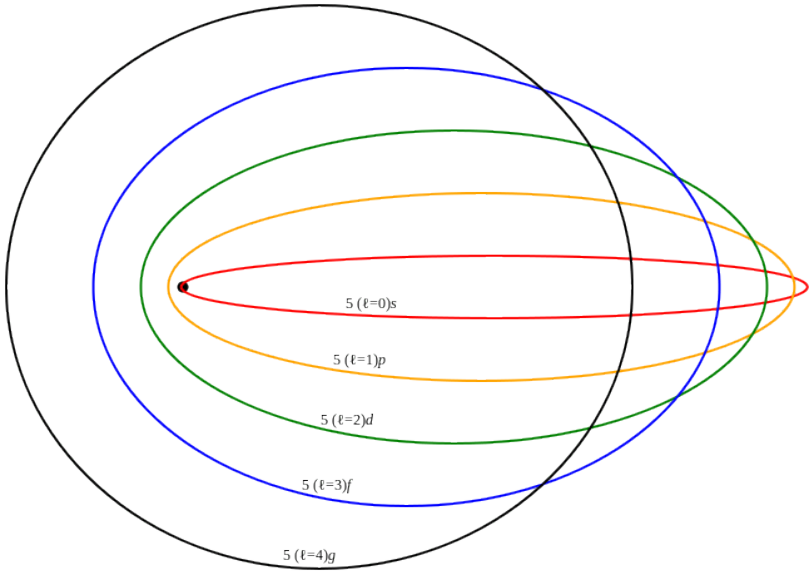


Figure 9.9.1: Elliptical orbits with the same energy and quantized angular momentum **in** Bohr-Sommerfield orbits for $n = 1,\ 2,\ 3$ (not to scale). (Public Domain; Pieter Kuiper)

The Bohr model was an important first step in the historical development of quantum mechanics. It introduced the quantization of atomic energy levels and gave quantitative agreement with the atomic hydrogen spectrum. With the Sommerfeld-Wilson generalization, it accounted as well for the degeneracy of hydrogen energy levels. Although the Bohr model was able to sidestep the atomic "Hindenberg disaster," it cannot avoid what we might call the "Heisenberg disaster." By this we mean that the assumption of well-defined electronic orbits around a nucleus is completely contrary to the basic premises of quantum mechanics. Another flaw in the Bohr picture is that the angular momenta are all too large by one unit, for example, the ground state actually has zero orbital angular momentum (rather than $\hbar$).

> The assumption of well-defined electronic orbits around a nucleus in the Bohr atom is completely contrary to the basic premises of quantum mechanics.

Quantum Mechanics of Hydrogenlike Atoms

In contrast to the particle in a box and the harmonic oscillator, the hydrogen atom is a real physical system that can be treated exactly by quantum mechanics. In addition to their inherent significance, these solutions suggest prototypes for atomic orbitals used in approximate treatments of complex atoms and molecules.

For an electron in the field of a nucleus of charge $+Ze$, the Schrödinger equation can be written

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{r}\right\}\psi(r) = E\psi(r) \tag{9.9.24}$$

It is convenient to introduce *atomic units* in which length is measured in bohrs:

$$a_0 = \frac{\hbar^2}{me^2} = 5.29 \times 10^{-11}\ \text{m} \equiv 1\ \text{bohr} \tag{9.9.25}$$

and energy in hartrees:

$$\frac{e^2}{a_0} = 4.358 \times 10^{-18}\ \text{J} = 27.211\ \text{eV} \equiv 1\ \text{hartree} \tag{9.9.26}$$

Electron volts (eV) are a convenient unit for atomic energies. One eV is defined as the energy an electron gains when accelerated across a potential difference of 1 volt. The ground state of the hydrogen atom has an energy of $-1/2$ hartree or -13.6 eV. Conversion to atomic units is equivalent to setting

$$\hbar = e = m = 1 \tag{9.9.27}$$

in all formulas containing these constants. Rewriting the Schrödinger equation in atomic units, we have

$$\left\{-\frac{1}{2}\nabla^2 - \frac{Z}{r}\right\}\psi(r) = E\psi(r) \tag{9.9.28}$$

Since the potential energy is spherically symmetrical (a function of r alone), it is obviously advantageous to treat this problem in spherical polar coordinates $r,\ \theta,\ \phi$. Expressing the Laplacian operator in these coordinates [cf. Eq (6-20)],

$$\begin{aligned} -\frac{1}{2}\left\{\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right\} \\ \times\psi(r,\ \theta,\ \phi) - \frac{Z}{r}\psi(r,\ \theta,\ \phi) = E\psi(r,\ \theta,\ \phi) \end{aligned} \tag{9.9.29}$$

Equation 9.9.29 shows that the second and third terms in the Laplacian represent the angular momentum operator $\hat{L}^2$. Clearly, Equation 9.9.29 will have separable solutions of the form

$$\psi(r,\ \theta,\ \phi) = R(r)Y_{\ell m}(\theta,\ \phi) \tag{9.9.30}$$

Substituting Equation 9.9.30 into Equation 9.9.29 and using the angular momentum eigenvalue Equation Equation ???, we obtain an ordinary differential equation for the radial function $R(r)$:

$$\left\{-\frac{1}{2r^2}\frac{d}{dr}r^2\frac{d}{dr} + \frac{\ell(\ell+1)}{2r^2} - \frac{Z}{r}\right\}R(r) = ER(r) \tag{9.9.31}$$

Note that in the domain of the variable r, the angular momentum contribution $\ell(\ell+1)/2r^2$ acts as an effective addition to the potential energy. It can be identified with *centrifugal force*, which pulls the electron outward, in opposition to the Coulomb attraction. Carrying out the successive differentiations in Equation 9.9.32 and simplifying, we obtain

$$\frac{1}{2}R''(r) + \frac{1}{r}R'(r) + \left[\frac{Z}{r} - \frac{\ell(\ell+1)}{2r^2} + E\right]R(r) = 0 \tag{9.9.32}$$

another second-order linear differential equation with non-constant coefficients. It is again useful to explore the asymptotic solutions to Equation 9.9.32, as $r \to \infty$. In the asymptotic approximation,

$$R''(r) - 2r|E|R(r) \approx 0 \tag{9.9.33}$$

having noted that the energy E is negative for bound states. Solutions to Equation 9.9.33 are

$$R(r) \approx \text{const}\, e^{\pm\sqrt{2|E|}r} \tag{9.9.34}$$

We reject the positive exponential on physical grounds, since $R(r) \to \infty$ as $r \to \infty$, in violation of the requirement that the wavefunction must be finite everywhere. Choosing the negative exponential and setting $E = -Z^2/2$ the ground state energy in the Bohr theory (in atomic units), we obtain

$$R(r) \approx \text{const}\, e^{-Zr} \tag{9.9.35}$$

It turns out, very fortunately, that this asymptotic approximation is also an *exact* solution of the Schrödinger equation (Equation \(\ref{29}\)) with $\ell = 0$, just what happened for the harmonic-oscillator problem in Chap. 5. The solutions to Equation 9.9.32, designated $R_{n\ell}(r)$, are labeled by n, known as the *principal quantum number*, as well as by the angular momentum ℓ, which is a parameter in the radial equation. The solution in Equation 9.9.35 corresponds to $R_{10}(r)$. This should be normalized according to the condition

$$\int_0^\infty [R_{10}(r)]^2\, r^2\, dr = 1 \tag{9.9.36}$$

A useful definite integral is

$$\int_0^\infty r^n\, e^{-\alpha r}\, dr = \frac{n!}{\alpha^{n+1}} \tag{9.9.37}$$

The normalized radial function is thereby given by

$$R_{10}(r) = 2Z^{3/2}e^{-Zr} \tag{9.9.38}$$

Since this function is nodeless, we identify it with the ground state of the hydrogenlike atom. Multipyling Equation 9.9.38 by the spherical harmonic $Y_{00} = 1/\sqrt{4\pi}$, we obtain the total wavefunction (Equation 9.9.30)

$$\psi_{100}(r,\theta,\phi) = \left(\frac{Z^3}{\pi}\right)^{1/2} e^{-Zr} \tag{9.9.39}$$

This is conventionally designated as the 1s function $\psi_{1s}(r)$.

Integrals in spherical-polar coordinates over a spherically-symmetrical integrand (like the 1s orbital) can be significantly simplified. We can do the reduction

$$\int_0^\infty \int_0^\pi \int_0^{2\pi} f(r)\, r^2\, \sin\theta\, dr\, d\theta\, d\phi = \int_0^\infty f(r)\, 4\pi r^2\, dr \tag{9.9.40}$$

since integration over θ and ϕ gives 4π, the total solid angle of a sphere. The normalization of the 1s wavefunction can thus be written as

$$\int_0^\infty [\psi_{1s}(r)]^2\, 4\pi r^2\, dr = 1 \tag{9.9.41}$$

Hydrogen Atom Ground State

There are a number of different ways of representing hydrogen-atom wavefunctions graphically. We will illustrate some of these for the 1s ground state. In atomic units,

$$\psi_{1s}(r) = \frac{1}{\sqrt{\pi}} e^{-r} \tag{9.9.42}$$

is a decreasing exponential function of a single variable r, and is simply plotted in Figure 3.

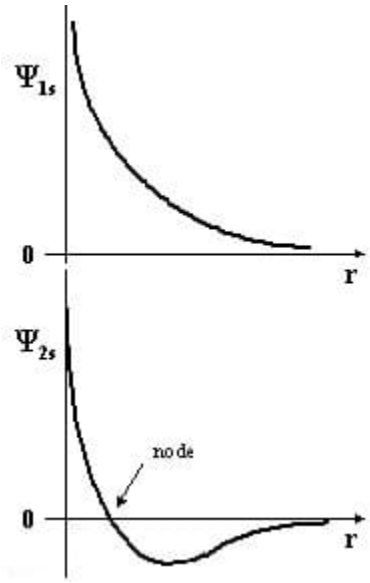


Figure 9.9.3. Wavefunctions for 1s and 2s orbitals for atomic hydrogen. The 2s-wavefunction (scaled by a factor of 2) has a node at $r = 2$ bohr. Taken from voh.chem.ucla.edu/vohtar/wint.../lecture3.html

Figure 9.9.3 gives a somewhat more pictorial representation, a three-dimensional contour plot of $\psi_{1s}(r)$ as a function of x and y in the x, y-plane.

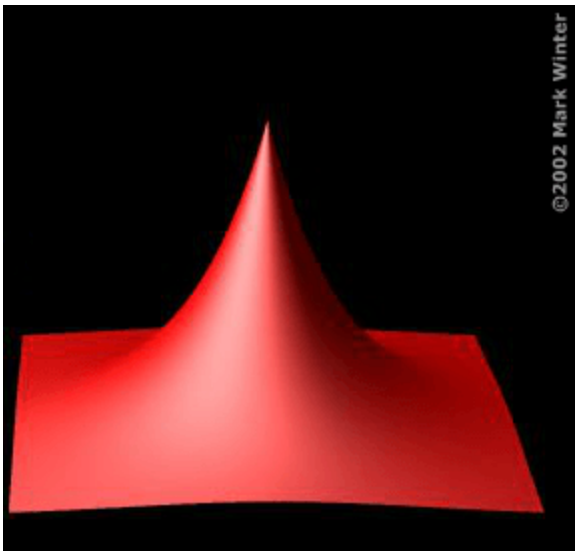


Figure 9.9.4: Contour map of 1s orbital in the x-, y-plane. Taken from winter.group.shef.ac.uk/orbit...s/wave-fn.html

According to Born's interpretation of the wavefunction, the probability per unit volume of finding the electron at the point $(r,\ \theta,\ \phi)$ is equal to the square of the normalized wavefunction

$$\rho_{1s}(r) = [\psi_{1s}(r)]^2 = \frac{1}{\pi} e^{-2r} \tag{9.9.43}$$

This is represented in Figure 5 by a scatter plot describing a possible sequence of observations of the electron position. Although results of individual measurements are not predictable, a statistical pattern does emerge after a sufficiently large number of measurements.

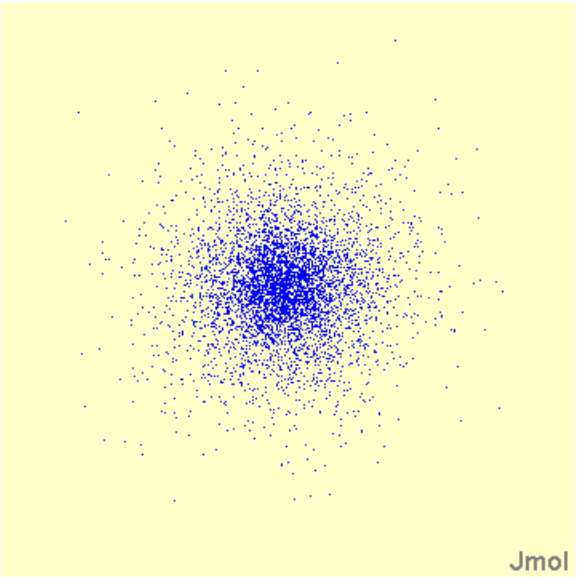


Figure 9.9.5: Scatter plot of electron position measurements in hydrogen 1s orbital. Taken from winter.group.shef.ac.uk/orbit...sity-dots.html

The probability density is normalized such that

$$\int_0^\infty \rho_{1s}(r)\, 4\pi r^2\, dr = 1 \tag{9.9.44}$$

In some ways $\rho(r)$ does *not* provide the best description of the electron distribution, since the region around $r = 0$, where the wavefunction has its largest values, is a relatively small fraction of the volume accessible to the electron. Larger radii r represent larger physical regions since, in spherical polar coordinates, a value of r is associated with a shell of volume $4\pi r^2\ dr$. A more significant measure is therefore the *radial distribution function*

$$D_{1s}(r) = 4\pi r^2[\psi_{1s}(r)]^2 \tag{9.9.45}$$

which represents the probability density within the entire shell of radius r, normalized such that

$$\int_0^\infty D_{1s}(r)\, dr = 1 \tag{9.9.46}$$

The functions $\rho_{1s}(r)$ and $D_{1s}(r)$ are both shown in Figure 9.9.6. Remarkably, the 1s RDF has its maximum at $r = a_0$, equal to the radius of the first Bohr orbit

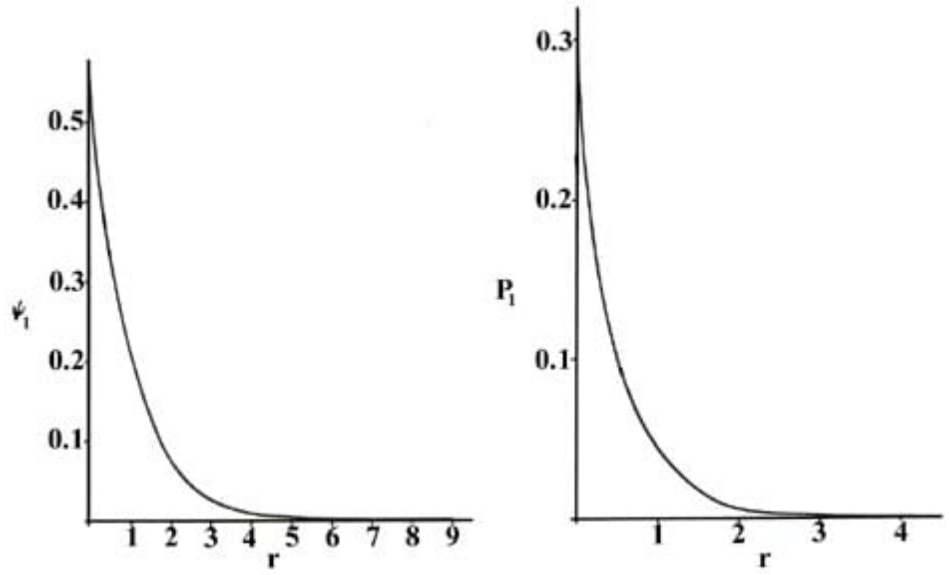


Figure 9.9.6: The wavefunction and probability distribution as functions of r for the n = 1 level of the H atom. The functions and the radius r are in atomic units in this and succeeding figures. Image used with permisison (Bader).

Atomic Orbitals

The general solution for $R_{n\ell}(r)$ has a rather complicated form which we give without proof:

$$R_{n\ell}(r) = N_{n\ell}\, \rho^\ell\, L_{n+\ell}^{2\ell+1}(\rho) e^{-\rho/2} \qquad \rho \equiv \frac{2Zr}{n} \tag{9.9.47}$$

Here L_β^α is an associated Laguerre polynomial and $N_{n\ell}$, a normalizing constant. The angular momentum quantum number ℓ is by convention designated by a code: *s* for $\ell = 0$, *p* for $\ell = 1$, *d* for $\ell = 2$, *f* for $\ell = 3$, *g* for $\ell = 4$, and so on. The first four letters come from an old classification scheme for atomic spectral lines: *sharp*, *principal*, *diffuse* and *fundamental*. Although these designations have long since outlived their original significance, they remain in general use. The solutions of the hydrogenic Schrödinger equation
in spherical polar coordinates can now be written in full

$$\psi_{n\ell m}(r,\, \theta,\, \phi) = R_{n\ell}(r) Y_{\ell m}(\theta,\, \phi) \tag{9.9.48}$$
$$n = 1,\, 2... \qquad \ell = 0,\, 1... n-1 \qquad m = 0,\, \pm 1,\, \pm 2... \pm \ell$$

where $Y_{\ell m}$ are the spherical harmonics tabulated in Chap. 6. Table 1 below enumerates all the hydrogenic functions we will actually need. These are called *hydrogenic atomic orbitals*, in anticipation of their later applications to the structure of atoms and molecules.

Table 1. Real hydrogenic functions in atomic units.

$$\psi_{1s} = \frac{1}{\sqrt{\pi}} e^{-r} \tag{9.9.49}$$

$$\psi_{2s} = \frac{1}{2\sqrt{2\pi}} \left(1 - \frac{r}{2}\right) e^{-r/2}$$	(9.9.50)
$$\psi_{2p_z} = \frac{1}{4\sqrt{2\pi}} z\, e^{-r/2}$$	(9.9.51)
ψ_{2p_x}, ψ_{2p_y} analogous	(9.9.52)
$$\psi_{3s} = \frac{1}{81\sqrt{3\pi}} (27 - 18r + 2r^2) e^{-r/3}$$	(9.9.53)
$$\psi_{3p_z} = \frac{\sqrt{2}}{81\sqrt{\pi}} (6 - r) z\, e^{-r/3}$$	(9.9.54)
ψ_{3p_x}, ψ_{3p_y} analogous	(9.9.55)
$$\psi_{3d_{z^2}} = \frac{1}{81\sqrt{6\pi}} (3z^2 - r^2) e^{-r/3}$$	(9.9.56)
$$\psi_{3d_{xz}} = \frac{\sqrt{2}}{81\sqrt{\pi}} zx\, e^{-r/3}$$	(9.9.57)
$\psi_{3d_{yz}}$, $\psi_{3d_{xy}}$ analogous	(9.9.58)
$$\psi_{3d_{x^2-y^2}} = \frac{1}{81\sqrt{\pi}} (x^2 - y^2) e^{-r/3}$$	(9.9.59)

The energy levels for a hydrogenic system are given by

$$E_n = -\frac{Z^2}{2n^2} \text{ hartrees} \tag{9.9.60}$$

and depends on the principal quantum number alone. Considering all the allowed values of ℓ and m, the level E_n has a degeneracy of n^2. Figure 7 shows an energy level diagram for hydrogen $(Z = 1)$. For $E \geq 0$, the energy is a continuum, since the electron is in fact a free particle. The continuum represents states of an electron and proton in interaction, but not bound into a stable atom. Figure $9.9.7$ also shows some of the transitions which make up the Lyman series in the ultraviolet and the Balmer series in the visible region.

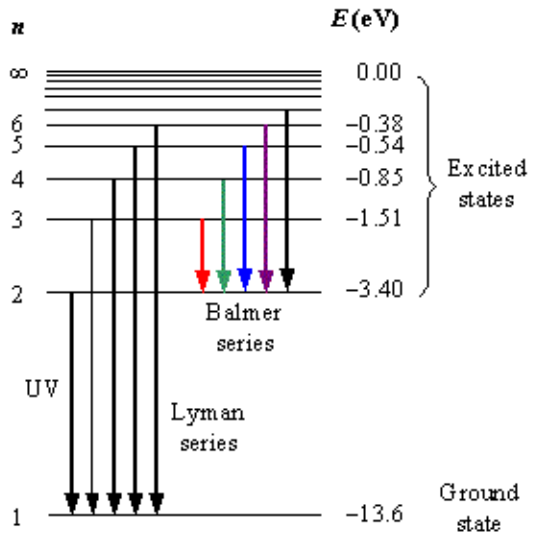


Figure 9.9.7: Energy levels of atomic hydrogen. Taken from http://eilat.sci.brooklyn.cuny.edu/c...hws/hw2d_c.htm

The ns orbitals are all spherically symmetrical, being associated with a constant angular factor, the spherical harmonic $Y_{00} = 1/\sqrt{4\pi}$. They have $n-1$ radial nodes—spherical shells on which the wavefunction equals zero. The 1s ground state is nodeless and the number of nodes increases with energy, in a pattern now familiar from our study of the particle-in-a-box and harmonic oscillator. The 2*s* orbital, with its radial node at $r = 2$ bohr, is also shown in Figure 9.9.3.

p- and d-Orbitals

The lowest-energy solutions deviating from spherical symmetry are the 2*p*-orbitals. Using Equations 9.9.47, 9.9.48 and the $\ell = 1$ spherical harmonics, we find three degenerate eigenfunctions:

$$\psi_{210}(r,\ \theta,\ \phi) = \frac{1}{4\sqrt{2\pi}} re^{-r/2}\cos\theta \tag{9.9.61}$$

and

$$\psi_{21\pm1}(r,\ \theta,\ \phi) = \mp\frac{1}{4\sqrt{2\pi}} re^{-r/2}\sin\theta e^{\pm i\phi} \tag{9.9.62}$$

The function ψ_{210} is real and contains the factor $r\cos\theta$, which is equal to the cartesian variable z. In chemical applications, this is designated as a $2p_z$ orbital:

$$\psi_{2p_z} = \frac{1}{4\sqrt{2\pi}} ze^{-r/2} \tag{9.9.63}$$

A contour plot is shown in Figure 9.9.8. Note that this function is cylindrically-symmetrical about the z-axis with a node in the x, y-plane. The $\psi_{21\pm1}$ are complex functions and not as easy to represent graphically. Their angular dependence is that of the spherical harmonics $Y_{1\pm1}$, shown in Figure 6-4. As noted in Chap. 4, any linear combination of degenerate eigenfunctions is an equally-valid alternative eigenfunction. Making use of the Euler formulas for sine and cosine

$$\cos\phi = \frac{e^{i\phi} + e^{-i\phi}}{2} \qquad \text{and} \qquad \sin\phi = \frac{e^{i\phi} - e^{-i\phi}}{2} \tag{9.9.64}$$

and noting that the combinations $\sin\theta\cos\phi$ and $\sin\theta\sin\phi$ correspond to the cartesian variables x and y, respectively, we can define the alternative 2*p* orbitals

$$\psi_{2p_x} = \frac{1}{\sqrt{2}}(\psi_{21-1} - \psi_{211}) = \frac{1}{4\sqrt{2\pi}} xe^{-r/2} \tag{9.9.65}$$

and

$$\psi_{2p_y} = -\frac{i}{\sqrt{2}}(\psi_{21-1} + \psi_{211}) = \frac{1}{4\sqrt{2\pi}} y e^{-r/2} \tag{9.9.66}$$

Clearly, these have the same shape as the 2p_z-orbital, but are oriented along the x- and y-axes, respectively. The threefold degeneracy of the *p*-orbitals is very clearly shown by the geometric equivalence the functions 2p_x, 2p_y and 2p_z, which is not obvious for the spherical harmonics. The functions listed in Table 1 are, in fact, the real forms for all atomic orbitals, which are more useful in chemical applications. All higher *p*-orbitals have analogous functional forms $x\ f(r)$, $y\ f(r)$ and $z\ f(r)$ and are likewise 3-fold degenerate.

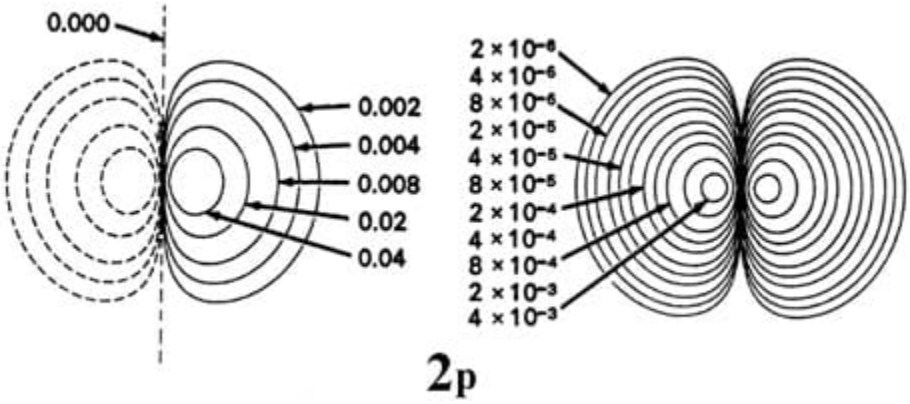


Figure 9.9.8: Contour plot of 2p_z orbital. (Bader).

The orbital ψ_{320} is, like ψ_{210}, a real function. It is known in chemistry as the d_{z^2}-orbital and can be expressed as a cartesian factor times a function of r:

$$\psi_{3d_{z^2}} = \psi_{320} = (3z^2 - r^2) f(r) \tag{9.9.67}$$

A contour plot is shown in Figure 9.9.9. This function is also cylindrically symmetric about the z-axis with *two* angular nodes—the conical surfaces with $3z^2 - r^2 = 0$. The remaining four 3*d* orbitals are complex functions containing the spherical harmonics $Y_{2\pm1}$ and $Y_{2\pm2}$ pictured in Figure 6-4. We can again construct real functions from linear combinations, the result being four geometrically equivalent "four-leaf clover" functions with two perpendicular planar nodes. These orbitals are designated $d_{x^2-y^2}$, d_{xy}, d_{zx} and d_{yz}. Two of them are shown in Figure 9. The d_{z^2} orbital has a different shape. However, it can be expressed in terms of two non-standard *d*-orbitals, $d_{z^2-x^2}$ and $d_{y^2-z^2}$. The latter functions, along with $d_{x^2-y^2}$ add to zero and thus constitute a linearly *dependent* set. *Two* combinations of these three functions can be chosen as independent eigenfunctions.

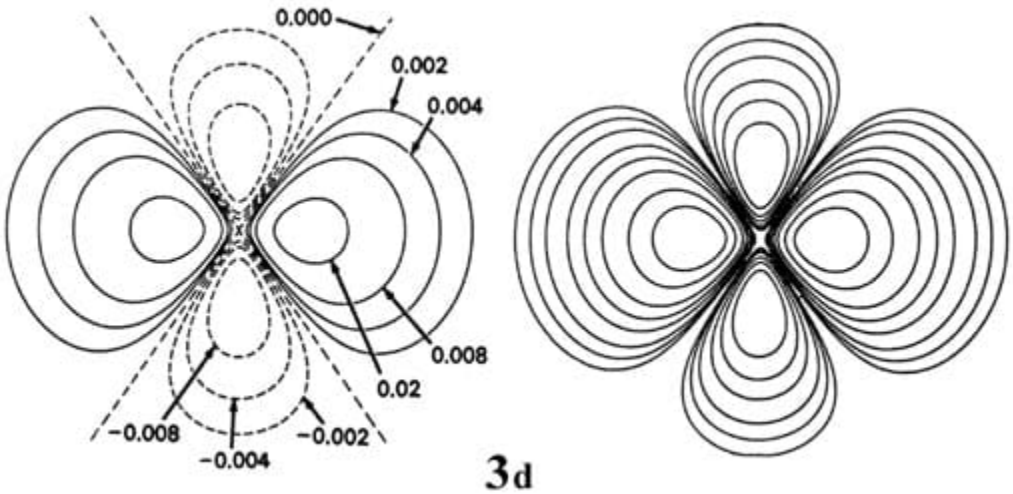


Figure 9.9.8: Contour plot of $3d_{xy}$ orbital.

Summary

The atomic orbitals listed in Table 1 are illustrated in Figure $9.9.20$. Blue and red indicate, respectively, positive and negative regions of the wavefunctions (the radial nodes of the 2s and 3s orbitals are obscured). These pictures are intended as stylized representations of atomic orbitals and should *not* be interpreted as quantitatively accurate.

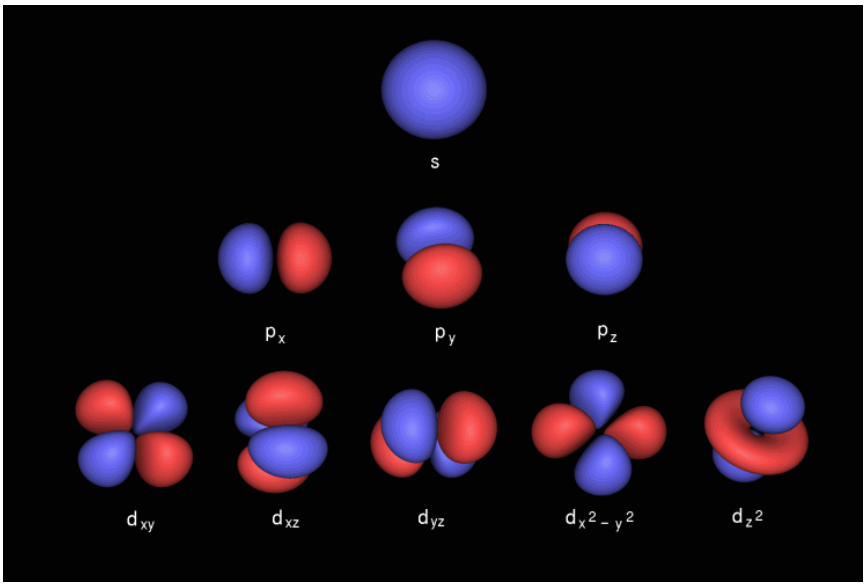


Figure $9.9.20$: Hydrogenic atomic orbitals. Taken from www.chemcomp.com/journal/molorbs.htm

The electron charge distribution in an orbital $\psi_{n\ell m}(\mathbf{r})$ is given by

$$\rho(\mathbf{r}) = |\psi_{n\ell m}(\mathbf{r})|^2 \tag{9.9.68}$$

which for the s-orbitals is a function of r alone. The radial distribution function can be defined, even for orbitals containing angular dependence, by

$$D_{n\ell}(r) = r^2[R_{n\ell}(r)]^2 \tag{9.9.69}$$

This represents the electron density in a shell of radius r, including all values of the angular variables θ, ϕ. Figure $9.9.11$ shows plots of the RDF for the first few hydrogen orbitals.

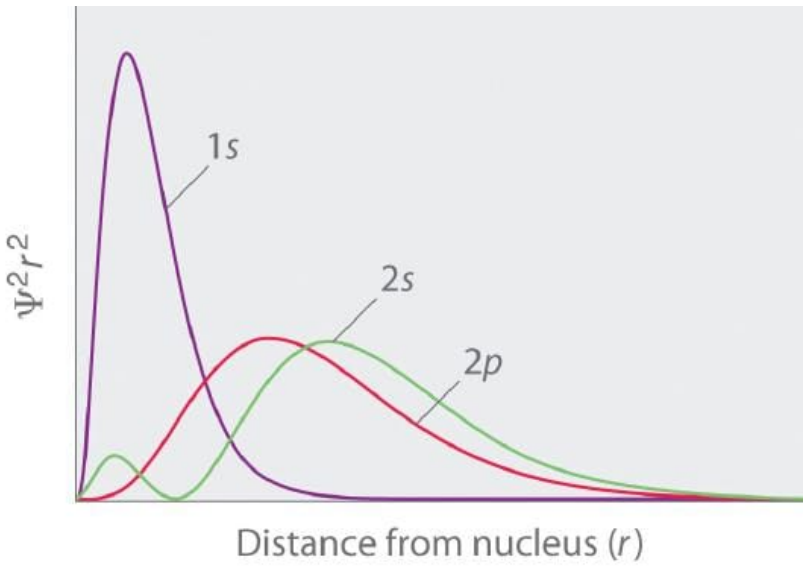


Figure $9.9.22$: Some radial distribution functions.

Contributors and Attributions

Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)

- Integrated by Daniel SantaLucia (Chemistry student at Hope College, Holland MI)

This page titled 9.9: Hydrogen Atom is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Chapter 7. Hydrogen Atom is licensed CC BY-NC-SA 4.0.

9.10: Helium Atom

The second element in the periodic table provides our first example of a quantum-mechanical problem which *cannot* be solved exactly. Nevertheless, as we will show, approximation methods applied to helium can give accurate solutions in perfect agreement with experimental results. In this sense, it can be concluded that quantum mechanics is correct for atoms more complicated than hydrogen. By contrast, the Bohr theory failed miserably in attempts to apply it beyond the hydrogen atom.

The helium atom has two electrons bound to a nucleus with charge Z = 2. The successive removal of the two electrons can be diagrammed as

$$\mathrm{He} \xrightarrow{I_1} \mathrm{He}^+ + e^- \xrightarrow{I_2} \mathrm{He}^{++} + 2e^- \tag{9.10.1}$$

The *first ionization energy* I_1, the minimum energy required to remove the first electron from helium, is experimentally 24.59 eV. The second ionization energy, I_2, is 54.42 eV. The last result can be calculated exactly since He^+ is a hydrogen-like ion. We have

$$I_2 = -E_{1s}(\mathrm{He}^+) = \frac{Z^2}{2n^2} = 2 \text{ hartrees} = 54.42 \text{ eV} \tag{9.10.2}$$

The energy of the three separated particles on the right side of Equation 9.10.1 is, by definition, zero. Therefore the ground-state energy of helium atom is given by $E_0 = -(I_1 + I_2) = -79.02 \text{ eV} = -2.90372 \text{ hartrees}$. We will attempt to reproduce this value, as close as possible, by theoretical analysis.

Schrödinger Equation and Variational Calculations

The Schrödinger equation for He atom, again using atomic units and assuming infinite nuclear mass, can be written

$$\left\{ -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{Z}{r_1} - \frac{Z}{r_2} + \frac{1}{r_{12}} \right\} \psi(\mathrm{r}_1, \mathrm{r}_2) = E\psi(\mathrm{r}_1, \mathrm{r}_2) \tag{9.10.3}$$

The five terms in the Hamiltonian represent, respectively, the kinetic energies of electrons 1 and 2, the nuclear attractions of electrons 1 and 2, and the repulsive interaction between the two electrons. It is this last contribution which prevents an exact solution of the Schrödinger equation and which accounts for much of the complication in the theory. In seeking an approximation to the ground state, we might first work out the solution in the absence of the $1/r_{12}$-term. In the Schrödinger equation thus simplified, we can separate the variables **r_1** and **r_2** to reduce the equation to two independent hydrogen-like problems. The ground state wavefunction (not normalized) for this hypothetical helium atom would be

$$\psi(\mathrm{r}_1, \mathrm{r}_2) = \psi_{1s}(r_1)\psi_{1s}(r_2) = e^{-Z(r_1+r_2)} \tag{9.10.4}$$

and the energy would equal $2 \times (-Z^2/2) = -4$ hartrees, compared to the experimental value of -2.90 hartrees. **Neglect of electron repulsion evidently introduces a very large error.**

A significantly improved result can be obtained with the functional form (Equation 9.10.4), but with Z replaced by a adjustable parameter α, thus:

$$\tilde{\psi}(r_1, r_2) = e^{-\alpha(r_1+r_2)} \tag{9.10.5}$$

Using this function in the variational principle [cf. Eq (4.53)], we have

$$\tilde{E} = \frac{\int \psi(r_1, r_2)\hat{H}\psi(r_1, r_2)d\tau_1\tau_2}{\int \psi(r_1, r_2)\psi(r_1, r_2)d\tau_1 d\tau_2} \tag{9.10.6}$$

where $\hat{H}$ is the full Hamiltonian as in Equation 9.10.3, including the $1/r_{12}$-term. The expectation values of the five parts of the Hamiltonian work out to

$$\left\langle -\frac{1}{2}\nabla_1^2 \right\rangle = \left\langle -\frac{1}{2}\nabla_2^2 \right\rangle = \frac{\alpha^2}{2} \tag{9.10.7}$$

$$\left\langle -\frac{Z}{r_1} \right\rangle = \left\langle -\frac{Z}{r_2} \right\rangle = -Z\alpha, \left\langle \frac{1}{r_{12}} \right\rangle = \frac{5}{8}\alpha \tag{9.10.8}$$

The sum of the integrals in Equation 9.10.8 gives the variational energy

$$\tilde{E}(\alpha) = \alpha^2 - 2Z\alpha + \frac{5}{8}\alpha \tag{9.10.9}$$

This will be always be an upper bound for the true ground-state energy. We can optimize our result by finding the value of α which *minimizes* the energy (Equation 9.10.9). We find

$$\frac{d\tilde{E}}{d\alpha} = 2\alpha - 2Z + \frac{5}{8} = 0 \tag{9.10.10}$$

giving the optimal value

$$\alpha = Z - \frac{5}{16} \tag{9.10.11}$$

This can be given a physical interpretation, noting that the parameter α in the wavefunction (Equation 9.10.5) represents an *effective* nuclear charge. Each electron partially shields the other electron from the positively-charged nucleus by an amount equivalent to $5/8$ of an electron charge. Substituting Equation 9.10.11 into Equation 9.10.9, we obtain the optimized approximation to the energy

$$\tilde{E} = -\left(Z - \frac{5}{16}\right)^2 \tag{9.10.12}$$

For helium ($Z = 2$), this gives -2.84765 hartrees, an error of about 2% ($E_0 = -2.90372$). Note that the inequality $\tilde{E} > E_0$ applies in an *algebraic* sense.

In the late 1920's, it was considered important to determine whether the helium computation could be improved, as a test of the validity of quantum mechanics for many electron systems. The table below gives the results for a selection of variational computations on helium.

wavefunction	parameters	energy
$e^{-Z(r_1+r_2)}$	$Z = 2$	-2.75
$e^{-\alpha(r_1+r_2)}$	$\alpha = 1.6875$	-2.84765
$\psi(r_1)\psi(r_2)$	best $\psi(r)$	-2.86168
$e^{-\alpha(r_1+r_2)}(1+cr_{12})$	best α, c	-2.89112
Hylleraas (1929)	10 parameters	-2.90363
Pekeris (1959)	1078 parameters	-2.90372

The third entry refers to the *self-consistent field* method, developed by Hartree. Even for the best possible choice of one-electron functions $\psi(r)$, there remains a considerable error. This is due to failure to include the variable r_{12} in the wavefunction. The effect is known as *electron correlation*.

The fourth entry, containing a simple correction for correlation, gives a considerable improvement. Hylleraas (1929) extended this approach with a variational function of the form

$$\psi(r_1, r_2, r_{12}) = e^{-\alpha(r_1+r_2)} \times \text{polynomial in} r_1, r_2, r_{12} \tag{9.10.13}$$

and obtained the nearly exact result with 10 optimized parameters. More recently, using modern computers, results in essentially perfect agreement with experiment have been obtained.

Spinorbitals and the Exclusion Principle

The simpler wavefunctions for helium atom in Equation 9.10.5, can be interpreted as representing two electrons in hydrogen-like

1s orbitals, designated as a $1s^2$ configuration. According to Pauli's exclusion principle, which states that no two electrons in an atom can have the same set of four quantum numbers, the two 1s electrons must have *different* spins, one spin-up or α, the other spin-down or β. A product of an orbital with a spin function is called a *spinorbital*. For example, electron 1 might occupy a spinorbital which we designate

$$\phi(1) = \psi_{1s}(1)\alpha(1) \text{or} \psi_{1s}(1)\beta(1) \tag{9.10.14}$$

Spinorbitals can be designated by a single subscript, for example, ϕ_a or ϕ_b, where the subscript stands for a *set* of four quantum numbers. In a two electron system the occupied spinorbitals ϕ_a and ϕ_b must be different, meaning that at least one of their four quantum numbers must be unequal. A two-electron spinorbital function of the form

$$\Psi(1,2) = \frac{1}{2}\Big(\phi_a(1)\phi_b(2) - \phi_b(1)\phi_a(2)\Big) \tag{9.10.15}$$

automatically fulflls the Pauli principle since it vanishes if $a = b$. Moreover, this function associates each electron equally with each orbital, which is consistent with the *indistinguishability* of identical particles in quantum mechanics. The factor $1/\sqrt{2}$ normalizes the two-particle wavefunction, assuming that ϕ_a and ϕ_b are normalized and mutually orthogonal. The function (Equation 9.10.15) is *antisymmetric* with respect to interchange of electron labels, meaning that

$$\Psi(2,1) = -\Psi(1,2) \tag{9.10.16}$$

This antisymmetry property is an elegant way of expressing the Pauli principle.

We note, for future reference, that the function in Equation 9.10.15 can be expressed as a 2×2 determinant:

$$\Psi(1,2) = \frac{1}{\sqrt{2}}\begin{vmatrix}\phi_a(1) & \phi_b(1)\\ \phi_a(2) & \phi_b(2)\end{vmatrix} \tag{9.10.17}$$

For the $1s^2$ configuration of helium, the two orbital functions are the same and Equation 9.10.15 can be written

$$\Psi(1,2) = \psi_{1s}(1)\psi_{1s}(2) \times \frac{1}{\sqrt{2}}\Big(\alpha(1)\beta(2) - \beta(1)\alpha(2)\Big) \tag{9.10.18}$$

For two-electron systems (but *not* for three or more electrons), the wavefunction can be factored into an orbital function times a spin function. The two-electron spin function

$$\sigma_{0,0}(1,2) = \frac{1}{\sqrt{2}}\Big(\alpha(1)\beta(2) - \beta(1)\alpha(2)\Big) \tag{9.10.19}$$

represents the two electron spins in opposing directions (antiparallel) with a total spin angular momentum of zero. The two subscripts are the quantum numbers S and M_S for the total electron spin. Eqution 9.10.18 is called the *singlet* spin state since there is only a single orientation for a total spin quantum number of zero. It is also possible to have both spins in the *same* state, provided the orbitals are different. There are three possible states for two parallel spins:

$$\sigma_{1,1}(1,2) = \alpha(1)\alpha(2) \tag{9.10.20}$$

$$\sigma_{1,0}(1,2) = \frac{1}{\sqrt{2}}\Big(\alpha(1)\beta(2) + \beta(2)\alpha(2)\Big) \tag{9.10.21}$$

$$\sigma_{1,-1}(1,2) = \beta(1)\beta(2) \tag{9.10.22}$$

These make up the *triplet* spin states, which have the three possible orientations of a total angular momentum of 1.

Excited States of Helium

The lowest excitated state of helium is represented by the electron configuration 1s 2s. The 1s 2*p* configuration has higher energy, even though the 2*s* and 2*p* orbitals in hydrogen are degenerate, because the 2*s* penetrates closer to the nucleus, where the potential energy is more negative. When electrons are in different orbitals, their spins can be either parallel or antiparallel. In order that the wavefunction satisfy the antisymmetry requirement (Equation 9.10.16), the two-electron orbital and spin functions must have *opposite* behavior under exchange of electron labels. There are four possible states from the 1s 2*s* configuration: a

singlet state

$$\Psi^{+}(1,2)=\frac{1}{\sqrt{2}}\Big(\psi_{1s}(1)\psi_{2s}(2)+\psi_{2s}(1)\psi_{1s}(2)\Big)\sigma_{0,0}(1,2) \tag{9.10.23}$$

and three triplet states

$$\Psi^{-}(1,2)=\frac{1}{\sqrt{2}}\Big(\psi_{1s}(1)\psi_{2s}(2)-\psi_{2s}(1)\psi_{1s}(2)\Big)\begin{cases}\sigma_{1,1}(1,2)\\ \sigma_{1,0}(1,2)\\ \sigma_{1,-1}(1,2)\end{cases} \tag{9.10.24}$$

Using the Hamiltonian in Equation 9.10.3, we can compute the approximate energies

$$E^{\pm}=\iint \Psi^{\pm}(1,2)\hat{H}\Psi^{\pm}(1,2)d\tau_1 d\tau_2 \tag{9.10.25}$$

After evaluating some fierce-looking integrals, this reduces to the form

$$E^{\pm}=I(1s)+I(2s)+J(1s,2s)\pm K(1s,2s) \tag{9.10.26}$$

in terms of the one electron integrals

$$I(a)=\int \psi_a(\mathrm{r})\left\{-\frac{1}{2}\nabla^2-\frac{Z}{r}\right\}\psi_a(\mathrm{r})d\tau \tag{9.10.27}$$

the Coulomb integrals

$$J(a,b)=\iint \psi_a(\mathrm{r}_1)^2\frac{1}{r_{12}}\psi_b(\mathrm{r}_2)^2 d\tau_1 d\tau_2 \tag{9.10.28}$$

and the exchange integrals

$$K(a,b)=\iint \psi_a(\mathrm{r}_1)\psi_b(\mathrm{r}_1)\frac{1}{r_{12}}\psi_a(\mathrm{r}_2)\psi_b(\mathrm{r}_2) d\tau_1 d\tau_2 \tag{9.10.29}$$

The Coulomb integral represents the repulsive potential energy for two interacting charge distributions $\psi_a(\mathbf{r}_1)^2$ and $\psi_b(\mathbf{r}_2)^2$. The exchange integral, which has no classical analog, arises because of the exchange symmetry (or antisymmetry) requirement of the wavefunction. Both *J* and *K* can be shown to be positive quantities. Therefore the lower sign in (22) represents the state of lower energy, making the triplet state of the configuration 1s 2s lower in energy than the singlet state. This is an almost universal generalization and contributes to Hund's rule, to be discussed in the next Chapter.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)
- Gerardo Ochoa (Hope College, Holland, MI)

This page titled 9.10: Helium Atom is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- 8: The Helium Atom is licensed CC BY-NC-SA 4.0.

9.11: Atomic Structure and The Periodic Law

Quantum mechanics can account for the periodic structure of the elements, by any measure a major conceptual accomplishment for any theory. Although accurate computations become increasingly more challenging as the number of electrons increases, the general patterns of atomic behavior can be predicted with remarkable accuracy.

Slater Determinants

According to the orbital approximation, which was introduced in the last Chapter, an *N*-electron atom contains *N* occupied spinorbitals, which can be designated ϕ_a, ϕ_b . . . ϕ_n. In accordance with the Pauli exclusion principle, no two of these spinorbitals can be identical. Also, every electron should be equally associated with every spinorbital. A very neat mathematical representation for these properties is a generalization of the two-electron wavefunction (8.13) or (8.15) called a *Slater determinant*

$$\Psi(1,2\ldots N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_a(1) & \phi_b(1) & \ldots & \phi_n(1) \\ \phi_a(2) & \phi_b(2) & \ldots & \phi_n(2) \\ & & \vdots & \\ \phi_a(N) & \phi_b(N) & \ldots & \phi_n(N) \end{vmatrix} \tag{9.11.1}$$

Since interchanging any two rows (or columns) of a determinant multiplies it by -1, the antisymmetry property (8.15) is fulfilled, for *every* pair of electrons.

The Hamiltonian for an atom with *N* electrons around a nucleus of charge *Z* can be written

$$\hat{H} = \sum_{i=1}^{N} \left\{ -\frac{1}{2} \bigtriangledown_i^2 - \frac{Z}{r_i} \right\} + \sum_{i<j}^{N} \frac{1}{r_{ij}} \tag{9.11.2}$$

The sum over electron repulsions is written so that each pair {*i*,*j*} is counted just once. The energy of the state represented by a Slater determinant (Equation 9.11.1) can be obtained after a lengthy derivation. We give just the final result

$$\tilde{E} = \sum_a I_a + \frac{1}{2} \sum_{a,b} (J_{ab} - K_{ab}) \tag{9.11.3}$$

where the sums run over all occupied spinorbitals. The one-electron, Coulomb and exchange integrals have the same form as those defined for helium atom in Eqs (8.22-24). The only difference is that an exchange integral equals zero unless the spins of orbitals *a* and *b* are both α or both β. The factor $^1/_2$ corrects for the double counting of pairs of spinorbitals in the second sum. The contributions with *a* = *b* can be omitted since $J_{aa} = K_{aa}$. This effectively removes the Coulomb interaction of an orbital with itself, which is spurious.

The Hartree-Fock or self-consistent field (SCF) method is a procedure for optimizing the orbital functions in the Slater determinant (1), so as to minimize the energy (Equation 9.11.3). SCF computations have been carried out for all the atoms of the periodic table, with predictions of total energies and ionization energies generally accurate in the $1-2\%$ range.

Aufbau Principles and Periodic Structure

Aufbau means "building-up." Aufbau principles determine the order in which atomic orbitals are filled as the atomic number is increased. For the hydrogen atom, the order of increasing orbital energy is given by 1s < 2*s* = 2*p* < 3s = 3*p* = 3*d*, etc. The dependence of energy on *n* alone leads to extensive degeneracy, which is however removed for orbitals in many-electron atoms. Thus 2*s* lies below 2*p*, as already observed in helium. Similarly, 3s, 3*p* and 3*d* increase energy in that order, and so on. The 4*s* is lowered sufficiently that it becomes comparable to 3*d*. The general ordering of atomic orbitals is summarized in the following scheme:

$$\begin{aligned} &1s < 2s < 2p < 3s < 3p < 4s \sim 3d < 4p < 5s \sim 4d \\ &< 5p < 6s \sim 5d \sim 4f < 6p < 7s \sim 6d \sim 5f \end{aligned} \tag{9.11.4}$$

and illustrated in Figure 1. This provides enough orbitals to fill the ground states of all the atoms in the periodic table. For orbitals designated as comparable in energy, e.g., 4s ~ 3*d*, the actual order depends which other orbitals are occupied. The sequence of orbitals pictured above increases in the order $n + \frac{1}{2}l$, except that l = 4 (rather than 3) is used for an *f*-orbital.

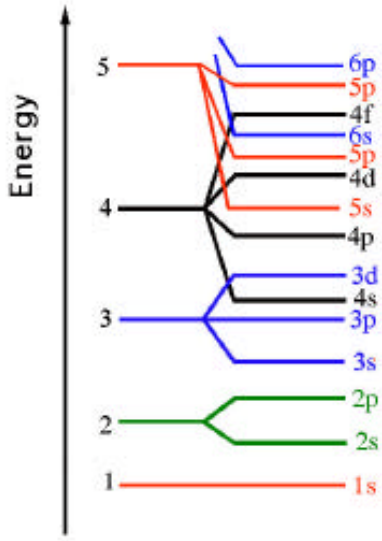


Figure 1: Approximate order of atomic orbital energy levels.

The tabulation below shows the ground-state electron configuration and term symbol for selected elements in the first part of the periodic table. From the term symbol, one can read off the total orbital angular momentum *L* and the total spin angular momentum S. The code for the total orbital angular momentum mirrors the one-electron notation, but using upper-case letters, as follows:

L = 0 1 2 3 4
S P D F G

The total spin S is designated, somewhat indirectly, by the spin multiplicity 2S + 1 written as a superscript *before* the S, P, D. . . symbol. For example ^{1}S (singlet S) ,^{1}P (singlet P). . . mean S = 0; ^{2}S (doublet S) ,^{2}P (doublet P). . . mean S = 1/2; ^{3}S (triplet S) ,^{3}P (triplet P). . . mean S = 1, and so on. Please do not confuse the spin quantum number S with the orbital designation S.

Atom	Z	Electron Configuration	Term Symbol
H	1	1s	$^2S_{1/2}$
He	2	$1s^2$	$^1S_0$
Li	3	[He]2s	$^2S_{1/2}$
Be	4	$[He]2s^2$	$^1S_0$
B	5	$[He]2s^2 2p$	$^2P_{1/2}$
C	6	$[He]2s^2 2p^2$	$^3P_0$
N	7	$[He]2s^2 2p^3$	$^4S_{3/2}$
O	8	$[He]2s^2 2p^4$	$^3P_2$
F	9	$[He]2s^2 2p^5$	$^2P_{3/2}$
Ne	10	$[He]2s^2 2p^6$	$^1S_0$

Atom	Z	Electron Configuration	Term Symbol
Na	11	[Ne]3s	$^2S_{1/2}$
Cl	17	[Ne]$3s^2 3p^5$	$^2P_{3/2}$
Ar	18	[Ne]$3s^2 3p^6$	$^1S_0$
K	19	[Ar]4s	$^2S_{1/2}$
Ca	20	[Ar]$4s^2$	$^1S_0$
Sc	21	[Ar]$4s^2 3d$	$^2D_{3/2}$
Ti	22	[Ar]$4s^2 3d^2$	$^3F_2$
V	23	[Ar]$4s^2 3d^3$	$^4F_{3/2}$
Cr	24	[Ar]$4s 3d^5$	$^7S_3$
Mn	25	[Ar]$4s^2 3d^5$	$^6S_{5/2}$
Fe	26	[Ar]$4s^2 3d^6$	$^5D_4$
Co	27	[Ar]$4s^2 3d^7$	$^4F_{9/2}$
Ni	28	[Ar]$4s^2 3d^8$	$^3F_4$
Cu	29	[Ar]$4s 3d^{10}$	$^2S_{1/2}$
Zn	30	[Ar]$4s^2 3d^{10}$	$^1S_0$
Ga	31	[Ar]$4s^2 3d^{10} 4p$	$^2P_{1/2}$
Br	35	[Ar]$4s^2 3d^{10} 4p^5$	$^2P_{3/2}$
Kr	36	[Ar]$3d^{10} 4s^2 4p^6$	$^1S_0$

The vector sum of the orbital and spin angular momentum is designated

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \tag{9.11.5}$$

The possible values of the total angular momentum quantum number *J* runs in integer steps between |*L* - S| and *L* + S. The *J* value is appended as a subscript on the term symbol, eg, $^1S_0$, $^2P_{1/2}$, $^2P_{3/2}$. The energy differences between *J* states is a result of *spin-orbit interaction*, a magnetic interaction between the circulating charges associated with orbital and spin angular momenta. For atoms of low atomic number, the spin-orbit coupling is a relatively small correction to the energy, but it can become increasingly significant for heavier atoms.

We will next consider in some detail the Aufbau of ground electronic states starting at the beginning of the periodic table. Hydrogen has one electron in an s-orbital so its total orbital angular momentum is also designated S. The single electron has s = 1/2, thus S = 1/2. The spin multiplicity 2S + 1 equals 2, thus the term symbol is written ^2S. In helium, a second electron can occupy the 1s shell, provided it has the opposite spin. The total spin angular momentum is therefore zero, as is the total orbital angular momentum. The term symbol is ^1S, as it will be for all other atoms with complete electron shells. In determining the total spin and orbital angular moments, we need consider only electrons outside of closed shells. Therefore lithium and beryllium are a reprise of hydrogen and helium. The angular momentum of boron comes from the single 2*p* electron, with *l* = 1 and s = 1/2, giving a ^2P state.

To build the carbon atom, we add a second 2*p* electron. Since there are three degenerate 2*p* orbitals, the second electron can go into either the already-occupied 2*p* orbital or one of the unoccupied 2*p* orbitals. Clearly, two electrons in different 2*p* orbitals will have less repulsive energy than two electrons crowded into the same 2*p* orbital. In terms of the Coulomb integrals, we would

expect, for example

$$J(2px, 2py) < J(2px, 2px) \tag{9.11.6}$$

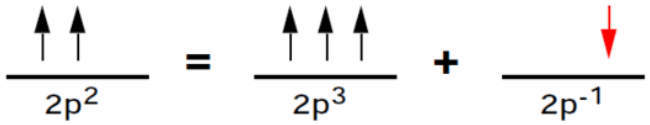
For nitrogen atom, with three 2*p* electrons, we expect, by the same line of reasoning, that the third electron will go into the remaining unoccupied 2*p* orbital. The half-filled $2p^3$ subshell has an interesting property. If the three occupied orbitals are $2p_x$, $2p_y$ and $2p_z$, then their total electron density is given by

$$\rho_{2p} = \psi^2_{2p_x} + \psi^2_{2p_y} + \psi^2_{2p_z} = \left(x^2 + y^2 + z^2\right) \times \text{function of r} = \text{function of r} \tag{9.11.7}$$

noting that $x^2 + y^2 + z^2 = r^2$. But spherical symmetry implies zero angular momentum, like an s-orbital. In fact, any half filled subshell, such as p^3, d^5, f^7, will contribute zero angular momentum. The same is, of course true as well for *filled* subshells, such as p^6, d^{10}, f^{14}. These are all S terms.

Another way to understand this vector cancelation of angular momentum is to consider the alternative representation of the degenerate 2*p*-orbitals: $2p_{-1}$; $2p_0$ and $2p_1$. Obviously, the *z*-components of angular momentum now add to zero, and since only this one component is observable, the total angular momentum must also be zero.

Returning to our unfinished consideration of carbon, the $2p^2$ subshell can be regarded, in concept, as a half-filled $2p^3$ subshell plus an electron "hole." The advantage of this picture is that the total orbital angular momentum must be equal to that of the hole, namely $l = 1$. This is shown below:



Thus the term symbol for the carbon ground state is P. It remains to determine the total spins of these subshells. Recall that exchange integrals K_{ab} are non-zero only if the orbitals *a* and *b* have the same spin. Since exchange integrals enter the energy formula (3) with negative signs, the more nonvanishing *K* integrals, the lower the energy. This is achieved by having the maximum possible number of electrons with *unpaired* spins. We conclude that S = 1 for carbon and S = 3/2 for nitrogen, so that the complete term symbols are ^{3}P and ^{4}S, respectively.

The allocation electrons among degenerate orbitals can be formalized by *Hund's rule*: For an atom in its ground state, the term with the highest multiplicity has the lowest energy.

Resuming Aufbau of the periodic table, oxygen with four 2*p* electrons must have one of the 2*p*-orbitals doubly occupied. But the remaining two electrons will choose unoccupied orbitals with parallel spins. Thus oxygen has, like carbon, a ^{3}P ground state. Fluorine can be regarded as a complete shell with an electron hole, thus a ^{2}P ground state. Neon completes the 2*s*2*p* shells, thus term symbol ^{1}S. The chemical stability and high ionization energy of all the noble-gas atoms can be attributed to their electronic structure of complete shells. The third row of the periodic table is filled in complete analogy with the second row. The similarity of the outermost electron shells accounts for the periodicity of chemical properties. Thus, the alkali metals Na and K belong in the same family as Li, the halogens Cl and Br are chemically similar to F, and so forth.

The transition elements, atomic numbers 21 to 30, present further challenges to our understanding of electronic structure. A complicating factor is that the energies of the 4*s* and 3*d* orbitals are very close, so that interactions among occupied orbitals often determines the electronic state. Ground-state electron configurations can be deduced from spectroscopic and chemical evidence, and confirmed by accurate self-consisent field computations. The 4*s* orbital is the first to be filled in K and Ca. Then come 3*d* electrons in Sc, Ti and V. A discontinuity occurs at Cr. The groundstate configuration is found to be $4s3d^5$, instead of the extrapolated $4s^23d^4$. This can be attributed to the enhanced stability of a half-filled $3d^5$-shell. All six electrons in the valence shells have parallel spins, maximizing the number of stabilizing exchange integrals and giving the observed ^{6}S term. An analogous discontinuity occurs for copper, in which the 4*s* subshell is again raided to complete the $3d^{10}$ subshell.

The order in which orbitals are filled is not necessarily consistent with the order in which electrons are removed. Thus, in all the positive ions of transition metals, the two 4*s*-electrons are removed first. The inadequacy of any simple generalizations about

orbital energies is demonstrated by comparing the three ground-state electron configurations: Ni $4s^2 3d^8$, Pd $5s^0 4d^{10}$ and Pt $6s5d^9$.

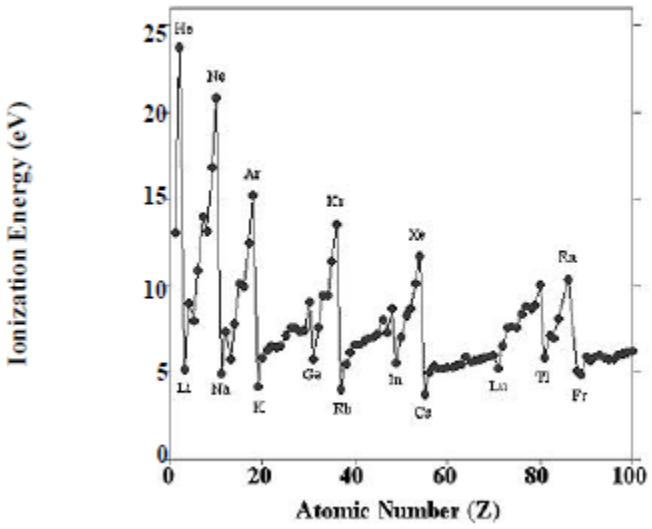


Figure 2. Periodic trends in ionization energy.

The periodic structure of the elements is evident for many physical and chemical properties, including chemical valence, atomic radius, electronegativity, melting point, density, and hardness. The classic prototype for periodic behavior is the variation of the first ionization energy with atomic number, which is plotted in in Figure 2.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)
- Caitlin Kozack (Hope College)

This page titled 9.11: Atomic Structure and The Periodic Law is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- 9: Atomic Structure and The Periodic Law is licensed CC BY-NC-SA 4.0.

9.12: The Chemical Bond

The Hydrogen Molecule

This four-particle system, two nuclei plus two electrons, is described by the Hamiltonian

$$\hat{H} = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{1}{2M_A}\nabla_A^2 - \frac{1}{2M_B}\nabla_B^2 - \frac{1}{r_{1A}} - \frac{1}{r_{2B}} - \frac{1}{r_{2A}} - \frac{1}{r_{1B}} + \frac{1}{r_{12}} + \frac{1}{R} \tag{9.12.1}$$

in terms of the coordinates shown in Figure 9.12.1. We note first that

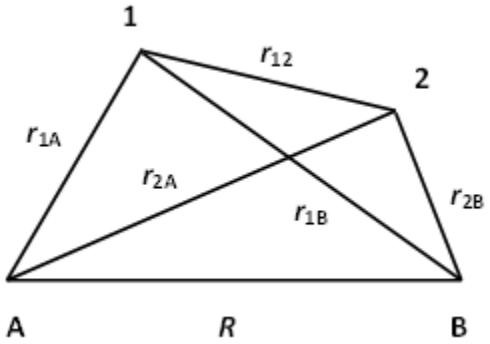


Figure 9.12.1: Coordinates used for hydrogen molecule.

the masses of the nuclei are much greater than those of the electrons,M_{proton} = 1836 atomic units, compared to m_{electron} = 1 atomic unit. Therefore nuclear kinetic energies will be negligibly small compared to those of the electrons. In accordance with the Born-Oppenheimer approximation, we can first consider the electronic Schrödinger equation

$$\hat{H}_{elec}\psi(r_1, r_2, R) = E_{elec}(R)\psi(r_1, r_2, R) \tag{9.12.2}$$

where

$$\hat{H} = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{1}{r_{1A}} - \frac{1}{r_{2B}} - \frac{1}{r_{2A}} - \frac{1}{r_{1B}} + \frac{1}{r_{12}} + \frac{1}{R} \tag{9.12.3}$$

The internuclear separation *R* occurs as a parameter in this equation so that the Schrödinger equation must, in concept, be solved for each value of the internuclear distance *R*. A typical result for the energy of a diatomic molecule as a function of *R* is shown in Figure 9.12.2. For a bound state, the energy minimum occurs at for $R = R_e$, known as the *equilibrium internuclear distance*. The depth of the potential well at R_e is called the *binding energy* or *dissociation energy* D_e. For the H_2 molecule, D_e = 4.746 eV and R_e=1.400 bohr = 0.7406 Å. Note that as $R \rightarrow 0$, $E(R) \rightarrow \infty$, since the 1/*R* nuclear repulsion will become dominant.

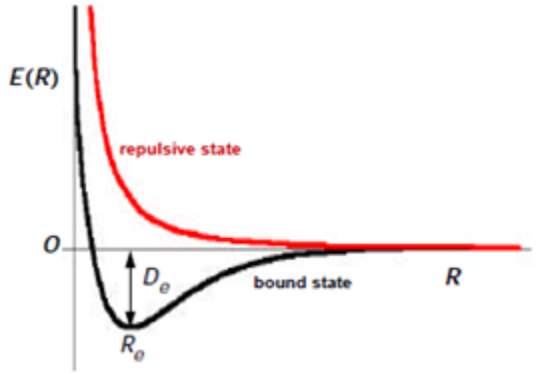


Figure 9.12.2. Energy curves for a diatomic molecule.

The more massive nuclei move much more slowly than the electrons. From the viewpoint of the nuclei, the electrons adjust almost instantaneously to any changes in the internuclear distance. The electronic energy $E_{elec}(R)$ therefore plays the role of a *potential energy* in the Schrödinger equation for nuclear motion

$$\left\{-\frac{1}{2M_A}\nabla_A^2 - \frac{1}{2M_B}\nabla_B^2 + V(R)\right\}\chi(r_A, r_B) = E\chi(r_A, r_B) \tag{9.12.4}$$

where

$$V(R) = E_{elec}(R) \tag{9.12.5}$$

from solution of Equation 9.12.2. Solutions of Equation 9.12.4 determine the vibrational and rotational energies of the molecule. These will be considered elsewhere. For the present, we are interested in the obtaining electronic energy from Equation 9.12.2 and 9.12.3. We will thus drop the subscript "elec" on $\hat{H}$ and $E(R)$ for the remainder this Chapter.

The first quantum-mechanical account of chemical bonding is due to Heitler and London in 1927, only one year after the Schrödinger equation was proposed. They reasoned that, since the hydrogen molecule H_2 was formed from a combination of hydrogen atoms *A* and *B*, a first approximation to its electronic wavefunction might be

$$\psi(r_1, r_2) = \psi_{1s}(r_{1A})\psi_{1s}(r_{2B}) \tag{9.12.6}$$

Using this function into the variational integral

$$\tilde{E}(R) = \frac{\int \psi\hat{H}\psi d\tau}{\int \psi^2 d\tau} \tag{9.12.7}$$

the value $R_e \approx 1.7$ bohr was obtained, indicating that the hydrogen atoms can indeed form a molecule. However, the calculated binding energy $D_e \approx 0.25$ eV, is much too small to account for the strongly-bound H_2 molecule. Heitler and London proposed that it was necessary to take into account the *exchange* of electrons, in which the electron labels in Equation 9.12.6 are reversed. The properly symmetrized function

$$\psi(r_1, r_2) = \psi_{1s}(r_{1A})\psi_{1s}(r_{2B}) + \psi_{1s}(r_{1B})\psi_{1s}(r_{2A}) \tag{9.12.8}$$

gave a much more realistic binding energy value of 3.20 eV, with R_e = 1.51 bohr. We have already used exchange symmetry (and antisymmetry) in our treatment of the excited states of helium. The variational function (Equation 9.12.8) was improved (Wang, 1928) by replacing the hydrogen 1s functions e^{-r} by $e^{-\zeta r}$. The optimized value ζ = 1.166 gave a binding energy of 3.782 eV. The quantitative breakthrough was the computation of James and Coolidge (1933). Using a 13-parameter function of the form

$$\psi(r_1, r_2) = e^{-\alpha(\xi_1+\xi_2)} \text{ x polynomial in}\{\xi_1, \xi_2, \eta_1, \eta_2, \rho\}, \xi_i \equiv \frac{r_{iA} + r_{iB}}{R}, \eta_i \equiv \frac{r_{iA} + r_{iB}}{R}, \rho \equiv \frac{r_{12}}{R} \tag{9.12.9}$$

they obtained R_e = 1.40 bohr, D_e = 4.720 eV. In a sense, this result provided a proof of the validity of quantum mechanics for molecules, in the same sense that Hylleraas' computation on helium was a proof for many-electron atoms.

The Valence Bond Theory

The basic idea of the Heitler-London model for the hydrogen molecule can be extended to chemical bonds between any two atoms. The orbital function (8) must be associated with the singlet spin function $\sigma_{0,0}(1, 2)$ in order that the overall wavefunction be antisymmetric. This is a quantum-mechanical realization of the concept of an electron-pair bond, first proposed by G. N. Lewis in 1916. It is also now explained why the electron spins must be paired, i.e., antiparallel. It is also permissible to combine an antisymmetric orbital function with a triplet spin function but this will, in most cases, give a repulsive state, as shown by the red curve in Figure 9.12.2.

According to valence-bond theory, unpaired orbitals in the valence shells of two adjoining atoms can combine to form a chemical bond if they overlap significantly and are symmetry compatible. A σ-bond is cylindrically symmetrical about the axis joining the atoms. Two s AO's, two p_z AO's or an s and a p_z can contribute to a σ-bond, as shown in Figure 9.12.3. The z-axis is chosen along the internuclear axis. Two p_x or two p_y AO's can form a π-bond, which has a nodal plane containing the internuclear axis. Examples of symmetry-incompatible AO's would be an *s* with a p_x or a p_x with a p_y. In such cases the overlap integral would vanish because of cancelation of positive and negative contributions. Some possible combinations of AO's forming σ and π bonds are shown in

Figure 9.12.3.

Bonding in the HCl molecule can be attributed to a combination of a hydrogen 1s with an unpaired $3p_z$ on chlorine. In Cl_2, a sigma bond is formed between the $3p_z$ AO's on each chlorine. As a first approximation, the other doubly-occupied AO's on chlorine-the inner shells and the valence-shell lone pairs-are left undisturbed.

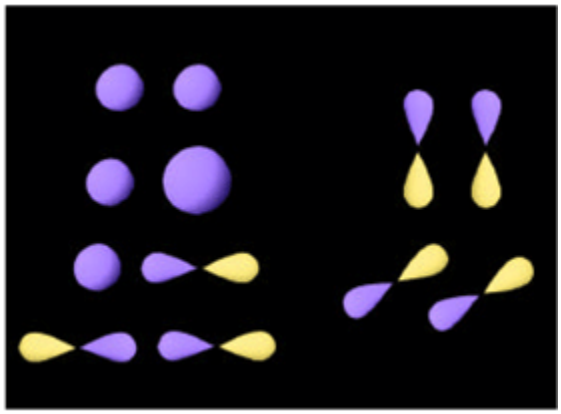


Figure 9.12.3. Formation of σ and π bonds.

The oxygen atom has two unpaired 2*p*-electrons, say $2p_x$ and $2p_y$. Each of these can form a σ-bond with a hydrogen 1s to make a water molecule. It would appear from the geometry of the *p*-orbitals that the HOH bond angle would be 90°. It is actually around 104.5°. We will resolve this discrepency shortly. The nitrogen atom, with three unpaired 2*p* electrons can form three bonds. In NH_3, each 2*p*-orbital forms a σ-bond with a hydrogen 1s. Again 90° HNH bond angles are predicted, compared with the experimental 107°. The diatomic nitrogen molecule has a triple bond between the two atoms, one σ bond from combining $2p_z$ AO's and two π bonds from the combinations of $2p_x$'s and $2p_y$'s, respectively.

Hybrid Orbitals and Molecular Geometry

To understand the bonding of carbon atoms, we must introduce additional elaborations of valence-bond theory. We can write the valence shell configuration of carbon atom as $2s^2 2p_x 2p_y$, signifying that two of the 2*p* orbitals are unpaired. It might appear that carbon would be divalent, and indeed the species CH_2 (carbene or methylene radical) does have a transient existence. But the chemistry of carbon is dominated by tetravalence. Evidently it is a good investment for the atom to promote one of the 2*s* electrons to the unoccupied $2p_z$ orbital. The gain in stability attained by formation of four bonds more than compensates for the small excitation energy. It can thus be understood why the methane molecule CH_4 exists. The molecule has the shape of a regular tetrahedron, which is the result of *hybridization*, mixing of the s and three *p* orbitals to form four sp^3 hybrid atomic orbitals. Hybrid orbitals can overlap more strongly with neighboring atoms, thus producing stronger bonds. The result is four C-H σ-bonds, identical except for orientation in space, with 109.5° H-C-H bond angles.

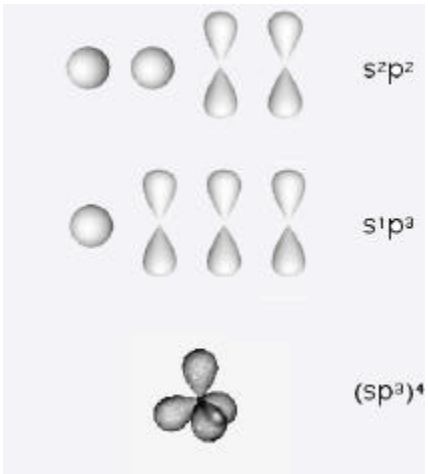
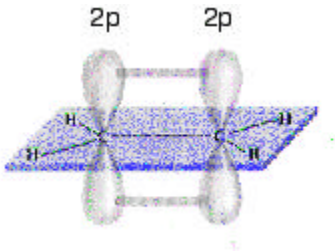
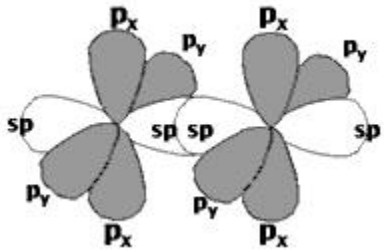


Figure 9.12.4. Promotion and hybridization of atomic orbitals in carbon atom.

Other carbon compounds make use of two alternative hybridization schemes. The s AO can form hybrids with *two* of the *p* AO's to give three sp^2 hybrid orbitals, with one *p*-orbital remaining unhybridized. This accounts for the structure of ethylene (ethene):



The C-H and C-C σ-bonds are all trigonal sp^2 hybrids, with 120° bond angles. The two unhybridized *p*-orbitals form a π-bond, which gives the molecule its rigid planar structure. The two carbon atoms are connected by a double bond, consisting of one σ and one π. The third canonical form of *sp*-hybridization occurs in C-C triple bonds, for example, acetylene (ethyne). Here, two of the *p* AO's in carbon remain unhybridized and can form two π-bonds, in addition to a σ-bond, with a neighboring carbon:



Acetylene H-C≡C-H is a linear molecule since *sp*-hybrids are oriented 180° apart.

The deviations of the bond angles in H_2O and NH_3 from 90° can be attributed to fractional hybridization. The angle H-O-H in water is 104.5° while H-N-H in ammonia is 107°. It is rationalized that the *p*-orbitals of the central atom acquire some s-character and increase their angles towards the tetrahedral value of 109.5°. Correspondingly, the lone pair orbitals must also become hybrids. Apparently, for both water and ammonia, a model based on tetrahedral orbitals on the central atoms would be closer to the actual behavior than the original selection of s- and *p*-orbitals. The hybridization is driven by repulsions between the electron densities of neighboring bonds.

Valence Shell Model

An elementary, but quite successful, model for determining the shapes of molecules is the *valence shell electron repulsion theory* (VSEPR), first proposed by Sidgewick and Powell and popularized by Gillespie. The local arrangement of atoms around each multivalent center in the molecule can be represented by $AX_{n-k}E_k$, where X is another atom and E is a lone pair of electrons. The geometry around the central atom is then determined by the arrangement of the *n* electron pairs (bonding plus nonbonding), which minimizes their mutual repulsion. The following geometric configurations satisfy this condition:

n *shape*
2 linear 5 trigonal bipyramid
3 trigonal planar 6 octahedral
4 tetrahedral 7 pentagonal bipyramid

The basic geometry will be distorted if the *n* surrounding pairs are not identical. The relative strength of repulsion between pairs follows the order E-E > E-X > X-X. In ammonia, for example, which is NH_3E, the shape will be tetrahedral to a first approximation. But the lone pair E will repel the N-H bonds more than they repel one another. Thus the E-N-H angle will increase from the tetrahedral value of 109.5°, causing the H-N-H angles to decrease slightly. The observed value of 107° is quite well accounted for. In water, OH_2E_2, the opening of the E-O-E angle will likewise cause a closing of H-O-H, and again, 104.5° seems like a reasonable value.

Valence-bond theory is about 90% successful in explaining much of the descriptive chemistry of ground states. VB theory fails to account for the triplet ground state of O_2 or for the bonding in electron-deffcient molecules such as diborane, B_2H_6. It is not very useful in consideration of excited states, hence for spectroscopy. Many of these deficiencies are remedied by molecular orbital theory, which we take up in the next Chapter.

This page titled 9.12: The Chemical Bond is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Catherine Calyore & S.M. Blinder.

- 10: The Chemical Bond by S. M. Blinder is licensed CC BY-NC-SA 4.0.

9.13: Molecular Orbital Theory

Molecular orbital theory is a conceptual extension of the orbital model, which was so successfully applied to atomic structure. As was once playfully remarked, "a molecule is nothing more than an atom with more nuclei." This may be overly simplistic, but we do attempt, as far as possible, to exploit analogies with atomic structure. Our understanding of atomic orbitals began with the exact solutions of a prototype problem – the hydrogen atom. We will begin our study of homonuclear diatomic molecules beginning with another exactly solvable prototype, the hydrogen molecule-ion H_2^+.

The Hydrogen Molecule-Ion

The simplest conceivable molecule would be made of two protons and one electron, namely H_2^+. This species actually has a transient existence in electrical discharges through hydrogen gas and has been detected by mass spectrometry. It also has been detected in outer space. The Schrödinger equation forHH_2^+ can be solved exactly within the Born-Oppenheimer approximation. For fixed internuclear distance *R*, this reduces to a problem of one electron in the field of two protons, designated A and B. We can write

$$\left\{-\frac{1}{2}\nabla^2\frac{1}{r_A}-\frac{1}{r_B}+\frac{1}{R}\right\}\psi(r)=E\psi(r) \tag{9.13.1}$$

where r_A and r_B are the distances from the electron to protons A and B, respectively. This equation was solved by Burrau (1927), after separating the variables in prolate spheroidal coordinates. We will write down these coordinates but give only a pictorial account of the solutions. The three prolate spheroidal coordinates are designated ξ, η, ϕ. the first two are defined by

$$\xi=\frac{r_A+r_B}{R} \tag{9.13.2}$$

and

$$\eta=\frac{r_A-r_B}{R} \tag{9.13.3}$$

while ϕ is the angle of rotation about the internuclear axis. The surfaces of constant ξ and η are, respectively, confocal ellipsoids and hyperboloids of revolution with foci at A and B. The two-dimensional analog should be familiar from analytic geometry, an ellipse being the locus of points such that the sum of the distances to two foci is a constant. Analogously, a hyperbola is the locus whose *difference* is a constant. Figure $9.13.1$ shows several surfaces of constant ξ, η and ϕ. The ranges of the three coordinates are $\xi \overset{\text{int}}{} \{1,\infty\}$, $\eta \overset{\text{int}}{} \{-1,1\}$, $\phi \overset{\text{int}}{} \{0,2\pi\}$. The prolate-spheroidal coordinate system conforms to the natural symmetry of the H_2^+ problem in the same way that spherical polar coordinates were the appropriate choice for the hydrogen atom.

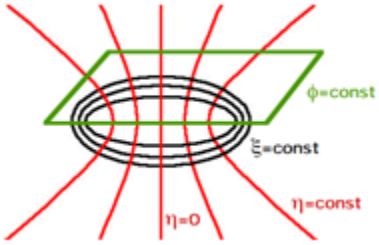


Figure $9.13.1$: Prolate spheroidal coordinates.

The first few solutions of the H_2^+ Schrödinger equation are sketched in Figure $9.13.2$, roughly in order of increasing energy. The ϕ -dependence of the wavefunction is contained in a factor

$$\Phi(\phi)=e^{i\lambda\phi}, \qquad \lambda=0,\pm1,\pm2,... \tag{9.13.4}$$

which is identical to the ϕ-dependence in atomic orbitals. In fact, the quantum number λ represents the component of orbital angular momentum along the internuclear axis, the only component which has a definite value in systems with axial (cylindrical) symmetry. The quantum number λ determines the basic shape of a diatomic molecular orbital, in the same way that ℓ did for an atomic orbital. An analogous code is used σ for λ = 0, π for $\lambda = \pm1$, δ for $\lambda = \pm2$, and so on. We are already familiar with σ- and π

-orbitals from valence-bond theory. A second classification of the H_2^+ eigenfunctions pertains to their symmetry with respect to inversion through the center of the molecule, also known as *parity*. If ψ(-r) = +ψ(r), the function is classified *gerade* or even parity, and the orbital designation is given a subscript *g*, as in σ_g or π_g. If ψ(-r) = -ψ(r), the function is classified as *ungerade* or odd parity, and we write instead σ_u or π_u. Atomic orbitals can also be classified by inversion symmetry. However, all *s* and *d* atomic orbitals are *g*, while all *p* and *f* orbitals are *u*, so no further designation is necessary. The molecular orbitals of a given symmetry are numbered in order of increasing energy, for example, 1σ_g, 2σ_g, 3σ_g.

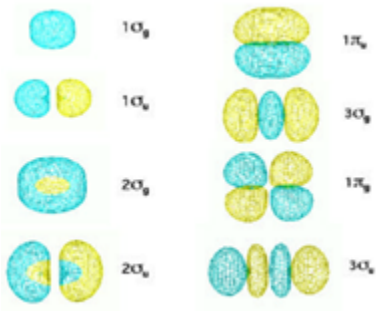


Figure 9.13.2: H_2^+ molecular orbitals.

The lowest-energy orbital, as we have come to expect, is nodeless. It obviously must have cylindrical symmetry (λ = 0) and inversion symmetry (*g*). It is designated 1σ_g since it is the first orbital of this classification. The next higher orbital has a nodal plane, with η = 0, perpendicular to the axis. This function still has cylindrical symmetry (σ) but now changes sign upon inversion (*u*). It is designated 1σ_u, as the first orbital of this type. The next higher orbital has an inner ellipsiodal node. It has the same symmetry as the lowest orbital and is designated 2σ_g. Next comes the 2σ_u orbital, with both planar and ellipsoidal nodes. Two degenerate π-orbitals come next, each with a nodal plane containing the internuclear axis, with ϕ=const. Their classification is 1π_u. The second 1π_u-orbital, not shown in Figure 9.13.2, has the same shape rotated by 90°. The 3σ_g orbital has two hyperbolic nodal surfaces, where η = $\pm$const. The 1π_g, again doubly-degenerate, has two nodal planes, η = 0 and ϕ=const. Finally, the 3σ_u, the last orbital we consider, has three nodal surfaces where η=const.

An molecular orbital is classified as a *bonding orbital* if it promotes the bonding of the two atoms. Generally a bonding molecular orbital has a significant accumulation of electron charge in the region between the nuclei and thus reduces their mutual repulsion. The 1σ_g, 2σ_g, 1π_u and 3σ_g are evidently bonding orbitals. An molecular orbital which does *not* significantly contribute to nuclear shielding is classified as an *antibonding orbital*. The 1σ_u, 2σ_u, 1π_g and 3σ_u belong in this category. Often an antibonding molecular orbital is designated by σ* or π*.

The actual ground state of H_2^+ has the 1σ_g orbital occupied. The equilibrium internuclear distance R_e is 2.00 bohr and the binding energy D_e is 2.79 eV, which represents quite a strong chemical bond. The 1σ_u is a repulsive state and a transition from the ground state results in dissociation of the molecule.

The LCAO Approximation

In Figure 9.13.3, the 1σ_g and 1σ_u orbitals are plotted as functions of *z*, along the internuclear axis. Both functions have cusps, discontinuities in slope, at the positions of the two nuclei A and B. The 1s orbitals of hydrogen atoms have the same cusps. The shape of the 1σ_g and 1σ_u suggests that they can be approximated by a sum and difference, respectively, of hydrogen 1s orbitals, such that

$$\psi(1\sigma_{g,u}) \approx \psi(1s_A) \pm \psi(1s_B) \tag{9.13.5}$$

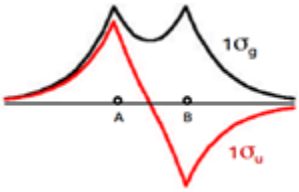


Figure 9.13.3: H_2^+ orbitals along internuclear axis.

This *linear combination of atomic orbitals* is the basis of the so-called LCAO approximation. The other orbitals pictured in Figure 9.13.2 can likewise be approximated as follows:

$$\psi(2\sigma_{g,u}) \approx \psi(2s_A) \pm \psi(2s_B) \tag{9.13.6}$$

$$\psi(3\sigma_{g,u}) \approx \psi(2p\sigma_A) \pm \psi(2p\sigma_B) \tag{9.13.7}$$

$$\psi(1\pi_{u,g}) \approx \psi(2p\pi_A) \pm \psi(2p\pi_B) \tag{9.13.8}$$

The $2p\sigma$ atomic orbital refers to $2p_z$, which has the axial symmetry of a σ-bond. Likewise $2p\pi$ refers to $2p_x$ or $2p_y$, which are positioned to form π-bonds. An alternative notation for diatomic molecular orbitals which specifies their atomic origin and bonding/antibonding character is the following:

$1\sigma_g$	$1\sigma_u$	$2\sigma_g$	$2\sigma_u$	$3\sigma_g$	$3\sigma_u$	$1\pi_u$	$1\pi_g$
σ1s	σ*1s	σ2s	σ*2s	σ2p	σ*2p	π2p	π*2p

Almost all applications of molecular-orbital theory are based on the LCAO approach, since the exact H_2^+ functions are far too complicated to work with.

The relationship between molecular orbitals and their constituent atomic orbitals can be represented in correlation diagrams, show in Figure $9.13.4$.

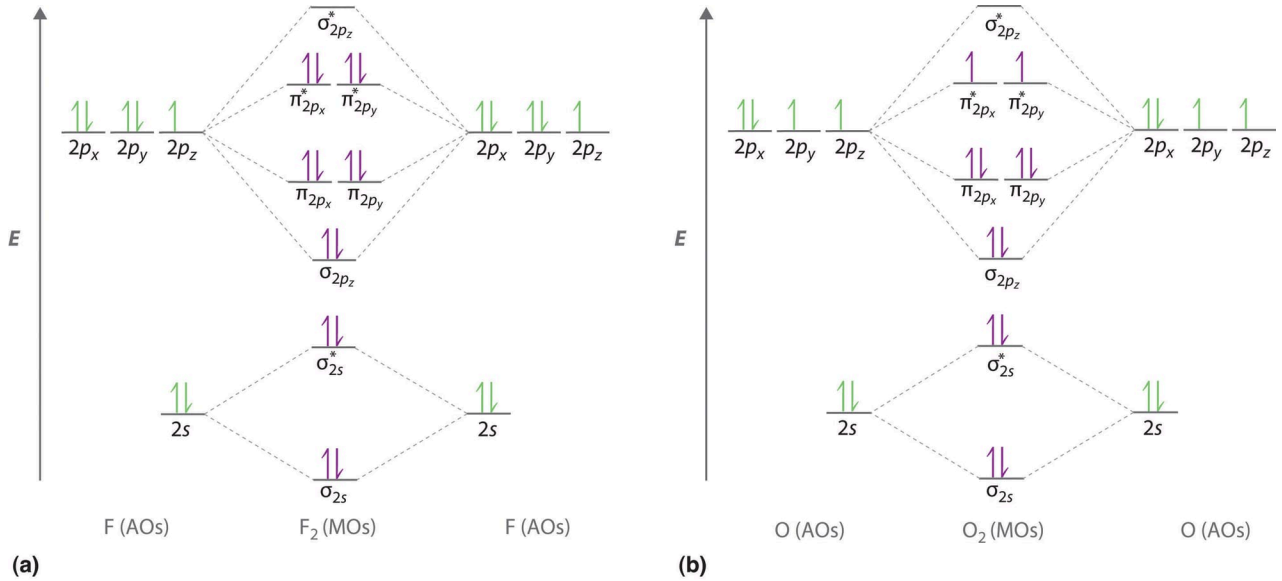


Figure $9.13.4$: Molecular Orbital Energy-Level Diagrams for Homonuclear Diatomic Molecules.(a) For F_2, with 14 valence electrons (7 from each F atom), all of the energy levels except the highest, $\sigma^\star_{2p_z}$ are filled. This diagram shows 8 electrons in bonding orbitals and 6 in antibonding orbitals, resulting in a bond order of 1. (b) For O_2, with 12 valence electrons (6 from each O atom), there are only 2 electrons to place in the $\left(\pi^\star_{np_x},\ \pi^\star_{np_y}\right)$ pair of orbitals. Hund's rule dictates that one electron occupies each orbital, and their spins are parallel, giving the O_2 molecule two unpaired electrons. This diagram shows 8 electrons in bonding orbitals and 4 in antibonding orbitals, resulting in a predicted bond order of 2.

MO Theory of Homonuclear Diatomic Molecules

A sufficient number of orbitals is available for the *Aufbau* of the ground states of all homonuclear diatomic species from H_2 to Ne_2. Table 1 summarizes the results. The most likely order in which the molecular orbitals are filled is given by

$$1\sigma_g < 1\sigma_u < 2\sigma_g < 2\sigma_u < 3\sigma_g \sim 1\pi_u < 1\pi_g < 3\sigma_u \tag{9.13.9}$$

The relative order of $3\sigma_g$ and $1\pi_u$ depends on which other molecular orbitals are occupied, much like the situation involving the 4s and 3*d* atomic orbitals. The results of photoelectron spectroscopy indicate that $1\pi_u$ is lower up to and including N_2, but $3\sigma_g$ is lower thereafter.

The term symbol $\Sigma, \Pi, \Delta \ldots$, analogous to the atomic S, P, D. . . symbolizes the axial component of the total orbital angular momentum. When a π-shell is filled (4 electrons) or half-filled (2 electrons), the orbital angular momentum cancels to zero and we

find a Σ term. The spin multiplicity is completely analogous to the atomic case. The total parity is again designated by a subscript *g* or *u*. Since the many electron wavefunction is made up of products of individual MO's, the total parity is odd only if the molecule contains an *odd* number of *u* orbitals. Thus a σ_u^2 or a π_u^2 subshell transforms like *g*.

For Σ terms, the superscript $\pm$ denotes the sign change of the wavefunction under a reflection in a plane containing the internuclear axis. This is equivalent to a sign change in the variable $\phi \rightarrow -\phi$. This symmetry is needed when we deal with spectroscopic selection rules. In a spin-paired π_u^2 subshell the triplet spin function is symmetric so that the orbital factor must be antisymmetric, of the form

$$\frac{1}{\sqrt{2}}\Big(\pi_x(1)\pi_y(2) - \pi_y(1)\pi_x(2)\Big) \tag{9.13.10}$$

This will change sign under the reflection, since $x \rightarrow x$ but $y \rightarrow -y$. We need only remember that a π_u^2 subshell will give the term symbol ${}^3\Sigma_g^-$.

The net bonding effect of the occupied molecular orbitals is determined by the *bond order*, half the excess of the number bonding minus the number antibonding. This definition brings the molecular orbital results into correspondence with the Lewis (or valence-bond) concept of single, double and triple bonds. It is also possible in molecular orbital theory to have a bond order of 1/2, for example, in H_2^+ which is held together by a single bonding orbital. A bond order of zero generally indicates no stable chemical bond, although helium and neon atoms can still form clusters held together by much weaker van der Waals forces. Molecular-orbital theory successfully accounts for the transient stability of a ${}^3\Sigma_u^+$ excited state of He_2, in which one of the antibonding electrons is promoted to an excited bonding orbital. This species has a lifetime of about 10^{-4} sec, until it emits a photon and falls back into the unstable ground state. Another successful prediction of molecular orbital theory concerns the relative binding energy of the positive ions N_2^+ and O_2^+, compared to the neutral molecules. Ionization weakens the N–N bond since a bonding electron is lost, but it strengthens the O–O bond since an antibonding electron is lost.

One of the earliest triumphs of molecular orbital theory was the prediction that the oxygen molecule is paramagnetic. Figure 9.13.5 shows that liquid O_2 is a magnetic substance, attracted to the region between the poles of a permanent magnet. The paramagnetism arises from the half-filled $1\pi_g^2$ subshell. According to Hund's rules the two electrons singly occupy the two degenerate $1\pi_g$ orbitals with their spins aligned *parallel*. The term symbol is ${}^3\Sigma_g^-$ and the molecule thus has a nonzero spin angular momentum and a net magnetic moment, which is attracted to an external magnetic field. Linus Pauling invented the paramagnetic oxygen analyzer, which is extensively used in medical technology.

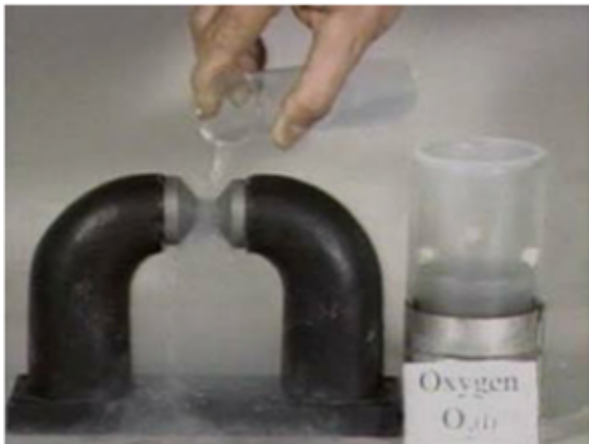


Figure 9.13.5. Demonstration showing blue liquid O_2 attracted to the poles of a permanent magnet. From jchemed.chem.wisc.edu/jcesoft.../PARANIO/9.HTM

Variational Computation of Molecular Orbitals

Thus far we have approached molecular orbital theory from a mainly descriptive point of view. To begin a more quantitative treatment, recall the LCAO approximation to the H_2^+ ground state, Equation 9.13.5, which can be written

$$\psi = c_A\psi_A + c_B\psi_B \tag{9.13.11}$$

Using this as a trial function in the variational principle, we have

$$E(c_A, c_B) = \frac{\int \psi \hat{H} \psi d\tau}{\int \psi^2 d\tau} \tag{9.13.12}$$

where $\hat{H}$ is the Hamiltonian from Equation 9.13.1. In fact, these equations can be applied more generally to construct *any* molecular orbital, not just solutions for H_2^+. In the general case, $\hat{H}$ will represent an effective one-electron Hamiltonian determined by the molecular environment of a given orbital. The energy expression involves some complicated integrals, but can be simplified somewhat by expressing it in a standard form. Hamiltonian matrix elements are defined by

$$H_{AA} = \int \psi_A \hat{H} \psi_A d\tau \tag{9.13.13}$$

$$H_{BB} = \int \psi_B \hat{H} \psi_B d\tau \tag{9.13.14}$$

$$H_{AB} = H_{BA} = \int \psi_A \hat{H} \psi_B d\tau \tag{9.13.15}$$

while the overlap integral is given by

$$S_{AB} = \int \psi_A \psi_B d\tau \tag{9.13.16}$$

Presuming the functions ψ_A and ψ_B to be normalized, the variational energy (Equation 9.13.12) reduces to

$$E(c_A c_B) = \frac{c_A^2 H_{AA} + 2c_A c_B H_{AB} + c_B^2 H_{BB}}{c_A^2 + 2c_A c_B S_{AB} + c_B^2} \tag{9.13.17}$$

To optimize the MO, we find the minimum of *E* wrt variation in c_A and c_B, as determined by the two conditions

$$\frac{\partial E}{\partial c_A} = 0, \qquad \frac{\partial E}{\partial c_B} = 0 \tag{9.13.18}$$

The result is a *secular equation* determining two values of the energy:

$$\begin{vmatrix} H_{AA} - E & H_{AB} - ES_{AB} \\ H_{AB} - ES_{AB} & H_{BB} - E \end{vmatrix} = 0 \tag{9.13.19}$$

For the case of a homonuclear diatomic molecule, for example H_2^+, the two Hamiltonian matrix elements H_{AA} and H_{BB} are equal, say to α. Setting $H_{AB} = \beta$ and $S_{AB} = S$, the secular equation reduces to

$$\begin{vmatrix} \alpha - E & \beta - ES \\ \beta - ES & \alpha - E \end{vmatrix} = (\alpha - E)^2 - (\beta - ES)^2 = 0 \tag{9.13.20}$$

with the two roots

$$E^{\pm} = \frac{\alpha \pm \beta}{1 \pm S} \tag{9.13.21}$$

The calculated integrals α and β are usually negative, thus for the bonding orbital

$$E^{+} = \frac{\alpha + \beta}{1 + S} \qquad (bonding) \tag{9.13.22}$$

while for the antibonding orbital

$$E^{-} = \frac{\alpha - \beta}{1 - S} \qquad (antibonding) \tag{9.13.23}$$

Note that $(E^- - \alpha) > (\alpha - E^+)$, thus the energy increase associated with antibonding is slightly greater than the energy decrease for bonding. For historical reasons, α is called a *Coulomb integral* and β, a *resonance integral*.

Heteronuclear Molecules

The variational computation leading to Equation 9.13.19 can be applied as well to the heteronuclear case in which the orbitals ψ_A and ψ_B are *not* equivalent. The matrix elements H_{AA} and H_{BB} are approximately equal to the energies of the atomic orbitals ψ_A and ψ_B, respectively, say E_A and E_B with $E_A > E_B$. It is generally true that $|E_A|, |E_B| \gg |H_{AB}|$. With these simplifications, secular equation can be written

$$\begin{vmatrix} E_A - E & H_{AB} - ES_{AB} \\ H_{AB} - ES_{AB} & E_B - E \end{vmatrix} = (E_A - E)(E_B - E) - (H_{AB} - ES_{AB})^2 = 0 \tag{9.13.24}$$

This can be rearranged to

$$E - E_A = \frac{(H_{AB} - ES_{AB})^2}{E - E_B} \tag{9.13.25}$$

To estimate the root closest to E_A, we can replace E by E_A on the right hand side of the equation. This leads to

$$E^- \approx E_A + \frac{(H_{AB} - E_A S_{AB})^2}{E_A - E_B} \tag{9.13.26}$$

and analogously for the other root,

$$E^+ \approx E_B - \frac{(H_{AB} - E_B S_{AB})^2}{E_A - E_B} \tag{9.13.27}$$

The following correlation diagram represents the relative energies of these atomic orbitals and MO's:

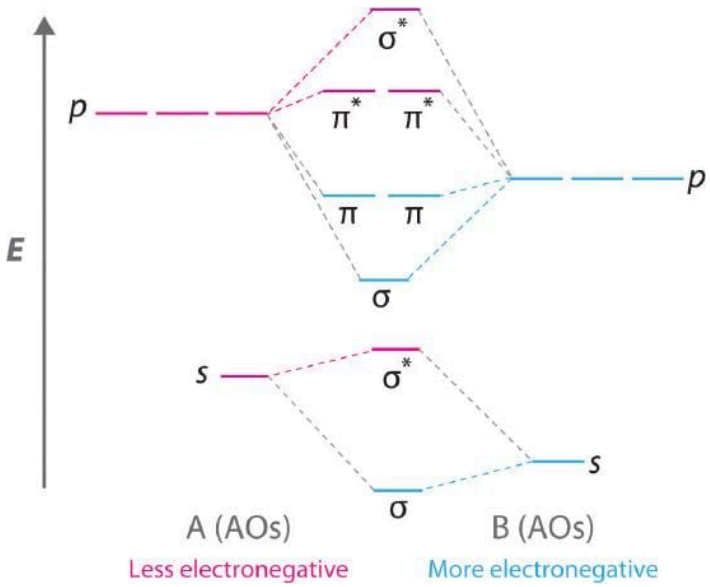


Figure 9.13.6: Molecular Orbital Energy-Level Diagram for a Heteronuclear Diatomic Molecule AB, Where $\chi_B > \chi_A$. The bonding molecular orbitals are closer in energy to the atomic orbitals of the more electronegative B atom. Consequently, the electrons in the bonding orbitals are not shared equally between the two atoms. On average, they are closer to the B atom, resulting in a polar covalent bond.

A simple analysis of Equations 9.13.24 implies that, in order for two atomic orbitals ψ_A and ψ_B to form effective molecular orbitals the following conditions must be met:

I. (The atomic orbitals must have compatible symmetry. For example, ψ_A and ψ_B can be either s or $p\sigma$ orbitals to form a σ-bond or both can be $p\pi$ (with the same orientation) to form a π-bond.
II. The charge clouds of ψ_A and ψ_B should overlap as much as possible. This was the rationale for hybridizing the *s* and *p* orbitals in carbon. A larger value of S_{AB} implies a larger value for H_{AB}.
II. The energies E_A and E_B must be of comparable magnitude. Otherwise, the denominator in (20) and (21) will be too large and the molecular orbitals will not differ significantly from the original AO's. A rough criterion is that E_A and E_B should be within about 0.2 hartree or 5 eV. For example, the chlorine 3*p* orbital has an energy of -13.0 eV, comfortably within range of the hydrogen 1s, with energy -13.6 eV. Thus these can interact to form a strong bonding (plus an antibonding) molecular orbital

in HCl. The chlorine 3s has an energy of -24.5 eV, thus it could not form an effective bond with hydrogen even if it were available.

Hückel Molecular Orbital Theory

Molecular orbital theory has been very successfully applied to large conjugated systems, especially those containing chains of carbon atoms with alternating single and double bonds. An approximation introduced by Hückel in 1931 considers only the delocalized *p* electrons moving in a framework of σ-bonds. This is, in fact, a more sophisticated version of a free-electron model. We again illustrate the model using butadiene CH_2=CH-CH=CH_2. From four *p* atomic orbitals with nodes in the plane of the carbon skeleton, one can construct four π molecular orbitals by an extension of the LCAO approach:

$$\psi = c_1\psi_1 + c_2\psi_2 + c_3\psi_3 + c_4\psi_4 \tag{9.13.28}$$

Applying the linear variational method, the energies of the molecular orbitals are the roots of the 4 x 4 secular equation

$$\begin{vmatrix} H_{11}-E & H_{12}-ES_{12} & \ldots \\ H_{12}-ES_{12} & H_{22}-E & \ldots \\ \ldots & \ldots & \ldots \end{vmatrix} = 0 \tag{9.13.29}$$

Four simplifying assumptions are now made

I. All overlap integrals S_{ij} are set equal to zero. This is quite reasonable since the *p*-orbitals are directed perpendicular to the direction of their bonds.
II. All resonance integrals H_{ij} between non-neighboring atoms are set equal to zero.
II. All resonance integrals H_{ij} between neighboring atoms are set equal to β.
IV. All coulomb integrals H_{ii} are set equal to α.

The secular equation thus reduces to

$$\begin{vmatrix} \alpha-E & \beta & 0 & 0 \\ \beta & \alpha-E & \beta & 0 \\ 0 & \beta & \alpha-E & \beta \\ 0 & 0 & \beta & \alpha-E \end{vmatrix} = 0 \tag{9.13.30}$$

Dividing by β^4 and defining

$$x = \frac{\alpha - E}{\beta} \tag{9.13.31}$$

the equation simplifies further to

$$\begin{vmatrix} x & 1 & 0 & 0 \\ 1 & x & 1 & 0 \\ 0 & 1 & x & 1 \\ 0 & 0 & 1 & x \end{vmatrix} = 0 \tag{9.13.32}$$

This is essentially the connection matrix for the molecule. Each pair of connected atoms is represented by 1, each non-connected pair by 0 and each diagonal element by x. Expansion of the determinant gives the 4th order polynomial equation

$$x^4 - 3x^2 + 1 = 0 \tag{9.13.33}$$

Noting that this is a quadratic equation in x^2, the roots are found to be $x^2 = (3 \pm \sqrt{5})/2$, so that $x = \pm 0.618, \pm 1.618$. This corresponds to the four MO energy levels

$$E = \alpha \pm 1.618\beta, \qquad \alpha \pm 0.618\beta \tag{9.13.34}$$

Since α and β are negative, the lowest molecular orbitals have

$$E(1\pi) = \alpha + 1.618\beta \tag{9.13.35}$$

and

$$E(2\pi) = \alpha + 0.618\beta \tag{9.13.36}$$

and the total π-electron energy of the $1\pi^2 2\pi^2$ configuration equals

$$E_\pi = 2(\alpha + 1.618\beta) + 2(\alpha + 0.618\beta) = 4\alpha + 4.472\beta \tag{9.13.37}$$

The simplest application of Hückel theory, to the ethylene molecule CH_2=CH_2 gives the secular equation

$$\begin{vmatrix} x & 1 \\ 1 & x \end{vmatrix} = 0 \tag{9.13.38}$$

This is easily solved for the energies $E = \alpha \pm \beta$. The lowest orbital has $E(1\pi) = \alpha + \beta$ and the $1\pi^2$ ground state has $E_\pi = 2(\alpha + \beta)$. If butadiene had two localized double bonds, as in its dominant valence-bond structure, its π-electron energy would be given by $E_\pi = 4(\alpha + \beta)$. Comparing this with the Hückel result (Equation 9.13.37), we see that the energy lies lower than the that of two double bonds by 0.48β. The thermochemical value is approximately -17 kJmol^{-1}. This stabilization of a conjugated system is known as the *delocalization energy*. It corresponds to the resonance-stabilization energy in valence-bond theory.

Aromatic systems provide the most significant applications of Hückel theory. For benzene, we find the secular equation

$$\begin{vmatrix} x & 1 & 0 & 0 & 0 & 1 \\ 1 & x & 1 & 0 & 0 & 0 \\ 0 & 1 & x & 1 & 0 & 0 \\ 0 & 0 & 1 & x & 1 & 0 \\ 0 & 0 & 0 & 1 & x & 1 \\ 1 & 0 & 0 & 0 & 1 & x \end{vmatrix} = 0 \tag{9.13.39}$$

with the **six roots** $x = \pm 2, \pm 1, \pm 1$. The energy levels are $E = \alpha \pm 2\beta$ and two-fold degenerate $E = \alpha \pm \beta$. With the three lowest molecular orbitals occupied, we have

$$E_\pi = 2(\alpha + 2\beta) + 4(\alpha + \beta) = 6\alpha + 8\beta \tag{9.13.40}$$

Since the energy of three localized double bonds is $6\alpha + 6\beta$, the delocalization energy equals 2β. The thermochemical value is -152 kJmol^{-1}.

Contributors

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)

This page titled 9.13: Molecular Orbital Theory is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Molecular Orbital Theory is licensed CC BY-NC-SA 4.0.

9.14: Molecular Symmetry

In many cases, the symmetry of a molecule provides a great deal of information about its quantum states, even without a detailed solution of the Schrödinger equation. A geometrical transformation which turns a molecule into an indistinguishable copy of itself is called a *symmetry operation*. A symmetry operation can consist of a rotation about an axis, a reflection in a plane, an inversion through a point, or some combination of these.

The Ammonia Molecule

We shall introduce the concepts of symmetry and group theory by considering a concrete example–the ammonia molecule NH_3. In any symmetry operation on NH_3, the nitrogen atom remains fixed but the hydrogen atoms can be permuted in 3!=6 different ways. The axis of the molecule is called a C_3 axis, since the molecule can be rotated about it into 3 equivalent orientations, 120° apart. More generally, a C_n axis has n equivalent orientations, separated by $2\pi/n$ radians. The axis of highest symmetry in a molecule is called the *principal axis*. Three mirror planes, designated $\sigma_1, \sigma_2, \sigma_3$, run through the principal axis in ammonia. These are designated as σ_v or *vertical* planes of symmetry. Ammonia belongs to the symmetry group designated C_{3v}, characterized by a three-fold axis with three vertical planes of symmetry.

Let us designate the orientation of the three hydrogen atoms in Figure $9.14.1$ as {1, 2, 3}, reading in clockwise order from the bottom. A counterclockwise rotation by 120°, designated

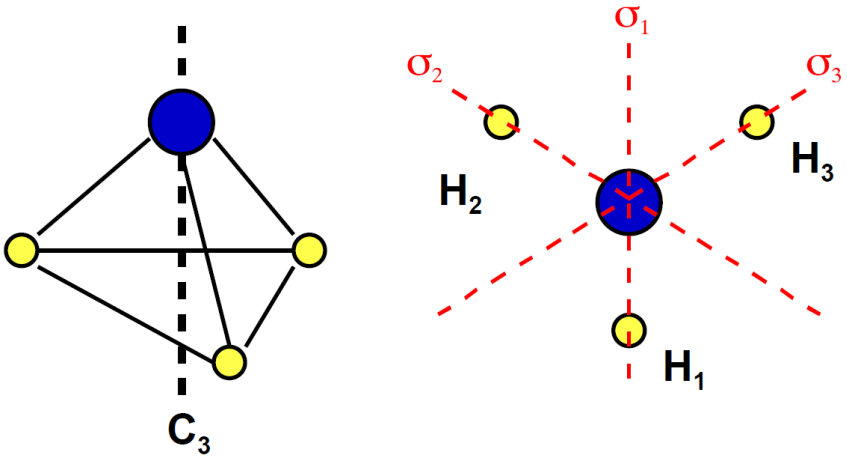


Figure $9.14.1$: Two views of the ammonia molecule.

by the operator C_3, produces the orientation {2, 3, 1}. A second counterclockwise rotation, designated C_3^2, produces {3, 1, 2}. Note that two successive counterclockwise rotations by 120° is equivalent to one clockwise rotation by 120°, so the last operation could also be designated C_3^{-1}. The three reflection operations $\sigma_1, \sigma_2, \sigma_3$, applied to the original configuration {1, 2, 3} produces {1, 3, 2}, {3, 2, 1} and {2, 1, 3}, respectively. Finally, we must include the identity operation, designated E, which leaves an orientation unchanged. The effects of the six possible operations of the symmetry group C_{3v} can be summarized as follows:

$$E\{1,2,3\} = \{1,2,3\} C_3\{1,2,3\} = \{2,3,1\} \tag{9.14.1}$$

$$C_3^2\{1,2,3\} = \{3,1,2\} \sigma_1\{1,2,3\} = \{1,3,2\} \tag{9.14.2}$$

$$\sigma_2\{1,2,3\} = \{3,2,1\} \sigma_3\{1,2,3\} = \{2,1,3\} \tag{9.14.3}$$

We have thus accounted for all 6 possible permutations of the three hydrogen atoms.

The successive application of two symmetry operations is equivalent to some single symmetry operation. For example, applying C_3, then σ_1 to our starting orientation, we have

$$\sigma_1 C_3\{1,2,3\} = \sigma_1\{2,3,1\} = \{2,1,3\} \tag{9.14.4}$$

But this is equivalent to the single operation σ_3. This can be represented as an algebraic relation among symmetry operators

$$\sigma_1 C_3 = \sigma_3 \tag{9.14.5}$$

Note that successive operations are applied in the order *right to left* when represented algebraically. For the same two operations in reversed order, we find

$$C_3\sigma_1\{1,2,3\} = C_3\{1,3,2\} = \{3,2,1\} = \sigma_2\{1,2,3\} \tag{9.14.6}$$

Thus symmetry operations do *not*, in general commute

$$AB \not\equiv BA \tag{9.14.7}$$

although they *may* commute, for example, C_3 and C_3^2.

The algebra of the group C_{3v} can be summarized by the following multiplication table.

1st / 2nd	E	C_3	C_3^2	σ_1	σ_2	σ_3
E	E	C_3	C_3^2	σ_1	σ_2	σ_3
C_3	C_3	C_3^2	E	σ_3	σ_1	σ_2
C_3^2	C_3^2	E	C_3	σ_2	σ_3	σ_1
σ_1	σ_1	σ_2	σ_3	E	C_3	C_3^2
σ_2	σ_2	σ_3	σ_1	C_3^2	E	C_3
σ_3	σ_3	σ_1	σ_2	C_3	C_3^2	E

(9.14.8)

Notice that each operation occurs once and only once in each row and each column.

Group Theory

In mathematics, a *group* is defined as a set of g elements $\mathcal{G} \equiv \{G_1, G_2 ... G_h\}$ together with a rule for combination of elements, which we usually refer to as a *product*. The elements must fulfill the following four conditions.

I. The product of any two elements of the group is another element of the group. That is $G_i G_j = G_k$ with $G_k^{\text{int}} \mathcal{G}$
II. Group multiplication obeys an associative law, $G_i(G_j G_k) = (G_i G_j)G_k \equiv G_i G_j G_k$
II. There exists an *identity element E* such that $EG_i = G_i E = G_i$ for all *i*.
IV. Every element G_i has a unique inverse G_i^{-1}, such that $G_i G_i^{-1} = G_i^{-1} G_i = E$ with Double exponent: use braces to clarify .

The number of elements *h* is called the *order* of the group. Thus C_{3v} is a group of order 6.

A set of quantities which obeys the group multiplication table is called a *representation* of the group. Because of the possible noncommutativity of group elements [cf. Eq (1)], simple numbers are not always adequate to represent groups; we must often use matrices. The group C_{3v} has three *irreducible representation*s, or IR's, which cannot be broken down into simpler representations. A trivial, but nonetheless important, representation of any group is the *totally symmetric representation*, in which each group element is represented by 1. The multiplication table then simply reiterates that $1 \times 1 = 1$. For C_{3v} this is called the A_1 representation:

$$A_1 : E = 1, C_3 = 1, C_3^2 = 1, \sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 1 \tag{9.14.9}$$

A slightly less trivial representation is A_2:

$$A_2 : E = 1, C_3 = 1, C_3^2 = 1, \sigma_1 = -1, \sigma_2 = -1, \sigma_3 = -1 \tag{9.14.10}$$

Much more exciting is the E representation, which requires 2×2 matrices:

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad C_3 = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$$
$$C_3^2 = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \qquad \sigma_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$
$$\sigma_2 = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \tag{9.14.11}$$

The operations C_3 and C_3^2 are said to belong to the same *class* since they perform the same geometric function, but for different orientations in space. Analogously, σ_1, σ_2 and σ_3 are obviously in the same class. E is in a class by itself. The class structure of the group is designated by $\{E, 2C_3, 3\sigma_v\}$. We state without proof that the number of irreducible representations of a group is equal to the number of classes. Another important theorem states that the sum of the squares of the dimensionalities of the irreducible representations of a group adds up to the order of the group. Thus, for C_{3v}, we find $1^2 + 1^2 + 2^2 = 6$.

The *trace* or *character* of a matrix is defined as the sum of the elements along the main diagonal:

$$\chi(M) \equiv \sum_k M_{kk} \tag{9.14.12}$$

For many purposes, it suffices to know just the characters of a matrix representation of a group, rather than the complete matrices. For example, the characters for the E representation of C_{3v} in Eq (4) are given by

$$\chi(E) = 2, \quad \chi(C_3) = -1, \quad \chi(C_3^2) = -1,$$
$$\chi(\sigma_1) = 0, \quad \chi(\sigma_2) = 0, \quad \chi(\sigma_3) = 0 \tag{9.14.13}$$

It is true in general that the characters for all operations in the same class are equal. Thus Eq (6) can be abbreviated to

$$\chi(E) = 2, \quad \chi(C_3) = -1, \quad \chi(\sigma_v) = 0 \tag{9.14.14}$$

For one-dimensional representations, such as A_1 and A_2, the characters are equal to the matrices themselves, so Equations 9.14.9 and 9.14.10 can be read as a table of characters.

The essential information about a symmetry group is summarized in its *character table*. We display here the character table for C_{3v}

C_{3v}	E	$2C_3$	$3\sigma_v$		
A_1	1	1	1	z	$z^2, x^2 + y^2$
A_2	1	1	-1		
E	2	-1	0	(x, y)	$(xy, x^2 - y^2), (xz, yz)$

(9.14.15)

The last two columns show how the cartesian coordinates x, y, z and their products transform under the operations of the group.

Group Theory and Quantum Mechanics

When a molecule has the symmetry of a group $\mathcal{G}$, this means that each member of the group commutes with the molecular Hamiltonian

$$[\hat{G}_i, \hat{H}] = 0 \quad i = 1...h \tag{9.14.16}$$

where we now explicitly designate the group elements G_i as operators on wavefunctions. As was shown in Chap. 4, commuting operators can have simultaneous eigenfunctions. A representation of the group of dimension d means that there must exist a set of d degenerate eigenfunctions of $\hat{H}$ that transform among themselves in accord with the corresponding matrix representation. For example, if the eigenvalue E_n is d-fold degenerate, the commutation conditions (Equation 9.14.9) imply that, for $i = 1...h$,

$$\hat{G}_i \hat{H} \psi_{nk} = \hat{H} \hat{G}_i \psi_{nk} = E_n \hat{G}_i \psi_{nk} \text{ for } k = 1...d \tag{9.14.17}$$

Thus each $\hat{G}_i \psi_{nk}$ is also an eigenfunction of $\hat{H}$ with the same eigenvalue E_n, and must therefore be represented as a linear combination of the eigenfunctions ψ_{nk}. More precisely, the eigenfunctions transform among themselves according to

$$\hat{G}_i\psi_{nk} = \sum_{m=1}^{d} D(G_i)_{km}\psi_{nm} \tag{9.14.18}$$

where $D(G_i)_{km}$ means the $\{k, m\}$ element of the matrix representing the operator $\hat{G}_i$.

The character of the identity operation E immediately shows the degeneracy of the eigenvalues of that symmetry. The C_{3v} character table reveals that NH_3, and other molecules of the same symmetry, can have only nondegenerate and two-fold degenerate energy levels. The following notation for symmetry species was introduced by Mulliken:

I. One dimensional representations are designated either A or B. Those symmetric wrt rotation by $2\pi/n$ about the C_n principal axis are labeled A, while those antisymmetric are labeled B.
II. Two dimensional representations are designated E; 3, 4 and 5 dimensional representations are designated T, F and G, respectively. These latter cases occur only in groups of high symmetry: cubic, octahedral and icosohedral.
II. In groups with a center of inversion, the subscripts g and u indicate even and odd parity, respectively.
IV. Subscripts 1 and 2 indicate symmetry and antisymmetry, respectively, wrt a C_2 axis perpendicular to C_n, or to a σ_v plane.
V. Primes and double primes indicate symmetry and antisymmetry to a σ_h plane.

For individual orbitals, the lower case analogs of the symmetry designations are used. For example, MO's in ammonia are classified a_1, a_2 or e.

For ammonia and other C_{3v} molecules, there exist three species of eigenfunctions. Those belonging to the classification A_1 are transformed into themselves by all symmetry operations of the group. The 1s, 2s and $2p_z$ AO's on nitrogen are in this category. The z-axis is taken as the 3-fold axis. There are no low-lying orbitals belonging to A_2. The nitrogen $2p_x$ and $2p_y$ AO's form a two-dimensional representation of the group C_{3v}. That is to say, any of the six operations of the group transforms either one of these AO's into a linear combination of the two, with coefficients given by the matrices (4). The three hydrogen 1s orbitals transform like a 3×3 representation of the group. If we represent the hydrogens by a column vector {H1,H2,H3}, then the six group operations generate the following algebra

$$\begin{aligned} E &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad C_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \\ C_3^2 &= \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \sigma_2 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{9.14.19}$$

Let us denote this representation by Γ. It can be shown that Γ is a *reducible* representation, meaning that by some unitary transformation the 3×3 matrices can be factorized into blockdiagonal form with 2×2 plus 1×1 submatrices. The reducibility of Γ can be deduced from the character table. The characters of the matrices (Equation 9.14.19) are

$$\Gamma: \qquad \chi(E) = 3, \quad \chi(C_3) = 0, \quad \chi_(\sigma_v) = 1 \tag{9.14.20}$$

The character of each of these permutation operations is equal to the number of H atoms left untouched: 3 for the identity, 1 for a reflection and 0 for a rotation. The characters of Γ are seen to equal the sum of the characters of A_1 plus E. This reducibility relation is expressed by writing

$$\Gamma = A_1 \oplus E \tag{9.14.21}$$

The three H atom 1s functions can be combined into LCAO functions which transform according to the IR's of the group. Clearly the sum

$$\psi = \psi_{1s}(1) + \psi_{1s}(2) + \psi_{1s}(3) \tag{9.14.22}$$

transforms like A_1. The two remaining linear combinations which transform like E must be orthogonal to (Equation 9.14.22) and to one another. One possible choice is

$$\psi' = \psi_{1s}(2) - \psi_{1s}(3), \quad \psi'' = 2\psi_{1s}(1) - \psi_{1s}(2) - \psi_{1s}(3) \tag{9.14.23}$$

Now, Equation 9.14.22 can be combined with the N 1s, 2s and $2p_z$ to form MO's of A_1 symmetry, while Equation 9.14.23 can be combined with the N $2p_x$ and $2p_y$ to form MO's of E symmetry. Note that no hybridization of AO's is predetermined, it emerges automatically in the results of computation.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)

This page titled 9.14: Molecular Symmetry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- 11: Molecular Symmetry is licensed CC BY-NC-SA 4.0.

9.15: Molecular Spectroscopy

Our most detailed knowledge of atomic and molecular structure has been obtained from spectroscopy-study of the emission, absorption and scattering of electromagnetic radiation accompanying transitions among atomic or molecular energy levels. Whereas atomic spectra involve only electronic transitions, the spectroscopy of molecules is more intricate because vibrational and rotational degrees of freedom come into play as well. Early observations of absorption or emission by molecules were characterized as band spectra-in contrast to the line spectra exhibited by atoms. It is now understood that these bands reflect closely-spaced vibrational and rotational energies augmenting the electronic states of a molecule. With improvements in spectroscopic techniques over the years, it has become possible to resolve individual vibrational and rotational transitions. This has provided a rich source of information on molecular geometry, energetics and dynamics. Molecular spectroscopy has also contributed significantly to analytical chemistry, environmental science, astrophysics, biophysics and biochemistry.

Reduced Mass

Consider a system of two particles of masses m_1 and m_2 interacting with a potential energy which depends only on the separation of the particles. The classical energy is given by

$$E = \frac{1}{2}m_1\dot{\vec{r}}_1^2 + \frac{1}{2}m_2\dot{\vec{r}}_2^2 + V(|\vec{r}_2 - \vec{r}_1|) \tag{9.15.1}$$

the dots signifying derivative wrt time. Introduce two new variables, the particle separation $\vec{r}$ and the position of the center of mass $\vec{R}$:

$$\vec{r} = \vec{r}_2 - \vec{r}_1, \qquad \vec{R} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{m} \tag{9.15.2}$$

where $m = m_1 + m_2$. In terms of the new coordinates

$$\vec{r}_1 = \vec{R} + \frac{m_2}{m}\vec{r}, \qquad \vec{r}_2 = \vec{R} - \frac{m_1}{m}\vec{r} \tag{9.15.3}$$

and

$$E = \frac{1}{2}m\dot{\vec{R}}^2 + \frac{1}{2}\mu\dot{\vec{r}}^2 + V(r) \tag{9.15.4}$$

where $r = |\vec{r}|$ and μ is called the $reduced\ mass$

$$\mu \equiv \frac{m_1 m_2}{m_1 + m_2} \tag{9.15.5}$$

An alternative relation for reduced mass is

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \tag{9.15.6}$$

reminiscent of the formula for resistance of a parallel circuit. Note that, if $m_2 \to \infty$, then $\mu \to m_1$. The term containing $\dot{\vec{R}}$ represents the kinetic energy of a single hypothetical particle of mass μ located at the center of mass $\vec{R}$. The remaining terms represent the $relative$ motion of the two particles. It was the appearance of a $single$ particle of effective mass μ moving in the potential field $V(r)$.

$$E_{rel} = \frac{1}{2}\mu\dot{\vec{r}}^2 + V(r) = \frac{\vec{p}^{\,2}}{2\mu} + V(r) \tag{9.15.7}$$

We can thus write the Schrödinger equation for the relative motion

$$\left\{-\frac{\hbar^2}{2\mu}\bigtriangledown^2 + V(r)\right\}\psi(\vec{r}) = E\psi(\vec{r}) \tag{9.15.8}$$

When we treated the hydrogen atom, it was assumed that the nuclear mass was infinite. In that case we can set $\mu = m$, the mass of an electron. The Rydberg constant for infinite nuclear mass was calculated to be

$$R_\infty = \frac{2\pi^2 me^4}{h^3 c} = 109,737\text{cm}^{-1} \tag{9.15.9}$$

If instead, we use the reduced mass of the electron-proton system

$$\mu = \frac{mM}{m+M} \approx \frac{1836}{1837}\, m \approx 0.999456\, m \tag{9.15.10}$$

.

This changes the Rydberg constant for hydrogen to

$$R_H \approx 109,677\, cm^{-1} \tag{9.15.11}$$

in perfect agreement with experiment.

In 1931, H. C. Urey evaporated four liters of hydrogen down to one milliliter and measured the spectrum of the residue. The result was a set of lines displaced slightly from the hydrogen spectrum. This amounted to the discovery of deuterium, or heavy hydrogen, for which Urey was awarded in 1934 Nobel Prize in Chemistry. Estimating the mass of the deuteron, $^2H_1$, as twice that of the proton, gives

$$R_D \approx 109,707\, cm^{-1} \tag{9.15.12}$$

Another interesting example involving reduced mass concerns positronium, a short-lived combination of an electron and a positron-the electron's antiparticle. The electron and position mutually annihilate with a half-life of approximately 10^{-7} sec. and positroium decays into gamma rays. The reduced mass of positronium is

$$\mu = \frac{m \times m}{m+m} = \frac{m}{2} \tag{9.15.13}$$

half the mass of the electron. Thus the ionization energy equals 6.80 eV, half that of hydrogen atom. Positron emission tomography (PET) provides a sensitive scanning technique for functioning living tissue, notably the brain. A compound containing a positron-emitting radionuclide, for example, ^{11}C, ^{13}N, ^{15}O or ^{18}F, is injected into the body. The emitted positrons attach to electrons to form short-lived positronium, and the annihilation radiation is monitored.

Vibration of Diatomic Molecules

A diatomic molecule with nuclear masses M_A, M_B has a reduced mass

$$\mu = \frac{M_A M_B}{M_A + M_B} \tag{9.15.14}$$



Figure 1. Jens Zorn sculpture depicting positronium annihilation. Outside University of Michigan Physics Building.

Solution of the electronic Schrödinger equation gives the energy as a function of internuclear distance $E_{elec}(R)$. This plays the role of a potential energy function for motion of the nuclei V(R), as sketched in Fig. 2. We can thus write the Schrödinger equation for vibration

$$\left\{-\frac{\hbar^2}{2\mu}\frac{d^2}{dR^2}+V(R)\right\}\chi(R)=E_\chi(R) \tag{9.15.15}$$

If the potential energy is expanded in a Taylor series about R = R_e

$$V(R)=V(R_e)+(R-R_e)V'(R_e)+\frac{1}{2}(R-R_e)^2V''(R_e)+... \tag{9.15.16}$$

An approximation for this expansion has the form of a harmonic oscillator with

$$V(R)\approx\frac{1}{2}k(R-R_e)^2 \tag{9.15.17}$$

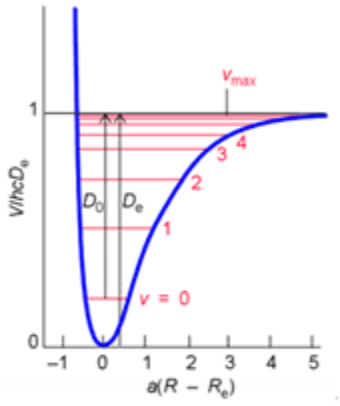


Figure 2. Vibrational energies of a diatomic molecule, as approximated by a Morse oscillator.

The energy origin can be chosen so V (R_e) = 0. At the minimum of the potential, V'(R_e) = 0. The best fit to the parabola (17) is obtained with a force constant set equal to

$$k\approx\frac{d^2V(R)}{dR^2}\,|_{R=R_e} \tag{9.15.18}$$

From the solution for the harmonic oscillator, we identify the ground state vibrational energy, with quantum number ν = 0

$$E_0 = \hbar\omega = \hbar\sqrt{\frac{k}{\mu}} \tag{9.15.19}$$

The actual dissociation energy from the ground vibrational state is then approximated by

$$D_0 \approx D_e - \frac{1}{2}\hbar\omega \tag{9.15.20}$$

In wavenumber units

$$hcD_0 \approx hcD_e - \frac{1}{2}\tilde{\nu}\, cm^{-1} \tag{9.15.21}$$

An improved treatment of molecular vibration must account for *anharmonicity*, deviation from a harmonic oscillator. Anharmonicity results in a finite number of vibrational energy levels and the possibility of dissociation of the molecule at sufficiently high energy. A very successful approximation for the energy of a diatomic molecule is the Morse potential:

$$V(R) = hcD_e\left\{1 - e^{a(R-R_e)}\right\}^2 \quad a = \left(\frac{\mu\omega^2}{2hcD_e}\right)^{\frac{1}{2}} \tag{9.15.22}$$

Note that V (R_e) = 0 at the minimum of the potential well. The Schrödinger equation for a Morse oscillator can be solved to give the energy levels

$$E_\upsilon = (\upsilon + \frac{1}{2})\hbar\omega - (\upsilon + \frac{1}{2})^2\hbar\omega x_e \tag{9.15.23}$$

or, expressed in wavenumber units,

$$hcE_\upsilon = (\upsilon + \frac{1}{2})\tilde{\nu} - (\upsilon + \frac{1}{2})^2 x_e\tilde{\nu} \tag{9.15.24}$$

Higher vibrational energy levels are spaced closer together, just as in real molecules. Vibrational transitions of diatomic molecules occur in the infrared, roughly in the range of 50-12,000 cm^{-1}. A molecule will absorb or emit radiation only if it has a non-zero dipole moment. Thus HCl is *infrared active* while H_2 and Cl_2 are not.

Vibration of Polyatomic Molecules

A molecule with N atoms has a total of 3N degrees of freedom for its nuclear motions, since each nucleus can be independently displaced in three perpendicular directions. Three of these degrees of freedom correspond to translational motion of the center of mass. For a nonlinear molecule, three more degrees of freedom determine the orientation of the molecule in space, and thus its rotational motion. This leaves 3N - 6 vibrational modes. For a linear molecule, there are just two rotational degrees of freedom, which leaves 3N - 5 vibrational modes. For example, the nonlinear molecule H_2O has three vibrational modes while the linear molecule CO_2 has four vibrational modes. The vibrations consist of coordinated motions of several atoms in such a way as to keep the center of mass stationary and nonrotating. These are called the *normal modes*. Each normal mode has a characteristic resonance frequency $\tilde{\nu}_i$, which is usually determined experimentally. To a reasonable approximation, each normal mode behaves as an independent harmonic oscillator of frequency $\tilde{\nu}_i$. The normal modes of H_2O and CO_2 are pictured below.

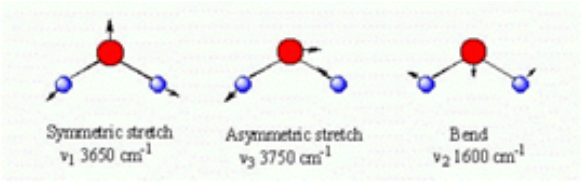


Figure 3. Normal modes of H_2O.

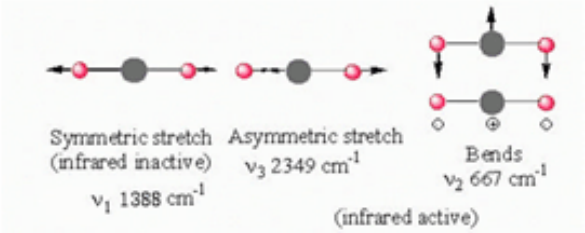


Figure 4. Normal modes of CO_2.

A normal mode will be infrared active only if it involves a change in the dipole moment. All three modes of H_2O are active. The symmetric stretch of CO_2 is inactive because the two C-O bonds, each of which is polar, exactly compensate. Note that the bending mode of CO_2 is doubly degenerate. Bending of adjacent bonds in a molecule generally involves less energy than bond stretching, thus bending modes generally have lower wavenumbers than stretching modes.

Rotation of Diatomic Molecules

The rigid rotor model assumes that the internuclear distance *R* is a constant. This is not a bad approximation since the amplitude of vibration is generally of the order of 1% of *R*. The Schrödinger equation for nuclear motion then involves the three-dimensional angular momentum operator, written $\hat{J}$ rather than $\hat{L}$ when it refers to molecular rotation. The solutions to this equation are already known and we can write

$$\frac{\hat{J}^2}{2\mu R^2} Y_{JM}(\theta,\phi) = E_J Y_{JM}(\theta,\phi) \quad J = 0, 1, 2... \quad M = 0, \pm 1... \pm J \tag{9.15.25}$$

where $Y_{JM}(\theta,\phi)$ are spherical harmonics in terms of the quantum numbers *J* and *M*, rather than *l* and *m*. Since the eigenvalues of $\hat{J}^2$ are $J(J+1)\hbar^2$, the rotational energy levels are

$$E_J = \frac{\hbar^2}{2I} J(J+1) \tag{9.15.26}$$

The moment of inertia is given by

$$I = \mu R^2 = M_A R_A^2 + M_B R_B^2 \tag{9.15.27}$$

where R_A and R_B are the distances from nuclei A and B, respectively, to the center of mass. In wavenumber units, the rotational energy is expressed

$$hcE_J = BJ(J+1)cm^{-1} \tag{9.15.28}$$

where *B* is the rotational constant. The rotational energy-level diagram is shown in Fig.5. Each level is (2J + 1)-fold degenerate. Again, only polar molecules can absorb or emit radiation in the course of rotational transitions. The radiation is in the microwave or far infrared region. The selection rule for rotational transitions is $\Delta J = \pm 1$.

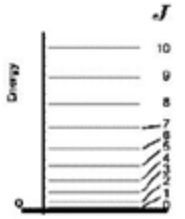


Figure 5. Rotational Energies $hcE_J = B\hat{J}(J+1)$

Molecular Parameters from Spectroscopy

Following is a table of spectroscopic constants for the four hydrogen halides:

	$\tilde{\nu}/\mathrm{cm}^{-1}$	B/cm^{-1}
$^1H^{19}F$	4138.32	20.956
$^1H^{35}Cl$	2990.95	10.593
$^1H^{81}Br$	2648.98	8.465
$^1H^{127}I$	2308.09	6.511

The force constant can be found from the vibrational constant. Equating the energy quantities $\hbar\omega = hc\tilde{\nu}$, we find

$$\omega = 2\pi c\tilde{\nu} = \sqrt{\frac{k}{\mu}} \tag{9.15.29}$$

Thus

$$k = (2\pi c\tilde{\nu})^2\mu \tag{9.15.30}$$

with

$$\mu = \frac{m_A m_B}{m_A + m_B} = \frac{M_A M_B}{M_A + M_B}u \tag{9.15.31}$$

where u = 1.66054 x 10^{-27} kg, the atomic mass unit. M_A and M_B are the conventional atomic weights of atoms A and B (on the scale ^{12}C = 12). Putting in numerical factors

$$k = 58.9 \times 10^{-6}(\tilde{\nu}/cm^{-1})^2\frac{M_A M_B}{M_A + M_B}N/m \tag{9.15.32}$$

This gives 958.6, 512.4, 408.4 and 311.4 N/m for HF, HCl, HBr and HI, respectively. These values do *not* take account of anharmonicity.

The internuclear distance *R* is determined by the rotational constant. By definition,

$$hcB = \frac{\hbar^2}{2I} \tag{9.15.33}$$

Thus

$$B = \frac{\hbar}{4\pi cI} \tag{9.15.34}$$

with

$$I = \mu R^2 = \frac{m_A m_B}{m_A + m_B}R^2 = \frac{M_A M_B}{M_A + M_B}uR^2 \;\; kg\,m^2 \tag{9.15.35}$$

Solving for R,

$$R = 410.6/\sqrt{\frac{M_A M_B}{M_A + M_B}(B/cm^{-1})}\ pm \tag{9.15.36}$$

For the hydrogen halides, HF, HCl, HBr, HI, we calculate *R* = 92.0, 127.9, 142.0, 161.5 pm, respectively.

Rotation of Nonlinear Molecules

A nonlinear molecule has three moments of inertia about three principal axes, designated I_a, I_b and I_c. The classical rotational energy can be written

$$E = \frac{J_a^2}{2I_a} + \frac{J_b^2}{2I_b} + \frac{J_c^2}{2I_c} \tag{9.15.37}$$

where J_a, J_b, J_c are the components of angular momentum about the principal axes. For a spherical rotor, such as CH_4 or SF_6, the three moments of inertia are equal to the same value *I*. The energy simplifies to $J^2/2I$ and the quantum-mechanical Hamiltonian is given by

$$\hat{H} = \frac{\hat{J}^2}{2I} \tag{9.15.38}$$

The eigenvalues are

$$E_J = \frac{\hbar^2}{2I}J(J+1) \quad J = 0, 1, 2.. \tag{9.15.39}$$

just as for a linear molecule. But the levels of a spherical rotor have degeneraciesof $(2J + 1)^2$ rather than $(2J + 1)$.

A symmetric rotor has two equal moments of inertia, say $I_c = I_b \neq I_a$. The molecules NH_3, CH_3Cl and C_6H_6 are examples. The Hamiltonian takes the form

$$\hat{H} = \frac{\hat{J}_a^2}{2I_a} + \frac{\hat{J}_b^2 + \hat{J}_c^2}{2I_b} = \frac{\hat{J}^2}{2I_b} + (\frac{1}{2I_a} - \frac{1}{2I_b})\hat{J}_a^2 \tag{9.15.40}$$

Since it is possible to have simultaneous eigenstates of $\hat{J}^2$ and one of its components $\hat{J}_a$, the energies of a symmetric rotor have the form

$$E_{JK} = \frac{J(J+1)}{2I_b} + (\frac{1}{2I_a} - \frac{1}{2I_b})K^2 \quad J = 0, 1, 2... \quad K = 0 \pm 1, \pm 2, ... \pm J \tag{9.15.41}$$

There is, in addition, the $(2J + 1)$-fold *M* degeneracy.

Electronic Excitations in Diatomic Molecules

The quantum states of molecules are composites of rotational, vibrational and electronic contributions. The energy spacing characteristic of these different degrees of freedom vary over many orders of magnitude, giving rise to very different spectroscopic techniques for studying rotational, vibrational and electronic transitions. Electronic excitations are typically of the order of several electron volts, 1 eV being equivalent to approximately 8000 cm^{-1} or 100 kJ mol^{-1}. As we have seen, typical energy differences are of the order of 1000 cm^{-1} for vibration and 10 cm^{-1} for rotation. Fig. 6 gives a general idea of the relative magnitudes of these energy contributions. Each electronic state has a vibrational structure, characterized by vibrational quantum numbers v and each vibrational state has a rotational structure, characterized by rotational quantum numbers *J* and *M*.

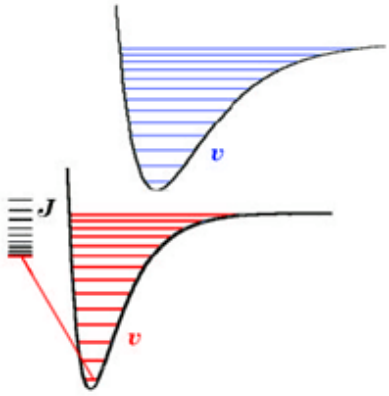


Figure 6. Schematic representation of the electronic ground state and an excited state of a diatomic molecule. Vibrational levels of the ground state are shown in red, those of the excited state, in blue. The rotational levels for $v = 0$ are also shown.

Every electronic transition in a molecule is accompanied by changes in vibrational and rotational states. Generally, in the liquid state, individual vibrational transitions are not resolved, so that electronic spectra consist of broad bands comprising a large number of overlapping vibrational and rotational transitions. Spectroscopy on the gas phase, however, can often resolve individual vibrational and even rotational transitions.

When a molecule undergoes a transition to a different electronic state, the electrons rearrange themselves much more rapidly than the nuclei. To a very good approximation, the electronic state can be considered to occur instantaneously, while the nuclear configuration remains fixed. This is known as the *Franck-Condon principle*. It has the same physical origin as the Born-Oppenheimer approximation, namely the great disparity in the electron and nuclear masses. On a diagram showing the energies of the ground and excited states as functions of internuclear distance, Franck-Condon behavior is characterized by *vertical transitions*, in which *R* remains approximately constant as the molecule jumps from one potential curve to the other.

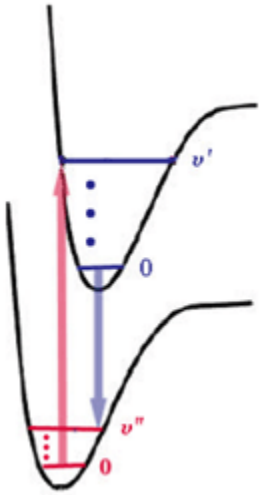


Figure 7. Franck-Condon principle

In a vibrational state υ = 0 the maximum of probability for the internuclear distance R is near the center of the potential well. For all higher values vibrational states, maxima of probability occur near the two turning points of the potential-where the total energy equals the potential energy. These correspond on the diagrams to the end points of the horizontal dashes inside the potential curve.

Contributors and Attributions

Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)

This page titled 9.15: Molecular Spectroscopy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- 16: Molecular Spectroscopy is licensed CC BY-NC-SA 4.0.

9.16: Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a versatile and highly-sophisticated spectroscopic technique which has been applied to a growing number of diverse applications in science, technology and medicine. This chapter will consider, for the most part, magnetic resonance involving protons.

Magnetic Properties of Nuclei

In all our previous work, it has been sufficient to treat nuclei as structureless point particles characterized fully by their mass and electric charge. On a more fundamental level, as was discussed in Chap. 1, nuclei are actually composite particles made of nucleons (protons and neutrons) and the nucleons themselves are made of quarks. The additional properties of nuclei which will now become relevant are their spin angular momenta and magnetic moments. Recall that electrons possess an intrinsic or spin angular momentum s which can have just two possible projections along an arbitrary spacial direction, namely $\pm\frac{1}{2}\hbar$. Since $\hbar$ is the fundamental quantum unit of angular momentum, the electron is classified as a particle of spin one-half. The electron's spin state is described by the quantum numbers $s = 1$ and $m_s = \pm 1$. A circulating electric charge produces a magnetic moment $\vec{\mu}$ proportional to the angular momentum **J**. Thus

$$\vec{\mu} = \gamma \vec{J} \tag{9.16.1}$$

where the constant of proportionality γ is known as the magnetogyric ratio. The z-component of $\vec{\mu}$ has the possible values

$$\vec{\mu}_z = \gamma \hbar m_J \text{where} m_J = -J, -J+1, ..., +J \tag{9.16.2}$$

determined by space quantization of the angular momentum J. The energy of a magnetic dipole in a magnetic field B is given by

$$E = -\vec{\mu} \cdot \vec{B} = -\vec{\mu}_z B \tag{9.16.3}$$

where magnetic field defines the z-axis. The SI unit of magnetic field (more correctly, magnetic induction) is the tesla, designated T. Electromagnets used in NMR produce fields in excess of 10 T. Small iron magnets have fields around .01 T, while some magnets containing rare-earth elements such as NIB (niobium-iron-boron) reach 0.2 T. The Earth's magnetic field is approximately 5×10^{-5}T (0.5 gauss in alternative units), dependent on geographic location. At the other extreme, a neutron star, which is really a giant nucleus, has a field predicted to be of the order of 10^8 T. The energy relation (3) determines the most conveniently units for magnetic moment, namely joules per tesla, J T^{-1}.

For orbital motion of an electron, where the angular momentum is l, the magnetic moment is given by

$$\vec{\mu}_z = -\frac{e\hbar}{2m} m_l = -\vec{\mu}_B m_l \tag{9.16.4}$$

where the minus sign reflects the *negative* electric charge. The Bohr *magneton* is defined by

$$\vec{\mu}_B = -\frac{e\hbar}{2m} = 9.274 \times 10^{-24} JT^{-1} \tag{9.16.5}$$

The magnetic moment produced by electron spin is written

$$\vec{\mu}_z = -g\vec{\mu}_B m_s \tag{9.16.6}$$

with introduction of the g-factor. Eq (4) implies g = 1 for orbital motion. For electron spin, however, g = 2 (more exactly, 2.0023). The factor 2 compensates for ms = 1 such that spin and l = 1 orbital magnetic moments 2 are both equal to one Bohr magneton.

Many nuclei possess spin angular momentum, analogous to that of the electron. The nuclear spin, designated I, has an integral or half-integral value: 0, 1 , 1, 3 , and so on. **Table 1** lists some nuclei of importance in chemical applications of NMR. The proton and the neutron both are spin 2 particles, like the electron. Complex nuclei have angular momenta which are resultants of the spins of their component nucleons. The deuteron ^{2}H, with I = 1, evidently has parallel proton and neutron spins. The ^{4}He nucleus has I = 0, as do ^{12}C, ^{16}O, ^{20}Ne, ^{28}Si and ^{32}S. These nuclei contain filled shells of protons and neutrons with the vector sum of the

component angular momenta equal to zero, analogous to closed shells of electrons in atoms and molecules. In fact, all even-even nuclei have spins of zero. Nuclear magnetic moments are of the order of a *nuclear magneton*

$$\vec{\mu}_N = \frac{e\hbar}{2M} = 5.051 \times 10^{-27} JT^{-1} \tag{9.16.7}$$

where M is the mass of the proton. The nuclear magneton is smaller than the Bohr magneton by a factor m/M ≈ 1836.

Table 1: Some common nuclei in NMR spectroscopy

Nuclide	I	g_I	μ/μ_N	$\gamma_I/10^7$	Abundance %
^1_0n	1/2	-3.8260	-1.9130	-18.324	
$^1_1$H	1/2	5.5857	2.7928	26.752	99.98
$^2_1$H	1	0.8574	0.8574	4.1067	0.0156
$^{11}_5$B	3/2	1.7923	2.6886	8.5841	80.4
$^{13}_6$C	1/2	1.4046	0.7023	6.7272	1.1
$^{14}_7$N	1	0.4036	0.4036	1.9328	99.635
$^{17}_8$O	5/2	-0.7572	-1.893	-3.627	0.037
$^{19}_9$F	1/2	5.2567	2.628	25.177	100
$^{31}_{15}$P	1/2	2.2634	1.2317	10.840	100

In analogy with Equations 9.16.2 and 9.16.6, nuclear moments are represented by

$$\vec{\mu}_z = g_I \vec{\mu}_N m_I = \hbar \gamma_I m_I \tag{9.16.8}$$

where g_I is the nuclear g-factor and γ_I, the magnetogyric ratio. Most nuclei have positive g-factors, as would be expected for a rotating positive electric charge. It was long puzzling that the neutron, although lacking electric charge, has a magnetic moment. It is now understood that the neutron is a composite of three charged quarks, *udd*. The negatively-charged d- quarks are predominantly in the outermost regions of the neutron, thereby producing a negative magnetic moment, like that of the electron. The g- factor for ^{17}O, and other nuclei dominated by unpaired neutron spins, is consequently also negative.

Nuclear Magnetic Resonance

The energy of a nuclear moment in a magnetic field, according to Equation 9.16.3, is given by

$$E_{m_I} = -\hbar \gamma_I m_I B \tag{9.16.9}$$

For a nucleus of spin I, the energy of a nucleus in a magnetic field is split into $2I + 1$ Zeeman levels. A proton and other nuclei with spin $\frac{1}{2}$ particles will have just two possible levels:

$$E_{\pm\frac{1}{2}} = \pm\frac{1}{2}\hbar\gamma B \tag{9.16.10}$$

with the α-spin state ($m_I = -\frac{1}{2}$) lower in energy than the β-spin state ($m_I = +\frac{1}{2}$) by

$$\Delta E = \hbar \gamma B \tag{9.16.11}$$

Fig. 1 shows the energy of a proton as a function of magnetic field. In zero field (B = 0), the two spin states are degenerate. In a field B, the energy splitting corresponds to a photon of energy $\Delta E = \hbar\omega = h\nu$ where

$$\omega_L = \gamma_B \text{ or } \nu_L = \gamma_B \tag{9.16.12}$$

known as the Larmor frequency of the nucleus. For the proton in a field of 1 T, ν_L = 42.576 MHz, as the proton spin orientation flips from $+\frac{1}{2}$ to $-\frac{1}{2}$. This transition is in the radiofrequency region of the electromagnetic spectrum. NMR spectroscopy consequently

exploits the technology of radiowave engineering.

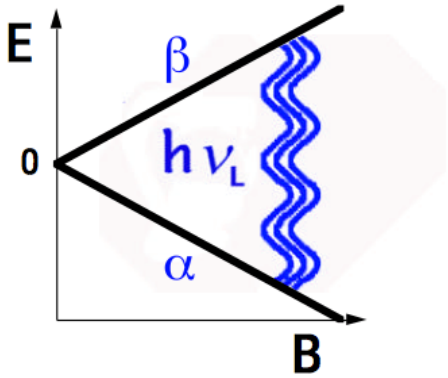


Figure 1. *Energies of spin $\frac{1}{2}$ in magnetic field showing NMR transition at Larmor frequency ν_L.*

A transition cannot occur unless the values of the radiofrequency and the magnetic field accurately fulfill Eq (12). This is why the technique is categorized as a resonance phenomenon. If some resonance condition is not satisfied, no radiation can be absorbed or emitted by the nuclear spins. In the earlier techniques of NMR spectroscopy, it was found more convenient keep the radiofrequency fixed and sweep over values of the magnetic field B to detect resonances. These have been largely supplanted by modern pulse techniques, to be described later.

The transition probability for the upward transition (absorption) is equal to that for the downward transition (stimulated emission). (The contribution of spontaneous emission is neglible at radiofrequencies.) Thus if there were equal populations of nuclei in the α and β spin states, there would be zero net absorption by a macroscopic sample. The possibility of observable NMR absorption depends on the lower state having at least a slight excess in population. At thermal equilibrium, the ratio of populations follows a Boltzmann distribution

$$\frac{N_\beta}{N_\alpha} = \frac{e^{\frac{-E_\beta}{kT}}}{e^{\frac{-E_\alpha}{kt}}} = e^{-\frac{\hbar\gamma B}{kT}} \tag{9.16.13}$$

Thus the relative population difference is given by

$$\frac{\Delta N}{N_\alpha} = \frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} \approx \frac{\hbar\gamma B}{2kT} \tag{9.16.14}$$

Since nuclear Zeeman energies are so small, the populations of the α and β spin states differ very slightly. For protons in a 1 T field, ΔN/N ≈ 3×10^{-6}. Although the population excess in the lower level is only of the order of parts per million, NMR spectroscopy is capable of detecting these weak signals. Higher magnetic fields and lower temperatures are favorable conditions for enhanced NMR sensitivity.

The Chemical Shift

NMR has become such an invaluable technique for studying the structure of atoms and molecules because nuclei represent ideal noninvasive probes of their electronic environment. If all nuclei of a given species responded at their characteristic Larmor frequencies, NMR might then be useful for chemical analysis, but little else. The real value of NMR to chemistry comes from minute differences in resonance frequencies dependent on details of the electronic structure around a nucleus. The magnetic field induces orbital angular momentum in the electron cloud around a nucleus, thus, in effect, partially shielding the nucleus from the external field B. The actual or local value of the magnetic field at the position of a nucleus is expressed

$$B_{loc} = (1-\sigma)B \tag{9.16.15}$$

where the fractional reduction of the field is denoted by σ, the shielding constant, typically of the order of parts per million. The actual resonance frequency of the nucleus in its local environment is then equal to

$$\nu = (1 - \sigma)\frac{\gamma B}{2\pi} \tag{9.16.16}$$

A classic example of this effect is the proton NMR spectrum of ethanol CH_3CH_2OH, shown in Fig. 2. The three peaks, with intensity ratios 3:2:1 can be identified with the three chemically-distinct environments in which the protons find themselves: three methyl protons (CH_3), two methylene protons (CH_2) and one hydroxyl proton (OH).

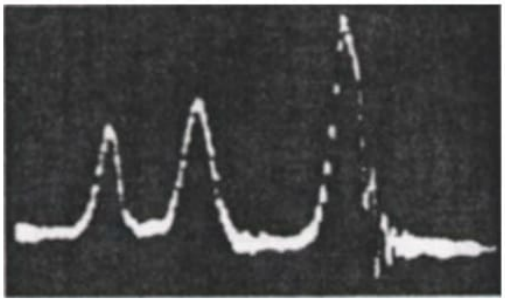


Figure 2. *Oscilloscope trace showing the first NMR spectrum of ethanol, taken at Stanford University in 1951. Courtesy Varian Associates, Inc.*

The variation in resonance frequency due to the electronic environment of a nucleus is called the chemical shift. Chemical shifts on the delta scale are defined by

$$\delta = \frac{\nu - \nu^O}{\nu^O} \times 10^6 \tag{9.16.17}$$

where ν_o represents the resonance frequency of a reference compound, usually tetramethylsilane $Si(CH_3)_4$, which is rich in highly-shielded chemically-equivalent protons, as well as being unreactive and soluble in many liquids. By definition δ = 0 for TMS and almost everything else is "downfield" with positive values of δ. Most compounds have delta values in the range of 0 to 12 (hydrogen halides have negative values, e.g. δ ≈ –13 for HI). The hydrogen atom has δ ≈ 13 while the bare proton would have δ ≈ 31. Conventionally, the δ-scale is plotted as increasing from right to left, in the opposite sense to the magnitude of the magnetic field. Nuclei with larger values of δ are said to be more deshielded, with the bare proton being the ultimate limit. Fig. 3 shows some typical values of δ for protons in some common organic compounds.

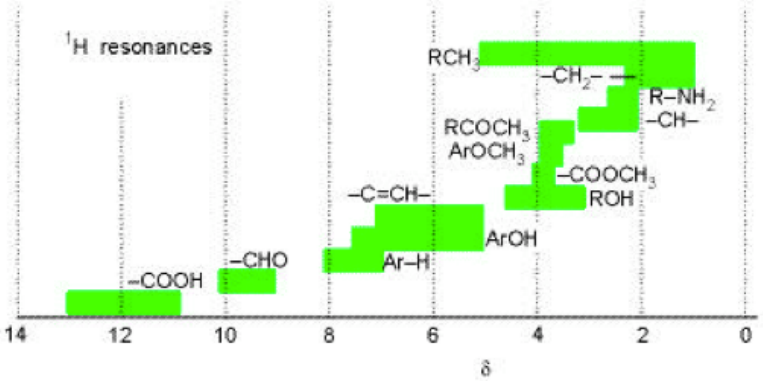


Figure 3. *Ranges of proton chemical shifts for common functional groups. From P. Atkins, Physical Chemistry, (Freeman, New York, 2002).*

Fig. 4 shows a high-resolution NMR spectrum of ethanol, including a δ-scale. The "fine structure" splittings of the three chemically-shifted components will be explained in the next Section. The chemical shift of a nucleus is very difficult to treat theoretically. However, certain empirical regularities, for example those represented in Fig. 3, provide clues about the chemical environment of the nucleus. We will not consider these in any detail except to remark that often increased deshielding of a nucleus (larger δ) can often be attributed to a more electronegative neighboring atom. For example the proton in the ethanol spectrum (Fig. 4) with δ ≈ 5 can be identified as the hydroxyl proton, since the oxygen atom can draw significant electron density from around the proton.

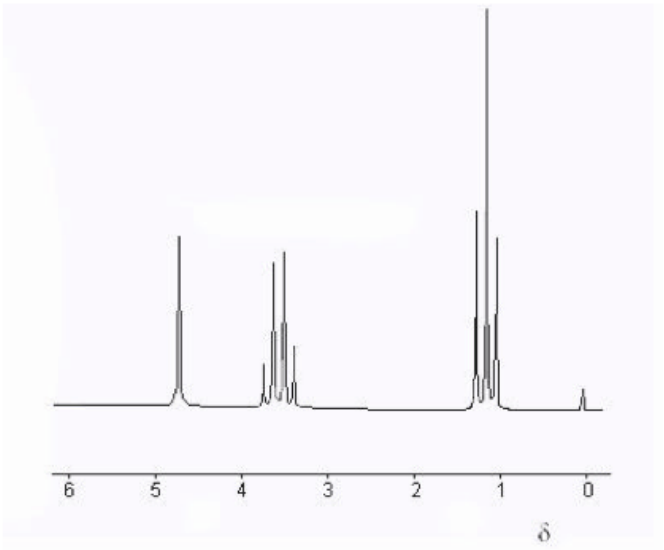


Figure 4. *High-resolution NMR spectrum of ethanol showing δ scale of chemical shifts. The line at δ = 0 corresponds to the TMS trace added as a reference.*

Neighboring groups can also contribute to the chemical shift of a given atom, particularly those with mobile π-electrons. For example, the ring current in a benzene ring acts as a secondary source of magnetic field. Depending on the location of a nucleus, this can contribute either shielding or deshielding of the external magnetic field, as shown in Fig. 5. The interaction of neighboring groups can be exploited to obtain structural information by using lanthanide shift reagents. Lanthanides (elements 58 through 71) contain 4f-electrons, which are not usually involved in chemical bonding and can give large paramagnetic contributions. Lanthanide complexes which bind to organic molecules can thereby spread out pro- ton resonances to simplify their analysis. A popular chelating complex is Eu(dpm)3, tris(dipivaloylmethanato)europium, where dpm is the group $(CH_3)_3C–CO=CH–CO–C(CH_3)_3$.

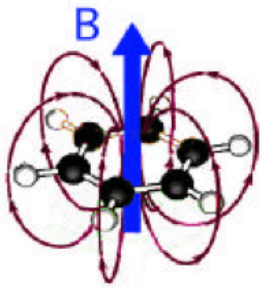
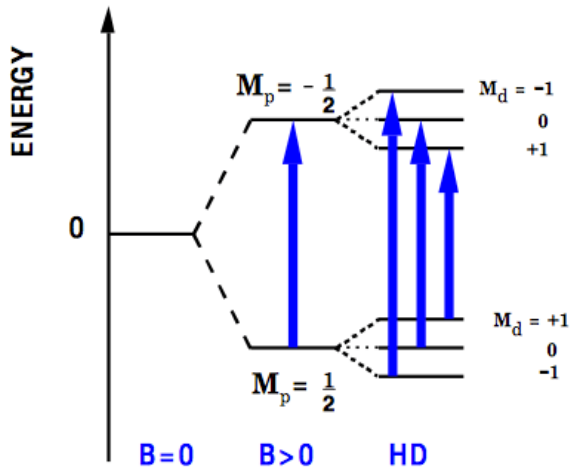


Figure 5. Magnetic field produced by ring current in benzene, shown as red loops. Where the arrows are parallel to the external field B, including protons directly attached to the ring, the effect is deshielding. However, any nuclei located within the return loops will experience a shielding effect.

Spin-Spin Coupling

Two of the resonances in the ethanol spectrum shown in Fig. 4 are split into closely-spaced multiplets—one triplet and one quartet. These are the result of spin-spin coupling between magnetic nuclei which are relatively close to one another, say within two or three bond separations. Identical nuclei in identical chemical environments are said to be equivalent. They have equal chemical shifts and do not exhibit spin-spin splitting. Nonequivalent magnetic nuclei, on the other hand, can interact and thereby affect one another's NMR frequencies. A simple example is the HD molecule, in which the spin-1proton can interact with the spin-1 deuteron, even though the atoms are chemically equivalent. The proton's energy is split into two levels by the external magnetic field, as shown in Fig. 1. The neighboring deuteron, itself a magnet, will also contribute to the local field at the pro- ton. The deuteron's three possible orientations in the external field, with MI = –1,0,+1, with different contributions to the magnetic

field at the proton, as shown in Fig. 6. The proton's resonance is split into three evenly spaced, equally intense lines (a triplet), with a separation of 42.9 Hz. Corespondingly the deuteron's resonance is split into a 42.9 Hz doublet by its interaction with the proton. These splittings are independent of the external field B, whereas chemical shifts are proportional to B. Fig. 6 represents the energy levels and NMR transitions for the proton in HD.



- ***Figure 6***. *Nuclear energy levels for proton in HD molecule. The two Zeeman levels of the proton when B > 0 are further split by interaction with the three possible spin orientations of the deuteron M_d= −1, 0, +1. The proton NMR transition, represented by blue arrows, is split into a triplet with separation 42.9 Hz.*

Nuclear-spin phenomena in the HD molecule can be compactly represented by a spin Hamiltonian

$$\hat{H} = -\hbar\gamma_H M_H(1-\sigma_H) - \hbar\gamma_D M_D(1-\sigma_D)B + hJ_{HD}I_H \cdot I_D \tag{9.16.18}$$

- The shielding constants σ_H and σ_D are, in this case, equal since the two nuclei are chemically identical. For sufficiently large magnetic fields B, the last term is effectively equal to $hJ_{HD}M_HM_D$. The spin-coupling constant J can be directly equated to the splitting expressed in Hz. We consider next the case of two equivalent protons, for example, the CH_2 group of ethanol. Each proton can have two possible spin states with $M_I = \pm\frac{1}{2}$, giving a total of four composite spin states. Just as in the case of 2 electron spins, these combine to give singlet and triplet nuclear-spin states with M = 0 and 1, respectively. Also, just as for electron spins, transitions between singlet and triplet states are forbidden. The triplet state allows NMR transitions with $\Delta M = \pm 1$ to give a single resonance frequency, while the singlet state is inactive. As a consequence, spin-spin splittings do not occur among identical nuclei. For example, the H_2 molecule shows just a single NMR frequency. And the CH_2 protons in ethanol do not show spin- spin interactions with one another. They can however cause a splitting of the neighboring CH_3 protons. Fig. 7 (left side) shows the four possible spin states of two equivalent protons, such as those in the methylene group CH_2, and the triplet with intensity ratios 1:2:1 which these produce in nearby protons. Also shown (right side) are the eight possible spin states for three equivalent protons, say those in a methyl group CH_3, and the quartet with intensity ratios 1:3:3:1 which these produce. In general, n equivalent protons will give a splitting pattern of n + 1 lines in the ratio of binomial coefficients 1:n:n(n − 1)/2 . . . The tertiary hydrogen in isobutane $(CH_3)_3CH*$, marked with an asterisk, should be split into 10 lines by the 9 equivalent methyl protons.

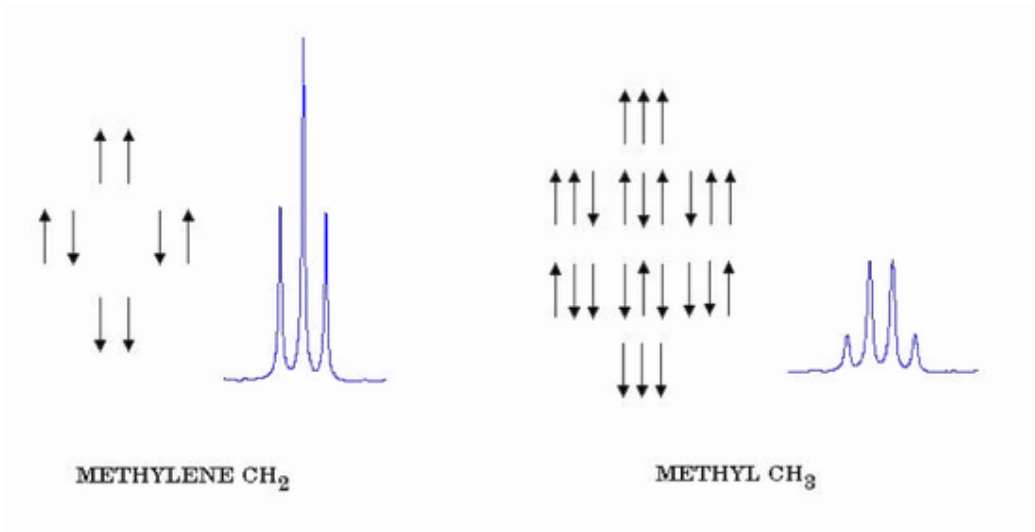


Figure 7. *Splitting patterns from methylene and methyl protons.*

The NMR spectrum of ethanol CH_3CH_2OH (Fig. 4) can now be interpreted. The CH_3 protons are split into a 1:2:1 triplet by spin-spin interaction with the neighboring CH_2. Conversely, the CH_2 protons are split into a 1:3:3:1 quartet by interaction with the CH_3. The OH (hydroxyl) proton evidently does not either cause or undergo spin-spin splitting. The explanation for this is hydrogen bonding, which involves rapid exchange of hydroxyl protons among neighboring molecules. If this rate of exchange is greater than or comparable to the NMR radiofrequency, then the splittings will be “washed out.” Only one line with a motion-averaged value of the chemical shift will be observed. NMR has consequently become a useful tool to study intramolecular motions.

Contributors and Attributions

- Seymour Blinder (Professor Emeritus of Chemistry and Physics at the University of Michigan, Ann Arbor)

This page titled 9.16: Nuclear Magnetic Resonance is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by S. M. Blinder.

- Chapter 14. Nuclear Magnetic Resonance is licensed CC BY-NC-SA 4.0.

10: MOLECULAR SPECTROSCOPY

Spectroscopy generally is defined as the area of science concerned with the absorption, emission, and scattering of electromagnetic radiation by atoms and molecules, which may be in the gas, liquid, or solid phase. Visible electromagnetic radiation is called light, although the terms light, radiation, and electromagnetic radiation can be used interchangeably. Spectroscopy played a key role in the development of quantum mechanics and is essential to understanding molecular properties and the results of spectroscopic experiments. It is used as a “stepping stone” to take us to the concepts of quantum mechanics and the quantum mechanical description of molecular properties in order to make the discussion more concrete and less abstract and mathematical.

10.1: The Electromagnetic Spectrum

10.2: Rotations Accompany Vibrational Transitions

10.3: Unequal Spacings in Vibration-Rotation Spectra

10.4: Unequal Spacings in Pure Rotational Spectra

10.5: Vibrational Overtones

10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information

10.7: Electronic Spectroscopy

10.7.1: Spectroscopy of the Particle-in-a-Box Model

10.7.2: Cyanine Dyes

10.7.3: Fluorescence and Phosphorescence

10.8: Rotational Spectra of Polyatomic Molecules

10.9: Normal Modes in Polyatomic Molecules

10.10: Irreducible Representation of Point Groups

10.11: The Selection Rule for the Rigid Rotor

10.12: The Harmonic Oscillator Selection Rule

10.13: Group Theory Determines Infrared Activity

10.E: Molecular Spectroscopy (Exercises)

This page titled 10: Molecular Spectroscopy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 13: Molecular Spectroscopy is licensed CC BY-NC-SA 4.0.

10.1: The Electromagnetic Spectrum

An important aspect of studying Physical Chemistry is to be able to recognize the interaction of molecules to the surroundings. Molecular Spectroscopy provides a clear image of how diatomic and polyatomic molecules interact by looking at the Frequency, Wavelength, Wave number, Energy, and molecular process. We will also be able to see the absorption properties of molecules in various regions from the electromagnetic spectrum.

Electromagnetic Radiation

Electromagnetic radiation—light—is a form of energy whose behavior is described by the properties of both waves and particles. Some properties of electromagnetic radiation, such as its refraction when it passes from one medium to another are explained best by describing light as a wave. Other properties, such as absorption and emission, are better described by treating light as a particle. The exact nature of electromagnetic radiation remains unclear, as it has since the development of quantum mechanics in the first quarter of the 20th century. Nevertheless, the dual models of wave and particle behavior provide a useful description for electromagnetic radiation.

Electromagnetic radiation consists of oscillating electric and magnetic fields that propagate through space along a linear path and with a constant velocity. In a vacuum electromagnetic radiation travels at the speed of light, c, which is $2.99792 \times 10^8\ m/s$. When electromagnetic radiation moves through a medium other than a vacuum its velocity, v, is less than the speed of light in a vacuum. The difference between v and c is sufficiently small (<0.1%) that the speed of light to three significant figures, $3.00 \times 10^8\ m/s$, is accurate enough for most purposes.

The oscillations in the electric and magnetic fields are perpendicular to each other, and to the direction of the wave's propagation. Figure 13.1.1 shows an example of plane-polarized electromagnetic radiation, consisting of a single oscillating electric field and a single oscillating magnetic field.

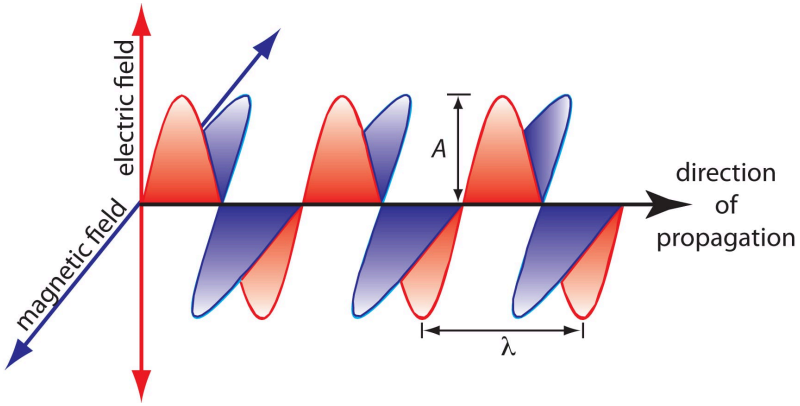


Figure 13.1.1 : Plane-polarized electromagnetic radiation showing the oscillating electric field in red and the oscillating magnetic field in blue. The radiation's amplitude, A, and its wavelength, λ, are shown. Normally, electromagnetic radiation is unpolarized, with oscillating electric and magnetic fields present in all possible planes perpendicular to the direction of propagation.

An electromagnetic wave is characterized by several fundamental properties, including its velocity, amplitude, frequency, phase angle, polarization, and direction of propagation.[2] For example, the amplitude of the oscillating electric field at any point along the propagating wave is

$$A_\mathrm{t} = A_\mathrm{e} \sin(2\pi\nu t + \phi)$$

where A_t is the magnitude of the electric field at time t, A_e is the electric field's maximum **amplitude**, ν is the wave's **frequency**—the number of oscillations in the electric field per unit time—and ϕ is a **phase angle**, which accounts for the fact that A_t need not have a value of zero at $t = 0$. The identical equation for the magnetic field is

$$A_\mathrm{t} = A_\mathrm{m} \sin(2\pi\nu t + \phi)$$

where A_m is the magnetic field's maximum amplitude.

The frequency and wavelength of electromagnetic radiation vary over many orders of magnitude. For convenience, we divide electromagnetic radiation into different regions—the **electromagnetic spectrum**—based on the type of atomic or molecular transition that gives rise to the absorption or emission of photons (Figure 13.1.2). The boundaries between the regions of the electromagnetic spectrum are not rigid, and overlap between spectral regions is possible.

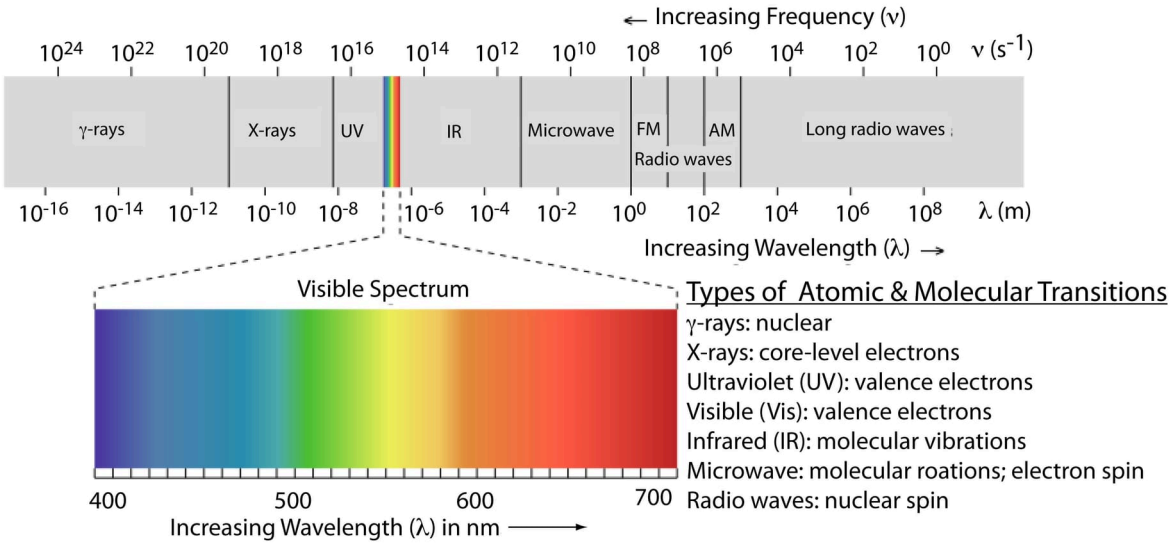


Figure 13.1.2 : The electromagnetic spectrum showing the boundaries between different regions and the type of atomic or molecular transition responsible for the change in energy. The colored inset shows the visible spectrum. Source: modified from Zedh (www.commons.Wikipedia.org).

Other Units

Other properties also are useful for characterizing the wave behavior of electromagnetic radiation. The **wavelength**, λ, is defined as the distance between successive maxima (Figure 13.1.1). For ultraviolet and visible electromagnetic radiation the wavelength is usually expressed in nanometers (1 nm = 10^{-9} m), and for infrared radiation it is given in microns (1 μm = 10^{-6} m). The relationship between wavelength and frequency is

$$\lambda = \frac{c}{\nu}$$

Another unit useful unit is the **wavenumber**, $\tilde{\nu}$, which is the reciprocal of wavelength

$$\tilde{\nu} = \frac{1}{\lambda}$$

Wavenumbers are frequently used to characterize infrared radiation, with the units given in cm^{-1}.

Example 13.1.1

In 1817, Josef Fraunhofer studied the spectrum of solar radiation, observing a continuous spectrum with numerous dark lines. Fraunhofer labeled the most prominent of the dark lines with letters. In 1859, Gustav Kirchhoff showed that the D line in the sun's spectrum was due to the absorption of solar radiation by sodium atoms. The wavelength of the sodium D line is 589 nm. What are the frequency and the wavenumber for this line?

Solution

The frequency and wavenumber of the sodium D line are

$$\nu = \frac{c}{\lambda} = \frac{3.00 \times 10^{8}\ \text{m/s}}{589 \times 10^{-9}\ \text{m}} = 5.09 \times 10^{14}\ \text{s}^{-1}$$

$$\tilde{\nu} = \frac{1}{\lambda} = \frac{1}{589 \times 10^{-9}\ \text{m}} \times \frac{1\ \text{m}}{100\ \text{cm}} = 1.70 \times 10^{4}\ \text{cm}^{-1}$$

Exercise 13.1.1

Another historically important series of spectral lines is the Balmer series of emission lines form hydrogen. One of the lines has a wavelength of 656.3 nm. What are the frequency and the wavenumber for this line?

Above, we defined several characteristic properties of electromagnetic radiation, including its energy, velocity, amplitude, frequency, phase angle, polarization, and direction of propagation. A spectroscopic measurement is possible only if the photon's interaction with the sample leads to a change in one or more of these characteristic properties. We can divide spectroscopy into two broad classes of techniques. In one class of techniques there is a transfer of energy between the photon and the sample. Table 13.1.1 provides a list of several representative examples.

Table 13.1.1 : Examples of Spectroscopic Techniques Involving an Exchange of Energy Between a Photon and the Sample

Type of Energy Transfer	Region of Electromagnetic Spectrum	Spectroscopic Technique
absorption	γ-ray	Mossbauer spectroscopy
	X-ray	X-ray absorption spectroscopy
	UV/Vis	UV/Vis spectroscopy atomic absorption spectroscopy
	IR	infrared spectroscopy raman spectroscopy
	Microwave	microwave spectroscopy
	Radio wave	electron spin resonance spectroscopy nuclear magnetic resonance spectroscopy
emission (thermal excitation)	UV/Vis	atomic emission spectroscopy
photoluminescence	X-ray	X-ray fluorescence
	UV/Vis	fluorescence spectroscopy phosphorescence spectroscopy atomic fluorescence spectroscopy
chemiluminescence	UV/Vis	chemiluminescence spectroscopy

Electromagnetic spectrum provides clearly information of molecules if they are rotational transitions, vibrational transitions, or electronic transitions. A molecule or a set of molecules can be read by the absorption of microwave radiation which provides transitions between rotational energy levels. In addition, if the molecules absorbs infrared radiation provides the transitions

between vibrational levels follows by transitions between rotational energy levels. Finally, when molecules absorbs visible and ultraviolet radiation gives transitions between electronic energy levels follows by simultaneous transitions between vibrational and rotational levels.

When given the energy level of the molecules along with wavelength, we can easily figure the frequency of the molecules where they fall in the electromagnetic spectrum regions:

$$\Delta E = E_u - E_l = h\nu$$

The above equation describes the energy change between upper state and lower state of energy.

- Frequency falls between 10^9 - 10^{11} which is in the microwave range correlates to the rotation of polyatomic molecules.
- Frequency falls between 10^{11} - 10^{13} which is in the far infrared range correlates to the rotation of small molecules.
- Frequency falls between 10^{13} - 10^{14} which is in the infrared range correlates to the vibrations of flexible bonds.
- Frequency falls between 10^{14} - 10^{16} which is in the visible and ultraviolet range correlates to the electronic transitions.

The powerful technique of figuring out the the frequency of the molecules can help us determine the bond length, temperature, probability distribution as you will learn later on from the degree of freedoms and how the process is undergo in specific a reaction.

This page titled 10.1: The Electromagnetic Spectrum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Delmar Larsen.

- 13.1: The Electromagnetic Spectrum by Delmar Larsen is licensed CC BY 4.0.

10.2: Rotations Accompany Vibrational Transitions

Each of the modes of vibration of diatomic molecules in the gas phase also contains closely-spaced (1-10 cm^{-1} difference) energy states attributable to rotational transitions that accompany the vibrational transitions. A molecule's rotation can be affected by its vibrational transition because there is a change in bond length, so these rotational transitions are expected to occur. Since vibrational energy states are on the order of 1000 cm^{-1}, the rotational energy states can be superimposed upon the vibrational energy states.

Selection Rules

Rotational and Vibration transitions (also known as rigid rotor and harmonic oscillator) of molecules help us identify how molecules interact with each other, their bond length as mentioned in previous section. In order to know each transitions, we have to consider other terms like wavenumber, force constant, quantum number, etc. There are rotational energy levels associated with all vibrational levels. From this, vibrational transitions can couple with rotational transitions to give **rovibrational spectra**. Rovibrational spectra can be analyzed to determine average bond length.

We treat the molecule's **vibrations** as those of a harmonic oscillator (ignoring anharmonicity). The energy of a vibration is quantized in discrete levels and given by

$$E_v = h\nu\left(v + \frac{1}{2}\right)$$

Where v is the vibrational quantum number and can have integer values 0, 1, 2..., and ν is the frequency of the vibration given by:

$$\nu = \frac{1}{2\pi}\left(\frac{k}{\mu}\right)^{\frac{1}{2}}$$

where k is the force constant and μ is the reduced mass of a diatomic molecule with atom masses m_1 and m_2, given by

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \tag{10.2.1}$$

We treat the molecule's **rotations** as those of a rigid rotor (ignoring centrifugal distortion from non-rigid rotor aspects). The energy of a rotation is also quantized in discrete levels given by

$$E_r = \frac{h^2}{8\pi^2 I}J(J+1)$$

In which I is the moment of inertia, given by

$$I = \mu r^2$$

where μ is the reduced mass (Equation 10.2.1) and r is the equilibrium bond length.

Experimentally, frequencies or wavenumbers are measured rather than energies, and dividing by h or hc gives more commonly seen term symbols, $F(J)$ using the rotational quantum number J and the rotational constant B in either frequency

$$F(J) = \frac{E_r}{h} = \frac{h}{8\pi^2 I}J(J+1) = BJ(J+1)$$

or wavenumbers

$$\tilde{F}(J) = \frac{E_r}{hc} = \frac{h}{8\pi^2 cI}J(J+1) = \tilde{B}J(J+1)$$

It is important to note in which units one is working since the rotational constant is always represented as B, whether in frequency or wavenumbers.

- **Vibrational Transition Selection Rules**: At room temperature, typically only the lowest energy vibrational state v= 0 is populated, so typically v_0 = 0 and Δv = +1. The full selection rule is technically that Δv = ±1, however here we assume energy can only go upwards because of the lack of population in the upper vibrational states.
- **Rotational Transition Selection Rules**: At room temperature, states with J≠0 can be populated since they represent the fine structure of vibrational states and have smaller energy differences than successive vibrational levels. Additionally, ΔJ = ±1 since a photon contains one quantum **of angular momentum** and we abide by the principle of **conservation of energy**. This is also the selection rule for rotational transitions.

These two selection rules mean that the transition ΔJ = 0 (i.e. J" = 0 and J' = 0, but $\nu_0 \neq 0$ is forbidden and the pure vibrational transition is not observed in most cases. The rotational selection rule gives rise to an R-branch (when ΔJ = +1) and a P-branch (when ΔJ = -1). Each line of the branch is labeled R(J) or P(J), where J represents the value of the lower state Figure 13.2.1).

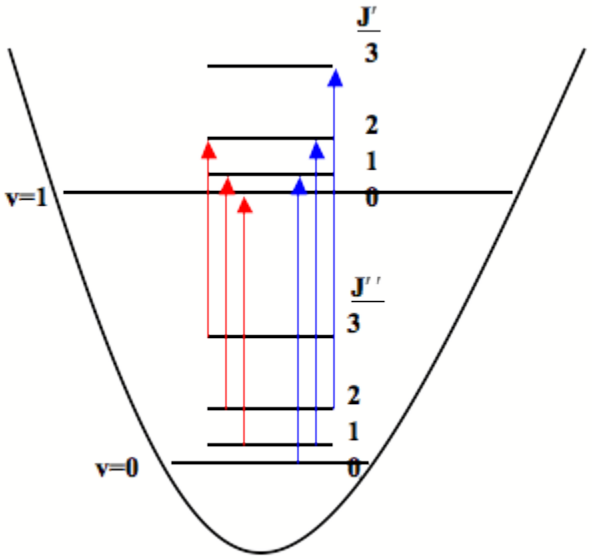


Figure 13.2.1 : Rotation-Vibration Transitions. Rotational transitions are on the order of 1-10 cm^{-1}, while vibrational transitions are on the order of 1000 cm^{-1}. The difference of magnitude between the energy transitions allow rotational levels to be superimposed within vibrational levels.

R-branch

When ΔJ = +1, i.e. the rotational quantum number in the ground state is one more than the rotational quantum number in the excited state – R branch (in French, *riche* or rich). To find the energy of a line of the R-branch:

$$\Delta E = h\nu_0 + hB\left[J(J+1) - J'(J'+1)\right] \tag{10.2.2}$$
$$= h\nu_0 + hB\left[(J+1)(J+2) - J(J+1)\right] \tag{10.2.3}$$
$$= h\nu_0 + 2hB(J+1) \tag{10.2.4}$$

P-branch

When ΔJ = -1, i.e. the rotational quantum number in the ground state is one less than the rotational quantum number in the excited state – P branch (in French, *pauvre* or poor). To find the energy of a line of the P-branch:

$$\Delta E = h\nu_0 + hB\left[J(J+1) - J'(J'+1)\right] \tag{10.2.5}$$
$$= h\nu_0 + hB\left[J(J-1) - J(J+1)\right] \tag{10.2.6}$$
$$= h\nu_0 - 2hBJ \tag{10.2.7}$$

Q-branch

When ΔJ = 0, i.e. the rotational quantum number in the ground state is the same as the rotational quantum number in the excited state – Q branch (simple, the letter between P and R). To find the energy of a line of the Q-branch:

$$\Delta E = h\nu_0 + hB[J(J+1) - J'(J'+1)] \tag{10.2.8}$$

$$= h\nu_0 \tag{10.2.9}$$

The Q-branch can be observed in polyatomic molecules and diatomic molecules with electronic angular momentum in the ground electronic state, e.g. nitric oxide, NO. Most diatomics, such as O_2, have a small moment of inertia and thus very small angular momentum and yield no Q-branch.

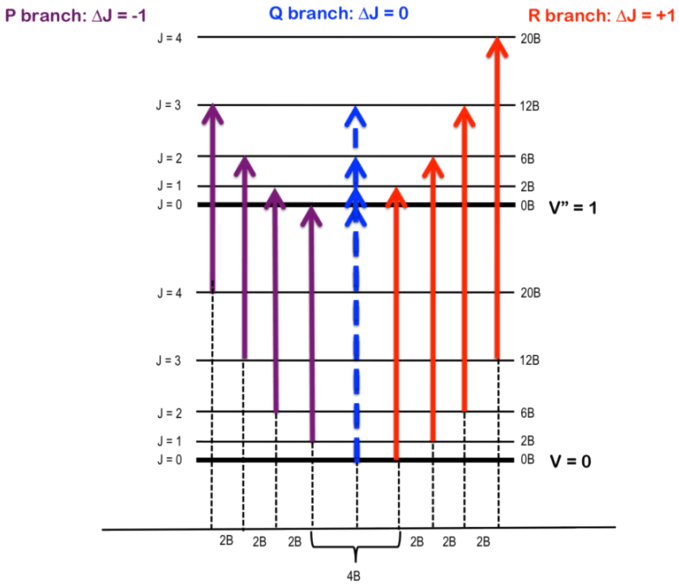


Figure 13.2.3 : Cartoon depiction of rotational energy levels, J, imposed on vibrational energy levels, v. The transitions between levels that would result in the P- and R-branches are depicted in purple and red, respectively, in addition to the theoretical Q-branch line in blue.

As seen in Figure 13.2.2 , the lines of the P-branch (represented by purple arrows) and R-branch (represented by red arrows) are separated by specific multiples of B (i.e, $2B$), thus the bond length can be deduced without the need for pure rotational spectroscopy.

The total nuclear energy of the combined rotation-vibration terms, $S(v, J)$, can be written as the sum of the vibrational energy and the rotational energy

$$S(v, J) = G(v) + F(J)$$

where $G(v)$ represents the energy of the harmonic oscillator, ignoring anharmonic components and $S(J)$ represents the energy of a rigid rotor, ignoring centrifugal distortion.

From this, we can derive

$$S(v, J) = \nu_0 \left(v + \frac{1}{2}\right) + BJ(J+1)$$

The spectrum we expect, based on the conditions described above, consists of lines equidistant in energy from one another, separated by a value of $2B$. The *relative* intensity of the lines is a function of the rotational populations of the ground states, i.e. the intensity is proportional to the number of molecules that have made the transition. The *overall* intensity of the lines depends on the vibrational transition dipole moment.

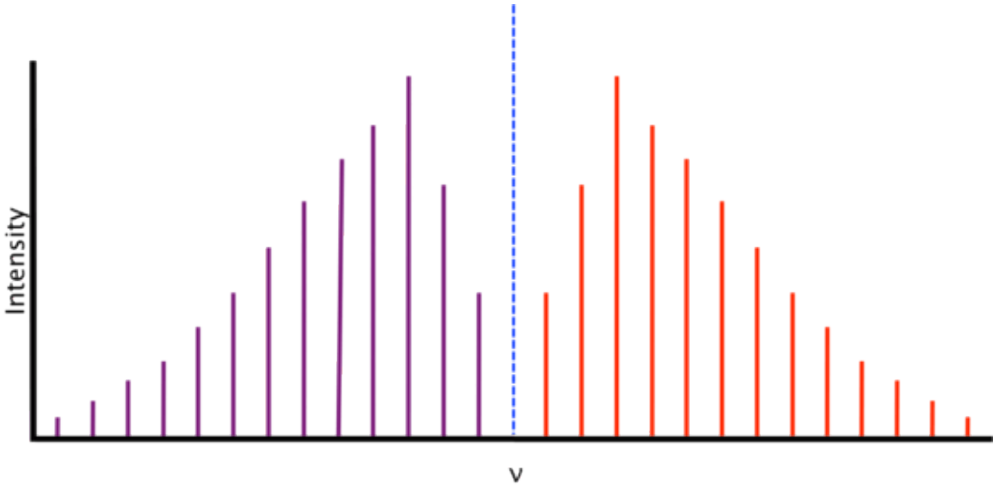


Figure 13.2.3 : A cartoon depiction of an ideal rovibrational spectrum.

In Figure 13.2.2 , between $P(1)$ and $R(0)$ lies the zero gap, where the the first lines of both the P- and R-branch are separated by $4B$, assuming that the rotational constant B is equal for both energy levels. The zero gap is also where we would expect the Q-branch, depicted as the dotted line, if it is allowed.

Advanced Concept: Occupations (Peak Intensities)

The relative intensity of the P- and R-branch lines depends on the thermal distribution of electrons; more specifically, they depend on the population of the lower J state. If we represent the population of the Jth upper level as N_J and the population of the lower state as N_0, we can find the population of the upper state relative to the lower state using the Boltzmann distribution:

$$\frac{N_J}{N_0} = (2J+1)e^{\left(-\frac{E_r}{kT}\right)}$$

(2J+1) gives the degeneracy of the Jth upper level arising from the allowed values of M_J (+J to –J). As J increases, the degeneracy factor increases and the exponential factor decreases until at high J, the exponential factor wins out and N_J/N_0 approaches zero at a certain level, J_{max}. Thus, when

$$\frac{d}{dJ}\left(\frac{N_J}{N_0}\right) = 0$$

by differentiation, we obtain

$$J_{max} = \left(\frac{kT}{2hB}\right)^{\frac{1}{2}} - \frac{1}{2}$$

This is the reason that rovibrational spectral lines increase in energy to a maximum as J increases, then decrease to zero as J continues to increase, as seen in Figure 13.2.2 .

From this relationship, we can also deduce that in heavier molecules, B will decrease because the moment of inertia will increase, and the decrease in the exponential factor is less pronounced. This results in the population distribution shifting to higher values of J. Similarly, as temperature increases, the population distribution will shift towards higher values of J.

This page titled 10.2: Rotations Accompany Vibrational Transitions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Joya Cooley.

- 13.2: Rotations Accompany Vibrational Transitions by Joya Cooley is licensed CC BY 4.0.

10.3: Unequal Spacings in Vibration-Rotation Spectra

We find that real rovibrational spectra do not exhibit the equal spacing expectations from the treatment in the previous section and look more like the idealized spectrum in Figure 13.3.1 . As energy increases, the R-branch lines become increasingly similar in energy (i.e., the lines move closer together) and as energy decreases, the P-branch lines become increasingly dissimilar in energy (i.e. the lines move farther apart). This is attributable to two phenomena: *rotational-vibrational coupling* and *centrifugal distortion*.

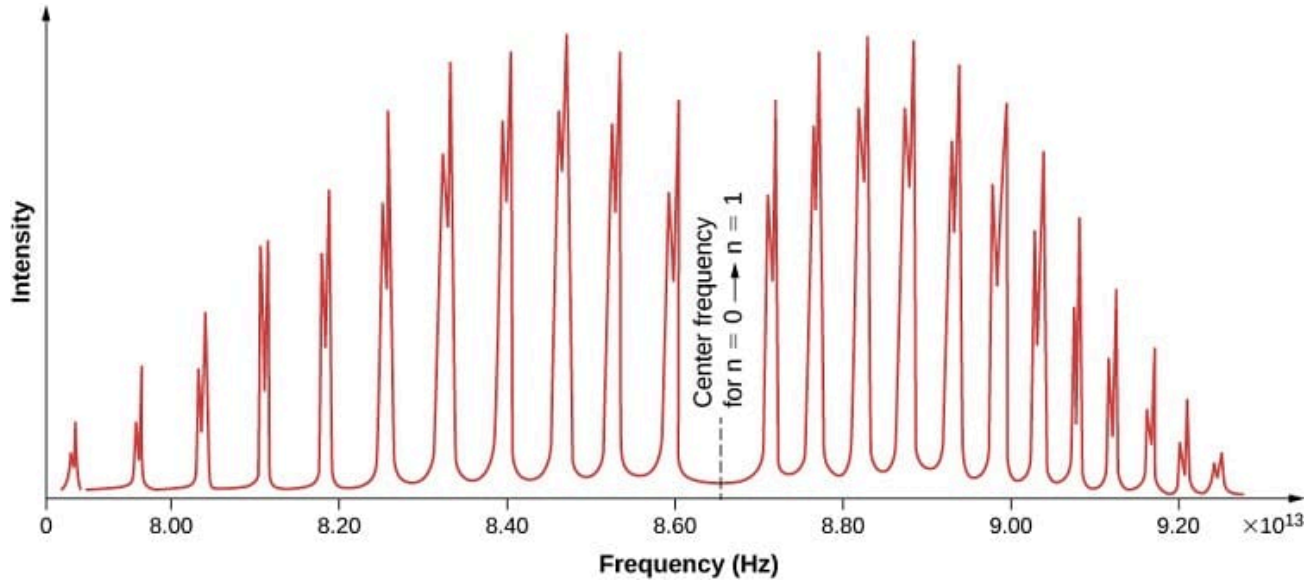


Figure 13.3.1 : Absorption spectrum of hydrogen chloride (HCl) from the $n = 0$ to $n = 1$ vibrational levels. The discrete peaks indicate a quantization of the angular momentum of the molecule. The bands to the left indicate a decrease in angular momentum, whereas those to the right indicate an increase in angular momentum. (CC-SA; OpenStax).

Experimental data shows that the P branch has closer spaces between the lines as the frequencies decrease while R branch has further spaces between the lines as the frequencies increase. On the spectrum, R region is on the left and P region is on the right separated by a large gap in between. There are other graphs show differently, but the regions can be labeled using their spacing characteristic discussed below.

Vibration-Rotation Interaction

Consider the rotational constant for a rigid rotator $\tilde{B} = h/8\pi^2 c\mu R_e^2$ where $\tilde{B}$ is the rotational constant, and R_e^2 is the bond length. It is clear from the equation that $\tilde{B}$ depends on R_e^2 in a inverse manner, so $\tilde{B}$ will decrease as R_e^2 increase, and vice versa.

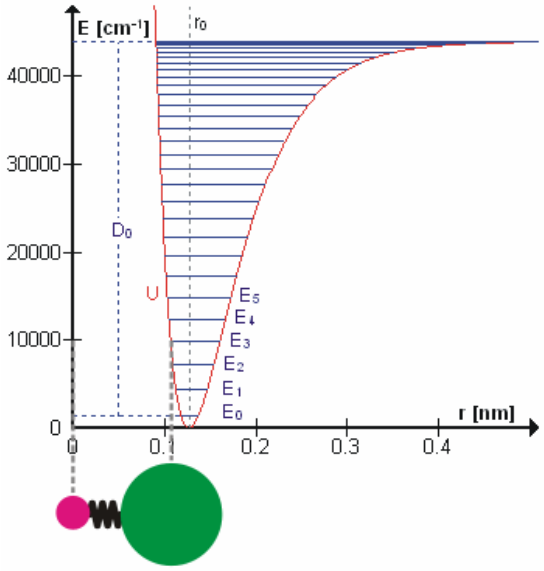


Figure 13.3.2 : Anharmonic Oscillator and Bond Length. Figure used with permission from Wikipedia.

Vibrational state and the bond length also possess a relationship, and that is bond length, R_e^2, will increase as the vibrational state, v increases. High vibrational states have large vibrational amplitudes (the amplitude can be visualized as the distance from one side to the other on the harmonic-oscillator parabola, and higher states has greater distance). On a potential graph, imagine a diatomic molecule with one of the atom fixed at the origin, the bond is along the x-axis, and the molecule vibrate toward and away from the origin along x-axis. The range at how much the bond can stretch or squeeze depend on the vibrational amplitude. So, at high vibrational state (thus high amplitude), the bond can squeeze in and stretch out more (just a little more) on the regular harmonic-oscillator potential, but it can stretch out greatly on an anharmonic oscillator. That's why bond length increase with vibrational state.

Note

The two relationship established above

- $\tilde{B} \propto R_e^2$ (inversely proportional)

and

- $R_e^2 \propto v$ (directly proportional)

Therefore,

- $\tilde{B} \propto v$ (inversely proportional)

The rotational constant dependent on the vibrational state is denoted as $\tilde{B}_v$, and the dependence of $\tilde{B}$ on v is the vibrational-rotational interaction.

Let's calculate the frequencies of R and P Branches with vibrational dependent rotational constants. The energies of the rigid rotator-harmonic oscillator is

$$\tilde{E}_{v,J} = G(v) + F(J) = \tilde{\nu}(v + \frac{1}{2}) + \tilde{B}J(J+1)$$

or in term of $\tilde{B}_v$

$$\tilde{E}_{v,J} = \tilde{\nu}\left(v + \frac{1}{2}\right) + \tilde{B}_v J(J+1)$$

According to the vibration-rotation interaction, $\tilde{B}_1 < \tilde{B}_0$ because $R_e^2(v=1) > R_e^2(v=0)$. Hence, the frequencies of rovibrational transitions from $v = 0 \rightarrow 1$ can be calculated as the following:

For R branch, allowed $J = 0, 1, 2, ...$

$$\tilde{\nu}_R(\Delta J = +1) = \tilde{E}_{1,J+1} - \tilde{E}_{0,J}$$

$$= \frac{3}{2}\tilde{\nu} + \tilde{B}_1(J+1)(J+2) - \frac{1}{2}\tilde{\nu} - \tilde{B}_0 J(J+1)$$

$$= \tilde{\nu} + 2\tilde{B}_1 + (3\tilde{B}_1 - \tilde{B}_0)J + (\tilde{B}_1 - \tilde{B}_0)J^2$$

For P branch, allowed $J = 1, 2, 3, ...$

$$\tilde{\nu}_P(\Delta J = -1) = \tilde{E}_{1,J-1} - \tilde{E}_{0,J}$$

$$= \tilde{\nu} - (\tilde{B}_1 + \tilde{B}_0)J + (\tilde{B}_1 - \tilde{B}_0)J^2$$

Decrease in Spacing of Lines in the R Branch with Increasing J

A closer look at

$$\tilde{\nu}_R = \tilde{\nu} + 2\tilde{B}_1 + (3\tilde{B}_1 - \tilde{B}_0)J + (\tilde{B}_1 - \tilde{B}_0)J^2$$

shows that the last term in the parentheses $(\tilde{B}_1 - \tilde{B}_0)$ will be always negative because $\tilde{B}_1 < \tilde{B}_0$, and it also multiplies with J^2, so the square term $(\tilde{B}_1 - \tilde{B}_0)J^2$ will give a larger negative value for increasing J. As a result, $\tilde{\nu}_R = \tilde{\nu}$ +(smaller and smaller value) as J increase. Because the rotational frequencies keep getting smaller, the spacing between lines in R branch decreases as J increases.

Increasing in Spacing of Lines in the P Branch with Decreasing J

Using the same analysis as above,

$$\tilde{\nu}_P = \tilde{\nu} - (\tilde{B}_1 + \tilde{B}_0)J + (\tilde{B}_1 - \tilde{B}_0)J^2$$

shows that the last term in the parentheses $(\tilde{B}_1 - \tilde{B}_0)$ will be always negative because $\tilde{B}_1 < \tilde{B}_0$, and it also multiplies with J^2, so the square term $(\tilde{B}_1 - \tilde{B}_0)J^2$ will give a larger negative value even for decreasing J. As a result, $\tilde{\nu}_R = \tilde{\nu}$ −(smaller and smaller value) as J decreases. Because the rotational frequencies keep getting larger, the spacing between lines in P branch increase as J decreases.

References

1. McQuarrie, Donald A. Quantum Chemistry. New York: University Science Books, 2007.
2. Hollas, M. J. *Modern spectroscopy*. (3rd ed.). Chichester: John Wiley & Sons, 1996.
3. Hollas, M. J. *Basic atomic and molecular spectroscopy*. Cambridge: The Royal Society of Chemistry, 2002.
4. Herzberg, G. *Molecular spectra and molecular structure*. (2nd ed.). New York: Prentice-Hall, 1950.
5. Fetterolf, Monty L. Enhanced Intensity Distribution Analysis of the Rotational–Vibrational Spectrum of HCl. J. Chem. Ed. **2007**, *84*, 1064. DOI: 10.1021/ed084p1062

This page titled 10.3: Unequal Spacings in Vibration-Rotation Spectra is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Dayton Dang & Joya Cooley.

- 13.3: Unequal Spacings in Vibration-Rotation Spectra by Dayton Dang, Joya Cooley is licensed CC BY-NC-SA 4.0.

10.4: Unequal Spacings in Pure Rotational Spectra

Vibrational energy which is a consequence of the oscillations/ vibrations of the nuclei along inter nuclear axis, is possible only when the distance between the nuclei is not fixed/ rigid; that means the separation between the two nuclei is flexible/ elastic (non-rigid rotator). Consequently, centrifugal force, when the molecule is rotating, tends to fly the reduced mass μ away from the axis of rotation. To keep the mass rotating about the axis, there must be some restoring force to counter balance the centrifugal force. The work done to supply this force is stored as potential energy. Therefore, unlike the case of rigid rotator, total energy of rotation in a molecule comprises of kinetic and the potential energy corresponding to centrifugal force of rotation.

Let r_e is the distance between the nuclei when the separation is taken to be rigid and r_c under the action of the centrifugal force. According to Hook's law, restoring force is proportional to change in internuclear distance $(r_c - r_e)$, or $= k(r_c - r_e)$ which in turn, will be equal to the centrifugal force.

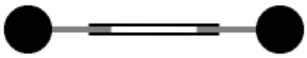


Figure 10.4.1: This animation can serve as an illustration of one of the aspects of the conservation of angular momentum. The animation is a schematic representation of two weights that are connected to each other. The two weights are circling around their common center of mass, the center of mass is stationary. The connection between the weight consists of pistons. When the connecting pistons contract, the rotation rate goes up. When the inward force exerted by the pistons relaxes the weight slide outwards, and the rotation rate goes down. (Public domain; Cleonis).

Centrifugal force,

$$F_c = \mu r_c \omega^2 = L^2/\mu {r_c}^3 \tag{10.4.1}$$

Equating the restoring force to the centrifugal force, one gets

$$k(r_c - r_e) = \frac{L^2}{\mu {r_c}^3} \tag{10.4.2}$$

$$r_c - r_e = \frac{L^2}{k\mu {r_c}^3} \cong \frac{L^2}{k\mu {r_e}^3} \tag{10.4.3}$$

Total energy, on adding kinetic energy, as expressed in Equation 10.4.3 to the potential energy,

$$\frac{1}{2}k(r_c - r_e)^2 \tag{10.4.4}$$

is given by

$$E_r = \frac{L^2}{2\mu {r_c}^2} + \frac{1}{2}k\,(r_c - r_e)^2 \tag{10.4.5}$$

Using $(r_c - r_e) = \Delta r$ and eliminating r_c from equation 10.4.5, one gets

$$E_r = \frac{L^2}{2\mu\ {r_e}^2(1+\Delta r/r_e)^2} + \frac{1}{2}k\Delta r^2 \tag{10.4.6}$$

$$= \frac{L^2(1-2\Delta r/r_e)}{2\mu\ {r_e}^2} + \frac{1}{2}k\Delta r^2 \tag{10.4.7}$$

using

$$(1+\Delta r/r_e)^{-2} \approx (1-2\Delta r/r_e) \tag{10.4.8}$$

$$E_r = \frac{L^2}{2\mu\, {r_e}^2} - \frac{L^4}{k\mu^2\, {r_e}^6} + \frac{1}{2}k\frac{L^4}{k^{2\mu 2}{r_e}^6} \tag{10.4.9}$$

$$= \frac{L^2}{2\mu\, {r_e}^2} - \frac{1}{2}\frac{L^4}{k\mu^2\, {r_e}^6} \tag{10.4.10}$$

$$= \frac{\hbar^2\ J(J+1)}{2\mu\, {r_e}^2} - \frac{\hbar^4\ J^2(J+1)^2}{2k\mu^2\, {r_e}^6} \tag{10.4.11}$$

Use of Equation 10.4.3 and of the relation regarding angular momentum of a rotor

$$L = \hbar\ \sqrt{J(J+1)} \tag{10.4.12}$$

has been made to obtain relation in Equation 10.4.11, which may be expressed in cm^{-1} as

$$F(J) = \underbrace{\tilde{B}J(J+1)}_{\text{rigid rotator term}} - \underbrace{\tilde{D}J^2(J+1)^2 cm^{-1}}_{\text{centrifugal stretching}} \tag{10.4.13}$$

where

$$\tilde{B} = \frac{\hbar}{4\pi\,\mu\, {r_e}^2 c}\ (\text{in cm}^{-1}) \tag{10.4.14}$$

and

$$\tilde{D} = \frac{\hbar^3}{4\pi k\,\mu^2\, {r_e}^6\, c}\ (\text{in cm}^{-1}) \tag{10.4.15}$$

First term in the Equation 10.4.13 is same as for the rigid rotator; second term is the consequence of the centrifugal stretching. Recall Equation 10.4.3 wherein k is the spring constant that, as we will see in the following section, plays the same role as in the vibrational motion. In other words, centrifugal stretching constant $\tilde{D}$ is not only measures the influence of centrifugal force, but also hints upon the interaction between the rotational and vibrational motions.

Since $\tilde{D}$ is positive, it is clear from Equation 10.4.13 that the energy levels for the non-rigid rotator are slightly lower on energy scale than those of rigid rotator for the corresponding J values; the magnitude of decrease in energy of the non-rigid rotator states increases with J as shown in the Figure 10.4.2.

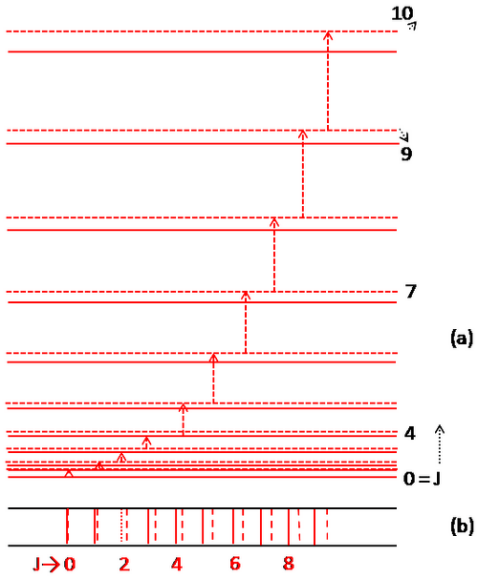


Figure $10.4.2$: Energy levels and the transitions of rigid rotator (broken red lines) & of non-rigid rotator (solid red lines). (b) resulting spectra from both the models

Consequently, on applying the selection rule $\Delta J = \pm 1$, rotational spectrum of a non-rigid rotator consists of a series of lines (red lines) wherein separation, unlike the case of spectral series (broken red lines) of a rigid rotator, between the consecutive rotational lines decreases with increase in J , as shown in the Figure $10.4.2$. It may be noted that value of D is very small compared to B with the result that the influence of D is significant only for very large J values.

For example, for HCl the values are $B \sim 10.4 \quad cm^{-1}$ and $D \sim 0.0004 \ cm^{-1}$. Usually, D is ignored in the calculations. Figure $10.4.2$ exaggerates the decrease in energy to visualize its effect. Nevertheless, non-rigid rotator is the model that describes the rotational motion more accurately and hence explains the spectral experimental observations not only in the microwave region but also the rotation-vibration spectra and the rotational structure of the electronic bands discussed in the later sections.

Contributors and Attributions

- 202.141.40.218/wiki/index.php...r_Spectroscopy

This page titled 10.4: Unequal Spacings in Pure Rotational Spectra is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

10.5: Vibrational Overtones

Although the harmonic oscillator proves useful for modeling lower energy levels, like $v = 0, 1$, it fails at higher numbers of v. It not only fails to properly model atomic bonds and dissociations, but also is unable to match spectra showing additional lines that are not accounted for in the harmonic oscillator model.

Anharmonicity

Until this point, we have been using the harmonic oscillator to describe the internuclear potential energy of the vibrational motion. Fundamental vibrational frequencies of a molecule corresponds to a transition of $\Delta v = \pm 1$. While this is a decent approximation, bonds do not behave like they do in the Harmonic Oscillator approximation (Figure $10.5.1$). For example, the parabolic potential of the Harmonic Oscillator approximation means that the vibrations can hold any amount of energy and bonds can stretch to any length, while in real molecules if the atoms are too far apart the bond will dissociate.

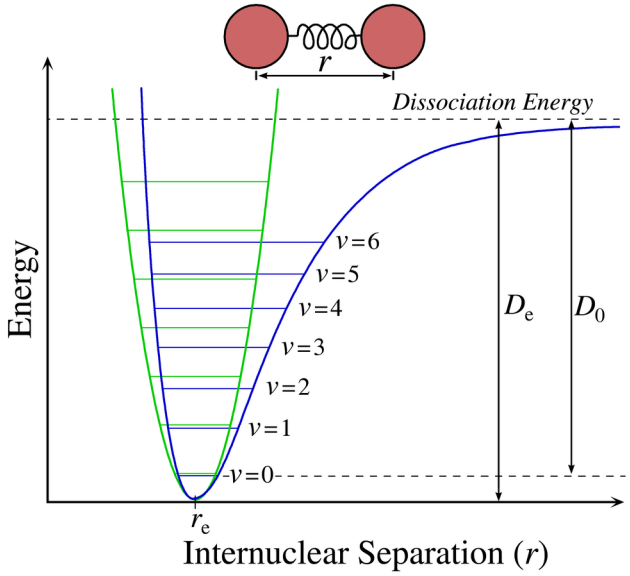


Figure $10.5.1$: Pictured above is the Harmonic Oscillator approximation (green parabola) superimposed on the anharmonic oscillator (blue curve) on a potential energy diagram. V(R) is the potential energy of a diatomic molecule and R is the radius between the centers of the two atoms. Towards the left is compression of the bond, towards the right is extension. (CC BY-SA 3.0; Created by Mark Somoza March 26 2006).

As you can see in Figure $10.5.1$, the harmonic oscillator potential well (in green) only roughly fits over the more accurate anharmonic oscillator well (in blue). The blue curve accounts for dissociation at large R values, which the green curve does not even remotely cover. However, this is just one important difference between the harmonic and anharmonic (real) oscillators.

The real potential energy can be expanded in the Taylor series.

$$V(R) = V(R_e) + \frac{1}{2!}\left(\frac{d^2V}{dR^2}\right)_{R=R_e}(R-R_e)^2 + \frac{1}{3!}\left(\frac{d^3V}{dR^3}\right)_{R=R_e}(R-R_e)^3 + \frac{1}{4!}\left(\frac{d^4V}{dR^4}\right)_{R=R_e}(R-R_e)^4 + \ldots \qquad (10.5.1)$$

This expansion was discussed in detail previously. The first term in the expansion is ignored since the derivative of the potential at R_e is zero (i.e., at the bottom of the well). The Harmonic Oscillator approximation only uses the next term, the quadratic term, in the series

$$V_{HO}(R) \approx V(R_e) + \frac{1}{2!}\left(\frac{d^2V}{dR^2}\right)_{R=R_e}(R-R_e)^2$$

or in terms of a spring constant, ignoring the absolute energy term($V(R_e)$) and defining r to equal the displacement from equilibrium ($r = R - R_e$), then we get the "standard" harmonic oscillator potential:

$$V_{HO}(R) = \frac{1}{2}kr^2$$

Alternatively, the expansion in Equation 10.5.1 can be shortened to the cubic term

$$V(x) = \frac{1}{2}kr^2 + \frac{1}{6}\gamma r^3 \tag{10.5.2}$$

where

- $V(x_0) = 0$, and $r = R - R_0$.
- k is the harmonic force constant, and
- γ is the first (i.e., cubic) *anharmonic* term

It is important to note that this approximation is only good for R near R_0.

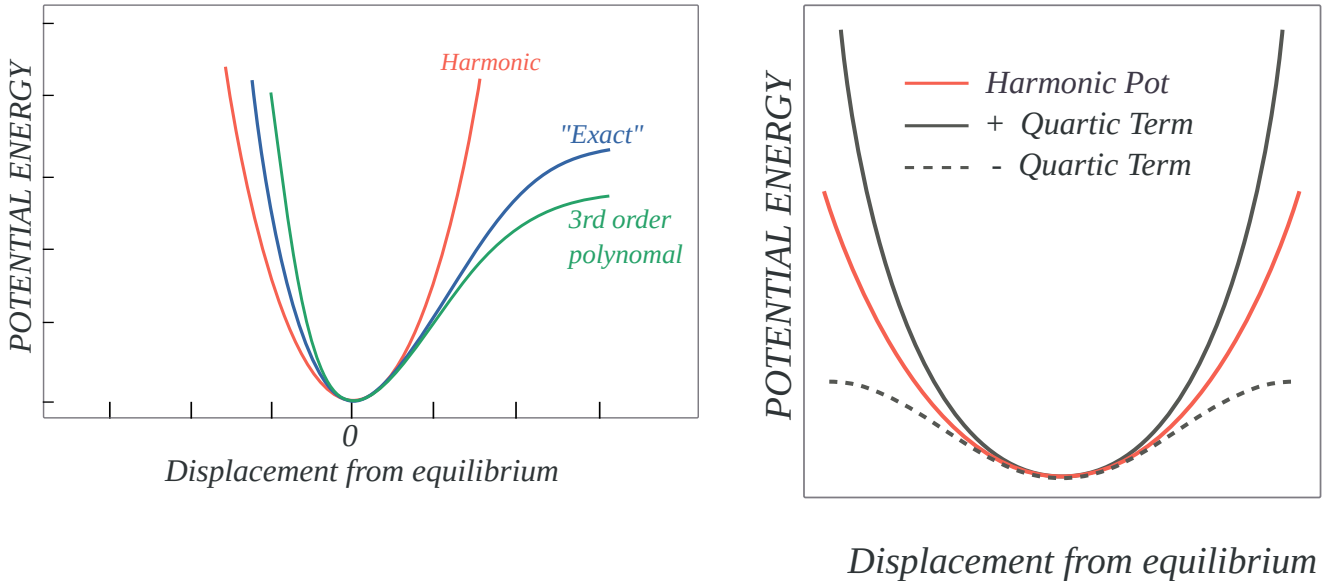


Figure 10.5.2 : (left) Deviation from simple harmonic potential approximation (red curve) of true/exact potential (blue curve) with cubic term (green). (right) Expansion with positive or negative quartic terms. (CC BY-NC; Ümit Kaya via LibreTexts)

The harmonic oscillator approximation gives by the following energies:

$$E_v = \tilde{\nu}\left(v + \frac{1}{2}\right)$$

When the cubic term in the expansion (Equation 10.5.2) is included, and the Schrödinger equation is solved using perturbation theory this gives:

$$E_v = \tilde{\nu}\left(v + \frac{1}{2}\right) - \tilde{\chi_e}\tilde{\nu}\left(v + \frac{1}{2}\right)^2$$

where $\tilde{\chi_e}$ is the **anharmonicity constant**. It is much smaller than 1, which makes sense because the terms in the Taylor series approach zero. This is why, although E_v technically includes all of the Taylor series, we only concern ourselves with the first and second terms. The rest are so small and barely add to the total and thus can be ignored. To get a more accurate approximation, more terms can be included, but otherwise, can be ignored. Almost all diatomics have experimentally determined $\frac{d^2V}{dx^2}$ for their lowest energy states. H_2, Li_2, O_2, N_2, and F_2 have had terms up to $n < 10$ determined of Equation 10.5.1.

Overtones

The Harmonic Oscillator approximation predicts that there will be only one line in the spectrum of a diatomic molecule, while experimental data shows there is in fact one dominant line--**the fundamental**--and other weaker lines. How can we account for

these extra lines?

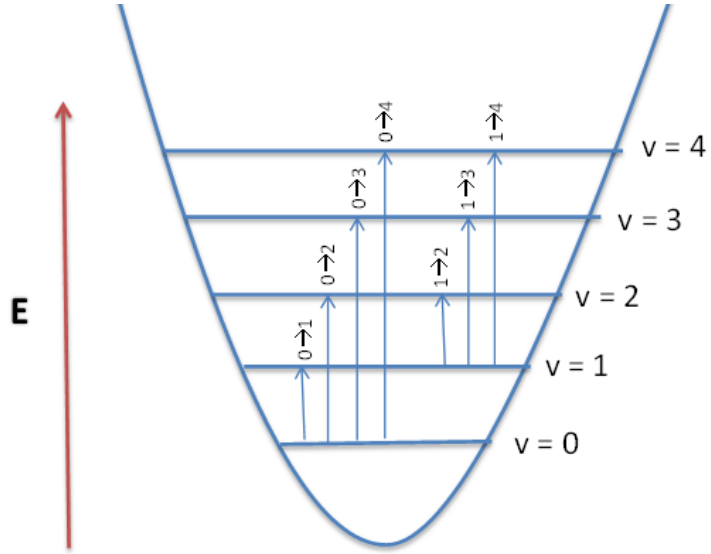


Figure $10.5.3$: A diagram for a vibrating diatomic molecule. The levels denoted by vibrational quantum numbers v represent the potential energy for the harmonic (quadratic) oscillator. The transition $0 \rightarrow 1$ is the fundamental, transitions $0 \rightarrow n$ (n>1) are called overtones, and transitions $1 \rightarrow n$ (n > 1) are called hot transitions (hot bands).

Any resonant frequency above the fundamental frequency is referred to as an *overtone*. In the IR spectrum, overtone bands are multiples of the fundamental absorption frequency. As you can recall, the energy levels in the Harmonic Oscillator approximation are evenly spaced apart. Energy is proportional to the frequency absorbed, which in turn is proportional to the wavenumber, the first overtone that appears in the spectrum will be twice the wavenumber of the fundamental. That is, first overtone $v = 0 \rightarrow 2$ is (approximately) twice the energy of the fundamental, $v = 0 \rightarrow 1$.

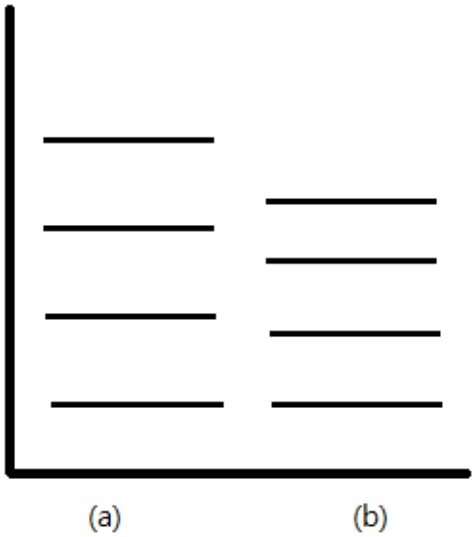


Figure $10.5.4$: An exaggerated view of the energy levels. (a) would be the harmonic oscillator, with the evenly spaced out energy levels. (b) would be the anharmonic, with the energy levels compressing until they converge.

The levels are not equally spaced, like in the harmonic oscillator, but decrease as v increases, until it ultimately converges, as implied by Figure $10.5.4$. Also as a result of anharmonicity, the $\Delta v = \pm 1$ selection rule is no longer valid and v can be any number. This leads to the observation of higher order transitions, or **overtones**, which result from the transition of the ground state to higher energy levels.

Anharmonic Oscillator Selection Rules

For the anharmonic oscillator, the selection rule is $\Delta V = \text{any number}$. That is, there are no selection rules (for state to state transitions)

Overtones occur when a vibrational mode is excited from $v = 0$ to $v = 2$ (the first overtone) or $v = 0$ to $v = 3$ (the second overtone). The fundamental transitions, $v = \pm 1$, are the most commonly occurring, and the probability of overtones rapidly decreases as $\Delta v > \pm 1$ gets bigger. Based on the harmonic oscillator approximation, the energy of the overtone transition will be approximately v times the fundamental associated with that particular transition. The anharmonic oscillator calculations show that the overtones are usually less than a multiple of the fundamental frequency. Overtones are generally not detected in larger molecules.

This is demonstrated with the vibrations of the diatomic HCl in the gas phase:

Table $10.5.1$ *: HCl vibrational spectrum.*

Transition	$\tilde{\nu}_{obs}$ [cm^{-1}]	$\tilde{\nu}_{obs}$ Harmonic [cm^{-1}]	$\tilde{\nu}_{obs}$ Anharmonic [cm^{-1}]
$0 \rightarrow 1$ (fundamental)	2885.9	2885.9	2,885.3
$0 \rightarrow 2$ (first overtone)	5668.0	5771.8	5,665.0
$0 \rightarrow 3$ (second overtone)	8347.0	8657.7	8,339.0
$0 \rightarrow 4$ (third overtone)	10 923.1	11 543.6	10,907.4
$0 \rightarrow 5$ (fourth overtone)	13 396.5	14 429.5	13,370

We can see from Table $10.5.1$ that the anharmonic frequencies correspond much better with the observed frequencies, especially as the vibrational levels increase. Because the energy levels and overtones are closer together in the anharmonic model, they are also more easily reached. This means that there is a higher chance of that level possibly being occupied, meaning it can show up as additional, albeit weaker intensity lines (the weaker intensity indicates a smaller probability of the transition occuring).

Exercise 13.5.1

HCl has a fundamental band at 2885.9 cm^{-1} and an overtone at 5668.1 cm^{-1} Calculate $\tilde{\nu}$ and $\tilde{\chi}_e$.

Exercise 13.5.2

Write out the Taylor series, and comment on the trend in the increasing terms. Using a test number x, please add terms 3, 4, and 5, then compare this to term 2. How do they compare? We have seen that the anharmonic terms increase the accuracy of our oscillator approximation. Why don't we care so much about terms past the second?

References

1. Infrared and Raman Spectra of Inorganic and Coordination Compounds. Part A: Theory and Applications in Inorganic Chemistry; Part B: Application in Coordination, Organometallic, and Bioinorganic Chemistry, 5th Edition (Nakamoto, Kazuo)
2. Lyle McAfee Journal of Chemical Education 2000 77 (9), 1122
3. Stuart, Barbara. *Infrared Spectroscopy: Fundamentals and Applications*. Wiley, 2004. Print.

This page titled 10.5: Vibrational Overtones is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Alexandra Holmes & Hannah Toru Shay.

- 13.5: Vibrational Overtones by Alexandra Holmes, Hannah Toru Shay is licensed CC BY 4.0.

10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information

Molecules can also undergo changes in electronic transitions during microwave and infrared absorptions. The energy level differences are usually high enough that it falls into the visible to UV range; in fact, most emissions in this range can be attributed to electronic transitions.

Electron Transitions are not Purely Electronic

We have thus far studied *rovibrational transitions*--that is, transitions involving both the vibrational and rotational states. Similarly, electronic transitions tend to accompany both rotational and vibrational transitions. These are often portrayed as an electronic potential energy cure with the vibrational level drawn on each curve. Additionally, each vibrational level has a set of rotational levels associated with it.

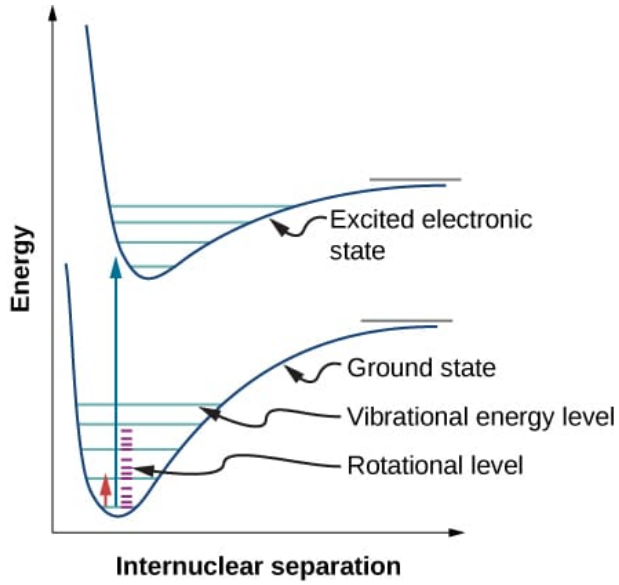


Figure 10.6.1: Three types of energy levels in a diatomic molecule: electronic, vibrational, and rotational. If the vibrational quantum number (n) changes by one unit, then the rotational quantum number (l) changes by one unit. (CC BY 3.0; OpenStax).

Recall that in the Born-Oppenheimer approximation, nuclear kinetic energies can be ignored (e.g., fixed) to solve for electronic wavefunctions and energies, which are much faster than rotation or vibration. As such, it is important to note that unlike rovibrational transitions, electronic transitions aren't dependent on rotational or transitional terms and are assumed to be separate. Therefore, when using an anharmonic oscillator-nonrigid rotator approximation (and excluding translation energy), the total energy of a diatomic is:

$$\tilde{E}_{total} = \tilde{\nu}_{el} + G(v) + F(J) \tag{10.6.1}$$

where $\tilde{\nu}_{el}$ is the electronic transition energy change in wavenumbers, $G(n)$ is the vibrational energy with energy level v (assuming anharmonic oscillator), and $F(J)$ is the rotational energy, assuming a nonrigid rotor. Equation 10.6.1 can be expanded accordingly:

$$\tilde{E}_{total} = \underbrace{\tilde{\nu}_{el}}_{\text{electronic}} + \underbrace{\tilde{\nu}_e\left(v+\frac{1}{2}\right) - \tilde{\chi}_e\tilde{\nu}_e\left(v+\frac{1}{2}\right)^2}_{\text{vibrational}} + \underbrace{\tilde{B}J(J+1) - \tilde{D}J^2(J+1)^2}_{\text{rotational}} \tag{10.6.2}$$

Notice that both the vibration constant ($\tilde{\nu}_e$) and anharmonic constant ($\tilde{\chi}_e$) are electronic state dependent (and hence the rotational constants would be too, but are ignored here). Since rotational energies tend to be so small compared to electronic, their effects are minimal and are typically ignored when we do calculations and are referred to as *vibronic* transitions.

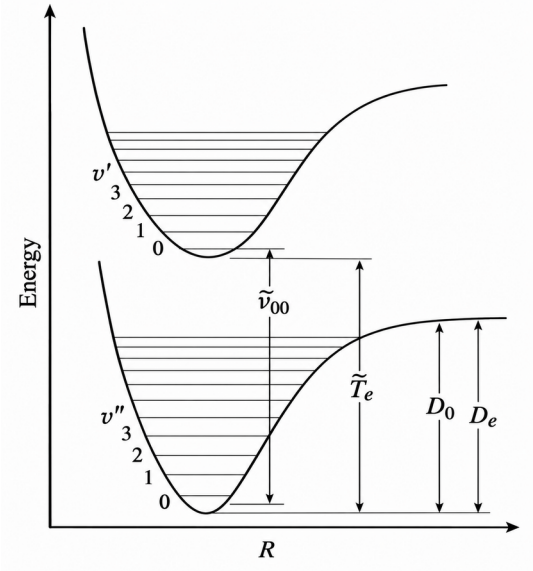


Figure 10.6.2: As you can see, electronic transitions involve two energy potential wells! The vibrational levels are included, but rotational levels are excluded. (CC BY 4.0; LibreTexts)

The eigenstate-to-eigenstate transitions (e.g., $1 \rightarrow 2$) possible are numerous and have absorption lines at

$$\tilde{\nu}_{obs} = \tilde{E}_2 - \tilde{E}_1 \tag{10.6.3}$$

and for simplification, we refer to constants associated with these states as $|'\rangle$ and $|''\rangle$, respectively. So Equation 10.6.3 is

$$\tilde{\nu}_{obs} = E''\tilde{(}v'') - E'\tilde{(}v')$$

Also important to note that typically vibronic transitions are usually the result of the vibrational $v' = 0$ vibratonal state. Within this assumption and excluding the rotational contributions (due to their low energies), Equation 10.6.2 can be used with Equation 10.6.3 to get

$$\tilde{\nu}_{obs} = \tilde{T}_{el} + \left(\frac{1}{2}\tilde{\nu}'_e - \frac{1}{4}\tilde{\chi}'_e\tilde{\nu}'_e\right) - \left(\frac{1}{2}\tilde{\nu}''_e - \frac{1}{4}\tilde{\chi}''_e\tilde{\nu}''_e\right) + \tilde{\nu}'_e v'' - \tilde{\chi}'_e\tilde{\nu}'_e v''(v''+1) \tag{10.6.4}$$

A common transition of importance is the $\tilde{\nu}_{00}$, which is the $0 \rightarrow 0$ transition and include no vibrational change. For this case, equation 10.6.4 is then

$$\tilde{\nu}_{00} = \tilde{T}_{el} + \left(\frac{1}{2}\tilde{\nu}'_e - \frac{1}{4}\tilde{\chi}'_e\tilde{\nu}'_e\right) - \left(\frac{1}{2}\tilde{\nu}''_e - \frac{1}{4}\tilde{\chi}''_e\tilde{\nu}''_e\right)$$

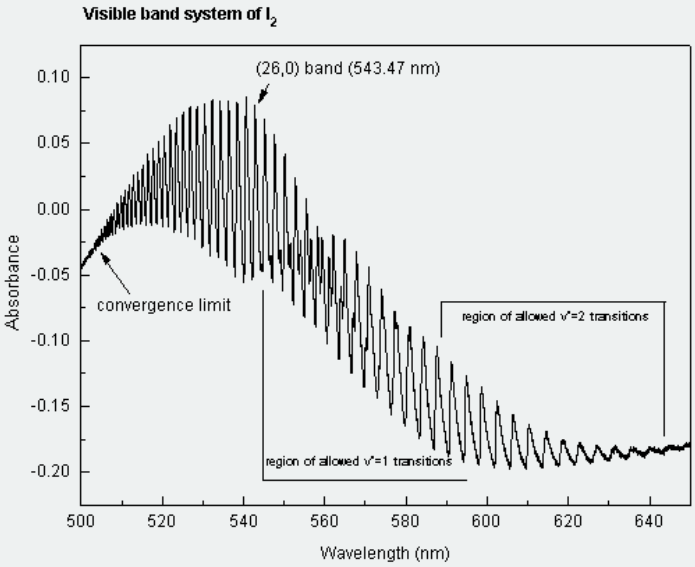
This is the lowest energy possible to observe in an electronic transition although it may be of low intensity as discussed in the following section.

Iodine

The absorption spectrum of iodine yields information about the excited state well rather than the ground state well (notice that equation 10.6.4 depends primarily on excited state parameters). In this experiment you will characterize the excited state well by extracting values for the following excited state parameters.

Recall that as v' increases, the vibrational energy spacing decreases. At the upper edge of the well, the vibrational energy spacing decreases to 0, which means that the energies form a continuum rather than being quantized. It is at this limit that **bond dissociation occurs**. The energy required to dissociate the bond is actually D'_o rather than D'_e because the molecule cannot have less than **the zero point energy**.

The vibrational-electronic spectrum of I_2 in the region from 500-650 nm displays a large number of well-defined bands which, for the most part, correspond to v'<-- 0 transitions connecting the v" = 0 vibrational level of the ground electronic state (denoted as $X^1\Sigma^+$) to many different vibrational levels v' of the excited $B^3\Pi$ electronic state. Under the conditions of this experiment (i.e., low resolution), the rotational lines within each band are not resolved. However, the peaks may be identified as R-branch band heads (1). For a molecule as heavy as I_2, the position of each band head is within a few tenths of one cm^{-1} of the band origin (2), and for the purposes of this experiment, the distinction between the two may be ignored. The general features of the absorption spectrum are shown below:



Each small bump, or peak, such as the (26,0) band labelled on the spectrum, corresponds to a transition between two vibrational levels and is called a *band*. Each band is comprised of several hundred *lines*, each of which involves different upper and lower rotational quantum numbers; as mentioned, these lines are not resolved in the present experiment. The region of maximum absorption in each band is caused by many of these lines falling together; it is called the *band head*. The set of all of these bands is referred to as the *visible band system* of I_2.

If the sample is hot, then excited vibrational levels of the ground state may be populated, and these also will absorb light. The *hot bands* arising from absorption from v"=1 and v"=2 are shown very approximately on the absorption spectrum above.

At a point called the *convergence limit*, the spacing between bands decreases to zero. Beyond this convergence limit, the spectrum is continuous because the excited state of the I_2 molecule is not bound. One of the purposes of this experiment is to identify this convergence limit accurately.

This page titled 10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 13.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information is licensed CC BY 4.0.

10.7: ELECTRONIC SPECTROSCOPY

This page explains what happens when organic compounds absorb UV or visible light, and why the wavelength of light absorbed varies from compound to compound.

Molecular Absorption of Light

When we were talking about the various sorts of molecular orbitals present in organic compounds earlier, you will have come across this diagram showing their relative energies:

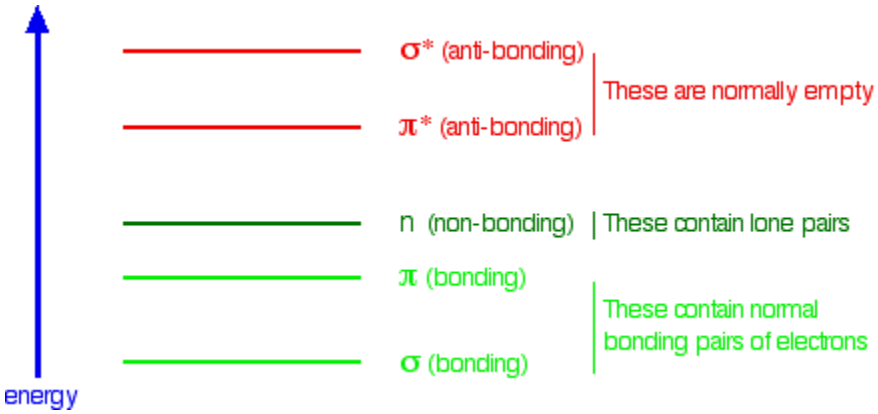


Figure 10.7.1: The diagram is not intended to be to scale - it just shows the relative placing of the different orbitals.

When light passes through the compound, energy from the light is used to promote an electron from a bonding or non-bonding orbital into one of the empty anti-bonding orbitals. The possible electron jumps that light might cause are:

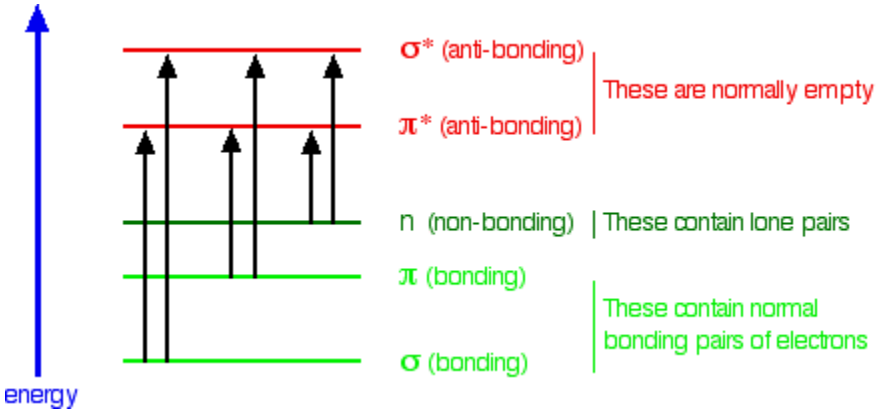


Figure 10.7.2: Promotion of an electron via different electron transitions

In each possible case, an electron is excited from a full orbital into an empty anti-bonding orbital. Each jump takes energy from the light, and a big jump obviously needs more energy than a small one. Each wavelength of light has a particular energy associated with it. If that particular amount of energy is just right for making one of these energy jumps, then that wavelength will be absorbed - its energy will have been used in promoting an electron.

An absorption spectrometer works in a range from about 200 nm (in the near ultra-violet) to about 800 nm (in the very near infra-red). Only a limited number of the possible electron jumps absorb light in that region. Look again at the possible jumps. This time, the important jumps are shown in black, and a less important one in grey. The grey dotted arrows show jumps which absorb light outside the region of the spectrum we are working in.

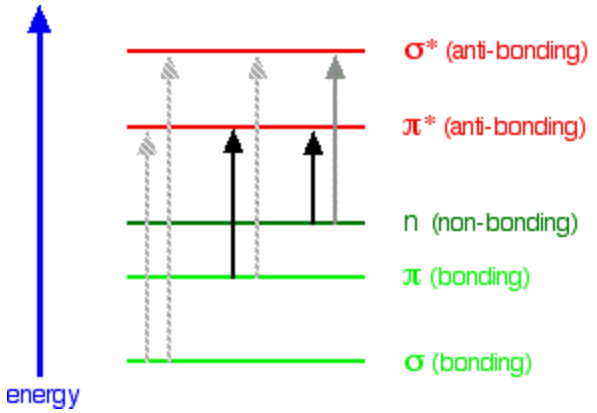


Figure 10.7.3: Promotion of an electron via different electron transitions scaled by energies.

Remember that bigger jumps need more energy and so absorb light with a shorter wavelength. The jumps shown with grey dotted arrows absorb UV light of wavelength less that 200 nm. The important jumps are:

- from π bonding orbitals to π anti-bonding orbitals;
- from non-bonding orbitals to π anti-bonding orbitals;
- from non-bonding orbitals to sigma anti-bonding orbitals.

That means that in order to absorb light in the region from 200 - 800 nm (which is where the spectra are measured), the molecule must contain either π bonds or atoms with non-bonding orbitals. Remember that a non-bonding orbital is a lone pair on, say, oxygen, nitrogen or a halogen. Groups in a molecule which absorb light are known as chromophores.

What does an absorption spectrum look like

The diagram below shows a simple UV-visible absorption spectrum for buta-1,3-diene - a molecule we will talk more about later. Absorbance (on the vertical axis) is just a measure of the amount of light absorbed. The higher the value, the more of a particular wavelength is being absorbed.

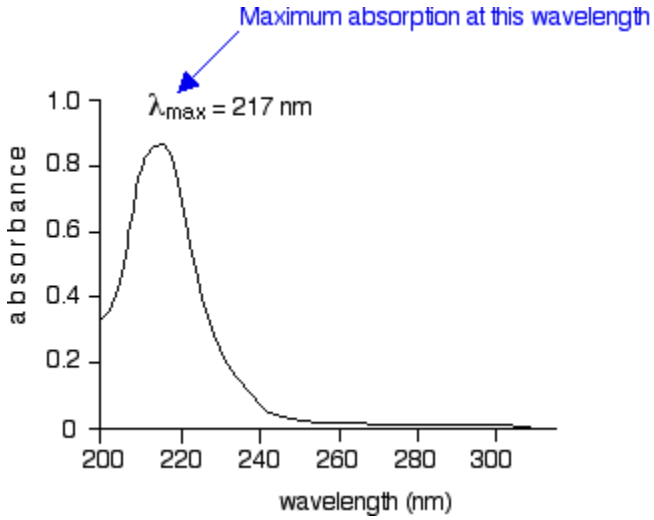


Figure 10.7.4: UV-visible absorption spectrum for buta-1,3-diene

You will see that absorption peaks at a value of 217 nm. This is in the ultra-violet and so there would be no visible sign of any light being absorbed - buta-1,3-diene is colorless. In buta-1,3-diene, CH_2=CH-CH=CH_2, there are no non-bonding electrons. That means that the only electron jumps taking place (within the range that the spectrometer can measure) are from π bonding to π anti-bonding orbitals.

A chromophore such as the carbon-oxygen double bond in ethanal, for example, obviously has π electrons as a part of the double bond, but also has lone pairs on the oxygen atom. That means that both of the important absorptions from the last energy diagram

are possible. You can get an electron excited from a π bonding to a π anti-bonding orbital, or you can get one excited from an oxygen lone pair (a non-bonding orbital) into a π anti-bonding orbital.

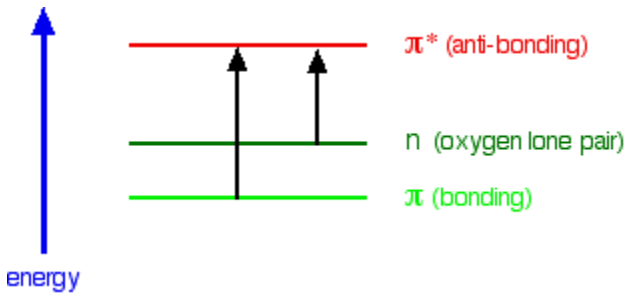


Figure 10.7.4: A simplified diagram showing $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions

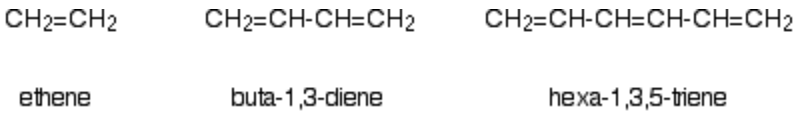
The non-bonding orbital has a higher energy than a π bonding orbital. That means that the jump from an oxygen lone pair into a π anti-bonding orbital needs less energy. That means it absorbs light of a lower frequency and therefore a higher wavelength. For example, ethanal can therefore absorb light of two different wavelengths:

- the π bonding to π anti-bonding absorption peaks at 180 nm. These $n \rightarrow \pi^*$ transitions involve moving an electron from a nonbonding electron pair to a antibonding *pi^*\) orbital. They tend to have molar absorbtivities less than 2000
- the non-bonding to π anti-bonding absorption peaks at 290 nm. These $\pi \rightarrow \pi^*$ transitions involve moving an electron from a bonding $\pi *$ orbital to an antibonding π^* orbital. They tend to have molar absorptivities on the order of 10,000.

Both of these absorptions are in the ultra-violet, but most spectrometers will not pick up the one at 180 nm because they work in the range from 200 - 800 nm.

Conjugation and Delocalization

Consider these three molecules:



Ethene contains a simple isolated carbon-carbon double bond, but the other two have conjugated double bonds. In these cases, there is delocalization of the π bonding orbitals over the whole molecule. Now look at the wavelengths of the light which each of these molecules absorbs.

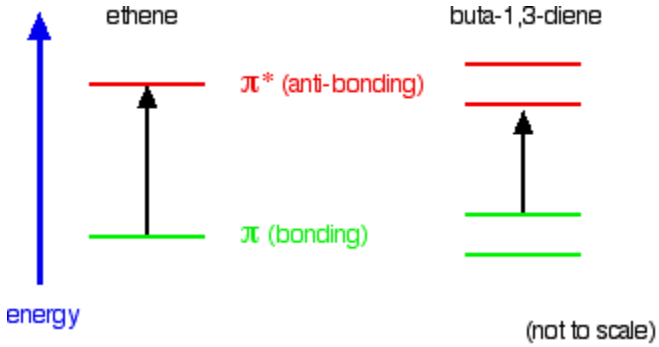
molecule	wavelength of maximum absorption (nm)
ethene	171
buta-1,3-diene	217
hexa-1,3,5-triene	258

All of the molecules give similar UV-visible absorption spectra - the only difference being that the absorptions move to longer and longer wavelengths as the amount of delocalization in the molecule increases. Why is this? You can actually work out what must be happening.

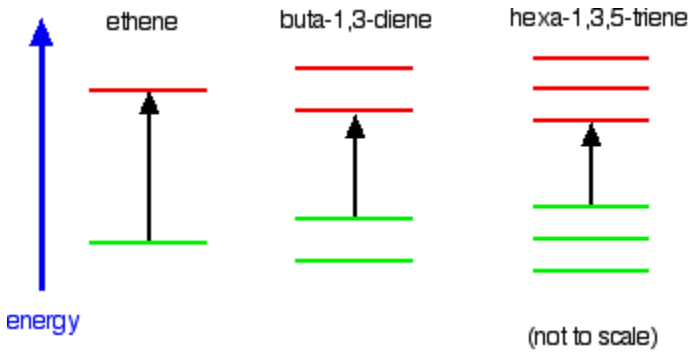
- The maximum absorption is moving to longer wavelengths as the amount of delocalization increases.
- Therefore maximum absorption is moving to shorter frequencies as the amount of delocalization increases.
- Therefore absorption needs less energy as the amount of delocalization increases.
- Therefore there must be less energy gap between the bonding and anti-bonding orbitals as the amount of delocalization increases.

. . . and that's what is happening.

Compare ethene with buta-1,3-diene. In ethene, there is one π bonding orbital and one π anti-bonding orbital. In buta-1,3-diene, there are two π bonding orbitals and two π anti-bonding orbitals. This is all discussed in detail on the introductory page that you should have read.



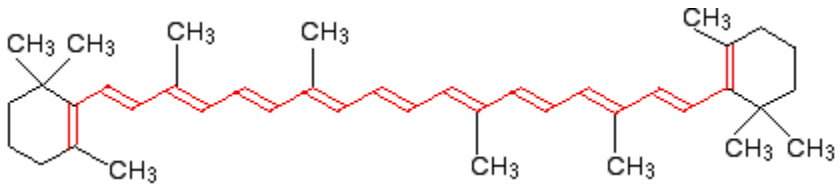
The highest occupied molecular orbital is often referred to as the HOMO - in these cases, it is a π bonding orbital. The lowest unoccupied molecular orbital (the LUMO) is a π anti-bonding orbital. Notice that the gap between these has fallen. It takes less energy to excite an electron in the buta-1,3-diene case than with ethene. In the hexa-1,3,5-triene case, it is less still.



If you extend this to compounds with really massive delocalization, the wavelength absorbed will eventually be high enough to be in the visible region of the spectrum, and the compound will then be seen as colored. A good example of this is the orange plant pigment, beta-carotene - present in carrots, for example.

Why is beta-carotene orange?

Beta-carotene has the sort of delocalization that we've just been looking at, but on a much greater scale with 11 carbon-carbon double bonds conjugated together. The diagram shows the structure of beta-carotene with the alternating double and single bonds shown in red.

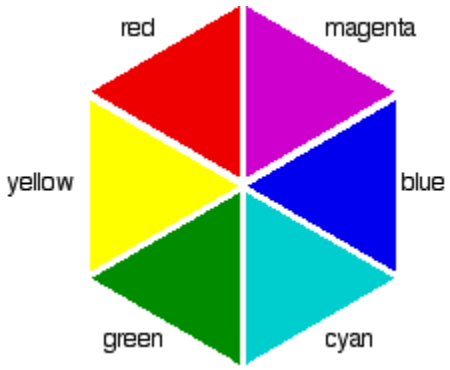


The more delocalization there is, the smaller the gap between the highest energy π bonding orbital and the lowest energy π anti-bonding orbital. To promote an electron therefore takes less energy in beta-carotene than in the cases we've looked at so far - because the gap between the levels is less. Remember that less energy means a lower frequency of light gets absorbed - and that's equivalent to a longer wavelength. Beta-carotene absorbs throughout the ultra-violet region into the violet - but particularly strongly in the visible region between about 400 and 500 nm with a peak about 470 nm. The wavelengths associated with the various colors are approximately:

color region	wavelength (nm)

violet	380 - 435
blue	435 - 500
cyan	500 - 520
green	520 - 565
yellow	565 - 590
orange	590 - 625
red	625 - 740

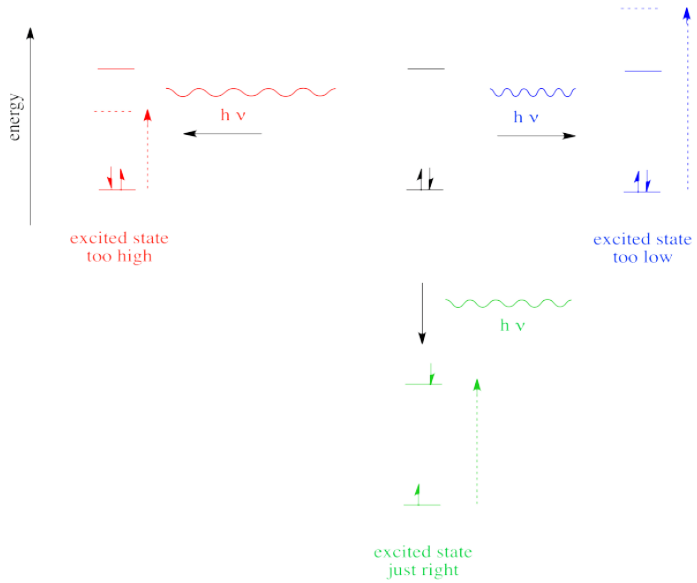
So if the absorption is strongest in the violet to cyan region, what color will you actually see? It is tempting to think that you can work it out from the colors that are left - and in this particular case, you wouldn't be far wrong. Unfortunately, it is not as simple as that! Sometimes what you actually see is quite unexpected. Mixing different wavelengths of light does not give you the same result as mixing paints or other pigments. You can, however, sometimes get some estimate of the color you would see using the idea of complementary colors. If you arrange some colors in a circle, you get a "color wheel". The diagram shows one possible version of this. An internet search will throw up many different versions!



colors directly opposite each other on the color wheel are said to be complementary colors. Blue and yellow are complementary colors; red and cyan are complementary; and so are green and magenta. Mixing together two complementary colors of light will give you white light. What this all means is that if a particular color is absorbed from white light, what your eye detects by mixing up all the other wavelengths of light is its complementary color. In the beta-carotene case, the situation is more confused because you are absorbing such a range of wavelengths. However, if you think of the peak absorption running from the blue into the cyan, it would be reasonable to think of the color you would see as being opposite that where yellow runs into red - in other words, orange.

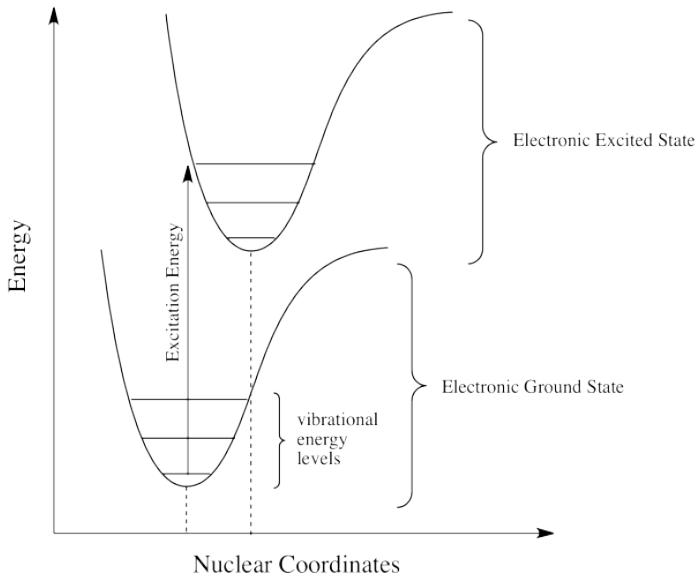
Franck-Condon: Electronic and Vibrational Coupling

So far, we have come across one big rule of photon absorbance. To be absorbed, a photon's energy has to match an energy difference within the compound that is absorbing it.



In the case of visible or ultraviolet light, the energy of a photon is roughly in the region that would be appropriate to promote an electron to a higher energy level. Different wavelengths would be able to promote different electrons, depending on the energy difference between an occupied electronic energy level and an unoccupied one. Other types of electromagnetic radiation would not be able to promote an electron, but they would be coupled to other events. For example, absorption of infrared light is tied to vibrational energy levels. Microwave radiation is tied to rotational energy levels in molecules. Thus, one reason a photon may or may not be absorbed has to do with whether its energy corresponds to the available energy differences within the molecule or ion that it encounters.

Photons face other limitations. One of these is a moderate variation on our main rule. It is called the Frank Condon Principle. According to this idea, when an electron is excited from its normal position, the ground state, to a higher energy level, the optimal positions of atoms in the molecule may need to shift. Because electronic motion is much faster than nuclear motion, however, any shifting of atoms needed to optimize positions as they should be in the excited state will have to wait until after the electron gets excited. In that case, when the electron lands and the atoms aren't yet in their lowest energy positions for the excited state, the molecule will find itself in an excited vibrational state as well as an excited electronic state.



That means the required energy for excitation does not just correspond to the difference in electronic energy levels; it is fine-tuned to reach a vibrational energy level, which is quantized as well.

- The Franck Condon Principle states that electronic transitions are vertical.

- A vertical transition is one in which non of the nuclei move while the electron journeys from one state to another.
- A vertical transition may begin in a vibrational ground state of an electronic ground state and end in a vibrational excited state of an electronic excited state.

Contributors and Attributions

- Chris P Schaller, Ph.D., (College of Saint Benedict / Saint John's University)
- Jim Clark (Chemguide.co.uk)

10.7.1: Spectroscopy of the Particle-in-a-Box Model

10.7.2: Cyanine Dyes

10.7.3: Fluorescence and Phosphorescence

This page titled 10.7: Electronic Spectroscopy is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 14.4: Electronic Spectroscopy is licensed CC BY-NC-SA 4.0.

10.7.1: Spectroscopy of the Particle-in-a-Box Model

The wavefunctions that describe electrons in atoms and molecules are called orbitals. An orbital is a wavefunction for a single electron. When we say an electron is in orbital n, we mean that it is described by a particular wavefunction Ψ_n and has energy E_n. All the properties of this electron can be calculated from Ψ_n as described in Chapter 3.

We will now use the particle-in-a-box model to explain the absorption spectra of the cyanine dyes. When an atom or molecule absorbs a photon, the atom or molecule goes from one energy level, designated by quantum number n_i, to a higher energy level, designated by n_f. We can also say that the molecule goes from one electronic state to another. This change is called a transition. Sometimes it is said that an electron goes from one orbital to another in a transition, but this statement is not general. It is valid for a particle-in-a-box, but not for real atoms and molecules, which are more complicated than the simple particle-in-a-box model.

The energy of the photon absorbed ($E_{photon} = h\nu$) matches the difference in the energy between the two states involved in the transition (ΔE_{states}). In general, the observed frequency or wavelength for a transition is calculated from the change in energy using the following equalities,

$$\Delta E_{states} = E_f - E_i = E_{photon} = h\nu = hc\bar{\nu} \tag{10.7.1.1}$$

Then, for the specific case of the particle-in-a-box,

$$E_{photon} = \Delta E_{states} = E_f - E_i = \frac{(n_f^2 - n_i^2)h^2}{8mL^2} \tag{10.7.1.2}$$

where n_f is the quantum number associated with the final state and n_i is the quantum number for the initial state. A negative value for E_{photon} means the photon is emitted as a result of the transition in states; a positive value means the photon is absorbed.

Generally the transition energy, E_{photon} or ΔE_{states}, is taken to correspond to the peak in the absorption spectrum. When high accuracy is needed for the electronic transition energy, the spectral line shape must be analyzed to account for rotational and vibrational motion as well as effects due to the solvent or environment. Contributions of rotational and vibrational motion to an absorption spectrum will be discussed in later chapters.

In a cyanine dye molecule that has three carbon atoms in the chain, there are six π-electrons. When light is absorbed, one of these electrons increases its energy by an amount $h\nu$ and jumps to a higher energy level. In order to use Equation 10.7.1.1, we need to know which energy levels are involved.

We assign the electrons to the lowest energy levels to create the ground-state lowest-energy electron configuration. We could put all six electrons in the $n = 1$ level, or we could put one electron in each of $n = 1$ through $n = 6$, or we could pair the electrons in $n = 1$ through $n = 3$, etc.

The Pauli Exclusion Principle says that each spatial wavefunction can describe, at most, two electrons, or in other words, that each energy level can have only two electrons assigned to it. Spatial refers to our 3-dimensional space, and a spatial wavefunction depends upon the spatial coordinates x, y, or z. We will discuss the Pauli Exclusion Principle more fully later, but you probably have encountered it in other courses.

Rather than appeal to the Pauli Exclusion Principle to assign the electrons to the energy levels, let's try an empirical approach and discover the Pauli Exclusion Principle as a result. Assign the electrons to the energy levels in different ways until you find an assignment that agrees with experiment.

When there is an even number of electrons, the lowest-energy transition is the energy difference between the highest occupied level (HOMO) and the lowest unoccupied level (LUMO). HOMO designates the highest-energy occupied molecular orbital, and LUMO designates the lowest-energy unoccupied molecular orbital. The term orbital refers to the wavefunction or energy level for one electron. All other transitions have a higher energy. For the case with all the electrons in the first energy level, the lowest-energy transition energy would be $h\nu = E_2–E_1$. With one electron in each of the first six levels, $h\nu = E_7–E_6$, and with the electrons paired, $h\nu = E_4–E_3$.

Example 10.7.1.1

Draw energy level diagrams indicating the HOMO, the LUMO, the electrons and the lowest energy transition for each of the three cases mentioned in the preceding paragraph.

Example 10.7.1.2

For the three ways of assigning the 6 electrons to the energy levels in Exercise 10.7.1.17, calculate the peak absorption wavelength λ for a cyanine dye molecule with 3 carbon atoms in the chain using a value for L of 0.849 nm, which is obtained by estimating bond lengths. Which wavelength agrees most closely with the experimental value of 309 nm for this molecule?

It turns out that the assignment that gives a reasonable wavelength for the absorption of a cyanine dye with 6 π electrons is $h\nu = E_4 – E_3$ as you concluded from Exercise 10.7.1.2. In this way we have "discovered" the Pauli Exclusion Principle, electrons should be paired in the same energy level whenever possible, and we accept it for now because it agrees with the experimental observations of the cyanine dye spectra.

In molecules with an odd number of electrons, it is possible to have transitions between the doubly occupied molecular orbitals and the singly occupied molecular orbital as well as from the singly occupied orbital to an unoccupied orbital.

This page titled 10.7.1: Spectroscopy of the Particle-in-a-Box Model is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by David M. Hanson, Erica Harvey, Robert Sweeney, Theresa Julia Zielinski via source content that was edited to the style and standards of the LibreTexts platform.

- 4.4: Spectroscopy of the Particle-in-a-Box Model by David M. Hanson, Erica Harvey, Robert Sweeney, Theresa Julia Zielinski is licensed CC BY-NC-SA 4.0. Original source: https://web.archive.org/web/20200619182410/http://www.chemeddl.org.

10.7.2: Cyanine Dyes

Cyanine dye molecules, which have the general structure shown in Figure $10.7.2.1$, are planar cations. The number of carbon atoms in the chain can vary, as can the nature of the end groups containing the nitrogen atoms. The R groups in the diagram represent H, CH_3, CH_3CH_2, or many other moieties including ring structures. Since these dyes are cations, they can be paired with many anions, e.g. I^-, iodide. The position (wavelength) and strength (absorption coefficient) of the absorption band depends upon the length of the carbon chain between the nitrogen atoms but is not affected very much by the nature of the end groups beyond the nitrogen atoms.

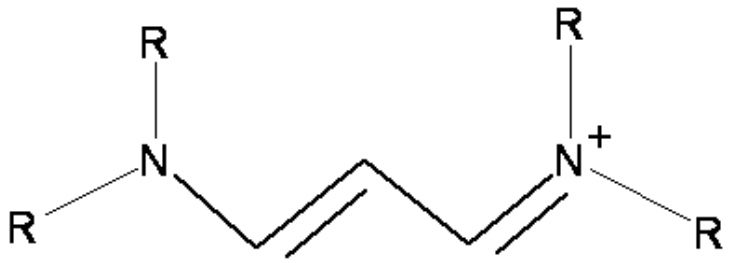


Figure $10.7.2.2$ if $R = CH_3$).

These molecules are called dye molecules because they have very intense absorption bands in the visible region of the spectrum as shown in Figure $10.7.2.2$. This strong absorption of light at particular wavelengths makes solutions of these molecules brightly colored. A solution of a dye shows the color of the light that is not absorbed. The strong absorption leads to many applications in technology. For example, dyes are used to color plastics, fabrics, and hair. They also can be used as filters to produce colored light and as a laser medium in medical applications.

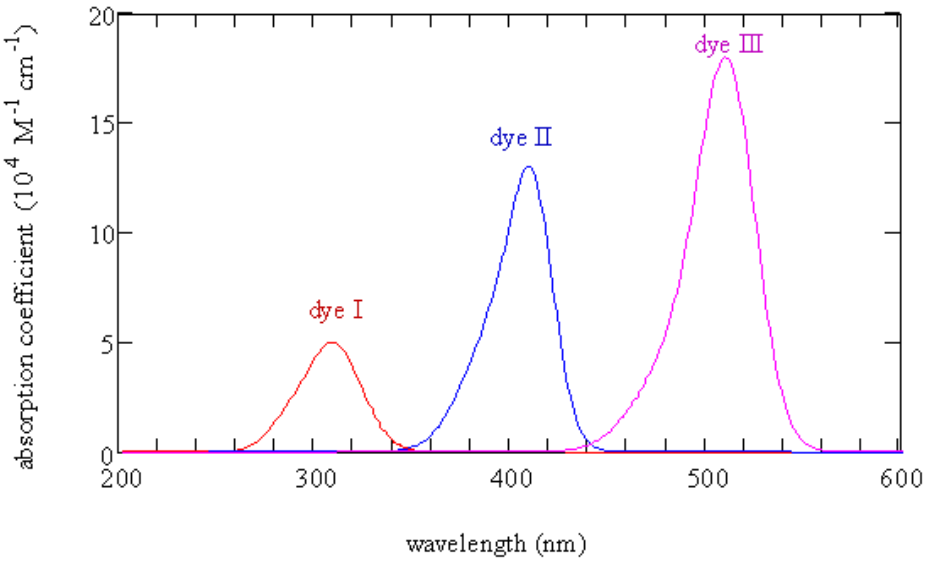


Figure $10.7.2.2$: Absorption spectra of 3 cyanine dyes constructed from data in the paper by W.T. Simpson, J. Chem. Phys. Series A, 1948, 16, p 1124. $R = CH_3$ for each dye. Dye I has 3 carbon atoms and the absorption maximum is at 309 nm, dye II has 5 carbon atoms and the absorption maximum, λ_{max}, is at 409 nm, and dye III has 7 carbon atoms and the absorption maximum is at 512 nm.

Example $10.7.2.1$

Draw the Lewis electron dot structure of dye I that produced the spectrum shown in Figure $10.7.2.2$ with the maximum absorption at 309 nm. Examine the resonance structures and determine whether all the carbon-carbon bonds are identical or whether some are single bonds and some are double bonds.

Example 10.7.2.2

Use Figure 10.7.2.2 to describe what happens to the maximum absorption coefficient and the wavelength of the peak absorption as the length of a cyanine dye molecule increases.

Example 10.7.2.3

The decadic molar absorption coefficient for dye III at λ = 512 nm is almost 200,000 in units of $1000 cm^2/mol$. If 0.1 gram of dye III (very small) were dissolved in 10 liters of water (very large), what fraction of light at 512 nm would be absorbed in a path length of 1 mm (very small)? What is the concentration of this solution? What insight do you gain from your results? (Note: the molar mass is 179 g/mol.)

The electrons and bonds in the cyanine dyes can be classified as sigma or pi. The probability densities for the sigma electrons are large along the lines connecting the nuclei, while the probability densities for the pi electrons are large above and below the plane containing the nuclei. In molecular orbital theory, the π electrons can be described by wavefunctions composed from p_z atomic orbitals, shown in Figure 10.7.2.3.

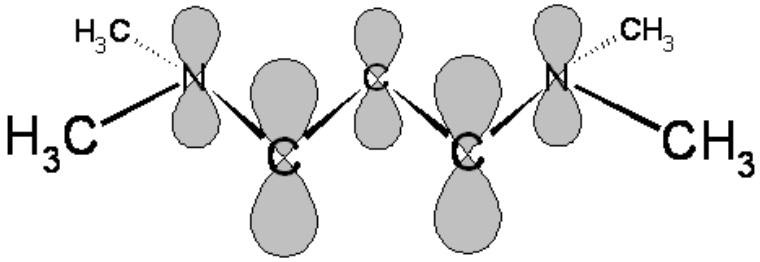


Figure 10.7.2.3: Carbon and nitrogen p_z-orbitals of a cyanine dye (dye I in Figure 10.7.2.2).

Example 10.7.2.4

Determine the number of pi electrons in each of the three molecules described in Figure 10.7.2.2.

The pi electrons in these molecules, one from each carbon atom and three from the two nitrogen atoms, are delocalized over the length of the molecule between the nitrogen atoms. When ultraviolet and visible light is absorbed by the cyanine dyes, the energy is used to cause transitions of the pi electrons from one energy level to another, as sketched in Figure 10.7.2.4. The longest wavelength transition occurs from the highest-energy occupied level to the lowest-energy unoccupied level. We will use Quantum Mechanics and a simple model, called the particle-in-a-box model, to explain why the longer molecules absorb at longer wavelengths and have larger absorption coefficients. This analysis will demonstrate that Quantum Mechanics is a quantitative theory. It provides both a qualitative understanding of chemical systems and numerical values for the properties of interest. Both are important for understanding molecules and their chemistry.

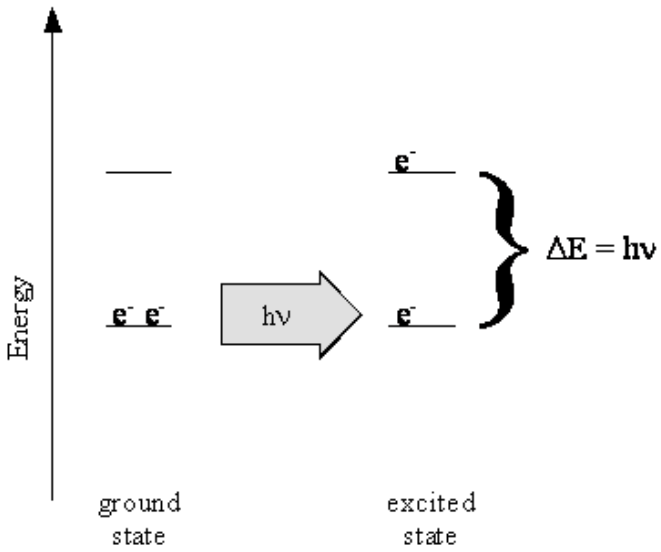


Figure 10.7.2.4: The transition of an electron from a lower energy level to a higher one absorbs energy and produces the observed absorption spectrum.

You can visualize the absorption of energy and the promotion of an electron from a lower energy level to a higher one like throwing a shirt from your closet floor to a shelf. There is an important difference however. You can see the shirt on the way from the floor to the shelf. You can tell when it left the floor, and when it arrived on the shelf. Such precise information cannot be obtained for the electron. We only know the probability that the electron is in the lower level and the probability that it is in the higher level as a function of time. We do not know exactly when during this period that the electron makes the transition from one energy level to the other.

We can imagine that the potential energy experienced by the pi electron varies along the chain as shown in Figure 10.7.2.5 effectively trapping the electron in the pi region of the molecule, i.e. in a one-dimensional box. At the end of the chain the potential energy rises to a large value. The particle-in-a-box model essentially consists of three approximations to the actual potential energy.

1. The chain of carbon atoms forms a one-dimensional space of some length L for the pi electrons.
2. The potential energy is constant along the chain; i.e. the oscillations are ignored.
3. The potential energy becomes infinite at some point slightly past the nitrogen atoms.

Since only changes in energy are meaningful, and an absolute zero of energy does not exist, the constant potential energy of the electron along the chain between the nitrogen atoms can be defined as zero. The particle-in-a-box potential energy also is shown in Figure 10.7.2.5.

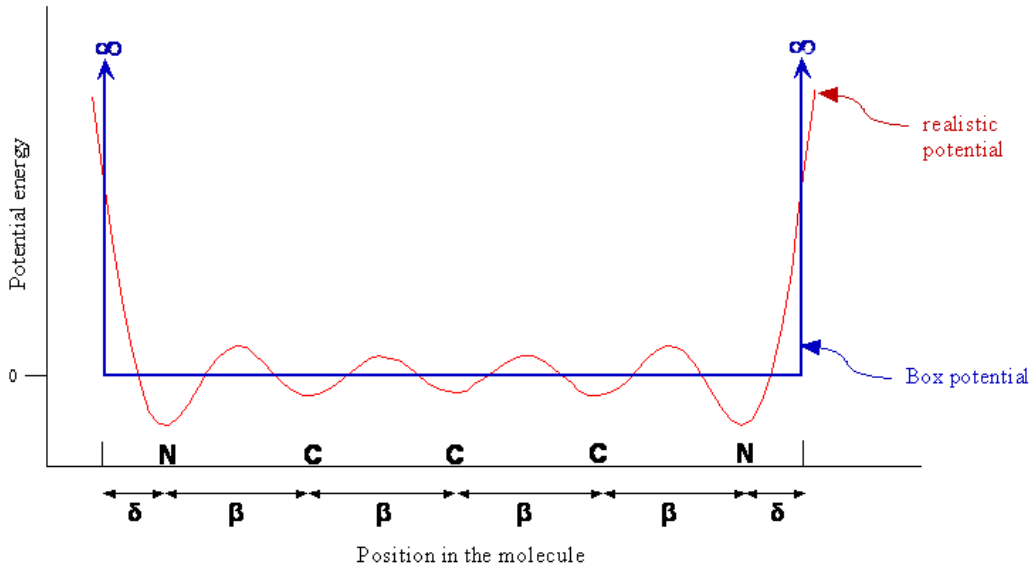


Figure 10.7.2.5: A diagram of the particle-in-a-box potential energy superimposed on a somewhat more realistic potential. The bond length is given by β, the overshoot by δ, and the length of the box by L = bβ + 2δ, where b is the number of bonds. For this example, dye I in Figure 10.7.2.1, L = 4β + 2δ.

The particle-in-a-box model allows us to relate the transition energy obtained from the position of the absorption maximum to the length of the conjugated part of the molecule, i.e. distance L between the infinite potential barriers at the ends of the molecule. From this distance for different series of dyes, we can obtain the average bond length β and the distance δ the box extends beyond a nitrogen atom for each series. If this model is reasonable, we expect the average bond lengths to be similar for each series and δ to vary from one series to another due to differences in the end groups attached to the nitrogen atoms. Some end groups might, due to their polarizability or electronegativity, allow the electrons to penetrate further past the nitrogen atoms than others. Analyzing the data in this way rather than using estimated bond lengths to predict transition energies was suggested by R.S. Moog. (J. Chem Educ. 1991, 68, 506-508.)

Example 10.7.2.5

In Figure 10.7.2.5, why does a realistic potential energy dip at each atom? Why is the dip larger for nitrogen than for carbon? Why does the potential energy increase sharply at the ends of the molecule?

This page titled 10.7.2: Cyanine Dyes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by David M. Hanson, Erica Harvey, Robert Sweeney, Theresa Julia Zielinski via source content that was edited to the style and standards of the LibreTexts platform.

- 4.2: Cyanine Dyes by David M. Hanson, Erica Harvey, Robert Sweeney, Theresa Julia Zielinski is licensed CC BY-NC-SA 4.0. Original source: https://web.archive.org/web/20200619182410/http://www.chemeddl.org.

10.7.3: Fluorescence and Phosphorescence

Singlet and Triplet Excited State

Understanding the difference between fluorescence and phosphorescence requires the knowledge of electron spin and the differences between singlet and triplet states. The Pauli Exclusion principle states that two electrons in an atom cannot have the same four quantum numbers (n, l, m_l, m_s) and only two electrons can occupy each orbital where they must have opposite spin states. These opposite spin states are called spin pairing. Because of this spin pairing, most molecules do not exhibit a magnetic field and are diamagnetic. In diamagnetic molecules, electrons are not attracted or repelled by the static electric field. Free radicals are paramagnetic because they contain unpaired electrons have magnetic moments that are attracted to the magnetic field.

A singlet state is defined as a state when all the electron spins are paired in the molecular electronic state and the electronic energy levels do not split when the molecule is exposed to a magnetic field. A doublet state occurs when there is an unpaired electron that can have two possible orientations when exposed to a magnetic field and imparts different energy to the system. A singlet or a triplet can form when one electron is excited to a higher energy level. In an excited singlet state, the electron is promoted with the same spin orientation as it had in the ground state (paired). In a triplet excited stated, the electron that is promoted has the same spin orientation (parallel) as the other unpaired electron. The difference between the spins of ground singlet, excited singlet, and excited triplet is shown in Figure $10.7.3.1$. Singlet, doublet and triplet is derived using the equation for multiplicity, 2S+1, where S is the total spin angular momentum (sum of all the electron spins). Individual spins are denoted as spin up (s = +1/2) or spin down (s = -1/2). If we were to calculated the S for the excited singlet state, the equation would be 2(+1/2 + -1/2)+1 = 2(0)+1 = 1, therefore making the center orbital in the figure a singlet state. If the spin multiplicity for the excited triplet state was calculated, we obtain 2(+1/2 + +1/2)+1 = 2(1)+1 =3, which gives a triplet state as expected.

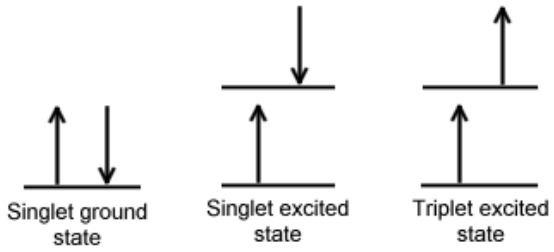


Figure $10.7.3.1$: Spin in the ground and excited states

The difference between a molecule in the ground and triplet excited state is that the electrons are diamagnetic in the ground state and paramagnetic in the excited triplet state.This difference in spin state makes the transition from singlet to triplet (or triplet to singlet) more improbable than the singlet-to-singlet transitions because it involves a spin flip. This singlet to triplet (or reverse) transition involves a change in *electronic* state. For this reason, the lifetime of the triplet state is longer than that of the singlet state by an approximately 10^4 fold difference. Absorption of radiation inducing the transition from ground to excited triplet state has a low probability, thus their absorption bands are less intense than singlet-singlet state absorption. The excited triplet state can be populated from the excited singlet state of certain molecules which results in phosphorescence. These spin multiplicities in ground and excited states can be used to explain transition in photoluminescence molecules by the Jablonski diagram.

Jablonski Diagrams

The Jablonski diagram that drawn below is a partial energy diagram that represents the energy of photoluminescent molecule in its different energy states. The lowest and darkest horizontal line represents the ground-state electronic energy of the molecule which is the singlet state labeled as S_o. At room temperature, majority of the molecules in a solution are in this state.

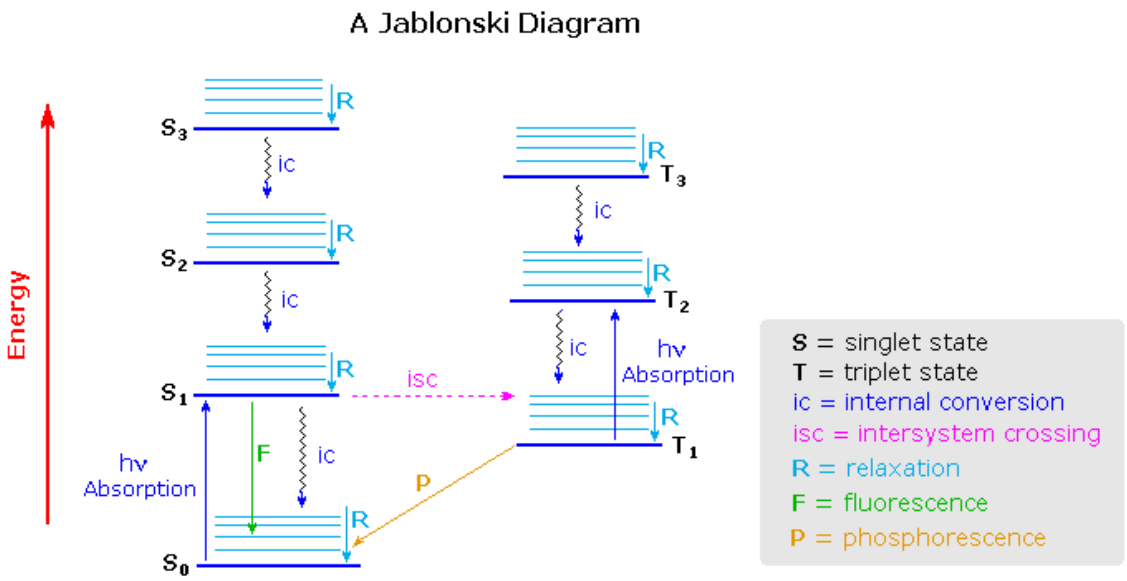


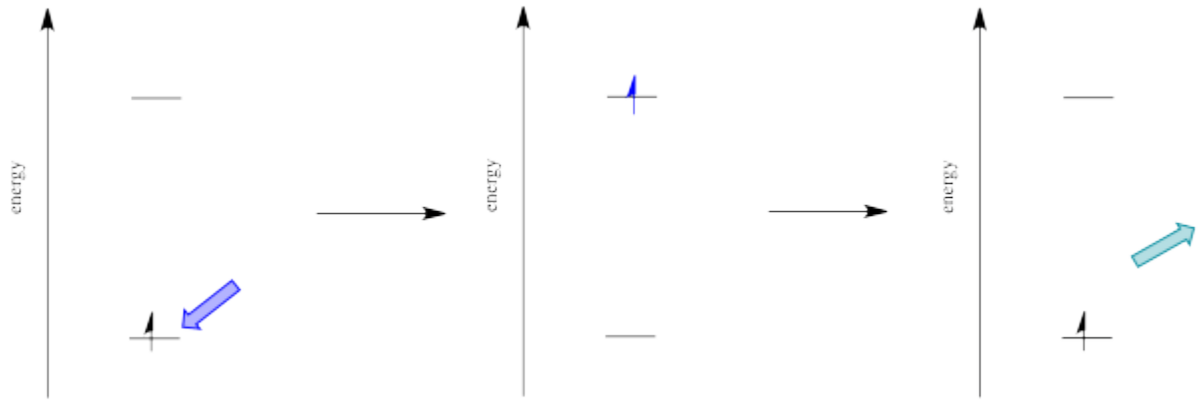
Figure 10.7.3.2: Partial Jablonski Diagram for Absorption, Fluorescence, and Phosphorescence. from Bill Reusch.

The upper lines represent the energy state of the three excited electronic states: S_1 and S_2 represent the electronic singlet state (left) and T_1 represents the first electronic triplet state (right). The upper darkest line represents the ground vibrational state of the three excited electronic state.The energy of the triplet state is lower than the energy of the corresponding singlet state.

There are numerous vibrational levels that can be associated with each electronic state as denoted by the thinner lines. Absorption transitions (blues lines in Figure 10.7.3.2) can occur from the ground singlet electronic state (S_o) to various vibrational levels in the singlet excited vibrational states. It is unlikely that a transition from the ground singlet electronic state to the triplet electronic state because the electron spin is parallel to the spin in its ground state (Figure 10.7.3.1). This transition leads to a change in multiplicity and thus has a low probability of occurring thus it is a partially **forbidden transition.** Molecules also undergo vibrational relaxation to lose any excess vibrational energy that remains when excited to the electronic states (S_1 and S_2) as demonstrated in light blue arrows labelled R in Figure 10.7.3.2. The knowledge of forbidden transition is used to explain and compare the peaks of absorption and emission.

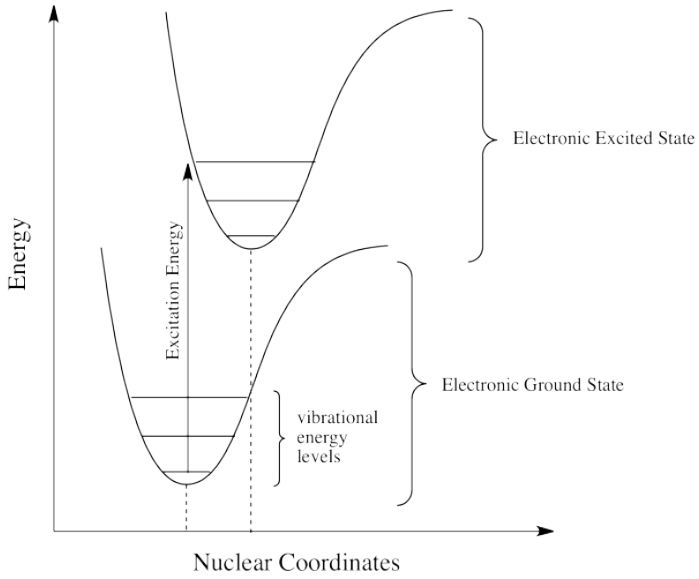
Relaxation and Fluorescence

Sometimes, when an excited state species relaxes, giving off a photon, the wavelength of the photon is different from the one that initially led to excitation. When this happens, the photon is invariably red-shifted; its wavelength is longer than the initial one. This situation is called "fluorescence".

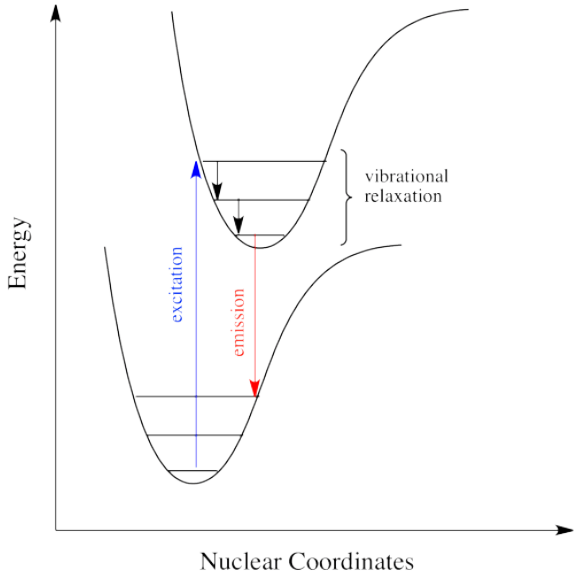


How can that be? Isn't energy quantized? How is the molecule suddenly taking a commission out of the energy the original photon brought with it? This discrepancy is related to the Franck-Condon principle from the previous page. When an electron is promoted to an electronic excited state, it often ends up in an excited vibrational state as well. Thus, some of the energy put into

electronic excitation is immediately passed into vibrational energy. Vibrational energy, however, doesn't just travel in photons. It can be gained or lost through molecular collisions and heat transfer.



The electron might simply drop down again immediately; a photon would be emitted of exactly the same wavelength as the one that was previously absorbed. On the other hand, if the molecule relaxes into a lower vibrational state, some of that initial energy will have been lost as heat. When the electron relaxes, the distance back to the ground state is a little shorter. The photon that is emitted will have lower energy and longer wavelength than the initial one.



Just how does a molecule undergo vibrational relaxation? Vibrational energy is the energy used to lengthen or shorten bonds, or to widen or squeeze bond angles. Given a large enough molecule, some of this vibrational energy could be transferred into bond lengths and angles further away from the electronic transition. Otherwise, if the molecule is small, it may transfer some of its energy in collisions with other molecules.

Note

There are many examples of energy being transferred this way in everyday life. In a game of pool, one billiard ball can transfer its energy to another, sending it toward the pocket. Barry Bonds can transfer a considerable amount of energy through his bat into a baseball, sending it out of the park, just as Serena Williams can send a whole lot of energy whizzing back at her sister.

Exercise 1

How does the energy of an electronic absorption compare to other processes? To find out, you might consider the excitation of an entire mole of molecules, rather than a single molecule absorbing a single photon. Calculate the energy in kJ/mol for the following transitions.

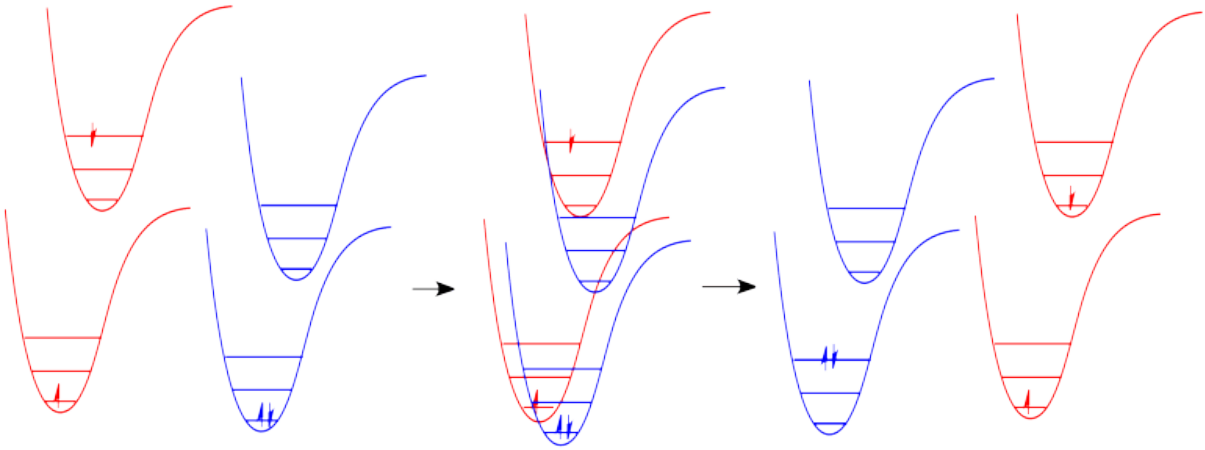
a. absorbance at 180 nm (ultraviolet)
b. absorbance at 476 nm (blue)
c. absorbance at 645 nm (red)

Exercise 2

How does the energy of an excitation between vibrational states compare to that of an electronic excitation? Typically, infrared absorptions are reported in cm^{-1}, which is simply what it looks like: the reciprocal of the wavelength in cm. Because wavelength and frequency are inversely related, wavenumbers are considered a frequency unit. Calculate the energy in kJ/mol for the following transitions.

a. absorbance at 3105 cm^{-1}
b. absorbance at 1695 cm^{-1}
c. absorbance at 963 cm^{-1}

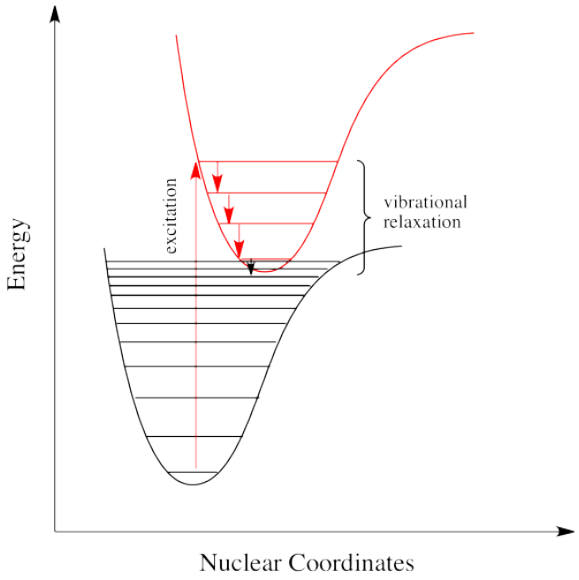
In molecules, as one molecule drops to a lower vibrational state, the other will hop up to a higher vibrational state with the energy it gains. In the drawing below, the red molecule is in an electronic excited and vibrational state. In a collision, it transfers some of its vibrational energy to the blue molecule. Energy can also be transferred from vibration to the kinetic energy of the molecules.



Radiationless Transitions: Internal Conversion

If electrons can get to a lower energy state, and give off a little energy at a time, by hopping down to lower and lower vibrational levels, do they need to give off a giant photon at all? Maybe they can relax all the way down to the ground state via vibrational

relaxation. That is certainly the case. Given lots of vibrational energy levels, and an excited state that is low enough in energy so that some of its lower vibrational levels overlap with some of the higher vibrational levels of the ground state, the electron can hop over from one state to the other, without releasing a photon.



This event is called a "radiationless transition", because it occurs without release of a photon. The electron simply slides over from a low vibrational state of the excited electronic state to a high vibrational state of the electronic ground state. If the electron simply keeps dropping a vibrational level at a time back to the ground state, the process is called **"internal conversion"**.

Internal conversion has an important consequence. Because the absorption of UV and visible light can result in energy transfer into vibrational states, much of the energy that is absorbed from these sources is converted into heat. That can be a good thing if you happen to be a marine iguana trying to warm up in the sun after a plunge in the icy Pacific. It can also be a tricky thing if you are a process chemist trying to scale up a photochemical reaction for commercial production of a pharmaceutical, because you have to make sure the system has adequate cooling available.

Radiationless Transitions: Intersystem Crossing

There is a very similar event, called "intersystem crossing", that leads to the electron getting caught between the excited state and the ground state. Just as, little by little, vibrational relaxation can lead the electron back onto the ground state energy surface, it can also lead the electron into states that are intermediate in energy.

For example, suppose an organic molecule undergoes electronic excitation. Generally, organic molecules have no unpaired electrons. Their ground states are singlet states. According to one of our selection rules for electronic excitation, the excited state must also have no unpaired electrons. In other words, the spin on the electron that gets excited is the same after excitation as it was before excitation.

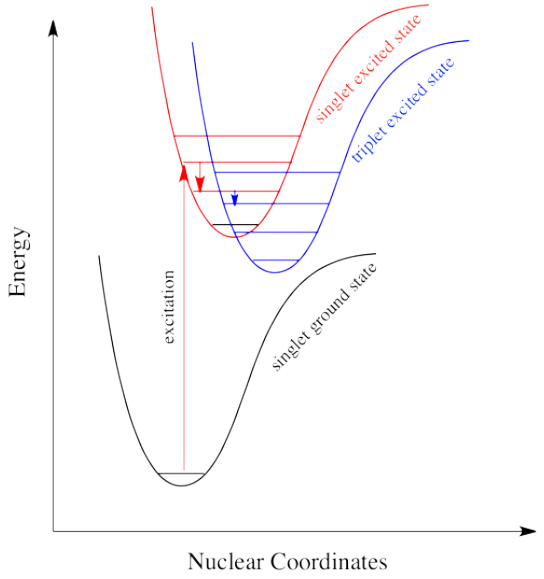
However, that's not the lowest possible energy state for that electron. When we think about atomic orbital filling, there is a rule that governs the spin on the electrons in degenerate orbitals: in the lowest energy state, **spin is maximized** (Hund's rule). In other words, when we draw a picture of the valence electron configuration of nitrogen, we show nitrogen's three p electrons each in its own orbital, with their spins parallel.



The picture with three unpaired electrons, all with parallel spins, shows a nitrogen in the quartet spin state. Having one of those spins point the other way would result in a different spin state. One pair of electrons in the p level would be spin-paired, one

up and one down, even though they are in different p orbitals. That would leave one electron without an opposite partner. The nitrogen would be in a doublet spin state. That is not what happens. The spin state on the left is **lower** in energy than the state on the right. That's just one of the rules of quantum mechanics (**Hund's rule**): maximize spin when orbitals are singly occupied.

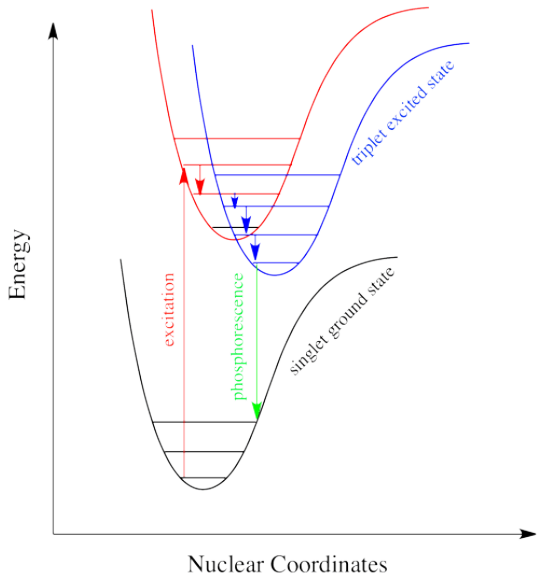
It's the same in a molecule with the triplet state lower in energy than the singlet state. Why didn't the electron get excited to the triplet state in the first place? That's against the rules. But sliding down vibrationally onto the triplet state from the singlet excited state is not, because it doesn't involve absorption of a photon.



Intersystem crossing can have important consequences in reaction chemistry because it allows access to triplet states that are not normally available in many molecules. Because triplet states feature unpaired electrons, their reactivity is often high as would be typical for radicals. That means an added suite of reactions can be accessed via this process.

Phosphorescence: A Radiationless Transition Followed by Emission

Intersystem crossing is one way a system can end up in a triplet excited state. Even though this state is lower in energy than a singlet excited state, it cannot be accessed directly via electronic excitation because that would violate the spin selection rule (\Delta S=0\). That's where the electron gets stuck, though. The quick way back down to the bottom is by emitting a photon, but because that would involve a change in spin state, it is not allowed. Realistically speaking, that means it takes a long time. By "a long time", we might mean a few seconds, several minutes, or possibly even hours. Eventually, the electron can drop back down, accompanied by the emission of a photon. This situation is called "phosphorescence".



Molecules that display phosphorescence are often incorporated into toys and shirts so that they will glow in the dark.

Contributors and Attributions

- Chris P Schaller, Ph.D., (College of Saint Benedict / Saint John's University)

This page titled 10.7.3: Fluorescence and Phosphorescence is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 14.7: Fluorescence and Phosphorescence is licensed CC BY-NC-SA 4.0.

10.8: Rotational Spectra of Polyatomic Molecules

Quick Review of Diatomic Rotation

As discussed previously, the Schrödinger equation for the angular motion of a rigid (i.e., having fixed bond length R) diatomic molecule is

$$\frac{\hbar^2}{2\mu}\left[\frac{1}{R^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{R^2\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right]|\psi\rangle = E|\psi\rangle$$

or

$$\frac{L^2}{2\mu R^2}|\psi\rangle = E|\psi\rangle$$

The Hamiltonian in this problem contains only the kinetic energy of rotation; no potential energy is present because the molecule is undergoing unhindered "free rotation". The angles θ and φ describe the orientation of the diatomic molecule's axis relative to a laboratory-fixed coordinate system, and μ is the reduced mass of the diatomic molecule

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

The eigenvalues corresponding to each eigenfunction are straightforward to find because H_{rot} is proportional to the L^2 operator whose eigenvalues have already been determined. The resultant rotational energies are given as:

$$E_J = \frac{\hbar^2 J(J+1)}{2\mu R^2} = BJ(J+1) \tag{10.8.1}$$

and are independent of M. B is the rotational constant. Thus each energy level is labeled by J and is $2J+1$-fold degenerate (because M ranges from $-J$ to J). The rotational energy in Equation 10.8.1 can be expressed in terms of the moment of inertia I

$$I = \sum_i m_i R_i^2 \tag{10.8.2}$$

where m_i is the mass of the ith atom and R is its distance from the center of mass of the molecule. This moment of inertia replaces μR^2 in the denominator of Equation 10.8.1:

$$E_J = \frac{\hbar^2 J(J+1)}{2I} = BJ(J+1) \tag{10.8.3}$$

Rotation of Polyatomic Molecules

In contrast to diatomic molecules (Equation 10.8.2), the rotational motions of polyatomic molecules in three dimensions are characterized by multiple moments of inertia. typically reflected in an 3×3 inertia tensor. It is common in rigid body mechanics to express in these moments of inertia in *lab-based Cartesian coordinates* via a notation that explicitly identifies the x, y, and z axes such as I_{xx} and I_{xy}, for the components of the inertia tensor.

$$I = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix} \tag{10.8.4}$$

The components of this tensor can be assembled into a matrix given by

$$I_{xx} = \sum_{k=1}^{N} m_k (y_k^2 + z_k^2)$$

$$I_{yy} = \sum_{k=1}^{N} m_k (x_k^2 + z_k^2)$$

$$I_{zz} = \sum_{k=1}^{N} m_k (x_k^2 + y_k^2)$$

$$I_{yx} = I_{xy} = -\sum_{k=1}^{N} m_k x_k y_k$$

$$I_{zx} = I_{xz} = -\sum_{k=1}^{N} m_k x_k z_k$$

$$I_{zy} = I_{yz} = -\sum_{k=1}^{N} m_k y_k z_k.$$

The rotational motions of polyatomic molecules are characterized by moments of inertia that are defined in a *molecule based coordinates* with axes that are labeled a, b, and c. Measured in the body frame the inertia matrix (Equation 10.8.4) is a constant real symmetric matrix, which can be decomposed into a diagonal matrix, given by

$$I = \begin{pmatrix} I_a & 0 & 0 \\ 0 & I_b & 0 \\ 0 & 0 & I_c \end{pmatrix}$$

These labels are assigned so that I_c is the **largest** principal moment of inertia with an order of the three moments set as

$$I_a < I_b < I_c$$

The rotational kinetic energy operator for a rigid non-linear polyatomic molecule is then expressed as

$$H_{rot} = \frac{J_a^2}{2I_a} + \frac{J_b^2}{2I_b} + \frac{J_c^2}{2I_c} \tag{10.8.5}$$

The components of the quantum mechanical angular momentum operators along the three principal axes are:

$$J_a = -i\hbar \cos\chi \left[\cot\theta \frac{\partial}{\partial\chi} - (\sin\theta)^{-1} \frac{\partial}{\partial\varphi} \right] - -i\hbar \sin\chi \frac{\partial}{\partial\theta} \tag{10.8.6}$$

$$J_b = i\hbar \sin\chi \left[\cot\theta \frac{\partial}{\partial\chi} - (\sin\theta)^{-1} \frac{\partial}{\partial\varphi} \right] - -i\hbar \cos\chi \frac{\partial}{\partial\theta} \tag{10.8.7}$$

$$J_c = -\frac{ih\partial}{\partial\chi} \tag{10.8.8}$$

The angles θ, φ, and χ are the **Euler angles** needed to specify the orientation of the rigid molecule relative to a laboratory-fixed coordinate system. The corresponding square of the total angular momentum operator J^2 can be obtained as

$$J^2 = J_a^2 + J_b^2 + J_c^2 \tag{10.8.9}$$

$$= -\frac{\partial^2}{\partial\theta^2} - \cot\theta \frac{\partial}{\partial\theta} - \left(\frac{1}{\sin\theta} \right) \left(\frac{\partial^2}{\partial\varphi^2} + \frac{\partial^2}{\partial\chi^2} - 2\cos\theta \frac{\partial^2}{\partial\varphi\partial\chi} \right) \tag{10.8.10}$$

and the component along the lab-fixed Z axis is

$$J_Z = -ih \frac{\partial}{\partial\varphi}.$$

Spherical Tops

When the three principal moment of inertia values are identical, the molecule is termed a spherical top. In this case, the total rotational energy Equation 10.8.5 can be expressed in terms of the total angular momentum operator J^2

$$H_{rot} = \frac{J^2}{2I}$$

As a result, the eigenfunctions of H_{rot} are those of J^2 (and J_a as well as J_Z both of which commute with J_2 and with one another; J_Z is the component of J along the lab-fixed Z-axis and commutes with J_a because

$$J_Z = -ih\frac{\partial}{\partial\varphi}$$

and

$$J_a = -ih\frac{\partial}{\partial\chi}$$

act on different angles. The energies associated with such eigenfunctions are

$$E(J,K,M) = \frac{\hbar^2 J(J+1)}{2I^2}$$

for all K (i.e., J a quantum numbers) ranging from -J to J in unit steps and for all M (i.e., J Z quantum numbers) ranging from -J to J. Each energy level is therefore $(2J+1)^2$ degenarate because there are $2J+1$ possible K values and $2J+1$ possible M values for each J. The eigenfunctions of J^2, J_Z and J_a, $|J,M,K>$ are given in terms of the set of rotation matrices $D_{J,M,K}$:

$$|J,M,K\rangle = \sqrt{\frac{2J+1}{8\pi^2}} D^*_{J,M,K}(\theta,\varphi,\chi)$$

which obey

$$J^2|J,M,K\rangle = \hbar^2 J(J+1)|J,M,K\rangle$$

$$J_a|J,M,K\rangle = \hbar K|J,M,K\rangle$$

$$J_Z|J,M,K\rangle = \hbar M|J,M,K\rangle$$

Symmetric Tops

Symmetrical tops are molecules with two rotational axes that have the same inertia and one unique rotational axis with a different inertia. Symmetrical tops can be divided into two categories based on the relationship between the inertia of the unique axis and the inertia of the two axes with equivalent inertia. If the unique rotational axis has a greater inertia than the degenerate axes the molecule is called an **oblate symmetrical top** (Figure 13.8.1). If the unique rotational axis has a lower inertia than the degenerate axes the molecule is called a **prolate symmetrical top**. For simplification think of these two categories as either frisbees for oblate tops or footballs for prolate tops.

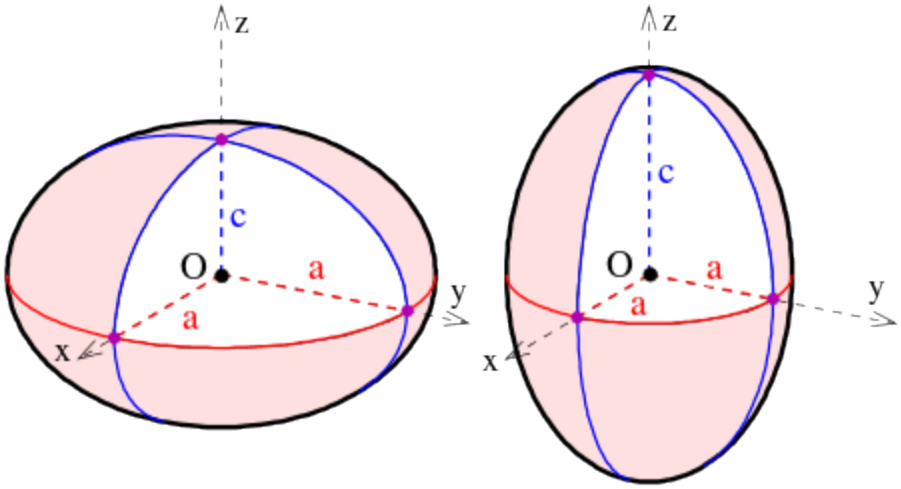


Figure 13.8.1 : The assignment of semi-axes on a spheroid. It is oblate if $c < a$ (left) and prolate if $c > a$ (right). (CC BY-SA 4.0; Ag2gaeh)

Again, the rotational kinetic energy, which is the full rotational Hamiltonian, can be written in terms of the total rotational angular momentum operator J^2 and the component of angular momentum along the axis with the unique principal moment of inertia.

For prolate tops, Equation 10.8.5 becomes

$$H_{rot} = \frac{J^2}{2I} + J_a^2 \left(\frac{1}{2I_a} - \frac{1}{2I} \right)$$

For oblate tops, Equation 10.8.5 becomes

$$H_{rot} = \frac{J^2}{2I} + J_c^2 \left(\frac{1}{2I_c} - \frac{1}{2I} \right)$$

As a result, the eigenfunctions of H_{rot} are those of J^2 and J_a or J_c (and of J_Z), and the corresponding energy levels.

The energies for prolate tops are

$$E(J,K,M) = \frac{h^2 J(J+1)}{2I^2} + h^2 K^2 \left(\frac{1}{2I_a} - \frac{1}{2I} \right)$$

and the energies for oblate tops are

$$E(J,K,M) = \frac{h^2 J(J+1)}{2I2} + h^2 K^2 \left(\frac{1}{2I_c} - \frac{1}{2I} \right)$$

again for K and M (i.e., J_a or J_c and J_Z quantum numbers, respectively) ranging from $-J$ to J in unit steps. Since the energy now depends on K, these levels are only $2J+1$ degenerate due to the $2J+1$ different M values that arise for each J value. The eigenfunctions $|J,M,K>$ are the same rotation matrix functions as arise for the spherical-top case.

Asymmetric Tops

The rotational eigenfunctions and energy levels of a molecule for which all three principal moments of inertia are distinct (a **asymmetric top**) can not easily be expressed in terms of the angular momentum eigenstates and the J, M, and K quantum numbers. However, given the three principal moments of inertia I_a, I_b, and I_c, a matrix representation of each of the three contributions to the general rotational Hamiltonian in Equation 10.8.5 can be formed within a basis set of the $\{|J,M,K\rangle\}$ rotation matrix functions. This matrix will **not** be diagonal because the $|J,M,K\rangle$ functions are not eigenfunctions of the asymmetric top H_{rot}. However, the matrix can be formed in this basis and subsequently brought to diagonal form by finding its eigenvectors {C n, J,M,K } and its eigenvalues $\{E_n\}$. The vector coefficients express the asymmetric top eigenstates as

$$\psi_n(\theta,\varphi,\chi) = \sum_{J,M,K} C_{n,J,M,K} |J,M,K\rangle$$

Because the total angular momentum J^2 still commutes with H_{rot}, each such eigenstate will contain only one J-value, and hence Ψ_n can also be labeled by a J quantum number:

$$\psi_{n,J}(\theta,\varphi,\chi) = \sum_{M,K} C_{n,J,M,K}|J,M,K\rangle$$

To form the only non-zero matrix elements of H_{rot} within the $|J,M,K\rangle$ basis, one can use the following properties of the rotation-matrix functions:

$$\langle j,\rangle = \langle j,\rangle = 1/2 < j,\rangle = h2[J(J+1) - K2],$$

$$\langle j,\rangle = h^2K^2$$

$$\langle j\rangle = -\langle j\rangle = h^2[J(J+1) - K(K\pm1)]1/2[J(J+1) - (K\pm1)(K\pm2)]1/2\langle j\rangle = 0$$

Each of the elements of J_c^2, J_a^2, and J_b^2 must, of course, be multiplied, respectively, by $1/2I_c$, $1/2I_a$, and $1/2I_b$ and summed together to form the matrix representation of H_{rot}. The diagonalization of this matrix then provides the asymmetric top energies and wavefunctions.

This page titled 10.8: Rotational Spectra of Polyatomic Molecules is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Jack Simons.

- 13.8: Rotational Spectra of Polyatomic Molecules by Jack Simons is licensed CC BY-NC-SA 4.0.

10.9: Normal Modes in Polyatomic Molecules

Normal modes are used to describe the different vibrational motions in molecules. Each mode can be characterized by a different type of motion and each mode has a certain symmetry associated with it. Group theory is a useful tool in order to determine what symmetries the normal modes contain and predict if these modes are IR and/or Raman active. Consequently, IR and Raman spectroscopy is often used for vibrational spectra.

Degrees of Freedom

In general, a normal mode is an independent motion of atoms in a molecule that occurs without causing movement to any of the other modes. Normal modes, as implied by their name, are orthogonal to each other. In order to discuss the quantum-mechanical equations that govern molecular vibrations it is convenient to convert Cartesian coordinates into so called normal coordinates. Vibrations in polyatomic molecules are represented by these normal coordinates.

A molecule can have three types of degrees of freedom and a total of 3N degrees of freedom, where N equals the number of atoms in the molecule. These degrees of freedom can be broken down into three categories.

- **Translational**: These are the simplest of the degrees of freedom. These entail the movement of the entire molecule's center of mass. This movement can be completely described by three orthogonal vectors and thus contains 3 degrees of freedom.
- **Rotational**: These are rotations around the center of mass of the molecule and like the translational movement they can be completely described by three orthogonal vectors. This again means that this category contains only 3 degrees of freedom. However, in the case of a linear molecule only two degrees of freedom are present due to the rotation along the bonds in the molecule having a negligible inertia.
- **Vibrational**: These are any other types of movement not assigned to rotational or translational movement and thus there are 3N – 6 degrees of vibrational freedom for a nonlinear molecule and 3N – 5 for a linear molecule. These vibrations include bending, stretching, wagging and many other aptly named internal movements of a molecule. These various vibrations arise due to the numerous combinations of different stretches, contractions, and bends that can occur between the bonds of atoms in the molecule.

Table 13.9.1 : Overview of degrees of freedom

	Total Degrees of Freedom	Translational degrees of freedom	Rotational degrees of freedom	Vibrational degrees of freedom
Nonlinear Molecules	3N	3	3	3N -**6**
Linear Molecules	3N	3	2	3N - 5

Each of these degrees of freedom is able to store energy. However, In the case of rotational and vibrational degrees of freedom, energy can only be stored in discrete amounts. This is due to the quantized break down of energy levels in a molecule described by quantum mechanics. In the case of rotations the energy stored is dependent on the rotational inertia of the gas along with the corresponding quantum number describing the energy level.

Example 13.9.1 : Ethane vs. Carbon Dioxide

Ethane, C_2H_6 has eight atoms ($N = 8$) and is a nonlinear molecule so of the $3N = 24$ degrees of freedom, three are translational and three are rotational. The remaining 18 degrees of freedom are internal (vibrational). This is consistent with:

$$3N - 6 = 3(8) - 6 = 18$$

Carbon Dioxide, CO_2 has three atoms ($N = 3$ and is a linear molecule so of the $3N = 9$ degrees of freedom, three are translational and two are rotational. The remaining 4 degrees of freedom are vibrational. This is consistent with:

$$3N - 5 = 3(3) - 5 = 4$$

The normal modes of vibration are: asymmetric, symmetric, wagging, twisting, scissoring, and rocking for polyatomic molecules.

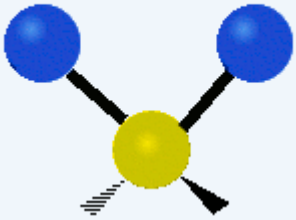
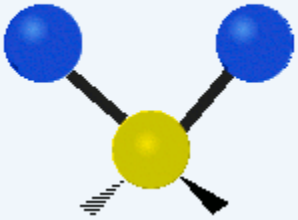
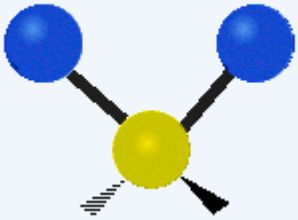
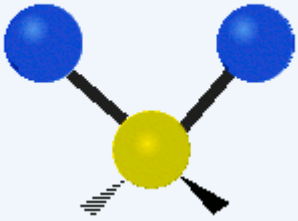
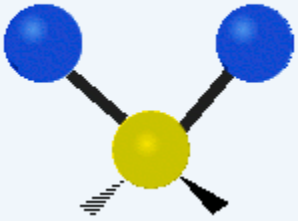
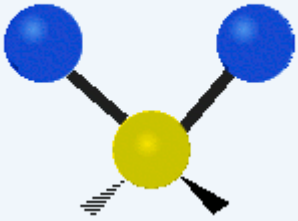
Symmetricical Stretching	Asymmetrical Stretching	Wagging
		
Twisting	**Scissoring**	**Rocking**
		

Figure 13.9.1 : *Six types of Vibrational Modes. Taken from publisher en.Wikipedia.org/wiki/Infrared_spectroscopy with permission from copyright holder.*

Normal Modes

If there is no external field present, the energy of a molecule does not depend on its orientation in space (its translational degrees of freedom) nor its center of mass (its rotational degrees of freedom). The potential energy of the molecule is therefore made up of its vibrational degrees of freedom only of $3N - 6$ (or $3N - 5$ for linear molecules). The difference in potential energy is given by:

$$\Delta V = V(q_1, q_2, q_3, ..., q_n) - V(0, 0, 0, ..., 0) \tag{10.9.1}$$

$$= \frac{1}{2} \sum_{i=1}^{N_{vib}} \sum_{j=1}^{N_{vib}} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right) q_i q_j \tag{10.9.2}$$

$$= \frac{1}{2} \sum_{i=1}^{N_{vib}} \sum_{j=1}^{N_{vib}} f_{ij} q_i q_j \tag{10.9.3}$$

where q represents the equilibrium displacement and N_{vib} the number of vibrational degrees of freedom.

For simplicity, the *anharmonic* terms are neglected in this equation (i.e., higher order terms are ignore). A theorem of classical mechanics states that the cross terms can be eliminated from the above equation (the details of the theorem are very complex and will not be discussed in detail). By using matrix algebra a new set of coordinates {Qj} can be found such that

$$\Delta V = \frac{1}{2} \sum_{j=1}^{N_{vib}} F_j Q_j^2 \tag{10.9.4}$$

Note that there are no cross terms in this new expression. These new coordinates are called **normal coordinates** or **normal**

modes. With these new normal coordinates in hand, the Hamiltonian operator for vibrations can be written as follows:

$$\hat{H}_{vib} = -\sum_{j=1}^{N_{vib}} \frac{\hbar^2}{2\mu_i} \frac{d^2}{dQ_j^2} + \frac{1}{2} \sum_{j=1}^{N_{vib}} F_j Q_j^2 \tag{10.9.5}$$

The total wavefunction is a product of the individual wavefunctions and the energy is the sum of independent energies. This leads to:

$$\hat{H}_{vib} = \sum_{j=1}^{N_{vib}} \hat{H}_{vib,j} = \sum_{j=1}^{N_{vib}} \left(\frac{-\hbar^2}{2\mu_j} \frac{d^2}{dQ_i^2} + \frac{1}{2} \sum_{j=1}^{N_{vib}} F_j Q_j^2 \right) \tag{10.9.6}$$

and the wavefunction is then

$$\begin{aligned} \psi_{vib} &= Q_1, Q_2, Q_3 ..., Q_{vib} \\ &= \psi_{vib,1}(Q_1)\psi_{vib,2}(Q_2)\psi_{vib,3}(Q_3), ..., \psi_{vib,N_{vib}}(Q_{N_{vib}}) \end{aligned}$$

and the total vibrational energy of the molecule is

$$E_{vib} = \sum_{j=1}^{N_{vin}} h\nu_j \left(v_j + \frac{1}{2} \right) \tag{10.9.7}$$

where $v_j = 0, 1, 2, 3....$

The consequence of the result stated in the above equations is that each vibrational mode can be treated as a harmonic oscillator approximation. There are N_{vib} harmonic oscillators corresponding to the total number of vibrational modes present in the molecule.

Pictorial description of normal coordinates using CO

The normal coordinate q is used to follow the path of a normal mode of vibration. As shown in Figure 13.9.2 the displacement of the C atom, denoted by $\Delta r_o(\mathrm{C})$, and the displacement of the O atom, denoted by $\Delta r_o(\mathrm{O})$, occur at the same frequency. The displacement of atoms is measured from the equilibrium distance in ground vibrational state, r_o.

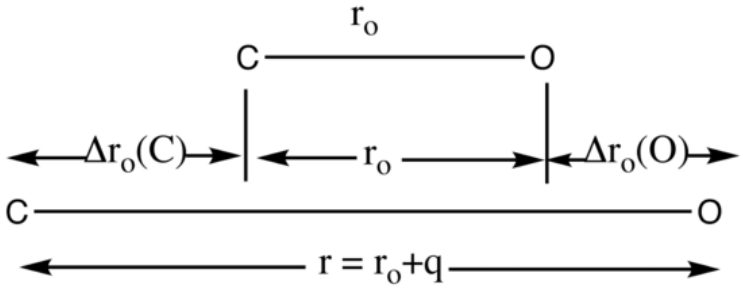


Figure 13.9.2 : The Normal coordinate for CO is equation to $\Delta r(C) + \Delta r(O)$

This page titled 10.9: Normal Modes in Polyatomic Molecules is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kristin Kowolik.

- 13.9: Normal Modes in Polyatomic Molecules by Kristin Kowolik is licensed CC BY-NC-SA 4.0.

10.10: Irreducible Representation of Point Groups

There exists an important fact about normal coordinates. Each of these coordinates belongs to an irreducible representation of the point the molecule under investigation. Vibrational wavefunctions associated with vibrational energy levels share this property as well. The normal coordinates and the vibration wavefunction can be categorized further according to the point group they belong to. From the character table predictions can be made for which symmetries can exist. The irreducible representation offers insight into the IR and/or Raman activity of the molecule in question.

Symmetry of normal modes

It is important to realize that every normal mode has a certain type of symmetry associated with it. Identifying the point group of the molecule is therefore an important step. With this in mind it is not surprising that every normal mode forms a basis set for an irreducible representation of the point group the molecule belongs to. For a molecule such as water, having a structure of XY_2, three normal coordinates can be determined. The two stretching modes are equivalent in symmetry and energy. The figure below shows the three normal modes for the water molecule:

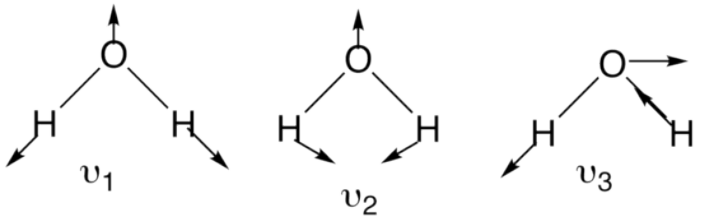


Figure 13.10.1 : Three normal modes of water

By convention, with nonlinear molecules, the symmetric stretch is denoted v_1 whereas the asymmetric stretch is denoted v_2. Bending motions are v_3. With linear molecules, the bending motion is ν_2 whereas asymmetric stretch is v_3.

The water molecule has C_{2v} symmetry and its symmetry elements are E, C_2, σ(xz) and σ(yz). To determine the symmetries of the three vibrations and how they each transform, symmetry operations will be performed. As an example, performing C_2 operations using the two normal mode v_2 and v_3 gives the following transformation:

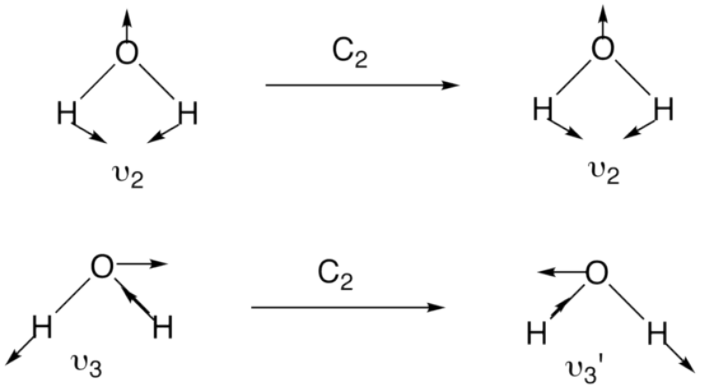


Figure 13.10.2 : Symmetry operations on the symmetric and asymmetric stretches of water

Once all the symmetry operations have been performed in a systematic manner for each modes the symmetry can be assigned to the normal mode using the character table for C_{2v}:

Table 13.10.2 : Character table for the C_{2v} point group

C_{2v}	E	C2	σ (xz)	σ (yz)	
v_1	1	1	1	1	**= a_1**

C_{2v}	E	C2	σ (xz)	σ (yz)	
ν_2	1	1	1	1	**= a_1**
ν_3	1	-1	-1	1	**= b_2**

Water has three normal modes that can be grouped together as the reducible representation

$$\Gamma_{vib} = 2a_1 + b_2.$$

Determination of normal modes becomes quite complex as the number of atoms in the molecule increases. Nowadays, computer programs that simulate molecular vibrations can be used to perform these calculations. The example of $[PtCl_4]^{2-}$ shows the increasing complexity (Figure 13.10.2). The molecule has five atoms and therefore 15 degrees of freedom, 9 of these are vibrational degrees of freedom. The nine normal modes are exemplified below along with the irreducible representation the normal mode belongs to (D_{4h} point group).

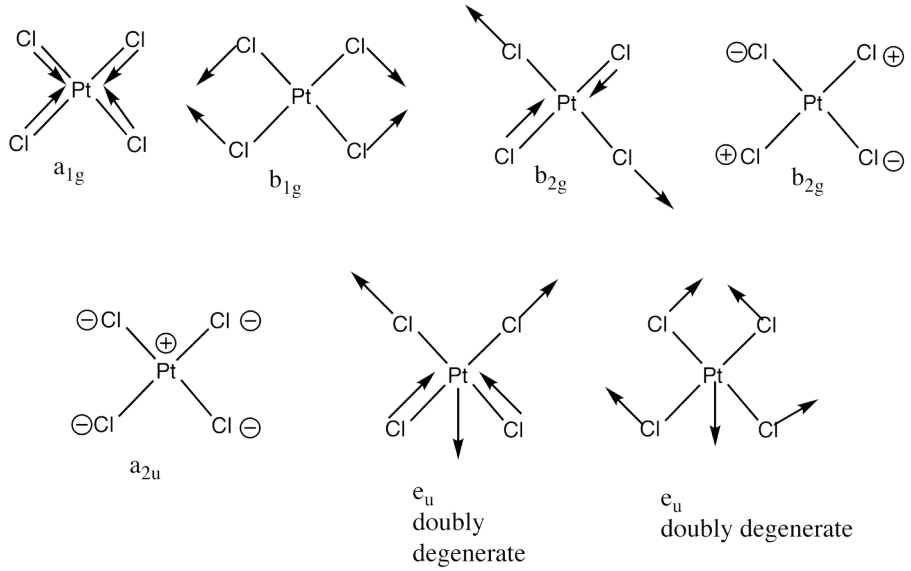


Figure 13.10.3 : A_{1g}, b_{1g} and e_u are stretching vibrations whereas b_{2g}, a_{2u}, b_{2u} and e_u are bending vibrations.

This page titled 10.10: Irreducible Representation of Point Groups is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kristin Kowolik.

- 13.10: Irreducible Representation of Point Groups by Kristin Kowolik is licensed CC BY-NC-SA 4.0.

10.11: The Selection Rule for the Rigid Rotor

A **selection rule** describes how the probability of transitioning from one level to another cannot be **zero**. It has two sub-pieces: a **gross selection rule** and a **specific selection rule**. A gross selection rule illustrates characteristic requirements for atoms or molecules to display a spectrum of a given kind, such as an IR spectroscopy or a microwave spectroscopy. Once the atom or molecules follow the gross selection rule, the specific selection rule must be applied to the atom or molecules to determine whether a certain transition in quantum number may happen or not.

Selection rules specify the possible transitions among quantum levels due to absorption or emission of electromagnetic radiation. Incident electromagnetic radiation presents an oscillating electric field $E_0 \cos(\omega t)$ that interacts with a transition dipole. The dipole operator is $\mu = e \cdot r$ where r is a vector pointing in a direction of space.

A dipole moment of a given state is

$$\mu_z = \int \Psi_1{}^* \mu_z \Psi_1 \, d\tau$$

A transition dipole moment is a transient dipolar polarization created by an interaction of electromagnetic radiation with a molecule

$$(\mu_z)_{12} = \int \Psi_1{}^* \mu_z \Psi_2 \, d\tau$$

In an experiment we present an electric field along the z axis (in the laboratory frame) and we may consider specifically the interaction between the transition dipole along the x, y, or z axis of the molecule with this radiation. If μ_z is zero then a transition is forbidden. The selection rule is a statement of when μ_z is non-zero.

> The selection rule is a statement of when μ_z is non-zero.

Rotational transitions

We can use the definition of the transition moment and the spherical harmonics to derive selection rules for a rigid rotator. Once again we assume that radiation is along the z axis.

$$(\mu_z)_{J,M,J',M'} = \int_0^{2\pi} \int_0^{\pi} Y_{J'}^{M'}(\theta,\phi) \mu_z Y_J^M(\theta,\phi) \sin\theta \, d\phi, d\theta$$

Notice that m must be non-zero in order for the transition moment to be non-zero. This proves that a molecule must have a permanent dipole moment in order to have a rotational spectrum. The spherical harmonics can be written as

$$Y_J^M(\theta,\phi) = N_{JM} P_J^{|M|}(\cos\theta) e^{iM\phi}$$

where N_{JM} is a normalization constant. Using the standard substitution of $x = \cos q$ we can express the rotational transition moment as

$$(\mu_z)_{J,M,J',M'} = \mu N_{JM} N_{J'M'} \int_0^{2\pi} e^{I(M-M')\phi} \, d\phi \int_{-1}^{1} P_{J'}^{|M'|}(x) P_J^{|M|}(x) dx$$

The integral over f is zero unless M = M' so $\Delta M = 0$ is part of the rigid rotator selection rule. Integration over ϕ for $M = M'$ gives 2π so we have

$$(\mu_z)_{J,M,J',M'} = 2\pi\mu N_{JM} N_{J'M'} \int_{-1}^{1} P_{J'}^{|M'|}(x) P_J^{|M|}(x) dx$$

We can evaluate this integral using the identity

$$(2J+1)x \, P_J^{|M|}(x) = (J - |M| + 1) P_{J+1}^{|M|}(x) + (J - |M|) P_{J-1}^{|M|}(x)$$

Substituting into the integral one obtains an integral which will vanish unless $J' = J + 1$ or $J' = J - 1$.

$$\int_{-1}^{1} P_{J'}^{|M'|}(x)\left(\frac{(J-|M|+1)}{(2J+1)}P_{J+1}^{|M|}(x)+\frac{(J-|M|)}{(2J+1)}P_{J-1}^{|M|}(x)\right)dx$$

This leads to the selection rule $\Delta J = \pm 1$ for absorptive rotational transitions. Keep in mind the physical interpretation of the quantum numbers J and M as the total angular momentum and z-component of angular momentum, respectively. As stated above in the section on electronic transitions, these selection rules also apply to the orbital angular momentum ($\Delta l = \pm 1$, $\Delta m = 0$).

This page titled 10.11: The Selection Rule for the Rigid Rotor is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- 13.12: The Selection Rule for the Rigid Rotor is licensed CC BY-SA 4.0.

10.12: The Harmonic Oscillator Selection Rule

Selection rules are a very important concept in spectroscopy and physical chemistry, as they directly complement the concept that systems on the atomic level are quantized. Quantization tells us how electrons, neutrons, and protons are set up, and selection rules tell us how, energetically speaking, they may move about. For example, the principal quantum number, n, determines the energy state of particles in the Particle in a Box model. With respect to real systems, the quantum number n describes electronic transitions, such as the excitation of an electron. There are other transitions, such as the rotational transition, which applies to the rigid rotator model and occurs where there are spherical harmonics. Vibrational transitions occur where there is a system that can be visualized as two masses connected by a spring, such as a diatomic molecule.

Selection rules can be very useful in spectroscopy for obtaining information about an unknown substance; any given substance has properties and behaviors that operate within the selection rules and determine the wavelengths of electromagnetic radiation (light) that cause transitions. By tracking these properties, chemists can use techniques like IR spectroscopy to determine the chemical structure of a molecule. For example, a selection rule might govern how a quantum number changes during a transition. That quantum number appears in equations that determine the energy transition based on the mass and other structural properties of a molecule. Thus, if one knows the selection rules and the energy transition, he or she can determine other properties about the molecule, such as its mass or its bond strength.

Example: Evaluating vibrational transitions using a transition moment integral.

Vibrational systems are described by the harmonic oscillator wavefunctions, which look like

$$\Psi_v(x) = N_v H_v(\alpha^{1/2}x)e^{-\alpha x^2/2}$$

A transition dipole moment can be written to describe a transition along the z-axis.

$$\mu_z = \int_{-\infty}^{\infty} N_v N_{v+1} H_{v+1}(\alpha^{1/2}x)e^{-\alpha x^2/2} H_{v+1}\mu_z(\alpha^{1/2}x)e^{-\alpha x^2/2}dx$$

μ is the dipole moment, so we pull it out of the integral.

$$\mu_z = \mu_z \int_{-\infty}^{\infty} N_v N_{v+1} H_{v+1}(\alpha^{1/2}x)e^{-\alpha x^2/2} H_{v+1}(\alpha^{1/2}x)e^{-\alpha x^2/2}dx$$

Using the idea that the transition moment μ_z changes on x, we express it as a derivative.

$$\mu_z = \frac{d\mu}{dx} \int_{-\infty}^{\infty} N_v N_{v+1} H_{v+1}(\alpha^{1/2}x)e^{-\alpha x^2/2} H_{v+1}(\alpha^{1/2}x)e^{-\alpha x^2/2}dx$$

To evaluate the integral, we take advantage of a special Hermite relationship: $\sqrt{\alpha}xH_v(\sqrt{\alpha}x) = vH_{v-1}(\sqrt{\alpha}x) + \frac{1}{2}H_{v+1}(\sqrt{\alpha}x)$

$$\mu_z = \frac{N_v N_{v+1}}{\sqrt{\alpha}}\left(\frac{d\mu}{dx}\right) \int_{-\infty}^{\infty} H_{v+1}(\sqrt{\alpha}x)e^{(}-\alpha x^2/2)[vH_{v-1}(\sqrt{\alpha}x) + \frac{1}{2}H_{v+1}(\sqrt{\alpha}x)]dx$$

By inspection, it is clear that the integral only allows transitions of $\Delta v = \pm 1$, because for any other value, the integral will be zero.

This page titled 10.12: The Harmonic Oscillator Selection Rule is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Alex Krueger.

- 13.13: The Harmonic Oscillator Selection Rule by Alex Krueger is licensed CC BY-NC-SA 4.0.

10.13: Group Theory Determines Infrared Activity

Determining if a Normal Modes is IR or Raman Active

A transition from $v \rightarrow v'$ is IR active if the **transition moment integral** contains the totally symmetric irreducible representation of the point group the molecule belongs to. The transition moment integral is derived from the one-dimensional harmonic oscillator. Using the definition of electric dipole moment μ, the integral is:

$$M(v \rightarrow v') = \langle \text{final wavefunction} | \vec{\mu} | \text{initial wavefunction} \rangle$$

or in terms of vibrational wavefunctions for a specific normal mode $|\phi(v)\rangle$

$$M(v \rightarrow v') = \langle \phi(v' \neq 0) | \vec{\mu} | \phi(v = 0) \rangle \tag{10.13.1}$$

assuming the transition from the $v = 0$ wavefunction to the $v' \neq 0$ wavefunction.

Now, consider the case that $\vec{\mu}$, is a constant and therefore independent of the vibration (i.e., the electric dipole moment does not change during the vibration). This it could be taken outside the integral in Equation 10.13.1 becomes

$$M(v \rightarrow v') = \vec{\mu} \langle \phi(v' \neq 0) | \phi(v = 0) \rangle \tag{10.13.2}$$

Since $|\phi(v = 0)\rangle$ and $|\phi(v =\neq 0)\rangle$ are mutually orthogonal to each other, the integral in Equation 10.13.1 will equal zero and the transition will **not** be allowed (i.e., it is forbidden). For the M to be nonzero, $\vec{\mu}$ must change during a vibration. This selection rule explains why homonuclear diatomic molecules do not produce an IR spectrum. There is no change in dipole moment resulting in a transition moment integral of zero and a transition that is forbidden.

For a transition to be Raman active, the same rules apply. The transition moment integral must contain the totally symmetric irreducible representation of the point group. The integral contains the polarizability tensor α (usually represented by a square matrix):

$$M(v \rightarrow v') = \langle \phi(v' \neq 0) | \alpha | \phi(v = 0) \rangle \tag{10.13.3}$$

Following a similar argument as above, α must be nonzero for the transition to be allowed and exhibits Raman scattering.

Character Table

For a molecule to be IR active the dipole moment has to change during the vibration. For a molecule to be Raman active the polarizability of the molecule has to change during the vibration. The reducible representation $\mathbf{\Gamma_{vib}}$ can also be found by determining the reducible representation of the 3N degrees of freedom of H_2O, Γ_{tot}. By applying Group Theory it is straightforward to find $\mathbf{\Gamma_{x,y,z}}$ as well as UMA (number of unmoved atoms). Again, using water as an example with C_{2v} symmetry where 3N = 9, $\mathbf{\Gamma_{tot}}$ can be determined:

C_{2v}	E	C_2	σ (xz)	σ (yz)	
$\mathbf{T_{x,y,z}}$	3	-1	1	1	
UMA	3	1	1	3	
$\mathbf{\Gamma_{tot}}$	**9**	**-1**	**1**	**3**	$\mathbf{=3a_1 + a_2 + 2b_1 + 3b_2}$
Note that Γ_{tot} contains nine degrees of freedom consistent with 3N = 9.					

Γ_{tot} contains $\Gamma_{translational}$, $\Gamma_{rotational}$ as well as $\Gamma_{vibrational}$. Γ_{trans} can be obtained by finding the irreducible representations corresponding to x,y and z in the right side of the character table, Γ_{rot} by finding the ones corresponding to R_x, R_y and R_z. Γ_{vib} can be obtained by Γ_{tot} - Γ_{trans} - Γ_{rot}.

$$\Gamma_{vib}(H_2O) = (3a_1 + a_2 + 2b_1 + 3b_2) - (a_1 + b_1 + b_2) - (a_2 + b_1 + b_2) = 2a_1 + b_2$$

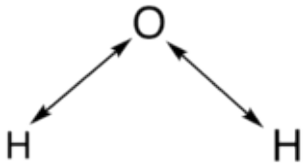
In order to determine which modes are IR active, a simple check of the irreducible representation that corresponds to x,y and z and a cross check with the reducible representation Γ_{vib} is necessary. If they contain the same irreducible representation, the mode is IR active.

For H_2O, z transforms as a_1, x as b_1 and y as b_2. The modes a1 and b2 are IR active since Γvib contains $2a_1 + b_2$.

In order to determine which modes are Raman active, the irreducible representation that corresponds to z^2, x^2-y^2, xy, xz and yz is used and again cross checked with Γvib. For H_2O, z^2 and x^2-y^2 transform as a_1, xy as a_2, xz as b_1 and yz as b_2.The modes a_1 and b_2 are also Raman active since Γ_{vib} contains both these modes.

The IR spectrum of H2O does indeed have three bands as predicted by Group Theory. The two symmetric stretches v1 and v2 occur at 3756 and 3657 cm-1 whereas the bending v3 motion occurs at 1595 cm-1.

In order to determine which normal modes are stretching vibrations and which one are bending vibrations, a stretching analysis can be performed. Then the stretching vibrations can be deducted from the total vibrations in order to obtain the bending vibrations. A double-headed arrow is drawn between the atom as depicted below:



Then a determination of how the arrows transform under each symmetry operation in C2v symmetry will yield the following results:

C_{2v}	E	C_2	σ (xz)	σ (yz)	
$\Gamma_{stretch}$	2	0	0	2	**= a_1 + b_2**

$$\Gamma_{bend} = \Gamma_{vib} - \Gamma_{stretch} = 2a_1 + b_2 - a_1 - b_2 = a_1$$

H_2O has two stretching vibrations as well as one bending vibration. This concept can be expanded to complex molecules such as $PtCl4^-$. Four double headed arrows can be drawn between the atoms of the molecule and determine how these transform in D_{4h} symmetry. Once the irreducible representation for $\Gamma_{stretch}$ has been worked out, Γ_{bend} can be determined by $\Gamma_{bend} = \Gamma_{vib} - \Gamma_{stretch}$.

Most molecules are in their zero point energy at room temperature. Therefore, most transitions do originate from the v=0 state. Some molecules do have a significant population of the v=1 state at room temperature and transitions from this thermally excited state are called **hot** bands.

References

1. Merlin, J.C., Cornard, J.P., *J. Chem. Educ.*, **2006**, 83 (9), p 1383. **DOI:** 10.1021/ed083p1393
2. McGuinn, C.J., *J. Chem. Educ.*, **1982**, 59 (10), p 813. **DOI:** 10.1021/ed059p813
3. Harris, D.C., Bertolucci, M.D., *Symmetry and Spectroscopy: An introduction to Vibrational and Electronic Spectroscopy*. Dover Publocations, Inc., New York, **1989**.
4. McQuarrie, D. A., Simon, J.D., *Physical Chemistry: A Molecular Approach*, University Science Books, Sausalito, California, **1997**; 518-521.
5. Housecroft, C.E., Sharpe, A.G., **Inorganic Chemistry.** Pearson Education Limited, England, **2008**, 107.
6. Atkins, P., dePaula, J., *Physical Chemistry*, W.H. Freeman and Company, New York, **2002**, 520-523.
7. Bishop, D.M., *Group Theory and Chemistry*, Dover Publications, Inc, New York, **1973**, 166.

This page titled 10.13: Group Theory Determines Infrared Activity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Kristin Kowolik.

- 13.14: Group Theory Determines Infrared Activity by Kristin Kowolik is licensed CC BY-NC-SA 4.0.

10.E: Molecular Spectroscopy (Exercises)

These are homework exercises to accompany Chapter 13 of McQuarrie and Simon's "Physical Chemistry: A Molecular Approach" Textmap.

Q13.1

1. Calculate the energy difference for radiation wavenumber of, $\nu = 1.00 cm^{-1}$ including the type of molecular process that absorbs of this radiation corresponding to.
2. Given the wavelength of 3.4 x 10^{-4}, which molecular process does it fall into?
3. What are other advantages of knowing the spectroscopy of molecules?
4. Do you think the spectroscopy and molecular processes of the molecules have relationship to temperature? Explain.
5. What is other form that you can express in above equation?

Q13.20

The vibrational term of a diatomic molecule is

$$G(v) = (v + \frac{1}{2})\nu_e - (v + \frac{1}{2})^2 \chi_e \nu_e \tag{10.E.1}$$

Show that the spacing between adjacent levels is given by

$$\Delta G = G(v+1) - G(v) = \nu_e - 2\chi_e \nu_e (v+1)$$

Show that the maximum vibrational quantum number, v_{max} is given by

$$v_{max} = \frac{1}{2\chi_e} - 1 \tag{10.E.2}$$

Use this result to show that the dissociation energy of the diatomic molecule can be written as

$$D_e = \frac{\nu_e - \nu_e \chi_e^2}{4\chi_e} \approx \frac{\nu_e}{4\chi_e} \tag{10.E.3}$$

Explain how the constants ν_e and χ_e can be evaluated from a plot of ΔG versus $v + 1$, called a Birge-Sponer plot. After finding these constants, determine the dissociation energy of the molecule. Use the following experimental data of H_2 to determine the dissociation energy of the molecule.

v	G(v) / cm^{-1}
0	4161.12
1	8087.11
2	11782.35
3	15250.36
4	18497.92
5	21505.65
6	24287.83
7	26830.97
8	29123.93
9	31150.19

v	G(v) / cm^{-1}
10	32886.65
11	34301.83
12	35351.01
13	35972.97

Explain why your plot is not linear for high values of v. How does this dissocation energy differ with the experimental value of $38269.48cm^{-1}$.

S13.20

$$\Delta G = G(v+1) - G(v) = \left(v+\frac{3}{2}\right)\nu_e - \left(v+\frac{3}{2}\right)^2 \chi_e\nu_e - \left(v+\frac{1}{2}\right)\nu_e + \left(v+\frac{1}{2}\right)^2 \chi_e\nu_e = \nu_e - \chi_e\nu_e(2v+2) = \nu_e - 2\chi_e\nu_e(v+1)$$

In the limit that $\Delta G \to 0, v \to v_{max}$, solving for v_{max} gives

$$0 = \nu_e - 2\chi_e\nu_e(v_{max}+1) \tag{10.E.4}$$

$$2\chi_e(v_{max}+1) = 1 \tag{10.E.5}$$

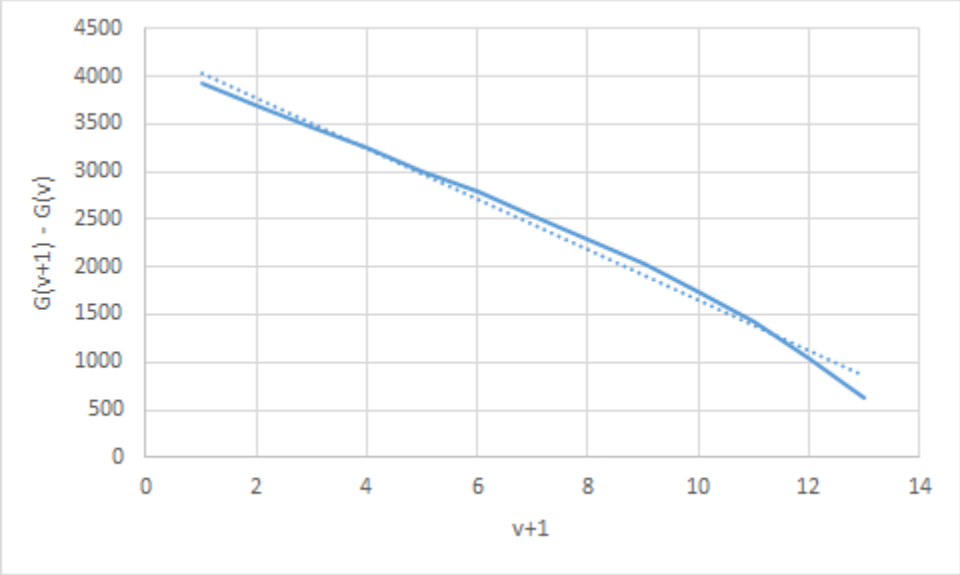
$$v_{max} = \frac{1}{2\chi_e} - 1 \tag{10.E.6}$$

The molecule dissociates in the limit $\Delta G \to 0$, so the dissociation energy is given by

$$D_e = G(v_{max}) = (\frac{1}{2\chi_e} - \frac{1}{2})\nu_e - (\frac{1}{2\chi_e} - \frac{1}{2})^2\chi_e\nu_e = \frac{\nu_e}{2\chi_e}(1-\chi_e) - \frac{\nu_e}{4\chi_e}(1-\chi_e)^2 = \frac{\nu_e}{4\chi_e}(2-2\chi_e-1+2\chi_e-\chi_e^2) = \frac{\nu_e}{4\chi_e}(1$$

The final step can be made if we assume χ_e to be very small.

Looking at the equation for ΔG, we conclude that a plot of ΔG versus $v+1$ will have an intercept of ν_e and a slope of $-2\chi_e\nu_e$. The experimental data points for H_2 can be plotted as such:



If we use a best linear fit, we determine that the intercept is $\nu_e = 4164.4cm^{-1}$ and the slope is $-2\chi_e\nu_e = -232.01cm^{-1}$. Thus, $\chi_e = 0.0279$. Then, the dissociation energy is

$$D_e = \frac{\nu_e}{4\chi_e} = 37400 cm^{-1} \tag{10.E.8}$$

The plot is not linear for large values of v because the potential curve is not well described by the anharmonic potential energy function.

Q13.21

An analysis of the vibrational spectrum of the ground-state homonuclear diatomic molecule Na_2 gives $\tilde{\nu}_e = 159.125\ \ cm^{-1}$ and $\tilde{\nu}_e\tilde{x}_e = 0.7255\ \ cm^{-1}$. Suggest an experimental method that can be used to determine there spectroscopic parameters. Use $\nu_{max} = \dfrac{1}{2\tilde{x}_e} - 1$ to determine the number of bound vibrational levels for ground state of Na_2.

S13.21

First solve for $\tilde{x}$:

$$\tilde{x} = \frac{0.7255\, cm^{-1}}{159.125\, cm^{-1}} = 4.5593 \times 10^{-3}\, cm^{-1} \tag{10.E.9}$$

Then plug it in to the equation:

$$\nu_{max} = \frac{1}{2 \times 4.5593 \times 10^{-3}\, cm^{-1}} - 1 = 109.67\, cm^{-1} \tag{10.E.10}$$

There are 109 bound vibrational levels for the ground state of Ne_2.

Q13.22

A Morse potential is a decent representation of internuclear potential and is modeled by

$$U(q) = D_e(1 - e^{-/betaq})^2 \tag{10.E.11}$$

where $q = R - R_e$. Prove the force constant for the Morse Potential is evaluated

$$k = 2D_e\beta^2 \tag{10.E.12}$$

If $D_e = 6.23 \times 10^{-19} J \cdot Molecule^{-1}$ and $\beta = 1.37 \times 10^{10} m^{-1}$ for a given compound, calculate the force constant for that compound.

S13.22

Using a Maclaurin series expansion of the Morse potential, the formula can be rewritten

$$U(q) = D_e(1 - e^{-\beta q})^2 = D_e(1 - (1 - \beta q + \frac{\beta^2 q^2}{2} O(q)^3))^2 = D_e(\beta^2 q^2 + O(q)^3) \tag{10.E.13}$$

We know that

$$U(q) = \frac{kq^2}{2} \tag{10.E.14}$$

assuming that $O(q)^3$ approaches zero, we get

$$\frac{kq^2}{2} = D_e(\beta^2 q^2 + O(q)^3) \implies k = 2D_e\beta^2 \tag{10.E.15}$$

Plugging in our given values, we get

$$l = 2(6.23 \times 10^{-19} J \cdot Molecule^{-1})(1.37 \times 10^{10})^2 = 234 N \cdot m^{-1} \tag{10.E.16}$$

Q13.23

Given that $D_e = 7.33 \times 10^{-19} J \cdot molecule^{-1}, \tilde{v}_e = 1580.0 cm^{-1}, and R_e = 121 pm$, find k and β for $^{16}O_2$.

S13.23

Assuming a a harmonic oscillator model, we can use the equation:

$$\tilde{v}_e = \frac{1}{2\pi c}\left(\frac{k}{\mu}\right)^{1/2} \tag{10.E.17}$$

We can find k from the parameters given in the problem. Solving for k gives

$$k = (2\pi c \tilde{v}_e)^2 \mu \tag{10.E.18}$$

$$= [2\pi(3 \times 10^{10} cm \cdot s^{-1})(1580.0 cm^{-1})]^2 (7.9975 amu)(1.661 x 10^{-27} kg \cdot amu^{-1}) \tag{10.E.19}$$

$$= 1176.3 N \cdot m^{-1} \tag{10.E.20}$$

Using the equation:

$$\beta = \left(\frac{k}{2D_e}\right)^{1/2} \tag{10.E.21}$$

We can solve for β using the information above

$$\beta = \left[\frac{1176.3 N \cdot m^{-1}}{2(7.33 \times 10^{-19} J)}\right]^{1/2} \tag{10.E.22}$$

$$= 2.84 \times 10^{10} m^{-1} \tag{10.E.23}$$

Q13.24

The fundamental line in the IR spectrum of $^{12}C^{16}O\ cm^{-1}$, and the first overtone occurs at 4260.0 cm^{-1}. Find $\tilde{v}_e$ and $\tilde{x}_e\tilde{v}_e$ for $^{12}C^{16}O$.

S13.24

The equations for the fundamental frequency and the overtone are

$$\tilde{v} = \tilde{v}_e - 2\tilde{x}_e\tilde{v}_e \tag{10.E.24}$$

and

$$\tilde{v} = 2\tilde{v}_e - 6\tilde{x}_e\tilde{v}_e, \tag{10.E.25}$$

respectively.

We can set the fundamental frequency equal to $2143.0\ cm^{-1}$ and the first overtone equal to $4260.0\ cm^{-1}$ which is given from the problem statement. Then, we want to solve for $\tilde{x}_e\tilde{v}_e$.

Multiply the fundamental frequency by 3 and subtract the overtone.

$$\tilde{v}_e = [3(2143.0) - 4260.0] cm^{-1} = 2169.0 cm^{-1} \tag{10.E.26}$$

Then multiply the fundamental frequency by 2 and subtract from the overtone.

$$\tilde{x}_e\tilde{v}_e = 13.0 cm^{-1} \tag{10.E.27}$$

Q13.24

Which of the following molecules exhibit a microwave rotational absorption spectrum: BF_2,SO_2,C_2F_2,NO_3^-

S13.24

SO_2

notes: microwave rotational absorption correlates to the rotation of polyatomic molecules. Would BF_2 and NO_3^- also fall in this range?

Q13-25 Slightly Incorrect

Calculate the fundamental and the first two overtones of

$$H^{35}Cl \tag{10.E.28}$$

given

$$\tilde{v_e} = 2990.946 cm^{-1} \tag{10.E.29}$$

$$\tilde{x_e}\tilde{v_e} = 52.819 cm^{-1} \tag{10.E.30}$$

S13-25

Use equation,

$$\tilde{v_{obs}} = \tilde{v_e}\nu - \tilde{x_e}\tilde{v_e}\nu(\nu+1) \tag{10.E.31}$$

with v =1,2...

The fundamental frequency is given by v=1 and the first two overtones are given by v=2 and v=3.

Fundamental:

$$\tilde{v_{obs}} = \tilde{v_e} - 2\tilde{x_e}\tilde{v_e} = 2990.946 cm^{-1} - 2(52.819 cm^{-1}) = 2885.308 cm^{-1} \tag{10.E.32}$$

First overtone:

$$\tilde{v_{obs}} = 2\tilde{v_e} - 6\tilde{x_e}\tilde{v_e} = 2990.946 cm^{-1} - 6(52.819 cm^{-1}) = 5664.978 cm^{-1} \tag{10.E.33}$$

Second overtone:

$$\tilde{v_{obs}} = 3\tilde{v_e} - 12\tilde{x_e}\tilde{v_e} = 3(2990.946 cm^{-1}) - 12(52.819 cm^{-1}) = 8339.01 cm^{-1} \tag{10.E.34}$$

Q13.26

Plot the

$$\frac{\nu_{obs}}{\nu} \tag{10.E.35}$$

versus

$$(\nu+1) \tag{10.E.36}$$

for the anharmonic oscillator approximation and use it to determine the values of

$$\nu_e \tag{10.E.37}$$

and

$$x_e\nu_e \tag{10.E.38}$$

for

$$H^{3}5Cl \tag{10.E.39}$$

. Use the information in Table13.4.

S13.26

$$\nu_{obs} = \nu_{obs}\nu - x_e\nu_e\nu(\nu+1) \tag{10.E.40}$$

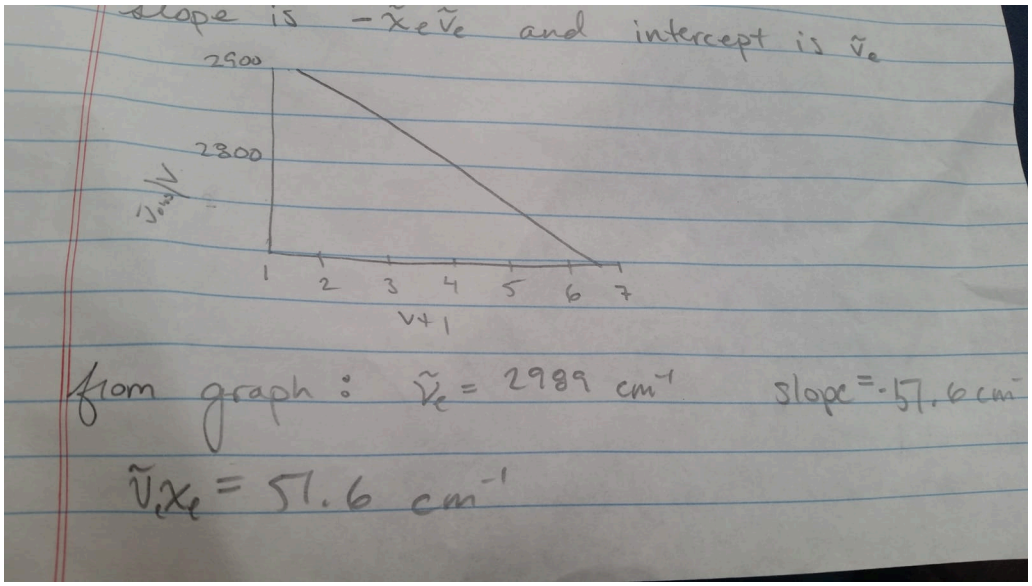
$$\frac{\nu_{obs}}{\nu} = \nu_e - x_e\nu_e(\nu+1) \tag{10.E.41}$$

The slope is

$$-x_e\nu_e \tag{10.E.42}$$

and the intercept is

$$\nu_e \tag{10.E.43}$$



The problem is nice and straight forward. It would be great for an explanation of why the slope is negative for an insight of the graph

Q13.31

Phosphorus nitride (PN) is a heteronuclear diatomic molecule that is a similar to N_2 with a triple bond, although is substantially weaker because P is a larger atomic radius than N. Using the following data, calculate $\tilde{\nu_e}'$ and $\tilde{\chi_e}'\ \tilde{\nu_e}'$ for PN.

IR Transition	$\nu_{\tilde{abs}}$ [cm^{-1}]
$0 \rightarrow 0$	39,699.10
$0 \rightarrow 1$	40,786.80
$0 \rightarrow 2$	41,858.90

S13.31

We will use the following equation:

$$\nu_{\tilde{abs}} = \nu_{\tilde{0,0}}' + \tilde{\nu_e}'\tilde{\nu}' - \tilde{\chi_e}'\tilde{\nu_e}'\tilde{\nu}'(\nu'+1) \tag{10.E.44}$$

Plugging in the appropriate values of ν', we get:

According to your table $\tilde{\nu_{0,0}}'$ should be 39,699.10 not 29,699.10.

- $\nu' = 0$:

$$\tilde{\nu_{0,0}}' = 29,699.10\ cm^{-1} \tag{1}$$

- $\nu' = 1$:

$$\tilde{\nu_{0,0}}' + \tilde{\nu_e}' - 2\tilde{\chi_e}'\tilde{\nu_e}' = 40,786.80\ cm^{-1} \tag{2}$$

- $\nu' = 2$:

$$\tilde{\nu_{0,0}}' + 2\tilde{\nu_e}' - 6\tilde{\chi_e}'\tilde{\nu_e}' = 41,858.90\ cm^{-1} \tag{3}$$

Solving as a system of equations, we subtract:

(2) - (1) = (4), (3) - (1) = (5), and finally (5) - 2x(4) = (6)

Equation (6) can be solved for $\tilde{\chi_e}'\tilde{\nu_e}'$, and then plugged back in to equation (4) or equation (5) to solve for $\tilde{\nu_e}'$. The answer is:

$$\tilde{\chi_e}'\tilde{\nu_e}' = 1103.3\ cm^{-1} \tag{10.E.45}$$

and

$$\tilde{\nu_e}' = 7.80\ cm^{-1} \tag{10.E.46}$$

from my calculations $\tilde{\nu_e}' = 1103.3$ and $\tilde{\chi_e}'\tilde{\nu_e}' = 7.80$ it looks like you had them switched

Q13.32

The frequencies of the first few vibronic transitions to an excited state of NaCl* are as follows:

Vibronic Transitions	$0 \rightarrow 0$	$0 \rightarrow 1$	$0 \rightarrow 2$
$\tilde{v}_{obs}/cm^{-}1$	12376	89706	15219

*not accurate values for NaCl

Use these data to calculate the values of $\tilde{v}_e$ and $\tilde{v}_e\tilde{x}_e$ for the excited state of NaCl

S13.32

$A = \tilde{v}_e', B = \tilde{v}_e'\tilde{x}_e'$

$12376 = \tilde{v}_0, 0$

$89706 = \tilde{v}_0, 0 + \tilde{v}_e' - 2\tilde{v}_e'\tilde{x}_e' - 12376 = \tilde{v}_0, 0 = 77330 = +\tilde{v}_e' - 2\tilde{v}_e'\tilde{x}_e'$ (1)

$15219 = \tilde{v}_0, 0 + 2\tilde{v}_e' - 6\tilde{v}_e'\tilde{x}_e' - 12376 = \tilde{v}_0, 0 = 2843 = 2\tilde{v}_e' - 6\tilde{v}_e'\tilde{x}_e'$ (2)

77330 + 2B = A (1a)

2843 = 2A - 6B (2a)

77330+2B = A

2843 - 154660 = -2B

B = 75908.5 cm^{-1} = $\tilde{v}_e'$

A = 229147 cm^{-1} = $\tilde{v}'_e\tilde{x}'_e$

Q13.33

Determine the number of translational, rotational, and vibrational degrees of freedom in

a. Xe
b. HCl
c. CS_2
d. hemoglobin containing 9272 molecules

The total number of degrees of freedom is 3N, where N is the number of atoms in the molecule. All molecules have three translational degrees of freedom. A nonlinear molecule has three rotational degrees of freedom and a linear molecule has two rotational degrees of freedom. A linear molecule has $3N-5$ vibrational degrees of freedom and a nonlinear molecule has 3N — 6 vibrational degrees of freedom.

S13.33

The number of translational degrees of freedom is 3, the number of rotational degrees of freedom is 2 for linear molecules and 3 for non-linear molecules. Therefore, the remaining degrees of freedom, which is equal to 3N-5 for linear molecules or 3N-6 for non-linear molecules, is equal to the number of vibrational degrees of freedom.

	Translation	Rotation	Vibration	Total
Xe	3	0	0	3
HCl	3	2	1	6
CS_2	3	2	4	9
hemoglobin	3	3	27810	27816

Q13.36

Classify each of the following molecules as either a prolate or an oblate symmetric top: XeF_4, $ClCH_3$, NH_3, and C_6H_6.

S13.36

- XeF_4 :oblate
- $ClCH_3$:prolate
- NH_3:prolate
- C_6H_6:oblate

Q13.37

Solve for the components of the moment of inertia of a trigonal planar molecule if all the masses are m, the bond lengths are unit length, and all bond angles are 120°.

S13.37

$$I_{xx} = \sum m_j y_j^2 = m(1)^2 + 2m(sin^2 30^\circ) = \frac{3}{2}m \tag{10.E.47}$$

$$I_{yy} = \sum m_j x_j^2 = m(0)^2 + 2m(cos^2 30^\circ) = \frac{3}{2}m \tag{10.E.48}$$

$$I_{zz} = \sum m_j x_j^2 + \sum m_j y_j^2 = 3m \tag{10.E.49}$$

Q13.38

This problem illustrates how the principal moments of inertia can be obtained as an eigenvalue problem. Consider a molecule where all the masses are unit masses and the long and short bond lengths are 2 and 1, respectively.

Show that

$$I_{xx} = 2\cos^2\theta + 8\sin^2\theta \tag{10.E.50}$$

$$I_{yy} = 8\cos^2\theta + 2\sin^2\theta \tag{10.E.51}$$

$$I_{xy} = -6\cos\theta\sin\theta \tag{10.E.52}$$

Solve the secular determinant equation for λ

$$\begin{vmatrix} I_{xx} - \lambda & I_{xy} \\ I_{xy} & I_{yy} - \lambda \end{vmatrix} = 0 \tag{10.E.53}$$

and compare the result with the values of I_{xx} and I_{yy} that you would obtain if you align the "molecule" and coordinate system such that $\theta = 180^\circ$. What does this comparison tell you? What are the values of I_{xx} and I_{yy} if $\theta = 90$?

S13.38

We use trigonometric functions to find the x and y components

$$I_{xx} = \sum m_i y_i^2 = 2(1)(2\sin\theta)^2 + 2(1)\left[\sin\left(\frac{\pi}{2}\right) - \theta\right]^2 = 8\sin^2\theta + 2\cos^2\theta \tag{10.E.54}$$

$$I_{yy} = \sum m_i x_i^2 = 2(1)(2\cos\theta)^2 + 2(1)\left[\cos\left(\frac{\pi}{2}\right) - \theta\right]^2 = 8\cos^2\theta + 2\sin^2\theta \tag{10.E.55}$$

$$I_{xy} = -\sum m_i x_i y_i = -(2\cos\theta)(2\sin\theta) - 2\left[\cos(\pi) + \theta\right]\left[\sin(\pi) + \theta\right] - \left[\sin\left(\frac{\pi}{2}\right) - \theta\right]\left[\cos\left(\frac{\pi}{2}\right) - \theta\right] - \left[\cos\left(\frac{\pi}{2}\right) + \theta\right]\left[\sin\left(\frac{\pi}{2}\right.\right.$$

$$= -6\cos\theta\sin\theta$$

The secular determinantal equation becomes

$$\begin{vmatrix} 8\sin^2\theta + 2\cos^2\theta - \lambda & -6\cos\theta\sin\theta \\ -6\cos\theta\sin\theta & 8\cos^2\theta + 2\sin^2\theta - \lambda \end{vmatrix} = 0 \tag{10.E.57}$$

Expanding the determinant gives

$$0 = \lambda^2 - \lambda(10)(\sin\theta^2 + \cos\theta^2) + 16(\sin\theta^2 + \cos\theta^2)^2 \tag{10.E.58}$$

$$0 = \lambda^2 - 10\lambda + 16 \tag{10.E.59}$$

$$\lambda = 5 \pm 3 \tag{10.E.60}$$

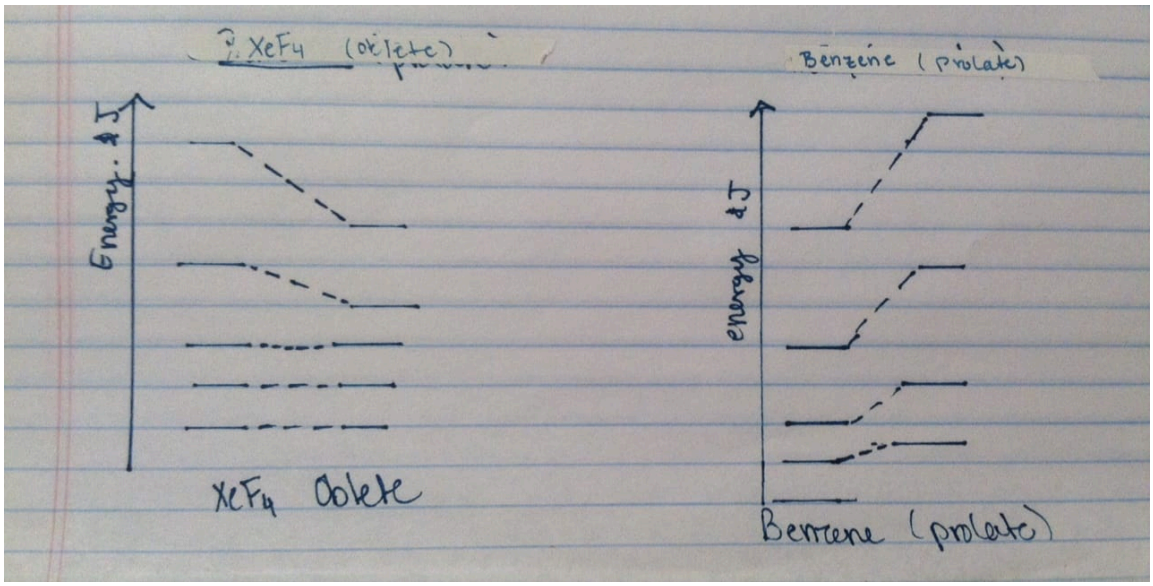
If $\theta = 180^\circ$ then $I_{xx} = 2, I_{yy} = 8, and I_{xy} = 0$. If $\theta = 90^\circ$ then $I_{xx} = 8, I_{yy} = 2, and I_{xy} = 0$. This tells us that the coordinate system does not affect the values of the principal momentum of inertia.

Q13.39

Sketch the energy level diagram of NH_3 (a prolate symmetric top molecule) and XeF_4 (an oblate symmetric top). How do they differ?

S13.39

The energies increase as the J levels increases for the prolate symmetry and the energies decrease as J increases for the oblate symmetry



This page titled 10.E: Molecular Spectroscopy (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

INDEX

A

absolute zero
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law

absorption properties
10.1: The Electromagnetic Spectrum

acetoacetate
8.5.6: Oxidation of fatty acids

acetone
8.5.6: Oxidation of fatty acids

activated complex
2.3.5: Collision Theory

activation energy
2.3.5: Collision Theory
6.8: Temperature Dependence

activation of a fatty acids
8.5.6: Oxidation of fatty acids

adenosine diphosphate
8.5.1: Basics of metabolism

adenosine monophosphate
8.5.1: Basics of metabolism

adenosine triphosphate
8.5.1: Basics of metabolism

ADP
8.5.1: Basics of metabolism

Aeration
8.2.10: Water Treatment

aerobic condtions
8.5.3: Oxidation of glucose -the glycolysis

aerobic respiration
8.5.3: Oxidation of glucose -the glycolysis

agave
8.1.4.4: Agave

albumen
8.1.1.4: Coagulation

amino acids
8.5.7: Degradation of amino acids

ammonium bicarbonate
8.1.5.7: Ammonium Bicarbonate

Amonton's law
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law

AMP
8.5.1: Basics of metabolism

amphiprotic specie
8.2.8: Water Chemistry

amylase
8.5.2: Digestion of food

anabolism
8.5.1: Basics of metabolism

anaerobic condtions
8.5.3: Oxidation of glucose -the glycolysis

anaerobic respiration
8.5.3: Oxidation of glucose -the glycolysis

angular momentum
10.8: Rotational Spectra of Polyatomic Molecules
9.8: Angular Momentum

angular motion
10.8: Rotational Spectra of Polyatomic Molecules

angular_momentum
10.4: Unequal Spacings in Pure Rotational Spectra

anharmonic effects
10: Molecular Spectroscopy

anharmonicity
10.5: Vibrational Overtones

aquatic chemistry
8.2: Aquatic Chemistry

Arrhenius equation
2.3.5: Collision Theory
6.8: Temperature Dependence

ATP
8.5.1: Basics of metabolism

ATP synthase
8.5.5: Oxidative Phosphorylation

ATP synthesis
8.5.5: Oxidative Phosphorylation

ATP yield
8.5.6: Oxidation of fatty acids

Author tag:OpenStax
2.1.3: The Kinetic-Molecular Theory
2.1.4: Non-Ideal Gas Behavior
2.2.1: Energy Basics
2.2.3: Spontaneity
2.2.5: The Second and Third Laws of Thermodynamics
2.3.1: Chemical Reaction Rates
2.3.2: Factors Affecting Reaction Rates
2.3.3: Rate Laws
2.3.4: Integrated Rate Laws
2.3.5: Collision Theory
2.3.6: Reaction Mechanisms
2.3.7: Catalysis

autonumheader:yes2
3: The First Law of Thermodynamics
3.1: Overview of Classical Thermodynamics
3.10: Enthalpy Changes for Chemical Equations are Additive
3.11: Heats of Reactions Can Be Calculated from Tabulated Heats of Formation
3.12: The Temperature Dependence of ΔH
3.3: Work and Heat are not State Functions
3.4: Energy is a State Function
3.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred
3.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion
3.7: Pressure-Volume Work
3.8: Heat Capacity is a Path Function
3.9: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition
3.E: The First Law of Thermodynamics (Exercises)
4: Entropy and The Second and 3rd Law of Thermodynamics
4.1: Energy Does not Determine Spontaneity
4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal
4.3: Unlike heat, Entropy is a State Function
4.4: The Second Law of Thermodynamics
4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes
4.6: Entropy Increases With Increasing Temperature
4.7: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale
4.8: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition
4.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions
4.E: Entropy and The Second Law of Thermodynamics (Exercises)
5: Helmholtz and Gibbs Energies
5.6: Thermodynamic Functions have Natural Variables
5.7: The Gibbs-Helmholtz Equation

average rate
2.3.1: Chemical Reaction Rates

avogadro's law
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law

B

baking powder
8.1.5.5: Baking Powder

baking soda
8.1.5.6: Sodium Bicarbonate

basic
8.3: Biochemical Cycles
8.3.2: Carbon Cycle
8.3.3: Nitrogen Cycle

bending vibrations
10.13: Group Theory Determines Infrared Activity

bimolecular reaction
2.3.6: Reaction Mechanisms
7.1: Reaction Mechanisms

biotransformation
8.4.6.12.1: Introduction to Biotransformation
Boyle's law
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
bromate
8.1.2.6: Flour Additives
brown fat
8.5.5: Oxidative Phosphorylation

C

calorie
2.2.1: Energy Basics
carbon cycle
8.3.2: Carbon Cycle
casein
8.1.1.4: Coagulation
catabolism
8.5.1: Basics of metabolism
Catabolism of glycerol
8.5.6: Oxidation of fatty acids
catalyst
2.3.2: Factors Affecting Reaction Rates
7.9: Catalysis
catalytic decomposition of ozone
7.9: Catalysis
cell structure
8.5.1: Basics of metabolism
centrifugal_force
10.4: Unequal Spacings in Pure Rotational Spectra
character table
10.10: Irreducible Representation of Point Groups
Charles's law
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
chemical thermodynamics
2.2.2: Enthalpy
chymotripsin
8.5.2: Digestion of food
circular integrals
4.3: Unlike heat, Entropy is a State Function
citrate
8.5.4: Citric acid cycle
citrate synthase
8.5.1: Basics of metabolism
citric acid cycle
8.5.4: Citric acid cycle
CoA
8.5.1: Basics of metabolism
coagulation
8.1.1.4: Coagulation
8.2.10: Water Treatment
Coenzyme A
8.5.1: Basics of metabolism
collision theory
2.3.5: Collision Theory
7.10: Theories of Reaction Rates
complex I
8.5.5: Oxidative Phosphorylation
complex II
8.5.5: Oxidative Phosphorylation
complex III
8.5.5: Oxidative Phosphorylation
complex IV
8.5.5: Oxidative Phosphorylation
complex V
8.5.5: Oxidative Phosphorylation
compressibility factor
2.1.4: Non-Ideal Gas Behavior
concentration profiles
7.2: Concentration Profiles for Some Simple Mechanisms
cooking
8.1: Chemistry of Cooking (Rodriguez-Velazquez)
cornstarch
8.1.1.3: Types of Thickening Agents
couverture
8.1.8.3: Couverture
cyanine dyes
10.7.2: Cyanine Dyes
Cytochrome c
8.5.5: Oxidative Phosphorylation
cytochrome c oxidase
8.5.5: Oxidative Phosphorylation
cytoplasm
8.5.1: Basics of metabolism
cytosol
8.5.1: Basics of metabolism
cytotoxicity assays
8.4.3.4: Different Cytotoxicity Assays

D

Dalton's Law of Partial Pressure
2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions
degrees of freedom
10.9: Normal Modes in Polyatomic Molecules
dextrose equivalent
8.1.4.9: Glucose/Dextrose
diatomic
10.8: Rotational Spectra of Polyatomic Molecules
diatomic molecules
10.2: Rotations Accompany Vibrational Transitions
10.5: Vibrational Overtones
10.E: Molecular Spectroscopy (Exercises)
diatomic_molecule
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
digestion
8.5.2: Digestion of food
dipole moment
10.11: The Selection Rule for the Rigid Rotor
dissociation energy
10.E: Molecular Spectroscopy (Exercises)
DNA replication
8.4.2.3: DNA and RNA Metabolism

E

egg
8.1.7: Eggs
electric dipole moment
10.13: Group Theory Determines Infrared Activity
electromagnetic radiation
10: Molecular Spectroscopy
10.1: The Electromagnetic Spectrum
electron transport
8.5.5: Oxidative Phosphorylation
electron transport chain
8.5.5: Oxidative Phosphorylation
electronic_transitions
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
elementary reaction
2.3.6: Reaction Mechanisms
7.1: Reaction Mechanisms
emission spectroscopy
10.1: The Electromagnetic Spectrum
endothermic process
2.2.1: Energy Basics
energy
2.2.1: Energy Basics
energy level diagram
10.E: Molecular Spectroscopy (Exercises)
energy levels
10.5: Vibrational Overtones
10.8: Rotational Spectra of Polyatomic Molecules
energy quantization
10: Molecular Spectroscopy
10.2: Rotations Accompany Vibrational Transitions
energy_levels
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
enthalpy
2.2.2: Enthalpy
enthalpy change
2.2.2: Enthalpy
entropy
2.2.4: Entropy
entropy of mixing
4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes
eutrophication
8.3.1: Biochemical Cycles
exact differential
4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal
exothermic process
2.2.1: Energy Basics
expansion work
2.2.2: Enthalpy
experimental data
10.E: Molecular Spectroscopy (Exercises)

F

F0c ring
8.5.5: Oxidative Phosphorylation
FAD
8.5.1: Basics of metabolism
FADH2
8.5.1: Basics of metabolism
fermentation
8.5.3: Oxidation of glucose -the glycolysis
first law of thermodynamics
2.2.2: Enthalpy
first order reaction
6.4: 1st order rate law
Flavin adenine dinucleotide
8.5.1: Basics of metabolism

flavin mononucleotide
8.5.5: Oxidative Phosphorylation
Flucculation
8.2.10: Water Treatment
fluorescence
10.7.3: Fluorescence and Phosphorescence
free energy
5.2: ΔA, ΔG, and Maximum Work
free expansion
4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes
frequency factor
2.3.5: Collision Theory
fumarate
8.5.4: Citric acid cycle

G

Gibbs Free Energy
2.2.6: Free Energy
Gibbs function
5.1: Free Energy Functions
glucogenic
8.5.7: Degradation of amino acids
gluten
8.1.1.4: Coagulation
glycolysis
8.5.3: Oxidation of glucose -the glycolysis
group theory
10.13: Group Theory Determines Infrared Activity
10.9: Normal Modes in Polyatomic Molecules
gum tragacanth
8.1.1.3: Types of Thickening Agents

H

harmonic oscillator
10.2: Rotations Accompany Vibrational Transitions
10.5: Vibrational Overtones
heat
2.2.1: Energy Basics
heat capacity
2.2.1: Energy Basics
Helmholtz function
5.1: Free Energy Functions
Hess's law
2.2.2: Enthalpy
heterogeneous catalyst
2.3.7: Catalysis
hidetop:yes
8.2: Aquatic Chemistry
8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems
8.2.10: Water Treatment
8.2.2: 2. Carbonate Equilibria in Natural Waters
8.2.3: 3. Redox Equilibria in Natural Waters
8.2.4: 4. Solids in Contact With Natural Waters
8.2.5: Fundamental Characteristics of Water
8.2.6: Natural Water
8.2.7: Water Biology
8.2.8: Water Chemistry
8.2.9: Water Physics
8.3: Biochemical Cycles
8.3.1: Biochemical Cycles
8.3.2: Carbon Cycle
8.3.3: Nitrogen Cycle
homogeneous catalyst
2.3.7: Catalysis
honey
8.1.4.5: Honey
Hook_law
10.4: Unequal Spacings in Pure Rotational Spectra
hydrocarbon
2.2.2: Enthalpy
Hydrosphere
8.2.6: Natural Water

I

ideal gas
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
ideal gas constant
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
ideal gas law
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
infrared spectroscopy
10.1: The Electromagnetic Spectrum
infrared_absorption
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
initial rate
2.3.1: Chemical Reaction Rates
inner membrane
8.5.1: Basics of metabolism
instantaneous rate
2.3.1: Chemical Reaction Rates
integrated rate law
2.3.4: Integrated Rate Laws
intermediate
2.3.6: Reaction Mechanisms
intermembrane space
8.5.1: Basics of metabolism
internal energy
2.2.2: Enthalpy
invert sugar
8.1.4.5: Honey
IR active
10.13: Group Theory Determines Infrared Activity
IR spectroscopy
10.9: Normal Modes in Polyatomic Molecules
irreducible representation
10.10: Irreducible Representation of Point Groups
isolated system
4.4: The Second Law of Thermodynamics

J

joule
2.2.1: Energy Basics
Joule expansion
4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes

K

ketoacidosis
8.5.6: Oxidation of fatty acids
ketogenesis
8.5.6: Oxidation of fatty acids
ketogenic
8.5.7: Degradation of amino acids
Ketone bodies
8.5.6: Oxidation of fatty acids
ketosis
8.5.6: Oxidation of fatty acids
kinetic energy
10.8: Rotational Spectra of Polyatomic Molecules
2.2.1: Energy Basics
kinetic molecular theory
2.1.3: The Kinetic-Molecular Theory
kinetic_energy
10.4: Unequal Spacings in Pure Rotational Spectra
Krebs cycle
8.5.4: Citric acid cycle

L

lactase
8.5.2: Digestion of food
lard
8.1.3.3: Major Fats and Oils Used in Bakeries
Lindemann mechanism
7.7: The Lindemann Mechanism
lipase
8.5.2: Digestion of food

M

Maillard reaction
8.1.4.3: The Application of Sugar
malate
8.5.4: Citric acid cycle
malate dehydrogenase
8.5.1: Basics of metabolism
matix
8.5.1: Basics of metabolism
metabolic pathways
8.5.1: Basics of metabolism
metabolism
8.5.1: Basics of metabolism
Method of Initial Rates
2.3.3: Rate Laws
6.6: The Method of Initial Rates
Michaelis constant
7.8: The Michaelis-Menten Mechanism
microstates
2.2.4: Entropy
microwave spectroscopy
10.11: The Selection Rule for the Rigid Rotor
microwave_absorption
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
milk
8.1.6.1: Introduction to Dairy Products
mitochondria
8.5.1: Basics of metabolism
mole fraction
2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions
molecular bonds
10.5: Vibrational Overtones
molecular energy
10.E: Molecular Spectroscopy (Exercises)

molecular properties
10: Molecular Spectroscopy
molecular spectroscopy
10.1: The Electromagnetic Spectrum
molecular vibrations
10.9: Normal Modes in Polyatomic Molecules
molecularity
2.3.6: Reaction Mechanisms
7.1: Reaction Mechanisms
molecules
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
moment of inertia
10.8: Rotational Spectra of Polyatomic Molecules
Montreal Protocol
7.9: Catalysis
Morse potential
10.E: Molecular Spectroscopy (Exercises)

N

NAD+
8.5.1: Basics of metabolism
NADH
8.5.1: Basics of metabolism
NADH dehydrogenase
8.5.5: Oxidative Phosphorylation
natural logarithm
2.4: Using logarithms - Log vs. Ln
Nicotinamide adenine dinucleotide
8.5.1: Basics of metabolism
non_rigid_rotator
10.4: Unequal Spacings in Pure Rotational Spectra
nonlinear molecules
10.9: Normal Modes in Polyatomic Molecules
nonspontaneous process
2.2.3: Spontaneity
normal coordinates
10.10: Irreducible Representation of Point Groups
normal modes
10.10: Irreducible Representation of Point Groups
10.9: Normal Modes in Polyatomic Molecules
nuclear_kinetic_energy
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
nucleus
8.5.1: Basics of metabolism

O

organelle
8.5.1: Basics of metabolism
oscillations
10.4: Unequal Spacings in Pure Rotational Spectra
outer membrane
8.5.1: Basics of metabolism
overall reaction order
2.3.3: Rate Laws
overtones
10.5: Vibrational Overtones
oxaloacetate
8.5.4: Citric acid cycle
8.5.7: Degradation of amino acids
oxidation
8.5.1: Basics of metabolism
oxidative deamination
8.5.7: Degradation of amino acids
oxidative phosphorylation
8.5.5: Oxidative Phosphorylation
ozone
7.9: Catalysis

P

partial pressure
2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions
Pasteurization
8.1.6.2: Milk
8.1.6.7: Cheese
pectins
8.1.1.3: Types of Thickening Agents
pentosan
8.1.2.9: Rye Flour
pepsin
8.5.2: Digestion of food
Permanent Hard Water
8.2.6: Natural Water
peroxide
8.5.5: Oxidative Phosphorylation
perturbation theory
10.5: Vibrational Overtones
phosphorescence
10.7.3: Fluorescence and Phosphorescence
phosphorylation
8.5.1: Basics of metabolism
physical chemistry
10: Molecular Spectroscopy
10.1: The Electromagnetic Spectrum
10.E: Molecular Spectroscopy (Exercises)
polarizability tensor
10.13: Group Theory Determines Infrared Activity
polyatomic
10.8: Rotational Spectra of Polyatomic Molecules
potential energy
10.5: Vibrational Overtones
10.8: Rotational Spectra of Polyatomic Molecules
2.2.1: Energy Basics
potential_energy
10.4: Unequal Spacings in Pure Rotational Spectra
Pressure Dependence of Gibbs Energy
5.3: Pressure Dependence of Gibbs Energy
pyruvate
8.5.3: Oxidation of glucose -the glycolysis

Q

QSAR
8.4.4.4: Other Approaches for Predictive Toxicity Modeling
Quantitative Structure Activity Relationships
8.4.4.4: Other Approaches for Predictive Toxicity Modeling
quantum number
10.E: Molecular Spectroscopy (Exercises)
quantum selection rules
10: Molecular Spectroscopy

R

Raman active
10.13: Group Theory Determines Infrared Activity
Raman spectroscopy
10.9: Normal Modes in Polyatomic Molecules
Rapid Equilibrium Approximation
7.2: Concentration Profiles for Some Simple Mechanisms
7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws
7.6: The Equilibrium Approximation
rate constant
2.3.3: Rate Laws
6.1: Reaction Rate
rate determining step
7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws
7.4: The Rate Determining Step Approximation
rate expression
2.3.1: Chemical Reaction Rates
rate law
2.3.3: Rate Laws
6.2: Rate Laws
rate of reaction
2.3.1: Chemical Reaction Rates
reaction mechanism
2.3.6: Reaction Mechanisms
7.1: Reaction Mechanisms
reaction order
2.3.3: Rate Laws
6.2: Rate Laws
reaction rate
6.1: Reaction Rate
reactive oxygen species
8.5.5: Oxidative Phosphorylation
reduction
8.5.1: Basics of metabolism
Rennet
8.1.1.4: Coagulation
reversible process
2.2.4: Entropy
rigid rotor
10.2: Rotations Accompany Vibrational Transitions
rigid_rotator
10.4: Unequal Spacings in Pure Rotational Spectra
root mean square velocity
2.1.3: The Kinetic-Molecular Theory
rotational motion
10.9: Normal Modes in Polyatomic Molecules
rotational transitions
10: Molecular Spectroscopy
10.11: The Selection Rule for the Rigid Rotor
10.2: Rotations Accompany Vibrational Transitions
rotational_spectrum
10.4: Unequal Spacings in Pure Rotational Spectra
rovibrational
10.3: Unequal Spacings in Vibration-Rotation Spectra

rovibrational spectra
10.2: Rotations Accompany Vibrational Transitions
rovibrational_transitions
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information
Rye Flour
8.1.2.9: Rye Flour

S

Schrödinger equation
10.8: Rotational Spectra of Polyatomic Molecules
Second Law of Thermodynamics
2.2.5: The Second and Third Laws of Thermodynamics
second order reaction
6.5: 2nd order Rate Laws
sedimentation
8.2.10: Water Treatment
selection rules
10.11: The Selection Rule for the Rigid Rotor
10.12: The Harmonic Oscillator Selection Rule
10.2: Rotations Accompany Vibrational Transitions
soft water
8.1.5.8: Water Hardness and pH
specific heat capacity
2.2.1: Energy Basics
spectra
10.5: Vibrational Overtones
spectral analysis
10.2: Rotations Accompany Vibrational Transitions
spectroscopic techniques
10.1: The Electromagnetic Spectrum
spectroscopy
10: Molecular Spectroscopy
10.E: Molecular Spectroscopy (Exercises)
spontaneous change
2.2.3: Spontaneity
Stage 1 of food catabolism
8.5.2: Digestion of food
stages of food catabolism
8.5.1: Basics of metabolism
standard conditions of temperature and pressure
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
standard enthalpy of combustion
2.2.2: Enthalpy
standard enthalpy of formation
2.2.2: Enthalpy
standard entropy
2.2.5: The Second and Third Laws of Thermodynamics
standard entropy change
2.2.5: The Second and Third Laws of Thermodynamics
standard free energy change
2.2.6: Free Energy
standard free energy of formation
2.2.6: Free Energy
standard molar volume
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
standard state
2.2.2: Enthalpy
state function
2.2.2: Enthalpy
4.3: Unlike heat, Entropy is a State Function
steady state approximation
7.2: Concentration Profiles for Some Simple Mechanisms
7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws
7.5: The Steady-State Approximation
STP
2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law
succinate
8.5.4: Citric acid cycle
succinate dehydrogenase
8.5.5: Oxidative Phosphorylation
Sucrase
8.5.2: Digestion of food
superoxide
8.5.5: Oxidative Phosphorylation
symmetric stretches
10.13: Group Theory Determines Infrared Activity

T

temperature
2.2.1: Energy Basics
Temperature Dependence of Gibbs Energy
5.4: Temperature Dependence of A and G
Temperature Dependence of Helmholtz Energy
5.4: Temperature Dependence of A and G
Temperature Dependence to Rates
6.8: Temperature Dependence
Temporary Hard Water
8.2.6: Natural Water
termolecular reaction
2.3.6: Reaction Mechanisms
7.1: Reaction Mechanisms
thermal energy
2.2.1: Energy Basics
thermochemistry
2.2.1: Energy Basics
Third Law of Thermodynamics
2.2.5: The Second and Third Laws of Thermodynamics
transamination
8.5.7: Degradation of amino acids
Transition dipole moment
10.2: Rotations Accompany Vibrational Transitions
transition moment integral
10.13: Group Theory Determines Infrared Activity
transition state theory
7.10: Theories of Reaction Rates
translational motion
10.9: Normal Modes in Polyatomic Molecules
tricarboxylic acid cycle
8.5.4: Citric acid cycle
trypsin
8.5.2: Digestion of food
Two Variables Change at Once
5.5: When Two Variables Change at Once

U

ubiquinone
8.5.5: Oxidative Phosphorylation
uncouplers
8.5.5: Oxidative Phosphorylation
unimolecular reaction
2.3.6: Reaction Mechanisms
7.1: Reaction Mechanisms
Unsaturated fatty acids
8.5.6: Oxidation of fatty acids
urea cycle
8.5.7: Degradation of amino acids
UV/Vis Spectroscopy
10.1: The Electromagnetic Spectrum

V

van der Waals equation
2.1.4: Non-Ideal Gas Behavior
vapor pressure of water
2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions
vibrational levels
10.E: Molecular Spectroscopy (Exercises)
vibrational motion
10.5: Vibrational Overtones
vibrational motions
10.9: Normal Modes in Polyatomic Molecules
vibrational transitions
10: Molecular Spectroscopy
10.2: Rotations Accompany Vibrational Transitions
vibrational wavefunctions
10.10: Irreducible Representation of Point Groups
vibrational_energy
10.4: Unequal Spacings in Pure Rotational Spectra
vibronic_transitions
10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information

W

water
8.2: Aquatic Chemistry
8.2.5: Fundamental Characteristics of Water
8.2.6: Natural Water
8.2.8: Water Chemistry
wheat flour
8.1.2.4: Flour Streams and Types of Wheat Flour
Whole grain
8.1.2.7: Whole Grain and Artisan Milling
work
2.2.1: Energy Basics

Y

yeast
8.1.5.2: Yeast

Z

zeroth order reaction
6.3: 0th order Rate Law

This page titled Index is shared under a license and was authored, remixed, and/or curated by LibreTexts.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a license and was authored, remixed, and/or curated by LibreTexts.

DETAILED LICENSING

Overview

Title: CHEM 2141: Survey of Physical Chemistry

Webpages: 347

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 51.3% (178 pages)
- CC BY-NC 4.0: 32.9% (114 pages)
- CC BY 4.0: 7.5% (26 pages)
- Undeclared: 4.9% (17 pages)
- Public Domain: 2.3% (8 pages)
- CC BY-SA 4.0: 1.2% (4 pages)

By Page

- CHEM 2141: Survey of Physical Chemistry - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Introduction - *CC BY 4.0*
    - 1.1: Units - *CC BY 4.0*
    - 1.2: Quantity Calculus - *CC BY 4.0*
    - 1.3: Dimensional Analysis - *CC BY 4.0*
    - 1.4: Chapter 1 Problem - *CC BY 4.0*
  - 2: General Chemistry Review - *Undeclared*
    - 2.1: Gases - *Undeclared*
      - 2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law - *CC BY 4.0*
      - 2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions - *CC BY 4.0*
      - 2.1.3: The Kinetic-Molecular Theory - *CC BY 4.0*
      - 2.1.4: Non-Ideal Gas Behavior - *CC BY 4.0*
    - 2.2: Thermodynamics - *Undeclared*
      - 2.2.1: Energy Basics - *CC BY 4.0*
      - 2.2.2: Enthalpy - *CC BY 4.0*
      - 2.2.3: Spontaneity - *CC BY 4.0*
      - 2.2.4: Entropy - *CC BY 4.0*
      - 2.2.5: The Second and Third Laws of Thermodynamics - *CC BY 4.0*
      - 2.2.6: Free Energy - *CC BY 4.0*
    - 2.3: Kinetics - *Undeclared*
      - 2.3.1: Chemical Reaction Rates - *CC BY 4.0*
      - 2.3.2: Factors Affecting Reaction Rates - *CC BY 4.0*
      - 2.3.3: Rate Laws - *CC BY 4.0*
      - 2.3.4: Integrated Rate Laws - *CC BY 4.0*
      - 2.3.5: Collision Theory - *CC BY 4.0*
      - 2.3.6: Reaction Mechanisms - *CC BY 4.0*
      - 2.3.7: Catalysis - *CC BY 4.0*
    - 2.4: Using logarithms - Log vs. Ln - *CC BY-NC-SA 4.0*
  - 3: The First Law of Thermodynamics - *CC BY-NC-SA 4.0*
    - 3.1: Overview of Classical Thermodynamics - *CC BY-NC-SA 4.0*
    - 3.2: Pressure-Volume Work - *CC BY 4.0*
    - 3.3: Work and Heat are not State Functions - *CC BY-NC-SA 4.0*
    - 3.4: Energy is a State Function - *CC BY-NC-SA 4.0*
    - 3.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred - *CC BY-NC-SA 4.0*
    - 3.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion - *CC BY-NC-SA 4.0*
    - 3.7: Pressure-Volume Work - *CC BY-NC-SA 4.0*
    - 3.8: Heat Capacity is a Path Function - *CC BY-NC-SA 4.0*
    - 3.9: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition - *CC BY-NC-SA 4.0*
    - 3.10: Enthalpy Changes for Chemical Equations are Additive - *CC BY-NC-SA 4.0*
    - 3.11: Heats of Reactions Can Be Calculated from Tabulated Heats of Formation - *CC BY-NC-SA 4.0*
    - 3.12: The Temperature Dependence of ΔH - *CC*

BY-NC-SA 4.0
- 3.13: Enthalpy is a State Function - *Undeclared*
- 3.E: The First Law of Thermodynamics (Exercises) - *CC BY-NC-SA 4.0*

◦ 4: Entropy and The Second and 3rd Law of Thermodynamics - *CC BY-NC-SA 4.0*
- 4.1: Energy Does not Determine Spontaneity - *CC BY-NC-SA 4.0*
- 4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal - *CC BY-NC-SA 4.0*
- 4.3: Unlike heat, Entropy is a State Function - *CC BY-NC-SA 4.0*
- 4.4: The Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
- 4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes - *CC BY-NC-SA 4.0*
- 4.6: Entropy Increases With Increasing Temperature - *CC BY-NC-SA 4.0*
- 4.7: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale - *CC BY-NC-SA 4.0*
- 4.8: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition - *CC BY-NC-SA 4.0*
- 4.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions - *CC BY-NC-SA 4.0*
- 4.E: Entropy and The Second Law of Thermodynamics (Exercises) - *CC BY-NC-SA 4.0*

◦ 5: Helmholtz and Gibbs Energies - *CC BY-NC-SA 4.0*
- 5.1: Free Energy Functions - *CC BY-NC-SA 4.0*
- 5.2: ΔA, ΔG, and Maximum Work - *CC BY-NC-SA 4.0*
- 5.3: Pressure Dependence of Gibbs Energy - *CC BY-NC-SA 4.0*
- 5.4: Temperature Dependence of A and G - *CC BY-NC-SA 4.0*
- 5.5: When Two Variables Change at Once - *CC BY-NC-SA 4.0*
- 5.6: Thermodynamic Functions have Natural Variables - *CC BY-NC-SA 4.0*
- 5.7: The Gibbs-Helmholtz Equation - *CC BY-NC-SA 4.0*
- 5.E: Gibbs Energy (Exercises) - *CC BY-NC-SA 4.0*

◦ 6: Chemical Kinetics I - *CC BY-NC-SA 4.0*
- 6.1: Reaction Rate - *CC BY-NC-SA 4.0*
- 6.2: Rate Laws - *CC BY-NC-SA 4.0*
- 6.3: 0th order Rate Law - *CC BY-NC-SA 4.0*
- 6.4: 1st order rate law - *CC BY-NC-SA 4.0*
- 6.5: 2nd order Rate Laws - *CC BY-NC-SA 4.0*
- 6.6: The Method of Initial Rates - *CC BY-NC-SA 4.0*
- 6.7: The Method of Half-Lives - *CC BY-NC-SA 4.0*
- 6.8: Temperature Dependence - *CC BY-NC-SA 4.0*
- 6.E: Chemical Kinetics I (Exercises) - *CC BY-NC-SA 4.0*
- 6.S: Chemical Kinetics I (Summary) - *CC BY-NC-SA 4.0*

◦ 7: Chemical Kinetics II - *CC BY-NC-SA 4.0*
- 7.1: Reaction Mechanisms - *CC BY-NC-SA 4.0*
- 7.2: Concentration Profiles for Some Simple Mechanisms - *CC BY-NC-SA 4.0*
- 7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws - *CC BY-NC-SA 4.0*
- 7.4: The Rate Determining Step Approximation - *CC BY-NC-SA 4.0*
- 7.5: The Steady-State Approximation - *CC BY-NC-SA 4.0*
- 7.6: The Equilibrium Approximation - *CC BY-NC-SA 4.0*
- 7.7: The Lindemann Mechanism - *CC BY-NC-SA 4.0*
- 7.8: The Michaelis-Menten Mechanism - *CC BY-NC-SA 4.0*
- 7.9: Catalysis - *CC BY-NC-SA 4.0*
- 7.10: Theories of Reaction Rates - *CC BY-NC-SA 4.0*
- 7.E: Chemical Kinetics II (Exercises) - *CC BY-NC-SA 4.0*
- 7.S: Chemical Kinetics II (Summary) - *CC BY-NC-SA 4.0*

◦ 8: Optional- Special topics - *Undeclared*
- 8.1: Chemistry of Cooking (Rodriguez-Velazquez) - *CC BY-NC-SA 4.0*
  - 8.1.1: Thickening and Concentrating Flavors - *CC BY-NC-SA 4.0*
    - 8.1.1.1: Viscosity - *CC BY-NC-SA 4.0*
    - 8.1.1.2: Thickening Agents - *CC BY-NC-SA 4.0*
    - 8.1.1.3: Types of Thickening Agents - *CC BY-NC-SA 4.0*
    - 8.1.1.4: Coagulation - *CC BY-NC-SA 4.0*
    - 8.1.1.5: Gelatinization - *CC BY-NC-SA 4.0*
    - 8.1.1.6: Crystallization - *CC BY-NC-SA 4.0*
    - 8.1.1.7: Non-traditional thickeners - *CC BY-NC-SA 4.0*
    - 8.1.1.8: Sauces - *CC BY-NC-SA 4.0*
    - 8.1.1.9: Low-temperature and sous-vide - *CC BY-NC-SA 4.0*
  - 8.1.2: Flour - *CC BY-NC-SA 4.0*
    - 8.1.2.1: Introduction - Understanding Ingredients - *CC BY-NC-SA 4.0*
    - 8.1.2.2: The History of Wheat Flour - *CC BY-NC-SA 4.0*
    - 8.1.2.3: Milling of Wheat - *CC BY-NC-SA 4.0*

- 8.1.2.4: Flour Streams and Types of Wheat Flour - *CC BY-NC-SA 4.0*
- 8.1.2.5: Flour Terms and Treatments - *CC BY-NC-SA 4.0*
- 8.1.2.6: Flour Additives - *CC BY-NC-SA 4.0*
- 8.1.2.7: Whole Grain and Artisan Milling - *CC BY-NC-SA 4.0*
- 8.1.2.8: Flour in Baking - *CC BY-NC-SA 4.0*
- 8.1.2.9: Rye Flour - *CC BY-NC-SA 4.0*
- 8.1.2.10: Other Grains and Flours - *CC BY-NC-SA 4.0*

- 8.1.3: Fat - *CC BY-NC-SA 4.0*
  - 8.1.3.1: Understanding Fats and Oils - *CC BY-NC-SA 4.0*
  - 8.1.3.2: Sources of Bakery Fats and Oils - *CC BY-NC-SA 4.0*
  - 8.1.3.3: Major Fats and Oils Used in Bakeries - *CC BY-NC-SA 4.0*
  - 8.1.3.4: Functions of Fat in Baking - *CC BY-NC-SA 4.0*
- 8.1.4: Sugar - *CC BY-NC-SA 4.0*
  - 8.1.4.1: Sugar Chemistry (ADD US) - *CC BY-NC-SA 4.0*
  - 8.1.4.2: Sugar Refining - *CC BY-NC-SA 4.0*
  - 8.1.4.3: The Application of Sugar - *CC BY-NC-SA 4.0*
  - 8.1.4.4: Agave - *CC BY-NC-SA 4.0*
  - 8.1.4.5: Honey - *CC BY-NC-SA 4.0*
  - 8.1.4.6: Malt - *CC BY-NC-SA 4.0*
  - 8.1.4.7: Maple Syrup (ADD US) - *CC BY-NC-SA 4.0*
  - 8.1.4.8: Sugar Substitutes (ADD US) - *CC BY-NC-SA 4.0*
  - 8.1.4.9: Glucose/Dextrose - *CC BY-NC-SA 4.0*
- 8.1.5: Leavening Agents - *CC BY-NC-SA 4.0*
  - 8.1.5.1: Introduction to Leavening Agents - *CC BY-NC-SA 4.0*
  - 8.1.5.2: Yeast - *CC BY-NC-SA 4.0*
  - 8.1.5.3: The Functions of Yeast - *CC BY-NC-SA 4.0*
  - 8.1.5.4: Using Yeast in Baking - *CC BY-NC-SA 4.0*
  - 8.1.5.5: Baking Powder - *CC BY-NC-SA 4.0*
  - 8.1.5.6: Sodium Bicarbonate - *CC BY-NC-SA 4.0*
  - 8.1.5.7: Ammonium Bicarbonate - *CC BY-NC-SA 4.0*
  - 8.1.5.8: Water Hardness and pH - *CC BY-NC-SA 4.0*
- 8.1.6: Dairy Products - *CC BY-NC-SA 4.0*
  - 8.1.6.1: Introduction to Dairy Products - *CC BY-NC-SA 4.0*
  - 8.1.6.2: Milk - *CC BY-NC-SA 4.0*
  - 8.1.6.3: Milk Products ADD US - *CC BY-NC-SA 4.0*
  - 8.1.6.4: Milk in bread baking - *CC BY-NC-SA 4.0*
  - 8.1.6.5: Yogurt - *CC BY-NC-SA 4.0*
  - 8.1.6.6: Lactose - *CC BY-NC-SA 4.0*
  - 8.1.6.7: Cheese - *CC BY-NC-SA 4.0*
- 8.1.7: Eggs - *CC BY-NC-SA 4.0*
  - 8.1.7.1: Eggs Grade - *CC BY-NC-SA 4.0*
  - 8.1.7.2: Composition and Nutrition - *CC BY-NC-SA 4.0*
  - 8.1.7.3: Egg Products - *CC BY-NC-SA 4.0*
  - 8.1.7.4: The Function of Eggs - *CC BY-NC-SA 4.0*
  - 8.1.7.5: Storing Eggs - *CC BY-NC-SA 4.0*
- 8.1.8: Chocolate - *CC BY-NC-SA 4.0*
  - 8.1.8.1: From the Cocoa Bean to the Finished Chocolate - *CC BY-NC-SA 4.0*
  - 8.1.8.2: Chocolate Produced for the Baking Industry - *CC BY-NC-SA 4.0*
  - 8.1.8.3: Couverture - *CC BY-NC-SA 4.0*
  - 8.1.8.4: Definitions and Regulations (ADD US) - *CC BY-NC-SA 4.0*
- 8.1.9: Spices - *CC BY-NC-SA 4.0*
  - 8.1.9.1: Elements of Taste - *CC BY-NC-SA 4.0*
  - 8.1.9.2: Introduction to Salt - *CC BY-NC-SA 4.0*
  - 8.1.9.3: Origins of Salt - *CC BY-NC-SA 4.0*
  - 8.1.9.4: Functions of Salt in Baking - *CC BY-NC-SA 4.0*
  - 8.1.9.5: Using Salt in Fermented Doughs - *CC BY-NC-SA 4.0*
  - 8.1.9.6: Storing Salt - *CC BY-NC-SA 4.0*
  - 8.1.9.7: Introduction to Spices and Other Flavorings - *CC BY-NC-SA 4.0*
  - 8.1.9.8: Seasoning and Flavoring - *CC BY-NC-SA 4.0*
  - 8.1.9.9: Herbs - *CC BY-NC-SA 4.0*
  - 8.1.9.10: Spices - *CC BY-NC-SA 4.0*
  - 8.1.9.11: Flavorings in Baking - *CC BY-NC-SA 4.0*
- 8.2: Aquatic Chemistry - *CC BY-NC-SA 4.0*
  - 8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems - *CC BY-NC-SA 4.0*
  - 8.2.2: 2. Carbonate Equilibria in Natural Waters - *CC BY-NC-SA 4.0*
  - 8.2.3: 3. Redox Equilibria in Natural Waters - *CC BY-NC-SA 4.0*
  - 8.2.4: 4. Solids in Contact With Natural Waters - *CC BY-NC-SA 4.0*
  - 8.2.5: Fundamental Characteristics of Water - *CC BY-NC-SA 4.0*

- 8.2.6: Natural Water - *CC BY-NC-SA 4.0*
- 8.2.7: Water Biology - *CC BY-NC-SA 4.0*
- 8.2.8: Water Chemistry - *CC BY-NC-SA 4.0*
- 8.2.9: Water Physics - *CC BY-NC-SA 4.0*
- 8.2.10: Water Treatment - *CC BY-NC-SA 4.0*

- 8.3: Biochemical Cycles - *CC BY-NC-SA 4.0*
  - 8.3.1: Biochemical Cycles - *CC BY-NC-SA 4.0*
  - 8.3.2: Carbon Cycle - *CC BY-NC-SA 4.0*
  - 8.3.3: Nitrogen Cycle - *CC BY-NC-SA 4.0*
- 8.4: Toxicology MSDT - *CC BY-NC 4.0*
  - 8.4.1: Pathophysiology - *CC BY-NC 4.0*
    - 8.4.1.1: What is Pathophysiology? - *CC BY-NC 4.0*
    - 8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity - *CC BY-NC 4.0*
    - 8.4.1.3: Cellular Response to Toxicant-Induced Injury - *CC BY-NC 4.0*
    - 8.4.1.4: Repair and Adaptation - *CC BY-NC 4.0*
    - 8.4.1.5: Patterns of Toxic Injury - *CC BY-NC 4.0*
    - 8.4.1.6: Final Evaluation - *CC BY-NC 4.0*
  - 8.4.2: Biochemistry and Molecular Genetics - *CC BY-NC 4.0*
    - 8.4.2.1: Introduction to Biomolecules and Cell Components - *CC BY-NC 4.0*
    - 8.4.2.2: Cell Structure and Subcellular Compartments - *CC BY-NC 4.0*
    - 8.4.2.3: DNA and RNA Metabolism - *CC BY-NC 4.0*
    - 8.4.2.4: Epigenetic Mechanisms - *CC BY-NC 4.0*
    - 8.4.2.5: Section 2 Final Evaluation - *CC BY-NC 4.0*
  - 8.4.3: Principles of Genetic Toxicology - *CC BY-NC 4.0*
    - 8.4.3.1: Introduction to Genetic-toxicology Assay - *CC BY-NC 4.0*
    - 8.4.3.2: Different Genetic Damages or Mutations - *CC BY-NC 4.0*
    - 8.4.3.3: Different Genetic-Toxicology Assays - *CC BY-NC 4.0*
    - 8.4.3.4: Different Cytotoxicity Assays - *CC BY-NC 4.0*
    - 8.4.3.5: Epigenetics Assay - *CC BY-NC 4.0*
    - 8.4.3.6: Section 3 Final Evaluation - *CC BY-NC 4.0*
  - 8.4.4: Applied Systems Toxicology - *CC BY-NC 4.0*
    - 8.4.4.1: Systems Toxicology - *CC BY-NC 4.0*
    - 8.4.4.2: Dose Level and Applied Toxicology - *CC BY-NC 4.0*
    - 8.4.4.3: Tools and Technologies in Systems Toxicology - *CC BY-NC 4.0*
    - 8.4.4.4: Other Approaches for Predictive Toxicity Modeling - *CC BY-NC 4.0*
    - 8.4.4.5: Technologies Used In Systems Biology/Toxicology - *CC BY-NC 4.0*
    - 8.4.4.6: Takeaways Summary - *CC BY-NC 4.0*
    - 8.4.4.7: Section 4 Final Evaluation - *CC BY-NC 4.0*
  - 8.4.5: Regulatory Toxicology - *CC BY-NC 4.0*
    - 8.4.5.1: Introduction to Regulatory Toxicology - *CC BY-NC 4.0*
    - 8.4.5.2: Global Regulatory Toxicology - *CC BY-NC 4.0*
    - 8.4.5.3: Topic 3- Regional Regulatory Toxicology - *CC BY-NC 4.0*
    - 8.4.5.4: National Regulatory Toxicology - *CC BY-NC 4.0*
    - 8.4.5.5: State Regulatory Toxicology - *CC BY-NC 4.0*
    - 8.4.5.6: Non-Governmental Regulatory Toxicology - *CC BY-NC 4.0*
    - 8.4.5.7: Section 5 Final Evaluation - *CC BY-NC 4.0*
  - 8.4.6: Principles of Toxicology - *CC BY-NC 4.0*
    - 8.4.6.1: Introduction to Toxicology - *CC BY-NC 4.0*
      - 8.4.6.1.1: What is toxicology? - *CC BY-NC 4.0*
      - 8.4.6.1.2: Basic Terminology - *CC BY-NC 4.0*
    - 8.4.6.2: Dose and Dose Response - *CC BY-NC 4.0*
      - 8.4.6.2.1: Dose and It's Impact on Toxicity - *CC BY-NC 4.0*
      - 8.4.6.2.2: The Dose Response Relationship - *CC BY-NC 4.0*
      - 8.4.6.2.3: Dose Estimates of Toxic Effects - *CC BY-NC 4.0*
      - 8.4.6.2.4: Determining the Safety of a Drug - *CC BY-NC 4.0*
      - 8.4.6.2.5: NOAEL and LOAEL - *CC BY-NC 4.0*
    - 8.4.6.3: Toxic Effects - *CC BY-NC 4.0*
      - 8.4.6.3.1: Types of Toxic Effects - *CC BY-NC 4.0*
      - 8.4.6.3.2: Factors Affecting Toxicity - *CC BY-NC 4.0*
      - 8.4.6.3.3: Systemic Toxic Effects - *CC BY-NC 4.0*
      - 8.4.6.3.4: Organ Specific Toxic Effects - *CC BY-NC 4.0*
    - 8.4.6.4: Interactions - *CC BY-NC 4.0*
      - 8.4.6.4.1: Interactions - *CC BY-NC 4.0*

- 8.4.6.5: Toxicity Testing Methods - *CC BY-NC 4.0*
  - 8.4.6.5.1: Testing and Assessing Toxicity - *CC BY-NC 4.0*
  - 8.4.6.5.2: Clinical Investigations and Other Types of Human Data - *CC BY-NC 4.0*
  - 8.4.6.5.3: Epidemiology Studies - *CC BY-NC 4.0*
- 8.4.6.6: Risk Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.1: Risk Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.2: Hazard Identification - *CC BY-NC 4.0*
  - 8.4.6.6.3: Dose-Response Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.4: Exposure Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.5: Risk Characterization - *CC BY-NC 4.0*
- 8.4.6.7: Exposure Standards and Guidelines - *CC BY-NC 4.0*
  - 8.4.6.7.1: Exposure Standards and Guidelines - *CC BY-NC 4.0*
  - 8.4.6.7.2: Regulation of Consumer Products and Drug Safety - *CC BY-NC 4.0*
  - 8.4.6.7.3: Environmental Exposure Standards/Guidelines - *CC BY-NC 4.0*
  - 8.4.6.7.4: Occupational (Workplace) Exposure Standards/ Guidelines/Approaches - *CC BY-NC 4.0*
- 8.4.6.8: Basic Physiology - *CC BY-NC 4.0*
  - 8.4.6.8.1: Introduction to Basic Physiology - *CC BY-NC 4.0*
  - 8.4.6.8.2: Homeostasis - *CC BY-NC 4.0*
  - 8.4.6.8.3: Organs and Organ Systems - *CC BY-NC 4.0*
  - 8.4.6.8.4: Tissues - *CC BY-NC 4.0*
  - 8.4.6.8.5: Cells - *CC BY-NC 4.0*
  - 8.4.6.8.6: Chemicals - *CC BY-NC 4.0*
- 8.4.6.9: Introduction to Toxicokinetics - *CC BY-NC 4.0*
  - 8.4.6.9.1: What is Toxicokinetics - *CC BY-NC 4.0*
- 8.4.6.10: Absorption - *CC BY-NC 4.0*
  - 8.4.6.10.1: Introduction to Absorption - *CC BY-NC 4.0*
  - 8.4.6.10.2: Gastrointestinal Tract - *CC BY-NC 4.0*
  - 8.4.6.10.3: Respiratory Tract - *CC BY-NC 4.0*
  - 8.4.6.10.4: Dermal Route - *CC BY-NC 4.0*
  - 8.4.6.10.5: Other Routes of Exposure - *CC BY-NC 4.0*
- 8.4.6.11: Distribution - *CC BY-NC 4.0*
  - 8.4.6.11.1: Introduction to Distribution - *CC BY-NC 4.0*
  - 8.4.6.11.2: Influence of Route of Exposure - *CC BY-NC 4.0*
  - 8.4.6.11.3: Disposition Models - *CC BY-NC 4.0*
  - 8.4.6.11.4: Structural Barriers to Distribution - *CC BY-NC 4.0*
  - 8.4.6.11.5: Storage Sites - *CC BY-NC 4.0*
- 8.4.6.12: Biotransformation - *CC BY-NC 4.0*
  - 8.4.6.12.1: Introduction to Biotransformation - *CC BY-NC 4.0*
  - 8.4.6.12.2: Chemical Reactions - *CC BY-NC 4.0*
  - 8.4.6.12.3: Biotransformation Sites - *CC BY-NC 4.0*
  - 8.4.6.12.4: Modifiers of Biotransformation - *CC BY-NC 4.0*
- 8.4.6.13: Excretion - *CC BY-NC 4.0*
  - 8.4.6.13.1: Introduction to Secretion - *CC BY-NC 4.0*
  - 8.4.6.13.2: Urinary Excretion - *CC BY-NC 4.0*
  - 8.4.6.13.3: Fecal Excretion - *CC BY-NC 4.0*
  - 8.4.6.13.4: Exhaled Air - *CC BY-NC 4.0*
  - 8.4.6.13.5: Other Routes - *CC BY-NC 4.0*
- 8.4.6.14: Cellular Toxicology - *CC BY-NC 4.0*
  - 8.4.6.14.1: Adaptation - *CC BY-NC 4.0*
  - 8.4.6.14.2: Cell Damage and Tissue Repair - *CC BY-NC 4.0*
  - 8.4.6.14.3: Cancer - *CC BY-NC 4.0*
  - 8.4.6.14.4: Neurotoxicity - *CC BY-NC 4.0*
- 8.4.6.15: Intuitive Toxicology and Risk Communication - *CC BY-NC 4.0*
  - 8.4.6.15.1: Intuitive Toxicology - *CC BY-NC 4.0*
  - 8.4.6.15.2: Risk Communication - *CC BY-NC 4.0*
- 8.4.6.16: Environmental Toxicology, Environmental Health, and One Health - *CC BY-NC 4.0*
  - 8.4.6.16.1: Environmental Toxicology - *CC BY-NC 4.0*

            - 8.4.6.16.2: Environmental Health - *CC BY-NC 4.0*
            - 8.4.6.16.3: One Health - *CC BY-NC 4.0*
        - 8.4.6.17: Conclusion - *CC BY-NC 4.0*
    - 8.5: Food to energy metabolic pathways - *Public Domain*
        - 8.5.1: Basics of metabolism - *Public Domain*
        - 8.5.2: Digestion of food - *Public Domain*
        - 8.5.3: Oxidation of glucose -the glycolysis - *Public Domain*
        - 8.5.4: Citric acid cycle - *Public Domain*
        - 8.5.5: Oxidative Phosphorylation - *Public Domain*
        - 8.5.6: Oxidation of fatty acids - *Public Domain*
        - 8.5.7: Degradation of amino acids - *Public Domain*
- 9: Quantum Basics - *CC BY-NC-SA 4.0*
    - 9.1: Atoms and Photons- Origin of the Quantum Theory - *CC BY-NC-SA 4.0*
    - 9.2: Waves and Particles - *CC BY-NC-SA 4.0*
    - 9.3: The Free Particles - *CC BY-SA 4.0*
    - 9.4: The Particle in a Box - *CC BY-SA 4.0*
    - 9.5: Quantum Mechanics of Some Simple Systems - *CC BY-NC-SA 4.0*
    - 9.6: Principles of Quantum Mechanics - *CC BY-NC-SA 4.0*
    - 9.7: Harmonic Oscillator - *CC BY-NC-SA 4.0*
    - 9.8: Angular Momentum - *CC BY-NC-SA 4.0*
    - 9.9: Hydrogen Atom - *CC BY-NC-SA 4.0*
    - 9.10: Helium Atom - *CC BY-NC-SA 4.0*
    - 9.11: Atomic Structure and The Periodic Law - *CC BY-NC-SA 4.0*
    - 9.12: The Chemical Bond - *CC BY-NC-SA 4.0*
    - 9.13: Molecular Orbital Theory - *CC BY-NC-SA 4.0*
    - 9.14: Molecular Symmetry - *CC BY-NC-SA 4.0*
    - 9.15: Molecular Spectroscopy - *CC BY-NC-SA 4.0*
    - 9.16: Nuclear Magnetic Resonance - *CC BY-NC-SA 4.0*
- 10: Molecular Spectroscopy - *CC BY-NC-SA 4.0*
    - 10.1: The Electromagnetic Spectrum - *CC BY 4.0*
    - 10.2: Rotations Accompany Vibrational Transitions - *CC BY 4.0*
    - 10.3: Unequal Spacings in Vibration-Rotation Spectra - *CC BY-NC-SA 4.0*
    - 10.4: Unequal Spacings in Pure Rotational Spectra - *CC BY-NC-SA 4.0*
    - 10.5: Vibrational Overtones - *CC BY 4.0*
    - 10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information - *CC BY-SA 4.0*
    - 10.7: Electronic Spectroscopy - *CC BY-NC-SA 4.0*
        - 10.7.1: Spectroscopy of the Particle-in-a-Box Model - *CC BY-NC-SA 4.0*
        - 10.7.2: Cyanine Dyes - *CC BY-NC-SA 4.0*
        - 10.7.3: Fluorescence and Phosphorescence - *CC BY-NC-SA 4.0*
    - 10.8: Rotational Spectra of Polyatomic Molecules - *CC BY-NC-SA 4.0*
    - 10.9: Normal Modes in Polyatomic Molecules - *CC BY-NC-SA 4.0*
    - 10.10: Irreducible Representation of Point Groups - *CC BY-NC-SA 4.0*
    - 10.11: The Selection Rule for the Rigid Rotor - *CC BY-SA 4.0*
    - 10.12: The Harmonic Oscillator Selection Rule - *CC BY-NC-SA 4.0*
    - 10.13: Group Theory Determines Infrared Activity - *CC BY-NC-SA 4.0*
    - 10.E: Molecular Spectroscopy (Exercises) - *CC BY-NC-SA 4.0*
- Back Matter - *Undeclared*
    - Index - *Undeclared*
    - Glossary - *Undeclared*
    - Detailed Licensing - *Undeclared*
    - Detailed Licensing - *Undeclared*

This page titled CHEM 2141: Survey of Physical Chemistry is shared under a license and was authored, remixed, and/or curated by LibreTexts.

DETAILED LICENSING

Overview

Title: CHEM 2141: Survey of Physical Chemistry

Webpages: 347

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 51.3% (178 pages)
- CC BY-NC 4.0: 32.9% (114 pages)
- CC BY 4.0: 7.5% (26 pages)
- Undeclared: 4.9% (17 pages)
- Public Domain: 2.3% (8 pages)
- CC BY-SA 4.0: 1.2% (4 pages)

By Page

- CHEM 2141: Survey of Physical Chemistry - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Introduction - *CC BY 4.0*
    - 1.1: Units - *CC BY 4.0*
    - 1.2: Quantity Calculus - *CC BY 4.0*
    - 1.3: Dimensional Analysis - *CC BY 4.0*
    - 1.4: Chapter 1 Problem - *CC BY 4.0*
  - 2: General Chemistry Review - *Undeclared*
    - 2.1: Gases - *Undeclared*
      - 2.1.1: Relating Pressure, Volume, Amount, and Temperature - The Ideal Gas Law - *CC BY 4.0*
      - 2.1.2: Stoichiometry of Gaseous Substances, Mixtures, and Reactions - *CC BY 4.0*
      - 2.1.3: The Kinetic-Molecular Theory - *CC BY 4.0*
      - 2.1.4: Non-Ideal Gas Behavior - *CC BY 4.0*
    - 2.2: Thermodynamics - *Undeclared*
      - 2.2.1: Energy Basics - *CC BY 4.0*
      - 2.2.2: Enthalpy - *CC BY 4.0*
      - 2.2.3: Spontaneity - *CC BY 4.0*
      - 2.2.4: Entropy - *CC BY 4.0*
      - 2.2.5: The Second and Third Laws of Thermodynamics - *CC BY 4.0*
      - 2.2.6: Free Energy - *CC BY 4.0*
    - 2.3: Kinetics - *Undeclared*
      - 2.3.1: Chemical Reaction Rates - *CC BY 4.0*
      - 2.3.2: Factors Affecting Reaction Rates - *CC BY 4.0*
      - 2.3.3: Rate Laws - *CC BY 4.0*
      - 2.3.4: Integrated Rate Laws - *CC BY 4.0*
      - 2.3.5: Collision Theory - *CC BY 4.0*
      - 2.3.6: Reaction Mechanisms - *CC BY 4.0*
      - 2.3.7: Catalysis - *CC BY 4.0*
    - 2.4: Using logarithms - Log vs. Ln - *CC BY-NC-SA 4.0*
  - 3: The First Law of Thermodynamics - *CC BY-NC-SA 4.0*
    - 3.1: Overview of Classical Thermodynamics - *CC BY-NC-SA 4.0*
    - 3.2: Pressure-Volume Work - *CC BY 4.0*
    - 3.3: Work and Heat are not State Functions - *CC BY-NC-SA 4.0*
    - 3.4: Energy is a State Function - *CC BY-NC-SA 4.0*
    - 3.5: An Adiabatic Process is a Process in which No Energy as Heat is Transferred - *CC BY-NC-SA 4.0*
    - 3.6: The Temperature of a Gas Decreases in a Reversible Adiabatic Expansion - *CC BY-NC-SA 4.0*
    - 3.7: Pressure-Volume Work - *CC BY-NC-SA 4.0*
    - 3.8: Heat Capacity is a Path Function - *CC BY-NC-SA 4.0*
    - 3.9: Relative Enthalpies Can Be Determined from Heat Capacity Data and Heats of Transition - *CC BY-NC-SA 4.0*
    - 3.10: Enthalpy Changes for Chemical Equations are Additive - *CC BY-NC-SA 4.0*
    - 3.11: Heats of Reactions Can Be Calculated from Tabulated Heats of Formation - *CC BY-NC-SA 4.0*
    - 3.12: The Temperature Dependence of ΔH - *CC*

- *BY-NC-SA 4.0*
  - 3.13: Enthalpy is a State Function - *Undeclared*
  - 3.E: The First Law of Thermodynamics (Exercises) - *CC BY-NC-SA 4.0*
- 4: Entropy and The Second and 3rd Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 4.1: Energy Does not Determine Spontaneity - *CC BY-NC-SA 4.0*
  - 4.2: Nonequilibrium Isolated Systems Evolve in a Direction That Increases Their Energy Dispersal - *CC BY-NC-SA 4.0*
  - 4.3: Unlike heat, Entropy is a State Function - *CC BY-NC-SA 4.0*
  - 4.4: The Second Law of Thermodynamics - *CC BY-NC-SA 4.0*
  - 4.5: We Must Always Devise a Reversible Process to Calculate Entropy Changes - *CC BY-NC-SA 4.0*
  - 4.6: Entropy Increases With Increasing Temperature - *CC BY-NC-SA 4.0*
  - 4.7: The 3rd Law of Thermodynamics Puts Entropy on an Absolute Scale - *CC BY-NC-SA 4.0*
  - 4.8: The Entropy of a Phase Transition can be Calculated from the Enthalpy of the Phase Transition - *CC BY-NC-SA 4.0*
  - 4.9: Standard Entropies Can Be Used to Calculate Entropy Changes of Chemical Reactions - *CC BY-NC-SA 4.0*
  - 4.E: Entropy and The Second Law of Thermodynamics (Exercises) - *CC BY-NC-SA 4.0*
- 5: Helmholtz and Gibbs Energies - *CC BY-NC-SA 4.0*
  - 5.1: Free Energy Functions - *CC BY-NC-SA 4.0*
  - 5.2: ΔA, ΔG, and Maximum Work - *CC BY-NC-SA 4.0*
  - 5.3: Pressure Dependence of Gibbs Energy - *CC BY-NC-SA 4.0*
  - 5.4: Temperature Dependence of A and G - *CC BY-NC-SA 4.0*
  - 5.5: When Two Variables Change at Once - *CC BY-NC-SA 4.0*
  - 5.6: Thermodynamic Functions have Natural Variables - *CC BY-NC-SA 4.0*
  - 5.7: The Gibbs-Helmholtz Equation - *CC BY-NC-SA 4.0*
  - 5.E: Gibbs Energy (Exercises) - *CC BY-NC-SA 4.0*
- 6: Chemical Kinetics I - *CC BY-NC-SA 4.0*
  - 6.1: Reaction Rate - *CC BY-NC-SA 4.0*
  - 6.2: Rate Laws - *CC BY-NC-SA 4.0*
  - 6.3: 0th order Rate Law - *CC BY-NC-SA 4.0*
  - 6.4: 1st order rate law - *CC BY-NC-SA 4.0*
  - 6.5: 2nd order Rate Laws - *CC BY-NC-SA 4.0*
  - 6.6: The Method of Initial Rates - *CC BY-NC-SA 4.0*
  - 6.7: The Method of Half-Lives - *CC BY-NC-SA 4.0*
  - 6.8: Temperature Dependence - *CC BY-NC-SA 4.0*
  - 6.E: Chemical Kinetics I (Exercises) - *CC BY-NC-SA 4.0*
  - 6.S: Chemical Kinetics I (Summary) - *CC BY-NC-SA 4.0*
- 7: Chemical Kinetics II - *CC BY-NC-SA 4.0*
  - 7.1: Reaction Mechanisms - *CC BY-NC-SA 4.0*
  - 7.2: Concentration Profiles for Some Simple Mechanisms - *CC BY-NC-SA 4.0*
  - 7.3: The Connection between Reaction Mechanisms and Reaction Rate Laws - *CC BY-NC-SA 4.0*
  - 7.4: The Rate Determining Step Approximation - *CC BY-NC-SA 4.0*
  - 7.5: The Steady-State Approximation - *CC BY-NC-SA 4.0*
  - 7.6: The Equilibrium Approximation - *CC BY-NC-SA 4.0*
  - 7.7: The Lindemann Mechanism - *CC BY-NC-SA 4.0*
  - 7.8: The Michaelis-Menten Mechanism - *CC BY-NC-SA 4.0*
  - 7.9: Catalysis - *CC BY-NC-SA 4.0*
  - 7.10: Theories of Reaction Rates - *CC BY-NC-SA 4.0*
  - 7.E: Chemical Kinetics II (Exercises) - *CC BY-NC-SA 4.0*
  - 7.S: Chemical Kinetics II (Summary) - *CC BY-NC-SA 4.0*
- 8: Optional- Special topics - *Undeclared*
  - 8.1: Chemistry of Cooking (Rodriguez-Velazquez) - *CC BY-NC-SA 4.0*
    - 8.1.1: Thickening and Concentrating Flavors - *CC BY-NC-SA 4.0*
      - 8.1.1.1: Viscosity - *CC BY-NC-SA 4.0*
      - 8.1.1.2: Thickening Agents - *CC BY-NC-SA 4.0*
      - 8.1.1.3: Types of Thickening Agents - *CC BY-NC-SA 4.0*
      - 8.1.1.4: Coagulation - *CC BY-NC-SA 4.0*
      - 8.1.1.5: Gelatinization - *CC BY-NC-SA 4.0*
      - 8.1.1.6: Crystallization - *CC BY-NC-SA 4.0*
      - 8.1.1.7: Non-traditional thickeners - *CC BY-NC-SA 4.0*
      - 8.1.1.8: Sauces - *CC BY-NC-SA 4.0*
      - 8.1.1.9: Low-temperature and sous-vide - *CC BY-NC-SA 4.0*
    - 8.1.2: Flour - *CC BY-NC-SA 4.0*
      - 8.1.2.1: Introduction - Understanding Ingredients - *CC BY-NC-SA 4.0*
      - 8.1.2.2: The History of Wheat Flour - *CC BY-NC-SA 4.0*
      - 8.1.2.3: Milling of Wheat - *CC BY-NC-SA 4.0*

- 8.1.2.4: Flour Streams and Types of Wheat Flour - *CC BY-NC-SA 4.0*
- 8.1.2.5: Flour Terms and Treatments - *CC BY-NC-SA 4.0*
- 8.1.2.6: Flour Additives - *CC BY-NC-SA 4.0*
- 8.1.2.7: Whole Grain and Artisan Milling - *CC BY-NC-SA 4.0*
- 8.1.2.8: Flour in Baking - *CC BY-NC-SA 4.0*
- 8.1.2.9: Rye Flour - *CC BY-NC-SA 4.0*
- 8.1.2.10: Other Grains and Flours - *CC BY-NC-SA 4.0*

- 8.1.3: Fat - *CC BY-NC-SA 4.0*
  - 8.1.3.1: Understanding Fats and Oils - *CC BY-NC-SA 4.0*
  - 8.1.3.2: Sources of Bakery Fats and Oils - *CC BY-NC-SA 4.0*
  - 8.1.3.3: Major Fats and Oils Used in Bakeries - *CC BY-NC-SA 4.0*
  - 8.1.3.4: Functions of Fat in Baking - *CC BY-NC-SA 4.0*
- 8.1.4: Sugar - *CC BY-NC-SA 4.0*
  - 8.1.4.1: Sugar Chemistry (ADD US) - *CC BY-NC-SA 4.0*
  - 8.1.4.2: Sugar Refining - *CC BY-NC-SA 4.0*
  - 8.1.4.3: The Application of Sugar - *CC BY-NC-SA 4.0*
  - 8.1.4.4: Agave - *CC BY-NC-SA 4.0*
  - 8.1.4.5: Honey - *CC BY-NC-SA 4.0*
  - 8.1.4.6: Malt - *CC BY-NC-SA 4.0*
  - 8.1.4.7: Maple Syrup (ADD US) - *CC BY-NC-SA 4.0*
  - 8.1.4.8: Sugar Substitutes (ADD US) - *CC BY-NC-SA 4.0*
  - 8.1.4.9: Glucose/Dextrose - *CC BY-NC-SA 4.0*
- 8.1.5: Leavening Agents - *CC BY-NC-SA 4.0*
  - 8.1.5.1: Introduction to Leavening Agents - *CC BY-NC-SA 4.0*
  - 8.1.5.2: Yeast - *CC BY-NC-SA 4.0*
  - 8.1.5.3: The Functions of Yeast - *CC BY-NC-SA 4.0*
  - 8.1.5.4: Using Yeast in Baking - *CC BY-NC-SA 4.0*
  - 8.1.5.5: Baking Powder - *CC BY-NC-SA 4.0*
  - 8.1.5.6: Sodium Bicarbonate - *CC BY-NC-SA 4.0*
  - 8.1.5.7: Ammonium Bicarbonate - *CC BY-NC-SA 4.0*
  - 8.1.5.8: Water Hardness and pH - *CC BY-NC-SA 4.0*
- 8.1.6: Dairy Products - *CC BY-NC-SA 4.0*
  - 8.1.6.1: Introduction to Dairy Products - *CC BY-NC-SA 4.0*
  - 8.1.6.2: Milk - *CC BY-NC-SA 4.0*
  - 8.1.6.3: Milk Products ADD US - *CC BY-NC-SA 4.0*
  - 8.1.6.4: Milk in bread baking - *CC BY-NC-SA 4.0*
  - 8.1.6.5: Yogurt - *CC BY-NC-SA 4.0*
  - 8.1.6.6: Lactose - *CC BY-NC-SA 4.0*
  - 8.1.6.7: Cheese - *CC BY-NC-SA 4.0*
- 8.1.7: Eggs - *CC BY-NC-SA 4.0*
  - 8.1.7.1: Eggs Grade - *CC BY-NC-SA 4.0*
  - 8.1.7.2: Composition and Nutrition - *CC BY-NC-SA 4.0*
  - 8.1.7.3: Egg Products - *CC BY-NC-SA 4.0*
  - 8.1.7.4: The Function of Eggs - *CC BY-NC-SA 4.0*
  - 8.1.7.5: Storing Eggs - *CC BY-NC-SA 4.0*
- 8.1.8: Chocolate - *CC BY-NC-SA 4.0*
  - 8.1.8.1: From the Cocoa Bean to the Finished Chocolate - *CC BY-NC-SA 4.0*
  - 8.1.8.2: Chocolate Produced for the Baking Industry - *CC BY-NC-SA 4.0*
  - 8.1.8.3: Couverture - *CC BY-NC-SA 4.0*
  - 8.1.8.4: Definitions and Regulations (ADD US) - *CC BY-NC-SA 4.0*
- 8.1.9: Spices - *CC BY-NC-SA 4.0*
  - 8.1.9.1: Elements of Taste - *CC BY-NC-SA 4.0*
  - 8.1.9.2: Introduction to Salt - *CC BY-NC-SA 4.0*
  - 8.1.9.3: Origins of Salt - *CC BY-NC-SA 4.0*
  - 8.1.9.4: Functions of Salt in Baking - *CC BY-NC-SA 4.0*
  - 8.1.9.5: Using Salt in Fermented Doughs - *CC BY-NC-SA 4.0*
  - 8.1.9.6: Storing Salt - *CC BY-NC-SA 4.0*
  - 8.1.9.7: Introduction to Spices and Other Flavorings - *CC BY-NC-SA 4.0*
  - 8.1.9.8: Seasoning and Flavoring - *CC BY-NC-SA 4.0*
  - 8.1.9.9: Herbs - *CC BY-NC-SA 4.0*
  - 8.1.9.10: Spices - *CC BY-NC-SA 4.0*
  - 8.1.9.11: Flavorings in Baking - *CC BY-NC-SA 4.0*
- 8.2: Aquatic Chemistry - *CC BY-NC-SA 4.0*
  - 8.2.1: 1. Acid-Base Chemistry of Natural Aquatic Systems - *CC BY-NC-SA 4.0*
  - 8.2.2: 2. Carbonate Equilibria in Natural Waters - *CC BY-NC-SA 4.0*
  - 8.2.3: 3. Redox Equilibria in Natural Waters - *CC BY-NC-SA 4.0*
  - 8.2.4: 4. Solids in Contact With Natural Waters - *CC BY-NC-SA 4.0*
  - 8.2.5: Fundamental Characteristics of Water - *CC BY-NC-SA 4.0*

- 8.2.6: Natural Water - *CC BY-NC-SA 4.0*
- 8.2.7: Water Biology - *CC BY-NC-SA 4.0*
- 8.2.8: Water Chemistry - *CC BY-NC-SA 4.0*
- 8.2.9: Water Physics - *CC BY-NC-SA 4.0*
- 8.2.10: Water Treatment - *CC BY-NC-SA 4.0*

8.3: Biochemical Cycles - *CC BY-NC-SA 4.0*
- 8.3.1: Biochemical Cycles - *CC BY-NC-SA 4.0*
- 8.3.2: Carbon Cycle - *CC BY-NC-SA 4.0*
- 8.3.3: Nitrogen Cycle - *CC BY-NC-SA 4.0*

8.4: Toxicology MSDT - *CC BY-NC 4.0*
- 8.4.1: Pathophysiology - *CC BY-NC 4.0*
  - 8.4.1.1: What is Pathophysiology? - *CC BY-NC 4.0*
  - 8.4.1.2: Outcomes of Targeted and Non-Targeted Toxicity - *CC BY-NC 4.0*
  - 8.4.1.3: Cellular Response to Toxicant-Induced Injury - *CC BY-NC 4.0*
  - 8.4.1.4: Repair and Adaptation - *CC BY-NC 4.0*
  - 8.4.1.5: Patterns of Toxic Injury - *CC BY-NC 4.0*
  - 8.4.1.6: Final Evaluation - *CC BY-NC 4.0*
- 8.4.2: Biochemistry and Molecular Genetics - *CC BY-NC 4.0*
  - 8.4.2.1: Introduction to Biomolecules and Cell Components - *CC BY-NC 4.0*
  - 8.4.2.2: Cell Structure and Subcellular Compartments - *CC BY-NC 4.0*
  - 8.4.2.3: DNA and RNA Metabolism - *CC BY-NC 4.0*
  - 8.4.2.4: Epigenetic Mechanisms - *CC BY-NC 4.0*
  - 8.4.2.5: Section 2 Final Evaluation - *CC BY-NC 4.0*
- 8.4.3: Principles of Genetic Toxicology - *CC BY-NC 4.0*
  - 8.4.3.1: Introduction to Genetic-toxicology Assay - *CC BY-NC 4.0*
  - 8.4.3.2: Different Genetic Damages or Mutations - *CC BY-NC 4.0*
  - 8.4.3.3: Different Genetic-Toxicology Assays - *CC BY-NC 4.0*
  - 8.4.3.4: Different Cytotoxicity Assays - *CC BY-NC 4.0*
  - 8.4.3.5: Epigenetics Assay - *CC BY-NC 4.0*
  - 8.4.3.6: Section 3 Final Evaluation - *CC BY-NC 4.0*
- 8.4.4: Applied Systems Toxicology - *CC BY-NC 4.0*
  - 8.4.4.1: Systems Toxicology - *CC BY-NC 4.0*
  - 8.4.4.2: Dose Level and Applied Toxicology - *CC BY-NC 4.0*
  - 8.4.4.3: Tools and Technologies in Systems Toxicology - *CC BY-NC 4.0*
  - 8.4.4.4: Other Approaches for Predictive Toxicity Modeling - *CC BY-NC 4.0*
  - 8.4.4.5: Technologies Used In Systems Biology/Toxicology - *CC BY-NC 4.0*
  - 8.4.4.6: Takeaways Summary - *CC BY-NC 4.0*
  - 8.4.4.7: Section 4 Final Evaluation - *CC BY-NC 4.0*
- 8.4.5: Regulatory Toxicology - *CC BY-NC 4.0*
  - 8.4.5.1: Introduction to Regulatory Toxicology - *CC BY-NC 4.0*
  - 8.4.5.2: Global Regulatory Toxicology - *CC BY-NC 4.0*
  - 8.4.5.3: Topic 3- Regional Regulatory Toxicology - *CC BY-NC 4.0*
  - 8.4.5.4: National Regulatory Toxicology - *CC BY-NC 4.0*
  - 8.4.5.5: State Regulatory Toxicology - *CC BY-NC 4.0*
  - 8.4.5.6: Non-Governmental Regulatory Toxicology - *CC BY-NC 4.0*
  - 8.4.5.7: Section 5 Final Evaluation - *CC BY-NC 4.0*
- 8.4.6: Principles of Toxicology - *CC BY-NC 4.0*
  - 8.4.6.1: Introduction to Toxicology - *CC BY-NC 4.0*
    - 8.4.6.1.1: What is toxicology? - *CC BY-NC 4.0*
    - 8.4.6.1.2: Basic Terminology - *CC BY-NC 4.0*
  - 8.4.6.2: Dose and Dose Response - *CC BY-NC 4.0*
    - 8.4.6.2.1: Dose and It's Impact on Toxicity - *CC BY-NC 4.0*
    - 8.4.6.2.2: The Dose Response Relationship - *CC BY-NC 4.0*
    - 8.4.6.2.3: Dose Estimates of Toxic Effects - *CC BY-NC 4.0*
    - 8.4.6.2.4: Determining the Safety of a Drug - *CC BY-NC 4.0*
    - 8.4.6.2.5: NOAEL and LOAEL - *CC BY-NC 4.0*
  - 8.4.6.3: Toxic Effects - *CC BY-NC 4.0*
    - 8.4.6.3.1: Types of Toxic Effects - *CC BY-NC 4.0*
    - 8.4.6.3.2: Factors Affecting Toxicity - *CC BY-NC 4.0*
    - 8.4.6.3.3: Systemic Toxic Effects - *CC BY-NC 4.0*
    - 8.4.6.3.4: Organ Specific Toxic Effects - *CC BY-NC 4.0*
  - 8.4.6.4: Interactions - *CC BY-NC 4.0*
    - 8.4.6.4.1: Interactions - *CC BY-NC 4.0*

- 8.4.6.5: Toxicity Testing Methods - *CC BY-NC 4.0*
  - 8.4.6.5.1: Testing and Assessing Toxicity - *CC BY-NC 4.0*
  - 8.4.6.5.2: Clinical Investigations and Other Types of Human Data - *CC BY-NC 4.0*
  - 8.4.6.5.3: Epidemiology Studies - *CC BY-NC 4.0*
- 8.4.6.6: Risk Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.1: Risk Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.2: Hazard Identification - *CC BY-NC 4.0*
  - 8.4.6.6.3: Dose-Response Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.4: Exposure Assessment - *CC BY-NC 4.0*
  - 8.4.6.6.5: Risk Characterization - *CC BY-NC 4.0*
- 8.4.6.7: Exposure Standards and Guidelines - *CC BY-NC 4.0*
  - 8.4.6.7.1: Exposure Standards and Guidelines - *CC BY-NC 4.0*
  - 8.4.6.7.2: Regulation of Consumer Products and Drug Safety - *CC BY-NC 4.0*
  - 8.4.6.7.3: Environmental Exposure Standards/Guidelines - *CC BY-NC 4.0*
  - 8.4.6.7.4: Occupational (Workplace) Exposure Standards/ Guidelines/Approaches - *CC BY-NC 4.0*
- 8.4.6.8: Basic Physiology - *CC BY-NC 4.0*
  - 8.4.6.8.1: Introduction to Basic Physiology - *CC BY-NC 4.0*
  - 8.4.6.8.2: Homeostasis - *CC BY-NC 4.0*
  - 8.4.6.8.3: Organs and Organ Systems - *CC BY-NC 4.0*
  - 8.4.6.8.4: Tissues - *CC BY-NC 4.0*
  - 8.4.6.8.5: Cells - *CC BY-NC 4.0*
  - 8.4.6.8.6: Chemicals - *CC BY-NC 4.0*
- 8.4.6.9: Introduction to Toxicokinetics - *CC BY-NC 4.0*
  - 8.4.6.9.1: What is Toxicokinetics - *CC BY-NC 4.0*
- 8.4.6.10: Absorption - *CC BY-NC 4.0*
  - 8.4.6.10.1: Introduction to Absorption - *CC BY-NC 4.0*
  - 8.4.6.10.2: Gastrointestinal Tract - *CC BY-NC 4.0*
  - 8.4.6.10.3: Respiratory Tract - *CC BY-NC 4.0*
  - 8.4.6.10.4: Dermal Route - *CC BY-NC 4.0*
  - 8.4.6.10.5: Other Routes of Exposure - *CC BY-NC 4.0*
- 8.4.6.11: Distribution - *CC BY-NC 4.0*
  - 8.4.6.11.1: Introduction to Distribution - *CC BY-NC 4.0*
  - 8.4.6.11.2: Influence of Route of Exposure - *CC BY-NC 4.0*
  - 8.4.6.11.3: Disposition Models - *CC BY-NC 4.0*
  - 8.4.6.11.4: Structural Barriers to Distribution - *CC BY-NC 4.0*
  - 8.4.6.11.5: Storage Sites - *CC BY-NC 4.0*
- 8.4.6.12: Biotransformation - *CC BY-NC 4.0*
  - 8.4.6.12.1: Introduction to Biotransformation - *CC BY-NC 4.0*
  - 8.4.6.12.2: Chemical Reactions - *CC BY-NC 4.0*
  - 8.4.6.12.3: Biotransformation Sites - *CC BY-NC 4.0*
  - 8.4.6.12.4: Modifiers of Biotransformation - *CC BY-NC 4.0*
- 8.4.6.13: Excretion - *CC BY-NC 4.0*
  - 8.4.6.13.1: Introduction to Secretion - *CC BY-NC 4.0*
  - 8.4.6.13.2: Urinary Excretion - *CC BY-NC 4.0*
  - 8.4.6.13.3: Fecal Excretion - *CC BY-NC 4.0*
  - 8.4.6.13.4: Exhaled Air - *CC BY-NC 4.0*
  - 8.4.6.13.5: Other Routes - *CC BY-NC 4.0*
- 8.4.6.14: Cellular Toxicology - *CC BY-NC 4.0*
  - 8.4.6.14.1: Adaptation - *CC BY-NC 4.0*
  - 8.4.6.14.2: Cell Damage and Tissue Repair - *CC BY-NC 4.0*
  - 8.4.6.14.3: Cancer - *CC BY-NC 4.0*
  - 8.4.6.14.4: Neurotoxicity - *CC BY-NC 4.0*
- 8.4.6.15: Intuitive Toxicology and Risk Communication - *CC BY-NC 4.0*
  - 8.4.6.15.1: Intuitive Toxicology - *CC BY-NC 4.0*
  - 8.4.6.15.2: Risk Communication - *CC BY-NC 4.0*
- 8.4.6.16: Environmental Toxicology, Environmental Health, and One Health - *CC BY-NC 4.0*
  - 8.4.6.16.1: Environmental Toxicology - *CC BY-NC 4.0*

        - 8.4.6.16.2: Environmental Health - *CC BY-NC 4.0*
        - 8.4.6.16.3: One Health - *CC BY-NC 4.0*
      - 8.4.6.17: Conclusion - *CC BY-NC 4.0*
  - 8.5: Food to energy metabolic pathways - *Public Domain*
    - 8.5.1: Basics of metabolism - *Public Domain*
    - 8.5.2: Digestion of food - *Public Domain*
    - 8.5.3: Oxidation of glucose -the glycolysis - *Public Domain*
    - 8.5.4: Citric acid cycle - *Public Domain*
    - 8.5.5: Oxidative Phosphorylation - *Public Domain*
    - 8.5.6: Oxidation of fatty acids - *Public Domain*
    - 8.5.7: Degradation of amino acids - *Public Domain*
- 9: Quantum Basics - *CC BY-NC-SA 4.0*
  - 9.1: Atoms and Photons- Origin of the Quantum Theory - *CC BY-NC-SA 4.0*
  - 9.2: Waves and Particles - *CC BY-NC-SA 4.0*
  - 9.3: The Free Particles - *CC BY-SA 4.0*
  - 9.4: The Particle in a Box - *CC BY-SA 4.0*
  - 9.5: Quantum Mechanics of Some Simple Systems - *CC BY-NC-SA 4.0*
  - 9.6: Principles of Quantum Mechanics - *CC BY-NC-SA 4.0*
  - 9.7: Harmonic Oscillator - *CC BY-NC-SA 4.0*
  - 9.8: Angular Momentum - *CC BY-NC-SA 4.0*
  - 9.9: Hydrogen Atom - *CC BY-NC-SA 4.0*
  - 9.10: Helium Atom - *CC BY-NC-SA 4.0*
  - 9.11: Atomic Structure and The Periodic Law - *CC BY-NC-SA 4.0*
  - 9.12: The Chemical Bond - *CC BY-NC-SA 4.0*
  - 9.13: Molecular Orbital Theory - *CC BY-NC-SA 4.0*
  - 9.14: Molecular Symmetry - *CC BY-NC-SA 4.0*
  - 9.15: Molecular Spectroscopy - *CC BY-NC-SA 4.0*
  - 9.16: Nuclear Magnetic Resonance - *CC BY-NC-SA 4.0*
- 10: Molecular Spectroscopy - *CC BY-NC-SA 4.0*
  - 10.1: The Electromagnetic Spectrum - *CC BY 4.0*
  - 10.2: Rotations Accompany Vibrational Transitions - *CC BY 4.0*
  - 10.3: Unequal Spacings in Vibration-Rotation Spectra - *CC BY-NC-SA 4.0*
  - 10.4: Unequal Spacings in Pure Rotational Spectra - *CC BY-NC-SA 4.0*
  - 10.5: Vibrational Overtones - *CC BY 4.0*
  - 10.6: Electronic Spectra Contain Electronic, Vibrational, and Rotational Information - *CC BY-SA 4.0*
  - 10.7: Electronic Spectroscopy - *CC BY-NC-SA 4.0*
    - 10.7.1: Spectroscopy of the Particle-in-a-Box Model - *CC BY-NC-SA 4.0*
    - 10.7.2: Cyanine Dyes - *CC BY-NC-SA 4.0*
    - 10.7.3: Fluorescence and Phosphorescence - *CC BY-NC-SA 4.0*
  - 10.8: Rotational Spectra of Polyatomic Molecules - *CC BY-NC-SA 4.0*
  - 10.9: Normal Modes in Polyatomic Molecules - *CC BY-NC-SA 4.0*
  - 10.10: Irreducible Representation of Point Groups - *CC BY-NC-SA 4.0*
  - 10.11: The Selection Rule for the Rigid Rotor - *CC BY-SA 4.0*
  - 10.12: The Harmonic Oscillator Selection Rule - *CC BY-NC-SA 4.0*
  - 10.13: Group Theory Determines Infrared Activity - *CC BY-NC-SA 4.0*
  - 10.E: Molecular Spectroscopy (Exercises) - *CC BY-NC-SA 4.0*
- Back Matter - *Undeclared*
  - Index - *Undeclared*
  - Glossary - *Undeclared*
  - Detailed Licensing - *Undeclared*
  - Detailed Licensing - *Undeclared*

This page titled CHEM 2141: Survey of Physical Chemistry is shared under a license and was authored, remixed, and/or curated by LibreTexts.