CHEM1102: DRUG DISCOVERY - FROM THE LABORATORY TO THE CLINIC



Joshua Schrier
Fordham University

LibreTexts™

Fordham University

Chem1102: Drug Discovery - From the Laboratory to the Clinic

Joshua Schrier

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 born 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 was 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.

This text was compiled on 01/11/2026

TABLE OF CONTENTS

Licensing

1: Atomic Structure

- 1.1: The Elements
- 1.2: Atomic Theory
- 1.3: The Structure of Atoms
- 1.4: Nuclei of Atoms
- 1.5: Atomic Masses
- 1.6: Arrangements of Electrons
- 1.7: The Periodic Table
- 1.8: Elements, Atoms, and the Periodic Table (Summary)
- 1.9: Elements, Atoms, and the Periodic Table (Exercises)

2: Ionic Bonding and Simple Ionic Compounds

- 2.1: Prelude to Ionic Bonding and Simple Ionic Compounds
- 2.2: Two Types of Bonding
- 2.3: Ions
- 2.4: Formulas for Ionic Compounds
- 2.5: Ionic Bonding and Simple Ionic Compounds (Exercises)
- 2.6: Ionic Bonding and Simple Ionic Compounds (Summary)

3: Covalent Bonding and Simple Molecular Compounds

- 3.1: Prelude to Covalent Bonding and Simple Molecular Compounds
- 3.2: Covalent Bonds
- 3.3: Covalent Compounds - Formulas and Names
- 3.4: Drawing Lewis Structures
- 3.5: Characteristics of Covalent Bonds
- 3.6: Characteristics of Molecules
- 3.7: Organic Chemistry
- 3.8: Covalent Bonding and Simple Molecular Compounds (Exercises)
- 3.9: Covalent Bonding and Simple Molecular Compounds (Summary)

4: Noncovalent bonds

- 4.1: Intermolecular Interactions

5: Organic Molecules

- 5.1: Structures and Names of Alkanes
- 5.2: Branched-Chain Alkanes
- 5.3: Condensed Structural and Line-Angle Formulas
- 5.4: IUPAC Nomenclature
- 5.5: Physical Properties of Alkanes
- 5.6: Alkenes- Structures and Names
- 5.7: Cis-Trans Isomers (Geometric Isomers)
- 5.8: Line Notation (SMILES and InChI)

6: Lipids

- 6.1: Prelude to Lipids
- 6.2: Fatty Acids
- 6.3: Fats and Oils
- 6.4: Membranes and Membrane Lipids
- 6.5: Steroids
- 6.6: Exercises
- 6.7: Lipids (Summary)

7: Conformations and Stereochemistry

- 7.1: Prelude to Conformations and Stereochemistry
- 7.2: Conformations of open-chain organic molecules
- 7.3: Chirality and stereoisomers
- 7.4: Naming chiral centers- the R and S system
- 7.5: Optical Activity
- 7.6: Compounds with multiple chiral centers
- 7.7: Stereochemistry of alkenes
- 7.8: Stereochemistry in biology and medicine

8: Acids and Bases

- 8.1: Prelude to Acids and Bases
- 8.2: Brønsted-Lowry Definition of Acids and Bases
- 8.3: Water - Both an Acid and a Base
- 8.4: The Strengths of Acids and Bases
- 8.5: Buffers
- 8.6: Acids and Bases (Exercises)
- 8.7: Acids and Bases (Summary)

9: Amino Acids, Proteins, and Enzymes

- 9.1: Prelude to Amino Acids, Proteins, and Enzymes
- 9.2: Properties of Amino Acids
- 9.3: Reactions of Amino Acids
- 9.4: Peptides
- 9.5: Proteins
- 9.6: X-ray Protein Crystallography
- 9.7: Enzymes
- 9.8: Enzyme Action
- 9.9: Enzyme Activity
- 9.10: Enzyme Inhibition
- 9.11: Enzyme Cofactors and Vitamins
- 9.12: Amino Acids, Proteins, and Enzymes (Exercises)
- 9.13: Amino Acids, Proteins, and Enzymes (Summary)

10: Equilibrium and Kinetics- Making Drugs Stick

- 10.1: Prelude--A Drug's Life (in your Body)
- 10.2: A Quick Review of Thermodynamics and Kinetics
- 10.3: Catalysis

- 10.4: Pharmacology as an Equilibrium and Kinetics Problem

11: Molecular Genetics

- 11.1: Prelude to Nucleic Acids
- 11.2: Nucleotides
- 11.3: Nucleic Acid Structure
- 11.4: Replication and Expression of Genetic Information
- 11.5: Protein Synthesis and the Genetic Code
- 11.6: Mutations and Genetic Diseases
- 11.7: Viruses
- 11.8: Nucleic Acids (Exercises)
- 11.9: Nucleic Acids (Summary)

12: Chemical and Biological Databases

- 12.1: Public Chemical Databases
- 12.2: Data Organization in PubChem as a Data Aggregator
- 12.3: Database Query Introduction
- 12.4: PubChem Web Interfaces for Text
- 12.5: Text Search in PubChem
- 12.6: Additional Data Retrieval Approaches in PubChem
- 12.7: Searching PubChem Using a Non-Textual Query

13: Computer-Aided Drug Design

- 13.1: Reading
- 13.2: How to Dock Your Own Drug

14: Does it work? Clinical Testing

- 14.1: Understanding Clinical Trial Types (Video)
- 14.2: Statistical Basics
- 14.3: Random Sampling
- 14.4: Clinical Studies
- 14.5: Types of clinical interventions
- 14.6: Evolution of new intervention products and sequence of study phases
- 14.7: Interpreting Clinical Research

15: What is it worth? Estimating Value

- 15.1: Types of economic analyses of health interventions
- 15.2: The Time Value of Money
- 15.3: Calculating the Relationship of Time and Value

16: Appendix

- 16.1: Getting started with Zotero

Index

Glossary

Detailed Licensing

Licensing

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

CHAPTER OVERVIEW

1: Atomic Structure

Just as a language has an alphabet from which words are built, chemistry has an alphabet from which matter is described. However, the chemical alphabet is larger than the one we use for spelling. You may have already figured out that the chemical alphabet consists of the chemical elements. Their role is central to chemistry, for they combine to form the millions and millions of known compounds.

1.1: The Elements
1.2: Atomic Theory
1.3: The Structure of Atoms
1.4: Nuclei of Atoms
1.5: Atomic Masses
1.6: Arrangements of Electrons
1.7: The Periodic Table
1.8: Elements, Atoms, and the Periodic Table (Summary)
1.9: Elements, Atoms, and the Periodic Table (Exercises)

1: Atomic Structure is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

1.1: The Elements

Learning Objectives

- Define a chemical element and give examples of the abundance of different elements.
- Represent a chemical element with a chemical symbol.

An element is a substance that cannot be broken down into simpler chemical substances. There are about 90 naturally occurring elements known on Earth. Using technology, scientists have been able to create nearly 30 additional elements that do not occur in nature. Today, chemistry recognizes 118 elements—some of which were created an atom at a time. Figure **1.1.1** shows some of the chemical elements.



Figure **1.1.1**: Samples of Elements. Gold is a yellowish solid, iron is a silvery solid, while mercury is a silvery liquid at room temperature. © Thinkstock

Abundance

The elements vary widely in abundance. In the universe as a whole, the most common element is hydrogen (about 90% of atoms), followed by helium (most of the remaining 10%). All other elements are present in relatively minuscule amounts, as far as we can detect.

Table **1.1.1**: Elemental Composition of Earth

Earth's Crust		Earth (overall)	
Element	**Percentage**	**Element**	**Percentage**
oxygen	46.1	iron	34.6
silicon	28.2	oxygen	29.5
aluminum	8.23	silicon	15.2
iron	5.53	magnesium	12.7
calcium	4.15	nickel	2.4
sodium	2.36	sulfur	1.9
magnesium	2.33	all others	3.7
potassium	2.09		
titanium	0.565		
hydrogen	0.14		
phosphorus	0.105		
all others	0.174		

Source: D. R. Lide, ed. *CRC Handbook of Chemistry and Physics*, 89th ed. (Boca Raton, FL: CRC Press, 2008–9), 14–17.

On the planet Earth, however, the situation is rather different. Oxygen makes up 46.1% of the mass of Earth's crust (the relatively thin layer of rock forming Earth's surface), mostly in combination with other elements, while silicon makes up 28.2%. Hydrogen, the most abundant element in the universe, makes up only 0.14% of Earth's crust. Table **1.1.1** lists the relative abundances of elements on Earth as a whole and in Earth's crust. Table **1.1.2** lists the relative abundances of elements in the human body. If you compare Table **1.1.1** and Table **1.1.2**, you will find disparities between the percentage of each element in the human body and on Earth. Oxygen has the highest percentage in both cases, but carbon, the element with the second highest percentage in the body, is relatively rare on Earth and does not even appear as a separate entry in Table **1.1.1**; carbon is part of the 0.174% representing "other" elements. How does the human body concentrate so many apparently rare elements?

Table **1.1.2**: Elemental Composition of a Human Body

Element	Percentage by Mass
oxygen	61
carbon	23
hydrogen	10
nitrogen	2.6
calcium	1.4
phosphorus	1.1
sulfur	0.20
potassium	0.20
sodium	0.14
chlorine	0.12
magnesium	0.027
silicon	0.026
iron	0.006
fluorine	0.0037
zinc	0.0033
all others	0.174

Source: D. R. Lide, ed. *CRC Handbook of Chemistry and Physics*, 89th ed. (Boca Raton, FL: CRC Press, 2008–9), 7–24.

The relative amounts of elements in the body have less to do with their abundances on Earth than with their availability in a form we can assimilate. We obtain oxygen from the air we breathe and the water we drink. We also obtain hydrogen from water. On the other hand, although carbon is present in the atmosphere as carbon dioxide, and about 80% of the atmosphere is nitrogen, we obtain those two elements from the food we eat, not the air we breathe.

LOOKING CLOSER: PHOSPHOROUS, THE CHEMICAL BOTTLENECK

There is an element that we need more of in our bodies than is proportionately present in Earth's crust, and *this* element is not easily accessible. Phosphorus makes up 1.1% of the human body but only 0.105% of Earth's crust. We need phosphorus for our bones and teeth, and it is a crucial component of all living cells. Unlike carbon, which can be obtained from carbon dioxide through photosynthesis, there is no phosphorus compound present in our surroundings that can serve as a convenient source. Phosphorus, then, is nature's bottleneck. Its availability limits the amount of life our planet can sustain.

Higher forms of life, such as humans, can obtain phosphorus by selecting a proper diet (plenty of protein); but lower forms of life, such as algae, must absorb it from the environment. When phosphorus-containing detergents were introduced in the 1950s, wastewater from normal household activities greatly increased the amount of phosphorus available to algae and other plant life. Lakes receiving this wastewater experienced sudden increases in growth of algae. When the algae died, concentrations of bacteria that ate the dead algae increased. Because of the large bacterial concentrations, the oxygen content of the water dropped, causing fish to die in large numbers. This process, called *eutrophication*, is considered a negative environmental impact.

Today, many detergents are made without phosphorus so the detrimental effects of eutrophication are minimized. You may even see statements to that effect on detergent boxes. It can be sobering to realize how much impact a single element can have on life—or the ease with which human activity can affect the environment.

Names and Symbols

Each element has a name. Some of these names date from antiquity, while others are quite new. Today, the names for new elements are proposed by their discoverers but must be approved by the International Union of Pure and Applied Chemistry, an international organization that makes recommendations concerning all kinds of chemical terminology.

> *Today, new elements are usually named after famous scientists.*

The names of the elements can be cumbersome to write in full, especially when combined to form the names of compounds. Therefore, each element name is abbreviated as a one- or two-letter chemical symbol. By convention, the first letter of a chemical symbol is a capital letter, while the second letter (if there is one) is a lowercase letter. The first letter of the symbol is usually the first letter of the element's name, while the second letter is some other letter from the name. Some elements have symbols that derive from earlier, mostly Latin names, so the symbols may not contain any letters from the English name. Table **1.1.3** lists the names and symbols of some of the most familiar elements.

Table **1.1.3**: Element Names and Symbols

aluminum	Al	magnesium	Mg
argon	Ar	manganese	Mn
arsenic	As	mercury	Hg*
barium	Ba	neon	Ne
bismuth	Bi	nickel	Ni
boron	B	nitrogen	N
bromine	Br	oxygen	O
calcium	Ca	phosphorus	P
carbon	C	platinum	Pt
chlorine	Cl	potassium	K*
chromium	Cr	silicon	Si
copper	Cu*	silver	Ag*
fluorine	F	sodium	Na*
gold	Au*	strontium	Sr
helium	He	sulfur	S
hydrogen	H	tin	Sn*
iron	Fe	tungsten	W†
iodine	I	uranium	U
lead	Pb*	zinc	Zn
lithium	Li	zirconium	Zr

*The symbol comes from the Latin name of element.

†The symbol for tungsten comes from its German name—*wolfram*.

Element names in languages other than English are often close to their Latin names. For example, gold is *oro* in Spanish and *or* in French (close to the Latin *aurum*), tin is *estaño* in Spanish (compare to *stannum*), lead is *plomo* in Spanish and *plomb* in French (compare to *plumbum*), silver is *argent* in French (compare to *argentum*), and iron is *fer* in French and *hierro* in Spanish (compare to *ferrum*). The closeness is even more apparent in pronunciation than in spelling.

Example 1.1.1

Write the chemical symbol for each element without consulting Table **1.1.3** "Element Names and Symbols".

a. bromine
b. boron
c. carbon
d. calcium
e. gold

Answer a

Br

Answer b

B

Answer c

C

Answer d

Ca

Answer e

Au

Exercise 1.1.1

Write the chemical symbol for each element without consulting Table 1.1.3.

a. manganese
b. magnesium
c. neon
d. nitrogen
e. silver

Answer a

Mn

Answer b

Mg

Answer c

Ne

Answer d

N

Answer e

Ag

Example 1.1.2

What element is represented by each chemical symbol?

a. Na
b. Hg
c. P
d. K
e. I

Answer a

sodium

Answer b

mercury

Answer c

phosphorus

Answer d

potassium

Answer e

iodine

? Exercise 1.1.2

What element is represented by each chemical symbol?

a. Pb
b. Sn
c. U
d. O
e. F

Answer a

lead

Answer b

tin

Answer c

uranium

Answer d

oxygen

Answer e

fluorine

Concept Review Exercises

1. What is an element?
2. Give some examples of how the abundance of elements varies.
3. Why are chemical symbols so useful? What is the source of the letter(s) for a chemical symbol?

Answers

1. An element is the basic chemical building block of matter; it is the simplest chemical substance.
2. Elements vary from being a small percentage to more than 30% of the atoms around us.
3. Chemical symbols are useful to concisely represent the elements present in a substance. The letters usually come from the name of the element.

Key Takeaways

- All matter is composed of elements.
- Chemical elements are represented by a one- or two-letter symbol.

1.1: The Elements is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.1: The Elements** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.2: Atomic Theory

Learning Objectives

- Explain how all matter is composed of atoms.
- Describe the modern atomic theory.
- Recognize which elements exist as diatomic molecules.

Take some aluminum foil. Cut it in half. Now you have two smaller pieces of aluminum foil. Cut one of the pieces in half again. Cut one of those smaller pieces in half again. Continue cutting, making smaller and smaller pieces of aluminum foil.

It should be obvious that the pieces are still aluminum foil; they are just becoming smaller and smaller. But how far can you take this exercise, at least in theory? Can you continue cutting the aluminum foil into halves forever, making smaller and smaller pieces? Or is there some limit, some absolute smallest piece of aluminum foil? (Thought experiments like this—and the conclusions based on them—were debated as far back as the fifth century BC.)

The modern atomic theory, proposed about 1803 by the English chemist John Dalton (Figure **1.2.1**), is a fundamental concept that states that all elements are composed of atoms. Previously, we defined an atom as the smallest part of an element that maintains the identity of that element. Individual atoms are extremely small; even the largest atom has an approximate diameter of only 5.4×10^{-10} m. With that size, it takes over 18 million of these atoms, lined up side by side, to equal the width of your little finger (about 1 cm).

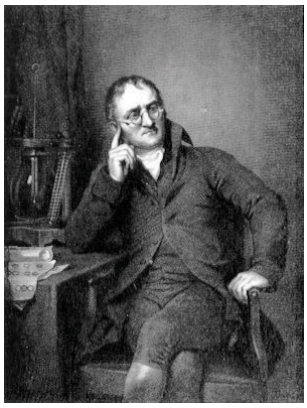


Figure **1.2.1** John Dalton was an English scientist who enunciated the modern atomic theory.

Most elements in their pure form exist as **individual atoms**. For example, a macroscopic chunk of iron (Fe) metal is composed, microscopically, of individual atoms. Some elements, however, exist as groups of atoms called molecules. Several important elements exist as two-atom combinations and are called **diatomic molecules**. In representing a diatomic molecule, we use the symbol of the element and include the ***subscript*** "2" to indicate that two atoms of that element are joined together. The elements that exist as diatomic molecules are hydrogen (H_2), nitrogen (N_2), fluorine (F_2), oxygen (O_2), iodine (I_2), chlorine (Cl_2) and bromine (Br_2). A few other elements can exist as 3-atom molecules like ozone (O_3) and 4-atom molecules like phosphorus (P_4). The most common form of the element sulfur is composed of molecules that consist of eight atoms of sulfur; its molecular formula is S_8 (Figure **1.2.2**).

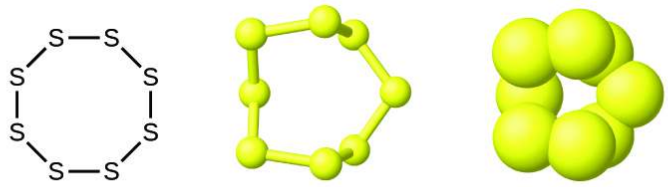


Figure **1.2.2**: A molecule of sulfur is composed of eight sulfur atoms and is therefore written as S_8. It can be represented as (a) a structural formula, (b) a ball-and-stick model, and (c) a space-filling model. Sulfur atoms are represented by yellow spheres.

Figure A shows eight sulfur atoms, symbolized with the letter S, that are bonded to each other to form an octagon. Figure B shows a 3-D, ball-and-stick model of the arrangement of the sulfur atoms. The shape is clearly not octagonal as it is represented in the structural formula. Figure C is a space-filling model that shows each sulfur atom is partially embedded into the sulfur atom it bonds with.

It is important to note that a subscript following a symbol and a number in front of a symbol do not represent the same thing; for example, H_2 and 2H represent distinctly different species. H_2 is a molecular formula; it represents a diatomic molecule of hydrogen, consisting of two atoms of the element that are chemically bonded together. The expression 2H, on the other hand, indicates two separate hydrogen atoms that are **not** combined as a unit. The expression $2H_2$ represents two molecules of diatomic hydrogen (Figure **1.2.3**).

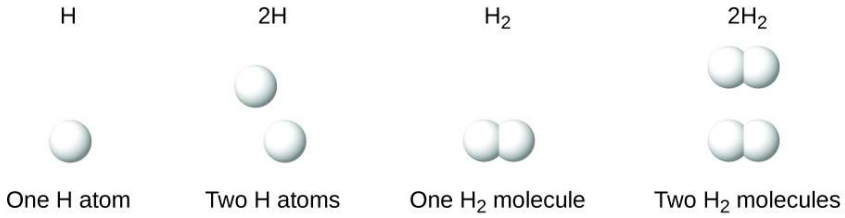


Figure **1.2.3**: The symbols H, 2H, H_2, and $2H_2$ represent very different entities.

This figure shows four diagrams. The diagram for H shows a single, white sphere and is labeled one H atom. The diagram for 2 H shows two white spheres that are not bonded together. It is labeled 2 H atoms. The diagram for H subscript 2 shows two white spheres bonded together. It is labeled one H subscript 2 molecule. The diagram for 2 H subscript 2 shows two sets of bonded, white spheres. It is labeled 2 H subscript 2 molecules.

✔ Example 1.2.1

Write the chemical formula of the following elements:

a. oxygen
b. carbon
c. potassium

Solution

Memorizing the diatomic molecules is worthwhile for our future endeavors. A mnemonic device to help in the memorization of the diatomic elements is as follows: **H**ave **N**o **F**ear **O**f **I**ce **C**old **B**eer. (hydrogen, nitrogen, fluorine, oxygen, iodine, chlorine and bromine).

a. Oxygen is one of the seven diatomic molecular elements. Its chemical formula is O_2.
b. Carbon is monatomic, hence its formula is C.
c. Potassium is monatomic hence its formula is K.

1.2.1 Exercise

Write the chemical formula of the following elements:

a. hydrogen
b. nitrogen
c. neon

Answer a

Hydrogen is one of the seven diatomic molecular elements. Its chemical formula is H_2.

Answer b

Nitrogen is one of the seven diatomic molecular elements. Its chemical formula is N_2.

Answer c

Neon is a monatomic element, hence its formula is Ne.

Looking Closer: Atomic Theory

Dalton's ideas are called the *modern* atomic theory because the concept of atoms is very old. The Greek philosophers Leucippus and Democritus originally introduced atomic concepts in the fifth century BC. (The word *atom* comes from the Greek word *atomos*, which means "indivisible" or "uncuttable.") Dalton had something that the ancient Greek philosophers didn't have, however; he had experimental evidence, such as the formulas of simple chemicals and the behavior of gases. In the 150 years or so before Dalton, natural philosophy had been maturing into modern science, and the scientific method was being used to study nature. So when Dalton announced a modern atomic theory, he was proposing a fundamental theory to describe many previous observations of the natural world; he was not just participating in a philosophical discussion.

Concept Review Exercises

1. What is the modern atomic theory?
2. What are atoms?

Answers

1. The modern atomic theory states that all matter is composed of atoms.
2. Atoms are the smallest parts of an element that maintain the identity of that element.

Key Takeaways

- **Atoms** are the ultimate building blocks of all matter.
- The modern atomic theory establishes the concepts of **atoms** and how they compose matter.
- The **diatomic** elements are hydrogen, nitrogen, fluorine, oxygen, iodine, chlorine and bromine.

1.2: Atomic Theory is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.2: Atomic Theory** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.3: The Structure of Atoms

Learning Objectives

- Describe the three main subatomic particles.
- State how the subatomic particles are arranged in atoms.

There have been several minor but important modifications to Dalton's atomic theory. For one thing, Dalton considered atoms to be indivisible. We know now that atoms not only can be divided but also are composed of three different kinds of particle, *subatomic particles*, with their own properties that are different from the chemical properties of atoms.

Subatomic Particles

The first subatomic particle was identified in 1897 and called the electron. It is an extremely tiny particle, with a mass of about 9.109×10^{-31} kg. Experiments with magnetic fields showed that the electron has a negative electrical charge. By 1920, experimental evidence indicated the existence of a second particle. A proton has the same amount of charge as an electron, but its charge is positive, not negative. Another major difference between a proton and an electron is mass. Although still incredibly small, the mass of a proton is 1.673×10^{-27} kg, which is almost 2,000 times greater than the mass of an electron. Because opposite charges attract each other (while like charges repel each other), protons attract electrons (and vice versa).

Finally, additional experiments pointed to the existence of a third particle. Evidence produced in 1932 established the existence of the neutron, a particle with about the same mass as a proton but with no electrical charge, it is *neutral*. We understand now that all atoms can be broken down into subatomic particles: protons, neutrons, and electrons. Table **1.3.1** lists some of their important characteristics and the symbols used to represent each particle.

Table **1.3.1**: Properties of the Subatomic Particles

Particle	Symbol	Mass (kg)	Relative Mass (proton = 1)	Relative Charge
proton	p^+	1.673×10^{-27}	1	+1
neutron	n^0	1.675×10^{-27}	1	0
electron	e^-	9.109×10^{-31}	0.00055	−1

The Nucleus

How are these subatomic particles arranged? Between 1909 and 1911, Ernest Rutherford, a Cambridge physicist, and his associates Hans Geiger and Ernest Marsden performed experiments that provided strong evidence concerning the internal structure of an atom. They took a very thin metal foil, such as gold or platinum, and aimed a beam of positively charged particles (called alpha particles, which are combinations of two protons and two neutrons) from a radioactive source toward the foil. Surrounding the foil was a detector—either a scintillator (a material that glows when hit by such particles) or some unexposed film (which is exposed where the particles hit it). The detector allowed the scientists to determine the distribution of the alpha particles after they interacted with the foil. Figure **1.3.1** shows a diagram of the experimental setup.

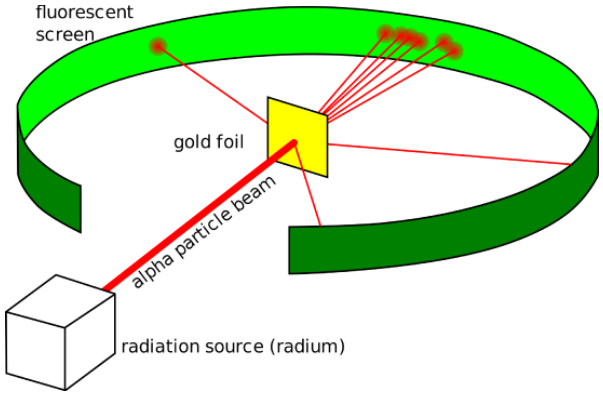


Figure 1.3.1: The Geiger-Marsden Experimental Setup. Experiments using this setup were used to investigate the structure of atoms. In this experiment, most of the particles traveled straight through the foil, but some alpha particles were deflected off to one side. Some were even deflected back toward the source. This was unexpected. Rutherford once said, "It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you." (CC SA-BY 3.0; Kurzon).

Rutherford proposed the following model to explain these experimental results. Protons and neutrons are concentrated in a central region he called the nucleus (plural, *nuclei*) of the atom. Electrons are outside the nucleus and orbit about it because they are attracted to the positive charge in the nucleus. Most of the mass of an atom is in the nucleus, while the orbiting electrons account for an atom's size. As a result, an atom consists largely of empty space. Rutherford called his description the "planetary model" of the atom. Figure 1.3.2 shows how this model explains the experimental results.

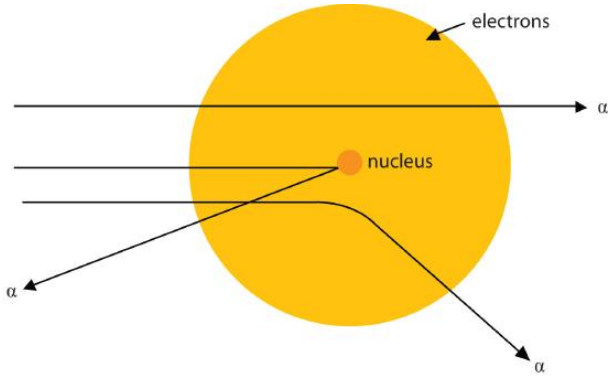


Figure 1.3.2: Rutherford's Metal Foil Experiments. Rutherford explained the results of the metal foil experiments by proposing that most of the mass and the positive charge of an atom are located in its nucleus, while the relatively low-mass electrons orbit about the nucleus. Most alpha particles go straight through the empty space, a few particles are deflected, and fewer still ricochet back toward the source. The nucleus is much smaller proportionately than depicted here.

> *The planetary model of the atom replaced the plum pudding model, which had electrons floating around aimlessly like plums in a "pudding" of positive charge.*

Rutherford's model is essentially the same model that we use today to describe atoms but with one important modification. The planetary model suggests that electrons occupy certain specific, circular orbits about the nucleus. We know now that this model is overly simplistic. A better description is that electrons form fuzzy clouds around nuclei. Figure 1.3.3 shows a more modern version of our understanding of atomic structure.

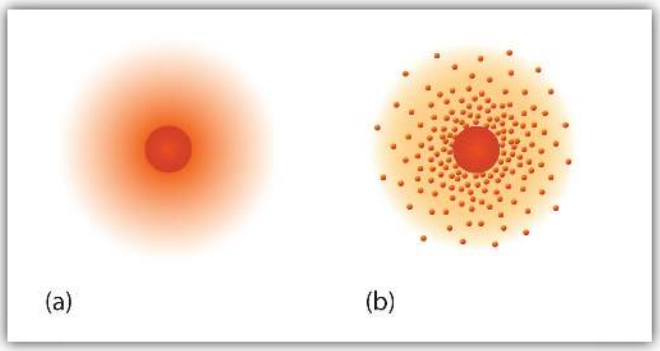


Figure **1.3.3**: A Modern Depiction of Atomic Structure. A more modern understanding of atoms, reflected in these representations of the electron in a hydrogen atom, is that electrons occupy regions of space about the nucleus; they are not in discrete orbits like planets around the sun. (a) The darker the color, the higher the probability that an electron will be at that point. (b) In a two-dimensional cross section of the electron in a hydrogen atom, the more crowded the dots, the higher the probability that an electron will be at that point. In both (a) and (b), the nucleus is in the center of the diagram.

Concept Review Exercises

1. What are the charges and the relative masses of the three subatomic particles?
2. Describe the structure of an atom in terms of its protons, neutrons, and electrons.

Answers

1. proton: +1, large; neutron: 0, large; electron: −1, small
2. Protons and neutrons are located in a central nucleus, while electrons orbit about the nucleus.

Key Takeaways

- Atoms are composed of three main subatomic particles: protons, neutrons, and electrons.
- Protons and neutrons are grouped together in the nucleus of an atom, while electrons orbit about the nucleus.

1.3: The Structure of Atoms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.3: The Structure of Atoms** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.4: Nuclei of Atoms

Learning Objectives

- Define and differentiate between the atomic number and the mass number of an element.
- Explain how isotopes differ from one another.

Now that we know how atoms are generally constructed, what do atoms of any particular element look like? How many protons, neutrons, and electrons are in a specific kind of atom? First, if an atom is electrically neutral overall, then the number of protons equals the number of electrons. Because these particles have the same but opposite charges, equal numbers cancel out, producing a neutral atom.

Atomic Number

In the 1910s, experiments with x-rays led to this useful conclusion: the magnitude of the positive charge in the nucleus of every atom of a particular element is the same. In other words, all atoms of the same element have the same number of protons. Furthermore, different elements have a different number of protons in their nuclei, so the number of protons in the nucleus of an atom is characteristic of a particular element. This discovery was so important to our understanding of atoms that the number of protons in the nucleus of an atom is called the atomic number (Z).

For example, hydrogen has the atomic number 1; all hydrogen atoms have 1 proton in their nuclei. Helium has the atomic number 2; all helium atoms have 2 protons in their nuclei. There is no such thing as a hydrogen atom with 2 protons in its nucleus; a nucleus with 2 protons would be a helium atom. The atomic number *defines* an element. Table **1.4.1** lists some common elements and their atomic numbers. Based on its **atomic number**, you can determine the **number of protons in the nucleus** of an atom. The largest atoms have over 100 protons in their nuclei.

Example 1.4.1

What is the number of protons in the nucleus of each element?

a. aluminum
b. iron
c. carbon

Answer a

According to Table 2.4.1, aluminum has an atomic number of 13. Therefore, every aluminum atom has 13 protons in its nucleus.

Answer b

Iron has an atomic number of 26. Therefore, every iron atom has 26 protons in its nucleus.

Answer c

Carbon has an atomic number of 6. Therefore, every carbon atom has 6 protons in its nucleus.

Exercise 1.4.1

What is the number of protons in the nucleus of each element? Use Table 2.4.1.

a. sodium
b. oxygen
c. chlorine

Answer a

Sodium has 11 protons in its nucleus.

Answer b

Oxygen has 8 protons in its nucleus.

Answer c

Chlorine has 17 protons in its nucleus

How many electrons are in an atom? Previously we said that for an electrically neutral atom, the **number of electrons equals the number of protons**, so the total opposite charges cancel. Thus, the **atomic number** of an element also gives the **number of electrons** in an atom of that element. (Later we will find that some elements may gain or lose electrons from their atoms, so those atoms will no longer be electrically neutral. Thus we will need a way to differentiate the number of electrons for those elements.)

Example 1.4.2

How many electrons are present in the atoms of each element?

a. sulfur
b. tungsten
c. argon

Answer a

The atomic number of sulfur is 16. Therefore, in a neutral atom of sulfur, there are 16 electrons.

Answer b

The atomic number of tungsten is 74. Therefore, in a neutral atom of tungsten, there are 74 electrons.

Answer c

The atomic number of argon is 18. Therefore, in a neutral atom of argon, there are 18 electrons.

Exercise 1.4.2

How many electrons are present in the atoms of each element?

a. magnesium
b. potassium
c. iodine

Answer a

Mg has 12 electrons.

Answer b

K has 19 electrons.

Answer c

I has 53 electrons.

Table **1.4.1**: Some Common Elements and Their Atomic Numbers

Element	Atomic Number	Element	Atomic Nmbers
aluminum (Al)	13	magnesium (Mg)	12
americium (Am)	95	manganese (Mn)	25
argon (Ar)	18	mercury (Hg)	80
barium (Ba)	56	neon (Ne)	10
beryllium (Be)	4	nickel (Ni)	28
bromine (Br)	35	nitrogen (N)	7
calcium (Ca)	20	oxygen (O)	8
carbon (C)	6	phosphorus (P)	15
chlorine (Cl)	17	platinum (Pt)	78
chromium (Cr)	24	potassium (K)	19
cesium (Cs)	55	radon (Rn)	86
fluorine (F)	9	silver (Ag)	47
gallium (Ga)	31	sodium (Na)	11
gold (Au)	79	strontium (Sr)	38
helium (He)	2	sulfur (S)	16
hydrogen (H)	1	titanium (Ti)	22
iron (Fe)	26	tungsten (W)	74
iodine (I)	53	uranium (U)	92
lead (Pb)	82	zinc (Zn)	30
lithium (Li)	3	zirconium (Zr)	40

Isotopes

How many neutrons are in atoms of a particular element? At first it was thought that the number of neutrons in a nucleus was also characteristic of an element. However, it was found that atoms of the same element can have *different* numbers of neutrons. Atoms of the same element (i.e., same atomic number, Z) that have different numbers of neutrons are called isotopes. For example, 99% of the carbon atoms on Earth have 6 neutrons and 6 protons in their nuclei; about 1% of the carbon atoms have 7 neutrons in their nuclei. Naturally occurring carbon on Earth, therefore, is actually a mixture of isotopes, albeit a mixture that is 99% carbon with 6 neutrons in each nucleus.

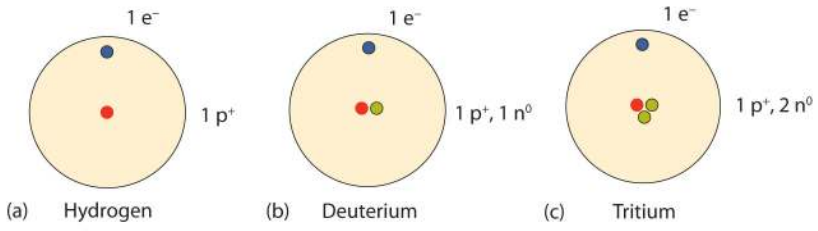


Figure **1.4.1**: Isotopes of Hydrogen. Most hydrogen atoms have only a proton in the nucleus (a). A small amount of hydrogen exists as the isotope deuterium, which has one proton and one neutron in its nucleus (b). A tiny amount of the hydrogen isotope tritium, with one proton and two neutrons in its nucleus, also exists on Earth (c). The nuclei and electrons are proportionately much smaller than depicted here.

An important series of isotopes is found with hydrogen atoms. Most hydrogen atoms have a nucleus with only a single proton. About 1 in 10,000 hydrogen nuclei, however, also has a neutron; this particular isotope is called *deuterium*. An extremely rare hydrogen isotope, *tritium*, has 1 proton and 2 neutrons in its nucleus. Figure **1.4.1** compares the three isotopes of hydrogen.

> The discovery of isotopes required a minor change in Dalton's atomic theory. Dalton thought that all atoms of the same element were exactly the same.

Most elements exist as mixtures of isotopes. In fact, there are currently over 3,500 isotopes known for all the elements. When scientists discuss individual isotopes, they need an efficient way to specify the number of neutrons in any particular nucleus. The *mass number* (A) of an atom is the sum of the numbers of protons and neutrons in the nucleus. Given the mass number for a nucleus (and knowing the atomic number of that particular atom), you can determine the number of neutrons by subtracting the atomic number from the mass number.

A simple way of indicating the mass number of a particular isotope is to list it as a superscript on the left side of an element's symbol. Atomic numbers are often listed as a subscript on the left side of an element's symbol. Thus, we might see

$$\begin{matrix}\text{mass number}\longrightarrow 56\\ \text{atomic number}\longrightarrow 26\end{matrix}\mathrm{Fe} \tag{1.4.1}$$

which indicates a particular isotope of iron. The 26 is the atomic number (which is the same for all iron atoms), while the 56 is the mass number of the isotope. To determine the number of neutrons in this isotope, we subtract 26 from 56: 56 − 26 = 30, so there are 30 neutrons in this atom.

✔ Example 1.4.3

How many protons and neutrons are in each atom?

a. $^{35}_{17}\mathrm{Cl}$
b. $^{127}_{53}\mathrm{I}$

Answer a

In $^{35}_{17}\mathrm{Cl}$ there are 17 protons, and 35 − 17 = 18 neutrons in each nucleus.

Answer b

In $^{127}_{53}\mathrm{I}$ there are 53 protons, and 127 − 53 = 74 neutrons in each nucleus.

? Exercise 1.4.3

How many protons and neutrons are in each atom?

a. $^{197}_{79}\mathrm{Au}$
b. $^{23}_{11}\mathrm{Na}$

Answer a

In $^{197}_{79}\mathrm{Au}$ there are 79 protons, and 197 − 79 = 118 neutrons in each nucleus.

Answer b

In $^{23}_{11}\mathrm{Na}$ there are 11 protons, and 23 − 11 = 12 neutrons in each nucleus.

It is not absolutely necessary to indicate the atomic number as a subscript because each element has its own unique atomic number. Many isotopes are indicated with a superscript only, such as $^{13}\mathrm{C}$ or $^{235}\mathrm{U}$. You may also see isotopes represented in print as, for example, carbon-13 or uranium-235.

Summary

The atom consists of discrete particles that govern its chemical and physical behavior. Each atom of an element contains the same number of protons, which is the **atomic number (Z)**. Neutral atoms have the same number of electrons and protons. Atoms of an element that contain different numbers of neutrons are called **isotopes**. Each isotope of a given element has the same atomic number but a different **mass number (A)**, which is the **sum of the numbers of protons and neutrons**.

Almost all of the mass of an atom is from the total protons and neutrons contained within a tiny (and therefore very dense) nucleus. The majority of the volume of an atom is the surrounding space in which the electrons reside. A representation of a carbon-12 atom is shown below in Figure **1.4.2**.

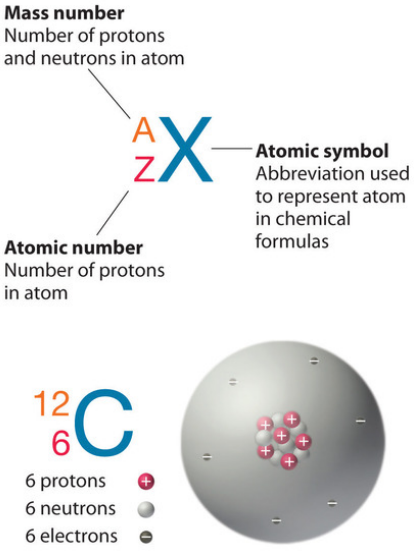


Figure **1.4.2**: Formalism used for identifying specific nuclide (any particular kind of nucleus)

The atomic symbol, X, is the abbreviation used to represent an atom in chemical formulas. The mass number, A, is the number of protons and neutrons in the atom, which is to the left of the X. The atomic number, Z, is the number of protons in the atom, which is to the left of the X and below the A.

Concept Review Exercises

1. Why is the atomic number so important to the identity of an atom?
2. What is the relationship between the number of protons and the number of electrons in an atom?
3. How do isotopes of an element differ from each other?
4. What is the mass number of an element?

Answers

1. The atomic number defines the identity of an element. It describes the number of protons in the nucleus.
2. In an electrically neutral atom, the number of protons equals the number of electrons.
3. Isotopes of an element have the same number of protons but have different numbers of neutrons in their nuclei.
4. The mass number is the sum of the numbers of protons and neutrons in the nucleus of an atom.

Key Takeaways

- Each element is identified by its atomic number. The atomic number provides the element's location on the periodic table
- The isotopes of an element have different masses and are identified by their mass numbers.

1.4: Nuclei of Atoms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.4: Nuclei of Atoms** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.5: Atomic Masses

Learning Objectives

- Define atomic mass and atomic mass unit.
- Calculate atomic mass.

Even though atoms are very tiny pieces of matter, they have mass. Their masses are so small, however, that chemists often use a unit other than grams to express them—the atomic mass unit.

Atomic Mass Unit

The atomic mass unit (abbreviated u, although amu is also used) is defined as 1/12 of the mass of a ^{12}C atom:

$$1\ \text{u} = \frac{1}{12}\ \text{the mass of}\ ^{12}\text{C atom} \tag{1.5.1}$$

It is equal to 1.661×10^{-24} g.

Masses of other atoms are expressed with respect to the atomic mass unit. For example, the mass of an atom of ^{1}H is 1.008 u, the mass of an atom of ^{16}O is 15.995 u, and the mass of an atom of ^{32}S is 31.97 u. Note, however, that these masses are for particular isotopes of each element. Because most elements exist in nature as a mixture of isotopes, any sample of an element will actually be a mixture of atoms having slightly different masses (because neutrons have a significant effect on an atom's mass). How, then, do we describe the mass of a given element? By calculating an average of an element's atomic masses, weighted by the natural abundance of each isotope, we obtain a weighted average mass called the atomic mass (also commonly referred to as the *atomic weight*) of an element.

Atomic Mass is the Weighted Average Mass of Isotopes

As stated above, most elements occur naturally as a mixture of two or more isotopes. Listed below (Table **1.5.1**) are the naturally occurring isotopes of selected elements along with the percent natural abundance of each.

Table **1.5.1**: Atomic Masses and Percent Abundances of Some Natural Isotopes

Element	Isotope (Symbol)	Percent Natural Abundance	Atomic Mass (amu)	Average Atomic Mass (amu)
Hydrogen	$^{1}_{1}H$	99.985	1.0078	1.0079
	$^{2}_{1}H$	0.015	2.0141	
	$^{3}_{1}H$	negligible	3.0160	
Carbon	$^{12}_{6}C$	98.89	12.000	12.011
	$^{13}_{6}C$	1.11	13.003	
	$^{14}_{6}C$	trace	14.003	
Oxygen	$^{16}_{8}O$	99.759	15.995	15.999
	$^{17}_{8}O$	0.037	16.995	
	$^{18}_{8}O$	0.204	17.999	
Chlorine	$^{35}_{17}Cl$	75.77	34.969	35.453
	$^{38}_{17}Cl$	24.23	36.966	
Copper	$^{63}_{29}Cu$	69.17	62.930	63.546
	$^{65}_{29}Cu$	30.83	64.928	

For some elements, one particular isotope is much more abundant than any other isotopes. For example, naturally occurring hydrogen is nearly all hydrogen-1, and naturally occurring oxygen is nearly all oxygen-16. For many other elements, however, more than one isotope may exist in substantial quantities. Chlorine (atomic number 17) is yellowish-green toxic gas. About three quarters of all chlorine atoms have 18 neutrons, giving those atoms a mass number of 35. About one quarter of all chlorine atoms have 20 neutrons, giving those atoms a mass number of 37. Were you to simply calculate the arithmetic average of the precise atomic masses, you would get approximately 36.

$$\frac{34.969\,u + 36.966\,u}{2} = 35.968\,u$$

As you can see, the average atomic mass given in the last column of the table above (35.453) is significantly lower. Why? The reason is that we need to take into account the natural abundance percentages of each isotope in order to calculate what is called the **weighted average**. The atomic mass *of an element is the weighted average of the atomic masses of the naturally occurring isotopes of that element.*

$$0.7577\,(34.969\,u) + 0.2423\,(36.966\,u) = 35.453\,u$$

The weighted average is determined by multiplying the percent of natural abundance by the actual mass of the isotope. This is repeated until there is a term for each isotope. For chlorine, there are only two naturally occurring isotopes so there are only two terms.

Atomic mass = (%1)(mass 1) + (%2)(mass 2) + ⋯

Another example: oxygen exists as a mixture that is 99.759% ^{16}O, 0.037% ^{17}O and 0.204% ^{18}O. The atomic mass of oxygen (use percent natural abundance data from Table 2.5.1) would be calculated as follows:

Atomic mass = (%1)(mass 1) + (%2)(mass 2) + (%3)(mass 3)

$$0.99759\,(15.995u) + 0.00037\,(16.995u) + 0.00204\,(17.999u) = 15.999u$$

To confirm your answer, compare the calculated value to the weighted mass displayed on the periodic table.

Example 1.5.1

Calculate the atomic mass of oxygen. Oxygen exists as a mixture of 3 isotopes. Their respective masses and natural abundance are shown below.

- ^{16}O: 15.995 u (99.759%)
- ^{17}O: 16.995 u (0.037%)
- ^{18}O: 17.999 u (0.204%)

Solution

Multiply the isotope abundance by the actual mass of the isotope, and then sum up the products.

$$0.99759\,(15.995\,u) + 0.00037\,(16.995\,u) + 0.00204\,(17.999\,u) = 15.999\,u$$

1.5.1Exercise

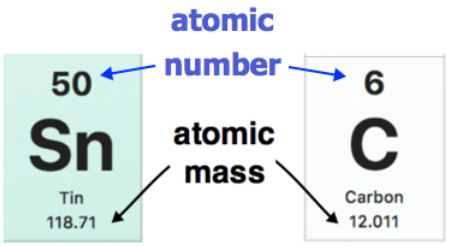
Calculate the atomic mass of copper. Copper exists as a mixture of 2 isotopes. Their respective masses and natural abundance are shown below.

- ^{63}Cu: 62.930 u (69.17%)
- ^{65}Cu: 64.928 u (30.83%)

Answer

63.546 u

The **atomic mass** of each element is found under the element symbol in the **periodic table.** Examples are shown below. The atomic mass of tin (Sn) is 118.71 u while the atomic mass of carbon (C) is 12.011 u. On the other hand, the **atomic number (Z)** of each element is found **above** the atomic symbol.



Atomic mass indicated on entries of the Periodic Table. (public Domain; Pubchem)

The **periodic table** is found in this link:

https://pubchem.ncbi.nlm.nih.gov/periodic-table/png/Periodic_Table_of_Elements_w_Atomic_Mass_PubChem.png

Example 1.5.2: Mass of Carbon

What is the average mass of a carbon atom in grams? The atomic mass is found in the Periodic Table. Please use two decimal places.

Solution

This is a simple one-step conversion, similar to conversions we did in Chapter 1. We use the fact that 1 u = 1.661×10^{-24} g:

$$12.01\ \cancel{u} \times \frac{1.661 \times 10^{-24}\ \text{g}}{1\ \cancel{u}} = 1.995 \times 10^{-23}\ \text{g}$$

This is an extremely small mass, which illustrates just how small individual atoms are.

? Exercise 1.5.2: Mass of Tin

What is the average mass of a tin atom in grams? The average atomic mass is found in the Periodic Table. Please use two decimal places.

Answer

$$118.71 \cancel{u} \times \frac{1.661 \times 10^{-24}\ \text{g}}{1\ \cancel{u}} = 1.972 \times 10^{-22}\ \text{g}$$

Concept Review Exercises

1. Define atomic mass. Why is it considered a weighted average?
2. What is an atomic mass unit?

Answers

1. The atomic mass is an average of an element's atomic masses, weighted by the natural abundance of each isotope of that element. It is a weighted average because different isotopes have different masses.
2. An atomic mass unit is 1/12th of the mass of a ^{12}C atom.

Key Takeaway

- Atoms have a mass that is based largely on the number of protons and neutrons in their nucleus.
- The atomic mass of each element in the Periodic Table is the weighted average of the mass of all its isotopes.

1.5: Atomic Masses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.5: Atomic Masses** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.6: Arrangements of Electrons

Learning Objectives

- Describe how electrons are grouped within atoms.

Although we have discussed the general arrangement of subatomic particles in atoms, we have said little about how electrons occupy the space about the nucleus. Do they move around the nucleus at random, or do they exist in some ordered arrangement?

The modern theory of electron behavior is called quantum mechanics. It makes the following statements about electrons in atoms:

- Electrons in atoms can have only certain specific energies. We say that the energies of the electrons are quantized.
- Electrons are organized according to their energies into sets called **shells** (labeled by the principle quantum number, ***n***). Generally the higher the energy of a shell, the farther it is (on average) from the nucleus. Shells do not have specific, fixed distances from the nucleus, but an electron in a higher-energy shell will spend more time farther from the nucleus than does an electron in a lower-energy shell.
- Shells are further divided into subsets of electrons called **subshells**. The first shell has only one subshell, the second shell has two subshells, the third shell has three subshells, and so on. The subshells of each shell are labeled, in order, with the letters *s*, *p*, *d*, and *f*. Thus, the first shell has only a single *s* subshell (called **1*s***), the second shell has **2*s*** and **2*p*** subshells, the third shell has **3*s***, **3*p***, and **3*d*** and so forth.

Table **1.6.1**: Shells and Subshells

Shell	Number of Subshells	Names of Subshells
1	1	*1s*
2	2	*2s* and *2p*
3	3	*3s*, *3p* and *3d*
4	4	*4s*, *4p*, *4d* and *4f*

- Different subshells hold a different maximum number of electrons. **Any *s* subshell can hold up to 2 electrons; *p*, 6; *d*, 10; and *f*, 14.**

Table **1.6.2**: Number of Electrons

Subshell	Maximum Number of Electrons
s	2
p	6
d	10
f	14

It is the arrangement of electrons into shells and subshells that most concerns us here, so we will focus on that.

We use numbers to indicate which shell an electron is in. As shown in Table **1.6.1**, the first shell, closest to the nucleus and with the lowest-energy electrons, is shell 1. This first shell has only one subshell, which is labeled 1*s* and can hold a maximum of 2 electrons. We combine the shell and subshell labels when referring to the organization of electrons about a nucleus and use a superscript to indicate how many electrons are in a subshell. Thus, because a hydrogen atom has its single electron in the *s* subshell of the first shell, we use $\mathbf{1}s^{\mathbf{1}}$ to describe the electronic structure of hydrogen. This structure is called an electron configuration. Electron configurations are shorthand descriptions of the arrangements of electrons in atoms. The electron configuration of a hydrogen atom is spoken out loud as “one-ess-one.”

Helium atoms have 2 electrons. Both electrons fit into the 1*s* subshell because *s* subshells can hold up to 2 electrons; therefore, the electron configuration for helium atoms is $\mathbf{1}s^2$ (spoken as “one-ess-two”).

The 1*s* subshell cannot hold 3 electrons (because an *s* subshell can hold a maximum of 2 electrons), so the electron configuration for a lithium atom cannot be $1s^3$. Two of the lithium electrons can fit into the 1*s* subshell, but the third electron must go into the second shell. The second shell has two subshells, *s* and *p*, which fill with electrons in that order. The 2*s* subshell holds a maximum of 2 electrons, and the 2*p* subshell holds a maximum of 6 electrons. Because lithium’s final electron goes into the 2*s* subshell, we write the electron configuration of a lithium atom as $\mathbf{1}s^2\mathbf{2}s^1$**.** The shell diagram for a lithium atom is shown below. The shell closest to the nucleus (first shell) has 2 dots representing the 2 electrons in **1*s***, while the outermost shell (**2*s***) has 1 electron.

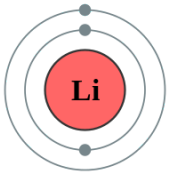


Figure **1.6.1**: Shell diagram of lithium (Li) atom.

The next largest atom, beryllium, has 4 electrons, so its electron configuration is $1s^22s^2$. Now that the 2*s* subshell is filled, electrons in larger atoms start filling the 2*p* subshell. Thus, the electron configurations for the next six atoms are as follows:

- B: $1s^22s^22p^1$
- C: $1s^22s^22p^2$
- N: $1s^22s^22p^3$
- O: $1s^22s^22p^4$
- F: $1s^22s^22p^5$
- Ne: $1s^22s^22p^6$

With neon, the 2*p* subshell is completely filled. Because the second shell has only two subshells, atoms with more electrons now must begin the third shell. The third shell has three subshells, labeled *s*, *p*, and *d*. The *d* subshell can hold a maximum of 10 electrons. The first two subshells of the third shell are filled in order—for example, the electron configuration of aluminum, with 13 electrons, is $1s^22s^22p^63s^23p^1$. However, a curious thing happens after the 3*p* subshell is filled: the 4*s* subshell begins to fill before the 3*d* subshell does. In fact, the exact ordering of subshells becomes more complicated at this point (after argon, with its 18 electrons), so we will not consider the electron configurations of larger atoms. A fourth subshell, the *f* subshell, is needed to complete the electron configurations for all elements. An *f* subshell can hold up to 14 electrons.

Electron filling always starts with **1*s***, the subshell closest to the nucleus. Next is **2*s*, 2*p*, 3*s*, 3*p*, 4*s*, 3*d*, 4*p*, 5s, 4d, 5p, 6s,** etc., shown in the electron shell filling order diagram in Figure **1.6.2**. Follow each arrow in order **from top to bottom**. The subshells you reach along each arrow give the ordering of filling of subshells in larger atoms.

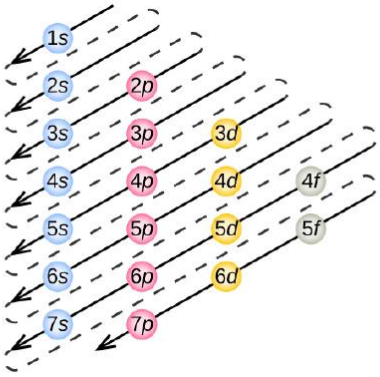


Figure **1.6.2**: The order of electron filling in an atom.

Example 1.6.1: Electronic Configuration of Phosphorus Atoms

Using Figure **1.6.2** as your guide, write the electron configuration of a neutral phosphorus atom. The atomic number of P is 15.

Solution

A neutral phosphorus atom has 15 electrons. Two electrons can go into the 1*s* subshell, 2 can go into the 2*s* subshell, and 6 can go into the 2*p* subshell. That leaves 5 electrons. Of those 5 electrons, 2 can go into the 3*s* subshell, and the remaining 3 electrons can go into the 3*p* subshell. Thus, the electron configuration of neutral phosphorus atoms is $1s^2 2s^2 2p^6 3s^2 3p^3$.

Exercise 1.6.1: Electronic Configuration of Chlorine Atoms

Using Figure **1.6.2** as your guide, write the electron configuration of a neutral chlorine atom. The atomic number of Cl is 17.

Answer

A neutral chlorine atom has 17 electrons. Two electrons can go into the 1*s* subshell, 2 can go into the 2*s* subshell, and 6 can go into the 2*p* subshell. That leaves 7 electrons. Of those 7 electrons, 2 can go into the 3*s* subshell, and the remaining 5 electrons can go into the 3*p* subshell. Thus, the electron configuration of neutral chlorine atoms is $1s^2 2s^2 2p^6 3s^2 3p^5$.

Since the arrangement of the periodic table is based on the electron configurations, Figure **1.6.3** provides an alternative method for determining the electron configuration. The filling order simply **begins at the top left,** with hydrogen (Z=1) and includes each subshell as you proceed in increasing atomic number (*Z*) order.

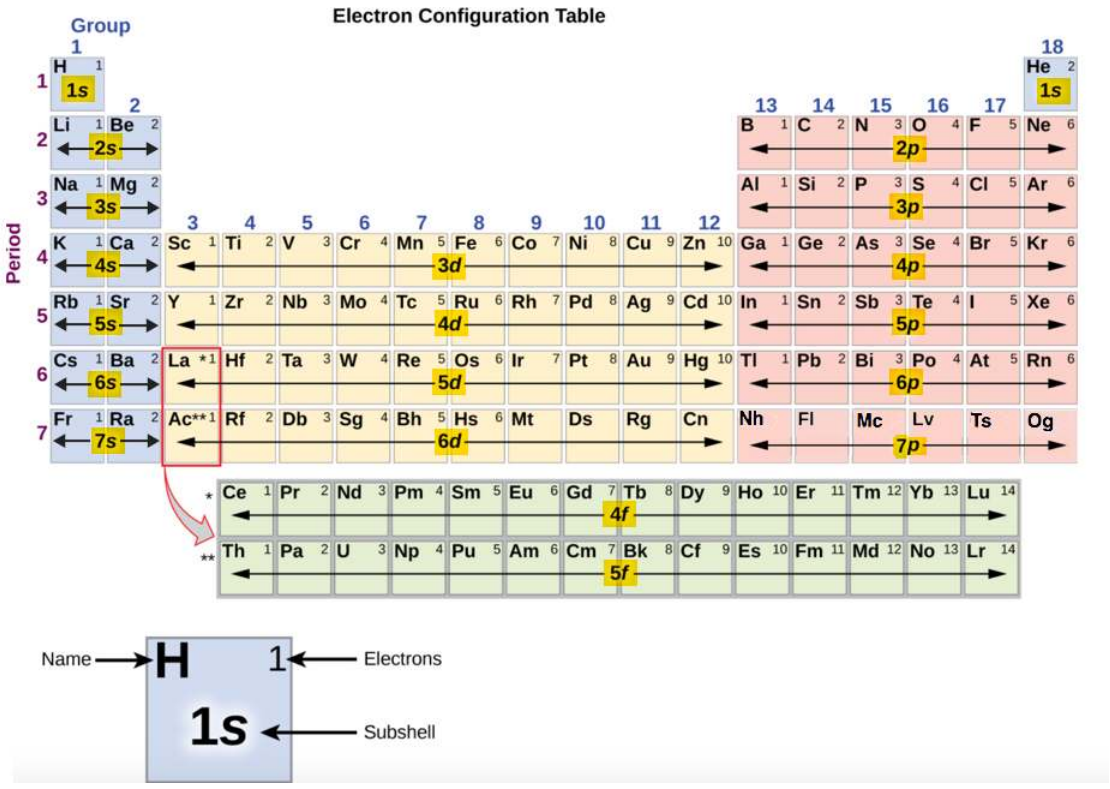


Figure **1.6.3**: This periodic table shows the electron configuration for each subshell. By “building up” from hydrogen, this table can be used to determine the electron configuration for any atom on the periodic table.

For example, the first row (Period 1) contains H and He only, because only two electrons are required to fill the 1s subshell. The second row s-block, contains only two elements, Li and Be, to fill the 2s subshell. This is followed by the second row p-block, containing 6 elements (B through Ne) since six electrons are required to fill the 2p subshell. The third row is similar to the second row elements. Two electrons are needed (Na and Mg) to fill the 3s subshell and six electrons are required (Al through Ar) to complete the 3p subshell. After filling the 3*p* block up to Ar, we see the next subshell will be 4s (K, Ca), followed by the 3*d* subshell, which are filled by ten electrons (Sc through Zn). The 4p subshell is filled next by six electrons (Ga through Kr). As you can see, the periodic table shown in Figure **1.6.3** provides a simple way to remember the order of filling the subshells in determining the electron configuration. The order of filling subshells is the same: **1s, 2*s*, 2*p*, 3*s*, 3*p*, 4*s*, 3*d*, 4*p*, 5s, 4d, 5p, 6s,** etc.

✔ Example 1.6.2: Aluminum

Using Figure **1.6.3** as your guide, write the electron configuration of neutral aluminum atom. The atomic number of Al is 13.

Solution

Aluminum has 13 electrons.

Start at Period 1 of the periodic table, Figure **1.6.3**. Place two electrons in the 1s subshell (**$1s^2$**).

Proceed to Period 2 (left to right direction). Place the next two electrons in the 2s subshell (**$2s^2$**) and the next six electrons in the 2p subshell (**$2p^6$**).

Proceed to Period 3 (left to right direction). Place the next two electrons in the 3s subshell (**$3s^2$**) and the last one electron in the 3p subshell (**$3p^1$**).

The electron configuration of Aluminum is **$1s^2 2s^2 2p^6 3s^2 3p^1$**

1.6.2 Exercise

Using Figure **1.6.3** as your guide, write the electron configuration of the atom that has 20 electrons

Answer

Start at Period 1 of Figure **1.6.3**. Place two electrons in the 1s subshell (**$1s^2$**).

Proceed to Period 2 (left to right direction). Place the next two electrons in the 2s subshell (**$2s^2$**) and the next six electrons in the 2p subshell (**$2p^6$**).

Proceed to Period 3 (left to right direction). Place the next two electrons in the 3s subshell (**$3s^2$**) and the next six electron in the 3p subshell (**$3p^6$**).

Proceed to Period 4. Place the remaining two electrons in the 4s subshell (**$4s^2$).**

The electron configuration is **$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2$**

Valence Electrons

In the study of chemical reactivity, we will find that the electrons in the outermost principal energy level are very important and so they are given a special name. **Valence electrons** are the electrons in the highest occupied principal energy level of an atom.

In the second period elements, the two electrons in the **1*s*** sublevel are called **inner-shell electrons** and are not involved directly in the element's reactivity or in the formation of compounds. Lithium has a single electron in the second principal energy level and so we say that lithium has one valence electron. Beryllium has two valence electrons. How many valence electrons does boron have? You must recognize that the second principal energy level consists of both the **2*s*** and the **2*p*** sublevels and so the answer is three. In fact, the number of valence electrons goes up by one for each step across a period until the last element is reached. Neon, with its configuration ending in **$2s^2 2p^6$**, has eight valence electrons.

The alkali metal sodium (atomic number 11) has one more electron than the neon atom. This electron must go into the lowest-energy subshell available, the 3*s* orbital, giving a $1s^2 2s^2 2p^6 3s^1$ configuration. The electrons occupying the outermost shell orbital(s) (highest value of *n*) are called valence electrons, and those occupying the inner shell orbitals are called **core electrons** (Figure

\PageIndex4). Since the core electron shells correspond to noble gas electron configurations, we can abbreviate electron configurations by writing the noble gas that matches the core electron configuration, along with the valence electrons in a condensed format. For our sodium example, the symbol [Ne] represents core electrons, ($1s^2 2s^2 2p^6$) and our abbreviated or condensed configuration is [Ne]$3s^1$.

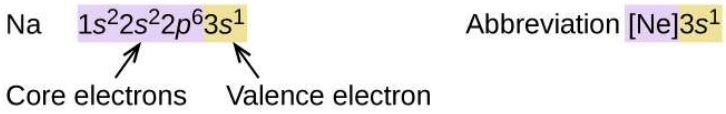


Figure **1.6.4**: A core-abbreviated electron configuration (right) replaces the core electrons with the noble gas symbol whose configuration matches the core electron configuration of the other element.

Similarly, the abbreviated configuration of lithium can be represented as [He]$2s^1$, where [He] represents the configuration of the helium atom, which is identical to that of the filled inner shell of lithium. Writing the configurations in this way emphasizes the similarity of the configurations of lithium and sodium. Both atoms, which are in the alkali metal family, have only one electron in a valence *s* subshell outside a filled set of inner shells.

Li : [He] $2s^1$
Na : [Ne] $3s^1$

A chemical reaction results from electron removal, electron addition, or electron sharing of the valence electrons of the different atoms. The path a specific element will take depends on where the electrons are in the atom and how many there are. Thus, it is convenient to separate electrons into two groups. Valence shell electrons (or, more simply, the *valence electrons*) are the electrons in the highest-numbered shell, or valence shell, while core electrons are the electrons in lower-numbered shells. We can see from the electron configuration of a carbon atom—$1s^2 2s^2 2p^2$—that it has 4 valence electrons ($2s^2 2p^2$) and 2 core electrons ($1s^2$). You will see in the next chapters that the chemical properties of elements are determined by the number of valence electrons.

Example 1.6.3

Examine the electron configuration of neutral phosphorus atoms in Example **1.6.1**, $1s^2 2s^2 2p^6 3s^2 3p^3$ and write the abbreviated notation.

Solution

Phosphorus has electron configuration, $1s^2 2s^2 2p^6 3s^2 3p^3$.

The highest-numbered shell is the third shell ($3s^2 3p^3$): **2 electrons** in the 3*s* subshell and **3 electrons** in the 3*p* subshell. That gives a total of **5 valence electrons**.

The 10 inner shell (core) electrons, $1s^2 2s^2 2p^6$ can be replaced by [Ne] (see Figure **1.6.3**). Abbreviated notation is : [Ne]$3s^2 3p^3$

Exercise 1.6.3

Examine the electron configuration of neutral calcium atom (Exercise **1.6.2**), $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2$, and write the abbreviated notation.

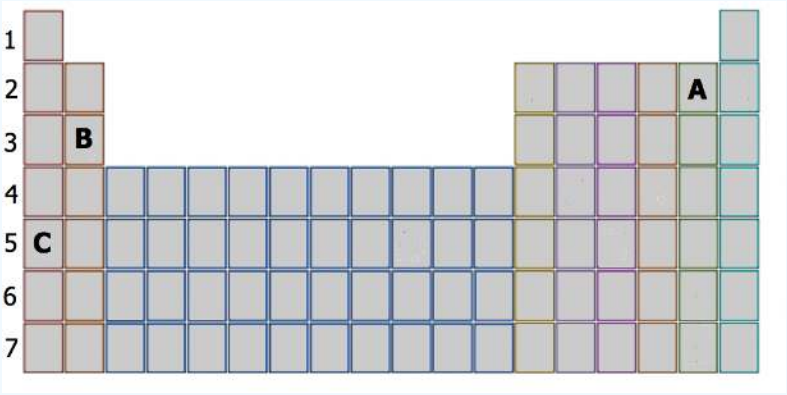
Answer

The highest-numbered shell is the fourth shell $4s^2$, which has **2 electrons** in the *4s* subshell. Hence, Calcium has **2 valence electrons**.

The 18 inner-shell (core) electrons, $1s^2 2s^2 2p^6 3s^2 3p^6$, can be replaced by [Ar], see Figure **1.6.3**. The abbreviated notation is: [Ar]$4s^2$

Example 1.6.4

Based on their respective locations in the periodic table (use Figure **1.6.3**), determine the number of valence electrons and the valence shell configuration of elements A, B and C.



Solution

Element A is located in Period 2, the *5th position* in *2p*-block. Before the electrons are placed in *2p* subshell, the *2s* subshell must be filled first. This means that A has ***two valence electrons*** in 2s (**$2s^2$**) and ***five valence electrons*** in 2p (**$2p^5$**). Answer: **$2s^2 2p^5$.** It has ***2 + 5 = 7*** **valence electrons.**

Element B is located in Period 3, the *2nd position* in *3s*-block. This means that B has ***two valence electrons*** in 3s (**$3s^2$**). Answer: **$3s^2$.**

Element C is located in Period 5, the *1st position* in *5s*-block). This means that there is only **one valence electron** in 5s (**$5s^1$**). Answer: **$5s^1$**.

1.6.4 Exercise

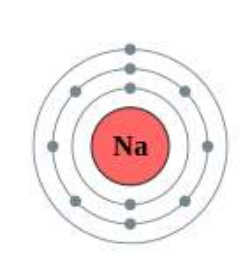
Using the location of Na is the periodic table (Figure **1.6.3**), draw the shell diagram of sodium atom.

Answer

Sodium (Na) is the first element in the 3rd row (Period 3) in the periodic table. This means that the first shell and second shells of Na atom are filled to the maximum number of electrons.

The first shell **(1s)** *is filled with* ***2 electrons***. The second shell (2s and 2p) has a **total of *8 electrons***. And, the third (last) shell has ***1 electron***.

The shell diagram of the Na atom is shown below. The shell nearest the nucleus (first shell) has 2 electrons (2 dots), the second shell has 8 electrons and the last (outermost) shell has 1 electron. (**2.8.1**)



Concept Review Exercises

1. How are electrons organized in atoms?
2. What information does an electron configuration convey?

3. What is the difference between core electrons and valence electrons?

Answers

1. Electrons are organized into shells and subshells around nuclei.
2. The electron configuration states the arrangement of electrons in shells and subshells.
3. Valence electrons are in the highest-numbered shell; all other electrons are core electrons.

Key Takeaway

- Electrons are organized into shells and subshells about the nucleus of an atom.
- The valence electrons determine the reactivity of an atom.

Exercises

1.6: Arrangements of Electrons is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.6: Arrangements of Electrons** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.7: The Periodic Table

Learning Objectives

- Explain how elements are organized into the periodic table.
- Describe how some characteristics of elements relate to their positions on the periodic table.

In the 19th century, many previously unknown elements were discovered, and scientists noted that certain sets of elements had similar chemical properties. For example, chlorine, bromine, and iodine react with other elements (such as sodium) to make similar compounds. Likewise, lithium, sodium, and potassium react with other elements (such as oxygen) to make similar compounds. Why is this so?

In 1864, Julius Lothar Meyer, a German chemist, organized the elements by atomic mass and grouped them according to their chemical properties. Later that decade, Dmitri Mendeleev, a Russian chemist, organized all the known elements according to similar properties. He left gaps in his table for what he thought were undiscovered elements, and he made some bold predictions regarding the properties of those undiscovered elements. When elements were later discovered whose properties closely matched Mendeleev's predictions, his version of the table gained favor in the scientific community. Because certain properties of the elements repeat on a regular basis throughout the table (that is, they are periodic), it became known as the periodic table.

> *Mendeleev had to list some elements out of the order of their atomic masses to group them with other elements that had similar properties.*

The periodic table is one of the cornerstones of chemistry because it organizes all the known elements on the basis of their chemical properties. A modern version is shown in Figure **1.7.1**. Most periodic tables provide additional data (such as atomic mass) in a box that contains each element's symbol. The elements are listed in order of atomic number.

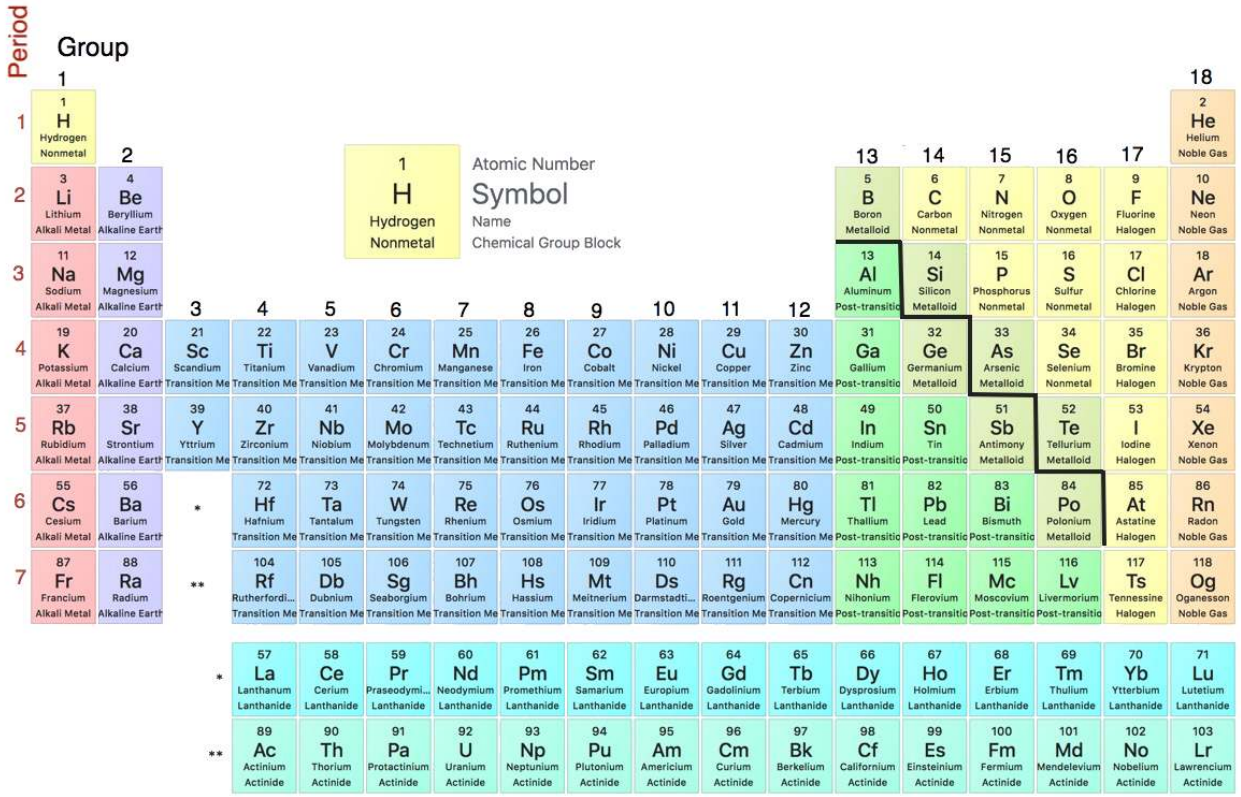


Figure **1.7.1**: Modern Periodic Table. (Public Domain; PubChem modified Leticia Colmenares).

Features of the Periodic Table

Elements that have similar chemical properties are grouped in columns called **groups** (or families). As well as being numbered, some of these groups have names—for example, *alkali metals* (the first column of elements), *alkaline earth metals* (the second column of elements), *halogens* (the next-to-last column of elements), and *noble gases* (the last column of elements).

Each row of elements on the periodic table is called a **period**. Periods have different lengths; the first period has only 2 elements (hydrogen and helium), while the second and third periods have 8 elements each. The fourth and fifth periods have 18 elements each, and later periods are so long that a segment from each is removed and placed beneath the main body of the table.

Metals, Nonmetals and Metalloids

Certain elemental properties become apparent in a survey of the periodic table as a whole. Every element can be classified as either a metal, a nonmetal, or a semimetal, as shown in Figure **1.7.2**. A metal is a substance that is shiny, typically (but not always) silvery in color, and an excellent conductor of electricity and heat. Metals are also malleable (they can be beaten into thin sheets) and ductile (they can be drawn into thin wires). A nonmetal is typically dull and a poor conductor of electricity and heat. Solid nonmetals are also very brittle. As shown in Figure **1.7.2**, **metals** occupy the left three-fourths of the periodic table, while **nonmetals** (except for hydrogen) are clustered in the upper right-hand corner of the periodic table. The elements with properties intermediate between those of metals and nonmetals are called **semimetals** (or **metalloids**). Elements adjacent to the **bold zigzag** line in the right-hand portion of the periodic table have semimetal properties.

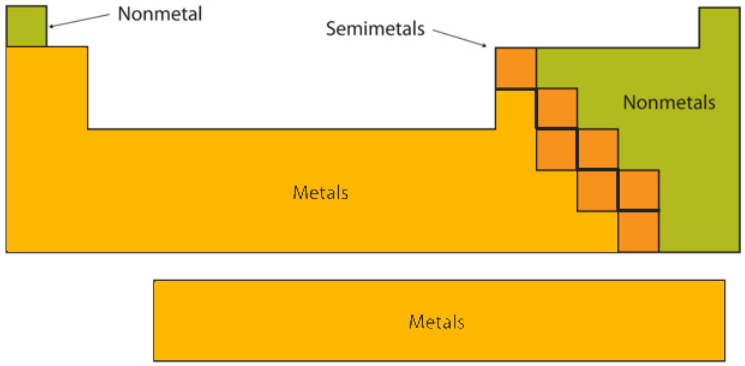


Figure **1.7.2**: Types of Elements. Elements are either metals, nonmetals, or semimetals. Each group is located in a different part of the periodic table.

Example 1.7.1

Based on its position in the periodic table, do you expect selenium (Se) to be a metal, a nonmetal, or a semimetal?

Solution

The atomic number of selenium is 34, which places it in period 4 and group 16. In Figure 2.7.1, selenium lies above and to the right of the diagonal line marking the boundary between metals and nonmetals, so it should be a nonmetal. Note, however, that because selenium is close to the metal-nonmetal dividing line, it would not be surprising if selenium were similar to a semimetal in some of its properties.

1.7.1 Exercise

Based on its location in the periodic table, do you expect indium (In) to be a nonmetal, a metal, or a semimetal?

Answer

metal

Representative, Transition and Inner-transition

Another way to categorize the elements of the periodic table is shown in Figure **1.7.3**. The first two columns on the left and the last six columns on the right are called the **main group** or **representative** elements. The ten-column block between these columns contains the **transition metals**. The two rows beneath the main body of the periodic table contain the inner transition metals. The elements in these two rows are also referred to as, respectively, the **lanthanide metals** and the **actinide metals**.

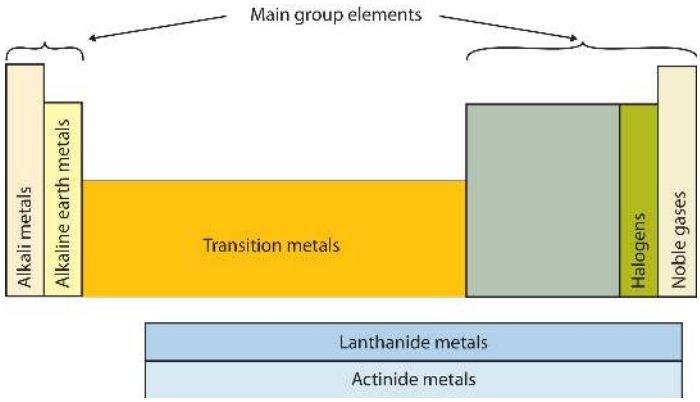


Figure **1.7.3**: Special Names for Sections of the Periodic Table. Some sections of the periodic table have special names. The elements lithium, sodium, potassium, rubidium, cesium, and francium are collectively known as alkali metals.

Alkali metals are the first column. Alkaline earth metals are the second. Halogens are the second to last column. Noble gases are the last column.

To Your Health: Transition Metals in the Body

Most of the elemental composition of the human body consists of main group elements. The first element appearing on the list that is not a main group element is iron, at 0.006 percentage by mass. Because iron has relatively massive atoms, it would appear even lower on a list organized in terms of percent by *atoms* rather than percent by mass.

Iron is a transition metal. Transition metals have interesting chemical properties, partially because some of their electrons are in *d* subshells. The chemistry of iron makes it a key component in the proper functioning of red blood cells.

Red blood cells are cells that transport oxygen from the lungs to cells of the body and then transport carbon dioxide from the cells to the lungs. Without red blood cells, animal respiration as we know it would not exist. The critical part of the red blood cell is a protein called *hemoglobin*. Hemoglobin combines with oxygen and carbon dioxide, transporting these gases from one location to another in the body. Hemoglobin is a relatively large molecule, with a mass of about 65,000 u.

The crucial atom in the hemoglobin protein is iron. Each hemoglobin molecule has four iron atoms, which act as binding sites for oxygen. It is the presence of this particular transition metal in your red blood cells that allows you to use the oxygen you inhale.

Other transition metals have important functions in the body, despite being present in low amounts. Zinc is needed for the body's immune system to function properly, as well as for protein synthesis and tissue and cell growth. Copper is also needed for several proteins to function properly in the body. Manganese is needed for the body to metabolize oxygen properly. Cobalt is a necessary component of vitamin B-12, a vital nutrient. These last three metals are not listed explicitly in Table 2.1.2, so they are present in the body in very small quantities. However, even these small quantities are required for the body to function properly.

As previously noted, the periodic table is arranged so that elements with similar chemical behaviors are in the same group. Chemists often make general statements about the properties of the elements in a group using descriptive names with historical origins. For example, the elements of Group 1 are known as the alkali metals, Group 2 are the alkaline earth metals, Group 17 are the halogens, and Group 18 are the noble gases.

Group 1: The Alkali Metals

The alkali metals are lithium, sodium, potassium, rubidium, cesium, and francium. Hydrogen is unique in that it is generally placed in Group 1, but it is not a metal. The compounds of the alkali metals are common in nature and daily life. One example is table salt (sodium chloride); lithium compounds are used in greases, in batteries, and as drugs to treat patients who exhibit manic-depressive, or bipolar, behavior. Although lithium, rubidium, and cesium are relatively rare in nature, and francium is so unstable and highly radioactive that it exists in only trace amounts, sodium and potassium are the seventh and eighth most abundant elements in Earth's crust, respectively.

Group 2: The Alkaline Earth Metals

The alkaline earth metals are beryllium, magnesium, calcium, strontium, barium, and radium. Beryllium, strontium, and barium are rare, and radium is unstable and highly radioactive. In contrast, calcium and magnesium are the fifth and sixth most abundant elements on Earth, respectively; they are found in huge deposits of limestone and other minerals.

Group 17: The Halogens

The halogens are fluorine, chlorine, bromine, iodine, and astatine. The name halogen is derived from the Greek words for "salt forming," which reflects that all the halogens react readily with metals to form compounds, such as sodium chloride and calcium chloride (used in some areas as road salt).

Compounds that contain the fluoride ion are added to toothpaste and the water supply to prevent dental cavities. Fluorine is also found in Teflon coatings on kitchen utensils. Although chlorofluorocarbon propellants and refrigerants are believed to lead to the depletion of Earth's ozone layer and contain both fluorine and chlorine, the latter is responsible for the adverse effect on the ozone layer. Bromine and iodine are less abundant than chlorine, and astatine is so radioactive that it exists in only negligible amounts in nature.

Group 18: The Noble Gases

The noble gases are helium, neon, argon, krypton, xenon, and radon. Because the noble gases are composed of only single atoms, they are called monatomic. At room temperature and pressure, they are **unreactive gases**. Because of their lack of reactivity, for many years they were called **inert gases** or rare gases. However, the first chemical compounds containing the noble gases were prepared in 1962. Although the noble gases are relatively minor constituents of the atmosphere, natural gas contains substantial amounts of helium. Because of its low reactivity, argon is often used as an unreactive (inert) atmosphere for welding and in light bulbs. The red light emitted by neon in a gas discharge tube is used in neon lights.

To Your Health: Radon

Radon is an invisible, odorless noble gas that is slowly released from the ground, particularly from rocks and soils whose uranium content is high. Because it is a noble gas, radon is not chemically reactive. Unfortunately, it is radioactive, and increased exposure to it has been correlated with an increased lung cancer risk.

Because radon comes from the ground, we cannot avoid it entirely. Moreover, because it is denser than air, radon tends to accumulate in basements, which if improperly ventilated can be hazardous to a building's inhabitants. Fortunately, specialized ventilation minimizes the amount of radon that might collect. Special fan-and-vent systems are available that draw air from below the basement floor, before it can enter the living space, and vent it above the roof of a house.

After smoking, radon is thought to be the second-biggest *preventable* cause of lung cancer in the United States. The American Cancer Society estimates that 10% of all lung cancers are related to radon exposure. There is uncertainty regarding what levels of exposure cause cancer, as well as what the exact causal agent might be (either radon or one of its breakdown products, many of which are also radioactive and, unlike radon, not gases). The US Environmental Protection Agency recommends testing every floor below the third floor for radon levels to guard against long-term health effects.

Why do elements in a given group have similar properties?

The periodic table is organized on the basis of similarities in elemental properties, but what explains these similarities? It turns out that the shape of the periodic table reflects the filling of subshells with electrons, as shown in Figure **1.7.4**. Starting with the first

period and going from left to right, the table reproduces the order of filling of the electron subshells in atoms. Furthermore, elements in the same *group* share the same valence shell electron configuration. For example, all elements in the first column have a single *s* electron in their valence shells, so their electron configurations can be described as ns^1 (where *n* represents the shell number). This last observation is crucial. Chemistry is largely the result of interactions between the valence electrons of different atoms. Thus, atoms that have **the same valence shell electron configuration** will have **similar chemistry.**

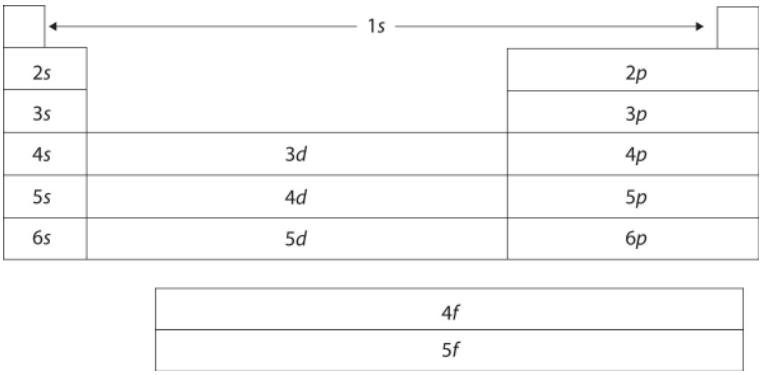


Figure **1.7.4** The Shape of the Periodic Table. The shape of the periodic table reflects the order in which electron shells and subshells fill with electrons.

Example 1.7.1

Using the variable *n* to represent the number of the valence electron shell, write the valence shell electron configuration for each group.

a. the alkaline earth metals
b. the column of elements headed by carbon

Answer a

The alkaline earth metals are in the second column of the periodic table. This column corresponds to the *s* subshell being filled with 2 electrons. Therefore, the valence shell electron configuration is ns^2.

Answer b

The electron configuration of carbon is $1s^22s^22p^2$. Its valence shell electron configuration is $2s^22p^2$. Every element in the same column should have a similar valence shell electron configuration, which we can represent as ns^2np^2.

Exercise 1.7.1

Using the variable *n* to represent the number of the valence electron shell, write the valence shell electron configuration for each group.

a. the halogens
b. the column of elements headed by oxygen

Answer a

The halogens are in the 17th column (or Group 7A) of the periodic table. This column corresponds to the *p* subshell being filled with 5 electrons. Therefore, the valence shell electron configuration is ns^2np^5.

Answer b

The column headed by O is the 16th column (or Group 6A). This column corresponds to the *p* subshell being filled with 4 electrons. Therefore, the valence shell electron configuration is ns^2np^4.

Valence Electrons and Group Number

The number of valence electrons of an element can be determined by the periodic table group (vertical column) in which the element is categorized. With the exception of groups 3–12 (the transition metals), the units digit of the group number identifies how many valence electrons are associated with a neutral atom of an element listed under that particular column.

Table **1.7.1**. The Group number and the number of valence electrons.

Periodic table group	Valence electrons
Group 1 (I) (alkali metals)	1
Group 2 (II) (alkaline earth metals)	2
Groups 3-12 (transition metals)	2*
Group 13 (III) (boron group)	3
Group 14 (IV) (carbon group)	4
Group 15 (V) (pnictogens)	5
Group 16 (VI) (chalcogens)	6
Group 17 (VII) (halogens)	7
Group 18 (VIII or 0) (noble gases)	8**

** The general method for counting valence electrons is generally not useful for transition metals.*

*** Except for helium, which has only two valence electrons.*

Atomic Radius

The periodic table is useful for understanding atomic properties that show periodic trends. One such property is the atomic radius (Figure **1.7.5**). The atomic radius is defined as one-half the distance between the nuclei of identical atoms that are bonded together. The units for atomic radii are picometers, equal to 10^{-12} meters. As an example, the internuclear distance between the two hydrogen atoms in an H_2 molecule is measured to be $74\ pm$. Therefore, the atomic radius of a hydrogen atom is $\frac{74}{2} = 37\ pm$.

As mentioned earlier, the higher the shell number, the farther from the nucleus the electrons in that shell are likely to be. In other words, the size of an atom is generally determined by the number of the valence electron shell. Therefore, as we go down a column on the periodic table, the atomic radius increases. As we go *across* a period on the periodic table, however, electrons are being added to the *same* valence shell; meanwhile, more protons are being added to the nucleus, so the positive charge of the nucleus is increasing. The increasing positive charge attracts the electrons more strongly, pulling them closer to the nucleus. Consequently, as we go across a period, from left to right, the atomic radius decreases. These trends are seen clearly in Figure **1.7.5**

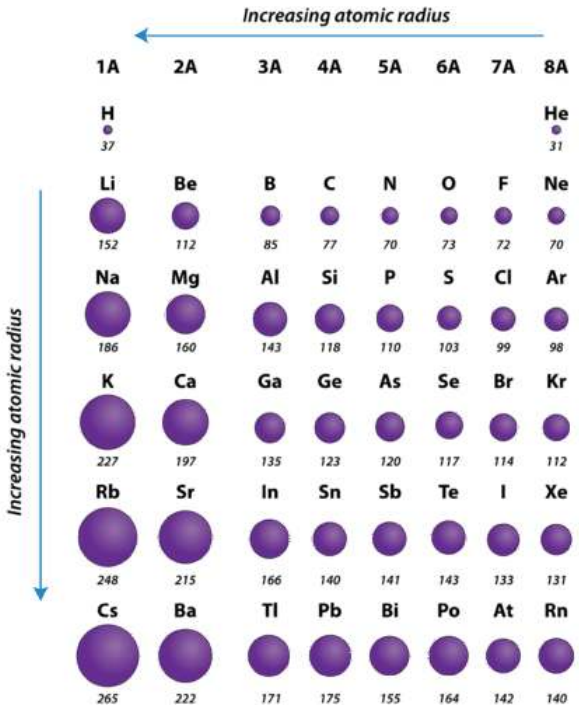


Figure **1.7.5** Trends on the Periodic Table. Atomic radii of the representative elements measured in picometers. The relative sizes of the atoms show several trends with regard to the structure of the periodic table. Atoms become larger going down a group and going from right to left across a period.

✔ Example 1.7.2

Using the periodic table (rather than Figure **1.7.5**), which atom is larger?

a. N or Bi
b. Mg or Cl

Answer a

Bi is below N in Group 5A in the periodic table and has electrons in higher-numbered shells, hence we expect that Bi atoms are larger than N atoms.

Answer b

Both Mg and Cl are in period 3 of the periodic table, but Cl lies farther to the right. Therefore we expect Mg atoms to be larger than Cl atoms.

? Exercise 1.7.2

Using the periodic table (rather than Figure **1.7.5**), which atom is larger?

a. Li or F
b. Na or K

Answer a

Li and F are on the same period, but F lies farther to the right. Therefore, we expect Li to be larger than F atoms.

Answer b

K lies below Na in Group 1A, hence has more electron shells, making it larger than Na.

Career Focus: Clinical Chemist

Clinical chemistry is the area of chemistry concerned with the analysis of body fluids to determine the health status of the human body. Clinical chemists measure a variety of substances, ranging from simple elements such as sodium and potassium to complex molecules such as proteins and enzymes, in blood, urine, and other body fluids. The absence or presence, or abnormally low or high amounts, of a substance can be a sign of some disease or an indication of health. Many clinical chemists use sophisticated equipment and complex chemical reactions in their work, so they not only need to understand basic chemistry, but also be familiar with special instrumentation and how to interpret test results.

Concept Review Exercises

1. How are the elements organized into the periodic table?
2. Looking at the periodic table, where do the following elements appear?
    a. the metals
    b. the nonmetals
    c. the halogens
    d. the transition metals
    e. the noble gases
3. Describe the trends in atomic radii as related to an element's position on the periodic table.

Answers

1. Elements are organized in order of increasing atomic number.
2. a. the left three-quarters of the periodic table (to the left of the zigzag band)
    b. the right quarter of the periodic table (to the right of the zigzag band)
    c. the next-to-last column of the periodic table
    d. the middle section of the periodic table
    e. the last column of the periodic table
3. As you go across the periodic table, atomic radii decrease; as you go down the periodic table, atomic radii increase.

Key Takeaways

- The chemical elements are arranged in a chart called the periodic table.
- Some characteristics of the elements are related to their position on the periodic table.
- The number of valence electrons of an element can be determined by the group (vertical column) number in the Periodic Table. Elements with the same number of valence electrons have similar chemical properties.

1.7: The Periodic Table is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.7: The Periodic Table** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.8: Elements, Atoms, and the Periodic Table (Summary)

To ensure that you understand the material in this chapter, you should review the meanings of the following bold terms and ask yourself how they relate to the topics in the chapter.

An element is a substance that cannot be broken down into simpler chemical substances. Only about 90 naturally occurring elements are known. They have varying abundances on Earth and in the body. Each element has a one- or two-letter **chemical symbol**.

The **modern atomic theory** states that the smallest piece of an element is an **atom**. Individual atoms are extremely small, on the order of 10^{-10} m across. Most elements exist in pure form as individual atoms, but some exist as **diatomic molecules**. Atoms themselves are composed of subatomic particles. The **electron** is a tiny subatomic particle with a negative charge. The **proton** has a positive charge and, while small, is much larger than the electron. The **neutron** is also much larger than an electron but has no electrical charge.

Protons, neutrons, and electrons have a specific arrangement in an atom. The protons and neutrons are found in the center of the atom, grouped together into a **nucleus**. The electrons are found in fuzzy clouds around the nucleus.

Each element has a characteristic number of protons in its nucleus. This number of protons is the **atomic number** of the element. An element may have different numbers of neutrons in the nuclei of its atoms; such atoms are referred to as **isotopes**. Two isotopes of hydrogen are deuterium, with a proton and a neutron in its nucleus, and tritium, with a proton and two neutrons in its nucleus. The sum of the numbers of protons and neutrons in a nucleus is called the **mass number** and is used to distinguish isotopes from each other.

Masses of individual atoms are measured in **atomic mass units**. An atomic mass unit is equal to 1/12th of the mass of a single carbon-12 atom. Because different isotopes of an element have different masses, the **atomic mass** of an element is a weighted average of the masses of all the element's naturally occurring isotopes.

The modern theory of electron behavior is called **quantum mechanics**. According to this theory, electrons in atoms can only have specific, or **quantized**, energies. Electrons are grouped into general regions called **shells**, and within these into more specific regions called **subshells**. There are four types of subshells, and each type can hold up to a maximum number of electrons. The distribution of electrons into shells and subshells is the **electron configuration** of an atom. Chemistry typically occurs because of interactions between the electrons of the outermost shell of different atoms, called the valence shell electrons. Electrons in inner shells are called core electrons.

Elements are grouped together by similar chemical properties into a chart called the **periodic table**. Vertical columns of elements are called **groups** or **families**. Some of the groups of elements have names, like the alkali metals, the alkaline earth metals, the halogens, and the noble gases. A horizontal row of elements is called a **period**. Periods and groups have differing numbers of elements in them. The periodic table separates elements into **metals**, **nonmetals**, and **semimetals**. The periodic table is also separated into **main group elements**, **transition metals**, **lanthanide elements**, and **actinide elements**. The lanthanide and actinide elements are also referred to as **inner transition metal elements**. The shape of the periodic table reflects the sequential filling of shells and subshells in atoms.

The periodic table helps us understand trends in some of the properties of atoms. One such property is the **atomic radius** of atoms. From top to bottom of the periodic table, atoms get bigger because electrons are occupying larger and bigger shells. From left to right across the periodic table, electrons are filling the same shell but are being attracted by an increasing positive charge from the nucleus, and thus the atoms get smaller.

1.8: Elements, Atoms, and the Periodic Table (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **2.S: Elements, Atoms, and the Periodic Table (Summary)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

1.9: Elements, Atoms, and the Periodic Table (Exercises)

02.E: Exercises II

1.9: Elements, Atoms, and the Periodic Table (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

2: Ionic Bonding and Simple Ionic Compounds

There are only 118 known chemical elements but tens of millions of known chemical compounds. Compounds can be very complex combinations of atoms, but many important compounds are fairly simple. Table salt, as we have seen, consists of only two elements: sodium and chlorine. Nevertheless, the compound has properties completely different from either elemental sodium (a chemically reactive metal) or elemental chlorine (a poisonous, green gas). We will see additional examples of such differences in this chapter as we consider how atoms combine to form compounds through **ionic interactions**.

2.1: Prelude to Ionic Bonding and Simple Ionic Compounds

2.2: Two Types of Bonding

2.3: Ions

2.4: Formulas for Ionic Compounds

2.5: Ionic Bonding and Simple Ionic Compounds (Exercises)

2.6: Ionic Bonding and Simple Ionic Compounds (Summary)

Template:HideTOC

2: Ionic Bonding and Simple Ionic Compounds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.1: Prelude to Ionic Bonding and Simple Ionic Compounds

We will see that the word salt has a specific meaning in chemistry, but to most people, this word refers to table salt. This kind of salt is used as a condiment throughout the world, but it was not always so abundant. Two thousand years ago, Roman soldiers received part of their pay as salt, which explains why the words salt and salary come from the same Latin root (salarium). Today, table salt is either mined or obtained from the evaporation of saltwater.

Table salt is sodium chloride (NaCl), which is a simple compound of two elements that are necessary for the human body to function properly. Sodium, for example, is important for nerve conduction and fluid balance. In fact, human blood is about a 0.9% sodium chloride solution, and a solution called normal saline is commonly administered intravenously in hospitals.

Although some salt in our diets is necessary to replenish the sodium and chloride ions that we excrete in urine and sweat, too much is unhealthy, and many people may be ingesting more salt than their bodies need. The RDI of sodium is 2,400 mg—the amount in about 1 teaspoon of salt—but the average intake of sodium in the United States is between 4,000 mg and 5,000 mg, partly because salt is a common additive in many prepared foods. Previously, the high ingestion of salt was thought to be associated with high blood pressure, but current research does not support this link. Even so, some doctors still recommend a low-salt diet (never a “no-salt” diet) for patients with high blood pressure, which may include using a salt substitute. Most salt substitutes use potassium instead of sodium, but some people complain that the potassium imparts a slightly bitter taste.

2.1: Prelude to Ionic Bonding and Simple Ionic Compounds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.0: Prelude to Ionic Bonding and Simple Ionic Compounds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

2.2: Two Types of Bonding

Learning Objectives

- Define the octet rule.
- Describe how ionic bonds are formed.

Atoms can join together by forming a chemical bond, which is a very strong attraction between two atoms. Chemical bonds are formed when electrons in different atoms interact with each other to make an arrangement that is more stable than when the atoms are apart.

What causes atoms to make a chemical bond with other atoms, rather than remaining as individual atoms? A clue comes by considering the noble gas elements, the rightmost column of the periodic table. These elements—helium, neon, argon, krypton, xenon, and radon—do not form compounds very easily, which suggests that they are especially stable as lone atoms. What else do the noble gas elements have in common? Except for helium, they all have eight valence electrons. Chemists have concluded that atoms are especially stable if they have eight electrons in their outermost shell. This useful rule of thumb is called the **octet rule**, and it is a key to understanding why compounds form.

> *Of the noble gases, only krypton, xenon, and radon have been found to make compounds.*

There are two ways for an atom that does not have an octet of valence electrons to obtain an octet in its outer shell. One way is the transfer of electrons between two atoms until all atoms have octets. Because some atoms will lose electrons and some atoms will gain electrons, there is no overall change in the number of electrons, but individual atoms acquire a nonzero electric charge. Those that lose electrons become positively charged, and those that gain electrons become negatively charged. Charged atoms are called ions. Because opposite charges attract (while like charges repel), these oppositely charged ions attract each other, forming **ionic bonds**. The resulting compounds are called **ionic compounds** and are the primary subject of this chapter.

The second way for an atom to obtain an octet of electrons is by sharing electrons with another atom. These shared electrons simultaneously occupy the outermost shell of more than one atom. The bond made by electron sharing is called a **covalent bond**.

Despite our focus on the octet rule, we must remember that for small atoms, such as hydrogen, helium, and lithium, the first shell is, or becomes, the outermost shell and hold only two electrons. Therefore, these atoms satisfy a "**duet rule**" rather than the octet rule.

Example 2.2.1

A sodium atom has one valence electron. Do you think it is more likely for a sodium atom to lose one electron or gain seven electrons to obtain an octet?

Solution

Although either event is possible, a sodium atom is more likely to lose its single valence electron. When that happens, it becomes an ion with a net positive charge. This can be illustrated as follows:

Table explaining the solution to the example.

Sodium atom		Sodium ion	
11 protons	11+	11 protons	11+
11 electrons	11−	10 electrons	10−
	0 overall charge		+1 overall charge

? Exercise 2.2.1

A fluorine atom has seven valence electrons. Do you think it is more likely for a fluorine atom to lose seven electrons or gain one electron to obtain an octet? Write the formula of the resulting ion.

Answer

The process that involves less number of electrons is more favorable. Fluorine would gain one electron. The formula of the resulting ion is F^-.

Key Takeaways

- Atoms have a tendency to have eight electrons in their valence shell.
- The attraction of oppositely charged ions is what makes ionic bonds.

Exercises

1. What is the octet rule?

2. How are ionic bonds formed?

3. Why is an ionic compound unlikely to consist of two positively charged ions?

4. Why is an ionic compound unlikely to consist of two negatively charged ions?

5. A calcium atom has two valence electrons. Do you think it will lose two electrons or gain six electrons to obtain an octet in its outermost electron shell? Write the formula of the resulting ion.

6. An aluminum atom has three valence electrons. Do you think it will lose three electrons or gain five electrons to obtain an octet in its outermost electron shell? Write the formula of the resulting ion.

7. A selenium atom has six valence electrons. Do you think it will lose six electrons or gain two electrons to obtain an octet in its outermost electron shell? Write the formula of the resulting ion.

8. An iodine atom has seven valence electrons. Do you think it will lose seven electrons or gain one electron to obtain an octet in its outermost electron shell? Write the formula of the resulting ion.

Answers

1. The octet rule is the concept that atoms tend to have eight electrons in their valence electron shell.

2. Ionic bonds are formed by the attraction between oppositely charged ions.

3. Positive charges repel each other, so an ionic compound is not likely between two positively charged ions.

4. Negative charges repel each other also. 5. Ca atom is more likely to lose two electrons. It will become Ca^{2+} ion. 6. An Al atom is more likely to lose three electrons. It will become Al^{3+} ion.

7. Selenium is more likely to gain two electrons. It will become Se^{2-} ion.

8. Iodine is more likely to gain one electron. It will become I^- ion.

2.2: Two Types of Bonding is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.1: Two Types of Bonding** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

2.3: Ions

Learning Objectives

- Define the two types of ions.
- Use Lewis diagrams to illustrate ion formation.

Most atoms do not have eight electrons in their valence electron shell. Some atoms have only a few electrons in their outer shell, while some atoms lack only one or two electrons to have an octet. In cases where an atom has three or fewer valence electrons, the atom may lose those valence electrons quite easily until what remains is a lower shell that contains an octet. Atoms that lose electrons acquire a positive charge as a result because they are left with fewer negatively charged electrons to balance the positive charges of the protons in the nucleus. Positively charged ions are called cations. Most metals become cations when they make ionic compounds.

Some atoms have nearly eight electrons in their valence shell and can gain additional valence electrons until they have an octet. When these atoms gain electrons, they acquire a negative charge because they now possess more electrons than protons. Negatively charged ions are called anions. Most nonmetals become anions when they make ionic compounds.

The names for positive and negative ions are pronounced CAT-eye-ons (cations) and ANN-eye-ons (anions), respectively.

Electron Transfer

We can use electron configurations to illustrate the electron transfer process between sodium atoms and chlorine atoms.

$$\text{Na: } 1s^2 2s^2 2p^6 3s^1$$

As demonstrated here, a sodium atom (Na) has one valence electron in the third principal energy level. It is likely to achieve an octet in its outermost shell by losing its one valence electron. The cation produced in this way, Na^+, is called the sodium ion to distinguish it from the element. The sodium ion, Na^+, has the electron configuration with an octet of electrons from the second principal energy level. It is now the same as that of the noble gas neon. The term **isoelectronic** refers to an atom and an ion of a different atom (or two different ions) that have the same electron configuration. The sodium ion is isoelectronic with the neon atom. The equation below illustrates this process.

$$\begin{array}{ccc} \mathbf{Na} & \rightarrow & \mathbf{Na^+ + e^-} \\ \mathbf{1s^2\ 2s^2\ 2p^6\ 3s^1} & & \mathbf{1s^2\ 2s^2\ 2p^6\ (octet)} \end{array}$$

Figure **2.3.1** is a graphical depiction of this process.

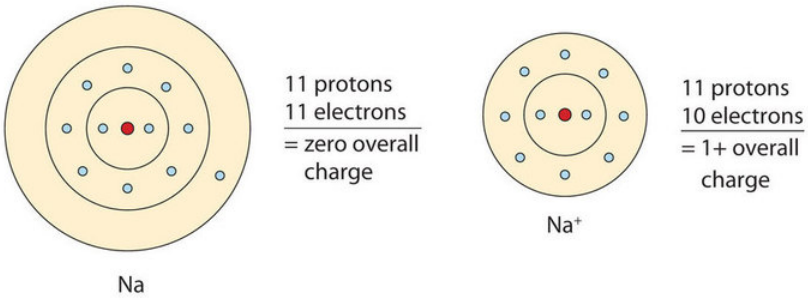


Figure **2.3.1**: The Formation of a Sodium Ion. On the left, a sodium atom has 11 electrons. On the right, the sodium ion only has 10 electrons and a 1+ charge.

Now, let's consider chlorine atom, Cl: $1s^2 2s^2 2p^6 3s^2 3p^5$

Only one more electron is needed to achieve an octet in chlorine's valence shell. When a chlorine atom gains an electron, its outermost principal energy level achieves an octet. In this case, the ion has the same outermost shell as the original atom, but now that shell has eight electrons in it. Once again, the octet rule has been satisfied. The resulting anion, Cl^-, is called the chloride ion;

note the slight change in the suffix (*-ide* instead of *-ine*) to create the name of this anion. This process is illustrated below. (In table salt, this electron comes from the sodium atom.)

$$\underset{1s^2\,2s^2\,2p^6\,3s^2\,3p^5}{Cl + e^-} \rightarrow \underset{1s^2\,2s^2\,2p^6\,3s^2\,3p^6(\text{octet})}{Cl^-}$$

Figure **2.3.2** is a graphical depiction of this process.

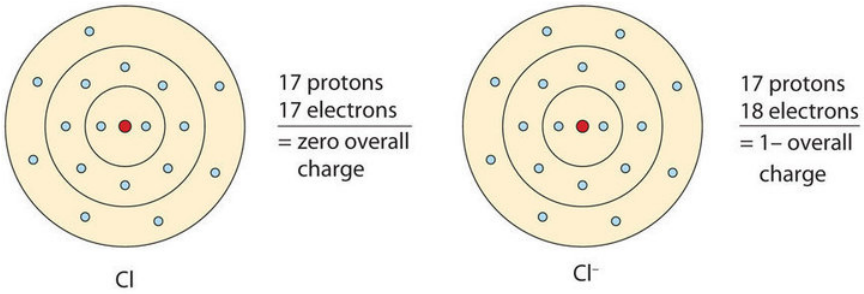


Figure **2.3.2**: The Formation of a Chlorine Ion. On the left, the chlorine atom has 17 electrons. On the right, the chloride ion has 18 electrons and has a 1− charge.

With two oppositely charged ions, there is an electrostatic attraction between them because opposite charges attract. The resulting combination is the compound sodium chloride. Notice that there are no leftover electrons. The number of electrons lost by the sodium atom (one) equals the number of electrons gained by the chlorine atom (one), so the compound is electrically neutral. In macroscopic samples of sodium chloride, there are billions and billions of sodium and chloride ions, although there is always the same number of cations and anions.

✔ Example 2.3.1

Write the electron configuration of aluminum atom (Z=13). How many electrons must Al lose/gain to achieve octet? Write the formula of the resulting ion and its electron configuration.

Solution

The electron configuration of Al atom is $1s^2 2s^2 2p^6 3s^2 3p^1$. The second shell has octet ($2s^2 2p^6$) while the valence shell has 3 electrons ($3s^2 3p^1$). Mg can achieve octet by losing the 3 valence electrons. The resulting cation is Al^{3+} with electron configuration, $1s^2 2s^2 2p^6$.

2.3.1 Exercise

Write the electron configuration of oxygen atom (Z=8). How many electrons must O lose/gain to achieve octet? Write the formula of the resulting ion and its electron configuration.

Answer

The electron configuration of O atom is $1s^2 2s^2 2p^4$. The second shell has six electrons ($2s^2 2p^4$) and needs two electrons to achieve octet. Oxygen will gain 2 electrons. The resulting anion is O^{2-} with electron configuration, $1s^2 2s^2 2p^6$.

In many cases, elements that belong to the same group (vertical column) on the periodic table form ions with the same charge because they have the same number of valence electrons. Thus, the periodic table becomes a tool for remembering the charges on many ions. For example, all ions made from alkali metals, the first column on the periodic table, have a 1+ charge. Ions made from alkaline earth metals, the second group on the periodic table, have a 2+ charge. On the other side of the periodic table, the next-to-last column, the halogens, form ions having a 1− charge. Figure **2.3.3** shows how the charge on many ions can be predicted by the location of an element on the periodic table. Note the convention of first writing the number and then the sign on a multiply charged ion. The barium cation is written Ba^{2+}, not Ba^{+2}.

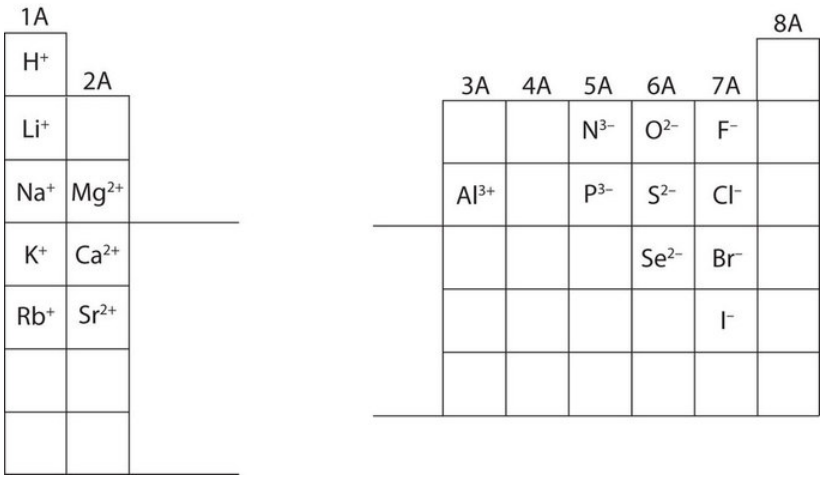


Figure **2.3.3**: Predicting Ionic Charges. The charge that an atom acquires when it becomes an ion is related to the structure of the periodic table. Within a group (family) of elements, atoms form ions of a certain charge.

Example 2.3.2

Which of these ions is *not likely* to form?

a. Mg^{+}
b. K^{+}

Solution

(a) Mg is in Group 2A and has two valence electrons. It achieves octet by losing two electrons to form Mg^{2+} cation. Losing only one electron to form Mg^{+} does not make an octet, hence, Mg^{+} is not likely to form.

2.3.2 Exercise

Which of these ions is *not likely* to form?

a. S^{3-}
b. N^{3-}

Answer

(a) S is in Group 6A and has six valence electrons. It achieves octet by gaining two electrons to form S^{2-} anion. Gaining three electrons to form S^{3-} does not make it octet, hence, S^{3-} is not likely to form.

Lewis Diagrams

Chemists use simple diagrams to show an atom's valence electrons and how they transfer. These diagrams have two advantages over the electron shell diagrams. First, they show only valence electrons. Second, instead of having a circle around the chemical symbol to represent the electron shell, they have up to eight dots around the symbol; each dot represents a valence electron. These dots are arranged to the right and left and above and below the symbol, **with no more than two dots on a side**. For example, the representation for sodium is as follows:



and the representation for chlorine is as follows:



> For the above diagrams, it does not matter what sides the dots are placed on in Lewis diagrams as long as each side has a maximum of two dots.

These diagrams are called **Lewis electron dot diagrams**, or simply Lewis diagrams, after Gilbert N. Lewis, the American chemist who introduced them. To write an element's Lewis dot symbol, place the dots representing its valence electrons, one at a time, around the element's chemical symbol. Up to four dots are placed above, below, to the left, and to the right of the symbol (in any order, as long as elements with four or fewer valence electrons have no more than one dot in each position). The next dots, for elements with more than four valence electrons, are again distributed one at a time, each paired with one of the first four. In other words, **place the dots singly on each side before pairing them**. The Lewis electron dot diagram of fluorine, for example, with seven valence electrons, is constructed as follows:

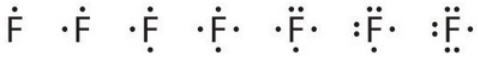
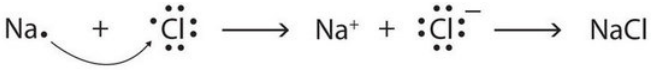


Figure **2.3.4** shows the electron configurations and Lewis diagrams of the elements lithium through neon, which is the entire second period of the periodic table. For the **main group** elements, the **number of valence electrons** is the **same as the group number** listed at the top of the periodic table.



Figure **2.3.4**: Lewis Electron Dot Diagrams of the Elements Lithium through Neon

The transfer of electrons can be illustrated easily with Lewis diagrams:



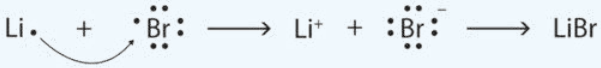
In representing the final formula, the dots are omitted.

Example 2.3.3

Starting with lithium and bromine atoms, use Lewis diagrams to show the formation of the ionic compound LiBr.

Solution

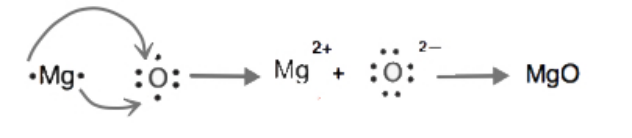
From the periodic table, we see that lithium is in the same column as sodium, so it will have the same valence shell electron configuration. That means that the neutral lithium atom will have the same Lewis diagram that the sodium atom has. Similarly, bromine is in the same column as chlorine, so it will have the same Lewis diagram that chlorine has. Therefore,



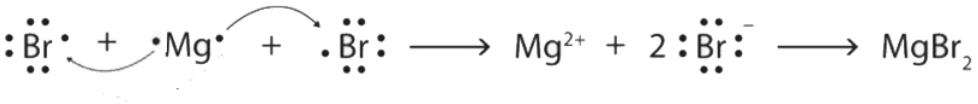
Exercise 2.3.3

Starting with magnesium and oxygen atoms, use Lewis diagrams to show the formation of the ionic compound MgO.

Answer



Some ionic compounds have different numbers of cations and anions. In those cases, electron transfer occurs between more than one atom. For example, here is the formation of $MgBr_2$:



Notice that in this example there are two bromide ions (1– charge) needed for every one magnesium ion (2+ charge) in order for the overall charge of the compound to equal zero. This is called **charge balance**. The number of each type of ion is indicated in the formula by the subscript.

Most of the elements that make ionic compounds form an ion that has a characteristic charge. For example, sodium makes ionic compounds in which the sodium ion always has a 1+ charge. Chlorine makes ionic compounds in which the chloride ion always has a 1− charge. Some elements, especially transition metals, can form ions of multiple charges. Figure **2.3.5** shows the characteristic charges for some of these ions. As we saw in Figure **2.3.1**, there is a pattern to the charges on many of the main group ions, but there is no simple pattern for transition metal ions (or for the larger main group elements).

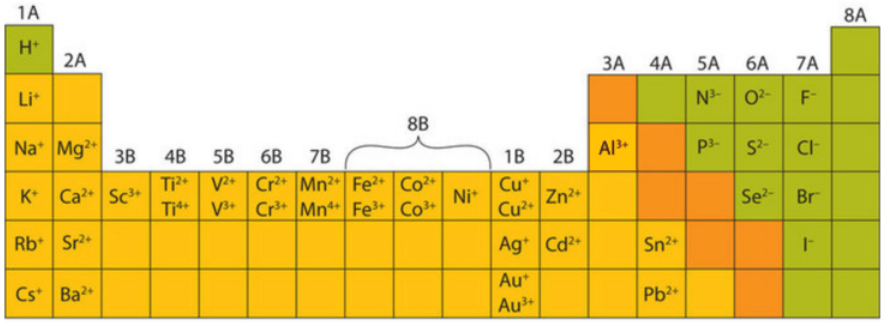


Figure **2.3.5**: Charges of the Monatomic Ions. Note that some atoms commonly form ions of variable charges.

Key Takeaways

- Ions can be positively charged or negatively charged.
- A Lewis diagram is used to show how electrons are transferred to make ions and ionic compounds.

Exercises

1. What are the two types of ions?
2. Use Lewis diagrams to illustrate the formation of an ionic compound from a potassium atom and an iodine atom.

3. When the following atoms become ions, what charges do they acquire?

a. Li
b. S
c. Ca
d. F

4. Identify each as a cation, an anion, or neither.

a. H^+
b. Cl^-
c. O_2
d. Ba^{2+}
e. CH_4
f. CS_2

5. Identify each as a cation, an anion, or neither.

a. NH_3
b. Br^-
c. H^-
d. Hg^{2+}

e. CCl_4
f. SO_3

6. Write the electron configuration for each ion.

a. Li^+
b. Mg^{2+}
c. F^-
d. S^{2-}

7. Write the electron configuration for each ion.

a. Na^+
b. Be^{2+}
c. Cl^-
d. O^{2-}

8. Draw Lewis diagrams for the ions listed in Exercise 6. Also include Lewis diagrams for the respective neutral atoms as a comparison.

9. Draw Lewis diagrams for the ions listed in Exercise 7. Also include Lewis diagrams for the respective neutral atoms as a comparison.

10. Using Lewis diagrams, show the electron transfer for the formation of LiF.

11. Using Lewis diagrams, show the electron transfer for the formation of MgO.

12. Using Lewis diagrams, show the electron transfer for the formation of Li_2O.

13. Using Lewis diagrams, show the electron transfer for the formation of CaF_2.

14. What characteristic charge do atoms in the first column of the periodic table have when they become ions?

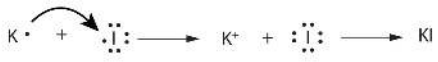
15. What characteristic charge do atoms in the second column of the periodic table have when they become ions?

16. What characteristic charge do atoms in the third-to-last column of the periodic table have when they become ions?

17. What characteristic charge do atoms in the next-to-last column of the periodic table have when they become ions?

Answers

1. Cations (positive charged) and anions (negative charged).

2. 

3.
a. 1+
b. 2−
c. 2+
d. 1−

4.
a. cation
b. anion
c. neither
d. cation
e. neither
f. neither

5.
a. neither
b. anion

c. anion
d. cation
e. neither
f. neither

6.

a. $1s^2$
b. $1s^22s^22p^6$
c. $1s^22s^22p^6$
d. $1s^22s^22p^63s^23p^6$

7.

a. $1s^22s^22p^6$
b. $1s^2$
c. $1s^22s^22p^63s^23p^6$
d. $1s^22s^22p^6$

8.

a. $Li\bullet$, Li^+

b. $\bullet Mg\bullet$, Mg^{2+}

c. $:\ddot{F}:$, $:\ddot{F}:^-$

d. $:\ddot{S}:$, $:\ddot{S}:^{2-}$

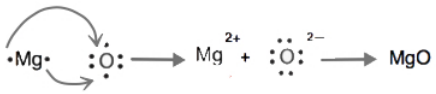
$Na\bullet$, Na^+

$\bullet Be\bullet$, Be^{2+}

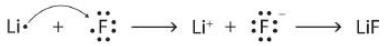
$:\ddot{Cl}:$, $:\ddot{Cl}:^-$

$:\ddot{O}:$, $:\ddot{O}:^{2-}$

9.

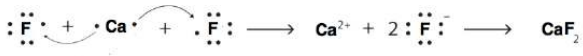


10.



11.

12.



13.



14. 1+

15. 2+

16. 2−

17. 1−

2.3: Ions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.2: Ions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

2.4: Formulas for Ionic Compounds

Learning Objectives

- Write the chemical formula for a simple ionic compound.
- Recognize polyatomic ions in chemical formulas.

We have already encountered some chemical formulas for simple ionic compounds. A chemical formula is a concise list of the elements in a compound and the ratios of these elements. To better understand what a chemical formula means, we must consider how an ionic compound is constructed from its ions.

Ionic compounds exist as alternating positive and negative ions in regular, three-dimensional arrays called crystals (Figure 2.4.1). As you can see, there are no individual NaCl "particles" in the array; instead, there is a continuous lattice of alternating sodium and chloride ions. However, we can use the ratio of sodium ions to chloride ions, expressed in the lowest possible whole numbers, as a way of describing the compound. In the case of sodium chloride, the ratio of sodium ions to chloride ions, expressed in lowest whole numbers, is 1:1, so we use NaCl (one Na symbol and one Cl symbol) to represent the compound. Thus, NaCl is the chemical formula for sodium chloride, which is a concise way of describing the relative number of different ions in the compound. A macroscopic sample is composed of myriads of NaCl pairs; each individual pair called a **formula unit**. Although it is convenient to think that NaCl crystals are composed of individual NaCl units, Figure 2.4.1 shows that no single ion is exclusively associated with any other single ion. Each ion is surrounded by ions of opposite charge.

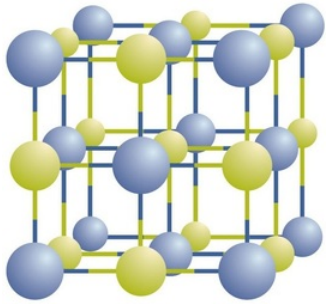


Figure 2.4.1: A Sodium Chloride Crystal. A crystal contains a three-dimensional array of alternating positive and negative ions. The precise pattern depends on the compound. A crystal of sodium chloride, shown here, is a collection of alternating sodium and chlorine ions.

The formula for an ionic compound follows several conventions. First, the **cation** is written **before** the **anion**. Because most metals form cations and most nonmetals form anions, formulas typically list the metal first and then the nonmetal. Second, **charges** are **not written** in a formula. Remember that in an ionic compound, the component species are ions, not neutral atoms, even though the formula does not contain charges. Finally, the proper formula for an ionic compound always has ***a net zero charge,*** meaning the total positive charge must equal the total negative charge. To determine the proper formula of any combination of ions, determine how many of each ion is needed to balance the total positive and negative charges in the compound.

> *This rule is ultimately based on the fact that matter is, overall, electrically neutral.*

By convention, assume that there is only one atom if a subscript is not present. We do not use 1 as a subscript.

If we look at the ionic compound consisting of lithium ions and bromide ions, we see that the lithium ion has a 1+ charge and the bromide ion has a 1− charge. Only one ion of each is needed to balance these charges. The formula for lithium bromide is LiBr.

When an ionic compound is formed from magnesium and oxygen, the magnesium ion has a 2+ charge, and the oxygen atom has a 2− charge. Although both of these ions have higher charges than the ions in lithium bromide, they still balance each other in a one-to-one ratio. Therefore, the proper formula for this ionic compound is MgO.

Now consider the ionic compound formed by magnesium and chlorine. A magnesium ion has a 2+ charge, while a chlorine ion has a 1− charge:

$$Mg^{2+}Cl^{-}$$

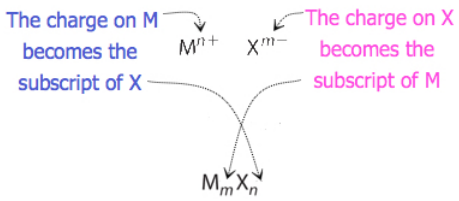
Combining one ion of each does not completely balance the positive and negative charges. The easiest way to balance these charges is to assume the presence of *two* chloride ions for each magnesium ion:

$$Mg^{2+}\ Cl^{-}\ Cl^{-}$$

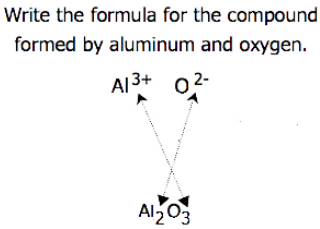
Now the positive and negative charges are balanced. We could write the chemical formula for this ionic compound as **MgClCl**, but the convention is to use a numerical subscript when there is more than one ion of a given type—**$MgCl_2$**. This chemical formula says that there are one magnesium ion and *two* chloride ions in this formula. (Do not read the "Cl_2" part of the formula as a molecule of the diatomic elemental chlorine. Chlorine does not exist as a diatomic element in this compound. Rather, it exists as two individual chloride ions.) By convention, the **lowest whole number ratio** is used in the formulas of ionic compounds. The formula **Mg_2Cl_4** has balanced charges with the ions in a 1:2 ratio, but it is not the lowest whole number ratio.

> By convention, the lowest whole-number ratio of the ions is used in ionic formulas. There are exceptions for certain ions, such as **Hg_2^{2+}**.

For compounds in which the ratio of ions is not as obvious, the subscripts in the formula can be obtained by **crossing charges**: use the absolute value of the charge on one ion as the subscript for the other ion. This method is shown schematically in Figure 3.3.2.



The image shows M with a charge of n+ and X with a charge of m-. The charge on the M becomes the subscript of X, and the charge on X becomes the subscript of M, making the final product M subscript m X subscript n.



The image asks to write the formula for the compound formed by aluminum and oxygen. It shows Al with a charge of 3+ and O with a charge of 2-. By crossing charges, the final result is Al subscript 2 O subscript 3.

Figure **2.4.2**: Crossing charges. One method for obtaining subscripts in the empirical formula is by crossing charges.

When crossing charges, it is sometimes necessary to reduce the subscripts to their simplest ratio to write the empirical formula. Consider, for example, the compound formed by Pb^{4+} and O^{2-}. Using the absolute values of the charges on the ions as subscripts gives the formula Pb_2O_4. This simplifies to its correct empirical formula **PbO_2**. The empirical formula has one Pb^{4+} ion and two O^{2-} ions.

✔ Example 2.4.1

Write the chemical formula for an ionic compound composed of each pair of ions.

a. the sodium ion and the sulfur ion
b. the aluminum ion and the fluoride ion
c. the 3+ iron ion and the oxygen ion

Solution

a. To obtain a valence shell octet, sodium forms an ion with a 1+ charge, while the sulfur ion has a 2− charge. Two sodium 1+ ions are needed to balance the 2− charge on the sulfur ion. Rather than writing the formula as $\mathbf{NaNaS}$, we shorten it by convention to $\mathbf{Na_2S}$.

b. The aluminum ion has a 3+ charge, while the fluoride ion formed by fluorine has a 1− charge. Three fluorine 1− ions are needed to balance the 3+ charge on the aluminum ion. This combination is written as $\mathbf{AlF_3}$.

c. Iron can form two possible ions, but the ion with a 3+ charge is specified here. The oxygen atom has a 2− charge as an ion. To balance the positive and negative charges, we look to the least common multiple—6: two iron 3+ ions will give 6+, while three 2− oxygen ions will give 6−, thereby balancing the overall positive and negative charges. Thus, the formula for this ionic compound is $\mathbf{Fe_2O_3}$. Alternatively, use the crossing charges method shown in Figure 3.3.2.

? Exercise 2.4.1

Write the chemical formula for an ionic compound composed of each pair of ions.

a. the calcium ion and the oxygen ion
b. the 2+ copper ion and the sulfur ion
c. the 1+ copper ion and the sulfur ion

Answer a:

CaO

Answer b:

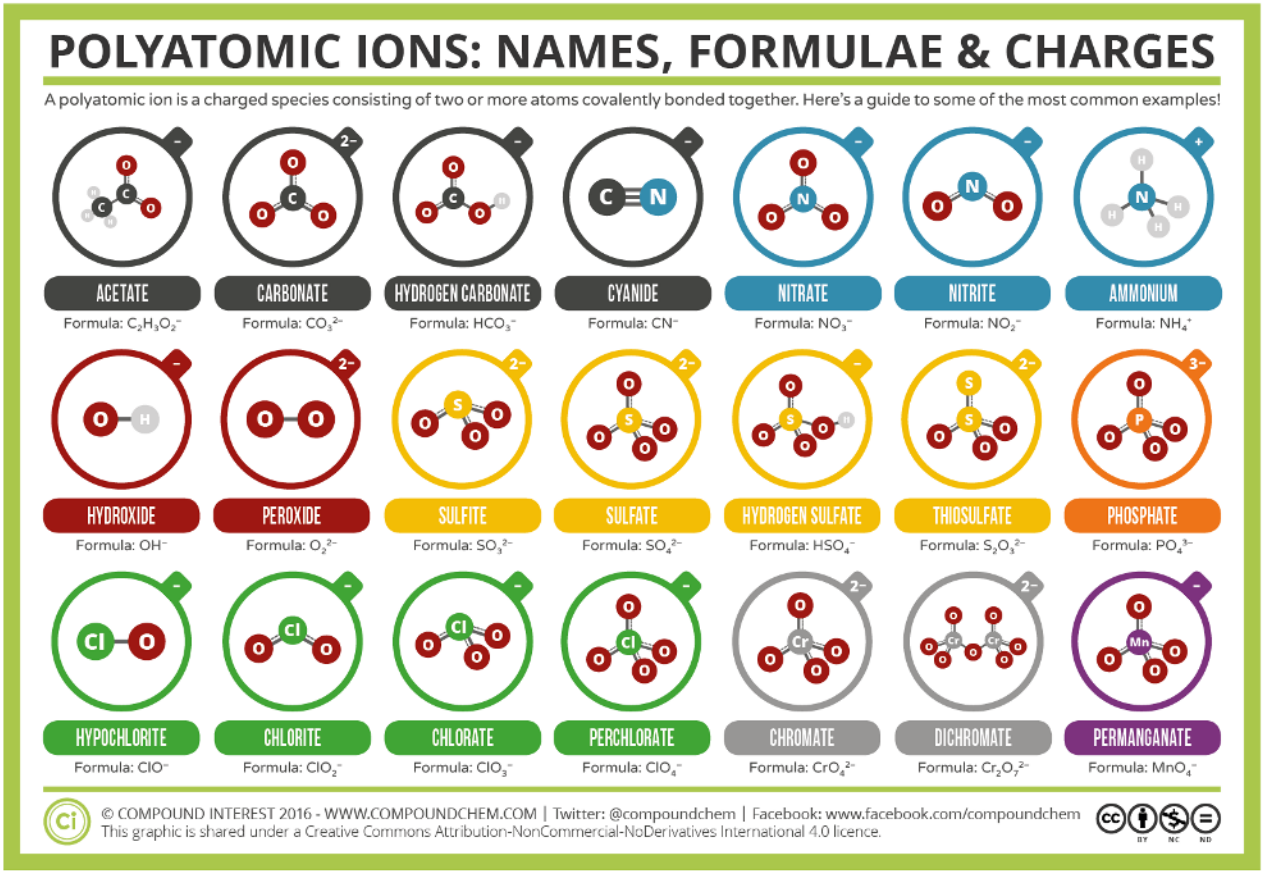
CuS

Answer c:

Cu_2S

Polyatomic Ions

Some ions consist of groups of atoms ***covalently*** bonded together and have an **overall electric charge**. Because these ions contain more than one atom, they are called polyatomic ions. The Lewis structures, names and formulas of some polyatomic ions are found in Table 3.3.1.



Some Polyatomic Ions

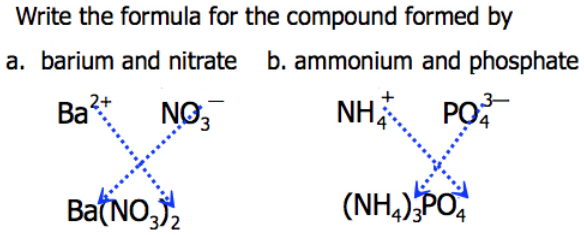
Polyatomic ions have defined formulas, names, and charges that cannot be modified in any way. Table **2.4.2** lists the ion names and ion formulas of the most common polyatomic ions. For example, $\mathbf{NO_3^-}$ is the nitrate ion; it has one nitrogen atom and three oxygen atoms and an overall 1− charge. Figure **2.4.2** lists the most common polyatomic ions.

Table **2.4.2**: Ion Names and Ion Formulas of Common Polyatomic Ions

Ion Name	Ion Formula
ammonium ion	NH_4^{+1}
hydroxide ion	OH^{-1}
cyanide ion	CN^{-1}
carbonate ion	CO_3^{-2}
bicarbonate or hydrogen carbonate	HCO_3^-
acetate ion	$C_2H_3O_2^{-1}$ **or** $CH_3CO_2^{-1}$
nitrate ion	NO_3^{-1}
nitrite ion	NO_2^{-1}
sulfate ion	SO_4^{-2}
sulfite ion	SO_3^{-2}
phosphate ion	PO_4^{-3}
phosphite ion	PO_3^{-3}

Note that only one polyatomic ion in this Table, the ammonium ion ($NH_4^{+1)}$, is a cation. This polyatomic ion contains one nitrogen and four hydrogens that collectively bear a +1 charge. The remaining polyatomic ions are all negatively-charged and, therefore, are classified as anions. However, only two of these, the hydroxide ion and the cyanide ion, are named using the "-ide" suffix that is typically indicative of negatively-charged particles. The remaining polyatomic anions, which all contain oxygen, in combination with another non-metal, exist as part of a series in which the number of oxygens within the polyatomic unit can vary. As has been repeatedly emphasized in several sections of this text, no two chemical formulas should share a common chemical name. A single suffix, "-ide," is insufficient for distinguishing the names of the anions in a related polyatomic series. Therefore, "-ate" and "-ite" suffixes are employed, in order to denote that the corresponding polyatomic ions are part of a series. Additionally, these suffixes also indicate the relative number of oxygens that are contained within the polyatomic ions. Note that all of the polyatomic ions whose names end in "-ate" contain one more oxygen than those polyatomic anions whose names end in "-ite." Unfortunately, much like the common system for naming transition metals, these suffixes only indicate the *relative* number of oxygens that are contained within the polyatomic ions. For example, the nitrate ion, which is symbolized as NO_3^{-1}, has one more oxygen than the nitrite ion, which is symbolized as NO_2^{-1}. However, the sulfate ion is symbolized as SO_4^{-2}. While both the nitrate ion and the sulfate ion share an "-ate" suffix, the former contains three oxygens, but the latter contains four. Additionally, both the nitrate ion and the sulfite ion contain three oxygens, but these polyatomic ions do not share a common suffix. Unfortunately, the relative nature of these suffixes mandates that the ion formula/ion name combinations of the polyatomic ions must simply be memorized.

The rule for constructing formulas for ionic compounds containing polyatomic ions is the same as for formulas containing monatomic (single-atom) ions: the positive and negative charges must balance. If more than one of a particular polyatomic ion is needed to balance the charge, the *entire formula* for the polyatomic ion must be enclosed in parentheses, and the numerical subscript is placed *outside* the parentheses. This is to show that the subscript applies to the entire polyatomic ion. Two examples are shown below:



The image asks a two-part question: Write the formula for the compound formed by (a) barium and nitrate, and (b) ammonium and phosphate. For (a), Ba has a charge of 2+ and NO_3 has a charge of -1. By crossing charges, the formula is $Ba(NO_3)_2$. For (b), NH_4 has a charge of +1 and PO_4 has a charge of 3-, so by crossing charges the formula is $(NH_4)_3PO_4$.

Example 2.4.2

Write the chemical formula for an ionic compound composed of each pair of ions.

a. the potassium ion and the sulfate ion
b. the calcium ion and the nitrate ion

Solution

a. Potassium ions have a charge of 1+, while sulfate ions have a charge of 2−. We will need two potassium ions to balance the charge on the sulfate ion, so the proper chemical formula is $\mathbf{K_2SO_4}$.
b. Calcium ions have a charge of 2+, while nitrate ions have a charge of 1−. We will need two nitrate ions to balance the charge on each calcium ion. The formula for nitrate must be enclosed in parentheses. Thus, we write $\mathbf{Ca(NO_3)_2}$ as the formula for this ionic compound.

Exercise 2.4.2

Write the chemical formula for an ionic compound composed of each pair of ions.

a. the magnesium ion and the carbonate ion
b. the aluminum ion and the acetate ion

Answer a:

Mg^{2+} and $CO_3^{2-} = MgCO_3$

Answer b:

Al^{3+} and $C_2H_3O_2^- = Al(C_2H_3O_2)_3$

Recognizing Ionic Compounds

There are two ways to recognize ionic compounds. First, compounds between **metal and nonmetal** elements are usually **ionic**. For example, $CaBr_2$ contains a metallic element (calcium, a group 2A metal) and a nonmetallic element (bromine, a group 7A nonmetal). Therefore, it is most likely an ionic compound. (In fact, it *is* ionic.) In contrast, the compound NO_2 contains two elements that are both nonmetals (nitrogen, from group 5A, and oxygen, from group 6A). It is not an ionic compound; it belongs to the category of covalent compounds discuss elsewhere. Also note that this combination of nitrogen and oxygen has no electric charge specified, so it is *not* the nitrite ion.

Second, if you recognize the formula of a **polyatomic ion** in a compound, the compound is **ionic**. For example, if you see the formula $\mathbf{Ba(NO_3)_2}$, you may recognize the "NO_3" part as the nitrate ion, $\mathbf{NO_3^-}$. (Remember that the convention for writing formulas for ionic compounds is not to include the ionic charge.) This is a clue that the other part of the formula, $\mathbf{Ba}$, is actually the $\mathbf{Ba^{2+}}$ ion, with the 2+ charge balancing the overall 2− charge from the two nitrate ions. Thus, this compound is also ionic.

Example 2.4.3

Identify each compound as ionic or not ionic.

a. Na_2O
b. PCl_3
c. NH_4Cl
d. OF_2

Solution

a. Sodium is a metal, and oxygen is a nonmetal; therefore, Na_2O is expected to be ionic.
b. Both phosphorus and chlorine are nonmetals. Therefore, PCl_3 is not ionic.
c. The NH_4 in the formula represents the ammonium ion, NH_4^+, which indicates that this compound is ionic.
d. Both oxygen and fluorine are nonmetals. Therefore, OF_2 is not ionic.

Exercise 2.4.3

Identify each compound as ionic or not ionic.

a. N_2O
b. $FeCl_3$
c. $(NH_4)_3PO_4$
d. $SOCl_2$

Answer a:

not ionic

Answer b:

ionic

Answer c:

ionic

Answer d:

not ionic

Looking Closer: Blood and Seawater

Science has long recognized that blood and seawater have similar compositions. After all, both liquids have ionic compounds dissolved in them. The similarity may be more than mere coincidence; many scientists think that the first forms of life on Earth arose in the oceans. A closer look, however, shows that blood and seawater are quite different. A 0.9% solution of sodium chloride approximates the salt concentration found in blood. In contrast, seawater is principally a 3% sodium chloride solution, over three times the concentration in blood. Here is a comparison of the amounts of ions in blood and seawater:

Table showing a comparison of the amounts of ions in blood and seawater.

Ion	Percent in Seawater	Percent in Blood
Na^+	2.36	0.322
Cl^-	1.94	0.366
Mg^{2+}	0.13	0.002
SO_4^{2-}	0.09	—
K^+	0.04	0.016
Ca^{2+}	0.04	0.0096
HCO_3^-	0.002	0.165
HPO_4^{2-}, $H_2PO_4^-$	—	0.01

Most ions are more abundant in seawater than they are in blood, with some important exceptions. There are far more hydrogen carbonate ions ($\mathbf{HCO_3^-}$) in blood than in seawater. This difference is significant because the hydrogen carbonate ion and some related ions have a crucial role in controlling the acid-base properties of blood. The amount of hydrogen phosphate ions—$\mathbf{HPO_4^{2-}}$ and $\mathbf{H_2PO_4^-}$—in seawater is very low, but they are present in higher amounts in blood, where they also affect acid-base properties. Another notable difference is that blood does not have significant amounts of the sulfate ion ($\mathbf{SO_4^{2-}}$), but this ion is present in seawater.

Key Takeaways

- Proper chemical formulas for ionic compounds balance the total positive charge with the total negative charge.
- Groups of atoms with an overall charge, called polyatomic ions, also exist.

EXERCISES

1. What information is contained in the formula of an ionic compound?

- Why do the chemical formulas for some ionic compounds contain subscripts, while others do not?

3. Write the chemical formula for the ionic compound formed by each pair of ions.

   a. Mg^{2+} and I^-
   b. Na^+ and O^{2-}

4. Write the chemical formula for the ionic compound formed by each pair of ions.

   a. Na^+ and Br^-
   b. Mg^{2+} and Br^-
   c. Mg^{2+} and S^{2-}

5. Write the chemical formula for the ionic compound formed by each pair of ions.

   a. K^+ and Cl^-
   b. Mg^{2+} and Cl^-
   c. Mg^{2+} and Se^{2-}

6. Write the chemical formula for the ionic compound formed by each pair of ions.

   a. Na^+ and N^{3-}
   b. Mg^{2+} and N^{3-}
   c. Al^{3+} and S^{2-}

7. Write the chemical formula for the ionic compound formed by each pair of ions.

   a. Li^+ and N^{3-}
   b. Mg^{2+} and P^{3-}

c. Li^+ and P^{3-}

8. Write the chemical formula for the ionic compound formed by each pair of ions.

a. Fe^{3+} and Br^-
b. Fe^{2+} and Br^-
c. Au^{3+} and S^{2-}
d. Au^+ and S^{2-}

9. Write the chemical formula for the ionic compound formed by each pair of ions.

a. Cr^{3+} and O^{2-}
b. Cr^{2+} and O^{2-}
c. Pb^{2+} and Cl^-
d. Pb^{4+} and Cl^-

10. Write the chemical formula for the ionic compound formed by each pair of ions.

a. Cr^{3+} and NO_3^-
b. Fe^{2+} and PO_4^{3-}
c. Ca^{2+} and CrO_4^{2-}
d. Al^{3+} and OH^-

11. Write the chemical formula for the ionic compound formed by each pair of ions.

a. NH_4^+ and NO_3^-
b. H^+ and $Cr_2O_7^{2-}$
c. Cu^+ and CO_3^{2-}
d. Na^+ and HCO_3^-

12. For each pair of elements, determine the charge for their ions and write the proper formula for the resulting ionic compound between them.

a. Ba and S
b. Cs and I

13. For each pair of elements, determine the charge for their ions and write the proper formula for the resulting ionic compound between them.

a. K and S
b. Sc and Br

14. Which compounds would you predict to be ionic?

a. Li_2O
b. $(NH_4)_2O$
c. CO_2
d. $FeSO_3$
e. C_6H_6
f. C_2H_6O

15. Which compounds would you predict to be ionic?

a. $Ba(OH)_2$
b. CH_2O
c. NH_2CONH_2
d. $(NH_4)_2CrO_4$
e. C_8H_{18}
f. NH_3

Answers

1. the ratio of each kind of ion in the compound

2. Sometimes more than one ion is needed to balance the charge on the other ion in an ionic compound.

3.
   a. MgI_2
   b. Na_2O

4.
   a. NaBr
   b. $MgBr_2$
   c. MgS

5.
   a. KCl
   b. $MgCl_2$
   c. MgSe

6.
   a. Na_3N
   b. Mg_3N_2
   c. Al_2S_3

7.
   a. Li_3N
   b. Mg_3P_2
   c. Li_3P

8.
   a. $FeBr_3$
   b. $FeBr_2$
   c. Au_2S_3
   d. Au_2S

9.
   a. Cr_2O_3
   b. CrO
   c. $PbCl_2$
   d. $PbCl_4$

10.
   a. $Cr(NO_3)_3$
   b. $Fe_3(PO_4)_2$
   c. $CaCrO_4$
   d. $Al(OH)_3$

11.
   a. NH_4NO_3
   b. $H_2Cr_2O_7$
   c. Cu_2CO_3
   d. $NaHCO_3$

12.

a. Ba^{2+}, S^{2-}, BaS
b. Cs^{+}, I^{-}, CsI

13.

a. K^{+}, S^{2-}, K_2S
b. Sc^{3+}, Br^{-}, $ScBr_3$

14.

a. ionic
b. ionic
c. not ionic
d. ionic
e. not ionic
f. not ionic

15.

a. ionic
b. not ionic
c. not ionic
d. ionic
e. not ionic
f. not ionic

2.4: Formulas for Ionic Compounds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.3: Formulas for Ionic Compounds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

2.5: Ionic Bonding and Simple Ionic Compounds (Exercises)

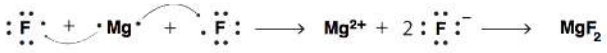
Additional Exercises

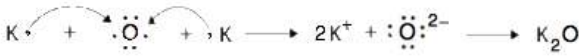
1. What number shell is the valence electron shell of a sodium atom? What number shell is the valence shell of a sodium ion? Explain the difference.
2. What number shell is the valence electron shell of a bromine atom? What number shell is the valence shell of a bromide ion? Explain the difference between these answers and the answers to Exercise 1.
3. What is the electron configuration of each ion?
    1. K^+
    2. Mg^{2+}
    3. F^-
    4. S^{2-}
4. What is the electron configuration of each ion?
    1. Li^+
    2. Ca^{2+}
    3. Cl^-
    4. O^{2-}
5. 1. If a sodium atom were to lose two electrons, what would be the electron configuration of the resulting cation?
    2. Considering that electron shells are typically separated by large amounts of energy, use your answer to Exercise 5a to suggest why sodium atoms do not form a 2+ cation.
6. 1. If a chlorine atom were to gain two electrons, what would be the electron configuration of the resulting anion?
    2. Considering that electron shells are typically separated by large amounts of energy, use your answer to Exercise 6a to suggest why chlorine atoms do not form a 2− anion.
7. Use Lewis diagrams and arrows to show the electron transfer that occurs during the formation of an ionic compound among Mg atoms and F atoms. (Hint: how many atoms of each will you need?)
8. Use Lewis diagrams and arrows to show the electron transfer that occurs during the formation of an ionic compound among K atoms and O atoms. (Hint: how many atoms of each will you need?)
9. Mercury forms two possible cations—Hg^{2+} and Hg_2^{2+}, the second of which is actually a two-atom cation with a 2+ charge.
    1. Using common names, give the probable names of these ions.
    2. What are the chemical formulas of the ionic compounds these ions make with the oxide ion, O^{2-}?
10. The uranyl ion (UO_2^{2+}) is a common water-soluble form of uranium. What is the chemical formula of the ionic compound uranyl nitrate? What is the chemical formula of the ionic compound uranyl phosphate?
11. The formal chemical name of the mineral *strengite* is iron(III) phosphate dihydrate. What is the chemical formula of strengite? What is the formula mass of strengite?
12. What is the formula mass of $MgSO_4 \cdot 7H_2O$?
13. What is the formula mass of $CaSO_4 \cdot ½H_2O$?
14. What mass does 20 formula units of NaCl have?
15. What mass does 75 formula units of K_2SO_4 have?
16. If an atomic mass unit equals 1.66×10^{-24} g, what is the mass in grams of one formula unit of NaCl?
17. If an atomic mass unit equals 1.66×10^{-24} g, what is the mass in grams of 5.00×10^{22} formula units of NaOH?
18. If an atomic mass unit equals 1.66×10^{-24} g, what is the mass in grams of 3.96×10^{23} formula units of $(NH_4)_2SO_4$?
19. Both tin and lead acquire 2+ or 4+ charges when they become ions. Use the periodic table to explain why this should not surprise you.
20. Which ion would you expect to be larger in size—In^{3+} or Tl^{3+}? Explain.
21. Which ion would you expect to be smaller in size—I^- or Br^-? Explain.
22. Which ion with a 2+ charge has the following electron configuration? $1s^2 2s^2 2p^6$
23. Which ion with a 3− charge has the following electron configuration? $1s^2 2s^2 2p^6$

Answers

1. For sodium, the valence shell is the third shell; for the sodium ion, the valence shell is the second shell because it has lost all its third shell electrons.
2. The valence shell for both bromine atom and bromide ion is n=4. This is because the valence shell of bromine atom can accommodate one more electron.
3. 1. $1s^2 2s^2 2p^6 3s^2 3p^6$
   2. $1s^2 2s^2 2p^6$
   3. $1s^2 2s^2 2p^6$
   4. $1s^2 2s^2 2p^6 3s^2 3p^6$
4. 1. $1s^2$
   2. $1s^2 2s^2 2p^6 3s^2 3p^6$
   3. $1s^2 2s^2 2p^6 3s^2 3p^6$
   4. $1s^2 2s^2 2p^6$
5. 1. $1s^2 2s^2 2p^5$
   2. It probably requires too much energy to form.
6. 1. $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1$

2. Gaining the second electron would probably require too much energy.

7. 

8. 

9. 1. mercuric and mercurous, respectively

2. HgO and Hg_2O, respectively

10. uranyl nitrate $UO_2(NO_3)_2$ and uranyl phosphate $(UO_2)_3(PO4)_2$

11. $FePO_4 \cdot 2H_2O$; 186.86 u

12. 246.51 u

13. 145.16 u

14. 1,169 u

15. 13,070 u

16. 9.701×10^{-23} g

17. 3.32 g

18. 86.9 g

19. Both tin and lead have two *p* electrons and two *s* electrons in their valence shells.

20. Tl^{3+} is larger because it is found lower on the periodic table

21. Br^- because it is higher up on the periodic table

22. Mg^{2+}

23. N^{3-}

2.5: Ionic Bonding and Simple Ionic Compounds (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.E: Ionic Bonding and Simple Ionic Compounds (Exercises)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

2.6: Ionic Bonding and Simple Ionic Compounds (Summary)

To ensure that you understand the material in this chapter, you should review the meanings of the following bold terms and ask yourself how they relate to the topics in the chapter.

Atoms combine into compounds by forming **chemical bonds**. A survey of stable atoms and molecules leads to the **octet rule**, which says that stable atoms tend to have eight electrons in their outermost, or valence, shell. One way atoms obtain eight electrons in the valence shell is for some atoms to lose electrons while other atoms gain them. When this happens, the atoms take on an electrical charge. Charged atoms are called **ions**. Ions having opposite charges attract each other. This attraction is called **ionic bonding**, and the compounds formed are called **ionic compounds**.

Positively charged ions are called **cations**, while negatively charged ions are called **anions**. The formation of both cations and anions can be illustrated using electron configurations. Because elements in a column of the periodic table have the same valence shell electron configuration, atoms in the same column of the periodic table tend to form ions having the same charge. **Electron dot diagrams**, or **Lewis diagrams**, can also be used to illustrate the formation of cations and anions.

Ionic compounds are represented in writing by a **chemical formula**, which gives the lowest ratio of cations and anions present in the compound. In a formula, the symbol of the cation is written first, followed by the symbol of the anion. **Formula unit** is considered the basic unit of an ionic compound because ionic compounds do not exist as discrete units. Instead, they exist as **crystals**, three-dimensional arrays of ions, with cations surrounded by anions and anions surrounded by cations. Chemical formulas for ionic compounds are determined by balancing the positive charge from the cation(s) with the negative charge from the anion(s). A subscript to the right of the ion indicates that more than one of that ion is present in the chemical formula.

Some ions are groups of atoms bonded together and having an overall electrical charge. These are called **polyatomic ions**. Writing formulas with polyatomic ions follows the same rules as with monatomic ions, except that when more than one polyatomic ion is present in a chemical formula, the polyatomic ion is enclosed in parentheses and the subscript is outside the right parenthesis. Ionic compounds typically form between metals and nonmetals or between polyatomic ions.

Names of ionic compounds are derived from the names of the ions, with the name of the cation coming first, followed by the name of the anion. If an element can form cations of different charges, there are two alternate systems for indicating the compound's name. In the **Stock system**, a roman numeral in parentheses indicates the charge on the cation. An example is the name for $FeCl_2$, which is iron(II) chloride. In the common system, the suffixes *-ous* and *-ic* are used to stand for the lower and higher possible charge of the cation, respectively. These suffixes are attached to a stem representing the element (which frequently comes from the Latin form of the element name). An example is the common name for $FeCl_2$, which is ferrous chloride.

The **formula mass** of an ionic compound is the sum of the masses of each individual atom in the formula. Care must be taken when calculating formula masses for formulas containing multiple polyatomic ions because the subscript outside the parentheses refers to all the atoms in the polyatomic ion.

Ionic compounds are hard, brittle and have very high melting points. When in solution, the ions separate and form **electrolyte** solutions.

2.6: Ionic Bonding and Simple Ionic Compounds (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.S: Ionic Bonding and Simple Ionic Compounds (Summary)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

CHAPTER OVERVIEW

3: Covalent Bonding and Simple Molecular Compounds

Ionic bonding results from the transfer of electrons among atoms or groups of atoms. In this chapter, we will consider another type of bonding—covalent bonding. We will examine how atoms share electrons to form these bonds, and we will begin to explore how the resulting compounds, such as cholesterol, are different from ionic compounds.

3.1: Prelude to Covalent Bonding and Simple Molecular Compounds
3.2: Covalent Bonds
3.3: Covalent Compounds - Formulas and Names
3.4: Drawing Lewis Structures
3.5: Characteristics of Covalent Bonds
3.6: Characteristics of Molecules
3.7: Organic Chemistry
3.8: Covalent Bonding and Simple Molecular Compounds (Exercises)
3.9: Covalent Bonding and Simple Molecular Compounds (Summary)

Template:HideTOC

3: Covalent Bonding and Simple Molecular Compounds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.1: Prelude to Covalent Bonding and Simple Molecular Compounds

Cholesterol ($C_{27}H_{46}O$), a compound that is sometimes in the news, is a white, waxy solid produced in the liver of every animal, including humans. It is important for building cell membranes and in producing certain hormones (chemicals that regulate cellular activity in the body). As such, it is necessary for life, but why is cholesterol the object of attention?

Besides producing cholesterol, we also ingest some whenever we eat meat or other animal-based food products. People who eat such products in large quantities, or whose metabolisms are unable to handle excess amounts, may experience an unhealthy buildup of cholesterol in their blood. Deposits of cholesterol, called plaque, may form on blood vessel walls, eventually blocking the arteries and preventing the delivery of oxygen to body tissues. Heart attacks, strokes, and other circulatory problems can result.

Most medical professionals recommend diets that minimize the amount of ingested cholesterol as a way of preventing heart attacks and strokes. Tests are available to measure cholesterol in the blood, and there are several drugs capable of lowering cholesterol levels.

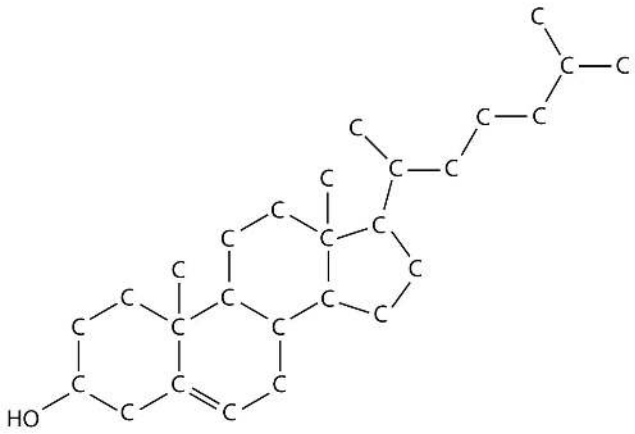


Figure 3.1.1: A Molecular Model of Cholesterol

3.1: Prelude to Covalent Bonding and Simple Molecular Compounds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.0: Prelude to Covalent Bonding and Simple Molecular Compounds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

3.2: Covalent Bonds

Learning Objectives

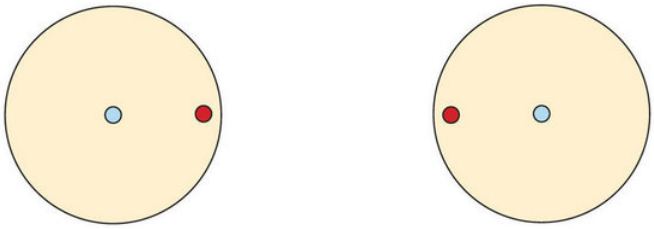
- To describe how a covalent bond forms.
- To apply the octet rule to covalent compounds

You have already seen examples of substances that contain covalent bonds. One substance mentioned previously was water (H_2O). You can tell from its formula that it is not an ionic compound; it is not composed of a metal and a nonmetal. Consequently, its properties are different from those of ionic compounds.

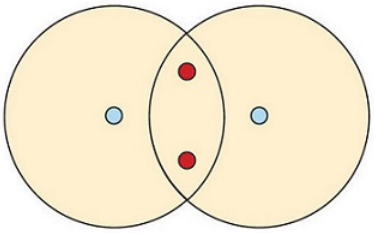
Electron Sharing

Previously, we discussed ionic bonding where electrons can be transferred from one atom to another so that both atoms have an energy-stable outer electron shell. Because most filled electron shells have eight electrons in them, chemists called this tendency the octet rule. However, there is another way an atom can achieve a full valence shell: atoms can *share* electrons.

This concept can be illustrated by using two hydrogen atoms, each of which has a single electron in its valence shell. (For small atoms such as hydrogen atoms, the valence shell will be the first shell, which holds only two electrons.) We can represent the two individual hydrogen atoms as follows:



In contrast, when two hydrogen atoms get close enough together to share their electrons, they can be represented as follows:



By sharing their valence electrons, both hydrogen atoms now have two electrons in their respective valence shells. Because each valence shell is now filled, this arrangement is more stable than when the two atoms are separate. The sharing of electrons between atoms is called a covalent bond, and the two electrons that join atoms in a covalent bond are called a bonding pair of electrons. A discrete group of atoms connected by covalent bonds is called a molecule—the smallest part of a compound that retains the chemical identity of that compound.

Chemists frequently use Lewis diagrams to represent covalent bonding in molecular substances. For example, the Lewis diagrams of two separate hydrogen atoms are as follows:



The Lewis diagram of two hydrogen atoms sharing electrons looks like this:



This depiction of molecules is simplified further by using a dash to represent a covalent bond. The hydrogen molecule is then represented as follows:



Remember that the dash, also referred to as a single bond, represents a *pair* of electrons.

The bond in a hydrogen molecule, measured as the distance between the two nuclei, is about 7.4×10^{-11} m, or 74 picometers (pm; 1 pm = 1×10^{-12} m). This particular bond length represents a balance between several forces: the attractions between oppositely charged electrons and nuclei, the repulsion between two negatively charged electrons, and the repulsion between two positively charged nuclei. If the nuclei were closer together, they would repel each other more strongly; if the nuclei were farther apart, there would be less attraction between the positive and negative particles.

Fluorine is another element whose atoms bond together in pairs to form *diatomic* (two-atom) molecules. Two separate fluorine atoms have the following electron dot diagrams:

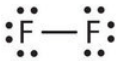


Each fluorine atom contributes one valence electron, making a single bond and giving each atom a complete valence shell, which fulfills the **octet** rule:



Two F's with two dots in between them, and two dots on the top, left/right side, and bottoms of each F.

The circles show that each fluorine atom has eight electrons around it. As with hydrogen, we can represent the fluorine molecule with a dash in place of the bonding electrons:



Two F's with a long dash in between them, and two dots on the top, left/right side, and bottoms of each F.

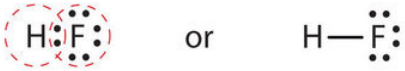
Each fluorine atom has six electrons, or three pairs of electrons, that are not participating in the covalent bond. Rather than being shared, they are considered to belong to a single atom. These are called nonbonding pairs (or lone pairs) of electrons.

Covalent Bonds between Different Atoms

Now that we have looked at electron sharing between atoms of the same element, let us look at covalent bond formation between atoms of different elements. Consider a molecule composed of one hydrogen atom and one fluorine atom:



Each atom needs one additional electron to complete its valence shell. By each contributing one electron, they make the following molecule:



An H and F connected by two dots, and the F has 2 dots on the top, right, and bottom. It also shows the same combination but with a long dash in between the H and F rather than two dots.

In this molecule, the hydrogen atom does not have nonbonding electrons, while the fluorine atom has six nonbonding electrons (three lone electron pairs). The circles show how the valence electron shells are filled for both atoms.

Example 3.2.1

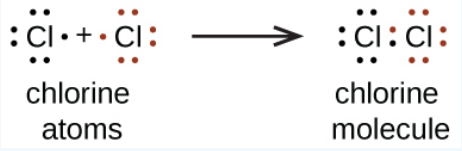
Use Lewis diagrams to indicate the formation of the following:

a. Cl_2
b. HBr

Solution

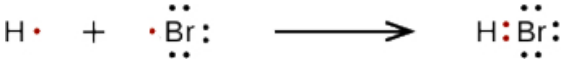
a. When two chlorine atoms form a chlorine molecule, they share one pair of electrons. In Cl_2 molecule, each chlorine atom is surrounded by an **octet** number of electrons.

The Lewis diagram for a Cl_2 molecule is similar to the one for F_2 (shown above).



Cl with 7 dots surrounding it plus Cl with 7 dots surrounding it means they are chlorine atoms. To turn into a chlorine molecule, have two dots in between each Cl and 6 dots surrounding each Cl on the sides that aren't connected.

b. When a hydrogen atom and a bromine atom form HBr, they share one pair of electrons. In the HBr molecule, H achieves a full valence of two electrons (**duet**) while Br achieves an **octet**. The Lewis diagram for HBr is similar to that for HF shown above.



H with one dot plus Br with 7 dots turns into H connected to Br by two dots, with 6 other dots surrounding Br.

? Exercise 3.2.1

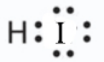
Draw the Lewis diagram for each compound.

a. a molecule composed of one chlorine atom and one fluorine atom
b. a molecule composed of one hydrogen atom and one iodine atom

Answer a:

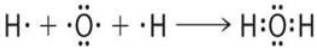


Answer b:



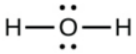
Covalent Bonds in Larger Molecules

The formation of a water molecule from two hydrogen atoms and an oxygen atom can be illustrated using Lewis dot symbols (shown below).



H with one dot + O with 6 dots + H with one dot turns into H connected to O with two dots connected to H with two dots. O also has two dots on top and top dots on bottom.

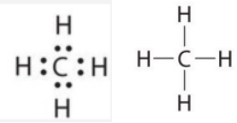
The structure on the right is the *Lewis electron structure*, or *Lewis structure*, for H_2O. With two bonding pairs and two lone pairs, the oxygen atom has now completed its **octet**. Moreover, by sharing a bonding pair with oxygen, each hydrogen atom now has a full valence shell of two electrons. Chemists usually indicate a bonding pair by a single line, as shown (below).



Water

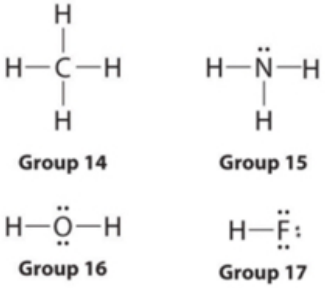
Other large molecules are constructed in a similar fashion, with some atoms participating in more than one covalent bond. For example, methane (CH_4), the central carbon atom bonded to four hydrogen atoms, can be represented using either of the Lewis

structures below. Again, sharing electrons between C and H atoms results in C achieving and octet while H achieving a duet number of electrons.



How Many Covalent Bonds Are Formed?

The number of bonds that an atom can form can often be predicted from the number of electrons needed to reach an octet (eight valence electrons). In the Lewis structure, the number of bonds formed by an element in a neutral compound is the same as the number of unpaired electrons it must share with other atoms to complete its octet of electrons. For example, each atom of a **group 4A** (14) element has four electrons in its outermost shell and therefore requires four more electrons to reach an octet. These four electrons can be gained by forming **four covalent bonds**, as illustrated here for carbon in CH_4 (methane). **Group 5A** (15) elements such as nitrogen have five valence electrons in the atomic Lewis symbol: one lone pair and three unpaired electrons. To obtain an octet, these atoms form **three covalent bonds**, as in NH_3 (ammonia). Oxygen and other atoms in **group 6A** (16) obtain an octet by forming **two covalent bonds**. Fluorine and the other halogens in **group 7A** (17) have seven valence electrons and can obtain an octet by forming **one covalent bond**.



Typically, the atoms of group 4A form 4 covalent bonds; group 5A form 3 bonds; group 6A form 2 bonds; and group 7A form one bond. The number of electrons required to obtain an octet determines the number of covalent bonds an atom can form. This is summarized in the table below. In each case, the sum of the number of bonds and the number of lone pairs is 4, which is equivalent to eight (octet) electrons.

This table shows atoms and their group numbers, and how many bonds and lone pairs each has.

Atom (Group number)	Number of Bonds	Number of Lone Pairs
Carbon (Group 14 or 4A)	4	0
Nitrogen (Group 15 or 5A)	3	1
Oxygen (Group 16 or 6A)	2	2
Fluorine (Group 17 or 7A)	1	3

Because hydrogen only needs two electrons to fill its valence shell, it follows the duet rule. It is an **exception to the octet rule.** Hydrogen only needs to form one bond. This is the reason why H is always a terminal atom and never a central atom. Figure **3.2.1** shows the number of covalent bonds various atoms typically form.

The transition elements and inner transition elements also do not follow the octet rule since they have d and f electrons involved in their valence shells.

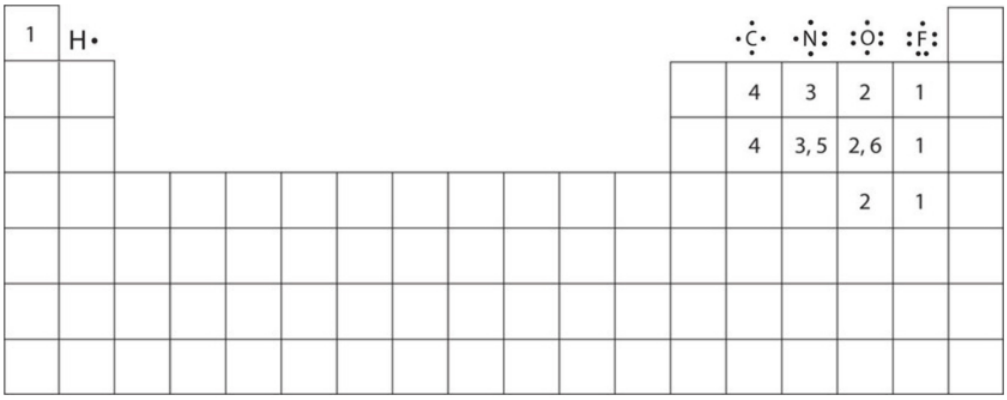
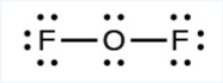


Figure **3.2.1**: How Many Covalent Bonds Are Formed? In molecules, there is a pattern to the number of covalent bonds that different atoms can form. Each block with a number indicates the number of covalent bonds formed by that atom in neutral compounds.

Example 3.2.2

Examine the Lewis structure of OF_2 below. Count the number of bonds formed by each element. Based on the element's location in the periodic table, does it correspond to the expected number of bonds shown in Table 4.1? Does the Lewis structure below follow the octet rule?



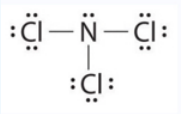
F surrounded by 6 dots, connected to O with a long dash. O surrounded by 4 dots and connects to another F with a long dash. F surrounded by 6 dots.

Solution

Yes. F (group 7A) forms one bond and O (group 6A) forms 2 bonds. Each atom is surrounded by 8 electrons. This structure satisfies the octet rule.

3.2.2 Exercise

Examine the Lewis structure of NCl_3 below. Count the number of bonds formed by each element. Based on the element's location in the periodic table, does it correspond to the expected number of bonds shown in Table 4.1? Does the Lewis structure below follow the octet rule?



Answer

Both Cl and N form the expected number of bonds. Cl (group 7A) has one bond and 3 lone pairs. The central atom N (group 5A) has 3 bonds and one lone pair. Yes, the Lewis structure of NCl_3 follows the octet rule.

Key Takeaways

- A covalent bond is formed between two atoms by sharing electrons.
- The number of bonds an element forms in a covalent compound is determined by the number of electrons it needs to reach octet.
- Hydrogen is an exception to the octet rule. H forms only one bond because it needs only two electrons.

3.2: Covalent Bonds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.1: Covalent Bonds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

3.3: Covalent Compounds - Formulas and Names

Learning Objectives

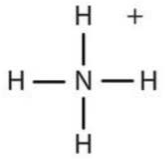
- Identify covalent and ionic compounds.
- Determine the chemical formula of a simple covalent compound from its name.
- Determine the name of a simple covalent compound from its chemical formula.

COVALENT AND IONIC COMPOUNDS

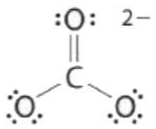
What elements make covalent bonds? Covalent bonds form when two or more nonmetals combine. For example, both hydrogen and oxygen are nonmetals, and when they combine to make water, they do so by forming covalent bonds. Compounds that are composed of only non-metals or semi-metals with non-metals will display covalent bonding and will be classified as molecular compounds.

As a general rule of thumb, compounds that involve a metal binding with either a non-metal or a semi-metal will display ionic bonding. Thus, the compound formed from sodium and chlorine will be ionic (a metal and a non-metal). Nitrogen monoxide (NO) will be a covalently bound molecule (two non-metals), silicon dioxide (SiO_2) will be a covalently bound molecule (a semi-metal and a non-metal) and $MgCl_2$ will be ionic (a metal and a non-metal).

A polyatomic ion is an ion composed of two or more atoms that have a charge as a group (poly = many). The ammonium ion (see figure below) consists of one nitrogen atom and four hydrogen atoms. Together, they comprise a single ion with a 1+ charge and a formula of NH_4^+. The carbonate ion (see figure below) consists of one carbon atom and three oxygen atoms and carries an overall charge of 2−. The formula of the carbonate ion is CO_3^{2-}.



ammonium ion



carbonate ion

The atoms of a polyatomic ion are tightly bonded together and so the entire ion behaves as a single unit. Several examples are found in Table 3.3.1. Nonmetal atoms in polyatomic ions are joined by covalent bonds, but the ion as a whole participates in ionic bonding. For example, ammonium chloride (NH_4Cl) has ionic bonding between a polyatomic ion, $\mathbf{NH_4^+}$, and $\mathbf{Cl^-}$ ions, but within the ammonium ion (NH_4^+), the nitrogen and hydrogen atoms are connected by covalent bonds (shown above).

Both ionic and covalent bonding are also found in calcium carbonate. Calcium carbonate ($CaCO_3$) has ionic bonding between calcium ion $\mathbf{Ca^{2+}}$ and a polyatomic ion, $\mathbf{CO_3^{2-}}$, but within the carbonate ion (CO_3^{2-}), the carbon and oxygen atoms are connected by covalent bonds (shown above).

Characteristics of Covalent (Molecular) Compounds

Compounds that contain covalent bonds (also called molecular compounds) exhibit different physical properties than ionic compounds. Because the attraction between molecules, which are electrically neutral, is weaker than that between electrically charged ions, covalent compounds generally have much lower melting and boiling points than ionic compounds (discussed in Section 3.6). For example, water (molecular compound) boils at 100 °C while sodium chloride (ionic compound) boils at 1413 °C. In fact, many covalent compounds are liquids or gases at room temperature, and, in their solid states, they are typically much softer than ionic solids. Furthermore, whereas ionic compounds are good conductors of electricity when dissolved in water, most covalent compounds, being electrically neutral, are poor conductors of electricity in any state. The attraction between molecules (called intermolecular forces) will be discussed in more detail in Section 8.1

Example 3.3.1

Is each compound formed from ionic bonds, covalent bonds, or both?

a. Na_2O
b. Na_3PO_4
c. N_2O_4

Answer a

The elements in Na_2O are a metal and a nonmetal, which form ionic bonds.

Answer b

Because sodium is a metal and we recognize the formula for the phosphate ion, we know that this compound is ionic. However, within the polyatomic phosphate ion, the atoms are held together by covalent bonds, so this compound contains both ionic and covalent bonds.

Answer c

The elements in N_2O_4| are both nonmetals, rather than a metal and a nonmetal. Therefore, the atoms form covalent bonds.

Exercise 3.3.1

Is each compound are formed from ionic bonds, covalent bonds, or both?

a. $Ba(OH)_2$
b. F_2
c. PCl_3

Answer a:

both

Answer b:

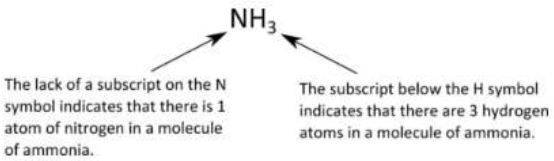
covalent

Answer c:

covalent

MOLECULAR FORMULAS

The chemical formulas for covalent compounds are referred to as **molecular formulas** because these compounds exist as separate, discrete molecules. Typically, a molecular formula begins with the nonmetal that is closest to the lower left corner of the periodic table, except that hydrogen is almost never written first (H_2O is the prominent exception). Then the other nonmetal symbols are listed. Numerical subscripts are used if there is more than one of a particular atom. For example, we have already seen CH_4, the molecular formula for methane. Below is the molecular formula of ammonia, NH_3.



NH_3. An arrow points to the N and says, "The lack of a subscript on the N symbol indicates that there is 1 atom of nitrogen in a molecule of ammonia". An arrow points to H and says, "The subscript below the H symbol indicates that there are 3 hydrogen atoms in a molecule of ammonia".

NAMING COVALENT COMPOUNDS

Naming *binary* (two-element) covalent compounds is similar to naming simple ionic compounds. The first element in the formula is simply listed using the name of the element. The second element is named by taking the stem of the element name and adding

the suffix *-ide*. A system of numerical prefixes is used to specify the number of atoms in a molecule. Table **3.3.1** lists these numerical prefixes. Normally, no prefix is added to the first element's name if there is only one atom of the first element in a molecule. If the second element is oxygen, the trailing vowel is usually omitted from the end of a polysyllabic prefix but not a monosyllabic one (that is, we would say "monoxide" rather than "monooxide" and "trioxide" rather than "troxide").

Table **3.3.1**: Numerical Prefixes for Naming Binary Covalent Compounds

Number of Atoms in Compound	Prefix on the Name of the Element
1	mono-*
2	di-
3	tri-
4	tetra-
5	penta-
6	hexa-
7	hepta-
8	octa-
9	nona-
10	deca-

*This prefix is not used for the first element's name.

Let us practice by naming the compound whose molecular formula is CCl_4. The name begins with the name of the first element—carbon. The second element, chlor*ine*, becomes chlor*ide*, and we attach the correct numerical prefix ("tetra-") to indicate that the molecule contains four chlorine atoms. Putting these pieces together gives the name *carbon tetrachloride* for this compound.

Example 3.3.2

Write the molecular formula for each compound.

a. chlorine trifluoride
b. phosphorus pentachloride
c. sulfur dioxide
d. dinitrogen pentoxide

Solution

If there is no numerical prefix on the first element's name, we can assume that there is only one atom of that element in a molecule.

a. ClF_3
b. PCl_5
c. SO_2
d. N_2O_5 (The *di-* prefix on nitrogen indicates that two nitrogen atoms are present.)

Exercise 3.3.2

Write the molecular formula for each compound.

a. nitrogen dioxide
b. dioxygen difluoride
c. sulfur hexafluoride

d. selenium monoxide

Answer a:

a. NO_2

Answer b:

O_2F_2

Answer c:

SF_6

Answer d:

SeO

Because it is so unreactive, sulfur hexafluoride is used as a spark suppressant in electrical devices such as transformers.

Example 3.3.3

Write the name for each compound.

a. BrF_5
b. S_2F_2
c. CO

Solution

a. bromine pentafluoride
b. disulfur difluoride
c. carbon monoxide

Exercise 3.3.3

Write the name for each compound.

a. CF_4
b. $SeCl_2$
c. SO_3

Answer a:

carbon tetrafluoride

Answer b:

selenium dichloride

Answer c:

sulfur trioxide

For some simple covalent compounds, we use common names rather than systematic names. We have already encountered these compounds, but we list them here explicitly:

- H_2O: water
- NH_3: ammonia
- CH_4: methane

Methane is the simplest organic compound. Organic compounds are compounds with carbon atoms and are named by a separate nomenclature system that we will introduce in in a separate section.

Key Takeaways

- The chemical formula of a simple covalent compound can be determined from its name.
- The name of a simple covalent compound can be determined from its chemical formula.

3.3: Covalent Compounds - Formulas and Names is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.2: Covalent Compounds - Formulas and Names** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

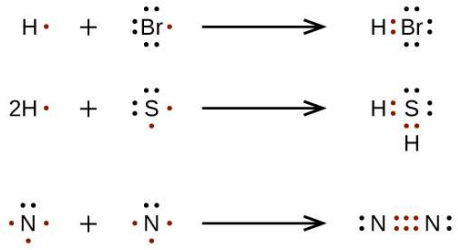
3.4: Drawing Lewis Structures

Learning Objectives

- To draw Lewis structures.
- To recognize molecules that are likely to have multiple covalent bonds.

DRAWING LEWIS STRUCTURES

For very simple molecules and molecular ions, we can write the Lewis structures by merely pairing up the unpaired electrons on the constituent atoms. See these examples:



For more complicated molecules and molecular ions, it is helpful to follow the step-by-step procedure outlined here:

1. Determine the total number of valence (outer shell) electrons among all the atoms. For cations, subtract one electron for each positive charge. For anions, add one electron for each negative charge.
2. Draw a skeleton structure of the molecule or ion, arranging the atoms around a central atom. (Generally, the least electronegative element should be placed in the center.) Connect each atom to the central atom with a single bond (one electron pair).
3. Distribute the remaining electrons as lone pairs on the terminal atoms (except hydrogen), completing an octet around each atom.
4. Place all remaining electrons on the central atom.
5. Rearrange the electrons of the outer atoms to make multiple bonds with the central atom in order to obtain octets wherever possible.

Let us determine the Lewis structures of OF_2 and HCN as examples in following this procedure:

1. Determine the total number of valence (outer shell) electrons in the molecule or ion. For a molecule, we add the number of valence electrons (use the main group number) on each atom in the molecule. This is the total number of electrons that must be used in the Lewis structure.

$$O + 2\,(F) = OF_2$$

$$6e^- + (2 \times 7e^-) = 20e^-$$

$$H + C + N = HCN$$

$$1e^- + 4e^- + 5e^- = 10e^-$$

2. Draw a skeleton structure of the molecule or ion, arranging the atoms around a central atom and connecting each atom to the central atom with a single (one electron pair) bond. Note that H and F can only form one bond, and are always on the periphery rather than the central atom.

F—O—F H—C—N

3. Distribute the remaining electrons as lone pairs on the terminal atoms (except hydrogen) to complete their valence shells with an octet of electrons.

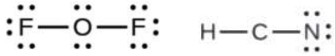
- In OF_2, six electrons are placed on each F.

- In HCN, six electrons placed on N



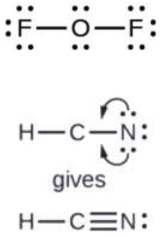
4. Place all remaining electrons on the central atom.

- In OF_2, 4 electrons are placed on O.
- In HCN: no electrons remain (the total valence of $10e^-$ is reached) so nothing changes.



5. Rearrange the electrons of the outer atoms to make multiple bonds with the central atom in order to obtain octets wherever possible.

- In OF_2, each atom has an octet as drawn, so nothing changes.
- In HCN, form two more C–N bonds

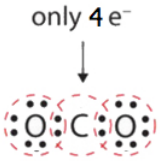


H dash C dash N with 6 dots around the N. 4 of the dots are pointing to the dash connecting the C and the N. This turns the formula into H dash C and three lines connecting the C and N together, with two dots on the outside of the N.

Finally, check to see if the total number of valence electrons are present in the Lewis structure. And then, inspect if the H atom has 2 electrons surrounding it and if each of the main group atoms is surrounded by 8 electrons.

MULTIPLE BONDS

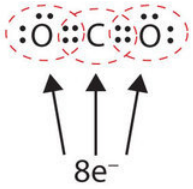
In many molecules, the octet rule would not be satisfied if each pair of bonded atoms shares only two electrons. Review HCN in Step 5 above. Another example is carbon dioxide (CO_2). CO_2 has a total valence of $4e^- + (2 \times 6e^-) = 16e^-$. Following steps 1 to 4, we draw the following:



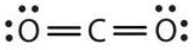
This does not give the carbon atom a complete octet; only four electrons are in its valence shell. This arrangement of shared electrons is far from satisfactory.



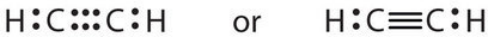
In this case, more than one pair of electrons must be shared between two atoms for both atoms to have an octet. A second electron pair from each oxygen atom must be shared with the central carbon atom shown by the arrows above. A lone pair from each O must be converted into a bonding pair of electrons.



In this arrangement, the carbon atom shares four electrons (two pairs) with the oxygen atom on the left and four electrons with the oxygen atom on the right. There are now eight electrons around each atom. Two pairs of electrons shared between two atoms make a double bond between the atoms, which is represented by a double dash:



Some molecules contain triple bonds (like HCN, shown above). Triple bonds are covalent bonds in which *three* pairs of electrons are shared by two atoms. Another compound that has a triple bond is acetylene (C_2H_2), whose Lewis diagram is as follows:



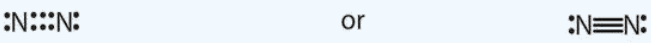
Example 3.4.1

Draw the Lewis diagram for each molecule.

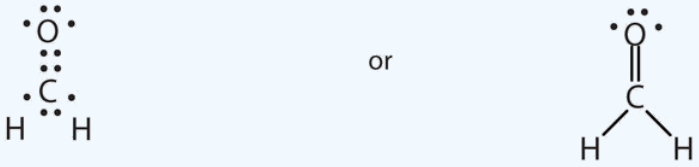
1. $\mathbf{N_2}$
2. $\mathbf{CH_2O}$ (The carbon atom is the central atom.) One application of CH_2O, also called formaldehyde, is the preservation of biological specimens. Aqueous solutions of CH_2O are called formalin and have a sharp, characteristic (pungent) odor.

Solution

1. The total number of electrons is 2 x 5 = 10 electrons. The bond between the two nitrogen atoms is a triple bond. The Lewis diagram for $N_2$is as follows:



2. The total number of electrons is 4 x 2(1) + 6 = 12 electrons. In CH_2O, the central atom is surrounded by two different types of atoms. The Lewis diagram that fills each atom's valence electron shell is as follows:

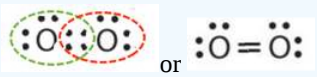


Exercise 3.4.1

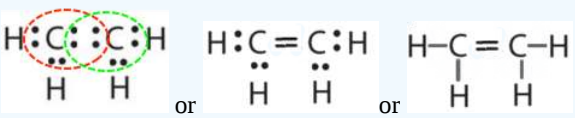
Draw the Lewis diagram for each molecule.

a. $\mathbf{O_2}$
b. $\mathbf{C_2H_4}$

Answer a:



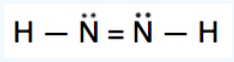
Answer b:

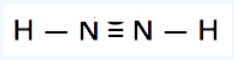


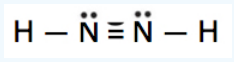
https://chem.libretexts.org/@go/page/255584

Example 3.4.2

Which is the correct Lewis structure for N_2H_2?

A. 

B. 

C. 

Solution

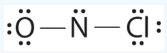
Lewis structure A is the correct answer. It has a total of (2 x $5e^-$) + (2 x $1e^-$) = $12e^-$. Each of the N atoms satisfy the octet requirement and the H atoms follow the duet rule.

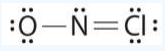
Structure B is electron deficient. It has only $10e^-$ instead of 12.

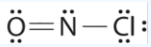
Structure C has 14 (2 extra) electrons. The N atoms do not satisfy the octet.

3.4.2 Exercise

Which is the correct Lewis structure for NOCl?

A. 

B. 

C. 

Answer

Structure A violates the octet rule; N is surrounded by only $6e^-$.

Structure B violates the octet rule; Cl has $10e^-$ around it. Furthermore, there are a total of $20e^-$ instead of $18e^-$.

Structure C is the correct structure. It has a total of $6e^-$ + $5e^-$ + $7e^-$ = $18e^-$. Each atom is surrounded by 8 electrons (octet rule).

Key Takeaways

- A Lewis structure shows the bonding and nonbonding electrons around individual atoms in a molecule.
- Some molecules must have multiple covalent bonds between atoms to satisfy the octet rule.
- A double bond contains four electrons and a triple bond contains six electrons.

3.4: Drawing Lewis Structures is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.3: Drawing Lewis Structures** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

3.5: Characteristics of Covalent Bonds

Learning Objectives

- Recognize bond characteristics of covalent compounds: bond length and bond polarity.
- Use electronegativity values to predict bond polarity.

Bond Length

We previously stated that the covalent bond in the hydrogen molecule (H_2) has a certain length (about 7.4×10^{-11} m). Other covalent bonds also have known bond lengths, which are dependent on both the identities of the atoms in the bond and whether the bonds are single, double, or triple bonds. Table **3.5.1** lists the approximate bond lengths for some single covalent bonds. The exact bond length may vary depending on the identity of the molecule but will be close to the value given in the table.

Table **3.5.1**: Approximate Bond Lengths of Some Single Bonds

Bond	Length ($\times 10^{-12}\ m$)
H–H	74
H–C	110
H–N	100
H–O	97
H–I	161
C–C	154
C–N	147
C–O	143
N–N	145
O–O	145

Table **3.5.2** compares the lengths of single covalent bonds with those of double and triple bonds between the same atoms. Without exception, as the number of covalent bonds between two atoms increases, the bond length decreases. With more electrons between the two nuclei, the nuclei can get closer together before the internuclear repulsion is strong enough to balance the attraction.

Table **3.5.2**: Comparison of Bond Lengths for Single and Multiple Bonds

Bond	Length ($\times 10^{-12}\,m$)
C–C	154
C=C	134
C≡C	120
C–N	147
C=N	128
C≡N	116
C–O	143
C=O	120
C≡O	113
N–N	145
N=N	123
N≡N	110
O–O	145
O=O	121

Electronegativity and Bond Polarity

Although we defined covalent bonding as electron sharing, the electrons in a covalent bond are not always shared equally by the two bonded atoms. Unless the bond connects two atoms of the same element, as in H_2, there will always be one atom that attracts the electrons in the bond more strongly than the other atom does, as in HCl, shown in Figure **3.5.1**. A covalent bond that has an equal sharing of electrons (Figure **3.5.1***a*) is called a **nonpolar covalent bond**. A covalent bond that has an unequal sharing of electrons, as in Figure **3.5.1***b*, is called a **polar covalent bond**.

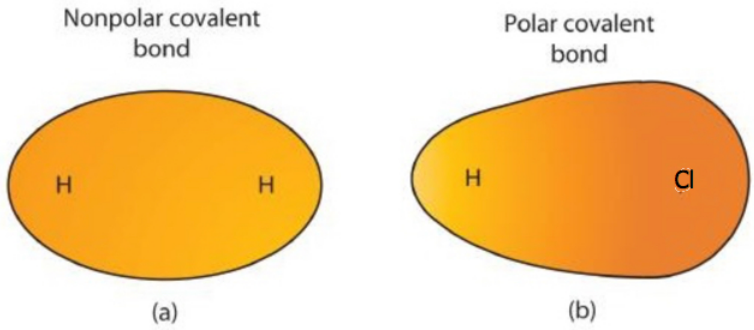


Figure **3.5.1** Polar versus Nonpolar Covalent Bonds. (a) The electrons in the covalent bond are equally shared by both hydrogen atoms. This is a nonpolar covalent bond. (b) The chlorine atom attracts the electrons in the bond more than the hydrogen atom does, leading to an imbalance in the electron distribution. This is a polar covalent bond.

The distribution of electron density in a polar bond is uneven. It is greater around the atom that attracts the electrons more than the other. For example, the electrons in the H–Cl bond of a hydrogen chloride molecule spend more time near the chlorine atom than near the hydrogen atom. Note that the shaded area around Cl in Figure **3.5.1***b* is much larger than it is around H.

This imbalance in electron density results in a buildup of partial negative charge (designated as δ−) on one side of the bond (Cl) and a partial positive charge (designated δ+) on the other side of the bond (H). This is seen in Figure **3.5.2***a*. The separation of charge in a polar covalent bond results in an electric dipole (two poles), represented by the arrow in Figure **3.5.2***b*. The direction of the arrow is pointed toward the δ− end while the + tail of the arrow indicates the δ+ end of the bond.

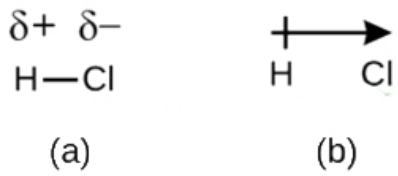


Figure **3.5.2**: (a) Unequal sharing of the bonding pair of electrons between H and Cl leads to partial positive charge on the H atom and partial negative charge on the Cl. Symbols δ+ and δ– indicate the polarity of the H–Cl bond. (b) The dipole is represented by an arrow with a cross at the tail. The cross is near the δ+ end and the arrowhead coincides with the δ–.

Any covalent bond between atoms of different elements is a polar bond, but the degree of polarity varies widely. Some bonds between different elements are only minimally polar, while others are strongly polar. Ionic bonds can be considered the ultimate in polarity, with electrons being transferred rather than shared. To judge the relative polarity of a covalent bond, chemists use **electronegativity**, which is a relative measure of how strongly an atom attracts electrons when it forms a covalent bond. There are various numerical scales for rating electronegativity. Figure **3.5.3** shows one of the most popular—the Pauling scale.

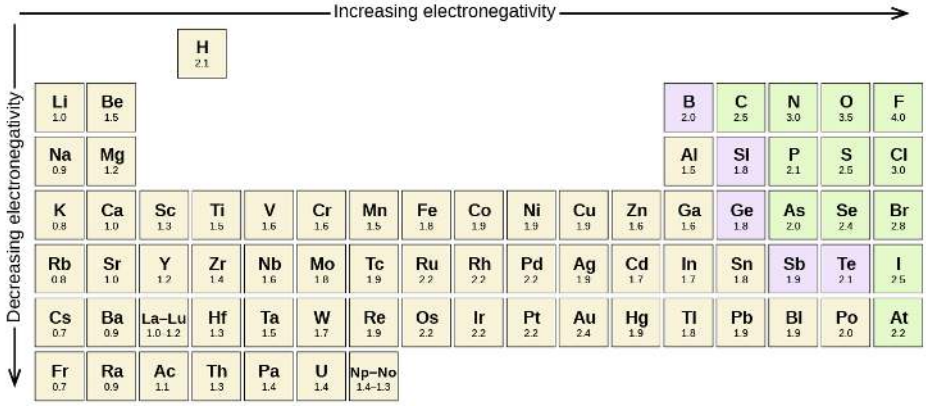
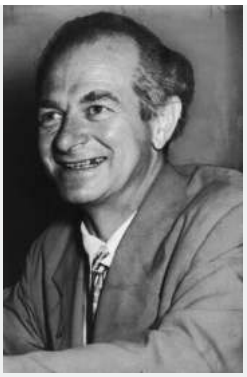


Figure **3.5.3** The electronegativity values derived by Pauling follow predictable periodic trends with the higher electronegativities toward the upper right of the periodic table. Fluorine has the highest value (4.0).

Looking Closer: Linus Pauling

Arguably the most influential chemist of the 20th century, Linus Pauling (1901–94) is the only person to have won two individual (that is, unshared) Nobel Prizes. In the 1930s, Pauling used new mathematical theories to enunciate some fundamental principles of the chemical bond. His 1939 book *The Nature of the Chemical Bond* is one of the most significant books ever published in chemistry.

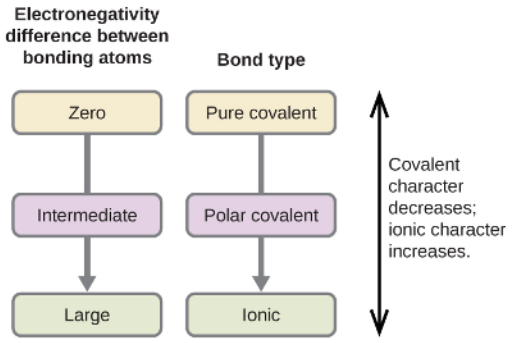
By 1935, Pauling's interest turned to biological molecules, and he was awarded the 1954 Nobel Prize in Chemistry for his work on protein structure. (He was very close to discovering the double helix structure of DNA when James Watson and James Crick announced their own discovery of its structure in 1953.) He was later awarded the 1962 Nobel Peace Prize for his efforts to ban the testing of nuclear weapons.



Linus Pauling was one of the most influential chemists of the 20th century.

In his later years, Pauling became convinced that large doses of vitamin C would prevent disease, including the common cold. Most clinical research failed to show a connection, but Pauling continued to take large doses daily. He died in 1994, having spent a lifetime establishing a scientific legacy that few will ever equal.

The polarity of a covalent bond can be judged by determining the difference in the electronegativities of the two atoms making the bond. The greater the difference in electronegativities, the greater the imbalance of electron sharing in the bond. Although there are no hard and fast rules, the general rule is if the difference in electronegativities is **less than about 0.4,** the bond is considered **nonpolar;** if the difference is **greater than 0.4,** the bond is considered **polar.** If the difference in electronegativities is large enough **(**generally **greater than about 1.8),** the resulting compound is considered **ionic** rather than covalent**.** An electronegativity difference of **zero,** of course, indicates a **nonpolar covalent bond.**



Bond Type	Electronegativity Difference
pure covalent	< 0.4
polar covalent	between 0.4 and 1.8
ionic	> 1.8

Figure **3.5.4**: As the electronegativity difference increases between two atoms, the bond becomes more ionic.

Example 3.5.1

Describe the electronegativity difference between each pair of atoms and the resulting polarity (or bond type).

a. C and H
b. H and H
c. Na and Cl
d. O and H

Solution

a. Carbon has an electronegativity of 2.5, while the value for hydrogen is 2.1. The difference is 0.4, which is rather small. The C–H bond is therefore considered nonpolar.
b. Both hydrogen atoms have the same electronegativity value—2.1. The difference is zero, so the bond is nonpolar.
c. Sodium's electronegativity is 0.9, while chlorine's is 3.0. The difference is 2.1, which is rather high, and so sodium and chlorine form an ionic compound.
d. With 2.1 for hydrogen and 3.5 for oxygen, the electronegativity difference is 1.4. We would expect a very polar bond. The sharing of electrons between O and H is unequal with the electrons more strongly drawn towards O.

? Exercise 3.5.1

Describe the electronegativity (EN) difference between each pair of atoms and the resulting polarity (or bond type).

a. C and O
b. K and Br
c. N and N
d. Cs and F

Answer a:

The EN difference is 1.0 , hence polar. The sharing of electrons between C and O is unequal with the electrons more strongly drawn towards O.

Answer b:

The EN difference is greater than 1.8, hence ionic.

Answer c:

Identical atoms have zero EN difference, hence nonpolar.

Answer d:

The EN difference is greater than 1.8, hence ionic.

Bond Polarity and Molecular Polarity

If there is only one bond in the molecule, the bond polarity determines the molecular polarity. Any diatomic molecule in which the two atoms are the same element must be a **nonpolar** molecule. A diatomic molecule that consists of a polar covalent bond, such as HF, is a **polar molecule**. A **polar molecule** *is a molecule in which one end of the molecule is slightly positive, while the other end is slightly negative*. The two electrically charged regions on either end of the molecule are called poles, similar to a magnet having a north and a south pole. Hence, a *molecule with two poles is called a* **dipole**. A simplified way to depict polar molecules like HF is pictured below (see figure below).

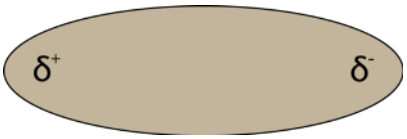


Figure 3.5.5: A molecular dipole results from the unequal distribution of electron density throughout a molecule.

When placed between oppositely charged plates, polar molecules orient themselves so that their positive ends are closer to the negative plate and their negative ends are closer to the positive plate (see Figure 4.4.6 below).

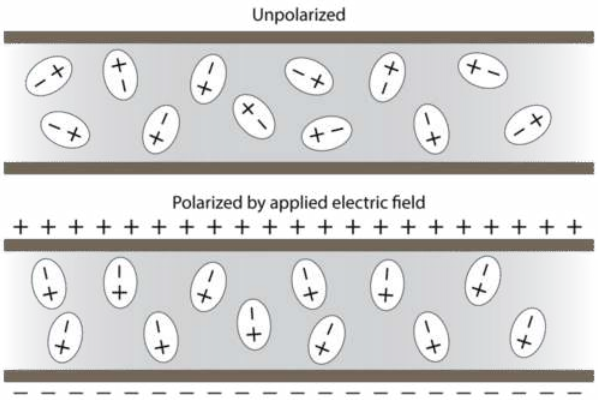


Figure 3.5.6: Polar molecules are randomly oriented in the absence of an applied electric field (top). In an electric field, the molecules orient themselves to maximize the attraction between opposite charges (bottom).

Experimental techniques involving electric fields can be used to determine if a certain substance is composed of polar molecules and to measure the degree of polarity.

For molecules with more than two atoms, the **molecular geometry** must also be taken into account when determining if the molecule is polar or nonpolar. (Figure **3.5.7**) is a comparison between carbon dioxide and water. Carbon dioxide (CO_2) is a **linear** molecule. The oxygen atoms are more electronegative than the carbon atom, so there are two individual dipoles pointing outward from the C atom to each O atom. However, since the dipoles are of equal strength and are oriented directly opposite each other, they cancel each other out, and the overall molecular polarity of CO_2 is zero. CO_2 is a **nonpolar** molecule. In H_2O, the orientation of the two O–H bonds is **bent**: one end of the molecule has a partial positive charge, and the other end has a partial negative charge. In short, the molecule itself is **polar**. The polarity of water has an enormous impact on its physical and chemical properties. For example, the boiling point of water (100°C) is high for such a small molecule due to the fact that polar molecules attract each other strongly. On the other hand, the nonpolar carbon dioxide becomes a gas at −77°C, almost 200° lower than the temperature at which water boils. More details will be presented in the next section.



Figure **3.5.7** Physical Properties and Polarity. The physical properties of water and carbon dioxide are affected by their polarities.

Key Takeaways

- Covalent bonds between different atoms have different bond lengths.
- Covalent bonds can be polar or nonpolar, depending on the electronegativity difference between the atoms involved.

3.5: Characteristics of Covalent Bonds is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.4: Characteristics of Covalent Bonds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

3.6: Characteristics of Molecules

Learning Objectives

- Determine the molecular mass of a molecule.
- Predict the general shape of a simple covalent molecule.
- Predict the polarity of the molecule.
- Compare the properties of ionic and molecular compounds.

Unlike the ions in ionic compounds, which are arranged in lattices called crystals, molecules of covalent compounds exist as discrete units with a characteristic mass and a certain three-dimensional shape.

Molecular Mass

The mass of a molecule—the molecular mass (sometimes called the molecular weight)—is simply the sum of the masses of its atoms. As with formula masses, it is important that you keep track of the number of atoms of each element in the molecular formula to obtain the correct molecular mass.

Example 3.6.1

What is the molecular mass of each covalent compound?

a. H_2O
b. C_6H_6
c. NO_2
d. N_2O_5

Solution

Use the atomic masses from the Periodic Table in Section 2.7.

The molecular formula H_2O indicates that there are two hydrogen atoms and one oxygen atom in each molecule. Summing the masses of these atoms,

This table finds the mass of atoms.

2 H:	2 × 1.01 =	2.02 amu
1 O:		+ 16.00 amu
Total:		18.02 amu

The molecular mass of H_2O is **18.02** amu.

This table finds the mass of atoms.

6 C:	6 × 12.01 =	72.06 amu
6 H:	6 × 1.01 =	+ 6.06 amu
Total:		78.12 amu

The molecular mass of C_6H_6 is **78.12** amu.

This table finds the mass of atoms.

1 N:		14.01 amu
2 O:	2 × 16.00 =	+ 32.00 amu
Total:		46.01 amu

The molecular mass of NO_2 is **46.01** amu.

This table finds the mass of atoms.

2 N:	2 × 14.01 =	28.02 amu
5 O:	5 × 16.00 =	+ 80.00 amu
Total:		108.02 amu

The molecular mass of N_2O_5 is **108.02** amu.

Note that the two different nitrogen and oxygen compounds in these examples have different molecular masses.

? Exercise 3.6.1

What is the molecular mass of each covalent compound?

a. C_2H_2
b. CO
c. CO_2
d. BF_3

Answer a:

26.04 amu

Answer b:

28.01 amu

Answer c:

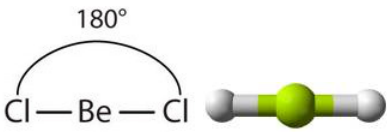
44.01 amu

Answer d:

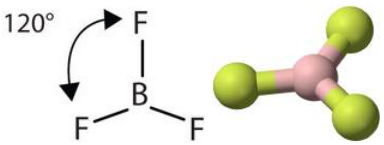
67.81 amu

Molecular Shape: VSEPR Theory

Unlike ionic compounds, with their extended crystal lattices, covalent molecules are discrete units with specific three-dimensional shapes. The shape of a molecule is determined by the fact that covalent bonds, which are composed of negatively charged electrons, tend to repel one another. This concept is called the **valence shell electron pair repulsion (VSEPR) theory**. For example, the two covalent bonds in $BeCl_2$ **stay as far from each other** as possible, ending up 180° apart from each other. The result is a ***linear*** molecule:



The three covalent bonds in BF_3 repel each other to form 120° angles in a plane, in a shape called ***trigonal planar***:



The molecules $BeCl_2$ and BF_3 actually violate the octet rule; however, such exceptions are rare and will not be discussed in this text.

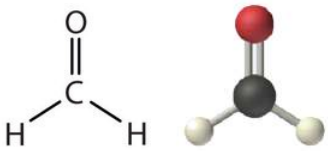
> Try sticking three toothpicks into a marshmallow or a gumdrop and see if you can find different positions where your "bonds" are farther apart than the planar 120° orientation.

The four covalent bonds in CCl_4 arrange themselves three dimensionally, pointing toward the corner of a tetrahedron and making bond angles of 109.5°. CCl_4 is said to have a ***tetrahedral*** shape:

Table describes how to use the number of atoms around the central atom to determine the geometry of a molecule, as well as examples.

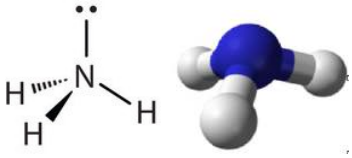
Atoms Around Central Atom	Geometry	Example
2 $\mathbf{AB_2}$	Linear	$\mathbf{BeCl_2}$
3 $\mathbf{AB_3}$	Trigonal Planar	$\mathbf{BF_3}$
4 $\mathbf{AB_4}$	Tetrahedral	$\mathbf{CCl_4}$

In determining the shapes of molecules, it is useful to first determine the Lewis diagram for a molecule. The shapes of molecules with multiple bonds are determined by treating the multiple bonds as one bond. Thus, formaldehyde (CH_2O) has a shape similar to that of BF_3. It is ***trigonal planar***.

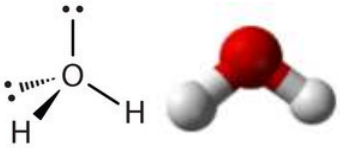


Molecules With Lone Pairs Around Central Atom

Molecules with lone electron pairs around the central atom have a shape based on the position of the atoms, not the electron pairs. For example, NH_3 has **one lone electron pair** and **three bonded electron pairs**. These four electron pairs repel each other and adopt a tetrahedral arrangement. However, the shape of the molecule is described in terms of the positions of the atoms, not the lone electron pairs. Thus, NH_3 is said to have a ***trigonal pyramidal*** shape, not a tetrahedral one.



Similarly, H_2O has **two lone pairs** of electrons around the central oxygen atom and **two bonded electron pairs**. Although the four electron pairs adopt a tetrahedral arrangement, the shape of the molecule is described by the positions of the atoms only. The shape of H_2O is ***bent*** with an approximate 109.5° angle.

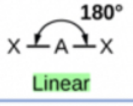
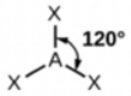
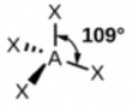


In summary, to determine the molecular geometry:

Step 1: Draw the Lewis structure.

Step 2: Count the number of bonds (a double/triple bond counts as one) and lone pairs around the **central atom**.

Step 3: Use Table 4.5.1 to determine the molecular geometry.

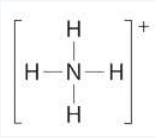
Number of bonds and lone pairs	Number of bonds 0 lone pair	1 lone pair	2 lone pairs
2	 Linear		
3	 Trigonal planar	 Bent or angular	Lone Pair
4	 Tetrahedral	 Trigonal pyramid	 Bent or angular

If a molecule has 2 bonds and 0 lone pairs, it is linear. If a molecule has 3 bonds and 0 lone pairs, it is trigonal planar. If a molecule has 3 bonds and 1 lone pair, it is bent or angular. If a molecule has 4 bonds and 0 lone pairs, it is tetrahedral. If a molecule has 4 bonds and 1 lone pair, it is trigonal pyramid. If a molecule has 4 bonds and 2 lone pairs, it is bent or angular.

Table 4.5.1: *The molecular geometry depends on the number of bonds and lone pairs around the* **central atom, A.**

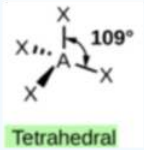
Example 3.6.1

What is the geometry of the ammonium ion, NH_4^+? Its Lewis structure is shown below. How is this different from ammonia, NH_3?

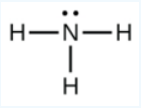


Solution

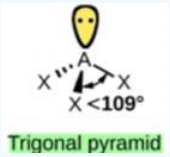
In ammonium ion, the central atom N has 4 bonds and no lone pair. It is equivalent to the below in Table 4.5.1. Hence, this is *tetrahedral*.



In ammonia (NH_3), shown below, N has 3 bonds and one lone pair.

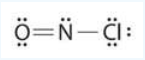


It is equivalent to the below in Table 4.5.1. Hence, the shape of this molecule is *trigonal pyramid*.



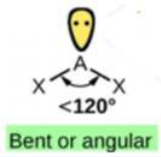
3.6.1 Exercise

What is the molecular shape of nitrosyl chloride, a highly corrosive, reddish-orange gas? Its Lewis structure is shown below.



Answer

Focus on the central atom, N. It has a double bond to O, count this as one bond. It also has a single bond to Cl. Thus, N has 2 bonds and one lone pair. These 3 electron pairs will spread out 120 degrees from each other. But, since the shape is defined by the arrangement of the atoms only, the shape is bent or angular. If you consult Table 4.5.1, this molecule is equivalent to the below. Hence, two bonds and one lone pair has a *bent or angular* shape.



Molecular Polarity

In general, a molecule is nonpolar if all its bonds are nonpolar. Examples are I_2, O_2, H_2, CH_4, C_2H_6 and C_3H_8.

In general, a molecule is polar if it contains polar bonds EXCEPT when the bond polarities cancel each other. As mentioned in Section 4.4, the shape of the CO_2 molecule (linear) orients the two C=O polar bonds **directly opposite** each other, thus **cancelling** each other's effect. **Carbon dioxide** (CO_2) is a ***nonpolar*** molecule.

On the other hand, water (as discussed above) is a bent molecule because of the two lone pairs on the central oxygen atom. Because of the bent shape, the dipoles do not cancel each other out and the water molecule is ***polar***. In the figure below, the individual H-O polar bonds represented by the two red arrows are not directly opposite each other. These two dipoles don't cancel each other out. In fact, the net dipole (blue arrow) points upward. There is a resultant partial positive charge at one end (between the two H atoms) and a partial negative charge on the other end (where O is located). The uneven distribution of charge or the overall dipole is shown by the blue arrow below (Figure 4.5.1). Hence, **water is *polar*** (has + and - poles) while **carbon dioxide is *nonpolar***.

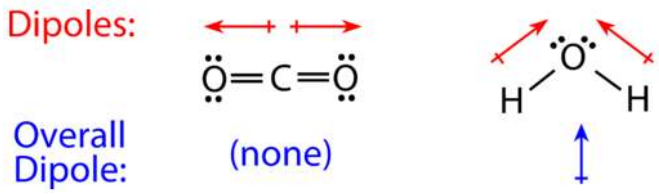


Figure 4.5.1: The molecular geometry of a molecule (linear vs. bent) affects its polarity.

Similarly, in BF_3 (trigonal planar), the effect of a B-F bond is cancelled by the sum of the other two B-F bonds (see video). Hence, a trigonal planar molecule (BF_3) is nonpolar because the bond polarities cancel each other, but a trigonal pyramidal molecule (NH_3) is polar.

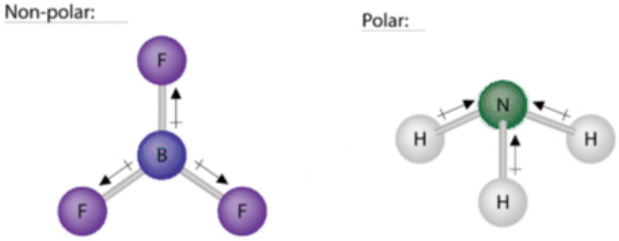


Figure 4.5.2: The molecular geometry of a molecule (trigonal planar vs. trigonal pyramid) affects its polarity.

Some other molecules are shown in the figure below. Notice that a tetrahedral molecule such as CCl_4 is ***nonpolar***. However, if the peripheral atoms are not of the same electronegativity, the bond polarities don't cancel and the molecule becomes ***polar***, as in CH_3Cl.

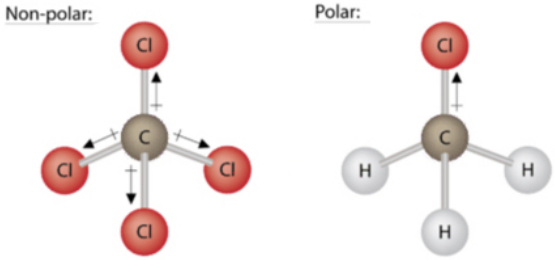


Figure 4.5.3: *The same molecular geometry but peripheral bonds are of different electronegativity. CCl_4 is nonpolar but CH_3Cl is polar.*

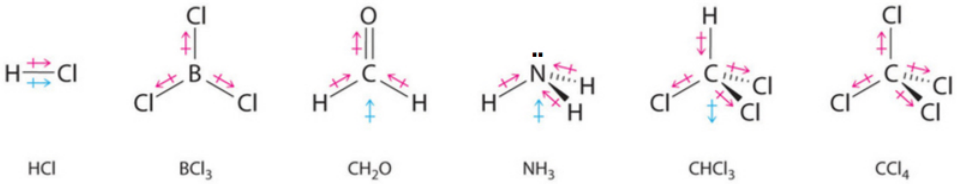


Figure **3.6.4**: Molecules with Polar Bonds. Individual bond dipole moments are indicated in red. Due to their different three-dimensional geometry, some molecules with polar bonds have a net dipole moment (HCl, CH_2O, NH_3, and $CHCl_3$), indicated in blue, whereas others do not because the bond dipoles cancel due to symmetry (BCl_3 and CCl_4).

Physical Properties of Molecular Compounds

The physical state and properties of a particular compound depend in large part on the type of chemical bonding it displays. **Molecular compounds**, sometimes called covalent compounds, display a wide range of physical properties due to the different types of intermolecular attractions such as different kinds of polar interactions. The melting and boiling points of molecular compounds are generally quite low compared to those of **ionic compounds**. This is because the energy required to disrupt the **intermolecular forces** (discussed in Section 8.1) between molecules is far less than the energy required to break the ionic bonds in a crystalline ionic compound. Since molecular compounds are composed of neutral molecules, their electrical conductivity is generally quite poor, whether in the solid or liquid state. Ionic compounds do not conduct electricity in the solid state because of their rigid structure, but conduct well when either molten or dissolved into a solution. The water solubility of molecular compounds is variable and depends primarily on the type of intermolecular forces involved. Substances that exhibit hydrogen bonding or dipole-dipole forces are generally water soluble, whereas those that exhibit only London dispersion forces are generally insoluble. Most, but not all, ionic compounds are quite soluble in water. The table below summarizes some of the differences between ionic and molecular compounds.

Table 4.5.2: Comparison of Ionic and Molecular Compounds

Property	Ionic Compounds	Molecular Compounds
Type of elements	Metal and nonmetal	Nonmetals only
Bonding	Ionic - transfer of electron(s) between atoms	Covalent - sharing of pair(s) of electrons between atoms
Representative unit	Formula unit	Molecule
Physical state at room temperature	Solid	Gas, liquid, or solid
Water solubility	Usually high	Variable
Melting and boiling temperatures	Generally high	Generally low
Electrical conductivity	Good when molten or in solution	Poor

In summary, covalent compounds are softer, have lower boiling and melting points, are more flammable, are less soluble in water and do not conduct electricity compared to ionic compounds. The individual melting and boiling points, solubility and other physical properties of molecular compounds depend on molecular polarity.

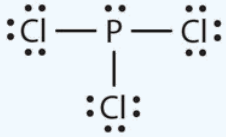
Example 3.6.2

Describe the shape of each molecule. Is it polar or nonpolar?

a. PCl_3
b. CO_2

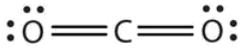
Solution

a. The Lewis diagram for PCl_3 is as follows:



Focus on the central atom, P that has 3 bonds and one lone pair. The four electron pairs arrange themselves tetrahedrally, but the lone electron pair is not considered in describing the molecular shape. Like NH_3, this molecule is *pyramidal*. The 3 P-Cl bonds don't cancel each other. This is *polar.*

- The Lewis diagram for CO_2 is as follows:



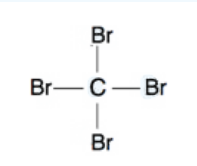
Focus on the central atom, C. The multiple bonds are treated as one group, hence C has 2 bonds and zero lone pair. CO_2 has only two groups of electrons that repel each other. They will direct themselves 180° apart from each other, so CO_2 molecules are *linear*. This is highly symmetrical, with the two opposite dipoles cancelling each other. The CO_2 molecule is *nonpolar*.

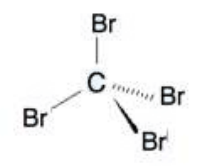
Exercise 3.6.2

Describe the shape of each molecule. Is it polar or nonpolar?

a. CBr_4
b. BCl_3

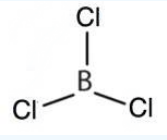
Answer a:



or


The Lewis structure shows 4 groups attached to the central atom, hence *tetrahedral*. All the 4 groups are identical and the shape is symmetrical. Hence, it is *nonpolar.*

Answer b:



The Lewis diagram shows 3 groups attached to the central atom, hence *trigonal planar*. All the 3 groups are identical and shape is symmetrical, hence, it is *nonpolar.*

Key Takeaways

- A molecule has a certain mass, called the molecular mass.
- Simple molecules have geometries that can be determined from VSEPR theory.
- Polar molecules result from differences in electronegativity of the atoms in the molecule.
- Dipoles that are directly opposite one another cancel each other out.

3.6: Characteristics of Molecules is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.5: Characteristics of Molecules** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

3.7: Organic Chemistry

Learning Objectives

- Define *organic chemistry*.
- Identify organic molecules as alkanes, alkenes, alkynes, alcohols, or carboxylic acids.

When methane was mentioned previously, we described it as the simplest organic compound. In this section, we introduce organic chemistry more formally. Organic chemistry is the study of the chemistry of carbon compounds. Carbon is singled out because it has a chemical diversity unrivaled by any other chemical element. Its diversity is based on the following:

- Carbon atoms bond reasonably strongly with other carbon atoms.
- Carbon atoms bond reasonably strongly with atoms of other elements.
- Carbon atoms make a large number of covalent bonds (four).

Curiously, elemental carbon is not particularly abundant. It does not even appear in the list of the most common elements in Earth's crust. Nevertheless, all living things consist of organic compounds. Most organic chemicals are covalent compounds, which is why we introduce organic chemistry here. By convention, compounds containing carbonate ions and bicarbonate ions, as well as carbon dioxide and carbon monoxide, are not considered part of organic chemistry, even though they contain carbon.

Alkanes

The simplest organic compounds are the **hydrocarbons**, compounds composed of carbon and hydrogen atoms only. Some hydrocarbons have only single bonds and appear as a chain (which can be a straight chain or can have branches) of carbon atoms also bonded to hydrogen atoms. These hydrocarbons are called **alkanes (saturated hydrocarbons)**. Each alkane has a characteristic, systematic name depending on the number of carbon atoms in the molecule. These names consist of a stem that indicates the number of carbon atoms in the chain plus the ending -ane. The stem meth- means one carbon atom, so methane is an alkane with one carbon atom. Similarly, the stem eth- means two carbon atoms; ethane is an alkane with two carbon atoms. Continuing, the stem prop- means three carbon atoms, so propane is an alkane with three carbon atoms. The stem but- means four carbon atoms; butane is an alkane with four carbon atoms. Figure **3.7.1** gives the Lewis structures, condensed structural formulas and molecular formulas of the four simplest alkanes. In the **condensed structural formula**, the covalent bonds are understood to exist between each carbon and the hydrogens associated with it, as well as between carbon atoms.

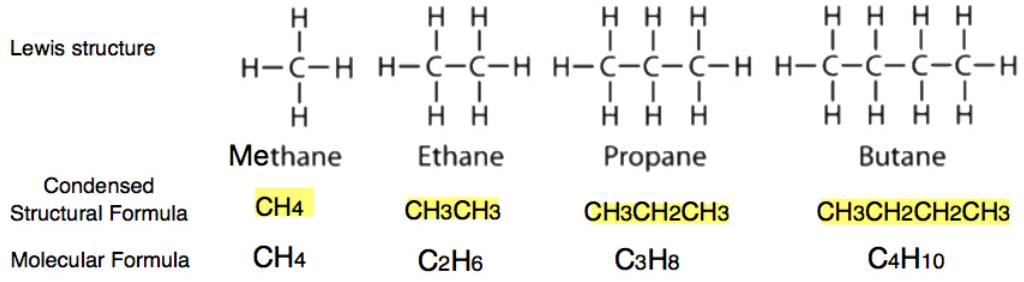


Figure **3.7.1**: Formulas and Molecular Models of the Four Simplest Alkanes. The four smallest alkanes are methane, ethane, propane and butane.

Alkenes

Some hydrocarbons have one or more carbon–carbon double bonds (denoted C=C). These hydrocarbons are called alkenes. Figure **3.7.2** shows the formulas and the molecular models of the two simplest alkenes. Note that the names of alkenes have the same stem as the alkane with the same number of carbon atoms in its chain but have the ending *-ene*. Thus, ethene is an alkene with two carbon atoms per molecule, and propene is a compound with three carbon atoms and one double bond.

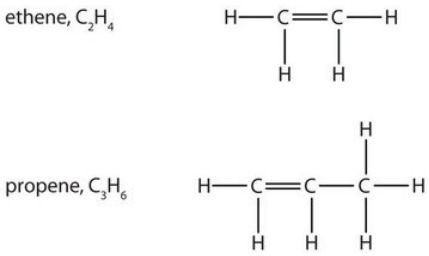
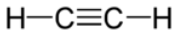


Figure **3.7.2**: Formulas and Molecular Models of the Two Simplest Alkenes. Ethene is commonly called ethylene, while propene is commonly called propylene.

Alkynes

Alkynes are hydrocarbons with a carbon–carbon triple bond (denoted C≡C) as part of their carbon skeleton. Below is the formula and the molecular model of the simplest alkyne and its systematic name. Its common name is acetylene. Its chemical formula is $\mathbf{C_2H_2}$.



Ethyne

The names for alkynes have the same stems as for alkanes but with the ending *-yne*.

To Your Health: Saturated and Unsaturated Fats

Hydrocarbons are not the only compounds that can have carbon–carbon double bonds. A group of compounds called fats can have them as well, and their presence or absence in the human diet is becoming increasingly correlated with health issues.

Fats are combinations of long-chain organic compounds (fatty acids) and glycerol ($C_3H_8O_3$). (For more information on fats, see Chapter 17) The long carbon chains can have either all single bonds, in which case the fat is classified as *saturated*, or one or more double bonds, in which case it is a *monounsaturated* or a *polyunsaturated* fat, respectively. Saturated fats are typically solids at room temperature; beef fat (tallow) is one example. Mono- or polyunsaturated fats are likely to be liquids at room temperature and are often called oils. Olive oil, flaxseed oil, and many fish oils are mono- or polyunsaturated fats.

Studies have linked higher amounts of saturated fats in people's diets with a greater likelihood of developing heart disease, high cholesterol, and other diet-related diseases. In contrast, increases in unsaturated fats (either mono- or polyunsaturated) have been linked to a lower incidence of certain diseases. Thus, there have been an increasing number of recommendations by government bodies and health associations to decrease the proportion of saturated fat and increase the proportion of unsaturated fat in the diet. Most of these organizations also recommend decreasing the total amount of fat in the diet.

Recently, certain fats called trans fats have been implicated in the presence of heart disease. These are fats from animal sources and are also produced when liquid oils are exposed to partial hydrogenation, an industrial process that increases their saturation. Trans fats are used in many prepared and fried foods. Because they bring with them the health risks that naturally occurring saturated fats do, there has been some effort to better quantify the presence of trans fats in food products. US law now requires that food labels list the amount of trans fat in each serving.

Since the early 1900's, the US Department of Agriculture has been providing science-based dietary guidelines for the public. The most current version is the MyPlate illustration that gives a simple, visual picture of how much of what kind of foods make up a good, balanced diet. It recommends minimizing daily intake of sugars, the "bad fats", trans and saturated fat, and sodium. "Good fats", unsaturated fats or oils, are not considered a food group but do contain essential nutrients and therefore are included as part of a healthy eating pattern. The difference as simple as the difference between a single and double carbon–carbon bond, good and bad fats, can have a significant impact on health.



MyPlate is a reminder that everything you eat and drink matters. Source: Image courtesy of the USDA, www.mypyramid.gov/downloads/MiniPoster.pdf.

Functional Groups

The carbon–carbon double and triple bonds are examples of functional groups in organic chemistry. A functional group is a specific structural arrangement of atoms or bonds that imparts a characteristic chemical reactivity to a molecule. Alkanes have no functional group. A carbon–carbon double bond is considered a functional group because carbon–carbon double bonds chemically react in specific ways that differ from reactions of alkanes (for example, under certain circumstances, alkenes react with water); a carbon–carbon triple bond also undergoes certain specific chemical reactions. In the remainder of this section, we introduce two other common functional groups.

If an OH group (also called a hydroxyl group) is substituted for a hydrogen atom in a hydrocarbon molecule, the compound is an alcohol. Alcohols are named using the parent hydrocarbon name but with the final *-e* dropped and the suffix *-ol* attached. The two simplest alcohols are methanol and ethanol. Figure **3.7.4** shows their formulas along with a molecular model of each.

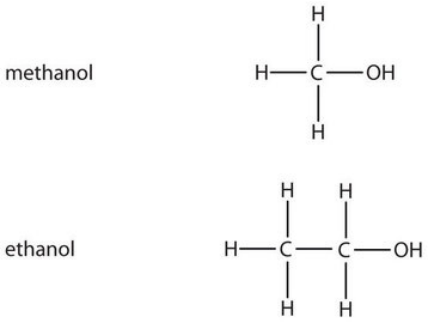


Figure **3.7.4**: The Two Simplest Organic Alcohol Compounds. Alcohols have an OH functional group in the molecule.

Cholesterol, described in the chapter-opening essay, has an alcohol functional group, as its name implies.

Alcohol

Ethanol (also called ethyl alcohol) is the alcohol in alcoholic beverages. Other alcohols include methanol (or methyl alcohol), which is used as a solvent and a cleaner, and isopropyl alcohol (or rubbing alcohol), which is used as a medicinal disinfectant. Neither methanol nor isopropyl alcohol should be ingested, as they are toxic even in small quantities.

Another important family of organic compounds has a carboxyl group, in which a carbon atom is double-bonded to an oxygen atom and to an OH group. Compounds with a carboxyl functional group are called carboxylic acids, and their names end in *-oic acid*. Figure **3.7.5** shows the formulas and the molecular models of the two simplest carboxylic acids, perhaps best known by the common names formic acid and acetic acid. The carboxyl group is sometimes written in molecules as COOH.

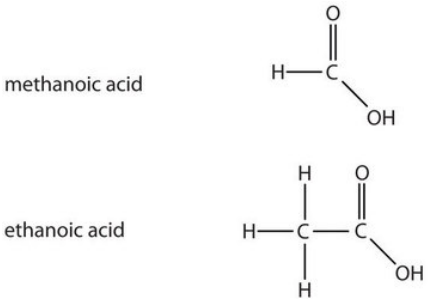


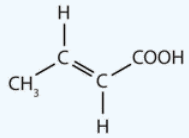
Figure **3.7.5**: The Two Smallest Organic Acids. The two smallest carboxylic acids are formic acid (found in the stingers of ants) and acetic acid (found in vinegar).

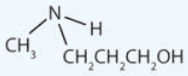
The condensed structures of methanoic acid and ethanoic acid are **HCOOH** and **CH_3COOH**, respectively.

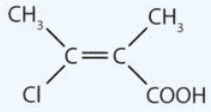
Many organic compounds are considerably more complex than the examples described here. Many compounds, such as cholesterol discussed in the chapter-opening essay, contain more than one functional group. The formal names can also be quite complex.

Example 3.7.1

Identify the functional group(s) in each molecule as a double bond, a triple bond, an alcohol, or a carboxyl.

a. 

b. 

c. 

d. **$CH_3CH_2CH_2CH_2OH$**

Answer a

This molecule has a double bond and a carboxyl functional group.

Answer b

This molecule has an alcohol functional group.

Answer c

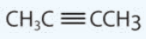
This molecule has a carbon-carbon double bond and a carboxyl functional group.

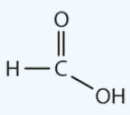
Answer d

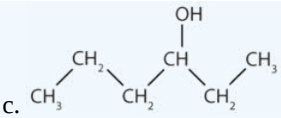
This molecule has an alcohol functional group.

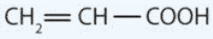
Exercise 3.7.1

Identify the functional group(s) in each molecule as a double bond, a triple bond, an alcohol, or a carboxyl.

a. 

b. 

c. 

d. 

Answer a:

triple bond (alkyne)

Answer b:

carboxyl group

Answer c:

alcohol group

Answer d:

double bond (alkene) and carboxyl group

Career Focus: Forensic Chemist

The main job of a forensic chemist is to identify unknown materials and their origins. Although forensic chemists are most closely associated in the public mind with crime labs, they are employed in pursuits as diverse as tracing evolutionary patterns in living organisms, identifying environmental contaminants, and determining the origin of manufactured chemicals.

In a crime lab, the forensic chemist has the job of identifying the evidence so that a crime can be solved. The unknown samples may consist of almost anything—for example, paint chips, blood, glass, cloth fibers, drugs, or human remains. The forensic chemist subjects them to a variety of chemical and instrumental tests to discover what the samples are. Sometimes these samples are extremely small, but sophisticated forensic labs have state-of-the-art equipment capable of identifying the smallest amount of unknown sample.

Another aspect of a forensic chemist's job is testifying in court. Judges and juries need to be informed about the results of forensic analyses, and it is the forensic chemist's job to explain those results. Good public-speaking skills, along with a broad background in chemistry, are necessary to be a successful forensic chemist.

Key Takeaways

- Organic chemistry is the study of the chemistry of carbon compounds.
- Organic molecules can be classified according to the types of elements and bonds in the molecules.

3.7: Organic Chemistry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.6: Organic Chemistry** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

3.8: Covalent Bonding and Simple Molecular Compounds (Exercises)

4.1: Covalent Bonds

Concept Review Exercises

1. How is a covalent bond formed between two atoms?
2. How does covalent bonding allow atoms in group 6A to satisfy the octet rule?

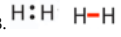
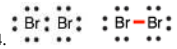
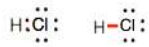
Answers

1. Covalent bonds are formed by two atoms sharing electrons.
2. The atoms in group 6A make two covalent bonds.

Exercises

1. Define *covalent bond*.
2. What is electron sharing?
3. Draw the Lewis diagram for the covalent bond in the H_2 molecule.
4. Draw the Lewis diagram for the covalent bond in the Br_2 molecule.
5. Draw the Lewis diagram for the covalent bond in the HCl molecule.
6. What is the difference between a molecule and a formula unit?
7. Why do hydrogen atoms not follow the octet rule when they form covalent bonds?
8. Draw the Lewis diagram for the covalent bonding in H_2S. How many bonding electrons and nonbonding electrons are in the molecule?
9. Draw the Lewis diagram for the covalent bonding in CF_4. How many bonding electrons and nonbonding electrons are in the molecule?
10. Draw the Lewis diagram for the covalent bonding in PCl_3. How many bonding electrons and nonbonding electrons are in the molecule?
11. How many covalent bonds does a hydrogen atom typically form? Why?
12. How many covalent bonds does an oxygen atom typically form? Why?
13. Tellurium atoms make covalent bonds. How many covalent bonds would a tellurium atom make? Predict the formula of a compound between tellurium and hydrogen.
14. Tin atoms make covalent bonds. How many covalent bonds would a tin atom make? Predict the formula of a compound between tin and hydrogen.
15. Astatine is a synthetic element, made one atom at a time in huge “atom-smasher” machines. It is in the halogen group on the periodic table. How many covalent bonds would an atom of this element form?
16. There have been reports that atoms of element 116 (Lv) were made by smashing smaller atoms together. Using the periodic table, determine what column element 116 would be in and suggest how many covalent bonds an atom of this element would form.

Answers

1. A covalent bond is formed when two atoms share electrons.
2. Electron sharing joins two atoms in a covalent bond. This is a more stable arrangement than 2 individual atoms.
3. 
4. 
5. 

6. A molecule is a discrete combination of atoms; a formula unit is the lowest ratio of ions in a crystal. 7. Hydrogen atoms follow the duet rule (not the octet rule). This is because it has only one shell and this shell can only hold 2 electrons.

8. 

bonding electrons: 4; nonbonding electrons: 4

9. 

bonding electrons: 8; nonbonding electrons: 24

10. 

bonding electrons: 6; nonbonding electrons: 20

11. Hydrogen atoms form only one covalent bond because they have only one valence electron to pair.
12. Oxygen atoms form 2 covalent bonds because oxygen atoms have 6 valence electrons (2 lone pairs plus 2 unpaired electrons that are shared to achieve octet).
13. two; H_2Te
14. four: SnH_4

15. one
16. two

4.2: Covalent Compounds - Formulas and Names

Concept Review Exercises

1. How do you recognize a covalent compound?
2. What are the rules for writing the molecular formula of a simple covalent compound?
3. What are the rules for naming a simple covalent compound?

Answers

1. A covalent compound is usually composed of two or more nonmetal elements.
2. It is just like an ionic compound except that the element further down and to the left on the periodic table is listed first and is named with the element name.
3. Name the first element first and then the second element by using the stem of the element name plus the suffix -ide. Use numerical prefixes if there is more than one atom of the first element; always use numerical prefixes for the number of atoms of the second element.

Exercises

1. Identify whether each compound has covalent bonds.
   a. NaI
   b. Na_2CO_3
   c. N_2O
   d. SiO_2
2. Identify whether each compound has covalent bonds.
   a. C_2H_6
   b. C_6H_5Cl
   c. $KC_2H_3O_2$
   d. $Ca(OH)_2$
3. Identify whether each compound has ionic bonds, covalent bonds, or both.
   a. Na_3PO_4
   b. K_2O
   c. $COCl_2$
   d. $CoCl_2$
4. Identify whether each compound has ionic bonds, covalent bonds, or both.
   a. $FeCl_3$
   b. $Fe(NO_3)_3$
   c. $(NH_2)_2CO$
   d. SO_3
5. Which is the correct molecular formula—H_4Si or SiH_4? Explain.
6. Which is the correct molecular formula—SF_6 or F_6S? Explain.
7. Write the name for each covalent compound.
   a. SiF_4
   b. NO_2
   c. CS_2
   d. P_2O_5
8. Write the name for each covalent compound.
   a. CO
   b. S_2O_3
   c. BF_3
   d. GeS_2
9. Write the formula for each covalent compound.
   a. iodine trichloride
   b. disulfur dibromide
   c. arsenic trioxide
   d. xenon hexafluoride
10. Write the formula for each covalent compound.
    a. boron trichloride
    b. carbon dioxide
    c. tetraphosphorus decoxide
    d. germanium dichloride
11. Write two covalent compounds that have common rather than systematic names.
12. What is the name of the simplest organic compound? What would its name be if it followed the nomenclature for binary covalent compounds?

Answers

1. a. no
   b. yes
   c. yes
   d. yes
2.

a. yes
b. yes
c. yes
d. yes

3. a. both
b. ionic
c. covalent
d. ionic

4.

a. ionic
b. both
c. covalent
d. covalent

5. SiH_4; except for water, hydrogen is almost never listed first in a covalent compound.

6. SF_6; the less electronegative atom (S) is written first

7. a. silicon tetrafluoride
b. nitrogen dioxide
c. carbon disulfide
d. diphosphorus pentoxide

8.

a. carbon monoxide
b. disulfur trioxide
c. boron trifluoride
d. germanium disulfide

9. a. ICl_3
b. S_2Br_2
c. AsO_3
d. XeF_6

10.

a. BCl_3
b. CO_2
c. P_4O_{10}
d. $GeCl_2$

11. H_2O and NH_3 (water and ammonia) (answers will vary)
12. CH_4; carbon tetrahydride

4.3: Drawing Lewis Structures

Exercises

1. What is one clue that a molecule has a multiple bond?

2. Draw the Lewis diagram for each of the following.

a. H_2O

b. NH_3

c. C_2H_6

d. CCl_4

3. Each molecule contains double bonds. Draw the Lewis diagram for each. The first element is the central atom.

a. CS_2
b. C_2F_4
c. $COCl_2$

4. Each molecule contains multiple bonds. Draw the Lewis diagram for each. Assume that the first element is the central atom, unless otherwise noted.

a. N_2
b. CO
c. HCN (The carbon atom is the central atom.)
d. POCl (The phosphorus atom is the central atom.)

5. Explain why hydrogen atoms do not form double bonds.

6. Why is it incorrect to draw a double bond in the Lewis diagram for MgO?

Answers

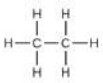
1. If single bonds between all atoms do not give all atoms (except hydrogen) an octet, multiple covalent bonds may be present.
2. a.



b.



c.



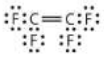
d.



3. a.



b.



c.

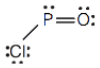


4. a. $\ddot{N}\equiv\ddot{N}$

b. $:C\equiv O:$

c. $H:C\equiv\ddot{N}$

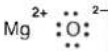
d.



5. Hydrogen can accept only one more electron; multiple bonds require more than one electron pair to be shared.

6. MgO is an ionic compound (Mg transfers two electrons to O). The electrons are not shared hence it's incorrect to draw a double bond.

This is the Lewis dot structure of MgO.



4.4: Characteristics of Covalent Bonds

Concept Review Exercises

1. What is the name for the distance between two atoms in a covalent bond?
2. What does the electronegativity of an atom indicate?
3. What type of bond is formed between two atoms if the difference in electronegativities is small? Medium? Large?

Answers

1. bond length
2. Electronegativity is a qualitative measure of how much an atom attracts electrons in a covalent bond.
3. nonpolar; polar; ionic

Exercises

1. Which is longer—a C–H bond or a C–O bond? (Refer to Table **3.8.1**.)
2. Which is shorter—an N–H bond or a C–H bond? (Refer to Table **3.8.1**.)
3. A nanometer is 10^{-9} m. Using the data in Table **3.8.1** and Table **3.8.2**, determine the length of each bond in nanometers.
   a. a C–O bond
   b. a C=O bond
   c. an H–N bond
   d. a C≡N bond
4. An angstrom (Å) is defined as 10^{-10} m. Using Table **3.8.1** and Table **3.8.2**, determine the length of each bond in angstroms.
   a. a C–C bond
   b. a C=C bond
   c. an N≡N bond
   d. an H–O bond
5. Refer to Exercise 3. Why is the nanometer unit useful as a unit for expressing bond lengths?
6. Refer to Exercise 4. Why is the angstrom unit useful as a unit for expressing bond lengths?
7. Using Figure **3.8.3**, determine which atom in each pair has the higher electronegativity.
   a. H or C
   b. O or Br

c. Na or Rb
d. I or Cl

8. Using Figure **3.8.3**, determine which atom in each pair has the lower electronegativity.

a. Mg or O
b. S or F
c. Al or Ga
d. O or I

9. Will the electrons be shared equally or unequally across each covalent bond? If unequally, to which atom are the electrons more strongly drawn?

a. a C–O bond
b. an F–F bond
c. an S–N bond
d. an I–Cl bond

10. Will the electrons be shared equally or unequally across each covalent bond? If unequally, to which atom are the electrons more strongly drawn?

a. a C–C bond
b. a S–Cl bond
c. an O–H bond
d. an H–H bond

11. Arrange the following bonds from least polar to most polar: H-F, H-N, H-O, H-C

12. Arrange the following bonds from least polar to most polar: C-F, C-N, C-O, C-C

Answers

1. A C–O bond is longer.

2. An H-N bond is shorter than an H-C bond.

3. a. 0.143 nm
b. 0.120 nm
c. 0.100 nm
d. 0.116 nm

4.

a. 1.54 Å
b. 1.34 Å
c. 1.10 Å
d. 0.97 Å

5. Actual bond lengths are very small, so the nanometer unit makes the expression of length easier to understand.

6. Actual bond lengths are very small, so the angstrom unit makes the expression of length easier to understand.

7. a. C
b. O
c. Na
d. Cl

8.

a. Mg
b. S
c. Al
d. I

9.

a. unequally toward the O
b. equally
c. unequally toward the N
d. unequally toward the Cl

10.

a. equally
b. unequally toward the Cl
c. unequally toward the O
d. equally

11. The electronegativity difference increases from 0.4; 0.9; 1.4; 1.9. Hence, the least to most polar: H-C, H-N, H-O, H-F

12. The electronegativity difference increases from 0; 0.5; 1.0; 1.5. Hence, the least to most polar: C-C, C-N, C-O, C-F

4.5: Characteristics of Molecules

Concept Review Exercises

1. How do you determine the molecular mass of a covalent compound?
2. How do you determine the shape of a molecule?
3. How do you determine whether a molecule is polar or nonpolar?

Answers

1. The molecular mass is the sum of the masses of the atoms in the formula.

2. The shape of a molecule is determined by the position of the atoms, which in turn is determined by the repulsion of the bonded and lone electron pairs around the central atom.

3. If all the bonds in a molecule are nonpolar, the molecule is nonpolar. If it contains identical polar bonds that are oriented symmetrically opposite each other (linear, trigonal planar or tetrahedral) then the molecule is nonpolar. If it contains polar bonds that don't cancel each other's effects, the molecule is polar.

Exercises

1. What is the molecular mass of each compound?
   a. H_2S
   b. N_2O_4
   c. ICl_3
   d. HCl
2. What is the molecular mass of each compound?
   a. O_2F_2
   b. CCl_4
   c. C_6H_6
   d. SO_3
3. Aspirin ($C_9H_8O_4$) is a covalent compound. What is its molecular mass?
4. Cholesterol ($C_{27}H_{46}O$) is a biologically important compound. What is its molecular mass?
5. What is the shape of each molecule? State whether it is polar or nonpolar.
   a. H_2S
   b. $COCl_2$
   c. SO_2
6. What is the shape of each molecule? State whether it is polar or nonpolar.
   a. NBr_3
   b. SF_2
   c. SiH_4
7. Predict the shape of nitrous oxide (N_2O), which is used as an anesthetic. A nitrogen atom is in the center of this three-atom molecule. Is this polar?
8. Predict the shape of acetylene (C_2H_2), which has the two carbon atoms in the middle of the molecule with a triple bond. What generalization can you make about the shapes of molecules that have more than one central atom?

Answers

1. a. 34.08 amu
   b. 92.02 amu
   c. 233.25 amu
   d. 36.46 amu
2. What is the molecular mass of each compound?
   a. 70.00 amu
   b. 153.81 amu
   c. 78.12 amu
   d. 80.06 amu
3. 180.17 amu
4. 386.73 amu
5. a. bent; polar
   b. trigonal planar; nonpolar
   c. bent; polar
6. a. pyramidal; polar
   b. bent; polar
   c. tetrahedral; nonpolar
7. linear; polar
8. linear; in a molecule with more than one central atom, the geometry around each central atom needs to be examined.

4.6: Organic Chemistry

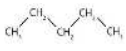
Concept Review Exercises

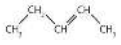
1. What is organic chemistry?
2. What is a functional group? Give at least two examples of functional groups.

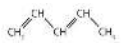
Answers

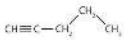
1. Organic chemistry is the study of the chemistry of carbon compounds.
2. A functional group is a specific structural arrangement of atoms or bonds that imparts a characteristic chemical reactivity to the molecule; alcohol group and carboxylic group (answers will vary).

Exercises

1. Give three reasons why carbon is the central element in organic chemistry.
2. Are organic compounds based more on ionic bonding or covalent bonding? Explain.
3. Identify the type of hydrocarbon in each structure.
   a. 

b. 

c. 

d. 

4. Identify the type of hydrocarbon in each structure.

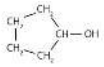
a. 

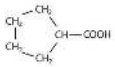
b. 

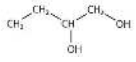
c.

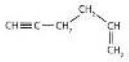
d. 

5. Identify the functional group(s) in each molecule.

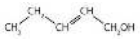
a. 

b. 

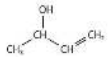
c. 

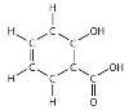
d. 

6. Identify the functional group(s) in each molecule.

a. 

b. 

c. 

d. 

7. How many functional groups described in this section contain carbon and hydrogen atoms only? Name them.

8. What is the difference in the ways the two oxygen atoms in the carboxyl group are bonded to the carbon atom?

Answers

1. Carbon atoms bond reasonably strongly with other carbon atoms. Carbon atoms bond reasonably strongly with atoms of other elements. Carbon atoms make a large number of covalent bonds (four).

2. Organic compounds are based on covalent bonding or electron sharing. The atoms C, H, O, N that make up organic compounds are all nonmetals.

3.

a. alkane
b. alkene
c. alkene
d. alkyne

4.

a. alkene
b. alkane
c. alkyne
d. alkene

5.

a. alcohol
b. carboxyl
c. alcohol
d. alkene and alkyne

6.

a. a carbon-carbon double bond and alcohol
b. carboxyl group
c. carbon-carbon double bond and alcohol
d. carbon-carbon double bond; alcohol and carboxyl group

7. two; carbon-carbon double bonds and carbon-carbon triple bonds

8. There are two oxygen atoms in a carboxyl group: one is double-bonded while the other is OH, single bonded to the same carbon atom.

Additional Exercises

Use the atomic masses found in Figure 2.7.1

1. An atomic mass unit equals 1.661×10^{-24} g. what is the mass in grams of each molecule of (a) H_2S (b) N_2O_4 (c) ICl_3 (d) NCl_3?
2. An atomic mass unit equals 1.661×10^{-24} g. What is the mass in grams of (a) O_2F_2 (b) CCl_4 (c) C_6H_6 (d) SO_3?
3. An atomic mass unit equals 1.661×10^{-24} g. What is the mass in grams of 5.00×10^{22} molecules of $C_9H_8O_4$?
4. An atomic mass unit equals 1.661×10^{-24} g. What is the mass in grams of 1.885×10^{20} molecules of $C_{27}H_{46}O$?
5. Acetic acid has the following structure:

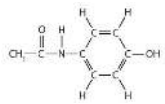


This molecule can lose a hydrogen ion (H^+) and the resulting anion can combine with other cations, such as Na^+:



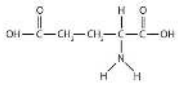
Name this ionic compound.

6. Formic acid (HCOOH) loses a hydrogen ion to make the formate ion ($HCOO^-$). Write the formula for each ionic compound: potassium formate, calcium formate, and ferric formate.
7. Cyanogen has the formula C_2N_2. Propose a bonding scheme that gives each atom the correct number of covalent bonds. (Hint: the two carbon atoms are in the center of a linear molecule.)
8. How many carbon–carbon single bonds, linked together, are needed to make a carbon chain that is 1.000 cm long?
9. How many carbon–carbon double bonds, linked together, are needed to make a carbon chain that is 1.000 cm long?
10. In addition to themselves, what other atoms can carbon atoms bond with and make covalent bonds that are nonpolar (or as nonpolar as possible)?
11. What is the greatest possible electronegativity difference between any two atoms? Use Figure 4.4 to find the answer.
12. Acetaminophen, a popular painkiller, has the following structure:



Name the recognizable functional groups in this molecule. Do you think there are other groups of atoms in this molecule that might qualify as functional groups?

13. Glutamic acid is the parent compound of monosodium glutamate (known as MSG), which is used as a flavor enhancer. Glutamic acid has the following structure:



Name the functional groups you recognize in this molecule. Do you think there are other groups of atoms in this molecule that might qualify as functional groups?

Answers

1.

a: 5.661×10^{-23} g

b: 1.528×10^{-22} g

c: 3.874×10^{-22} g

d: 1.999×10^{-22} g

2.

a: 1.163×10^{-22} g

b: 2.555×10^{-22} g

c: 1.298×10^{-22} g

d: 1.330×10^{-22} g

3. 14.96 g

4. 0.1211 g

5. sodium acetate

6.

a. KHCOO

b. $Ca(HCOO)_2$

c. $Fe(HCOO)_3$

7. :N≡C–C≡N:
8. 6.49×10^7 C-C bonds
9. 7.46×10^7 C=C bonds
10. Hydrogen atoms make relatively nonpolar bonds with carbon atoms.
11. The greatest electronegativity difference is 3.2, between F and Rb.
12. alcohol; the ring with double bonds, and the O=C-NH are also likely functional groups.
13. carboxyl and $-NH_2$ functional groups

Additional Questions

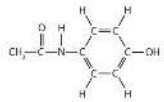
1. An atomic mass unit equals 1.661×10^{-24} g. What is the mass in grams of each molecule of (a) H_2S (b) N_2O_4 (c) ICl_3 (d) NCl_3?
2. An atomic mass unit equals 1.661×10^{-24} g. What is the mass in grams of (a) O_2F_2 (b) CCl_4 (c) C_6H_6 (d) SO_3?
3. An atomic mass unit equals 1.661×10^{-24} g. What is the mass in grams of 5.00×10^{22} molecules of $C_9H_8O_4$?
4. An atomic mass unit equals 1.661×10^{-24} g. What is the mass in grams of 1.885×10^{20} molecules of $C_{27}H_{46}O$?
5. Acetic acid has the following structure:



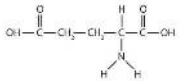
   This molecule can lose a hydrogen ion (H^+) and the resulting anion can combine with other cations, such as Na^+:



   Name this ionic compound.
6. Formic acid (HCOOH) loses a hydrogen ion to make the formate ion ($HCOO^-$). Write the formula for each ionic compound: potassium formate, calcium formate, and ferric formate.
7. Cyanogen has the formula C_2N_2. Propose a bonding scheme that gives each atom the correct number of covalent bonds. (Hint: the two carbon atoms are in the center of a linear molecule.)
8. The molecular formula C_3H_6 represents not only propene, a compound with a carbon–carbon double bond, but also a molecule that has all single bonds. Draw the molecule with formula C_3H_6 that has all single bonds.
9. How many carbon–carbon single bonds, linked together, are needed to make a carbon chain that is 1.000 cm long?
10. How many carbon–carbon double bonds, linked together, are needed to make a carbon chain that is 1.000 cm long?
11. In addition to themselves, what other atoms can carbon atoms bond with and make covalent bonds that are nonpolar (or as nonpolar as possible)?
12. What is the greatest possible electronegativity difference between any two atoms? Use Figure 4.4 to find the answer.
13. Acetaminophen, a popular painkiller, has the following structure:



   Name the recognizable functional groups in this molecule. Do you think there are other groups of atoms in this molecule that might qualify as functional groups?
14. Glutamic acid is the parent compound of monosodium glutamate (known as MSG), which is used as a flavor enhancer. Glutamic acid has the following structure:



   Name the functional groups you recognize in this molecule. Do you think there are other groups of atoms in this molecule that might qualify as functional groups?

Answers

1. 1a: 5.75×10^{-23} g; 1b: 1.53×10^{-22} g; 1c: 3.88×10^{-22} g; 1d: 6.06×10^{-23} g

3. 14.96 g

5. sodium acetate

7. N≡C–C≡N

9. 6.49×10^7 bonds

11. Hydrogen atoms make relatively nonpolar bonds with carbon atoms.

13. alcohol; the N–H group, the ring with double bonds, and the C=O are also likely functional groups.

3.8: Covalent Bonding and Simple Molecular Compounds (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.1: Covalent Bonds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **4.2: Covalent Compounds - Formulas and Names** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **4.3: Drawing Lewis Structures** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **4.4: Characteristics of Covalent Bonds** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **4.5: Characteristics of Molecules** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **4.6: Organic Chemistry** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **4.E: Covalent Bonding and Simple Molecular Compounds (Exercises)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

3.9: Covalent Bonding and Simple Molecular Compounds (Summary)

To ensure that you understand the material in this chapter, you should review the meanings of the following bold terms in the following summary and ask yourself how they relate to the topics in the chapter.

Atoms can share pairs of valence electrons to obtain a valence shell octet. This sharing of electrons is a **covalent bond**. A species formed from covalently bonded atoms is a **molecule** and is represented by a **molecular formula**, which gives the number of atoms of each type in the molecule. The two electrons shared in a covalent bond are called a **bonding pair of electrons**. The electrons that do not participate in covalent bonds are called **nonbonding pairs** (or **lone pairs**) **of electrons**. A covalent bond consisting of one pair of shared electrons is called a **single bond**.

Covalent bonds occur between nonmetal atoms. Naming simple covalent compounds follows simple rules similar to those for ionic compounds. However, for covalent compounds, numerical prefixes are used as necessary to specify the number of atoms of each element in the compound.

In some cases, more than one pair of electrons is shared to satisfy the octet rule. Two pairs of electrons are shared by two atoms to make a **double bond**. Three pairs of atoms are shared to make a **triple bond**. Single, double, and triple covalent bonds may be represented by one, two, or three dashes, respectively, between the symbols of the atoms.

The distance between two covalently bonded atoms is the **bond length**. Bond lengths depend on the types of atoms participating in the bond as well as the number of electron pairs being shared. A covalent bond can be a **polar covalent bond** if the electron sharing between the two atoms is unequal. If the sharing is equal, the bond is a **nonpolar covalent bond**. Because the strength of an atom's attraction for electrons in a bond is rated by the atom's **electronegativity**, the difference in the two atoms' electronegativities indicates how polar a covalent bond between those atoms will be.

The mass of a molecule is called its **molecular mass** and is the sum of the masses of the atoms in the molecule. The shape of a molecule can be predicted using **valence shell electron pair repulsion (VSEPR)**, which uses the fact that the negative electrons in covalent bonds repel each other as much as possible. Molecules with polar bonds are **polar** except when the bond polarities cancel due to symmetry.

Organic chemistry is the chemistry of carbon compounds. Carbon forms covalent bonds with other carbon atoms and with the atoms of many other elements. The simplest organic compounds are **hydrocarbons**, which consist solely of carbon and hydrogen. Hydrocarbons containing only single bonds are called **alkanes (saturated hydrocarbons)**. Hydrocarbons containing carbon–carbon double bonds are **alkenes**, while hydrocarbons with carbon–carbon triple bonds are **alkynes**. Carbon-carbon double and triple bonds are examples of **functional groups**, atoms or bonds that impart a characteristic chemical function to the molecule. Other functional groups include the alcohol functional group (OH) and the **carboxyl functional group** (COOH). They are the characteristic functional group in organic compounds called **alcohols** and **carboxylic acids**.

3.9: Covalent Bonding and Simple Molecular Compounds (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **4.S: Covalent Bonding and Simple Molecular Compounds (Summary)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

CHAPTER OVERVIEW

4: Noncovalent bonds

In addition to ionic and covalent bonding, there are also a variety of non-covalent interactions that stabilize the interactions between different molecules. Although each of these interactions tend to be weak, there are many of them present, and they are responsible for many of the interactions that take place in biological environments.

4.1: Intermolecular Interactions

4: Noncovalent bonds is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

4.1: Intermolecular Interactions

Learning Objectives

- Define *phase*.
- Identify the types of interactions between molecules.

A phase is a certain form of matter that includes a specific set of physical properties. That is, the atoms, the molecules, or the ions that make up the phase do so in a consistent manner throughout the phase. Science recognizes three stable phases: the ***solid** phase*, in which individual particles can be thought of as in contact and held in place; the ***liquid** phase*, in which individual particles are in contact but moving with respect to each other; and the ***gas** phase*, in which individual particles are separated from each other by relatively large distances (see Figure 8.1.1). Not all substances will readily exhibit all phases. For example, carbon dioxide does not exhibit a liquid phase unless the pressure is greater than about six times normal atmospheric pressure. Other substances, especially complex organic molecules, may decompose at higher temperatures, rather than becoming a liquid or a gas.

> *For many substances, there are different arrangements the particles can take in the solid phase, depending on temperature and pressure.*

Which phase a substance adopts depends on the pressure and the temperature it experiences. Of these two conditions, temperature variations are more obviously related to the phase of a substance. When it is very cold, H_2O exists in the solid form as ice. When it is warmer, the liquid phase of H_2O is present. At even higher temperatures, H_2O boils and becomes steam.

Pressure changes can also affect the presence of a particular phase (as we indicated for carbon dioxide), but its effects are less obvious most of the time. We will mostly focus on the temperature effects on phases, mentioning pressure effects only when they are important. Most chemical substances follow the same pattern of phases when going from a low temperature to a high temperature: the solid phase, then the liquid phase, and then the gas phase. However, the temperatures at which these phases are present differ for all substances and can be rather extreme. Table **4.1.1** shows the temperature ranges for solid, liquid, and gas phases for three substances. As you can see, there is extreme variability in the temperature ranges.

Table **4.1.1**: Temperature Ranges for the Three Phases of Various Substances

Substance	Solid Phase Below	Liquid Phase Above	Gas Phase Above
hydrogen (H_2)	−259°C	−259°C	−253°C
water (H_2O)	0°C	0°C	100°C
sodium chloride (NaCl)	801°C	801°C	1413°C

> *The melting point of a substance is the temperature that separates a solid and a liquid. The boiling point of a substance is the temperature that separates a liquid and a gas.*

What accounts for this variability? Why do some substances become liquids at very low temperatures, while others require very high temperatures before they become liquids? It all depends on the strength of the **intermolecular forces (IMF)** between the **particles** of substances and the **kinetic energies** (KE) of its molecules. (Although ionic compounds are not composed of discrete molecules, we will still use the term *intermolecular* to include interactions between the ions in such compounds.) Substances that experience **strong intermolecular interactions** require **higher temperatures to become liquids and, finally, gases.** Substances that experience **weak intermolecular interactions** do not need much energy (as measured by temperature) to become liquids and gases and will exhibit these phases at **lower temperatures.**

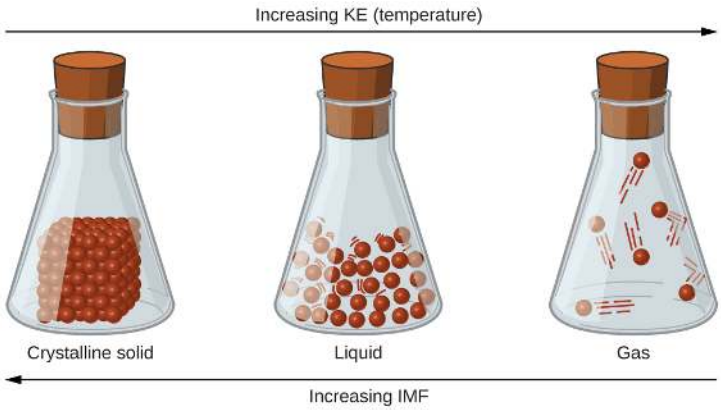


Figure **4.1.1**: Transitions between solid, liquid, and gaseous states of a substance occur when conditions of temperature or pressure favor the associated changes in intermolecular forces. (Note: The space between particles in the gas phase is much greater than shown.) (CC BY 4.0; OpenStax)

Covalent Network Materials

Substances with the highest melting and boiling points have **covalent network** bonding. This type of intermolecular interaction is actually a covalent bond. In these substances, all the atoms in a sample are covalently bonded to one another; in effect, the entire sample is essentially **one giant molecule**. Many of these substances are solid over a large temperature range because it takes a lot of energy to disrupt all the covalent bonds at once. One example of a substance that shows covalent network bonding is diamond (Figure **4.1.2**). Diamond is composed entirely of carbon atoms, each bonded to four other carbon atoms in a tetrahedral geometry. Melting a covalent network solid is not accomplished by overcoming the relatively weak intermolecular forces. Rather, all of the covalent bonds must be broken, a process that requires extremely high temperatures. Diamond, in fact, does not melt at all. Instead, it vaporizes to a gas at temperatures above 3,500°C. Diamond is extremely hard and is one of the few materials that can cut glass.

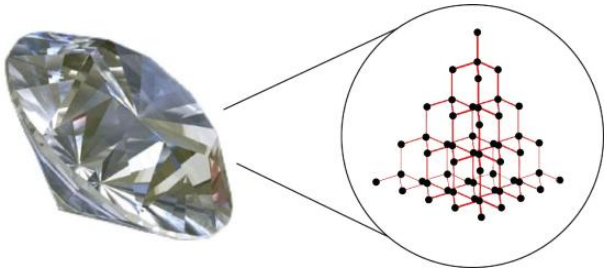


Figure **4.1.2**: Diamond. Diamond, a form of pure carbon, has covalent network bonding. It takes a very high temperature—over 3,500°C—for diamond to leave the solid state. Source: Photo © Thinkstock

Ionic Compounds

The strongest force between any two particles is the ionic bond, in which two ions of opposing charge are attracted to each other. Thus, ionic interactions between particles are another type of intermolecular interaction. Substances that contain ionic interactions are relatively strongly held together, so these substances typically have high melting and boiling points. Sodium chloride (Figure **4.1.3**) is an example of a substance whose particles experience ionic interactions (Table **4.1.1**). These attractive forces are sometimes referred to as ion-ion interactions.

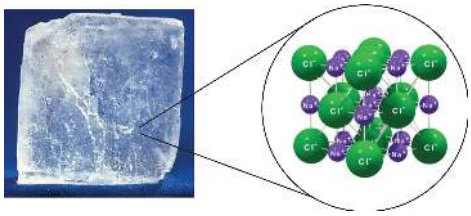


Figure **4.1.3**: Sodium Chloride. Solid NaCl is held together by ionic interactions. Source: Photo © Thinkstock

Covalent Molecular Compounds

There are two different covalent structures: molecular and network. Covalent network compounds like SiO_2 (quartz) have structures of atoms in a network like diamond described earlier. In this section, we are dealing with the molecular type that contains individual molecules. The bonding between atoms in the individual molecule is covalent but the attractive forces between the molecules are called ***intermolecular forces (IMF)***.

In contrast to *intramolecular* forces (see Figure 8.1.4), such as the covalent bonds that hold atoms together in molecules and polyatomic ions, intermolecular forces hold molecules together in a **liquid or solid**. Intermolecular forces are generally much weaker than covalent bonds. For example, it requires 927 kJ to overcome the intramolecular forces and break both O–H bonds in 1 mol of water, but it takes only about 41 kJ to overcome the intermolecular attractions and convert 1 mol of liquid water to water vapor at 100°C. (Despite this seemingly low value, the intermolecular forces in liquid water are among the strongest such forces known!) Given the large difference in the strengths of intra- and intermolecular forces, changes between the solid, liquid, and gaseous states almost invariably occur for molecular substances ***without breaking covalent bonds.***

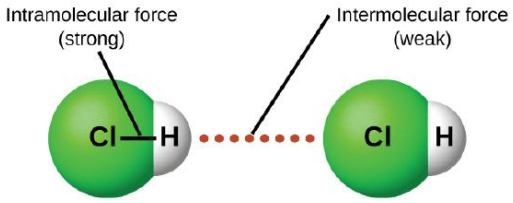


Figure **4.1.4**: **Intramolecular** forces keep a molecule intact. **Intermolecular** forces hold multiple molecules together and determine many of a substance's properties. (CC BY 4.0; OpenStax)

In this section, we will discuss the three types of IMF in molecular compounds: dipole-dipole, hydrogen bonding and London dispersion forces.

Dipole-dipole Intermolecular Forces

As discussed in Section 4.4, covalent bond that has an unequal sharing of electrons is called a **polar covalent** bond. (A covalent bond that has an equal sharing of electrons, as in a covalent bond with the same atom on each side, is called a **nonpolar covalent** bond.) A molecule with a net unequal distribution of electrons in its covalent bonds is a polar molecule. HF is an example of a polar molecule (see Figure 8.1.5).



Figure **4.1.5**: Polar Covalent Bonds. The electrons in the HF molecule are not equally shared by the two atoms in the bond. Because the fluorine atom has nine protons in its nucleus, it attracts the negatively charged electrons in the bond more than the hydrogen atom does with its one proton in its nucleus. Thus, electrons are more strongly attracted to the fluorine atom, leading to an imbalance in the electron distribution between the atoms. The fluorine side of the bond picks up a partial overall negative charge (represented by the δ− [illegible] hydrogen side of the bond has an overall partial positive charge (represented by the δ+ in the diagram). Such a bond is called a *polar covalent bond*.

Empty alt attribute

The charge separation in a **polar covalent bond** is not as extreme as is found in ionic compounds, but there is a related result: oppositely charged ends of different molecules will attract each other. This type of intermolecular interaction is called a **dipole-dipole interaction**. Many molecules with polar covalent bonds experience dipole-dipole interactions.

The covalent bonds in some molecules are oriented in space in such a way that the bonds in the molecules cancel each other out. The individual bonds are polar, but due to molecular symmetry, the overall molecule is not polar; rather, the molecule is ***nonpolar.*** Such molecules experience little or no dipole-dipole interactions. Carbon dioxide (CO_2) and carbon tetrachloride (CCl_4) are examples of such molecules (Figure **4.1.6**).

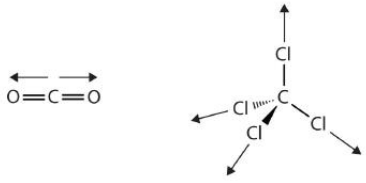


Figure **4.1.6**: Nonpolar Molecules. Although the individual bonds in both CO_2 and CCl_4 are polar, their effects cancel out because of the spatial orientation of the bonds in each molecule. As a result, such molecules experience little or no dipole-dipole interaction.

Recall from the Sections 4.4 and 4.5, on chemical bonding and molecular geometry that *polar* molecules have a partial positive charge on one side and a partial negative charge on the other side of the molecule—a separation of charge called a ***dipole.*** Consider a polar molecule such as hydrogen chloride, HCl. In the HCl molecule, the more electronegative Cl atom bears the partial negative charge, whereas the less electronegative H atom bears the partial positive charge. An attractive force between HCl molecules results from the attraction between the positive end of one HCl molecule and the negative end of another. This attractive force is called a dipole-dipole attraction—the electrostatic force between the partially positive end of one polar molecule and the partially negative end of another, as illustrated in Figure **4.1.7**.

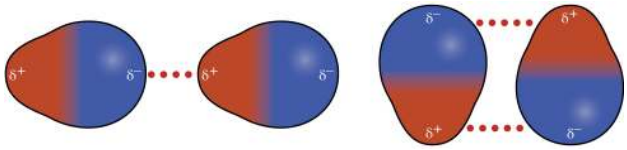


Figure **4.1.7**: This image shows two arrangements of polar molecules, such as HCl, that allow an attraction between the partial negative end of one molecule and the partial positive end of another.

Two pairs of molecules are shown where each molecule has one larger blue side labeled “delta sign, negative sign” and a smaller red side labeled “delta sign, positive sign. In the first pair, the red sides of the two molecules both face to the left and the blue side to the right. A horizontal dotted line lies in between the two. In the second pair, the molecules face up and down, with the red and blue ends aligning. A horizontal dotted line lies between the red and blue ends facing upward and another lies between the red and blue ends facing downward.

The effect of a dipole-dipole attraction is apparent when we compare the properties of HCl molecules to nonpolar F_2 molecules. Both HCl and F_2 consist of the same number of atoms and have approximately the same molecular mass. At a temperature of 150 K, molecules of both substances would have the same average kinetic energy. However, the dipole-dipole attractions between HCl molecules are sufficient to cause them to “stick together” to form a liquid, whereas the relatively weaker dispersion forces between nonpolar F_2 molecules are not, and so this substance is gaseous at this temperature. The higher normal boiling point of HCl (188 K) compared to F_2 (85 K) is a reflection of the greater strength of dipole-dipole attractions between HCl molecules, compared to the attractions between nonpolar F_2 molecules. We will often use values such as boiling or freezing points as indicators of the relative strengths of IMFs of attraction present within different substances.

Empty alt attribute

Example 4.1.1

Predict which will have the higher boiling point: N_2 or CO. Explain your reasoning.

Solution

CO and N_2 are both diatomic molecules with masses of about 28 amu, so they experience similar London dispersion forces. Because CO is a polar molecule, it experiences dipole-dipole attractions. Because N_2 is nonpolar, its molecules cannot exhibit dipole-dipole attractions. The dipole-dipole attractions between CO molecules are comparably stronger than the dispersion forces between nonpolar N_2 molecules, so CO is expected to have the higher boiling point.

4.1.1 Exercise

Predict which will have the higher boiling point: **ICl** or **Br_2**. Explain your reasoning.

Answer

ICl. ICl and Br_2 have similar masses (~160 amu) and therefore experience similar London dispersion forces. ICl is polar and thus also exhibits dipole-dipole attractions; Br_2 is nonpolar and does not. The relatively stronger dipole-dipole attractions require more energy to overcome, so ICl will have the higher boiling point.

Hydrogen Bonding Intermolecular Forces

Molecules with **hydrogen atoms** bonded to electronegative atoms such as **O, N, and F** tend to exhibit unusually strong intermolecular interactions due to a particularly strong type of dipole-dipole attraction called **hydrogen bonding**. The very large difference in electronegativity between the H atom (2.1) and the atom to which it is bonded (4.0 for an F atom, 3.5 for an O atom, or 3.0 for a N atom), combined with the very small size of a H atom and the relatively small sizes of F, O, or N atoms, leads to ***highly concentrated partial charges*** with these atoms. Because the hydrogen atom does not have any electrons other than the ones in the

Empty alt attribute

covalently bonded, its positively charged nucleus is almost completely exposed, allowing strong attractions to other nearby lone pairs of electrons.

Examples of hydrogen bonds include HF···HF, H_2O···HOH, and H_3N···HNH_2, in which the **hydrogen bonds** are denoted by **dots**. Figure **4.1.8** illustrates hydrogen bonding between water molecules. The physical properties of water, which has two O–H bonds, are strongly affected by the presence of hydrogen bonding between water molecules. Most molecular compounds that have a mass similar to water are gases at room temperature. However, because of the strong hydrogen bonds, water molecules are able to stay condensed in the liquid state.

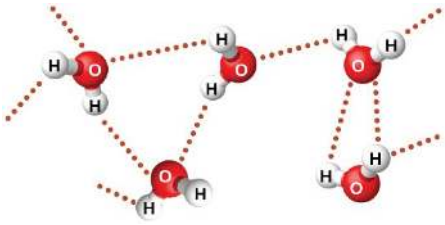


Figure **4.1.8**: Water molecules participate in multiple hydrogen-bonding interactions with nearby water molecules.
Five water molecules are shown near one another, but not touching. A dotted line lies between many of the hydrogen atoms on one molecule and the oxygen atom on another molecule.

A **hydrogen bond** is an intermolecular attractive force in which *a* **hydrogen atom**, that is covalently bonded to a small, highly electronegative atom, is attracted to a **lone pair of electrons** on an atom in a neighboring molecule. Figure **4.1.9** shows how methanol (CH_3OH) molecules experience hydrogen bonding. Methanol contains both a hydrogen atom attached to O; methanol can thus form hydrogen bonds by acting as either a hydrogen bond donor or a hydrogen bond acceptor (lone pair). The hydrogen-bonded structure of methanol is as follows:

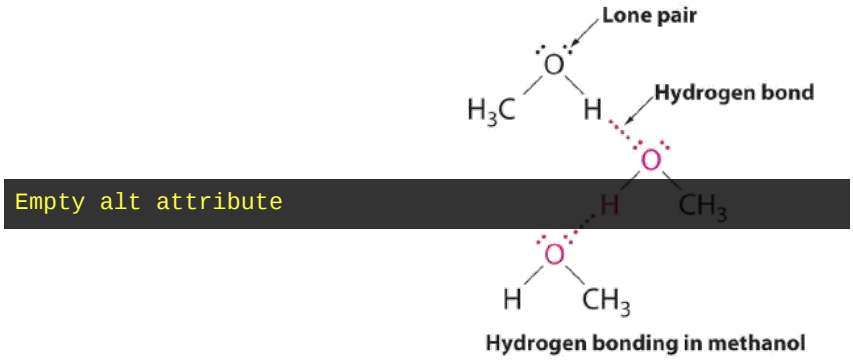


Figure **4.1.9**: Methanol molecules participate in multiple hydrogen-bonding interactions with nearby water molecules.

Despite use of the word “bond,” keep in mind that hydrogen bonds are *intermolecular* attractive forces, not *intramolecular* attractive forces (covalent bonds). Hydrogen bonds are much weaker than covalent bonds, but are generally much stronger than other dipole-dipole attractions and dispersion forces.

Effect of Hydrogen Bonding on Boiling Points

Consider the compounds dimethylether (CH_3OCH_3), ethanol (CH_3CH_2OH), and propane ($CH_3CH_2CH_3$). Their boiling points, not necessarily in order, are −42.1 °C, −24.8 °C, and 78.4 °C. Match each compound with its boiling point. Explain your reasoning.

Solution

The shapes of CH_3OCH_3, CH_3CH_2OH, and $CH_3CH_2CH_3$ are similar, as are their molar masses (46 g/mol, 46 g/mol, and 44 g/mol, respectively), so they will exhibit similar dispersion forces. Since $CH_3CH_2CH_3$ is nonpolar, it may exhibit *only* dispersion forces. Because CH_3OCH_3 is polar, it will also experience dipole-dipole attractions. Finally, CH_3CH_2OH has an −OH group, and so it will experience the uniquely strong dipole-dipole attraction known as hydrogen bonding. So the ordering in terms of strength of IMFs, and thus boiling points, is $CH_3CH_2CH_3$ < CH_3OCH_3 < CH_3CH_2OH. The boiling point of propane is −42.1 °C, the boiling point of dimethylether is −24.8 °C, and the boiling point of ethanol is 78.5 °C.

Empty alt attribute

Exercise 4.1.2

Ethane (CH_3CH_3) has a melting point of −183 °C and a boiling point of −89 °C. Predict the melting and boiling points for methylamine (CH_3NH_2). Explain your reasoning.

Answer

The melting point and boiling point for methylamine are predicted to be significantly greater than those of ethane. CH_3CH_3 and CH_3NH_2 are similar in size and mass, but methylamine possesses an −NH group and therefore may exhibit hydrogen bonding. This greatly increases its IMFs, and therefore its melting and boiling points. It is difficult to predict values, but the known values are a melting point of −93 °C and a boiling point of −6 °C.

Hydrogen Bonding and DNA

Deoxyribonucleic acid (DNA) is found in every living organism and contains the genetic information that determines the organism's characteristics, provides the blueprint for making the proteins necessary for life, and serves as a template to pass this information on to the organism's offspring. A DNA molecule consists of two (anti-)parallel chains of repeating nucleotides, which form its well-known double helical structure, as shown in Figure **4.1.10**.

Empty alt attribute

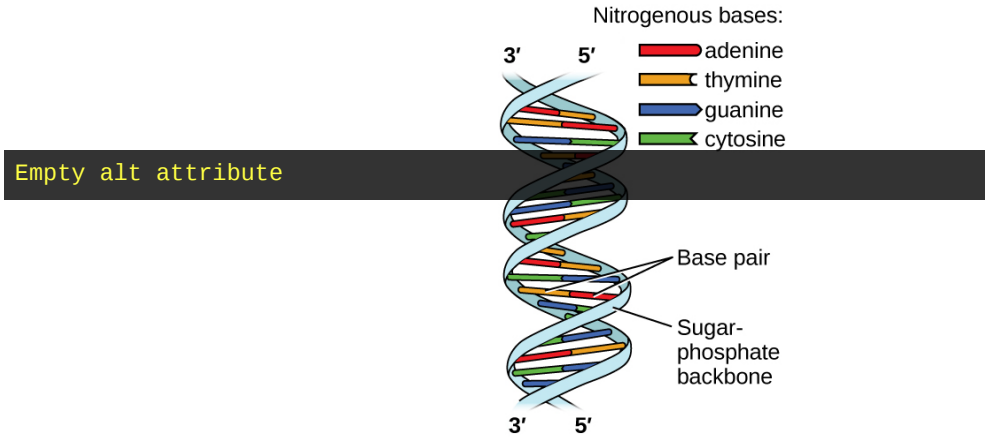


Figure **4.1.10**: Two separate DNA molecules form a double-stranded helix in which the molecules are held together via hydrogen bonding. (credit: modification of work by Jerome Walker, Dennis Myts)

Each nucleotide contains a (deoxyribose) sugar bound to a phosphate group on one side, and one of four nitrogenous bases on the other. Two of the bases, cytosine (C) and thymine (T), are single-ringed structures known as pyrimidines. The other two, adenine (A) and guanine (G), are double-ringed structures called purines. These bases form complementary base pairs consisting of one purine and one pyrimidine, with adenine pairing with thymine, and cytosine with guanine. Each base pair is held together by hydrogen bonding. A and T share two hydrogen bonds, C and G share three, and both pairings have a similar shape and structure Figure **4.1.11**

Empty alt attribute

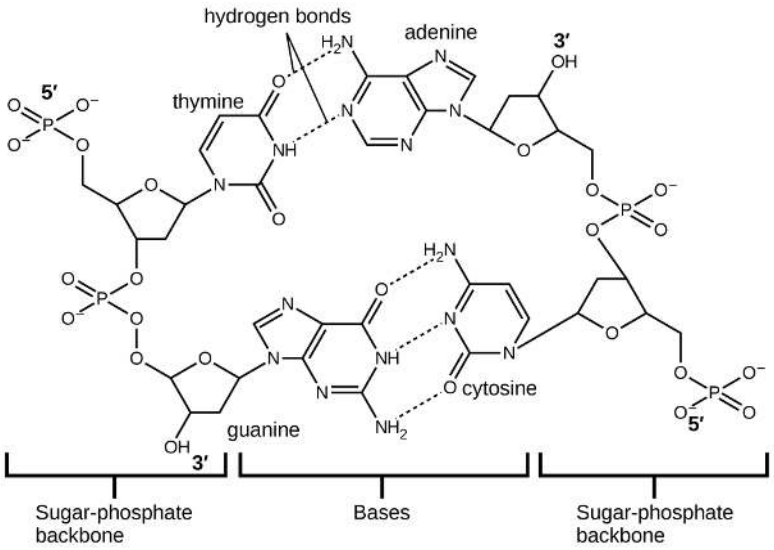


Figure **4.1.11**: The geometries of the base molecules result in maximum hydrogen bonding between adenine and thymine (AT) and between guanine and cytosine (GC), so-called "complementary base pairs."

The cumulative effect of millions of hydrogen bonds effectively holds the two strands of DNA together. Importantly, the two strands of DNA can relatively easily "unzip" down the middle since hydrogen bonds are relatively weak compared to the covalent bonds that hold the atoms of the individual DNA molecules together. This allows both strands to function as a template for replication.

London Dispersion Forces

Finally, there are forces between all molecules that are caused by electrons being in different places in a molecule at any one time, which sets up a temporary separation of charge that disappears almost as soon as it appears. These are very weak intermolecular interactions and are called **dispersion forces (or London forces)**. (An alternate name is London dispersion forces.) Molecules that experience no other type of intermolecular interaction will at least experience dispersion forces. Substances that experience only dispersion forces are typically soft in the solid phase and have relatively low melting points. Examples include *waxes*, which are long hydrocarbon chains that are solids at room temperature because the molecules have so many electrons. The resulting dispersion forces between these molecules make them assume the solid phase at normal temperatures.

Dispersion forces that develop between atoms in different molecules can attract the two molecules to each other. The forces are relatively weak, however, and become significant only when the molecules are very close. **Larger and heavier atoms and molecules** exhibit **stronger dispersion forces** than do smaller and lighter atoms and molecules. F_2 and Cl_2 are gases at room temperature (reflecting weaker attractive forces); Br_2 is a liquid, and I_2 is a solid (reflecting stronger attractive forces). Trends in observed melting and boiling points for the halogens clearly demonstrate this effect, as seen in Table 8.1.2.

Table **4.1.2**: **Melting and Boiling Points of the Halogens**

Halogen	Molar Mass	Atomic Radius	Melting Point	Boiling Point
fluorine, F_2	38 g/mol	72 pm	53 K	85 K
chlorine, Cl_2	71 g/mol	99 pm	172 K	238 K
bromine, Br_2	160 g/mol	114 pm	266 K	332 K
iodine, I_2	254 g/mol	133 pm	387 K	457 K
astatine, At_2	420 g/mol	150 pm	575 K	610 K

The increase in melting and boiling points with increasing atomic/molecular size may be rationalized by considering how the strength of dispersion forces is affected by the electronic structure of the atoms or molecules in the substance. In a larger atom, the valence electrons are, on average, farther from the nuclei than in a smaller atom. Thus, they are less tightly held and can more easily form the temporary dipoles that produce the attraction. The measure of how easy or difficult it is for another electrostatic charge (for example, a nearby ion or polar molecule) to distort a molecule's charge distribution (its electron cloud) is known as **polarizability**. A molecule

that has a charge cloud that is easily distorted is said to be very polarizable and will have large dispersion forces; one with a charge cloud that is difficult to distort is not very polarizable and will have small dispersion forces.

Example 4.1.3

Order the following compounds of a group 14 element and hydrogen from lowest to highest boiling point: CH_4, SiH_4, GeH_4, and SnH_4. Explain your reasoning.

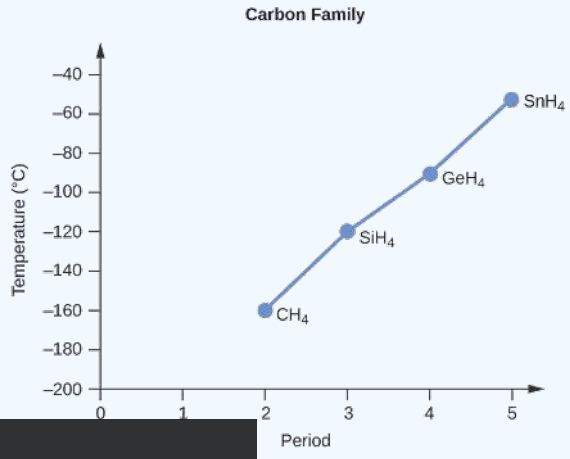
Solution

Empty alt attribute

Applying the principles discussed in the chapter on chemical bonding and molecular geometry, all of these compounds are predicted to be nonpolar, so they may experience only dispersion forces: the smaller the molecule, the less polarizable and the weaker the dispersion forces; the larger the molecule, the larger the dispersion forces. The molar masses of CH_4, SiH_4, GeH_4, and SnH_4 are approximately 16 g/mol, 32 g/mol, 77 g/mol, and 123 g/mol, respectively. Therefore, CH_4 is expected to have the lowest boiling point and SnH_4 the highest boiling point. The ordering from lowest to highest boiling point is expected to be

$CH_4 < SiH_4 < GeH_4 < SnH_4$

A graph of the actual boiling points of these compounds versus the period of the group 14 elementsshows this prediction to be correct:



Short alt text: alt

A line graph, titled “Carbon Family,” is shown where the y-axis is labeled “Temperature, (degree sign C)” and has values of “negative 200” to “negative 40” from bottom to top in increments of 20. The x-axis is labeled “Period” and has values of “0” to “5” in increments of 1. The first point on the graph is labeled “C H subscript 4” and is at point “2, negative 160.” The second point on the graph is labeled “S i H subscript 4” and is at point “3, negative 120” while the third point on the graph is labeled “G e H subscript 4” and is at point “4, negative 100.” The fourth point on the graph is labeled “S n H subscript 4” and is at point “5, negative 60.”

4.1.3 Exercise

Order the following hydrocarbons from lowest to highest boiling point: C_2H_6, C_3H_8, and C_4H_{10}.

Answer

All of these compounds are nonpolar and only have London dispersion forces: the larger the molecule, the larger the dispersion forces and the higher the boiling point. The ordering from lowest to highest boiling point is therefore

$$C_2H_6 < C_3H_8 < C_4H_{10}.$$

Applications: Geckos and Intermolecular Forces

Geckos have an amazing ability to adhere to most surfaces. They can quickly run up smooth walls and across ceilings that have no toe-holds, and they do this without having suction cups or a sticky substance on their toes. And while a gecko can lift its feet easily as it walks along a surface, if you attempt to pick it up, it sticks to the surface. How are geckos (as well as spiders and some

other insects) able to do this? Although this phenomenon has been investigated for hundreds of years, scientists only recently uncovered the details of the process that allows geckos' feet to behave this way.

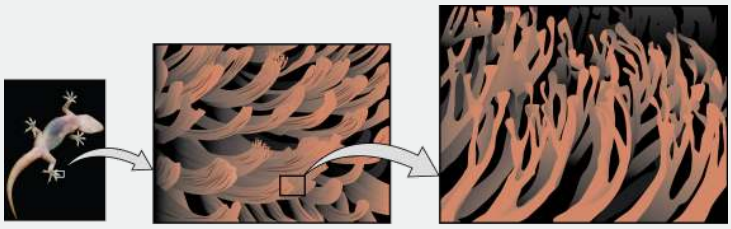


Figure **4.1.12**: *Geckos' toes contain large numbers of tiny hairs (setae), which branch into many triangular tips (spatulae). Geckos adhere to surfaces because of van der Waals attractions between the surface and a gecko's millions of spatulae. By changing how the spatulae contact the surface, geckos can turn their stickiness "on" and "off." (credit photo: modification of work by "JC*+A!"/Flickr)*

Three figures are shown. The first is a photo of the bottom of a gecko's foot. The second is bigger version which shows the setae. The third is a bigger version of the setae and shows the spatulae.

Geckos' toes are covered with hundreds of thousands of tiny hairs known as *setae*, with each seta, in turn, branching into hundreds of tiny, flat, triangular tips called *spatulae*. The huge numbers of spatulae on its setae provide a gecko, shown in Figure 8.1.12, with a large total surface area for sticking to a surface. In 2000, Kellar Autumn, who leads a multi-institutional gecko research team, found that geckos adhered equally well to both polar silicon dioxide and nonpolar gallium arsenide. This proved that geckos stick to surfaces because of dispersion forces—weak intermolecular attractions arising from temporary, synchronized charge distributions between adjacent molecules. Although dispersion forces are very weak, the total attraction over millions of spatulae is large enough to support many times the gecko's weight.

In 2014, two scientists developed a model to explain how geckos can rapidly transition from "sticky" to "non-sticky." Alex Greaney and Congcong Hu at Oregon State University described how geckos can achieve this by changing the angle between their spatulae and the surface. Geckos' feet, which are normally nonsticky, become sticky when a small shear force is applied. By curling and uncurling their toes, geckos can alternate between sticking and unsticking from a surface, and thus easily move across it. Further investigations may eventually lead to the development of better adhesives and other applications.

Boiling Points and Bonding Types

In order for a substance to enter the gas phase, its particles must completely overcome the intermolecular forces holding them together. Therefore, a comparison of boiling points is essentially equivalent to comparing the strengths of the attractive intermolecular forces exhibited by the individual molecules. For small molecular compounds, London dispersion forces are the weakest intermolecular forces. Dipole-dipole forces are somewhat stronger, and hydrogen bonding is a particularly strong form of dipole-dipole interaction. However, when the mass of a nonpolar molecule is sufficiently large, its dispersion forces can be stronger than the dipole-dipole forces in a lighter polar molecule. Thus, nonpolar Cl_2 has a higher boiling point than polar HCl.

Table **4.1.3**: Intermolecular Forces and Boiling Points

```
Empty alt attribute
```

Substance	Strongest Intermolecular Force	Boiling Point (°C)
H_2	dispersion	-253
Ne	dispersion	-246
O_2	dispersion	-183
Cl_2	dispersion	-34
HCl	dipole-dipole	-85
HBr	dipole-dipole	-66
H_2S	dipole-dipole	-61
NH_3	hydrogen bonding	-33
HF	hydrogen bonding	20
H_2O	hydrogen bonding	100

Example 4.1.4: Intermolecular Forces

What intermolecular forces besides dispersion forces, if any, exist in each substance? Are any of these substances solids at room temperature?

a. potassium chloride (KCl)
b. ethanol (C_2H_5OH)
c. bromine (Br_2)

Solution

a. Potassium chloride is composed of ions, so the intermolecular interaction in potassium chloride is ionic forces. Because ionic interactions are strong, it might be expected that potassium chloride is a solid at room temperature.
b. Ethanol has a hydrogen atom attached to an oxygen atom, so it would experience hydrogen bonding. If the hydrogen bonding is strong enough, ethanol might be a solid at room temperature, but it is difficult to know for certain. (Ethanol is actually a liquid at room temperature.)
c. Elemental bromine has two bromine atoms covalently bonded to each other. Because the atoms on either side of the covalent bond are the same, the electrons in the covalent bond are shared equally, and the bond is a nonpolar covalent bond. Thus, diatomic bromine does not have any intermolecular forces other than dispersion forces. It is unlikely to be a solid at room temperature unless the dispersion forces are strong enough. Bromine is a liquid at room temperature.

Exercise 4.1.4

What intermolecular forces besides dispersion forces, if any, exist in each substance? Are any of these substances solids at room temperature?

a. methylamine (CH_3NH_2)
b. calcium sulfate ($CaSO_4$)
c. carbon monoxide (CO)

Answer

a. dipole-dipole, hydrogen bonding
b. ionic forces (solid at room temperature)
c. dipole-dipole

Key Takeaways

- A phase is a form of matter that has the same physical properties throughout.
- Empty alt attribute
- Dipole-dipole attractions result from the electrostatic attraction of the partial negative end of one dipolar molecule for the partial positive end of another.
- Hydrogen bonds are a special type of dipole-dipole attraction that results when hydrogen is bonded to one of the three most electronegative elements: F, O, or N.
- The temporary dipole that results from the motion of the electrons in an atom can induce a dipole in an adjacent atom and give rise to the London dispersion force. London forces increase with increasing molecular size.

4.1: Intermolecular Interactions is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **8.1: Intermolecular Interactions** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

https://chem.libretexts.org/@go/page/255591

```
Empty alt attribute
```


https://chem.libretexts.org/@go/page/255591

```
Empty alt attribute
```


https://chem.libretexts.org/@go/page/255591

CHAPTER OVERVIEW

5: Organic Molecules

Organic chemistry is the study of molecules containing carbon, hydrogen, and oxygen (with a smattering of other elements). This small selection of elements gives rise to tremendously diverse structures and properties.

5.1: Structures and Names of Alkanes
5.2: Branched-Chain Alkanes
5.3: Condensed Structural and Line-Angle Formulas
5.4: IUPAC Nomenclature
5.5: Physical Properties of Alkanes
5.6: Alkenes- Structures and Names
5.7: Cis-Trans Isomers (Geometric Isomers)
5.8: Line Notation (SMILES and InChI)

5: Organic Molecules is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

5.1: Structures and Names of Alkanes

Learning Objectives

- To identify and name simple (straight-chain) alkanes given formulas and write formulas for straight-chain alkanes given their names.

We begin our study of organic chemistry with the hydrocarbons, the simplest organic compounds, which are composed of carbon and hydrogen atoms only. As we noted, there are several different kinds of hydrocarbons. They are distinguished by the types of bonding between carbon atoms and the properties that result from that bonding. Hydrocarbons with only carbon-to-carbon single bonds (C–C) and existing as a continuous chain of carbon atoms also bonded to hydrogen atoms are called alkanes (or saturated hydrocarbons). *Saturated*, in this case, means that each carbon atom is bonded to four other atoms (hydrogen or carbon)—the most possible; there are no double or triple bonds in the molecules.

> The word *saturated* has the same meaning for hydrocarbons as it does for the dietary fats and oils: the molecule has no carbon-to-carbon double bonds (C=C).

We previously introduced the three simplest alkanes—methane (CH_4), ethane (C_2H_6), and propane (C_3H_8) and they are shown again in Figure **5.1.1**.

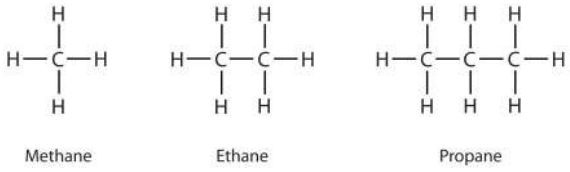


Figure **5.1.1**: The Three Simplest Alkanes

The flat representations shown do not accurately portray bond angles or molecular geometry. Methane has a tetrahedral shape that chemists often portray with wedges indicating bonds coming out toward you and dashed lines indicating bonds that go back away from you. An ordinary solid line indicates a bond in the plane of the page. Recall that the VSEPR theory correctly predicts a tetrahedral shape for the methane molecule (Figure **5.1.2**).

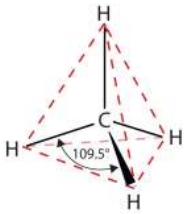


Figure **5.1.2**: The Tetrahedral Methane Molecule

Methane (CH_4), ethane (C_2H_6), and propane (C_3H_8) are the beginning of a series of compounds in which any two members in a sequence differ by one carbon atom and two hydrogen atoms—namely, a CH_2 unit. The first 10 members of this series are given in Table **5.1.1**.

Table **5.1.1**: The First 10 Straight-Chain Alkanes

Name	Molecular Formula (C_nH_{2n+2})	Condensed Structural Formula	Number of Possible Isomers
methane	CH_4	CH_4	—
ethane	C_2H_6	CH_3CH_3	—
propane	C_3H_8	$CH_3CH_2CH_3$	—
butane	C_4H_{10}	$CH_3CH_2CH_2CH_3$	2
pentane	C_5H_{12}	$CH_3CH_2CH_2CH_2CH_3$	3
hexane	C_6H_{14}	$CH_3CH_2CH_2CH_2CH_2CH_3$	5
heptane	C_7H_{16}	$CH_3CH_2CH_2CH_2CH_2CH_2CH_3$	9
octane	C_8H_{18}	$CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_3$	18
nonane	C_9H_{20}	$CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_3$	35
decane	$C_{10}H_{22}$	$CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_3$	75

Consider the series in Figure **5.1.3**. The sequence starts with C_3H_8, and a CH_2 unit is added in each step moving up the series. Any family of compounds in which adjacent members differ from each other by a definite factor (here a CH_2 group) is called a homologous series. The members of such a series, called *homologs*, have properties that vary in a regular and predictable manner. The principle of *homology* gives organization to organic chemistry in much the same way that the periodic table gives organization to inorganic chemistry. Instead of a bewildering array of individual carbon compounds, we can study a few members of a homologous series and from them deduce some of the properties of other compounds in the series.

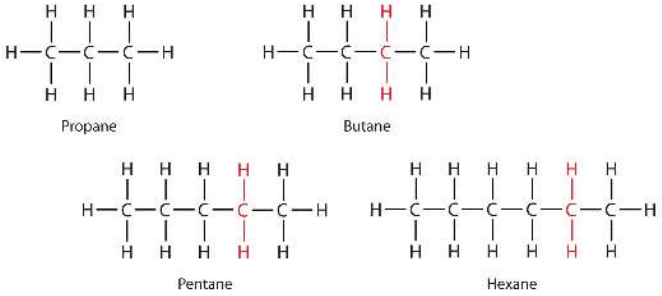


Figure **5.1.3**: Members of a Homologous Series. Each succeeding formula incorporates one carbon atom and two hydrogen atoms more than the previous formula.

The principle of homology allows us to write a general formula for alkanes: C_nH_{2n+2}. Using this formula, we can write a molecular formula for any alkane with a given number of carbon atoms. For example, an alkane with eight carbon atoms has the molecular formula $C_8H_{(2\times 8)+2} = C_8H_{18}$.

Key Takeaway

- Simple alkanes exist as a homologous series, in which adjacent members differ by a CH_2 unit.

5.1: Structures and Names of Alkanes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **12.2: Structures and Names of Alkanes** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

5.2: Branched-Chain Alkanes

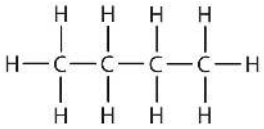
Learning Objectives

- To learn how alkane molecules can have branched chains and recognize compounds that are isomers.

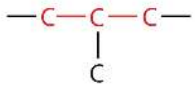
We can write the structure of butane (C_4H_{10}) by stringing four carbon atoms in a row,

–C–C–C–C–

and then adding enough hydrogen atoms to give each carbon atom four bonds:



The compound butane has this structure, but there is another way to put 4 carbon atoms and 10 hydrogen atoms together. Place 3 of the carbon atoms in a row and then branch the fourth one off the middle carbon atom:



Now we add enough hydrogen atoms to give each carbon four bonds.

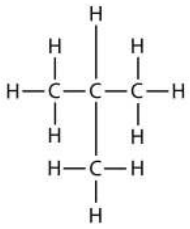


Figure **5.2.1**).

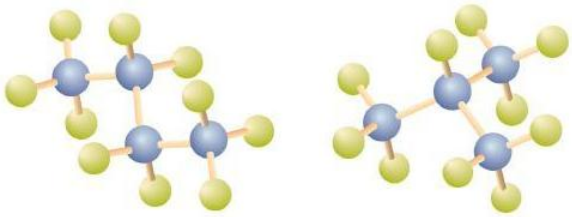
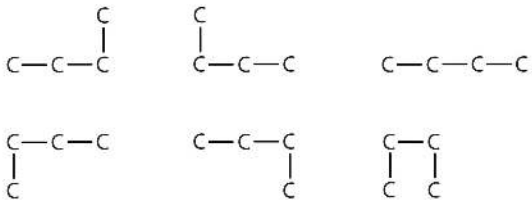


Figure **5.2.1**: Butane and Isobutane. The ball-and-stick models of these two compounds show them to be isomers; both have the molecular formula C_4H_{10}.

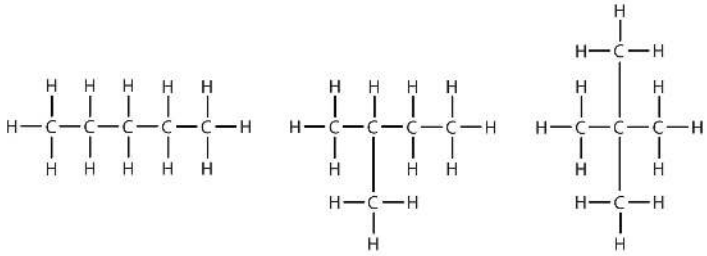
Notice that C_4H_{10} is depicted with a bent chain in Figure **5.2.1**. The four-carbon chain may be bent in various ways because the groups can rotate freely about the C–C bonds. However, this rotation does not change the identity of the compound. It is important to realize that bending a chain does *not* change the identity of the compound; all of the following represent the same compound, butane:



The structure of isobutane shows a continuous chain of three carbon atoms only, with the fourth attached as a branch off the middle carbon atom of the continuous chain, which is different from the structures of butane (compare the two structures in Figure **5.2.1**.

Unlike C_4H_{10}, the compounds methane (CH_4), ethane (C_2H_6), and propane (C_3H_8) do not exist in isomeric forms because there is only one way to arrange the atoms in each formula so that each carbon atom has four bonds.

Next beyond C_4H_{10} in the homologous series is pentane. Each compound has the same molecular formula: C_5H_{12}. (Table 12.1.1 from the previous section has a column identifying the number of possible isomers for the first 10 straight-chain alkanes.) The compound at the far left is pentane because it has all five carbon atoms in a continuous chain. The compound in the middle is isopentane; like isobutane, it has a one CH_3 branch off the second carbon atom of the continuous chain. The compound at the far right, discovered after the other two, was named neopentane (from the Greek *neos*, meaning "new"). Although all three have the same molecular formula, they have different properties, including boiling points: pentane, 36.1°C; isopentane, 27.7°C; and neopentane, 9.5°C.



A continuous (unbranched) chain of carbon atoms is often called a *straight chain* even though the tetrahedral arrangement about each carbon gives it a zigzag shape. Straight-chain alkanes are sometimes called *normal alkanes*, and their names are given the prefix *n*-. For example, butane is called *n*-butane. We will not use that prefix here because it is not a part of the system established by the International Union of Pure and Applied Chemistry.

5.2: Branched-Chain Alkanes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **12.3: Branched-Chain Alkanes** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

5.3: Condensed Structural and Line-Angle Formulas

Learning Objectives

- Write condensed structural formulas for alkanes given complete structural formulas.
- Draw line-angle formulas given structural formulas.

We use several kinds of formulas to describe organic compounds. A *molecular formula* shows only the kinds and numbers of atoms in a molecule. For example, the molecular formula C_4H_{10} tells us there are 4 carbon atoms and 10 hydrogen atoms in a molecule, but it doesn't distinguish between butane and isobutane. A structural formula shows all the carbon and hydrogen atoms and the bonds attaching them. Thus, structural formulas identify the specific isomers by showing the order of attachment of the various atoms.

Unfortunately, structural formulas are difficult to type/write and take up a lot of space. Chemists often use condensed structural formulas to alleviate these problems. The condensed formulas show hydrogen atoms right next to the carbon atoms to which they are attached, as illustrated for butane:



The ultimate condensed formula is a line-angle formula, in which carbon atoms are implied at the corners and ends of lines, and each carbon atom is understood to be attached to enough hydrogen atoms to give each carbon atom four bonds. For example, we can represent pentane ($CH_3CH_2CH_2CH_2CH_3$) and isopentane [$(CH_3)_2CHCH_2CH_3$] as follows:



Parentheses in condensed structural formulas indicate that the enclosed grouping of atoms is attached to the adjacent carbon atom.

Key Takeaways

- Condensed chemical formulas show the hydrogen atoms (or other atoms or groups) right next to the carbon atoms to which they are attached.
- Line-angle formulas imply a carbon atom at the corners and ends of lines. Each carbon atom is understood to be attached to enough hydrogen atoms to give each carbon atom four bonds.

5.3: Condensed Structural and Line-Angle Formulas is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **12.4: Condensed Structural and Line-Angle Formulas** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

5.4: IUPAC Nomenclature

Learning Objectives

- To name alkanes by the IUPAC system and write formulas for alkanes given IUPAC names

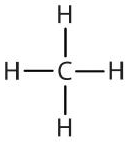
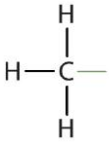
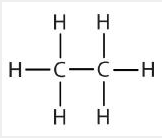
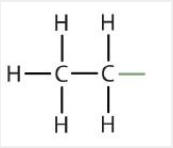
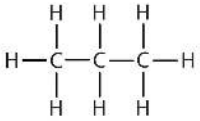
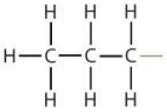
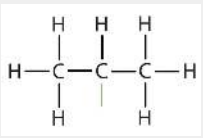
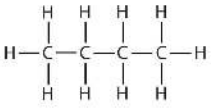
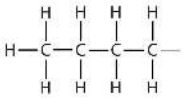
As noted in previously, the number of isomers increases rapidly as the number of carbon atoms increases. There are 3 pentanes, 5 hexanes, 9 heptanes, and 18 octanes. It would be difficult to assign unique individual names that we could remember. A systematic way of naming hydrocarbons and other organic compounds has been devised by the International Union of Pure and Applied Chemistry (IUPAC). These rules, used worldwide, are known as the IUPAC System of Nomenclature. (Some of the names we used earlier, such as isobutane, isopentane, and neopentane, do not follow these rules and are called *common names*.) A stem name (Table **5.4.1**) indicates the number of carbon atoms in the longest continuous chain (LCC). Atoms or groups attached to this carbon chain, called *substituents*, are then named, with their positions indicated by numbers. For now, we will consider only those substituents called alkyl groups.

Table **5.4.1**: Stems That Indicate the Number of Carbon Atoms in Organic Molecules

Stem	Number
meth-	1
eth-	2
prop-	3
but-	4
pent-	5
hex-	6
hept-	7
oct-	8
non-	9
dec-	10

An alkyl group is a group of atoms that results when one hydrogen atom is removed from an alkane. The group is named by replacing the *-ane* suffix of the parent hydrocarbon with *-yl*. For example, the $-CH_3$ group derived from methane (CH_4) results from subtracting one hydrogen atom and is called a *methyl group*. The alkyl groups we will use most frequently are listed in Table **5.4.2**. Alkyl groups are not independent molecules; they are parts of molecules that we consider as a unit to name compounds systematically.

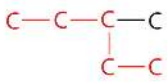
Table **5.4.2**: Common Alkyl Groups

Parent Alkane		Alkyl Group		Condensed Structural Formula
methane		methyl		$CH_3–$
ethane		ethyl		$CH_3CH_2–$
propane		propyl		$CH_3CH_2CH_2–$
		isopropyl		$(CH_3)_2CH–$
butane		butyl*		$CH_3CH_2CH_2CH_2–$

*There are four butyl groups, two derived from butane and two from isobutane. We will introduce the other three where appropriate.

Simplified IUPAC rules for naming alkanes are as follows (demonstrated in Example **5.4.1**).

1. **Name alkanes according to the LCC (longest continuous chain) of carbon atoms in the molecule (rather than the total number of carbon atoms).** This LCC, considered the parent chain, determines the base name, to which we add the suffix *-ane* to indicate that the molecule is an alkane.
2. **If the hydrocarbon is branched, number the carbon atoms of the LCC.** Numbers are assigned in the direction that gives the lowest numbers to the carbon atoms with attached substituents. Hyphens are used to separate numbers from the names of substituents; commas separate numbers from each other. (The LCC need not be written in a straight line; for example, the LCC in the following has five carbon atoms.)

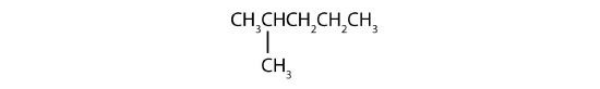
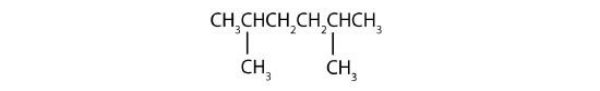
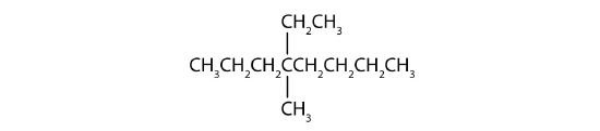


3. **Place the names of the substituent groups in alphabetical order before the name of the parent compound.** If the same alkyl group appears more than once, the numbers of all the carbon atoms to which it is attached are expressed. If the same group appears more than once on the same carbon atom, the number of that carbon atom is repeated as many times as the group appears. Moreover, the number of identical groups is indicated by the Greek prefixes *di-*, *tri-*, *tetra-*, and so on. These prefixes are *not* considered in determining the alphabetical order of the substituents. For example, ethyl is listed before dimethyl; the di- is simply ignored. The last alkyl group named is prefixed to the name of the parent alkane to form one word.

When these rules are followed, every unique compound receives its own exclusive name. The rules enable us to not only name a compound from a given structure but also draw a structure from a given name. The best way to learn how to use the IUPAC system is to put it to work, not just memorize the rules. It's easier than it looks.

Example 5.4.1

Name each compound.

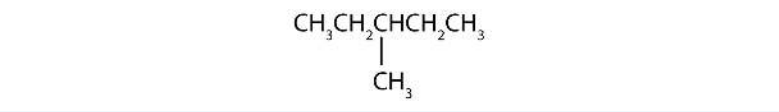
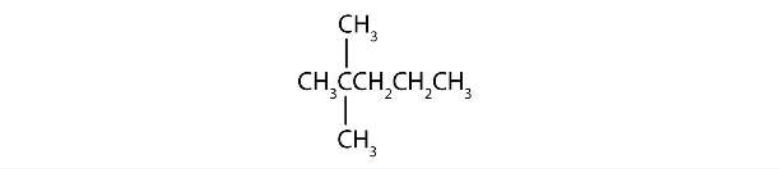
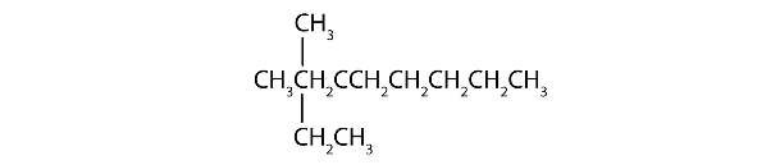
1. 
2. 
3. 

Solution

1. The LCC has five carbon atoms, and so the parent compound is pentane (rule 1). There is a methyl group (rule 2) attached to the second carbon atom of the pentane chain. The name is therefore 2-methylpentane.
2. The LCC has six carbon atoms, so the parent compound is hexne (rule 1). Methyl groups (rule 2) are attached to the second and fifth carbon atoms. The name is 2,5-dimethylhexane.
3. The LCC has eight carbon atoms, so the parent compound is octane (rule 1). There are methyl and ethyl groups (rule 2), both attached to the fourth carbon atom (counting from the *right* gives this carbon atom a lower number; rule 3). The correct name is thus 4-ethyl-4-methyloctane.

Exercise 5.4.1

Name each compound.

1. 
2. 
3. 

Example 5.4.2

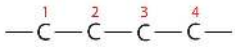
Draw the structure for each compound.

a. 2,3-dimethylbutane
b. 4-ethyl-2-methylheptane

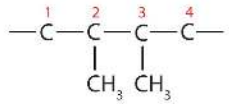
Solution a

In drawing structures, always start with the parent chain.

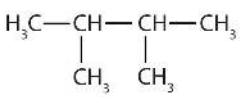
The parent chain is butane, indicating four carbon atoms in the LCC.



Then add the groups at their proper positions. You can number the parent chain from either direction as long as you are consistent; just don't change directions before the structure is done. The name indicates two methyl (CH_3) groups, one on the second carbon atom and one on the third.

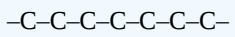


Finally, fill in all the hydrogen atoms, keeping in mind that each carbon atom must have four bonds.

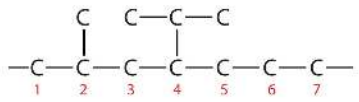


Solution b

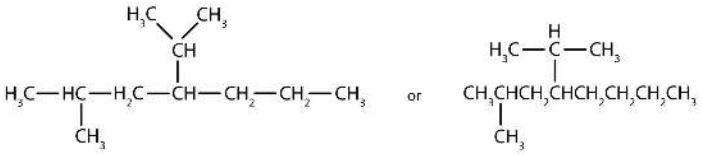
The parent chain is heptane in this case, indicating seven carbon atoms in the LCC.



Adding the groups at their proper positions gives



Filling in all the hydrogen atoms gives the following condensed structural formulas:



Note that the bonds (dashes) can be shown or not; sometimes they are needed for spacing.

Exercise 5.4.2

Draw the structure for each compound.

a. 4-ethyloctane
b. 3-ethyl-2-methylpentane
c. 3,3,5-trimethylheptane

Key Takeaway

- Alkanes have both common names and systematic names, specified by IUPAC.

5.4: IUPAC Nomenclature is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **12.5: IUPAC Nomenclature** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

5.5: Physical Properties of Alkanes

Learning Objectives

- To identify the physical properties of alkanes and describe trends in these properties.

Because alkanes have relatively predictable physical properties and undergo relatively few chemical reactions other than combustion, they serve as a basis of comparison for the properties of many other organic compound families. Let's consider their physical properties first.

Table **5.5.1** describes some of the properties of some of the first 10 straight-chain alkanes. Because alkane molecules are nonpolar, they are insoluble in water, which is a polar solvent, but are soluble in nonpolar and slightly polar solvents. Consequently, alkanes themselves are commonly used as solvents for organic substances of low polarity, such as fats, oils, and waxes. Nearly all alkanes have densities less than 1.0 g/mL and are therefore less dense than water (the density of H_2O is 1.00 g/mL at 20°C). These properties explain why oil and grease do not mix with water but rather float on its surface.

Table **5.5.1**: Physical Properties of Some Alkanes

Molecular Name	Formula	Melting Point (°C)	Boiling Point (°C)	Density (20°C)*	Physical State (at 20°C)
methane	CH_4	–182	–164	0.668 g/L	gas
ethane	C_2H_6	–183	–89	1.265 g/L	gas
propane	C_3H_8	–190	–42	1.867 g/L	gas
butane	C_4H_{10}	–138	–1	2.493 g/L	gas
pentane	C_5H_{12}	–130	36	0.626 g/mL	liquid
hexane	C_6H_{14}	–95	69	0.659 g/mL	liquid
octane	C_8H_{18}	–57	125	0.703 g/mL	liquid
decane	$C_{10}H_{22}$	–30	174	0.730 g/mL	liquid

*Note the change in units going from gases (grams per liter) to liquids (grams per milliliter). Gas densities are at 1 atm pressure.



Figure 5.5.1: Oil Spills. Crude oil coats the water's surface in the Gulf of Mexico after the *Deepwater Horizon* oil rig sank following an explosion. The leak was a mile below the surface, making it difficult to estimate the size of the spill. One liter of oil can create a slick 2.5 hectares (6.3 acres) in size. This and similar spills provide a reminder that hydrocarbons and water don't mix. Source: Photo courtesy of NASA Goddard / MODIS Rapid Response Team, NASA.gov, Topics, Earth Features, Oil Spill(opens in new window) [www.nasa.gov].

Looking Closer: Gas Densities and Fire Hazards

Table 5.5.1 indicates that the first four members of the alkane series are gases at ordinary temperatures. Natural gas is composed chiefly of methane, which has a density of about 0.67 g/L. The density of air is about 1.29 g/L. Because natural gas is less dense than air, it rises. When a natural-gas leak is detected and shut off in a room, the gas can be removed by opening an upper window. On the other hand, bottled gas can be either propane (density 1.88 g/L) or butanes (a mixture of butane and isobutane; density about 2.5 g/L). Both are much heavier than air (density 1.2 g/L). If bottled gas escapes into a building, it collects near the floor. This presents a much more serious fire hazard than a natural-gas leak because it is more difficult to rid the room of the heavier gas.

Also shown in Table 5.5.1 are the boiling points of the straight-chain alkanes increase with increasing molar mass. This general rule holds true for the straight-chain homologs of all organic compound families. Larger molecules have greater surface areas and consequently interact more strongly; more energy is therefore required to separate them. For a given molar mass, the boiling points of alkanes are relatively low because these nonpolar molecules have only weak dispersion forces to hold them together in the liquid state.

Looking Closer: An Alkane Basis for Properties of Other Compounds

An understanding of the physical properties of the alkanes is important in that petroleum and natural gas and the many products derived from them—gasoline, bottled gas, solvents, plastics, and more—are composed primarily of alkanes. This understanding is also vital because it is the basis for describing the properties of other organic and biological compound families. For example, large portions of the structures of lipids consist of nonpolar alkyl groups. Lipids include the dietary fats and fatlike compounds called phospholipids and sphingolipids that serve as structural components of living tissues. These compounds have both polar and nonpolar groups, enabling them to bridge the gap between water-soluble and water-insoluble phases. This characteristic is essential for the selective permeability of cell membranes.

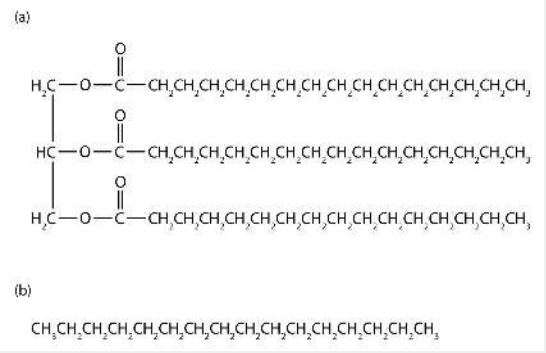


Figure 5.5.2: Tripalmitin (a), a typical fat molecule, has long hydrocarbon chains typical of most lipids. Compare these chains to hexadecane (b), an alkane with 16 carbon atoms

Key Takeaway

- Alkanes are nonpolar compounds that are low boiling and insoluble in water.

5.5: Physical Properties of Alkanes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

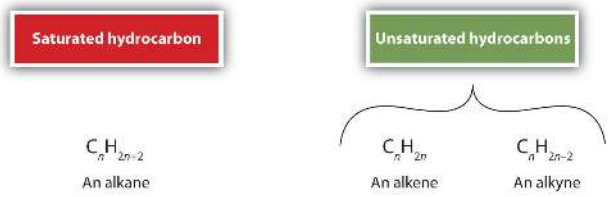
- **12.6: Physical Properties of Alkanes** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

5.6: Alkenes- Structures and Names

Learning Objectives

- To name alkenes given formulas and write formulas for alkenes given names.

As noted before, alkenes are hydrocarbons with carbon-to-carbon double bonds ($R_2C{=}CR_2$) and alkynes are hydrocarbons with carbon-to-carbon triple bonds (R–C≡C–R). Collectively, they are called unsaturated hydrocarbons because they have fewer hydrogen atoms than does an alkane with the same number of carbon atoms, as is indicated in the following general formulas:

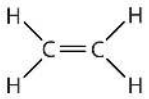


Some representative alkenes—their names, structures, and physical properties—are given in Table **5.6.1**.

Table **5.6.1**: Physical Properties of Some Selected Alkenes

IUPAC Name	Molecular Formula	Condensed Structural Formula	Melting Point (°C)	Boiling Point (°C)
ethene	C_2H_4	$CH_2{=}CH_2$	–169	–104
propene	C_3H_6	$CH_2{=}CHCH_3$	–185	–47
1-butene	C_4H_8	$CH_2{=}CHCH_2CH_3$	–185	–6
1-pentene	C_5H_{10}	$CH_2{=}CH(CH_2)_2CH_3$	–138	30
1-hexene	C_6H_{12}	$CH_2{=}CH(CH_2)_3CH_3$	–140	63
1-heptene	C_7H_{14}	$CH_2{=}CH(CH_2)_4CH_3$	–119	94
1-octene	C_8H_{16}	$CH_2{=}CH(CH_2)_5CH_3$	–102	121

We used only condensed structural formulas in Table **5.6.1**. Thus, $CH_2{=}CH_2$ stands for



The double bond is shared by the two carbons and does not involve the hydrogen atoms, although the condensed formula does not make this point obvious. Note that the molecular formula for ethene is C_2H_4, whereas that for ethane is C_2H_6.

The first two alkenes in Table **5.6.1**, ethene and propene, are most often called by their common names—ethylene and propylene, respectively (Figure **5.6.1**). Ethylene is a major commercial chemical. The US chemical industry produces about 25 billion kilograms of ethylene annually, more than any other synthetic organic chemical. More than half of this ethylene goes into the manufacture of polyethylene, one of the most familiar plastics. Propylene is also an important industrial chemical. It is converted to plastics, isopropyl alcohol, and a variety of other products.

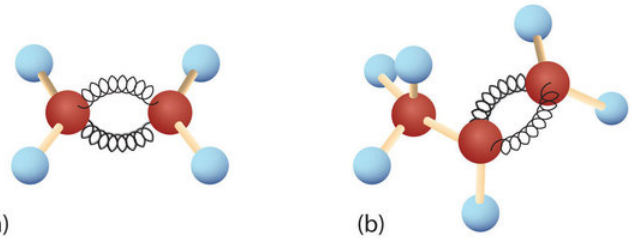
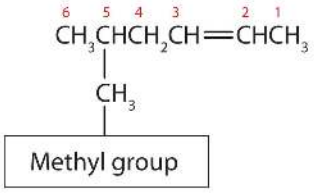


Figure **5.6.1**: Ethene and Propene. The ball-and-spring models of ethene/ethylene (a) and propene/propylene (b) show their respective shapes, especially bond angles.

Although there is only one alkene with the formula C_2H_4 (ethene) and only one with the formula C_3H_6 (propene), there are several alkenes with the formula C_4H_8.

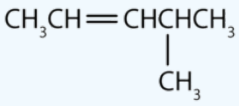
Here are some basic rules for naming alkenes from the International Union of Pure and Applied Chemistry (IUPAC):

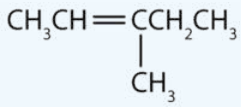
1. The longest chain of carbon atoms *containing the double bond* is considered the parent chain. It is named using the same stem as the alkane having the same number of carbon atoms but ends in *-ene* to identify it as an alkene. Thus the compound CH_2=$CHCH_3$ is *propene*.
2. If there are four or more carbon atoms in a chain, we must indicate the position of the double bond. The carbons atoms are numbered so that the first of the two that are doubly bonded is given the lower of the two possible numbers.The compound CH_3CH=$CHCH_2CH_3$, for example, has the double bond between the second and third carbon atoms. Its name is 2-pentene (not 3-pentene).
3. Substituent groups are named as with alkanes, and their position is indicated by a number. Thus, the structure below is 5-methyl-2-hexene. Note that the numbering of the parent chain is always done in such a way as to give the double bond the lowest number, even if that causes a substituent to have a higher number. *The double bond always has priority in numbering.*



Example 5.6.1

Name each compound.

a. 

b. 

Solution

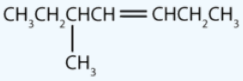
a. The longest chain containing the double bond has five carbon atoms, so the compound is a *pentene* (rule 1). To give the first carbon atom of the double bond the lowest number (rule 2), we number from the left, so the compound is a 2-pentene. There is a methyl group on the fourth carbon atom (rule 3), so the compound's name is 4-methyl-2-pentene.

b. The longest chain containing the double bond has five carbon atoms, so the parent compound is a *pentene* (rule 1). To give the first carbon atom of the double bond the lowest number (rule 2), we number from the left, so the compound is a 2-pentene. There is a methyl group on the third carbon atom (rule 3), so the compound's name is 3-methyl-2-pentene.

Exercise 5.6.1

Name each compound.

a. $CH_3CH_2CH_2CH_2CH_2CH{=}CHCH_3$

b. 

Answer

Just as there are cycloalkanes, there are *cycloalkenes*. These compounds are named like alkenes, but with the prefix *cyclo-* attached to the beginning of the parent alkene name.

Example 5.6.2

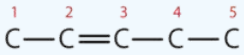
Draw the structure for each compound.

a. 3-methyl-2-pentene
b. cyclohexene

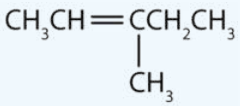
Solution

a.

First write the parent chain of five carbon atoms: C–C–C–C–C. Then add the double bond between the second and third carbon atoms:

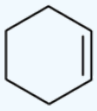


Now place the methyl group on the third carbon atom and add enough hydrogen atoms to give each carbon atom a total of four bonds.



b

First, consider what each of the three parts of the name means. *Cyclo* means a ring compound, *hex* means 6 carbon atoms, and *-ene* means a double bond.



Exercise 5.6.2

Draw the structure for each compound.

a. 2-ethyl-1-hexene
b. cyclopentene

Key Takeaway

- Alkenes are hydrocarbons with a carbon-to-carbon double bond.

5.6: Alkenes- Structures and Names is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **13.1: Alkenes- Structures and Names** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

5.7: Cis-Trans Isomers (Geometric Isomers)

Learning Objectives

- Recognize that alkenes that can exist as cis-trans isomers.
- Classify isomers as cis or trans.
- Draw structures for cis-trans isomers given their names.

There is free rotation about the carbon-to-carbon single bonds (C–C) in alkanes. In contrast, the structure of alkenes requires that the carbon atoms of a double bond and the two atoms bonded to each carbon atom all lie in a single plane, and that each doubly bonded carbon atom lies in the center of a triangle. This part of the molecule's structure is rigid; rotation about doubly bonded carbon atoms is *not* possible without rupturing the bond. Look at the two chlorinated hydrocarbons in Figure **5.7.1**.

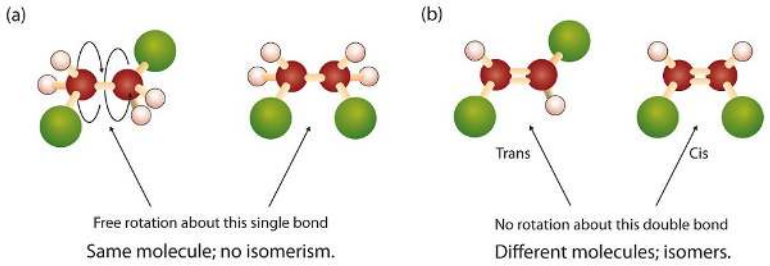


Figure **5.7.1**: *Rotation about Bonds. In 1,2-dichloroethane (a), free rotation about the C–C bond allows the two structures to be interconverted by a twist of one end relative to the other. In 1,2-dichloroethene (b), restricted rotation about the double bond means that the relative positions of substituent groups above or below the double bond are significant.*

In 1,2-dichloroethane (part (a) of Figure **5.7.1**), there is free rotation about the C–C bond. The two models shown represent exactly the same molecule; they are *not* isomers. You can draw structural formulas that look different, but if you bear in mind the possibility of this free rotation about single bonds, you should recognize that these two structures represent the same molecule:



In 1,2-dichloroethene (Figure **5.7.1***b*), however, restricted rotation about the double bond means that the relative positions of substituent groups above or below the double bond become significant. This leads to a special kind of isomerism. The isomer in which the two chlorine (Cl) atoms lie on the same side of the molecule is called the cis isomer (Latin *cis*, meaning "on this side") and is named cis-1,2-dichloroethene. The isomer with the two Cl atoms on opposite sides of the molecule is the trans isomer (Latin *trans*, meaning "across") and is named *trans*-1,2-dichloroethene. These two compounds are cis-trans isomers (or geometric isomers), compounds that have different configurations (groups permanently in different places in space) because of the presence of a rigid structure in their molecule.

Consider the alkene with the condensed structural formula $CH_3CH{=}CHCH_3$. We could name it 2-butene, but there are actually two such compounds; the double bond results in cis-trans isomerism (Figure **5.7.2**).

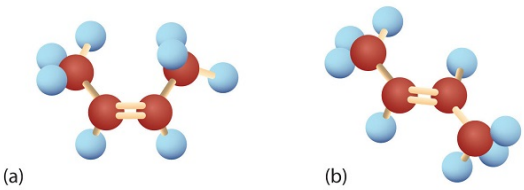


Figure **5.7.2**: Ball-and-Spring Models of (a) Cis-2-Butene and (b) Trans-2-Butene. Cis-trans isomers have different physical, chemical, and physiological properties.

Cis-2-butene has both methyl groups on the same side of the molecule. *Trans*-2-butene has the methyl groups on opposite sides of the molecule. Their structural formulas are as follows:



***Figure* 5.7.3**: *Models of (left) Cis-2-Butene and (right) Trans-2-Butene.*

Note, however, that the presence of a double bond does **not** necessarily lead to cis-trans isomerism (Figure **5.7.4**). We can draw two *seemingly* different propenes:



***Figure* 5.7.4**: *Different views of the propene molecule (flip vertically). These are not isomers.*

However, these two structures are not really different from each other. If you could pick up either molecule from the page and flip it over top to bottom, you would see that the two formulas are identical. Thus there are two requirements for cis-trans isomerism:

1. Rotation must be restricted in the molecule.
2. There must be **two nonidentical groups** on **each** doubly bonded carbon atom.

In these propene structures, the second requirement for cis-trans isomerism is not fulfilled. One of the doubly bonded carbon atoms does have two different groups attached, but the rules require that *both* carbon atoms have two different groups. In general, the following statements hold true in cis-trans isomerism:

- Alkenes with a $C{=}CH_2$ unit do not exist as cis-trans isomers.
- Alkenes with a $C{=}CR_2$ unit, where the two R groups are the same, do not exist as cis-trans isomers.
- Alkenes of the type R–CH=CH–R can exist as cis and trans isomers; cis if the two R groups are on the same side of the carbon-to-carbon double bond, and trans if the two R groups are on opposite sides of the carbon-to-carbon double bond.

Advanced Note: E/Z Isomerization

If a molecule has a C=C bond with one non-hydrogen group attached to each of the carbons, cis/trans nomenclature descried above is enough to describe it. However, if you have three different groups (or four), then the cis/trans approach is insufficient to describe the different isomers, since we do not know which two of the three groups are being described. For example, if you have a C=C bond, with a methyl group and a bromine on one carbon , and an ethyl group on the other, it is neither trans nor cis, since it is not clear whether the ethyl group is trans to the bromine or the methyl. This is addressed with a more advanced E,Z Convention [E] [E] [E] discussed elsewhere.

Cis-trans isomerism also occurs in cyclic compounds. In ring structures, groups are unable to rotate about any of the ring carbon–carbon bonds. Therefore, groups can be either on the same side of the ring (cis) or on opposite sides of the ring (trans). For our purposes here, we represent all cycloalkanes as planar structures, and we indicate the positions of the groups, either above or below the plane of the ring.



Example 5.7.1

Which compounds can exist as cis-trans (geometric) isomers? Draw them.

1. CHCl=CHBr
2. CH_2=$CBrCH_3$
3. $(CH_3)_2C$=$CHCH_2CH_3$
4. CH_3CH=$CHCH_2CH_3$

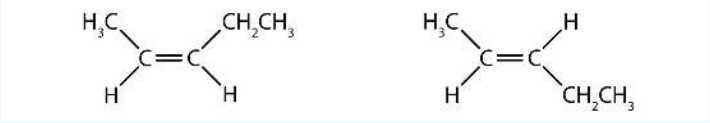
Solution

All four structures have a double bond and thus meet rule 1 for cis-trans isomerism.

1. This compound meets rule 2; it has two nonidentical groups on *each* carbon atom (H and Cl on one and H and Br on the other). It exists as both cis and trans isomers:

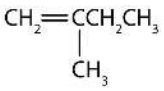


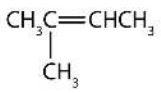
2. This compound has two hydrogen atoms on one of its doubly bonded carbon atoms; it fails rule 2 and does not exist as cis and trans isomers.
3. This compound has two methyl (CH_3) groups on one of its doubly bonded carbon atoms. It fails rule 2 and does not exist as cis and trans isomers.
4. This compound meets rule 2; it has two nonidentical groups on *each* carbon atom and exists as both cis and trans isomers:



Exercise 5.7.1

Which compounds can exist as cis-trans isomers? Draw them.

a. CH_2=$CHCH_2CH_2CH_3$
b. CH_3CH=$CHCH_2CH_3$
c. CH_3CH_2CH=$CHCH_2CH_3$
d. 

e. 

Key Takeaway

- Cis-trans (geometric) isomerism exists when there is restricted rotation in a molecule and there are two nonidentical groups on *each* doubly bonded carbon atom.

5.7: Cis-Trans Isomers (Geometric Isomers) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **13.2: Cis-Trans Isomers (Geometric Isomers)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

5.8: Line Notation (SMILES and InChI)

Learning Objectives:

- Explain what SMILES, SMARTS and SMIRKS are.
- Explain what InChI and InChIKey are.
- Review SMILES specification rules.
- Compare and contrast SMILES and InChI.
- Demonstrate how to interpret SMILES, SMARTS, InChI strings into their corresponding chemical structures.

Introduction

Line notations represent structures as a linear string of characters. They are widely used in Cheminformatics because computers can easily process linear strings of data. Examples of line notations include the Wiswesser Line-Formula Notation (WLN)[1], Sybyl Line Notation (SLN)[2,3] and Representation of structure diagram arranged linearly (ROSDAL)[4,5]. Currently, the most widely used linear notations are the Simplified Molecular-Input Line-Entry System (SMILES)[6-9] and the IUPAC Chemical Identifier (InChI)[10-13], which are described below. In this class we will focus on SMILES and InChI line notation.

SMILES

The **Simplified Molecular-Input Line-Entry System** (SMILES)[6-9] is a line notation for describing chemical structures using short ASCII strings. SMILES is like a connection table in that it identifies the nodes and edges of a molecular graph. SMILES was developed in the late 1980s and implemented by Daylight Chemical Information Systems (Santa Fe, NM), but it is still widely used today. A detailed information on SMILES can be found in Chapter 3[14] of the Daylight Theory Manual as well as the SMILES tutorial[15].

SMILES Specification Rules

In SMILES, hydrogen are typically implicitly implied and atoms are represented by their atomic symbol enclosed in brackets unless they are elements of the “organic subset” (B, C, N, O, P, S, F, Cl, Br, and I), which do not require brackets unless they are charged. So gold would be [Au] but chlorine would be Cl. If hydrogens are explicitly implied brackets are used. A formal charge is represented by one of the symbols + or -. Single, double, triple, and aromatic bonds are represented by the symbols, -, =, #, and :, respectively. Single and aromatic bonds may be, and usually are, omitted. Here are some examples of SMILES strings.

Table **5.8.1**: Common notations used in SMILES strings, note, *single and aromatic bonds are often omitted

Function	Symbol		Function	Symbol
single bond*	-		Positive charge	[C+]
double bond	=		Negative charge	[C-]
triple bond	#		aromatic carbon	c (lower case c)
aromatic bond*	:			

Table 2.4.2 shows some common SMILES strings. Note the following conventions

- **Branches** are specified by enclosures in parentheses and can be nested or stacked, as shown in these examples.

- **Rings** are represented by breaking one single or aromatic bond in each ring, and designating this ring-closure point with a digit immediately following the atoms connected through the broken bond. Atoms in aromatic rings are specified by lower cases letters.
- **Aromatic Rings** use lower case c
- Although the carbon-carbon bonds in these two SMILES are omitted, it is possible to deduce that the omitted bonds are single bonds (for cyclohexane) and aromatic bonds (for benzene). One can also represent an aromatic compound as a non-aromatic, KeKulé structure. For example, the following is a valid SMILES string for benzene.
- C1=CC=CC=C1 Benzene (C6H6)

Table 5.8.2: Smiles Strings for some common molecules, note there are several ways to represent aromaticity

SMILES	Name (formula)		SMILES	Name (formula)		SMILES	Name(formula)
C	Methane (CH_4)		COC	Dimethyl ether (CH_3OCH_3)		CC(C)CO	Isobutyl alcohol (CH_3-$CH(CH_3)$-CH_2-OH)
CC	Ethane (CH_3CH_3)		CCO	Ethanol (CH_3CH_2OH)		CC(CCC(=O)N)CN	5-amino-4-methylpentanamide
C=C	Ethene (CH_2CH_2)		CC=O	Acetaldehyde (CH_3-CH=O)		C1CCCCC1	Cyclohexane (C_6H_{12})
C#C	Ethyne (CHCH)		CC(=O)[O-]	Acetate		c1ccccc1	Benzene (C_6H_6) (aromatic representation)
C#N	Hydrogen Cyanide (HCN)		[C-]#N	Cyanide anion		C1=CC=CC=C1	Benzene (C_6H_6) (KeKulé representation)

Note that aromaticity is not a measurable physical quantity, but a concept without a unanimous mathematical definition. As a result, different aromaticity detection algorithms often disagree with each other on whether a given molecule is aromatic or not, *making it difficult to interchange information between databases that use different aromaticity detection algorithms for SMILES generation.*

Also note that a ring structure can have multiple potential ring-closure points. For example, a six-membered ring has six bonds, each of which can be a ring-closure point. As a result, a ring compound may be represented by many different but equally valid SMILES strings. Actually, it is very common that there are a lot of SMILES strings that represent the same structure, whether it has a ring or not, because one can start with any atom in a molecule to derive a SMILES string. Therefore, it is necessary to select a "unique SMILES" for a molecule among many possibilities. Because this is done through a process called "canonicalization", this unique SMILES string is also called the "canonical SMILES".

Isomeric SMILES

Isomeric SMILES allow for the specification of the isotopism and stereochemistry of a molecule. Information on isotopism is indicated by the integral atomic mass preceding the atomic symbol. The atomic mass must be specified inside square brackets. For example, C-13 methane can be represented by "[13CH4]". Configuration around double bonds is specified by "directional bonds" (characters / and \). For example, E- and Z-1,2-difluoroethene can be represented by the following isomeric SMILES:

- F/C=C/F or F\C=C\F (E)-1,2-difluoroethene (trans isomer)
- F/C=C\F or F\C=C/F (Z)-1,2-difluoroethene (cis isomer)

Configuration around tetrahedral centers are indicated by the symbols “@” or “@@”

- C[C@@H](C(=O)O)N L-Alanine
- C[C@H](C(=O)O)N D-Alanine

More detailed information on chirality specification can be found in Chapter 3[14] of the Daylight Theory Manual.

Limitations of SMILES

SMILES is proprietary and it is not an open project. This has led different chemical software developers to use different SMILES generation algorithms, resulting in different SMILES versions for the same compound. Therefore, SMILES strings obtained from different databases or research groups are not interchangeable unless they used the same software to generate the SMILES strings. With an aim to address this interchangeability issue of SMILES, an open-source project has launched to develop an open, standard version of the SMILES language called OpenSMILES. However, the most noticeable community effort in this area is development of InChI, which is described in next section.

SMARTS

SMiles **AR**bitrary **T**arget **S**pecification (SMARTS) notation allows one to search in certain databases (like PubChem) for generic structures. It is a language used for describing molecular patterns. SMARTS is useful for substructure searching, which finds a particular pattern (subgraph) in a molecule. SMARTS are straightforward extensions of SMILES. All SMILES symbols and properties are legal in SMARTS. SMARTS includes logical operators and additional molecular descriptors. Detailed information on SMARTS is given in the SMARTS specification document in the Daylight theory manual and SMARTS tutorial.

SMIRKS

Another extension of SMILES is SMIRKS, which is a line notation for generic reactions. A generic reaction represents a group of reactions that undergo the same set of atom and bond changes. Note that SMILES and SMARTS can be used to represent reactions, using the “>” symbol between the reactants, products, and agents, as described in the SMILES and SMARTS specification documents. (Therefore, these SMILES and SMARTS that describe reactions are often called reaction SMILES and reaction SMARTS, respectively.) On the other hand, SMIRKS is used to represent *types* of reactions (e.g., S_N2 reaction). More detailed information on SMIRKS is given in the SMIRKS specification document and SMIRKS tutorial.

InChI

Since 1919 the International Union of Pure and Applied Chemistry (IUPAC) has been the international authority on chemical nomenclature and terminology. IUPAC currently consists of members from 57 national adhering organizations (NAOs) whose recommendations are made public through the IUPAC journal Pure and Applied Chemistry and the IUPAC Color books. As we entered the new millennial the leadership of IUPAC recognized the need to extend chemical nomenclature into the digital realm of computer databases and software agents, and in March of 2000 during a meeting at the U.S. Naval Academy started a project with the U.S. National Institute of Standards and Technology (NIST), to build a machine readable nomenclature standard, the InChI. It was originally called INChI for IUPAC/NIST Chemical Identifier, but was changed to InChI (International Chemical Identifier) as although it was built with efforts from NIST, it was not appropriate for NIST as a government agency to place its name as a recommendation for the identifier. In 2010 the InChI Trust was formed and development of the standard is continuing the purview of the InChI Trust and the IUPAC InChI subcommittee.

InChI is an open, freely available non-proprietary computer generated chemical identifier that is based on a hierarchical layered line notation (see below). The first three layers essentially deal with the information within the simplified connection table, and the additional layers are added as needed, and deal with complexities like isomers, isotopic distributions and the other types of issues brought up in section 2.3 of this chapter, and these layers are extensible. A standard InChI has a predefined number of layers, and these can be extended to non-standard InChI's that can have new layers relating to define additional information, that is what is meant by extensible layes. Unlike SMILES, InChI is a canonical line notation and so is a unique identifier that is built upon a set of nomenclature rules. That is, although there are canonical SMILES built through a canonicalization algorithm, there can be more than one canonicalization algorithm for SMILES, and so you can have more than one SMILES string for the same structure.

Students may be familiar with the American Chemical Society's Chemical Abstract Service (CAS) registry number, which is supposed to be a unique identifier based on the registry system, but issues can arise (see below, other identifiers, and problem). Also, a CAS registry number is associated with a compound that has been published in the primary literature or patents, and the CAS system bases its identifiers on the registry system, not the structure of the molecule. That is, InChI is not a registry system, it is a type of nomenclature that describes the structure of a molecule, and you can make an InChI for a molecule that does not exist, as long as you specify its structure.

The most recent version of InChI (and its documentation) can be obtained at the InChI Trust Download site.

InChI: A Layered Notation

The power of a layered notation is that it gets to the essence of what is a molecule? For example, we think water is H_2O, but if you look at a real sample of water you will note that some of the hydrogens are protium (one proton) and others are deuterium (a proton and a neutron), and in fact the ratio of deuterium to protium in ground water samples can vary from one region of the US to another, and thus the molar mass of samples of water can vary. In fact IUPAC has now adopted an "interval atomic weight notation system" for some elements whose atomic mass varies across samples, and this can affect physical properties of a sample, like the vapor pressure of water. So water is water, but not all water is the same, and the question becomes, do you care? If you are uploading data to a database and know the isotopic distribution you care, but if you do not know it, you do not care, but in both cases, your data deals with water. Through a layered notation system you can have an isotopic layer to describe your water if you care, but you don't need it if you don't care. This leads to one of the issues that comes up with the layered notation, in that you can have different InChI's for a compound, depending on the kind of information you want in the name, that is how many and what kind of layers you use. This leads to the standard InChI, which is an InChI that has defined layers and is thus canonical, and starts with a number followed by the letter "s" to indicate the version of the standard.

Standard InChI

Standard InChI version 1.05 was released in January 2017 and has 6 core layers (and several sublayers within the core layers) and starts with InChI=1S/... . Each layer in the InChI string is separated by a "/" and the "main layer" is essentially the connection table. The InChI software generates both standard and nonstandard InChI, with the standard InChI having "fixed options" that ensures interoperability between databases and software agents. The standard InChI (version 1.05) has the following layers:

1. Main Layer
    1. Chemical Formula Layer (based on Hill Notation)
    2. Connections- bonds between atoms and may have sublayers, with the last one dealing with mobile hydrogens.
2. Charge Layer
    1. Component Charge
    2. Protons
3. Stereochemical Layer
    1. Double Bond sp^2 (Z/E) Sterochemistry
    2. Tetrahedral Sterochemistry
4. Isotopic Layer
5. Fixed Hydrogen Layer (binds mobile hydrogens)
6. Polymer Layer (actually a new experimental layer) and does not affect the content of the earlier layers.

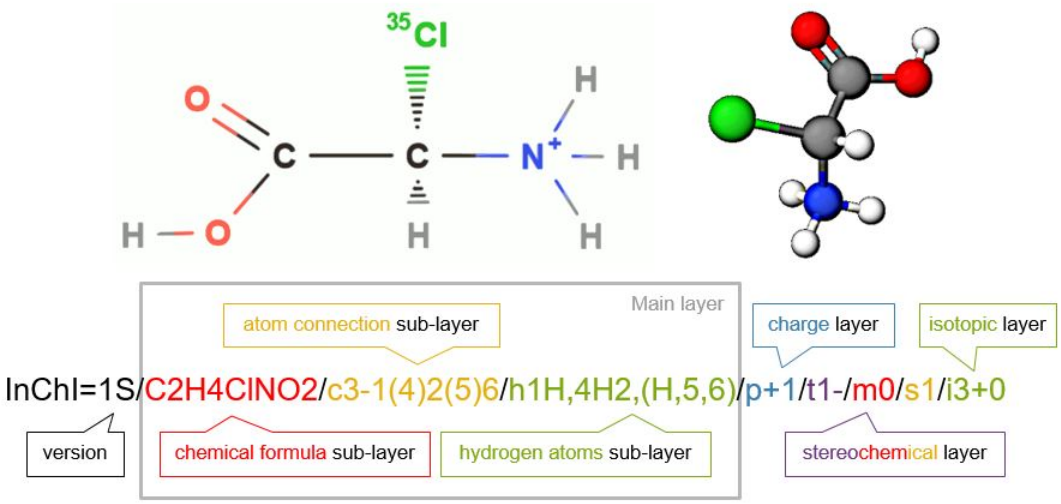


Figure 5.8.1: The main layers for a standard InChI of [(R)-carboxy(chloro)methyl]azanium, the protonated form of 2-(^{35}Cl)chloro-R-glycine . Note each layer or sublayer is separated by a forward slash [/].

NonStandard InChI

Note a nonstandard InChI does not start with InChI=1S/... but with a InChI=1/.... and has additional layers that approach different facets of a molecule's structure or features. In fact a company could create their own lawyer for a nonstandard information and encode into it proprietary information that they wished to keep private. The nonstandard InChI may not be canonical, but can handle facets of information information that a standard can not, in fact for a standard to be canonical different tautomers must have the same InChI, or you have two InChIs for the same molecule

So defining specific tautomers is one use of a nonstandard InChI as can be seen in the case of 4,5-Dihydro-1,3-Oxazol-3-lum. Figure 2 shows the two tautomeric forms of this molecule which must have the same standard InChI or it would not be canonical (you would have two InChIs for the same molecules). If you want to define just one of the tautomers, you need to use a nonstandard InChI and add a fixed hydrogen layer (in red). Although these are two ways of drawing the same molecule, one form may be favored over the other in certain environments and so there may be data indicative of the behavior of one of these form and not the other, and thus there may be a need to distinguish between the tautomers.

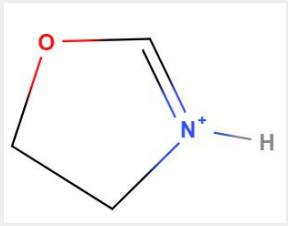
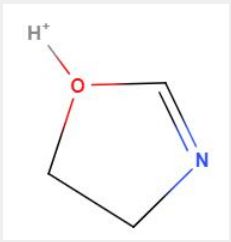
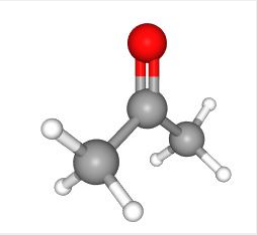
InChI=1S/C3H5NO/c1-2-5-3-4-1/h3H,1-2H2/p+1	
	
InChI=1/C3H5NO/c1-2-5-3-4-1/h3H,1-2H2/p+1/fC3H6NO/h4H/q+1	InChI=1/C3H5NO/c1-2-5-3-4-1/h3H,1-2H2/p+1/fC3H6NO/h5H/q+1

Figure 5.8.2: The above two structures are tautomeric drawings of the same molecule and thus have the same standard InChI. If you were interested in just one of the structures you could use a nonstandard InChI with a fixed hydrogen layer (in red). Borrowed from section 6.2 of InChI Trust FAQ.

Drawbacks of InChI

InChIs are not meant to be human readable but to contain molecular information that computers can read within the layers, so unlike SMILES you can't really read even a simple InChI (see Figure 3), never mind a complex one (figure 4).

SMILES	InChI
CC(=O)C	InChI=1S/C3H6O/c1-3(2)4/h1-2H3

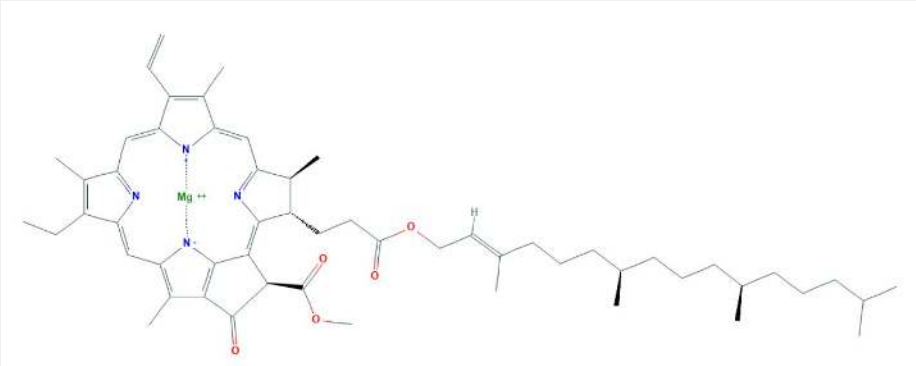


Acetone

Figure **5.8.3***: Canonical SMILES and InChI for Acetone (source: PubChem)*

Another drawback of InChI is just like an IUPAC systematic name, they are of variable length and become real long (figure 4). The problem with the variable length is it makes InChI impractical as a database registry number, and the length is often too long for internet search engines to handle.

IUPAC Systematic Name	InChI
magnesium;methyl (3*R*,21*S*,22*S*)-16-ethenyl-11-ethyl-12,17,21,26-tetramethyl-4-oxo-22-[3-oxo-3-[(*E*,7*R*,11*R*)-3,7,11,15-tetramethylhexadec-2-enoxy]propyl]-23,25-diaza-7,24-diazanidahexacyclo[18.2.1.1^{5,8}.1^{10,13}.1^{15,18}.0^{2,6}]hexacosa-1,5,8(26),9,11,13(25),14,16,18,20(23)-decaene-3-carboxylate	InChI=1S/C55H73N4O5.Mg/c1-13-39-35(8)42-28-44-37(10)41(24-25-48(60)64-27-26-34(7)23-17-22-33(6)21-16-20-32(5)19-15-18-31(3)4)52(58-44)50-51(55(62)63-12)54(61)49-38(11)45(59-53(49)50)30-47-40(14-2)36(9)43(57-47)29-46(39)56-42;/h13,26,28-33,37,41,51H,1,14-25,27H2,2-12H3,(H-,56,57,58,59,61);/q-1;+2/p-1/b34-26+;/t32-,33-,37+,41+,51-;/m1./s1



Chlorophyll A

Figure **5.8.4***:* On the top left is the IUPAC systematic name for chlorophyll A and on the right is it's InChI (source: PubChem).

There is an additional issue with the InChI in that some of the characters interfere with web search queries and thus the InChI itself is not appropriate for web searches. To solve these problems a hashed InChI Key has been developed which is of constant length and enables web searches. The hashed key is also of constant length, making it better suited for databases.

InChI Keys

The InChI suite will generate a hashed version of the InChI, the InChI Key. The hash function generates a standard key of 27 characters that stores information in four parts (see figure 5). The InChIKey may be a standard or nonstandard key as indicated by the version, but all keys are of the same length and format.

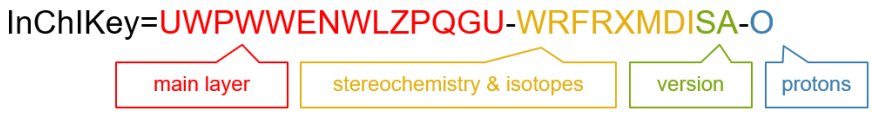


Figure 5.8.5: InChI key for 2-(^{35}Cl)chloro-R-glycine (molecule in figure 1).

The hash function is a one-way conversion (figure 6), that is, if you have an InChI you can generate the key, but if you have the key you can not generate the InChI. The key can function as an identifier is you made it registry number where you would need a look up table to know the molecule it is associated with.



Figure 5.8.6: The one way InChIKey generation function.

If two different chemical compound databases have the same chemical (InChI)they will generate the same standard InChIKeys and thus it is customary for databases and other information sources like Wikipedia chemboxes to generate standard InChIKeys, and they effectively function as a standard "registry number", that is, if two chemicals in different databases have the same standard InChIKey, they are the same chemical. On the other hand if you had a non-standard InChI the non-standard layers would induce variability of the key and so you could not compare across databases.

InChIKeys and Web Searching

The molecule (R)-2-(^{35}Cl)chloroglycine probably does not exist and was created to demonstrate the layers of an InChI and the correlating key. If you do a web search of the entire key (UWPWWENWLZPQGU-WRFRXMDISA-0) you do not get any hits, but if you search just the main layer you get several hits. What you are doing is essentially looking for any molecule with the same simplified connection table, that is, all stereoisomers, or isotopic labels. One of the hits is for [(*S*)-carboxy(chloro)methyl]azanium which is the other isomer. If you go to properties they are all computed and none were deposited to PubChem by vendors or contributors, and so this molecule has probably never been synthesized. Under 5.2 Related Compounds/Exact Same Parent you also get the non-protonated form (2-chloro-L-glycine) , of which there is published information. It is also of interest that the search of first part of the InChIKey also turned up an article in Russian on the L isomer, (you may need to download the pdf to see the actual bonds).

InChI OER

The InChI Trust runs an Open Education Resource (OER) where you can find material on InChI https://www.inchi-trust.org/oer/. The InChI OER is a repository where anyone can upload and tag material on InChI, or link to and tag existing material on the use of InChI. Once material is posted within the OER it can be searched through a filter system.

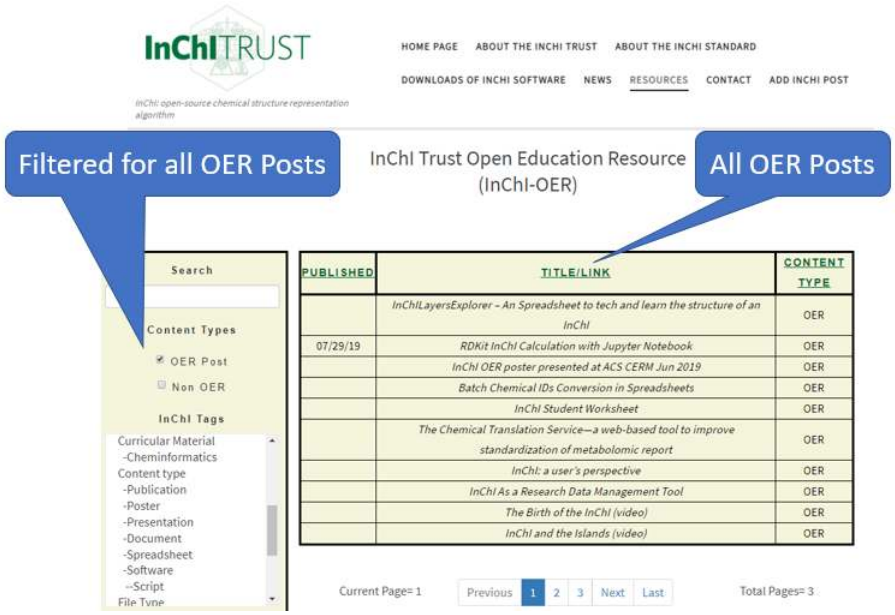


Figure 5.8.7: InChI OER tag filter and associated content. The default setting is to show all OER site material, clicking non-OER will extend the filter to include off site material like publications which have records that have been submitted to the OER.

InChI Layers Explorer

In this activity we will use the InChI OER to obtain an Excel spreadsheet that breaks an InChI into layers, and start to analyze how cheminformatics functionality can be integrated into common tools like spreadsheets. Go to the InChI OER and in the filter click "Spreadsheet" (middle of figure 5.8.8). This filters the content to items that are tagged "spreadsheet" and also removes any tag that is not associated with one of those content items. Now move down to tag category "File Type" and while holding the <ctrl> key, click Excel (right figure 5.8.8). You now get a list of excel spreadsheets (figure 5.8.9).

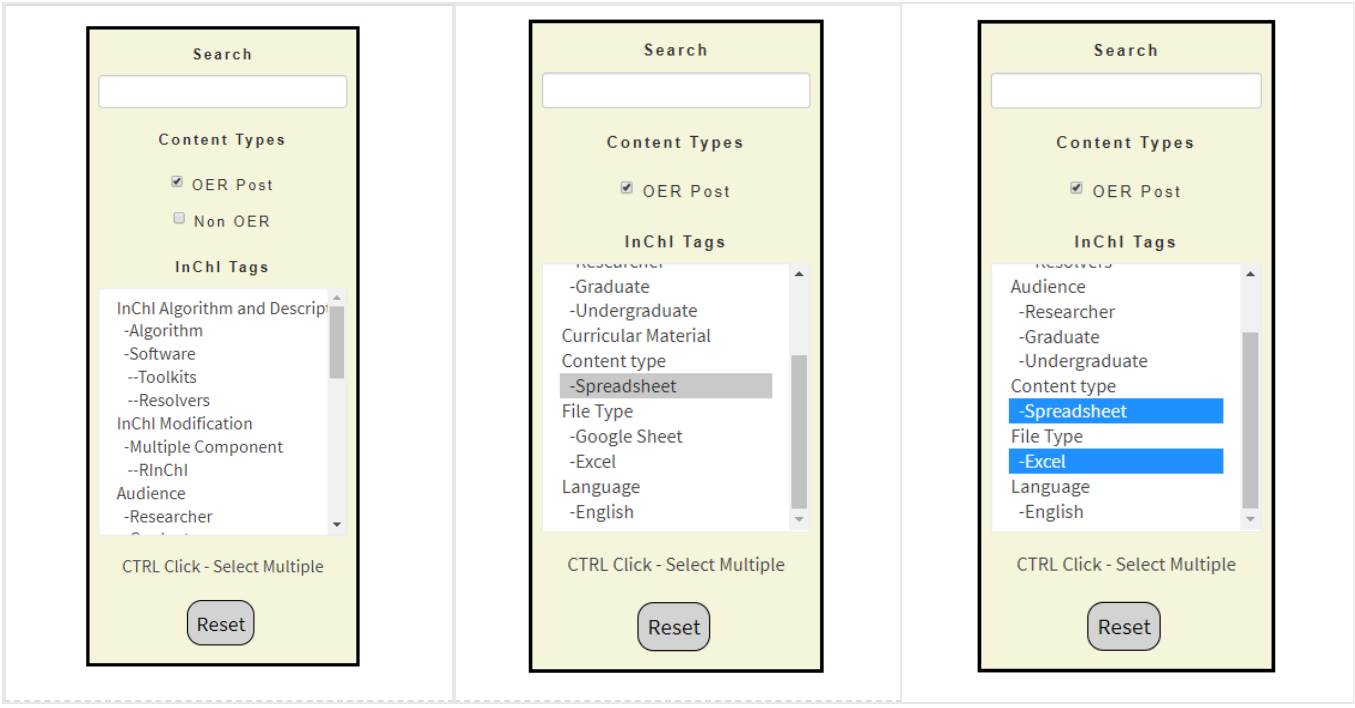


Figure 5.8.8: InChI OER Tag Filter.

On the left is the default setting and all content loaded to the site is displayed in the window (right side of figure 5.8.7. In the middle the filter for spreadsheets is activated, and you can see there are two types that have been uploaded, Google Sheets and Excel sheets. On the right both Spreadsheet and Excel have been activated, and so only spreadsheets in Excel are displayed and the content view is reduced to those items that are tagged both "Spreadsheet" and "Excel" (Figure 5.8.9)

PUBLISHED	TITLE/LINK	CONTENT TYPE
	InChILayersExplorer – An Spreadsheet to tech and learn the structure of an InChI	OER
	Batch Chemical IDs Conversion in Spreadsheets	OER
	Identifier conversion on an Excel spreadsheet	OER

Figure 5.8.9: At the time this page was created there were three items uploaded to the InChI OER that were tagged as Excel Spreadsheets.

Click on the InChILayersExplorer and you go to it's content page. This page will have a description of the content and a green information box (Figure 5.8.10), and in the information box is a "Download Publication Files", that allows you to obtain the spreadsheet.

INFORMATION	
Content Type	OER
Uploaded By	Jordi Cuadros
Download Publication Files	http://www.inchi-trust.org/wp/wp-content/uploads/2019/06/InChILayersExplorer.xlsx
License	CC BY 3.0 Unported
Content Status	publish
Number of Comments	No Comments
Date Published	
Content Tags	Audience, Content type, Excel, File Type, Graduate, InChI Algorithm and Description, Researcher, Spreadsheet, Undergraduate

Figure 5.8.10*: Green Information box for the InChILayersExplorer*

Now click on the link in the "Download Publications File" field and you will have a copy of the InChI Layers Explorer, which you should open and enable editing.

Activity 5.8.1

Using the InChILayersExplorer show the difference between the InChI for (R)-thalidomide and (S)-thalidomide. Note, the goal of this activity is not to answer the question, but to gain an understanding on how the InChILayersExplorer works, which is in effect a "smart spreadsheet" that communicates with database web APIs via webservices functions. One of the skills we hope you can gain from this class is enough familiarity with how code works so if you see new code, you can hack in and figure how it works. Be sure to enable the spreadsheet after you download it. This spreadsheet communicates with the NCI Chemical Resolver (section 2.7.)

1. Type (R)-thalidomide in the yellow region (type over CoA), OK, it fails, now try the (S) isomer, and it still fails, so now try thalidomide without specifying an isomer. OK, so you have the InChI for thalidomide, but there is nothing in the stereochemical layer, as you have not specified the stereochemistry. These spreadsheet uses the Chemical Identifier Resolver of the NIH which will be covered in section 2.6.2.1.1), which can be accessed directly at https://cactus.nci.nih.gov/chemical/structure and is shwon in figure 5.8.11. Now lets start by searching for (R)-thalidomide directly in the resolver (figure 5.8.11).

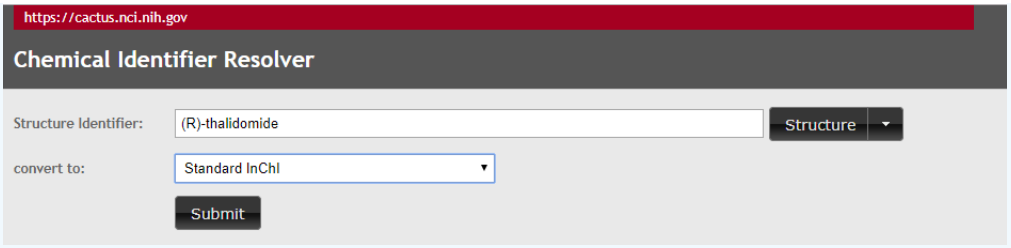


Figure 5.8.11: NCI/CADD Chemical Resolver set up to find standard InChI for (R)-thalidomid

As you may have guessed, neither (R) or (S) works, but "thalidomide" does (incidentally, you have to hit submit, not Structure), and so this resolver will not provide information on the isomers of thalidomide. So now do a web search of (R)-thalidomide, and paste in its key (UEJJHQNACJXSKW-SECBINFHSA-N), and note the stereochemical layer [/t9-/m1/s1] is the only part that is different. Now repeating for (S)-thalidomide.

You should get the following results

Table 5.8.3

Compound	InChI Key	Stereochemical layer
thalidomide	UEJJHQNACJXSKW-UHFFFAOYSA-N	none
(R)-thalidomide	UEJJHQNACJXSKW-SECBINFHSA-N	/t9-/m1/s1
(S)-thalidomide	UEJJHQNACJXSKW-VIFPVBQESA-N	/t9-/m0/s1

Note, if you click on the merged cells that generates the InChI (Rows 7-8) you see the following code.

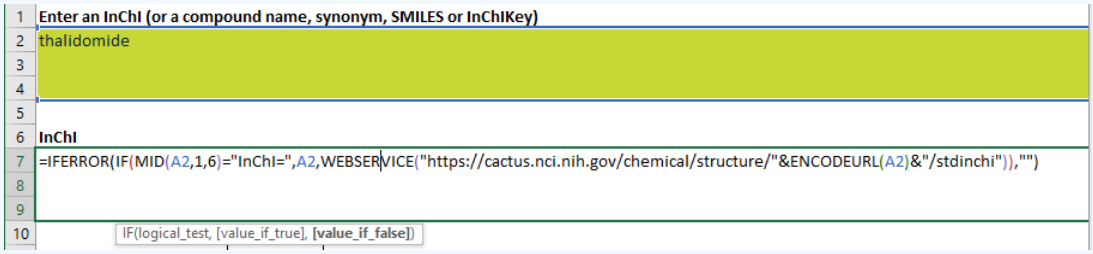
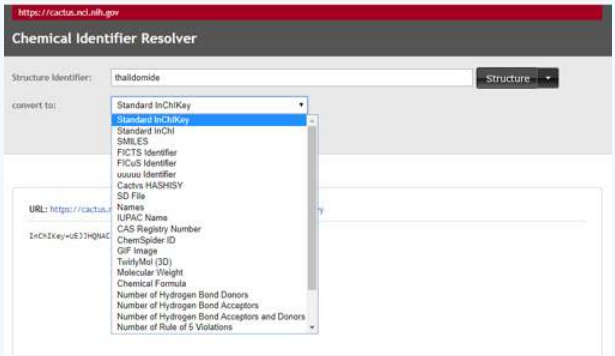


Figure 5.8.11: Code in spreadsheet that uses WEBSERVICE function to get InChI from NCI/CADD chemical resolver

Now open up a browser tab and paste in the following URL:

https://cactus.nci.nih.gov/chemical/structure/thalidomide/stdinchi

Now go back to the NCI Chemical Resolver and click the dropdown box of the "convert to" field (figure5.8.12 and try another option, say "TwirlyMol(3D).



5.8.12: Dropdown menu of NCI Chemical resolver showing some of the options.

Can you figure out the URL that uses the NCI Chemical Resolver to give the 3D molecule in a webpage? Once you have done this, can you identify a problem that has resulted from these molecular representations. Hint, think of adding two more columns to table **5.8.3**, one for 2D and one for 3D images. What is the issue when you draw the 3D image that does not arise when you draw the 2D?

References and Further Reading

1. (1) Wiswesser, W. J. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 88.
2. (2) Ash, S.; Cline, M. A.; Homer, R. W.; Hurst, T.; Smith, G. B. *J. Chem. Inf. Comput. Sci.***1997**, *37*, 71.
3. (3) Homer, R. W.; Swanson, J.; Jilek, R. J.; Hurst, T.; Clark, R. D. *J. Chem Inf. Model.***2008**, *48*, 2294.
4. (4) Barnard, J. M.; Jochum, C. J.; Welford, S. M. *Acs Symposium Series* **1989**, *400*, 76.
5. (5) Rohbeck, H. G. In *Software Development in Chemistry 5*; Gmehling, J., Ed.; Springer Berlin Heidelberg: 1991, p 49.
6. (6) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
7. (7) Weininger, D.; Weininger, A.; Weininger, J. L. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97.
8. (8) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 237.
9. (9) SMILES: Simplified Molecular Input Line Entry System (http://www.daylight.com/smiles/) (Accessed on 6/30/2015).
10. (10) Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. *J. Cheminform.* **2013**, *5*, 7.
11. (11) Heller, S.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. *J. Cheminform.* **2015**, *7*, 23.
12. (12) The IUPAC International Chemical Identifier (InChI) (http://www.iupac.org/home/publications/e-resources/inchi.html) (Accessed on 6/29/2015).
13. (13) InChI Trust (http://www.inchi-trust.org/) (Accessed on 6/29/2015).
14. (14) Daylight Theory Manual, Chapter 3: SMILES - A Simplified Chemical Language (http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html) (Accessed on 6/23/2015).
15. (15) Daylight SMILES Tutorial (http://www.daylight.com/dayhtml_tutorials/languages/smiles/index.html) (Accessed on 6/23/2015).

Contributors

Robert E. Belford (University of Arkansas Little Rock; Department of Chemistry). The breadth, depth and veracity of this work is the responsibility of Robert E. Belford, rebelford@ualr.edu. You should contact him if you have any concerns. This material has both original contributions, and content built upon prior contributions of the LibreTexts Community and other resources, including but not limited to:

- Sunghwan Kim
- Material Adapted from 2017 Cheminformatics OLCC

5.8: Line Notation (SMILES and InChI) is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

6: Lipids

The lipids are a large and diverse group of naturally occurring organic compounds that are related by their solubility in nonpolar organic solvents (e.g., ether, chloroform, acetone and benzene) and general insolubility in water. There is great structural variety among the lipids, as will be demonstrated in the following sections.

6.1: Prelude to Lipids
6.2: Fatty Acids
6.3: Fats and Oils
6.4: Membranes and Membrane Lipids
6.5: Steroids
6.6: Exercises
6.7: Lipids (Summary)

Template:HideTOC

6: Lipids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.1: Prelude to Lipids

On July 11, 2003, the Food and Drug Administration amended its food labeling regulations to require that manufacturers list the amount of *trans* fatty acids on Nutrition Facts labels of foods and dietary supplements, effective January 1, 2006. This amendment was a response to published studies demonstrating a link between the consumption of *trans* fatty acids and an increased risk of heart disease. *Trans* fatty acids are produced in the conversion of liquid oils to solid fats, as in the creation of many commercial margarines and shortenings. They have been shown to increase the levels of low-density lipoproteins (LDLs)—complexes that are often referred to as bad cholesterol—in the blood. In this chapter, you will learn about fatty acids and what is meant by a *trans* fatty acid, as well as the difference between fats and oils. You will also learn what cholesterol is and why it is an important molecule in the human body.

Fats and oils, found in many of the foods we eat, belong to a class of biomolecules known as lipids. Gram for gram, they pack more than twice the caloric content of carbohydrates: the oxidation of fats and oils supplies about 9 kcal of energy for every gram oxidized, whereas the oxidation of carbohydrates supplies only 4 kcal/g. Although the high caloric content of fats may be bad news for the dieter, it says something about the efficiency of nature's designs. Our bodies use carbohydrates, primarily in the form of glucose, for our *immediate* energy needs. Our capacity for storing carbohydrates for later use is limited to tucking away a bit of glycogen in the liver or in muscle tissue. We store our *reserve* energy in lipid form, which requires far less space than the same amount of energy stored in carbohydrate form. Lipids have other biological functions besides energy storage. They are a major component of the membranes of the 10 trillion cells in our bodies. They serve as protective padding and insulation for vital organs. Furthermore, without lipids in our diets, we would be deficient in the fat-soluble vitamins A, D, E, and K.

Lipids are not defined by the presence of specific functional groups, as carbohydrates are, but by a physical property—solubility. Compounds isolated from body tissues are classified as lipids if they are more soluble in organic solvents, such as dichloromethane, than in water. By this criterion, the lipid category includes not only fats and oils, which are esters of the trihydroxy alcohol glycerol and fatty acids, but also compounds that incorporate functional groups derived from phosphoric acid, carbohydrates, or amino alcohols, as well as steroid compounds such as cholesterol (Figure 6.1.1 presents one scheme for classifying the various kinds of lipids). We will discuss the various kinds of lipids by considering one subclass at a time and pointing out structural similarities and differences as we go.

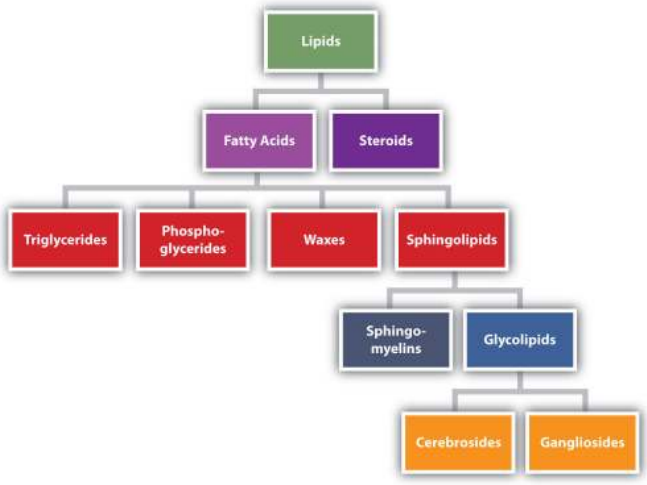


Figure 6.1.1: Lipid Organization Based on Structural Relationships

Lipid categorized into fatty acids and steroids. Fatty acids are further separated into triglycerides, phospho-glycerides, waxes, and sphingolipids. Sphingolipids are separated into sphingo-myelins and glycolipids. Glycolipids are separated into cerebrosides and gangliosides.

6.1: Prelude to Lipids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **17.0: Prelude to Lipids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

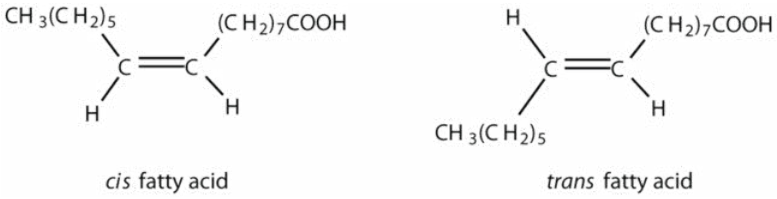
https://chem.libretexts.org/@go/page/255602

6.2: Fatty Acids

Learning Objectives

- To recognize the structures of common fatty acids and classify them as saturated, monounsaturated, or polyunsaturated.

Fatty acids are carboxylic acids that are structural components of fats, oils, and all other categories of lipids, except steroids. More than 70 have been identified in nature. They usually contain an even number of carbon atoms (typically 12–20), are generally unbranched, and can be classified by the presence and number of carbon-to-carbon double bonds. Thus, saturated fatty acids contain no carbon-to-carbon double bonds, monounsaturated fatty acids contain one carbon-to-carbon double bond, and polyunsaturated fatty acids contain two or more carbon-to-carbon double bonds.



Cis fatty acids have two H atoms on the same side of the plane of the fatty acid while trans has the two H atoms on opposite sides.

Table **6.2.1** lists some common fatty acids and one important source for each. The atoms or groups around the double bonds in unsaturated fatty acids can be arranged in either the cis or trans isomeric form. Naturally occurring fatty acids are generally in the cis configuration.

Table **6.2.1**: Some Common Fatty Acids Found in Natural Fats

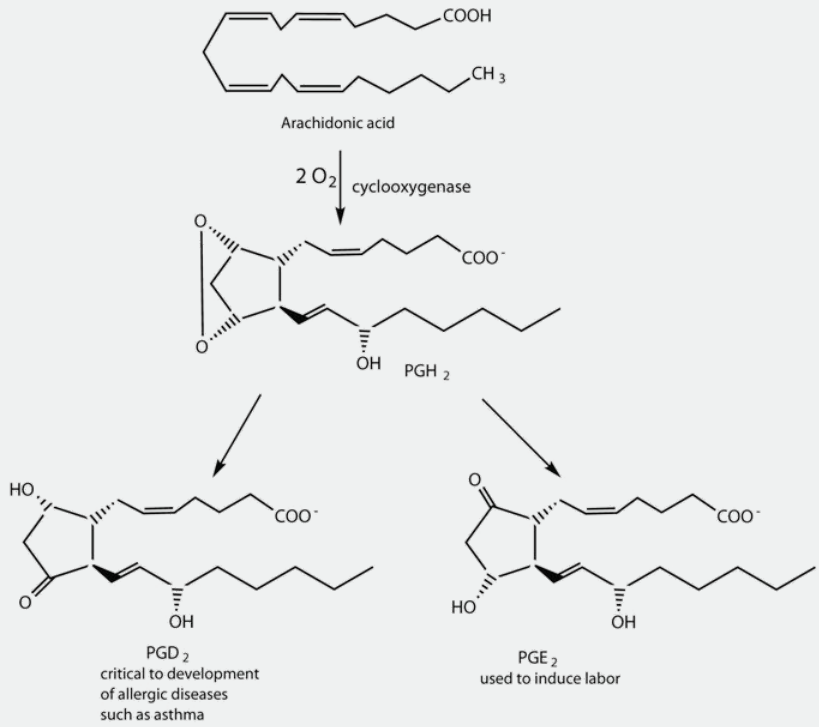
Name	Abbreviated Structural Formula	Condensed Structural Formula	Melting Point (°C)	Source
lauric acid	$C_{11}H_{23}COOH$	$CH_3(CH_2)_{10}COOH$	44	palm kernel oil
myristic acid	$C_{13}H_{27}COOH$	$CH_3(CH_2)_{12}COOH$	58	oil of nutmeg
palmitic acid	$C_{15}H_{31}COOH$	$CH_3(CH_2)_{14}COOH$	63	palm oil
palmitoleic acid	$C_{15}H_{29}COOH$	$CH_3(CH_2)_5CH=CH(CH_2)_7COOH$	0.5	macadamia oil
stearic acid	$C_{17}H_{35}COOH$	$CH_3(CH_2)_{16}COOH$	70	cocoa butter
oleic acid	$C_{17}H_{33}COOH$	$CH_3(CH_2)_7CH=CH(CH_2)_7COOH$	16	olive oil
linoleic acid	$C_{17}H_{31}COOH$	$CH_3(CH_2)_3(CH_2CH=CH)_2(CH_2)_7COOH$	−5	canola oil
α-linolenic acid	$C_{17}H_{29}COOH$	$CH_3(CH_2CH=CH)_3(CH_2)_7COOH$	−11	flaxseed
arachidonic acid	$C_{19}H_{31}COOH$	$CH_3(CH_2)_4(CH_2CH=CH)_4(CH_2)_2COOH$	−50	liver

Two polyunsaturated fatty acids—linoleic and α-linolenic acids—are termed essential fatty acids because humans must obtain them from their diets. Both substances are required for normal growth and development, but the human body does not synthesize them. The body uses linoleic acid to synthesize many of the other unsaturated fatty acids, such as arachidonic acid, a precursor for the synthesis of prostaglandins. In addition, the essential fatty acids are necessary for the efficient transport and metabolism of cholesterol. The average daily diet should contain about 4–6 g of the essential fatty acids.

To Your Health: Prostaglandins

Prostaglandins are chemical messengers synthesized in the cells in which their physiological activity is expressed. They are unsaturated fatty acids containing 20 carbon atoms and are synthesized from arachidonic acid—a polyunsaturated fatty acid—when needed by a particular cell. They are called *prostaglandins* because they were originally isolated from semen found in the prostate gland. It is now known that they are synthesized in nearly all mammalian tissues and affect almost all organs in the body. The five major classes of prostaglandins are designated as PGA, PGB, PGE, PGF, and PGI. Subscripts are attached at the end of these abbreviations to denote the number of double bonds outside the five-carbon ring in a given prostaglandin.

The prostaglandins are among the most potent biological substances known. Slight structural differences give them highly distinct biological effects; however, all prostaglandins exhibit some ability to induce smooth muscle contraction, lower blood pressure, and contribute to the inflammatory response. Aspirin and other nonsteroidal anti-inflammatory agents, such as ibuprofen, obstruct the synthesis of prostaglandins by inhibiting cyclooxygenase, the enzyme needed for the initial step in the conversion of arachidonic acid to prostaglandins.



Their wide range of physiological activity has led to the synthesis of hundreds of prostaglandins and their analogs. Derivatives of PGE_2 are now used in the United States to induce labor. Other prostaglandins have been employed clinically to lower or increase blood pressure, inhibit stomach secretions, relieve nasal congestion, relieve asthma, and prevent the formation of blood clots, which are associated with heart attacks and strokes.

Although we often draw the carbon atoms in a straight line, they actually have more of a zigzag configuration (Figure **6.2.2*a***). Viewed as a whole, however, the saturated fatty acid molecule is relatively straight (Figure **6.2.2*b***). Such molecules pack closely together into a crystal lattice, maximizing the strength of dispersion forces and causing fatty acids and the fats derived from them to have relatively high melting points. In contrast, each *cis* carbon-to-carbon double bond in an unsaturated fatty acid produces a pronounced bend in the molecule, so that these molecules do not stack neatly. As a result, the intermolecular attractions of unsaturated fatty acids (and unsaturated fats) are weaker, causing these substances to have lower melting points. Most are liquids at room temperature.

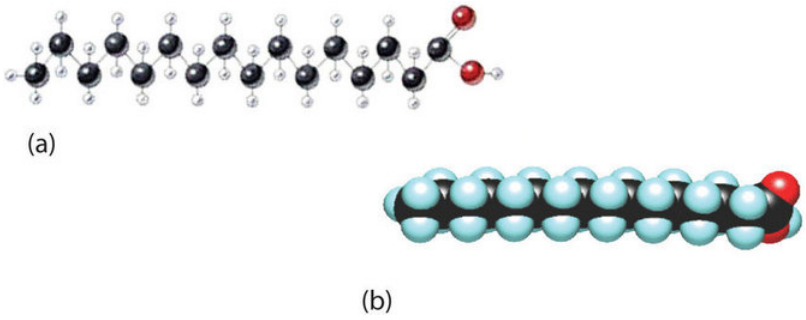
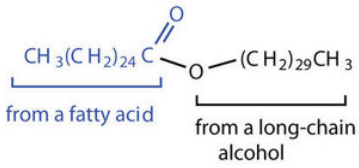


Figure **6.2.2**: The Structure of Saturated Fatty Acids. (a) There is a zigzag pattern formed by the carbon-to-carbon single bonds in the ball-and-stick model of a palmitic acid molecule. (b) A space-filling model of palmitic acid shows the overall straightness of a saturated fatty acid molecule.

Waxes are esters formed from long-chain fatty acids and long-chain alcohols. Most natural waxes are mixtures of such esters. Plant waxes on the surfaces of leaves, stems, flowers, and fruits protect the plant from dehydration and invasion by harmful microorganisms. Carnauba wax, used extensively in floor waxes, automobile waxes, and furniture polish, is largely myricyl cerotate, obtained from the leaves of certain Brazilian palm trees. Animals also produce waxes that serve as protective coatings, keeping the surfaces of feathers, skin, and hair pliable and water repellent. In fact, if the waxy coating on the feathers of a water bird is dissolved as a result of the bird swimming in an oil slick, the feathers become wet and heavy, and the bird, unable to maintain its buoyancy, drowns.



Myricyl cerotate
(found in carnauba wax)

Summary

Fatty acids are carboxylic acids that are the structural components of many lipids. They may be saturated or unsaturated. Most fatty acids are unbranched and contain an even number of carbon atoms. Unsaturated fatty acids have lower melting points than saturated fatty acids containing the same number of carbon atoms.

6.2: Fatty Acids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **17.1: Fatty Acids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

6.3: Fats and Oils

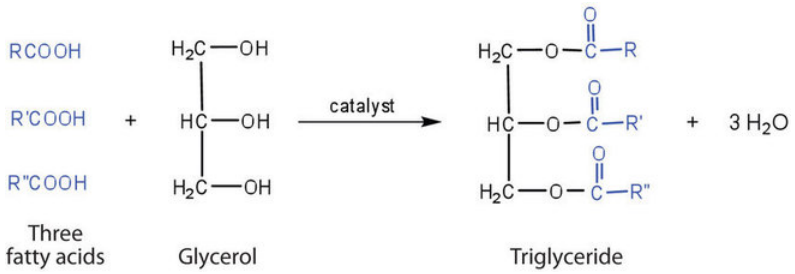
Learning Objectives

- Explain why fats and oils are referred to as triglycerides.
- Explain how the fatty acid composition of the triglycerides determines whether a substance is a fat or oil.
- Describe the importance of key reactions of triglycerides, such as hydrolysis, hydrogenation, and oxidation.

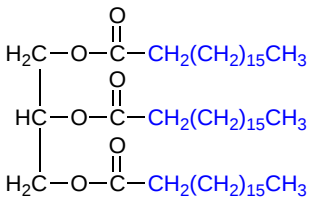
Fats and oils are the most abundant lipids in nature. They provide energy for living organisms, insulate body organs, and transport fat-soluble vitamins through the blood.

Structures of Fats and Oils

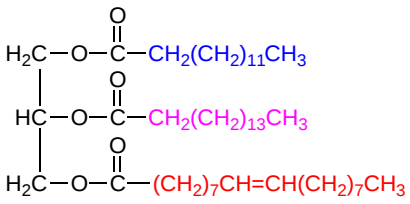
Fats and oils are called triglycerides (or *triacylcylgerols*) because they are esters composed of three fatty acid units joined to *glycerol*, a trihydroxy alcohol:



If all three OH groups on the glycerol molecule are esterified with the same fatty acid, the resulting ester is called a *simple triglyceride*. Although simple triglycerides have been synthesized in the laboratory, they rarely occur in nature. Instead, a typical triglyceride obtained from naturally occurring fats and oils contains two or three different fatty acid components and is thus termed a *mixed triglyceride*.



Tristearin
a simple triglyceride



a mixed triglyceride

A triglyceride is called a fat if it is a solid at 25°C; it is called an oil if it is a liquid at that temperature. These differences in melting points reflect differences in the degree of unsaturation and number of carbon atoms in the constituent fatty acids. Triglycerides obtained from animal sources are usually solids, while those of plant origin are generally oils. Therefore, we commonly speak of animal fats and vegetable oils.

No single formula can be written to represent the naturally occurring fats and oils because they are highly complex mixtures of triglycerides in which many different fatty acids are represented. Table **6.3.1** shows the fatty acid compositions of some common

fats and oils. The composition of any given fat or oil can vary depending on the plant or animal species it comes from as well as on dietetic and climatic factors. To cite just one example, lard from corn-fed hogs is more highly saturated than lard from peanut-fed hogs. Palmitic acid is the most abundant of the saturated fatty acids, while oleic acid is the most abundant unsaturated fatty acid.

Table **6.3.1**: Average Fatty Acid Composition of Some Common Fats and Oils (%)*

	Lauric	Myristic	Palmitic	Stearic	Oleic	Linoleic	Linolenic
Fats							
butter (cow)	3	11	27	12	29	2	1
tallow		3	24	19	43	3	1
lard		2	26	14	44	10	
Oils							
canola oil			4	2	62	22	10
coconut oil†	47	18	9	3	6	2	
corn oil			11	2	28	58	1
olive oil			13	3	71	10	1
peanut oil			11	2	48	32	
soybean oil			11	4	24	54	7

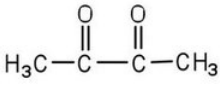
*Totals less than 100% indicate the presence of fatty acids with fewer than 12 carbon atoms or more than 18 carbon atoms.

†Coconut oil is highly saturated. It contains an unusually high percentage of the low-melting C_8, C_{10}, and C_{12} saturated fatty acids.

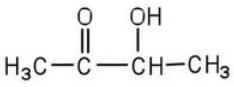
Terms such as *saturated fat* or *unsaturated oil* are often used to describe the fats or oils obtained from foods. Saturated fats contain a high proportion of saturated fatty acids, while unsaturated oils contain a high proportion of unsaturated fatty acids. The high consumption of saturated fats is a factor, along with the high consumption of cholesterol, in increased risks of heart disease.

Physical Properties of Fats and Oils

Contrary to what you might expect, *pure* fats and oils are colorless, odorless, and tasteless. The characteristic colors, odors, and flavors that we associate with some of them are imparted by foreign substances that are lipid soluble and have been absorbed by these lipids. For example, the yellow color of butter is due to the presence of the pigment carotene; the taste of butter comes from two compounds—diacetyl and 3-hydroxy-2-butanone—produced by bacteria in the ripening cream from which the butter is made.



Diacetyl

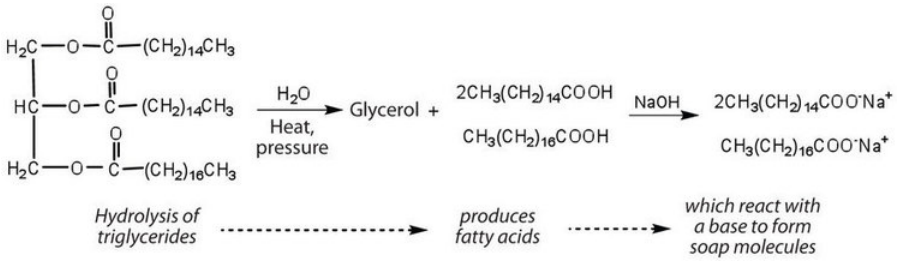


3-hydroxy-2-butanone

Fats and oils are lighter than water, having densities of about 0.8 g/cm^3. They are poor conductors of heat and electricity and therefore serve as excellent insulators for the body, slowing the loss of heat through the skin.

Chemical Reactions of Fats and Oils

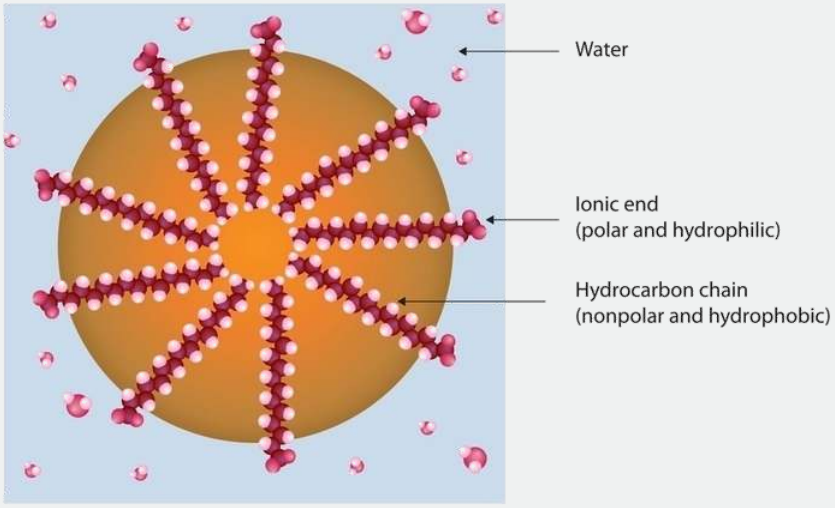
Fats and oils can participate in a variety of chemical reactions—for example, because triglycerides are esters, they can be hydrolyzed in the presence of an acid, a base, or specific enzymes known as lipases. The hydrolysis of fats and oils in the presence of a base is used to make soap and is called saponification. Today most soaps are prepared through the hydrolysis of triglycerides (often from tallow, coconut oil, or both) using water under high pressure and temperature [700 lb/in^2 (~50 atm or 5,000 kPa) and 200°C]. Sodium carbonate or sodium hydroxide is then used to convert the fatty acids to their sodium salts (soap molecules):



Looking Closer: Soaps

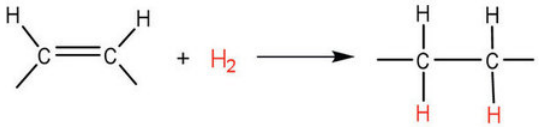
Ordinary soap is a mixture of the sodium salts of various fatty acids, produced in one of the oldest organic syntheses practiced by humans (second only to the fermentation of sugars to produce ethyl alcohol). Both the Phoenicians (600 BCE) and the Romans made soap from animal fat and wood ash. Even so, the widespread production of soap did not begin until the 1700s. Soap was traditionally made by treating molten lard or tallow with a slight excess of alkali in large open vats. The mixture was heated, and steam was bubbled through it. After saponification was completed, the soap was precipitated from the mixture by the addition of sodium chloride (NaCl), removed by filtration, and washed several times with water. It was then dissolved in water and reprecipitated by the addition of more NaCl. The glycerol produced in the reaction was also recovered from the aqueous wash solutions.

Pumice or sand is added to produce scouring soap, while ingredients such as perfumes or dyes are added to produce fragrant, colored soaps. Blowing air through molten soap produces a floating soap. Soft soaps, made with potassium salts, are more expensive but produce a finer lather and are more soluble. They are used in liquid soaps, shampoos, and shaving creams.



Dirt and grime usually adhere to skin, clothing, and other surfaces by combining with body oils, cooking fats, lubricating greases, and similar substances that act like glues. Because these substances are not miscible in water, washing with water alone does little to remove them. Soap removes them, however, because soap molecules have a dual nature. One end, called the *head*, carries an ionic charge (a carboxylate anion) and therefore dissolves in water; the other end, the *tail*, has a hydrocarbon structure and dissolves in oils. The hydrocarbon tails dissolve in the soil; the ionic heads remain in the aqueous phase, and the soap breaks the oil into tiny soap-enclosed droplets called *micelles*, which disperse throughout the solution. The droplets repel each other because of their charged surfaces and do not coalesce. With the oil no longer “gluing” the dirt to the soiled surface (skin, cloth, dish), the soap-enclosed dirt can easily be rinsed away.

The double bonds in fats and oils can undergo hydrogenation and also oxidation. The hydrogenation of vegetable oils to produce semisolid fats is an important process in the food industry. Chemically, it is essentially identical to the catalytic hydrogenation reaction described for alkenes.

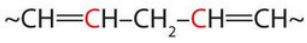


In commercial processes, the number of double bonds that are hydrogenated is carefully controlled to produce fats with the desired consistency (soft and pliable). Inexpensive and abundant vegetable oils (canola, corn, soybean) are thus transformed into margarine and cooking fats. In the preparation of margarine, for example, partially hydrogenated oils are mixed with water, salt, and nonfat dry milk, along with flavoring agents, coloring agents, and vitamins A and D, which are added to approximate the look, taste, and nutrition of butter. (Preservatives and antioxidants are also added.) In most commercial peanut butter, the peanut oil has been partially hydrogenated to prevent it from separating out. Consumers could decrease the amount of saturated fat in their diet by using the original unprocessed oils on their foods, but most people would rather spread margarine on their toast than pour oil on it.

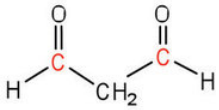
Many people have switched from butter to margarine or vegetable shortening because of concerns that saturated animal fats can raise blood cholesterol levels and result in clogged arteries. However, during the hydrogenation of vegetable oils, an isomerization reaction occurs that produces the *trans* fatty acids mentioned in the opening essay. However, studies have shown that *trans* fatty acids also raise cholesterol levels and increase the incidence of heart disease. *Trans* fatty acids do not have the bend in their structures, which occurs in *cis* fatty acids and thus pack closely together in the same way that the saturated fatty acids do. Consumers are now being advised to use polyunsaturated oils and soft or liquid margarine and reduce their total fat consumption to less than 30% of their total calorie intake each day.

Fats and oils that are in contact with moist air at room temperature eventually undergo oxidation and hydrolysis reactions that cause them to turn rancid, acquiring a characteristic disagreeable odor. One cause of the odor is the release of volatile fatty acids by hydrolysis of the ester bonds. Butter, for example, releases foul-smelling butyric, caprylic, and capric acids. Microorganisms present in the air furnish lipases that catalyze this process. Hydrolytic rancidity can easily be prevented by covering the fat or oil and keeping it in a refrigerator.

Another cause of volatile, odorous compounds is the oxidation of the unsaturated fatty acid components, particularly the readily oxidized structural unit



in polyunsaturated fatty acids, such as linoleic and linolenic acids. One particularly offensive product, formed by the oxidative cleavage of both double bonds in this unit, is a compound called *malonaldehyde*.



Malonaldehyde

Rancidity is a major concern of the food industry, which is why food chemists are always seeking new and better antioxidants, substances added in very small amounts (0.001%–0.01%) to prevent oxidation and thus suppress rancidity. Antioxidants are compounds whose affinity for oxygen is greater than that of the lipids in the food; thus they function by preferentially depleting the supply of oxygen absorbed into the product. Because vitamin E has antioxidant properties, it helps reduce damage to lipids in the body, particularly to unsaturated fatty acids found in cell membrane lipids.

Summary

Fats and oils are composed of molecules known as triglycerides, which are esters composed of three fatty acid units linked to glycerol. An increase in the percentage of shorter-chain fatty acids and/or unsaturated fatty acids lowers the melting point of a fat or oil. The hydrolysis of fats and oils in the presence of a base makes soap and is known as saponification. Double bonds present in unsaturated triglycerides can be hydrogenated to convert oils (liquid) into margarine (solid). The oxidation of fatty acids can form compounds with disagreeable odors. This oxidation can be minimized by the addition of antioxidants.

6.3: Fats and Oils is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **17.2: Fats and Oils** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

6.4: Membranes and Membrane Lipids

Learning Objectives

- Identify the distinguishing characteristics of membrane lipids.
- Describe membrane components and how they are arranged.

All living cells are surrounded by a cell membrane. Plant cells (Figure **6.4.1*A***) and animal cells (Figure **6.4.1*B***) contain a cell nucleus that is also surrounded by a membrane and holds the genetic information for the cell. Everything between the cell membrane and the nuclear membrane—including intracellular fluids and various subcellular components such as the mitochondria and ribosomes—is called the cytoplasm. The membranes of all cells have a fundamentally similar structure, but membrane function varies tremendously from one organism to another and even from one cell to another within a single organism. This diversity arises mainly from the presence of different proteins and lipids in the membrane.

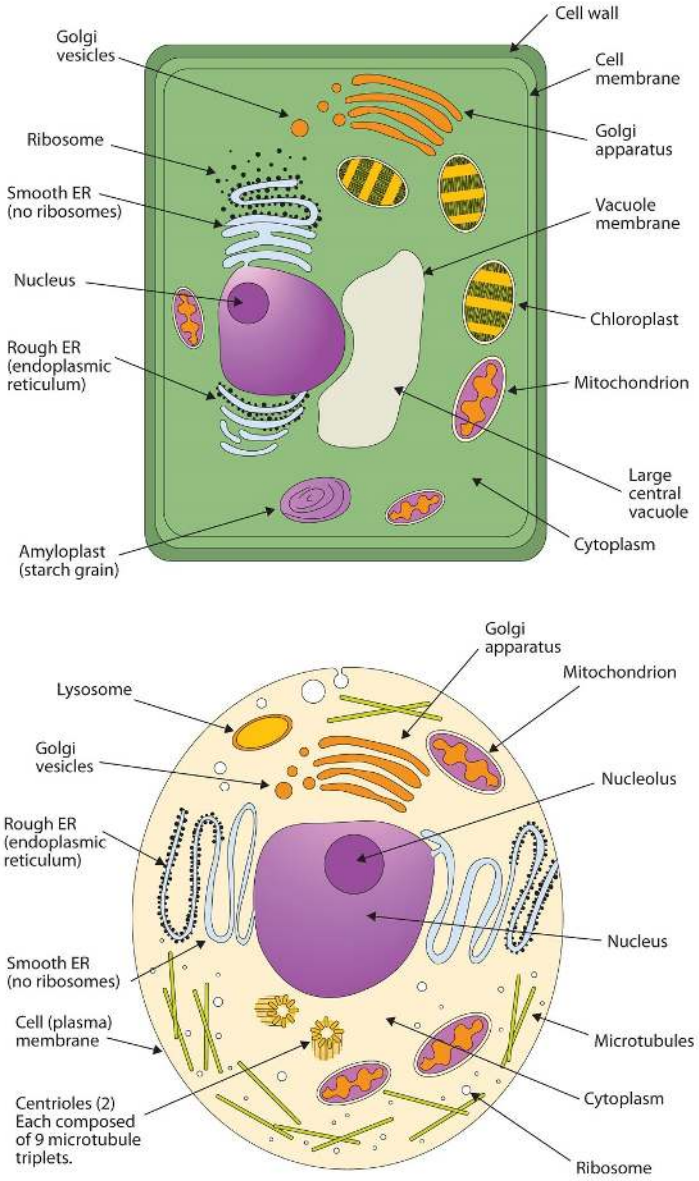


Figure **6.4.1**: (A) An Idealized Plant Cell. Not all the structures shown here occur in every type of plant cell. (B) An Idealized Animal Cell. The structures shown here will seldom all be found in a single animal cell.

The lipids in cell membranes are highly polar but have dual characteristics: part of the lipid is ionic and therefore dissolves in water, whereas the rest has a hydrocarbon structure and therefore dissolves in nonpolar substances. Often, the ionic part is referred to as hydrophilic, meaning “water loving,” and the nonpolar part as hydrophobic, meaning “water fearing” (repelled by water). When allowed to float freely in water, polar lipids spontaneously cluster together in any one of three arrangements: micelles, monolayers, and bilayers (Figure **6.4.2**).

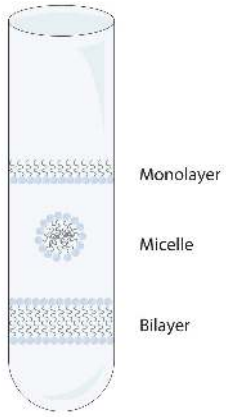


Figure **6.4.2**: Spontaneously Formed Polar Lipid Structures in Water: Monolayer, Micelle, and Bilayer

Micelles are aggregations in which the lipids’ hydrocarbon tails—being hydrophobic—are directed toward the center of the assemblage and away from the surrounding water while the hydrophilic heads are directed outward, in contact with the water. Each micelle may contain thousands of lipid molecules. Polar lipids may also form a monolayer, a layer one molecule thick on the surface of the water. The polar heads face into water, and the nonpolar tails stick up into the air. Bilayers are double layers of lipids arranged so that the hydrophobic tails are sandwiched between an inner surface and an outer surface consisting of hydrophilic heads. The hydrophilic heads are in contact with water on either side of the bilayer, whereas the tails, sequestered inside the bilayer, are prevented from having contact with the water. Bilayers like this make up every cell membrane (Figure **6.4.3**).

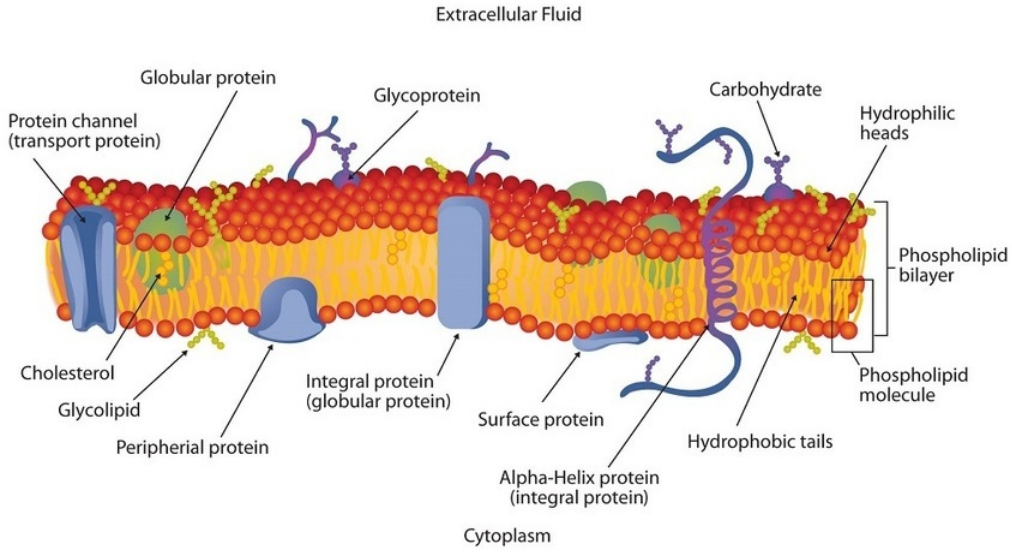


Figure **6.4.3**: Schematic Diagram of a Cell Membrane. The membrane enclosing a typical animal cell is a phospholipid bilayer with embedded cholesterol and protein molecules. Short oligosaccharide chains are attached to the outer surface.

In the bilayer interior, the hydrophobic tails (that is, the fatty acid portions of lipid molecules) interact by means of dispersion forces. The interactions are weakened by the presence of unsaturated fatty acids. As a result, the membrane components are free to mill about to some extent, and the membrane is described as fluid.

The lipids found in cell membranes can be categorized in various ways. Phospholipids are lipids containing phosphorus. Glycolipids are sugar-containing lipids. The latter are found exclusively on the outer surface of the cell membrane, acting as distinguishing surface markers for the cell and thus serving in cellular recognition and cell-to-cell communication. Sphingolipids

are phospholipids or glycolipids that contain the unsaturated amino alcohol sphingosine rather than glycerol. Diagrammatic structures of representative membrane lipids are presented in Figure **6.4.4**.

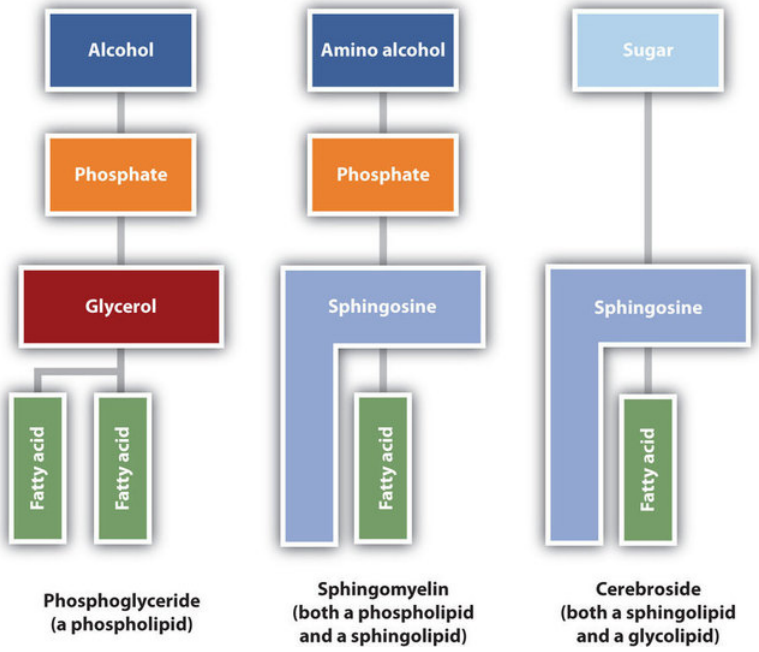


Figure **6.4.4**: Component Structures of Some Important Membrane Lipids

Phosphoglyceride are composed of alcohol, phosphate, glycerol and 2 fatty acids. A sphingomyelin contains phosphate, sphingosine, and a fatty acid. A cerbroside contains sugar, sphingosine and a fatty acid.

Phosphoglycerides (also known as glycerophospholipids) are the most abundant phospholipids in cell membranes. They consist of a glycerol unit with fatty acids attached to the first two carbon atoms, while a phosphoric acid unit, esterified with an alcohol molecule (usually an amino alcohol, as in part (a) of Figure **6.4.5**) is attached to the third carbon atom of glycerol (part (b) of Figure **6.4.5**). Notice that the phosphoglyceride molecule is identical to a triglyceride up to the phosphoric acid unit (part (b) of Figure **6.4.5**).

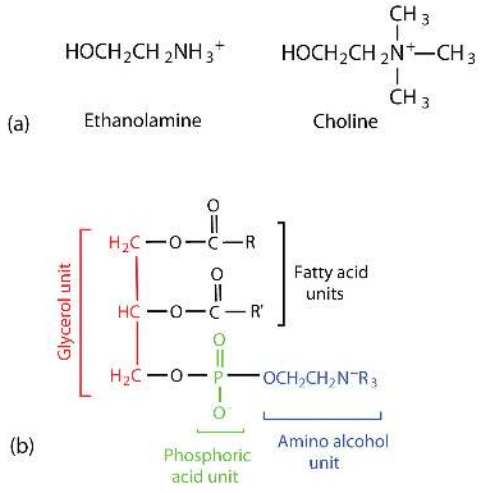
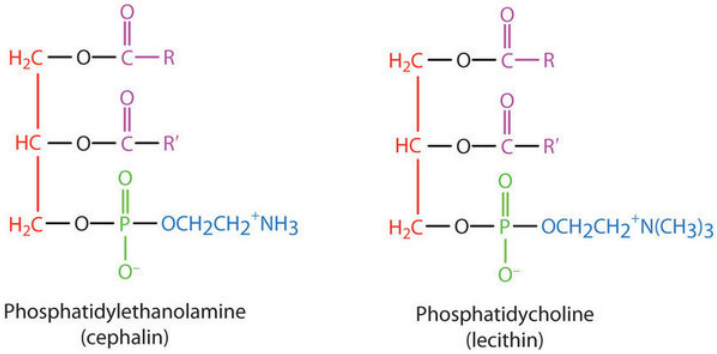


Figure **6.4.5**: Phosphoglycerides. (a) Amino alcohols are commonly found in phosphoglycerides, which are evident in its structural formula (b).

Structural formula of ethanolamine and choline are shown. The structural formula of a phosphoglyceride is shown with the glycerol unit, phosphoric unit, and amino alcohol unit highlighted in different colors.

There are two common types of phosphoglycerides. Phosphoglycerides containing ethanolamine as the amino alcohol are called *phosphatidylethanolamines* or *cephalins*. Cephalins are found in brain tissue and nerves and also have a role in blood clotting. Phosphoglycerides containing choline as the amino alcohol unit are called *phosphatidylcholines* or *lecithins*. Lecithins occur in all

living organisms. Like cephalins, they are important constituents of nerve and brain tissue. Egg yolks are especially rich in lecithins. Commercial-grade lecithins isolated from soybeans are widely used in foods as emulsifying agents. An emulsifying agent is used to stabilize an emulsion—a dispersion of two liquids that do not normally mix, such as oil and water. Many foods are emulsions. Milk is an emulsion of butterfat in water. The emulsifying agent in milk is a protein called *casein*. Mayonnaise is an emulsion of salad oil in water, stabilized by lecithins present in egg yolk.



Sphingomyelins, the simplest sphingolipids, each contain a fatty acid, a phosphoric acid, sphingosine, and choline (Figure **6.4.6**). Because they contain phosphoric acid, they are also classified as phospholipids. Sphingomyelins are important constituents of the myelin sheath surrounding the axon of a nerve cell. Multiple sclerosis is one of several diseases resulting from damage to the myelin sheath.

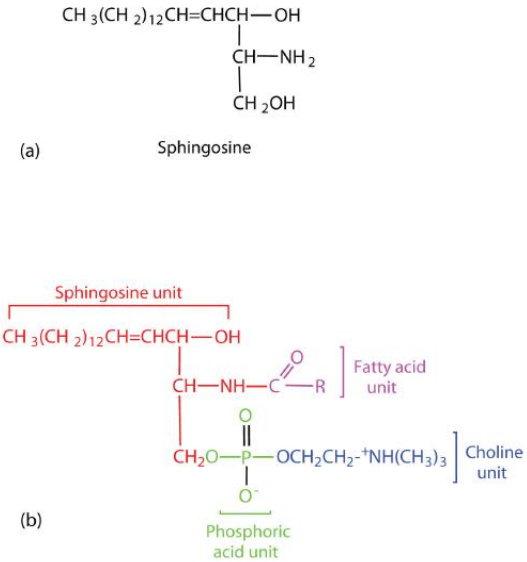


Figure **6.4.6**: Sphingolipids. (a) Sphingosine, an amino alcohol, is found in all sphingolipids. (b) A sphingomyelin is also known as a phospholipid, as evidenced by the phosphoric acid unit in its structure.

Structural formula of sphingosine is shown. The general structure of a sphingolipid is shown with the sphingosine unit, fatty acid unit, phosphoric acid unit, and choline unit highlighted in different colors.

Most animal cells contain sphingolipids called cerebrosides (Figure **6.4.7**). Cerebrosides are composed of sphingosine, a fatty acid, and galactose or glucose. They therefore resemble sphingomyelins but have a sugar unit in place of the choline phosphate group. Cerebrosides are important constituents of the membranes of nerve and brain cells.

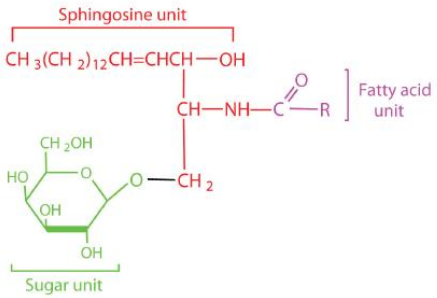


Figure 6.4.7: Cerebrosides. Cerebrosides are sphingolipids that contain a sugar unit.
General structure of a cerebroside with its sugar unit, sphingosine unit, and fatty acid unit highlighted in different colors.

The sphingolipids called gangliosides are more complex, usually containing a branched chain of three to eight monosaccharides and/or substituted sugars. Because of considerable variation in their sugar components, about 130 varieties of gangliosides have been identified. Most cell-to-cell recognition and communication processes (e.g., blood group antigens) depend on differences in the sequences of sugars in these compounds. Gangliosides are most prevalent in the outer membranes of nerve cells, although they also occur in smaller quantities in the outer membranes of most other cells. Because cerebrosides and gangliosides contain sugar groups, they are also classified as glycolipids.

Membrane Proteins

If membranes were composed only of lipids, very few ions or polar molecules could pass through their hydrophobic "sandwich filling" to enter or leave any cell. However, certain charged and polar species do cross the membrane, aided by proteins that move about in the lipid bilayer. The two major classes of proteins in the cell membrane are integral proteins, which span the hydrophobic interior of the bilayer, and peripheral proteins, which are more loosely associated with the surface of the lipid bilayer (Figure 6.4.3). Peripheral proteins may be attached to integral proteins, to the polar head groups of phospholipids, or to both by hydrogen bonding and electrostatic forces.

Small ions and molecules soluble in water enter and leave the cell by way of channels through the integral proteins. Some proteins, called *carrier proteins*, facilitate the passage of certain molecules, such as hormones and neurotransmitters, by specific interactions between the protein and the molecule being transported.

Summary

Lipids are important components of biological membranes. These lipids have dual characteristics: part of the molecule is hydrophilic, and part of the molecule is hydrophobic. Membrane lipids may be classified as phospholipids, glycolipids, and/or sphingolipids. Proteins are another important component of biological membranes. Integral proteins span the lipid bilayer, while peripheral proteins are more loosely associated with the surface of the membrane.

6.4: Membranes and Membrane Lipids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **17.3: Membranes and Membrane Lipids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

6.5: Steroids

Learning Objectives

- To identify the functions of steroids produced in mammals.

All the lipids discussed so far are *saponifiable*, reacting with aqueous alkali to yield simpler components, such as glycerol, fatty acids, amino alcohols, and sugars. Lipid samples extracted from cellular material, however, also contain a small but important fraction that does not react with alkali. The most important nonsaponifiable lipids are the steroids. These compounds include the *bile salts*, *cholesterol* and related compounds, and certain *hormones* (such as cortisone and the sex hormones).

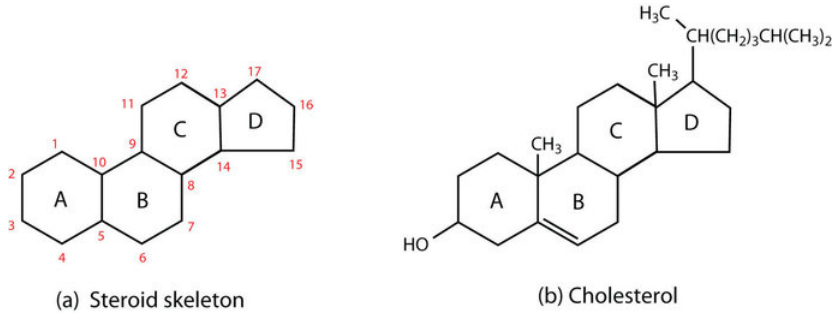


Figure **6.5.1** Steroids. (a) The four-fused-ring steroid skeleton uses letter designations for each ring and the numbering of the carbon atoms. (b) The cholesterol molecule follows this pattern.

Steroids occur in plants, animals, yeasts, and molds but not in bacteria. They may exist in free form or combined with fatty acids or carbohydrates. All steroids have a characteristic structural component consisting of four fused rings. Chemists identify the rings by capital letters and number the carbon atoms as shown in Figure **6.5.1*a***. Slight variations in this structure or in the atoms or groups attached to it produce profound differences in biological activity.

Cholesterol

Cholesterol (Figure **6.5.1*b***) does not occur in plants, but it is the most abundant steroid in the human body (240 g is a typical amount). Excess cholesterol is believed to be a primary factor in the development of atherosclerosis and heart disease, which are major health problems in the United States today. About half of the body's cholesterol is interspersed in the lipid bilayer of cell membranes. Much of the rest is converted to cholic acid, which is used in the formation of bile salts. Cholesterol is also a precursor in the synthesis of sex hormones, adrenal hormones, and vitamin D.

Excess cholesterol not metabolized by the body is released from the liver and transported by the blood to the gallbladder. Normally, it stays in solution there until being secreted into the intestine (as a component of bile) to be eliminated. Sometimes, however, cholesterol in the gallbladder precipitates in the form of gallstones (Figure **6.5.2**). Indeed, the name *cholesterol* is derived from the Greek *chole*, meaning "bile," and *stereos*, meaning "solid."



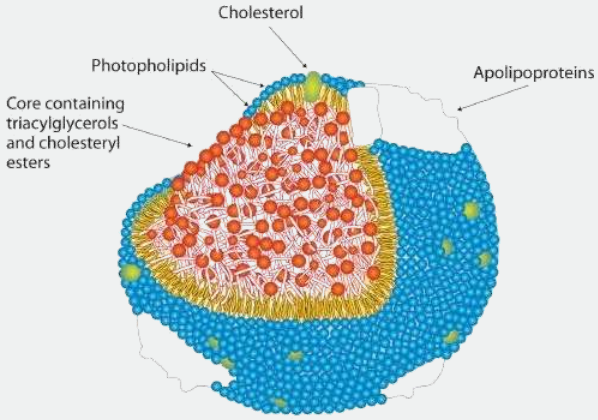
Figure **6.5.2**: Numerous small gallstones made up largely of cholesterol, all removed in one patient. Grid scale 1 mm

To Your Health: Cholesterol and Heart Disease

Heart disease is the leading cause of death in the United States for both men and women. The Centers for Disease Control and Prevention reported that heart disease claimed 631,636 lives in the United States (26% of all reported deaths) in 2006.

Scientists agree that elevated cholesterol levels in the blood, as well as high blood pressure, obesity, diabetes, and cigarette smoking, are associated with an increased risk of heart disease. A long-term investigation by the National Institutes of Health showed that among men ages 30 to 49, the incidence of heart disease was five times greater for those whose cholesterol levels were above 260 mg/100 mL of serum than for those with cholesterol levels of 200 mg/100 mL or less. The cholesterol content of blood varies considerably with age, diet, and sex. Young adults average about 170 mg of cholesterol per 100 mL of blood, whereas males at age 55 may have cholesterol levels at 250 mg/100 mL or higher because the rate of cholesterol breakdown decreases with age. Females tend to have lower blood cholesterol levels than males.

To understand the link between heart disease and cholesterol levels, it is important to understand how cholesterol and other lipids are transported in the body. Lipids, such as cholesterol, are not soluble in water and therefore cannot be transported in the blood (an aqueous medium) unless they are complexed with proteins that are soluble in water, forming assemblages called *lipoproteins*. Lipoproteins are classified according to their density, which is dependent on the relative amounts of protein and lipid they contain. Lipids are less dense than proteins, so lipoproteins containing a greater proportion of lipid are less dense than those containing a greater proportion of protein.



Research on cholesterol and its role in heart disease has focused on serum levels of low-density lipoproteins (LDLs) and high-density lipoproteins (HDLs). One of the most fascinating discoveries is that high levels of HDLs reduce a person's risk of developing heart disease, whereas high levels of LDLs increase that risk. Thus the serum LDL:HDL ratio is a better predictor of heart disease risk than the overall level of serum cholesterol. Persons who, because of hereditary or dietary factors, have high LDL:HDL ratios in their blood have a higher incidence of heart disease.

How do HDLs reduce the risk of developing heart disease? No one knows for sure, but one role of HDLs appears to be the transport of excess cholesterol to the liver, where it can be metabolized. Therefore, HDLs aid in removing cholesterol from blood and from the smooth muscle cells of the arterial wall.

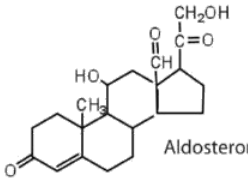
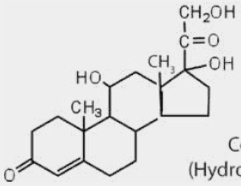
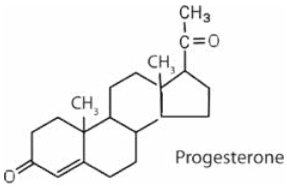
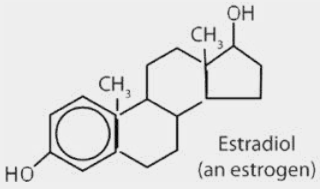
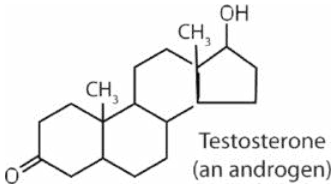
Dietary modifications and increased physical activity can help lower total cholesterol and improve the LDL:HDL ratio. The average American consumes about 600 mg of cholesterol from animal products each day and also synthesizes approximately 1 g of cholesterol each day, mostly in the liver. The amount of cholesterol synthesized is controlled by the cholesterol level in the blood; when the blood cholesterol level exceeds 150 mg/100 mL, the rate of cholesterol biosynthesis is halved. Hence, if cholesterol is present in the diet, a feedback mechanism suppresses its synthesis in the liver. However, the ratio of suppression is not a 1:1 ratio; the reduction in biosynthesis does not equal the amount of cholesterol ingested. Thus, dietary substitutions of unsaturated fat for saturated fat, as well as a reduction in consumption of *trans* fatty acids, is recommended to help lower serum cholesterol and the risk of heart disease.

Steroid Hormones

Hormones are chemical messengers that are released in one tissue and transported through the circulatory system to one or more other tissues. One group of hormones is known as steroid hormones because these hormones are synthesized from cholesterol, which is also a steroid. There are two main groups of steroid hormones: adrenocortical hormones and sex hormones.

The adrenocortical hormones, such as aldosterone and cortisol (Table **6.5.1**), are produced by the adrenal gland, which is located adjacent to each kidney. Aldosterone acts on most cells in the body, but it is particularly effective at enhancing the rate of reabsorption of sodium ions in the kidney tubules and increasing the secretion of potassium ions and/or hydrogen ions by the tubules. Because the concentration of sodium ions is the major factor influencing water retention in tissues, aldosterone promotes water retention and reduces urine output. Cortisol regulates several key metabolic reactions (for example, increasing glucose production and mobilizing fatty acids and amino acids). It also inhibits the inflammatory response of tissue to injury or stress. Cortisol and its analogs are therefore used pharmacologically as immunosuppressants after transplant operations and in the treatment of severe skin allergies and autoimmune diseases, such as rheumatoid arthritis.

Table **6.5.1**: Representative Steroid Hormones and Their Physiological Effects

Hormone	Effect
 Aldosterone	regulates salt metabolism; stimulates kidneys to retain sodium and excrete potassium
 Cortisol (Hydrocortisone)	stimulates the conversion of proteins to carbohydrates
 Progesterone	regulates the menstrual cycle; maintains pregnancy
 Estradiol (an estrogen)	stimulates female sex characteristics; regulates changes during the menstrual cycle
 Testosterone (an androgen)	stimulates and maintains male sex characteristics

The sex hormones are a class of steroid hormones secreted by the gonads (ovaries or testes), the placenta, and the adrenal glands. Testosterone and androstenedione are the primary male sex hormones, or *androgens*, controlling the primary sexual characteristics of males, or the development of the male genital organs and the continuous production of sperm. Androgens are also responsible for the development of secondary male characteristics, such as facial hair, deep voice, and muscle strength. Two kinds of sex

hormones are of particular importance in females: progesterone, which prepares the uterus for pregnancy and prevents the further release of eggs from the ovaries during pregnancy, and the estrogens, which are mainly responsible for the development of female secondary sexual characteristics, such as breast development and increased deposition of fat tissue in the breasts, the buttocks, and the thighs. Both males and females produce androgens and estrogens, differing in the amounts of secreted hormones rather than in the presence or absence of one or the other.

Sex hormones, both natural and synthetic, are sometimes used therapeutically. For example, a woman who has had her ovaries removed may be given female hormones to compensate. Some of the earliest chemical compounds employed in cancer chemotherapy were sex hormones. For example, estrogens are one treatment option for prostate cancer because they block the release and activity of testosterone. Testosterone enhances prostate cancer growth. Sex hormones are also administered in preparation for sex-change operations, to promote the development of the proper secondary sexual characteristics. Oral contraceptives are synthetic derivatives of the female sex hormones; they work by preventing ovulation.

Bile Salts

Bile is a yellowish green liquid (pH 7.8–8.6) produced in the liver. The most important constituents of bile are bile salts, which are sodium salts of amidelike combinations of bile acids, such as cholic acid (part (a) of Figure **6.5.3**) and an amine such as the amino acid glycine (part (b) of Figure **6.5.3**). They are synthesized from cholesterol in the liver, stored in the gallbladder, and then secreted in bile into the small intestine. In the gallbladder, the composition of bile gradually changes as water is absorbed and the other components become more concentrated.

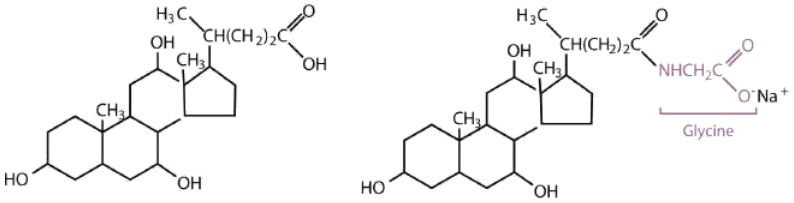


Figure **6.5.3** Bile Acids. (a) Cholic acid is an example of a bile acid. (b) Sodium glycocholate is a bile salt synthesized from cholic acid and glycine.

Because they contain both hydrophobic and hydrophilic groups, bile salts are highly effective detergents and emulsifying agents; they break down large fat globules into smaller ones and keep those smaller globules suspended in the aqueous digestive environment. Enzymes can then hydrolyze fat molecules more efficiently. Thus, the major function of bile salts is to aid in the digestion of dietary lipids.

> Surgical removal is often advised for a gallbladder that becomes infected, inflamed, or perforated. This surgery does not seriously affect digestion because bile is still produced by the liver, but the liver's bile is more dilute and its secretion into the small intestine is not as closely tied to the arrival of food.

Summary

Steroids have a four-fused-ring structure and have a variety of functions. Cholesterol is a steroid found in mammals that is needed for the formation of cell membranes, bile acids, and several hormones. Bile salts are secreted into the small intestine to aid in the digestion of fats.

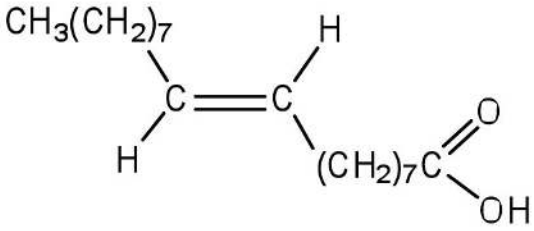
6.5: Steroids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **17.4: Steroids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

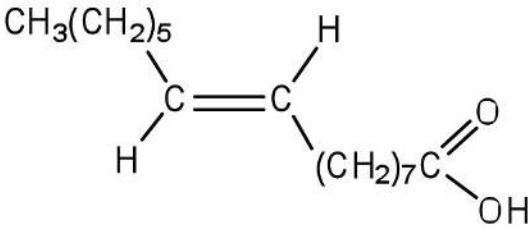
6.6: Exercises

Additional Exercises

1. The melting point of elaidic acid is 52°C.
   a. What trend is observed when comparing the melting points of elaidic acid, oleic acid, and stearic acid? Explain.
   b. Would you expect the melting point of palmitelaidic acid to be lower or higher than that of elaidic acid? Explain.

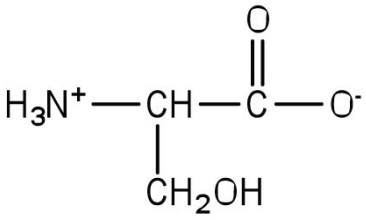


Elaidic acid



Palmitelaidic acid

2. Examine the labels on two brands of margarine and two brands of shortening and list the oils used in the various brands.
3. Draw a typical lecithin molecule that incorporates glycerol, palmitic acid, oleic acid, phosphoric acid, and choline. Circle all the ester bonds.
4. In cerebrosides, is the linkage between the fatty acid and sphingosine an amide bond or an ester bond? Justify your answer.
5. Serine is an amino acid that has the following structure. Draw the structure for a phosphatidylserine that contains a palmitic acid and a palmitoleic acid unit.

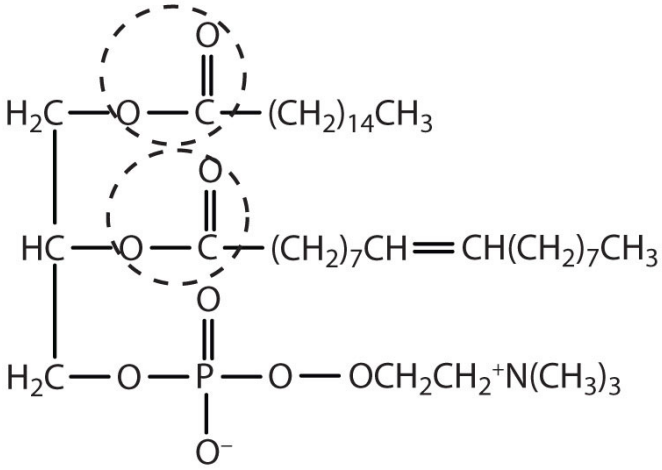


6. Explain whether each compound would be expected to diffuse through the lipid bilayer of a cell membrane.
   a. potassium chloride
   b. $CH_3CH_2CH_2CH_2CH_2CH_3$
   c. fructose
7. Identify the role of each steroid hormone in the body.
   a. progesterone
   b. aldosterone
   c. testosterone
   d. cortisol
8. How does the structure of cholic acid differ from that of cholesterol? Which compound would you expect to be more polar? Why?
9. a. What fatty acid is the precursor for the prostaglandins?
   b. Identify three biological effects of prostaglandins.
10. Why is it important to determine the ratio of LDLs to HDLs, rather than just the concentration of serum cholesterol?

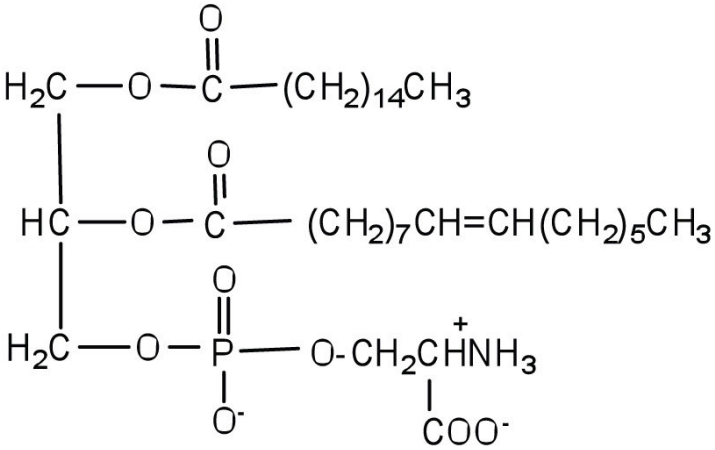
Answers

1. a. Stearic acid has the highest melting point, followed by elaidic acid, and then oleic acid with the lowest melting point. Elaidic acid is a *trans* fatty acid, and the carbon chains can pack together almost as tightly as those of the saturated stearic acid. Oleic acid is a *cis* fatty acid, and the bend in the hydrocarbon chain keeps these carbon chains from packing as closely together; fewer interactions lead to a much lower melting point.
   b. The melting point of palmitelaidic acid should be lower than that of elaidic acid because it has a shorter carbon chain (16, as compared to 18 for elaidic acid). The shorter the carbon chain, the lower the melting point due to a decrease in intermolecular interactions.

3.



5.



7. a. regulates the menstrual cycle and maintains pregnancy
   b. regulates salt metabolism by stimulating the kidneys to retain sodium and excrete potassium
   c. stimulates and maintains male sex characteristics
   d. stimulates the conversion of proteins to carbohydrates

9. a. arachidonic acid
   b. induce smooth muscle contraction, lower blood pressure, and contribute to the inflammatory response

6.6: Exercises is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.7: Lipids (Summary)

To ensure that you understand the material in this chapter, you should review the meanings of the bold terms in the following summary and ask yourself how they relate to the topics in the chapter.

Lipids, found in the body tissues of all organisms, are compounds that are more soluble in organic solvents than in water. Many of them contain **fatty acids**, which are carboxylic acids that generally contain an even number of 4–20 carbon atoms in an unbranched chain. **Saturated fatty acids** have no carbon-to-carbon double bonds. **Monounsaturated fatty acids** have a single carbon-to-carbon double bond, while **polyunsaturated fatty acids** have more than one carbon-to-carbon double bond. Linoleic and linolenic acid are known as **essential fatty acids** because the human body cannot synthesize these polyunsaturated fatty acids. The lipids known as **fats** and **oils** are triacylglycerols, more commonly called **triglycerides**—esters composed of three fatty acids joined to the trihydroxy alcohol glycerol. Fats are triglycerides that are solid at room temperature, and oils are triglycerides that are liquid at room temperature. Fats are found mainly in animals, and oils found mainly in plants. *Saturated triglycerides* are those containing a higher proportion of saturated fatty acid chains (fewer carbon-to-carbon double bonds); *unsaturated triglycerides* contain a higher proportion of unsaturated fatty acid chains.

Saponification is the hydrolysis of a triglyceride in a basic solution to form glycerol and three carboxylate anions or soap molecules. Other important reactions are the hydrogenation and oxidation of double bonds in unsaturated fats and oils.

Phospholipids are lipids containing phosphorus. In **phosphoglycerides**, the phosphorus is joined to an amino alcohol unit. Some phosphoglycerides, like lecithins, are used to stabilize an **emulsion**—a dispersion of two liquids that do not normally mix, such as oil and water. **Sphingolipids** are lipids for which the precursor is the amino alcohol sphingosine, rather than glycerol. A **glycolipid** has a sugar substituted at one of the OH groups of either glycerol or sphingosine. All are highly polar lipids found in cell membranes.

Polar lipids have dual characteristics: one part of the molecule is ionic and dissolves in water; the rest has a hydrocarbon structure and dissolves in nonpolar substances. Often, the ionic part is referred to as **hydrophilic** (literally, “water loving”) and the nonpolar part as **hydrophobic** (“water fearing”). When placed in water, polar lipids disperse into any one of three arrangements: *micelles*, *monolayers*, and *bilayers*. **Micelles** are aggregations of molecules in which the hydrocarbon tails of the lipids, being hydrophobic, are directed inward (away from the surrounding water), and the hydrophilic heads that are directed outward into the water. **Bilayers** are double layers arranged so that the hydrophobic tails are sandwiched between the two layers of hydrophilic heads, which remain in contact with the water.

Every living cell is enclosed by a *cell membrane* composed of a lipid bilayer. In animal cells, the bilayer consists mainly of phospholipids, glycolipids, and the steroid cholesterol. Embedded in the bilayer are **integral proteins**, and **peripheral proteins** are loosely associated with the surface of the bilayer. Everything between the cell membrane and the membrane of the cell nucleus is called the **cytoplasm**.

Most lipids can be saponified, but some, such as **steroids**, cannot be saponified. The steroid **cholesterol** is found in animal cells but never in plant cells. It is a main component of all cell membranes and a precursor for hormones, vitamin D, and bile salts. Bile salts are the most important constituents of **bile**, which is a yellowish-green liquid secreted by the gallbladder into the small intestine and is needed for the proper digestion of lipids.

6.7: Lipids (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **17.S: Lipids (Summary)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

CHAPTER OVERVIEW

7: Conformations and Stereochemistry

Stereochemistry is the study of how bonds are oriented in three-dimensional space. It is difficult to overstate the importance of stereochemistry in nature, and in the fields of biology and medicine in particular. As Pasteur so convincingly demonstrated, life itself is chiral: living things recognize different stereoisomers of organic compounds and process them accordingly.

7.1: Prelude to Conformations and Stereochemistry
7.2: Conformations of open-chain organic molecules
7.3: Chirality and stereoisomers
7.4: Naming chiral centers- the R and S system
7.5: Optical Activity
7.6: Compounds with multiple chiral centers
7.7: Stereochemistry of alkenes
7.8: Stereochemistry in biology and medicine

This page titled 7: Conformations and Stereochemistry is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg.

7.1: Prelude to Conformations and Stereochemistry

Molecular models are your friend!

*Because this chapter deals extensively with concepts that are inherently three-dimensional in nature, it will be very important for you to use a molecular modeling kit that is specifically intended for organic chemistry. Many of the ideas we will be exploring can be extremely confusing if you are limited to the two dimensions of this page. Be prepared to follow along with these discussions in **three** dimensions, with a molecular model in your hands!*

Introduction: Louis Pasteur and the discovery of molecular chirality

In 1848, a 25 year old chemist named Louis Pasteur made a startling - and some thought brash - claim to the scientific community. Pasteur was inexperienced, to say the least: he had only earned his doctorate the previous year, and had just started his first job as an assistant to a professor at the *Ecole normale superieure*, a university in Paris. Jean-Baptiste Biot, a highly respected physicist who had already made major contributions to scientific fields as diverse as meteorites, magnetism, and optics, was intrigued but unconvinced by Pasteur's claim. He invited the young man to come to his laboratory and reproduce his experiments.



(Photo credit: https://www.flickr.com/photos/nate/)

Decades earlier, Biot had discovered that aqueous solutions of some biologically-derived substances, such as tartaric acid, quinine, morphine, and various sugars, were optically active: that is, the plane of polarized light would rotate in either a positive (clockwise, or right-handed) or negative (counter-clockwise, or left-handed) direction when passed through the solutions. Nobody understood the source of this optical property. One of the biological substances known to be optically active was a salt of tartaric acid, a compound found in abundance in grapes and a major by-product of the wine-making industry.

The compound was dextrorotatory in solution – in other words, it rotated plane-polarized light in the positive (right-handed, or clockwise) direction. Curiously, though, chemists had also found that another form of processed tartaric acid was optically *in*active, despite that fact that it appeared to be identical to the optically active acid in every other respect. The optically inactive compound was called '*acide racemique*', from the Latin *racemus*, meaning 'bunch of grapes'.

Louis Pasteur's claims had to do with experiments he said he had done with the 'racemic' acid. Jean-Babtise Biot summoned Pasteur to his laboratory, and presented him with a sample of racemic acid which he himself had already confirmed was optically inactive. With Biot watching over his shoulder, and using Biot's reagents, Pasteur prepared the salt form of the acid, dissolved it in water, and left the aqueous solution in an uncovered flask to allow crystals to slowly form as the water evaporated.

Biot again summoned Pasteur to the lab a few days later when the crystallization was complete. Pasteur placed the crystals under a microscope, and began to painstakingly examine their shape, just as he had done in his original experiments. He had recognized that the crystals, which had a regular shape, were asymmetric: in other words, they could not be superimposed on their mirror

image. Scientists referred to asymmetric crystals and other asymmetric objects as being 'chiral', from the Greek word for 'hand'. Your hands are chiral objects, because although your right hand and your left hand are mirror images of one another, they cannot be superimposed. That is why you cannot fit your right hand in a left-handed glove.

More importantly, Pasteur had claimed that the chiral crystals he was seeing under the lens of his microscope were of two different types, and the two types were mirror images of each other: about half were what he termed 'right handed' and half were 'left-handed'. He carefully separated the right and left-handed crystals from each other, and presented the two samples to Biot. The eminent scientist then took what Pasteur told him were the left-handed crystals, dissolved them in water, and put the aqueous solution in a polarimeter, an instrument that measures optical rotation. Biot knew that the processed tartaric acid he had provided Pasteur had been optically inactive. He also knew that unprocessed tartaric acid from grapes had right-handed optical activity, whereas left-handed tartaric acid was unheard of. Before his eyes, however, he now saw that the solution was rotating light to the left. He turned to his young colleague and exclaimed, " *Mon cher enfant, j'ai tant aime´ les sciences dans ma vie que cela me fait battre le coeur*!' (My dear child, I have loved science so much during my life that this makes my heart pound!)

Biot had good reason to be so profoundly excited. Pasteur had just conclusively demonstrated, for the first time, the concept of **molecular chirality**: molecules themselves - not just macroscopic objects like crystals - could exhibit chirality, and could be separated into distinct right-handed and left-handed 'stereoisomers'. Tying together ideas from physics, chemistry, and biology, he had shown that nature could be chiral *at the molecular level*, and in doing do he had introduced to the world a new subfield which came to be known as 'stereochemistry'.

About ten years after his demonstration of molecular chirality, Pasteur went on to make another observation with profound implications for biological chemistry. It was already well known that 'natural' tartaric acid (the right-handed kind from grapes) could be fermented by bacteria. Pasteur discovered that the bacteria were selective with regard to the chirality of tartaric acid: no fermentation occurred when the bacteria were provided with pure left-handed acid, and when provided with racemic acid they specifically fermented the right-handed component, leaving the left-handed acid behind.

Pasteur was not aware, at the time of the discoveries described here, the details of the structural features of tartaric acid at the molecular level that made the acid chiral, although he made some predictions concerning the bonding patterns of carbon which turned out to be remarkably accurate. In the more than 150 years since Pasteur's initial tartaric acid work, we have greatly expanded our understanding of molecular chirality, and it is this knowledge that makes up the core of this chapter. Put simply, stereochemistry is the study of how bonds are oriented in three-dimensional space. It is difficult to overstate the importance of stereochemistry in nature, and in the fields of biology and medicine in particular. As Pasteur so convincingly demonstrated, life itself is chiral: living things recognize different stereoisomers of organic compounds and process them accordingly.

Organic Chemistry With a Biological Emphasis by Tim Soderberg (University of Minnesota, Morris)

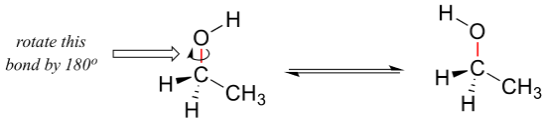
7.1: Prelude to Conformations and Stereochemistry is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

- **3.1: Prelude to Conformations and Stereochemistry** has no license indicated.

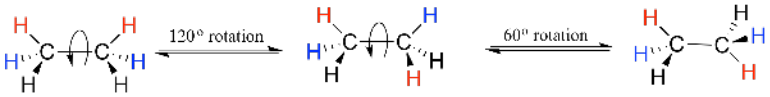
7.2: Conformations of open-chain organic molecules

Before we begin our exploration of stereochemistry and chirality, we first need to consider the subject of **conformational isomerism**, which has to do with rotation about single bonds.

We learned in section 2.1 that single bonds in organic molecules are free to rotate, due to the 'end-to-end' (sigma) nature of their orbital overlap. Consider the carbon-oxygen bond in ethanol, for example: with a 180° rotation about this bond, the shape of the molecule would look quite different:

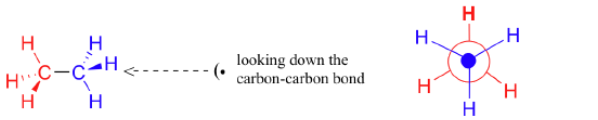


Or ethane: rotation about the carbon-carbon sigma bond results in many different possible three-dimensional arrangements of the atoms.



These different arrangements, resulting from sigma bond rotation, are referred to in organic chemistry as **conformations**. Any one specific conformation is called a **conformational isomer**, or **conformer**.

In order to better visualize different conformations of a molecule, it is convenient to use a drawing convention called the **Newman projection**. In a Newman projection, we look lengthwise down a specific bond of interest – in this case, the carbon-carbon bond in ethane. We depict the 'front' atom as a dot, and the 'back' atom as a larger circle.



'staggered' conformation
(Newman projection)

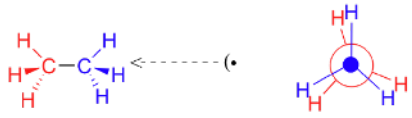
Interactive mode of ethane conformationsl

The six carbon-hydrogen bonds are shown as solid lines protruding from the two carbons. Note that we do *not* draw bonds as solid or dashed wedges in a Newman projection.

Looking down the C-C bond in this way, the angle formed between a C-H bond on the front carbon and a C-H bond on the back carbon is referred to as a **dihedral angle**. (The dihedral angle between the hour hand and the minute hand on a clock is 0° at noon, 90° at 3:00, and so forth).

The lowest energy conformation of ethane, shown in the figure above, is called the '**staggered'** conformation: all of the dihedral angles are 60°, and the distance between the front and back C-H bonds is maximized.

If we now rotate the front CH_3 group 60° clockwise, the molecule is in the highest energy '**eclipsed**' conformation, where the dihedral angles are all 0° (we stagger the bonds slightly in our Newman projection drawing so that we can see them all).

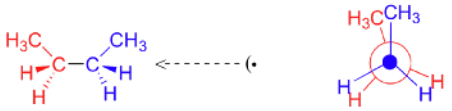


'eclipsed' conformation

The energy of the eclipsed conformation, where the electrons in the front and back C-H bonds are closer together, is approximately 12 kJ/mol higher than that of the staggered conformation.

Another 60° rotation returns the molecule to a second staggered conformation. This process can be continued all around the 360° circle, with three possible eclipsed conformations and three staggered conformations, in addition to an infinite number of conformations in between these two extremes.

Now let's consider butane, with its four-carbon chain. There are now three rotating carbon-carbon bonds to consider, but we will focus on the middle bond between C_2 and C_3. Below are two representations of butane in a conformation which puts the two CH_3 groups (C_1 and C_4) in the eclipsed position, with the two C-C bonds at a 0° dihedral angle.



eclipsed (A)

Interactive model of butane conformations

If we rotate the front, (blue) carbon by 60° clockwise, the butane molecule is now in a staggered conformation.



gauche

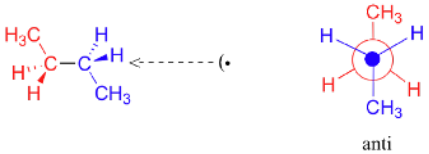
This is more specifically referred to as the **gauche** conformation of butane. Notice that although they are staggered, the two methyl groups are not as far apart as they could possibly be.

A further rotation of 60° gives us a second eclipsed conformation (B) in which both methyl groups are lined up with hydrogen atoms.



eclipsed (B)

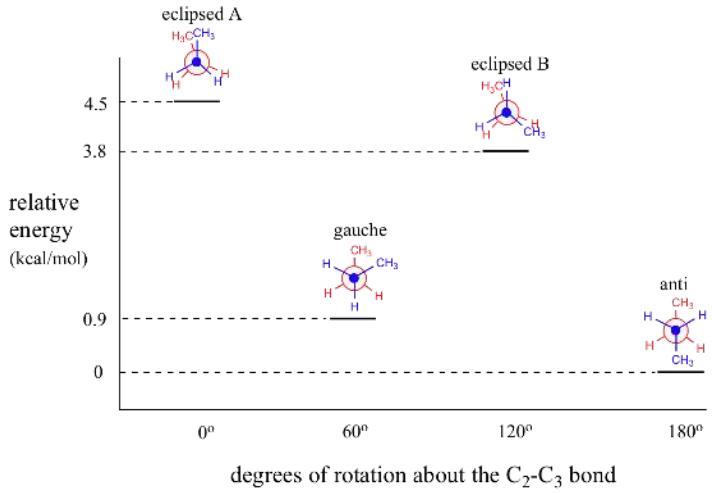
One more 60 rotation produces another staggered conformation called the **anti** conformation, where the two methyl groups are positioned opposite each other (a dihedral angle of 180°).



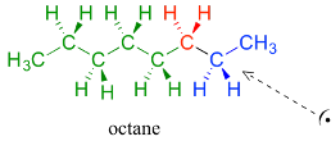
anti

As with ethane, the staggered conformations of butane are energy 'valleys', and the eclipsed conformations are energy 'peaks'. However, in the case of butane there are two different valleys, and two different peaks. The gauche conformation is a higher energy valley than the anti conformation due to **steric strain**, which is the repulsive interaction caused by the two bulky methyl groups being forced too close together. Clearly, steric strain is lower in the anti conformation. In the same way, steric strain causes the eclipsed A conformation - where the two methyl groups are as close together as they can possibly be - to be higher in energy than the two eclipsed B conformations.

The diagram below summarizes the relative energies for the various eclipsed, staggered, and gauche conformations.



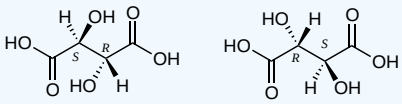
Because the anti conformation is lowest in energy (and also simply for ease of drawing), it is conventional to draw open-chain alkanes in a 'zigzag' form, which implies anti conformation at all carbon-carbon bonds. The figure below shows, as an example, a Newman projection looking down the C_2-C_3 bond of octane.





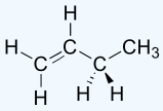
? Exercise 3.1

Using free rotation around C-C single bonds, show that (R,S) and (S,R)-tartaric acid are identical molecules.



Exercise 3.2

Draw a Newman projection, looking down the C_2-C_3 bond, of 1-butene in the conformation shown below (C_2 should be your *front* carbon).



Solutions to exercises

Online lectures from Khan Academy

Newman projections part I

Newman projections part II

Contributors and Attributions

- Organic Chemistry With a Biological Emphasis by Tim Soderberg (University of Minnesota, Morris)
- Layne Morsch (University of Illinois Springfield)

This page titled 7.2: Conformations of open-chain organic molecules is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg.

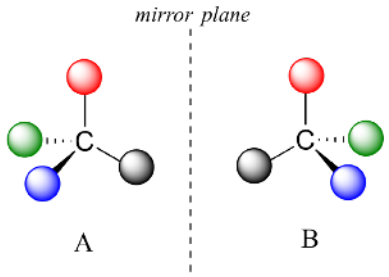
- **3.2: Conformations of open-chain organic molecules** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

7.3: Chirality and stereoisomers

We turn now to concept of chirality that formed the basis of the story about Louis Pasteur in the beginning of this chapter. Recall that the term **chiral**, from the Greek work for 'hand', refers to anything which cannot be superimposed on its own mirror image. Your hands, of course, are chiral - you cannot superimpose your left hand on your right, and you cannot fit your left hand into a right-handed glove (which is also a chiral object). Another way of saying this is that your hands do not have a **mirror plane of symmetry**: you cannot find any plane which bisects your hand in such a way that one side of the plane is a mirror image of the other side. *Chiral objects do not have a plane of symmetry*.

Your face, on the other hand is **achiral** - lacking chirality - because, some small deviations notwithstanding, you could superimpose your face onto its mirror image. If someone were to show you a mirror image photograph of your face, you could line the image up, point-for-point, with your actual face. Your face has a plane of symmetry, because the left side is the mirror image of the right side.

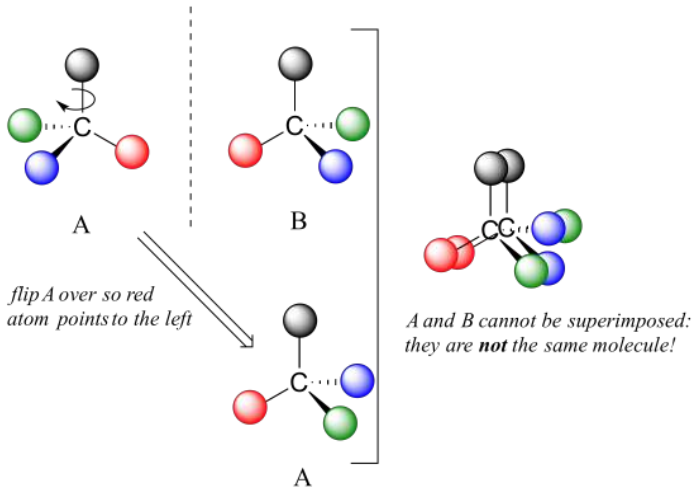
What Pasteur, Biot, and their contemporaries did not yet fully understand when Pasteur made his discovery of molecular chirality was the *source* of chirality at the molecular level. It stood to reason that a chiral molecule is one that does not contain a plane of symmetry, and thus cannot be superimposed on its mirror image. We now know that chiral molecules contain one or more **chiral centers**, which are almost always tetrahedral (sp^3-hybridized) carbons with four different substituents. Consider the cartoon molecule A below: a tetrahedral carbon, with four different substituents denoted by balls of four different colors (for the time being, don't worry about exactly what these substituents could be - we will see real examples very soon).



Another image

The mirror image of A, which we will call B, is drawn on the right side of the figure, and an imaginary mirror is in the middle. Notice that every point on A lines up through the mirror with the same point on B: in other words, if A looked in the mirror, it would see B looking back.

Now, if we flip compound A over and try to superimpose it point for point on compound B, we find that we cannot do it: if we superimpose any two colored balls, then the other two are misaligned.

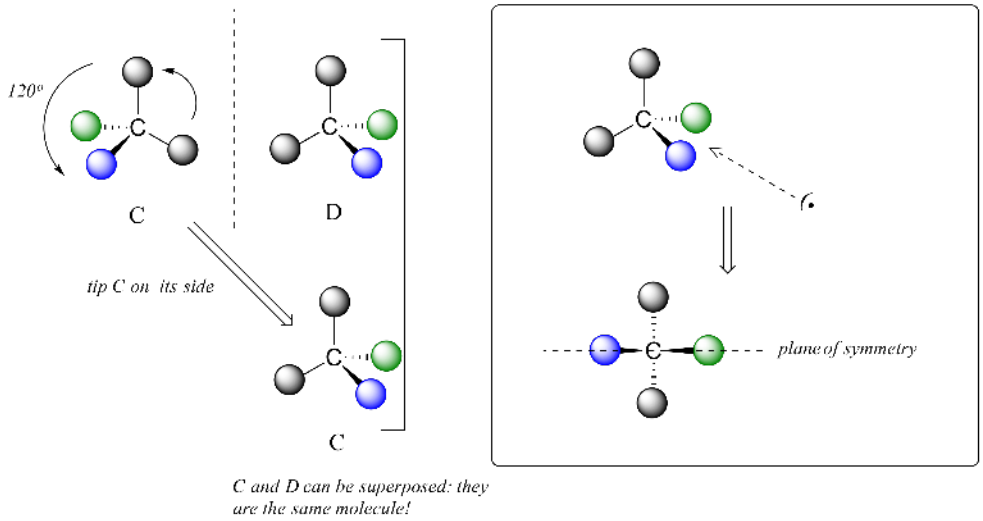


A is not superimposable on its mirror image (B), thus by definition A is a chiral molecule. It follows that B also is not superimposable on *its* mirror image (A), and thus it is also a chiral molecule. Also notice in the figure below (and convince yourself with models) that neither A nor B has an internal plane of symmetry.

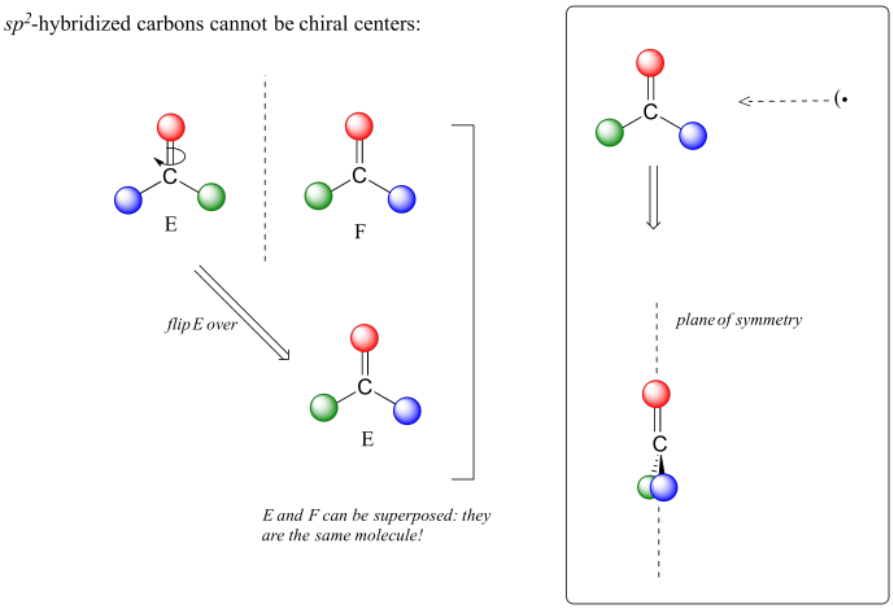
A and B are **stereoisomers**: molecules with the same molecular formula and the same bonding arrangement, but *a different arrangement of atoms in space*. There are two types of stereoisomers: enantiomers and diastereomers. **Enantiomers** are pairs of stereoisomers which are mirror images of each other: thus, A and B are enantiomers. It should be self-evident that a chiral molecule will always have one (and *only* one) enantiomer: enantiomers come in pairs. Enantiomers have identical physical properties (melting point, boiling point, density, and so on). However, enantiomers do differ in how they interact with polarized light (we will learn more about this soon) and they may also interact in very different ways with other chiral molecules - proteins, for example. We will begin to explore this last idea in later in this chapter, and see many examples throughout the remainder of our study of biological organic chemistry.

Diastereomers are stereoisomers which are *not* mirror images of each other. For now, we will concentrate on understanding enantiomers, and come back to diastereomers later.

We defined a chiral center as a tetrahedral carbon with four different substituents. If, instead, a tetrahedral carbon has two identical substituents (two black atoms in the cartoon figure below), then of course it still has a mirror image (*everything* has a mirror image, unless we are talking about a vampire!) However, it is superimposable on its mirror image, and has a plane of symmetry.



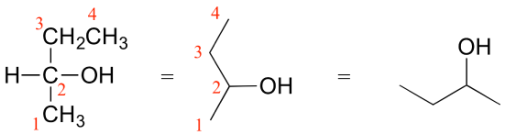
This molecule is **achiral** (lacking chirality). Using the same reasoning, we can see that a trigonal planar (sp^2-hybridized) carbon is also not a chiral center.



Notice that structure E can be superimposed on F, its mirror image - all you have to do is pick E up, flip it over, and it is the same as F. This molecule has a plane of symmetry, and is achiral.

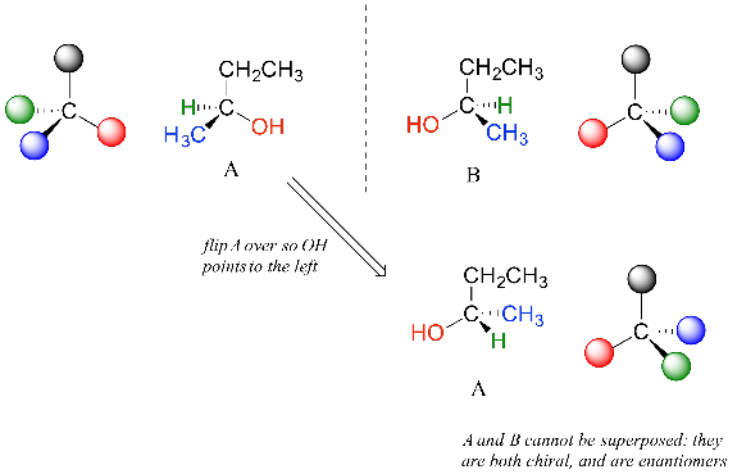
Let's apply our general discussion to real molecules. For now, we will limit our discussion to molecules with a single chiral center. It turns out that tartaric acid, the subject of our chapter introduction, has *two* chiral centers, so we will come back to it later.

Consider 2-butanol, drawn in two dimensions below.



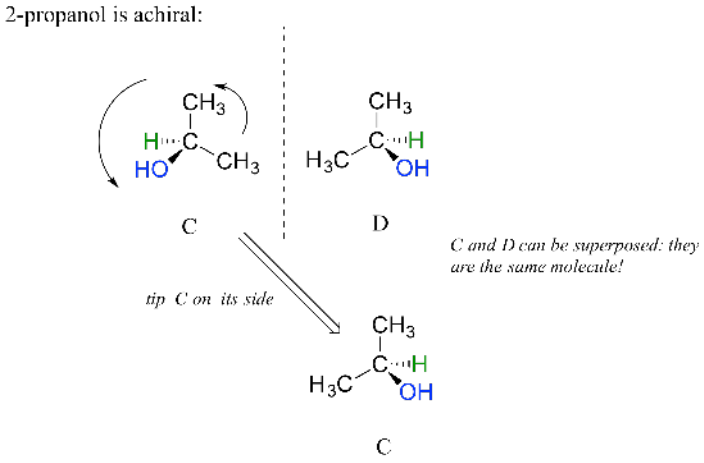
2-butanol
(drawn in 2D)

Carbon #2 is a chiral center: it is sp^3-hybridized and tetrahedral (even though it is not drawn that way above), and the four things attached to is are different: a hydrogen, a methyl ($-CH_3$) group, an ethyl ($-CH_2CH_3$) group, and a hydroxyl (OH) group. Let's draw the bonding at C_2 in three dimensions, and call this structure A. We will also draw the mirror image of A, and call this structure B.



When we try to superimpose A onto B, we find that we cannot do it. A and B are both chiral molecules, and they are enantiomers of each other.

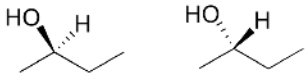
2-propanol, unlike 2-butanol, is *not* a chiral molecule. Carbon #2 is bonded to two identical substituents (methyl groups), and so it is not a chiral center.



Notice that 2-propanol is superimposable on its own mirror image.

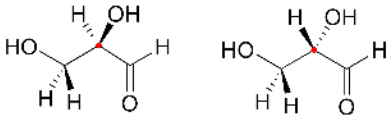
When we look at very simple molecules like 2-butanol, it is not difficult to draw out the mirror image and recognize that it is not superimposable. However, with larger, more complex molecules, this can be a daunting challenge in terms of drawing and three-dimensional visualization. The easy way to determine if a molecule is chiral is simply to look for the presence of one or more chiral centers: molecules with chiral centers will (almost always) be chiral. We insert the 'almost always' caveat here because it is possible to come up with the exception to this rule - we will have more to say on this later, but don't worry about it for now.

Here's another trick to make your stereochemical life easier: if you want to draw the enantiomer of a chiral molecule, it is not necessary to go to the trouble of drawing the point-for-point mirror image, as we have done up to now for purposes of illustration. Instead, keep the carbon skeleton the same, and simply reverse the solid and dashed wedge bonds on the chiral carbon: that accomplishes the same thing. You should use models to convince yourself that this is true, and also to convince yourself that swapping *any two substituents* about the chiral carbon will result in the formation of the enantiomer.

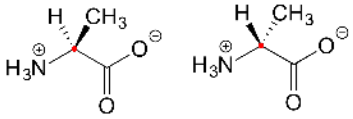


two enantiomers of 2-butanol

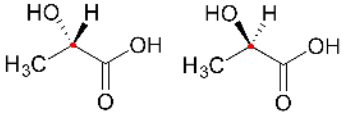
Here are four more examples of chiral biomolecules, each one shown as a pair of enantiomers, with chiral centers marked by red dots.



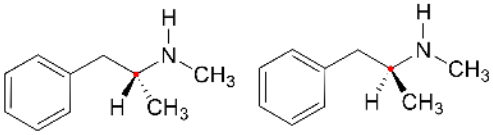
glyceraldehyde



alanine

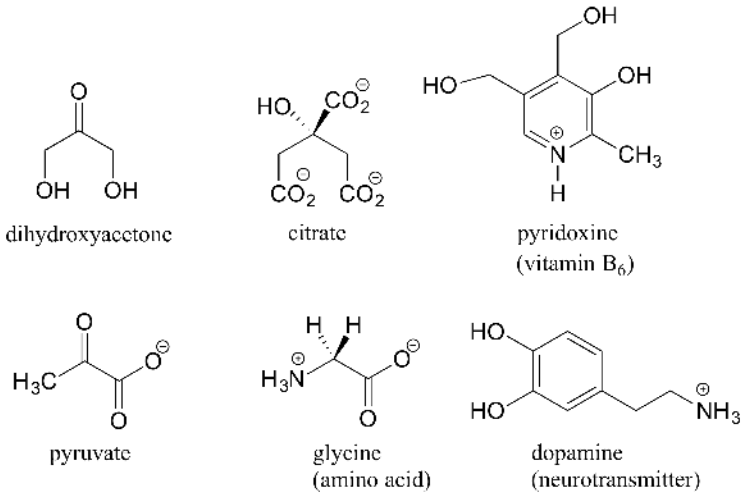


lactic acid



methamphetamine

Here are some examples of ***achiral*** biomolecules - convince yourself that none of them contain a chiral center:

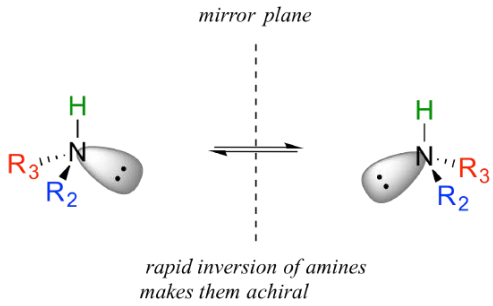


When looking for chiral centers, it is important to recognize that the question of whether or not the dashed/solid wedge drawing convention is used is irrelevant. Chiral molecules are sometimes drawn without using wedges (although obviously this means that stereochemical information is being omitted). Conversely, wedges may be used on carbons that are *not* chiral centers – look, for example, at the drawings of glycine and citrate in the figure above.

Can a chiral center be something other than a tetrahedral carbon with four different substituents? The answer to this question is 'yes' - however, these alternative chiral centers are very rare in the context of biological organic chemistry, and outside the scope of our discussion here.

You may also have wondered about amines: shouldn't we consider a secondary or tertiary amine to be a chiral center, as they are tetrahedral and attached to four different substituents, if the lone-pair electrons are counted as a 'substituent'? Put another way, isn't an amine non-superimposable on its mirror image?

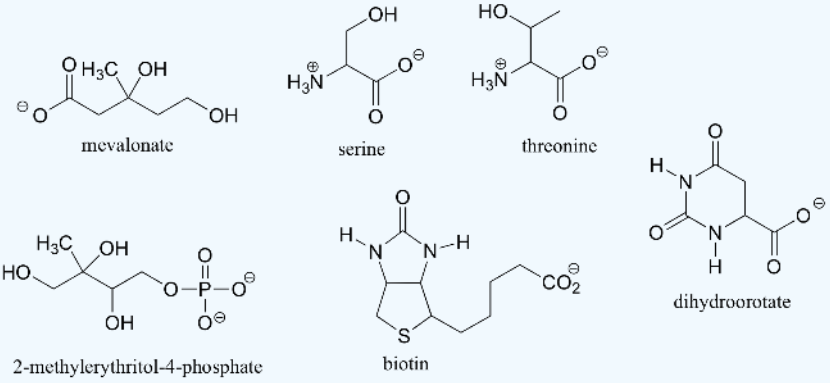
The answer: yes it is, in the static picture, but in reality, the nitrogen of an amine is rapidly and reversibly inverting, or turning inside out, at room temperature.



If you have trouble picturing this, take an old tennis ball and cut it in half. Then, take one of the concave halves and flip it inside out, then back again: this is what the amine is doing. The end result is that the two 'enantiomers' if the amine are actually two rapidly interconverting forms of the same molecule, and thus the amine itself is not a chiral center. This inversion process does *not* take place on a tetrahedral carbon, which of course has no lone-pair electrons.

? Exercise 3.8

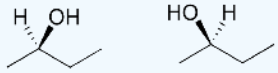
Locate all of the chiral centers (there may be more than one in a molecule). Remember, hydrogen atoms bonded to carbon usually are not drawn in the line structure convention - but they are still there!



Exercise 3.9

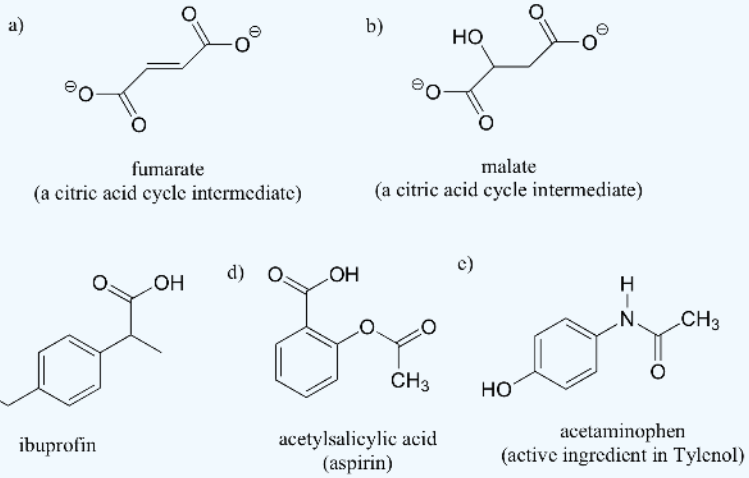
a) Draw two enantiomers of i) mevalonate and ii) serine.

b) Are the two 2-butanol structures below enantiomers?



Exercise 3.10

Label the molecules below as chiral or achiral, and locate all chiral centers.



Solutions to exercises

Khan Academy video tutorials

Chirality

Enantiomers

This page titled 7.3: Chirality and stereoisomers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg via source content that was edited to the style and standards of the LibreTexts platform.

- **3.4: Chirality and stereoisomers** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

7.4: Naming chiral centers- the R and S system

Chemists need a convenient way to distinguish one stereoisomer from another. The **Cahn-Ingold-Prelog system** is a set of rules that allows us to unambiguously define the stereochemical configuration of any stereocenter, using the designations '***R*** ' (from the Latin *rectus*, meaning right-handed) or ' ***S*** ' (from the Latin *sinister*, meaning left-handed).

The rules for this system of stereochemical nomenclature are, on the surface, fairly simple.

Rules for assigning an R/S designation to a chiral center

1: Assign priorities to the four substituents, with #1 being the highest priority and #4 the lowest. Priorities are based on the atomic number.

2: Trace a circle from #1 to #2 to #3.

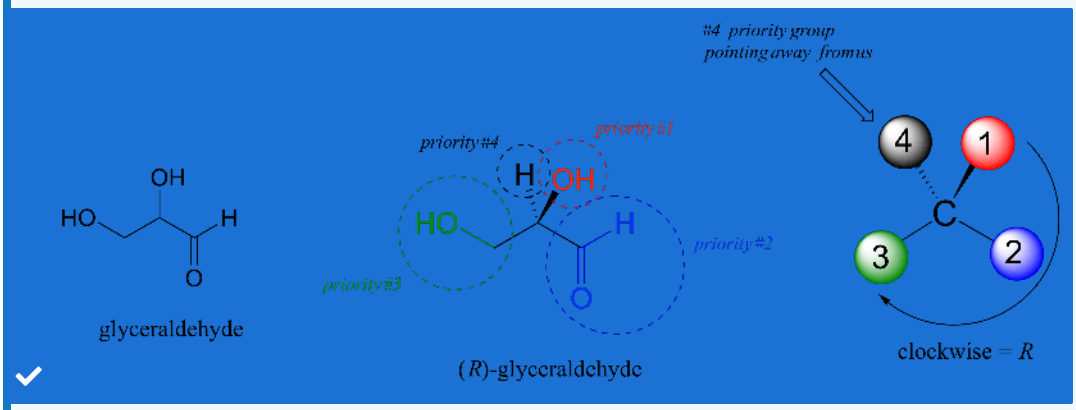
3: Determine the orientation of the #4 priority group. If it is oriented into the plane of the page (away from you), go to step 4a. If it is oriented out of the plane of the page (toward you) go to step 4b.

4a: *(#4 group pointing away from you*): a clockwise circle in part 2 corresponds to the *R* configuration, while a counterclockwise circle corresponds to the *S* configuration.

4b: *(#4 group pointing toward you*): a clockwise circle in part 2 corresponds to the *S* configuration, while a counterclockwise circle corresponds to the *R* configuration.

We'll use the 3-carbon sugar glyceraldehyde as our first example. The first thing that we must do is to assign a **priority** to each of the four substituents bound to the chiral center. We first look at the atoms that are directly bonded to the chiral center: these are H, O (in the hydroxyl), C (in the aldehyde), and C (in the CH_2OH group).

Assigning R/S configuration to glyceraldehyde

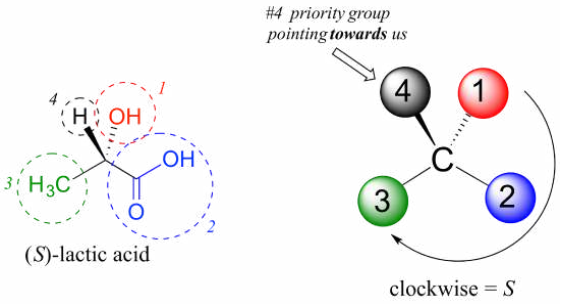


Two priorities are easy: hydrogen, with an atomic number of 1, is the lowest (#4) priority, and the hydroxyl oxygen, with atomic number 8, is priority #1. Carbon has an atomic number of 6. Which of the two 'C' groups is priority #2, the aldehyde or the CH_2OH? To determine this, we move one more bond away from the chiral center: for the aldehyde we have a *double* bond to an oxygen, while on the CH_2OH group we have a *single* bond to an oxygen. If the atom is the same, double bonds have a higher priority than single bonds. Therefore, the aldehyde group is assigned #2 priority and the CH_2OH group the #3 priority.

With our priorities assigned, we look next at the #4 priority group (the hydrogen) and see that it is pointed back away from us, into the plane of the page - thus step 4a from the procedure above applies. Then, we trace a circle defined by the #1, #2, and #3 priority groups, in increasing order. The circle is clockwise, which by step 4a tells us that this carbon has the '*R*' configuration, and that this molecule is (*R*)-glyceraldehyde. Its enantiomer, by definition, must be (*S*)-glyceraldehyde.

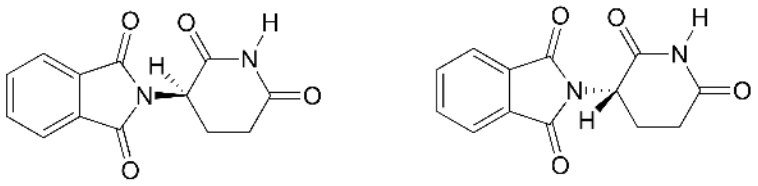
Next, let's look at one of the enantiomers of lactic acid and determine the configuration of the chiral center. Clearly, H is the #4 substituent and OH is #1. Owing to its three bonds to oxygen, the carbon on the acid group takes priority #2, and the methyl group

takes #3. The #4 group, hydrogen, happens to be drawn pointing *toward* us (out of the plane of the page) in this figure, so we use step 4b: The circle traced from #1 to #2 to #3 is clockwise, which means that the chiral center has the *S* configuration.



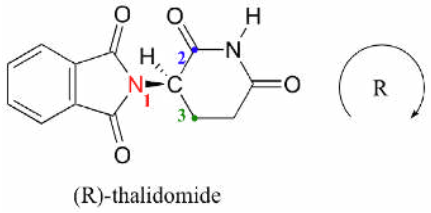
Interactive model of (S)-alanine

The drug thalidomide is an interesting - but tragic - case study in the importance of stereochemistry in drug design. First manufactured by a German drug company and prescribed widely in Europe and Australia in the late 1950's as a sedative and remedy for morning sickness in pregnant women, thalidomide was soon implicated as the cause of devastating birth defects in babies born to women who had taken it. Thalidomide contains a chiral center, and thus exists in two enantiomeric forms. It was marketed as a **racemic mixture**: in other words, a 50:50 mixture of both enantiomers.



two enantiomers of thalidomide

Let's try to determine the stereochemical configuration of the enantiomer on the left. Of the four bonds to the chiral center, the #4 priority is hydrogen. The nitrogen group is #1, the carbonyl side of the ring is #2, and the $-CH_2$ side of the ring is #3.



The hydrogen is shown pointing away from us, and the prioritized substituents trace a clockwise circle: this is the *R* enantiomer of thalidomide. The other enantiomer, of course, must have the S configuration.

Although scientists are still unsure today how thalidomide works, experimental evidence suggests that it was actually the *R* enantiomer that had the desired medical effects, while the *S* enantiomer caused the birth defects. Even with this knowledge, however, pure (*R*)-thalidomide is not safe, because enzymes in the body rapidly convert between the two enantiomers - we will see how that happens in chapter 12.

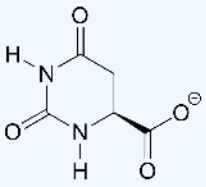
As a historical note, thalidomide was never approved for use in the United States. This was thanks in large part to the efforts of Dr. Frances Kelsey, a Food and Drug officer who, at peril to her career, blocked its approval due to her concerns about the lack of adequate safety studies, particularly with regard to the drug's ability to enter the bloodstream of a developing fetus. Unfortunately, though, at that time clinical trials for new drugs involved widespread and unregulated distribution to doctors and their patients across the country, so families in the U.S. were not spared from the damage caused.

Very recently a close derivative of thalidomide has become legal to prescribe again in the United States, with strict safety measures enforced, for the treatment of a form of blood cancer called multiple myeloma. In Brazil, thalidomide is used in the treatment of

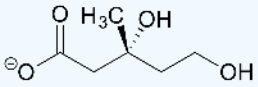
leprosy - but despite safety measures, children are still being born with thalidomide-related defects.

Exercise 3.11

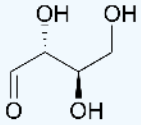
Determine the stereochemical configurations of the chiral centers in the biomolecules shown below.



dihydroorotate



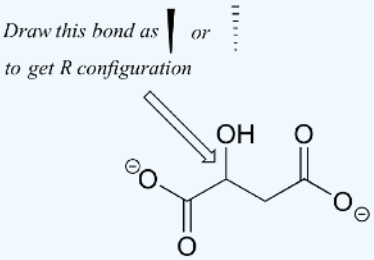
mevalonate



(D)-erythrose

Exercise 3.12

Should the (*R*) enantiomer of malate have a solid or dashed wedge for the C-O bond in the figure below?



(*R*)-malate

Exercise 3.13

Using solid or dashed wedges to show stereochemistry, draw the (*R*) enantiomer of ibuprofen and the (*S*) enantiomer of 2-methylerythritol-4-phosphate (structures are shown earlier in this chapter without stereochemistry).

Solutions to exercises

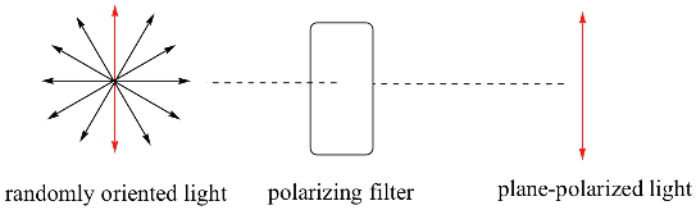
Khan Academy video tutorial on the R-S naming system

This page titled 7.4: Naming chiral centers- the R and S system is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg via source content that was edited to the style and standards of the LibreTexts platform.

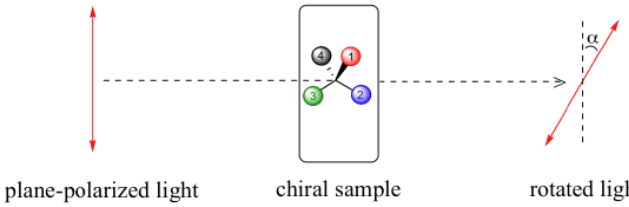
- **3.5: Naming chiral centers- the R and S system** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

7.5: Optical Activity

Chiral molecules, as we learned in the introduction to this chapter, have an interesting optical property. You may know from studying physics that light waves are oscillating electric and magnetic fields. In ordinary light, the oscillation is randomly oriented in an infinite number of planes. When ordinary light is passed through a polarizer, all planes of oscillation are filtered out except one, resulting in **plane-polarized light**.



A beam of plane-polarized light, when passed through a sample of a chiral compound, interacts with the compound in such a way that the angle of oscillation will rotate. This property is called **optical activity**.



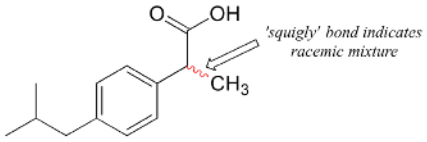
If a compound rotates plane polarized light in the clockwise (+) direction, it is said to be **dextrorotatory**, while if it rotates light in the counterclockwise (-) direction it is **levorotatory**. The magnitude of the observed optical activity is dependent on temperature, the wavelength of light used, solvent, concentration of the chiral sample, and the path length of the sample tube (path length is the length that the plane-polarized light travels through the chiral sample). Typically, optical activity measurements are made in a 1 decimeter (10 cm) path-length sample tube at 25 °C, using as a light source the so-called “D-line” from a sodium lamp, which has a wavelength of 589 nm. The **specific rotation** of a pure chiral compound at 25° is expressed by the expression:

$$[\alpha]_{\mathrm{D}}^{25} = \frac{\alpha_{\mathrm{obs}}}{lc}$$

where α_{obs} is the observed rotation, l is path length in decimeters, and c is the concentration of the sample in grams per 100 mL. In other words, the specific rotation of a chiral compound is the optical rotation that is observed when 1 g of the compound is dissolved in enough of a given solvent to make 100 mL solution, and the rotation is measured in a 1 dm cuvette at 25 ^{o}C using light from a sodium lamp.

Every chiral molecule has a characteristic specific rotation, which is recorded in the chemical literature as a physical property just like melting point or density. Different enantiomers of a compound will always rotate plane-polarized light with an equal but opposite magnitude. (*S*)-ibuprofen, for example, has a specific rotation of +54.5^{o} (dextrorotatory) in methanol, while (*R*)-ibuprofen has a specific rotation of -54.5^{o}. *There is no relationship between chiral compound's R/S designation and the direction of its specific rotation.* For example, the *S* enantiomer of ibuprofen is dextrorotatory, but the *S* enantiomer of glyceraldehyde is levorotatory.

A 50:50 mixture of two enantiomers (a racemic mixture) will have no observable optical activity, because the two optical activities cancel each other out. In a structural drawing, a 'squigly' bond from a chiral center indicates a mixture of both *R* and *S* configurations.



(+/-) ibuprofen
(racemic)

Chiral molecules are often labeled according to whether they are dextrorotatory or levorotatory as well as by their *R/S* designation. For example, the pure enantiomers of ibuprofen are labeled (*S*)-(+)-ibuprofen and (*R*)-(-)-ibuprofen, while (±)-ibuprofen refers to the racemic mixture, which is the form in which the drug is sold to consumers.

? Exercise 3.14

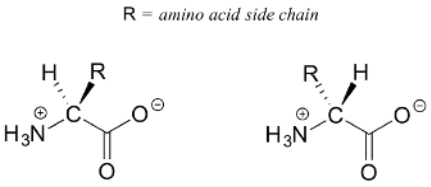
The specific rotation of (*R*)-limonene is +11.5º in ethanol. What is the expected observed rotation of a sample of 6.00 g (*S*)-limonene dissolved in ethanol to a total volume of 80.0 mL in a 1.00 dm (10.0 cm) pathlength cuvette?

Exercise 3.15

The specific rotation of (*S*)-carvone is +61°, measured 'neat' (pure liquid sample, no solvent). The optical rotation of a mixture of *R* and *S* carvone is measured at -23°. Which enantiomer is in excess in the mixture?

Solutions to exercises

All of the twenty natural amino acids except glycine have a chiral center at their alpha-carbon (recall that basic amino acid structure and terminology was introduced in section 1.3). Virtually all of the amino acids found in nature, both in the form of free amino acids or incorporated into peptides and proteins, have what is referred to in the biochemical literature as the 'L' configuration:



L-amino acids
(common in nature)

D-amino acids
(rare in nature)

All but one of the 19 L-amino acids have *S* stereochemistry at the a-carbon, using the rules of the *R/S* naming system.

D-amino acids are very rare in nature, but we will learn about an interesting example of a peptide containing one D-amino acid residue later in chapter 12.

? Exercise 3.16

Which of the 20 common L-amino acids found in nature has the *R* configuration? Refer to the amino acid table for structures.

Solutions to exercises

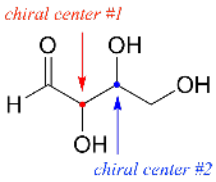
Khan Academy video tutorial on optical acitivity

This page titled 7.5: Optical Activity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg via source content that was edited to the style and standards of the LibreTexts platform.

- **3.6: Optical Activity** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

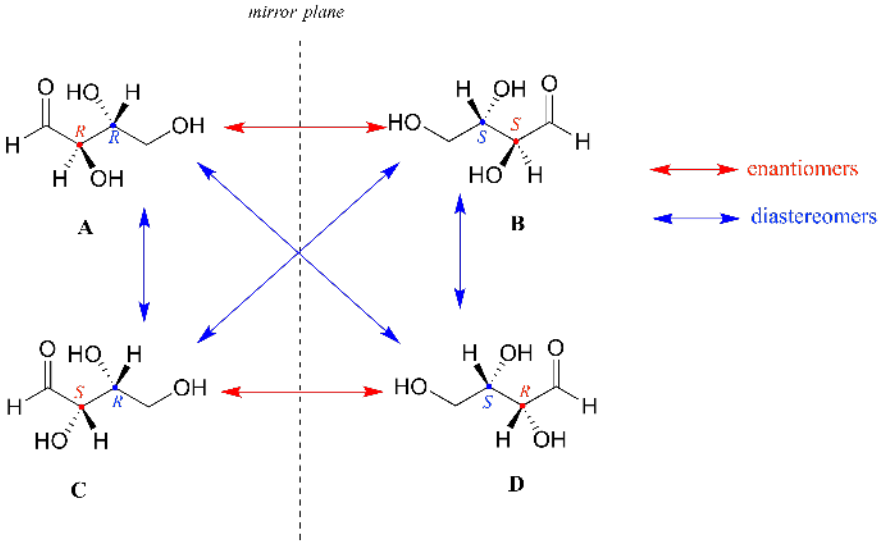
7.6: Compounds with multiple chiral centers

So far, we have been analyzing compounds with a single chiral center. Next, we turn our attention to those which have multiple chiral centers. We'll start with some stereoisomeric four-carbon sugars with two chiral centers.



To avoid confusion, we will simply refer to the different stereoisomers by capital letters.

Look first at compound A below. Both chiral centers in have the *R* configuration (you should confirm this for yourself!). The mirror image of Compound A is compound B, which has the *S* configuration at both chiral centers. If we were to pick up compound A, flip it over and put it next to compound B, we would see that they are *not* superimposable (again, confirm this for yourself with your models!). A and B are nonsuperimposable mirror images: in other words, enantiomers.



Now, look at compound C, in which the configuration is *S* at chiral center 1 and *R* at chiral center 2. Compounds A and C are stereoisomers: they have the same molecular formula and the same bond connectivity, but a different arrangement of atoms in space (recall that this is the definition of the term 'stereoisomer). However, they are *not* mirror images of each other (confirm this with your models!), and so they are *not* enantiomers. By definition, they are **diastereomers** of each other.

Notice that compounds C and B also have a diastereomeric relationship, by the same definition.

So, compounds A and B are a pair of enantiomers, and compound C is a diastereomer of both of them. Does compound C have its own enantiomer? Compound D is the mirror image of compound C, and the two are not superimposable. Therefore, C and D are a pair of enantiomers. Compound D is also a diastereomer of compounds A and B.

This can also seem very confusing at first, but there some simple shortcuts to analyzing stereoisomers:

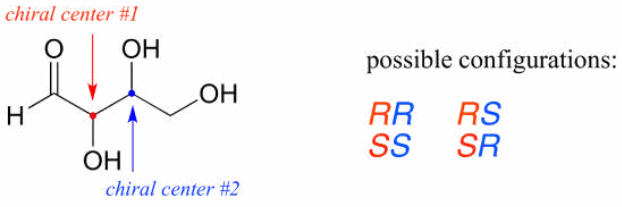
Stereoisomer shortcuts

If ***all*** of the chiral centers are of opposite R/S configuration between two stereoisomers, they are enantiomers.

If ***at least one, but not all*** of the chiral centers are opposite between two stereoisomers, they are diastereomers.

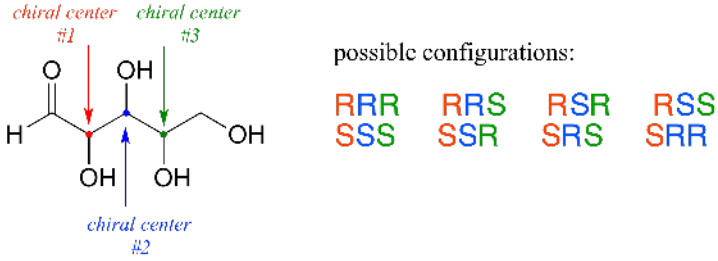
(*Note: these shortcuts to not take into account the possibility of additional stereoisomers due to alkene groups: we will come to that later)*

Here's another way of looking at the four stereoisomers, where one chiral center is associated with red and the other blue. Pairs of enantiomers are stacked together.

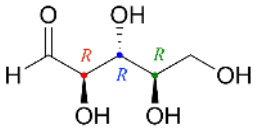


We know, using the shortcut above, that the enantiomer of RR must be SS - both chiral centers are different. We also know that RS and SR are diastereomers of RR, because in each case one - but not both - chiral centers are different.

Now, let's extend our analysis to a sugar molecule with three chiral centers. Going through all the possible combinations, we come up with eight total stereoisomers - four pairs of enantiomers.

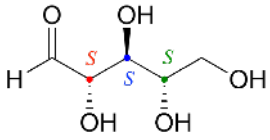


Let's draw the *RRR* stereoisomer. Being careful to draw the wedge bonds correctly so that they match the *RRR* configurations, we get:



RRR stereoisomer

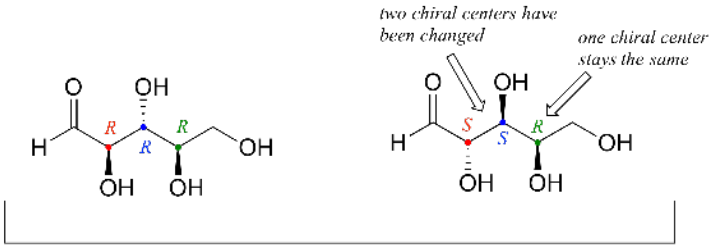
Now, using the above drawing as our model, drawing any other stereoisomer is easy. If we want to draw the enantiomer of *RRR*, we don't need to try to visualize the mirror image, we just start with the *RRR* structure and invert the configuration at *every* chiral center to get *SSS*.



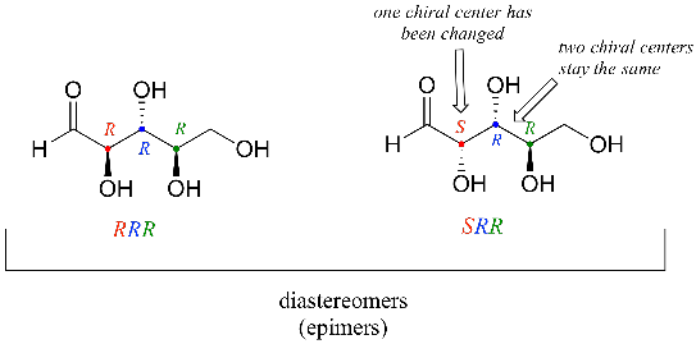
SSS stereoisomer

Try making models of *RRR* and *SSS* and confirm that they are in fact nonsuperimposable mirror images of each other.

There are six diastereomers of *RRR*. To draw one of them, we just invert the configuration of at least one, but not all three, of the chiral centers. Let's invert the configuration at chiral center 1 and 2, but leave chiral center 3 unchanged. This gives us the *SSR* configuration.



One more definition at this point: diastereomers which differ at only a single chiral center are called **epimers**. For example, *RRR* and *SRR* are epimers:



The *RRR* and *SSR* stereoisomers shown earlier are diastereomers but *not* epimers because they differ at *two* of the three chiral centers.

? Exercise 3.17

a) Draw the structure of the *enantiomer* of the *SRS* stereoisomer of the sugar used in the previous example.

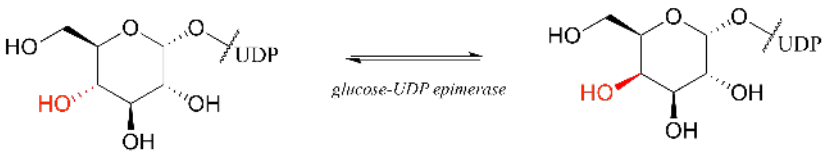
b) List (using the *XXX* format, not drawing the structures) all of the epimers of *SRS*.

c) List all of the stereoisomers that are diastereomers, but not epimers, of *SRS*.

Solutions to exercises

The epimer term is useful because in biochemical pathways, compounds with multiple chiral centers are isomerized at one specific center by enzymes known as **epimerases**. Two examples of epimerase-catalyzed reactions are below.

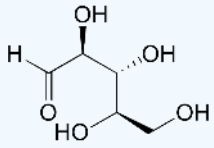




We know that enantiomers have identical physical properties and equal but opposite magnitude specific rotation. Diastereomers, in theory at least, have *different* physical properties – we stipulate 'in theory' because sometimes the physical properties of two or more diastereomers are so similar that it is very difficult to distinguish between them. In addition, the specific rotation values of diastereomers are unrelated – they could be the same sign or opposite signs, similar in magnitude or very dissimilar.

? Exercise 3.18

The sugar below is one of the stereoisomers that we have been discussing.



?

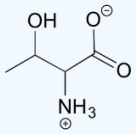
The only problem is, it is drawn with the carbon backbone in a different orientation from what we have seen. Determine the configuration at each chiral center to determine which stereoisomer it is.

Exercise 3.19

Draw the enantiomer of the xylulose-5-phosphate structure in the previous figure.

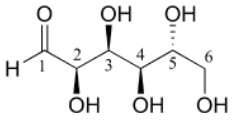
Exercise 3.20

The structure of the amino acid D-threonine, drawn without stereochemistry, is shown below. D-threonine has the (S) configuration at both of its chiral centers. Draw D-threonine, it's enantiomer, and its two diastereomers.



Solutions to exercises

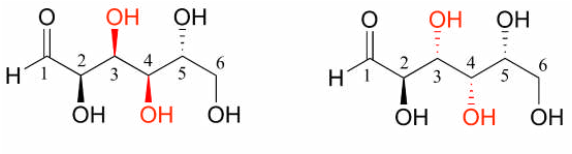
Here is some more practice in identifying isomeric relationships. D-glucose is the monosaccharide that serves as the entrance point for the glycolysis pathway and as a building block for the carbohydrate biopolymers starch and cellulose. The 'D' in D-glucose stands for *dextrarotatory* and is part of the specialized nomenclature system for sugars, which we will not concern ourselves with here. The open-chain structure of the sugar is shown below.



D-glucose

Because D-glucose has four chiral centers, it can exist in a total of 2^4 = 16 different stereoisomeric forms: it has one enantiomer and 14 diastereomers.

Now, let's compare the structures of the two sugars D-glucose and D-gulose, and try to determine their relationship.

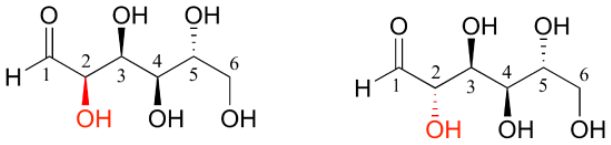


(diastereomers)

https://chem.libretexts.org/@go/page/255615

The two structures have the same molecular formula and the same connectivity, therefore they must be stereoisomers. They each have four chiral centers, and the configuration is different at two of these centers (at carbons #3 and #4). They are diastereomers.

Now, look at the structures of D-glucose and D-mannose.

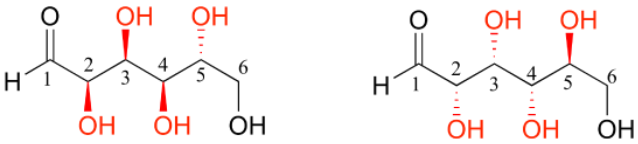


D-glucose D-mannose

(epimers)

Here, everything is the same except for the configuration of the chiral center at carbon #2. The two sugars differ at only one of the four chiral centers, so again they are diastereomers, and more specifically they are epimers.

D-glucose and L-glucose are enantiomers, because they differ at all four chiral centers.

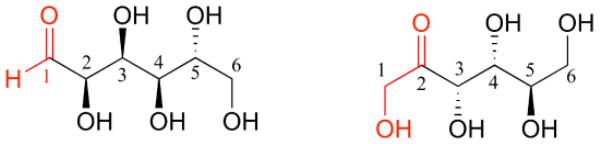


D-glucose L-glucose

(enantiomers)

D-glucose is the enantiomer commonly found in nature.

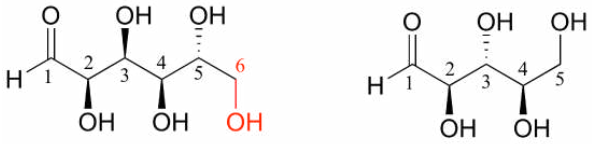
D-glucose and D-fructose are not stereoisomers, because they have different bonding connectivity: glucose has an aldehyde group, while fructose has a ketone. The two sugars do, however, have the same molecular formula, so by definition they are constitutional isomers.



D-glucose D-fructose

(constitutional isomers)

D-glucose and D-ribose are not isomers of any kind, because they have different molecular formulas.

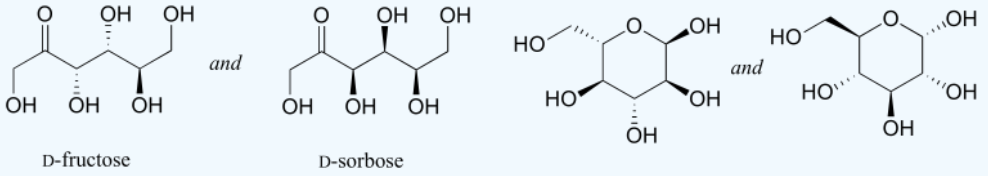


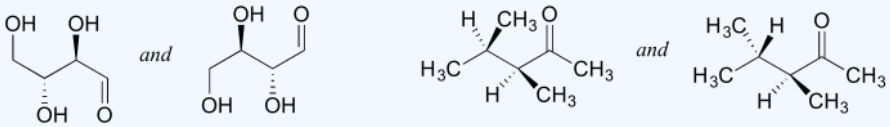
D-glucose D-ribose

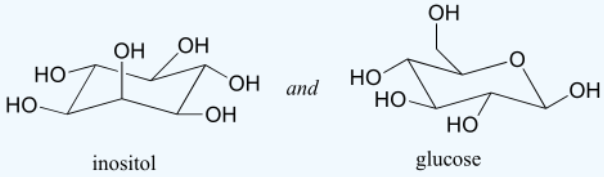
(not isomers)

? Exercise 3.21

Identify the relationship between each pair of structures. Your choices are: not isomers, constitutional isomers, diastereomers but not epimers, epimers, enantiomers, or same molecule

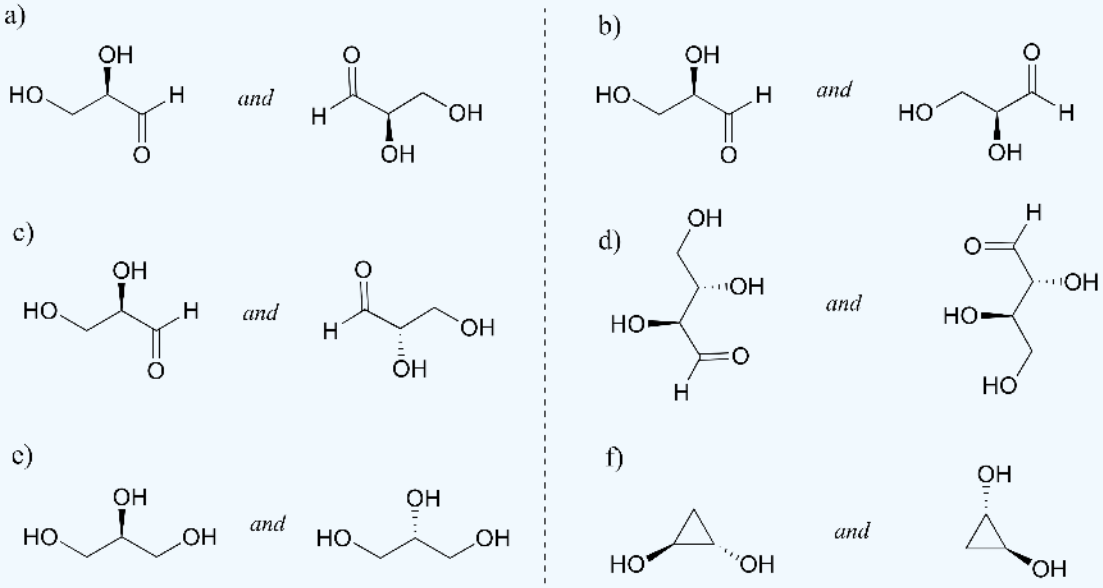






Exercise 3.22

Identify the relationship between each pair of structures. *Hint* - figure out the configuration of each chiral center.



Solutions to exercises

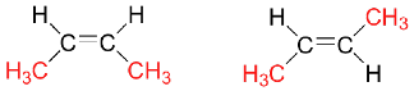
Khan Academy video tutorial on stereoisomeric relationships

This page titled 7.6: Compounds with multiple chiral centers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg.

- **3.7: Compounds with multiple chiral centers** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

7.7: Stereochemistry of alkenes

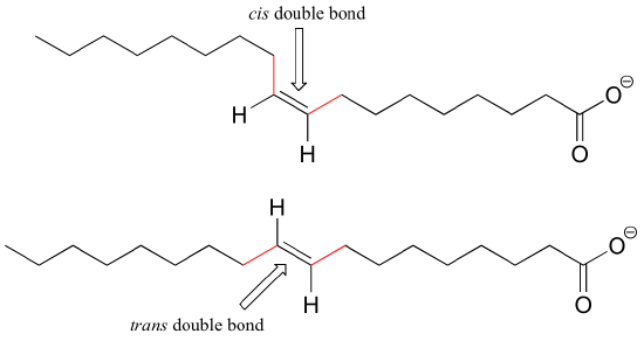
When we talk about stereochemistry, we are not always talking about chiral compounds and chiral centers. Consider *cis*- and *trans*-2-butene:



cis-2-butene *trans*-2-butene

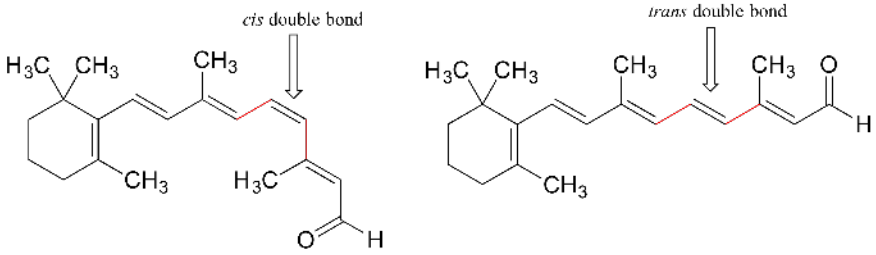
Each can be superimposed on its own mirror image, and neither is chiral (also, note the lack of a chiral center!) However, they both have the same molecular formula and the same bonding connectivity, so by definition they are stereoisomers of each other. Because they are not mirror images, they must be diastereomers. An alkene group which can exist in two stereoisomeric forms is referred to as **stereogenic**.

Alkene groups in natural unsaturated fatty acids are normally *cis*, but *trans*-fatty acids (which are thought to be associated with heart disease and other health problems) are found in some food products.



cis and *trans* fatty acids

Retinal is a light-sensitive molecule, derived from vitamin A, that is found in the rod cells of the eye. When light enters the eye through the retina, one form of retinal is converted to a diastereomer when a *cis* double bond is converted to *trans* (we"ll learn how this happens in chapter 13). This changes the shape of the molecule and the way that it binds to the vision protein rhodopsin, which in turn initiates a chain of events that leads to a signal being sent to the vision center of the brain.

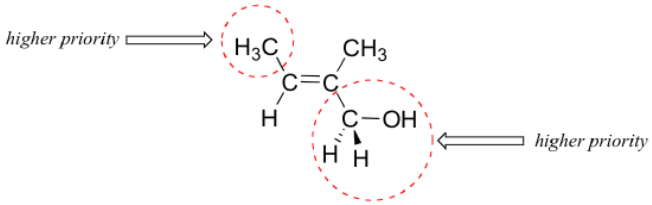


11-*cis*-retinal all-*trans*-retinal

While the terms *cis* and *trans* are quite clear in the examples above, in some cases they can be ambiguous, and a more rigorous stereochemical designation is required. To unambiguously designate alkene stereochemistry, it is best to use the designators '*E*' and 'Z' rather than *trans* and *cis*. To use this naming system, we first decide which is the higher priority group on each carbon of the double bond, using the same priority rules that we learned for the *R/S* system. If the higher-priority groups are one the same side of the double bond, it is a *Z*-alkene, and if they are on the opposite side it is an *E*-alkene. A memory device that many students find helpful is the phrase *'Z = zame zide'*.

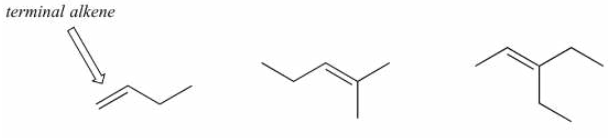
Shown below is an example of an *E*-alkene: notice that, although the two methyl groups are on the same side relative to one another, the alkene has *E* stereochemistry according to the rules of the *E/Z* system because one of the methyl groups takes a higher

priority (relative to a hydrogen) and the other takes lower priority (relative to a primary alcohol). The *cis/trans* terms would be ambiguous for this compound.



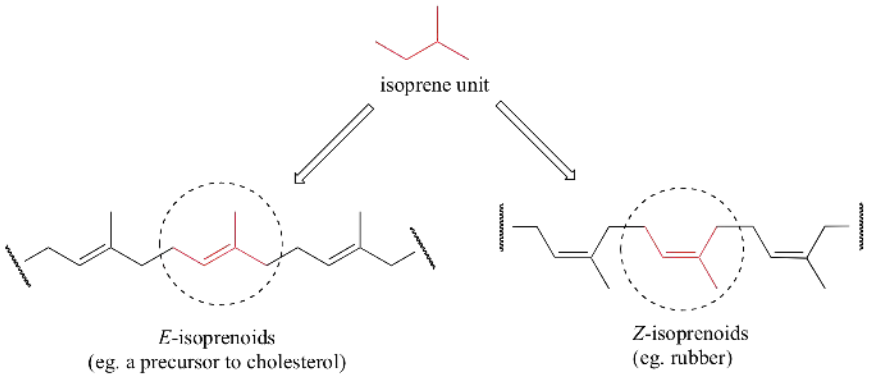
an *E*-alkene

Not all alkenes can be labeled *E* or *Z:* if one (or both) of the double-bonded carbons has identical substituents, the alkene is not stereogenic, and thus cannot be assigned an E or Z configuration. Terminal alkenes, in which one of the alkene carbons is bonded to two hydrogen atoms, are the most commonly seen type of nonstereogenic alkene.



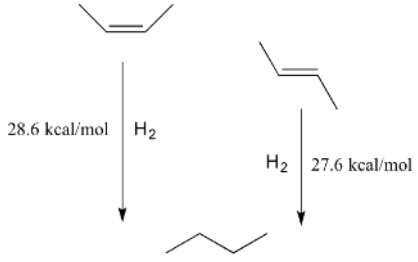
nonstereogenic alkenes (cannot be labelled *E* or *Z)*

Natural rubber is a polymer composed of five-carbon isoprenoid building blocks linked with *Z* stereochemistry. The same isoprenpoid building blocks can also be connected with *E* stereochemistry, leading to a polymer that is a precursor to cholesterol and many other natural isoprenoid compounds found in all forms of life.



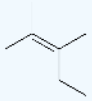
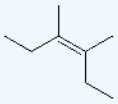
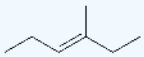
Alkenes located inside a five- or six-membered ring, such as in cyclohexene, are not generally labeled *E* or *Z*, simply because the closed geometry of the ring allows for only one stereochemical possibility. (*E*)-cyclohexene is not physically possible, so it is not necessary to include the (*Z*) designator for cyclohexene. Larger rings, however, can hypothetically have *E* or *Z* alkene groups: two actual examples are included in exercise 3.26 below.

As a general rule, alkenes with the bulkiest groups on opposite sides of the double bond are more stable, due to reduced steric strain. The *trans* (*E)* diastereomer of 2-butene, for example, is slightly lower in energy than the *cis (Z)* diastereomer, as seen by their relative heats of hydrogenation to butane (see section 2.2C for a reminder of the meaning of 'heat of hydrogenation'.)



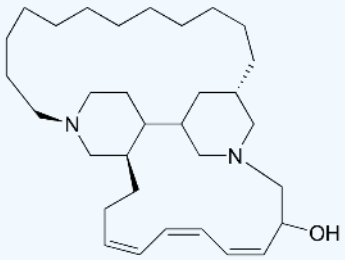
Exercise 3.25

Label the alkene groups below as *E*, *Z*, or *N* (for a nonstereogenic alkene).

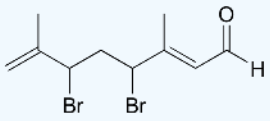
a)  b)  c)  d) 

Exercise 3.26

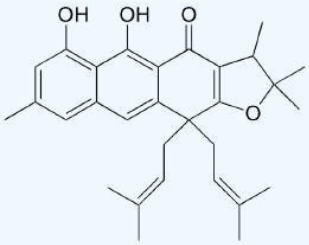
The compounds shown below were all isolated from natural sources and their structures reported in a 2007 issue of the *Journal of Natural Products*, an American Chemical Society publication. Label all alkene groups that are not inside 5- or 6-membered rings as *E*, *Z*, or *N* (for a nonstereogenic alkene).



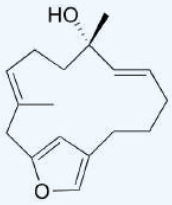
from marine sponge
J. Nat. Prod. **2007**, *70*, 538



from algea
J. Nat. Prod. **2007**, *70*, 596



from Nketto plant in Cameroon
J. Nat. Prod. **2007**, *70*, 600



from tree bark in Thailand
J. Nat. Prod. **2007**, *70*, 659

How many possible stereoisomers?

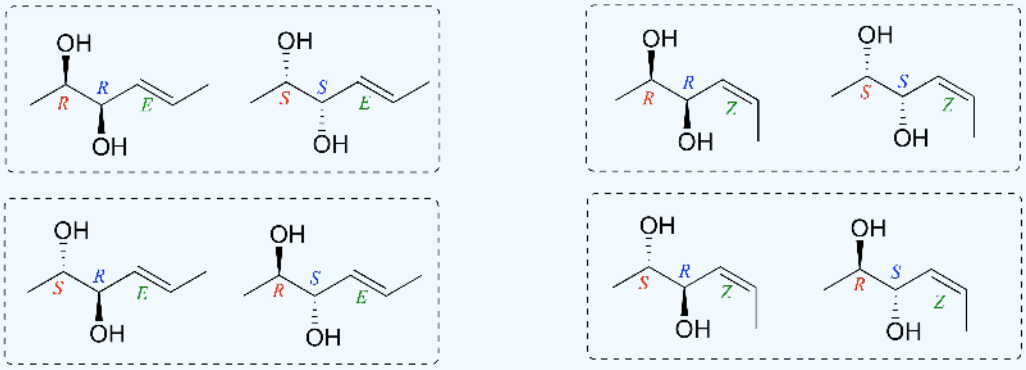
How do we know how many stereoisomers are possible for a given structure? There is actually a straightforward way to figure this out. All we need to do is count the number of chiral centers and stereogenic alkene groups, the use this following rule:

number of stereoisomeric forms = 2^n

. . . where *n* = the number of chiral centers ***plus*** the number of stereogenic alkene groups

(the rare exception to this rule is when a meso *form is possible - in this case, the rule becomes 2^n-1)*

Consider for example a molecule with two chiral centers and one stereogenic alkene. By the rule stated above, we know right away that there must be eight possible stereoisomers. Drawing out all the possibilities, we see:



8 stereoisomers of a compound with two chiral centers plus one stereogenic alkene

We see that, for example, *RRE* has one enantiomer, the *SSE* compound. The six other stereoisomers are all diastereomers of *RRE*.

It needs to be stressed that the enantiomer of the *RRE* compound is the *SSE* compound, ***not*** the *SSZ* compound. Remember, the *E/Z* relationship is diastereomeric, not enantiomeric. Use models to convince yourself that the *RRE* and the *SSE* isomers are mirror images of each other, while *RRE* and *SSZ* compounds are not. In general, to get the enantiomer of a compound, we invert all chiral centers *but leave all stereogenic alkenes the same.*

? Exercise 3.27

Draw the enantiomer of each the compounds below, and assign configurations to all chiral centers and stereogenic alkenes. How many diastereomers are possible for each of the structures you drew?



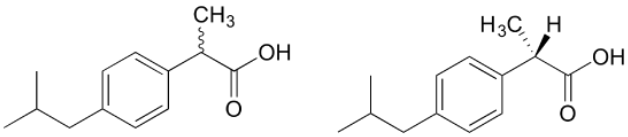
This page titled 7.7: Stereochemistry of alkenes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg.

- **3.10: Stereochemistry of alkenes** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

7.8: Stereochemistry in biology and medicine

While challenging to understand and visualize, the stereochemistry concepts we have explored in this chapter are integral to the study of living things. The vast majority of biological molecules contain chiral centers and/or stereogenic alkene groups. Most importantly, proteins are chiral, which of course includes all of the the enzymes which catalyze the chemical reactions of a cell, the receptors which transmit information within or between cells, and the antibodies which bind specifically to potentially harmful invaders. You know from your biology classes that proteins, because they fold up into a specific three dimensional shape, are able to very specifically recognize and bind to other organic molecules. The ligand or substrate bound by a particular protein could be a small organic molecule such as pyruvate all the way up to a large biopolymer such as a specific region of DNA, RNA, or another protein. Because they are chiral molecules, *proteins are very sensitive to the stereochemistry of their ligands*: a protein may bind specifically to (*R*)-glyceraldehyde, for example, but not bind to (*S*)-glyceraldehyde, just as your right hand will not fit into a left-handed baseball glove (see end of chapter for a link to an animation illustrating this concept).

The over-the-counter painkiller ibuprofen is currently sold as a racemic mixture, but only the *S* enantiomer is effective, due to the specific way it is able to bind to and inhibit the action of prostaglandin H_2 synthase, an enzyme in the body's inflammation response process.



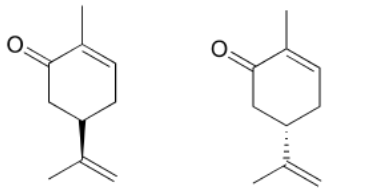
racemic ibuprofen

(*S*)-ibuprofen
(active enantiomer)

The *R* enantiomer of ibuprofen does not bind to prostaglandin H_2 synthase in the same way as the *S* enantiomer, and as a consequence does not exert the same inhibitory effect on the enzyme's action (see *Nature Chemical Biology* **2011**, *7*, 803 for more details). Fortunately, (*R*)-ibuprofen apparently does not cause any harmful side effects, and is in fact isomerized gradually by an enzyme in the body to (*S*)-ibuprofen.

Earlier in this chapter we discussed the tragic case of thalidomide, and mentioned that it appears that it is specifically the *S* enantiomer which caused birth defects. Many different proposals have been made over the past decades to try to explain the teratogenic (birth defect-causing) effect of the drug, but a clear understanding still evades the scientific community. In 2010, however, a team in Japan reported evidence that thalidomide binds specifically to a protein called 'thereblon'. Furthermore, when production of thereblon is blocked in female zebra fish, developmental defects occur in her offspring which are very similar to the defects caused by the administration of thalidomide, pointing to the likelihood that thalidomide binding somehow inactivates the protein, thus initiating the teratogenic effect. (http://news.sciencemag.org/2010/03/t...-partner-crime)

You can, with a quick trip to the grocery store, directly experience the biological importance of stereoisomerism. Carvone is a chiral, plant-derived molecule that contributes to the smell of spearmint in the *R* form and caraway (a spice) in the *S* form.



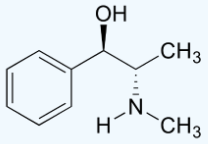
(*R*)-carvone
(spearmint odor)

(*S*)-carvone
(caraway odor)

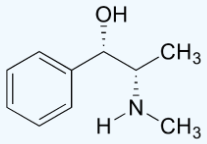
Although details are not known, the two enantiomers presumably interact differently with one or more smell receptor proteins in your nose, generating the transmission of different chemical signals to the olfactory center of your brain.

Exercise 3.28

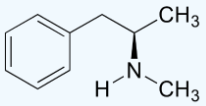
Ephedrine, found in the Chinese traditional medicine *ma huang*, is a stimulant and appetite suppressant. Both pseudoephedrine and levomethamphetamine are active ingredients in over-the-counter nasal decongestants. Methamphetamine is a highly addictive and illegal stimulant, and is usually prepared in illicit 'meth labs' using pseudoephedrine as a starting point.



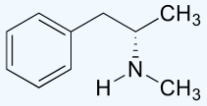
ephedrine



pseudoephedrine



methamphetamine



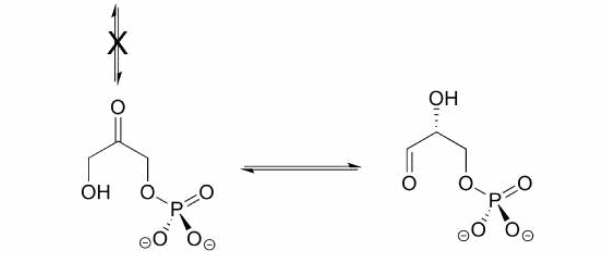
levomethamphetamine

What is the relationship between ephedrine and pseudoephedrine? Between methamphetamine and levomethamphetamine? Between pseudoephedrine and methamphetamine? Your choices are: not isomers, constitutional isomers, diastereomers, enantiomers, or same molecule.

Solutions to exercises

Enzymes are very specific with regard to the stereochemistry of the reactions they catalyze. *When the product of a biochemical reaction contains a chiral center or a stereogenic alkene, with very few exceptions only one stereoisomer of the product is formed.* In the glycolysis pathway, for example, the enzyme triose-phosphate isomerase catalyzes the reversible interconversion between dihydroxyacetone (which is achiral) and (*R*)-glyceraldehyde phosphate. The (*S*)-glyceraldehyde enantiomer is not formed by this enzyme in the left-to-right reaction, and is not used as a starting compound in the right-to-left reaction - it does not 'fit' in the active site of the enzyme.

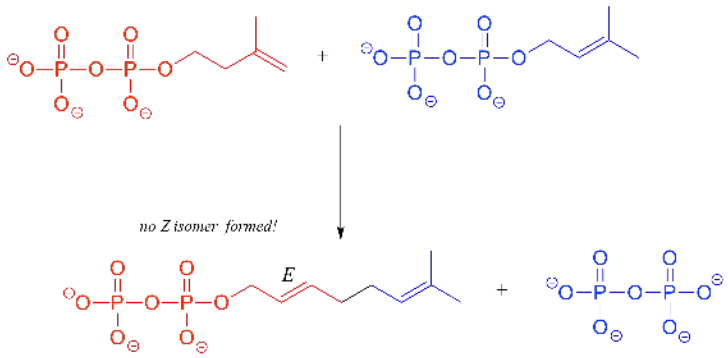
(*S*)-glyceraldehyde phosphate



dihydroxyacetone phosphate
(achiral)

(*R*)-glyceraldehyde phosphate

In the isoprenoid biosynthesis pathway, two five-carbon building-block molecules combine to form a ten-carbon chain containing an *E*-alkene group. The enzyme does *not* catalyze formation of the *Z* diastereomer.



In chapters 9-17 of this book, and continuing on into your study of biological and organic chemistry, you will be learning about how enzymes are able to achieve these feats of stereochemical specificity. If you take a more advanced class in organic synthesis, you will also learn how laboratory chemists are figuring out ingenious ways to exert control over the stereochemical outcomes of nonenzymatic reactions, an area of chemistry that is particularly important in the pharmaceutical industry.

This page titled 7.8: Stereochemistry in biology and medicine is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg.

- **3.11: Stereochemistry in biology and medicine** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

CHAPTER OVERVIEW

8: Acids and Bases

Many of us are familiar with the group of chemicals called acids. But do you know what it takes for a compound to be an acid? Actually, there are several different definitions of acid that chemistry uses, and each definition is appropriate under different circumstances. Less familiar—but just as important to chemistry and ultimately to us—is the group of chemicals known as bases. Both acids and bases are important enough that we devote an entire chapter to them—their properties and their reactions.

8.1: Prelude to Acids and Bases

8.2: Brønsted-Lowry Definition of Acids and Bases

8.3: Water - Both an Acid and a Base

8.4: The Strengths of Acids and Bases

8.5: Buffers

8.6: Acids and Bases (Exercises)

8.7: Acids and Bases (Summary)

Template:HideTOC

8: Acids and Bases is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.1: Prelude to Acids and Bases

One of the most concentrated acids in the body is stomach acid, which can be approximated as a 0.05 M hydrochloric acid solution. Special cells in the stomach wall secrete this acid, along with special enzymes, as part of the digestion process. In a laboratory, a 0.05 M solution of hydrochloric acid would dissolve some metals. How does the stomach survive the presence of such a reactive acid?

Actually, the stomach has several mechanisms for withstanding this chemical onslaught. First, the lining of the stomach is coated with a thin layer of mucus that contains some bicarbonate ions (HCO_3^-). These react with the hydrochloric acid to produce water, carbon dioxide, and harmless chloride ions. If any acid penetrates through the mucus, it can attack the surface layer of stomach cells, called the *gastric epithelium*. Cells in the gastric epithelium are being constantly shed, so damaged cells are quickly removed and replaced with healthy cells.



Figure **8.1.1**: Stomach with acidic digestive juices. (CC BY-SA 4.0; BruceBlaus).

However, if the gastric epithelium is destroyed faster than it can be replaced, the acid may reach the wall of the stomach, resulting in ulcers. If an ulcer grows large enough, it can expose blood vessels in the stomach wall, causing bleeding. In extreme situations, the loss of blood through a severe ulcer can threaten a person's health.

Ulcers can also result from the presence of a certain bacterium—*Helicobacter pylori*—in the stomach. The mechanism for this ulcer formation is not the same as that for ulcers caused by stomach acid and is not completely understood. However, there are two main treatments for ulcers: (1) *antacids* to react chemically with excess hydrochloric acid in the stomach and (2) *antibiotics* to destroy the *H. pylori* bacteria in the stomach.

8.1: Prelude to Acids and Bases is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.0: Prelude to Acids and Bases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

8.2: Brønsted-Lowry Definition of Acids and Bases

Learning Objectives

1. Recognize a compound as a Brønsted-Lowry acid or a Brønsted-Lowry base.
2. Illustrate the proton transfer process that defines a Brønsted-Lowry acid-base reaction.

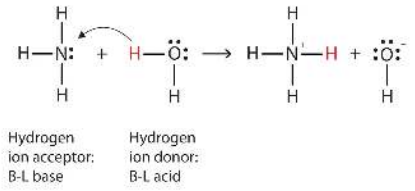
Ammonia (NH_3) increases the hydroxide ion concentration in aqueous solution by reacting with water rather than releasing hydroxide ions directly. In fact, the Arrhenius definitions of an acid and a base focus on hydrogen ions and hydroxide ions. Are there more fundamental definitions for acids and bases?

In 1923, the Danish scientist Johannes Brønsted and the English scientist Thomas Lowry independently proposed new definitions for acids and bases. Rather than considering both hydrogen and hydroxide ions, they focused on the hydrogen ion only. A Brønsted-Lowry acid is a compound that supplies a hydrogen ion in a reaction. A Brønsted-Lowry base, conversely, is a compound that accepts a hydrogen ion in a reaction. Thus, the Brønsted-Lowry definitions of an acid and a base focus on the movement of hydrogen ions in a reaction, rather than on the production of hydrogen ions and hydroxide ions in an aqueous solution.

Let us use the reaction of ammonia in water to demonstrate the Brønsted-Lowry definitions of an acid and a base. Ammonia and water molecules are reactants, while the ammonium ion and the hydroxide ion are products:

$$NH_3(aq) + H_2O(\ell) \rightarrow NH_4^+(aq) + OH^-(aq) \tag{8.2.1}$$

What has happened in this reaction is that the original water molecule has donated a hydrogen ion to the original ammonia molecule, which in turn has accepted the hydrogen ion. We can illustrate this as follows:



Because the water molecule donates a hydrogen ion to the ammonia, it is the Brønsted-Lowry acid, while the ammonia molecule—which accepts the hydrogen ion—is the Brønsted-Lowry base. Thus, ammonia acts as a base in both the Arrhenius sense and the Brønsted-Lowry sense.

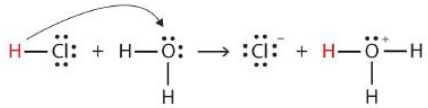
Is an Arrhenius acid like hydrochloric acid still an acid in the Brønsted-Lowry sense? Yes, but it requires us to understand what really happens when HCl is dissolved in water. Recall that the hydrogen *atom* is a single proton surrounded by a single electron. To make the hydrogen *ion*, we remove the electron, leaving a bare proton. Do we *really* have bare protons floating around in aqueous solution? No, we do not. What really happens is that the H^+ ion attaches itself to H_2O to make H_3O^+, which is called the *hydronium ion*. For most purposes, H^+ and H_3O^+ represent the same species, but writing H_3O^+ instead of H^+ shows that we understand that there are no bare protons floating around in solution. Rather, these protons are actually attached to solvent molecules.

> A proton in aqueous solution may be surrounded by more than one water molecule, leading to formulas like $H_5O_2^+$ or $H_9O_4^+$ rather than H_3O^+. It is simpler, however, to use H_3O^+.

With this in mind, how do we define HCl as an acid in the Brønsted-Lowry sense? Consider what happens when HCl is dissolved in H_2O:

$$HCl(g) + H_2O(\ell) \rightarrow H_3O^+(aq) + Cl^-(aq) \tag{8.2.2}$$

We can depict this process using Lewis electron dot diagrams:



Now we see that a hydrogen ion is transferred from the HCl molecule to the H_2O molecule to make chloride ions and hydronium ions. As the hydrogen ion donor, **HCl acts as a Brønsted-Lowry acid**; as a hydrogen ion acceptor, **H_2O is a Brønsted-Lowry base**. So HCl is an acid not just in the Arrhenius sense but also in the Brønsted-Lowry sense. Moreover, by the Brønsted-Lowry definitions, H_2O is a base in the formation of aqueous HCl. So the Brønsted-Lowry definitions of an acid and a base classify the dissolving of HCl in water as a reaction between an acid and a base—although the Arrhenius definition would not have labeled H_2O a base in this circumstance.

All Arrhenius acids and bases are Brønsted-Lowry acids and bases as well. However, not all Brønsted-Lowry acids and bases are Arrhenius acids and bases.

✔ Example 8.2.1

Aniline ($C_6H_5NH_2$) is slightly soluble in water. It has a nitrogen atom that can accept a hydrogen ion from a water molecule just like the nitrogen atom in ammonia does. Write the chemical equation for this reaction and identify the Brønsted-Lowry acid and base.

Solution

$C_6H_5NH_2$ and H_2O are the reactants. When $C_6H_5NH_2$ accepts a proton from H_2O, it gains an extra H and a positive charge and leaves an OH^- ion behind. The reaction is as follows:

$$C_6H_5NH_2(aq) + H_2O(\ell) \rightarrow C_6H_5NH_3^+(aq) + OH^-(aq)$$

Because $C_6H_5NH_2$ accepts a proton, it is the Brønsted-Lowry base. The H_2O molecule, because it donates a proton, is the Brønsted-Lowry acid.

? Exercise 8.2.1

Caffeine ($C_8H_{10}N_4O_2$) is a stimulant found in coffees and teas. When dissolved in water, it can accept a proton from a water molecule. Write the chemical equation for this process and identify the Brønsted-Lowry acid and base.

Answer

$C_8H_{10}N_4O_2(aq) + H_2O(\ell) \rightarrow C_8H_{11}N_4O_2^+(aq) + OH^-(aq)$

B-L base B-L acid

The Brønsted-Lowry definitions of an acid and a base can be applied to chemical reactions that occur in solvents other than water. The following example illustrates.

✔ Example 8.2.2

Sodium amide ($NaNH_2$) dissolves in methanol (CH_3OH) and separates into sodium ions and amide ions (NH_2^-). The amide ions react with methanol to make ammonia and the methoxide ion (CH_3O^-). Write a balanced chemical equation for this process and identify the Brønsted-Lowry acid and base.

Solution

The equation for the reaction is between NH_2^- and CH_3OH to make NH_3 and CH_3O^- is as follows:

$$NH_2^-(solv) + CH_3OH(\ell) \rightarrow NH_3(solv) + CH_3O^-(solv)$$

The label *(solv)* indicates that the species are dissolved in some solvent, in contrast to *(aq)*, which specifies an aqueous (H_2O) solution. In this reaction, we see that the NH_2^- ion accepts a proton from a CH_3OH molecule to make an NH_3 molecule. Thus, as the proton acceptor, NH_2^- is the Brønsted-Lowry base. As the proton donor, CH_3OH is the Brønsted-Lowry acid.

? Exercise 8.2.2

Pyridinium chloride (C_5H_5NHCl) dissolves in ethanol (C_2H_5OH) and separates into pyridinium ions ($C_5H_5NH^+$) and chloride ions. The pyridinium ion can transfer a hydrogen ion to a solvent molecule. Write a balanced chemical equation for this process and identify the Brønsted-Lowry acid and base.

Answer

$C_5H_5NH^+(solv) + C_2H_5OH(\ell) \rightarrow C_5H_5N(solv) + C_2H_5OH_2^+(solv)$

B-L acid B-L base

Application in Everyday Life

Many people like to put lemon juice or vinegar, both of which are acids, on cooked fish (Figure **8.2.1**). It turns out that fish have volatile amines (bases) in their systems, which are neutralized by the acids to yield involatile ammonium salts. This reduces the odor of the fish, and also adds a "sour" taste that we seem to enjoy.



$CH_3COOH + NH_2CH_2CH_2CH_2CH_2NH_2 \longrightarrow CH_3COO^- + NH_3^+CH_2CH_2CH_2CH_2NH_2$

Acetic acid + Putrescine → Acetate ion + Putrescinium ion

Figure **8.2.1**: A neutralization reaction takes place between citric acid or acetic acid (proton donors) in lemons or vinegar and putrescine (proton acceptor) in the flesh of fish. (CC BY 4.0; OpenStax)

Pickling is a method used to preserve vegetables using a naturally produced acidic environment. The vegetable, such as a cucumber, is placed in a sealed jar submerged in a brine solution. The brine solution favors the growth of beneficial bacteria and suppresses the growth of harmful bacteria. The beneficial bacteria feed on starches in the cucumber and produce lactic acid as a waste product in a process called fermentation. The lactic acid eventually increases the acidity of the brine to a level that kills any harmful bacteria, which require a basic environment. Without the harmful bacteria consuming the cucumbers they are able to last much longer than if they were unprotected. A byproduct of the pickling process changes the flavor of the vegetables with the acid making them taste sour.

To Your Health: Brønsted-Lowry Acid-Base Reactions in Pharmaceuticals

There are many interesting applications of Brønsted-Lowry acid-base reactions in the pharmaceutical industry. For example, drugs often need to be water soluble for maximum effectiveness. However, many complex organic compounds are not soluble or are only slightly soluble in water. Fortunately, those drugs that contain proton-accepting nitrogen atoms (and there are a lot of them) can be

reacted with dilute hydrochloric acid [HCl(aq)]. The nitrogen atoms—acting as Brønsted-Lowry bases—accept the hydrogen ions from the acid to make an ion, which is usually much more soluble in water. The modified drug molecules can then be isolated as chloride salts:

\[\mathrm{RN(sl\: aq) + H^+(aq) \rightarrow RNH^+(aq) \xrightarrow{Cl^-(aq)} RNHCl(s)} \label{Eq3} \]

where RN represents some organic compound containing nitrogen. The label *(sl aq)* means "slightly aqueous," indicating that the compound RN is only slightly soluble. Drugs that are modified in this way are called *hydrochloride salts*. Examples include the powerful painkiller codeine, which is commonly administered as codeine hydrochloride. Acids other than hydrochloric acid are also used. Hydrobromic acid, for example, gives *hydrobromide salts*. Dextromethorphan, an ingredient in many cough medicines, is dispensed as dextromethorphan hydrobromide. The accompanying figure shows another medication (lidocaine) as a hydrochloride salt.



Figure used with permission from Wikipedia.

Conjugate Acid-Base Pairs

According to the Bronsted-Lowry theory of acids and bases, an acid is a proton donor and a base is a proton acceptor. Once an acid has given up a proton, the part that remains is called the acid's **conjugate base**. This species is a base because it can accept a proton (to re-form the acid). The conjugate base of HF (first example below) is fluoride ion, F^-.

$$\text{Acid} = H^+ + \text{Conjugate base of Acid}^-$$

$$HF \rightleftharpoons H^+ + F^-$$

$$H_2O \rightleftharpoons H^+ + OH^-$$

$$NH_4^+ \rightleftharpoons H^+ + NH_3$$

Similarly, the part of the base that remains after a base accepts a proton is called the base's **conjugate acid**. This species is an acid because it can give up a proton (and thus re-form the base). The conjugate acid of fluoride ion, F^- (first example below) is HF.

$$H^+ + \text{Base} = \text{Conjugate acid of Base}^+$$

$$H^+ + F^- \rightleftharpoons HF$$

$$H^+ + OH^- \rightleftharpoons H_2O$$

$$H^+ + H_2O \rightleftharpoons H_3O^+$$

$$H^+ + NH_3 \rightleftharpoons NH_4^+$$

To summarize, the conjugate base of HF is fluoride ion, F^-, and the conjugate acid of fluoride ion, F^-, is HF. The HF/F- pair is referred to as a conjugate acid-base pair. The difference in the formulas of a conjugate acid-base pair (example: HF and F^-) is H^+. The table below lists conjugate acid-base pairs for your reference so that you can figure out the strategy of identifying them. For any given acid or base, you should be able to give its conjugate base or conjugate acid. The formula of an acid's conjugate base is generated by **removing a proton** (H^+) from the acid formula. The formula of the base's conjugate acid is formed by **adding a proton** (H^+) to the formula of the base.

Table **8.2.1**. Conjugate acid-base pairs.

Conjugate Acid	Conjugate Base
H_3O^+	H_2O
H_2O	OH^-
H_2SO_4	HSO_4^-
HSO_4^-	SO_4^{2-}
NH_4^+	NH_3
NH_3	NH_2^-
CH_3COOH	CH_3COO^-
$CH_3NH_3^+$	CH_3NH_2

Example 8.2.3 : Conjugate Pairs

Write the formula of the conjugate base of (a) HCl and (b) HCO_3^-.

Write the formula of the conjugate acid of (c) CH_3NH_2 and (d) OH^-.

Solution:

A conjugate base is formed by removing a proton (H^+). A conjugate acid is formed by accepting a proton (H^+).

a. After HCl donates a proton, a Cl^- ion is produced, and so Cl^- is the conjugate base.
b. After hydrogen carbonate ion, HCO_3^-, donates a proton, its conjugate base, CO_3^{2-} is produced.
c. After accepting a proton (H^+), CH_3NH_2 is converted to $CH_3NH_3^+$, its conjugate acid.
d. After accepting a proton (H^+), OH^- is converted to H_2O, its conjugate acid.

8.2.3Exercise : Conjugate Pairs

Write the formula of the conjugate base of (a) HNO_2 and (b) H_2CO_3.

Write the formula of the conjugate acid of (c) C_6H5NH_2 and (d) HCO_3^-.

Answer

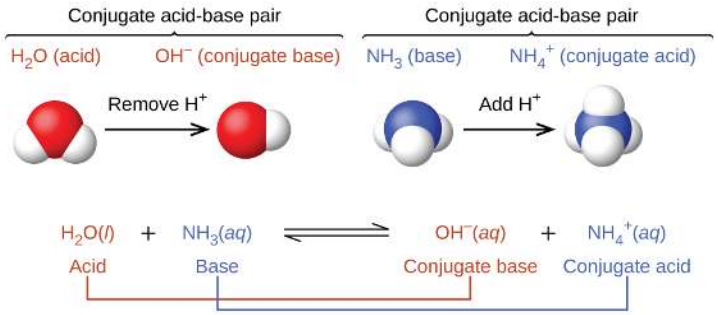
a. NO_2^- is the conjugate base of HNO_2.

b. HCO_3^- is the conjugate base of H_2CO_3

c. $C_6H5NH_3^+$ is the conjugate acid of C_6H5NH_2.

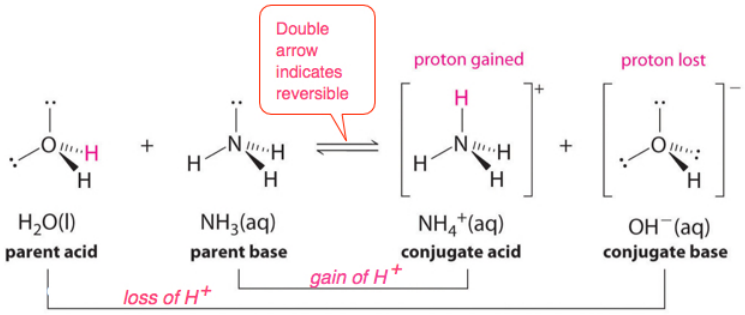
d. H_2CO_3 is the conjugate acid of HCO_3^-

In the reaction of ammonia with water to give ammonium ions and hydroxide ions, ammonia acts as a base by accepting a proton from a water molecule, which in this case means that water is acting as an acid. In the reverse reaction, an ammonium ion acts as an acid by donating a proton to a hydroxide ion, and the hydroxide ion acts as a base. The **conjugate acid–base pairs** for this reaction are $\boldsymbol{NH_4^+/NH_3}$ and $\boldsymbol{H_2O/OH^-}$. This means that the conjugate acid of the base NH_3 is NH_4^+ while the conjugate base of the acid NH_4^+ is NH_3. Similarly, the conjugate base of the acid H_2O is OH^-, and the conjugate acid of the base OH^- is H_2O.



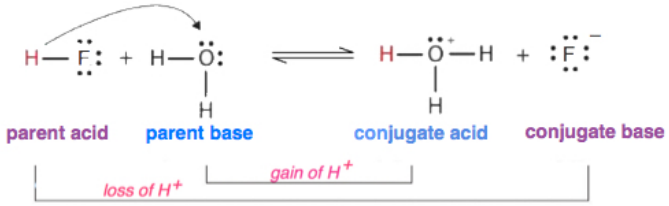
This figure has three parts in two rows. In the first row, two diagrams of acid-base pairs are shown. On the left, a space filling model of H subscript 2 O is shown with a red O atom at the center and two smaller white H atoms attached in a bent shape. Above this model is the label “H subscript 2 O (acid)” in purple. An arrow points right, which is labeled “Remove H superscript plus.” To the right is another space filling model with a single red O atom to which a single smaller white H atom is attached. The label in purple above this model reads, “O H superscript negative (conjugate base).” Above both of these red and white models is an upward pointing bracket that is labeled “Conjugate acid-base pair.” To the right is a space filling model with a central blue N atom to which three smaller white H atoms are attached in a triangular pyramid arrangement. A label in green above reads “N H subscript 3 (base).” An arrow labeled “Add H superscript plus” points right. To the right of the arrow is another space filling model with a blue central N atom and four smaller white H atoms in a tetrahedral arrangement. The green label above reads “N H subscript 3 superscript plus (conjugate acid).” Above both of these blue and white models is an upward pointing bracket that is labeled “Conjugate acid-base pair.” The second row of the figure shows the chemical reaction, H subscript 2 O (l) is shown in purple, and is labeled below in purple as “acid,” plus N H subscript 3 (a q) in green, labeled below in green as “base,” followed by a double sided arrow arrow and O H superscript negative (a q) in purple, labeled in purple as “conjugate base,” plus N H subscript 4 superscript plus (a q)” in green, which is labeled in green as “conjugate acid.” The acid on the left side of the equation is connected to the conjugate base on the right with a purple line. Similarly, the base on the left is connected to the conjugate acid on the right side.

In the **forward** reaction, the parent acid is H_2O and and the parent base is NH_3 (shown in the illustration below). The acid H_2O loses a proton (H^+) to form its conjugate base OH^-. The base NH_3 gains a proton, to produce its conjugate acid NH_4^+. In the **reverse** reaction, the acid NH_4^+ loses a proton (H^+) to form its conjugate base NH_3. The base OH^- gains a proton, to produce its conjugate acid H_2O.



When hydrogen fluoride (HF) dissolves in water and ionizes, protons are transferred from hydrogen fluoride (parent acid) molecules to water (parent base) molecules, yielding hydronium ions (conjugate acid of water) and fluoride ions (conjugate base of HF):

$$HF + H_2O \rightleftharpoons H_3O^+ + F^-$$



https://chem.libretexts.org/@go/page/255620

Example 8.2.3

Identify the conjugate acid-base pairs in this equilibrium.

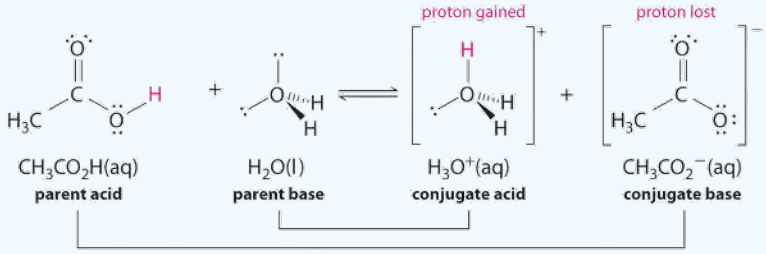
$$CH_3CO_2H + H_2O \rightleftharpoons H_3O^+ + CH_3CO_2^-$$

Solution

Similarly, in the reaction of acetic acid with water, acetic acid **donates** a proton to water, which acts as the base. In the reverse reaction, H_3O^+ is the acid that donates a proton to the acetate ion, which acts as the base.

Once again, we have two conjugate acid–base pairs:

- the parent acid and its conjugate base ($CH_3CO_2H/CH_3CO_2^-$) and
- the parent base and its conjugate acid (H_3O^+/H_2O).



Example 8.2.4

Identify the conjugate acid-base pairs in this equilibrium.

$$(CH_3)_3N + H_2O \rightleftharpoons (CH_3)_3NH^+ + OH^-$$

Solution

One pair is H_2O and OH^-, where H_2O has one more H^+ and is the conjugate acid, while OH^- has one less H^+ and is the conjugate base.

The other pair consists of $(CH_3)_3N$ and $(CH_3)_3NH^+$, where $(CH_3)_3NH^+$ is the conjugate acid (it has an additional proton) and $(CH_3)_3N$ is the conjugate base.

Exercise 8.2.4

Identify the conjugate acid-base pairs in this equilibrium.

$$NH_2^- + H_2O \rightleftharpoons NH_3 + OH^-$$

Answer:

H_2O (acid) and OH^- (base); NH_2^- (base) and NH_3 (acid)

The use of conjugate acid-base pairs allows us to make a very simple statement about relative strengths of acids and bases. *The stronger an acid, the weaker its conjugate base*, and, conversely, *the stronger a base, the weaker its conjugate acid.*

Key Takeaways

- A Brønsted-Lowry acid is a proton donor, and a Brønsted-Lowry base is a proton acceptor.
- Brønsted-Lowry acid-base reactions are essentially proton transfer reactions.

8.2: Brønsted-Lowry Definition of Acids and Bases is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.2: Brønsted-Lowry Definition of Acids and Bases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

8.3: Water - Both an Acid and a Base

Learning Objectives

- To write chemical equations for water acting as an acid and as a base.

Water (H_2O) is an interesting compound in many respects. Here, we will consider its ability to behave as an acid or a base.

In some circumstances, a water molecule will accept a proton and thus act as a **Brønsted-Lowry base**. We saw an example in the dissolving of HCl in H_2O:

$$\mathrm{HCl} + \mathrm{H_2O}(\ell) \rightarrow \mathrm{H_3O^+}(aq) + \mathrm{Cl^-}(aq) \tag{8.3.1}$$

In other circumstances, a water molecule can donate a proton and thus act as a **Brønsted-Lowry acid**. For example, in the presence of the amide ion (see Example 4 in Section 10.2), a water molecule donates a proton, making ammonia as a product:

$$H_2O(\ell) + NH_2^-(aq) \rightarrow OH^-(aq) + NH_3(aq) \tag{8.3.2}$$

In this case, NH_2^- is a Brønsted-Lowry base (the proton acceptor).

So, depending on the circumstances, H_2O can act as either a Brønsted-Lowry acid or a Brønsted-Lowry base. Water is not the only substance that can react as an acid in some cases or a base in others, but it is certainly the most common example—and the most important one. A substance that can either donate or accept a proton, depending on the circumstances, is called an **amphiprotic** compound.

A water molecule can act as an acid or a base even in a sample of pure water. About 6 in every 100 million (6 in 10^8) water molecules undergo the following reaction:

$$H_2O(\ell) + H_2O(\ell) \rightarrow H_3O^+(aq) + OH^-(aq) \tag{8.3.3}$$

This process is called the **autoionization of water** (Figure 8.3.1) and occurs in every sample of water, whether it is pure or part of a solution. Autoionization occurs to some extent in any amphiprotic liquid. (For comparison, liquid ammonia undergoes autoionization as well, but only about 1 molecule in a million billion (1 in 10^{15}) reacts with another ammonia molecule.)

Note

It is rare to truly have pure water. Water exposed to air will usually be slightly acidic because dissolved carbon dioxide gas, or carbonic acid, decreases the pH slightly below 7. Alternatively, dissolved minerals, like calcium carbonate (limestone), can make water slightly basic.

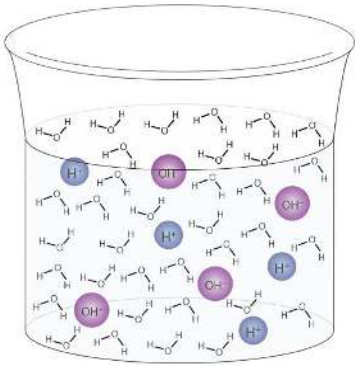


Figure 8.3.1 Autoionization. A small fraction of water molecules—approximately 6 in 100 million—ionize spontaneously into hydronium ions and hydroxide ions. This picture necessarily overrepresents the amount of autoionization that really occurs in pure water.

Example 8.3.1

Identify water as either a Brønsted-Lowry acid or a Brønsted-Lowry base.

1. $H_2O(\ell) + NO_2^-(aq) \rightarrow HNO_2(aq) + OH^-(aq)$
2. $HC_2H_3O_2(aq) + H_2O(\ell) \rightarrow H_3O^+(aq) + C_2H_3O_2^-(aq)$

Solution

1. In this reaction, the water molecule donates a proton to the NO_2^- ion, making $OH^-(aq)$. As the proton donor, H_2O acts as a Brønsted-Lowry acid.
2. In this reaction, the water molecule accepts a proton from $HC_2H_3O_2$, becoming $H_3O^+(aq)$. As the proton acceptor, H_2O is a Brønsted-Lowry base.

Exercise 8.3.2

Identify water as either a Brønsted-Lowry acid or a Brønsted-Lowry base.

1. $HCOOH(aq) + H_2O(\ell) \rightarrow H_3O^+(aq) + HCOO^-(aq)$
2. $H_2O(\ell) + PO_4^{3-}(aq) \rightarrow OH^-(aq) + HPO_4^{2-}(aq)$

Answer

1. H_2O acts as the proton acceptor (Brønsted-Lowry base)
2. H_2O acts as the proton donor (Brønsted-Lowry acid)

Key Takeaway

- Water molecules can act as both an acid and a base, depending on the conditions.

8.3: Water - Both an Acid and a Base is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.3: Water - Both an Acid and a Base** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

8.4: The Strengths of Acids and Bases

Learning Objectives

- Describe the difference between strong and weak acids and bases.
- Describe how a chemical reaction reaches chemical equilibrium.
- Define the pH scale and use it to describe acids and bases.

Acids and bases do not all demonstrate the same degree of chemical activity in solution. Different acids and bases have different strengths.

Strong and Weak Acids

Let us consider the strengths of acids first. A small number of acids ionize completely in aqueous solution. For example, when HCl dissolves in water, every molecule of HCl separates into a hydronium ion and a chloride ion:

$$\mathrm{HCl(g) + H_2O(l) \xrightarrow{\sim 100\,\%} H_3O^+(aq) + Cl^-(aq)} \tag{8.4.1}$$

HCl(aq) is one example of a **strong acid**, which is a compound that is essentially **100% ionized** in aqueous solution. There are very few strong acids. The important ones are listed in Table 8.4.1.

Table 8.4.1: Strong Acids and Bases (All in Aqueous Solution)

Acids	Bases
HCl	LiOH
HBr	NaOH
HI	KOH
HNO_3	$Mg(OH)_2$
H_2SO_4	$Ca(OH)_2$
$HClO_4$	

By analogy, a strong base is a compound that is essentially 100% ionized in aqueous solution. As with acids, there are only a few strong bases, which are also listed in Table 8.4.1.

If an acid is **not listed** in Table 8.4.1, it is likely a **weak acid**, which is a compound that is not 100% ionized in aqueous solution. Similarly, a weak base is a compound that is not 100% ionized in aqueous solution. For example, acetic acid ($\mathrm{HC_2H_3O_2}$) is a weak acid. The ionization reaction for acetic acid is as follows:

$$\mathrm{HC_2H_3O_2(aq) + H_2O(\ell) \rightarrow H_3O^+(aq) + C_2H_3O_2^-(aq)} \tag{8.4.2}$$

Depending on the concentration of $HC_2H_3O_2$, the ionization reaction may occur only for 1%–5% of the acetic acid molecules.

Looking Closer: Household Acids and Bases

Many household products are acids or bases. For example, the owner of a swimming pool may use muriatic acid to clean the pool. Muriatic acid is another name for hydrochloric acid [$\mathrm{HCl(aq)}$]. Vinegar has already been mentioned as a dilute solution of acetic acid [$\mathrm{HC_2H_3O_2(aq)}$]. In a medicine chest, one may find a bottle of vitamin C tablets; the chemical name of vitamin C is ascorbic acid ($\mathrm{HC_6H_7O_6}$).

One of the more familiar household bases is ammonia (NH_3), which is found in numerous cleaning products. As we mentioned previously, ammonia is a base because it increases the hydroxide ion concentration by reacting with water:

$$\mathrm{NH_3(aq) + H_2O(\ell) \rightarrow NH_4^+(aq) + OH^-(aq)} \tag{8.4.3}$$

Many soaps are also slightly basic because they contain compounds that act as Brønsted-Lowry bases, accepting protons from water and forming excess hydroxide ions. This is one reason that soap solutions are slippery.



Figure 8.4.1: (left) Bottles of alkaline drain cleaners containing sodium hydroxide can dissolve greases and hair. (right) Chemical burns caused by sodium hydroxide solution photographed 44 hours after exposure. Figures used with permission from Wikipedia.

Perhaps the most dangerous household chemical is the lye-based drain cleaner. Lye is a common name for sodium hydroxide, although it is also used as a synonym for potassium hydroxide. Lye is an extremely caustic chemical that can react with grease, hair, food particles, and other substances that may build up and form a clog in a pipe. Unfortunately, lye can also attack tissues and other substances in our bodies. Thus, when we use lye-based drain cleaners, we must be very careful not to touch any of the solid drain cleaner or spill the water it was poured into. Safer, nonlye drain cleaners use peroxide compounds to react on the materials in the clog and clear the drain.

Chemical Equilibrium in Weak Acids and Bases

The behavior of **weak acids and bases** illustrates a key concept in chemistry. Does the chemical reaction describing the ionization of a weak acid or base just stop when the acid or base is done ionizing? Actually, no. Rather, the reverse process—the reformation of the molecular form of the acid or base—occurs, ultimately at the same rate as the ionization process. For example, the ionization of the weak acid $\mathbf{HC_2H_3O_2\,(aq)}$ is as follows:

$$\mathbf{HC_2H_3O_2\,(aq) + H_2O(\ell) \rightarrow H_3O^+\,(aq) + C_2H_3O_2^-\,(aq)} \tag{8.4.4}$$

The reverse process also begins to occur:

$$\mathbf{H_3O^+\,(aq) + C_2H_3O_2^-\,(aq) \rightarrow HC_2H_3O_2\,(aq) + H_2O(\ell)} \tag{8.4.5}$$

Eventually, there is a balance between the two opposing processes, and no additional change occurs. The chemical reaction is better represented at this point with a double arrow:

$$\mathbf{HC_2H_3O_2\,(aq) + H_2O(\ell) \rightleftharpoons H_3O^+\,(aq) + C_2H_3O_2^-\,(aq)} \tag{8.4.6}$$

The $\rightleftharpoons$ implies that **both the forward and reverse reactions** are occurring, and their effects cancel each other out. A process at this point is considered to be at **chemical equilibrium** (or equilibrium). It is important to note that the processes do not stop. They balance out each other so that there is no further net change; that is, chemical equilibrium is a *dynamic equilibrium*.

Example 8.4.1: Partial Ionization

Write the equilibrium chemical equation for the partial ionization of each weak acid or base.

a. $HNO_2(aq)$
b. $C_5H_5N(aq)$

Solution

a. $HNO_2(aq) + H_2O(\ell) \leftrightarrows NO_2^-(aq) + H_3O^+(aq)$
b. $C_5H_5N(aq) + H_2O(\ell) \leftrightarrows C_5H_5NH^+(aq) + OH^-(aq)$

? Exercise 8.4.1

Write the equilibrium chemical equation for the partial ionization of each weak acid or base.

a. $HF_{(aq)}$
b. $AgOH_{(aq)}$
c. $CH_3NH_2(aq)$

Answer

a. $HF(aq) + H_2O(\ell) \leftrightarrows F^-(aq) + H_3O^+(aq)$

b. $AgOH(aq) \leftrightarrows Ag^+(aq) + OH^-(aq)$

c. $CH_3NH_2(aq) + H_2O(\ell) \leftrightarrows CH_3NH_3^+(aq) + OH^-(aq)$

Acid Ionization Constant, K_a

The ionization for a general weak acid, **HA**, can be written as follows:

$$\mathrm{HA}\,(aq) \rightleftharpoons \mathrm{H^+}\,(aq) + \mathrm{A^-}\,(aq)$$

Because the acid is weak, an equilibrium expression can be written. An **acid ionization constant** (K_a) is the equilibrium constant for the ionization of an acid.

$$K_a = \frac{[\mathrm{H^+}]\,[\mathrm{A^-}]}{[\mathrm{HA}]}$$

The acid ionization represents the fraction of the original acid that has been ionized in solution. Therefore, the numerical value of K_a is a reflection of the strength of the acid. Weak acids with relatively higher K_a values are stronger than acids with relatively lower K_a values. Because strong acids are essentially 100% ionized, the concentration of the acid in the denominator is nearly zero and the K_a value approaches infinity. For this reason, K_a values are generally reported for weak acids only.

The table below is a listing of acid ionization constants for several acids. Note that polyprotic acids have a distinct ionization constant for each ionization step, with each successive ionization constant being smaller than the previous one.

Table 8.4.2: Acid Ionization Constants at 25°C

Name of Acid	Ionization Equation	K_a
Sulfuric acid	$\mathrm{H_2SO_4 \rightleftharpoons H^+ + HSO_4^-}$ $\mathrm{HSO_4^- \rightleftharpoons H^+ + SO_4^{2-}}$	very large 1.3×10^{-2}
Hydrofluoric acid	$\mathrm{HF \rightleftharpoons H^+ + F^-}$	7.1×10^{-4}
Nitrous acid	$\mathrm{HNO_2 \rightleftharpoons H^+ + NO_2^-}$	4.5×10^{-4}
Benzoic acid	$\mathrm{C_6H_5COOH \rightleftharpoons H^+ + C_6H_5COO^-}$	6.5×10^{-5}
Acetic acid	$\mathrm{CH_3COOH \rightleftharpoons H^+ + CH_3COO^-}$	1.8×10^{-5}
Carbonic acid	$\mathrm{H_2CO_3 \rightleftharpoons H^+ + HCO_3^-}$ $\mathrm{HCO_3^- \rightleftharpoons H^+ + CO_3^{2-}}$	4.2×10^{-7} 4.8×10^{-11}

Hydrofluoric acid $HF_{(aq)}$ reacts directly with glass (very few chemicals react with glass). Hydrofluoric acid is used in glass etching.

Strong and Weak Bases and Base Ionization Constant, K_b

As with acids, bases can either be strong or weak, depending on their extent of ionization. A **strong base** is a base, which ionizes completely in an aqueous solution. The most common strong bases are soluble metal hydroxide compounds such as potassium

hydroxide. Some metal hydroxides are not as strong simply because they are not as soluble. Calcium hydroxide is only slightly soluble in water, but the portion that does dissolve also dissociates into ions.

A **weak base** is a base that ionizes only slightly in an aqueous solution. Recall that a base can be defined as a substance, which accepts a hydrogen ion from another substance. When a weak base such as ammonia is dissolved in water, it accepts an H^+ ion from water, forming the hydroxide ion and the conjugate acid of the base, the ammonium ion.

$$NH_3\,(aq) + H_2O\,(l) \rightleftharpoons NH_4^+\,(aq) + OH^-\,(aq)$$

The equilibrium greatly favors the reactants and the extent of ionization of the ammonia molecule is very small.

An equilibrium expression can be written for the reactions of weak bases with water. Because the concentration of water is extremely large and virtually constant, the water is not included in the expression. A **base ionization constant** (K_b) is the equilibrium constant for the ionization of a base. For ammonia the expression is:

$$K_b = \frac{[NH_4^+]\,[OH^-]}{[NH_3]}$$

The numerical value of K_b is a reflection of the strength of the base. Weak bases with relatively higher K_b values are stronger than bases with relatively lower K_b values. Table 8.4.3 is a listing of base ionization constants for several weak bases.

Table 8.4.3: Base Ionization Constants at 25°C

Name of Base	Ionization Equation	K_b
Methylamine	$CH_3NH_2 + H_2O \rightleftharpoons CH_3NH_3^+ + OH^-$	5.6×10^{-4}
Ammonia	$NH_3 + H_2O \rightleftharpoons NH_4^+ + OH^-$	1.8×10^{-5}
Pyridine	$C_5H_5N + H_2O \rightleftharpoons C_5H_5NH^+ + OH^-$	1.7×10^{-9}
Acetate ion	$CH_3COO^- + H_2O \rightleftharpoons CH_3COOH + OH^-$	5.6×10^{-10}
Fluoride ion	$F^- + H_2O \rightleftharpoons HF + OH^-$	1.4×10^{-11}
Urea	$H_2NCONH_2 + H_2O \rightleftharpoons H_2NCONH_3^+ + OH^-$	1.5×10^{-14}

The Ion-Product of Water

As we have already seen, H_2O can act as an acid or a base. Within any given sample of water, some H_2O molecules are acting as acids, and other H_2O molecules are acting as bases. The chemical equation is as follows:

$$\underbrace{H_2O}_{\text{acid}} + \underbrace{H_2O}_{\text{base}} \rightleftharpoons H_3O^+ + OH^- \tag{8.4.7}$$

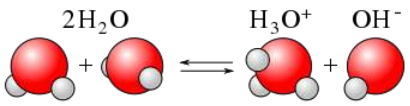


Figure 8.4.2: Autoionization of water, giving hydroxide and hydronium ions.

Similar to a weak acid, the autoionization of water is an equilibrium process, and is more properly written as follows:

$$H_2O(\ell) + H_2O(\ell) \rightleftharpoons H_3O^+\,(aq) + OH^-\,(aq) \tag{8.4.8}$$

We often use the simplified form of the reaction:

$$H_2O\,(l) \rightleftharpoons H^+\,(aq) + OH^-\,(aq)$$

The equilibrium constant for the autoionization of water is referred to as the ion-product for water and is given the symbol Kw.

$$K_w = [H^+][OH^-]$$

The **ion-product of water (*Kw*)** is the mathematical product of the concentration of hydrogen ions and hydroxide ions. Note that H_2O is not included in the ion-product expression because it is a pure liquid. The value of Kw is very small, in accordance with a

reaction that favors the reactants. At 25°C, the experimentally determined value of K_w in pure water is 1.0×10^{-14}.

$$K_w = [H^+][OH^-] = 1.0 \times 10^{-14}$$

In a sample of pure water, the concentrations of hydrogen and hydroxide ions are equal to one another. Pure water or any other aqueous solution in which this ratio holds is said to be **neutral.** To find the molarity of each ion, the square root of Kw is taken.

$$[H^+] = [OH^-] = 1.0\times10^{-7}$$

The product of these two concentrations is 1.0×10^{-14}

$$[H^+]\times[OH^-] = (1.0 \times 10^{-7})(1.0 \times 10^{-7}) = 1.0 \times 10^{-14}$$

- For acids, the concentration of H^+ or $[H^+]$) is greater than 1.0×10^{-7} M
- For bases, the concentration of OH^- or $[OH^-]$ is greater than 1.0×10^{-7} M.

Aqueous HCl is an example of acidic solution. Hydrogen chloride (HCl) ionizes to produce H^+ and Cl^- ions upon dissolving in water. This **increases the concentration of H^+ ions** in the solution. According to **Le Chatelier's principle**, the equilibrium represented by

$$H_2O\,(l) \rightleftharpoons H^+\,(aq) + OH^-\,(aq)$$

$$HCl\,(g) \longrightarrow H^+\,(aq) + Cl^-\,(aq)$$

is **forced to the left**, towards the reactant. As a result, the **concentration of the hydroxide ion decreases**.

Now, consider KOH (aq), a **basic solution**. Solid potassium hydroxide (KOH) dissociates in water to yield potassium ions and hydroxide ions.

$$KOH(s) \rightarrow K^+(aq) + OH^-(aq)$$

The **increase in concentration of the OH^- ions** will cause a decrease in the concentration of the H^+ ions.

No matter whether the aqueous solution is an acid, a base, or neutral:and the ion-product of $[H^+][OH^-]$ remains constant.

- For acidic solutions, $[H^+]$) is greater than $[OH^-]$.
- For basic solutions, [OH−] is greater than $[H^+]$.
- For neutral solutions, $[H_3O^+] = [OH-] = 1.0\times10^{-7}$M

This means that if you know $[H^+]$ for a solution, you can calculate what $[OH^-]$) has to be for the product to equal 1.0×10^{-14}, or if you know $[OH^-]$), you can calculate $[H^+]$. This also implies that as one concentration goes up, the other must go down to compensate so that their product always equals the value of K_w.

$$K_w = [H_3O^+][OH^-] = 1.0 \times 10^{-14} \tag{8.4.9}$$

Example 8.4.2

Hydrochloric acid (HCl) is a strong acid, meaning it is 100% ionized in solution. What is the $[H^+]$ and the $[OH^-]$ in a solution of 2.0×10^{-3} M HCl?

Solution

Step 1: List the known values and plan the problem.

Known

- $[HCl] = 2.0\times10^{-3}$ M
- $Kw = 1.0\times10^{-14}$

Unknown

- $[H^+]$=?M
- $[OH^-]$=?M

Because HCl is 100% ionized, the concentration of H+ ions in solution will be equal to the original concentration of HCl. Each HCl molecule that was originally present ionizes into one H+ ion and one Cl− ion. The concentration of OH− can then be determined from the [H+] and Kw.

Step 2: Solve.

$$[H^+]=2.0\times10^{-3}\,M$$

$$Kw = [H^+][OH^-] = 1.0\times10^{-14}$$

$$[OH^-] = Kw/[H^+] = 1.0\times10^{-14}/2.0\times10^{-3} = 5.0\times10^{-12}\,M$$

Step 3: Think about your result.

The $[H^+]$ is much higher than the $[OH^-]$ because the solution is acidic. As with other equilibrium constants, the unit for Kw is customarily omitted.

8.4.2 Exercise

Sodium hydroxide (NaOH) is a strong base. What is the $[H^+]$ and the $[OH^-]$ in a 0.001 M NaOH solution at 25 °C?

Answer

$[OH^-]$ = 0.001M or 1 x 10^{-3}M; $[H^+]=1\times10^{-11}$M.

The pH Scale

One qualitative measure of the strength of an acid or a base solution is the pH scale, which is based on the concentration of the hydronium (or hydrogen) ion in aqueous solution.

$$pH = -\log[H^+]$$

or

$$pH = -\log[H_3O^+]$$

Figure 8.4.3 illustrates this relationship, along with some examples of various solutions. Because hydrogen ion concentrations are generally less than one (for example $1.3 \times 10^{-3}\ M$), the log of the number will be a negative number. To make pH even easier to work with, pH is defined as the ***negative* log of** $[H_3O^+]$, which will give a positive value for pH.

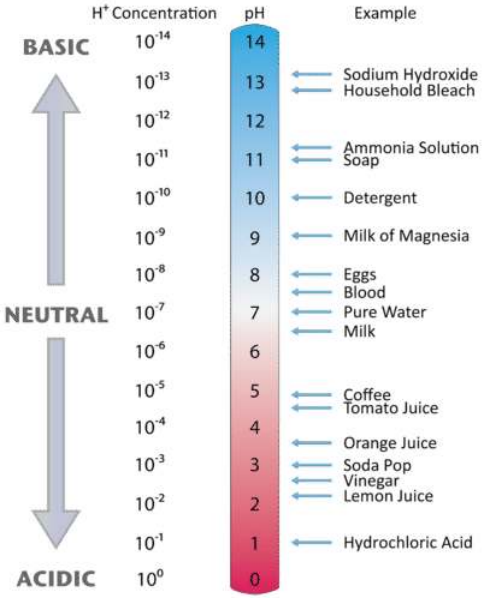


Figure 8.4.3: The relationship between $[H^+]$ and pH *values for several common materials.*

https://chem.libretexts.org/@go/page/255622

A neutral (neither acidic nor basic) solution has a pH of 7. A pH below 7 means that a solution is acidic, with lower values of pH corresponding to increasingly acidic solutions. A pH greater than 7 indicates a basic solution, with higher values of pH corresponding to increasingly basic solutions. Thus, given the pH of several solutions, you can state which ones are acidic, which ones are basic, and which are more acidic or basic than others. These are summarized in Table \(\PageIndex{4}.

Table **8.4.4**: Acidic, Basic and Neutral pH Values

Acidic, Basic and Neutral pH Values

Classification	Relative Ion Concentrations	pH at 25 °C
acidic	$[H^+] > [OH^-]$	pH < 7
neutral	$[H^+] = [OH^-]$	pH = 7
basic	$[H^+] < [OH^-]$	pH > 7

Example 8.4.3

Find the pH, given the $[H^+]$ of the following:

a. 1×10^{-3} M
b. 2.5×10^{-11} M
c. 4.7×10^{-9} M

Solution

pH = - log $[H_3O^+]$

Substitute the known quantity into the equation and solve. Use a scientific calculator for b and c.

a. pH = - log $[\underline{1} \times 10^{-3}]$ = 3.$\underline{0}$ (1 decimal place since 1 has 1 significant figure)
b. pH = - log $[\underline{2.5} \times 10^{-11}]$ = 10.$\underline{60}$ (2 decimal places since 2.5 has 2 significant figures)
c. pH = - log $[\underline{4.7} \times 10^{-9}]$ = 8.$\underline{33}$ (2 decimal places since 4.7 has 2 significant figures)

Note on significant figures:

Because the number(s) before the decimal point in the pH value relate to the power on 10, the number of digits *after* the decimal point (underlined) is what determines the number of significant figures in the final answer.

8.4.3 Exercise

Find the pH, given $[H^+]$ of the following:

a. 5.8×10^{-4} M
b. 1.0×10^{-7} M

Answer

a. 3.24

b. 7.00

Table **8.4.5** lists the pH of several common solutions. The most acidic among the listed solutions is battery acid with the lowest pH value (0.3). The most basic is 1M NaOH solution with the highest pH value of 14.0. Notice that some biological fluids (stomach acid and urine) are nowhere near neutral. You may also notice that many food products are slightly acidic. They are acidic because they contain solutions of weak acids. If the acid components of these foods were strong acids, the food would likely be inedible.

Table **8.4.5**: The pH Values of Some Common Solutions

Solution	pH
battery acid	0.3
stomach acid	1–2
lemon or lime juice	2.1
vinegar	2.8–3.0
Coca-Cola	3
wine	2.8–3.8
beer	4–5
coffee	5
milk	6
urine	6
pure H_2O	7
(human) blood	7.3–7.5
sea water	8
antacid (milk of magnesia)	10.5
NH_3 (1 M)	11.6
bleach	12.6
NaOH (1 M)	14.0

Example 8.4.4

Label each solution as acidic, basic, or neutral based only on the stated ***pH***.

a. milk of magnesia, pH = 10.5
b. pure water, pH = 7
c. wine, pH = 3.0

Solution

a. With a pH greater than 7, milk of magnesia is basic. (Milk of magnesia is largely $Mg(OH)_2$.)
b. Pure water, with a pH of 7, is neutral.
c. With a pH of less than 7, wine is acidic.

8.4.4 Exercise

Identify each substance as acidic, basic, or neutral based only on the stated ***pH***.

a. human blood with ***pH*** = 7.4
b. household ammonia with ***pH*** = 11.0
c. cherries with ***pH*** = 3.6

Answer

a. slightly basic

b. basic

c. acidic

Acid Rain

Normal rainwater has a pH between 5 and 6 due to the presence of dissolved CO_2 which forms carbonic acid:

$$\mathrm{H_2O\,(l) + CO_2\,(g) \longrightarrow H_2CO_3\,(aq)} \tag{8.4.10}$$

$$\mathrm{H_2CO_3\,(aq) \rightleftharpoons H^+\,(aq) + HCO_3^-\,(aq)} \tag{8.4.11}$$

Acid rain is rainwater that has a pH of less than 5, due to a variety of nonmetal oxides, including CO_2, SO_2, SO_3, NO, and NO_2 being dissolved in the water and reacting with it to form not only carbonic acid, but sulfuric acid and nitric acid. The formation and subsequent ionization of sulfuric acid are shown here:

$$\mathrm{H_2O\,(l) + SO_3\,(g) \longrightarrow H_2SO_4\,(aq)} \tag{8.4.12}$$

$$\mathrm{H_2SO_4\,(aq) \longrightarrow H^+\,(aq) + HSO_4^-\,(aq)} \tag{8.4.13}$$

Carbon dioxide is naturally present in the atmosphere because we and most other organisms produce it as a waste product of metabolism. Carbon dioxide is also formed when fires release carbon stored in vegetation or when we burn wood or fossil fuels. Sulfur trioxide in the atmosphere is naturally produced by volcanic activity, but it also stems from burning fossil fuels, which have traces of sulfur, and from the process of "roasting" ores of metal sulfides in metal-refining processes. Oxides of nitrogen are formed in internal combustion engines where the high temperatures make it possible for the nitrogen and oxygen in air to chemically combine.

Acid rain is a particular problem in industrial areas where the products of combustion and smelting are released into the air without being stripped of sulfur and nitrogen oxides. In North America and Europe until the 1980s, it was responsible for the destruction of forests and freshwater lakes, when the acidity of the rain actually killed trees, damaged soil, and made lakes uninhabitable for all but the most acid-tolerant species. Acid rain also corrodes statuary and building facades that are made of marble and limestone (Figure 8.4.4). Regulations limiting the amount of sulfur and nitrogen oxides that can be released into the atmosphere by industry and automobiles have reduced the severity of acid damage to both natural and manmade environments in North America and Europe. It is now a growing problem in industrial areas of China and India.



(a) (b)

Figure 8.4.4: (a) Acid rain makes trees more susceptible to drought and insect infestation, and depletes nutrients in the soil. (b) It also is corrodes statues that are carved from marble or limestone. (credit a: modification of work by Chris M Morris; credit b: modification of work by "Eden, Janine and Jim"/Flickr)

Two photos are shown. Photograph a on the left shows the upper portion of trees against a bright blue sky. The tops of several trees at the center of the photograph have bare branches and appear to be dead. Image b shows a statue of a man that appears to from the revolutionary war era in either marble or limestone.

Key Takeaways

- Acids and bases can be strong or weak depending on the extent of ionization in solution.
- Most chemical reactions reach equilibrium at which point there is no net change.
- The ion-product of $[H^+][OH^-]$ in an aqueous solution remains constant.
- A ***p*H** value is simply the *negative* of the logarithm of the H^+ ion concentration ($-\log[H^+]$).
- The pH scale is used to succinctly communicate the acidity or basicity of a solution.
- A solution is **acidic** if $p\mathrm{H} < 7$.
- A solution is **basic** if $p\mathrm{H} > 7$.
- A solution is **neutral** if $p\mathrm{H} = 7$.

8.4: The Strengths of Acids and Bases is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.4: The Strengths of Acids and Bases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

8.5: Buffers

Learning Objectives

- To define *buffer* and describe how it reacts with an acid or a base.

Weak acids are relatively common, even in the foods we eat. But we occasionally come across a strong acid or base, such as stomach acid, that has a strongly acidic pH of 1–2. By definition, strong acids and bases can produce a relatively large amount of hydrogen or hydroxide ions and, as a consequence, have a marked chemical activity. In addition, very small amounts of strong acids and bases can change the pH of a solution very quickly. If 1 mL of stomach acid [which we will approximate as 0.05 M HCl(aq)] is added to the bloodstream, and if no correcting mechanism is present, the pH of the blood would go from about 7.4 to about 4.9—a pH that is not conducive to continued living. Fortunately, the body has a mechanism for minimizing such dramatic pH changes.

The mechanism involves a **buffer**, a solution that resists dramatic changes in pH. A *buffer* (or *buffered*) solution is one that resists a change in its pH when H^+ or OH^- ions are added or removed owing to some other reaction taking place in the same solution. Buffers do so by being composed of certain pairs of solutes: either a **weak acid plus its conjugate base** or a **weak base plus its conjugate acid**.

For example, a buffer can be composed of dissolved **acetic acid ($HC_2H_3O_2$, a weak acid) and sodium acetate** ($NaC_2H_3O_2$). Sodium acetate is a salt that dissociates into sodium ions and acetate ions in solution. For as long as acetic acid and acetate ions are present in significant amounts a solution, this can resist dramatic pH changes. Another example of a buffer is a solution containing **ammonia (NH_3, a weak base) and ammonium chloride** (NH_4Cl). Ammonium acetate is also a salt that dissociates into ammonium ions and chloride ions in solution. The presence of ammonium ions with ammonia molecules satisfies the requisite condition for a buffer solution.

How Buffers Work

The essential component of a buffer system is a **conjugate acid-base pair** whose concentration is fairly high in relation to the concentrations of added H^+ or OH^- it is expected to buffer against. Let us use an acetic acid–sodium acetate buffer to demonstrate how buffers work. If a strong base—a source of OH^-(aq) ions—is added to the buffer solution, those hydroxide ions will react with the acetic acid in an acid-base reaction:

$$HC_2H_3O_2\,(aq) + OH^-\,(aq) \longrightarrow H_2O(\ell) + C_2H_3O_2^-\,(aq) \tag{8.5.1}$$

Rather than changing the pH dramatically by making the solution basic, the added hydroxide ions react to make water, and the pH does not change much.

> Many people are aware of the concept of buffers from *buffered aspirin*, which is aspirin that also has magnesium carbonate, calcium carbonate, magnesium oxide, or some other salt. The salt acts like a base, while aspirin is itself a weak acid.

If a strong acid—a source of H^+ ions—is added to the buffer solution, the H^+ ions will react with the anion from the salt. Because $HC_2H_3O_2$ is a weak acid, it is not ionized much. This means that if lots of hydrogen ions and acetate ions (from sodium acetate) are present in the same solution, they will come together to make acetic acid:

$$H^+\,(aq) + C_2H_3O_2^-\,(aq) \longrightarrow HC_2H_3O_2\,(aq) \tag{8.5.2}$$

Rather than changing the pH dramatically and making the solution acidic, the added hydrogen ions react to make molecules of a weak acid. Figure 8.5.1 illustrates both actions of a buffer.

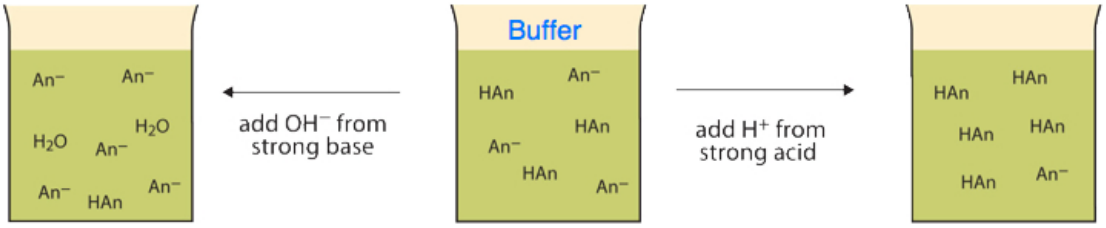


Figure 8.5.1: The Action of Buffers. Buffers can react with both strong acids (top) and strong bases (bottom) to minimize large changes in pH.

A simple buffer system might be a 0.2 *M* solution of sodium acetate; the conjugate pair here is acetic acid HAc and its conjugate base, the acetate ion Ac^-. The idea is that this conjugate pair "pool" will be available to gobble up any small ($\leq 10^{-3}$ *M*) addition of H+ or OH^- that may result from other processes going on in the solution.

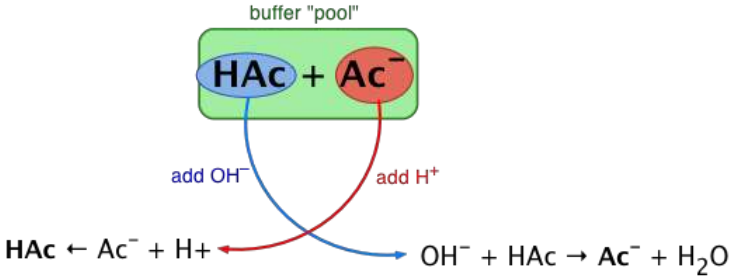


Figure 8.5.2: How HAc/Ac buffer works

Buffers work well only for limited amounts of added strong acid or base. Once either solute is all reacted, the solution is no longer a buffer, and rapid changes in pH may occur. We say that a buffer has a certain capacity. Buffers that have more solute dissolved in them to start with have larger capacities, as might be expected.

Buffers made from weak bases and salts of weak bases act similarly. For example, in a buffer containing NH_3 and NH_4Cl, ammonia molecules can react with any excess hydrogen ions introduced by strong acids:

$$NH_{3(aq)} + H^+_{(aq)} \rightarrow NH^+_{4(aq)} \tag{8.5.3}$$

while the ammonium ion [NH_4^+(aq)] can react with any hydroxide ions introduced by strong bases:

$$NH^+_{4(aq)} + OH^-_{(aq)} \rightarrow NH_{3(aq)} + H_2O_{(\ell)} \tag{8.5.4}$$

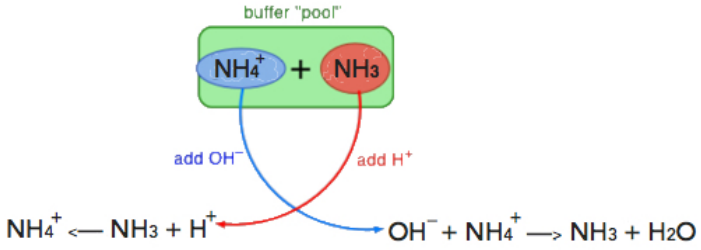


Figure 8.5.3: How NH_4^+/NH_3 buffer works

Example 8.5.1

Which solute combinations can make a buffer solution? Assume all are aqueous solutions.

a. $HCHO_2$ and $NaCHO_2$
b. HCl and NaCl
c. CH_3NH_2 and CH_3NH_3Cl

d. NH_3 and NaOH

Solution

a. Formic acid ($HCHO_2$) is a weak acid, while $NaCHO_2$ is a salt supplying —formate ion (CHO_2^-), the conjugate base of $HCHO_2$. The combination of these two solutes would make a buffer solution.
b. Hydrochloric acid (HCl) is a strong acid, not a weak acid, so the combination of these two solutes would not make a buffer solution.
c. Methylamine (CH_3NH_2) is like ammonia, a weak base. The compound CH_3NH_3Cl is a salt supplying $CH_3NH_3^+$, the conjugate acid of CH_3NH_2. The combination of these two solutes would make a buffer solution.
d. Ammonia (NH_3) is a weak base, but NaOH is a strong base. The combination of these two solutes would not make a buffer solution.

? Exercise 8.5.1

Which solute combinations can make a buffer solution? Assume all are aqueous solutions.

a. $NaHCO_3$ and NaCl
b. H_3PO_4 and NaH_2PO_4
c. NH_3 and $(NH_4)_3PO_4$
d. NaOH and NaCl

Answer

b. H_3PO_4 (weak acid) and $H_2PO_4^-$ (conjugate base of H_3PO_4) make a buffer.

c. NH_3 (weak base) and NH_4^+ (conjugate acid of NH_3) make a buffer

✓ Food and Drink App: The Acid That Eases Pain

Although medicines are not exactly "food and drink," we do ingest them, so let's take a look at an acid that is probably the most common medicine: acetylsalicylic acid, also known as aspirin. Aspirin is well known as a pain reliever and antipyretic (fever reducer).

The structure of aspirin is shown in the accompanying figure. The acid part is circled; it is the H atom in that part that can be donated as aspirin acts as a Brønsted-Lowry acid. Because it is not given in Table 10.5.1, acetylsalicylic acid is a weak acid. However, it is still an acid, and given that some people consume relatively large amounts of aspirin daily, its acidic nature can cause problems in the stomach lining, despite the stomach's defenses against its own stomach acid.

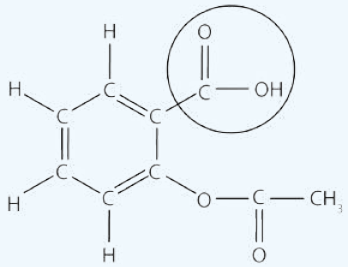


Figure 8.5.4 The Molecular Structure of Aspirin. The circled atoms are the acid part of the molecule.

Because the acid properties of aspirin may be problematic, many aspirin brands offer a "buffered aspirin" form of the medicine. In these cases, the aspirin also contains a buffering agent-usually MgO-that regulates the acidity of the aspirin to minimize its acidic side effects.

As useful and common as aspirin is, it was formally marketed as a drug starting in 1899. The US Food and Drug Administration (FDA), the governmental agency charged with overseeing and approving drugs in the United States, wasn't formed until 1906. Some have argued that if the FDA had been formed before aspirin was introduced, aspirin may never have gotten approval due to its potential for side effects-gastrointestinal bleeding, ringing in the ears, Reye's syndrome (a liver

problem), and some allergic reactions. However, recently aspirin has been touted for its effects in lessening heart attacks and strokes, so it is likely that aspirin is here to stay.

Buffer solutions are essential components of all living organisms.

- Our blood is buffered to maintain a pH of 7.4 that must remain unchanged as metabolically-generated CO_2 (carbonic acid) is added and then removed by our lungs.
- Buffers in the oceans, in natural waters such as lakes and streams, and within soils help maintain their environmental stability against acid rain and increases in atmospheric CO_2.
- Many industrial processes, such as brewing, require buffer control, as do research studies in biochemistry and physiology that involve enzymes, are active only within certain pH ranges.

The pH in living systems (Figure \(\PageIndex{1}\)) is maintained by buffer systems.

Table 7.3.2: pH in Living Systems

Compartment	pH
Gastric Acid	1
Lysosomes	4.5
Granules of Chromaffin Cells	5.5
Human Skin	5.5
Urine	6
Neutral H_2O at 37 °C	6.81
Cytosol	7.2
Cerebrospinal Fluid	7.3
Blood	7.43-7.45
Mitochondrial Matrix	7.5
Pancreas Secretions	8.1

Human blood has a buffering system to minimize extreme changes in pH. One buffer in blood is based on the presence of HCO_3^- and H_2CO_3 [H_2CO_3 is another way to write CO_2(aq)]. With this buffer present, even if some stomach acid were to find its way directly into the bloodstream, the change in the pH of blood would be minimal. Inside many of the body's cells, there is a buffering system based on phosphate ions.

Medicine: The Buffer System in Blood

The normal pH of human blood is about 7.4. The carbonate buffer system in the blood uses the following equilibrium reaction:

$$\mathbf{CO_2}(g) + \mathbf{2\,H_2O}(l) \rightleftharpoons \mathbf{H_2CO_3}(aq) \rightleftharpoons \mathbf{HCO_3^-}(aq) + \mathbf{H_3O^+}(aq)$$

The concentration of carbonic acid, H_2CO_3 is approximately 0.0012 *M*, and the concentration of the hydrogen carbonate ion, $\mathbf{HCO_3^-}$, is around 0.024 *M*. Using the Henderson-Hasselbalch equation and the pK_a of carbonic acid at body temperature, we can calculate the pH of blood:

$$\mathbf{pH} = \mathbf{p}K_\mathbf{a} + \log\frac{[\mathbf{base}]}{[\mathbf{acid}]} = 6.1 + \log\frac{0.024}{0.0012} = 7.4$$

The fact that the H_2CO_3 concentration is significantly lower than that of the $\mathbf{HCO_3^-}$ ion may seem unusual, but this imbalance is due to the fact that most of the by-products of our metabolism that enter our bloodstream are acidic. Therefore, there must be a larger proportion of base than acid, so that the capacity of the buffer will not be exceeded.

Lactic acid is produced in our muscles when we exercise. As the lactic acid enters the bloodstream, it is neutralized by the HCO_3^- ion, producing H_2CO_3. An enzyme then accelerates the breakdown of the excess carbonic acid to carbon dioxide and water, which can be eliminated by breathing. In fact, in addition to the regulating effects of the carbonate buffering system on the pH of blood, the body uses breathing to regulate blood pH. If the pH of the blood decreases too far, an increase in breathing removes CO_2 from the blood through the lungs driving the equilibrium reaction such that $[H_3O^+]$ is lowered. If the blood is too alkaline, a lower breath rate increases CO_2 concentration in the blood, driving the equilibrium reaction the other way, increasing $[H^+]$ and restoring an appropriate pH.

Career Focus: Blood Bank Technology Specialist

At this point in this text, you should have the idea that the chemistry of blood is fairly complex. Because of this, people who work with blood must be specially trained to work with it properly.

A blood bank technology specialist is trained to perform routine and special tests on blood samples from blood banks or transfusion centers. This specialist measures the pH of blood, types it (according to the blood's ABO+/− type, Rh factors, and other typing schemes), tests it for the presence or absence of various diseases, and uses the blood to determine if a patient has any of several medical problems, such as anemia. A blood bank technology specialist may also interview and prepare donors to give blood and may actually collect the blood donation.

Blood bank technology specialists are well trained. Typically, they require a college degree with at least a year of special training in blood biology and chemistry. In the United States, training must conform to standards established by the American Association of Blood Banks.

Key Takeaway

- A buffer is a solution that resists sudden changes in pH.

Concept Review Exercise

1. Explain how a buffer prevents large changes in pH.

Answer

1. A buffer has components that react with both strong acids and strong bases to resist sudden changes in pH.

Exercises

1. Describe a buffer. What two related chemical components are required to make a buffer?
2. Can a buffer be made by combining a strong acid with a strong base? Why or why not?
3. Which solute combinations can make a buffer? Assume all are aqueous solutions.
    1. HCl and NaCl
    2. HNO_2 and $NaNO_2$
    3. NH_4NO_3 and HNO_3
    4. NH_4NO_3 and NH_3
4. Which solute combinations can make a buffer? Assume all are aqueous solutions.
    1. H_3PO_4 and Na_3PO_4
    2. $NaHCO_3$ and Na_2CO_3
    3. $NaNO_3$ and $Ca(NO_3)_2$
    4. HN_3 and NH_3
5. For each combination in Exercise 3 that is a buffer, write the chemical equations for the reactions of the buffer components when a strong acid and a strong base is added.
6. For each combination in Exercise 4 that is a buffer, write the chemical equations for the reaction of the buffer components when a strong acid and a strong base is added.
7. The complete phosphate buffer system is based on four substances: H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, and PO_4^{3-}. What different buffer solutions can be made from these substances?

8. Explain why NaBr cannot be a component in either an acidic or a basic buffer.
9. Explain why $Mg(NO_3)_2$ cannot be a component in either an acidic or a basic buffer.

Answers

1. A buffer resists sudden changes in pH. It has a weak acid or base and a salt of that weak acid or base.
2. No. Combining a strong acid and a strong base will produce salt and water. Excess strong acid or strong base will not act as a buffer.
3. 1. not a buffer
   2. buffer
   3. not a buffer
   4. buffer

4. 1. not a buffer

2. buffer

3. not a buffer

4. not buffer

5. 3b: strong acid: $H^+ + NO_2^- \rightarrow HNO_2$; strong base: $OH^- + HNO_2 \rightarrow H_2O + NO_2^-$; 3d: strong acid: $H^+ + NH_3 \rightarrow NH_4^+$; strong base: $OH^- + NH_4^+ \rightarrow H_2O + NH_3$
6. 4b: strong acid: $H^+ + CO_3^{2-} \rightarrow HCO3^-$; strong base: $OH^- + HCO_3^- \rightarrow H_2O + CO_3^{2-}$;
7. Buffers can be made by combining H_3PO_4 and $H_2PO_4^-$, $H_2PO_4^-$ and HPO_4^{2-}, and HPO_4^{2-} and PO_4^{3-}.
8. NaBr splits up into two ions in solution, Na^+ and Br^-. Na^+ will not react with any added base knowing that NaOH is a strong base. Br- will not react with any added acid knowing that HBr is a strong acid. Because NaBr will not react with any added base or acid, it does not resist change in pH and is not a buffer.
9. $Mg(NO_3)_2$ includes two types of ions, Mg^{2+} and NO_3^-. $Mg(OH)_2$ is strong base and completely dissociates (100% falls apart), so Mg^{2+} will not react with any added base (0% combines with OH^-). HNO_3 is strong acid and completely dissociates (100% falls apart), so NO_3^- will not react with any added acid (0% combines with H^+). Because $Mg(NO_3)_2$ will not react with any added base or acid, it does not resist change in pH and is not a buffer.

8.5: Buffers is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.5: Buffers** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

8.6: Acids and Bases (Exercises)

10.1: Arrhenius Definition of Acids and Bases

Concept Review Exercises

1. Give the Arrhenius definitions of an acid and a base.
2. What is neutralization?

Answers

1. Arrhenius acid: a compound that increases the concentration of hydrogen ion (H^+) in aqueous solution; Arrhenius base: a compound that increases the concentration of hydroxide ion (OH^-) in aqueous solution.
2. the reaction of an acid and a base

Exercises

1. Give two examples of Arrhenius acids.
2. Give two examples of Arrhenius bases.
3. List the general properties of acids.
4. List the general properties of bases.
5. Name each compound. (For acids, look up the name in Table 10.1.1. For bases, use the rules for naming ionic compounds from Chapter 3.)

    a. HBr(aq)

    b. $Ca(OH)_2$(aq)

    c. HNO_3(aq)

    d. $Fe(OH)_3$(aq)

6. Name each compound.

    a. HI(aq)

    b. $Cu(OH)_2$(aq)

    c. H_3PO_4(aq)

    d. CsOH(aq)

7. Write a balanced chemical equation for the neutralization of $Ba(OH)_2$(aq) with HNO_3(aq).

8. Write a balanced chemical equation for the neutralization of H_2SO_4(aq) with $Cr(OH)_3$(aq).

9. Gastric juice, the digestive fluid produced in the stomach, contains hydrochloric acid, HCl. Milk of Magnesia, a suspension of solid $Mg(OH)_2$ in an aqueous medium, is sometimes used to neutralize excess stomach acid. Write a complete balanced equation for the neutralization reaction.

10. Identify the salt produced in each acid-base reaction below. Then, balance the equation.

    a. $2HCl + Sr(OH)_2 \rightarrow 2H_2O$ + ??

    b. KNO_3; $HNO_3 + KOH \rightarrow$?? + H_2O

    c. $HF + Ca(OH)_2$ ---> ?? + H_2O

11. How many moles of sodium hydroxide (NaOH) are needed to neutralize 0.844 mol of acetic acid ($HC_2H_3O_2$)? (Hint: begin by writing a balanced chemical equation for the process.)

12. How many moles of perchloric acid ($HClO_4$) are needed to neutralize 0.052 mol of calcium hydroxide [$Ca(OH)_2$]? (Hint: begin by writing a balanced chemical equation for the process

13. Hydrazoic acid (HN_3) can be neutralized by a base.

a. Write the balanced chemical equation for the reaction between hydrazoic acid and calcium hydroxide.

b. How many milliliters of 0.0245 M $Ca(OH)_2$ are needed to neutralize 0.564 g of HN_3?

14. Citric acid ($H_3C_6H_5O_7$) has three hydrogen atoms that can form hydrogen ions in solution.

a. Write the balanced chemical equation for the reaction between citric acid and sodium hydroxide.

b. If an orange contains 0.0675 g of $H_3C_6H_5O_7$, how many milliliters of 0.00332 M NaOH solution are needed to neutralize the acid?

15. Magnesium hydroxide [$Mg(OH)_2$] is an ingredient in some antacids. How many grams of $Mg(OH)_2$ are needed to neutralize the acid in 158 mL of 0.106 M HCl(aq)? It might help to write the balanced chemical equation first.

16. Aluminum hydroxide [$Al(OH)_3$] is an ingredient in some antacids. How many grams of $Al(OH)_3$ are needed to neutralize the acid in 96.5 mL of 0.556 M H_2SO_4(aq)? It might help to write the balanced chemical equation first.

17. Write the balanced chemical equation for the reaction between HBr and $Ca(OH)_2$. What volume of 0.010 M HBr solution is be required to neutralize 25 mL of a 0.0100*M* $Ca(OH)_2$ solution?

18. Write the balanced chemical equation for the reaction between HNO_3 and KOH. What volume of 0.5M HNO_3 is required to neutralize 60 mL of 0.4M KOH solution?

Answers

1. HCl and HNO_3 (answers will vary)
2. NaOH and $Ca(OH)_2$ (answers will vary)
3. sour taste, react with metals, react with bases, and turn litmus red
4. bitter taste, feels slippery, react with acids and turn litmus blue
5. a. hydrobromic acid

b. calcium hydroxide

c. nitric acid

d. iron(III) hydroxide

6. a. hydroiodic acid

b. cupric hydroxide

c. phosphoric acid

d. cesium hydroxide

7. $2HNO_3(aq) + Ba(OH)_2(aq) \rightarrow Ba(NO_3)_2(aq) + 2H_2O$

8. $3H_2SO_4(aq) + 2Cr(OH)_3(aq) \rightarrow Cr_2(SO_4)_3(aq) + 6H_2O$

9. $Mg(OH)_2 + 2HCl$ --> $MgCl_2 + 2H_2O$

10. a. $SrCl_2$; $2HCl + Sr(OH)_2 \rightarrow 2H_2O + SrCl_2$

b. KNO_3; $HNO_3 + KOH \rightarrow KNO_3 + H_2O$

c. CaF_2; $2HF + Ca(OH)_2 \rightarrow CaF_2 + 2H_2O$

11. 0.844 mol

12. 0.104 mol

13. Part 1: $2HN_3(aq) + Ca(OH)_2 \rightarrow Ca(N_3)_2 + 2H_2O$

Part 2: 268 mL

14. Part 1: $H_3C_6H_5O_7(aq) + 3NaOH(aq) \rightarrow Na_3C_6H_5O_7(aq) + 3H_2O$

Part 2: 317.5 mL

15. 0.488 g

16. 2.79 g

17. $2HBr + Ca(OH)_2 \rightarrow CaBr_2 + 2H_2O$; 50 mL HBr

18. $HNO_3 + KOH \rightarrow KNO_3 + H_2O$; 48 mL HNO_3

10.2: Brønsted-Lowry Definition of Acids and Bases

Concept Review Exercise

1. Give the definitions of a Brønsted-Lowry acid and a Brønsted-Lowry base.

Answer

1. A Brønsted-Lowry acid is a proton donor, while a Brønsted-Lowry base is a proton acceptor.

Exercises

1. Label each reactant as a Brønsted-Lowry acid or a Brønsted-Lowry base.

   $HCl(aq) + NH_3(aq) \rightarrow NH_4^+(aq) + Cl^-(aq)$

2. Label each reactant as a Brønsted-Lowry acid or a Brønsted-Lowry base.

   $H_2O(\ell) + N_2H_4(aq) \rightarrow N_2H_5^+(aq) + OH^-(aq)$

3. Explain why a Brønsted-Lowry acid can be called a proton donor.
4. Explain why a Brønsted-Lowry base can be called a proton acceptor.
5. Write the chemical equation of the reaction of ammonia in water and label the Brønsted-Lowry acid and base.
6. Write the chemical equation of the reaction of methylamine (CH_3NH_2) in water and label the Brønsted-Lowry acid and base.
7. Demonstrate that the dissolution of HNO_3 in water is actually a Brønsted-Lowry acid-base reaction by describing it with a chemical equation and labeling the Brønsted-Lowry acid and base.
8. Identify the Brønsted-Lowry acid and base in the following chemical equation:

   $C_3H_7NH_2(aq) + H_3O^+(aq) \rightarrow C_3H_7NH_3^+(aq) + H_2O(\ell)$

9. Identify the Brønsted-Lowry acid and the Brønsted-Lowry base in each of the following equations

   1. $\mathbf{NO_2^- + H_2O \longrightarrow HNO_2 + OH^-}$
   2. $\mathbf{HBr + H_2O \longrightarrow H_3O^+ + Br^-}$
   3. $\mathbf{HS^- + H_2O \longrightarrow H_2S + OH^-}$
   4. $\mathbf{H_2PO_4^- + OH^- \longrightarrow HPO_4^{2-} + H_2O}$
   5. $\mathbf{H_2PO_4^- + HCl \longrightarrow H_3PO_4 + Cl^-}$

10. Write the chemical equation for the reaction that occurs when cocaine hydrochloride ($C_{17}H_{22}ClNO_4$) dissolves in water and donates a proton to a water molecule. (When hydrochlorides dissolve in water, they separate into chloride ions and the appropriate cation.)

11. If codeine hydrobromide has the formula $C_{18}H_{22}BrNO_3$, what is the formula of the parent compound codeine?

Answers

1. HCl: Brønsted-Lowry acid; NH_3: Brønsted-Lowry base
2. H_2O: Brønsted-Lowry acid; N_2H_4: Brønsted-Lowry base
3. A Brønsted-Lowry acid gives away an H^+ ion—nominally, a proton—in an acid-base reaction.
4. A Brønsted-Lowry base accepts an H^+ ion (a proton) in an acid-base reaction.
5. $NH_3 + H_2O \rightarrow NH_4^+ + OH^-$ (here NH_3 = Brønsted-Lowry base; H_2O = Brønsted-Lowry acid)
6. $CH_3NH_2 + H_2O \rightarrow CH_3NH_3^+ + OH^-$ (here CH_3NH_2 = Brønsted-Lowry base; H_2O = Brønsted-Lowry acid)
7. $HNO_3 + H_2O \rightarrow H_3O^+ + NO_3^-$ (here HNO_3 = Brønsted-Lowry acid; H_2O = Brønsted-Lowry base)

8. $C_3H_7NH_2(aq) + H_3O^+(aq) \rightarrow C_3H_7NH_3^+(aq) + H_2O(\ell)$ (here H_3O^+ = Brønsted-Lowry acid; $C_3H_7NH_2$ = Brønsted-Lowry base)

10.3: Water - Both an Acid and a Base

Concept Review Exercises

1. Explain how water can act as an acid.
2. Explain how water can act as a base.

Answers

1. Under the right conditions, H_2O can donate a proton, making it a Brønsted-Lowry acid.
2. Under the right conditions, H_2O can accept a proton, making it a Brønsted-Lowry base.

Exercises

1. Is $H_2O(\ell)$ acting as an acid or a base?

   $H_2O(\ell) + NH_4^+(aq) \rightarrow H_3O^+(aq) + NH_3(aq)$

2. Is $H_2O(\ell)$ acting as an acid or a base?

   $CH_3^-(aq) + H_2O(\ell) \rightarrow CH_4(aq) + OH^-(aq)$

3. In the aqueous solutions of some salts, one of the ions from the salt can react with water molecules. In some $C_2H_3O_2^-$ solutions, the following reaction can occur:

   $C_2H_3O_2^-(aq) + H_2O(\ell) \rightarrow HC_2H_3O_2(aq) + OH^-(aq)$

   Is H_2O acting as an acid or a base in this reaction?

4. In the aqueous solutions of some salts, one of the ions from the salt can react with water molecules. In some NH_4^+ solutions, the following reaction can occur:

   $NH_4^+(aq) + H_2O \rightarrow NH_3(aq) + H_3O^+(aq)$

   Is H_2O acting as an acid or a base in this reaction?

5. Why is pure water considered neutral?

Answers

1. base
2. acid
3. acid
4. base

5. When water ionizes, equal amounts of H^+ (acid) and OH^- (base) are formed, so the solution is neither acidic nor basic: $H_2O(\ell) \rightarrow H^+(aq) + OH^-(aq)$

10.4: The Strengths of Acids and Bases

Concept Review Exercises

1. Explain the difference between a strong acid or base and a weak acid or base.
2. Explain what is occurring when a chemical reaction reaches equilibrium.
3. Define *pH*.

AnswerS

1. A strong acid or base is 100% ionized in aqueous solution; a weak acid or base is less than 100% ionized.
2. The overall reaction progress stops because the reverse process balances out the forward process.
3. pH is a measure of the hydrogen ion concentration.

Exercises

1. Name a strong acid and a weak acid. (Hint: use Table 10.4.1.)
2. Name a strong base and a weak base. (Hint: use Table 10.4.1.)
3. Is each compound a strong acid or a weak acid? Assume all are in aqueous solution. (Hint: use Table 10.4.1.)
    1. HF
    2. $HC_2H_3O_2$
    3. HCl
    4. $HClO_4$
4. Is each compound a strong acid or a weak acid? Assume all are in aqueous solution. (Hint: use Table 10.4.1.)
    1. H_2SO_4
    2. HSO_4^-
    3. HPO_4^{2-}
    4. HNO_3
5. Is each compound a strong base or a weak base? Assume all are in aqueous solution. (Hint: use Table 10.4.1.)
    1. NH_3
    2. NaOH
    3. $Mg(OH)_2$
    4. $Cu(OH)_2$
6. Is each compound a strong base or a weak base? Assume all are in aqueous solution. (Hint: use Table 10.4.1.)
    1. KOH
    2. H_2O
    3. $Fe(OH)_2$
    4. $Fe(OH)_3$
7. Write the chemical equation for the equilibrium process for each weak acid in Exercise 3.
8. Write the chemical equation for the equilibrium process for each weak acid in Exercise 4.
9. Write the chemical equation for the equilibrium process for each weak base in Exercise 5.
10. Write the chemical equation for the equilibrium process for each weak base in Exercise 6.
11. Which is the stronger acid—HCl(aq) or HF(aq)?
12. Which is the stronger base—KOH(aq) or $Ni(OH)_2$(aq)?
13. Consider the two acids in Exercise 11. For solutions that have the same concentration, which one would you expect to have a lower pH?
14. Consider the two bases in Exercise 12. For solutions that have the same concentration, which one would you expect to have a higher pH?
15. Consider the list of substances in Table \PageIndex3\PageIndex3.2"The pH Values of Some Common Solutions". What is the most acidic substance on the list that you have encountered recently?
16. Consider the list of substances in Table \PageIndex3\PageIndex3.2"The pH Values of Some Common Solutions". What is the most basic substance on the list that you have encountered recently?
17. Indicate whether solutions with the following pH values are acidic, basic, or neutral:
    1. pH = 9.4
    2. pH = 7.0
    3. pH = 1.2
    4. pH = 6.5

Answers

1. strong acid: HCl; weak acid: $HC_2H_3O_2$ (answers will vary)
2. strong base: NaOH; weak base: NH_3 (answers will vary)
3. 1. weak
    2. weak

3. strong
4. strong

4. 1. strong
2. weak
3. weak
4. strong

5. 1. weak
2. strong
3. strong
4. weak

6. 1. strong
2. weak
3. weak
4. weak

7. 3a: $HF(aq) \leftrightarrows H^+(aq) + F^-(aq)$; 3b: $HC_2H_3O_2(aq) \leftrightarrows H^+(aq) + C_2H_3O_2^-(aq)$

8. 4b: $HSO_4^-(aq) \leftrightarrows H^+(aq) + SO_4^{2-}(aq)$; 4c: $HPO_4^{2-}(aq) \leftrightarrows H^+(aq) + PO_4^{3-}(aq)$

9. 5a: $NH_3(aq) + H_2O \leftrightarrows NH_4^+(aq) + OH^-(aq)$; 5d: $Cu(OH)_2(aq) \leftrightarrows Cu^{2+}(aq) + 2OH^-(aq)$

10. 6b: $H_2O + H_2O \leftrightarrows H_3O^+(aq) + OH^-(aq)$; 6c: $Fe(OH)_2(aq) \leftrightarrows Fe^{2+}(aq) + 2OH^-(aq)$; 6d: $Fe(OH)3(aq) \leftrightarrows Fe^{3+}(aq) + 3OH^-(aq)$

11. HCl(aq)

12. KOH(aq)

13. HCl(aq)

14. KOH(aq)

15. (answers will vary)

16. (answers will vary)

17. 1. basic

2. neutral

3. acidic (strongly)

4. acidic (mildly)

10.5: Buffers

Concept Review Exercise

1. Explain how a buffer prevents large changes in pH.

Answer

1. A buffer has components that react with both strong acids and strong bases to resist sudden changes in pH.

Exercises

1. Describe a buffer. What two related chemical components are required to make a buffer?
2. Can a buffer be made by combining a strong acid with a strong base? Why or why not?
3. Which solute combinations can make a buffer? Assume all are aqueous solutions.
    1. HCl and NaCl
    2. HNO_2 and $NaNO_2$
    3. NH_4NO_3 and HNO_3

4. NH_4NO_3 and NH_3

4. Which solute combinations can make a buffer? Assume all are aqueous solutions.
    1. H_3PO_4 and Na_3PO_4
    2. $NaHCO_3$ and Na_2CO_3
    3. $NaNO_3$ and $Ca(NO_3)_2$
    4. HN_3 and NH_3
5. For each combination in Exercise 3 that is a buffer, write the chemical equations for the reactions of the buffer components when a strong acid and a strong base is added.
6. For each combination in Exercise 4 that is a buffer, write the chemical equations for the reaction of the buffer components when a strong acid and a strong base is added.
7. The complete phosphate buffer system is based on four substances: H_3PO_4, $H_2PO_4^-$, HPO_4^{2-}, and PO_4^{3-}. What different buffer solutions can be made from these substances?
8. Explain why NaBr cannot be a component in either an acidic or a basic buffer.
9. Explain why $Mg(NO_3)_2$ cannot be a component in either an acidic or a basic buffer.

Answers

1. A buffer resists sudden changes in pH. It has a weak acid or base and a salt of that weak acid or base.
2. No. Combining a strong acid and a strong base will produce salt and water. Excess strong acid or strong base will not act as a buffer.
3. 1. not a buffer
    2. buffer
    3. not a buffer
    4. buffer

4. 1. not a buffer

2. buffer

3. not a buffer

4. not buffer

5. 3b: strong acid: $H^+ + NO_2^- \rightarrow HNO_2$; strong base: $OH^- + HNO_2 \rightarrow H_2O + NO_2^-$; 3d: strong acid: $H^+ + NH_3 \rightarrow NH_4^+$; strong base: $OH^- + NH_4^+ \rightarrow H_2O + NH_3$
6. 4b: strong acid: $H^+ + CO_3^{2-} \rightarrow HCO3^-$; strong base: $OH^- + HCO_3^- \rightarrow H_2O + CO_3^{2-}$;
7. Buffers can be made by combining H_3PO_4 and $H_2PO_4^-$, $H_2PO_4^-$ and HPO_4^{2-}, and HPO_4^{2-} and PO_4^{3-}.
8. NaBr splits up into two ions in solution, Na^+ and Br^-. Na^+ will not react with any added base knowing that NaOH is a strong base. Br- will not react with any added acid knowing that HBr is a strong acid. Because NaBr will not react with any added base or acid, it does not resist change in pH and is not a buffer.
9. $Mg(NO_3)_2$ includes two types of ions, Mg^{2+} and NO_3^-. $Mg(OH)_2$ is strong base and completely dissociates (100% falls apart), so Mg^{2+} will not react with any added base (0% combines with OH^-). HNO_3 is strong acid and completely dissociates (100% falls apart), so NO_3^- will not react with any added acid (0% combines with H^+). Because $Mg(NO_3)_2$ will not react with any added base or acid, it does not resist change in pH and is not a buffer.

Additional Exercises

1. The properties of a 1.0 M HCl solution and a 1.0 M $HC_2H_3O_2$ solution are compared. Measurements show that the hydrochloric acid solution has a higher osmotic pressure than the acetic acid solution. Explain why.
2. Of a 0.50 M HNO_3 solution and a 0.50 M $HC_2H_3O_2$ solution, which should have the higher boiling point? Explain why.
3. The reaction of sulfuric acid [$H_2SO_4(aq)$] with sodium hydroxide [NaOH(aq)] can be represented by two separate steps, with only one hydrogen ion reacting in each step. Write the chemical equation for each step.
4. The reaction of aluminum hydroxide [$Al(OH)_3(aq)$] with hydrochloric acid [HCl(aq)] can be represented by three separate steps, with only one hydroxide ion reacting in each step. Write the chemical equation for each step.

5. A friend brings you a small sample of an unknown chemical. Assuming that the chemical is soluble in water, how would you determine if the chemical is an acid or a base?
6. A neutral solution has a hydrogen ion concentration of about 1×10^{-7} M. What is the concentration of the hydroxide ion in a neutral solution?
7. The Lewis definitions of an acid and a base are based on electron pairs, not protons. A Lewis acid is an electron pair acceptor, while a Lewis base is an electron pair donor. Use Lewis diagrams to show that

   $H^+(aq) + OH^-(aq) \rightarrow H_2O(\ell)$

   is an acid-base reaction in the Lewis sense as well as in the Arrhenius and Brønsted-Lowry senses.
8. Given the chemical reaction

   $NH_3(g) + BF_3(g) \rightarrow NH_3—BF_3(s)$

   show that the reaction illustrated by this equation is an acid-base reaction if we use the Lewis definitions of an acid and a base (see Exercise 7). The product contains a bond between the N and B atoms.

Answers

1. HCl is a strong acid and yields more ions in solution. $HC_2H_3O_2$ is a weak acid and undergoes partial ionization in solution.
2. HNO_3 is a strong acid while $HC_2H_3O_2$ is a weak acid. HNO_3 dissociates 100% and its solution contains more ions. The more ions the solution contains the lower is its vapor pressure; the higher temperature is required for it to boil.
3. $H_2SO_4 + NaOH \rightarrow NaHSO_4 + H_2O$; $NaHSO_4 + NaOH \rightarrow Na_2SO_4 + H_2O$
4. $Al(OH)_3 + HCl \rightarrow Al(OH)_2Cl + H_2O$; $Al(OH)_2Cl + HCl \rightarrow Al(OH)Cl_2 + H_2O$; $Al(OH)Cl_2 + HCl \rightarrow AlCl3 + H_2O$
5. One way is to add it to $NaHCO_3$; if it bubbles, it is an acid. Alternatively, add the sample to litmus and look for a characteristic color change (red for acid, blue for base).
6. In a neutral solution, $[OH^-] = [H^+] = 1.0 \times 10^{-7}$ M
7. The O atom is donating an electron pair to the H^+ ion, making the base an electron pair donor and the acid an electron pair acceptor.
8. The N atom is donating a lone pair to B in BF_3, Hence NH_3 is the Lewis base and BF_3 is the Lewis acid.

8.6: Acids and Bases (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.E: Acids and Bases (Exercises)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **10.1: Arrhenius Definition of Acids and Bases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **10.2: Brønsted-Lowry Definition of Acids and Bases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **10.3: Water - Both an Acid and a Base** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **10.4: The Strengths of Acids and Bases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **10.5: Buffers** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

8.7: Acids and Bases (Summary)

To ensure that you understand the material in this chapter, you should review the meanings of the bold terms in the following summary and ask yourself how they relate to the topics in the chapter.

The earliest chemical definition of an acid, the **Arrhenius definition**, says that an **acid** is a compound that increases the amount of hydrogen ion (H^+) in aqueous solution. An **Arrhenius base** is a compound that increases the amount of hydroxide ion (OH^-) in aqueous solution. While most bases are named as ionic hydroxide compounds, aqueous acids have a naming system unique to acids. Acids and bases react together in a characteristic chemical reaction called **neutralization**, in which the products are water and a salt. The principles of stoichiometry, along with the balanced chemical equation for a reaction between an acid and a base, can be used to determine how much of one compound will react with a given amount of the other.

A **Brønsted-Lowry acid** is any substance that donates a proton to another substance. A **Brønsted-Lowry base** is any substance that accepts a proton from another substance. The reaction of ammonia with water to make ammonium ions and hydroxide ions can be used to illustrate Brønsted-Lowry acid and base behavior.

Some compounds can either donate or accept protons, depending on the circumstances. Such compounds are called **amphiprotic**. Water is one example of an amphiprotic compound. One result of water being amphiprotic is that a water molecule can donate a proton to another water molecule to make a hydronium ion and a hydroxide ion. This process is called the **autoionization of water** and occurs in any sample of water.

Not all acids and bases are equal in chemical strength. A **strong acid** is an acid whose molecules are all dissociated into ions in aqueous solution. Hydrochloric acid is an example of a strong acid. Similarly, a **strong base** is a base whose molecules are dissociated into ions in aqueous solution. Sodium hydroxide is an example of a strong base. Any acid or base whose molecules are not all dissociated into ions in aqueous solution is a **weak acid** or a **weak base**. Solutions of weak acids and weak bases reach a **chemical equilibrium** between the un-ionized form of the compound and the dissociated ions. It is a dynamic equilibrium because acid and base molecules are constantly dissociating into ions and reassociating into neutral molecules.

The **pH** scale is a scale used to express the concentration of hydrogen ions in solution. A neutral solution, neither acidic nor basic, has a pH of 7. Acidic solutions have a pH lower than 7, while basic solutions have a pH higher than 7.

Buffers are solutions that resist dramatic changes in pH when an acid or a base is added to them. They contain a weak acid and a salt of that weak acid, or a weak base and a salt of that weak base. When a buffer is present, any strong acid reacts with the anion of the salt, forming a weak acid and minimizing the presence of hydrogen ions in solution. Any strong base reacts with the weak acid, minimizing the amount of additional hydroxide ions in solution. However, buffers only have limited **capacity**; there is a limit to the amount of strong acid or strong base any given amount of buffer will react with.

8.7: Acids and Bases (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **10.S: Acids and Bases (Summary)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

CHAPTER OVERVIEW

9: Amino Acids, Proteins, and Enzymes

Proteins may be defined as compounds of high molar mass consisting largely or entirely of chains of amino acids. Their masses range from several thousand to several million daltons (Da). In addition to carbon, hydrogen, and oxygen atoms, all proteins contain nitrogen and sulfur atoms, and many also contain phosphorus atoms and traces of other elements. Proteins serve a variety of roles in living organisms and are often classified by these biological roles. Muscle tissue is largely protein, as are skin and hair. Proteins are present in the blood, in the brain, and even in tooth enamel. Each type of cell in our bodies makes its own specialized proteins, as well as proteins common to all or most cells. We begin our study of proteins by looking at the properties and reactions of amino acids, which is followed by a discussion of how amino acids link covalently to form peptides and proteins. We end the chapter with a discussion of enzymes—the proteins that act as catalysts in the body.

9.1: Prelude to Amino Acids, Proteins, and Enzymes
9.2: Properties of Amino Acids
9.3: Reactions of Amino Acids
9.4: Peptides
9.5: Proteins
9.6: X-ray Protein Crystallography
9.7: Enzymes
9.8: Enzyme Action
9.9: Enzyme Activity
9.10: Enzyme Inhibition
9.11: Enzyme Cofactors and Vitamins
9.12: Amino Acids, Proteins, and Enzymes (Exercises)
9.13: Amino Acids, Proteins, and Enzymes (Summary)

Template:HideTOC

9: Amino Acids, Proteins, and Enzymes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

9.1: Prelude to Amino Acids, Proteins, and Enzymes

The 1923 Nobel Prize in Medicine or Physiology was awarded to Frederick Grant Banting and John James Richard Macleod for their discovery of the protein *insulin*. In 1958, the Nobel Prize in Chemistry was awarded to Frederick Sanger for his discoveries concerning the structure of proteins and, in particular, the structure of insulin. What is so important about insulin that two Nobel Prizes have been awarded for work on this protein?

Insulin is a hormone that is synthesized in the pancreas. Insulin stimulates the transport of glucose into cells throughout the body and the storage of glucose as glycogen. People with diabetes do not produce insulin or use it properly. The isolation of insulin in 1921 led to the first effective treatment for these individuals.



Figure **9.1.1**: Insulin pump, showing an infusion set loaded into spring-loaded insertion device. A reservoir is attached to the infusion set (shown here removed from the pump). (Public Domain; User:David-i98).

9.1: Prelude to Amino Acids, Proteins, and Enzymes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.0: Prelude to Amino Acids, Proteins, and Enzymes** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

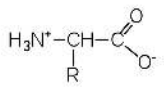
9.2: Properties of Amino Acids

Learning Objectives

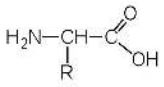
- To recognize amino acids and classify them based on the characteristics of their side chains.

The proteins in all living species, from bacteria to humans, are constructed from the same set of 20 amino acids, so called because each contains an amino group attached to a carboxylic acid. The amino acids in proteins are α-amino acids, which means the amino group is attached to the α-carbon of the carboxylic acid. Humans can synthesize only about half of the needed amino acids; the remainder must be obtained from the diet and are known as essential amino acids. However, two additional amino acids have been found in limited quantities in proteins: Selenocysteine was discovered in 1986, while pyrrolysine was discovered in 2002.

The amino acids are colorless, nonvolatile, crystalline solids, melting and decomposing at temperatures above 200°C. These melting temperatures are more like those of inorganic salts than those of amines or organic acids and indicate that the structures of the amino acids in the solid state and in neutral solution are best represented as having both a negatively charged group and a positively charged group. Such a species is known as a zwitterion.



α-Amino acid drawn as a zwitterion

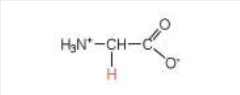
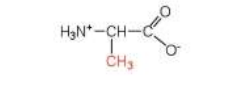
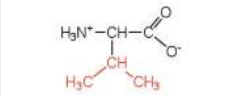
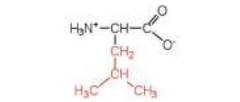
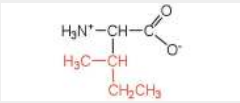
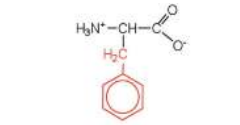
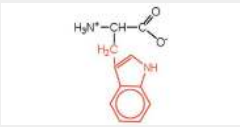
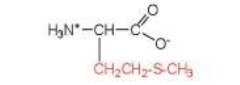
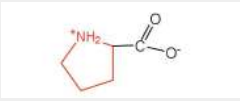
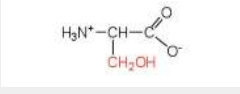
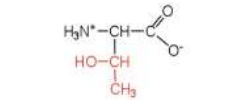
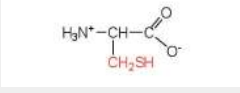


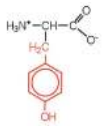
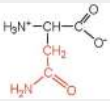
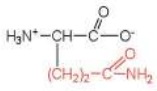
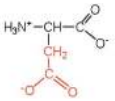
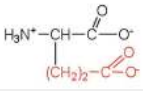
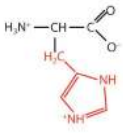
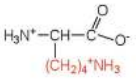
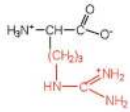
α-Amino acid drawn as an uncharged molecule; not an accurate respresentation of amino acid structure

Classification

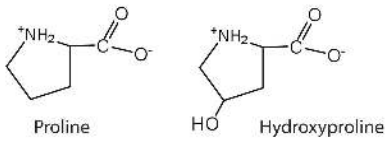
In addition to the amino and carboxyl groups, amino acids have a side chain or R group attached to the α-carbon. Each amino acid has unique characteristics arising from the size, shape, solubility, and ionization properties of its R group. As a result, the side chains of amino acids exert a profound effect on the structure and biological activity of proteins. Although amino acids can be classified in various ways, one common approach is to classify them according to whether the functional group on the side chain at neutral pH is nonpolar, polar but uncharged, negatively charged, or positively charged. The structures and names of the 20 amino acids, their one- and three-letter abbreviations, and some of their distinctive features are given in Table **9.2.1**.

Table **9.2.1**: Common Amino Acids Found in Proteins

Common Name	Abbreviation	Structural Formula (at pH 6)	Molar Mass	Distinctive Feature
Amino acids with a nonpolar R group				
glycine	gly (G)		75	the only amino acid lacking a chiral carbon
alanine	ala (A)		89	—
valine	val (V)		117	a branched-chain amino acid
leucine	leu (L)		131	a branched-chain amino acid
isoleucine	ile (I)		131	an essential amino acid because most animals cannot synthesize branched-chain amino acids
phenylalanine	phe (F)		165	also classified as an aromatic amino acid
tryptophan	trp (W)		204	also classified as an aromatic amino acid
methionine	met (M)		149	side chain functions as a methyl group donor
proline	pro (P)		115	contains a secondary amine group; referred to as an *α-imino acid*
Amino acids with a polar but neutral R group				
serine	ser (S)		105	found at the active site of many enzymes
threonine	thr (T)		119	named for its similarity to the sugar threose
cysteine	cys (C)		121	oxidation of two cysteine molecules yields *cystine*

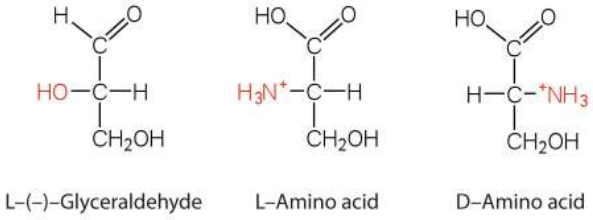
Common Name	Abbreviation	Structural Formula (at pH 6)	Molar Mass	Distinctive Feature
tyrosine	tyr (Y)		181	also classified as an aromatic amino acid
asparagine	asn (N)		132	the amide of aspartic acid
glutamine	gln (Q)		146	the amide of glutamic acid
Amino acids with a negatively charged R group				
aspartic acid	asp (D)		132	carboxyl groups are ionized at physiological pH; also known as aspartate
glutamic acid	glu (E)		146	carboxyl groups are ionized at physiological pH; also known as glutamate
Amino acids with a positively charged R group				
histidine	his (H)		155	the only amino acid whose R group has a pK_a (6.0) near physiological pH
lysine	lys (K)		147	—
arginine	arg (R)		175	almost as strong a base as sodium hydroxide

The first amino acid to be isolated was asparagine in 1806. It was obtained from protein found in asparagus juice (hence the name). Glycine, the major amino acid found in gelatin, was named for its sweet taste (Greek *glykys*, meaning “sweet”). In some cases an amino acid found in a protein is actually a derivative of one of the common 20 amino acids (one such derivative is hydroxyproline). The modification occurs *after* the amino acid has been assembled into a protein.



Configuration

Notice in Table 9.2.1 that glycine is the only amino acid whose α-carbon is *not* chiral. Therefore, with the exception of glycine, the amino acids could theoretically exist in either the D- or the L-enantiomeric form and rotate plane-polarized light. As with sugars, chemists used L-glyceraldehyde as the reference compound for the assignment of absolute configuration to amino acids. Its structure closely resembles an amino acid structure except that in the latter, an amino group takes the place of the OH group on the chiral carbon of the L-glyceraldehyde and a carboxylic acid replaces the aldehyde. Modern stereochemistry assignments using the Cahn-Ingold-Prelog priority rules used ubiquitously in chemistry show that all of the naturally occurring chiral amino acids are S except Cys which is R.



We learned that all naturally occurring sugars belong to the D series. It is interesting, therefore, that nearly all known plant and animal proteins are composed entirely of L-amino acids. However, certain bacteria contain D-amino acids in their cell walls, and several antibiotics (e.g., actinomycin D and the gramicidins) contain varying amounts of D-leucine, D-phenylalanine, and D-valine.

Summary

Amino acids can be classified based on the characteristics of their distinctive side chains as nonpolar, polar but uncharged, negatively charged, or positively charged. The amino acids found in proteins are L-amino acids.

9.2: Properties of Amino Acids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

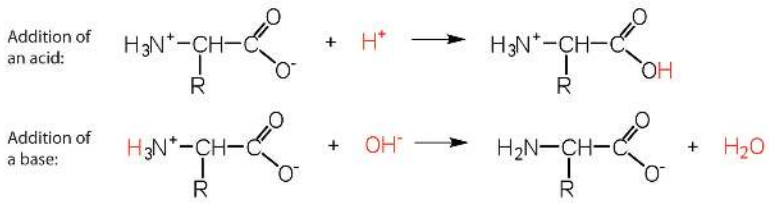
- **18.1: Properties of Amino Acids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.3: Reactions of Amino Acids

Learning Objectives

- To explain how an amino acid can act as both an acid and a base.

The structure of an amino acid allows it to act as both an acid and a base. An amino acid has this ability because at a certain pH value (different for each amino acid) nearly all the amino acid molecules exist as zwitterions. If acid is added to a solution containing the zwitterion, the carboxylate group captures a hydrogen (H^+) ion, and the amino acid becomes positively charged. If base is added, ion removal of the H^+ ion from the amino group of the zwitterion produces a negatively charged amino acid. In both circumstances, the amino acid acts to maintain the pH of the system—that is, to remove the added acid (H^+) or base (OH^-) from solution.

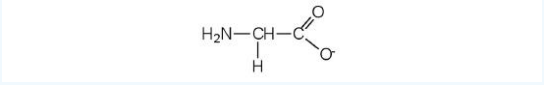


Example 9.3.1

a. Draw the structure for the anion formed when glycine (at neutral pH) reacts with a base.
b. Draw the structure for the cation formed when glycine (at neutral pH) reacts with an acid.

Solution

a. The base removes H^+ from the protonated amine group.



- The acid adds H^+ to the carboxylate group.



Exercise 9.3.1

a. Draw the structure for the cation formed when valine (at neutral pH) reacts with an acid.
b. Draw the structure for the anion formed when valine (at neutral pH) reacts with a base.

The particular pH at which a given amino acid exists in solution as a zwitterion is called the isoelectric point (pI). At its pI, the positive and negative charges on the amino acid balance, and the molecule as a whole is electrically neutral. The amino acids whose side chains are always neutral have isoelectric points ranging from 5.0 to 6.5. The basic amino acids (which have positively charged side chains at neutral pH) have relatively high examples. Acidic amino acids (which have negatively charged side chains at neutral pH) have quite low examples (Table 9.3.1).

Table 9.3.1: ExampIes of Some Representative Amino Acids

Amino Acid	Classification	pI
alanine	nonpolar	6.0
valine	nonpolar	6.0
serine	polar, uncharged	5.7
threonine	polar, uncharged	6.5
arginine	positively charged (basic)	10.8
histidine	positively charged (basic)	7.6
lysine	positively charged (basic)	9.8
aspartic acid	negatively charged (acidic)	3.0
glutamic acid	negatively charged (acidic)	3.2

Amino acids undergo reactions characteristic of carboxylic acids and amines. The reactivity of these functional groups is particularly important in linking amino acids together to form peptides and proteins, as you will see later in this chapter. Simple chemical tests that are used to detect amino acids take advantage of the reactivity of these functional groups. An example is the ninhydrin test in which the amine functional group of α-amino acids reacts with ninhydrin to form purple-colored compounds. Ninhydrin is used to detect fingerprints because it reacts with amino acids from the proteins in skin cells transferred to the surface by the individual leaving the fingerprint.



Ninhydrin



Summary

Amino acids can act as both an acid and a base due to the presence of the amino and carboxyl functional groups. The pH at which a given amino acid exists in solution as a zwitterion is called the *isoelectric point* (pI).

9.3: Reactions of Amino Acids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

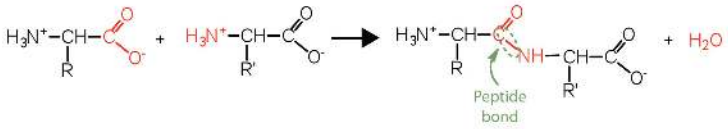
- **18.2: Reactions of Amino Acids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.4: Peptides

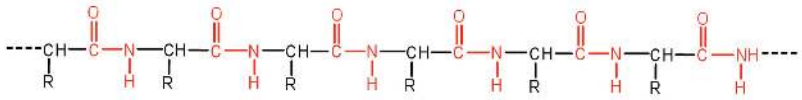
Learning Objectives

- Explain how a peptide is formed from individual amino acids.
- Explain why the sequence of amino acids in a protein is important.

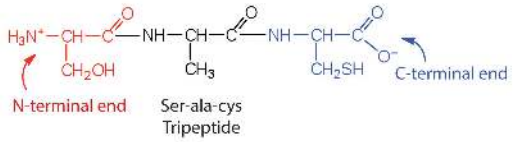
Two or more amino acids can join together into chains called peptides. Previously, we discussed the reaction between ammonia and a carboxylic acid to form an amide. In a similar reaction, the amino group on one amino acid molecule reacts with the carboxyl group on another, releasing a molecule of water and forming an amide linkage:



An amide bond joining two amino acid units is called a peptide bond. Note that the product molecule still has a reactive amino group on the left and a reactive carboxyl group on the right. These can react with additional amino acids to lengthen the peptide. The process can continue until thousands of units have joined, resulting in large proteins.



A chain consisting of only two amino acid units is called a *dipeptide*; a chain consisting of three is a *tripeptide*. By convention, peptide and protein structures are depicted with the amino acid whose amino group is free (the N-terminal end) on the left and the amino acid with a free carboxyl group (the C-terminal end) to the right.



The general term *peptide* refers to an amino acid chain of unspecified length. However, chains of about 50 amino acids or more are usually called proteins or polypeptides. In its physiologically active form, a protein may be composed of one or more polypeptide chains.

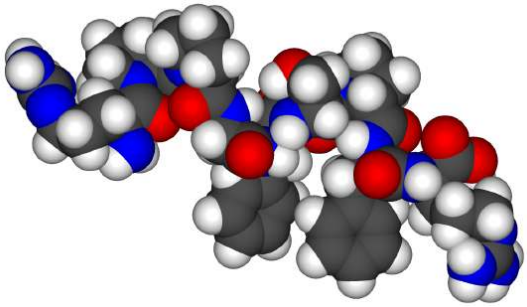


Figure **9.4.1**: Space-filling model of bradykinin. (Public Domain; Fvasconcellos)

For peptides and proteins to be physiologically active, it is not enough that they incorporate certain amounts of specific amino acids. The order, or *sequence*, in which the amino acids are connected is also of critical importance. Bradykinin is a nine-amino acid peptide (Figure **9.4.1**) produced in the blood that has the following amino acid sequence:

arg-pro-pro-gly-phe-ser-pro-phe-arg

This peptide lowers blood pressure, stimulates smooth muscle tissue, increases capillary permeability, and causes pain. When the order of amino acids in bradykinin is reversed,

arg-phe-pro-ser-phe-gly-pro-pro-arg

the peptide resulting from this synthesis shows none of the activity of bradykinin.

Just as millions of different words are spelled with our 26-letter English alphabet, millions of different proteins are made with the 20 common amino acids. However, just as the English alphabet can be used to write gibberish, amino acids can be put together in the *wrong sequence* to produce nonfunctional proteins. Although the correct sequence is ordinarily of utmost importance, it is not always absolutely required. Just as you can sometimes make sense of incorrectly spelled English words, a protein with a small percentage of “incorrect” amino acids may continue to function. However, it rarely functions as well as a protein having the correct sequence. There are also instances in which seemingly minor errors of sequence have disastrous effects. For example, in some people, every molecule of hemoglobin (a protein in the blood that transports oxygen) has a single incorrect amino acid unit out of about 300 (a single valine replaces a glutamic acid). That “minor” error is responsible for sickle cell anemia, an inherited condition that usually is fatal.

Summary

The amino group of one amino acid can react with the carboxyl group on another amino acid to form a peptide bond that links the two amino acids together. Additional amino acids can be added on through the formation of addition peptide (amide) bonds. A sequence of amino acids in a peptide or protein is written with the N-terminal amino acid first and the C-terminal amino acid at the end (writing left to right).

9.4: Peptides is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.3: Peptides** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.5: Proteins

Learning Objectives

- Describe the four levels of protein structure.
- Identify the types of attractive interactions that hold proteins in their most stable three-dimensional structure.
- Explain what happens when proteins are denatured.
- Identify how a protein can be denatured.

Each of the thousands of naturally occurring proteins has its own characteristic amino acid composition and sequence that result in a unique three-dimensional shape. Since the 1950s, scientists have determined the amino acid sequences and three-dimensional conformation of numerous proteins and thus obtained important clues on how each protein performs its specific function in the body.

Proteins are compounds of high molar mass consisting largely or entirely of chains of amino acids. Because of their great complexity, protein molecules cannot be classified on the basis of specific structural similarities, as carbohydrates and lipids are categorized. The two major structural classifications of proteins are based on far more general qualities: whether the protein is (1) fiberlike and insoluble or (2) globular and soluble. Some proteins, such as those that compose hair, skin, muscles, and connective tissue, are fiberlike. These fibrous proteins are insoluble in water and usually serve structural, connective, and protective functions. Examples of fibrous proteins are keratins, collagens, myosins, and elastins. Hair and the outer layer of skin are composed of keratin. Connective tissues contain collagen. Myosins are muscle proteins and are capable of contraction and extension. Elastins are found in ligaments and the elastic tissue of artery walls.

Globular proteins, the other major class, are soluble in aqueous media. In these proteins, the chains are folded so that the molecule as a whole is roughly spherical. Familiar examples include egg albumin from egg whites and serum albumin in blood. Serum albumin plays a major role in transporting fatty acids and maintaining a proper balance of osmotic pressures in the body. Hemoglobin and myoglobin, which are important for binding oxygen, are also globular proteins.

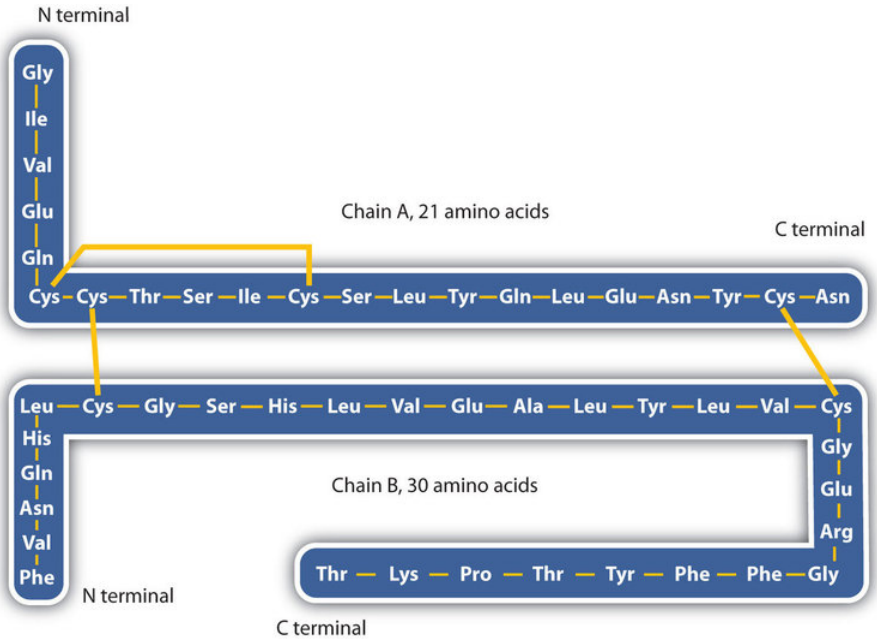


Figure **9.5.1**: Primary Structure of Human Insulin. Human insulin, whose amino acid sequence is shown here, is a hormone that is required for the proper metabolism of glucose.

Levels of Protein Structure

The structure of proteins is generally described as having four organizational levels. The first of these is the primary structure, which is the number and sequence of amino acids in a protein's polypeptide chain or chains, beginning with the free amino group

and maintained by the peptide bonds connecting each amino acid to the next. The primary structure of insulin, composed of 51 amino acids, is shown in Figure **9.5.1**.

A protein molecule is not a random tangle of polypeptide chains. Instead, the chains are arranged in unique but specific conformations. The term secondary structure refers to the fixed arrangement of the polypeptide backbone. On the basis of X ray studies, Linus Pauling and Robert Corey postulated that certain proteins or portions of proteins twist into a spiral or a helix. This helix is stabilized by *intrachain* hydrogen bonding between the carbonyl oxygen atom of one amino acid and the amide hydrogen atom four amino acids up the chain (located on the next turn of the helix) and is known as a right-handed α-helix. X ray data indicate that this helix makes one turn for every 3.6 amino acids, and the side chains of these amino acids project outward from the coiled backbone (Figure **9.5.2**). The α-keratins, found in hair and wool, are exclusively α-helical in conformation. Some proteins, such as gamma globulin, chymotrypsin, and cytochrome c, have little or no helical structure. Others, such as hemoglobin and myoglobin, are helical in certain regions but not in others.

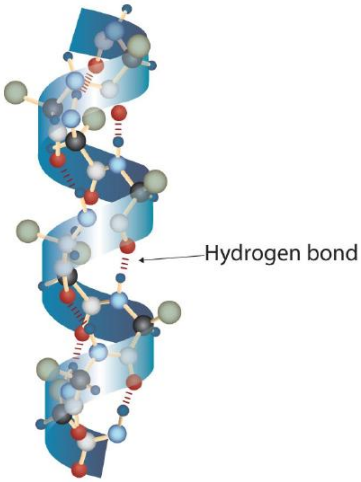


Figure **9.5.2** A Ball-and-Stick Model of an α-Helix. This ball-and-stick model shows the intrachain hydrogen bonding between carbonyl oxygen atoms and amide hydrogen atoms. Each turn of the helix spans 3.6 amino acids. Note that the side chains (represented as green spheres) point out from the helix.

Another common type of secondary structure, called the *β-pleated sheet conformation*, is a sheetlike arrangement in which two or more extended polypeptide chains (or separate regions on the same chain) are aligned side by side. The aligned segments can run either parallel or antiparallel—that is, the N-terminals can face in the same direction on adjacent chains or in different directions—and are connected by *interchain* hydrogen bonding (Figure **9.5.3**). The β-pleated sheet is particularly important in structural proteins, such as silk fibroin. It is also seen in portions of many enzymes, such as carboxypeptidase A and lysozyme.

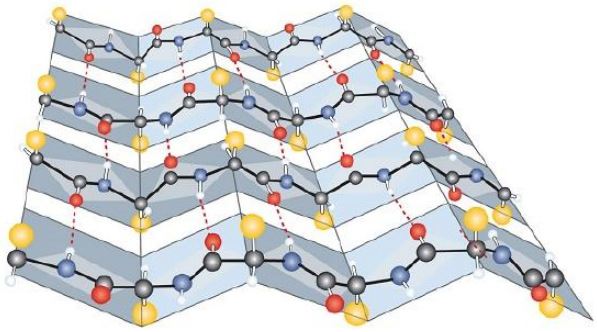


Figure **9.5.3**: A Ball-and-Stick Model of the β-Pleated Sheet Structure in Proteins. The side chains extend above or below the sheet and alternate along the chain. The protein chains are held together by interchain hydrogen bonding.

Tertiary structure refers to the unique three-dimensional shape of the protein as a whole, which results from the folding and bending of the protein backbone. The tertiary structure is intimately tied to the proper biochemical functioning of the protein. Figure **9.5.4** shows a depiction of the three-dimensional structure of insulin.

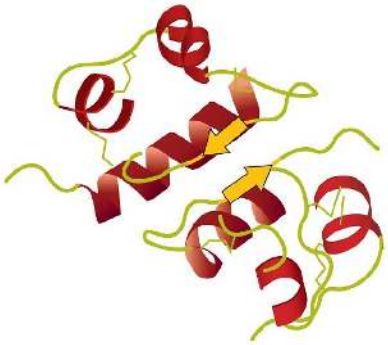
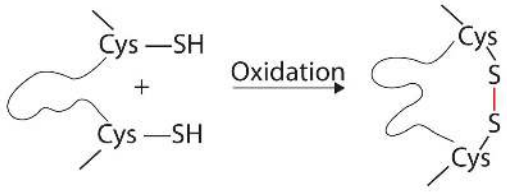


Figure 9.5.4: A Ribbon Model of the Three-Dimensional Structure of Insulin. The spiral regions represent sections of the polypeptide chain that have an α-helical structure, while the broad arrows represent β-pleated sheet structures.

Four major types of attractive interactions determine the shape and stability of the tertiary structure of proteins. You studied several of them previously.

1. Ionic bonding. Ionic bonds result from electrostatic attractions between positively and negatively charged side chains of amino acids. For example, the mutual attraction between an aspartic acid carboxylate ion and a lysine ammonium ion helps to maintain a particular folded area of a protein (part (a) of Figure 9.5.5).
2. Hydrogen bonding. Hydrogen bonding forms between a highly electronegative oxygen atom or a nitrogen atom and a hydrogen atom attached to another oxygen atom or a nitrogen atom, such as those found in polar amino acid side chains. Hydrogen bonding (as well as ionic attractions) is extremely important in both the intra- and intermolecular interactions of proteins (part (b) of Figure 9.5.5).
3. Disulfide linkages. Two cysteine amino acid units may be brought close together as the protein molecule folds. Subsequent oxidation and linkage of the sulfur atoms in the highly reactive sulfhydryl (SH) groups leads to the formation of cystine (part (c) of Figure 9.5.5). Intrachain disulfide linkages are found in many proteins, including insulin (yellow bars in Figure 9.5.1) and have a strong stabilizing effect on the tertiary structure.



4. Dispersion forces. Dispersion forces arise when a normally nonpolar atom becomes momentarily polar due to an uneven distribution of electrons, leading to an instantaneous dipole that induces a shift of electrons in a neighboring nonpolar atom. Dispersion forces are weak but can be important when other types of interactions are either missing or minimal (part (d) of Figure 9.5.5). This is the case with fibroin, the major protein in silk, in which a high proportion of amino acids in the protein have nonpolar side chains. The term *hydrophobic interaction* is often misused as a synonym for dispersion forces. Hydrophobic interactions arise because water molecules engage in hydrogen bonding with other water molecules (or groups in proteins capable of hydrogen bonding). Because nonpolar groups cannot engage in hydrogen bonding, the protein folds in such a way that these groups are buried in the interior part of the protein structure, minimizing their contact with water.

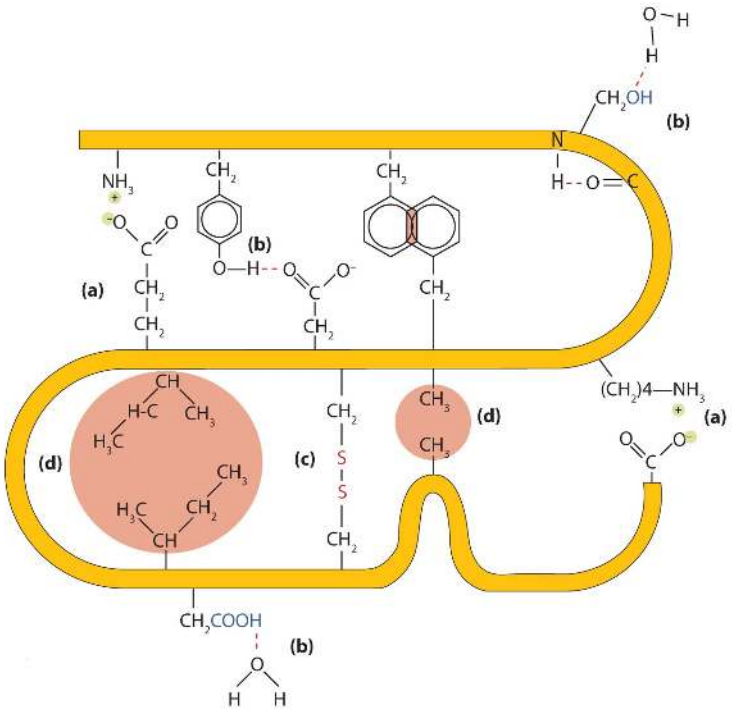


Figure 9.5.5: Tertiary Protein Structure Interactions. Four interactions stabilize the tertiary structure of a protein: (a) ionic bonding, (b) hydrogen bonding, (c) disulfide linkages, and (d) dispersion forces.

When a protein contains more than one polypeptide chain, each chain is called a *subunit*. The arrangement of multiple subunits represents a fourth level of structure, the quaternary structure of a protein. Hemoglobin, with four polypeptide chains or subunits, is the most frequently cited example of a protein having quaternary structure (Figure 9.5.6). The quaternary structure of a protein is produced and stabilized by the same kinds of interactions that produce and maintain the tertiary structure. A schematic representation of the four levels of protein structure is in Figure 9.5.7.

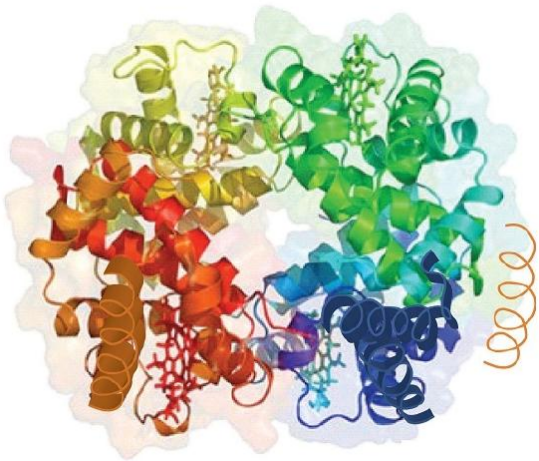


Figure 9.5.6 The Quaternary Structure of Hemoglobin. Hemoglobin is a protein that transports oxygen throughout the body.

Source: Image from the RCSB PDB (www.pdb.org(opens in new window)) of PDB ID 1I3D (R.D. Kidd, H.M. Baker, A.J. Mathews, T. Brittain, E.N. Baker (2001) Oligomerization and ligand binding in a homotetrameric hemoglobin: two high-resolution crystal structures of hemoglobin Bart's (gamma(4)), a marker for alpha-thalassemia. Protein Sci. 1739–1749).

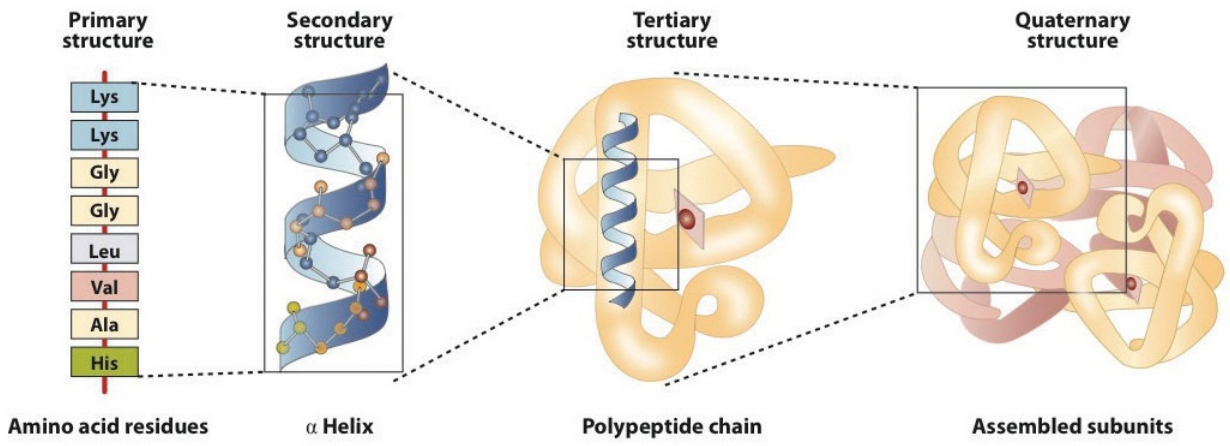


Figure **9.5.7**: Levels of Structure in Proteins

The *primary structure* consists of the specific amino acid sequence. The resulting peptide chain can twist into an α-helix, which is one type of *secondary structure*. This helical segment is incorporated into the *tertiary structure* of the folded polypeptide chain. The single polypeptide chain is a subunit that constitutes the *quaternary structure* of a protein, such as hemoglobin that has four polypeptide chains.

Denaturation of Proteins

The highly organized structures of proteins are truly masterworks of chemical architecture. But highly organized structures tend to have a certain delicacy, and this is true of proteins. Denaturation is the term used for any change in the three-dimensional structure of a protein that renders it incapable of performing its assigned function. A denatured protein cannot do its job. (Sometimes denaturation is equated with the precipitation or coagulation of a protein; our definition is a bit broader.) A wide variety of reagents and conditions, such as heat, organic compounds, pH changes, and heavy metal ions can cause protein denaturation (Figure **9.5.1**).

Figure **9.5.1**: Protein Denaturation Methods

Method	Effect on Protein Structure
Heat above 50°C or ultraviolet (UV) radiation	Heat or UV radiation supplies kinetic energy to protein molecules, causing their atoms to vibrate more rapidly and disrupting relatively weak hydrogen bonding and dispersion forces.
Use of organic compounds, such as ethyl alcohol	These compounds are capable of engaging in intermolecular hydrogen bonding with protein molecules, disrupting intramolecular hydrogen bonding within the protein.
Salts of heavy metal ions, such as mercury, silver, and lead	These ions form strong bonds with the carboxylate anions of the acidic amino acids or SH groups of cysteine, disrupting ionic bonds and disulfide linkages.
Alkaloid reagents, such as tannic acid (used in tanning leather)	These reagents combine with positively charged amino groups in proteins to disrupt ionic bonds.

Anyone who has fried an egg has observed denaturation. The clear egg white turns opaque as the albumin denatures and coagulates. No one has yet reversed that process. However, given the proper circumstances and enough time, a protein that has unfolded under sufficiently gentle conditions can refold and may again exhibit biological activity (Figure **9.5.8**). Such evidence suggests that, at least for these proteins, the primary structure determines the secondary and tertiary structure. A given sequence of amino acids seems to adopt its particular three-dimensional arrangement naturally if conditions are right.

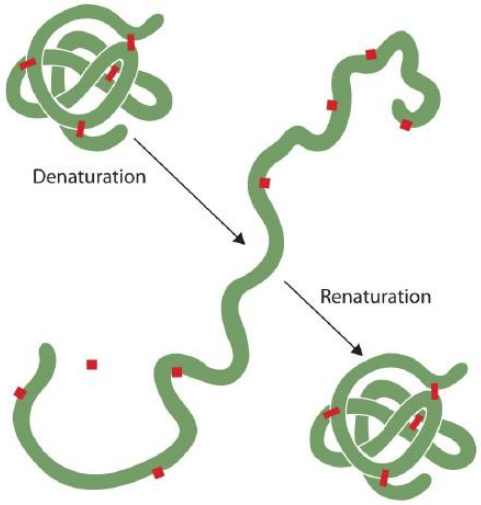


Figure 9.5.8: Denaturation and Renaturation of a Protein. The denaturation (unfolding) and renaturation (refolding) of a protein is depicted. The red boxes represent stabilizing interactions, such as disulfide linkages, hydrogen bonding, and/or ionic bonds.

The primary structures of proteins are quite sturdy. In general, fairly vigorous conditions are needed to hydrolyze peptide bonds. At the secondary through quaternary levels, however, proteins are quite vulnerable to attack, though they vary in their vulnerability to denaturation. The delicately folded globular proteins are much easier to denature than are the tough, fibrous proteins of hair and skin.

Summary

Proteins can be divided into two categories: fibrous, which tend to be insoluble in water, and globular, which are more soluble in water. A protein may have up to four levels of structure. The primary structure consists of the specific amino acid sequence. The resulting peptide chain can form an α-helix or β-pleated sheet (or local structures not as easily categorized), which is known as secondary structure. These segments of secondary structure are incorporated into the tertiary structure of the folded polypeptide chain. The quaternary structure describes the arrangements of subunits in a protein that contains more than one subunit. Four major types of attractive interactions determine the shape and stability of the folded protein: ionic bonding, hydrogen bonding, disulfide linkages, and dispersion forces. A wide variety of reagents and conditions can cause a protein to unfold or denature.

9.5: Proteins is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.4: Proteins** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.6: X-ray Protein Crystallography

X-ray protein crystallography is a technique by which it is possible to determine the three dimensional positions of each atom in a protein. Now over 100 years old, x-ray crystallography was first used to determine the three dimensional structures of inorganic materials, then small organic molecules, and finally macromolecules like DNA and proteins. To date, about 100,000 protein structures have been published in the Protein Data Bank, with almost 10,000 added every year. To use this technique, the crystallographer obtains protein crystals, records the diffraction pattern formed by x-rays passed through the crystals, and then interprets the data using a computer. The result is a atomic-resolution model of a protein.

History

Though crystal symmetry was explored in the late 1600s by Danish scientist Nicolas Steno, and continuing efforts by René Just Haüy and William Hallowes Miller in 1839 firmly established that a crystal is a ordered lattice, it wasn't until the discovery of x-rays in 1895 and the proof of their diffraction by Max von Laue in 1912 that crystallography as a science began.

After the use of x-ray crystallography to deduce the lattice structure of table salt in 1914, the father and son team of William Henry Bragg and William Lawrence Bragg shared the 1915 Nobel Prize in Physics for the development of Bragg's law,

$$n\lambda = 2d\sin(\theta), \tag{9.6.1}$$

which relates an x-ray diffraction pattern with the three-dimensional structure of a crystal.

The field has received numerous Nobel Prizes over the years, including in chemistry in 1964 to Dorothy Crowfoot Hodgkin, who solved the structures of the small molecules cholesterol, penicillin, and vitamin B12, and in chemistry in 1962 to Max Perutz and John Cowdery Kendrew for their work on sperm whale myoglobin. David Chilton Phillips solved the first structure of an enzyme, lysozyme, in 1965.

The early 70s saw the birth of the Research Collaboratory for Structural Bioinformatics' Protein Data Bank. The PDB began with 13 structures in 1976 and has grown to the "single worldwide archive of structural data of biological macromolecules".

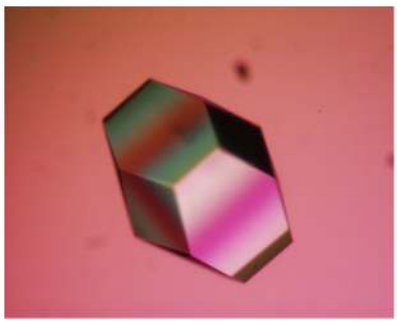
Technique

Obtaining crystals

The first and least certain step in crystallography of a protein is obtaining crystals of the protein of interest. Obtaining suitable amounts of the protein of interest is usually carried out in a straightforward manner using established molecular biology techniques such as molecular cloning and affinity chromatography. However, the crystallization step remains the bottleneck for this technique, with some proteins (particularly proteins that exist in the aliphatic environment of the plasma membrane) remaining intransigent to crystallization even in the face of the most diligent crystallographers. Thus, for each protein of interest, a large number of crystallization conditions must be tried, necessitating a relatively large amount (milligrams) of the pure protein.

Protein production and purification

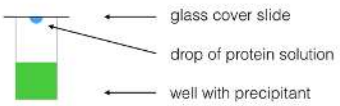
To produce suitable amounts of protein, contemporary crystallographers turn to molecular biology's old friend *Escherichia coli*. A gene which codes for the protein of interest is cloned into a small, circular piece of DNA known as an expression plasmid. The expression of the gene is typically under the control of an inducible promoter, and is regulated by the researcher rather than the bacteria. Cells are transformed with the expression plasmid, grown to high density, and induced to express the protein of interest. The cells are lysed chemically with detergents or physically with sonication, and the protein is purified, typically via affinity chromatography. High purity (greater than 95%) is desirable. Often, it takes multiple experiments before the method that obtains maximum protein is found.



Crystallization

The concentrated protein solution obtained is then subjected to a wide variety of crystallization conditions. Since we have no way of knowing *a priori* which set of conditions is right for obtaining crystals of a given protein, many different conditions are tried in parallel using a technique called drop diffusion.

In this technique, a small quantity (typically a microliter) of concentrated protein solution is mixed with an equal volume of precipitant. This drop is separated by air from a large volume of precipitant solution. The drop is hypotonic to the precipitant and slowly equilibrates to the concentration of the large volume of precipitant. Concomitantly, the concentration of protein increases. If this process occurs at just the right rate, the protein precipitates out of solution into an ordered lattice structure: a protein crystal.



It is often said that this part of crystallography is more of an art than a science, and indeed there is little theoretical guidance available to the crystallographer who wishes to crystallize a new protein. Patience, and to some extent, luck, determine the sucess or failure of the crystallization of any particular protein.

Obtaining x-ray diffraction data

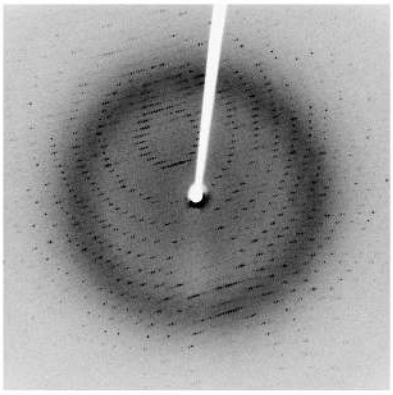
Once crystals of suitable size and composition are obtained, it is necessary to bombard the crystal with x-rays and observe the diffraction pattern. An x-ray diffractometer works in a similar manner to a light microscope. In a light microscope, the subject is irradiated with visible light (400 nm < λ < 700 nm$), which is diffracted by a lens onto the retina, producing a macroscopic image of a microscopic object. Molecules such as proteins are much smaller than microscopic structures like cells, and, as such, require that a shorter wavelength of radiation be used during diffraction. X-rays, where $100 pm < λ < 10,000 pm$, are the perfect size to diffract around atoms (32--225 pm), bonds (74--267 pm), and molecules (100 pm to hundreds of Angstroms). However, x-rays are difficult to focus in a manner analogous to the way a lens focuses visible light. Crystallographers employ computational methods to capture the x-ray scattering pattern (pictured at right) and infer the three-dimensional positions of atoms in a molecule.

X-ray sources

Traditionally, x-ray crystallographers filtered and directed the x-rays generated by radioactive cesium in their diffractometers, but today it is much more common to use synchrotron radiation to irradiate samples. Synchrotrons, huge hollow rings used to accelerate electrons for use in studies of subatomic particles, produce huge amounts of tunable (different wavelengths) x-ray radiation that is perfect for irradiating crystals.

Sample preparation

The crystal is suspended in aqueous solution containing a cryoprotectant in the eye of a small loop. The crystal and loop are cooled with a continuous stream of liquid nitrogen to prevent chemical damage by the x-rays. X-rays are directed through the crystal, and the diffraction pattern at any given moment is recorded by a detector. The crystal is rotated sightly and a new diffraction pattern is obtained. This process is repeated through 360 degrees along one axis (typically rotations through a smaller angle on another axis are also recorded to avoid blind spots) until the instrument has recorded a diffraction pattern for each position.



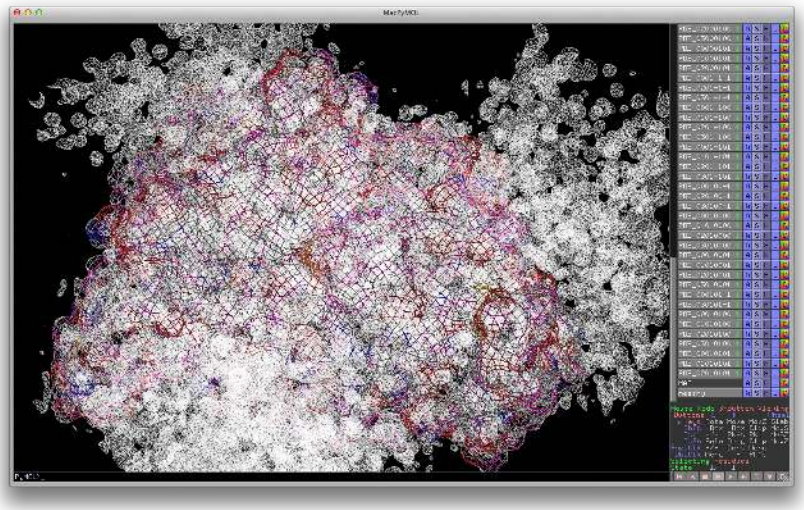
X-ray scattering

As an incident x-ray (electromagnetic wave) overlaps with an electron, it is elastically scattered, generating a secondary wave that has the same wavelength, but different direction, than the incident wave (thus the wave is "scattered" or "diffracted"). Due to the symmetry of the crystal and its many repeated units, these secondary waves interfere constructively at only one point along a circle drawn around the atom that scattered them. It is that point, described by Bragg's Law, that appears as a dark spot on the detector. An example diffraction pattern, from a SARS protease, is displayed at right.

Obtaining an electron density map

The data recorded by the detector during diffraction are now subjected to computational analysis. First, each spot in each diffraction image is indexed, integrated, merged, and scaled by a computer, producing a single text file from thousands of images. The position of each spot depends on the properties of the crystal, and as such is different for every protein. The process of converting the reciprocal space-representation of the crystal into an interpretable electron density map is known as phasing.

Shown below is the software PyMOL displaying the electron density map (white) for Protein Data Bank structure 4BLL, a peroxidase from the model organism *Pleurotus ostreatus*, overlayed with the model from the PDB structure (pink).



Fundamental challenge of crystallography

The problem is this: the detector is only able to record the position and intensity of an x-ray when it hits the detector. An x-ray has both intensity, which is related to the amplitude of the wave, and phase, which is related to the point at which the x-ray was scattered. The crystallographer would dearly like to know the phase of an x-ray, because this information along with with intensity, but the detector is not capable of capturing phase information due the quantum mechanical nature of x-rays and electrons.

Overcoming the challenge: phasing

Crystallographers use several methods for recovering phase information from diffraction data. Common techniques include:

- direct methods
- molecular replacement
- anomalous x-ray scattering
- multiple isomorphous replacement

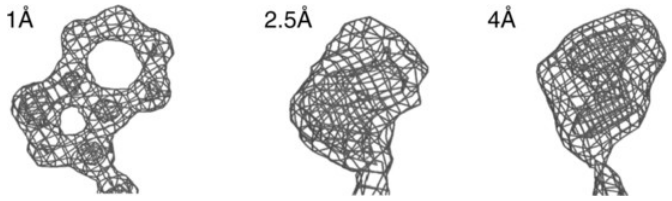
Direct methods use the Sayre equation to determine phases directly from the diffraction data. These methods are only viable for small (less than 1000 atoms) molecules and are not typically used in protein crystallography. The 1985 Nobel Prize in chemistry was awarded to Hauptman and Karle, who developed these methods.

The technique of molecular replacement uses the solved crystal structure of a homologous protein to provide a "seed" electron density map that can then be refined by a computer. Molecular replacement is used extensively in labs that solve the crystal structures of several mutants of a given protein.

Anomalous x-ray scattering relies on protein production in a host that is incapable of producing the amino acid methionine. The host instead uses a synthetic amino acid, selenomethionine, in which methionene's sulfur atom has been replaced by selenium. The positions of any selenomethionines in the diffraction pattern can be solved using different x-ray wavelengths and direct methods, and the rest of the structure can be solved using the position of the selenomethionines as a reference.

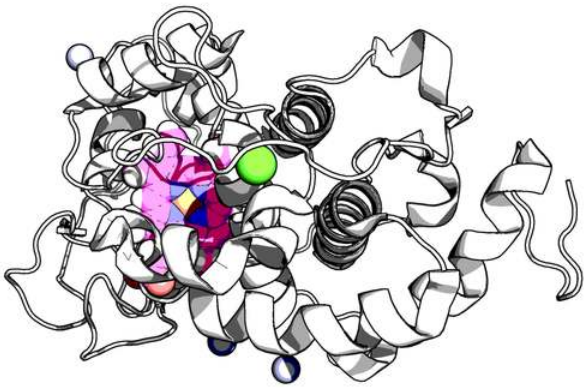
Multiple isomorphous replacement has largely been superseded by anomalous x-ray scattering and works in a similar way, except using metal ions instead of synthetic amino acids for initial phasing.

Once phases have been recovered, it is possible to mathematically reconstruct the positions of electrons within the crystal using Fourier synthesis. (The diffraction data is the Fourier transform of the electron density in the unit cell.) A computer applying these operations with correct phases constructs a three-dimensional map of electron density map than can be viewed with molecular visualization software. The resolution of the data determines the resolution of the model, as depicted below with electron density maps of tryptophan at three different resolutions.



Obtaining a three-dimensional model

With sufficient resolution (less than 1.5 A), it is possible to automatically generate a model based on the electron density map and known bond angles and lengths, and known sizes of atoms. In practice, not all crystallography data is of such high quality. Often, the crystallographer uses molecular visualization software to manually fit a chemical model to the electron density data. The result is a model that can be viewed with molecular visualization software. An example, draw in PyMOL from PDB structure 4BLL, is below.



Further reading

1. Bragg WL (1914). "The analysis of crystals by the X-ray spectrometer". Proc. R. Soc. Lond. A89 (613): 468
2. Glusker JP and Trueblood KN, 1972. *Crystal structure analysis: a primer*. Oxford University Press. [Reprint: OUP Oxford, May 27, 2010]
3. Drenth J. Principles of Protein X-Ray Crystallography. Springer, Apr 5, 2007
4. Hauptman H, 1997. "Phasing methods for protein crystallography". Curr. Opin. Struct. Biol. 7 (5): 672–80
5. Chernov AA (2003). "Protein crystals and their growth". J. Struct. Biol. 142 (1): 3–21

9.6: X-ray Protein Crystallography is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

- **X-ray Protein Crystallography** has no license indicated. Original source: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-013-electromagnetics-and-applications-spring-2009.

9.7: Enzymes

Learning Objectives

- Explain the functions of enzymes.
- Explain how enzymes are classified and named.

A catalyst is any substance that increases the *rate* or speed of a chemical reaction without being changed or consumed in the reaction. Enzymes are biological catalysts, and nearly all of them are proteins. The reaction rates attained by enzymes are truly amazing. In their presence, reactions occur at rates that are a million (10^6) or more times faster than would be attainable in their absence. What is even more amazing is that enzymes perform this function at body temperature (~37°C) and physiological pH (pH ~7), rather than at the conditions that are typically necessary to increase reaction rates (high temperature or pressure, the use of strong oxidizing or reducing agents or strong acids or bases, or a combination of any of these). In addition, enzymes are highly specific in their action; that is, each enzyme catalyzes only one type of reaction in only one compound or a group of structurally related compounds. The compound or compounds on which an enzyme acts are known as its substrates.

Hundreds of enzymes have been purified and studied in an effort to understand how they work so effectively and with such specificity. The resulting knowledge has been used to design drugs that inhibit or activate particular enzymes. An example is the intensive research to improve the treatment of or find a cure for acquired immunodeficiency syndrome (AIDS). AIDS is caused by the human immunodeficiency virus (HIV). Researchers are studying the enzymes produced by this virus and are developing drugs intended to block the action of those enzymes without interfering with enzymes produced by the human body. Several of these drugs have now been approved for use by AIDS patients.

Table **9.7.1**: Classes of Enzymes

Class	Type of Reaction Catalyzed	Examples
oxidoreductases	oxidation-reduction reactions	Dehydrogenases catalyze oxidation-reduction reactions involving hydrogen and reductases catalyze reactions in which a substrate is reduced.
transferases	transfer reactions of groups, such as methyl, amino, and acetyl	Transaminases catalyze the transfer of amino group, and kinases catalyze the transfer of a phosphate group.
hydrolases	hydrolysis reactions	Lipases catalyze the hydrolysis of lipids, and proteases catalyze the hydrolysis of proteins
lyases	reactions in which groups are removed without hydrolysis or addition of groups to a double bond	Decarboxylases catalyze the removal of carboxyl groups.
isomerases	reactions in which a compound is converted to its isomer	Isomerases may catalyze the conversion of an aldose to a ketose, and mutases catalyze reactions in which a functional group is transferred from one atom in a substrate to another.
ligases	reactions in which new bonds are formed between carbon and another atom; energy is required	Synthetases catalyze reactions in which two smaller molecules are linked to form a larger one.

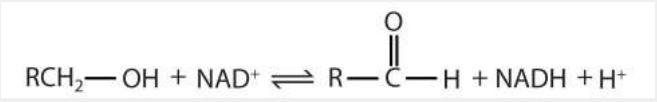
The first enzymes to be discovered were named according to their source or method of discovery. The enzyme *pepsin*, which aids in the hydrolysis of proteins, is found in the digestive juices of the stomach (Greek *pepsis*, meaning “digestion”). *Papain*, another

enzyme that hydrolyzes protein (in fact, it is used in meat tenderizers), is isolated from papayas. As more enzymes were discovered, chemists recognized the need for a more systematic and chemically informative identification scheme. In the current numbering and naming scheme, under the oversight of the Nomenclature Commission of the International Union of Biochemistry, enzymes are arranged into six groups according to the general type of reaction they catalyze (Table **9.7.1**), with subgroups and secondary subgroups that specify the reaction more precisely.

Figure* 9.7.1*: *Structure of the alcohol dehydrogenase protein (E.C.1.1.1.1) (EE ISOZYME) complexed wtih nicotinamide adenini dinulceotide (NAD) and zinc (PDB: 1CDO).*

Each enzyme is assigned a four-digit number, preceded by the prefix EC—for enzyme classification—that indicates its group, subgroup, and so forth. This is demonstrated in Table **9.7.2** for alcohol dehydrogenase. Each enzyme is also given a name consisting of the root of the name of its substrate or substrates and the *-ase* suffix. Thus urease is the enzyme that catalyzes the hydrolysis of urea.

Table **9.7.2**: Assignment of an Enzyme Classification Number

Alcohol Dehydrogenase: EC 1.1.1.1	
The first digit indicates that this enzyme is an oxidoreductase; that is, an enzyme that catalyzes an oxidation-reduction reaction. The second digit indicates that this oxidoreductase catalyzes a reaction involving a primary or secondary alcohol. The third digit indicates that either the coenzyme NAD^+ or $NADP^+$ is required for this reaction.	
The fourth digit indicates that this was the first enzyme isolated, characterized, and named using this system of nomenclature.	
The systematic name for this enzyme is *alcohol:NAD^+ oxidoreductase*, while the recommended or common name is alcohol dehydrogenase.	
Reaction catalyzed:	

Summary

An enzyme is a biological catalyst, a substance that increases the rate of a chemical reaction without being changed or consumed in the reaction. A systematic process is used to name and classify enzymes.

9.7: Enzymes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.5: Enzymes** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.8: Enzyme Action

Learning Objectives

- To describe the interaction between an enzyme and its substrate.

Enzyme-catalyzed reactions occur in at least two steps. In the first step, an enzyme molecule (E) and the substrate molecule or molecules (S) collide and react to form an intermediate compound called the *enzyme-substrate* (E–S) *complex*. (This step is reversible because the complex can break apart into the original substrate or substrates and the free enzyme.) Once the E–S complex forms, the enzyme is able to catalyze the formation of product (P), which is then released from the enzyme surface:

$$S + E \rightarrow E\text{–}S \tag{9.8.1}$$

$$E\text{–}S \rightarrow P + E \tag{9.8.2}$$

Hydrogen bonding and other electrostatic interactions hold the enzyme and substrate together in the complex. The structural features or functional groups on the enzyme that participate in these interactions are located in a cleft or pocket on the enzyme surface. This pocket, where the enzyme combines with the substrate and transforms the substrate to product is called the active site of the enzyme (Figure 9.8.1).

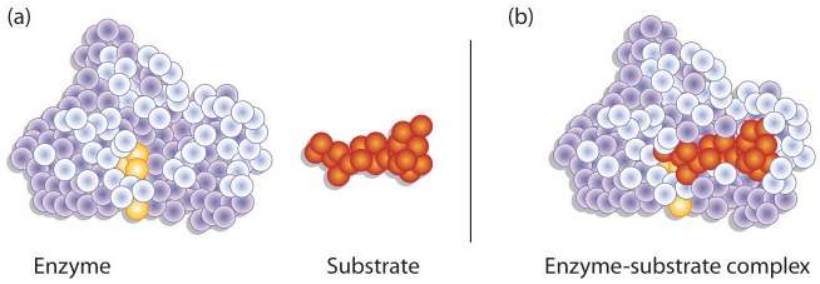


Figure 9.8.1: Substrate Binding to the Active Site of an Enzyme. The enzyme dihydrofolate reductase is shown with one of its substrates: $NADP^+$ (a) unbound and (b) bound. The $NADP^+$ (shown in red) binds to a pocket that is complementary to it in shape and ionic properties.

The active site of an enzyme possesses a unique conformation (including correctly positioned bonding groups) that is complementary to the structure of the substrate, so that the enzyme and substrate molecules fit together in much the same manner as a key fits into a tumbler lock. In fact, an early model describing the formation of the enzyme-substrate complex was called the lock-and-key model (Figure 9.8.2). This model portrayed the enzyme as conformationally rigid and able to bond only to substrates that exactly fit the active site.

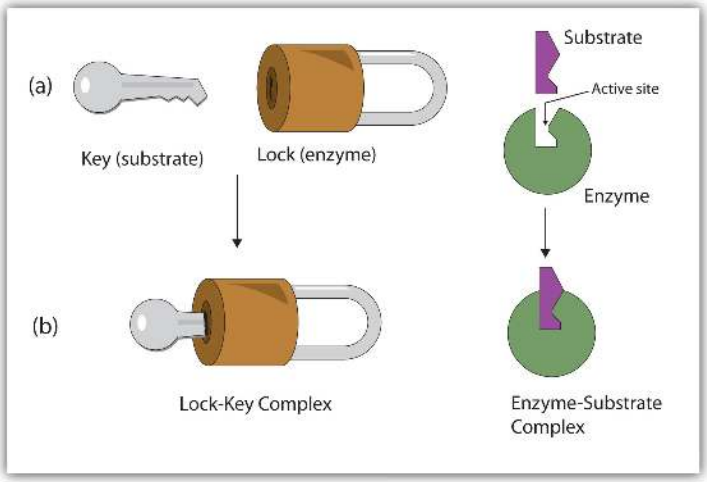


Figure 9.8.2: The Lock-and-Key Model of Enzyme Action. (a) Because the substrate and the active site of the enzyme have complementary structures and bonding groups, they fit together as a key fits a lock. (b) The catalytic reaction occurs while the two are bonded together in the enzyme-substrate complex.

Working out the precise three-dimensional structures of numerous enzymes has enabled chemists to refine the original lock-and-key model of enzyme actions. They discovered that the binding of a substrate often leads to a large conformational change in the enzyme, as well as to changes in the structure of the substrate or substrates. The current theory, known as the induced-fit model, says that enzymes can undergo a change in conformation when they bind substrate molecules, and the active site has a shape complementary to that of the substrate only *after* the substrate is bound, as shown for hexokinase in Figure 9.8.3. After catalysis, the enzyme resumes its original structure.

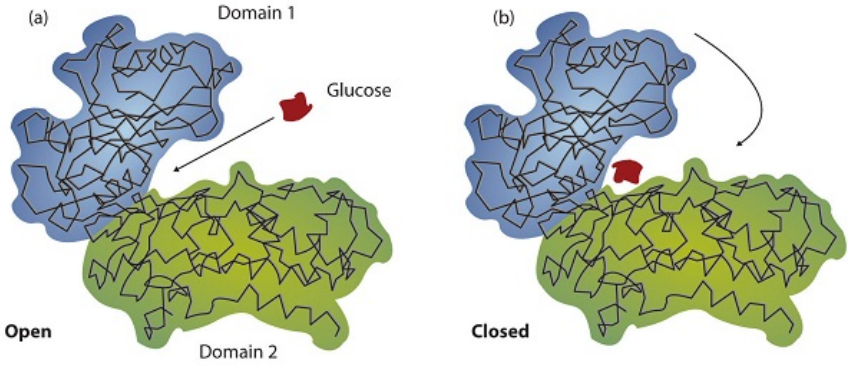


Figure 9.8.3: The Induced-Fit Model of Enzyme Action. (a) The enzyme hexokinase without its substrate (glucose, shown in red) is bound to the active site. (b) The enzyme conformation changes dramatically when the substrate binds to it, resulting in additional interactions between hexokinase and glucose.

The structural changes that occur when an enzyme and a substrate join together bring specific parts of a substrate into alignment with specific parts of the enzyme's active site. Amino acid side chains in or near the binding site can then act as acid or base catalysts, provide binding sites for the transfer of functional groups from one substrate to another or aid in the rearrangement of a substrate. The participating amino acids, which are usually widely separated in the primary sequence of the protein, are brought close together in the active site as a result of the folding and bending of the polypeptide chain or chains when the protein acquires its tertiary and quaternary structure. Binding to enzymes brings reactants close to each other and aligns them properly, which has the same effect as increasing the concentration of the reacting compounds.

✔ Example 9.8.1

a. What type of interaction would occur between an OH group present on a substrate molecule and a functional group in the active site of an enzyme?
b. Suggest an amino acid whose side chain might be in the active site of an enzyme and form the type of interaction you just identified.

Solution

a. An OH group would most likely engage in hydrogen bonding with an appropriate functional group present in the active site of an enzyme.

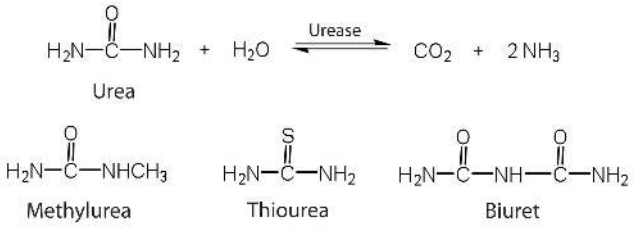
b. Several amino acid side chains would be able to engage in hydrogen bonding with an OH group. One example would be asparagine, which has an amide functional group.

? Exercise 9.8.1

a. What type of interaction would occur between an COO^- group present on a substrate molecule and a functional group in the active site of an enzyme?

b. Suggest an amino acid whose side chain might be in the active site of an enzyme and form the type of interaction you just identified.

One characteristic that distinguishes an enzyme from all other types of catalysts is its *substrate specificity*. An inorganic acid such as sulfuric acid can be used to increase the reaction rates of many different reactions, such as the hydrolysis of disaccharides, polysaccharides, lipids, and proteins, with complete impartiality. In contrast, enzymes are much more specific. Some enzymes act on a single substrate, while other enzymes act on any of a group of related molecules containing a similar functional group or chemical bond. Some enzymes even distinguish between D- and L-stereoisomers, binding one stereoisomer but not the other. Urease, for example, is an enzyme that catalyzes the hydrolysis of a single substrate—urea—but not the closely related compounds methyl urea, thiourea, or biuret. The enzyme carboxypeptidase, on the other hand, is far less specific. It catalyzes the removal of nearly any amino acid from the carboxyl end of any peptide or protein.



Enzyme specificity results from the uniqueness of the active site in each different enzyme because of the identity, charge, and spatial orientation of the functional groups located there. It regulates cell chemistry so that the proper reactions occur in the proper place at the proper time. Clearly, it is crucial to the proper functioning of the living cell.

Summary

A substrate binds to a specific region on an enzyme known as the active site, where the substrate can be converted to product. The substrate binds to the enzyme primarily through hydrogen bonding and other electrostatic interactions. The induced-fit model says that an enzyme can undergo a conformational change when binding a substrate. Enzymes exhibit varying degrees of substrate specificity.

9.8: Enzyme Action is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.6: Enzyme Action** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.9: Enzyme Activity

Learning Objectives

- To describe how pH, temperature, and the concentration of an enzyme and its substrate influence enzyme activity.

The single most important property of enzymes is the ability to increase the rates of reactions occurring in living organisms, a property known as *catalytic activity*. Because most enzymes are proteins, their activity is affected by factors that disrupt protein structure, as well as by factors that affect catalysts in general. Factors that disrupt protein structure include temperature and pH; factors that affect catalysts in general include reactant or substrate concentration and catalyst or enzyme concentration. The activity of an enzyme can be measured by monitoring either the rate at which a substrate disappears or the rate at which a product forms.

Concentration of Substrate

In the presence of a given amount of enzyme, the rate of an enzymatic reaction increases as the substrate concentration increases until a limiting rate is reached, after which further increase in the substrate concentration produces no significant change in the reaction rate (part (a) of Figure **9.9.1**). At this point, so much substrate is present that essentially all of the enzyme active sites have substrate bound to them. In other words, the enzyme molecules are saturated with substrate. The excess substrate molecules cannot react until the substrate already bound to the enzymes has reacted and been released (or been released without reacting).

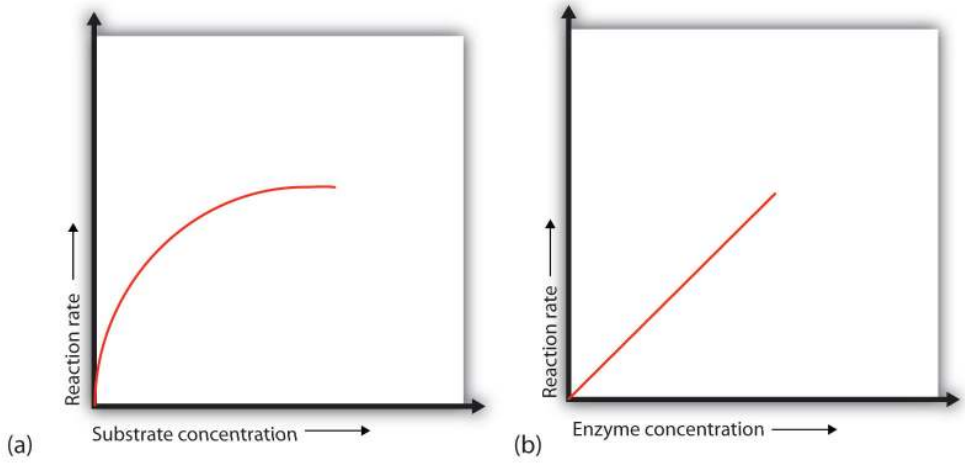


Figure **9.9.1**: Concentration versus Reaction Rate. (a) This graph shows the effect of substrate concentration on the rate of a reaction that is catalyzed by a fixed amount of enzyme. (b) This graph shows the effect of enzyme concentration on the reaction rate at a constant level of substrate.

Let's consider an analogy. Ten taxis (enzyme molecules) are waiting at a taxi stand to take people (substrate) on a 10-minute trip to a concert hall, one passenger at a time. If only 5 people are present at the stand, the rate of their arrival at the concert hall is 5 people in 10 minutes. If the number of people at the stand is increased to 10, the rate increases to 10 arrivals in 10 minutes. With 20 people at the stand, the rate would still be 10 arrivals in 10 minutes. The taxis have been "saturated." If the taxis could carry 2 or 3 passengers each, the same principle would apply. The rate would simply be higher (20 or 30 people in 10 minutes) before it leveled off.

Concentration of Enzyme

When the concentration of the enzyme is significantly lower than the concentration of the substrate (as when the number of taxis is far lower than the number of waiting passengers), the rate of an enzyme-catalyzed reaction is directly dependent on the enzyme concentration (part (b) of Figure **9.9.1**). This is true for any catalyst; the reaction rate increases as the concentration of the catalyst is increased.

Temperature

A general rule of thumb for most chemical reactions is that a temperature rise of 10°C approximately doubles the reaction rate. To some extent, this rule holds for all enzymatic reactions. After a certain point, however, an increase in temperature causes a decrease in the reaction rate, due to denaturation of the protein structure and disruption of the active site (part (a) of Figure **9.9.2**). For many proteins, denaturation occurs between 45°C and 55°C. Furthermore, even though an enzyme may appear to have a maximum reaction rate between 40°C and 50°C, most biochemical reactions are carried out at lower temperatures because enzymes are not stable at these higher temperatures and will denature after a few minutes.

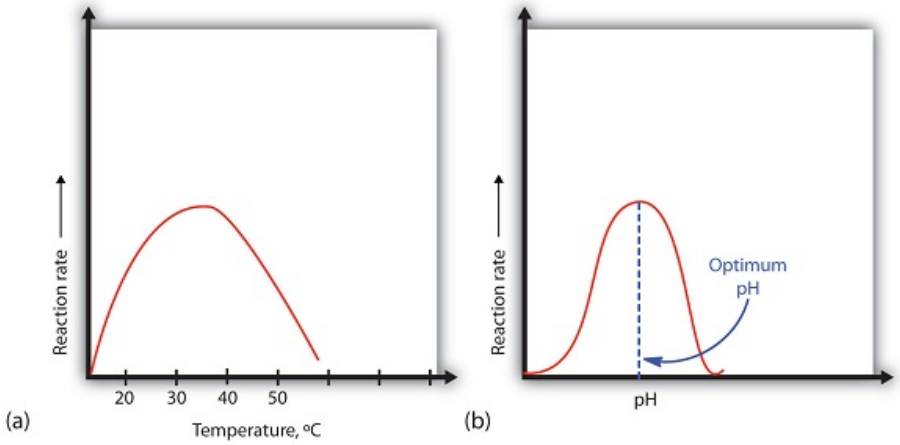


Figure **9.9.2**: Temperature and pH versus Concentration. (a) This graph depicts the effect of temperature on the rate of a reaction that is catalyzed by a fixed amount of enzyme. (b) This graph depicts the effect of pH on the rate of a reaction that is catalyzed by a fixed amount of enzyme.

At 0°C and 100°C, the rate of enzyme-catalyzed reactions is nearly zero. This fact has several practical applications. We sterilize objects by placing them in boiling water, which denatures the enzymes of any bacteria that may be in or on them. We preserve our food by refrigerating or freezing it, which slows enzyme activity. When animals go into hibernation in winter, their body temperature drops, decreasing the rates of their metabolic processes to levels that can be maintained by the amount of energy stored in the fat reserves in the animals' tissues.

Hydrogen Ion Concentration (pH)

Because most enzymes are proteins, they are sensitive to changes in the hydrogen ion concentration or pH. Enzymes may be denatured by extreme levels of hydrogen ions (whether high or low); *any* change in pH, even a small one, alters the degree of ionization of an enzyme's acidic and basic side groups and the substrate components as well. Ionizable side groups located in the active site must have a certain charge for the enzyme to bind its substrate. Neutralization of even one of these charges alters an enzyme's catalytic activity.

An enzyme exhibits maximum activity over the narrow pH range in which a molecule exists in its properly charged form. The median value of this pH range is called the optimum pH of the enzyme (part (b) of Figure **9.9.2**). With the notable exception of gastric juice (the fluids secreted in the stomach), most body fluids have pH values between 6 and 8. Not surprisingly, most enzymes exhibit optimal activity in this pH range. However, a few enzymes have optimum pH values outside this range. For example, the optimum pH for pepsin, an enzyme that is active in the stomach, is 2.0.

Summary

Initially, an increase in substrate concentration leads to an increase in the rate of an enzyme-catalyzed reaction. As the enzyme molecules become saturated with substrate, this increase in reaction rate levels off. The rate of an enzyme-catalyzed reaction increases with an increase in the concentration of an enzyme. At low temperatures, an increase in temperature increases the rate of an enzyme-catalyzed reaction. At higher temperatures, the protein is denatured, and the rate of the reaction dramatically decreases. An enzyme has an optimum pH range in which it exhibits maximum activity.

9.9: Enzyme Activity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.7: Enzyme Activity** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

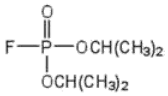
9.10: Enzyme Inhibition

Learning Objectives

- Explain what an enzyme inhibitor is.
- Distinguish between reversible and irreversible inhibitors.
- Distinguish between competitive and noncompetitive inhibitors.

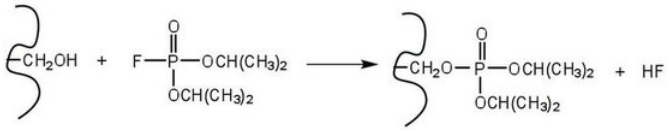
Previously, we noted that enzymes are inactivated at high temperatures and by changes in pH. These are nonspecific factors that would inactivate any enzyme. The activity of enzymes can also be regulated by more specific inhibitors. Many compounds are poisons because they bind covalently to particular enzymes or kinds of enzymes and inactivate them (Table **9.10.1**).

Table **9.10.1**: Poisons as Enzyme Inhibitors

Poison	Formula	Example of Enzyme Inhibited	Action
arsenate	AsO_4^{3-}	glyceraldehyde 3-phosphate dehydrogenase	substitutes for phosphate
iodoacetate	ICH_2COO^-	triose phosphate dehydrogenase	binds to cysteine **SH** group
diisopropylfluoro-phosphate (DIFP; a nerve poison)		acetylcholinesterase	binds to serine **OH** group

Irreversible Inhibition: Poisons

An irreversible inhibitor inactivates an enzyme by bonding covalently to a particular group at the active site. The inhibitor-enzyme bond is so strong that the inhibition cannot be reversed by the addition of excess substrate. The nerve gases, especially Diisopropyl fluorophosphate (DIFP), irreversibly inhibit biological systems by forming an enzyme-inhibitor complex with a specific OH group of serine situated at the active sites of certain enzymes. The peptidases trypsin and chymotrypsin contain serine groups at the active site and are inhibited by DIFP.



Reversible Inhibition

A reversible inhibitor inactivates an enzyme through noncovalent, more easily reversed, interactions. Unlike an irreversible inhibitor, a reversible inhibitor can dissociate from the enzyme. Reversible inhibitors include competitive inhibitors and noncompetitive inhibitors. (There are additional types of reversible inhibitors.) A competitive inhibitor is any compound that bears a structural resemblance to a particular substrate and thus competes with that substrate for binding at the active site of an enzyme. The inhibitor is not acted on by the enzyme but does prevent the substrate from approaching the active site.

The degree to which a competitive inhibitor interferes with an enzyme's activity depends on the relative concentrations of the substrate and the inhibitor. If the inhibitor is present in relatively large quantities, it will initially block most of the active sites. But because the binding is reversible, some substrate molecules will eventually bind to the active site and be converted to product. Increasing the substrate concentration promotes displacement of the inhibitor from the active site. Competitive inhibition can be completely reversed by adding substrate so that it reaches a much higher concentration than that of the inhibitor.

Studies of competitive inhibition have provided helpful information about certain enzyme-substrate complexes and the interactions of specific groups at the active sites. As a result, pharmaceutical companies have synthesized drugs that competitively inhibit metabolic processes in bacteria and certain cancer cells. Many drugs are competitive inhibitors of specific enzymes.

A classic example of competitive inhibition is the effect of malonate on the enzyme activity of succinate dehydrogenase (Figure 9.10.1). Malonate and succinate are the anions of dicarboxylic acids and contain three and four carbon atoms, respectively. The malonate molecule binds to the active site because the spacing of its carboxyl groups is not greatly different from that of succinate. However, no catalytic reaction occurs because malonate does not have a CH_2CH_2 group to convert to CH=CH. This reaction will also be discussed in connection with the Krebs cycle and energy production.

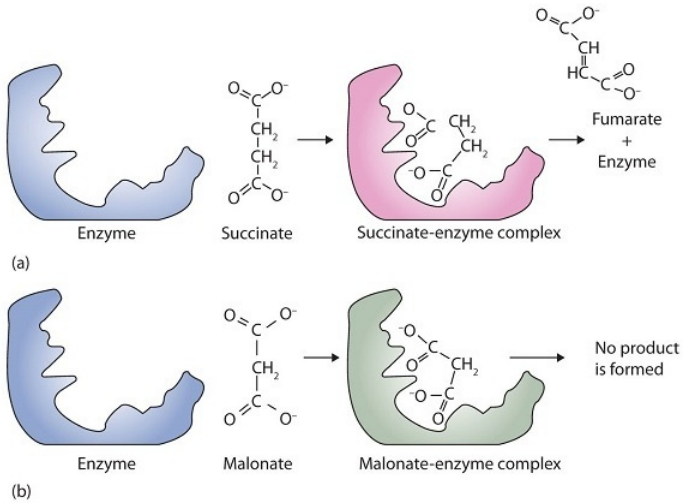


Figure 9.10.1: Competitive Inhibition. (a) Succinate binds to the enzyme succinate dehydrogenase. A dehydrogenation reaction occurs, and the product—fumarate—is released from the enzyme. (b) Malonate also binds to the active site of succinate dehydrogenase. In this case, however, no subsequent reaction occurs while malonate remains bound to the enzyme.

To Your Health: Penicillin

Chemotherapy is the strategic use of chemicals (that is, drugs) to destroy infectious microorganisms or cancer cells without causing excessive damage to the other, healthy cells of the host. From bacteria to humans, the metabolic pathways of all living organisms are quite similar, so the search for safe and effective chemotherapeutic agents is a formidable task. Many well-established chemotherapeutic drugs function by inhibiting a critical enzyme in the cells of the invading organism.

An *antibiotic* is a compound that kills bacteria; it may come from a natural source such as molds or be synthesized with a structure analogous to a naturally occurring antibacterial compound. Antibiotics constitute no well-defined class of chemically related substances, but many of them work by effectively inhibiting a variety of enzymes essential to bacterial growth.

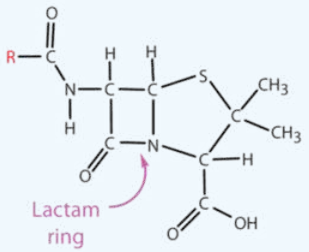
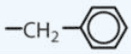
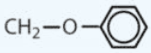
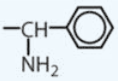
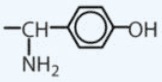
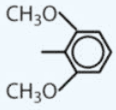
Penicillin, one of the most widely used antibiotics in the world, was fortuitously discovered by Alexander Fleming in 1928, when he noticed antibacterial properties in a mold growing on a bacterial culture plate. In 1938, Ernst Chain and Howard Florey began an intensive effort to isolate penicillin from the mold and study its properties. The large quantities of penicillin needed for this research became available through development of a corn-based nutrient medium that the mold loved and through the discovery of a higher-yielding strain of mold at a United States Department of Agriculture research center near Peoria, Illinois. Even so, it was not until 1944 that large quantities of penicillin were being produced and made available for the treatment of bacterial infections.

Penicillin functions by interfering with the synthesis of cell walls of reproducing bacteria. It does so by inhibiting an enzyme—transpeptidase—that catalyzes the last step in bacterial cell-wall biosynthesis. The defective walls cause bacterial cells to burst. Human cells are not affected because they have cell membranes, not cell walls.

Several naturally occurring penicillins have been isolated. They are distinguished by different R groups connected to a common structure: a four-member cyclic amide (called a lactam ring) fused to a five-member ring. The addition of appropriate organic compounds to the culture medium leads to the production of the different kinds of penicillin.

The penicillins are effective against gram-positive bacteria (bacteria capable of being stained by Gram's stain) and a few gram-negative bacteria (including the intestinal bacterium *Escherichia coli*). They are effective in the treatment of diphtheria, gonorrhea, pneumonia, syphilis, many pus infections, and certain types of boils. Penicillin G was the earliest penicillin to be used on a wide scale. However, it cannot be administered orally because it is quite unstable; the acidic pH of the stomach

converts it to an inactive derivative. The major oral penicillins—penicillin V, ampicillin, and amoxicillin—on the other hand, are acid stable.

Penicillin Structure	R Group	Drug Name
		penicillin G
		penicillin V
		ampicillin
		amoxicillin
		methicillin

Some strains of bacteria become resistant to penicillin through a mutation that allows them to synthesize an enzyme—penicillinase—that breaks the antibiotic down (by cleavage of the amide linkage in the lactam ring). To combat these strains, scientists have synthesized penicillin analogs (such as methicillin) that are not inactivated by penicillinase.

Some people (perhaps 5% of the population) are allergic to penicillin and therefore must be treated with other antibiotics. Their allergic reaction can be so severe that a fatal coma may occur if penicillin is inadvertently administered to them. Fortunately, several other antibiotics have been discovered. Most, including aureomycin and streptomycin, are the products of microbial synthesis. Others, such as the semisynthetic penicillins and tetracyclines, are made by chemical modifications of antibiotics; and some, like chloramphenicol, are manufactured entirely by chemical synthesis. They are as effective as penicillin in destroying infectious microorganisms. Many of these antibiotics exert their effects by blocking protein synthesis in microorganisms.

Initially, antibiotics were considered miracle drugs, substantially reducing the number of deaths from blood poisoning, pneumonia, and other infectious diseases. Some seven decades ago, a person with a major infection almost always died. Today, such deaths are rare. Seven decades ago, pneumonia was a dreaded killer of people of all ages. Today, it kills only the very old or those ill from other causes. Antibiotics have indeed worked miracles in our time, but even miracle drugs have limitations. Not long after the drugs were first used, disease organisms began to develop strains resistant to them. In a race to stay ahead of resistant bacterial strains, scientists continue to seek new antibiotics. The penicillins have now been partially displaced by related compounds, such as the cephalosporins and vancomycin. Unfortunately, some strains of bacteria have already shown resistance to these antibiotics.

Some reversible inhibitors are noncompetitive. A noncompetitive inhibitor can combine with either the free enzyme or the enzyme-substrate complex because its binding site on the enzyme is distinct from the active site. Binding of this kind of inhibitor alters the three-dimensional conformation of the enzyme, changing the configuration of the active site with one of two results. Either the enzyme-substrate complex does not form at its normal rate, or, once formed, it does not yield products at the normal rate. Because the inhibitor does not structurally resemble the substrate, the addition of excess substrate does *not* reverse the inhibitory effect.

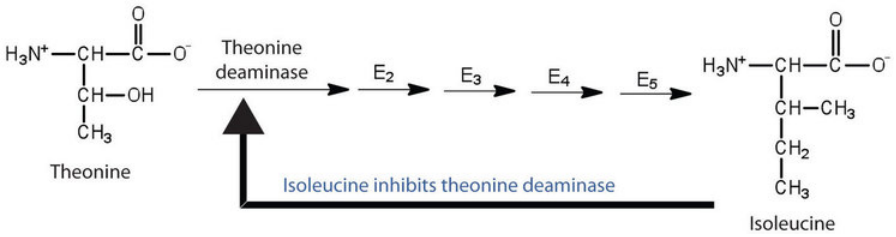


Figure 9.10.2: Feedback Inhibition of Threonine Deaminase by Isoleucine. Threonine deaminase is the first enzyme in the conversion of threonine to isoleucine. Isoleucine inhibits threonine deaminase through feedback inhibition.

Feedback inhibition is a normal biochemical process that makes use of noncompetitive inhibitors to control some enzymatic activity. In this process, the final product inhibits the enzyme that catalyzes the first step in a series of reactions. Feedback inhibition is used to regulate the synthesis of many amino acids. For example, bacteria synthesize isoleucine from threonine in a series of five enzyme-catalyzed steps. As the concentration of isoleucine increases, some of it binds as a noncompetitive inhibitor to the first enzyme of the series (threonine deaminase), thus bringing about a decrease in the amount of isoleucine being formed (Figure **9.10.2**).

Summary

An irreversible inhibitor inactivates an enzyme by bonding covalently to a particular group at the active site. A reversible inhibitor inactivates an enzyme through noncovalent, reversible interactions. A competitive inhibitor competes with the substrate for binding at the active site of the enzyme. A noncompetitive inhibitor binds at a site distinct from the active site.

9.10: Enzyme Inhibition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.8: Enzyme Inhibition** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

https://chem.libretexts.org/@go/page/255635

9.11: Enzyme Cofactors and Vitamins

Learning Objectives

- To explain why vitamins are necessary in the diet.

Many enzymes are simple proteins consisting entirely of one or more amino acid chains. Other enzymes contain a nonprotein component called a **cofactor** that is necessary for the enzyme's proper functioning. There are two types of cofactors: inorganic ions [e.g., zinc or Cu(I) ions] and organic molecules known as coenzymes. Most **coenzymes** are vitamins or are derived from vitamins.

Vitamins are organic compounds that are essential in very small (trace) amounts for the maintenance of normal metabolism. They generally cannot be synthesized at adequate levels by the body and must be obtained from the diet. The absence or shortage of a vitamin may result in a vitamin-deficiency disease. In the first half of the 20th century, a major focus of biochemistry was the identification, isolation, and characterization of vitamins. Despite accumulating evidence that people needed more than just carbohydrates, fats, and proteins in their diets for normal growth and health, it was not until the early 1900s that research established the need for trace nutrients in the diet.

Table **9.11.1**: Fat-Soluble Vitamins and Physiological Functions

Vitamin	Physiological Function	Effect of Deficiency
vitamin A (retinol)	formation of vision pigments; differentiation of epithelial cells	night blindness; continued deficiency leads to total blindness
vitamin D (cholecalciferol)	increases the body's ability to absorb calcium and phosphorus	osteomalacia (softening of the bones); known as rickets in children
vitamin E (tocopherol)	fat-soluble antioxidant	damage to cell membranes
vitamin K (phylloquinone)	formation of prothrombin, a key enzyme in the blood-clotting process	increases the time required for blood to clot

Because organisms differ in their synthetic abilities, a substance that is a vitamin for one species may not be so for another. Over the past 100 years, scientists have identified and isolated 13 vitamins required in the human diet and have divided them into two broad categories: the *fat-soluble vitamins*, which include vitamins A, D, E, and K, and the *water-soluble vitamins*, which are the B complex vitamins and vitamin C. All fat-soluble vitamins contain a high proportion of hydrocarbon structural components. There are one or two oxygen atoms present, but the compounds as a whole are nonpolar. In contrast, water-soluble vitamins contain large numbers of electronegative oxygen and nitrogen atoms, which can engage in hydrogen bonding with water. Most water-soluble vitamins act as coenzymes or are required for the synthesis of coenzymes. The fat-soluble vitamins are important for a variety of physiological functions. The key vitamins and their functions are found in Tables **9.11.1** and **9.11.2**.

Table **9.11.2**: Water-Soluble Vitamins and Physiological Functions

Vitamin	Coenzyme	Coenzyme Function	Deficiency Disease
vitamin B_1 (thiamine)	thiamine pyrophosphate	decarboxylation reactions	beri-beri
vitamin B_2 (riboflavin)	flavin mononucleotide or flavin adenine dinucleotide	oxidation-reduction reactions involving two hydrogen atoms	—
vitamin B_3 (niacin)	nicotinamide adenine dinucleotide or nicotinamide adenine dinucleotide phosphate	oxidation-reduction reactions involving the hydride ion (H^-)	pellagra
vitamin B_6 (pyridoxine)	pyridoxal phosphate	variety of reactions including the transfer of amino groups	—
vitamin B_{12} (cyanocobalamin)	methylcobalamin or deoxyadenoxylcobalamin	intramolecular rearrangement reactions	pernicious anemia
biotin	biotin	carboxylation reactions	—
folic acid	tetrahydrofolate	carrier of one-carbon units such as the formyl group	anemia
pantothenic Acid	coenzyme A	carrier of acyl groups	—
vitamin C (ascorbic acid)	none	antioxidant; formation of collagen, a protein found in tendons, ligaments, and bone	scurvy

Vitamins C and E, as well as the provitamin β-carotene can act as antioxidants in the body. Antioxidants prevent damage from free radicals, which are molecules that are highly reactive because they have unpaired electrons. Free radicals are formed not only through metabolic reactions involving oxygen but also by such environmental factors as radiation and pollution.

> *β-carotene is known as a provitamin because it can be converted to vitamin A in the body.*

Free radicals react most commonly react with lipoproteins and unsaturated fatty acids in cell membranes, removing an electron from those molecules and thus generating a new free radical. The process becomes a chain reaction that finally leads to the oxidative degradation of the affected compounds. Antioxidants react with free radicals to stop these chain reactions by forming a more stable molecule or, in the case of vitamin E, a free radical that is much less reactive (vitamin E is converted back to its original form through interaction with vitamin C).

Summary

Vitamins are organic compounds that are essential in very small amounts for the maintenance of normal metabolism. Vitamins are divided into two broad categories: fat-soluble vitamins and water-soluble vitamins. Most water-soluble vitamins are needed for the formation of coenzymes, which are organic molecules needed by some enzymes for catalytic activity.

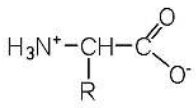
9.11: Enzyme Cofactors and Vitamins is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.9: Enzyme Cofactors and Vitamins** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.12: Amino Acids, Proteins, and Enzymes (Exercises)

18.1: Properties of Amino Acids

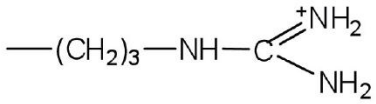
Concept Review Exercises

1. What is the general structure of an α-amino acid?
2. Identify the amino acid that fits each description.
   a. also known as aspartate
   b. almost as strong a base as sodium hydroxide
   c. does not have a chiral carbon
3. 

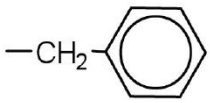
4. a. aspartic acid
   b. arginine
   c. glycine
5. Write the side chain of each amino acid.
   a. serine
   b. arginine
   c. phenylalanine
6. Write the side chain of each amino acid.
   a. aspartic acid
   b. methionine
   c. valine
7. Draw the structure for each amino acid.
   a. alanine
   b. cysteine
   c. histidine
8. Draw the structure for each amino acid.
   a. threonine
   b. glutamic acid
   c. leucine
9. Identify an amino acid whose side chain contains a(n)
   a. amide functional group.
   b. aromatic ring.
   c. carboxyl group.
10. Identify an amino acid whose side chain contains a(n)
    a. OH group
    b. branched chain
    c. amino group

1. a. CH_2OH^-

b.

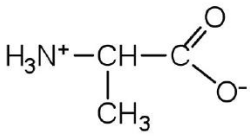


c.

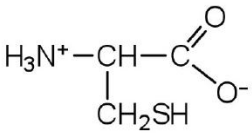


2.

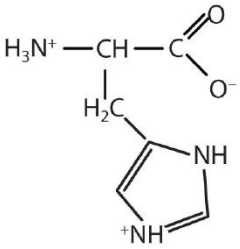
3. a.



b.



c.



4.

5. a. asparagine or glutamine
   b. phenylalanine, tyrosine, or tryptophan
   c. aspartic acid or glutamic acid

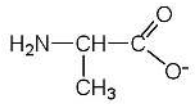
18.2: Reactions of Amino Acids

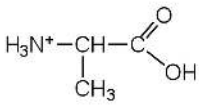
Concept Review Exercises

1. Define each term.
   a. zwitterion
   b. isoelectric point
2. Draw the structure for the anion formed when alanine (at neutral pH) reacts with a base.
3. Draw the structure for the cation formed when alanine (at neutral pH) reacts with an acid.

Answers

1. a. an electrically neutral compound that contains both negatively and positively charged groups
   b. the pH at which a given amino acid exists in solution as a zwitterion

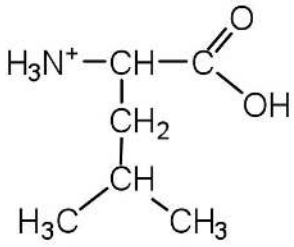
2. 

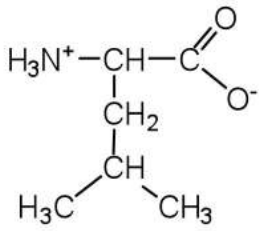
3. 

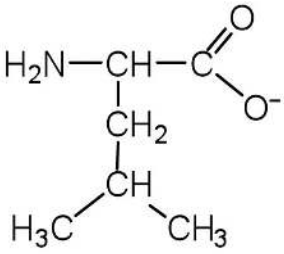
Exercises

1. Draw the structure of leucine and determine the charge on the molecule in a(n)
   a. acidic solution (pH = 1).
   b. neutral solution (pH = 7).
   c. a basic solution (pH = 11)
2. Draw the structure of isoleucine and determine the charge on the molecule in a(n)
   a. acidic solution (pH = 1).
   b. neutral solution (pH = 7).
   c. basic solution (pH = 11).

Answer

1. a. 

   b. 

   c. 

18.3: Peptides

Concept Review Exercises

1. Distinguish between the N-terminal amino acid and the C-terminal amino acid of a peptide or protein.
2. Describe the difference between an amino acid and a peptide.
3. Amino acid units in a protein are connected by peptide bonds. What is another name for the functional group linking the amino acids?

Answers

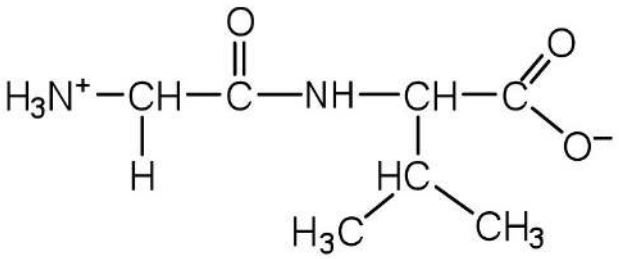
1. The N-terminal end is the end of a peptide or protein whose amino group is free (not involved in the formation of a peptide bond), while the C-terminal end has a free carboxyl group.
2. A peptide is composed of two or more amino acids. Amino acids are the building blocks of peptides.
3. amide bond

Exercises

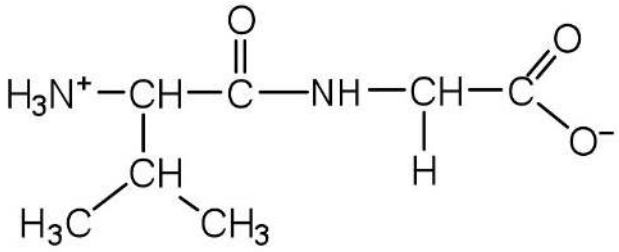
1. Draw the structure for each peptide.
   a. gly-val
   b. val-gly
2. Draw the structure for cys-val-ala.
3. Identify the C- and N-terminal amino acids for the peptide lys-val-phe-gly-arg-cys.
4. Identify the C- and N-terminal amino acids for the peptide asp-arg-val-tyr-ile-his-pro-phe.

Answers

1. a.



   b.



3. C-terminal amino acid: cys; N-terminal amino acid: lys

18.4: Proteins

Concept Review Exercises

1. What is the predominant attractive force that stabilizes the formation of secondary structure in proteins?
2. Distinguish between the tertiary and quaternary levels of protein structure.
3. Briefly describe four ways in which a protein could be denatured.

Answers

1. hydrogen bonding
2. Tertiary structure refers to the unique three-dimensional shape of a single polypeptide chain, while quaternary structure describes the interaction between multiple polypeptide chains for proteins that have more than one polypeptide chain.
3. (1) heat a protein above 50°C or expose it to UV radiation; (2) add organic solvents, such as ethyl alcohol, to a protein solution; (3) add salts of heavy metal ions, such as mercury, silver, or lead; and (4) add alkaloid reagents such as tannic acid

Exercises

1. Classify each protein as fibrous or globular.
   a. albumin
   b. myosin
   c. fibroin
2. Classify each protein as fibrous or globular.
   a. hemoglobin
   b. keratin
   c. myoglobin
3. What name is given to the predominant secondary structure found in silk?
4. What name is given to the predominant secondary structure found in wool protein?
5. A protein has a tertiary structure formed by interactions between the side chains of the following pairs of amino acids. For each pair, identify the strongest type of interaction between these amino acids.
   a. aspartic acid and lysine
   b. phenylalanine and alanine
   c. serine and lysine
   d. two cysteines
6. A protein has a tertiary structure formed by interactions between the side chains of the following pairs of amino acids. For each pair, identify the strongest type of interaction between these amino acids.
   a. valine and isoleucine
   b. asparagine and serine
   c. glutamic acid and arginine
   d. tryptophan and methionine
7. What level(s) of protein structure is(are) ordinarily disrupted in denaturation? What level(s) is(are) not?
8. Which class of proteins is more easily denatured—fibrous or globular?

Answers

1. a. globular
   b. fibrous
   c. fibrous
3. β-pleated sheet
5. a. ionic bonding
   b. dispersion forces
   c. dispersion forces
   d. disulfide linkage
7. Protein denaturation disrupts the secondary, tertiary, and quaternary levels of structure. Only primary structure is unaffected by denaturation.

18.5: Enzymes

Concept Review Exercise

In the small intestine, sucrose is hydrolyzed to form glucose and fructose in a reaction catalyzed by sucrase.

1. Identify the substrate in this reaction.
2. Name the enzyme.

Answers

1. sucrose
2. sucrase

Exercises

1. Identify the substrate catalyzed by each enzyme.
    a. lactase
    b. cellulase
    c. peptidase
2. Identify the substrate catalyzed by each enzyme.
    a. lipase
    b. amylase
    c. maltase
3. Identify each type of enzyme.
    a. decarboxylase
    b. protease
    c. transaminase
4. Identify each type of enzyme.
    a. dehydrogenase
    b. isomerase
    c. lipase

Answers

1. a. lactose
    b. cellulose
    c. peptides
3. a. lyase
    b. hydrolase
    c. transferase

18.6: Enzyme Action

Concept Review Exercises

1. Distinguish between the lock-and-key model and induced-fit model of enzyme action.
2. Which enzyme has greater specificity—urease or carboxypeptidase? Explain.

Answers

1. The lock-and-key model portrays an enzyme as conformationally rigid and able to bond only to substrates that exactly fit the active site. The induced fit model portrays the enzyme structure as more flexible and is complementary to the substrate only after the substrate is bound.
2. Urease has the greater specificity because it can bind only to a single substrate. Carboxypeptidase, on the other hand, can catalyze the removal of nearly any amino acid from the carboxyl end of a peptide or protein.

Exercises

1. What type of interaction would occur between each group present on a substrate molecule and a functional group of the active site in an enzyme?
   a. COOH
   b. NH_3^+
   c. OH
   d. $CH(CH_3)_2$
2. What type of interaction would occur between each group present on a substrate molecule and a functional group of the active site in an enzyme?
   a. SH
   b. NH_2
   c. C_6H_5
   d. COO^-
3. For each functional group in Exercise 1, suggest an amino acid whose side chain might be in the active site of an enzyme and form the type of interaction you identified.
4. For each functional group in Exercise 2, suggest an amino acid whose side chain might be in the active site of an enzyme and form the type of interaction you identified.

Answers

1. a. hydrogen bonding
   b. ionic bonding
   c. hydrogen bonding
   d. dispersion forces
3. a. The amino acid has a polar side chain capable of engaging in hydrogen bonding; serine (answers will vary).
   b. The amino acid has a negatively charged side chain; aspartic acid (answers will vary).
   c. The amino acid has a polar side chain capable of engaging in hydrogen bonding; asparagine (answers will vary).
   d. The amino acid has a nonpolar side chain; isoleucine (answers will vary).

18.7: Enzyme Activity

Concept Review Exercises

1. The concentration of substrate X is low. What happens to the rate of the enzyme-catalyzed reaction if the concentration of X is doubled?
2. What effect does an increase in the enzyme concentration have on the rate of an enzyme-catalyzed reaction?

Answers

1. If the concentration of the substrate is low, increasing its concentration will increase the rate of the reaction.
2. An increase in the amount of enzyme will increase the rate of the reaction (provided sufficient substrate is present).

Exercises

1. In non-enzyme-catalyzed reactions, the reaction rate increases as the concentration of reactant is increased. In an enzyme-catalyzed reaction, the reaction rate initially increases as the substrate concentration is increased but then begins to level off, so that the increase in reaction rate becomes less and less as the substrate concentration increases. Explain this difference.
2. Why do enzymes become inactive at very high temperatures?
3. An enzyme has an optimum pH of 7.4. What is most likely to happen to the activity of the enzyme if the pH drops to 6.3? Explain.
4. An enzyme has an optimum pH of 7.2. What is most likely to happen to the activity of the enzyme if the pH increases to 8.5? Explain.

Answers

1. In an enzyme-catalyzed reaction, the substrate binds to the enzyme to form an enzyme-substrate complex. If more substrate is present than enzyme, all of the enzyme binding sites will have substrate bound, and further increases in substrate concentration cannot increase the rate.
3. The activity will decrease; a pH of 6.3 is more acidic than 7.4, and one or more key groups in the active site may bind a hydrogen ion, changing the charge on that group.

18.8: Enzyme Inhibition

Concept Review Exercises

1. What is the difference between a cofactor and a coenzyme?
2. How are vitamins related to coenzymes?

Answers

1. A coenzyme is one type of cofactor. Coenzymes are organic molecules required by some enzymes for activity. A cofactor can be either a coenzyme or an inorganic ion.
2. Coenzymes are synthesized from vitamins.

Exercises

1. Identify each vitamin as water soluble or fat soluble.
   a. vitamin D
   b. vitamin C
   c. vitamin B_{12}
2. Identify each vitamin as water soluble or fat soluble.
   a. niacin
   b. cholecalciferol
   c. biotin
3. What vitamin is needed to form each coenzyme?
   a. pyridoxal phosphate
   b. flavin adenine dinucleotide
   c. coenzyme A
   d. nicotinamide adenine dinucleotide
4. What coenzyme is formed from each vitamin?
   a. niacin
   b. thiamine
   c. cyanocobalamin
   d. pantothenic acid

5. What is the function of each vitamin or coenzyme?
   a. flavin adenine dinucleotide
   b. vitamin A
   c. biotin
6. What is the function of each vitamin or coenzyme?
   a. vitamin K
   b. pyridoxal phosphate
   c. tetrahydrofolate

Answers

1. a. fat soluble
   b. water soluble
   c. water soluble
3. a. vitamin B_6 or pyridoxine
   b. vitamin B_2 or riboflavin
   c. pantothenic acid
   d. vitamin B_3 or niacin
5. a. needed by enzymes that catalyze oxidation-reduction reactions in which two hydrogen atoms are transferred
   b. needed for the formation of vision pigments
   c. needed by enzymes that catalyze carboxylation reactions

18.9: Enzyme Cofactors and Vitamins

Concept Review Exercises

1. What are the characteristics of an irreversible inhibitor?
2. In what ways does a competitive inhibitor differ from a noncompetitive inhibitor?

Answers

1. It inactivates an enzyme by bonding covalently to a particular group at the active site.
2. A competitive inhibitor structurally resembles the substrate for a given enzyme and competes with the substrate for binding at the active site of the enzyme. A noncompetitive inhibitor binds at a site distinct from the active site and can bind to either the free enzyme or the enzyme-substrate complex.

Exercises

1. What amino acid is present in the active site of all enzymes that are irreversibly inhibited by nerve gases such as DIFP?
2. Oxaloacetate (OOCCH2COCOO) inhibits succinate dehydrogenase. Would you expect oxaloacetate to be a competitive or noncompetitive inhibitor? Explain.

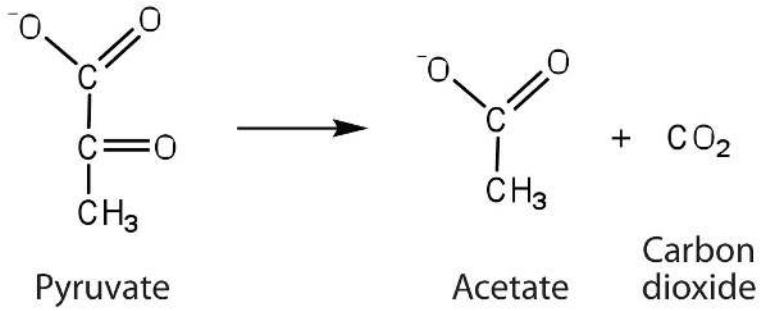
Answer

1. serine

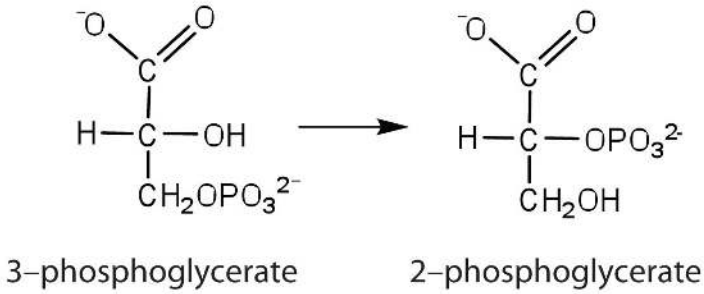
Additional Exercises

1. Draw the structure of the amino acid γ-aminobutyric acid (GABA). Would you expect to find GABA in the amino acid sequence of a protein? Explain.
2. Draw the structure of the amino acid homocysteine (R group = CH_2CH_2SH). Would you expect to find homocysteine in the amino acid sequence of a protein? Justify your answer.
3. Write equations to show how leucine can act as a buffer (that is, how it can neutralize added acid or base).

4. Write equations to show how isoleucine can act as a buffer (that is, how it can neutralize added acid or base).
5. Glutathione (γ-glutamylcysteinylglycine) is a tripeptide found in all cells of higher animals. It contains glutamic acid joined in an unusual peptide linkage involving the carboxyl group of the R group (known as γ-carboxyl group), rather than the usual carboxyl group (the α-carboxyl group). Draw the structure of glutathione.
6. Draw the structure of the pentapeptide whose sequence is arg-his-gly-leu-asp. Identify which of the amino acids have R groups that can donate or gain hydrogen ions.
7. Bradykinin is a peptide hormone composed of nine amino acids that lowers blood pressure. Its primary structure is arg-pro-pro-gly-phe-ser-pro-phe-arg. Would you expect bradykinin to be positively charged, negatively charged, or neutral at a pH of 6.0? Justify your answer.
8. One of the neurotransmitters involved in pain sensation is a peptide called substance P, which is composed of 11 amino acids and is released by nerve-cell terminals in response to pain. Its primary structure is arg-pro-lys-pro-gln-gln-phe-phe-gly-leu-met. Would you expect this peptide to be positively charged, negatively charged, or neutral at a pH of 6.0? Justify your answer.
9. Carbohydrates are incorporated into *glycoproteins*. Would you expect the incorporation of sugar units to increase or decrease the solubility of a protein? Justify your answer.
10. Some proteins have phosphate groups attached through an ester linkage to the OH groups of serine, threonine, or tyrosine residues to form *phosphoproteins*. Would you expect the incorporation of a phosphate group to increase or decrease the solubility of a protein? Justify your answer.
11. Refer to Table 18.5 and determine how each enzyme would be classified.
    a. the enzyme that catalyzes the conversion of ethanol to acetaldehyde
    b. the enzyme that catalyzes the breakdown of glucose 6-phosphate to glucose and inorganic phosphate ion (water is also a reactant in this reaction)
12. Refer to Table 18.5 and determine how each enzyme would be classified.
    a. the enzyme that catalyzes the removal of a carboxyl group from pyruvate to form acetate

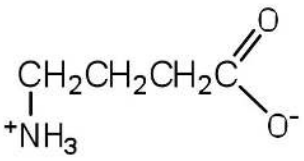


    b. the enzyme that catalyzes the rearrangement of 3-phosphoglycerate to form 2-phosphoglycerate



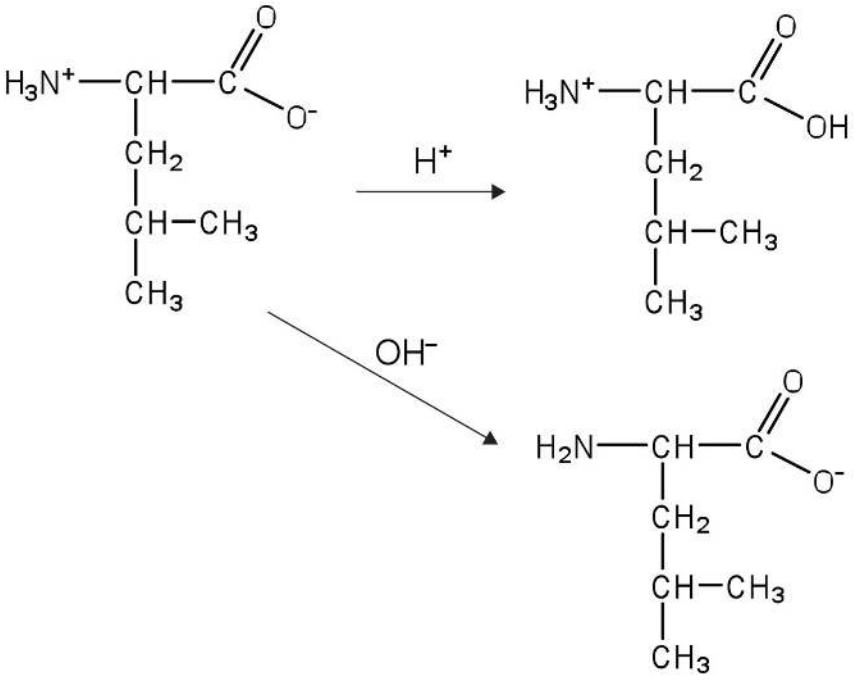
13. The enzyme lysozyme has an aspartic acid residue in the active site. In acidic solution, the enzyme is inactive, but activity increases as the pH rises to around 6. Explain why.
14. The enzyme lysozyme has a glutamic acid residue in the active site. At neutral pH (6–7), the enzyme is active, but activity decreases as the pH rises. Explain why.
15. The activity of a purified enzyme is measured at a substrate concentration of 1.0 μM and found to convert 49 μmol of substrate to product in 1 min. The activity is measured at 2.0 μM substrate and found to convert 98 μmol of substrate to product/minute.
    a. When the substrate concentration is 100 μM, how much substrate would you predict is converted to product in 1 min? What if the substrate concentration were increased to 1,000 μM (1.0 mM)?
    b. The activities actually measured are 676 μmol product formed/minute at a substrate concentration of 100 μM and 698 μmol product formed/minute at 1,000 μM (1.0 mM) substrate. Is there any discrepancy between these values and those you predicted in Exercise 15a? Explain.
16. A patient has a fever of 39°C. Would you expect the activity of enzymes in the body to increase or decrease relative to their activity at normal body temperature (37°C)?
17. Using your knowledge of factors that influence enzyme activity, describe what happens when milk is pasteurized.

Answers



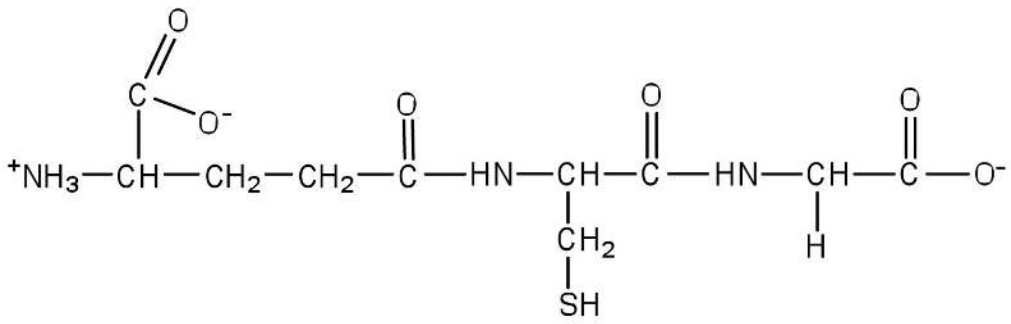
1.

This amino acid would not be found in proteins because it is not an α-amino acid.



3.

5.



7. Bradykinin would be positively charged; all of the amino acids, except for arginine, have R groups that do not become either positively or negatively charged. The two arginines are R groups that are positively charged at neutral pH, so the peptide would have an overall positive charge.

9. Carbohydrates have many OH groups attached, which can engage in hydrogen bonding with water, which increases the solubility of the proteins.

11.

a. oxidoreductase
b. hydrolase

13. The enzyme is active when the carboxyl group in the R group of aspartic acid does not have the hydrogen attached (forming COO^-); the hydrogen is removed when the pH of the solution is around pH 6 or higher.

15.

a. at 100 µM, you would predict that the rate would increase 100 times to 4,900 µmol of substrate to product in 1 min; at 1.0 mM, you would predict an increase to 49,000 µmol of substrate to product in 1 min.
b. There is a great discrepancy between the predicted rates and actual rates; this occurs because the enzyme becomes saturated with substrate, preventing a further increase in the rate of the reaction (the reaction is no longer linear with respect to substrate concentration because it is at very low concentrations).

17. When milk is pasteurized, it is heated to high temperatures. These high temperatures denature the proteins in bacteria, so they cannot carry out needed functions to grow and multiply.

9.12: Amino Acids, Proteins, and Enzymes (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.1: Properties of Amino Acids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **18.2: Reactions of Amino Acids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **18.3: Peptides** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **18.4: Proteins** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

- **18.5: Enzymes** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **18.6: Enzyme Action** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **18.7: Enzyme Activity** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **18.9: Enzyme Cofactors and Vitamins** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **18.8: Enzyme Inhibition** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

9.13: Amino Acids, Proteins, and Enzymes (Summary)

To ensure that you understand the material in this chapter, you should review the meanings of the bold terms in the following summary and ask yourself how they relate to the topics in the chapter.

A **protein** is a large biological polymer synthesized from **amino acids**, which are carboxylic acids containing an α-amino group. Proteins have a variety of important roles in living organisms, yet they are made from the same 20 L-amino acids. About half of these amino acids, the **essential amino acids**, cannot be synthesized by the human body and must be obtained from the diet. In the solid state and in neutral solutions, amino acids exist as **zwitterions**, species that are charged but electrically neutral. In this form, they behave much like inorganic salts. Each amino acid belongs to one of four classes depending on the characteristics of its R group or amino acid side chain: nonpolar, polar but neutral, positively charged, and negatively charged. Depending on the conditions, amino acids can act as either acids or bases, which means that proteins act as buffers. The pH at which an amino acid exists as the zwitterion is called the **isoelectric point (pI)**.

The amino acids in a protein are linked together by **peptide bonds**. Protein chains containing 10 or fewer amino acids are usually referred to as **peptides**, with a prefix such as di- or tri- indicating the number of amino acids. Chains containing more than 50 amino acid units are referred to as *proteins* or **polypeptides**. Proteins are classified globular or fibrous, depending on their structure and resulting solubility in water. **Globular proteins** are nearly spherical and are soluble in water; **fibrous proteins** have elongated or fibrous structures and are not soluble in water.

Protein molecules can have as many as four levels of structure. The **primary structure** is the sequence of amino acids in the chain. The **secondary structure** is the arrangement of adjacent atoms in the peptide chain; the most common arrangements are α-helices or β-pleated sheets. The **tertiary structure** is the overall three-dimensional shape of the molecule that results from the way the chain bends and folds in on itself. Proteins that consist of more than one chain have **quaternary structure**, which is the way the multiple chains are packed together.

Four types of intramolecular and intermolecular forces contribute to secondary, tertiary, and quaternary structure: (1) **hydrogen bonding** between an oxygen or a nitrogen atom and a hydrogen atom bound to an oxygen atom or a nitrogen atom, either on the same chain or on a neighboring chain; (2) **ionic bonding** between one positively charged side chain and one negatively charged side chain; (3) **disulfide linkages** between cysteine units; and (4) **dispersion forces** between nonpolar side chains.

Because of their complexity, protein molecules are delicate and easy to disrupt. A *denatured* protein is one whose conformation has been changed, in a process called **denaturation**, so that it can no longer do its physiological job. A variety of conditions, such as heat, ultraviolet radiation, the addition of organic compounds, or changes in pH can denature a protein.

An **enzyme** is an organic catalyst produced by a living cell. Enzymes are such powerful catalysts that the reactions they promote occur rapidly at body temperature. Without the help of enzymes, these reactions would require high temperatures and long reaction times.

The molecule or molecules on which an enzyme acts are called its **substrates**. An enzyme has an **active site** where its substrate or substrates bind to form an enzyme-substrate complex. The reaction occurs, and product is released:

$$E + S \rightarrow E\text{–}S \rightarrow E + P$$

The original **lock-and-key model** of enzyme and substrate binding pictured a rigid enzyme of unchanging configuration binding to the appropriate substrate. The newer **induced-fit model** describes the enzyme active site as changing its conformation after binding to the substrate.

Most enzymes have maximal activity in a narrow pH range centered on an **optimum pH**. In this pH range, the enzyme is correctly folded, and catalytic groups in the active site have the correct charge (positive, negative, or neutral). For most enzymes, the optimum pH is between 6 and 8.

Substances that interfere with enzyme function are called inhibitors. An **irreversible inhibitor** inactivates enzymes by forming covalent bonds to the enzyme, while a **reversible inhibitor** inactivates an enzyme by a weaker, noncovalent interaction that is easier to disrupt. A **competitive inhibitor** is a reversible inhibitor that is structurally similar to the substrate and binds to the active site. When the inhibitor is bound, the substrate is blocked from the active site and no reaction occurs. Because the binding of such

an inhibitor is reversible, a high substrate concentration will overcome the inhibition because it increases the likelihood of the substrate binding. A **noncompetitive inhibitor** binds reversibly at a site distinct from the active site. Thus, it can bind to either the enzyme or the enzyme-substrate complex. The inhibitor changes the conformation of the active site so that the enzyme cannot function properly. Noncompetitive inhibitors are important in **feedback inhibition**, in which the amount of product produced by a series of reactions is carefully controlled. The final product in a series of reactions acts as a noncompetitive inhibitor of the initial enzyme.

Simple enzymes consist entirely of one or more amino acid chains. Complex enzymes are composed of one or more amino acid chains joined to **cofactors**—inorganic ions or organic **coenzymes**. **Vitamins** are organic compounds that are essential in very small amounts for the maintenance of normal metabolism and generally cannot be synthesized at adequate levels by the body. Vitamins are divided into two broad categories: *fat-soluble* vitamins and *water-soluble* vitamins. Many of the water-soluble vitamins are used for the synthesis of coenzymes.

9.13: Amino Acids, Proteins, and Enzymes (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **18.S: Amino Acids, Proteins, and Enzymes (Summary)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

CHAPTER OVERVIEW

10: Equilibrium and Kinetics- Making Drugs Stick

The intermolecular interaction energy between a drug and a protein controls how quickly and how strongly it can bind. Chemists describe these processes as "kinetics" and "equilibrium" properties. In this chapter, we'll discuss basic aspects of these concepts using qualitative energy diagrams.

10.1: Prelude--A Drug's Life (in your Body)

10.2: A Quick Review of Thermodynamics and Kinetics

10.3: Catalysis

10.4: Pharmacology as an Equilibrium and Kinetics Problem

10: Equilibrium and Kinetics- Making Drugs Stick is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

10.1: Prelude--A Drug's Life (in your Body)

Antibiotic Cholesterol Bacterium Death Gene Cholesterol Cyclooxygenase abiotic transformation α-cleavage work

How does aspirin zap a headache? What happens after you rub some cortisone cream on a patch of poison ivy-induced rash on your arm? How do decongestant medicines such as Sudafed® dry up your nasal passages when you have a cold? As medicines find their way to their "job sites" in the body, hundreds of things happen along the way. One action triggers another, and medicines work to either mask a symptom, like a stuffy nose, or fix a problem, like a bacterial infection.

A Model for Success

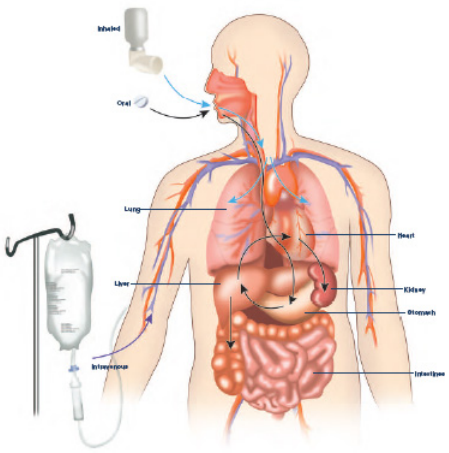
Turning a molecule into a good medicine is neither easy nor cheap. The Center for the Study of Drug Development at Tufts University in Boston estimates that it takes over $800 million and a dozen years to sift a few promising drugs from about 5,000 failures. Of this small handful of candidate drugs, only one will survive the rigors of clinical testing and end up on pharmacy shelves.

That's a huge investment for what may seem a very small gain and, n part, it explains the high cost of many prescription drugs. Sometimes, problems do not show up until after a drug reaches the market and many people begin taking the drug routinely. These problems range from irritating side effects, such as dry mouth or drowsiness, to life-threatening problems like serious bleeding or blood clots. The outlook might be brighter if pharmaceutical scientists could do a better job of predicting how potential drugs will act in the body (a science called pharmacodynamics), as well as what side effects the drugs might cause.

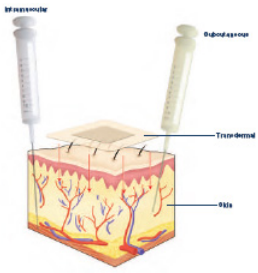
One approach that can help is computer modeling of a drug's properties. Computer modeling can help scientists at pharmaceutical and biotechnology companies filter out, and abandon early on, any candidate drugs that are likely to behave badly in the body. This can save significant amounts of time and money.

Computer software can examine the atom-by-atom structure of a molecule and determine how durable the chemical is likely to be inside a body's various chemical neighborhoods. Will the molecule break down easily? How well will the small intestines take it in? Does it dissolve easily in the watery environment of the fluids that course through the human body? Will the drug be able to penetrate the blood-brain barrier? Computer tools not only drive up the success rate for finding candidate drugs, they can also lead to the development of better medicines with fewer safety concerns.





A drug's life in the body. Medicines taken by mouth (oral) pass through the liver before they are absorbed into the bloodstream. Other forms of drug administration bypass the liver, entering the blood directly.



Drugs enter different layers of skin via intramuscular, subcutaneous, or transdermal delivery methods.

Scientists have names for the four basic stages of a medicine's life in the body: absorption, distribution, metabolism, and excretion. The entire process is sometimes abbreviated ADME. The first stage is *absorption*. Medicines can enter the body in many different ways, and they are absorbed when they travel from the site of administration into the body's circulation. A few of the most common ways to administer drugs are oral (swallowing an aspirin tablet), intramuscular (getting a flu shot in an arm muscle), subcutaneous (injecting insulin just under the skin), intravenous (receiving chemotherapy through a vein), or transdermal (wearing a skin patch), A drug faces its biggest hurdles during absorption. Medicines taken by mouth are shuttled via a special blood vessel leading from the digestive tract to the liver, where a large amount may be destroyed by metabolic enzymes in the so-called "first-pass effect." Other routes of drug administration bypass the liver, entering the bloodstream directly or via the skin or lungs.

Once a drug gets absorbed, the next stage is *distribution*. Most often, the bloodstream carries medicines throughout the body. During this step, side effects can occur when a drug has an effect in an organ other than the target organ. For a pain reliever, the target organ might be a sore muscle in the leg; irritation of the stomach could be a side effect. Many factors influence distribution, such as the presence of protein and fat molecules in the blood that can put drug molecules out of commission by grabbing onto them.

Drugs destined for the central nervous system (the brain and spinal cord) face an enormous hurdle: a nearly impenetrable barricade called the blood-brain barrier. This blockade is built from a tightly woven mesh of capillaries cemented together to protect the brain from potentially dangerous substances such as poisons or viruses. Yet pharmacologists have devised various ways to sneak some drugs past this barrier.

After a medicine has been distributed throughout the body and has done its job, the drug is broken down, or *metabolized*. The breaking down of a drug molecule usually involves two steps that take place mostly in the body's chemical processing plant, the liver. The liver is a site of continuous and frenzied, yet carefully controlled, activity. Everything that enters the bloodstream—whether swallowed, injected, inhaled, absorbed through the skin, or produced by the body itself—is carried to this largest internal organ. There, substances are chemically pummeled, twisted, cut apart, stuck together, and transformed.

Medicines and Your Genes



How you respond to a drug may be quite different from how your neighbor does. Why is that? Despite the fact that you might be about the same age and size, you probably eat different foods, get different amounts of exercise, and have different medical histories. But your genes, which are different from those of anyone else in the world, are really what make you unique. In part, your genes give you many obvious things, such as your looks, your mannerisms, and other characteristics that make you who you are. Your genes can also affect how you respond to the medicines you take. Your genetic code instructs your body how to make hundreds of thousands of different molecules called proteins. Some proteins determine hair color, and some of them are enzymes that process, or metabolize, food or medicines. Slightly different, but normal, variations in the human genetic code can yield proteins that work better or worse when they are metabolizing many different types of drugs and other substances. Scientists use the term pharmacogenetics to describe research on the link between genes and drug response.

One important group of proteins whose genetic code varies widely among people are "sulfation" enzymes, which perform chemical reactions in your body to make molecules more water-soluble, so they can be quickly excreted in the urine. Sulfation enzymes

metabolize many drugs, but they also work on natural body molecules, such as estrogen. Differences in the genetic code for sulfation enzymes can significantly alter blood levels of the many different kinds of substances metabolized by these enzymes. The same genetic differences may also put some people at risk for developing certain types of cancers whose growth is fueled by hormones like estrogen.

Pharmacogeneticist Rebecca Blanchard of Fox Chase Cancer Center in Philadelphia has discovered that people of different ethnic backgrounds have slightly different "spellings" of the genes that make sulfation enzymes. Lab tests revealed that sulfation enzymes manufactured from genes with different spellings metabolize drugs and estrogens at different rates. Blanchard and her coworkers are planning to work with scientists developing new drugs to include pharmacognetic testing in the early phases of screening new medicines.

The biotransformations that take place in the liver are performed by the body's busiest proteins, its enzymes. Every one of your cells has a variety of enzymes, drawn from a repertoire of hundreds of thousands. Each enzyme specializes in a particular job. Some break molecules apart, while others link small molecules into long chains. With drugs, the first step is usually to make the substance easier to get rid of in urine.

Many of the products of enzymatic break-down, which are called metabolites, are less chemically active than the original molecule. For this reason, scientists refer to the liver as a "detoxifying" organ. Occasionally, however, drug metabolites can have chemical activities of their own—sometimes as powerful as those of the original drug. When prescribing certain drugs, doctors must take into account these added effects. Once liver enzymes are finished working on a medicine, the now-inactive drug undergoes the final stage of its time in the body, *excretion*, as it exits via the urine or feces.

Perfect Timing

Pharmacokinetics is an aspect of pharmacology that deals with the absorption, distribution, and excretion of drugs. Because they are following drug actions in the body, researchers who specialize in pharmacokinetics must also pay attention to an additional dimension: time.

Pharmacokinetics research uses the tools of mathematics. Although sophisticated imaging methods can help track medicines as they travel through the body, scientists usually cannot actually see where a drug is going. To compensate, they often use mathematical models and precise measures of body fluids, such as blood and urine, to determine where a drug goes and how much of the drug or a break-down product remains after the body processes it. Other sentinels, such as blood levels of liver enzymes, can help predict how much of a drug is going to be absorbed.



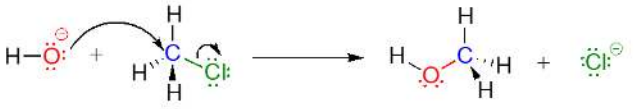
Studying pharmacokinetics also uses chemistry, since the interactions between drug and body molecules are really just a series of chemical reactions. Understanding the chemical encounters between drugs and biological environments, such as the bloodstream and the oily surfaces of cells, is necessary to predict how much of a drug will be taken in by the body. This concept, broadly termed bioavailability, is a critical feature that chemists and pharmaceutical scientists keep in mind when designing and packaging medicines. No matter how well a drug works in a laboratory simulation, the drug is not useful if it can't make it to its site of action.

10.1: Prelude--A Drug's Life (in your Body) is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

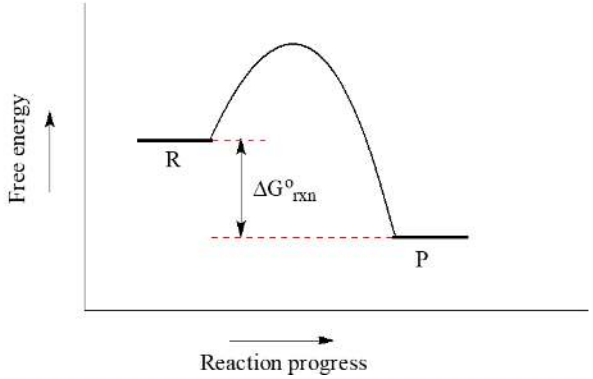
- **1.2: A Drug's Life** is licensed Public Domain.

10.2: A Quick Review of Thermodynamics and Kinetics

You may recall from general chemistry that it is often convenient to illustrate the energetics of a chemical reaction with a reaction coordinate diagram. In a reaction coordinate diagram, the vertical axis represents the overall potential energy of the reactants, while the horizontal axis is the 'reaction coordinate', tracing from left to right the progress of the reaction from starting reactants R to final products P. (Many students find it helpful to envision the horizontal axis in an reaction coordinate diagram as being analogous to the progress bar at the bottom of a youtube video). The diagram for a typical one-step nucleophilic substitution reaction such as that between hydroxide and methyl chloride might look like this:



reactants (R) **products (P)**



Despite its apparent simplicity, this diagram conveys some very important ideas about the thermodynamics and kinetics of the reaction. Recall that when we talk about the thermodynamics of a reaction, we are concerned primarily with the difference in energy between reactants (R) and products (P): whether the reaction as a whole is uphill or downhill. When we talk about kinetics, on the other hand, we are concerned with the rate of the reaction: how fast it goes from reactants to products, regardless of whether that transformation is energetically uphill or downhill.

Thermodynamics

First, a quick review of some key thermodynamics terms. Recall that the standard Gibbs free-energy change of a reaction (ΔG^o) is the difference in energy between reactants and products at standard conditions. Gibbs free-energy change is a combination of enthalpy change (ΔH^o) and entropy change (ΔS^o):

$$\Delta G = \Delta H^o - T\Delta S^o \tag{10.2.1}$$

where

- T is the temperature in Kelvin (recall that the Kelvin temperature is simply the Celsius temperature plus 273.15).
- Enthalpy change (ΔH^o) is the heat released or absorbed by the reaction.
- Entropy change (ΔS^o) is the change in disorder from reactants to products. In a reaction in which one molecule cleaves into two smaller molecules, for example, disorder increases, so ΔS^o is positive.

The equilibrium constant (K_{eq}) for a reaction is an expression of the relative concentrations of reactants and products after the reaction has reached equilibrium. The equilibrium constant is defined as:

this means that w molecules of compound A react with x molecules of compound B)

$$wA + xB \rightleftharpoons yC + zD$$

$$Keq = \frac{[products]}{[reactants]} = \frac{[C]^y\,[D]^z}{[A]^w[B]^x}$$

Now, let's review what the above energy diagram tells us about the thermodynamics of the reaction. Note that the energy level of the products is lower than that of the reactants. This tells us that the Gibbs free-energy change for the reaction is negative, and the step is exergonic, or energy releasing. We can also say the reaction is 'thermodynamically favorable', or, more informally, 'downhill'.

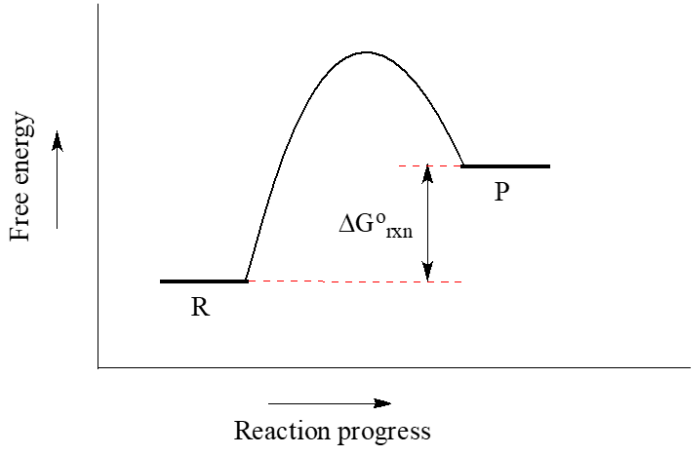
Recall from General Chemistry that the standard Gibbs free energy change for a reaction can be related to the reaction's equilibrium constant (Keq) by the equation:

$$\Delta G^o = -RT \ln K_{eq} \tag{10.2.2}$$

where R is the gas constant (8.314 J/mol×K) and T is the temperature in Kelvin (K).

If you do the math, you see that a negative value for ΔG_{rnx}° (an exergonic reaction) corresponds to K_{eq} being greater than 1, an equilibrium constant which favors product formation.

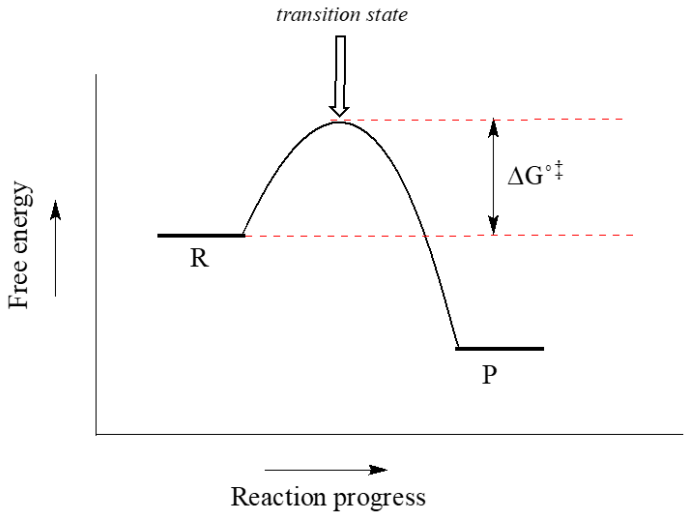
Conversely, an endergonic reaction is one in which the products are higher in energy than the reactants, and energy is absorbed. An endergonic reaction has a positive value of ΔG_{rnx}°, and a K_{eq} between 0 and 1.



Acid-base reactions provide convenient examples of thermodynamically favorable and unfavorable reactions. The reaction of a strong acid like HCl with a strong base like hydroxide ion, for example, is highly favorable, and has an equilibrium constant much greater than one. The reaction of a weak acid such as acetic acid with a weak base such as water, on the other hand, is unfavorable and has an equilibrium constant that is a very small (much less than 1) positive number: we can visualize this in a reaction coordinate diagram as an 'uphill' reaction, in which ΔG_{rnx}° is positive.

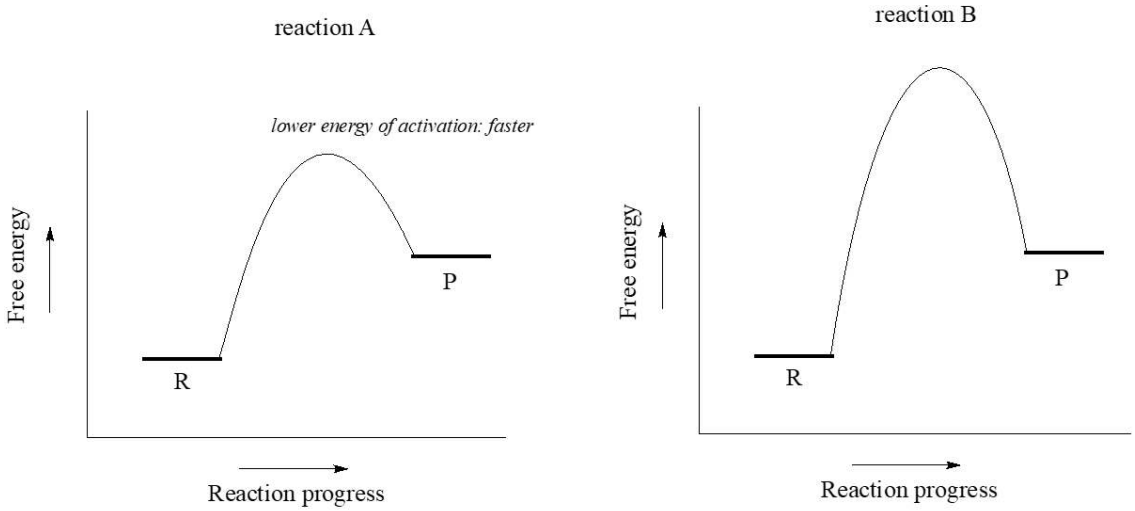
When talking about exergonic, or 'downhill' reactions, chemists sometimes use the term 'driving force' to describe the chemical factor or factors that drive the reaction from higher energy reactant to lower energy product. Using the 'downhill' analogy again, when water flows downhill, the driving force is gravity. In an exergonic chemical reaction, the driving force generally is based on a combination of two factors: a) the stability of positive and negative charges in the product relative to those in the reactant, and b) the total bond energy in the product relative to the reactant. That may not make a lot of sense right now, but keep it in the back of your mind and we will come back to the idea of driving force when we study different reaction types in greater detail.

Now, let's move to kinetics. Look again at the diagram for an exergonic reaction: although it is 'downhill' overall, it isn't a straight downhill run.



First, an 'energy barrier' must be overcome to get to the product side. The height of this energy barrier is called the standard free energy of activation ($\Delta G^{\circ}_{+}{}^{+}$). The activation energy, in combination with the temperature at which the reaction is being run, determines the rate of a reaction: the higher the activation energy, the slower the reaction. At the very top of the energy barrier, the reaction is at its transition state (TS), which you should recall is defined as the highest energy structure in the transition from reactant to product.

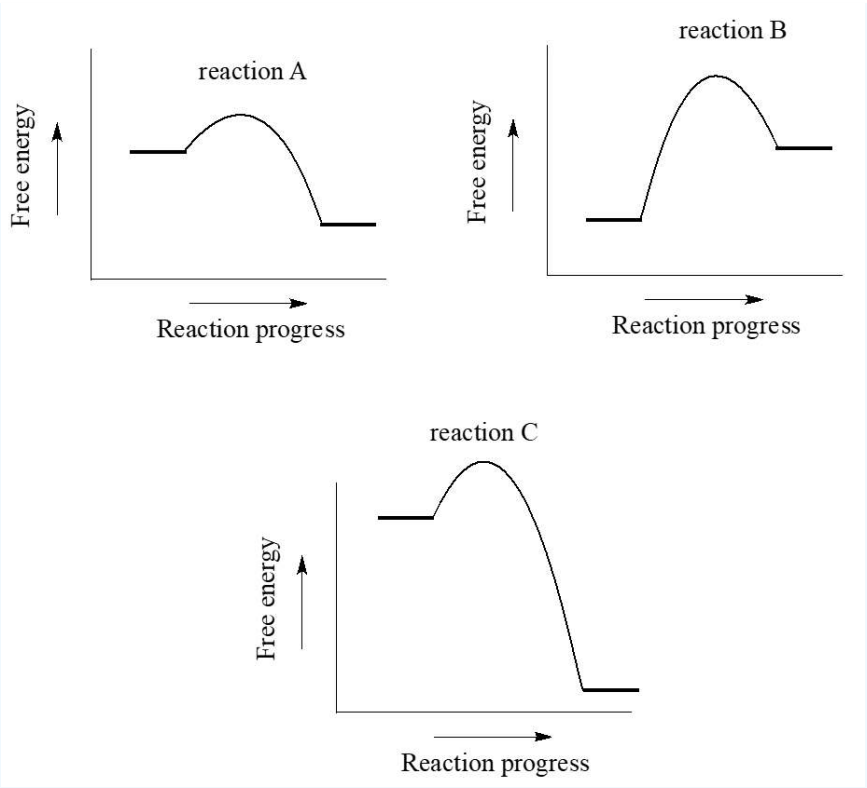
Consider the hypothetical reaction reaction coordinate diagrams below.



Both reaction A and reaction B are slightly endergonic, or uphill: ΔG°_{rxn} for both is positive, meaning that K_{eq} for both is between 0 and 1. However, the energy of activation is higher for reaction B. From this observation, we know that reaction A will proceed faster than reaction B in both forward and reverse directions (temperature and other conditions being equal), so reaction A will reach equilibrium in less time.

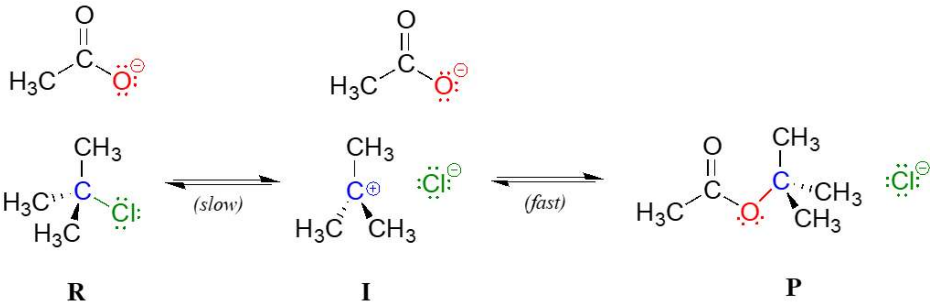
? Exercise 6.3.1

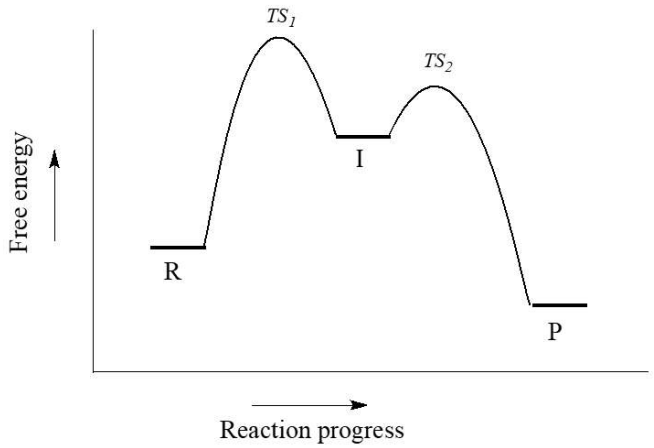
Consider the hypothetical reaction coordinate diagrams below, and assume that they are on the same scale.



a. Which of the diagrams describe(s) a reaction with $K_{eq} < 1$?
b. Which of the diagrams describes the fastest reaction?
c. Which of the diagrams describes the reaction with the highest value of K_{eq}?
d. Which of the diagrams describes the reaction with the largest $\Delta G^{\circ +}_{rxn}$ for the reverse reaction?
e. Copy the diagram for your answer to part (d), and add a label which graphically illustrates the value of $\Delta G^{\circ +}_{rxn}$ for the reaction in the reverse direction.

We turn our attention next to a two-step reaction mechanism, such as the nucleophilic substitution reaction between acetate and tert-butyl chloride. The reaction coordinate diagram for this reaction looks somewhat different from what we have seen until now:





Because there are two steps involved, there are also two transition states and two activation energies to consider, as well as a carbocation intermediate (denoted by the letter I). The first, bond-breaking step from R to I, passing over transition state TS_1, can be depicted as a highly endergonic (uphill) reaction, because the carbocation-chloride ion pair is significantly higher in energy than the reactants. The second step, attack on the carbocation electrophile by the acetate nucleophile and formation of the new carbon-oxygen bond, is a highly exergonic step that passes over a second, lower energy transition state TS_2. The intermediate (I) is thus depicted as an energy 'valley' (a local energy minimum) situated between the two energy peaks TS_1 and TS_2.

Notice that the activation energy for the first step is higher than the activation energy for the second, meaning that the first step is slower. This should make intuitive sense, because the first step involves bond-breaking, separation of charge, and formation of a carbocation, which is high in energy due to lacking a complete octet. Conversely, the second step involves bond-forming and neutralization of charge. In a multi-step reaction, the slowest step - the step with the highest energy of activation - is referred to as the rate-determining step (rds). The rds can be thought of as the 'bottleneck' of the reaction: a factor which affects the rds will affect the overall rate of the reaction. Conversely, a factor which affects a much faster step will not significantly affect the rate of the overall reaction.

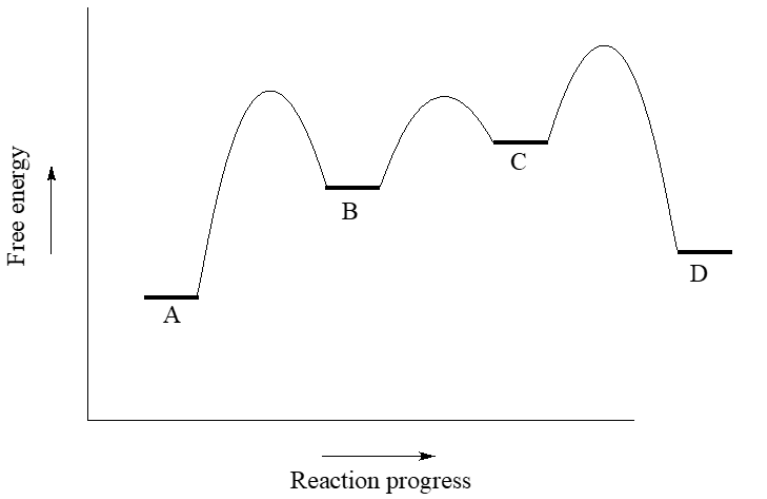
? Exercise 6.3.2

Imagine that you are trying to extinguish a burning campfire using buckets of water filled from a faucet some distance from the fire. It takes 20 seconds to fill a bucket at the faucet, and two seconds to carry the bucket to the fire and dump it on the flames. You have plenty of people to carry the buckets, but only one faucet so you can only fill one bucket at a time.

a. If you double the speed at which you carry the buckets by running instead of walking, what effect will this have on how fast you get water to the fire?
b. If, instead, you realize you can double the speed at which you fill up the buckets by using a second faucet, what effect will this have on how fast you get water to the fire?
c. What is the rate-determining step for the process?

? Exercise 6.3.3

Use the reaction coordinate diagram below to answer the questions.



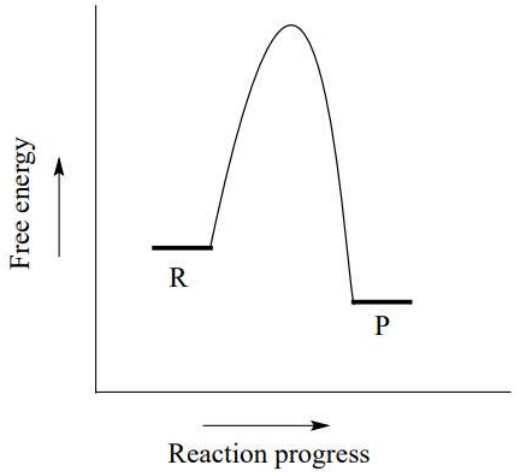
a. Is the overall reaction endergonic or exergonic in the forward (A to D) direction?
b. How many steps does the reaction mechanism have?
c. How many intermediates does the reaction mechanism have?
d. Redraw the diagram and add a label showing the activation energy for the rate-determining step of the forward reaction.
e. Add a label showing ΔG°_{rxn} for the reverse reaction (D to A).
f. What is the fastest reaction step, considering both the forward and reverse directions?

This page titled 10.2: A Quick Review of Thermodynamics and Kinetics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg.

- **6.3: A Quick Review of Thermodynamics and Kinetics** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

10.3: Catalysis

Consider a hypothetical reaction $\mathbf{R} \rightarrow \mathbf{P}$ described by the diagram below.



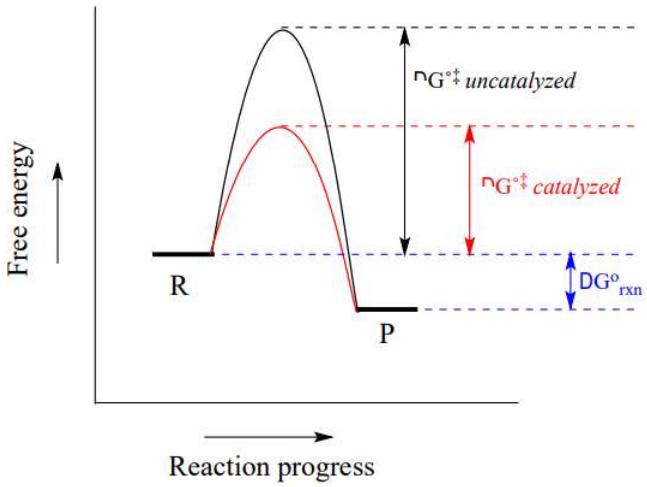
We notice two things about this reaction: it is exergonic, and it has a high activation energy. What this means is that although it is thermodynamically favorable, it is also slow: in other words, equilibrium favors product over reactants, but it will take a long time to reach equilibrium.

There are three ways that we could increase the rate of the reaction. First, we could add energy to the system by increasing the temperature, which gives the reacting molecules more energy to pass over the transition state. Increasing the temperature will increase the value of the rate constant k in the rate expression:

$$\text{rate} = k[R] \tag{10.3.1}$$

In the laboratory, many organic reactions are run at high temperatures for this very purpose. We could also increase the concentration of the reaction R, which would increase the rate of the reaction without increasing the value of k.

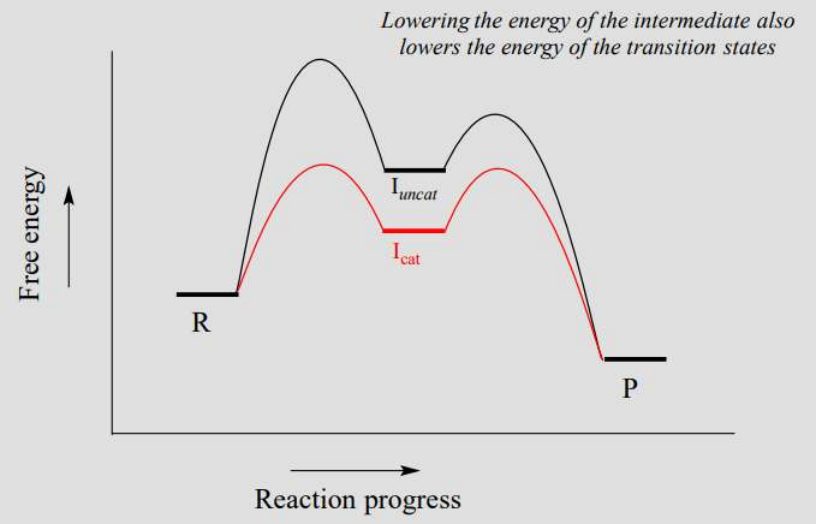
When talking about the biochemical reactions happening in a living cell, however, increasing the reaction temperature or reactant concentration is not a reasonable option. As an alternative, we could provide a new route from point R to point P in which the activation energy is lower. The role of a catalyst is to accelerate a reaction by stabilizing the transition state, and thus lowering the activation energy.



Catalysts, while they participate in the mechanism, are not consumed, so one catalyst molecule can catalyze multiple reaction cycles. Notice also that while the catalyst lowers the energy of the transition state (and thus the activation energy), it has no effect

on ΔG_{rxn}. A catalyst increases the rate of a reaction, but does not get consumed in the reaction and does not alter the equilibrium constant. In other words, a catalyst affects the kinetics of a reaction, but not the thermodynamics. Catalysts play a hugely important role in biochemical reactions.

Most organic reactions involve more than a single mechanistic step. Below is a reaction coordinate diagram illustrating rate acceleration of a two-step reaction by a catalyst:



Notice that the catalyst lowers the energy of the intermediate species. A concept known as the Hammond Postulate (the details of which are beyond our scope here) tells us that when a catalyst lowers the energy of an intermediate, it also lowers the energy of the adjacent transition states. Note in this diagram above that the energy barrier for the rate-determining first step is much lower in the catalyzed reaction - thus, the overall reaction is faster. When studying how an enzyme catalyzes a biochemical reaction, chemists often are actually looking at how the enzyme interacts with - and stabilizes - an intermediate species following a rate-determining step. The Hammond postulate tells us that an understanding of enzyme-intermediate interactions will also apply to enzyme-transition state interactions.

Acids and bases as are commonly used as catalysts in organic chemistry, and chemists have come up with a huge arsenal of catalysts, many of them metals, to speed up the rates of useful laboratory reactions. Almost all biochemical reactions are catalyzed by enzymes, which are protein catalysts. In the introduction to this chapter, we heard the story of the discovery of a heat-stable DNA polymerizing enzyme which turned out to be very useful to the scientific world.

How do enzymes accomplish their role as biochemical catalysts? Recall from section 1.3 that enzymes have an active site pocket in which substrates (reactant molecules) are bound. It is inside these active site pockets that most biochemical reactions take place. Enzymes achieve catalysis in the active site by some combination of the following:

1. By positioning two reacting molecules close to each other in the active site, in precisely the orientation necessary for them to react. Compare this to an uncatalyzed reaction in which completion depends on the two reactant molecules happening to collide, by chance, in the correct orientation.
2. By binding substrates in such a way that they assume the proper conformation necessary for a reaction to occur.
3. By increasing the reactivity of the substrates: making acidic protons more acidic, nucleophiles more nucleophilic, electrophiles more electrophilic, and leaving groups better at leaving. Very often, these feats are accomplished with acidic and/or basic amino acid side chains lining the active site pocket. As we go on to study many different types of biochemical reactions, we get a better of how this works.
4. By stabilizing the transition states of the slower, rate-determining steps of the reaction. If a transition state has a negative charge, for example, the enzyme might provide a positively charged amino acid side chain, or a bound metal cation such as Zn^{+2}, as a stabilizing factor. A lower-energy transition state, of course, means a lower activation energy and a faster reaction step.

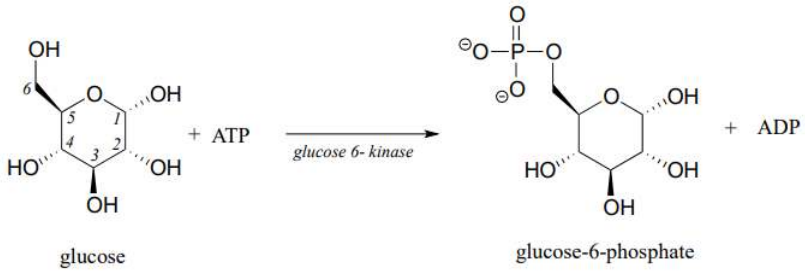
Enzymes are capable of truly amazing rate acceleration. Typical enzymes will speed up a reaction by anywhere from a million to a billion times, and the most efficient enzyme currently known to scientists is believed to accelerate its reaction by a factor of about 1017 over the uncatalyzed reaction (see Chemical and Engineering News, March 13, 2000, p. 42 for an interesting discussion about this nucleotide biosynthesis enzyme called 'orotidine monophosphate decarboxylase').

At this point, it is not necessary for you to fully understand the four ideas listed above: just keep them in mind as we go on to study a variety of common biological organic reactions and see in greater detail how enzymes have evolved to catalyze them.

? Exercise 6.4.1

Table sugar, or sucrose, is a high-energy dietary compound, as are the fats in vegetable oil. Conversion of these compounds, along with oxygen gas $\mathbf{O_2}$, to water and carbon dioxide releases a lot of energy. If they are both so high in energy (in other words, thermodyamically unstable), how can they sit for years on your kitchen shelf without reacting?

Another very important point to keep in mind about enzymes is the specificity with which they catalyze reactions. We have already discussed, in chapter 3, the idea that enzymes exert a very high level of control over the stereochemistry of a reaction: if two or more stereoisomeric products could potentially form in a reaction, an enzyme will likely only catalyze the formation of one stereoisomer, with negligible formation of other side products. Likewise, enzymes demonstrate remarkable control of regiochemistry in their reactions. The glycolysis pathway enzyme glucose-6-kinase, for example, transfers a phosphate group specifically to the hydroxyl group on carbon #6 of glucose, and not to any of the other four hydroxyl groups. We'll look more closely at this reaction and others like it in chapter 9.



glucose

glucose-6-phosphate

Finding ways to maintain control over stereochemistry and regiochemistry is a constant challenge for synthetic organic chemists working with non-enzymatic reactions, and the techniques that have been developed in this arena are a big part of what you will study if you go on to take a more advanced course focusing on organic synthesis.

Finally, we will encounter many biochemical reactions in this book in which the enzyme catalyzing the reaction does so with the assistance of a coenzyme. A coenzyme is a small (relative to a protein) molecule that binds in the active site of a partner enzyme and participates in some manner with the reaction being catalyzed. Table 6 at the back of the book shows the structures of several coenzymes commonly seen in biochemical reactions.

This page titled 10.3: Catalysis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Tim Soderberg.

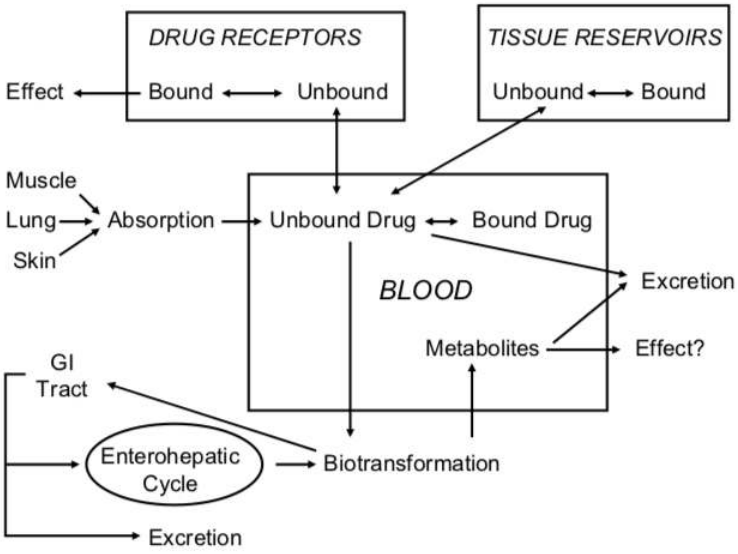
- **6.4: Catalysis** by Tim Soderberg is licensed CC BY-NC-SA 4.0. Original source: https://digitalcommons.morris.umn.edu/chem_facpubs/1/.

10.4: Pharmacology as an Equilibrium and Kinetics Problem

A drug is a chemical agent which can affect living processes. For purposes of this course we will mainly be talking about small molecules which affect cellular processes. Most of these are Xenobiotics (Gr. xenos - stranger) chemicals that are not synthesized by the body, but introduced into it from outside. There is inevitably a certain amount of ambiguity in this definition: Is oxygen or water a drug? How about Vitamin C in a glass of orange juice? How about an injection of Vitamin C to treat scurvy?

Pharmacology (Gr. pharmakon - a drug or poison, logos - word or discourse) is the science dealing with actions of drugs on the body (pharmacodynamics) and the fate of drugs in the body (pharmacokinetics). It overlaps with pharmacy, the science of preparation of drugs; much of it deals with therapeutics, the treatment of disease (by whatever means). Toxicology is the branch of pharmacology dealing with the "undesirable" effects of drugs on biological processes (in the case of a nerve gas the bad effect may be a desired one).

In order for a drug to work, it must enter the body and somehow be distributed in such a way that it gets to its site of action. In most cases the site of action is a macromolecular "receptor" located in the target tissue. Most drug effects are temporary, because the body has systems for drug detoxification and elimination. We will consider these issues broadly for now and go into more depth in individual lectures. As you read, refer to the figure below:



Overview of Pharmacokinetics - "What the body does to the drug"

1. The drug may enter the body in a variety of ways: as an oral liquid, pill, or capsule; as an inhaled vapor or aerosol; absorbed through intact skin or a mucous membrane; injected into muscle, subcutaneous tissue, spinal fluid, or directly into the bloodstream. As we shall see, the physical properties of the drug and the specific way it is prepared greatly influence the speed of absorption.
2. If the drug is given orally and swallowed, it must be absorbed from the GI tract into the portal circulation. If it is absorbed from the skin, mouth, lungs or muscle it will go directly into the systemic circulation. If drug is injected directly into the bloodstream (e.g., intravenous injection), 100% of it is available for distribution to tissues. This is not usually the case for other modes of administration. For example, drug which is absorbed via the portal circulation must first pass through the liver which is the primary site of drug metabolism (biotransformation). Some of the drug may therefore be metabolized before it ever reaches the systemic blood. In this case,"**first-pass**" **metabolism** reduces the **bioavailability** to less than 100%.
3. Once the drug is in the bloodstream a portion of it may exist as free drug, dissolved in plasma water. Some drug will be reversibly taken up by red cells and some will be reversibly bound to plasma proteins. For many drugs, the bound forms can account for 95-98% of the total. This is important because it is the free drug which traverses cell membranes and produces the effect. It is also important because protein-bound drug can act as a reservoir which releases drug slowly and thus prolongs its action.

4. The unbound drug may then follow its concentration gradient and distribute into peripheral tissues. In some cases, the tissue contains the target site and in others the tissue is not affected by the drug. Sites of non-specific binding act as further reservoirs for the drug. This total **volume of distribution** determines the equilibrium concentration of drug after a specified dose.
5. Tissue-bound drug eventually reenters the bloodstream where it perfuses the liver and kidneys. The liver metabolizes most drugs into inactive or less active compounds which are more readily excreted. These metabolites and some of the parent compound may be excreted in the bile and eventually may pass out of the body in the feces. Alternatively, some of the drug may be reabsorbed again, farther down the GI tract (the so-called enterohepatic cycle). Any biotransformed drug which is not excreted in bile passes back into the systemic circulation.
6. Parent drug and metabolites in the bloodstream may then be excreted: most are filtered by the kidney, where a portion undergoes reabsorption, and the remainder is excreted in the urine. Some drugs are actively secreted into the renal tubule. Another route of excretion is the lung: Drugs like alcohol and the anesthetic gases are eliminated by this route. Smaller amounts of drug are eliminated in the sweat, tears and breast milk.
7. Biotransformation may sometimes produce metabolites with a great deal of activity. Occasionally, we administer a parent drug which is inactive (a pro-drug) and only the metabolite has activity. [How might this be useful?]

Overview of Pharmacodynamics - "What the drug does to the body"

As stated above, the majority of drugs bind to specific receptors on the surface or interior of cells, but there are many other cellular components and non-specific sites which can serve as sites of drug action.

1. **Water** can be a target. Osmotic diuretics like mannitol are not reabsorbed by the kidney, and the osmotic load they create in the renal tubule obligates the loss of water. Laxatives like magnesium sulfate work in the intestine by the same principle.
2. **Hydrogen** ions can be targets. Ammonium chloride is sometimes used to acidify the urine. When it is taken orally, the liver metabolizes ammonium ion to urea, while the chloride is excreted in the urine. The loss of Cl- obligates the loss of H+ in the urine, thus the pH is lowered.
3. **Metal** ions can be targets. Chelating agents like EDTA may be used to bind divalent cations like Pb++. Metal ions are most frequently drug targets in cases of poisoning.
4. **Enzymes** are targets of many therapeutically useful drugs. Drugs may inhibit enzymes by competitive, non-competitive, or irreversible blockade at a substrate or cofactor binding site. Digitalis glycosides increase myocardial contractility by inhibiting the membrane enzyme, Na+-K+ - ATPase. Antimicrobial and antineoplastic drugs commonly work by inhibiting enzymes which are critical to the functioning of the cell. In order to be effective, these drugs must have at least **someselective toxicity** toward bacterial or tumor cells. This usually means that there is a unique metabolic pathway in these cells or some difference in enzyme selectivity for a common metabolic pathway. An example of this is the inhibition of folate synthesis by sulfonamides. These drugs are effective antibacterial agents because the bacteria depend upon folate synthesis, while the host doesn't. This example will be covered in detail in one of our case discussions.
5. **Nucleic acids** are targets for antimetabolites and some antibiotics. In the case of 5- fluorouracil, the compound acts as a counterfeit substitute for uracil and becomes incorporated into a faulty mRNA. Antisense oligonucleotides are another very specific way to interfere with a restricted part of the genome.
6. Some drugs, like general anesthetics, appear to act by non-specific **binding to a macromolecular receptor** target. These drugs are thought to alter the function of membrane proteins, in part, by disordering the structure of the surrounding lipid membranes. Their lack of specificity is reflected in very low chemical structural requirements. The general anesthetics include compounds as chemically diverse as nitrogen, xenon, halogenated ethers, and steroids. They exhibit very little stereoselectivity, that is, there are not marked differences in anesthetic activity between enantiomers.
7. Finally, we have the drugs which act by binding to **specific receptors**. As you will see in lectures 2 and 6, these drugs have both high structural specificity and stereoselectivity, i.e. relatively small changes in chemical structure can radically alter the activity of these drugs.

Let us finish with some important definitions. These are concepts which we will return to repeatedly throughout the course.

- **Agonist** is a drug which binds to its "receptor" and produces its characteristic effect. A drug may be a **full agonist** or **partial agonist**, depending on the maximal effect it produces. An antagonist binds to the receptor without causing an effect, thereby preventing an active substance from gaining access. Antagonists, like enzyme inhibitors, may be competitive, non-competitive or irreversible.

- **Dose-Response**. The sine qua non of drug effect. Simply put, as the dose of drug increases, the response should increase. [What if the response increases, then decreases as the dose is raised?] The curve generated is usually sigmoidal when effect is plotted against log dose (Dr. Strichartz will discuss the theoretical basis for this). Effect may be measured as a graded variable (change in blood pressure, force of contraction) or as a quantal variable (number dead/alive). The slope of the curve is characteristic of the particular drug-receptor interaction. When two drugs act by the same receptor mechanism, we expect to see two parallel log-dose response curves.
- **ED_{50}**. The median effective dose, or the dose which produces a response in 50% of subjects. If the response is death (lethality) we call it the **LD_{50}**. The EC50 refers to concentration rather than dose. Similar abbreviations are used for other response levels: ED_{99}, LD_1, etc.
- **Potency**. A terribly misused word – the lay public uses it to mean "effectiveness." The potency of a drug refers to the dose (actually the molar concentration) required to produce a specific intensity of effect. [We usually specify the ED50, why?] If the ED50of drug A and B are 5 and 10 mg, respectively, the Relative Potency of A is twice that of B. Relative potency specifically applies to the comparison of drugs which act by the same mechanism, and therefore have parallel dose-response curves.
- **Efficacy**. Also called **Maximal Efficacy** or **Intrinsic Activity**. This is the maximum effect of which the drug is capable. A potent drug may have a low efficacy, and a highly efficacious drug may have a low potency. For the clinician, efficacy is much more important than potency (within limits). Who cares if the pill contains 5 or 10 mg of drug?
- **Affinity**. This refers to the strength of binding between drug and receptor. It is quantified by the dissociation constant kD (covered in the next lecture).
- **Selectivity**. This refers to the separation between desired and undesired effects of a drug. In the ideal case, a drug is completely **specific**, and an effective dose does not elicit any undesired effect. Penicillin is an example of a highly selective drug, since it works specifically by inhibiting cell wall synthesis, and (other than allergic responses) it has very little effect on human cells at normal doses. Unfortunately, many therapeutic agents, like digoxin and theophylline, produce dose-related side effects near their therapeutic dose range. For some drugs like cancer chemotherapeutic agents, their selectivity is their dose-limiting property, i.e., they are given to kill tumor cells until they produce toxicity in normal cells as well.
- **Therapeutic Window**. For every drug, there exists some concentration which is just barely effective (the **Effective Concentration**) and some dose which is just barely toxic (the **Toxic Concentration**). Between them is the therapeutic window where most safe and effective treatment will occur.
- **Therapeutic Index**. This is the ratio of toxic to effective doses at the level of 50% response: TD_{50}/ED_{50}. In animal toxicology studies, it is usually the LD_{50}/ED_{50}. Another measure sometimes utilized is the **Certain Safety Factor**, which is TD_1/ED_{99}.

10.4: Pharmacology as an Equilibrium and Kinetics Problem is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

11: Molecular Genetics

The blueprint for the reproduction and the maintenance of each organism is found in the nuclei of its cells, concentrated in elongated, threadlike structures called chromosomes. These complex structures, consisting of DNA and proteins, contain the basic units of heredity, called genes. The number of chromosomes (and genes) varies with each species. Human body cells have 23 pairs of chromosomes having 20,000–40,000 different genes.

Sperm and egg cells contain only a single copy of each chromosome; that is, they contain only one member of each chromosome pair. Thus, in sexual reproduction, the entire complement of chromosomes is achieved only when an egg and sperm combine. A new individual receives half its hereditary material from each parent. Calling the unit of heredity a “gene” merely gives it a name. But what really are genes and how is the information they contain expressed? One definition of a gene is that it is a segment of DNA that constitutes the code for a specific polypeptide. If genes are segments of DNA, we need to learn more about the structure and physiological function of DNA. We begin by looking at the small molecules needed to form DNA and RNA (ribonucleic acid)—the nucleotides.

11.1: Prelude to Nucleic Acids
11.2: Nucleotides
11.3: Nucleic Acid Structure
11.4: Replication and Expression of Genetic Information
11.5: Protein Synthesis and the Genetic Code
11.6: Mutations and Genetic Diseases
11.7: Viruses
11.8: Nucleic Acids (Exercises)
11.9: Nucleic Acids (Summary)

Template:HideTOC

11: Molecular Genetics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

11.1: Prelude to Nucleic Acids

Following the initial isolation of insulin in 1921, diabetic patients could be treated with insulin obtained from the pancreases of cattle and pigs. Unfortunately, some patients developed an allergic reaction to this insulin because its amino acid sequence was not identical to that of human insulin. In the 1970s, an intense research effort began that eventually led to the production of genetically engineered human insulin—the first genetically engineered product to be approved for medical use. To accomplish this feat, researchers first had to determine how insulin is made in the body and then find a way of causing the same process to occur in nonhuman organisms, such as bacteria or yeast cells. Many aspects of these discoveries are presented in this chapter on nucleic acids.



Figure **11.1.1**: A vial of insulin. It has been given a trade name, Actrapid, by the manufacturer. (Public Domain; Mr Hyde).

11.1: Prelude to Nucleic Acids is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.0: Prelude to Nucleic Acids** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

11.2: Nucleotides

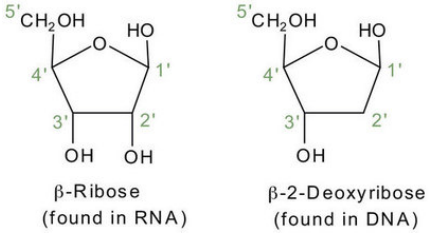
Learning Objectives

- To identify the different molecules that combine to form nucleotides.

The repeating, or monomer, units that are linked together to form nucleic acids are known as nucleotides. The deoxyribonucleic acid (DNA) of a typical mammalian cell contains about 3×10^9 nucleotides. Nucleotides can be further broken down to phosphoric acid (H_3PO_4), a pentose sugar (a sugar with five carbon atoms), and a nitrogenous base (a base containing nitrogen atoms).

$$\text{nucleic acids} \xrightarrow[\text{down into}]{\text{can be broken}} \text{nucleotides} \xrightarrow[\text{down into}]{\text{can be broken}} H_3PO_4 + \text{nitrogen base} + \text{pentose sugar}$$

If the pentose sugar is ribose, the nucleotide is more specifically referred to as a *ribonucleotide*, and the resulting nucleic acid is ribonucleic acid (RNA). If the sugar is 2-deoxyribose, the nucleotide is a *deoxyribonucleotide*, and the nucleic acid is DNA.



The nitrogenous bases found in nucleotides are classified as pyrimidines or purines. Pyrimidines are heterocyclic amines with two nitrogen atoms in a six-member ring and include uracil, thymine, and cytosine. Purines are heterocyclic amines consisting of a pyrimidine ring fused to a five-member ring with two nitrogen atoms. Adenine and guanine are the major purines found in nucleic acids (Figure **11.2.1**).

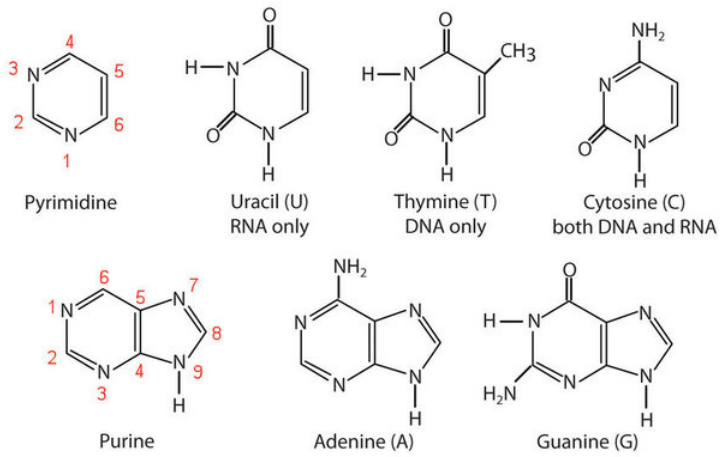


Figure **11.2.1**: The Nitrogenous Bases Found in DNA and RNA

The formation of a bond between C1′ of the pentose sugar and N1 of the pyrimidine base or N9 of the purine base joins the pentose sugar to the nitrogenous base. In the formation of this bond, a molecule of water is removed. Table **11.2.1** summarizes the similarities and differences in the composition of nucleotides in DNA and RNA.

> The numbering convention is that primed numbers designate the atoms of the pentose ring, and unprimed numbers designate the atoms of the purine or pyrimidine ring.

Table **11.2.1**: Composition of Nucleotides in DNA and RNA

Composition	DNA	RNA
purine bases	adenine and guanine	adenine and guanine
pyrimidine bases	cytosine and thymine	cytosine and uracil
pentose sugar	2-deoxyribose	ribose
inorganic acid	phosphoric acid (H_3PO_4)	H_3PO_4

The names and structures of the major ribonucleotides and one of the deoxyribonucleotides are given in Figure **11.2.2**.

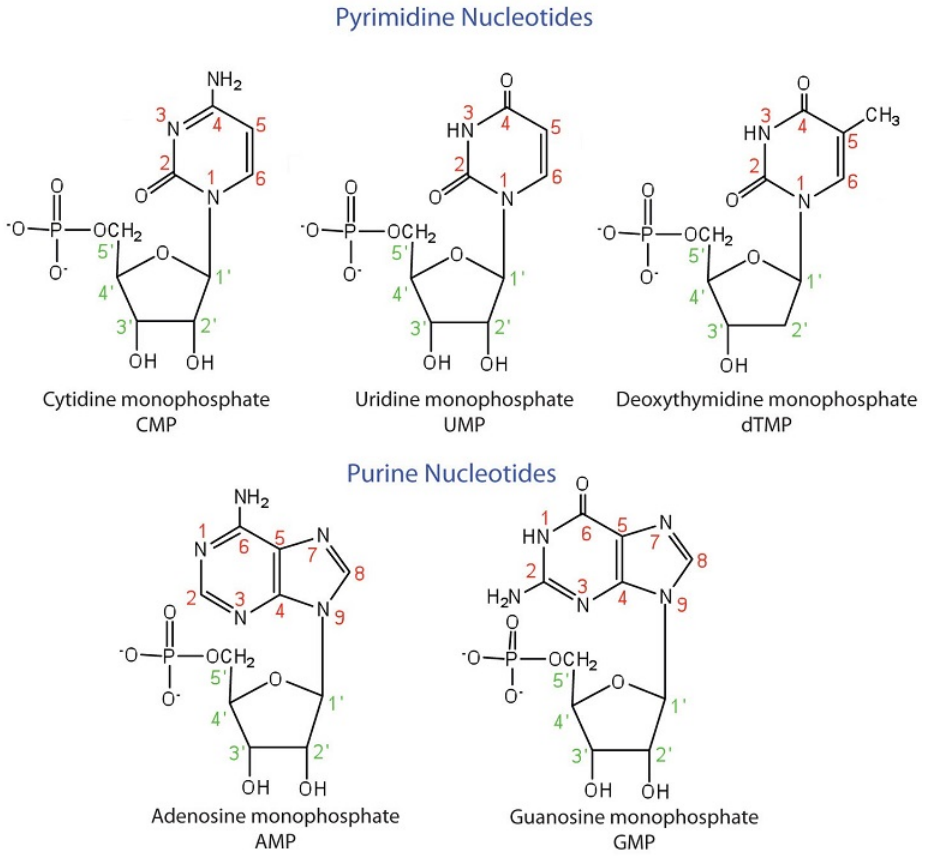


Figure **11.2.2**: The Pyrimidine and Purine Nucleotides

Apart from being the monomer units of DNA and RNA, the nucleotides and some of their derivatives have other functions as well. Adenosine diphosphate (ADP) and adenosine triphosphate (ATP), shown in Figure **11.2.3**, have a role in cell metabolism. Moreover, a number of coenzymes, including flavin adenine dinucleotide (FAD), nicotinamide adenine dinucleotide (NAD^+), and coenzyme A, contain adenine nucleotides as structural components.

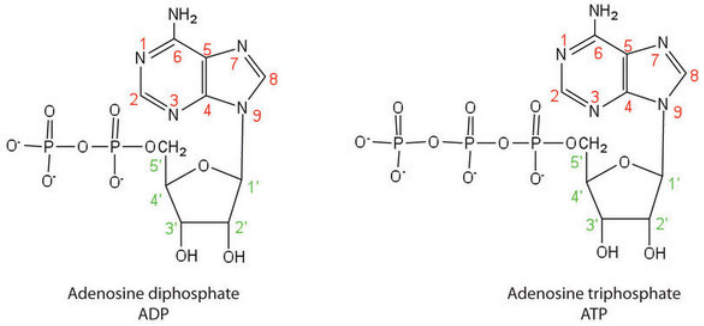


Figure **11.2.3**: Structures of Two Important Adenine-Containing Nucleotides

Summary

Nucleotides are composed of phosphoric acid, a pentose sugar (ribose or deoxyribose), and a nitrogen-containing base (adenine, cytosine, guanine, thymine, or uracil). Ribonucleotides contain ribose, while deoxyribonucleotides contain deoxyribose.

11.2: Nucleotides is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.1: Nucleotides** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

11.3: Nucleic Acid Structure

Skills to Develop

- Identify the two types of nucleic acids and the function of each type.
- Describe how nucleotides are linked together to form nucleic acids.
- Describe the secondary structure of DNA and the importance of complementary base pairing.

Nucleic acids are large polymers formed by linking nucleotides together and are found in every cell. Deoxyribonucleic acid (DNA) is the nucleic acid that stores genetic information. If all the DNA in a typical mammalian cell were stretched out end to end, it would extend more than 2 m. Ribonucleic acid (RNA) is the nucleic acid responsible for using the genetic information encoded in DNA to produce the thousands of proteins found in living organisms.

Primary Structure of Nucleic Acids

Nucleotides are joined together through the phosphate group of one nucleotide connecting in an ester linkage to the OH group on the third carbon atom of the sugar unit of a second nucleotide. This unit joins to a third nucleotide, and the process is repeated to produce a long nucleic acid chain (Figure **11.3.1**). The backbone of the chain consists of alternating phosphate and sugar units (2-deoxyribose in DNA and ribose in RNA). The purine and pyrimidine bases branch off this backbone.

> *Each phosphate group has one acidic hydrogen atom that is ionized at physiological pH. This is why these compounds are known as nucleic acids.*

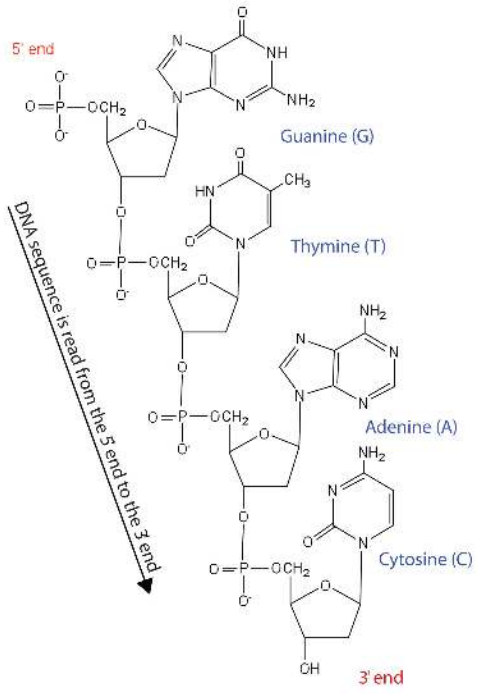


***Figure* 11.3.1** *Structure of a Segment of DNA. A similar segment of RNA would have OH groups on each C2', and uracil would replace thymine.*

Like proteins, nucleic acids have a primary structure that is defined as the sequence of their nucleotides. Unlike proteins, which have 20 different kinds of amino acids, there are only 4 different kinds of nucleotides in nucleic acids. For amino acid sequences in proteins, the convention is to write the amino acids in order starting with the N-terminal amino acid. In writing nucleotide sequences for nucleic acids, the convention is to write the nucleotides (usually using the one-letter abbreviations for the bases, shown in Figure **11.3.1**) starting with the nucleotide having a free phosphate group, which is known as the 5′ end, and indicate the nucleotides in order. For DNA, a lowercase *d* is often written in front of the sequence to indicate that the monomers are

deoxyribonucleotides. The final nucleotide has a free OH group on the 3′ carbon atom and is called the *3′ end*. The sequence of nucleotides in the DNA segment shown in Figure **11.3.1** would be written 5′-dG-dT-dA-dC-3′, which is often further abbreviated to dGTAC or just GTAC.

Secondary Structure of DNA

The three-dimensional structure of DNA was the subject of an intensive research effort in the late 1940s to early 1950s. Initial work revealed that the polymer had a regular repeating structure. In 1950, Erwin Chargaff of Columbia University showed that the molar amount of adenine (A) in DNA was always equal to that of thymine (T). Similarly, he showed that the molar amount of guanine (G) was the same as that of cytosine (C). Chargaff drew no conclusions from his work, but others soon did.

At Cambridge University in 1953, James D. Watson and Francis Crick announced that they had a model for the secondary structure of DNA. Using the information from Chargaff's experiments (as well as other experiments) and data from the X ray studies of Rosalind Franklin (which involved sophisticated chemistry, physics, and mathematics), Watson and Crick worked with models that were not unlike a child's construction set and finally concluded that DNA is composed of two nucleic acid chains running antiparallel to one another—that is, side-by-side with the 5′ end of one chain next to the 3′ end of the other. Moreover, as their model showed, the two chains are twisted to form a double helix—a structure that can be compared to a spiral staircase, with the phosphate and sugar groups (the backbone of the nucleic acid polymer) representing the outside edges of the staircase. The purine and pyrimidine bases face the inside of the helix, with guanine always opposite cytosine and adenine always opposite thymine. These specific base pairs, referred to as complementary bases, are the steps, or treads, in our staircase analogy (Figure **11.3.2**).

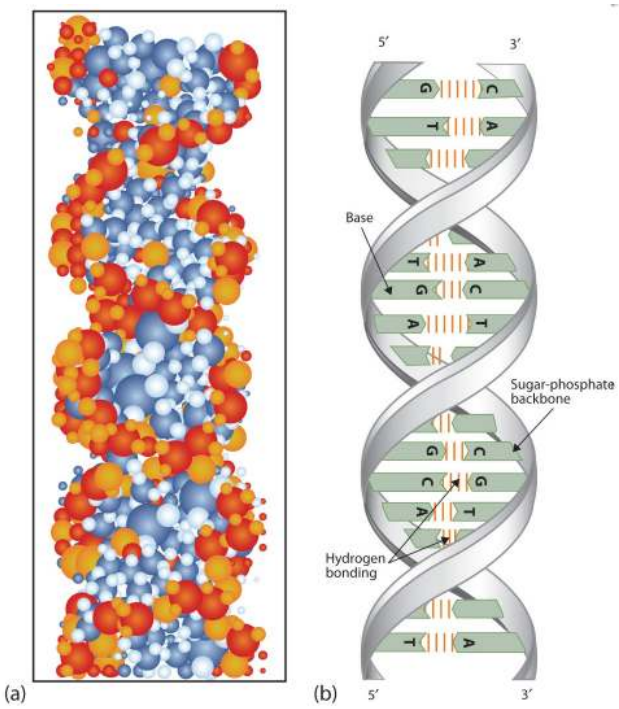


***Figure* 11.3.2** *DNA Double Helix. (a) This represents a computer-generated model of the DNA double helix. (b) This represents a schematic representation of the double helix, showing the complementary bases.*

The structure proposed by Watson and Crick provided clues to the mechanisms by which cells are able to divide into two identical, functioning daughter cells; how genetic data are passed to new generations; and even how proteins are built to required specifications. All these abilities depend on the pairing of complementary bases. Figure **11.3.3** shows the two sets of base pairs and illustrates two things. First, a pyrimidine is paired with a purine in each case, so that the long dimensions of both pairs are identical (1.08 nm).

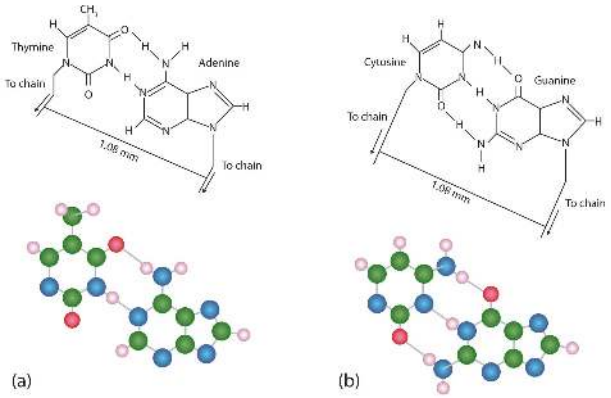


***Figure* 11.3.3** *Complementary Base Pairing. Complementary bases engage in hydrogen bonding with one another: (a) thymine and adenine; (b) cytosine and guanine.*

If two pyrimidines were paired or two purines were paired, the two pyrimidines would take up less space than a purine and a pyrimidine, and the two purines would take up more space, as illustrated in Figure **11.3.4**. If these pairings were ever to occur, the structure of DNA would be like a staircase made with stairs of different widths. For the two strands of the double helix to fit neatly, a pyrimidine must always be paired with a purine. The second thing you should notice in Figure **11.3.3** is that the correct pairing enables formation of three instances of hydrogen bonding between guanine and cytosine and two between adenine and thymine. The additive contribution of this hydrogen bonding imparts great stability to the DNA double helix.

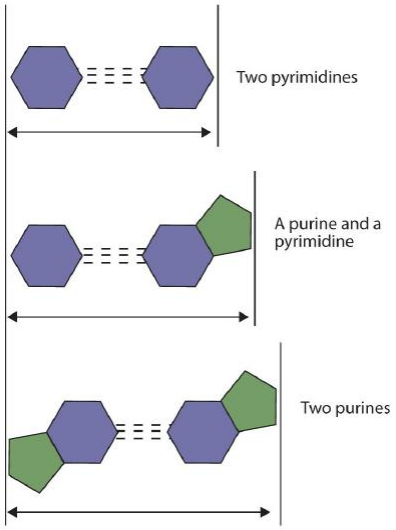


***Figure* 11.3.4** *Difference in Widths of Possible Base Pairs*

Summary

- DNA is the nucleic acid that stores genetic information. RNA is the nucleic acid responsible for using the genetic information in DNA to produce proteins.
- Nucleotides are joined together to form nucleic acids through the phosphate group of one nucleotide connecting in an ester linkage to the OH group on the third carbon atom of the sugar unit of a second nucleotide.
- Nucleic acid sequences are written starting with the nucleotide having a free phosphate group (the 5′ end).
- Two DNA strands link together in an antiparallel direction and are twisted to form a double helix. The nitrogenous bases face the inside of the helix. Guanine is always opposite cytosine, and adenine is always opposite thymine.

Concept Review Exercises

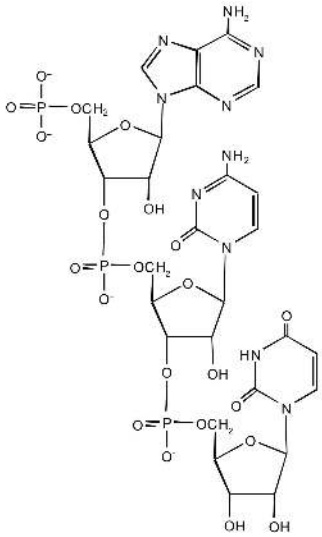
1.
   a. Name the two kinds of nucleic acids.
   b. Which type of nucleic acid stores genetic information in the cell?
2. What are complementary bases?
3. Why is it structurally important that a purine base always pair with a pyrimidine base in the DNA double helix?

Answers

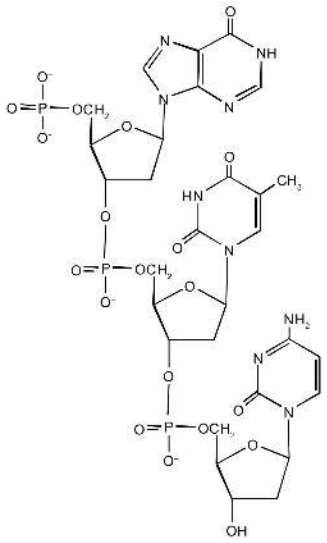
1.
   a. deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)
   b. DNA
2. the specific base pairings in the DNA double helix in which guanine is paired with cytosine and adenine is paired with thymine
3. The width of the DNA double helix is kept at a constant width, rather than narrowing (if two pyrimidines were across from each other) or widening (if two purines were across from each other).

Exercises

1. For this short RNA segment,
   a. identify the 5′ end and the 3′ end of the molecule.
   b. circle the atoms that comprise the backbone of the nucleic acid chain.
   c. write the nucleotide sequence of this RNA segment.



2. For this short DNA segment,
   a. identify the 5′ end and the 3′ end of the molecule.
   b. circle the atoms that comprise the backbone of the nucleic acid chain.
   c. write the nucleotide sequence of this DNA segment.



3. Which nitrogenous base in DNA pairs with each nitrogenous base?

   a. cytosine
   b. adenine
   c. guanine
   d. thymine

4. Which nitrogenous base in RNA pairs with each nitrogenous base?

   a. cytosine
   b. adenine
   c. guanine
   d. thymine

5. How many hydrogen bonds can form between the two strands in the short DNA segment shown below?

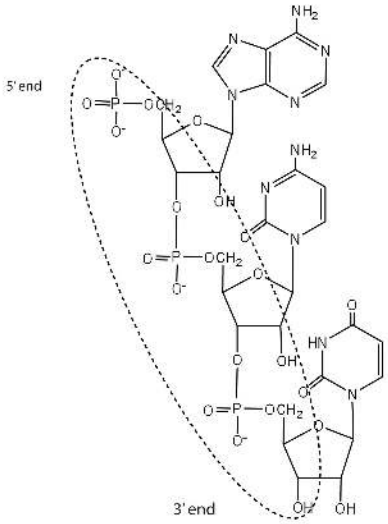
   5′ ATGCGACTA 3′ 3′ TACGCTGAT 5′

6. How many hydrogen bonds can form between the two strands in the short DNA segment shown below?

   5′ CGATGAGCC 3′ 3′ GCTACTCGG 5′

Answers

1.



c. ACU

3.

a. guanine
b. thymine
c. cytosine
d. adenine

5. 22 (2 between each AT base pair and 3 between each GC base pair)

11.3: Nucleic Acid Structure is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

11.4: Replication and Expression of Genetic Information

Learning Objectives

- Describe how a new copy of DNA is synthesized.
- Describe how RNA is synthesized from DNA.
- Identify the different types of RNA and the function of each type of RNA.

We previously stated that deoxyribonucleic acid (DNA) stores genetic information, while ribonucleic acid (RNA) is responsible for transmitting or expressing genetic information by directing the synthesis of thousands of proteins found in living organisms. But how do the nucleic acids perform these functions? Three processes are required: (1) *replication*, in which new copies of DNA are made; (2) *transcription*, in which a segment of DNA is used to produce RNA; and (3) *translation*, in which the information in RNA is translated into a protein sequence.

Replication

New cells are continuously forming in the body through the process of cell division. For this to happen, the DNA in a dividing cell must be copied in a process known as replication. The complementary base pairing of the double helix provides a ready model for how genetic replication occurs. If the two chains of the double helix are pulled apart, disrupting the hydrogen bonding between base pairs, each chain can act as a *template*, or pattern, for the synthesis of a new complementary DNA chain.

The nucleus contains all the necessary enzymes, proteins, and nucleotides required for this synthesis. A short segment of DNA is "unzipped," so that the two strands in the segment are separated to serve as templates for new DNA. DNA polymerase, an enzyme, recognizes each base in a template strand and matches it to the complementary base in a free nucleotide. The enzyme then catalyzes the formation of an ester bond between the 5′ phosphate group of the nucleotide and the 3′ OH end of the new, growing DNA chain. In this way, each strand of the original DNA molecule is used to produce a duplicate of its former partner (Figure **11.4.1**). Whatever information was encoded in the original DNA double helix is now contained in each replicate helix. When the cell divides, each daughter cell gets one of these replicates and thus all of the information that was originally possessed by the parent cell.

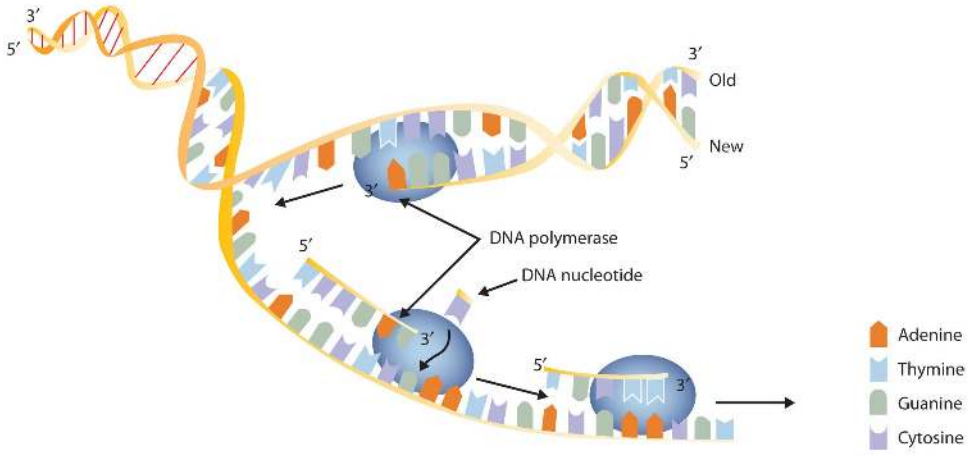


Figure **11.4.1**: A Schematic Diagram of DNA Replication. DNA replication occurs by the sequential unzipping of segments of the double helix. Each new nucleotide is brought into position by DNA polymerase and is added to the growing strand by the formation of a phosphate ester bond. Thus, two double helixes form from one, and each consists of one old strand and one new strand, an outcome called *semiconservative replications*. (This representation is simplified; many more proteins are involved in replication.)

Example 11.4.1

A segment of one strand from a DNA molecule has the sequence 5′-TCCATGAGTTGA-3′. What is the sequence of nucleotides in the opposite, or complementary, DNA chain?

Solution

Knowing that the two strands are antiparallel and that T base pairs with A, while C base pairs with G, the sequence of the complementary strand will be 3′-AGGTACTCAACT-5′ (can also be written as TCAACTCATGGA).

? Exercise 11.4.1

A segment of one strand from a DNA molecule has the sequence 5′-CCAGTGAATTGCCTAT-3′. What is the sequence of nucleotides in the opposite, or complementary, DNA chain?

What do we mean when we say information is encoded in the DNA molecule? An organism's DNA can be compared to a book containing directions for assembling a model airplane or for knitting a sweater. Letters of the alphabet are arranged into words, and these words direct the individual to perform certain operations with specific materials. If all the directions are followed correctly, a model airplane or sweater is produced.

In DNA, the particular sequences of nucleotides along the chains encode the directions for building an organism. Just as *saw* means one thing in English and *was* means another, the sequence of bases CGT means one thing, and TGC means something different. Although there are only four letters—the four nucleotides—in the genetic code of DNA, their sequencing along the DNA strands can vary so widely that information storage is essentially unlimited.

Transcription

For the hereditary information in DNA to be useful, it must be "expressed," that is, used to direct the growth and functioning of an organism. The first step in the processes that constitute DNA expression is the synthesis of RNA, by a template mechanism that is in many ways analogous to DNA replication. Because the RNA that is synthesized is a complementary copy of information contained in DNA, RNA synthesis is referred to as transcription. There are three key differences between replication and transcription:

1. RNA molecules are much shorter than DNA molecules; only a portion of one DNA strand is copied or transcribed to make an RNA molecule.
2. RNA is built from ribonucleotides rather than deoxyribonucleotides.
3. The newly synthesized RNA strand does not remain associated with the DNA sequence it was transcribed from.

The DNA sequence that is transcribed to make RNA is called the *template strand*, while the complementary sequence on the other DNA strand is called the *coding* or *informational strand*. To initiate RNA synthesis, the two DNA strands unwind at specific sites along the DNA molecule. Ribonucleotides are attracted to the uncoiling region of the DNA molecule, beginning at the 3′ end of the template strand, according to the rules of base pairing. Thymine in DNA calls for adenine in RNA, cytosine specifies guanine, guanine calls for cytosine, and adenine requires uracil. RNA polymerase—an enzyme—binds the complementary ribonucleotide and catalyzes the formation of the ester linkage between ribonucleotides, a reaction very similar to that catalyzed by DNA polymerase (Figure 11.4.2). Synthesis of the RNA strand takes place in the 5′ to 3′ direction, antiparallel to the template strand. Only a short segment of the RNA molecule is hydrogen-bonded to the template strand at any time during transcription. When transcription is completed, the RNA is released, and the DNA helix reforms. The nucleotide sequence of the RNA strand formed during transcription is identical to that of the corresponding coding strand of the DNA, except that U replaces T.

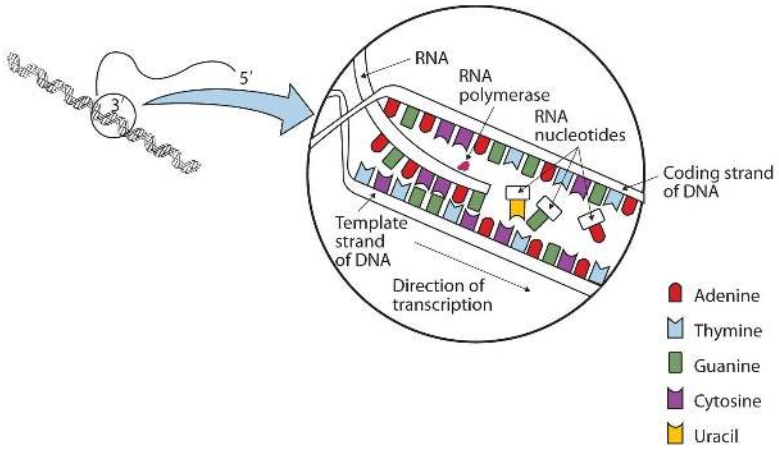


Figure 11.4.2: A Schematic Diagram of RNA Transcription from a DNA Template. The representation of RNA polymerase is proportionately much smaller than the actual molecule, which encompasses about 50 nucleotides at a time.

Example 11.4.2

A portion of the template strand of a gene has the sequence 5′-TCCATGAGTTGA-3′. What is the sequence of nucleotides in the RNA that is formed from this template?

Solution

Four things must be remembered in answering this question: (1) the DNA strand and the RNA strand being synthesized are antiparallel; (2) RNA is synthesized in a 5′ to 3′ direction, so transcription begins at the 3′ end of the template strand; (3) ribonucleotides are used in place of deoxyribonucleotides; and (4) thymine (T) base pairs with adenine (A), A base pairs with uracil (U; in RNA), and cytosine (C) base pairs with guanine (G). The sequence is determined to be 3′-AGGUACUCAACU-5′ (can also be written as 5′-UCAACUCAUGGA-3′).

Exercise 11.4.2

A portion of the template strand of a gene has the sequence 5′-CCAGTGAATTGCCTAT-3′. What is the sequence of nucleotides in the RNA that is formed from this template?

Three types of RNA are formed during transcription: *messenger RNA* (mRNA), *ribosomal RNA* (rRNA), and *transfer RNA* (tRNA). These three types of RNA differ in function, size, and percentage of the total cell RNA (Table 11.4.1). mRNA makes up only a small percent of the total amount of RNA within the cell, primarily because each molecule of mRNA exists for a relatively short time; it is continuously being degraded and resynthesized. The molecular dimensions of the mRNA molecule vary according to the amount of genetic information a given molecule contains. After transcription, which takes place in the nucleus, the mRNA passes into the cytoplasm, carrying the genetic message from DNA to the ribosomes, the sites of protein synthesis. Elsewhere, we shall see how mRNA directly determines the sequence of amino acids during protein synthesis.

Table 11.4.1: Properties of Cellular RNA in Escherichia coli

Type	Function	Approximate Number of Nucleotides	Percentage of Total Cell RNA
mRNA	codes for proteins	100–6,000	~3
rRNA	component of ribosomes	120–2900	83
tRNA	adapter molecule that brings the amino acid to the ribosome	75–90	14

Ribosomes are cellular substructures where proteins are synthesized. They contain about 65% rRNA and 35% protein, held together by numerous noncovalent interactions, such as hydrogen bonding, in an overall structure consisting of two globular particles of unequal size.

Molecules of tRNA, which bring amino acids (one at a time) to the ribosomes for the construction of proteins, differ from one another in the kinds of amino acid each is specifically designed to carry. A set of three nucleotides, known as a codon, on the mRNA determines which kind of tRNA will add its amino acid to the growing chain. Each of the 20 amino acids found in proteins has at least one corresponding kind of tRNA, and most amino acids have more than one.

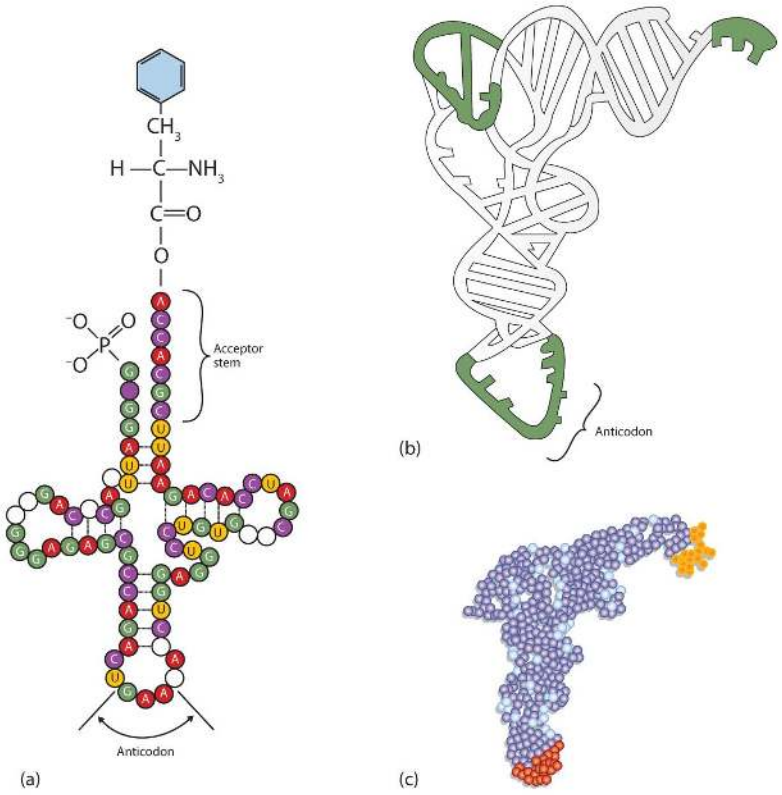


Figure 11.4.3: Transfer RNA. (a) In the two-dimensional structure of a yeast tRNA molecule for phenylalanine, the amino acid binds to the acceptor stem located at the 3′ end of the tRNA primary sequence. (The nucleotides that are not specifically identified here are slightly altered analogs of the four common ribonucleotides A, U, C, and G.) (b) In the three-dimensional structure of yeast phenylalanine tRNA, note that the anticodon loop is at the bottom and the acceptor stem is at the top right. (c) This shows a space-filling model of the tRNA.

The two-dimensional structure of a tRNA molecule has three distinctive loops, reminiscent of a cloverleaf (Figure 11.4.3). On one loop is a sequence of three nucleotides that varies for each kind of tRNA. This triplet, called the anticodon, is complementary to and pairs with the codon on the mRNA. At the opposite end of the molecule is the acceptor stem, where the amino acid is attached.

Summary

- In DNA replication, each strand of the original DNA serves as a template for the synthesis of a complementary strand.
- DNA polymerase is the primary enzyme needed for replication.
- In transcription, a segment of DNA serves as a template for the synthesis of an RNA sequence.
- RNA polymerase is the primary enzyme needed for transcription.
- Three types of RNA are formed during transcription: mRNA, rRNA, and tRNA.

11.4: Replication and Expression of Genetic Information is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.3: Replication and Expression of Genetic Information** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

11.5: Protein Synthesis and the Genetic Code

Learning Objectives

- describe the characteristics of the genetic code.
- describe how a protein is synthesized from mRNA.

One of the definitions of a gene is as follows: a segment of deoxyribonucleic acid (DNA) carrying the code for a specific polypeptide. Each molecule of messenger RNA (mRNA) is a transcribed copy of a gene that is used by a cell for synthesizing a polypeptide chain. If a protein contains two or more different polypeptide chains, each chain is coded by a different gene. We turn now to the question of how the sequence of nucleotides in a molecule of ribonucleic acid (RNA) is translated into an amino acid sequence.

How can a molecule containing just 4 different nucleotides specify the sequence of the 20 amino acids that occur in proteins? If each nucleotide coded for 1 amino acid, then obviously the nucleic acids could code for only 4 amino acids. What if amino acids were coded for by groups of 2 nucleotides? There are 4^2, or 16, different combinations of 2 nucleotides (AA, AU, AC, AG, UU, and so on). Such a code is more extensive but still not adequate to code for 20 amino acids. However, if the nucleotides are arranged in groups of 3, the number of different possible combinations is 4^3, or 64. Here we have a code that is extensive enough to direct the synthesis of the primary structure of a protein molecule.



***Video**: NDSU Virtual Cell Animations project animation "Translation". For more information, see VCell, NDSU(opens in new window) [vcell.ndsu.nodak.edu]*

The genetic code can therefore be described as the identification of each group of three nucleotides and its particular amino acid. The sequence of these triplet groups in the mRNA dictates the sequence of the amino acids in the protein. Each individual three-nucleotide coding unit, as we have seen, is called a *codon*.

Protein synthesis is accomplished by orderly interactions between mRNA and the other ribonucleic acids (transfer RNA [tRNA] and ribosomal RNA [rRNA]), the ribosome, and more than 100 enzymes. The mRNA formed in the nucleus during transcription is transported across the nuclear membrane into the cytoplasm to the ribosomes—carrying with it the genetic instructions. The process in which the information encoded in the mRNA is used to direct the sequencing of amino acids and thus ultimately to synthesize a protein is referred to as translation.

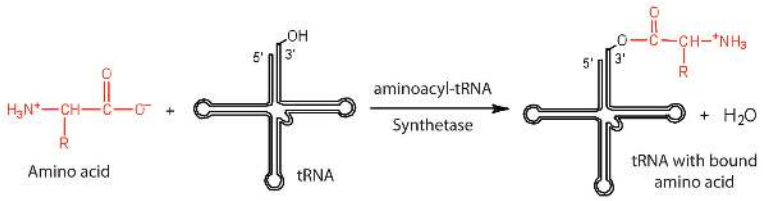


Figure 11.5.1: Binding of an Amino Acid to Its tRNA

Before an amino acid can be incorporated into a polypeptide chain, it must be attached to its unique tRNA. This crucial process requires an enzyme known as aminoacyl-tRNA synthetase (Figure 11.5.1). There is a specific aminoacyl-tRNA synthetase for each amino acid. This high degree of specificity is vital to the incorporation of the correct amino acid into a protein. After the amino acid molecule has been bound to its tRNA carrier, protein synthesis can take place. Figure 11.5.2 depicts a schematic stepwise representation of this all-important process.

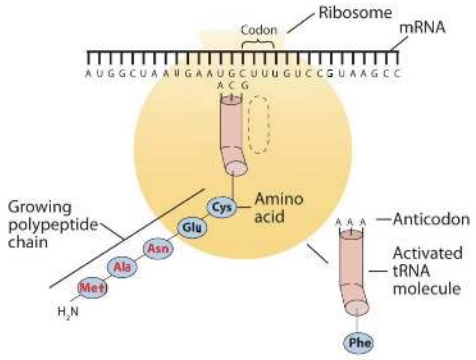


Figure 11.5.2a: The Elongation Steps in Protein Synthesis - Protein synthesis is already in progress at the ribosome. The growing polypeptide chain is attached to the tRNA that brought in the previous amino acid (in this illustration, cys.)

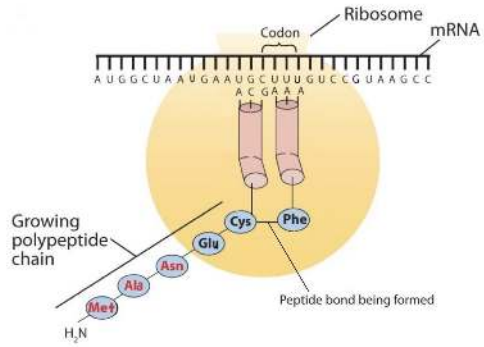


Figure 11.5.2b: The Elongation Steps in Protein Synthesis - An activated tRNA, which has the anticodon AAA, binds to the ribosome next to the previous bound tRNA and interacts with the mRNA molecule though basepairing of the codon and anticodon. The amino acid Phe is being incorporated into the polypeptide chain by the formation of a peptide linkage between the carboxyl group of Cys and the amino acid group of the Phe. This reaction is catalyzed by the enzyme peptidyl transferase, a component of the ribosome.

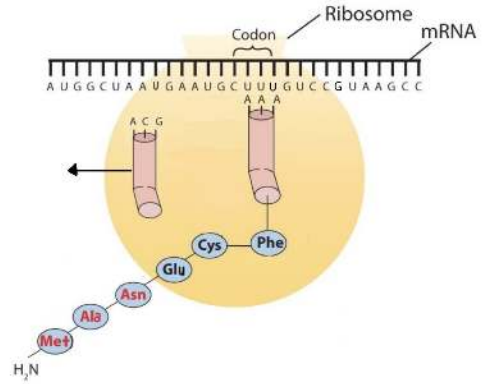


Figure **11.5.2*c***: The Elongation Steps in Protein Synthesis - The Cys-Phe linkage is now complete, and the growing polypeptide chain remains attached to the tRNA for Phe.

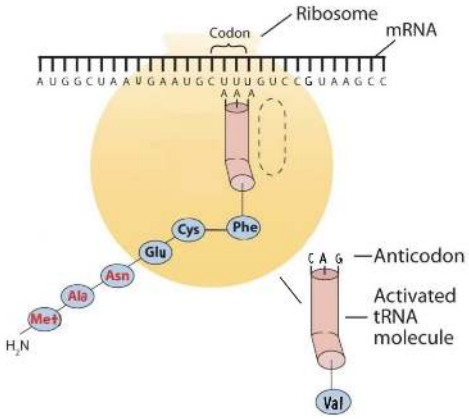


Figure **11.5.2*d***: The Elongation Steps in Protein Synthesis - The ribosome moves to the right along the mRNA strand. This shift brings the next codon, GUC, into its correct position on the surface of the ribosome. Note that an activated tRNA molecule, containing the next amino acid to be attached to the chain is moving to the ribosome. Steps (b)-(d) will be repeated until the ribosome reaches a stop codon.

Early experimenters were faced with the task of determining which of the 64 possible codons stood for each of the 20 amino acids. The cracking of the genetic code was the joint accomplishment of several well-known geneticists—notably Har Khorana, Marshall Nirenberg, Philip Leder, and Severo Ochoa—from 1961 to 1964. The genetic dictionary they compiled, summarized in Figure **11.5.3**, shows that 61 codons code for amino acids, and 3 codons serve as signals for the termination of polypeptide synthesis (much like the period at the end of a sentence). Notice that only methionine (AUG) and tryptophan (UGG) have single codons. All other amino acids have two or more codons.

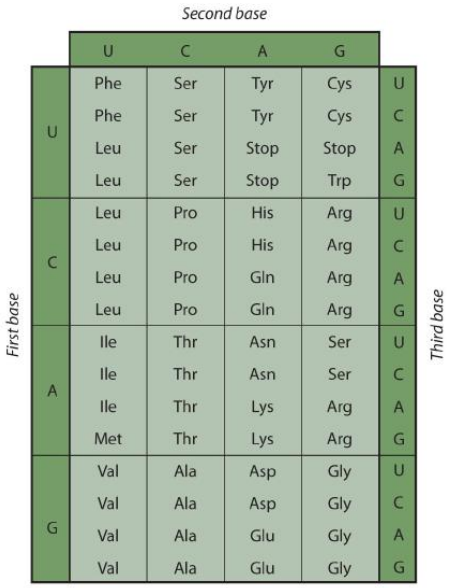


Figure **11.5.3**: The Genetic Code

✔ Example 11.5.1: Using the Genetic Code

A portion of an mRNA molecule has the sequence 5′-AUGCCACGAGUUGAC-3′. What amino acid sequence does this code for?

Solution

Use Figure **11.5.3** to determine what amino acid each set of three nucleotides (codon) codes for. Remember that the sequence is read starting from the 5′ end and that a protein is synthesized starting with the N-terminal amino acid. The sequence 5′-AUGCCACGAGUUGAC-3′ codes for met-pro-arg-val-asp.

? Exercise 11.5.4

A portion of an RNA molecule has the sequence 5′-AUGCUGAAUUGCGUAGGA-3′. What amino acid sequence does this code for?

Further experimentation threw much light on the nature of the genetic code, as follows:

1. The code is virtually universal; animal, plant, and bacterial cells use the same codons to specify each amino acid (with a few exceptions).
2. The code is “degenerate”; in all but two cases (methionine and tryptophan), more than one triplet codes for a given amino acid.
3. The first two bases of each codon are most significant; the third base often varies. This suggests that a change in the third base by a mutation may still permit the correct incorporation of a given amino acid into a protein. The third base is sometimes called the “wobble” base.
4. The code is continuous and nonoverlapping; there are *no* nucleotides between codons, and adjacent codons do not overlap.
5. The three termination codons are read by special proteins called release factors, which signal the end of the translation process.
6. The codon AUG codes for methionine and is also the initiation codon. Thus methionine is the first amino acid in each newly synthesized polypeptide. This first amino acid is usually removed enzymatically before the polypeptide chain is completed; the vast majority of polypeptides do not begin with methionine.

Summary

In translation, the information in mRNA directs the order of amino acids in protein synthesis. A set of three nucleotides (codon) codes for a specific amino acid.

11.5: Protein Synthesis and the Genetic Code is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.4: Protein Synthesis and the Genetic Code** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

11.6: Mutations and Genetic Diseases

Learning Objectives

- To describe the causes of genetic mutations and how they lead to genetic diseases.

We have seen that the sequence of nucleotides in a cell's deoxyribonucleic acid (DNA) is what ultimately determines the sequence of amino acids in proteins made by the cell and thus is critical for the proper functioning of the cell. On rare occasions, however, the nucleotide sequence in DNA may be modified either spontaneously (by errors during replication, occurring approximately once for every 10 billion nucleotides) or from exposure to heat, radiation, or certain chemicals. Any chemical or physical change that alters the nucleotide sequence in DNA is called a mutation. When a mutation occurs in an egg or sperm cell that then produces a living organism, it will be inherited by all the offspring of that organism.

Common types of mutations include **substitution** (a different nucleotide is substituted), **insertion** (the addition of a new nucleotide), and **deletion** (the loss of a nucleotide). These changes within DNA are called point mutations because only one nucleotide is substituted, added, or deleted (Figure **11.6.1**). Because an insertion or deletion results in a frame-shift that changes the reading of subsequent codons and, therefore, alters the entire amino acid sequence that follows the mutation, insertions and deletions are usually more harmful than a substitution in which only a single amino acid is altered.

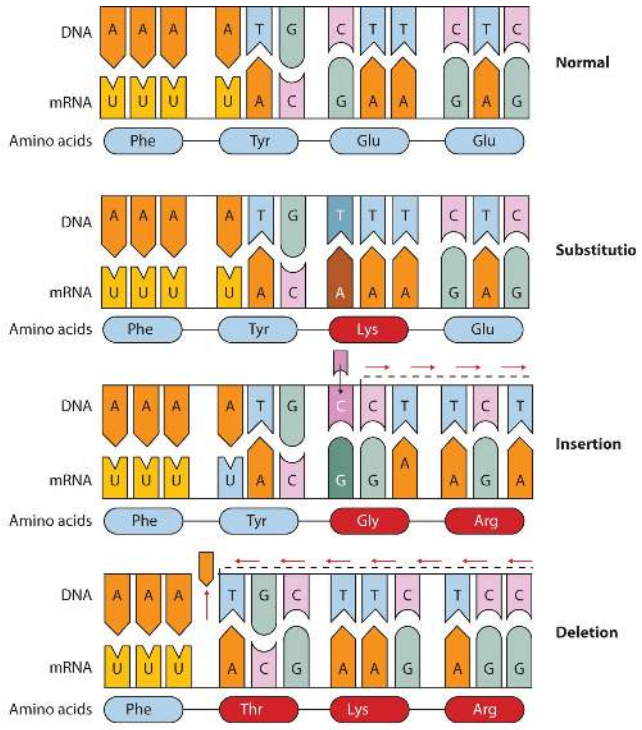


Figure **11.6.1**: Three Types of Point Mutations

The chemical or physical agents that cause mutations are called mutagens. Examples of physical mutagens are ultraviolet (UV) and gamma radiation. Radiation exerts its mutagenic effect either directly or by creating free radicals that in turn have mutagenic effects. Radiation and free radicals can lead to the formation of bonds between nitrogenous bases in DNA. For example, exposure to UV light can result in the formation of a covalent bond between two adjacent thymines on a DNA strand, producing a thymine dimer (Figure **11.6.2**). If not repaired, the dimer prevents the formation of the double helix at the point where it occurs. The genetic disease *xeroderma pigmentosum* is caused by a lack of the enzyme that cuts out the thymine dimers in damaged DNA. Individuals affected by this condition are abnormally sensitive to light and are more prone to skin cancer than normal individuals.

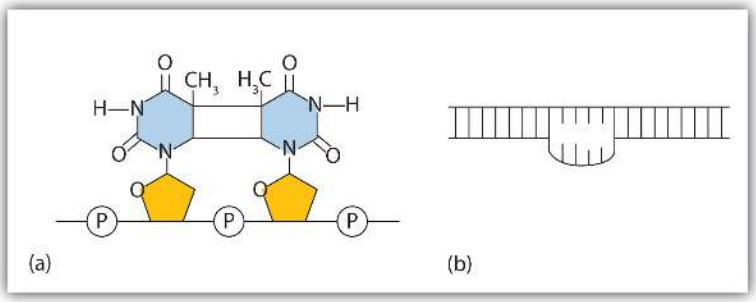


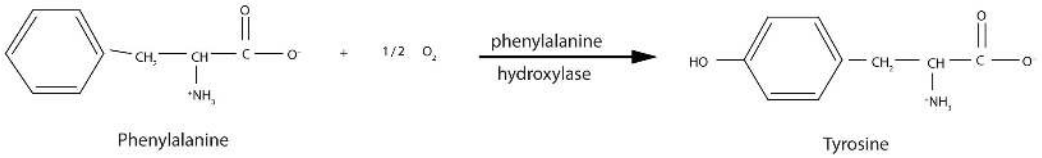
Figure **11.6.2**: An Example of Radiation Damage to DNA. (a) The thymine dimer is formed by the action of UV light. (b) When a defect in the double strand is produced by the thymine dimer, this defect temporarily stops DNA replication, but the dimer can be removed, and the region can be repaired by an enzyme repair system.

Sometimes gene mutations are beneficial, but most of them are detrimental. For example, if a point mutation occurs at a crucial position in a DNA sequence, the affected protein will lack biological activity, perhaps resulting in the death of a cell. In such cases the altered DNA sequence is lost and will not be copied into daughter cells. Nonlethal mutations in an egg or sperm cell may lead to metabolic abnormalities or hereditary diseases. Such diseases are called *inborn errors of metabolism* or genetic diseases. A partial listing of genetic diseases is presented in Figure **11.6.1**, and two specific diseases are discussed in the following sections. In most cases, the defective gene results in a failure to synthesize a particular enzyme.

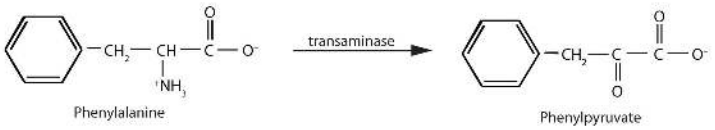
Figure **11.6.1**: Some Representative Genetic Diseases in Humans and the Protein or Enzyme Responsible

Disease	Responsible Protein or Enzyme
alkaptonuria	homogentisic acid oxidase
galactosemia	galactose 1-phosphate uridyl transferase, galactokinase, or UDP galactose epimerase
Gaucher disease	glucocerebrosidase
gout and Lesch-Nyhan syndrome	hypoxanthine-guanine phosphoribosyl transferase
hemophilia	antihemophilic factor (factor VIII) or Christmas factor (factor IX)
homocystinuria	cystathionine synthetase
maple syrup urine disease	branched chain α-keto acid dehydrogenase complex
McArdle syndrome	muscle phosphorylase
Niemann-Pick disease	sphingomyelinase
phenylketonuria (PKU)	phenylalanine hydroxylase
sickle cell anemia	hemoglobin
Tay-Sachs disease	hexosaminidase A
tyrosinemia	fumarylacetoacetate hydrolase or tyrosine aminotransferase
von Gierke disease	glucose 6-phosphatase
Wilson disease	Wilson disease protein

PKU results from the absence of the enzyme phenylalanine hydroxylase. Without this enzyme, a person cannot convert phenylalanine to tyrosine, which is the precursor of the neurotransmitters dopamine and norepinephrine as well as the skin pigment melanin.



When this reaction cannot occur, phenylalanine accumulates and is then converted to higher than normal quantities of phenylpyruvate. The disease acquired its name from the high levels of phenylpyruvate (a phenyl ketone) in urine. Excessive amounts of phenylpyruvate impair normal brain development, which causes severe mental retardation.



PKU may be diagnosed by assaying a sample of blood or urine for phenylalanine or one of its metabolites. Medical authorities recommend testing every newborn's blood for phenylalanine within 24 h to 3 weeks after birth. If the condition is detected, mental retardation can be prevented by immediately placing the infant on a diet containing little or no phenylalanine. Because phenylalanine is plentiful in naturally produced proteins, the low-phenylalanine diet depends on a synthetic protein substitute plus very small measured amounts of naturally produced foods. Before dietary treatment was introduced in the early 1960s, severe mental retardation was a common outcome for children with PKU. Prior to the 1960s, 85% of patients with PKU had an intelligence quotient (IQ) less than 40, and 37% had IQ scores below 10. Since the introduction of dietary treatments, however, over 95% of children with PKU have developed normal or near-normal intelligence. The incidence of PKU in newborns is about 1 in 12,000 in North America.

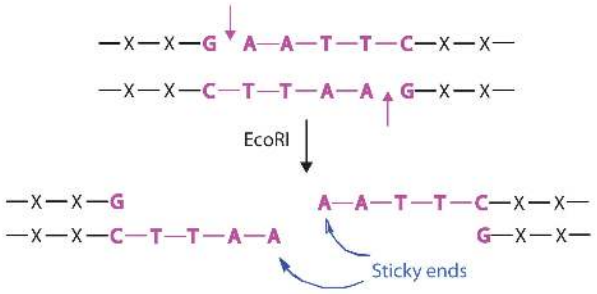
> *Every state in the United States has mandated that screening for PKU be provided to all newborns.*

Several genetic diseases are collectively categorized as *lipid-storage diseases*. Lipids are constantly being synthesized and broken down in the body, so if the enzymes that catalyze lipid degradation are missing, the lipids tend to accumulate and cause a variety of medical problems. When a genetic mutation occurs in the gene for the enzyme hexosaminidase A, for example, gangliosides cannot be degraded but accumulate in brain tissue, causing the ganglion cells of the brain to become greatly enlarged and nonfunctional. This genetic disease, known as *Tay-Sachs disease*, leads to a regression in development, dementia, paralysis, and blindness, with death usually occurring before the age of three. There is currently no treatment, but Tay-Sachs disease can be diagnosed in a fetus by assaying the amniotic fluid (amniocentesis) for hexosaminidase A. A blood test can identify Tay-Sachs carriers—people who inherit a defective gene from only one rather than both parents—because they produce only half the normal amount of hexosaminidase A, although they do not exhibit symptoms of the disease.

Looking Closer: Recombinant DNA Technology

More than 3,000 human diseases have been shown to have a genetic component, caused or in some way modulated by the person's genetic composition. Moreover, in the last decade or so, researchers have succeeded in identifying many of the genes and even mutations that are responsible for specific genetic diseases. Now scientists have found ways of identifying and isolating genes that have specific biological functions and placing those genes in another organism, such as a bacterium, which can be easily grown in culture. With these techniques, known as *recombinant DNA technology*, the ability to cure many serious genetic diseases appears to be within our grasp.

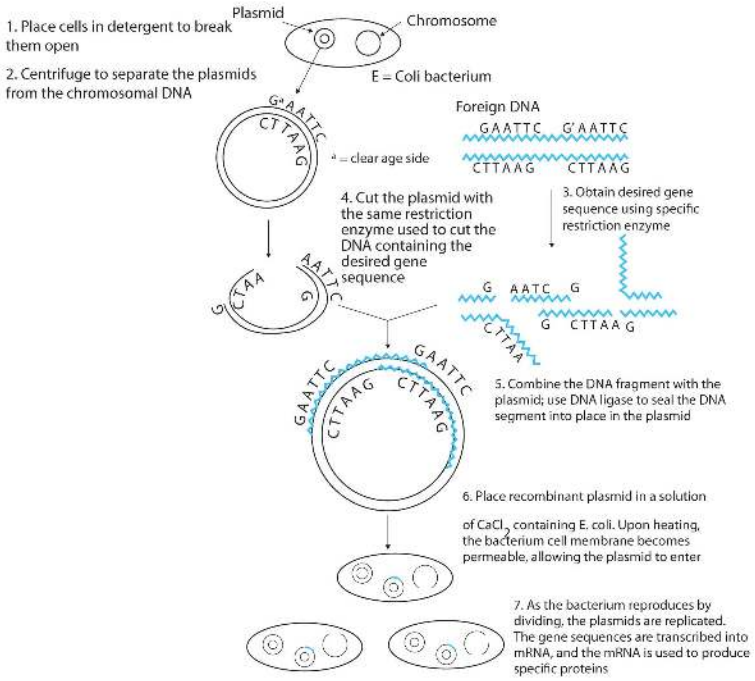
Isolating the specific gene or genes that cause a particular genetic disease is a monumental task. One reason for the difficulty is the enormous amount of a cell's DNA, only a minute portion of which contains the gene sequence. Thus, the first task is to obtain smaller pieces of DNA that can be more easily handled. Fortunately, researchers are able to use *restriction enzymes* (also known as restriction endonucleases), discovered in 1970, which are enzymes that cut DNA at specific, known nucleotide sequences, yielding DNA fragments of shorter length. For example, the restriction enzyme *EcoRI* recognizes the nucleotide sequence shown here and cuts both DNA strands as indicated:



Once a DNA strand has been fragmented, it must be cloned; that is, multiple identical copies of each DNA fragment are produced to make sure there are sufficient amounts of each to detect and manipulate in the laboratory. Cloning is accomplished by inserting the individual DNA fragments into phages (bacterial viruses) that can enter bacterial cells and be replicated. When a bacterial cell infected by the modified phage is placed in an appropriate culture medium, it forms a colony of cells, all containing copies of the original DNA fragment. This technique is used to produce many bacterial colonies, each containing a different DNA fragment. The result is a *DNA library*, a collection of bacterial colonies that together contain the entire genome of a particular organism.

The next task is to screen the DNA library to determine which bacterial colony (or colonies) has incorporated the DNA fragment containing the desired gene. A short piece of DNA, known as a *hybridization probe*, which has a nucleotide sequence complementary to a known sequence in the gene, is synthesized, and a radioactive phosphate group is added to it as a "tag." You might be wondering how researchers are able to prepare such a probe if the gene has not yet been isolated. One way is to use a segment of the desired gene isolated from another organism. An alternative method depends on knowing all or part of the amino acid sequence of the protein produced by the gene of interest: the amino acid sequence is used to produce an approximate genetic code for the gene, and this nucleotide sequence is then produced synthetically. (The amino acid sequence used is carefully chosen to include, if possible, many amino acids such as methionine and tryptophan, which have only a single codon each.)

After a probe identifies a colony containing the desired gene, the DNA fragment is clipped out, again using restriction enzymes, and spliced into another replicating entity, usually a plasmid. *Plasmids* are tiny mini-chromosomes found in many bacteria, such as *Escherichia coli* (*E. coli*). A recombined plasmid would then be inserted into the host organism (usually the bacterium *E. coli*), where it would go to work to produce the desired protein.



Proponents of recombinant DNA research are excited about its great potential benefits. An example is the production of human growth hormone, which is used to treat children who fail to grow properly. Formerly, human growth hormone was available only in tiny amounts obtained from cadavers. Now it is readily available through recombinant DNA technology. Another gene that has been cloned is the gene for epidermal growth factor, which stimulates the growth of skin cells and can be used to speed the healing of burns and other skin wounds. Recombinant techniques are also a powerful research tool, providing enormous aid to scientists as they map and sequence genes and determine the functions of different segments of an organism's DNA.

In addition to advancements in the ongoing treatment of genetic diseases, recombinant DNA technology may actually lead to cures. When appropriate genes are successfully inserted into *E. coli*, the bacteria can become miniature pharmaceutical factories, producing great quantities of insulin for people with diabetes, clotting factor for people with hemophilia, missing enzymes, hormones, vitamins, antibodies, vaccines, and so on. Recent accomplishments include the production in *E. coli* of recombinant DNA molecules containing synthetic genes for tissue plasminogen activator, a clot-dissolving enzyme that can rescue heart attack victims, as well as the production of vaccines against hepatitis B (humans) and hoof-and-mouth disease (cattle).

Scientists have used other bacteria besides *E. coli* in gene-splicing experiments and also yeast and fungi. Plant molecular biologists use a bacterial plasmid to introduce genes for several foreign proteins (including animal proteins) into plants. The bacterium is *Agrobacterium tumefaciens*, which can cause tumors in many plants, but which can be treated so that its tumor-causing ability is eliminated. One practical application of its plasmids would be to enhance a plant's nutritional value by transferring into it the gene necessary for the synthesis of an amino acid in which the plant is normally deficient (for example, transferring the gene for methionine synthesis into pinto beans, which normally do not synthesize high levels of methionine).

> Restriction enzymes have been isolated from a number of bacteria and are named after the bacterium of origin. *EcoRI* is a restriction enzyme obtained from the R strain of *E. coli*. The roman numeral I indicates that it was the first restriction enzyme obtained from this strain of bacteria.

Summary

- The nucleotide sequence in DNA may be modified either spontaneously or from exposure to heat, radiation, or certain chemicals and can lead to mutations.
- Mutagens are the chemical or physical agents that cause mutations.
- Genetic diseases are hereditary diseases that occur because of a mutation in a critical gene.

11.6: Mutations and Genetic Diseases is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.5: Mutations and Genetic Diseases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

11.7: Viruses

Learning Objectives

- To explain how viruses reproduce in cells.

Viruses are visible only under an electron microscope. They come in a variety of shapes, ranging from spherical to rod shaped. The fact that they contain either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA)—*but never both*—allows them to be divided into two major classes: DNA viruses and RNA viruses (Figure **11.7.1**).

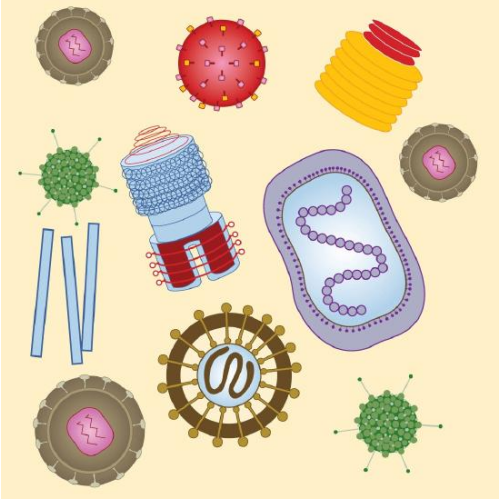


Figure **11.7.1**: Viruses. Viruses come in a variety of shapes that are determined by their protein coats.

Most *RNA viruses* use their nucleic acids in much the same way as the DNA viruses, penetrating a host cell and inducing it to replicate the viral RNA and synthesize viral proteins. The new RNA strands and viral proteins are then assembled into new viruses. Some RNA viruses, however, called retroviruses (Figure **11.7.2**), synthesize DNA in the host cell, in a process that is the reverse of the DNA-to-RNA transcription that normally occurs in cells. The synthesis of DNA from an RNA template is catalyzed by the enzyme reverse transcriptase.

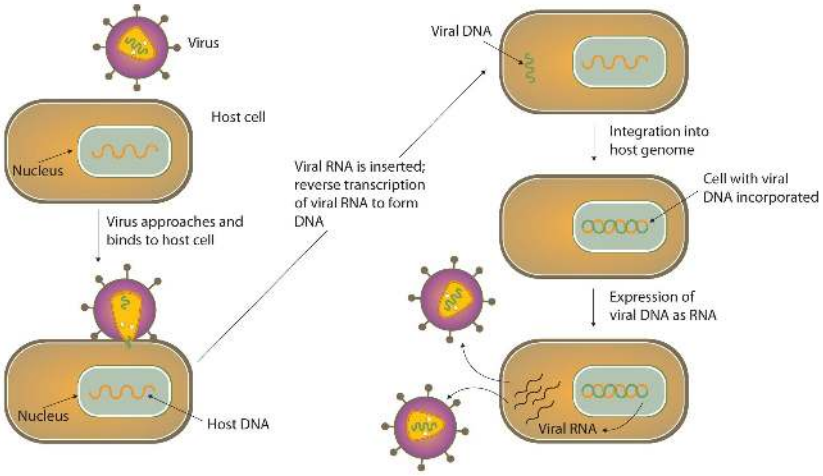
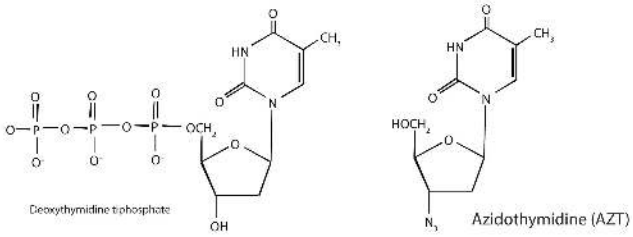
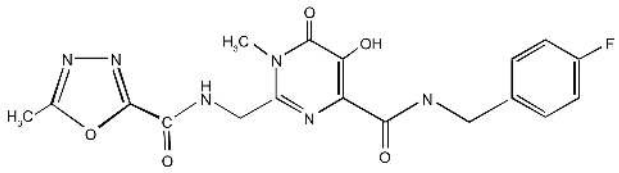


Figure **11.7.2**: Life Cycle of a Retrovirus

In 1987, **azidothymidine** (AZT, also known as zidovudine or the brand name Retrovir) became the first drug approved for the treatment of AIDS. It works by binding to reverse transcriptase in place of deoxythymidine triphosphate, after which, because AZT does not have a 3′OH group, further replication is blocked. In the past 10 years, several other drugs have been approved that also act by inhibiting the viral reverse transcriptase.



Raltegravir (Isentress) is a newer anti-AIDS drug that was approved by the FDA in October 2007. This drug inhibits the integrase enzyme that is needed to integrate the HIV DNA into cellular DNA, an essential step in the production of more HIV particles.



A major problem in treating HIV infections is that the virus can become resistant to any of these drugs. One way to combat the problem has been to administer a “cocktail” of drugs, typically a combination of two reverse transcriptase inhibitors along with a protease inhibitor. These treatments can significantly reduce the amount of HIV in an infected person.

Career Focus: Genetics Counselor

A genetics counselor works with individuals and families who have birth defects or genetic disorders or a family history of a disease, such as cancer, with a genetic link. A genetics counselor may work in a variety of health-care settings (such as a hospital) to obtain family medical and reproductive histories; explain how genetic conditions are inherited; explain the causes, diagnosis, and care of these conditions; interpret the results of genetic tests; and aid the individual or family in making decisions regarding genetic diseases or conditions. A certified genetics counselor must obtain a master’s degree from an accredited program. Applicants to these graduate programs usually have an undergraduate degree in biology, psychology, or genetics.



Photo courtesy of the United States National Institutes for Health, commons.wikimedia.org/wiki/File:Geneticcounseling.jpg.

Summary

Viruses are very small infectious agents that contain either DNA or RNA as their genetic material. The human immunodeficiency virus (HIV) causes acquired immunodeficiency syndrome (AIDS).

11.7: Viruses is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.6: Viruses** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

11.8: Nucleic Acids (Exercises)

19.1: Nucleotides

Concept Review Exercises

1. Identify the three molecules needed to form the nucleotides in each nucleic acid.
   a. DNA
   b. RNA
2. Classify each compound as a pentose sugar, a purine, or a pyrimidine.
   a. adenine
   b. guanine
   c. deoxyribose
   d. thymine
   e. ribose
   f. cytosine

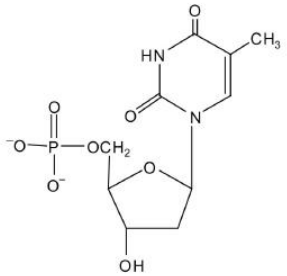
Answers

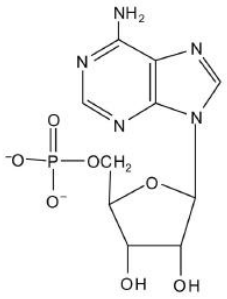
1. a. nitrogenous base (adenine, guanine, cytosine, and thymine), 2-deoxyribose, and H_3PO_4
   b. nitrogenous base (adenine, guanine, cytosine, and uracil), ribose, and H_3PO_4
2. a. purine
   b. purine
   c. pentose sugar
   d. pyrimidine
   e. pentose sugar
   f. pyrimidine

Exercises

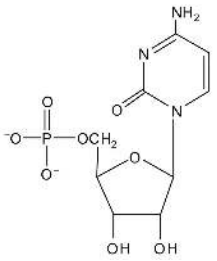
1. What is the sugar unit in each nucleic acid?
   a. RNA
   b. DNA
2. Identify the major nitrogenous bases in each nucleic acid.
   a. DNA
   b. RNA
3. For each structure, circle the sugar unit and identify the nucleotide as a ribonucleotide or a deoxyribonucleotide.

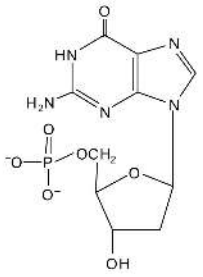
   a.



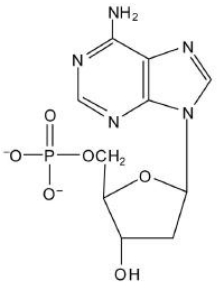
b. 

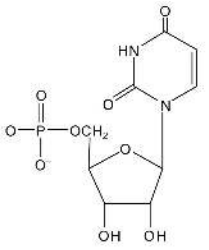
4. For each structure, circle the sugar unit and identify the nucleotide as a ribonucleotide or a deoxyribonucleotide.

a. 

b. 

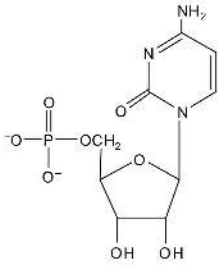
5. For each structure, circle the nitrogenous base and identify it as a purine or pyrimidine.

a. 

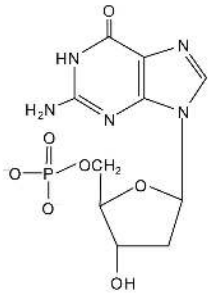
b. 

6. For each structure, circle the nitrogenous base and identify it as a purine or pyrimidine.

a.



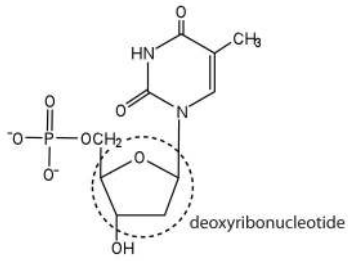
b.



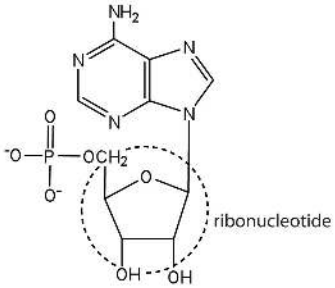
Answers

1. a. ribose
   b. deoxyribose

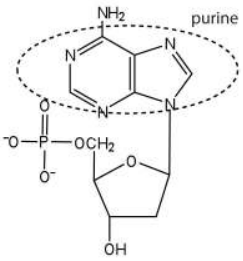
3. a.



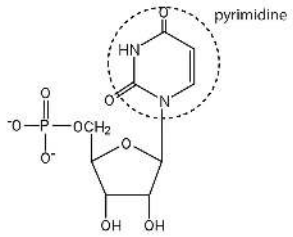
b.



a.



b.



19.2: Nucleic Acid Structure

Concept Review Exercises

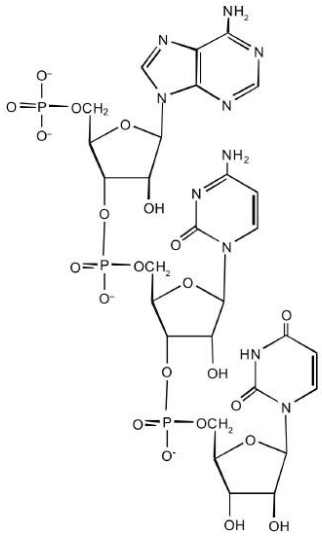
1. a. Name the two kinds of nucleic acids.
   b. Which type of nucleic acid stores genetic information in the cell?
2. What are complementary bases?
3. Why is it structurally important that a purine base always pair with a pyrimidine base in the DNA double helix?

Answers

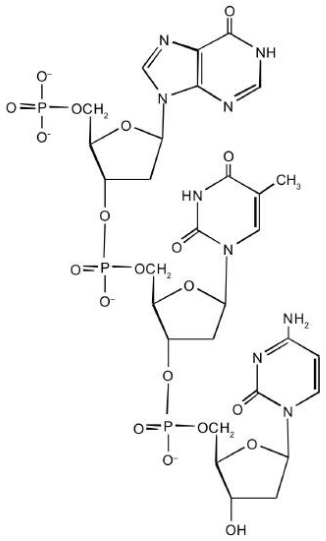
1. a. deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)
   b. DNA
2. the specific base pairings in the DNA double helix in which guanine is paired with cytosine and adenine is paired with thymine
3. The width of the DNA double helix is kept at a constant width, rather than narrowing (if two pyrimidines were across from each other) or widening (if two purines were across from each other).

Exercises

1. For this short RNA segment,
   a. identify the 5′ end and the 3′ end of the molecule.
   b. circle the atoms that comprise the backbone of the nucleic acid chain.
   c. write the nucleotide sequence of this RNA segment.



2. For this short DNA segment,
   a. identify the 5′ end and the 3′ end of the molecule.
   b. circle the atoms that comprise the backbone of the nucleic acid chain.
   c. write the nucleotide sequence of this DNA segment.

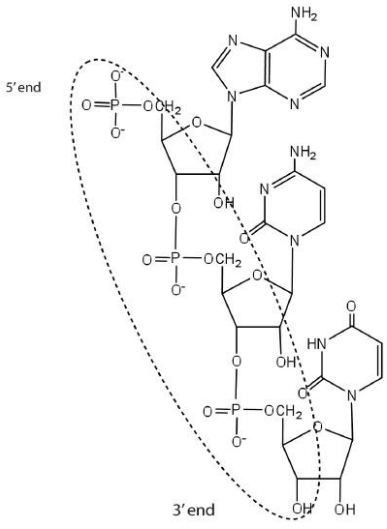


3. Which nitrogenous base in DNA pairs with each nitrogenous base?
    a. cytosine
    b. adenine
    c. guanine
    d. thymine
4. Which nitrogenous base in RNA pairs with each nitrogenous base?
    a. cytosine
    b. adenine
    c. guanine
    d. thymine
5. How many hydrogen bonds can form between the two strands in the short DNA segment shown below?

    5′ ATGCGACTA 3′ 3′ TACGCTGAT 5′
6. How many hydrogen bonds can form between the two strands in the short DNA segment shown below?

    5′ CGATGAGCC 3′ 3′ GCTACTCGG 5′

Answers

1. 

    c. ACU

3. a. guanine
   b. thymine
   c. cytosine
   d. adenine

5. 22 (2 between each AT base pair and 3 between each GC base pair)

19.3: Replication and Expression of Genetic Information

Concept Review Exercises

1. In DNA replication, a parent DNA molecule produces two daughter molecules. What is the fate of each strand of the parent DNA double helix?
2. What is the role of DNA in transcription? What is produced in transcription?
3. Which type of RNA contains the codon? Which type of RNA contains the anticodon?

Answers

1. Each strand of the parent DNA double helix remains associated with the newly synthesized DNA strand.
2. DNA serves as a template for the synthesis of an RNA strand (the product of transcription).
3. codon: mRNA; anticodon: tRNA

Exercises

1. Describe how replication and transcription are similar.
2. Describe how replication and transcription differ.
3. A portion of the coding strand for a given gene has the sequence 5′-ATGAGCGACTTTGCGGGATTA-3′.
   a. What is the sequence of complementary template strand?
   b. What is the sequence of the mRNA that would be produced during transcription from this segment of DNA?
4. A portion of the coding strand for a given gene has the sequence 5′-ATGGCAATCCTCAAACGCTGT-3′.
   a. What is the sequence of complementary template strand?
   b. What is the sequence of the mRNA that would be produced during transcription from this segment of DNA?

Answers

1. Both processes require a template from which a complementary strand is synthesized.

3.

a. 3′-TACTCGCTGAAACGCCCTAAT-5′
b. 5′-AUGAGCGACUUUGCGGGAUUA-3′

19.4: Protein Synthesis and the Genetic Code

Concept Review Exercises

1. What are the roles of mRNA and tRNA in protein synthesis?
2. What is the initiation codon?
3. What are the termination codons and how are they recognized?

Answers

1. mRNA provides the code that determines the order of amino acids in the protein; tRNA transports the amino acids to the ribosome to incorporate into the growing protein chain.
2. AUG
3. UAA, UAG, and UGA; they are recognized by special proteins called release factors, which signal the end of the translation process.

Exercises

1. Write the anticodon on tRNA that would pair with each mRNA codon.
   a. 5′-UUU-3′
   b. 5′-CAU-3′
   c. 5′-AGC-3′
   d. 5′-CCG-3′
2. Write the codon on mRNA that would pair with each tRNA anticodon.
   a. 5′-UUG-3′
   b. 5′-GAA-3′
   c. 5′-UCC-3′
   d. 5′-CAC-3′
3. The peptide hormone oxytocin contains 9 amino acid units. What is the minimum number of nucleotides needed to code for this peptide?
4. Myoglobin, a protein that stores oxygen in muscle cells, has been purified from a number of organisms. The protein from a sperm whale is composed of 153 amino acid units. What is the minimum number of nucleotides that must be present in the mRNA that codes for this protein?
5. Use Figure **11.8.3** to identify the amino acids carried by each tRNA molecule in Exercise 1.
6. Use Figure **11.8.3** to identify the amino acids carried by each tRNA molecule in Exercise 2.
7. Use Figure **11.8.3** to determine the amino acid sequence produced from this mRNA sequence: 5′-AUGAGCGACUUUGCGGGAUUA-3′.
8. Use Figure **11.8.3** to determine the amino acid sequence produced from this mRNA sequence: 5′-AUGGCAAUCCUCAAACGCUGU-3′

Answers

1. a. 3′-AAA-5′
   b. 3′-GUA-5′
   c. 3′-UCG-5′
   d. 3′-GGC-5′
3. 27 nucleotides (3 nucleotides/codon)
5. 1a: phenyalanine; 1b: histidine; 1c: serine; 1d: proline
7. met-ser-asp-phe-ala-gly-leu

19.5: Mutations and Genetic Diseases

Concept Review Exercises

1. a. What effect can UV radiation have on DNA?
   b. Is UV radiation an example of a physical mutagen or a chemical mutagen?
2. a. What causes PKU?
   b. How is PKU detected and treated?

Answers

1. a. It can lead to the formation of a covalent bond between two adjacent thymines on a DNA strand, producing a thymine dimer.
   b. physical mutagen
2. a. the absence of the enzyme phenylalanine hydroxylase
   b. PKU is diagnosed by assaying a sample of blood or urine for phenylalanine or one of its metabolites; treatment calls for an individual to be placed on a diet containing little or no phenylalanine.

Exercises

1. A portion of the coding strand of a gene was found to have the sequence 5′-ATGAGCGACTTTCGCCCATTA-3′. A mutation occurred in the gene, making the sequence 5′-ATGAGCGACCTTCGCCCATTA-3′.
   a. Identify the mutation as a substitution, an insertion, or a deletion.
   b. What effect would the mutation have on the amino acid sequence of the protein obtained from this mutated gene (use Figure 19.14)?
2. A portion of the coding strand of a gene was found to have the sequence 5′-ATGGCAATCCTCAAACGCTGT-3′. A mutation occurred in the gene, making the sequence 5′-ATGGCAATCCTCAACGCTGT-3′.
   a. Identify the mutation as a substitution, an insertion, or a deletion.
   b. What effect would the mutation have on the amino acid sequence of the protein obtained from this mutated gene (use Figure 19.14)?
3. a. What is a mutagen?
   b. Give two examples of mutagens.
4. For each genetic disease, indicate which enzyme is lacking or defective and the characteristic symptoms of the disease.
   a. PKU
   b. Tay-Sachs disease

Answers

1. a. substitution
   b. Phenylalanine (UUU) would be replaced with leucine (CUU).
3. a. a chemical or physical agent that can cause a mutation
   b. UV radiation and gamma radiation (answers will vary)

19.6: Viruses

Questions

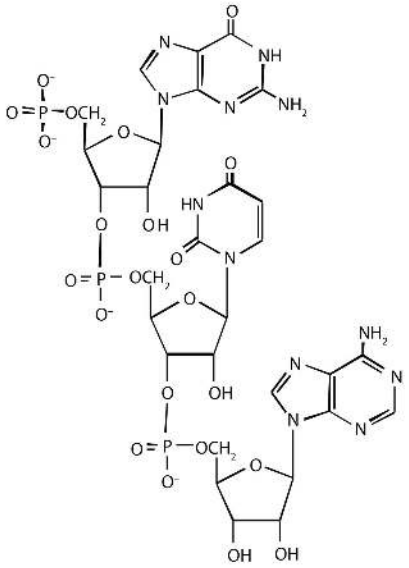
1. Describe the general structure of a virus.
2. How does a DNA virus differ from an RNA virus?
3. Why is HIV known as a retrovirus?
4. Describe how a DNA virus invades and destroys a cell.
5. a. Describe how an RNA virus invades and destroys a cell.
   b. How does this differ from a DNA virus?
6. What HIV enzyme does AZT inhibit?
7. What HIV enzyme does raltegravir inhibit?

Answers

1. A virus consists of a central core of nucleic acid enclosed in a protective shell of proteins. There may be lipid or carbohydrate molecules on the surface.
2. A DNA virus has DNA as its genetic material, while an RNA virus has RNA as its genetic material.
3. In a cell, a retrovirus synthesizes a DNA copy of its RNA genetic material.
4. The DNA virus enters a host cell and induces the cell to replicate the viral DNA and produce viral proteins. These proteins and DNA assemble into new viruses that are released by the host cell, which may die in the process.
5. -
6. reverse transcriptase
7. -

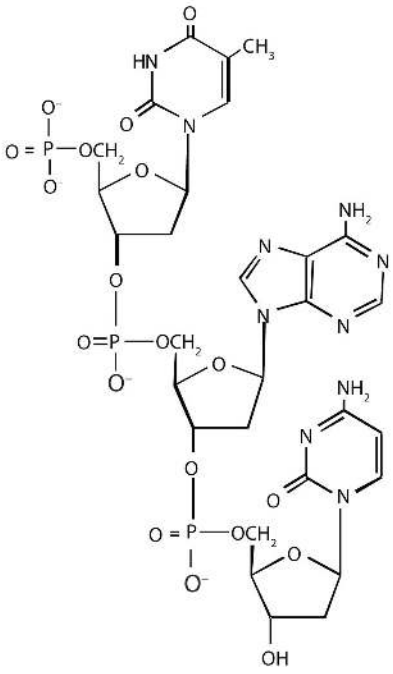
Additional Exercises

1. For this nucleic acid segment,



   a. classify this segment as RNA or DNA and justify your choice.
   b. determine the sequence of this segment, labeling the 5′ and 3′ ends.

2. For this nucleic acid segment,



   a. classify this segment as RNA or DNA and justify your choice.
   b. determine the sequence of this segment, labeling the 5′ and 3′ ends.

3. One of the key pieces of information that Watson and Crick used in determining the secondary structure of DNA came from experiments done by E. Chargaff, in which he studied the nucleotide composition of DNA from many different species. Chargaff noted that the molar quantity of A was always approximately equal to the molar quantity of T, and the molar quantity of C was always approximately equal to the molar quantity of G. How were Chargaff's results explained by the structural model of DNA proposed by Watson and Crick?

4. Suppose Chargaff (see Exercise 3) had used RNA instead of DNA. Would his results have been the same; that is, would the molar quantity of A approximately equal the molar quantity of T? Explain.
5. In the DNA segment

   5′-ATGAGGCATGAGACG-3′ (coding strand) 3′-TACTCCGTACTCTGC-5′ (template strand)

   a. What products would be formed from the segment's replication?
   b. Write the mRNA sequence that would be obtained from the segment's transcription.
   c. What is the amino acid sequence of the peptide produced from the mRNA in Exercise 5b?
6. In the DNA segment

   5′-ATGACGGTTTACTAAGCC-3′ (coding strand) 3′-TACTGCCAAATGATTCGG-5′ (template strand)

   a. What products would be formed from the segment's replication?
   b. Write the mRNA sequence that would be obtained from the segment's transcription.
   c. What is the amino acid sequence of the peptide produced from the mRNA in Exercise 6b?
7. A hypothetical protein has a molar mass of 23,300 Da. Assume that the average molar mass of an amino acid is 120.

   a. How many amino acids are present in this hypothetical protein?
   b. What is the minimum number of codons present in the mRNA that codes for this protein?
   c. What is the minimum number of nucleotides needed to code for this protein?
8. Bradykinin is a potent peptide hormone composed of nine amino acids that lowers blood pressure.

   a. The amino acid sequence for bradykinin is arg-pro-pro-gly-phe-ser-pro-phe-arg. Postulate a base sequence in the mRNA that would direct the synthesis of this hormone. Include an initiation codon and a termination codon.
   b. What is the nucleotide sequence of the DNA that codes for this mRNA?
9. A particular DNA coding segment is ACGTTAGCCCCAGCT.

   a. Write the sequence of nucleotides in the corresponding mRNA.
   b. Determine the amino acid sequence formed from the mRNA in Exercise 9a during translation.
   c. What amino acid sequence results from each of the following mutations?

      a. replacement of the underlined guanine by adenine
      b. insertion of thymine immediately after the underlined guanine
      c. deletion of the underlined guanine
10. A particular DNA coding segment is TACGACGTAACAAGC.

    a. Write the sequence of nucleotides in the corresponding mRNA.
    b. Determine the amino acid sequence formed from the mRNA in Exercise 10a during translation.
    c. What amino acid sequence results from each of the following mutations?

       a. replacement of the underlined guanine by adenine
       b. replacement of the underlined adenine by thymine
11. Two possible point mutations are the substitution of lysine for leucine *or* the substitution of serine for threonine. Which is likely to be more serious and why?
12. Two possible point mutations are the substitution of valine for leucine *or* the substitution of glutamic acid for histidine. Which is likely to be more serious and why?

Answers

1.

   a. RNA; the sugar is ribose, rather than deoxyribose
   b. 5′-GUA-3′

3. In the DNA structure, because guanine (G) is always paired with cytosine (C) and adenine (A) is always paired with thymine (T), you would expect to have equal amounts of each.

5.

a. Each strand would be replicated, resulting in two double-stranded segments.
b. 5′-AUGAGGCAUGAGACG-3′
c. met-arg-his-glu-thr

7.

a. 194
b. 194
c. 582

9.

a. 5′-ACGUUAGCCCCAGCU-3′
b. thr-leu-ala-pro-ala
c. a. thr-leu-thr-pro-ala
b. thr-leu-val-pro-ser
c. thr-leu-pro-gin

11. substitution of lysine for leucine because you are changing from an amino acid with a nonpolar side chain to one that has a positively charged side chain; both serine and threonine, on the other hand, have polar side chains containing the OH group.

11.8: Nucleic Acids (Exercises) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.1: Nucleotides** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **19.2: Nucleic Acid Structure** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **19.3: Replication and Expression of Genetic Information** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **19.4: Protein Synthesis and the Genetic Code** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **19.5: Mutations and Genetic Diseases** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.
- **19.6: Viruses** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

11.9: Nucleic Acids (Summary)

To ensure that you understand the material in this chapter, you should review the meanings of the bold terms in the following summary and ask yourself how they relate to the topics in the chapter.

A cell's hereditary information is encoded in **chromosomes** in the cell's nucleus. Each chromosome is composed of proteins and **deoxyribonucleic acid (DNA)**. The chromosomes contain smaller hereditary units called **genes,** which are relatively short segments of DNA. The hereditary information is expressed or used through the synthesis of **ribonucleic acid (RNA)**. Both **nucleic acids**—DNA and RNA—are polymers composed of monomers known as **nucleotides**, which in turn consist of phosphoric acid (H_3PO_4), a nitrogenous base, and a pentose sugar.

The two types of *nitrogenous bases* most important in nucleic acids are **purines**—adenine (A) and guanine (G)—and **pyrimidines**—cytosine (C), thymine (T), and uracil (U). DNA contains the nitrogenous bases adenine, cytosine, guanine, and thymine, while the bases in RNA are adenine, cytosine, guanine, and uracil. The sugar in the nucleotides of RNA is ribose; the one in DNA is 2-deoxyribose. The sequence of nucleotides in a nucleic acid defines the primary structure of the molecule.

RNA is a single-chain nucleic acid, whereas DNA possesses two nucleic-acid chains intertwined in a secondary structure called a **double helix**. The sugar-phosphate backbone forms the outside the double helix, with the purine and pyrimidine bases tucked inside. Hydrogen bonding between **complementary bases** holds the two strands of the double helix together; A always pairs with T and C always pairs with G.

Cell growth requires **replication**, or reproduction of the cell's DNA. The double helix unwinds, and hydrogen bonding between complementary bases breaks so that there are two single strands of DNA, and each strand is a *template* for the synthesis of a new strand. For protein synthesis, three types of RNA are needed: *messenger RNA* (mRNA), *ribosomal RNA* (rRNA), and *transfer RNA* (tRNA). All are made from a DNA template by a process called **transcription**. The double helix uncoils, and ribonucleotides base-pair to the deoxyribonucleotides on one DNA strand; however, RNA is produced using *uracil* rather than thymine. Once the RNA is formed, it dissociates from the template and leaves the nucleus, and the DNA double helix reforms.

Translation is the process in which proteins are synthesized from the information in mRNA. It occurs at structures called **ribosomes**, which are located outside the nucleus and are composed of rRNA and protein. The 64 possible three-nucleotide combinations of the 4 nucleotides of DNA constitute the **genetic code** that dictates the sequence in which amino acids are joined to make proteins. Each three-nucleotide sequence on mRNA is a **codon**. Each kind of tRNA molecule binds a specific amino acid and has a site containing a three-nucleotide sequence called an **anticodon**.

The general term for any change in the genetic code in an organism's DNA is **mutation**. A change in which a single base is substituted, inserted, or deleted is a **point mutation**. The chemical and/or physical agents that cause mutations are called **mutagens**. Diseases that occur due to mutations in critical DNA sequences are referred to as **genetic diseases**.

Viruses are infectious agents composed of a tightly packed central core of nucleic acids enclosed by a protective shell of proteins. Viruses contain either DNA or RNA as their genetic material but not both. Some RNA viruses, called **retroviruses**, synthesize DNA in the host cell from their RNA genome. The human immunodeficiency virus (HIV) causes acquired immunodeficiency syndrome (AIDS).

11.9: Nucleic Acids (Summary) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **19.S: Nucleic Acids (Summary)** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/introduction-to-chemistry-general-organic-and-biological.

CHAPTER OVERVIEW

12: Chemical and Biological Databases

12.1: Public Chemical Databases

12.2: Data Organization in PubChem as a Data Aggregator

12.3: Database Query Introduction

12.4: PubChem Web Interfaces for Text

12.5: Text Search in PubChem

12.6: Additional Data Retrieval Approaches in PubChem

12.7: Searching PubChem Using a Non-Textual Query

12: Chemical and Biological Databases is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.1: Public Chemical Databases

PubChem: chemical information repository at the U.S. NIH

PubChem (https://pubchem.ncbi.nlm.nih.gov)[1,2,3] is a public repository of information on small molecules and their biological activities, developed and maintained by the National Library of Medicine (NLM), an institute within the U.S. National Institutes of Health (NIH). Since its launch in 2004 as a component of the NIH's Molecular Libraries Roadmap Initiatives, it has been rapidly growing, and now serves as a key chemical information resource for researchers in many biomedical science areas, including cheminformatics, chemical biology, and medicinal chemistry. Detailed information on PubChem can be found in these three papers:

- PubChem Substance and Compound databases

  S. Kim *et al.*, Nucleic Acids Research **2016**, *44*, D1202-D1213

  (https://doi.org/10.1093/nar/gkv951)

- PubChem BioAssay: 2017 update

  Wang Y. *et al*. Nucleic Acids Research **2017**, *45*, D955-D963

  (https://doi.org/10.1093/nar/gkw1118)

- Getting the most out of PubChem for virtual screening

  S. Kim, Expert Opin. Drug Discov. **2016**, *11*, 843-855

  (http://dx.doi.org/10.1080/17460441.2016.1216967)

As of February 2017, PubChem contains more than 235 million depositor-provided substances, 94 million unique chemical structures, and one million biological assays, which cover about 10 thousand protein target sequences. For efficient use of this vast amount of data, PubChem provides various search and analysis tools. Some of these search tools will be used later in this course for demonstration purposes.

ChemSpider: a chemical database integrated with RSC's publishing process

ChemSpider (http://www.chemspider.com/)[4,5] is a free chemical structure database, containing information on 34 million structures collected from ~500 data sources. It also provides information on chemical reactions through ChemSpider SyntheticPages (CSSP)[6]. ChemSpider uses a crowdsourcing approach that allows registered users for manual comment and correction of ChemSpider records. Owned by the Royal Society of Chemistry (RSC), which publishes ~40 peer-reviewed chemistry journals, ChemSpider is integrated with the RSC publishing process, whereby new chemicals identified in newly published RSC articles also become available in ChemSpider.

ChEMBL: literature-extracted biological activity information

ChEMBL (https://www.ebi.ac.uk/chembl/)[7,8] is a large bioactivity database, developed and maintained by the European Bioinformatics Institute (EBI), which is part of the European Molecular Biology Laboratory (EMBL). The core activity data in ChEMBL are "manually" extracted from the full text of peer-reviewed scientific publications in select chemistry journals, such as *Journal of Medicinal Chemistry*, *Bioorganic Medicinal Chemistry Letters*, and *Journal of Natural products*. From each publication, details of the compounds tested, the assays performed and any target information for these assays are abstracted. ChEMBL also integrates screening results and bioactivity data from other public databases (such as PubChem BioAssay) and information on approved drugs from the U.S. FDA Orange Book[9] and the NLM's DailyMed[10].

ChEBI: a dictionary of small molecular entity

ChEBI (https://www.ebi.ac.uk/chebi/)[11,12] stands for "Chemical Entities of Biological Interest". It is a freely available database of "small" molecular entities, developed at the European Bioinformatics Institute (EBI). The molecular entities in ChEBI are either natural or synthetic products used to intervene the processes of living organisms. As a rule, however, ChEBI does not contain macromolecules directly encoded by genome (e.g., nucleic acids, proteins, and peptides derived from protein by cleavage). ChEBI

provides “standardized” descriptions of molecular entities that enable other databases to annotate their entries in a consistent fashion. ChEBI focuses on high-quality manual annotation, non-redundancy, and provision of a chemical ontology rather than full coverage of the vast chemical space. Note that both ChEMBL and ChEBI are developed and maintained by the EMBL-EBI. While ChEMBL focuses on “bioactivity” of a large number of bioactive molecules (currently ~2.0 millions), ChEBI is a “dictionary” that provides high-quality standardized descriptions for a relatively small number of molecules (currently ~50 thousands).

NIST Webbook: thermodynamic and spectroscopic data of chemicals

The U.S. National Institutes of Standards and Technology (NIST) compiles chemical and physical property data for chemical species and distributes them through the web site called the NIST Chemistry WebBook (http://webbook.nist.gov)[13,14]. These data include thermochemical data (e.g., enthalpy of formation, heat capacity, and vapor pressure), reaction thermochemistry data (e.g., enthalpy of reaction and free energy of reaction), spectroscopic data (e.g., IR and UV/Vis spectra), gas chromatographic data, ion energetics data, and so on.

DrugBank: comprehensive information on drug molecules

DrugBank[15,16,17] (http://www.drugbank.ca/) is a comprehensive online database containing biochemical and pharmacological information about ~8,000 drug molecules, including U.S. Food and Drug Administration (FDA)-approved small-molecule drugs and biotech drugs (e.g., protein/peptide drugs) as well as experimental drugs. DrugBank provides a wide range of drug information, including drug targets, mechanism of action, adverse drug reactions, food-drug and drug-drug interactions, experimental and theoretical ADMET properties (*i.e.*, Absorption, Distribution, Metabolism, Excretion, and Toxicity), and many others. Most of these data are curated from primary literature sources, by domain-specific experts and skilled biocurators.

HMDB: the Human Metabolome Database

The Human Metabolome Database (HMDB) (http://www.hmdb.ca)[18,19,20] is comprehensive information on human metabolites and human metabolism data. This database contains curated information derived from scientific literature, as well as experimentally determined metabolite concentrations in human tissue or biofluid (e.g., urine, blood, cerebrospinal fluid and so on). Reference Mass spectra (MS) and nuclear magnetic resonance (NMR) spectra for metabolites are also provided when available. In addition to data for “detected” metabolites (those with measured concentrations or experimental confirmation of their existence), the HMDB also provides information on “expected” metabolites (those for which biochemical pathways are known or human intake/exposure is frequent but the compound has yet to be detected in the body).

TOXNET: a collection of toxicological information

TOXNET (http://toxnet.nlm.nih.gov/)[21,22,23,24] maintained by the National Library of Medicine (NLM) at NIH, is a group of databases covering toxicology, hazardous chemicals, toxic releases, environmental and occupational health, risk assessment. Currently, 16 databases are integrated into the TOXNET system, and users can search all these databases either at once or individually. While all the 16 databases provide valuable information, three of them may be worth mentioning in the context of this course.

- ChemIDPlus[25,26] is a dictionary of over 400,000 chemical records (names, synonyms, and structures) and provides access to the structure and nomenclature files used for the identification of chemical substances in the TOXNET system and other NLM databases.
- The Hazardous Substances Data Bank (HSDB)[27,28] focuses on the toxicology of potentially hazardous chemicals, providing information on human exposure, industrial hygiene, emergency handling procedures, environmental fate, regulatory requirements, nanomaterials, and related areas. All HSDB data are referenced and derived from a core set of books, government documents, technical reports and selected primary journal literature. Importantly, HSDB is peer-reviewed by the Scientific Review Panel (SRP), a committee of experts in the major subject areas within the data bank's scope.
- The Comparative Toxicogenomics Database (CTD)[29,30] contains manually curated data describing interactions of chemicals with genes/proteins and diseases. This database provides insight into the molecular mechanisms underlying variable susceptibility for environmentally influenced diseases.

A brief overview of TOXNET and its databases can be found in the TOXNET Fact Sheet[22] and a recent paper by Fowler and Schnall[24].

Protein Data Bank (PDB): a key source for protein-bound ligand structures

The Protein Data Bank (PDB) is an archive of the experimentally determined 3-D structures of large biological molecules such as proteins and nucleic acids. These structures were determined primarily by using X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. While PDB is not a small molecule database, it contains the 3-D structures of many proteins with small-molecule ligands bound to them. PDB allows users to search for proteins that an input small molecule binds to. Considering that it is not possible to experimentally determine how small molecules (such as drug or toxic chemicals) actually bind to their target proteins in a living organism, PDB is the most widely used resource for experimentally determined protein-bound structures of small molecules. The PDB are maintained by the Worldwide PDB (wwPDB)[31], and freely accessible via the websites of its member organizations: PDBe (PDB in Europe)[32,33], PDBj (PDB Japan)[34,35], RCSB PDB (Research Collaboratory for Structural Bioinformatics PDB)[36,37].

References

1. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L. Y.; He, J. E.; He, S. Q.; Shoemaker, B. A.; Wang, J. Y.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2016**, *44*, D1202.
2. Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; Zhang, J. *Nucleic Acids Res.* **2017**, *45*, D955.
3. Kim, S. *Expert Opinion on Drug Discovery* **2016**, *11*, 843.
4. ChemSpider (http://www.chemspider.com) (Accessed on 2/17/2017).
5. Pence, H. E.; Williams, A. *J. Chem. Educ.* **2010**, *87*, 1123.
6. ChemSpider SyntheticPages (CSSP) (http://cssp.chemspider.com/) (Accessed on 2/17/2017).
7. ChEMBL (https://www.ebi.ac.uk/chembl/) (Accessed on 2/17/2017).
8. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magariños, M. P.; Overington, J. P.; Papadatos, G.; Smit, I.; Leach, A. R. *Nucleic Acids Res.* **2017**, *45*, D945.
9. Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations (http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm) (Accessed on 2/17/2017).
10. DailyMed (http://dailymed.nlm.nih.gov/) (Accessed on 2/17/2017).
11. ChEBI (https://www.ebi.ac.uk/chebi/) (Accessed on 2/17/2017).
12. Hastings, J.; de Matos, P.; Dekker, A.; Ennis, M.; Harsha, B.; Kale, N.; Muthukrishnan, V.; Owen, G.; Turner, S.; Williams, M.; Steinbeck, C. *Nucleic Acids Res.* **2013**, *41*, D456.
13. NIST Chemistry Webbook (http://webbook.nist.gov/chemistry/) (Accessed on 2/19/2017).
14. Linstrom, P. J.; Mallard, W. G. *J. Chem. Eng. Data* **2001**, *46*, 1059.
15. DrugBank (http://www.drugbank.ca/) (Accessed on 2/19/2017).
16. About DrugBank (http://www.drugbank.ca/about) (Accessed on 2/19/2017).
17. Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y. F.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B. S.; Zhou, Y.; Wishart, D. S. *Nucleic Acids Res.* **2014**, *42*, D1091.
18. The Human Metabolome Database (HMDB) (http://www.hmdb.ca/) (Accessed on 2/19/2017).
19. About the Human Metabolome Database (HMDB) (http://www.hmdb.ca/about) (Accessed on 2/19/2017).
20. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic Acids Res.* **2013**, *41*, D801.
21. ToxNet (http://toxnet.nlm.nih.gov/) (Accessed on 2/19/2017).
22. Factsheet - Toxicology Data Network (TOXNET) (http://www.nlm.nih.gov/pubs/factsheets/toxnetfs.html) (Accessed on 2/19/2017).
23. Wexler, P. *Toxicology* **2001**, *157*, 3.
24. Fowler, S.; Schnall, J. G. *Am. J. Nurs.* **2014**, *114*, 61.
25. ChemIDplus (http://chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp) (Accessed on 2/19/2017).
26. Fact Sheet - ChemIDplus (http://www.nlm.nih.gov/pubs/factsheets/chemidplusfs.html) (Accessed on 2/19/2017).
27. Hazardous Substances Data Bank (HSDB) (http://toxnet.nlm.nih.gov/newtoxnet/hsdb.htm) (Accessed on 2/19/2017).

28. Fact Sheet - Hazardous Substances Data Bank (HSDB) (http://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html) (Accessed on 2/19/2017).
29. Comparative Toxicogenomics Database (CTD) (http://toxnet.nlm.nih.gov/newtoxnet/ctd.htm) (Accessed on 2/19/2017).
30. Fact Sheet - Comparative Toxicogenomics Database (CTD) (http://www.nlm.nih.gov/pubs/factsheets/ctdfs.html) (Accessed on 2/19/2017).
31. Worldwide Protein Data Bank (wwPDB) (http://www.wwpdb.org/) (Accessed on 2/19/2017).
32. Protein Data Bank in Europe (PDBe) (http://www.ebi.ac.uk/pdbe/) (Accessed on 2/19/2017).
33. Gutmanas, A.; Alhroub, Y.; Battle, G. M.; Berrisford, J. M.; Bochet, E.; Conroy, M. J.; Dana, J. M.; Montecelo, M. A. F.; van Ginkel, G.; Gore, S. P.; Haslam, P.; Hendrickx, P. M. S.; Hirshberg, M.; Lagerstedt, I.; Mir, S.; Mukhopadhyay, A.; Oldfield, T. J.; Patwardhan, A.; Rinaldi, L.; Sahni, G.; Sanz-Garcia, E.; Sen, S.; Slowley, R. A.; Velankar, S.; Wainwright, M. E.; Kleywegt, G. J. *Nucleic Acids Res.* **2014**, *42*, D285.
34. Protein Data Bank Japan (PDBj) (http://pdbj.org/) (Accessed on 2/19/2017).
35. Kinjo, A. R.; Suzuki, H.; Yamashita, R.; Ikegawa, Y.; Kudou, T.; Igarashi, R.; Kengaku, Y.; Cho, H.; Standley, D. M.; Nakagawa, A.; Nakamura, H. *Nucleic Acids Res.* **2012**, *40*, D453.
36. RCSB Protein Data Bank (RCSB PDB) (http://www.rcsb.org/pdb/) (Accessed on 2/19/2017).
37. Rose, P. W.; Prlic, A.; Bi, C. X.; Bluhm, W. F.; Christie, C. H.; Dutta, S.; Green, R. K.; Goodsell, D. S.; Westbrook, J. D.; Woo, J.; Young, J.; Zardecki, C.; Berman, H. M.; Bourne, P. E.; Burley, S. K. *Nucleic Acids Res.***2015**, *43*, D345.

12.1: Public Chemical Databases is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.2: Data Organization in PubChem as a Data Aggregator

PubChem Aggregator Overview

PubChem (https://pubchem.ncbi.nlm.nih.gov) is a data aggregator, meaning that it collects data from other data sources. As of February 2017, PubChem's data are from more than 500 organizations, including government agencies, university labs, pharmaceutical companies, substance vendors, and other databases. An up-to-date list of PubChem's data sources is available at the PubChem Sources page (https://pubchem.ncbi.nlm.nih.gov/sources). To better understand the features of this page, read this article on PubChem Blog:

Data Sources Page

(http://go.usa.gov/xk7xU)

PubChem organizes its data into three inter-linked databases: Substance, Compound, and BioAssay

(See **Table 1**), which can be searched from either the PubChem home page (https://pubchem.ncbi.nlm.nih.gov) or the web page of one of the three PubChem databases.

Table 1. Three inter-linked databases in PubChem.

Database	URL	Identifier
Substance	https://www.ncbi.nlm.nih.gov/pcsubstance	SID
Compound	https://www.ncbi.nlm.nih.gov/pccompound	CID
BioAssay	https://www.ncbi.nlm.nih.gov/pcassay	AID

Individual data contributors deposit information on chemical substances to the Substance database (https://www.ncbi.nlm.nih.gov/pcsubstance). Different data contributors may provide information on the same molecule, hence the same chemical structure may appear multiple times in the Substance database. To provide a non-redundant view, chemical structures in the Substance database are normalized through a process called "standardization" and the unique chemical structures are identified and stored in the Compound database (https://www.ncbi.nlm.nih.gov/pccompound). The difference between the Substance and Compound databases is explained in more detail in this blog post.

Compounds and Substances

What is the difference between a substance and a compound in PubChem?
(http://1.usa.gov/1nl9ePL)

Descriptions of biological experiments on chemical substances are stored in the BioAssay database (https://www.ncbi.nlm.nih.gov/pcassay). The unique identifiers used to locate records in these three databases are called SID (Substance ID), CID (Compound ID), and AID (Assay ID) for the Substance, Compound, and BioAssay databases, respectively.

All information in the Substance database is submitted by individual data depositors. However, the Compound database does contain information that are not submitted by data depositors, but annotated by the PubChem team. [In the context of scientific databases, annotation refers to the process of adding extra information to a database entry (for example, a compound in the Compound database and an assay in the BioAssay database)]. The annotated information is always presented with its provenance information (that is, the source of the information). The list of all the annotation sources used in PubChem is available at the PubChem Sources page (https://pubchem.ncbi.nlm.nih.gov/sources). From this page, one may download all the annotations from a particular source.

12.2: Data Organization in PubChem as a Data Aggregator is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.3: Database Query Introduction

Basic Searches

After visiting a website that provides chemical information such as PubChem, there was most likely a text field available that allows the user to input an alphanumeric name, number or combination of both to retrieve a chemical. Simple text input searches may seem like the most basic way to look for something in a database and often they are, but many allow for inputing characters to refine how the search is performed. Using a simple google search as an example, you can change how the search is performed by simply putting your terms into parenthesis. This tells the search engine that everything in parenthesis must be found before being included in the results. There are some common practices for these alterations that are used among many different search engines, but this should not be interpreted to think that they will all work the same.

Before performing a series of searches, look for documentation provided by the host of the search engine to see what alterations can be made so that the results found have more relevancy to what the user seeks. The reason for this is simply due to the fact that chemical databases can contain a lot of data and this will save on trying to sort through thousands of results. Review section 1.3 of this course to see some other alterations that are often allowed in text searches. Take note of using boolean characters along with some of the different ways chemicals can be represented in chapter 2 of this course.

Custom Search Parameters

Advanced chemical searches may need to have several custom parameters used to narrow down a large list of chemicals into a smaller more relevant pool of results. The user interface for a custom search usually works similar to filling out an electronic form. After navigating to the advanced search on a website, the user will be presented with many options for defining a very specific search. Many of these advanced searches will contain checkboxes for selecting things like functional groups, experimental properties and other details for the compound. The user may also be able to specify ranges for things like density, number of atoms or other properties. Keywords can be valuable in searching the context of a chemical page to find a structure meeting certain categories.

12.3: Database Query Introduction is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.4: PubChem Web Interfaces for Text

PubChem Homepage

The PubChem homepage (https://pubchem.ncbi.nlm.nih.gov) provides a search interface that allow users to perform any term/keyword/identifier search against all three major databases of PubChem[1,2,3]: Compound, Substance, BioAssay. If a search returns multiple hits, they are presented on an Entrez DocSum page and will be explained in more detail later in this chapter. If the search returns a single record, the user will be directed to the web page that presents information on that record. This page is called the Compound Summary, Substance Record, or BioAssay Record page, depending on the record type (i.e., compound, substance, or assay). In addition, the PubChem homepage provides launch points to various PubChem services, tools, help documents, and more. In general, the PubChem homepage is a central location for all PubChem services.

Entrez Search and Retrieval System

NCBI's Entrez[4,5,6,7] is a database retrieval system that integrates PubChem's three major databases as well as other NCBI's major databases, including PubMed, Nucleotide and Protein Sequences, Protein Structures, Genome, Taxonomy, BioSystems, Gene Expression Omnibus (GEO) and many others. Entrez provides users with an integrated view of biomedical data and their relationships. This section focuses on search and retrieval of PubChem data using the Entrez system. A more detailed description on the Entrez system is given in the following documents:

- The Entrez Search and Retrieval System (http://www.ncbi.nlm.nih.gov/books/NBK184582/)
- Entrez Help (https://www.ncbi.nlm.nih.gov/books/NBK3836/)

Entry points to Entrez

One can search the PubChem databases through Entrez, by initiating a search from the NCBI home page (http://www.ncbi.nlm.nih.gov). By default, if a specific database is not selected in the search menu, Entrez searches all Entrez databases available, and lists the number of records in each database that are returned for this "global query". The following link directs you to the global query result page for the term "AIDS" against all databases integrated in the Entrez system.

https://www.ncbi.nlm.nih.gov/gquery/?term=AIDS

Simply by selecting one of the three PubChem databases from the global query results page (under the Chemical section), one can see the query results specific to that database.

Alternatively, one can start from the PubChem home page (http://pubchem.ncbi.nlm.nih.gov), where a search of one of the three PubChem databases may be initiated through the search box at the top. It is also possible to initiate an Entrez search against a PubChem database from the following pages:

- https://www.ncbi.nlm.nih.gov/pccompound/ (to search the Compound database)
- https://www.ncbi.nlm.nih.gov/pcsubstance/ (to search the Substance database)
- https://www.ncbi.nlm.nih.gov/pcassay/ (to search the BioAssay database)

Entrez DocSums

If an Entrez search for a query against any of the three PubChem databases returns a single record, the user will be directed to the Compound Summary, Substance Record, or BioAssay Record page for that record (depending on whether the record is a compound, substance, or assay). If it returns multiple records, Entrez will display a document summary report (also called "DocSum" page). The following link directs you to the DocSum page for a search for the term "lipitor" against the PubChem Compound database:

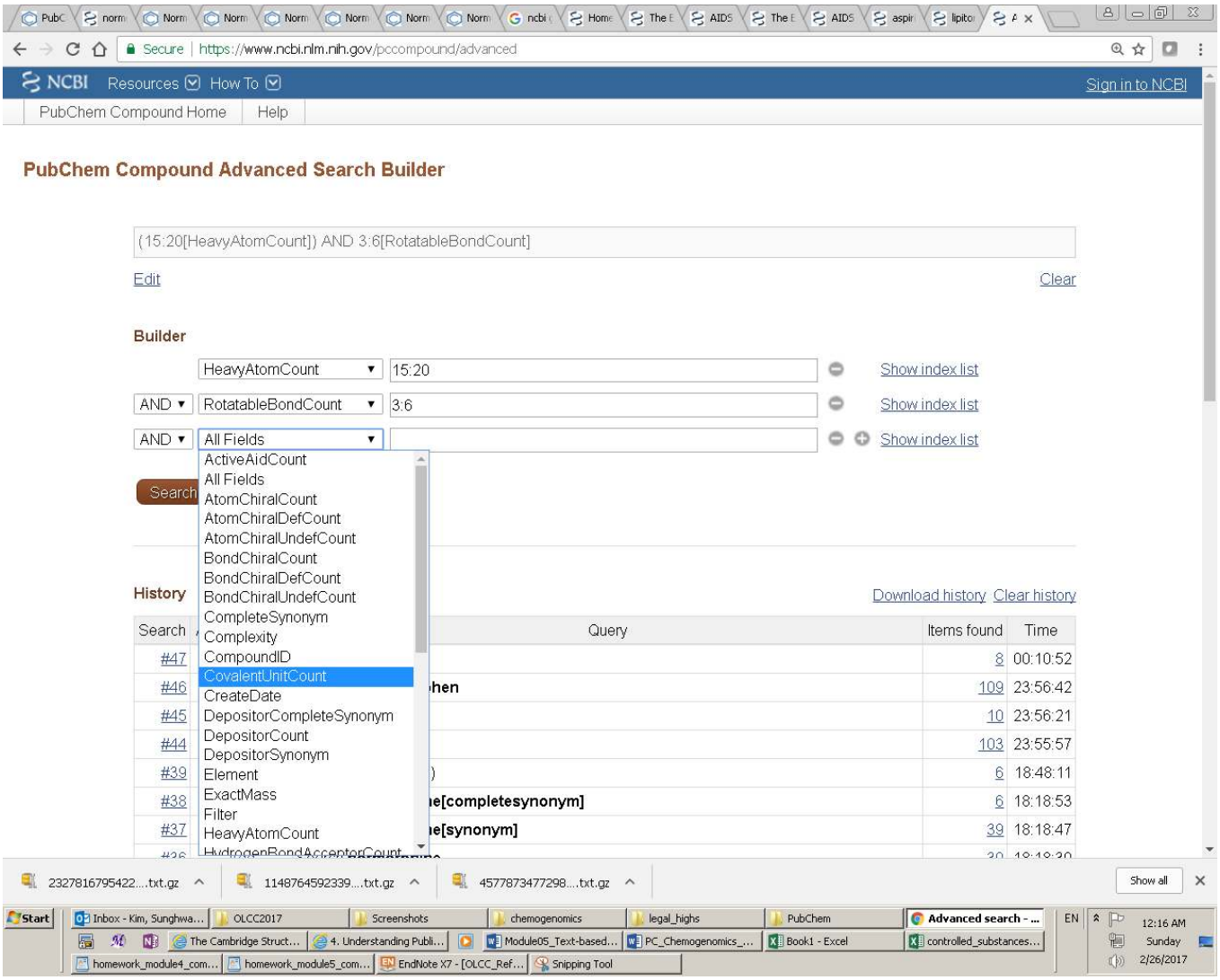
https://www.ncbi.nlm.nih.gov/pccompound?term=lipitor

In this example, the DocSum page displays a list of the compound records returned from the search. For each record, some data-specific information is provided with a link to the summary page for that record. The DocSum page contains controls to change the

display type, to sort the results by various means, or to export the page to a file or printer. Additional controls that operate on a query result list are available on the right column of the DocSum page. The DocSum page for the other two PubChem databases look similar to this example for the Compound database.

Entrez Indices

Entrez indices, tied to individual records in an Entrez database, include information on particular aspects (often referred to as fields) of the records. These indices may have text, numeric or date values, and some indices may have multiple values for each record. The available fields and their indexed terms in any Entrez database can be found from the drop-down menus on the Advanced Search Builder page (which can be accessed by clicking the “Advanced” link next to the “Go” button on the PubChem Home page).



When the user enters a query in the Entrez search interface, the Entrez indices are matched directly to that query. By default, in an Entrez search with a simple query, all indexed fields are matched against the query, usually resulting in the largest number of returned records including many unwanted results. One can narrow the search to a particular indexed field, by adding the index name in brackets after the term itself (e.g., “**lipitor[synonym]**”). For numeric indices, a search for a range of values can be done by using minimum and maximum values separated by a colon and followed by the bracketed index name (e.g., “**100:105[MolecularWeight]**”). Multiple indices may be searched simultaneously using Entrez’s Boolean operators (e.g., “**AND**”, “**OR**” and “**NOT**”).

A complete list of the Entrez indices available for the three PubChem databases can be retrieved in the XML format, using the eInfo functionality in E-Utilities (which will be covered in Module 7):

- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pccompound (for Compound)
- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcsubstance (for Substance)
- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcassay (for BioAssay).

Additional information on the PubChem Entrez indices is available in the “Indices and Filters in Entrez” section of the help documentation:

https://pubchem.ncbi.nlm.nih.gov/help.html#PubChem_Index

Entrez Links

Entrez links are cross links or associations between records in different Entrez databases, or within the same database. These links may be applied to an entire search result list (via the “find related data” section at the right column of a DocSum page) or to an individual record (via links at the bottom of each record presented on the DocSum page). The Entrez links provide a way to discover relevant information in other Entrez databases based on a user’s specific interests. Equivalently, one may think of this as a way to transform an identifier list from one database to another based on a particular criterion. Note that there are limits to how many records may be used as input in a link operation. To process a large amount of input records and/or to expect a large amount of output records associated with the input records, one should use the FLink tool (https://www.ncbi.nlm.nih.gov/Structure/flink/flink.cgi).

A complete list of the Entrez links available for the three PubChem databases can be retrieved in the XML format through these links

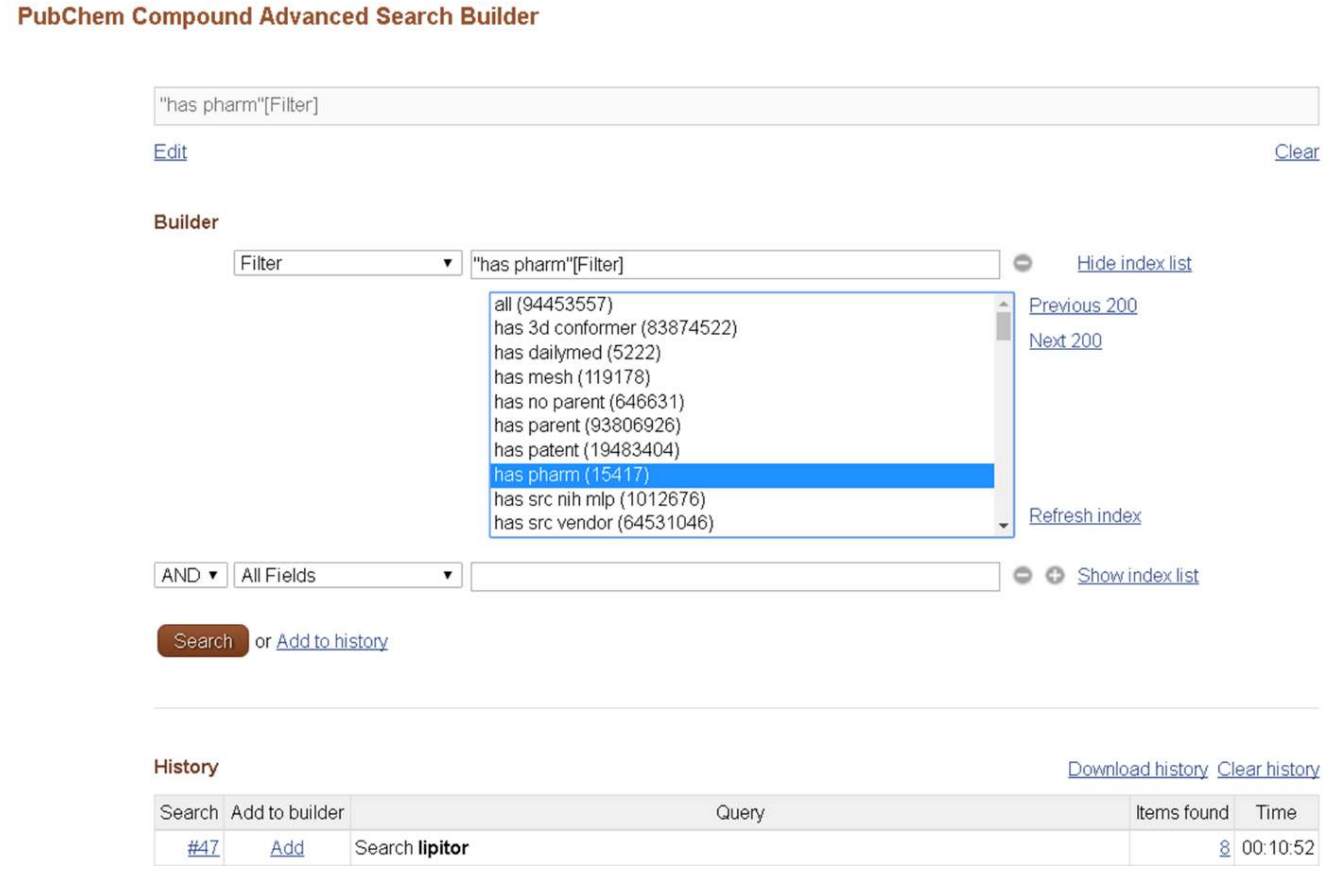
- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pccompound (for Compound)
- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcsubstance (for Substance)
- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/einfo.fcgi?db=pcassay (for BioAssay).

Entrez Filters

Entrez filters are essentially Boolean bits (true or false) for all records in a database that indicate whether or not a given record has a particular property. The Entrez filters may be used to subset other Entrez searches according to this property, by adding the filter to the query string.

Entrez filters are closely related to links in that the majority of Entrez filters in the PubChem databases are generated automatically based on whether PubChem records have Entrez links to a given database. However, some special filters, such as the "lipinski rule of 5" filter, or the “all” filter, are not link-based.

The Entrez filters available for each Entrez database may be found on the Advanced Search Builder page by selecting “Filter” from the “All Fields” dropdown and clicking “Show index list”.



More detailed description of the Entrez filters available for the three PubChem databases are given in the "Indices and Filters in Entrez" section of the help documentation:

https://pubchem.ncbi.nlm.nih.gov/help.html#PubChem_Index

Entrez History

Entrez has a history mechanism (Entrez history) that automatically keeps track of a user's searches, temporarily caches them (for eight hours), and allows one to combine search result sets with Boolean logic (i.e., "AND", "OR", and "NOT"). The Entrez history allows one to limit a search to a subset of records returned from a previous search. Use of Entrez history can help users avoid sending and receiving (potentially) very large lists of identifiers. In addition, through the Entrez history, one can use the search results as an input to various PubChem tools for further manipulation and analysis.

References

(1) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L. Y.; He, J. E.; He, S. Q.; Shoemaker, B. A.; Wang, J. Y.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2016**, *44*, D1202.

(2) Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; Zhang, J. *Nucleic Acids Res.* **2017**, *45*, D955.

(3) Kim, S. *Expert Opinion on Drug Discovery* **2016**, *11*, 843.

(4) Schuler, G. D.; Epstein, J. A.; Ohkawa, H.; Kans, J. A. *Methods Enzymol.* **1996**, *266*, 141.

(5) McEntyre, J. *Trends in genetics : TIG* **1998**, *14*, 39.

(6) The Entrez Search and Retrieval System (https://www.ncbi.nlm.nih.gov/books/NBK184582/) (Accessed on.

(7) Entrez Help (https://www.ncbi.nlm.nih.gov/books/NBK3836/) (Accessed on.

12.4: PubChem Web Interfaces for Text is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.5: Text Search in PubChem

Basics

Text search allows one to find chemical structures using one or more textual keywords, which may be chemical names (e.g., “aspirin”) or any word or phrase that describe molecules of interest (e.g., “cyclooxygenase inhibitors”). One can perform a text search from the PubChem homepage, by providing a text query in the search box. If the query is **a phrase or a name with non-alphanumeric characters**, **double quotes should be used around the query**. Various indices can be individually searched by suffixing a text query with an appropriate index enclosed by square brackets (for example, the query ***“N-(4-hydroxyphenyl)acetamide”[iupacname]***). Numeric range searches of appropriate index fields can be performed using a “:” delimiter (for example, the query ***100.5:200[molecularweight]*** for a molecular weight range search between 100.5 and 200.0 g/mol). One can see what search indices are available in PubChem from the drop-down menu on the “PubChem Compound Advanced Search Builder”, which can be accessed by clicking the “advanced” link (next to the “Go” button) on the PubChem homepage. Queries may be combined using the Boolean operators “**AND**”, “**OR**”, and “**NOT**”. These Boolean operators must be capitalized.

Depositor-supplied synonyms

Conceptually, data in a database are stored in the same way as we would record them in a table or excel spreadsheet. The rows in the table correspond to compounds, and the columns correspond to properties or descriptions for those compounds (e.g., melting and boiling points, chemical names, toxicity, bioactivity, target proteins, and so on). These columns are commonly called “data fields”. You may want to perform a search against all data fields or only a particular field. To search the (depositor-provided) chemical name field of the records in the PubChem Compound database, a chemical name query needs to be suffixed with either of the “[synonym]” or “[completesynonym]” index. The “[synonym]” index invokes search for molecules whose names contain the query chemical name as a part (that is, **partial matching**), and the “[completesynonym]” index invokes search for those whose names completely match the query (that is, **exact matching**). If no index is given after the query, PubChem will search all data fields. Compare the following searches for “aspirin” against the PubChem Compound database.

- **aspirin[completesynonym] (1 hit, as of Feb. 26, 2017)**
  https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin%5Bcompletesynonym%5D
- **aspirin[synonym] (98 hits)**
  https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin%5Bsynonym%5D
- **aspirin (103 hits)**
  https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin

Note that the URLs for these searches contain the query strings (following the string “?term=”), and that the square brackets enclosing the Entrez indices “completesynonym” and “synonym” are replaced with the strings “%5B” and “%5D”. Because the first query resulted in only one hit, the user is directed to the Compound Summary page for the hit compound (CID 2244). On the other hand, because the other two queries result in multiple hits, the results are presented on the DocSum pages.

When either “[completesynonym]” or “[synonym]” is used, it is the **“depositor-provided synonyms”** fields of the compound records in PubChem that is searched for the query string. The depositor-provided synonyms field for a compound contains a filtered list of chemical names (synonyms) provided by individual data providers for the substances associated with that compound. These synonyms are presented in the “Depositor-provided synonyms” section on a Compound Summary page. To see the variety of synonyms for a compound, check the following link [to the Depositor-provided synonyms” section of the Compound Summary page for CID 2244 (aspirin)]:

https://pubchem.ncbi.nlm.nih.gov/compound/2244#section=Depositor-Supplied-Synonyms

For CID 2244, there are more than 700 depositor-supplied synonyms. These synonyms include not only those commonly used in chemistry class (e.g., common names, IUPAC names, CAS registry numbers) but also those used in many other places (e.g., database identifiers, chemical vendor catalogues, the name of products that contains the chemical, code numbers internally used in a company, and so on).

As mentioned above, the search for aspirin with the “[completesynonym]” index specified returns only one compound (CID 2244). It means that one of many names of this compound exactly matches the query string “aspirin”. On the other hand, the search for aspirin with the “[synonym]” index returns additional 97 compounds. It means that at least one of the names of each these compound partially match the query string (that is, the compound contains the string “aspirin” in one of its names). Interestingly, the results from the last two queries include acetaminophen (CID 1983), which is the active ingredient of Tylenol. Check the following link to the depositor-provided synonyms section of CID 1983 to see what synonyms of Tylenol contains the string “aspirin”:

https://pubchem.ncbi.nlm.nih.gov/compound/1983#section=Depositor-Supplied-Synonyms

Some of the synonyms of Tylenol contains the phrase “aspirin-free” or “non-aspirin”. Note that Tylenol was returned from a search for “aspirin” (through partial matching using the [synonym] index).

MeSH Synonyms

The National Library of Medicine (NLM)’s Medical Subject Headings (MeSH)[1,2]

is a controlled vocabulary thesaurus of medical terms arranged in a hierarchical structure. It is used for indexing scientific articles from biomedical journals for PubMed and cataloging medical books, documents, and audiovisual materials, in order to facilitate retrieval of medical information at various levels of specificity.

Many of MeSH terms are chemical names (e.g., for drugs, nutrients, metabolites, toxic chemicals, and so on). PubChem performs an automated annotations of PubChem records with MeSH terms (by means of chemical name matching), creating associations between PubChem records and PubMed articles that share the same MeSH annotation. The MeSH term that match a (depositor-provided) synonym of a compound in PubChem is presented with its entry terms under the “MeSH Synonyms” section of the Compound Summary page of that compound.

Go to the Compound Summary page for CID 171511 via the following link to check the MeSH synonyms and Depositor-supplied Synonyms sections.

References

(1) Medical Subject Headings (MeSH) (https://www.nlm.nih.gov/mesh/)

(2) Medical Subject Headings (MeSH®) Fact Sheet (https://www.nlm.nih.gov/pubs/factsheets/mesh.html)

12.5: Text Search in PubChem is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.6: Additional Data Retrieval Approaches in PubChem

Classification Browser

The PubChem Classification Browser, which allows the user to navigate or search PubChem records associated to a hierarchical classification system of interest, is available via URL:

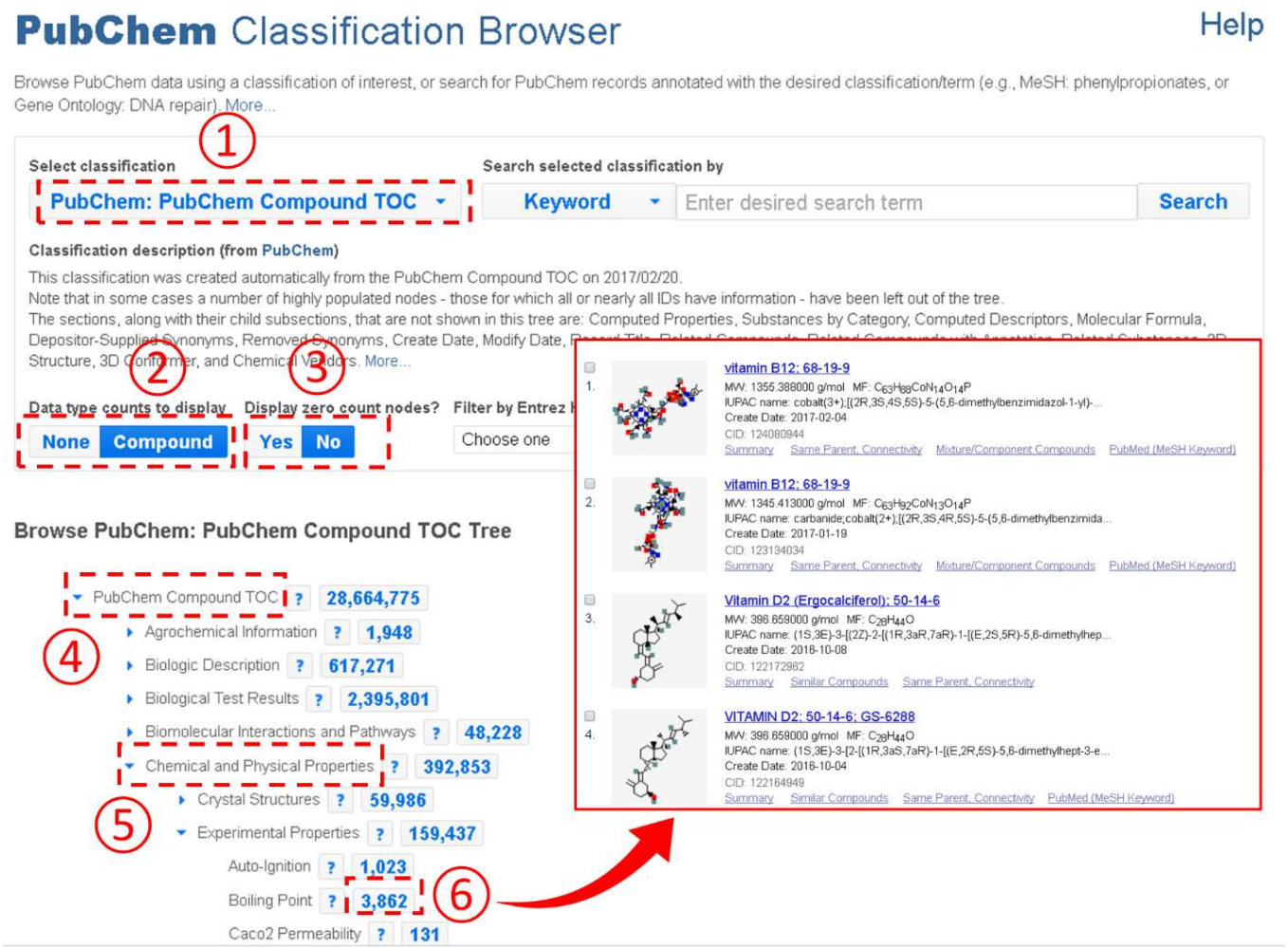
http://pubchem.ncbi.nlm.nih.gov/classification

The Classification Browser can also be accessed from the PubChem home page (through the "Services" menu at the top or the "Classification" icon on the right column of the page). Currently, the Classification Browser can retrieve records annotated with terms in the following classification systems:

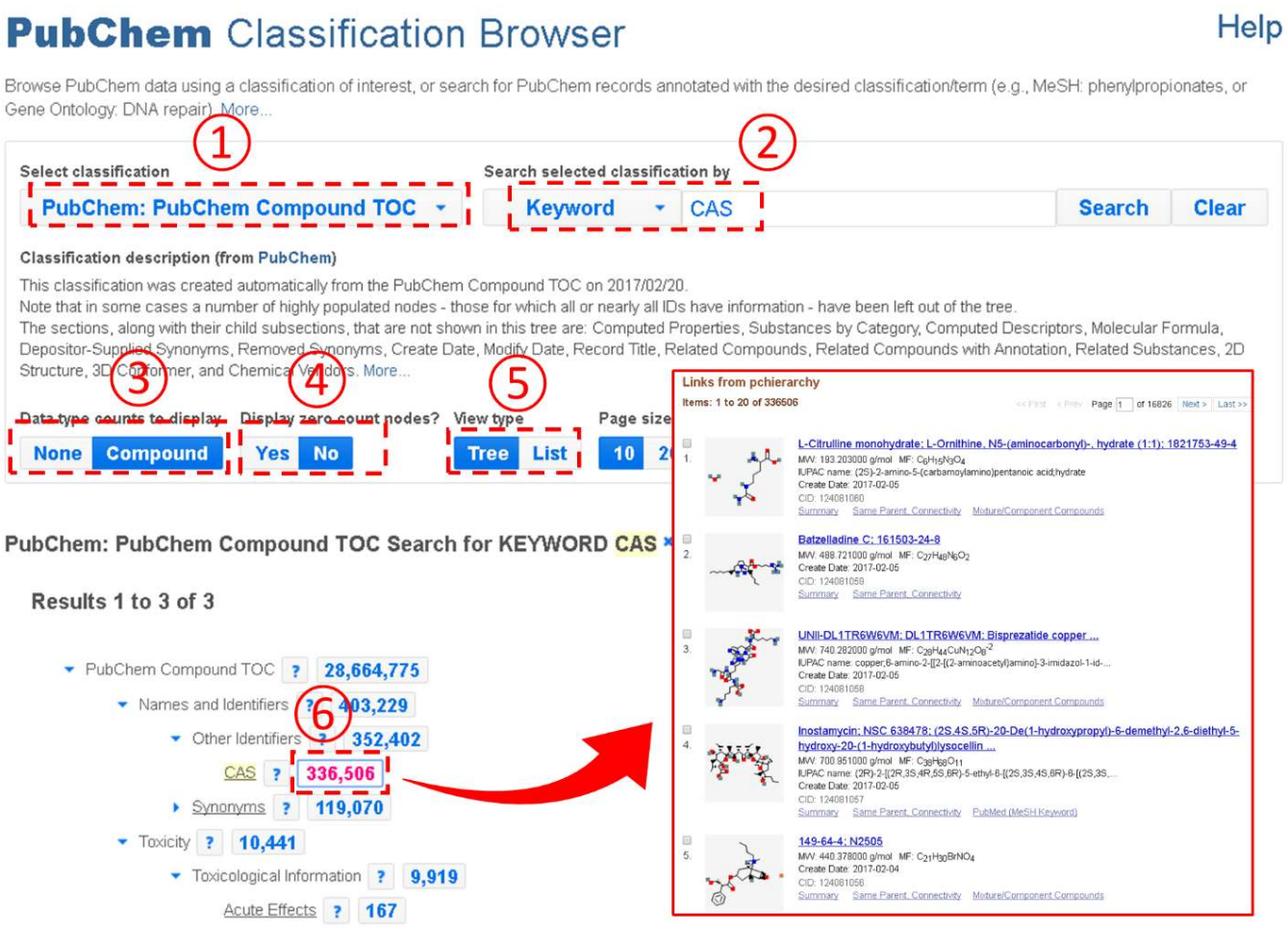
- MeSH (Medical Subject Headings)
- ChEBI
- FDA Pharmacological Classification
- KEGG
- LIPID MAPS
- World Health Organization (WHO)'s Anatomical Therapeutic Chemical (ATC)
- World Intellectual Property Organization (WIPO)'s IPC (International Patent Classification)

The Classification Browser provides a powerful way to quickly and visually find a desired subset of PubChem records. The output can be displayed in Tree view or List view.

An important feature of the Classification Browser is that **the Table of Contents presented on the Compound Summary is integrated into the Classification Browser, allowing users to quickly retrieve compounds with a particular type of information available**. For example, the figure below shows how to retrieve all compounds with the boiling point information from PubChem.



In the example above, users need to expand the Table of Contents tree to locate the boiling point node. However, this task may not be easy to some users who do not have prior knowledge about where the node that they want to find is located in the Table of Contents tree system. To assist these users, the Classification Brower supports a keyword search against the node names and descriptions of the classification trees. For example, the example below shows how to retrieve compounds with the CAS Registry number. Note that this task involves a search for the term “CAS”.



The Classification Browser also supports the **PubChem BioAssay Classification Tree**, providing an additional approach to browse, search, and access the BioAssay data. More detailed information on the Classification Browser is available at the URL:

http://pubchem.ncbi.nlm.nih.gov//classification/docs/classification_help.html

Identifier Exchange Service

The Identifier Exchange Service can be found at the following URL: http://pubchem.ncbi.nlm.nih.gov/idexchange

This service allows the user to convert one type of identifiers for a given set of chemical structures into a different type of identifiers for identical or similar chemical structures. Currently, it supports seven types of identifiers: CID, SID, InChI, InChIKey, SMILES, synonyms, Registry ID. When Registry ID is selected as an input or output identifier type, the DSN (Data Source Name) should also be provided.

The input identifier list may be provided using a string, a text file, or Entrez history. When a service request is submitted, it will be queued on PubChem servers. Once the actual task starts to run, the input identifiers will be converted into CIDs (called input CIDs) during the computation, and the CIDs (called output CIDs) that satisfy the condition specified by one of the following operation types will be retrieved:

- **Same CID**: Same CIDs as input CIDs.
- **Same, Stereochemistry**: CIDs that have same stereo centers as input CIDs.
- **Same, Isotopes**: CIDs that have the same isotopes as input CIDs.
- **Same, Connectivity**: CIDs that have the same connectivity as input CIDs.
- **Same parent**: CIDs that have the same parents as input CIDs.
- **Same parent, Stereochemistry**: CIDs that have the same stereo centers and parents as input CIDs.

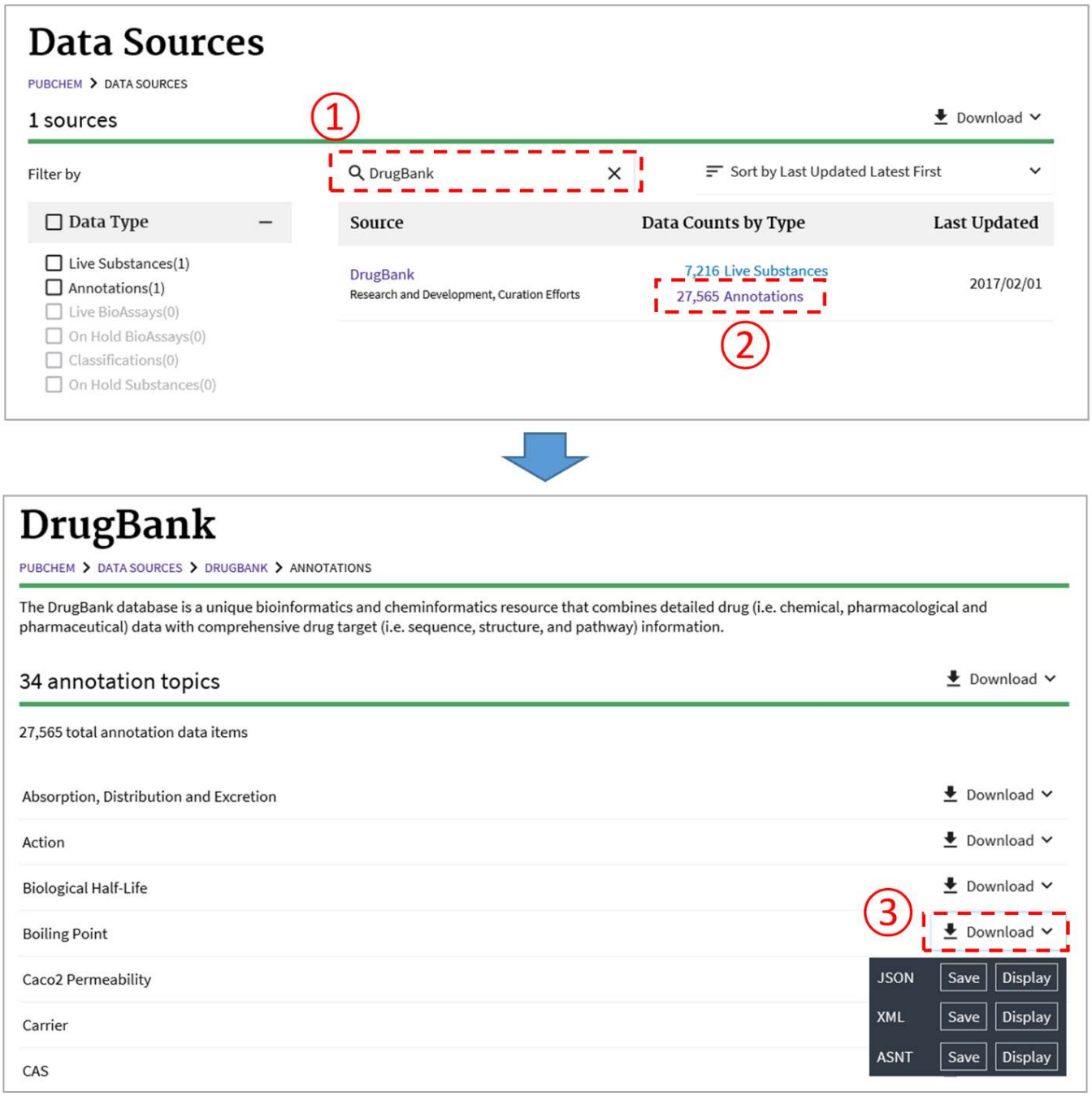
- **Same parent, Isotopes**: CIDs that have the same isotopes and parents as input CIDs.
- **Same parent, Connectivity**: CIDs that have the same connectivity and parents as input CIDs.
- **Similar 2D compounds**: CIDs similar to the input CIDs in PubChem's 2-D similarity.
- **Similar 3D conformers**: CIDs similar to the input CIDs in PubChem's 3-D similarity.

These output CIDs are then converted into the identifier type specified by the user and written into a file or sent to **Entrez history**. In practice, the identifier exchange service may be used as a quick approach to search the PubChem Compound database using multiple queries, although this type of task may be performed programmatically (for example, using PUG-REST,[1] which will be discussed in Module 7). A more detailed information is available at the URL:

http://pubchem.ncbi.nlm.nih.gov//idexchange/idexchange-help.html

The PubChem Data Sources page

As discussed in Section 3.4, the PubChem Data Sources page (https://pubchem.ncbi.nlm.nih.gov/sources/) helps users determine who provided what information. This page can be used to retrieve the data provided by a data depositor or to download the annotations collected from a data source. For example, the following figure illustrates how to download the boiling point data collected from DrugBank.[2]



To obtain a particular kind of annotated information (e.g., boiling points) through the PubChem Data Sources page, one may need to know "in advance" which depositors provide that information. This can be done through a PUG-REST request[1] (to be discussed in detail in Module 7). For example, the following PUG-REST request returns all data sources that provide the boiling point information for chemicals.

https://pubchem.ncbi.nlm.nih.gov/rest/pug/annotations/heading/boiling%20point/TXT

On the other hand, one may want to know what kind of information is provided by a given data source. This can also be done using a PUG-REST request:

https://pubchem.ncbi.nlm.nih.gov/rest/pug/annotations/sourcename/DrugBank/TXT

This example retrieves all types of annotations collected from DrugBank.

References

(1) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Bryant, S. H. *Nucleic Acids Res.* **2015**, *43*, W605.

(2) Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y. F.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B. S.; Zhou, Y.; Wishart, D. S. *Nucleic Acids Res.* **2014**, *42*, D1091.

12.6: Additional Data Retrieval Approaches in PubChem is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.7: Searching PubChem Using a Non-Textual Query

This section describes various searches that can be performed in PubChem.[1,2,3] Currently PubChem has three different search interfaces:

1. PubChem homepage (http://pubchem.ncbi.nlm.nih.gov)
2. PubChem Chemical Structure Search (https://pubchem.ncbi.nlm.nih.gov/search/search.cgi)
3. PubChem Search (https://pubchem.ncbi.nlm.nih.gov/search/).

As explained in Section 4.1, the PubChem homepage provides a search interface for all three primary databases (e.g., Substance, Compound, and BioAssay). However, the search box on the PubChem homepage can accepts textual keywords only, and it is difficult to input non-textual queries (such as chemical structures). The PubChem Chemical Structure Search allows users to perform various searches using both textual and non-textual queries. This search interface is integrated with PubChem Sketcher,[4] which enables users to provide the 2-D structure of a molecule as a query for chemical structure search. While the PubChem Chemical Structure Search is limited to search for chemical structures, the PubChem Search allows users to search for bioassays, bioactivities, patents, and targets as well as chemical structures, but it is still in beta testing. In this module, we use the Chemical Structure Search for chemical structure search.

Molecular formula search

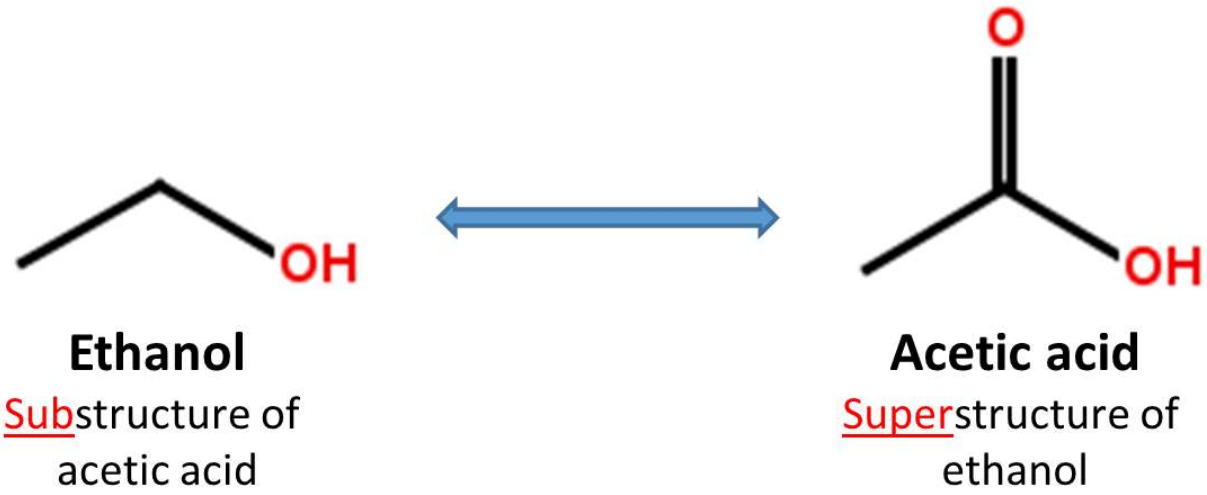
Molecular formula search allows one to find molecules that contain a certain number and type of elements. Typically, molecular formula search returns by default molecules that exactly match the queried stoichiometry. For example, a query of "C_6H_6" will return all structures containing six carbon atoms, six hydrogen atoms and nothing else. However, molecular formula search implemented in some databases, including PubChem Chemical Structure Search, has an option to allow other elements in returned hits (e.g., C_6H_6O or $C_6H_6N_2O$ for the "C_6H_6" query).

Identity search

Identity search is to locate a particular chemical structure that is "identical" to the query chemical structure. Although identity search seems conceptually straightforward, one should keep in mind that the word "identical" can have different notions. For example, if a molecule exists as multiple tautomeric forms in equilibrium, do you want to consider all these tautomers identical and search the database for all of them? If your query molecule has a chiral stereo center, should you consider both R- and S-forms in your search? In your identity search, do you want to include isotopically substituted species of the provided query molecule as well as the query itself? Depending on how to deal with these nuances of chemical structures, identical search will return different results. The identity search in the PubChem Chemical Structure Search allows users to choose a desired degree of "sameness" from several predefined options. To see these options, one need to expand the options section by clicking the "plus" button next to the "option" section heading.

Substructure and superstructure search

When a chemical structure occurs as a part of a bigger chemical structure, the former is called a *substructure* and the latter is referred to as a *superstructure*. For example, ethanol is a substructure of acetic acid, and acetic acid is a superstructure of ethanol.



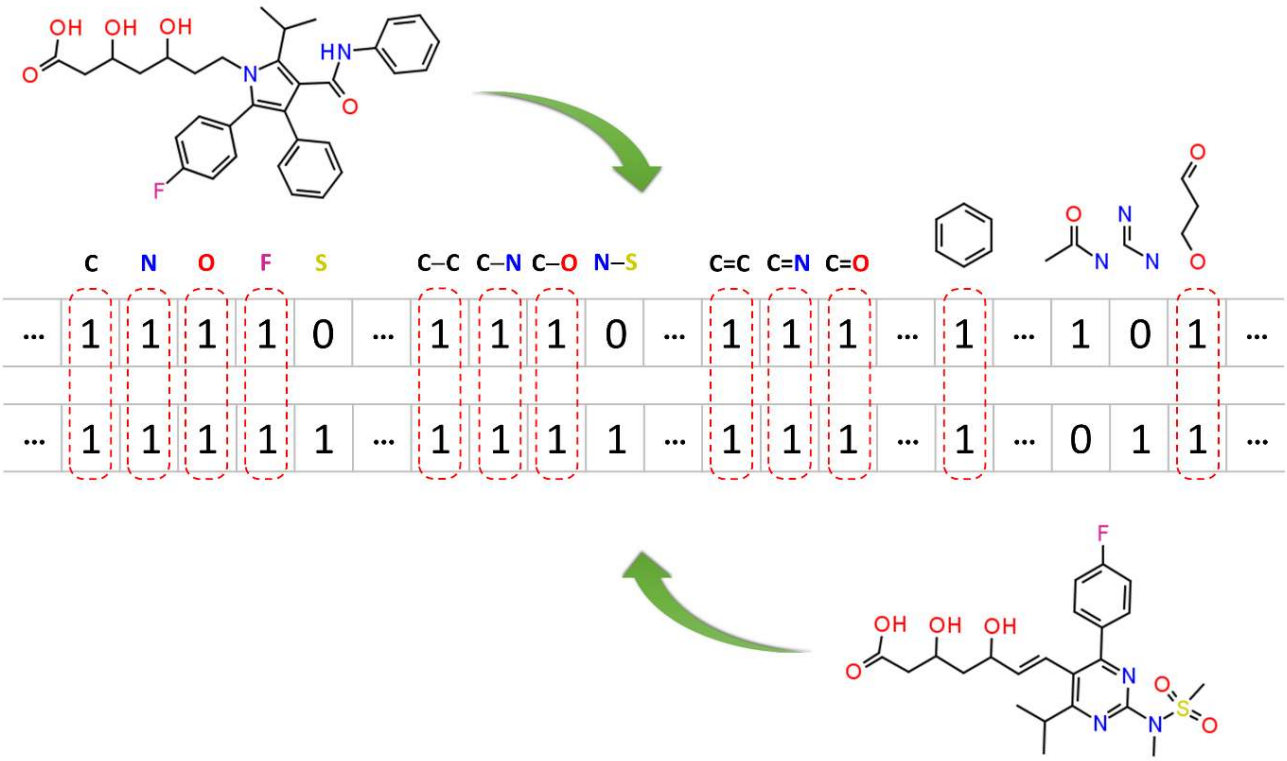
In substructure search, one provides an input substructure as a query to find molecules that contain the query substructure (that is, superstructures that contain the query substructure). On the contrary, superstructure search returns molecules that comprise or make up the provided chemical structure query (that is, substructures that is contained in the query superstructure). It should be noted that substructure search does *not* give you substructures of the query and that superstructure search does *not* return superstructures of the query.

It is possible to include explicit hydrogen atoms as part of the pattern being searched. For example, if you choose to do so, the SMILES queries [CH2][CH2][OH] and [CH3][CH][OH] will return molecules whose formula are R-CH_2-CH_2-OH and CH_3-CH(R)-OH, respectively. Substructure/superstructure searches implemented in some databases remove by default explicit hydrogens from the query molecule prior to search, the two SMILES queries [CH2][CH2][OH] and [CH3][CH][OH] may give you the same result as what the SMILES query CCO does, unless you specify that explicit hydrogens should be included in pattern matching.

In addition to explicit hydrogen atoms, there are additional factors that may affect results of substructure/superstructure searches, for example, whether to ignore stereochemistry, isotopism, tautomerism, formal charge, and so on.

Similarity search

Molecular similarity (also called chemical similarity or chemical structure similarity) is a fundamental concept in cheminformatics, playing an important role in computational methods for predicting properties of chemical compounds as well as designing chemicals with desired properties. The underlying assumption in these computational methods is that structurally similar molecules are likely to have similar biological and physicochemical properties (commonly called the similarity principle).[5] Molecular similarity is a straightforward and easy-to-understand concept, but there is no absolute, mathematical definition of molecular similarity that everyone agrees on. As a result, there are a virtually infinite number of molecular similarity methods, which quantify molecular similarity. Similarity search uses a molecular similarity method to find molecules similar to the query structure.



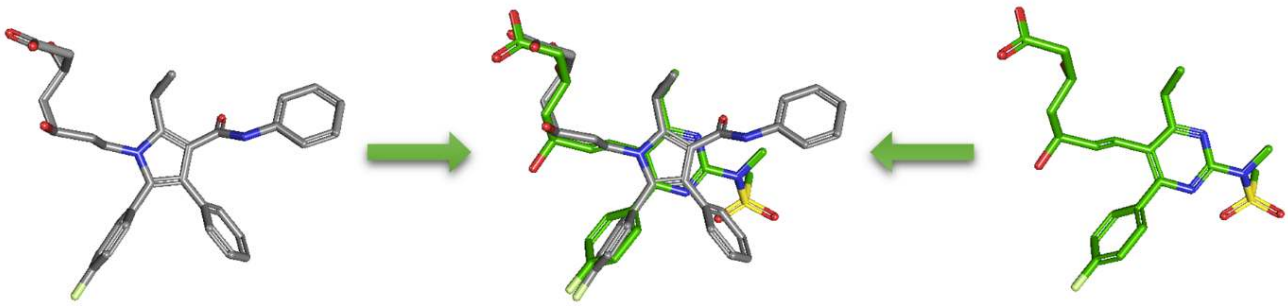
Molecular similarity methods can be broadly classified into two-dimensional (2-D) and three-dimensional (3-D) similarity methods. Typically, 2-D similarity methods use so-called molecular fingerprints. The most common types of molecular fingerprints are structural keys, which encode structural information of a molecule into a binary string (*that is*, a string of 0's and 1's). The position of each number in this string corresponds to a particular fragment. If the molecule has a particular fragment, the corresponding bit position is set to 1, and otherwise to 0. Note that there are many different ways to design molecular fingerprints, depending on what fragments are included in the fingerprint definition. PubChem uses its own fingerprint called PubChem subgraph fingerprints.

In 2-D similarity methods, structural similarity between two molecules is estimated by comparing their molecular fingerprints. Their similarity is quantified as a so-called similarity score or similarity coefficient. While several different methods can be used for computation of a similarity score, the underlying ideas are the same as each other: if the two fingerprints have 1's at the same position, it means that both compounds have the same fragment, and if the molecules share more common fragments, they are considered to be more similar. In conjunction with the PubChem subgraph fingerprints, PubChem 2-D similarity method use the Tanimoto coefficient[6-8]

$$Tanimoto = \frac{N_{AB}}{N_A + N_B - N_{AB}} \tag{12.7.1}$$

where N_A and N_B are the number of bits set in the fingerprints for molecules A and B, respectively, and N_{AB} is the number of bits set in both fingerprints. The Tanimoto score ranges from 0 (for no similarity) to 1 (for identical molecules). 2-D Similarity search returns molecules whose similarity scores with the query molecule are greater than or equal to a given Tanimoto cut-off value.

PubChem 3-D similarity method



As an alternative to 2-D similarity search, 3-D similarity search can also be performed using the "3D conformer" tab in PubChem Chemical Structure Search. 3-D similarity methods use the 3-D structures (that is, conformations) of molecules. PubChem's 3-D similarity method is based on the atom-centered Gaussian-shape comparison method by Grant and coworkers,[9,10,11,12] implemented in the Rapid Overlay of Chemical Structures (ROCS).[13,14] While the underlying mathematics of this approach is beyond the scope of this module, what this method essentially does is to find the "best" alignment of the 3-D structures of two molecules, which gives the maximized overlap between them. The 3-D similarity method quantifies the 3-D molecular similarity using three metrics.

- **Shape-Tanimoto (ST)**: quantifies steric shape similarity between two conformers.
- **Color-Tanimoto (CT)**: quantifies the overlap of functional groups between two conformers, such as hydrogen bond donors and acceptors, cations, anions, rings, and hydrophobes.
- **Combo-Tanimoto (ComboT)**: the sum of ST and CT scores between two conformers. It takes into account the shape similarity (ST) and functional group similarity (CT) simultaneously.

Because both the ST and CT scores range from 0 (for no similarity) to 1 (for identical molecules), the ComboT score may have a value from 0 to 2 (without normalization to unity). Note that the ST, CT and ComboT scores between two molecules can be evaluated in two different molecular superpositions: (1) in the ST- or shape-optimized superpositions, and (2) in the CT- or feature-optimization superpositions. In the ST-optimization approach, the shape overlap between the molecules (that is, the ST score) are maximized and the single-point CT score is evaluated at that superposition. On the contrary, the CT-optimization considers both ST and CT scores to find the best superposition between molecules, and the single-point ST score is computed at that superposition.

The 3-D similarity method used in PubChem requires the 3-D structures of molecules. PubChem generates a conformer ensemble containing up to 500 conformers for each compound that satisfy the following conditions[15,16,17]:

- Not too big or too flexible (with £ 50 non-hydrogen atoms and ≤ 15 rotatable bonds).
- Have only a single covalent unit (i.e., not a salt or a mixture).
- Consist of only supported elements (H, C, N, O, F, Si, P, S, Cl, Br, and I).
- Contain only atom types recognized by the MMFF94s force field.
- Fewer than six undefined atom or bond stereo centers.

About 90% of compounds in PubChem have computationally generated conformer models. Although each compound has up to 500 conformers (depending on the molecular size and flexibility), many PubChem tools and services support up to 10 conformers per compound. It should be emphasized that these conformers are not energy-minimized but sampled from the conformational space of a given molecule in such a way that the sampled conformers represent the overall diversity of shape and feature of the molecule.15,16,17 These conformer models aim to generate bioactive conformers, which would be found in protein-ligand complexes. For this reason, these conformers are often very different from their experimental structures determined in the gas phase.

References

(1) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L. Y.; He, J. E.; He, S. Q.; Shoemaker, B. A.; Wang, J. Y.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2016**, *44*, D1202.

(2) Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; Zhang, J. *Nucleic Acids Res.* **2017**, *45*, D955.

(3) Kim, S. *Expert Opinion on Drug Discovery* **2016**, *11*, 843.

(4) Ihlenfeldt, W. D.; Bolton, E. E.; Bryant, S. H. *J. Cheminform.* **2009**, *1*, 20.

(5) *Concepts and Applications of Molecular Similarity*; Johnson, M. A.; Maggiora, G. M., Eds.; John Wiley & Sons, Inc.: New York, NY, 1990.

(6) Chen, X.; Reynolds, C. H. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1407.

(7) Holliday, J. D.; Hu, C. Y.; Willett, P. *Combinatorial Chemistry & High Throughput Screening* **2002**, *5*, 155.

(8) Holliday, J. D.; Salim, N.; Whittle, M.; Willett, P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 819.

(9) Grant, J. A.; Pickup, B. T. *Journal of Physical Chemistry* **1995**, *99*, 3503.

(10) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. *Journal of Computational Chemistry* **1996**, *17*, 1653.

(11) Grant, J. A.; Pickup, B. T. *Journal of Physical Chemistry* **1996**, *100*, 2456.

(12) Grant, J. A.; Pickup, B. T. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer Academic Publishers: Dordrecht, 1997, p 150.

(13) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. *Journal of Medicinal Chemistry* **2005**, *48*, 1489.

(14) 3.1.0 ed.; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2010.

(15) Bolton, E. E.; Chen, J.; Kim, S.; Han, L. Y.; He, S. Q.; Shi, W. Y.; Simonyan, V.; Sun, Y.; Thiessen, P. A.; Wang, J. Y.; Yu, B.; Zhang, J.; Bryant, S. H. *J. Cheminform.* **2011**, *3*, 32.

(16) Bolton, E. E.; Kim, S.; Bryant, S. H. *J. Cheminform.* **2011**, *3*, 4.

(17) Kim, S.; Bolton, E. E.; Bryant, S. H. *J. Cheminform.* **2013**, *5*, 1.

12.7: Searching PubChem Using a Non-Textual Query is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

13: Computer-Aided Drug Design

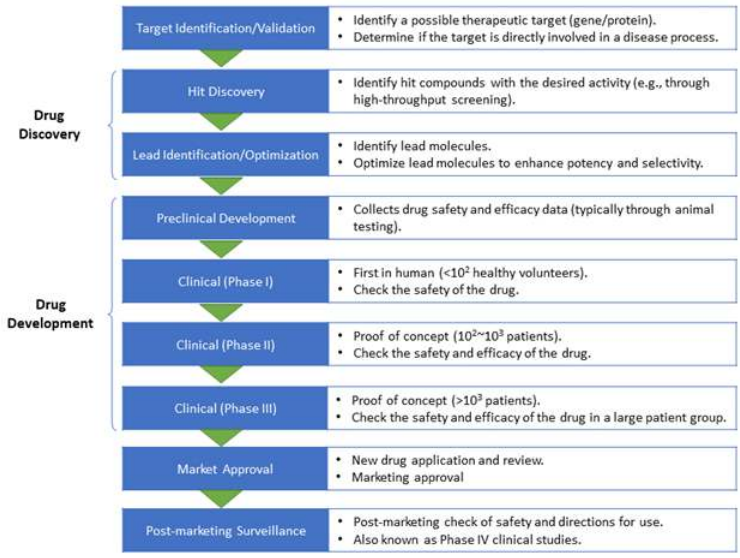
13.1: Reading
13.2: How to Dock Your Own Drug

13: Computer-Aided Drug Design is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

13.1: Reading

7.1. Drug discovery & development process

The drug discovery and development process is very resource-intensive and time-consuming. Bringing a new drug to market (from drug discovery through clinical trials to marketing approval) typically takes between 10 and 15 years and costs $1.395 billion (2013 dollars) on average [1]. Figure 1 shows a schematic diagram of the drug discovery and development process.



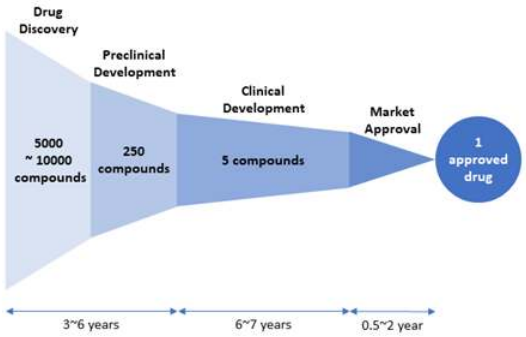
Copy and Paste Caption here

Figure 1. Drug discovery and development process.

The process presented in Figure 1 is simplified, ignoring many details within each stage. In reality, the drug discovery and development process is much more complicated, as illustrated in this Figure (https://www.nature.com/articles/nrd.2017.217/figures/1) [2].

This chapter describes some computational approaches used in the drug discovery stage. To further discuss computer-aided (or computer-assisted) drug discovery, it is necessary to learn the following commonly used terms.

- **Actives**
  Substances that meet a threshold level of activity in a primary screen, which typically measures the activity of compounds against the target at a single concentration. Because the activity was measured only at a single concentration, it is not possible to tell whether a compound can interact with the target in a dose-response way. Often the structure and purity of screening substances are not confirmed.
  - **Hits**
    Hits are compounds with intrinsic activity (IC50, EC50, etc.) against the target and they are characterized through secondary assays (which measure compounds' activity at multiple concentrations). In general, because hits have limited potency and/or selectivity, they are not suitable for in vivo studies (in animals or humans). However, they provide a starting point for structure activity relationship (SAR) analysis to help improve potency/selectivity and other drug properties.
- **Leads**
  A Lead represents a compound series that shows a relationship between chemical structure and target-based activity (in biochemical and cell-based assays). Compounds within the series have physicochemical properties, potency and selectivity deemed appropriate for in vivo evaluation.
- **Drug candidates**
  Compounds with strong therapeutic potential and whose activity and specificity have been optimized through the lead optimization step. These compounds move to the preclinical development stage for in vivo animal testing.



Copy and Paste Caption here

Figure 2. Drug attrition rate. Of 5,000–10,000 compounds experimentally screened in the drug discovery stage, approximately 250 will enter preclinical testing and 5 will enter clinical testing. Only one compound of them will be approved. [Adapted from "New Drug Approvals in 2011" (http://phrma-docs.phrma.org/sites/default/files/pdf/nda2011.pdf)].

7.2. What is virtual screening?

Virtual screening (VS) is a computational technique used in drug discovery to prioritize the compound selection from a large compound library for subsequent experimental assays. VS aims to ensure that those molecules with the largest a priori probabilities of being active against the target (protein/gene/disease) are tested first in a lead discovery program. VS is a cost-effective approach that complements high-throughput screening and it is routinely used in drug discovery projects.

Depending on the information available about the target and/or its ligands at the beginning of the VS campaign, VS can be broadly divided into two main approaches: structure-based and ligand-based approaches. The structure-based approaches, such as molecular docking, use the 3-D structure of the target macromolecule (protein/gene) to dock the candidate molecules and rank them based on their predicted binding affinity or complementarity to the binding site. On the other hand, the ligand-based approaches use a set of known actives against the target to identify database compounds that are likely to be active based on similarity/commonality between the known actives and database compounds. Because the structure-based approaches require the 3-D structural information of the target macromolecules, they are not a feasible option when the 3-D structure of the target is not available. In contrast, the ligand-based approaches can be used regardless of whether the target 3-D structure is available or not.

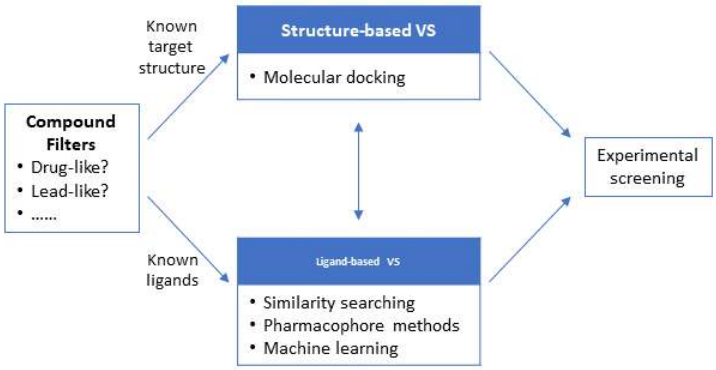


Figure 3. A schematic diagram of the virtual screening process, which involves many computational approaches to prioritize the compound selection for further experiments

.

7.3. Compound Filters

When screening a large compound library for drug discovery, it is desirable to identify problematic compounds that are not likely to lead to successful drug discovery campaign and exclude them from further computational or experimental screening. For example, if a molecule is *not* absorbed very well by the human body, that molecule may not be a good candidate for an orally administered drug, even if it shows a good activity against the target protein in *in vitro* testing. Therefore, one may want to exclude this compound from further testing.

Lipinski's rule of five (also known as the Pfizer's rule of five) (RO5) [3] is a rule of thumb to evaluate “drug-likeness” and helps determine whether a compound has good solubility and permeability that would make the compound a likely orally administered drug in human. According to RO5, drug-like compounds have:

- No more than 5 hydrogen-bond donors
- No more than 10 hydrogen-bond acceptors
- Molecular weight of 500 or less
- Calculated logP (CLogP) less than 5

Lipinski and coworkers [3] analyzed 2245 compounds that had reached phase II trials or higher and formulated RO5 based on the observation that most orally administered drugs are relatively small and moderately lipophilic.

While RO5 provides a fast and simple way to estimate oral bioavailability of molecules, it is a rule of thumb, not accurate and hard-set criteria. For example, ~16% of orally administrated drugs violate at least one criterion of RO5 [4]. For this reason, several variants of RO5 have been proposed as explained in detail in a recent review article [5].

An interesting concept related to drug-likeness is “lead-likeness”[6-9]. In the analysis of 470 lead-drug pairs, Hann et al [6]. found that, on average, lead compounds had lower molecular weight and logP and fewer hydrogen bond acceptors and more hydrogen bond donors, compared to the respective drugs developed from them. Therefore, it can be problematic to apply drug-likeness filters to compound libraries designed for lead discovery. This led to the development of “lead-likeness” filters. An example is Congreve’s rule of 3 (RO3) for lead-like compounds [10], which include:

- The number of hydrogen bond donors ≤3
- The number of hydrogen bond acceptors ≤ 3
- Molecular weight < 300
- ClogP ≤ 3

In general, the lead optimization step in drug discovery involves the modification of the lead molecule to give the final drug and this typically increases the molecular “complexity”, as reflected in the differences between RO3 and RO5. For this reason, some people argue that lead-like filters, rather than drug-like filters, should be used when performing virtual screening [6,7]. More detailed discussion about drug-likeness and lead-likeness can be found elsewhere [6-9].

7.4. Structure-based virtual screening

Structural-based approaches, exemplified by molecular docking, require the 3-D structure of the target macromolecule, which can be either experimentally determined through X-ray crystallography or computationally predicted through homology modeling. In molecular docking, the most likely binding mode for each compound is identified and assign a priority order to the molecules. While a large number of molecular docking methods have been proposed, all of them have two essential components to solve the molecular docking problem:

· **docking algorithm**, used to possible protein-ligand geometries (also called “poses”), and

· **scoring function**, which is a mathematical formula used to score or rank these poses.

Docking algorithms

Molecular docking methods can be classified into rigid-body docking and flexible docking, depending on the degree to which the flexibility of ligands and their macromolecule target is considered during the docking process. The rigid-body docking uses “fixed” structures of the ligands and target, treating them as rigid bodies. While the earliest docking programs used the rigid-body docking approach, more recent programs can explore the conformational space of the ligands (by changing the torsional angles of the ligands during the docking process). Some programs also permit conformational flexibility to the protein. Because rigid-body docking is faster than flexible docking, rigid docking may be preferred for initial quick screening a very large compound library. However, poses from such initial rigid docking need to be refined and optimized through flexible docking methods. In addition, because of the advance of computational resources and efficiency, flexible docking is becoming more popular. There are several approaches to take ligand flexibility in molecular docking, including:

· **Systematic methods**: incorporate ligand flexibility by gradually changing structural parameters of the ligands (such as torsional angles, translational and rotational degrees of freedom). Because large conformational space prohibits an exhaustive systematic search, some algorithms use heuristics to focus on regions on conformational space that are likely to contain good poses. While effective in conformational search, it can find a local minimum, rather than the global minimum.

· **Stochastic methods**: makes random changes to the ligand structure to perform the conformational search. It generates an ensemble of conformations that populates a wide range of the energy landscape. While this avoids trapping the final structure at a local energy minimum and increases the probability of finding the global minimum. However, it covers a broader range of the energy landscape, it is computationally more expensive.

· **Genetic algorithms** [11,12]: use the concepts of evolution and natural selection. It begins with encoding structural parameters of the initial structure in a vector (called a chromosome) and generating an ensemble (or population) of chromosomes using the random search algorithm. Each chromosome in this population is evaluated and the most "adapted" ones (with the lowest energy values) are selected as "templates" for the generation of next population. Each iteration of this process will lead to lower-energy binding poses than those from the previous iteration and after a reasonable number of iterations, the chromosome population will converge to a chromosome corresponding to the global energy minimum.

Scoring Functions

Scoring functions [13] are used to evaluate the binding affinity between the ligand and its target. different scoring functions have been developed over the years, and they can be broadly classified into four groups: (1) force-field-based, (2) empirical scoring functions, (3) knowledge-based functions, and (4) consensus-scoring functions.

· **Force-field scoring functions**

Force-field scoring functions use force field parameters used in molecular mechanics calculations. These parameters are derived from experimental data and ab initio quantum mechanical calculations. The force field scorning functions estimate the binding energy by summing the contributions of various bonded terms (e.g., stretching, bending, and torsional forces) and non-bonded terms (e.g., electrostatic and van der Waals interactions). An example of this type of scoring functions is the one used by DOCK [14], whose energy parameters are taken from the AMBER force fields [15,16].

· **Empirical scoring functions**

Empirical scoring functions [17-19] express the binding energy of a protein-ligand complex as a weighted sum of terms that represent physical events involved in the complex formation, such hydrogen bonding, hydrophobic contact, desolvation effects (due to the loss of solvent (water) molecules that stabilizes the ligand), and entropy penalty (due to the loss of ligand flexibility upon the complex formation). The weight and parameters in each term are empirically determined through regression analysis of a set of protein-ligand complexes with known binding affinities. While the performance of empirical scoring functions relies on the accuracy of the experimental data used to develop them, they are faster than force-field-based scoring functions. Empirical scoring functions are used in some popular molecular docking programs like Surflex [20] and FlexX [21,22].

· **Knowledge-based functions**

Knowledge-based functions [23-27] exploit information contained in experimentally determined 3-D structures of protein-ligand complexes. For each ligand-protein atom pair, an interaction potential is generated, which gives the pairwise interaction energy as a function of their separation. These potentials are generated through statistical analysis of the interatomic distance distribution observed in known protein-ligand complexes. The underlying assumption in this approach is that interatomic contacts observed more often in the data set are likely to represent favorable contacts that increase the binding affinity, while those contacts occurring with less frequency are unfavorable and likely to decrease the binding affinity. The final score is computed from these individual interactions.

· **Consensus scoring functions**

Because each scoring function has its strength and weakness, consensus scoring [28-33] has been gaining popularity more recently, which simultaneously uses multiple scoring approaches together to achieve improved accuracies. Many consensus-scoring strategies have been proposed and examples are MultiScore [29], X-CSCORE [30], GFscore [31], supervised consensus scoring (SCS) [32], and SeleX-CS [33].

7.5. Ligand-based virtual screening

Ligand-based approaches use a set of active compounds that are known to interact with the target protein. These approaches are based on the Similar Property Principle, which states structurally similar molecules are likely to have similar (physicochemical and biological) properties. In contrast to structure-based approaches, the ligand-based approaches can be applied when the structure of the target macromolecule is not known. Examples of ligand-based approaches are:

· Pharmacophore methods: identify the pharmacophore pattern common to a set of known actives and uses this pattern in a subsequent substructure search. IUPAC defines a pharmacophore as “the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response.”[34]

· Machine learning methods: use prediction models developed from a training set containing known actives and known inactives. These methods will be discussed in Chapter 8.

· Similarity methods: find molecules structurally similar to known active molecules, based on similarity measures. This topic was discussed in Chapter 6.

7.6. Prediction of ADMET Properties

During the drug discovery and development process, it is very important to consider not only how tightly a potential drug molecule can bind to the target protein (pharmacodynamics), but also how the molecule reach its site of action within the body (pharmacokinetics). This involves the absorption of the drug molecule by the body and the drug transport to the target organ/tissue. While a sufficient number of drug molecules should be available in the body to give the desired therapeutic effect but they must ultimately be removed from the body through metabolism and excretion. In addition, neither the drug nor its metabolites should be toxic. In pharmacology, these properties are often referred to as ADMET properties (which stands for Absorption, Distribution, Metabolism, Excretion, and Toxicity).

During the drug discovery process, various computational tools are routinely used to predict a wide range of ADMET properties of compounds [35-38], including:

· Solubility in water, which affects oral bioavailability of the drug.

· Caco-2 cell monolayer permeability, which is an experimental model for evaluating the intestinal absorption of drugs.

· Permeability through blood-brain barrier between the systemic circulation and the brain.

· Interaction with cytochrome P450 (CYP) proteins involved in drug metabolism.

· Binding affinity to Human ether-a-go-go related gene (hERG) protein, which is responsible for cardiotoxicity of many drugs.

· Interaction with P-glycoprotein efflux pump, involved in the active transport of various compounds out of cells [39].

· Plasma-protein binding, which help estimates the amount of free drugs that can cross membranes.

7.7. Further Reading

· Principles of early drug discovery

https://doi.org/10.1111/j.1476-5381.2010.01127.x

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3058157/

· Recognizing Pitfalls in Virtual Screening: A Critical Review
https://doi.org/10.1021/ci200528d

· ADMET IN SILICO MODELLING: TOWARDS PREDICTION PARADISE?
https://doi.org/10.1038/nrd1032

· Computational Methods in Drug Discovery

https://doi.org/10.1124/pr.112.007336

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3880464/

· Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions

https://doi.org/10.1039/C0CP00151A

· Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges
https://doi.org/10.3389/fphar.2018.01089

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6165880/

References

1. DiMasi JA, Grabowski HG, Hansen RW: **Innovation in the pharmaceutical industry: New estimates of R&D costs.** *J Health Econ* 2016, **47:**20-33.

2. Wagner J, Dahlem AM, Hudson LD, Terry SF, Altman RB, Gilliland CT, DeFeo C, Austin CP: **A dynamic map for learning, communicating, navigating and improving therapeutic development.** *Nat Rev Drug Discov* 2018, **17:**151-153.

3. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.** *Adv Drug Deliv Rev* 1997, **23:**3-25.

4. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL: **Quantifying the chemical beauty of drugs.** *Nat Chem* 2012, **4:**90-98.

5. Mignani S, Rodrigues J, Tomas H, Jalal R, Singh PP, Majoral JP, Vishwakarma RA: **Present drug-likeness filters in medicinal chemistry during the hit and lead optimization process: how far can they be simplified?** *Drug Discov Today* 2018, **23:**605-615.

6. Hann MM, Leach AR, Harper G: **Molecular complexity and its impact on the probability of finding leads for drug discovery.** *J Chem Inf Comput Sci* 2001, **41:**856-864.

7. Teague SJ, Davis AM, Leeson PD, Oprea T: **The design of leadlike combinatorial libraries.** *Angew Chem-Int Edit* 1999, **38:**3743-3748.

8. Oprea TI, Davis AM, Teague SJ, Leeson PD: **Is there a difference between leads and drugs? A historical perspective.** *J Chem Inf Comput Sci* 2001, **41:**1308-1315.

9. Oprea TI, Allu TK, Fara DC, Rad RF, Ostopovici L, Bologa CG: **Lead-like, drug-like or "pub-like": how different are they?** *J Comput-Aided Mol Des* 2007, **21:**113-119.

10. Congreve M, Carr R, Murray C, Jhoti H: **A rule of three for fragment-based lead discovery?** *Drug Discov Today* 2003, **8:**876-877.

11. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ: **Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.** *J Comput Chem* 1998, **19:**1639-1662.

12. Jones G, Willett P, Glen RC, Leach AR, Taylor R: **Development and validation of a genetic algorithm for flexible docking.** *J Mol Biol* 1997, **267:**727-748.

13. Huang SY, Grinter SZ, Zou XQ: **Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions.** *Phys Chem Chem Phys* 2010, **12:**12899-12908.

14. Meng EC, Shoichet BK, Kuntz ID: **AUTOMATED DOCKING WITH GRID-BASED ENERGY EVALUATION.** *J Comput Chem* 1992, **13:**505-524.

15. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P: **A NEW FORCE-FIELD FOR MOLECULAR MECHANICAL SIMULATION OF NUCLEIC-ACIDS AND PROTEINS.** *J Am Chem Soc* 1984, **106:**765-784.

16. Weiner SJ, Kollman PA, Nguyen DT, Case DA: **AN ALL ATOM FORCE-FIELD FOR SIMULATIONS OF PROTEINS AND NUCLEIC-ACIDS.** *J Comput Chem* 1986, **7:**230-252.

17. Murray CW, Auton TR, Eldridge MD: **Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model.** *J Comput-Aided Mol Des* 1998, **12:**503-519.

18. Guedes IA, Pereira FSS, Dardenne LE: **Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges.** *Front Pharmacol* 2018, **9:**18.

19. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP: **Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes.** *J Comput-Aided Mol Des* 1997, **11:**425-445.

20. Jain AN: **Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine.** *J Med Chem* 2003, **46:**499-511.

21. Kramer B, Rarey M, Lengauer T: **Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking.** *Proteins* 1999, **37:**228-241.

22. Rarey M, Kramer B, Lengauer T, Klebe G: **A fast flexible docking method using an incremental construction algorithm.** *J Mol Biol* 1996, **261:**470-489.

23. Muegge I, Martin YC: **A general and fast scoring function for protein-ligand interactions: A simplified potential approach.** *J Med Chem* 1999, **42:**791-804.

24. Mitchell JBO, Laskowski RA, Alex A, Thornton JM: **BLEEP - Potential of mean force describing protein-ligand interactions: I. Generating potential.** *J Comput Chem* 1999, **20:**1165-1176.

25. Mitchell JBO, Laskowski RA, Alex A, Forster MJ, Thornton JM: **BLEEP - Potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data.** *J Comput Chem* 1999, **20:**1177-1185.

26. Huang SY, Zou XQ: **An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials.** *J Comput Chem* 2006, **27:**1866-1875.

27. Huang SY, Zou XQ: **An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function.** *J Comput Chem* 2006, **27:**1876-1882.

28. O'Boyle NM, Liebeschuetz JW, Cole JC: **Testing Assumptions and Hypotheses for Rescoring Success in Protein-Ligand Docking.** *J Chem Inf Model* 2009, **49:**1871-1878.

29. Terp GE, Johansen BN, Christensen IT, Jorgensen FS: **A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities.** *J Med Chem* 2001, **44:**2333-2343.

30. Wang RX, Lai LH, Wang SM: **Further development and validation of empirical scoring functions for structure-based binding affinity prediction.** *J Comput-Aided Mol Des* 2002, **16:**11-26.

31. Betzi S, Suhre K, Chetrit B, Guerlesquin F, Morelli X: **GFscore: A general nonlinear consensus scoring function for high-throughput docking.** *J Chem Inf Model* 2006, **46:**1704-1712.

32. Teramoto R, Fukunishi H: **Supervised consensus scoring for docking and virtual screening.** *J Chem Inf Model* 2007, **47:**526-534.

33. Bar-Haim S, Aharon A, Ben-Moshe T, Marantz Y, Senderowitz H: **SeleX-CS: A New Consensus Scoring Algorithm for Hit Discovery and Lead Optimization.** *J Chem Inf Model* 2009, **49:**623-633.

34. Wermuth G, Ganellin CR, Lindberg P, Mitscher LA: **Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998).** *Pure Appl Chem* 1998, **70:**1129-1143.

35. Norinder U, Bergstrom CAS: **Prediction of ADMET properties.** *ChemMedChem* 2006, **1:**920-937.

36. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA: **Toward in silico structure-based ADMET prediction in drug discovery.** *Drug Discov Today* 2012, **17:**44-55.

37. Gleeson MP: **Generation of a set of simple, interpretable ADMET rules of thumb.** *J Med Chem* 2008, **51:**817-834.

38. van de Waterbeemd H, Gifford E: **ADMET in silico modelling: Towards prediction paradise?** *Nat Rev Drug Discov* 2003, **2:**192-204.

39. Li D, Chen L, Li YY, Tian S, Sun HY, Hou TJ: **ADMET Evaluation in Drug Discovery. 13. Development of in Silico Prediction Models for P-Glycoprotein Substrates.** *Mol Pharm* 2014, **11:**716-726.

Contact Bob Belford, rebelford@ualr.edu if you have any questions.

13.1: Reading is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

13.2: How to Dock Your Own Drug

Molecular Docking Experiments

This tutorial explains how to perform molecular docking experiments using Autodock Vina (molecular docking software) and UCSF Chimera (molecular visualization software), both of which are freely available for academic users. In this tutorial, we will use the 3-D structure of muscle glycogen phosphorylase from rabbit (PDB ID: 1c8k) as a target macromolecule template and three ligand molecules (stored in the files generated using the jupyter notebook for Chapter 7 (lecture07-virtual-screening_v0.ipynb)).

Step 0: Installation of Autodock Vina and Chimera

(0-A) Download Autodock Vina (http://vina.scripps.edu/download.html) and install it on your computer.

(0-B) Download UCSF Chimera (http://www.cgl.ucsf.edu/chimera/download.html) and install it on your computer.

· For both programs, make sure that you download a correct version, depending on the operation system of your computer (Windows, Mac, or Linux) and whether it is 64-bit or 32-bit.)

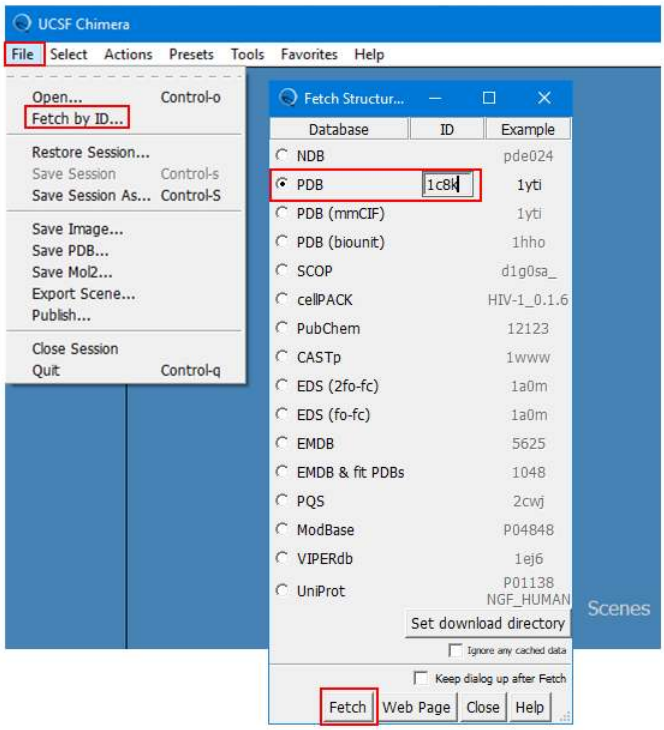
(0-C) Read the tutorial "UCSF Chimera – Getting Started" (https://www.cgl.ucsf.edu/Outreach/Tutorials/GettingStarted.html) to learn how to use chimera. Other tutorials are also available at https://www.cgl.ucsf.edu/chimera/tutorials.html.

(0-D) Watch this YouTube video (https://www.youtube.com/watch?v=cLGQ-951FDk). This video explains how to perform a docking experiment using Autodock Vina and Chimera.

Step 1: Loading the 3-D structure of the co-crystallized target-ligand complex.

(1A) Open Chimera.

(1B) Select "FILE" --> "Fetch by ID" --> "PDB", provide the PDB ID ("1c8k"), and click "Fetch" at the bottom. This will load the 3-D structure of "1c8k".



(1C) Save this structure as "pygm_1c8k.pdb" (for your reference) by selecting "FILE" --> "Save PDB …" and providing the filename. [Here, "pygm" is the symbol of the gene that encodes the target protein (glycogen phosphorylase, muscle associated).]

Step 2: Prepare a ligand structure for docking, using the inhibitor co-crystallized with the target in 1c8k. This docking example will allow us to compare a predicted pose of the inhibitor with the experimental pose. Note that the 1c8k protein-ligand complex has

two small molecules bound to the target protein. We need to figure out which one is the inhibitor.

(2A) Review the ligand information for PDB ID 1c8k, available at: https://www.rcsb.org/structure/1C8K. This protein-ligand complex has two small molecules bound to the protein: Flavopyridol (CPB) and Vitamin B6 Phosphate (PLP). According to the abstract available on this page, CPB is an inhibitor and PLP is a substrate of the protein. In this docking experiment, we will dock ligands to the inhibition site of the protein.

(2B) Investigate the structure of the inhibition site. Five amino acid residues with close contacts to the CPB ligand are already shown by default. Place the mouse pointer on each of these residues and record their names and IDs. (For example, PHE285, GLU382, …)

(2C) Delete all residues and ligand/water molecules, **except for CPB**. This can be done by selecting:

(2C-1) Select --> Residue --> CPB

(2C-2) Select --> Invert (all models)

(2C-3) Actions --> Atoms/Bonds --> Delete

(2D) Save the ligand molecule by following these steps:

(2D-1) select File --> Save PDB…

(2D-2) provide the file name "**pygm_1c8k_cpb**"

(2D-3) uncheck "Use untransformed coordinates"

(2D-4) press "Save"

(2E) Close the current session by selecting: File --> Close session.

Step 3: Prepare a protein structure for docking.

(3A) Load the 1c8k structure [in the same way as Step (1A)-(1B)].

(3B) Delete all ligand/water molecules, **except for protein**. This can be done by selecting:

(3B-1) Select --> Structure --> Protein

(3B-2) Select --> Invert (all models)

(3B-3) Actions --> Atoms/Bonds --> Delete

(3C) Save the protein structure by following these steps:

(3C-1) select File --> Save PDB…

(3C-2) provide the file name "**pygm_1c8k_protein.pdb**"

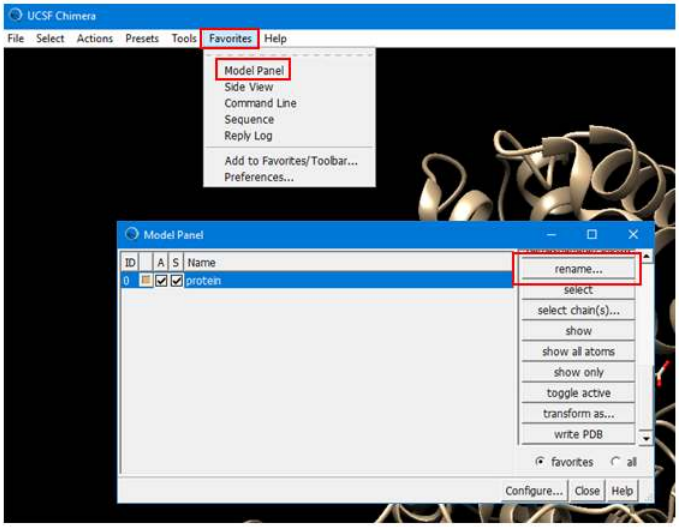
(3C-3) uncheck "Use untransformed coordinates"

(3C-4) press "Save"

NOTE: DO NOT CLOSE THE CURRENT SESSION NOW. WE WILL USE THIS PROTEIN STRUCTURE TO SET UP THE DOCKING EXPERIMENT.

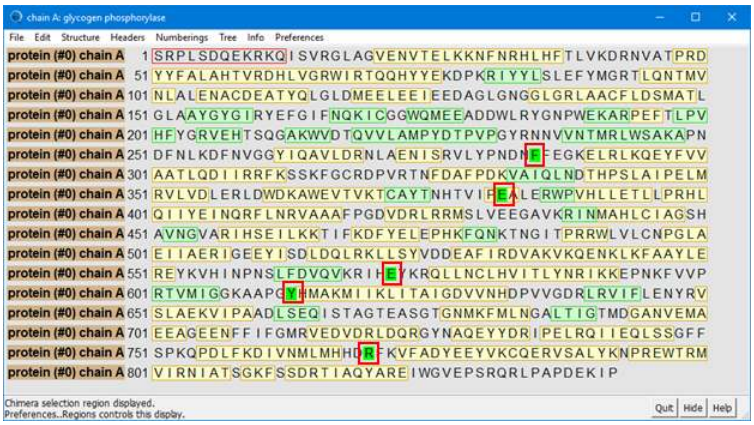
Step 4: Set up a molecular docking experiment.

(4A) Open the Model Panel by selecting: Favorites --> Model Panel (Alternatively, Tools --> General Controls --> Model Panel.) Rename the existing protein structure to "protein".



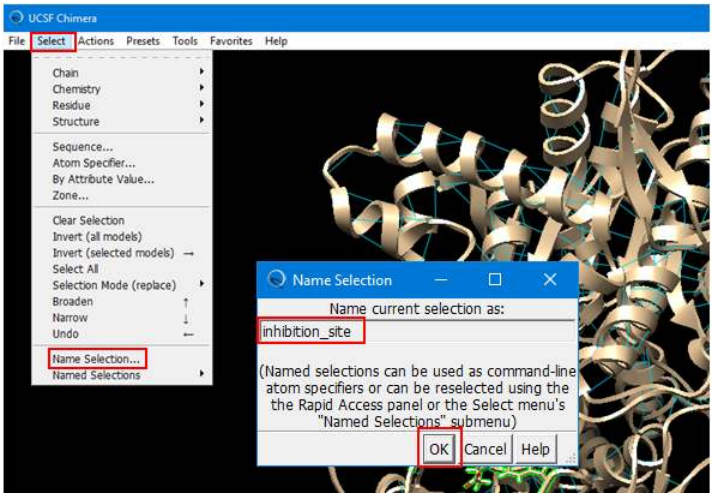
(4B) Create a named selection for the inhibition site, using the residue names/IDs from **Step 2B**.

(4B-1) Open the Sequence window by choosing: Tools --> Sequence --> Sequence and then choosing “protein”, which will be used as a template for docking experiments.



(4B-2) Select the five residues identified in Step 2B. When the mouse point is placed on a residue, their symbol and ID will be shown at the right-bottom of the windown (e.g., ILE 584). Use “control+drag” to add new region and “shift+drag” to add to region. Selected residues will be highlighted in green.

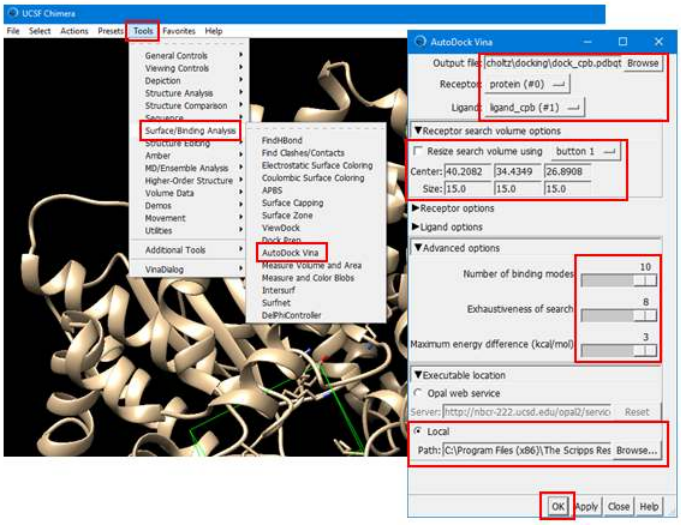
(4B-3) Go to the main Chimera window and choose: Select --> Name selection and provide “inhibition_site” as an alias for the five selected residues. After naming this selection, “inhibition_site” will be listed under the “Named Selections”. This will be used later to display the inhibition site.



(4C) Without closing the current session (which has the protein structure), open the ligand file prepared in Step 2 ("pygm_1c8k_cpb.pdb").

(4D) Check the location of the ligand by using the "Model Panel", which can be accessed by selecting: Favorites --> Model Panel (alternatively, by selecting: Tools --> General Controls --> Model Panel. Note that the ligand is not necessarily positioned at the binding pocket (because the "Use untransformed coordinates" box was unchecked when the ligand was saved). Rename the ligand as "ligand_cpb".

(4E) Open the Autodock Vina window by selecting: Tools --> Surface/Binding Analysis --> Autodock Vina.



(4E-1) Set the output file name to "dock_cpb.pdbqt". [Note that the file extension is "pdbqt"].

(4E-2) Select "protein" as the receptor and "ligand_cpb" as the ligand.

(4E-3) Define the simulator box using the "Receptor Search Box Options". This box defines the location of the binding site for the ligand and it tells the docking program where it should search to predict the best binding pose of the ligand. Once you check the box for "Resize search volume using Button X" (where X can be 1, 2, or 3 or Ctrl+1, Ctrl+2, Ctrl3, depending on your machine), you can draw, move, resize the simulator box. [To resize the box, you need to click and drag one of the faces of the box while holding down the left mouse button]. Make sure that all residues to which the CPB ligand was bound should be contained in the simulator box. If you have difficulty in setting up the box, use the following numbers:

Center: 40.2082, 34.4349, 26.8908

Size: 15.00, 15.00, 15.00

[These numbers were used to get the docking results shown in the images included in this tutorial.]

(4E-4) Select “Advanced Option” and adjust the parameters to the largest possible values.

· 10 for Number of binding mode

· 8 for Exhaustiveness of search

· 3 for maximum energy difference (kcal/mol)

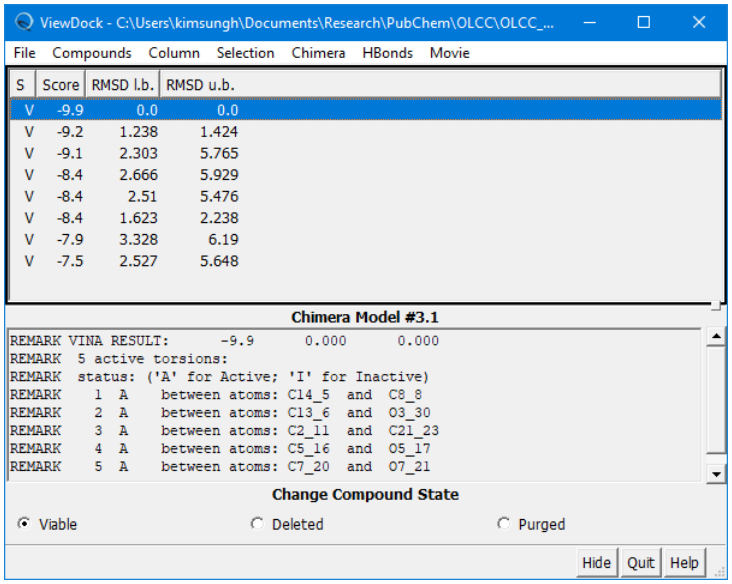
In this exercise, we are testing only a few ligands, which does not take too much computational resources, so it is okay to adjust these parameters to the maximum possible values. However, if you are performing a large-scale screening, they need to be adjusted accordingly.

(4E-5) Select “Executable Location” and then “Local” and provide the path of the “vina.exe” installed on you PC. It should be something similar to:

C:\Program Files (x86)\The Scripps Research Institute\Vina\vina.exe

(4E-6) Press “OK” to start the docking experiment. This will take a few minutes, depending on your machine. You may encounter some warning messages, but the computation will continue until it finds the requested number of poses.

(4E-7) Once the docking experiment is done, the **View Dock** window will be presented, which shows the list of poses predicted, along with additional information.



Step 5: Review the docking results

(5A) Look into the 10 predicted poses by selecting them from the ViewDock window.

(5B) Overlap the experimentally determined pose with the best predicted pose by opening the structure of PDB 1c8k [see **Step (1B)**].

(5C) Show only the CPB ligands (from the docking and the PDB), by taking these steps:

(5C-1) Select --> Residue --> CPB

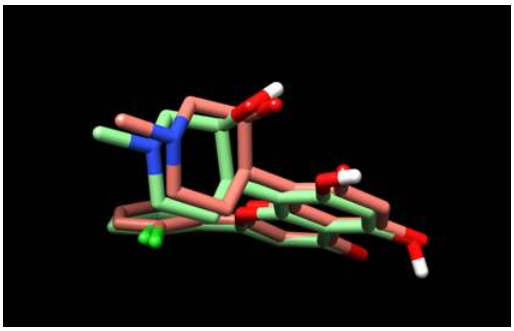
(5C-2) Select --> Invert (all models)

(5C-3) Actions --> Atoms/Bonds --> Hide

(5C-4) Actions --> Ribbon --> Hide

The resulting view will be similar to this:

Here, the predicted pose is in green and the experimental pose is in red.



(5D) Save the current view as “dock_cpb_img.png” by selecting File --> Save Image… and providing the file name.
(5E) Show the protein and other ligands that were hidden for image generation [in **(5C)-(5D)**], by taking these steps:

(5E-1) Select --> Structure --> Protein

(5E-2) Actions --> Ribbon --> Show

(5E-3) Select --> “Clear Selection”.

Step 6: Now dock one of the ligand molecules returned from the jupyter notebook.

(6A) Open the file containing the hit compound with the highest rank (pygm_lig0_11779854.mol).

(6B) Check the location of the ligand using the Model Panel [see **Step (4A)**] and rename the ligand as “ligand0”.

(6C) Perform the docking experiment using “ligand0”, in the similar manner to **Step (4E)**.

· Make sure that the **output file** should be set differently from the previous one in order not to overwrite the previous data. (For example, set the output file name “**dock_ligand0.pdbqt**”.

· Make sure that “**ligand0**” is selected as the ligand.

· Use the same simulator box location and size.

· Use the same values for other parameters.

Note: the ligands contained the .mol files are already energy-minimized and ready for docking. However, if a ligand does not have a 3-D structure (e.g., only the 2-D structure or SMILES string is available), a reasonable 3-D structure should be generated using the “Minimize Structure” and “Dock Prep” tools in Chimera.

(6D) Visualize the binding pose

(6D-1) Hide the protein structure by doing:

· Select --> Structure --> Protein, and

· Actions --> Ribbon --> Hide

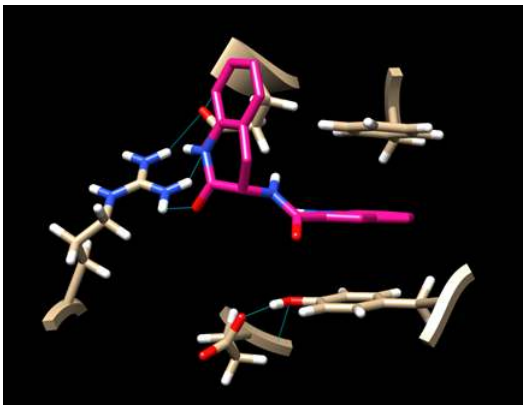
(6D-2) Show only the residues at the binding site by doing:

· Select --> Named Selections --> inhibition_site

· Actions --> Atoms/Bonds --> Show

· Actions --> Ribbon --> Show

(6D-3) Now you have a view similar to this:

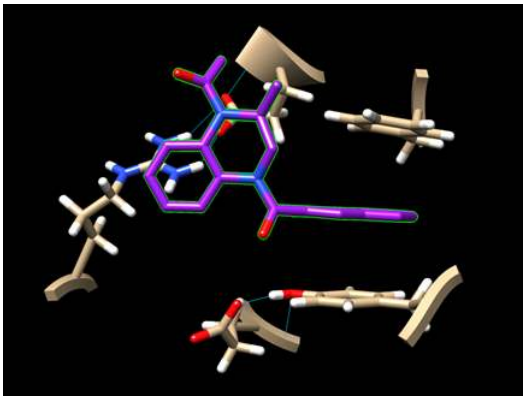


An aromatic ring of the ligand is stacked with the rings of TYR 163 and PHE 285, indicating some hydrophobic interaction. There is also a hydrogen bond between the ligand and ARG 770.

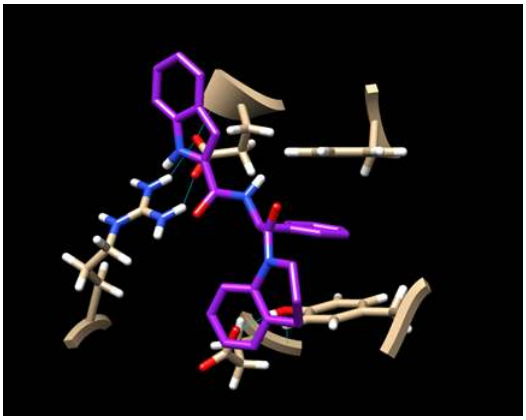
Save this image as “dock_ligand0_img.png”.

Step 7: For the remaining two ligands, repeat the docking experiment and generate the image of their best binding poses.

· dock_ligand1_img.png (for pygm_lig1_118078858.mol)



· dock_ligand2_img.png (for pygm_lig2_53013349.mol)



13.2: How to Dock Your Own Drug is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

14: Does it work? Clinical Testing

Clinical trials are used to ascertain whether drugs are safe, effective, and the extent to which they provide benefit to patients. This chapter contains some basic information on how clinical trials are designed and analyzed.

14.1: Understanding Clinical Trial Types (Video)
14.2: Statistical Basics
14.3: Random Sampling
14.4: Clinical Studies
14.5: Types of clinical interventions
14.6: Evolution of new intervention products and sequence of study phases
14.7: Interpreting Clinical Research

Sources:

Phases of a Clinical Trial infographic from the Lupus Foundation, https://lupustrials.org/about-trials/phases-of-a-trial/

14: Does it work? Clinical Testing is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

14.1: Understanding Clinical Trial Types (Video)



14.1: Understanding Clinical Trial Types (Video) is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

14.2: Statistical Basics

Data are all around us. Researchers collect data on the effectiveness of a medication for lowering cholesterol. Pollsters report on the percentage of Americans who support gun control. Economists report on the average salary of college graduates. There are many other areas where data are collected. In order to be able to understand data and how to summarize it, we need to understand statistics.

Suppose you want to know the average net worth of a current U.S. Senator. There are 100 Senators, so it is not that hard to collect all 100 values, and then summarize the data. If instead you want to find the average net worth of all current Senators and Representatives in the U.S. Congress, there are only 435 members of Congress. So even though it will be a little more work, it is not that difficult to find the average net worth of all members. Now suppose you want to find the average net worth of everyone in the United States. This would be very difficult, if not impossible. It would take a great deal of time and money to collect the information in a timely manner before all of the values have changed. So instead of getting the net worth of every American, we have to figure out an easier way to find this information. The net worth is what you want to measure, and is called a variable. The net worth of every American is called the population. What we need to do is collect a smaller part of the population, called a sample. In order to see how this works, let's formalize the definitions.

Variable: Any characteristic that is measured from an object or individual.

Population: A set of measurements or observations from all objects under study
Sample: A set of measurements or observations from some objects under study (a subset of a population)

Example 14.2.1: Stating Populations and Samples

Determine the population and sample for each situation.

1. A researcher wants to determine the length of the lifecycle of a bark beetle. In order to do this, he breeds 1000 bark beetles and measures the length of time from birth to death for each bark beetle.

Population: The set of lengths of lifecycle of all bark beetles

Sample: The set of lengths of lifecycle of 1000 bark beetles

2. The National Rifle Association wants to know what percent of Americans support the right to bear arms. They ask 2500 Americans whether they support the right to bear arms.

Population: The set of responses from all Americans to the question, "Do you support the right to bear arms?"

Sample: The set of responses from 2500 Americans to the question, "Do you support the right to bear arms?"

3. The Pew Research Center asked 1000 mothers in the U.S. what their highest attained education level was.

Population: The set of highest education levels of all mothers in the U.S.

Sample: The set of highest education level of 1000 mothers in the U.S.

It is very important that you understand what you are trying to measure before you actually measure it. Also, please note that the population is a set of measurements or observations, and not a set of people. If you say the population is all Americans, then you have only given part of the story. More important is what you are measuring from all Americans. The question is, do you want to measure their race, their eye color, their income, their education level, the number of children they have, or other variables? Therefore, it is very important to state what you measured or observed, and from whom or what the measurements or observations were taken. Once you know what you want to measure or observe, and the source from which you want to take measurements or observations, you need to collect the data.

A **data set** is a collection of values called data points or data values. **N** represents the number of data points in a population, while **n** represents the number of data points in a sample. A data value that is much higher or lower than all of the other data values is called an **outlier**. Sometimes outliers are just unusual data values that are very interesting and should be studied further, and sometimes they are mistakes. You will need to figure out which is which.

In order to collect the data, we have to understand the types of variables we can collect. There are actually two different types of variables. One is called qualitative and the other is called quantitative.

Qualitative (Categorical) Variable: A variable that represents a characteristic. Qualitative variables are not inherently numbers, and so they cannot be added, multiplied, or averaged, but they can be represented graphically with graphs such as a bar graph.

Examples: gender, hair color, race, nationality, religion, course grade, year in college, etc.

Quantitative (Numerical) Variable: A variable that represents a measurable quantity. Quantitative variables are inherently numbers, and so can they be added, multiplied, averaged, and displayed graphically.

Examples: Height, weight, number of cats owned, score of a football game, etc.

Quantitative variables can be further subdivided into other categories – continuous and discrete.

Continuous Variable: A variable that can take on an uncountable number of values in a range. In other words, the variable can be any number in a range of values. Continuous variables are usually things that are measured.

Examples: Height, weight, foot size, time to take a test, length, etc.

Discrete Variable: A variable that can take on only specific values in a range. Discrete variables are usually things that you count.

Examples: IQ, shoe size, family size, number of cats owned, score in a football game, etc.

Example 14.2.2: Determining Variable Types

Determine whether each variable is quantitative or qualitative. If it is quantitative, then also determine if it is continuous or discrete.

1. Length of race

Quantitative and continuous, since this variable is a number and can take on any value in an interval.

2. Opinion of a person about the President

Qualitative, since this variable is not a number.

3. House color in a neighborhood

Qualitative, since this variable is not a number.

4. Number of houses that are in foreclosure in a state

Quantitative and discrete, since this variable is a number but can only be certain values in an interval.

5. Weight of a baby at birth

Quantitative and continuous, since this variable is a number and can take on any value in an interval.

6. Highest education level of a mother

Qualitative, since the variable is not a number.

This page titled 14.2: Statistical Basics is shared under a CC BY-SA license and was authored, remixed, and/or curated by Maxie Inigo, Jennifer Jameson, Kathryn Kozak, Maya Lanzetta, & Kim Sonier.

- **1.1: Statistical Basics** by Maxie Inigo, Jennifer Jameson, Kathryn Kozak, Maya Lanzetta, & Kim Sonier is licensed CC BY-SA 4.0. Original source: https://www.coconino.edu/open-source-textbooks#college-mathematics-for-everyday-life-by-inigo-jameson-kozak-lanzetta-and-sonier.

14.3: Random Sampling

Now that you know that you have to take samples in order to gather data, the next question is how best to gather a sample? There are many ways to take samples. Not all of them will result in a representative sample. Also, just because a sample is large does not mean it is a good sample. As an example, you can take a sample involving one million people to find out if they feel there should be more gun control, but if you only ask members of the National Rifle Association (NRA) or the Coalition to Stop Gun Violence, then you may get biased results. You need to make sure that you ask a cross-section of individuals. Let's look at the types of samples that can be taken. Do realize that no sample is perfect, and may not result in a representation of the population.

Census: An attempt to gather measurements or observations from all of the objects in the entire population.

A true census is very difficult to do in many cases. However, for certain populations, like the net worth of the members of the U.S. Senate, it may be relatively easy to perform a census. We should be able to find out the net worth of each and every member of the Senate since there are only 100 members. But, when our government tries to conduct the national census every 10 years, you can believe that it is impossible for them to gather data on each and every American.

The best way to find a sample that is representative of the population is to use a random sample. There are several different types of random sampling. Though it depends on the task at hand, the best method is often simple random sampling which occurs when you randomly choose a subset from the entire population.

Simple Random Sample: Every sample of size n has the same chance of being chosen, and every individual in the population has the same chance of being in the sample.

An example of a simple random sample is to put all of the names of the students in your class into a hat, and then randomly select five names out of the hat.

Stratified Sampling: This is a method of sampling that divides a population into different groups, called strata, and then takes random samples inside each strata.

An example where stratified sampling is appropriate is if a university wants to find out how much time their students spend studying each week; but they also want to know if different majors spend more time studying than others. They could divide the student body into the different majors (strata), and then randomly pick a number of people in each major to ask them how much time they spend studying. The number of people asked in each major (strata) does not have to be the same.

Systematic Sampling: This method is where you pick every kth individual, where k is some whole number. This is used often in quality control on assembly lines.

For example, a car manufacturer needs to make sure that the cars coming off the assembly line are free of defects. They do not want to test every car, so they test every 100th car. This way they can periodically see if there is a problem in the manufacturing process. This makes for an easier method to keep track of testing and is still a random sample.

Cluster Sampling: This method is like stratified sampling, but instead of dividing the individuals into strata, and then randomly picking individuals from each strata, a cluster sample separates the individuals into groups, randomly selects which groups they will use, and then takes a census of every individual in the chosen groups.

Cluster sampling is very useful in geographic studies such as the opinions of people in a state or measuring the diameter at breast height of trees in a national forest. In both situations, a cluster sample reduces the traveling distances that occur in a simple random sample. For example, suppose that the Gallup Poll needs to perform a public opinion poll of all registered voters in Colorado. In order to select a good sample using simple random sampling, the Gallup Poll would have to have all the names of all the registered voters in Colorado, and then randomly select a subset of these names. This may be very difficult to do. So, they will use a cluster sample instead. Start by dividing the state of Colorado up into categories or groups geographically. Randomly select some of these groups. Now ask all registered voters in each of the chosen groups. This makes the job of the pollsters much easier, because they will not have to travel over every inch of the state to get their sample but it is still a random sample.

Quota Sampling: This is when the researchers deliberately try to form a good sample by *creating* a cross-section of the population under study.

For an example, suppose that the population under study is the political affiliations of all the people in a small town. Now, suppose that the residents of the town are 70% Caucasian, 25% African American, and 5% Native American. Further, the residents of the town are 51% female and 49% male. Also, we know information about the religious affiliations of the townspeople. The residents of the town are 55% Protestant, 25% Catholic, 10% Jewish, and 10% Muslim. Now, if a researcher is going to poll the people of this town about their political affiliation, the researcher should gather a sample that is representative of the entire population. If the researcher uses quota sampling, then the researcher would try to artificially create a cross-section of the town by insisting that his sample should be 70% Caucasian, 25% African American, and 5% Native American. Also, the researcher would want his sample to be 51% female and 49% male. Also, the researcher would want his sample to be 55% Protestant, 25% Catholic, 10% Jewish, and 10% Muslim. This sounds like an admirable attempt to create a good sample, but this method has major problems with selection bias.

The main concern here is when does the researcher stop profiling the people that he will survey? So far, the researcher has cross-sectioned the residents of the town by race, gender, and religion, but are those the only differences between individuals? What about socioeconomic status, age, education, involvement in the community, etc.? These are all influences on the political affiliation of individuals. Thus, the problem with quota sampling is that to do it right, you have to take into account all the differences among the people in the town. If you cross-section the town down to every possible difference among people, you end up with single individuals, so you would have to survey the whole town to get an accurate result. The whole point of creating a sample is so that you do not have to survey the entire population, so what is the point of quota sampling?

Note: The Gallup Poll did use quota sampling in the past, but does not use it anymore.

Convenience Sampling

As the name of this sampling technique implies, the basis of convenience sampling is to use whatever method is easy and convenient for the investigator. This type of sampling technique creates a situation where a random sample is not achieved. Therefore, the sample will be biased since the sample is not representative of the entire population.

For example, if you stand outside the Democratic National Convention in order to survey people exiting the convention about their political views. This may be a convenient way to gather data, but the sample will not be representative of the entire population.

Of all of the sampling types, a random sample is the best type. Sometimes, it may be difficult to collect a perfect random sample since getting a list of all of the individuals to randomly choose from may be hard to do.

Example 14.3.1: Which Type of Sample?

Determine if the sample type is simple random sample, stratified sample, systematic sample, cluster sample, quota sample, or convenience sample.

Solution

1. A researcher wants to determine the different species of trees that are in the Coconino National Forest. She divides the forest using a grid system. She then randomly picks 20 different sections and records the species of every tree in each of the chosen sections.

This is a cluster sample, since she randomly selected some of the groups, and all individuals in the chosen groups were surveyed.

2. A pollster stands in front of an organic foods grocery store and asks people leaving the store how concerned they are about pesticides in their food.

This is a convenience sample, since the person is just standing out in front of one store. Most likely the people leaving an organic food grocery store are concerned about pesticides in their food, so the sample would be biased.

3. The Pew Research Center wants to determine the education level of mothers. They randomly ask mothers to say if they had some high school, graduated high school, some college, graduated from college, or advance degree.

This is a simple random sample, since the individuals were picked randomly.

4. Penn State wants to determine the salaries of their graduates in the majors of agricultural sciences, business, engineering, and education. They randomly ask 50 graduates of agricultural sciences, 100 graduates of business, 200 graduates of engineering, and 75 graduates of education what their salaries are.

This is a stratified sample, since all groups were used, and then random samples were taken inside each group.

5. In order for the Ford Motor Company to ensure quality of their cars, they test every 130th car coming off the assembly line of their Ohio Assembly Plant in Avon Lake, OH.

This is a systematic sample since they picked every 130th car.

6. A town council wants to know the opinion of their residents on a new regional plan. The town is 45% Caucasian, 25% African American, 20% Asian, and 10% Native American. It also is 55% Christian, 25% Jewish, 12% Islamic, and 8% Atheist. In addition, 8% of the town did not graduate from high school, 12% have graduated from high school but never went to college, 16% have had some college, 45% have obtained bachelor's degree, and 19% have obtained a post-graduate degree. So the town council decides that the sample of residents will be taken so that it mirrors these breakdowns.

This is a quota sample, since they tried to pick people who fit into these subcategories.

This page titled 14.3: Random Sampling is shared under a CC BY-SA license and was authored, remixed, and/or curated by Maxie Inigo, Jennifer Jameson, Kathryn Kozak, Maya Lanzetta, & Kim Sonier.

- **1.2: Random Sampling** by Maxie Inigo, Jennifer Jameson, Kathryn Kozak, Maya Lanzetta, & Kim Sonier is licensed CC BY-SA 4.0. Original source: https://www.coconino.edu/open-source-textbooks#college-mathematics-for-everyday-life-by-inigo-jameson-kozak-lanzetta-and-sonier.

14.4: Clinical Studies

Now you know how to collect a sample, next you need to learn how to conduct a study. We will discuss the basics of studies, both observational studies and experiments.

Observational Study: This is where data is collected from just observing what is happening. There is no treatment or activity being controlled in any way. Observational studies are commonly conducted using surveys, though you can also collect data by just watching what is happening such as observing the types of trees in a forest.

Survey: Surveys are used for gathering data to create a sample. There are many different kinds of surveys, but overall, a survey is a method used to ask people questions when interested in the responses. Examples of surveys are Internet and T.V. surveys, customer satisfaction surveys at stores or restaurants, new product surveys, phone surveys, and mail surveys. The majority of surveys are some type of public opinion poll.

Experiment: This is an activity where the researcher controls some aspect of the study and then records what happens. An example of this is giving a plant a new fertilizer, and then watching what happens to the plant. Another example is giving a cancer patient a new medication, and monitoring whether the medication stops the cancer from growing. There are many ways to do an experiment, but a clinical study is one of the more popular ways, so we will look at the aspects of this.

Clinical Study: This is a method of collecting data for a sample and then comparing that to data collected for another sample where one sample has been given some sort of treatment and the other sample has not been given that treatment (control). *Note: There are occasions when you can have two treatments, and no control. In this case you are trying to determine which treatment is better.*

Example 14.4.1: Clinical Study Examples

Here are examples of clinical studies.

1. A researcher may want to study whether or not smoking increases a person's chances of heart disease.
2. A researcher may want to study whether a new antidepressant drug will work better than an old antidepressant drug.
3. A researcher may want to study whether taking folic acid before pregnancy will decrease the risk of birth defects.

Clinical Study Terminology:

Treatment Group: This is the group of individuals who are given some sort of treatment. The word treatment here does not necessarily mean medical treatment. The treatment is the cause, which may produce an effect that the researcher is interested in.

Control Group: This is the group of individuals who are not given the treatment.

Sometimes, they may be given some old treatment, or sometimes they will not be given anything at all. Other times, they may be given a placebo (see below).

✔ Example 14.4.2: Treatment/Control Group Examples

Determine the treatment group, control group, treatment, and control for each clinical study in Example 14.4.1.

Solution

1. A researcher may want to study whether or not smoking increases a person's chances of heart disease.

The treatment group is the people in the study who smoke and the treatment is smoking. The control group is the people in the study who do not smoke and the control is not smoking.

2. A researcher may want to study whether a new antidepressant drug will work better than an old antidepressant drug.

The treatment group is the people in the study who take the new antidepressant drug and the treatment is taking the new antidepressant drug. The control group is the people in the study who take the old antidepressant drug and the control is taking the old antidepressant drug. *Note: In this case the control group is given some treatment since you should not give a person with depression a non-treatment.*

3. A researcher may want to study whether taking folic acid before pregnancy will decrease the risk of birth defects.

The treatment group is the women who take folic acid before pregnancy and the treatment is taking folic acid. The control group is the women who do not take folic acid before pregnancy and the control is not taking the folic acid.

Note: In this case, you may choose to do an observational study of women who did or did not take folic acid during pregnancy so that you are not inducing women to avoid folic acid during pregnancy which could be harmful to their baby.

Confounding Variables: These are other possible causes that may produce the effect of interest rather than the treatment under study. Researchers minimize the effect of confounding variables by comparing the results from the treatment group versus the control group.

Controlled Study: Any clinical study where the researchers compare the results of a treatment group versus a control group.

Placebo: A placebo is sometimes used on the control group in a study to mimic the treatment that the treatment group is receiving. The idea is that if a placebo is used, then the people in the control group and in the treatment group will all think that they are receiving the treatment. However, the control group is merely receiving something that looks like the treatment, but should have no effect on the outcome. An example of a placebo could be a sugar pill if the treatment is a drug in pill form.

Example 14.4.3: Placebo Examples

For each situation in Example **14.4.1**, identify if a placebo is necessary to use.

1. A researcher may want to study whether or not smoking increases a person's chances of heart disease.

In this example, it is impossible to use a placebo. The treatment group is comprised of people who smoke and the control group is comprised of people who do not smoke. There is no way to get the control group to think that they are smoking as well as the treatment group.

2. A researcher may want to study whether a new antidepressant drug will work better than an old antidepressant drug.

In this example, a placebo is not needed since we are comparing the results of two different antidepressant drugs.

3. A researcher may want to study whether taking folic acid before pregnancy will decrease the risk of birth defects.

In this example, the control group could be given a sugar pill instead of folic acid. However, they may think that they are taking folic acid and so the psychological effect on a person's health can be being measured. This way, when we compare the results of taking folic acid versus taking a sugar pill, we can see if there were any dramatic differences in the results.

Blind Study: Usually, when a placebo is used in a study, the people in the study will not know if they received the treatment or the placebo until the study is completed. In other words, the people in the study do not know if they are in the treatment group or in the control group. This type of study is called a blind study. *Note: When researchers use a placebo in a blind study, the people in the study are told ahead of time that they may be getting the actual treatment, or they may be getting the placebo.*

Double-Blind Study: Sometimes when researchers are conducting a very extensive study using many healthcare workers, the researchers will not tell the people in the study or the healthcare workers which patients will receive the treatment and which patients will receive the placebo. In other words, the healthcare workers who are administering the treatment or placebo to the people in the study do not know which people are in the treatment group and which people are in the control group. This type of study is called a double-blind study.

Randomized Controlled Study: Any clinical study in which the treatment group and the control group are selected randomly from the population.

Whether you are doing an observational study or an experiment, you need to figure out what to do with the data. You will have many data values that you collected, and it sometimes helps to calculate numbers from these data values. Whether you are talking about the population or the sample, determines what we call these numbers.

Parameter: A numerical value calculated from a population

Statistic: A numerical value calculated from a sample, and used to estimate the parameter

Some examples of parameters that can be estimated from statistics are the percentage of people who strongly agree to a question and mean net worth of all Americans. The statistic would be the percentage of people asked who strongly agree to a question, and the mean net worth of a certain number of Americans.

Notation for Parameter and Statistics:

Parameters are usually denoted with Greek letters. This is not to make you learn a new alphabet. It is because there just are not enough letters in our alphabet. Also, if you see a letter you do not know, then you know that the letter represents a parameter. Examples of letters that are used are μ (mu), σ (sigma), ρ (rho), and p (yes this is our letter because there is not a good choice in the Greek alphabet).

Statistics are usually denoted with our alphabet, and in some cases we try to use a letter that would be equivalent to the Greek letter. Examples of letter that are used are $\bar{x}$ (x-bar), s, r, and $\hat{p}$ (p-hat, since we already used p for the parameter).

In addition, N is used to denote the size of the population and n is used to denote the size of the sample.

Sampling Error: This is the difference between a parameter and a statistic. There will always be some error between the two since a statistic is an estimate of a parameter. Sampling error is attributed to chance error and sample bias.
Chance Error: The error inherent in taking information from a sample instead of from the whole population. This comes from the fact that two different samples from the same population will likely give two different statistics.

Sample Bias: The error from using a sample that does not represent the population. To avoid this, use some sort of random sample.
Sampling Rate: The fraction of the total population that is in the sample. This can be denoted by n/N.

This page titled 14.4: Clinical Studies is shared under a CC BY-SA license and was authored, remixed, and/or curated by Maxie Inigo, Jennifer Jameson, Kathryn Kozak, Maya Lanzetta, & Kim Sonier.

- **1.3: Clinical Studies** by Maxie Inigo, Jennifer Jameson, Kathryn Kozak, Maya Lanzetta, & Kim Sonier is licensed CC BY-SA 4.0. Original source: https://www.coconino.edu/open-source-textbooks#college-mathematics-for-everyday-life-by-inigo-jameson-kozak-lanzetta-and-sonier.

14.5: Types of clinical interventions

Interventions can be classified into two broad categories: (1) preventive interventions are those that prevent disease from occurring and thus reduce the incidence (new cases) of disease, and (2) therapeutic interventions are those that treat, mitigate, or postpone the effects of disease, once it is under way, and thus reduce the case fatality rate or reduce the disability or morbidity associated with a disease. Some interventions may have both effects.

2.1 Preventive interventions

2.1.1 Vaccines

Vaccines are administered to individuals, usually before they have encountered the infectious agent against which the vaccine is targeted, in order to protect them when they are naturally exposed to the agent. Many are among the most cost-effective interventions, because, after a single dose or a series of doses of the vaccine, an individual may acquire long-term protection against the agent. They work by inducing a variety of immune mechanisms, through the humoral and/or cellular immune systems. The immunological responses and associated immunological memory induced by vaccination confer protection from later infections, though a booster vaccination may be necessary if the interval between the original vaccination and exposure to the agent is long. Most vaccines have to be administered before the infectious agent is encountered naturally, and thus field trials of such vaccines will involve the enrolment of healthy individuals and often involve infants or very young children—though the vaccine may be given at a later age if the age of natural infection is at later ages, for example, for most sexually transmitted infections (STIs), or if a new infectious agent, to which no one has been previously exposed, enters a community such as a new strain of influenza.

Not all vaccines are targeted at persons without previous exposure to the infectious agent. For example, there is substantial research to develop vaccines against parasitic diseases. The mode of action of some of these vaccines is to prevent parasitic proliferation within the host after invasion (and hence curtailment of disease), and some vaccines against vector-borne diseases are even targeted to prevent replication of the forms of the infection in the vector, so that onward transmission to humans is prevented.

For infectious diseases that affect both high-income countries (HICs) and LMICs, the first trials of new vaccines are usually conducted in HICs. This is because currently most new vaccines are developed and produced in HICs (though this situation is changing), and it is generally accepted that at least early clinical studies should be conducted in the country of vaccine manufacture. However, the results of trials in HICs may not be directly applicable to LMICs for a variety of reasons such as differing prevalences of other infections or of nutritional deficiencies, which might interfere with the mode of action of the vaccine. Thus, there will often be a need for further trials of the vaccine in LMICs, even if efficacy has been established in HICs. In addition, there has been increased focus in recent years on the development of vaccines against infectious agents that only, or almost only, occur in LMICs, such as malaria or visceral leishmaniasis, or where the overwhelming disease burden is in such countries, such as tuberculosis (TB) or HIV infection. For vaccines against these agents, the first major field trials to assess efficacy are likely to be conducted in LMICs.

2.1.2 Nutritional interventions

Food and nutrition are major determinants of human health and disease. Particularly in low-income countries and deprived populations in middle-income countries, under-nutrition remains a major cause of disease. Severe malnutrition, such as kwashiorkor or marasmus, is life-threatening, but milder forms of malnutrition are major risk factors that adversely influence the susceptibility to, and the outcome of, many infectious and other diseases, as well as cognitive development. In addition to calorie and protein deficiencies, specific deficiencies in micronutrients, such as iron, folate, zinc, iodine, and vitamin A, may be important determinants of severe diseases. Trials to address these problems may involve the regular provision of high-protein/calorie diets or supplementation to individuals with specific micronutrients, involving repeated visits to the same persons over several years, the frequency of administration depending on the nature of the supplement(s). Other trials, often with the intervention being applied at a community level, may involve food fortification (for example, iron, iodine, vitamin D) and experiments to change agricultural practices or eating or food preparation habits to increase the intake of particular micronutrients.

2.1.3 Maternal and neonatal interventions

A mother's health and well-being during pregnancy and around the time of delivery, including access to appropriate care, are critical determinants of maternal mortality and neonatal and child health in the early years of life, and possibly for much longer. Preventive interventions before or during pregnancy include family planning, treatment of infections, such as syphilis and malaria, good nutrition, including micronutrients, good antenatal monitoring and care, and access to skilled care at the time of delivery and post-partum. Trials of maternal interventions may involve both community-based studies, with the early identification of pregnancies and the instigation of preventive interventions to avoid pregnancy complications, or may be hospital- or health centre-based, directed at improving the performance of the health system in caring for women during and after pregnancy and at the time of birth.

Interventions directed to the neonate are also important, such as exclusive breastfeeding and care practices, such as 'kangaroo mother care', a method of care of preterm infants, involving infants being carried, usually by the mother, with skin-to-skin contact.

2.1.4 Education and behaviour change

Some interventions directed at preventing disease are based solely upon changing human behaviour (for example, anti-smoking campaigns or campaigns to promote breastfeeding). Nearly all health interventions must have an associated educational component for their effective deployment, but the extent of educational effort required ranges from the provision of simple information (for example, when and where a clinic for immunization will be held) to efforts at increasing understanding (for example, of the importance of male circumcision for the prevention of HIV) and to attempts to change lifestyles (for example, diet or sexual habits). Education to increase knowledge and impart new skills may be necessary but is rarely sufficient to induce behaviour change. Individuals must also have the capacity, willingness, and motivation to act on the knowledge and to use the skills. The design and implementation of an educational intervention, and other 'complex' interventions (Craig et al., 2008), will usually need to be researched through careful investigations in the community, using the kinds of methods discussed in Chapters 9 and 15.

Examples of educational components of disease control programmes include:

- educating children or mothers about the causes of the disease, such as diarrhoea, and how to prevent it
- promoting adherence to long-term treatment such as for HIV infection or TB
- developing effective participation in programmes that:
  - need broad coverage to maximize the effects of immunization or drug distribution
  - require people to recognize disease symptoms for early treatment
  - necessitate active co-operation in home improvements or insecticide programmes
  - involve direct action and responsibility in deploying vector, or intermediate host, traps
  - need community efforts for environmental improvements such as developing and maintaining improved water supplies or better disposal methods for faeces.

Organizing trials of behaviour change interventions are among the most challenging, and there are few examples illustrating the design of replicable interventions that achieve lasting behavioural change in the context of a trial. For example, changing tobacco smoking behaviour at a population level required decades of concerted, multifaceted campaigns. However, attempts to reduce diarrhoeal diseases and respiratory infections through the promotion of hand-washing with soap have produced encouraging results.

2.1.5 Environmental alterations

Alterations to the environment directed at reducing the transmission of infections are central to the control of many infectious diseases, particularly those that are transmitted through water, such as cholera, or through the faecal–oral route such as many gastrointestinal infections. Environmental interventions to reduce human faecal and urine contamination include latrine construction, provision of sewage systems, clean water supplies, and protected food storage. Other environmental interventions tackle indoor or outdoor air pollution or involve the disposal of contaminants such as pesticides or heavy metals. Many of these interventions require substantial educational efforts and lifestyle changes. They are also interventions that typically have to be applied to whole communities, rather than to individuals in a community, so that, in trials, the unit of randomization is the community or, in some instances, the household.

2.1.6 Vector and intermediate host control

Some major communicable diseases in developing countries depend on vector and intermediate hosts for their transmission. For different infections, the vectors include mosquitoes, tsetse flies, triatomine bugs, sandflies, ticks, and snails. There are a wide variety of control measures to reduce transmission of these infections through attacking the vectors or the reservoirs of infection. Most interventions require a good understanding of the vector or intermediate host, its life cycle, and the environmental conditions that it requires to propagate infections. Control measures may include the application of insecticides or larvicides, new or improved selective biological agents against disease vectors, engineering techniques for reducing vector habitats, community involvement in eliminating vector breeding sites and in deploying traps, housing and screening improvement for reducing human–vector contact, and strategies involving combinations of methods with, for example, the objective of reducing or delaying insecticide resistance. For many of these methods, intermediate process indicators, such as reduction in vector density, can be used for the assessment of impact, but it is often also necessary to determine the impact of the measures on the health status of the population. For example, for malaria, many different approaches to vector control have been used, based upon attacking the mosquito in various stages of its life cycle. These include control of breeding sites to reduce vector density by drainage and waterway engineering and application of specific larvicides and biological agents; the use of mosquito netting, screens, and repellents for personal protection from bites; aerosol distribution of insecticides to reduce adult mosquito densities; and different approaches to killing adult mosquitoes, through either spraying residual insecticides, such as with dichlorodiphenyltrichloroethane (DDT), on the internal walls of houses where mosquitoes rest after a blood meal or through the use of insecticide-treated bed-nets (ITNs) that kill and/or repel mosquitoes seeking a blood meal. These different approaches require quite different study designs. Residual insecticide on the walls of houses offers relatively little direct protection to those in the treated household, as the mosquitoes take up the insecticide while resting *after* a blood meal. The protection is to those in other households whom these mosquitoes would have bitten for their next blood meal. To reduce transmission in high transmission areas, virtually all households in the neighbourhood must be sprayed. The higher the intensity of transmission, the more difficult it is to achieve sufficient coverage. The use of ITNs, developed as an intervention against malaria over the last two decades, leads to reductions in transmission, clinical disease, and overall childhood mortality. Trials of these kinds of intervention often involve communities, rather than individuals, as the unit of randomization. These trials are especially challenging to design, because some vectors, such as mosquitoes, may have a flight range that may lead to the ‘contamination’ of intervention communities, with vectors coming in from outside of the community.

2.1.7 Drugs for the prevention of disease

Drugs or other interventions may be used for the prevention of infection (prophylaxis) or disease consequent on infection. An example of the former would be isoniazid prophylaxis to HIV-infected individuals to reduce their risk of TB, and of the latter, the treatment of HIV-infected individuals with antiretroviral drugs to slow the progression of their disease. Sometimes, the use of drugs for prophylaxis or to reduce disease progression does not involve individual diagnosis, but community or group diagnosis is needed to identify groups that should receive the treatment. For example, mass administration of anti-helminthic treatment to schoolchildren is sometimes administered in this way. Whether requiring specific diagnosis or not, therapeutic or preventive agents are usually taken on an individual basis, though sometimes agents can be distributed to everyone in a community through the water supply (for example, fluoride against dental caries) or in food (for example, historically, diethylcarbamazine for filariasis and chloroquine for malaria in medicated salt). Mass treatment of school-age children in areas highly endemic for the infection with an anti-schistosomal drug every year or two may be sufficient to virtually eliminate serious disease consequences of infection with *Schistosoma mansoni*.

Prophylaxis may be aimed at preventing or limiting infection, particularly in those at high risk for a limited period of time (for example, anti-malarials taken by those who are temporarily visiting malaria-endemic areas). The value of such an approach is limited by the duration of action of the agent (which determines the frequency with which it must be taken), by adverse reactions, and sometimes by the role of the intervention in stimulating the development of drug-resistant organisms. For some purposes, prophylaxis may be used by permanent residents of endemic areas (for example, anti-malarials in pregnancy).

Drugs also may be used prophylactically for treatment of preclinical infection (for example, during the incubation period before the onset of symptoms, as for the *gambiense* type of trypanosomiasis) or for treatment of subclinical infection (for example, ivermectin against onchocerciasis, and praziquantel against schistosomiasis).

Strategies for the use of such interventions include the mass treatment of entire populations or the targeted treatment of identifiable subgroups (such as school-age children) in areas where the infection is highly prevalent. Generally, such treatment is applied for

the benefit of the individuals treated, but the objective may also be to reduce the transmission of the agent in the community more generally. When the prevalence is very high and the treatment is cheap, treating all those in a defined population may be more cost-effective than screening the whole population and then treating only those found infected.

2.1.8 Injury prevention

Injuries are major causes of death and disability, especially in LMICs. They disproportionately affect the young and have a large economic impact on society. For children and young people, road traffic accidents, drowning, fires, poisoning, interpersonal violence, and war are leading global causes of serious injuries, but often these are not considered 'health problems' and are not sufficiently integrated into public health thinking. Yet there are many potential interventions that might lead to reductions in deaths and disabilities from injuries, such as traffic calming or infrastructural changes to separate pedestrians from fast-moving vehicles to reduce motor vehicle injuries, and improving the security of water sources to reduce drowning accidents; there is great need for more trials of interventions directed at reducing injuries.

2.2 Therapeutic interventions

2.2.1 Treatment of infectious diseases

The mechanism of action of a drug used for disease control will influence the design of field trials to evaluate its impact. Most drugs employed against infectious disease are used to kill or inhibit the replication or spread of the pathogen in the host. Strategies for disease control that use such agents may involve case detection (which requires an appropriate case definition and a diagnostic method), followed by treatment that is designed to reduce morbidity and mortality. Often, the public health success of this approach depends critically upon case finding, and, for diseases such as TB and leprosy, it depends also on case holding, i.e. being able to follow and treat each patient at regular intervals over sufficient time to eliminate the agent from the individual. Case finding and treatment may also reduce transmission of an agent if cases are the main reservoirs of infection, if case detection methods locate a high proportion of prevalent cases, and if the treatment is sufficiently effective.

2.2.2 Surgical and radiation treatment

RCTs of surgical and radiation treatments are usually done as clinical trials; field trials of these interventions are relatively uncommon. However, procedures, such as cataract extraction or simple inguinal hernia repair, are examples of where field trials have been usefully undertaken. In general, the only distinctive feature that may set these apart, in terms of study design, from other field trials is the issue of 'blinding' (see Chapter 11, Section 4). For some forms of surgery, 'sham' operations have been used in clinical studies and perhaps could be considered in field trials. In general, however, randomized trials of these procedures will have to be conducted without blinding.

2.2.3 Diagnostics to guide therapy

The efficient treatment of most diseases requires first that they be accurately diagnosed. Often the diagnosis is made on the basis of clinical symptoms and signs, but the imprecision of this method for many conditions is increasingly recognized. There is an urgent need for new, or improved, sensitive and specific diagnostic tests for many infectious and chronic diseases, that are both simple to use and cheap. For example, intervention strategies that depend upon case finding and treatment usually require suitable diagnostic tests. Specific studies may be necessary to measure the specificity, sensitivity, and predictive values of different diagnostic tests, as these properties will impact on the likely effectiveness of a case finding and treatment intervention. For example, the development and widespread introduction of rapid diagnostic tests for malaria, to replace microscopy or the presumptive treatment of fever, has been an important innovation in malaria control and has also focused attention on the need for improved diagnostic methods and appropriate treatment of non-malarial fevers.

Field trials to evaluate the performance characteristics of diagnostics are not discussed specifically in this book, other than in the context that they may be incorporated as part of an intervention strategy to improve the control of a specific disease. The design of studies to evaluate the properties of diagnostics has been discussed elsewhere (Peeling et al., 2010).

2.2.4 Control of chronic diseases

Chronic conditions may have an infectious aetiology (for example, HIV, TB) or may have environmental or other causes (for example, cardiovascular diseases and many cancers). Many chronic diseases, once diagnosed, may not be curable, but they can be

controlled by a combination of education/behaviour change interventions, plus regular, often daily, use of pharmaceuticals. The nature of the clinical care required is often more complicated than required for acute conditions, such as diarrhoea and pneumonia, which, once diagnosed, usually require a single course of treatment. Interventions for chronic disease often must include screening of communities to identify cases; assessment of each case for the stage of the disease and possible attendant complications that are likely to require a variety of laboratory tests; and developing a long-term treatment and assessment plan. The treatment of such conditions often requires long-term monitoring, with a dependence on reliable laboratory results and a system to track the clinical and laboratory findings within a single individual over time. Trials of such interventions must often be conducted over several years, or even decades, to completely assess treatment efficacy.

2.3 Other forms of intervention

2.3.1 Legislation, legal action, taxation, and subsidies

Enforcement of anti-pollution laws, food labelling, and legal restrictions have an important role to play in public health. Behaviour may be strongly influenced by legal restrictions, and increasing prices through taxation have been shown to be effective in reducing tobacco and alcohol consumption, for example. However, it is difficult to design randomized trials of such interventions, because the interventions usually have to be implemented at the national level, making it very difficult to identify a suitable control group.

There has been increasing interest recently in providing various types of subsidies to individuals to change their health-related behaviour (often known as conditional cash transfers). Examples include incentives for children to remain in school, or to health care providers to provide services of at least a certain minimum quality (performance incentives). Some of these interventions have been evaluated through RCTs, and there is further scope for using such approaches.

2.3.2 Health systems interventions

Increasing recognition of the importance of interventions that operate at health systems level, such as policy implementation, financing, educational reform, and strengthening of leadership, management, and governance, has led to a variety of health sector training programmes, organization changes, decentralization and devolution, and various incentives and personnel policies. Most of these efforts have been introduced on a system-wide basis, with little thought about the value of rigorous assessment. But, with adequate planning, rigorous evaluation of these kinds of interventions should be possible through randomized trials, especially by making use of the ‘stepped wedge’ approach of a phased introduction of measures in different communities over a period of time (Brown and Lilford, 2006). Many health systems research studies may be considered as implementation research, and most could be considered as complex interventions, as discussed in Sections 2.3.3 and 2.3.4.

2.3.3 Implementation research

Within the context of field trials, implementation research does not aim to develop new interventions but focuses on optimizing the delivery of existing interventions that have previously been shown to be efficacious when implemented well. Implementation research explores the challenges of how best to implement research findings in the real world and how to contextualize interventions for specific settings. Hence, an example of an implementation research trial was one where a comparison was made of the costs and effectiveness of health workers delivering antiretroviral therapy to patients who attend a central clinic or hospital, compared with lay workers delivering the antiretrovirals to patients in their homes and only referring them to the clinic if they reported problems on a screening questionnaire (Jaffar et al., 2009).

A general reference on implementation research is Werner (2004).

2.3.4 Complex interventions

The design of a trial to evaluate the efficacy of a new vaccine or drug is relatively straightforward, in the sense that there are many past examples of such evaluations to draw upon when planning a new study. However, the evaluation of some interventions, such as the deployment of a new procedure in the health service or in public health practice, may involve consideration of several interacting components, including, for example, educational components and behavioural change. Such interventions pose special problems for evaluation, and these kinds of intervention have been called ‘complex’. Many of the extra problems relate to the difficulty of standardizing the design and delivery of the interventions, their sensitivity to features of the local context, the organizational and logistical difficulty of applying experimental methods to service or policy change, and the length and complexity of the causal chains linking intervention with outcome.

In 2000, the UK Medical Research Council published a *Framework for development and evaluation of RCTs for complex interventions to improve health* to help researchers and research funders to recognize and adopt appropriate methods. These guidelines were updated and revised subsequently and can be downloaded from the Internet (<http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance>).

Box 2.1 is reproduced from the guidelines and summarizes the steps in developing and evaluating trials involving complex interventions.

Box 2.1 The development–evaluation–implementation process

Developing, piloting, evaluating, reporting, and implementing a complex intervention can be a lengthy process. All of the stages are important, and too strong a focus on the main evaluation, to the neglect of adequate development and piloting work, or proper consideration of the practical issues of implementation, will result in weaker interventions that are harder to evaluate, less likely to be implemented, and less likely to be worth implementing.'

Developing an intervention

Questions to ask yourself include: Are you clear about what you are trying to do—what outcome you are aiming for, and how you will bring about change? Does your intervention have a coherent theoretical basis? Have you used this theory systematically to develop the intervention? Can you describe the intervention fully, so that it can be implemented properly for the purposes of your evaluation and replicated by others? Does the existing evidence—ideally collated in a systematic review—suggest that it is likely to be effective or cost-effective? Can it be implemented in a research setting, and is it likely to be widely implementable if the results are favourable?If you are unclear about the answers to these questions, further development work is needed, before you begin your evaluation. If you are evaluating a policy or a service change as it is being implemented, rather than carrying out an experimental intervention study, you still need to be clear about the rationale for the change and the likely size and type of effects, in order to design the evaluation appropriately.

Piloting and feasibility

Questions to ask yourself include: Have you done enough piloting and feasibility work to be confident that the intervention can be delivered as intended? Can you make safe assumptions about effect sizes and variability, and rates of recruitment and retention in the main evaluation study?

Evaluating the intervention

Questions to ask yourself include: What design are you going to use, and why? Is an experimental design preferable, and, if so, is it feasible? If a conventional parallel group RCT is not possible, have you considered alternatives such as cluster randomization or a stepped wedge design? If the effects of the intervention are expected to be large or too rapid to be confused with secular trends, and selection biases are likely to be weak or absent, then an observational design may be appropriate. Have you set up procedures for monitoring the delivery of the intervention and overseeing the conduct of the evaluation?

Including a process evaluation is a good investment to explain discrepancies between expected and observed outcomes, to understand how the context influences outcomes, and to provide insights to aid implementation. Including an economic evaluation will likewise make the results of the evaluation much more useful for decision makers.

Reporting

Questions to ask yourself include: Have you reported your evaluation appropriately, and have you updated your systematic review? It is important to provide a detailed account of the intervention, as well as a standard report of the evaluation methods and findings, to enable replication studies or wider-scale implementation. The results should ideally be presented in the context of an updated systematic review of similar interventions.

Implementation

Questions to ask yourself include: Are your results accessible to decision makers, and have you presented them in a persuasive way? Are your recommendations detailed and explicit?

Strategies to encourage implementation of evaluation findings should be based on a scientific understanding of the behaviours that need to change, the relevant decision-making processes, and the barriers and facilitators of change. If the intervention is

translated into routine practice, monitoring should be undertaken to detect adverse events or long-term outcomes that could not be observed directly in the original evaluation, or to assess whether the effects observed in the study are replicated in routine practice.

Reproduced with permission from Medical Research Council, *Developing and evaluating complex interventions: new guidance*, Copyright © MRC, available from <http://www.mrc.ac.uk/complexinterventionsguidance>. This box is distributed under the terms of the Creative Commons Attribution Non Commercial 4.0 International licence (CC-BY-NC), a copy of which is available at creativecommons.org/licenses/by-nc/.

14.5: Types of clinical interventions is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **2.2: Types of intervention** by Drue H. Barrett, Angus Dawson, Leonard W. Ortmann is licensed CC BY-NC 4.0. Original source: https://library.oapen.org/handle/20.500.12657/33127.

14.6: Evolution of new intervention products and sequence of study phases

Many intervention products, and especially drugs and vaccines, are likely to originate from basic research in laboratories. Such products must go through a long series of tests, before they can be considered for use in the kinds of field trials that are the focus of this book. Before any human use, a new product will be tested in the laboratory for its activity and toxicity in various *in vitro* and animal test systems. If it successfully passes through these stages, studies of safety, toxicity, and activity may be conducted in a small number of human volunteers, with careful clinical monitoring. A series of further studies, each including increasing numbers of subjects, must be carried out before a new product can be introduced for widespread use. Trials in humans usually go through a series of sequential 'phases' of progressively increasing size to establish first the safety and mode of action and then, in later phases, the efficacy against the target disease(s) and safety in a larger number of subjects.

3.1 Clinical studies: Phases I to IV

Phase I studies are exploratory first-in-human trials and may involve the administration of small, then larger, doses of the study product to a small number of healthy human subjects (ten to 50) to gather preliminary data on the product's pharmacokinetics (where the product and its metabolites go within the body and in what concentrations) and pharmacodynamics (what the drug does in the body). These studies can help to establish the dosage and frequency that are safe and necessary to have an effect. These trials are designed to make an initial assessment of the safety and tolerability of the drug or vaccine in a small number of, usually healthy, volunteers.

Phase II trials are conducted for products that have shown no significant safety problems in Phase I trials. They involve progressively larger numbers of participants (for example, initially tens of subjects, but later studies may involve 100s) and are designed to assess how well the intervention works (therapeutic drugs would involve studies in patients, whereas vaccines would be assessed for immunogenicity in healthy volunteers), as well as to check for safety in a larger number of healthy volunteers (vaccines) or in patients (therapeutic drugs). Phase II trials may also be designed to evaluate what doses and the number of doses of the intervention should be given, and what the intervals should be between doses. Usually, a product will be evaluated in a number of different Phase II trials, evaluating its performance under different circumstances, for example, a malaria vaccine might be initially trialled in adults but then tested in progressively younger groups until tested in the final target population of infants.

Phase III trials aim to provide a definitive assessment of the efficacy of the intervention against the primary outcome(s) of interest. They also provide safety data in a larger group of subjects. These trials usually involve large numbers of individuals (e.g. 1000–3000 or more) and are studies that are conducted to produce the evidence of efficacy and safety required to submit a product to a licensing authority. For this reason, they are sometimes called 'pivotal' trials.

Phase IV studies are conducted after the intervention has been shown to be efficacious in Phase III trials and are conducted to assess the safety and effectiveness of an intervention when used under routine health service conditions, or close to these conditions (rather than in the special circumstances of a controlled trial). Where they involve a regulated product, such as a drug or vaccine, they are usually post-registration or post-licensure studies. Safety issues that are important, but which arise in a relatively small proportion of individuals, may only become apparent through Phase IV studies, once there is widespread use of an intervention. Phase IV studies sometimes take the form of randomized trials where the safety and effectiveness are assessed by comparing the results of administering the product to some individuals or communities, but not to others (allocated at random). However, such trials may be difficult to conduct, once a product has been licensed by the national regulatory authority, and then non-randomized assessments must be made, such as through 'before versus after studies' or case-control investigations. Many trials of strategies of how best to use drugs or vaccines can also be considered as Phase IV studies, such as a comparison of intermittent preventive therapy (IPT) using anti-malarial drugs given to all young children, compared to teaching their mothers to recognize and treat their children if they have possible falciparum malaria.

The main focus of the book will be on large-scale Phase III trials conducted 'in the field' (i.e. outside clinical facilities), but there is also a specific chapter on Phase IV studies (see Chapter 22).

Although similar terms are often used for the 'phase' of trials conducted to test the effectiveness or efficacy of interventions that do not use an investigational product, such as behaviour change interventions or incentives, these have much less well-defined, or universally agreed, phases, and it is not uncommon for the first RCT of such an intervention to be the equivalent of a Phase III trial of a drug or vaccine.

3.2 Registration of new interventions

Legal registration procedures are mandated in most countries before a drug or vaccine can be put into general use, and these procedures normally require documentation of the safety and efficacy of the intervention, based on RCTs involving many hundreds of subjects. Further guidance on the rules and regulations for assessing the safety and efficacy of products for use in human beings can be found at the website of the US Food and Drug Administration (<http://www.fda.gov>).

3.3 'Proof of principle' trials

The purposes of field trials may change as experience with an intervention accumulates. Sometimes, particularly in early trials of a new intervention, the purpose of the study is analytic to demonstrate an effect or to establish a principle, with little consideration as to whether the intervention is practicable at the population level for disease control. An example might be the use of a malaria vaccine that must be administered monthly to be effective. Such studies are sometimes called 'explanatory' or 'proof of principle' trials (Schwartz and Lellouch, 1967). Once an effect against the disease under study has been demonstrated, there might then be greater impetus to develop new formulations of the intervention or different schedules that would be more practicable for application in a disease control programme. Subsequent, and generally larger, trials are conducted, in which the purpose is to establish the benefit of an intervention applied under the circumstances of general use. These studies are often called 'pragmatic' trials (Schwartz and Lellouch, 1967).

3.4 Trials of intervention delivery strategies

Although new products developed through basic science research may serve as the impetus for field trials, some interventions or intervention strategies are developed directly as a result of field studies and experience such as a vaccine strategy for smallpox eradication and the use of tsetse fly traps for the control of trypanosomiasis transmission. Thus, trials may be needed not only of the product itself, but also of the way that product is used or delivered. Trials like these would involve intervention 'packages' which might include, for example, the same drug or vaccine, but provided with different educational approaches or delivery methods. Sometimes, an intervention that has been shown to be effective must be added into an ongoing disease control programme that involves other kinds of interventions. For example, it is expected that, when effective malaria vaccines become available, they will be added to other malaria control methods, based on a combination of vector control, case finding, and treatment strategies. Further studies of how best to integrate these interventions into an overall strategy will have to be worked out. In addition, policy and planning decisions about disease control will have to be guided by appropriate cost-effectiveness analyses.

14.6: Evolution of new intervention products and sequence of study phases is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **2.3: Evolution of new intervention products and sequence of study phases** by Drue H. Barrett, Angus Dawson, Leonard W. Ortmann is licensed CC BY-NC 4.0. Original source: https://library.oapen.org/handle/20.500.12657/33127.

14.7: Interpreting Clinical Research

Now that you are familiar with the different forms of nutrition research, the next step is understanding how to interpret and synthesize the information. To synthesize information on a certain topic, the two most popular methods are systematic literature review and meta-analysis. While there are differences between these two methods, they are similar in that they aim to draw a conclusion from the body of research evidence rather than from one study. We will focus on systematic literature reviews because of their rising popularity in biomedical literature.

A systematic literature review does what the name implies, it systematically reviews the literature related to a certain research question. For example, the research question might be, "Does chocolate decrease blood pressure?" There is a method for performing the review established ahead of time that details answers to questions such as:

- How will journal articles be identified?
- Using what databases?
- What search terms will be used?

The end product is a conclusion based on the evidence in the identified journal articles. The figure below illustrates how a systematic literature review identifies articles to utilize.

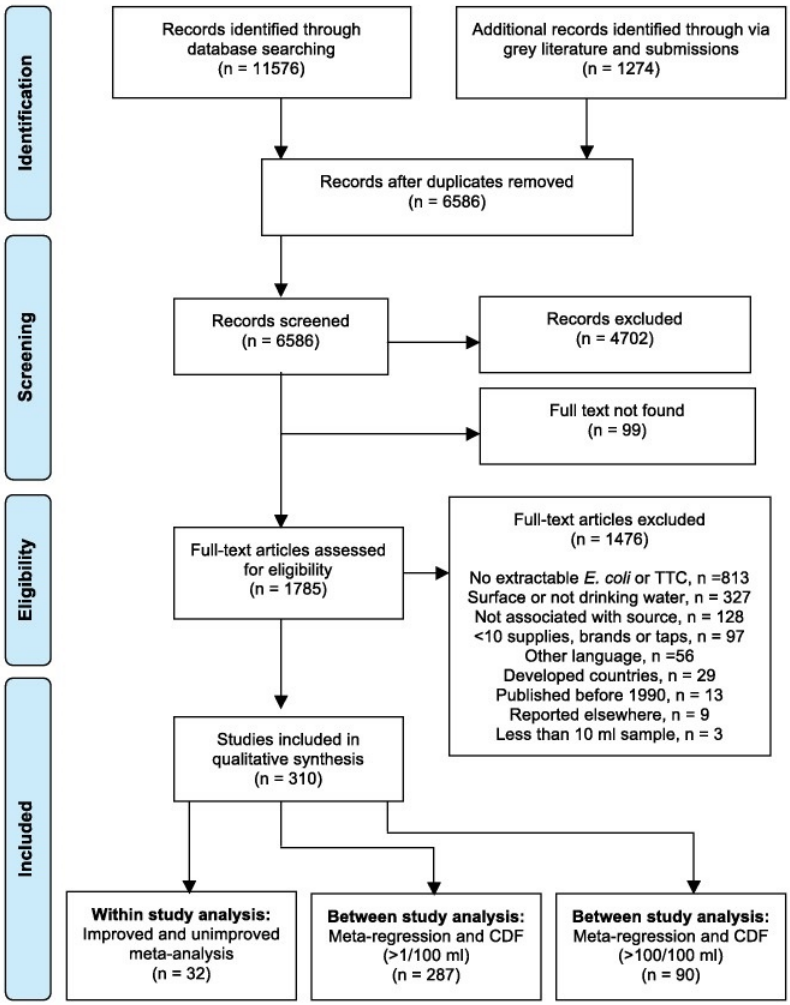


Figure **14.7.1**: An example of the process of a systematic literature review. (CC BY 4.0 International; Bain R, Cronk R, Wright J, Yang H, Slaymaker T, Bartram J (2014). "Fecal Contamination of Drinking-Water in Low- and Middle-Income Countries: A Systematic Review and Meta-Analysis". *PLOS Medicine*. DOI:10.1371/journal.pmed.1001644)

The video below, by the Cochrane Collaboration that performs many systematic literature reviews, describes what systematic literature reviews are.

Video **14.7.1**: This video explains why systematic reviews are important and how they are done. This includes an explanation of how the effects of interventions are compared in order to provide evidence. https://www.youtube.com/watch?v=egJlW4vkb1Y

Because they synthesize the findings from multiple trials or studies, systematic literature reviews are considered the highest level of nutrition research evidence and are therefore shown at the top of the strength of nutrition research pyramid below. In vitro studies are shown at the bottom of the triangle because they are the weakest form of evidence overall. It should be pointed out that most systematic literature reviews only consider epidemiological studies and clinical trials and do not include animal studies or in vitro evidence. While these studies are of less strength, they should still be considered.

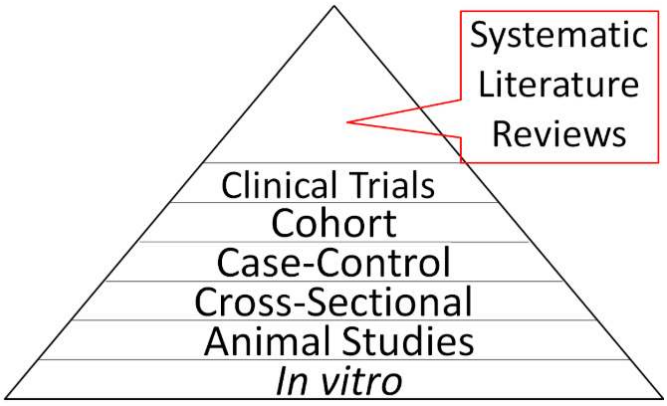


Figure **14.7.2**: Strength of scientific research, a higher position on the pyramid means greater strength of that form of evidence.

There are a couple of other factors to consider relating to the different forms of research. First, epidemiological studies cannot show causality (i.e. smoking causes lung cancer), but instead identify relationships or associations (i.e. smoking is associated with lung cancer occurrence). Clinical trials/human studies are the best form of primary research because their findings should be directly translatable to patients. So why use other forms of research? The description below should help explain why the other forms of research are also important.

In general, a certain sequence of studies in nutrition research is often followed as shown in the figure below.

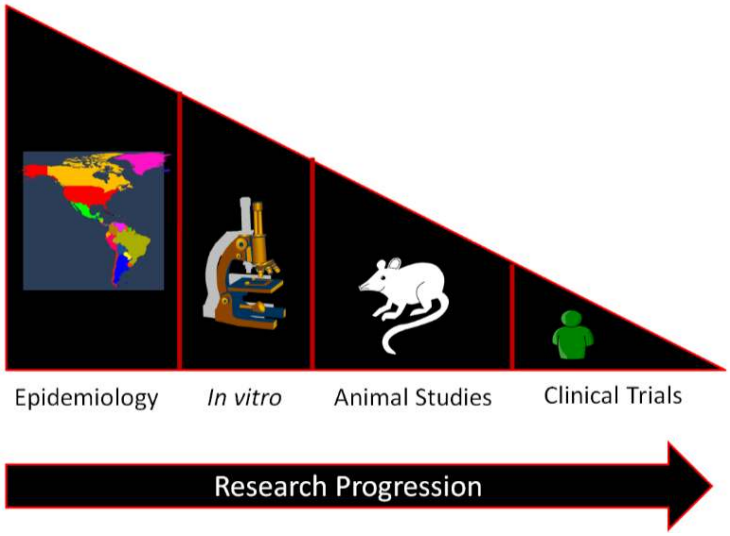


Figure 14.7.3: The progression of nutrition research. The shape indicates the selection of the most promising and safest hypotheses being moved to the next level. Thus, there is a decrease in the overall number of hypotheses represented by the height of the figure) being investigated by the next form of research.

Epidemiological studies find relationships between food/food components and a specific health outcome. This relationship is then investigated by in vitro studies and then, some of the most promising move to animal studies. Then the most promising and safe food/food components are moved into clinical trials/human studies. If it is an individual compound, there will be smaller trials designed to see if the compound is safe before it is moved into larger clinical trials to determine whether the food/food component results in beneficial health outcomes. The overall effect of this process is to select the most promising and safe food/food components for the clinical trials/human studies. This allows time and money to be used more efficiently, because while clinical trials/human studies are the best form of research, they are also normally the most expensive and time-consuming.

Researchers nevertheless have been tempted to skip directly to clinical trials in the past rather than following the research progression. To illustrate what happens in these situations, the following describes a couple of examples when "normal" research progression hasn't been followed.

Beta-Carotene and Lung Cancer

In the early 1980s there was a lot of excitement among researchers over the epidemiological evidence showing that higher dietary consumption of the carotenoid, beta-carotene, decreased lung cancer risk1. By the mid 1980's, two large, randomized, double-blind placebo-controlled trials began to determine whether high-dose beta-carotene supplementation (far higher than dietary intake) could decrease lung cancer incidence in high-risk populations before in vitro or animal studies had investigated this relationship. The research community was shocked when these two studies were terminated early in the mid 1990s because of significant increases in lung cancer incidence among smokers receiving beta-carotene supplements2,3. In vitro and animal studies completed after the clinical trial found that as beta-carotene intake shifts from normal dietary levels to high, supplement-type levels, the effect on lung cancer development also shifts from beneficial to detrimental in combination with smoke or carcinogen exposure4. Thus, if the normal research progression had been followed, it is likely that a lower dose of beta-carotene would have been used or trials wouldn't have been undertaken at all.

Selenium, Vitamin E, and Prostate Cancer

Another example of when the research progression was not followed was the relationship between selenium, vitamin E, and prostate cancer. Two clinical trials had found secondary outcomes that suggested that selenium and vitamin E supplementation may decrease prostate cancer risk 5,6. A secondary outcome is one that was not the primary outcome that the clinical trial was designed to find, thus they need to be interpreted with some caution. Rather than examining the relationship using in vitro and animal studies, a clinical trial was undertaken to determine if selenium and vitamin E supplementation alone, or in combination, could decrease the development of prostate cancer incidence7. This clinical trial was also terminated early because of a nonsignificant increase in prostate cancer in those receiving vitamin E8, and a nonsignificant increase in diabetes among those receiving

selenium9. In vitro studies and animal models performed since the clinical trial was undertaken and after its termination suggest that vitamin E is not effective and that another form of selenium could have possibly been more effective10.

In addition to the lessons learned about the sequence of research in nutrition, these studies add to growing evidence that suggests that single-agent interventions, even in combination, may not be an effective strategy for improving health. The common nutrition research approach was, after epidemiology finds an association or relationship, is to use in vitro and animal studies to identify the specific compound in a certain food that is responsible. This has been termed the reductionist approach because it takes something complex (food) and reduces it down to its simpler components. However, there is growing evidence that this may be a flawed approach. Some nutrition researchers feel that more focus should be on the food itself, rather than trying to discern exactly what is responsible for the beneficial health outcomes. Because it may not be one or 2 compounds alone that are responsible for the effect, it might be difficult to determine from the multitude of nutrients in foods which are responsible for the beneficial effect. This will mean changes in the overall research approach, especially at the human intervention studies/clinical trials level, because in most cases there is not a way to give a “placebo food”11,12.

References

1. Peto R, Doll R, Buckley JD, Sporn MB. Can dietary beta-carotene materially reduce human cancer rates? Nature 290, 201-208, 1981.
2. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. the alpha-tocopherol, beta carotene cancer prevention study group. N Engl J Med. 330, 1029-1035, 1994.
3. Goodman GE, Thornquist MD, Balmes J, Cullen MR, Meyskens FL,Jr., Omenn GS, et al. The beta-carotene and retinol efficacy trial: Incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping beta-carotene and retinol supplements. J Natl Cancer Inst 96, 1743-1750, 2004.
4. Lindshield BL, Erdman JW. (2006) Carotenoids. In: Bowman BA, Russell RM, editors. Present Knowledge in Nutrition. Washington, D.C.: International Life Sciences Institute. pp. 184-197.
5. Clark LC, Dalkin B, Krongrad A, Combs GF,Jr., Turnbull BW, Slate EH, et al. Decreased incidence of prostate cancer with selenium supplementation: Results of a double-blind cancer prevention trial. Br J Urol 81, 730-734, 1998.
6. Heinonen OP, Albanes D, Virtamo J, Taylor PR, Huttunen JK, Hartman AM, et al. Prostate cancer and supplementation with alpha-tocopherol and beta-carotene: Incidence and mortality in a controlled trial. J Natl Cancer Inst 90, 440-446, 1998.
7. Lippman SM, Goodman PJ, Klein EA, Parnes HL, Thompson IM,Jr., Kristal AR, et al. Designing the selenium and vitamin E cancer prevention trial (SELECT). J Natl Cancer Inst 97, 94-102, 2005.
8. Klein EA, Thompson IM, Tangen CM, Crowley JJ, Lucia MS, Goodman PH, Minasian LM, Ford LG, Parnes HL, Gaziano JM, Karp DD, Lieber MM, Walther PJ, Klotz L, Parsons JK, Chin JL, Darke AK, Lippman SM, Goodman GE, Meysken FL, Baker LH. Vitamin E and the Risk of Prostate Cancer: The Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA 306, 1549-1556, 2011.
9. Lippman SM, Klein EA, Goodman PJ, Lucia MS, Thompson IM, Ford LG, et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: The selenium and vitamin E cancer prevention trial (SELECT). JAMA 301, 39-51, 2009.
10. Lindshield BL, Ford NA, Canene-Adams K., Diamond AM, Wallig MA, Erdman JW Jr. Selenium, but not lycopene or vitamin E, decreases growth of transplantable Dunning R3327-H rat prostate tumors. PLoS One. 2010.
11. Gann PH, Khachik F. Tomatoes or lycopene versus prostate cancer: Is evolution anti-reductionist? J Natl Cancer Inst 95, 1563-1565, 2003.
12. Gann PH. Randomized trials of antioxidant supplementation for cancer prevention: First bias, now chance--next, cause. JAMA 301, 102-103, 2009.

This page titled 14.7: Interpreting Clinical Research is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Brian Lindshield.

- **1.7: Interpreting Research** by Brian Lindshield is licensed CC BY-NC-SA 4.0. Original source: http://goo.gl/vOAnR.

CHAPTER OVERVIEW

15: What is it worth? Estimating Value

This chapter introduces some of the methods that are used to estimate the value of pharmaceuticals, which are in turn used to make business investment, pricing, and healthcare provider decisions.

15.1: Types of economic analyses of health interventions

15.2: The Time Value of Money

15.3: Calculating the Relationship of Time and Value

15: What is it worth? Estimating Value is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

15.1: Types of economic analyses of health interventions

The main types of economic analyses are cost-effectiveness analysis (CEA), cost-utility analysis (CUA), and cost-benefit analyses (CBA). How the results of these different kinds of analysis are expressed is shown in Table 19.1. CEA and CUA are those most commonly used in the analysis of health interventions. The problem with CBA is that it requires putting a monetary value on a life saved.

Analysis of the costs involved in providing the health interventions under comparison in a trial is needed for all three types of analysis. Measurement of these costs can be made in the context of an intervention study, provided due account is taken of the fact that the costs associated with an intervention in a trial may be different from those which would apply if the intervention was applied in a public health programme. It is important therefore to separate out any trial-specific costs that would not be incurred in more widespread deployment of the intervention. For example, often checks are made in a trial that the intervention has been delivered to participants in the appropriate fashion at an appropriate time. Such checks might not be made, or not be made with the same rigour, in the context of the deployment of the intervention in the routine public health system. However, there may be additional costs in the public health deployment of an intervention that would not be incurred in a trial. For example, drugs or vaccines for use in a trial are often donated, whereas, for public health use, they may have to be purchased.

Table 19.1 Types of economic analysis

Type of analysis	Costs	Outcome (effect)	Results expressed as:
Cost	Monetary units (commonly US $)	Not relevant	$ per unit of output (for example, $ per fully vaccinated child)
CEA	Monetary units (commonly US $)	Effect of intervention (for example, cases prevented)	$ per effect (for example, $ per case prevented)
CUA	Monetary units (commonly US $)	Premature mortality and disability averted (measured in DALYs) or healthy life time gained (QALYs)	$ per DALY averted or QALY gained
CBA	Monetary units (commonly US $)	Monetary units (for example, value of a statistical life)	Benefit–cost ratio or net present value (for example, money value of benefits–costs

DALY, disability-adjusted life-year; QALY, quality-adjusted life-year.

2.1 Cost-effectiveness analysis

CEA has been the most commonly employed type of economic analysis used in relation to randomized trials of health interventions. CEA compares the costs to accomplish a specific technical goal by a new method with the costs of the present method such as the costs per case of a particular disease diagnosed by the new method with the costs per case of disease diagnosed using the current diagnostic method, or the costs of the prevention of a death from a given cause by the new intervention compared to the costs of the prevention of a death with the present intervention. Note that it is the incremental cost-effectiveness ratio that captures the value of the new method being examined, i.e. the difference in costs between the new method and the present method, divided by the difference in effects between the new method and the present method. This summary measure thus captures the extra cost per additional unit of effect and begs the question ‘is it worth it?’.

2.2 Cost-utility analysis

For CUA, the effects of an intervention are expressed as a measure of ‘utility’. Simply, the utility is a measure of the impact of the intervention on the health status of the individual or population, commonly stated as a combined measure of mortality (amount of life lost due to premature death) and morbidity (amount of life lived with disability, weighted according to its seriousness and

duration). Commonly used utility measures are the disability-adjusted life-year (DALY) and the quality-adjusted life-year (QALY) (Hyder et al., 2012).

2.2.1 Disability-adjusted life-years and quality-adjusted life-years

The DALY was first given prominence in the *World development report 1993* (World Bank, 1993) and has become the most widely used composite measure of population health in LMICs. It built on earlier work by the Ghana Health Assessment Project Team (1981) who introduced the similar concept of 'amount of healthy life lost', combining measures of the effects of a disease, in terms of life lost both from mortality (expected years of life remaining had the disease not occurred) and from morbidity (severity and duration of disability).

DALYs are calculated by combining the years of life lost (YLL) from premature mortality with the years of life lived with disability (YLD), weighted according to a severity grading. Thus:

$$\mathbf{DALY = YLL + YLD} \tag{15.1.1}$$

As originally formulated, the DALY directly incorporated three social value choices: (1) life expectancy values, (2) discount rates for future life, and (3) variable weighting for life lived at different ages. The recent *Global burden of disease report* for 2010, however, has dropped both discounting and age weighting (Murray et al., 2012).

A related measure, the QALY, was introduced in 1976 to provide a guide for individuals to select among alternative tertiary health care interventions (Zeckhauser and Shepard, 1976). The idea was to develop a measure of quality of life that would enable investigators to compare expected outcomes from different interventions, a measure that valued possible health states both for their impact on the quality of life and for their duration. The measure sums the time an individual spends in different health states, using weights on a scale of 0 (in a state equivalent to being dead) to 1 (perfectly healthy) for each health state; it is the sum of arithmetic products of the duration of time spent in a state and a measure of the quality of life in that state. QALYs in modified forms have come into widespread use in the UK (by the UK National Institute for Health and Clinical Excellence), Europe, and the USA (by the US Agency for Healthcare Quality and Research).

Despite distinctly different origins, DALYs and QALYs, with appropriate formulation and comparable parameters, can be considered equivalent indicators to assess intervention utility. However, there are many versions of both DALYs and QALYs, and it is very important to know exactly what is being counted in the study under consideration.

2.3 Cost–benefit analysis

CBA goes a step beyond CEA or CUA and expresses both costs and effects (or utility) of interventions in monetary terms. It directly compares the monetary costs of an intervention with the monetary benefits from the intervention. If the monetary benefits from an intervention exceed the monetary costs, the decision is straightforward in purely economic terms—implement the intervention. For most sectors, other than health, CBA is the standard form of economic analysis, and it lies at the centre of decision making in these sectors. For example, the decision to build a new road would be based on considerations of the cost of building the road, compared to the economic benefits it would bring (which might include reduction of wear and tear on vehicles, increased speed of delivery of people and goods, increase in trade, and also reduction in injuries and deaths from accidents). The aspect that has impeded its use in the health sector is that, in order to use CBA, a monetary value must be placed on human life. Some argue that this is done implicitly in any decision process, but there has been a reluctance to do this explicitly. Nevertheless, it should be recognized that decisions are regularly being taken in both public and private sectors that implicitly do place a monetary value on life. There are several different approaches to valuing human life which may give marked different results (Australian Safety and Compensation Council, 2008; Viscusi and Aldy, 2003), but further discussion is beyond the scope of this book.

Sometimes, a narrower perspective may be taken with respect to CBA. For example, in consideration of whether the public health service should introduce a vaccine against pneumonia, the costs assessed may be limited to those for the health system. If the vaccine reduces the incidence of pneumonia, the costs of delivering the vaccine to the at-risk population could be compared with the reduction of health service costs from fewer cases of pneumonia to treat. If there is a clear benefit, simply based on a CBA that only considers costs spent and saved by the health system, the decision about the introduction of the vaccine may be relatively straightforward. It becomes more complicated if there is not a monetary saving to the health system (for example, it costs more to deliver the vaccine than the saving in health service costs), but there is a reduction in mortality and/or morbidity in the population.

In fact, a ‘true’ CBA requires a comprehensive and comparable range of inputs and outcomes, all expressed in monetary terms and, for fatal diseases, that would include putting an explicit monetary value on human life at different ages.

15.1: Types of economic analyses of health interventions is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- **19.2: Types of economic analyses** by Drue H. Barrett, Angus Dawson, Leonard W. Ortmann is licensed CC BY-NC 4.0. Original source: https://library.oapen.org/handle/20.500.12657/33127.

15.2: The Time Value of Money

Learning Objectives

1. Explain the value of liquidity.
2. Demonstrate how time creates distance, risk, and opportunity cost.
3. Demonstrate how time affects liquidity.
4. Analyze how time affects value.

Part of the planning process is evaluating the possible future results of a decision. Since those results will occur some time from now (i.e., in the future), it is critical to understand how time passing may affect those benefits and costs—not only the probability of their occurrence, but also their value when they do. Time affects value because time affects liquidity.

Liquidity is valuable, and the liquidity of an asset affects its value: all things being equal, the more liquid an asset is, the better. This relationship—how the passage of time affects the liquidity of money and thus its value—is commonly referred to as the **time value of money**, which can actually be calculated concretely as well as understood abstractly.

Suppose you went to Mexico, where the currency is the peso. Coming from the United States, you have a fistful of dollars. When you get there, you are hungry. You see and smell a taco stand and decide to have a taco. Before you can buy the taco, however, you have to get some pesos so that you can pay for it because the right currency is needed to trade in that market. You have wealth (your fistful of dollars), but you don't have wealth that is liquid. In order to change your dollars into pesos and acquire liquidity, you need to exchange currency. There is a fee to exchange your currency: a **transaction cost**, which is the cost of simply making the trade. It also takes a bit of time, and you could be doing other things, so it creates an opportunity cost (see Chapter 2). There is also the chance that you won't be able to make the exchange for some reason, or that it will cost more than you thought, so there is a bit of risk involved. Obtaining liquidity for your wealth creates transaction costs, opportunity costs, and risk.

Figure 4.2.1



© 2010 Jupiterimages Corporation

In general, transforming not-so-liquid wealth into liquid wealth creates transaction costs, opportunity costs, and risk, all of which take away from the value of wealth. Liquidity has value because it can be used without any additional costs.

One dimension of difference between not-so-liquid wealth and liquidity is time. Cash flows (CF) in the past are sunk, cash flows in the present are liquid, and cash flows in the future are not yet liquid. You can only make choices with liquid wealth, not with cash that you don't have yet or that has already been spent. Separated from your liquidity and your choices by time, there is an opportunity cost: if you had liquidity now, you could use it for consumption or investment and benefit from it now. There is also risk, as there is always some uncertainty about the future: whether or not you will actually get your cash flows and just how much they'll be worth when you do.

The further in the future cash flows are, the farther away you are from your liquidity, the more opportunity cost and risk you have, and the more that takes away from the present value (PV) of your wealth, which is not yet liquid. In other words, time puts distance between you and your liquidity, and that creates costs that take away from value. The more time there is, the larger its effect on the value of wealth.

Financial plans are expected to happen in the future, so financial decisions are based on values some distance away in time. You could be trying to project an amount at some point in the future—perhaps an investment payout or college tuition payment. Or perhaps you are thinking about a series of cash flows that happen over time—for example, annual deposits into and then withdrawals from a retirement account. To really understand the time value of those cash flows, or to compare them in any reasonable way, you have to understand the relationships between the nominal or face values in the future and their equivalent, present values (i.e., what their values would be if they were liquid today). The equivalent present values today will be less than the nominal or face values in the future because that distance over time, that separation from liquidity, costs us by discounting those values.

KEY TAKEAWAY

- Liquidity has value because it enables choice.
- Time creates distance or delay from liquidity.
- Distance or delay creates risk and opportunity costs.
- Time affects value by creating distance, risk, and opportunity costs.
- Time discounts value.

EXERCISE

1. How does the expression "a bird in the hand is worth two in the bush" relate to the concept of the time value of money?
2. In what four ways can "delay to liquidity" affect the value of your wealth?

This page titled 15.2: The Time Value of Money is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous.

- **4.2: The Time Value of Money** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/individual-finance.

15.3: Calculating the Relationship of Time and Value

Learning Objectives

1. Identify the factors you need to know to relate a present value to a future value.
2. Write the algebraic expression for the relationship between present and future value.
3. Discuss the use of the algebraic expression in evaluating the relationship between present and future values.
4. Explain the importance of understanding the relationships among the factors that affect future value.

Financial calculation is not often a necessary skill since it is easier to use calculators, spreadsheets, and software. However, understanding the calculations is important in understanding the relationships between time, risk, opportunity cost, and value.

To do the math, you need to know

- what the future cash flows (CF) will be,
- when the future cash flows will be,
- the rate at which time affects value (e.g., the costs per time period, or the magnitude [the size or amount] of the effect of time on value).

It is usually not difficult to forecast the timing and amounts of future cash flows. Although there may be some uncertainty about them, gauging the rate at which time affects money can require some judgment. That rate, commonly called the **discount rate** because time discounts value, is the opportunity cost of not having liquidity. Opportunity cost derives from forgone choices or sacrificed alternatives, and sometimes it is not clear what those might have been (see Chapter 2). It is an important judgment call to make, though, because the rate will directly affect the valuation process.

Figure 4.3.1



© 2010 Jupiterimages Corporation

At times, the alternatives are clear: you could be putting the liquidity in an account earning 3 percent, so that's your opportunity cost of not having it. Or you are paying 6.5 percent on a loan, which you wouldn't be paying if you had enough liquidity to avoid having to borrow; so that's your opportunity cost. Sometimes, however, your opportunity cost is not so clear.

Say that today is your twentieth birthday. Your grandparents have promised to give you $1,000 for your twenty-first birthday, one year from today. If you had the money today, what would it be worth? That is, how much would $1,000 worth of liquidity one year from now be worth today?

That depends on the cost of its not being liquid today, or on the opportunity costs and risks created by not having liquidity today. If you had $1,000 today, you could buy things and enjoy them, or you could deposit it in an interest-bearing account. So on your twenty-first birthday, you would have more than $1,000. You would have the $1,000 plus whatever interest it had earned. If your bank pays 4 percent per year (interest rates are always stated as annual rates) on your account, then you would earn $40 of interest in the next year, or $1,000 × .04. So on your twenty-first birthday you would have $1,040.

$1,000+(1,000×0.04)=$1,000×(1+0.04)=$1,040

Figure 4.3.2

Today	Interest Rate	Time (years)	One Year from Now
1,000	0.04	1	$1{,}040 = 1{,}000 \times (1 + 0.04)$

If you left that amount in the bank until your twenty-second birthday, you would have

1,040+(1,040×0.04)=1,040×(1+0.04) =[1,000×(1+0.04)] × (1+0.04) =1,000×(1+0.04) 2 =1,081.60.

To generalize the computation, if your **present value**, or PV, is your value today, *r* is the rate at which time affects value or discount rate (in this case, your interest rate), and if *t* is the number of time periods between you and your liquidity, then the **future value**, or FV, of your wealth would be

Figure 4.3.3

Today	Interest Rate	Time (years)	One Year from Now
1,000	0.04	1	$1{,}040 = 1{,}000 \times (1 + 0.04)^1$
1,000	0.04	2	$1{,}081.60 = 1{,}000 \times (1 + 0.04)^2$
PV	r	t	$FV = PV \times (1 + r)^t$

PV× (1+r) t =FV.

In this case,

1,000× (1.04) 1 =1,040 and 1,000× (1.04) 2 =1,081.60.

Assuming there is little chance that your grandparents will not be able to give this gift, there is negligible risk. Your only cost of not having liquidity now is the opportunity cost of having to delay consumption or not earning the interest you could have earned.

The cost of delayed consumption is largely derived from a subjective valuation of whatever is consumed, or its **utility** or satisfaction. The more value you place on having something, the more it "costs" you not to have it, and the more the time that you are without it affects its value.

Assuming that if you had the money today you would save it (as it's much harder to quantify your joy from consumption), by having to wait to get it until your twenty-first birthday—and *not* having it today—you miss out on $40 it could have earned.

So, what would that nominal $1,000 (that future value that you get one year from now) actually be worth today? The rate at which time affects your value is 4 percent because that's what having a choice (spend it or invest it) could earn for you if only you had received the $1,000. That's your opportunity cost. That's what it costs you to not have liquidity. Since

PV× (1+r) t =FV, then PV=FV/[(1+r) t], so PV=1,000/(1.04 1)=961.5385.

Your gift is worth $961.5385 today (its present value). If your grandparents offered to give you your twenty-first birthday gift on your twentieth birthday, they could give you $961.5385 today, which would be the equivalent value to you of getting $1,000 one year from now.

It is important to understand the relationships between time, risk, opportunity cost, and value. This equation describes that relationship:

PV× (1+r) t =FV.

The "r" is more formally called the "discount rate" because it is the rate at which your liquidity is discounted by time, and it includes not only opportunity costs but also risk. (On some financial calculators, "r" is displayed as "I" or "i.")

The "t" is how far away you are from your liquidity over time.

Studying this equation yields valuable insights into the relationship it describes. Looking at the equation, you can observe the following relationships.

The more time (t) separating you from your liquidity, the more time affects value. The less time separating you from your liquidity, the less time affects value (as *t* decreases, PV increases).

As *t* increases	the PV of your FV liquidity decreases
As *t* decreases	the PV of your FV liquidity increases

The greater the rate at which time affects value (r), or the greater the opportunity cost and risk, the more time affects value. The less your opportunity cost or risk, the less your value is affected.

As *r* increases	the PV of your FV liquidity decreases
As *r* decreases	the PV of your FV liquidity increases

Figure 4.6 presents examples of these relationships.

	Today	Interest Rate	Time (years)	Future Value
Example A	1,000	0.04	1	1,040 $= PV \times 104\%$
Greater effect *r increases*	1,000	0.10	1	1,100 $= PV \times 110\%$
More time *t increases*	1,000	0.04	3	1,124.86 $= PV \times 104^3\%$
Less effect *r decreases*	1,000	0.01	1	1,010 $= PV \times 101\%$
Less time *t decreases*	1,000	0.04	0.5	1,019.80 $= PV \times 104^{0.5}\%$
Example B	961.54 $= FV \div 104\%$	0.04	1	1,000
Greater effect *r increases*	909.09 $= FV \div 110\%$	0.10	1	1,000
More time *t increases*	889.00 $= FV \div 104^3\%$	0.04	3	1,000
Less effect *r decreases*	990.10 $= FV \div 101\%$	0.01	1	1,000
Less time *t decreases*	980.58 $= FV \div 104^{.05}\%$	0.04	0.5	1,000

Figure 4.3.4 :Present Values, Interest Rates, Time, and Future Values

The strategy implications of this understanding are simple, yet critical. All things being equal, it is more valuable to have liquidity (get paid, or have positive cash flow) *sooner* rather than later and give up liquidity (pay out, or have negative cash flow) *later* rather than sooner.

If possible, accelerate incoming cash flows and decelerate outgoing cash flows: get paid sooner, but pay out later. Or, as Popeye's pal Wimpy used to say, "I'll give you 50 cents tomorrow for a hamburger today."

KEY TAKEAWAYS

- To relate a present (liquid) value to a future value, you need to know
  - what the present value is or the future value will be,
  - when the future value will be,
  - the rate at which time affects value: the costs per time period, or the magnitude of the effect of time on value.
- The relationship of
  - present value (PV),
  - future value (FV),
  - risk and opportunity cost (the discount rate, r), and
  - time (t), may be expressed as
  - $PV \times (1 + r)^t = FV$.
- The above equation yields valuable insights into these relationships:
  - The more time (t) creates distance from liquidity, the more time affects value.
  - The greater the rate at which time affects value (r), or the greater the opportunity cost and risk, the more time affects value.
  - The closer the liquidity, the less time affects value.
  - The less the opportunity cost or risk, the less value is affected.

- To maximize value, get paid sooner and pay later.

Exercises

1. In My Notes or your financial planning journal, identify a future cash flow. Calculate its present value and then calculate its future value based on the discount rate and time to liquidity. Repeat the process for other future cash flows you identify. What pattern of relationships do you observe between time and value?
2. Try the Time Value of Money calculator at http://www.money-zine.com/Calculators/Investment-Calculators/Time-Value-of-Money-Calculator/. How do the results compare with your calculations in Exercise 1?
3. View the TeachMeFinance.com animated audio slide show on "The Time Value of Money" at teachmefinance.com/timevalueofmoney.html. This slide show will walk you through an example of how to calculate the present and future values of money. How is each part of the formula used in that lesson equivalent to the formula presented in this text?
4. To have liquidity, when should you increase positive cash flows and decrease negative cash flows, and why?

This page titled 15.3: Calculating the Relationship of Time and Value is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous.

- **4.3: Calculating the Relationship of Time and Value** by Anonymous is licensed CC BY-NC-SA 3.0. Original source: https://2012books.lardbucket.org/books/individual-finance.

CHAPTER OVERVIEW

16: Appendix

- Getting started with Zotero https://www.zotero.org/support/quick_start_guide
- Fordham ACS Journal Login: https://fordham.libguides.com/AmericanChemSoc

- **Topic hierarchy**

  16.1: Getting started with Zotero

16: Appendix is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

16.1: Getting started with Zotero

16.1: Getting started with Zotero is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Index

A

acid strength
8.4: The Strengths of Acids and Bases
alkane
5: Organic Molecules
alkene
5: Organic Molecules
atomic structure
1: Atomic Structure
atoms
1: Atomic Structure

B

base strength
8.4: The Strengths of Acids and Bases
bile
6.5: Steroids

C

census
14.3: Random Sampling
charge balance
2.3: Ions
cholesterol
6.5: Steroids
Cis isomers
5: Organic Molecules
Clinical study
14: Does it work? Clinical Testing
14.6: Evolution of new intervention products and sequence of study phases
Complex interventions
14.5: Types of clinical interventions
Compound Interest
15: What is it worth? Estimating Value
covalent bond
3: Covalent Bonding and Simple Molecular Compounds
3.2: Covalent Bonds

D

denaturation
9.5: Proteins
diastereomers
7.6: Compounds with multiple chiral centers
Diisopropyl fluorophosphate (DIFP)
9.10: Enzyme Inhibition
discount rate
15.3: Calculating the Relationship of Time and Value
Drugs for the prevention of disease
14.5: Types of clinical interventions
dynamic equilibrium
8.4: The Strengths of Acids and Bases

E

ED50
10.4: Pharmacology as an Equilibrium and Kinetics Problem
Environmental alterations
14.5: Types of clinical interventions
Enzyme Classification Number
9.7: Enzymes
enzyme inhibition
9.10: Enzyme Inhibition
enzymes
9.7: Enzymes
epimerases
7.6: Compounds with multiple chiral centers

F

fats
6.3: Fats and Oils
fatty acids
6.2: Fatty Acids
future value
15.3: Calculating the Relationship of Time and Value

H

Health systems interventions
14.5: Types of clinical interventions

I

Implementation research
14.5: Types of clinical interventions
InChI
5: Organic Molecules
insulin
9.1: Prelude to Amino Acids, Proteins, and Enzymes
ionic bonding
2: Ionic Bonding and Simple Ionic Compounds
IUPAC
5: Organic Molecules

L

legal action
14.5: Types of clinical interventions
Legislation
14.5: Types of clinical interventions

M

Maternal and neonatal interventions
14.5: Types of clinical interventions
membrane lipids
6.4: Membranes and Membrane Lipids
membranes
6.4: Membranes and Membrane Lipids
molecules
3: Covalent Bonding and Simple Molecular Compounds

N

net present value
15: What is it worth? Estimating Value
nucleic acids
11: Molecular Genetics
nucleotide
11.2: Nucleotides
Nutritional interventions
14.5: Types of clinical interventions

O

oils
6.3: Fats and Oils
organic chemistry
5: Organic Molecules

P

periodic table
1: Atomic Structure
pharmacology
10.4: Pharmacology as an Equilibrium and Kinetics Problem
poisons
9.10: Enzyme Inhibition
present value
15.3: Calculating the Relationship of Time and Value
purine
11.2: Nucleotides
pyrimidine
11.2: Nucleotides

Q

qaly
15: What is it worth? Estimating Value
quota sampling
14.3: Random Sampling

R

racemic mixture
7.4: Naming chiral centers- the R and S system
reaction coordinate diagram
10.2: A Quick Review of Thermodynamics and Kinetics
Renaturation
9.5: Proteins

S

simple random sample
14.3: Random Sampling
SMILES
5: Organic Molecules
statistics
14: Does it work? Clinical Testing
steroids
6.5: Steroids
subsidies
14.5: Types of clinical interventions

T

taxation
14.5: Types of clinical interventions
thalidomide
7.4: Naming chiral centers- the R and S system

Therapeutic Interventions
14.5: Types of clinical interventions
time value of money
15.2: The Time Value of Money
transaction cost
15.2: The Time Value of Money
triglyceride
6.3: Fats and Oils

U

utility
15.3: Calculating the Relationship of Time and Value

V

Vaccines
14.5: Types of clinical interventions
Vector and intermediate host control
14.5: Types of clinical interventions
VSEPR
3.6: Characteristics of Molecules

W

Waxes
6.2: Fatty Acids

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Chem1102: Drug Discovery - From the Laboratory to the Clinic

Webpages: 127

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 4.0: 66.9% (85 pages)
- Undeclared: 26% (33 pages)
- CC BY-SA 4.0: 2.4% (3 pages)
- CC BY-NC 4.0: 2.4% (3 pages)
- CC BY-NC-SA 3.0: 1.6% (2 pages)
- Public Domain: 0.8% (1 page)

By Page

- Chem1102: Drug Discovery - From the Laboratory to the Clinic - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Atomic Structure - *Undeclared*
    - 1.1: The Elements - *CC BY-NC-SA 4.0*
    - 1.2: Atomic Theory - *CC BY-NC-SA 4.0*
    - 1.3: The Structure of Atoms - *CC BY-NC-SA 4.0*
    - 1.4: Nuclei of Atoms - *CC BY-NC-SA 4.0*
    - 1.5: Atomic Masses - *CC BY-NC-SA 4.0*
    - 1.6: Arrangements of Electrons - *CC BY-NC-SA 4.0*
    - 1.7: The Periodic Table - *CC BY-NC-SA 4.0*
    - 1.8: Elements, Atoms, and the Periodic Table (Summary) - *CC BY-NC-SA 4.0*
    - 1.9: Elements, Atoms, and the Periodic Table (Exercises) - *CC BY-NC-SA 4.0*
  - 2: Ionic Bonding and Simple Ionic Compounds - *CC BY-NC-SA 4.0*
    - 2.1: Prelude to Ionic Bonding and Simple Ionic Compounds - *CC BY-NC-SA 4.0*
    - 2.2: Two Types of Bonding - *CC BY-NC-SA 4.0*
    - 2.3: Ions - *CC BY-NC-SA 4.0*
    - 2.4: Formulas for Ionic Compounds - *CC BY-NC-SA 4.0*
    - 2.5: Ionic Bonding and Simple Ionic Compounds (Exercises) - *CC BY-NC-SA 4.0*
    - 2.6: Ionic Bonding and Simple Ionic Compounds (Summary) - *CC BY-NC-SA 4.0*
  - 3: Covalent Bonding and Simple Molecular Compounds - *CC BY-NC-SA 4.0*
    - 3.1: Prelude to Covalent Bonding and Simple Molecular Compounds - *CC BY-NC-SA 4.0*
    - 3.2: Covalent Bonds - *CC BY-NC-SA 4.0*
    - 3.3: Covalent Compounds - Formulas and Names - *CC BY-NC-SA 4.0*
    - 3.4: Drawing Lewis Structures - *CC BY-NC-SA 4.0*
    - 3.5: Characteristics of Covalent Bonds - *CC BY-NC-SA 4.0*
    - 3.6: Characteristics of Molecules - *CC BY-NC-SA 4.0*
    - 3.7: Organic Chemistry - *CC BY-NC-SA 4.0*
    - 3.8: Covalent Bonding and Simple Molecular Compounds (Exercises) - *CC BY-NC-SA 4.0*
    - 3.9: Covalent Bonding and Simple Molecular Compounds (Summary) - *CC BY-NC-SA 4.0*
  - 4: Noncovalent bonds - *Undeclared*
    - 4.1: Intermolecular Interactions - *CC BY-NC-SA 4.0*
  - 5: Organic Molecules - *Undeclared*
    - 5.1: Structures and Names of Alkanes - *CC BY-NC-SA 4.0*
    - 5.2: Branched-Chain Alkanes - *CC BY-NC-SA 4.0*
    - 5.3: Condensed Structural and Line-Angle Formulas - *CC BY-NC-SA 4.0*
    - 5.4: IUPAC Nomenclature - *CC BY-NC-SA 4.0*
    - 5.5: Physical Properties of Alkanes - *CC BY-NC-SA 4.0*
    - 5.6: Alkenes- Structures and Names - *CC BY-NC-SA 4.0*
    - 5.7: Cis-Trans Isomers (Geometric Isomers) - *CC BY-NC-SA 4.0*

- 5.8: Line Notation (SMILES and InChI) - *Undeclared*

- 6: Lipids - *CC BY-NC-SA 4.0*
  - 6.1: Prelude to Lipids - *CC BY-NC-SA 4.0*
  - 6.2: Fatty Acids - *CC BY-NC-SA 4.0*
  - 6.3: Fats and Oils - *CC BY-NC-SA 4.0*
  - 6.4: Membranes and Membrane Lipids - *CC BY-NC-SA 4.0*
  - 6.5: Steroids - *CC BY-NC-SA 4.0*
  - 6.6: Exercises - *CC BY-NC-SA 4.0*
  - 6.7: Lipids (Summary) - *CC BY-NC-SA 4.0*
- 7: Conformations and Stereochemistry - *CC BY-NC-SA 4.0*
  - 7.1: Prelude to Conformations and Stereochemistry - *Undeclared*
  - 7.2: Conformations of open-chain organic molecules - *CC BY-NC-SA 4.0*
  - 7.3: Chirality and stereoisomers - *CC BY-NC-SA 4.0*
  - 7.4: Naming chiral centers- the R and S system - *CC BY-NC-SA 4.0*
  - 7.5: Optical Activity - *CC BY-NC-SA 4.0*
  - 7.6: Compounds with multiple chiral centers - *CC BY-NC-SA 4.0*
  - 7.7: Stereochemistry of alkenes - *CC BY-NC-SA 4.0*
  - 7.8: Stereochemistry in biology and medicine - *CC BY-NC-SA 4.0*
- 8: Acids and Bases - *CC BY-NC-SA 4.0*
  - 8.1: Prelude to Acids and Bases - *CC BY-NC-SA 4.0*
  - 8.2: Brønsted-Lowry Definition of Acids and Bases - *CC BY-NC-SA 4.0*
  - 8.3: Water - Both an Acid and a Base - *CC BY-NC-SA 4.0*
  - 8.4: The Strengths of Acids and Bases - *CC BY-NC-SA 4.0*
  - 8.5: Buffers - *CC BY-NC-SA 4.0*
  - 8.6: Acids and Bases (Exercises) - *CC BY-NC-SA 4.0*
  - 8.7: Acids and Bases (Summary) - *CC BY-NC-SA 4.0*
- 9: Amino Acids, Proteins, and Enzymes - *CC BY-NC-SA 4.0*
  - 9.1: Prelude to Amino Acids, Proteins, and Enzymes - *CC BY-NC-SA 4.0*
  - 9.2: Properties of Amino Acids - *CC BY-NC-SA 4.0*
  - 9.3: Reactions of Amino Acids - *CC BY-NC-SA 4.0*
  - 9.4: Peptides - *CC BY-NC-SA 4.0*
  - 9.5: Proteins - *CC BY-NC-SA 4.0*
  - 9.6: X-ray Protein Crystallography - *Undeclared*
  - 9.7: Enzymes - *CC BY-NC-SA 4.0*
  - 9.8: Enzyme Action - *CC BY-NC-SA 4.0*
  - 9.9: Enzyme Activity - *CC BY-NC-SA 4.0*
  - 9.10: Enzyme Inhibition - *CC BY-NC-SA 4.0*
  - 9.11: Enzyme Cofactors and Vitamins - *CC BY-NC-SA 4.0*
  - 9.12: Amino Acids, Proteins, and Enzymes (Exercises) - *CC BY-NC-SA 4.0*
  - 9.13: Amino Acids, Proteins, and Enzymes (Summary) - *CC BY-NC-SA 4.0*
- 10: Equilibrium and Kinetics- Making Drugs Stick - *Undeclared*
  - 10.1: Prelude--A Drug's Life (in your Body) - *Public Domain*
  - 10.2: A Quick Review of Thermodynamics and Kinetics - *CC BY-NC-SA 4.0*
  - 10.3: Catalysis - *CC BY-NC-SA 4.0*
  - 10.4: Pharmacology as an Equilibrium and Kinetics Problem - *CC BY-NC-SA 4.0*
- 11: Molecular Genetics - *CC BY-NC-SA 4.0*
  - 11.1: Prelude to Nucleic Acids - *CC BY-NC-SA 4.0*
  - 11.2: Nucleotides - *CC BY-NC-SA 4.0*
  - 11.3: Nucleic Acid Structure - *CC BY-NC-SA 4.0*
  - 11.4: Replication and Expression of Genetic Information - *CC BY-NC-SA 4.0*
  - 11.5: Protein Synthesis and the Genetic Code - *CC BY-NC-SA 4.0*
  - 11.6: Mutations and Genetic Diseases - *CC BY-NC-SA 4.0*
  - 11.7: Viruses - *CC BY-NC-SA 4.0*
  - 11.8: Nucleic Acids (Exercises) - *CC BY-NC-SA 4.0*
  - 11.9: Nucleic Acids (Summary) - *CC BY-NC-SA 4.0*
- 12: Chemical and Biological Databases - *Undeclared*
  - 12.1: Public Chemical Databases - *Undeclared*
  - 12.2: Data Organization in PubChem as a Data Aggregator - *Undeclared*
  - 12.3: Database Query Introduction - *Undeclared*
  - 12.4: PubChem Web Interfaces for Text - *Undeclared*
  - 12.5: Text Search in PubChem - *Undeclared*
  - 12.6: Additional Data Retrieval Approaches in PubChem - *Undeclared*
  - 12.7: Searching PubChem Using a Non-Textual Query - *Undeclared*
- 13: Computer-Aided Drug Design - *Undeclared*
  - 13.1: Reading - *Undeclared*
  - 13.2: How to Dock Your Own Drug - *Undeclared*
- 14: Does it work? Clinical Testing - *Undeclared*
  - 14.1: Understanding Clinical Trial Types (Video) - *Undeclared*
  - 14.2: Statistical Basics - *CC BY-SA 4.0*
  - 14.3: Random Sampling - *CC BY-SA 4.0*

- 14.4: Clinical Studies - *CC BY-SA 4.0*
- 14.5: Types of clinical interventions - *CC BY-NC 4.0*
- 14.6: Evolution of new intervention products and sequence of study phases - *CC BY-NC 4.0*
- 14.7: Interpreting Clinical Research - *CC BY-NC-SA 4.0*

15: What is it worth? Estimating Value - *Undeclared*

- 15.1: Types of economic analyses of health interventions - *CC BY-NC 4.0*
- 15.2: The Time Value of Money - *CC BY-NC-SA 3.0*
- 15.3: Calculating the Relationship of Time and Value - *CC BY-NC-SA 3.0*

16: Appendix - *Undeclared*

- 16.1: Getting started with Zotero - *Undeclared*

Back Matter - *Undeclared*

- Index - *Undeclared*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*