SS1_2025_BIS2A_FACCIOTTI



Marc Facciotti
University of California, Davis



University of California, Davis

SS1_2025_BIS2a_Facciotti

Marc Facciotti

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 03/04/2026

TABLE OF CONTENTS

Licensing

1: Readings

- Reading_01
- Reading_02
- Reading_03
- Reading_04
- Reading_05
- Reading_06
- Reading_07
- Reading_08
- Reading_09
- Reading_10
- Reading_11
- Reading_12
- Reading_13
- Reading_14
- Reading_15
- Reading_16
- Reading_17
- Reading_18
- Reading_19
- Reading_20
- Reading_21
- Reading_22
- Reading_23
- Reading_24
- Reading_25
- Reading_26

Index

Glossary

Detailed Licensing

Licensing

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

CHAPTER OVERVIEW

1: Readings

Reading_01
Reading_02
Reading_03
Reading_04
Reading_05
Reading_06
Reading_07
Reading_08
Reading_09
Reading_10
Reading_11
Reading_12
Reading_13
Reading_14
Reading_15
Reading_16
Reading_17
Reading_18
Reading_19
Reading_20
Reading_21
Reading_22
Reading_23
Reading_24
Reading_25
Reading_26

This page titled 1: Readings is shared under a not declared license and was authored, remixed, and/or curated by Marc Facciotti.

Reading_01

Learning Objectives Associated with *SS1_2025_Bis2a_Facciotti_Reading_01*

- *GC.2 Identify the locations of the three core atomic elements (electrons, neutrons, and protons) from a basic model for an atom and describe their basic properties.*
- *GC.12 Recognize the symbols and names of the six core elements in biomolecules: C, H, N, O, P, S.*
- *GC.13 Interpret (identity elements and bond types) chemical and structural formulas (including 2- and 3-dimensional, condensed, and line-angle) of biomolecules.*
- GC.8 Define and correctly use electronegativity and explain how this concept can be used to predict the types and characteristics/properties of bonds that may be formed by two atoms.
- *GC.3 Identify ionic, covalent, polar covalent bonds, hydrogen bonds, and Van der Waals interactions in different types of molecular models.*
- *GC.4 Learn to use the vocabulary of biology and also to recognize that certain terms may mean different things in different contexts. Be able to discern the meaning of a term in context.*

Biology

Biology is the scientific study of life. Studying biology is an opportunity to ask exciting questions about the world that surrounds us. It is an opportunity to probe humanity's deepest questions about our origins, our planet's history, and our connections to other living beings (big and small/extant or extinct). It is also an opportunity to dive into a world of problem solving and to think hard about solutions for improving health care, maintaining sustainable food supplies, and producing renewable energy technologies.

Studying biology also helps us to understand and solve everyday problems. For instance, understanding the biochemical reactions that transform your food into something useful for your body, how energy is stored within food items, and how this energy transfers from your food to your body, teaches you about how your diet and exercise influence your health. Understanding what the microbes which live in, on, and around us, do may help you decide whether to buy products labeled "antimicrobial" or "probiotic". Understanding the biochemical principles that explain what happens to eggs as they cook can help us understand how similar physical processes may be central to the cellular stress response and associated diseases. Understanding of the genetic and biochemical mechanisms that link genetic information to physical traits can help explain your eye color.

Studying biology even helps us understand things that are “out of this world”. For instance, understanding the requirements for life can help us look for life on Mars or deep in Earth’s crust. When we understand how to “rewire” cellular decision-making networks, we may gain the ability to regenerate functional limbs or organs from someone’s own tissue, or reprogram diseased tissues back to health. All of these examples represent a small fraction of the multitude that exist in the natural world. There are many exciting opportunities to explore. Mastering a few basic principles helps you understand and think more deeply about an array of topics. Keep this notion in mind throughout the course.

Biology: an interdisciplinary science

Questions in biology span size scales over ten orders of magnitude, from the atomic makeup and chemical behavior of individual molecules to planetary-scale systems of interacting ecologies. Whatever the scale of interest, to develop a deep and functional understanding of biology, **we must first appreciate biological concepts**. This involves integrating important ideas and tools from across the spectrum of science, including chemistry, physics, and mathematics. Biology is truly an interdisciplinary science.

The potential impacts that can come from studying biology are broad

Some people may think studying biology is only about medicine—however, it can lead to or influence many careers. Biology has applications that are both vast and wide-ranging. Applications include treating (human or other animal) patients, improving agricultural practices, developing new building materials, writing new energy policies, remedying global climate change, creating new works of art—the list goes on and on. For the curious, biology has plenty of unexplored mysteries.

As you study biology, appreciate its exciting questions and topics and be open-minded. Even though course topics may not always seem related at first, they likely are, in the end. Being open-minded helps you discover and appreciate the connections

between the course's topics and your interests. Discovering how seemingly different topics interrelate can give you a deeper appreciation for the things you enjoy and maybe even spark a new passion.

General Biology—from molecules to cells

BIS2A focuses on the cell, one of the most fundamental units of life. Cells can be as simple as the disease-causing bacterium Mycoplasma genitalium, whose genome encodes just 525 genes (only 382 of which are essential for life), or as complex as a cell belonging to the multicellular plant Oryza sativa (rice), whose genome likely encodes ~51,000 genes. However, despite this diversity, all cells share fundamental properties. In BIS2A, we explore basic problems that all cells must deal with. We study the building blocks of cells, some of their key biochemical properties, the encoding and expression of biological information in genetic material, and how all this combines to make a living system. We will also discuss ways in which living systems exchange matter, energy, and information with their environment (including other living things). We focus primarily on core principles common to all life on Earth, and because of biology's massive breadth, we put these ideas into a variety of contexts throughout the quarter.

Evolution and Natural Selection

Brief overview



Evolution and **natural selection** are core concepts in biology that are typically invoked to help explain the diversity of and relationships between life on Earth, both extant and extinct.

The first idea you need to grasp is that evolution can be simply defined as the development/change of something over time. In the automotive industry, the shapes and features of cars can be said to evolve (change in time). In fashion, it can be said that style evolves. In biology, life and, in particular, **reproducing populations** of organisms with **different traits** evolve.

The second thing to understand is that natural selection is a process by which nature filters organisms in a population. What is the filter? Here it becomes a little more complicated (but only a little). The simplest explanation is that the selective filter is just a combination of all living and nonliving factors in an environment, which influence how successfully an organism can reproduce. The factors that influence the ability of an organism to reproduce are known as **selective pressures**. A small but important complication is that these factors are not the same everywhere; they change in time and by location. Thus, the selective pressures that create the filter are constantly changing (sometimes rapidly, sometimes slowly), and organisms in the same **reproducing population** could experience different pressures at different times and in different locations.

The theory of evolution by natural selection puts these two ideas together; it stipulates that change in biology happens over time and that the variation in a population is constantly subjected to selection based on how differences in traits influence reproduction. But what are these characteristics or **traits**? What traits/features/functions can be subject to selection? The short answer is: just about anything associated with an organism for which variation exists in a population and for which this variation leads to a differential likelihood of generating offspring will probably be subject to filtering by natural selection. We also call these traits heritable **phenotypes**. Organisms in a population that have phenotypes, which enable them to pass the selective filter more efficiently than others, are said to have a **selective advantage** and/or **greater fitness**.

It is important to reiterate that while the phenotypes carried by individual organisms may be subject to selection, the process of evolution by natural selection both requires and acts on **phenotypic variation** within **populations**. If neither variation nor populations in which that variation can reside exist, there is no opportunity or need for selection. Everything is and stays the same.

Common misconceptions and a course specific note

Finally, we draw your attention to a critical point and common misconception among beginning students in biology. This misconception can arise when, for the sake of discussion, we decide to anthropomorphize nature by giving it an intellect. For example, we may try to build an example for evolution by natural selection by proposing that a surplus of a particular food exists in an environment and there is an organism close by that is starving. It would be correct to reason that if the organism could eat that

food, that this might give it a selective advantage over other organisms that cannot. If later we find examples of organisms that can eat that surplus food, it might be tempting to say that nature evolved this new phenotype to "solve the problem of utilizing the surplus food". The process of evolution by natural selection, however, happens randomly and without direction. That is, nature does NOT identify “problems” that are limiting fitness. Nature does NOT identify features that would make an organism more successful and then start creating diverse solutions that meet this need. The generation of variation is not guided. Variation happens and natural selection filters what works best. The observation that an organism exists that can eat the surplus food is not a reflection of nature actively solving a problem, but rather, a reflection of whatever processes that led to phenotypic variation in an ancestral population that created—among many other variants—a phenotype that increased fitness (possibly because the ancestral organisms were able to eat the surplus food).

This point of the preceding paragraph is particularly important to understand in the context of general biology at UC Davis because of the way we will be utilizing the Design Challenge to understand biology (more on this later in the reading). While the Design Challenge is intended to help focus our attention on functions under selection and their relationship to determining fitness, it can be easy—if we aren’t attentive—to lapse into language that would suggest that nature purposefully designs solutions to solve specific problems. Always remember that we are looking retrospectively at what nature has selected and that we are attempting to understand why a specific phenotype may have been selected over many other possibilities. In doing so, we will be inferring or hypothesizing to the best of our ability (which is sometimes wrong) a sensible reason to explain why a phenotype might have provided a selective advantage. We are NOT saying that the phenotype evolved TO provide a specific selective advantage. The distinction between these two ideas may be subtle, but it is critical!

Possible NB Discussion Point

Examine the following statement: "*Natural selection acts for the good of the species*." Discuss what you think about this statement - perhaps invoking some of the reading above.

Introduction to Biomolecules for General Biology

Some context and motivation

Many general biology courses are concerned with getting students to start learning how living systems "work". To accomplish this we need to understand the "stuff" biological systems. We want to know what "stuff" organisms are made of and how the properties of this "stuff" contribute to organismal forms and to their behaviors. This requires us to dig into a little chemistry — the science of the "stuff" that makes up the world we know. By doing so, we can show that many of the biological processes we care about arise directly from the chemical properties of the "stuff" that makes up living systems. Developing a functional understanding of some basic chemical concepts can, therefore, be tremendously useful in thinking about how to solve problems in medicine, energy, and environment by attacking them at their molecular levels. For these reasons, we begin a study of biomolecules and their chemistry.

Importance of chemical composition

Most instructors of general biology will want you to learn how to classify major classes of biological macromolecules into groups by looking at their chemical composition and, based on this composition, to also infer some properties they might have. For example, you will learn to recognize the molecular structures of carbohydrates. You'll learn to do so by spotting chemical "parts" on carbohydrates called hydroxyl groups and appreciate that those chemical "parts" can form special types of interactions called "hydrogen bonds". You'll discover how hydrogen bonds give carbohydrates special properties that govern how they interact with water and other molecules and how these interactions give different properties to carbohydrates like simple sugars

and starches. This will shed light on the functional roles these molecules play in living systems, how they are built, and how they can be broken down. We will also look at other types of common macromolecules in biology in the same way.

Linking molecular structure to function

In biology, structure is often said to "encode" function. That is, the shape and physical properties of something governs what it can do. We'll want to make this general idea a bit more specific for you and begin to help you form the related conceptual ideas that will help you explain how the chemical structure and properties of biomolecules can contribute to their functions.

For example, biological membranes have the biological function of defining the boundaries between the inside and outside of a cell or between the inside and outside of compartments within a cell. Membranes also serve as selective barriers that allow certain things to pass into and out of a cell while excluding others substances from crossing the barrier. These are very important functions indeed! The chemical structures of membranes determine these functional properties. You will learn, for instance, that most membranes are composed of two layers of molecules called phospholipids and that phospholipids can be schematically broken down into two chemically distinct subregions, a so-called hydrophilic head group and a hydrophobic tail group. Each of these chemical subregions and the chemical "parts" they are made of, contribute directly to determining the structure and permeability of cell membranes.

Likewise, you will also learn about important biological molecules called enzymes. Enzymes help catalyze many of the chemical reactions in an organism, including how they process the food they eat, how an organism senses and interacts with its environment, and how it replicates (among many other functions). We will learn that the three dimensional structures of enzymes encode their functions and that their structures are determined to a large degree by the number, type, and order of chemicals called amino acids. The properties of these amino acids (and thus how they behave in an enzyme) are directly related to their own chemical "parts".

In sum, while structure "encodes" biological function, structure is itself usually determined by chemistry. Knowing a little bit of chemistry can be powerful in biology.

Objectives and strategy

On these notes, we set out to learn a little bit of useful chemistry for biology.

- First, we want to learn some basic vocabulary, symbols and models that help us communicate about chemistry in biology.
- Second, we want to learn a few basic ideas about what the "stuff-of-life" is made of and some basic properties of that stuff that help us understand biological phenomena.
- Third, we want to learn to recognize some of the most common chemical structures in biology and something about their properties.
- Fourth, we want to understand how some of the chemistries of the "stuff-of-life" can be altered and how this might change the function of a biomolecule.
- Fifth, we want to develop a vocabulary and learn key ideas that help us talk about how things change at the molecular level in biology.

If you've already taken some chemistry, some of the ideas will be familiar to you. If you haven't taken chemistry - or it's been a long time since you did - don't worry. In general biology, these topics aren't typically covered as theoretically or in as much detail as they might be in a formal chemistry class. Here, we want to extract the most important ideas/concepts that help us function as general biologists. The core ideas and concepts aren't too complicated. However, there will likely be a lot of new vocabulary to learn and a good amount of practice required to get comfortable thinking about and building mental models of things that happen at the molecular scale. Success comes from practice. So, practice early and often. Strive to learn the new vocabulary quickly and take advantage of practice opportunities to think about molecules. That'll make it much easier to then think about and add new concepts as we go move forward in the course.

The Periodic Table

The periodic table is a tool for organizing information about the different elements found in nature. Devised by Russian chemist Dmitri Mendeleev (1834–1907) in 1869, the table groups elements that, because of some commonalities of their atomic structure, share certain chemical properties. The atomic structure of elements is responsible for their physical properties including whether they exist as gases, solids, or liquids under specific conditions and their chemical reactivity, a term that refers to their ability to combine and to bond chemically with each other and other elements.

In the periodic table, shown below, the elements are organized and displayed according to their atomic number and are arranged in a series of rows and columns based on shared chemical and physical properties. Besides providing the atomic number for each element, the periodic table also displays the element's atomic mass. Looking at carbon, for example, its symbol (C) and name appear, and its atomic number of six (in the upper right-hand corner showing the number of protons in the neutral nucleus) and its atomic mass of 12.11 (sum of the mass of electrons, protons, and neutrons).

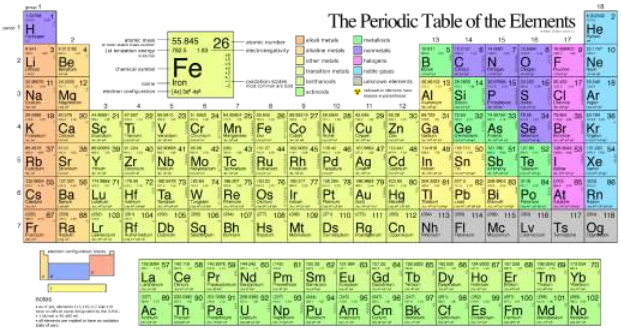


Figure: The periodic table shows the atomic mass and atomic number of each element. The atomic number appears above the symbol for the element and the approximate atomic mass appears to the left.
Source: By 2012rc (self-made using inkscape) [Public domain], via Wikimedia Commons Modified by Marc T. Facciotti - 2016

Electronegativity

MCAT Connection

Molecules are collections of atoms that associate with one another through bonds. It is reasonable to expect — and true empirically — that different atoms will exhibit different physical properties, including abilities to interact/bond with other atoms. We describe one such property, the tendency of an atom to attract electrons, by the chemical concept and term, electronegativity. While chemists have developed several methods for measuring electronegativity, Linus Pauling created the one most commonly taught to biologists.

A description of how Pauling electronegativity can be calculated is typically beyond the scope of introductory biology. It is important to know, however, that electronegativity values have been experimentally and/or theoretically determined for nearly all elements in the periodic table. The values are unitless. The larger the electronegativity value, the greater the tendency an atom has to attract electrons. Using this scale, one can quantitatively compare the electronegativity of different atoms. For instance, by using Table 1 below, you could report that oxygen atoms (O) are more electronegative than phosphorous atoms (P).

Element	Pauling Electronegativity
F	3.98
O	3.44
Cl	3.16
N	3.04
S	2.58
C	2.55
H	2.20
P	2.19
Na	0.93
Fr	0.70

Table 1. *Pauling electronegativity values for select elements of relevance to most general biology courses as well as elements at the two extremes (highest and lowest) of the electronegativity scale.*
Attribution: Marc T. Facciotti (original work)

The utility of the Pauling electronegativity scale in general biology is to provide a chemical basis for explaining the bonds that form between the commonly occurring elements in biological systems and to explain some key interactions that we observe routinely. We develop our understanding of electronegativity-based arguments about bonds and molecular interactions by comparing the electronegativities of two atoms. Recall, the larger the electronegativity, the stronger the "pull" an atom exerts on nearby electrons.

We can consider, for example, the common interaction between oxygen (O) and hydrogen (H). Let us assume that O and H are interacting (forming a bond) and write that interaction as O-H, where the dash between the letters represents the interaction between the two atoms. To understand this interaction better, we can compare the relative electronegativity of each atom. Examining the table above, we see that O has an electronegativity of 3.44, and H has an electronegativity of 2.20.

Based on the concept of electronegativity we can surmise that the oxygen (O) atom will tend to "pull" the electrons away from the hydrogen (H) when they are interacting. This will give rise to a slight but significant partial negative charge around the O atom (because of the higher tendency of the electrons to associate with the O atom). This also results in a slight partial positive charge around the H atom (because of the decrease in the probability of finding an electron around the hydrogen atom). Since the electrons distribute unevenly between the two atoms AND, by consequence, the electric charge also distributes unevenly, we describe this interaction or bond as **polar**. There are two poles in effect: the more negative pole near the oxygen and the more positive pole near the hydrogen.

To extend the utility of this concept, we can now ask how an interaction between oxygen (O) and hydrogen (H) differs from an interaction between sulfur (S) and hydrogen (H). That is, how does O-H differ from S-H? If we examine the table above, we see that the difference in electronegativity between O and H is 1.24 (3.44 - 2.20 = 1.24) and that the difference in electronegativity between S and H is 0.38 (2.58 – 2.20 = 0.38). We can therefore conclude that an O-H bond is more polar than an S-H bond. We will discuss the consequences of these differences in subsequent chapters.

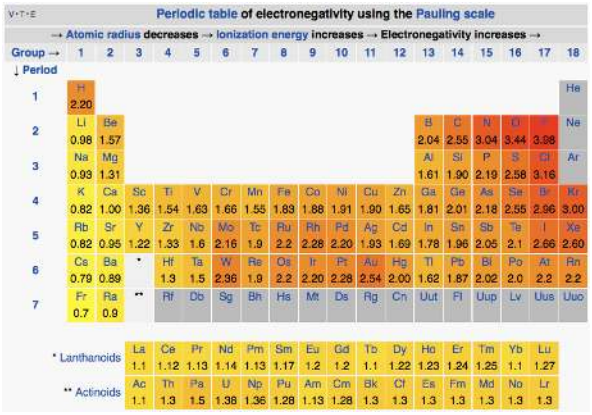


Figure 2. *The periodic table with the electronegativities of each atom listed.*
Attribution: By DMacks (https://en.wikipedia.org/wiki/Electronegativity) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

An examination of the periodic table of the elements (Figure 2) illustrates the relationship between electronegativity and some physical properties used to organize the elements into the table. Certain trends are plain. For instance, those atoms with the largest electronegativity tend to reside in the upper right-hand corner of the periodic table, such as Fluorine (F), Oxygen (O) and Chlorine (Cl), while elements with the smallest electronegativity tend to be found at the other end of the table, in the lower left, such as Francium (Fr), Cesium (Cs) and Radium (Ra).

You can find more information on electronegativity in the Chemistry LibreTexts.

The main use of the concept of electronegativity in general biology will therefore be to provide a conceptual grounding for discussing the different types of chemical interactions that occur between atoms in nature. We will focus primarily on **covalent bonds,** and several non-covalent interactions called **ionic bonds, hydrogen bonds, and Van der Waals forces**.

Covalent bonds and non-covalent molecular interactions

In general biology, we focus primarily on **covalent bonds**, **ionic bonds**, **hydrogen bonds** and Van der Waals forces. We expect students to be able to recognize each different bond type in molecular models. In addition, for commonly seen bonds in biology, we expect student to provide a chemical explanation, rooted in ideas like electronegativity, for how these bonds contribute to the chemistry of biological molecules.

MCAT Connection

Ionic bonds

Ionic bonds are electrostatic interactions formed between ions of opposite charges. For instance, in Chemistry we learn that in sodium chloride (NaCl) positively charged sodium ions and negatively charged chloride ions associate via electrostatic (+ attracts -) interactions to make crystals of sodium chloride, or table salt, creating a crystalline molecule with zero net charge. The origins of these interactions may arise from the association of neutral atoms whose difference in electronegativities is sufficiently high. If we imagine that a neutral sodium atom and a neutral chlorine atom approach one another, it is possible that at close distances, due to the relatively large difference in electronegativity between the two atoms, that an electron from the neutral sodium atom is transferred to the neutral chlorine atom, resulting in a negatively charged chloride ion and a positively charged sodium ion. These ions can now interact via an ionic bond.

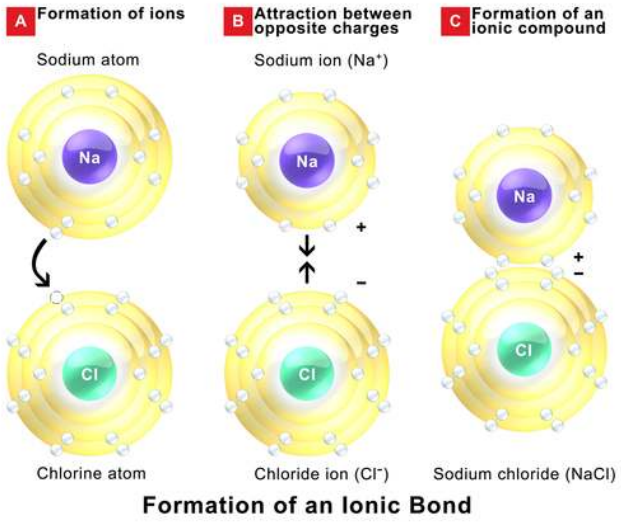
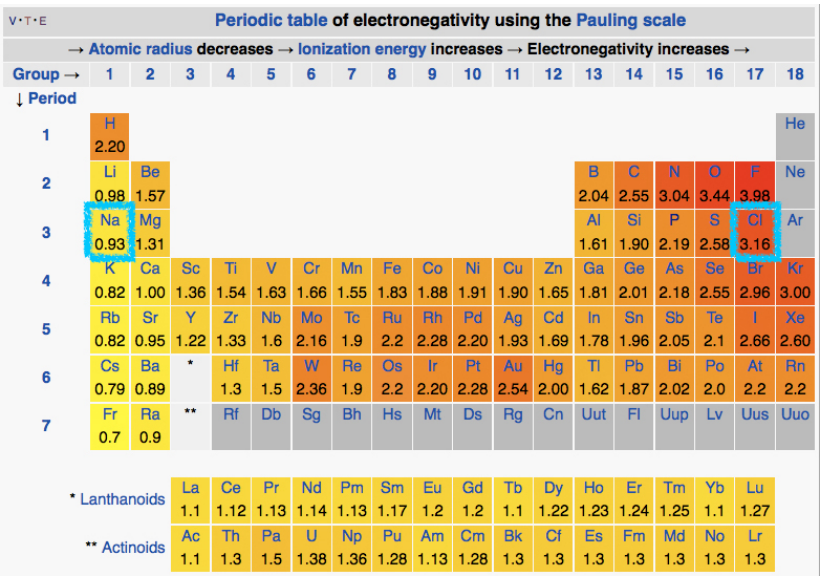


Figure 1. *The formation of an ionic bond between sodium and chlorine is depicted. In panel A, a sufficient difference in electronegativity between sodium and chlorine induces the transfer of an electron from the sodium to the chlorine, forming two ions, as illustrated in panel B. In panel C, the two ions associate via an electrostatic interaction. Attribution: By Bruce Blaus (own work) [CC BY-SA 4.0 (http://creativecommons.org/licenses/by-sa/4.0)], via Wikimedia Commons*

This movement of electrons from one atom to another is referred to as electron transfer. In the example above, when sodium loses an electron, it now has 11 protons, 12 neutrons, and 10 electrons, leaving it with an overall charge of +1 (summing charges: 11 protons at +1 charge each and 10 electrons at -1 charge each = +1). Once charged, the sodium atom is referred to as a sodium ion. Likewise, based on its electronegativity, a neutral chlorine (Cl) atom tends to gain an electron to create an ion with 17 protons, 18 neutrons, and 18 electrons, giving it a net negative (–1) charge. It is now referred to as a chloride ion.

We can interpret the electron transfer above using the concept of electronegativity. Begin by comparing the electronegativities of sodium and chlorine by examining the periodic table of elements below. We see that chlorine is located in the upper-right corner of the table, while sodium is in the upper left. Comparing the electronegativity values of chlorine and sodium directly, we see that the chlorine atom is more electronegative than is sodium. The difference in the electronegativity of chlorine (3.16) and sodium (0.93) is 2.23 (using the scale in the table below). Given that we know an electron transfer will take place between these two elements, we can conclude that differences in electronegativities of ~2.2 are large enough to cause an electron to transfer between two atoms and that interactions between such elements are likely through ionic bonds.



Periodic table of electronegativity using the Pauling scale

→ Atomic radius decreases → Ionization energy increases → Electronegativity increases →

Group →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
↓ Period																		
1	H 2.20																	He
2	Li 0.98	Be 1.57											B 2.04	C 2.55	N 3.04	O 3.44	F 3.98	Ne
3	Na 0.93	Mg 1.31											Al 1.61	Si 1.90	P 2.19	S 2.58	Cl 3.16	Ar
4	K 0.82	Ca 1.00	Sc 1.36	Ti 1.54	V 1.63	Cr 1.66	Mn 1.55	Fe 1.83	Co 1.88	Ni 1.91	Cu 1.90	Zn 1.65	Ga 1.81	Ge 2.01	As 2.18	Se 2.55	Br 2.96	Kr 3.00
5	Rb 0.82	Sr 0.95	Y 1.22	Zr 1.33	Nb 1.6	Mo 2.16	Tc 1.9	Ru 2.2	Rh 2.28	Pd 2.20	Ag 1.93	Cd 1.69	In 1.78	Sn 1.96	Sb 2.05	Te 2.1	I 2.66	Xe 2.60
6	Cs 0.79	Ba 0.89	*	Hf 1.3	Ta 1.5	W 2.36	Re 1.9	Os 2.2	Ir 2.20	Pt 2.28	Au 2.54	Hg 2.00	Tl 1.62	Pb 1.87	Bi 2.02	Po 2.0	At 2.2	Rn 2.2
7	Fr 0.7	Ra 0.9	**	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Uut	Fl	Uup	Lv	Uus	Uuo

* Lanthanoids	La 1.1	Ce 1.12	Pr 1.13	Nd 1.14	Pm 1.13	Sm 1.17	Eu 1.2	Gd 1.2	Tb 1.1	Dy 1.22	Ho 1.23	Er 1.24	Tm 1.25	Yb 1.1	Lu 1.27
** Actinoids	Ac 1.1	Th 1.3	Pa 1.5	U 1.38	Np 1.36	Pu 1.28	Am 1.13	Cm 1.28	Bk 1.3	Cf 1.3	Es 1.3	Fm 1.3	Md 1.3	No 1.3	Lr 1.3

Figure 2. *The periodic table of the elements listing electronegativity values for each element. The elements sodium and chlorine are boxed with a teal boundary. Attribution: By DMacks (https://en.wikipedia.org/wiki/Electronegativity) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons—Modified by Marc T. Facciotti*

Possible NB Discussion Point

The atoms in a 5 in. x 5 in. brick of table salt (NaCl) sitting on your kitchen counter are held together almost entirely by ionic bonds. Based on that observation, how would you characterize the strength of ionic bonds? Now consider that same brick of table salt after having been thrown into an average backyard swimming pool. After a couple of hours, the brick would be completely dissolved, and the sodium and chloride ions would be uniformly distributed throughout the pool. What might you conclude about the strength of ionic bonds from this observation? Propose a reason why NaCl's ionic bonds seemingly behave differently in air and water? What is the significance of this observation to biology?

Covalent bonds

We can also invoke the concept of electronegativity to help describe the interactions between atoms that have differences in electronegativity too small for the atoms to form an ionic bond. These types of interactions often result in a bond called a covalent bond. In these bonds, electrons are shared between two atoms—in contrast to an ionic interaction in which electrons remain on each atom of an ion or are transferred between species that have highly different electronegativities.

We start exploring the covalent bond by looking at an example where the difference in electronegativity is zero. Consider a very common interaction in biology, the interaction between two carbon atoms. In this case, each atom has the same electronegativity, 2.55; the difference in electronegativity is therefore zero. If we build our mental model of this interaction using the concept of electronegativity, we realize that each carbon atom in the carbon-carbon pair has the same tendency to "pull" electrons to it. In this case, when a bond is formed, neither of the two carbon atoms will tend to "pull" (a good anthropomorphism) electrons from the other. They will "share" (another anthropomorphism) the electrons equally, instead.

Aside: bounding example

The two examples above—(1) the interaction of sodium and chlorine, and (2) the interaction between two carbon atoms—frame a discussion by "bounding," or asymptotic analysis (see earlier reading). We examined what happens to a physical system when considering two extremes. In this case, the extremes were in electronegativity differences between interacting atoms. The interaction of sodium and chlorine illustrated what happens when two atoms have a large difference in electronegativities, and the carbon-carbon example illustrated what happens when that difference is zero. Once we create those mental goal posts describing what happens at the extremes, it is then easier to imagine what might happen in between—in this case, what happens when the difference in electronegativity is between 0 and 2.2. We do that next.

When the sharing of electrons between two covalently bonded atoms is nearly equal, we call these bonds nonpolar covalent bonds. If by contrast, the sharing of electrons is not equal between the two atoms (likely due to a difference in electronegativities between the atoms), we call these bonds **polar covalent** bonds.

In a **polar covalent** bond, the electrons are unequally shared by the atoms and are attracted to one nucleus more than to the other. Because of the unequal distribution of electrons between atoms in a polar covalent bond, a slightly positive (indicated by $\delta+$) or slightly negative (indicated by

δ–) charge develops at each pole of the bond. The slightly positive (δ+) charge will develop on the less electronegative atom, as electrons get pulled more towards the slightly more electronegative atom. A slightly negative (δ–) charge will develop on the more electronegative atom. Since there are two poles (the positive and negative poles), the bond is said to possess a **dipole**.

Examples of nonpolar covalent and polar covalent bonds in biologically relevant molecules

Nonpolar covalent bonds

Molecular oxygen

Molecular oxygen (O_2) is made from an association between two atoms of oxygen. Since the two atoms share the same electronegativity, the bonds in molecular oxygen are nonpolar covalent.

Methane

Another example of a nonpolar covalent bond is the C-H bond found in the methane gas (CH_4). Unlike the case of molecular oxygen where the two bonded atoms share the same electronegativity, carbon and hydrogen do not have the same electronegativity; C = 2.55 and H = 2.20—the difference in electronegativity is 0.35.

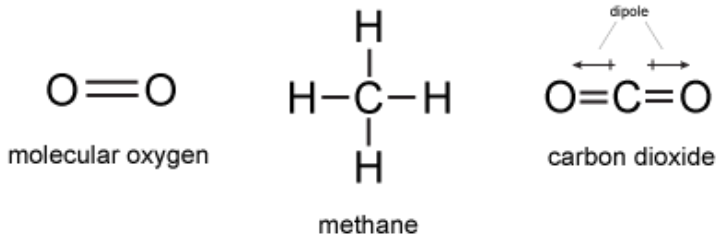


Figure 3. *Molecular line drawings of molecular oxygen, methane, and carbon dioxide. Attribution: Marc T. Facciotti (own work)*

Some of you may now be confused. If there is a difference in electronegativity between the two atoms, is the bond not, by definition, polar? The answer is both yes and no. It depends on the definition of polar that the speaker/writer is using. Since this is an example of how taking shortcuts in the use of specific vocabulary can sometimes lead to confusion, we take a moment to discuss this here. See the mock exchange between a student and an instructor below for clarification:

1. Instructor: "In biology, we often say that the C-H bond is nonpolar."
2. Student: "But there is an electronegativity difference between C and H, so it would appear that C should have a slightly stronger tendency to attract electrons. This electronegativity difference should create a small, negative charge around the carbon and a small, positive charge around the hydrogen."
3. Student: "Since there is a non-uniform distribution of charge across the bond, it would seem that, by definition, this should be considered a polar bond."
4. Instructor: "In fact, the bond does have some small polar character."
5. Student: "So, then it's polar? I'm confused."
6. Instructor: "It has some small amount of polar character, but it turns out that for most of the common chemistry that we will encounter in biology that this small amount of polar character is insufficient to lead to "interesting" chemistry. So, while the bond is, strictly speaking, slightly polar, from a practical standpoint it is effectively nonpolar. We therefore call it nonpolar."
7. Student: "That's needlessly confusing; how am I supposed to know when you mean strictly 100% nonpolar, slightly polar, or functionally polar when you use the same word to describe two of those three things?"
8. Instructor: "Yup, it sucks. The fix is that I need to be as clear as I can when I talk with you about how I am using the term "polarity." I also need to inform you that you will find this linguistic shortcut (and others) used when you go out into the field, and I encourage you to start learning to recognize what is intended by the context of the conversation.

A real-world analogy of this same problem might be the use of the word "newspaper". It can be used in a sentence to refer to the company that publishes some news, OR it can refer to the actual item that the company produces. In this case, the disambiguation is easily made by native English speakers, as they can determine the correct meaning from the context; non-native speakers may be more confused. Don't worry. As you see more examples of technical word use in science, you'll learn to read then intended meanings from contexts too."

Aside:

How large should the difference in electronegativity be in order to create a bond that is "polar enough" that we decide to call it polar in biology? Of course, the exact value depends on a number of factors, but as a loose rule of thumb, we sometimes use a difference of 0.4 as a guesstimate.

This extra information is purely for your information. You will usually not be asked to assign polarity based on this guesstimate in a general biology course (though you may want to make sure with your individual instructor). You should, however, appreciate the concept of how

polarity can be determined by using the concept of electronegativity. You should also appreciate the functional consequences of polarity (more on this in other sections) and the nuances associated with these terms (such as those in the discussion above).

Polar covalent bonds

The **polar covalent** bond can be illustrated by examining the association between O and H in water (H_2O). Oxygen has an electronegativity of 3.44, while hydrogen has an electronegativity of 2.20. The difference in electronegativity is 1.24. It turns out that this size of electronegativity difference is large enough that the dipole across the molecule contributes to chemical phenomenon of interest.

This is a good point to mention another common source of student confusion regarding the use of the term polar. Water has polar bonds. This statement refers specifically to the individual O-H bonds. Each of these bonds has a dipole. However, students will also hear that water is a polar molecule. This is also true. This latter statement is referring to the fact that the sum of the two bond dipoles creates a dipole across the whole molecule. However, it is also true that a molecule may be nonpolar but still have some polar bonds. The typical example given to illustrate this case is that of carbon dioxide (CO_2) - this molecule is shown in the figure above. While the CO_2 molecule has two polar C-O bonds, these diploes are equal in magnitude and point in opposite directions. When the bond dipoles in CO_2 are added together to determine the molecule's dipole they cancel one another. This leads a molecule with no dipole even though it has individual bonds that are polar.

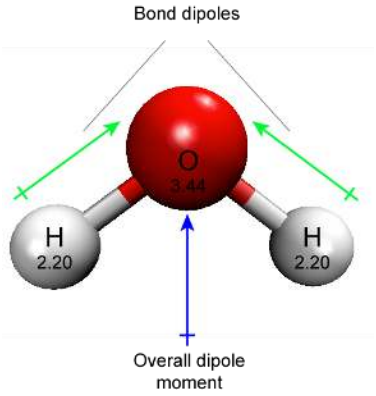


Figure 4. *A water molecule has two polar O-H bonds. Since the distribution of charge in the molecule is asymmetric (due to the number and relative orientations of the bond dipoles), the molecule is also polar. The element name and electronegativities are reported in the respective sphere. Attribution: Marc T. Facciotti (own work)*

The continuum of bonds between covalent and ionic

The discussion of bond types above highlights that in nature you will see bonds on a continuum from completely nonpolar covalent to purely ionic, depending on the atoms that are interacting. As you proceed through your studies, you will further discover that in larger, multi-atom molecules, the localization of electrons around an atom is also influenced by multiple factors. For instance, other atoms that are also bonded nearby will exert an influence on the electron distribution around a nucleus in a way that is not easily accounted for by invoking simple arguments of pairwise comparisons of electronegativity. Local electrostatic fields produced by other non-bonded atoms may also have an influence. Reality is always more complicated than are our models. However, if the models allow us to reason and predict with "good enough" precision or to understand some key underlying concepts that can be extended later, they are quite useful.

Key bonds in general biology

In general biololgy, we are usually concerned with the chemical behavior of and bonds between atoms in biomolecules. Fortunately, biological systems are composed of a relatively small number of common elements (e.g., C, H, N, O, P, S, etc.) and some key ions (e.g., Na+, Cl-, Ca^{2}+, K+, etc.). Start recognizing commonly occurring bonds and the chemical properties that we often see them showing. Some common bonds include C-C, C-O, C-H, N-H, C=O, C-N, P-O, O-H, S-H, and some variants. These will be discussed further in the context of functional groups. The task is not as daunting as it seems.

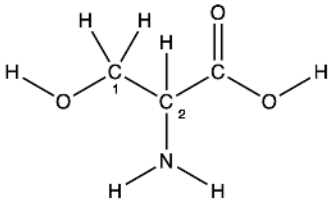
Note: Common Point of student confusion

In this reading we have been talking about the polarity of bonds. That is, we have been learning how to describe the polarity of a single bond joining two atoms (i.e. how are the electrons shared between two atoms distributed about the respective nuclei?). In biology we also sometimes talk about the polarity of a molecule. The polarity of a molecule is different than the polarity of a bond within the molecule. The latter is asking whether the whole molecule has a net dipole. The molecule's dipole can be roughly thought of as the sum of all of its bond dipoles. For example, let us examine a molecule of CO2 depicted in the figure above. If we ask whether one of the C=O bonds is polar we would conclude that it is since the oxygen is significantly more electronegative that the carbon to which it is covalently bonded. However, if we ask whether the molecule O=C=O is polar we would concluded that it is not. Why? Look at the figure of CO2 above. Each CO bond has a dipole. However, these two dipoles are pointed in directly opposite directions. If we add these two bond dipoles together to get the net dipole of the

molecule we get nothing - the two bond dipoles "cancel" one another out. By contrast, if we examine the structure of water above, we also see that each O-H bond has a dipole. In this case when we ask whether the molecule has a net dipole (done by adding the bond dipoles together) we see that the answer is yes. The sum of the the two bond dipoles still yields a net dipole moment. We therefore say that this molecule is polar. We can do this same exercise for parts of molecules so long as we define what specific part we are looking at.

Possible NB Discussion Point

Imagine that you were able to shrink yourself down to the size of an atom and see things like electrons and protons. Describe what you would see if you were standing on Carbon 1 in the molecule below and looking in different directions towards the bound oxygen, hydrogens or carbon 2. Compare and contrast what you expect to see along each bond.



Hydrogen Bonds

When hydrogen forms a polar covalent bond with an atom of higher electronegativity, the region around the hydrogen will have a fractional positive charge (termed δ^+). When this fractional positive charge encounters a partial negative charge (termed δ^-) from another electronegative atom to which the hydrogen is NOT bound, AND it is presented to that negative charge in a suitable orientation, a special kind of interaction called a hydrogen bond can form. While chemists are still debating the exact nature of the hydrogen bond, in general biology, we typically conceive of it as a weak electrostatic interaction between the δ^+ of the hydrogen and the δ^- charge on an electronegative atom. We call the molecule that contributes the partially charged hydrogen atom the "**hydrogen bond donor**" and the atom with the partial negative charge the "**hydrogen bond acceptor**." We will ask you to learn to recognize common biological hydrogen bond donors and acceptors and to identify putative hydrogen bonds from models of molecular structures.

Hydrogen bonds are common in biology both within and between many biomolecules. Hydrogen bonds are also critical interactions between biomolecules and their solvent, water. It is common, as seen in the figure below, to represent hydrogen bonds in figures with dashed lines.

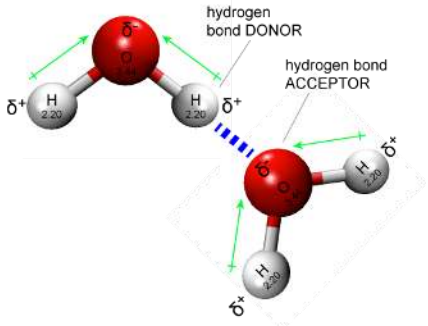


Figure 1: *Two water molecules are depicted forming a hydrogen bond (drawn as a dashed blue line). The water molecule on top "donates" a partially charged hydrogen while the water molecule on the bottom accepts that partial charge by presenting a complementary negatively charged oxygen atom. Attribution: Marc T. Facciotti (original work)*

Dipoles, Van der Waals Forces, and Pi Interactions

In addition to ionic and hydrogen bonds, there are several other types of non-covalent molecular interactions that we encounter in biology. Key among these are dipole-dipole interactions, Van der Waals forces and pi interactions. In this section we briefly describe each of these interactions and some of their underlying basis. Developing a deep and comprehensive theoretical understanding of these interaction types requires a dive into more advanced chemistry. We don't do that. Rather, we try to provide a more descriptive understanding of these phenomena that will hopefully be useful for interpreting common molecular interactions in biology. Recall that with respect to chemistry, our goals in general biology are relatively modest. We want students to recognize different chemical interactions between biomolecules, to appreciate that these interactions arise from the unique chemical properties of the elements that make up the molecules, and to appreciate how environmental and chemical factors can change these interactions. If you can identify obvious biological scenarios in which different interactions can take place, you're doing great!

Dipole-Dipole Interactions

Dipole-dipole interactions are, as the name suggests, simply interactions between two dipoles. Recall how the differences in electronegativities between elements can explain the creation of polar covalent bonds. We describe these polar covalent bonds as permanent dipoles, "hard-coded" by the properties of the elements bonded together. The dipole-dipole interaction is an interaction between two permanent dipoles. If partial charges carrying the same (+ or -) sign interact (i.e. positive interacts with positive or negative interacts with negative) we say that the interacting molecules experience a *repulsive* dipole-dipole force which pushes the molecules away from one another. If partial charges of opposite sign (i.e. positive interacts with negative) interact, we say that the interacting molecules experience an *attractive* dipole-dipole force which attracts the molecules to one another.

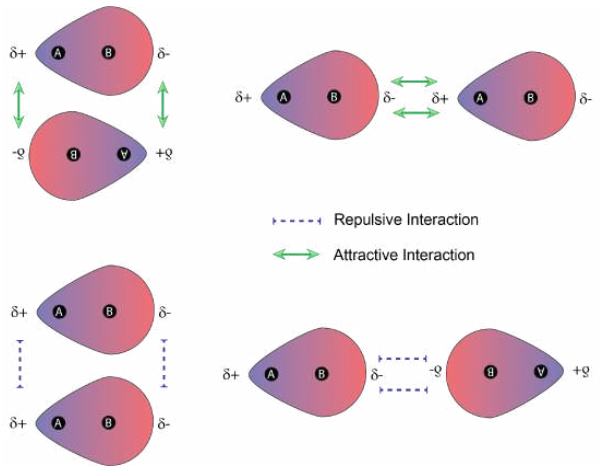


Figure 1: Dipole-dipole interactions. *Permanent dipoles established through the covalent interaction between atoms A and B can interact via dipole-dipole interactions. Atom B is more electronegative than atom A and thus recruits electrons near it causing an imbalance in charge around the molecule (this is depicted by an oblong electron cloud around nuclei for atoms A and B) - more negative is red; more positive is blue. The imbalance of charge creates a dipole with partial negative charges (delta-) and partial positive charges (delta+). The figure depicts various ways in which two of these dipoles can interact with one another that, depending on orientation, lead to either attractive or repulsive dipole-dipole interactions.*
Attribution: Marc T. Facciotti (original work)

The core idea underlying the formation of dipole-dipole interactions - the interaction between two permanent dipoles - should sound familiar. Recall that we describe a hydrogen bond as an electrostatic interaction between a partially charged hydrogen (the positive end of a bond dipole) with a partial negative charge from the negative end of a different bond dipole. While the hydrogen bond has some special properties not discussed in this text, you can think of it as a sub-type of attractive dipole-dipole interaction.

For a deeper dive into dipole-dipole interactions, see this LibreText Chemistry reading.

Van der Waals Forces

All molecules can experience Van der Waals forces, a type of molecular interaction found when molecules get very close together, typically at distances between 4-5 Angstroms. Just for reference, recall that 1 Angstrom = 10^{-10} meters. Van der Waals forces are, yet again, based on the attraction or repulsion of electrical poles. However, unlike the dipole-dipole interactions discussed above that arise from the interaction between permanent dipoles in molecules, the Van der Waals forces arise from the spontaneous and/or induced transient polarization of molecules. The local polarization (i.e. polarization on a part of a molecule) may last only a short time as electrons dynamically redistribute. When two molecules are close together one of them may spontaneously form a transient dipole (or more accurately, multipole). In response, the second molecule may "sense" the partial charge nearby and react by adjusting its own charge distribution in response, thus becoming polar itself. We say that the first dipole/multipole induces the formation of the second. If the two molecules remain between 4 and 5 Angstroms long enough, this process can repeat, and even synchronize, leading to molecular attraction at very short distances. At distances closer than ~4 Angstroms, electron clouds can overlap and this creates repulsive interaction. While all molecules can engage in Van der Waals interactions, in introductory biology we usually introduce students to these interactions in a discussion of lipid membrane structures. As you will soon discover, Nature creates biological membranes by packing many lipid molecules together at distances that allow many simultaneous Van der Waals interactions to occur. Collectively, these many small and transient interactions between lipid molecules contribute to the stability of membrane structures.

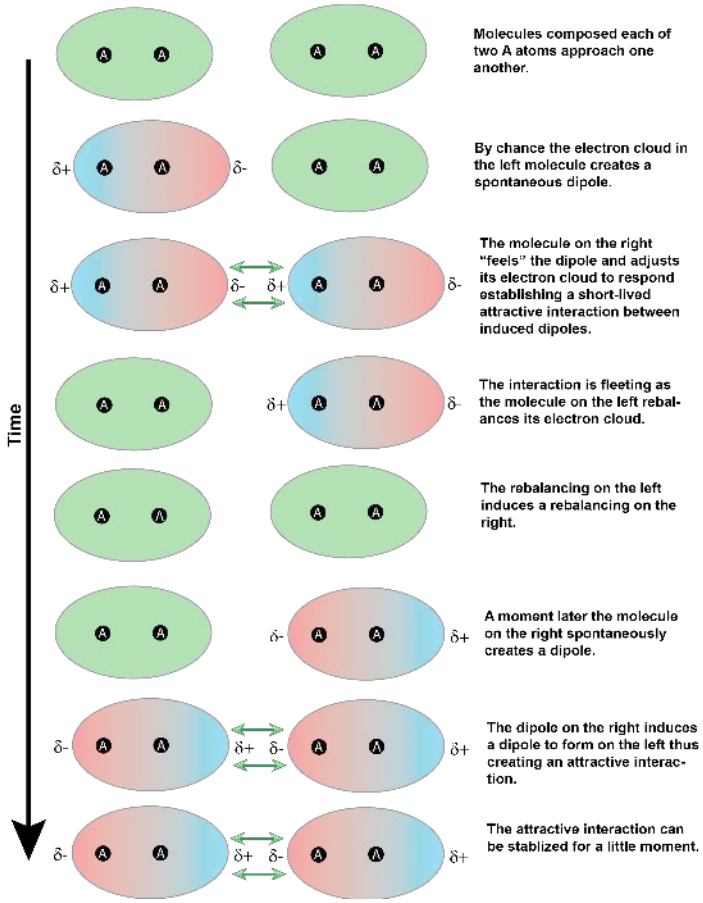


Figure 1: Schematic of dynamic, induced dipoles involved in Van der Waals forces. *Molecules composed of atoms A are depicted approaching one another near the top of the figure. The dynamics (change in time) of these molecules is depicted row-by-row as they change in through arbitrary jumps in time. This dynamic nature is critical to the formation of Van der Waals interactions. Attribution: Marc T. Facciotti (original work)*

Pi Interactions

Pi interactions are a type of molecular interaction that biologists typically encounter in when discussing stabilizing interactions in nucleic acid and protein structures. In a course of general biology, you may also encounter pi interactions in a discussion of protein-DNA interactions. These types of interactions derive their name from the involvement of pi bonds, a specific type of covalent bond between two atoms in which neighboring electron orbitals are close enough to overlap. We'll leave the underlying discussion of molecular orbital theory for your chemistry course and just say that we usually associate pi bonds with double or triple covalent bonds. In biology, these types of bonds occur in many kinds of molecules, particularly those with so called conjugated pi systems including aromatic ring structures like those seen in some amino acids, vitamins and cofactors, and nucleic acids.

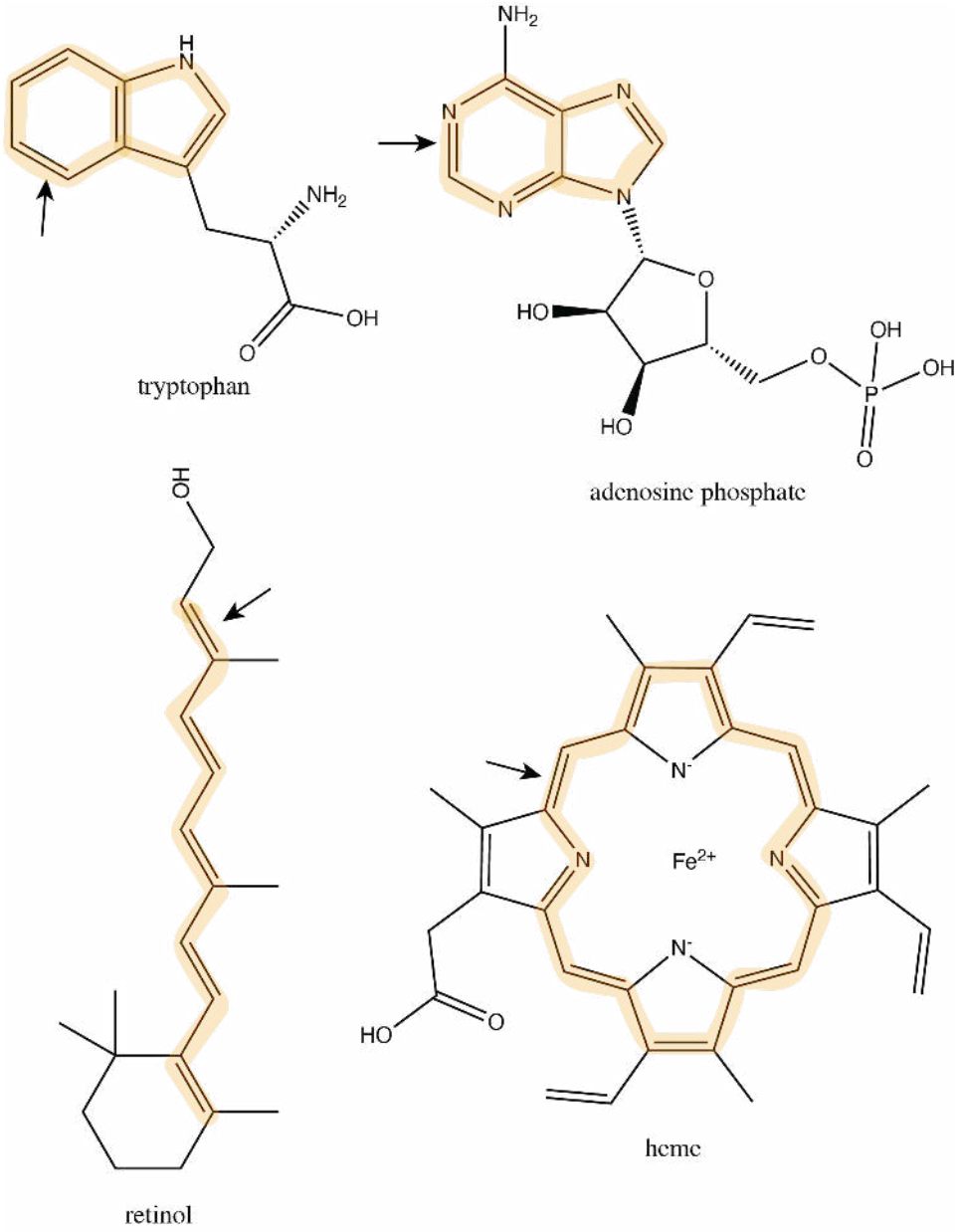


Figure 1: Examples of molecules with conjugated pi systems. *Each of the biomolecules has at least a portion of it that contains so called conjugated pi bonds that can engage in pi interactions with other molecules. With the exception of the retinol molecules the conjugated pi systems are present in planar ring structures. The retinol's conjugated pi system is in the linear hyrodrocarbon portion of the molecule. Arrows point to examples of double bonds. Yellow highlights the systems of delocalized electrons. Attribution: Marc T. Facciotti (original work)*

The distribution of electrons within these pi systems can create regions of more negative and more positive charge and thus creating areas that may “attract” or “repel” other charged or partially charged molecules depending on their relative alignments to one another. Again, we can reserve a deeper discussion of pi systems for your upper division chemistry classes. For now, simply appreciate that - once again - the unique properties of the elements that make up molecules contribute to how electrons distribute within those molecules. The often uneven distributions of electrons about the molecule can create local positive and negative charges and when these regions of partial charges on different molecules (or different parts of molecules) come together in appropriate orientations, that electrostatic interactions (attractive and repulsive) can happen.

A TAKE-HOME POINT ON MOLECULAR INTERACTIONS

Hopefully, you appreciate a common theme to our discussion of non-covalent molecular interactions. Whether we consider ionic bonds, hydrogen bonds, dipole-dipole interactions, Van der Waals forces, or pi interactions, all share the feature of being interactions between full or partial electrostatic charges. The key differences between each of these types of interactions types have to do with how the charges arise on molecules (i.e. the atomic basis for the charge) and/or how the charges interact. This depends, of course, on the underlying unique chemical properties of each element and how they behave at the subatomic level with one another - a suitable topic for discussion in your chemistry class.

Reading_01 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_02

Learning Objectives Associated with SS1_2025_Bis2a_Facciotti_Reading_02

GC.1 Hypothesize what potential types of bonds common functional groups may form.
GC.9 Given an image, identify and name the chemical structure of the following functional groups: amino, carboxyl, hydroxyl, methyl, carbonyl, thiol, and phosphate.
MS.2 Identify and name polar and nonpolar functional groups.
GC.10 Explain how different functional groups influence the chemical properties of biomolecules.
GC.22 Identify functional groups that can form hydrogen bonds and which parts are good hydrogen bond acceptors and/or donors.
GC.23 Predict how water could interact with a given biomolecule based on the properties of the biomolecule's functional groups.
GC.24 Describe the key chemical properties of water and the biological importance of the hydrogen bonds that form between water and various biomolecules.
GC.14 Describe the concept of solubility and how the solubility of a compound is related to the relationship between its chemical properties and those of the solvent - particularly when water is the solvent.
GC.5 Compare and contrast how we use the terms "polar" and "nonpolar" to both describe the distribution of charges in a molecule and the water solubility of a compound.

Functional groups



A functional group is a specific group of atoms within a molecule responsible for a characteristic of that molecule. Many biologically active molecules contain one or more functional groups. We will review the major functional groups found in biological molecules. These include: **hydroxyl, methyl, carbonyl, carboxyl, amino, sulfanyl (aka. thiol), and phosphate** (see Figure 1).

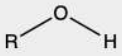
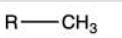
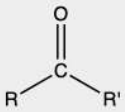
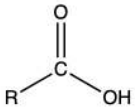
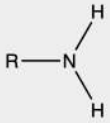
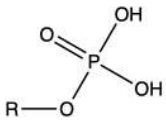
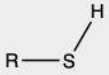
Functional Group	Structure	Properties
Hydroxyl		Polar
Methyl		Nonpolar
Carbonyl		Polar
Carboxyl		Charged, ionizes to release H^+. Since carboxyl groups can release H^+ ions into solution, they are considered acidic.
Amino		Charged, accepts H^+ to form NH_3^+. Since amino groups can remove H^+ from solution, the are considered basic.
Phosphate		Charged, ionizes to release H^+. Since phosphate groups can release H^+ ions into solution, the are considered acidic.
Sulfanyl		Moderately polar (less than -OH)

Figure 1. *The functional groups shown here are found in many different biological molecules. "R" represents any other atom or extension of the molecule.*
Attribution: Marc T. Facciotti (own work adapted from previous image of unknown source)

A functional group may take part in a variety of chemical reactions. We depict some important functional groups commonly found in biological molecules above: hydroxyl, methyl, carbonyl, carboxyl, amino, sulfanyl (aka. thiol), and phosphate. These groups play an important role in the formation of molecules like DNA, proteins, carbohydrates, and lipids. Functional groups can sometimes have polar or nonpolar properties depending on their atomic composition and organization. The term polar describes something that has a property that is not symmetric about it — it can have different poles (more or less of something at different places). With bonds and molecules, the property we care about is usually the distribution of electrons and therefore the electric charge between the atoms. In a nonpolar bond or molecule, electrons and charge will be relatively evenly distributed. In a polar bond or molecule, electrons will concentrate in some areas than others. An example of a nonpolar group is the methane molecule (see discussion in Bond Types Chapter for more detail). Among the polar functional groups is the carboxyl group found in amino acids, some amino acid side chains, and the fatty acids that form triglycerides and phospholipids.

Nonpolar functional groups

Methyl R-CH_3

The methyl group is the only nonpolar functional group in our class list above. The methyl group comprises a carbon atom bound to three hydrogen atoms. In this class, we will treat these C-H bonds as effectively nonpolar covalent bonds (more on this in the Bond Types chapter). This means that methyl groups cannot form hydrogen bonds and will not interact with polar compounds such as water.

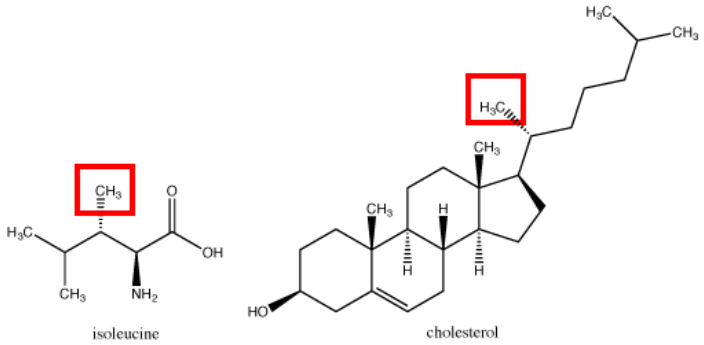


Figure 2. *The amino acid isoleucine is on the left, and cholesterol is on the right. Each has a methyl group circled in red. Attribution: created by Marc T. Facciotti (own work adapted from Erin Easlon)*

A variety of biologically relevant compounds contain methyl groups like those highlighted above. Sometimes, the compound can have a methyl group but still be a polar compound overall because of the presence of other functional groups with polar properties (see the discussion on polar functional groups below).

As we learn more about other functional groups, we will add to the list of nonpolar functional groups. Stay alert!

Polar functional groups

Hydroxyl R-OH

A hydroxyl (alcohol group) is an -OH group covalently bonded another atom. In biological molecules, the hydroxyl group is often (but not always) found bound to a carbon atom, as depicted below. The oxygen atom is much more electronegative than either the hydrogen or the carbon, which will cause the electrons in the covalent bonds to spend more time around the oxygen than around the C or H. Therefore, the O-H and O-C bonds in the hydroxyl group will be polar covalent bonds. Figure 3 depicts the partial charges, δ^+ and δ^-, associated with the hydroxyl group.

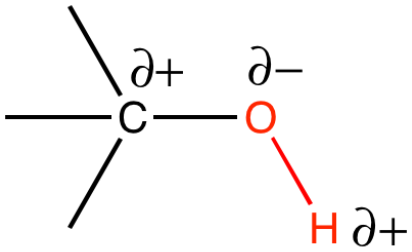


Figure 3. *The hydroxyl functional group shown here consists of an oxygen atom bound to a carbon atom and a hydrogen atom. These bonds are polar covalent, meaning the electron involved in forming the bonds is not shared equally between the C-O and O-H bonds. Attribution: created by Marc T. Facciotti (own work)*

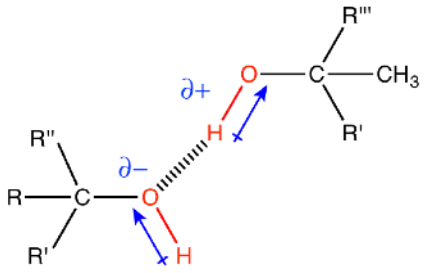


Figure 4. *The hydroxyl functional groups can form hydrogen bonds, shown as a dotted line. The hydrogen bond will form between the δ^- of the oxygen atom and the δ^+ of the hydrogen atom. Dipoles are shown in blue arrows. Attribution: Marc T. Facciotti (original work)*

Hydroxyl groups are very common in biological molecules. Hydroxyl groups appear on carbohydrates (A), on some amino acids (B), and on nucleic acids (C). Can you find any hydroxyl group(s) in the phospholipid in (D)?

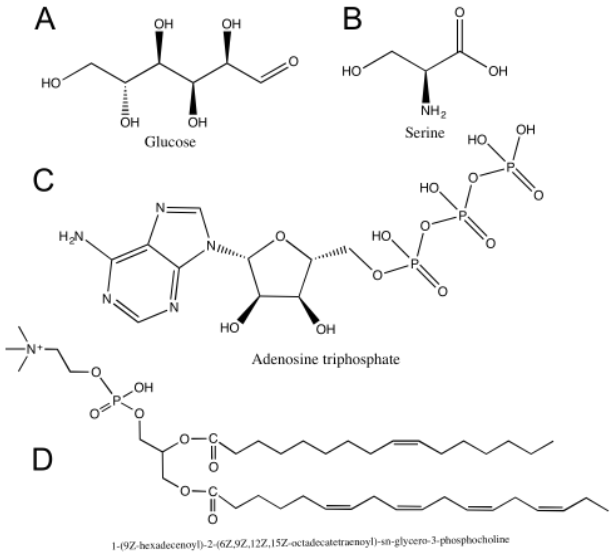


Figure 5. Hydroxyl groups appear on carbohydrates (A, glucose), on some amino acids (B, Serine), and on nucleotides (C, adenosine triphosphate). D is a phospholipid. *Attribution: created by Marc T. Facciotti (own work)*

Carboxyl R-COOH

Carboxylic acid is a combination of a carbonyl group and a hydroxyl group attached to the same carbon, resulting in new characteristics. The carboxyl group can ionize, which means it can act as an acid and release the hydrogen atom from the hydroxyl group as a free proton (H^+). This results in a delocalized negative charge on the remaining oxygen atoms. Carboxyl groups can switch back and forth between protonated (R-COOH) and deprotonated ($R\text{-}COO^-$) states depending on the pH of the solution.

The carboxyl group is very versatile. In its protonated state, it can form hydrogen bonds with other polar compounds. In its deprotonated state, it can form ionic bonds with other positively charged compounds. This will have several biological consequences that will be explored more when we discuss enzymes.

Can you identify all the carboxyl groups on the macromolecules shown above in Figure 5?

Amino R-NH_3

The amino group consists of a nitrogen atom attached by single bonds to hydrogen atoms. An organic compound that contains an amino group is called an amine. Like oxygen, nitrogen is also more electronegative than both carbon and hydrogen, which results in the amino group displaying some polar character.

Amino groups can also act as bases, which means that the nitrogen atom can bond to a fourth hydrogen atom, as shown in Figure 6. Once this occurs, the nitrogen atom gains a positive charge and can now take part in ionic bonds.

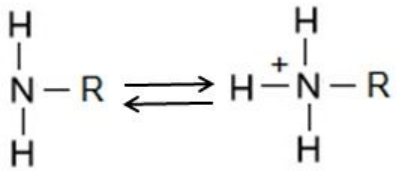


Figure 6. *The amine functional group can exist in a deprotonated or protonated state. When protonated, the nitrogen atom is bound to three hydrogen atoms and has a positive charge. The deprotonated form of this group is neutral. Attribution: created by Erin Easlon (own work)*

Phosphate R-PO_4^{2-}

A phosphate group is a phosphorus atom covalently bound to four oxygen atoms and contains one P=O bond and three P-O^- bonds. The oxygen atoms are more electronegative than the phosphorous atom, resulting in polar covalent bonds. Therefore, these oxygen atoms can form hydrogen bonds with nearby hydrogen atoms that also have a δ^+(hydrogen atoms bound to another electronegative atom). Phosphate groups also contain a negative charge and can take part in ionic bonds.

Phosphate groups are common in nucleic acids and on phospholipids (the term "phospho" referring to the phosphate group on the lipid). In Figure 7 are images of a nucleotide, deoxyadenosine monphosphate (left), and a phosphoserine (right).

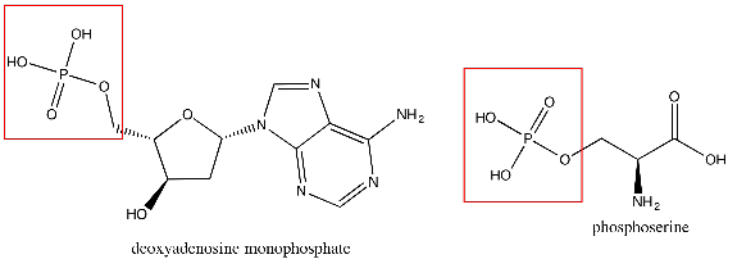
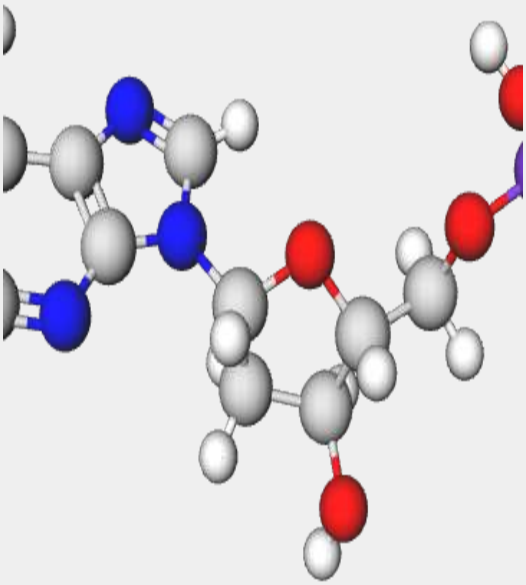
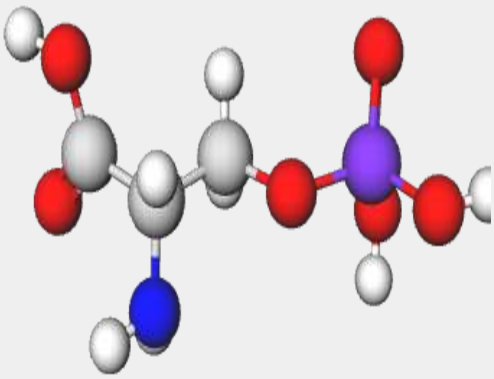


Figure 7. *A nucleotide, deoxyadenosine monphosphate, is on the left, and phosphoserine is on the right. Each has a phosphate group circled in red.*
Attribution: created by Marc T. Facciotti (own work)

You may also find it useful to get used to thinking about these molecules in three dimensions. The interactive figures below (try spinning the molecules) depict the two molecules above, deoxyadenosine monophosphate and phosphoserine as three-dimensional

models. Getting used to three-dimensional representations of biomolecules and interacting with these models can help you form more detailed mental models of what biomolecules look like and how they might interact in "real life".

Deoxyadenosine monophosphate	Phosphoserine
	

Possible NB Discussion Point

Which functional group is actually composed of two other functional groups? After naming the functional groups, discuss how the replacement of the functional groups adjacent to each other results in a new functional group with some amazing properties.

Water

Water is a unique substance whose special properties are intimately tied to the processes of life. Life originally evolved in a watery environment, and most of an organism's cellular chemistry and metabolism occur inside the water-solvated contents of the cell. Water solvates or "wets" the cell and the molecules in it, plays a key role as a reactant or product in an innumerable number of biochemical reactions, and mediates the interactions between molecules in and out of the cell. Many of water's important properties derive from the molecule's polar nature, which derives from the asymmetric arrangement of its polar covalent bonds between hydrogen and oxygen.

When starting the study of biology, the ubiquitous role of water in nearly all biological processes is easy to overlook by getting caught up in the details of specific processes, proteins, the roles of nucleic acids, and in your excitement for molecular machines (it'll happen). It turns out, however, that water plays key roles in all of those processes and we will need to stay continuously aware of the role that water is playing if we are to develop a more functional understanding. Be on the lookout and also pay attention when your instructor points this out.

In a liquid state, individual water molecules interact with one another through a network of dynamic hydrogen bonds that are being constantly forming and breaking. Water also interacts with other molecules that have charged functional groups and/or functional groups with hydrogen bond donors or acceptors. A substance with sufficient polar or charged character may dissolve or be highly miscible in water and is referred to as being hydrophilic (hydro- = “water”; -philic = “loving”). Molecules with more nonpolar characters such as oils and fats do not interact well with water and separate from it rather than dissolve in it. We call these nonpolar compounds hydrophobic (hydro- = “water”; -phobic = “fearing”). We will consider some of the energetic components of these types of reactions in other another chapter.

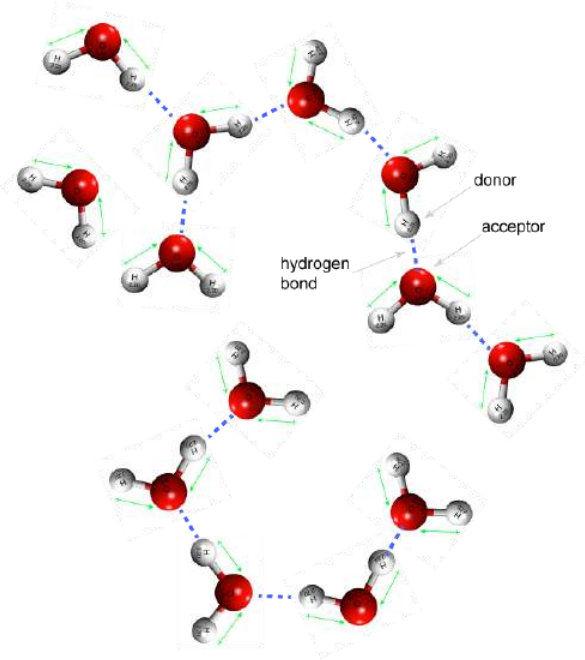


Figure 1. *In a liquid state water forms a dynamic network of hydrogen bonds between individual molecules. Shown are one donor-acceptor pair.*
Attribution: Marc T. Facciotti (original work)

Water's solvent properties

Since water is a polar molecule with slightly positive and slightly negative charges, ions and polar molecules can readily dissolve in it. Therefore, we refer to water as a solvent of other polar molecules and ionic compounds. Charges (or partial charges) associated with these molecules (the solutes) will interact electrostatically with water’s partial charges. Polar bonds with the potential to donate or accept hydrogen bonds will form hydrogen bonds with water. Water molecules that interact directly with individual solute molecules will have their motions slightly constrained as will other nearby molecules. We refer to the layer or partially constrained waters surrounding a solute particle as a hydration layer, hydration shell or sphere of hydration.

When ionic salts are added to water, the individual ions interact with the polar regions of the water molecules, and the ionic bonds are likely disrupted in the process called dissociation. Dissociation occurs when atoms or groups of atoms break off from molecules and form ions. Consider table salt (NaCl, or sodium chloride). A dry block of NaCl is held together by ionic bonds and is difficult to dissociate. When NaCl crystals are added to water, however, the molecules of NaCl dissociate into Na^+ and Cl^- ions, and spheres of hydration form around the ions. The positively charged sodium ion is surrounded by the partially negative charge of the water molecule’s oxygen. The negatively charged chloride ion is surrounded by the partially positive charge of the hydrogen on the water molecule. One may imagine a model in which the ionic bonds in the crystal are "traded" for many smaller scale ionic bonds with the polar groups on water molecules.

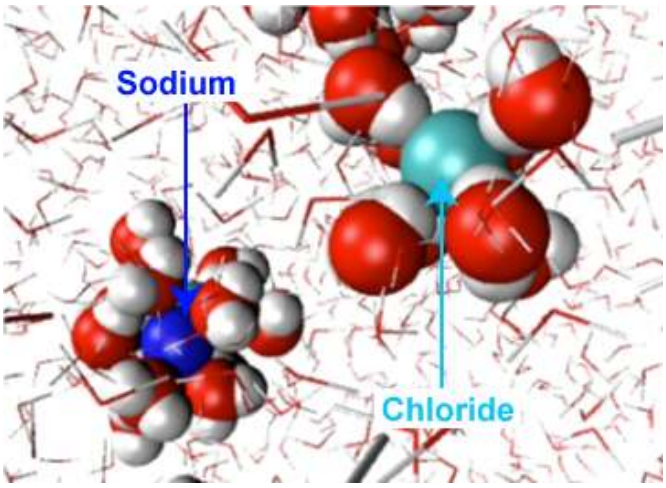


Figure 2. *When table salt (NaCl) is mixed in water, spheres of hydration are formed around the ions. This figure depicts a sodium ion (dark blue sphere) and a chloride ion (light blue sphere) solvated in a "sea" of water. Note how the dipoles of the water molecules surrounding the ions are aligned such that complementary charges/partial charges are associating with one another (i.e., the partial positive charges on the water molecules align with the negative chloride ion whereas the partial negative charges on the oxygen of water align with the positively charged sodium ion).*
Attribution: Ting Wang - UC Davis (original work modified by Marc T. Facciotti)

Possible NB Discussion Point

A pharmaceutical company wants to develop a new antibiotic that is more water soluble than an existing antibiotic. Their strategy will be to add various functional groups to the existing antibiotic and then test the water solubility of the resulting antibiotic. The scientists are trying to decide which functional group(s) to try first. Which one(s) would you recommend and why?

Tags recommended by the template: article:topic

Reading_02 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_03

Learning Objectives Associated with SS1_2025_Bis2a_Facciotti_Reading_03

GC.21 Define and correctly use the law of conservation of mass.
GC.25 Explain the role of water in condensation reactions and hydrolytic reactions.
GC.16 Discuss the concept of reaction reversibility and its relationship to equilibrium.
GC.17 Describe the equilibrium state.
GC.18 Relate the magnitude of the equilibrium constant to the equilibrium position.
GC.19 Explain the effect of changing the concentration of a reactant or product on a system initially at equilibrium - apply Le Chatelier's principle.
GC.35 Apply the concept of chemical equilibrium and the equilibrium constant to describe the progress of a chemical reaction, initially out of equilibrium, towards equilibrium, and finally at equilibrium in terms of "forward" and "reverse" reaction rates and concentrations of chemical reactants.
GC.27 Define and correctly use pH and understand the relationship between pH and the hydrogen ion concentration.
GC.20 Use the definition of pH to determine the difference in $[H^+]$ concentration between two aqueous solutions.
GC.15 Define and correctly use the acid dissociation constant, Ka.
GC.33 Define and correctly use pKa.
GC.32 Explain the relationship between pKa and pH and apply these concepts to predict the protonation state of a functional group at different pH values.
MS.4 Predict the potential influence of changes of pH on the structure and function of biomolecules.

Characteristic Chemical Reactions

MCAT Connection

Chemical reactions occur when two or more atoms bond together to form molecules or when bonded atoms break apart. We call the substances that "go in" to a chemical reaction **reactants** (by convention, we usually list these on the left side of a chemical equation), and the substances found that "come out" of the reaction **products** (by convention, we usually list these on the right side of a chemical equation). An arrow drawn between reactants and products typically shows the direction of the chemical reaction. By convention, for one-way reactions (a.k.a. unidirectional), reactants are listed on the left and products on the right of the single-headed arrow. However, you should be able to identify reactants and products of unidirectional reactions that are written in any orientation (e.g. right-to-left; top-to-bottom, diagonal right-to-left, around a circular arrow, etc.) by using the arrow to orient yourself.

In chemical reactions, the atoms and elements present in the reactant(s) must all also be present in the product(s). Similarly, there can be nothing present in the products that was not present in the reactants. This is because chemical reactions are governed by the law of conservation of mass, which states that matter cannot be created nor destroyed in a chemical reaction. This means that when you examine a chemical reaction, you must try to account for everything that goes in AND make sure that you can find it all in the stuff that comes out!

Just as you can express mathematical calculations in equations such as 2 + 7 = 9, you can use chemical equations to show how reactants become products. By convention, chemical equations are typically read or written from left to right. Reactants on the left are separated from products on the right by a single- or double-headed arrow indicating the direction in which the chemical reaction proceeds. For example, the chemical reaction in which one atom of nitrogen and three atoms of hydrogen produce ammonia would be written as:

$$\mathrm{N} + 3\,\mathrm{H} \longrightarrow \mathrm{NH_3}. \tag{1}$$

Correspondingly, the breakdown of ammonia into its components would be written as:

$$\mathrm{NH_3} \longrightarrow \mathrm{N} + 3\,\mathrm{H}\cdot \tag{2}$$

Note that in either direction, you find 1 N and 3 Hs on both sides of the equation.

Possible NB Discussion Point

In General Biology courses, it is important to appreciate the law of the conservation of mass in the context of biological processes. In chemistry, you will take a quantitative approach to this topic, learning to balance equations, and making sure that the total number of atoms and the total charge does not change. In General Biology, we take a more qualitative approach to the topic. Do you think this leads to confusion? Should we place more emphasis on balancing equations in General Biology?

Reversibility

While all chemical reactions can technically proceed in both directions, some reactions tend to favor one direction over the other. Depending on the degree to which a reaction spontaneously proceed in either both or one direction a different name can be given to characterize the reactions **reversibility**. Some chemical reactions, such as the one shown above, proceed mostly in one direction with the "reverse" direction happening on such long time scales or with such low probability that, for practical purposes, we ignore the "reverse" reaction. These unidirectional reactions are also called **irreversible** reactions and are depicted with a single-headed (unidirectional) arrow. By contrast, **reversible reactions** are those that can readily proceed in either direction. Reversible reactions are usually depicted by a chemical equation with a double-headed arrow pointing toward both the reactants and the products. In practice, you will find a continuum of chemical reactions; some proceed mostly in one direction and nearly never reverse, while others change direction easily depending on various factors like the relative concentrations of reactants and products. These terms are just ways of describing reactions with different equilibrium points.

Use of vocabulary

You may have realized that the terms "reactants" and "products" are relative to the direction of the reaction. If you have a reaction that is reversible, though, the products of running the reaction in one direction become the reactants of the reverse. You can label the same compound with two different terms. That can be a bit confusing. So, what is one to do in such cases? The answer is that if you want to use the terms "reactants" and "products", you must be clear about the direction of reaction that you are referring to - even for when discussing reversible reactions. The choice of terms, "reactants" or "products" that you use will communicate to others the directionality of the reaction that you are considering.

Let's look at an example of a reversible reaction in biology and discuss an important extension of these core ideas that arises in a biological system. In human blood, excess hydrogen ions (H^+) bind to bicarbonate ions (HCO_3^-), forming an equilibrium state with carbonic acid (H_2CO_3). This reaction is readily reversible. If carbonic acid were added to this system, some of it would be converted to bicarbonate and hydrogen ions as the chemical system sought out equilibrium.

$$HCO_3^{-+} H^+ \rightleftharpoons H_2CO_3 \tag{3}$$

The example above examines and "idealized" chemical systems as it might occur in a test-tube. In biological systems, however, equilibrium for a single reaction is rarely reached as it might be in the test-tube. In biological systems, reactions do not occur in isolation. Rather, the concentrations of the reactants and/or products are constantly changing, often with a product of one reaction being a reactant for another reaction. These linked reactions form what are known as biochemical pathways. The immediate example below illustrates this point. While the reaction between the bicarbonate/proton and carbonic acid is highly reversible, it turns out that, physiologically, this reaction is usually "pulled" toward the formation of carbonic acid. Why? As shown below, carbonic acid becomes a reactant for another biochemical reaction—the conversion of carbonic acid to CO_2 and H_2O. This conversion reduces the concentration of H_2CO_3, thus pulling the reaction between bicarbonate and H^+ to the right. Moreover, a third, unidirectional reaction, the removal of CO_2 and H_2O from the system, also pulls the reaction further to the right. These kinds of reactions are important contributors to maintaining the H^+ homeostasis of our blood.

$$HCO_3^- + H^+ \rightleftharpoons H_2CO_3 \rightleftharpoons CO_2 + H_{20} \rightarrow \text{waste} \tag{4}$$

The reaction involving the synthesis of carbonic acid is actually linked to its breakdown into CO_2 and H_2O. These products are then removed from the system/body when they are exhaled. Together, the breakdown of carbonic acid and the act of exhaling the

products pull the first reaction to the right.

Synthesis reactions

Many macromolecules are made from smaller subunits, or building blocks, called monomers. Monomers covalently link to form larger molecules known as polymers. Often, the synthesis of polymers from monomers will also produce water molecules as products of the reaction. This type of reaction is known as **dehydration synthesis** or **condensation** reaction.

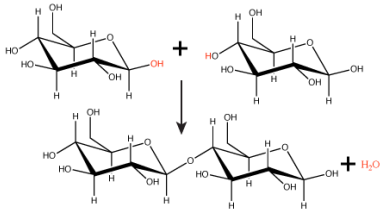
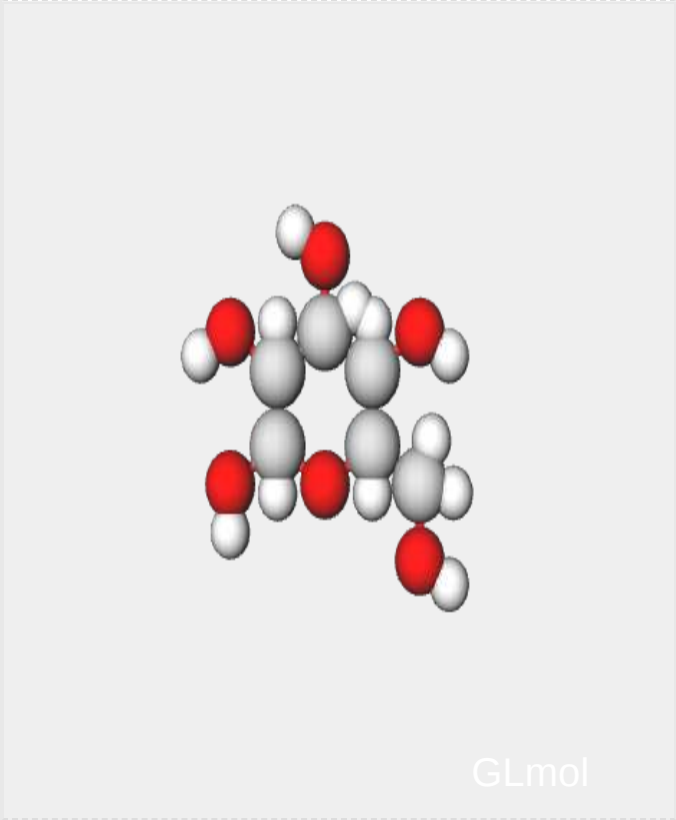
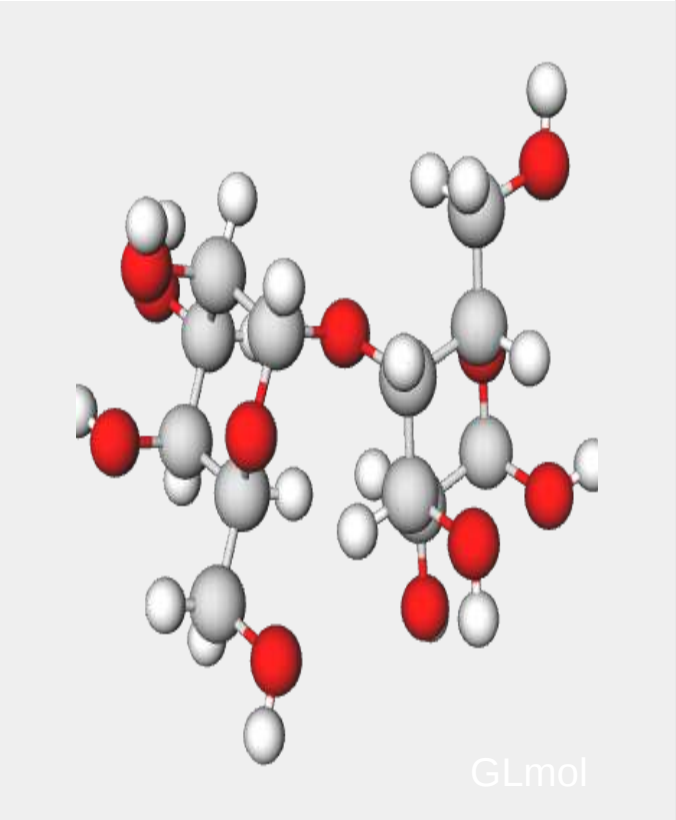


Figure 1. *In the dehydration synthesis reaction depicted above, two molecules of glucose are linked together to form the disaccharide maltose. In the process, a water molecule is formed. Attribution: Marc T. Facciotti (original work)*

Interactive Figure 1. *The molecules of glucose and maltose depicted as 3D interactive molecules.*

Glucose	Maltose
	

In a dehydration synthesis reaction (Figure 1), the hydrogen of one monomer combines with the hydroxyl group of another monomer, releasing a molecule of water. At the same time, the monomers share electrons and form covalent bonds. As additional monomers join, this chain of repeating monomers forms a polymer. Different types of monomers can combine in many

configurations, giving rise to a diverse group of macromolecules. Even one kind of monomer can combine in a variety of ways to form several different polymers; for example, glucose monomers are the constituents of starch, glycogen, and cellulose.

In the carbohydrate monomer example above, the polymer is formed by a dehydration reaction; this type of reaction is also used to add amino acids to a growing peptide chain and nucleotides to the growing DNA or RNA polymer. Visit the modules on Amino Acids, Lipids, and Nucleic Acids to see if you can identify the water molecules that are removed when a monomer is added to the growing polymer.

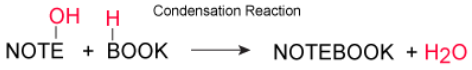


Figure 2. *This depicts, using words, (decorated with functional groups colored in red) a generic dehydration synthesis/condensation reaction. Attribution: Marc T. Facciotti (original work)*

Hydrolysis reactions

Polymers are broken down into monomers in a reaction known as hydrolysis. A hydrolysis reaction includes a water molecule as a reactant (Figure 3). During these reactions, a polymer can be broken into two components: one product carries a hydrogen ion (H^+) from the water, while the second product carries the water's remaining hydroxide (OH^-).

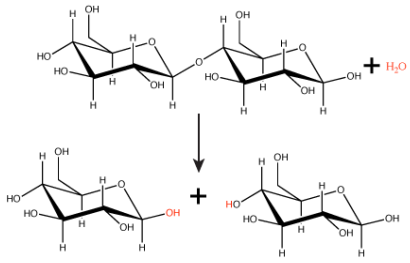


Figure 3. *In the hydrolysis reaction shown here, the disaccharide maltose is broken down to form two glucose monomers with the addition of a water molecule. Note that this reaction is the reverse of the synthesis reaction shown in Figure 1 above. Attribution: Marc T. Facciotti (original work)*

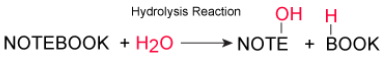


Figure 4. *This depicts using words (decorated with functional groups colored in red) a generic hydrolysis reaction. Attribution: Marc T. Facciotti (original work)*

Dehydration synthesis and hydrolysis reactions are catalyzed, or “sped up,” by specific enzymes. Note that both dehydration synthesis and hydrolysis reactions involve the making and breaking of bonds between the reactants—a reorganization of the bonds between the atoms in the reactants. In biological systems (our bodies included), food in the form of molecular polymers is hydrolyzed into smaller molecules by water via enzyme-catalyzed reactions in the digestive system. This allows for the smaller nutrients to be absorbed and reused for a variety of purposes. In the cell, monomers derived from food may then be reassembled into larger polymers that serve new functions.

Helpful links:

Visit this site to see visual representations of dehydration synthesis and hydrolysis.
Example of Hydrolysis with Enzyme Action is shown in this 3 minute video entitled: Hydrolysis of Sucrose by Sucrase.

Exchange/transfer reactions

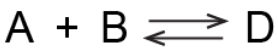
We will also encounter reactions termed exchange reactions. In these types of reactions, "parts" of molecules are transferred between one another—bonds are broken to release a part of a molecule and bonds are formed between the released part and another molecule. These enzyme-catalyzed reactions are usually reasonably complex multi-step chemical processes.



Figure 5. *An exchange reaction in which both synthesis and hydrolysis can occur, chemical bonds are both formed and broken, is depicted using a word analogy.*

Chemical equilibrium—Part 1: forward and reverse reactions

Understanding the concept of chemical equilibrium is critical to following several of the discussions that we have in general biology and indeed throughout biology and the sciences. It is difficult to completely describe the concept of chemical equilibrium without reference to the energy of a system, but for the sake of simplicity, let's try anyway and reserve the discussion of energy for another chapter. Let us, rather, begin developing our understanding of equilibrium by considering the reversible reaction below:



Hypothetical reaction #1: A hypothetical reaction involving compounds A, B and D. If we read this from left to right, we would say that A and B come together to form a larger compound: D. Reading the reaction from right to left, we would say that compound D breaks down into smaller compounds: A and B.

We first need to define what is meant by a "reversible reaction." The term "reversible" simply means that a reaction can proceed in both directions. That is, the things on the left side of the reaction equation can react together to become the things on the right of the equation, AND the things on the right of the equation can also react together to become the things on the left side of the equation. Reactions that only proceed in one direction are called irreversible reactions.

To start our discussion of equilibrium, we begin by considering a reaction that we posit is readily reversible. In this case, it is the reaction depicted above: the imaginary formation of compound D from compounds A and B. Since it is a reversible reaction, we could also call it the decomposition of D into A and B. Let us, however, imagine an experiment in which we watch the reaction proceed from a starting point where only A and B are present.

Example #1: Left-balanced reaction

Hypothetical reaction #1: time course

Concentration	t=0	t=1	t=5	t=10	t=15	t=20	t=25	t=30	t=35	t=40
[A]	100	90	80	70	65	62	60	60	60	60
[B]	100	90	80	70	65	62	60	60	60	60
[C]	0	10	20	30	45	38	40	40	40	40

At time t = 0 (before the reaction starts), the reaction has 100 concentration units of compounds A and B and zero units of compound D. We now allow the reaction to proceed and observe the individual concentrations of the three compounds over time (t=1, 5, 10, 15, 20, 25, 30, 35, and 40 time units). As A and B react, D forms. In fact, one can see D forming from t=0 all the way to t=25. After that time, however, the concentrations of A, B and D stop changing. Once the reaction reaches the point where the concentrations of the components stop changing, we say that the reaction has reached equilibrium. Notice that the concentrations of A, B, and D are not equal at equilibrium. In fact, the reaction seems left balanced so that there is more A and B than D.

Note: **Common student misconception warning**

Many students fall victim to the misconception that the concentrations of a reaction's reactants and products must be equal at equilibrium. Given that the term equilibrium sounds a lot like the word "equal," this is not surprising. But as the experiment above tries to illustrate, this is NOT correct!

Example #2: right-balanced reaction

We can examine a second hypothetical reaction, the synthesis of compound **J** from the compounds **E** and **F**.

$$\mathbf{E} + \mathbf{F} \rightleftharpoons \mathbf{J}$$

Hypothetical reaction #2: *A hypothetical reaction involving compounds E, F and J. If we read this from left to right, we would say that E and F come together to form a larger compound: J. Reading the reaction from right to left, we would say that compound J breaks down into smaller compounds: E and F.*

The structure of hypothetical reaction #2 looks identical to that of hypothetical reaction #1, which we considered above—two things come together to make one bigger thing. We just need to assume, in this case, that E, F, and J have different properties from A, B, and D. Let's imagine a similar experiment to the one described above and examine this data:

Hypothetical reaction #2: time course

	t=0	t=1	t=5	t=10	t=15	t=20	t=25	t=30	t=35	t=40
Concentration of E	100	80	62	45	30	24	20	18	18	18
Concentration of F	100	80	62	45	30	24	20	18	18	18
Concentration of J	0	20	38	55	70	76	80	82	82	82

In this case, the reaction also reaches equilibrium. This time, however, equilibrium occurs at around t=30. After that point, the concentrations of E, F, and J do not change. Note again that the concentrations of **E**, **F**, and **J** are not equal at equilibrium. In contrast to hypothetical reaction #1 (the ABD reaction), this time the concentration of J, the thing on the right side of the arrows, is at a higher concentration than E and F. We say that, for this reaction, equilibrium lies to the right.

Four more points need to be made at this juncture.

- Point 1: Whether equilibrium for a reaction lies to the left or the right will be a function of the properties of the components of the reaction and the environmental conditions that the reaction is taking place in (e.g., temperature, pressure, etc.).
- Point 2: We can also talk about equilibrium using concepts of energy, and we will do this soon, just not yet.
- Point 3: While hypothetical reactions #1 and #2 appear to reach a point where the reaction has "stopped," you should imagine that reactions are still happening even after equilibrium has been reached. At equilibrium the "forward" and "reverse" reactions are just happening at the same rate. That is, in example #2, at equilibrium J is forming from E and F at the same rate that it is breaking down into E and F. This explains how the concentrations of the compounds aren't changing despite the fact that the reactions are still happening.
- Point 4: From this description of equilibrium, we can define something we call the equilibrium constant. Typically, the constant is represented by an uppercase K and may be written as K_{eq}. In terms of concentrations, K_{eq} is written as the mathematical product of the reaction product concentrations (stuff on the right) divided by the mathematical product of the reactant concentrations (stuff on the left). For example, $K_{eq,1}$ = [D]/[A][B], and $K_{eq,2}$ = [J]/[E][F]. The square brackets "[]" indicate the "concentration of" whatever is inside the bracket.

Possible NB Discussion Point

The following statement is true: *A chemical equilibrium can be established starting with equal concentrations of reactants and products.* Can you think of and describe other starting conditions for which a chemical equilibrium can also be established? Are there any starting conditions for which a chemical equilibrium can NOT be established?

The Role of Acid/Base Chemistry in General Biology

We have learned that the behavior of chemical functional groups depends on the composition, order, and properties of their constituent atoms. We will see that pH, a measure of the hydrogen ion concentration of a solution, can alter the chemical properties of some key biological functional groups in ways that change how they interact with other molecules and thus their biological role.

For example, depending on the pH, some functional groups on the amino acid that make up proteins can exist in different chemical states. We will learn that the chemical state of these functional groups can have a profound effect on the shape of the protein or on its ability to carry out chemical reactions. As we move through the course, we will see many examples of this type of chemistry in different contexts.

In pure water, hydrogen ions are spontaneously generated by the dissociation (ionization) of a small percentage of water molecules into equal numbers of hydrogen (H^+) ions and hydroxide (OH^-) ions. The OH^- that result from the ionization of water departs into the sea of water molecules interacting with other molecules through polar interactions, while the now "free" (unbonded) H^+ ions produced by the ionization associates with water molecules (line two of the figure below) to create a new molecule called a hydronium ion, H_3O^+. At some point, the hydroxide ion from line 1 in the figure below will rejoin with a proton and reform another water molecule. This process of dissociation and re-association between hydroxide and hydrogen ions happens continuously at equilibrium.

While most H^+ ions in solution really exist as H_3O^+ ions, we usually represent the H_3O^+ in figures or equations more simply as H^+. Why? Because it is easier. Just remember that since nearly all chemistry in biology happens in water that when you see H^+ referred to in the text, figures, or in equations, it usually represents H_3O^+.

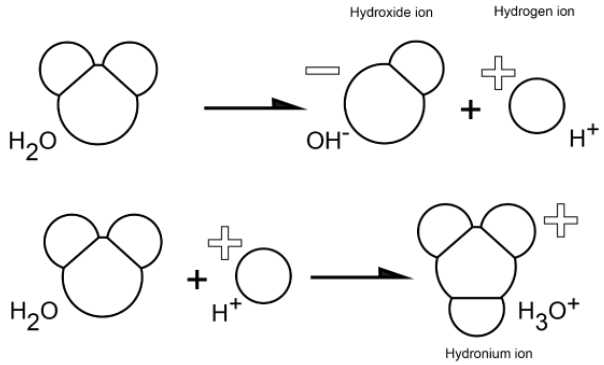


Figure 1: *Water spontaneously dissociates into a proton and hydroxyl group. The proton will combine with a water molecule forming a hydronium ion.*
Attribution: Marc T. Facciotti

While some paradoxes to this rule can be found in the chemistry of concentrated solutions, in General Biology, it is convenient to formally define **pH** as:

$$pH = -\log_{10}[H^+] \quad (5)$$

In the equation above, the square brackets surrounding $[H^+]$ indicate concentration. If necessary, try a math review at wiki-logarithm or kahn-logarithm. Also see: definition-concentration or wiki-concentration. The **pH** of a solution is therefore a measure of the concentration of hydrogen ions in a solution (or the number of hydronium ions).

The pH is typically reported on a logarithmic **pH scale** that ranges from 0 to 14 (Figure 2). We define pH=7.0 as neutral. We call anything with a pH below 7.0 **acidic** and any reported pH above 7.0 **alkaline** or **basic**. Extremes in pH in either direction from 7.0 are often considered inhospitable to life, although examples exist to the contrary. pH levels in the human body usually range between 6.8 and 7.4, except in the stomach where the pH is more acidic, typically between 1 and 2. Some microbial species like *Sulfolobus acidocaldarius* thrive in hyper acidic environments (pH < 3) while others like *Natronomonas pharaonis* have been found living in lakes with pH > 11. These organisms are classified as "extremophiles" for their abilities to thrive in extreme environments. Proteins from these organisms are sometimes used in industrial processes where their ability to withstand environmental stress is a valued property.

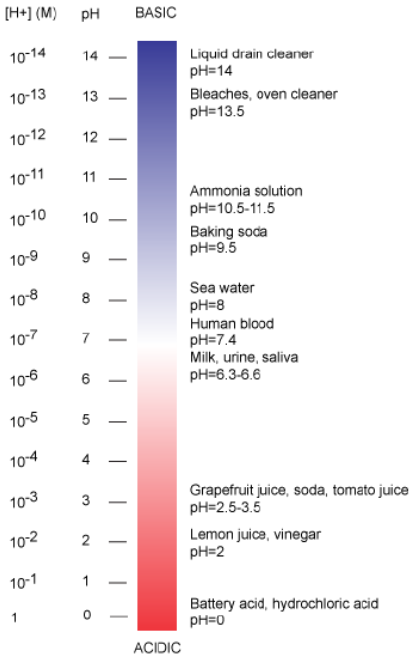


Figure 2: *The pH scale ranging from acidic to basic with various biological compounds or substances that exist at that particular pH. Attribution: Marc T. Facciotti*

For Additional Information

Watch this video for an expanded explanation of pH and its relationship to [H+] and the logarithmic scale.



Let's work out an example to see how the pH scale works.

For reference: 1 mole (mol) of a substance (which can be atoms, molecules, ions, etc.), is defined as being equal to 6.02 x 10^{23} particles of the substance. Therefore, 1 mole of water is equal to 6.02 x 10^{23} water molecules.

Mathematically this can be written as:

1 mol = 6.02×10^{23} particles in a substance

1 mol H_2O = 6.02×10^{23} water molecules

Example: The concentration of hydrogen ions dissociating from pure water is approximately 1×10^{-7} moles H^+ ions per liter of water. The pH is calculated as the negative of the base 10 logarithm of this unit of concentration. The $\log_{10}$ of 1×10^{-7} is -7.0, and the negative of this number yields a pH of 7.0 (neutral pH).

Mathematically this can be represented as:

$pH = -\log_{10}[H^+]$

$pH = -\log_{10}[1 \times 10^{-7}]$

pH = 7.0 (neutral pH)

The figure below provides another way to visualize the inverse relationship between proton and hydroxide ion concentrations by graphically illustrating how proton concentration decreases as pH increases while the hydroxide ion concentration simultaneously increases.

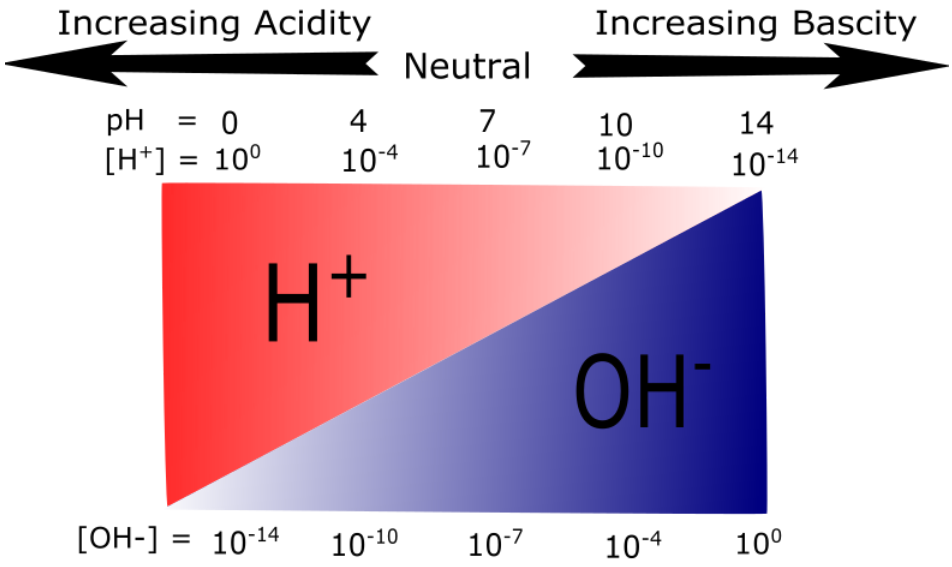


Figure 3: A graphical representation of acidity and basicity. This figure illustrates the relationship between H^+ and OH^- concentrations on the pH scale. At low pH values H^+ ions are plentiful. As the pH increases the relative abundance of OH^- ions increase while H^+ abundance decreases.

Attribution: Mary O. Aina

The inverse relationship between pH and the concentration of protons confuses many students - take the time to convince yourself that you "get it." One way could be to predict whether different pH values are acidic or basic and then do the calculations to make sure. Start by trying these practice questions.

Acids and Bases

Acids and bases are molecules that can influence the pH of a solution. In General Biology it is often convenient to use the Brønsted-Lowry definition of acids and bases. Using this formalism we define:

Acids = molecules that can donate a proton to another molecule (including water to form a hydronium ion)
Bases = molecules that can accept a proton from another molecule (including hydronium ions)

When protons from acidic molecules dissociate from their "parent" they increase the H^+ concentration and thereby lower the pH of the solution. By contrast, when a base absorbs a "free" proton from a solution onto the "parent" molecule, the decrease in proton concentration in solution results in a shift to higher pH values.

Generically we can represent acids and bases as follows:

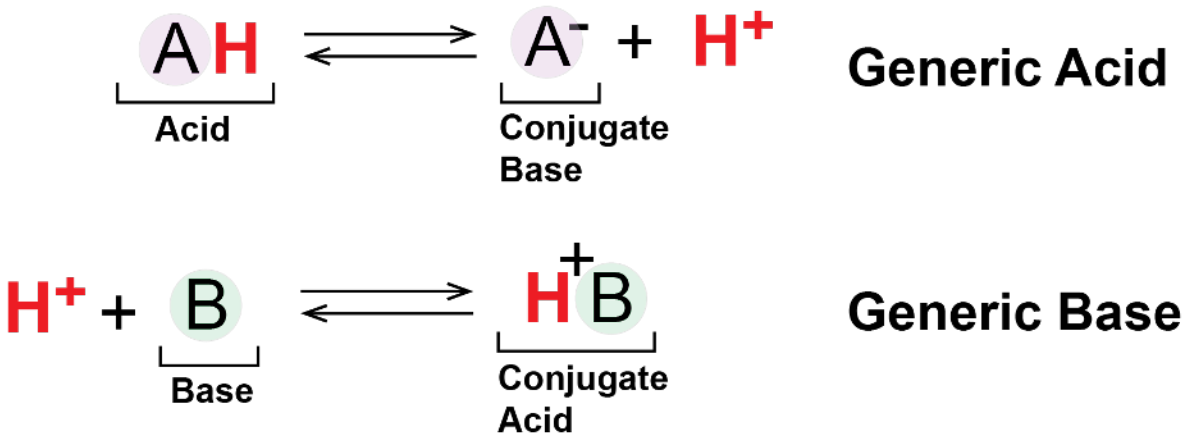


Figure 4: Generic Acids and Bases. This figure shows the behavior of Brønsted-Lowry acids and bases. The acid (A in a light purple circle) starts in a protonated form bound to an H^+ ion, drawn as a red H. The acid deprotonates, shedding its H^+ into solution or to another molecule. Meanwhile the base (B in a light green circle) begins deprotonated and absorbs a proton (red $H^{+)}$ from solution or other molecule.
Attribution: Marc T. Facciotti

In the figure above, the molecule A^- - the deprotonated form of the acid AH - can also be referred to as the **conjugate base** of the acid AH. Likewise the molecule BH^+ - the protonated form of the base B - can be referred to as the **conjugate acid** of the base B.

We call acids that completely dissociate into A^- and H^+ ions at equilibrium **strong acids**. These reactions are characterized by an equilibrium position that lies far to the right (favoring product formation) and their chemical equations are often drawn with a single arrow separating reactants and products. By contrast, acids that do NOT completely dissociate into A^- and H^+ ions at equilibrium are called **weak acids**. Depending on the pH, it is common to find both protonated and deprotonated forms of the weak acid (or both the acid and it's conjugate base) in solution at the same time. The chemical equations representing these reactions are therefore usually depicted with double arrows, indicating that the protonation/deprotonation of A^-/AH, respectively, is reversible.

Two important examples of weak acids/bases in biology are the carboxyl and amino functional groups. At physiological pH values (around pH = 7) the carboxyl group tends to behave as an acid by donating it's proton to solution or other molecules. Under the same conditions, the amino group tends to act as a base, absorbing protons from solution or other molecules. As we will soon see, these and other protonation/deprotonation reactions play key roles in many biological processes.

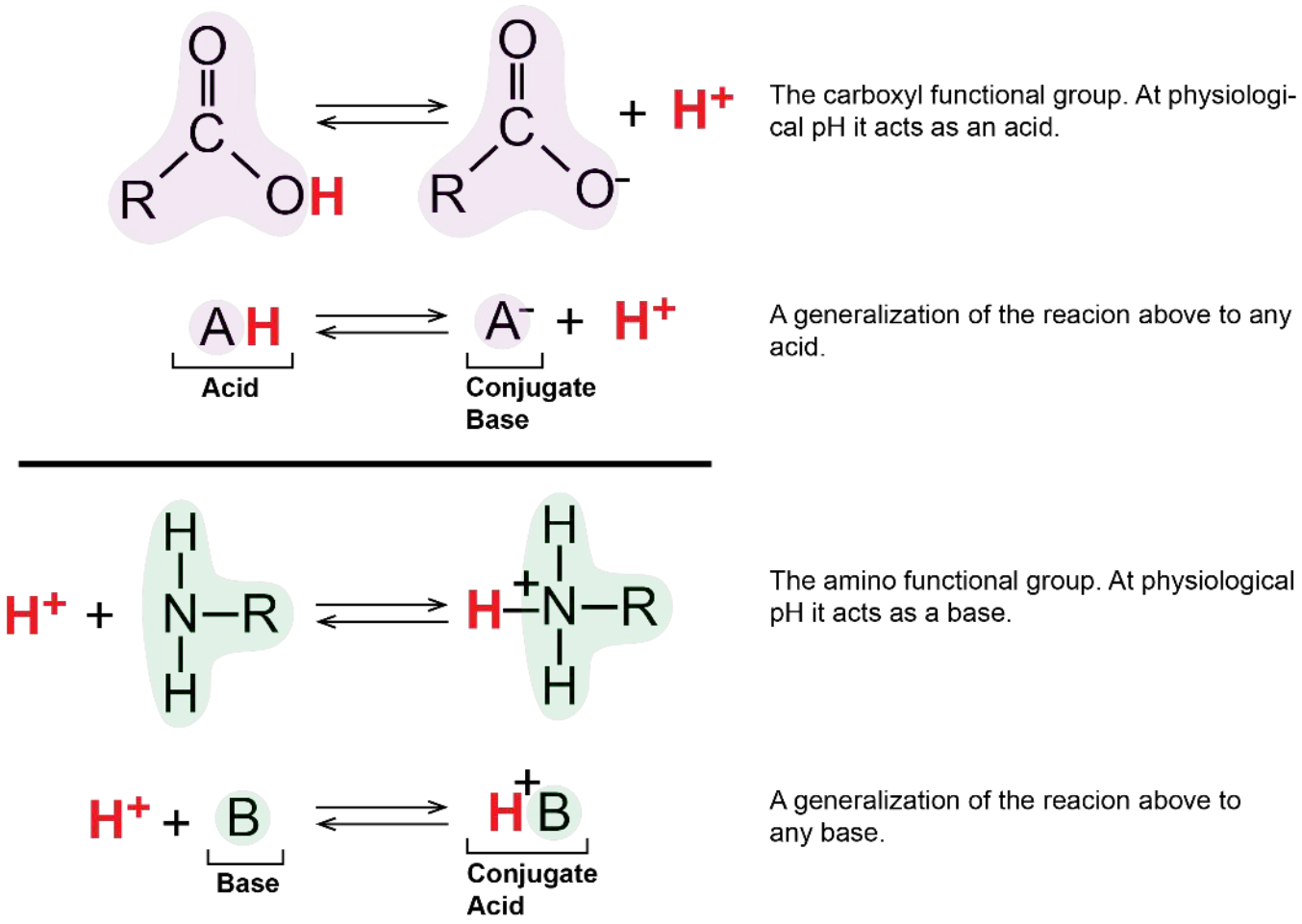


Figure 5: *The carboxylic acid group acts as an acid by releasing a proton. This can increase the number of protons in solution and thus decrease the pH. The amino group acts as a base by accepting hydrogen ions, which can decrease the number of hydrogen ions in solutions, thus increasing the pH.*
Attribution: Marc T. Facciotti (original work)

pKa

MCAT Connection

pK$_a$ is defined as the negative $\log_{10}$ of the dissociation constant of an acid, its K_a.

$$\mathbf{pK_a = -\log_{10}[K_a]}$$

The pK_a is a quantitative measure of how readily an acid gives up a proton to a solution and is thus a measure of the "strength" of the acid. Strong acids have a small pK_a, weak acids have a larger pK_a.

As noted, the carboxylic acid functional group R-COOH is found in many biomolecules. This functional group has a pK_a between 2-4 in aqueous solution and is considered a **weak** acid. At many biologically relevant pH values, the carboxylic acid only partially dissociates into H^+ cations and R-COO^- anions. In a population of molecules containing the carboxylic acid functional group, it is not uncommon - at biologically relevant pH values - to simultaneously find molecules in both the protonated (R-COOH) and deprotonated (R-COO^-) forms. By contrast, HCl (hydrogen chloride), a common **strong** acid, has a pKa << 0. This means that it will fully dissociate into H^+ and Cl^- at all biologically relevant pH values. HCl is a strong acid and will almost never be found in the protonated form.

One of the key ways we depict the difference between a strong acid or base and a weak acid or base in chemical equations is the use of a single arrow (strong acid/base) versus a double arrow (weak acid/base). The full dissociation of proton(s) from strong acids at equilibrium means that the likelihood of protons returning to the molecule they dissociated from is very small. The reaction goes mostly one way, hence the single arrow. In weak acids, the likelihood of the forward reaction happening is more similar to the likelihood of the reverse reaction happening. The consequence of this is partial dissociation at equilibrium, hence the double arrow in the chemical equation.

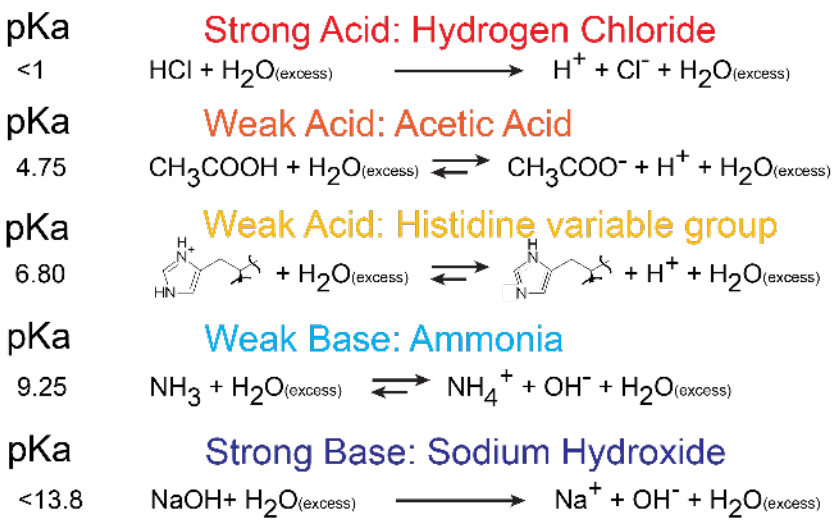


Figure 1. *An example of strong acids, weak acids, strong bases, and weak bases in their biologically relevant protonated and deprotonated states. The value of their pKa is shown on the left. Attribution: Marc T. Facciotti.*

In General Biology some instructors may ask you to relate pH and pKa to each other when discussing the protonation state of a weak acid or weak base in, for example, amino acids. How can we use the information given in this module to answer the question: **Will the functional groups on the amino acid Glutamate be protonated or deprotonated at a pH of 2, at a pH of 8, or at a pH of 11?**

To answer this kind of question, we need to create a relationship between pH and pKa. The relationship between pKa and pH is mathematically represented by the Henderson-Hasselbach equation shown below, where $[A^-]$ represents the deprotonated form of the acid and [HA] represents the protonated form of the acid.

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$

Figure 2. *The Henderson-Hasselbach equation*

This equation has three core "parts":

1. The pH;
2. The pK_a; and
3. $\log_{10}[A^-]/[HA]$.

Part 1 tells you about the proton concentration. Part 2 tells you about a property of the acid - how likely it is to "give up" its protons to solution. Part 3 tells you about how much of the acid is in its deprotonated form $[A^-]$ and in its protonated form [HA]. In most experimental conditions we usually assume that the pK_a doesn't change (after all, it's a property of the molecule). So, this equation tells us that the pH and the ratio of deprotonated $[A^-]$ and protonated [HA] form of the acid are related to one another. If you are able to independently control the pH by adding more acid or base, you can control the ratio of deprotonated $[A^-]$ and protonated [HA] forms of the acid. You can, of course, use the Henderson-Hasselbach equation to solve the problem of knowing

the protonation state of Glutamate's functional groups at different pHs. However, we can also develop an intuition about the relationship between these three quantities.

Titrations can help develop an intuitive understanding

Another useful way to develop an intuitive understanding of the relationship between pH, pK_a and the protonation states of functional groups is to think about the results of a titration. The titration experiment typically involves the slow, stepwise, and gradual addition of one reagent (e.g. Reagent #1) into a mixture of other molecules (e.g. Solution #1). The experimenter slowly adds Reagent #1 (independent variable) into Solution #1 and makes observations of one or more properties (dependent variables) of the mixture after every step. Depending on the reagents, observations can be things like a change in color, change in viscosity, change in taste, or change in pH. The experimental data is usually plotted in a graph with increasing Reagent #1 on the x-axis and the measured solution property (e.g pH, color, viscosity, etc.) plotted on the y-axis.

Interpreting a Titration Graph

Below is a graph showing the titration of a solution of acetic acid (Solution #1). Acetic acid, the acid found in vinegar, can also be represented by the chemical formula CH_3COOH and contains a single carboxyl functional group. In this experiment, the acetic acid is titrated with a base, represented as “OH” (Reagent #1) in the figure. The graph when read from left to right reports the change in pH of the solution when base (OH) is slowly added. Examining the graph, you see three phases of pH change:

(a) Adding between 0 to about 3 equivalents of OH^- causes a rapid rise of the pH. At the molecular level, this rise can be explained by each additional base equivalent reacting with H_3O^+ to create two neutral water molecules. This reduces the $[H^+]$ and therefore increases the pH. At this pH, the molecule (CH_3COOH) has the "strength" to hold onto its H^+ ions.

(b) When between 3 and 7 equivalents of OH are added, the pH hardly changes; it stays stable around a value equivalent to the pK_a of CH_3COOH (4.76). In this zone of the chart, acetic acid molecules start to "let go" of their H^+ ions when OH is added. Each OH added “grabs” a proton from solution which is replaced by a proton “released” by CH_3COOH molecules in solution. When the pH = pK_a (around 5 equivalents of OH^-), the Henderson-Hasselbach equation tells us that 50% of the acetic acid molecules in solution are protonated and 50% are deprotonated. As more equivalents of OH are added, more acetic acid molecules become deprotonated until all acetic acid molecules become deprotonated.

(c) Once this happens, at 7 or more equivalents of OH, no more protonated molecules are available to neutralize added OH. Adding more OH therefore starts rapidly raising the pH again.

Building a mental picture of this process can be a powerful tool to help you intuitively solve the problem that started this discussion. If you want to intuit the protonation state of a functional group, with a known pK_a, at a given pH, you can start by imagining the situation when the pH = pK_a. At this point, you know that the functional group is 50% protonated and 50% deprotonated. Therefore, if the pH in question is below the pKa moving from the pH = pK_a to the target pH requires adding protons to the solution. When the solution becomes more acidic, there will be more H^+ ions ready to protonate deprotonated functional groups and thus increasing the amount of protonated functional groups. On the titration curve below, this is like starting at the center of the flat portion of the curve and moving to the left. When the target pH is higher than the pK_a the solution must become more basic in the shift from the starting pH to the target pH. Here, there will be fewer H^+ ions than at the beginning, meaning that the functional groups will start to deprotonate thus increasing the proportion of unprotonated functional groups.

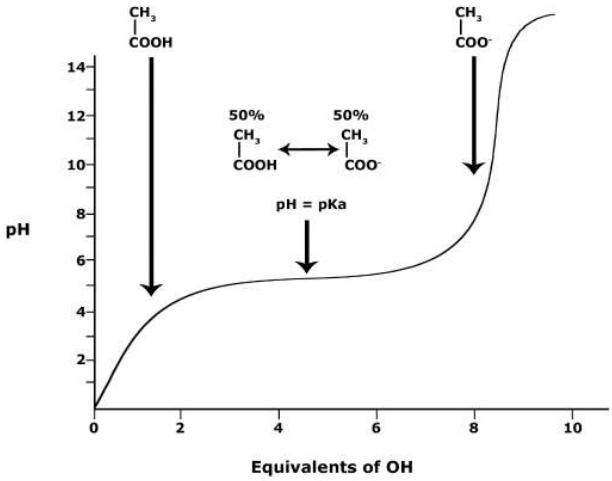


Figure 3. *This graph depicts the protonation state of acetic acid as the pH changes. At a pH below the pKa, the acid is protonated. At a pH above the pKa the acid is deprotonated. If the pH equals the pKa, the acid is 50% protonated and 50% deprotonated. Attribution: Ivy Jose*

This movie provides a visual demonstration of the explanation given above.



We close this section by returning to the original question: **Will the functional groups on the amino acid glutamate be protonated or deprotonated at a pH of 2, at a pH of 8, at a pH of 11?** In the example of acetic acid above, we developed an understanding of how to relate the protonation state of one functional group to its pK_a and the pH of the solution its in to the balance/ratio of protonated to deprotonated state for a single functional group.

In biology, however, you will often be interested in the behavior of molecules with multiple functional groups, each with their own pK_a values. An free amino acid like glutamate has three different functional groups, each with their own pK_a. So, answering the question about the protonation state of glutamate at different pH values requires you to assess the protonation state and deprotonation state of each of those functional groups independently. The figure below shows a titration experiment for amino acid glutamate. Like the plot above, it shows the relationship between the pH and pK_a for each of the three ionizable functional groups of glutamate. The protonation state of functional group can be evaluated independently at a specified pH value to ultimately determine the protonation state of the whole molecule.

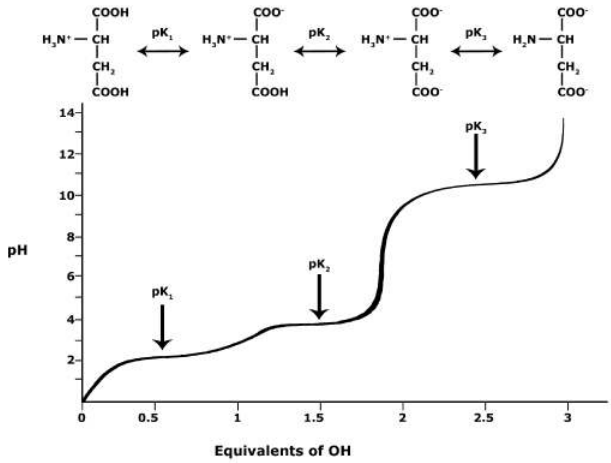


Figure 4. *This graph depicts the protonation state of glutamate as the pH changes. At a pH below the pKa for each functional group on the amino acid, the functional group is protonated. At a pH above the pKa for the functional group it is deprotonated. If the pH equals the pKa, the functional group is 50% protonated and 50% deprotonated.*
Attribution: Ivy Jose

Interactive Exercise

Quick Reference: pKa and pH Comparison Chart

pK_a	**pH**
$pK_a = -\log_{10}[K_a]$	$pH = -\log_{10}[H^+]$
• Reports on a **property of a molecule**. It says something about the likelihood of a proton will *dissociate* from it.	• Reports on a **property of a solution**: the $[H^+]$ in solution. ○ Acid increases the $[H^+]$ in solution, by donating hydrogen ion(s). This lowers the pH. ○ Base lowers the $[H^+]$ in solution by binding hydrogen ion(s). This raises the pH.
• Can be associated with the strength of an acid. ○ Low pK_a = strong acid ○ High pK_a = weak acid	• Indicates whether a solution is basic OR acidic. ○ High pH (>7) = Basic ○ Low pH (<7) = Acidic ○ pH= 7 is neutral
• While pK_a depends largely on the physical properties of a molecule, it can also be influenced/changed by the local environment the molecule finds itself in.	• The pH can be changed by addition or removal of various chemical factors in solution.

Additional pH resources

Here are some additional links on pH and pKa to help learn the material. Note that there is an additional module devoted to pKa.

ChemLibreText Links

- Determining and calculating pH
- pH and pKa
  - This has an expanded discussion about pH that is a bit more detailed than how we present the concept.

Khan Academy Links

- What is pH
- strong acids and bases
- weak acids
- weak acid-base equilibria
- pH and pKa

Simulations

- Acid-base simulation. -------
  - ----------embed this in document---------

Extra pH Links

- https://chemclass-ol.org/logarithms-and-ph
- https://web.mst.edu/~gbert/logs/pH.html
- http://www.biology.arizona.edu/biomath/tutorials/applications/pH.html
- Intro to acids, bases, pH

Reading_03 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_04

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_04

MS.6 *Be able to classify common biomolecules as a lipid, protein, carbohydrate, or nucleic acid.*
MS.5 *Create illustrations that serve as models of the three-dimensional structures of nucleic acids, proteins, carbohydrates, lipids, and phospholipid bilayers. Generate multiple models to span several levels of detail and abstraction, such as specific molecule structure all the way to a general function in a cell.*
MS.22 *Create a "parts-level" abstract sketch of a generic glycerophospholipid that includes: (a) lipid tails, (b) glycerol, (c) phosphate, and (d) a "decoration" (e.g. ethanolamine). This need not be an atom-by-atom structure but rather a cartoon depicting the order and types of bonds between structural elements.*
MS.20 *Draw how water molecules might interact with hydrophobic parts of molecules.*
MS.21 *Describe how the chemical structures and properties of lipids contribute to their functions (e.g. formation of membrane structures, contribution to membrane fluidity, etc.).*
MS.23 *Compare and contrast the influence of different phospholipid chain lengths and degree of saturation in membrane lipids on membrane fluidity.*
MS.24 *Create simple illustrations of carbohydrates that include the major functional groups, the formation of glycosidic bonds, and the potential interactions of carbohydrates with water molecules or other biomolecules.*
MS.25 *Diagram the pentose and hexose sugars, be able to number their carbon atoms, and identify the key functional groups on each molecule.*
MS.8 *Describe the structure and all the parts of an amino acid and a polypeptide.*
MS.3 *If given the R-groups of amino acids, be able to categorize them as nonpolar, polar, acidic or basic at physiological pH.*
MS.7 *Given two or more amino acids, pKa values for functional groups, and a specified pH, draw a figure that depicts the amino acids linked by peptide bonds, identifies backbone and side-chains, products of the condensation reactions, and water molecules interacting with the structure.*
MS.9 *Understand how the 1° sequence influences the 2°, 3° and 4° structures of a protein.*
MS.11 *Relate basic structures, such as alpha-helices and beta-pleated sheets to the tertiary structure of a protein.*
MS.12 *Describe and discuss the types of bonds and the portions of the amino acids (carboxyl group, amino group, R group, alpha carbon) that are responsible for the formation of the 1°, 2°, 3°, and 4° structure of a protein.*
MS.13 *Create simple cartoon models depicting secondary, tertiary, and quaternary protein structures.*
MS.15 *Understand how changes to amino acids in the binding pocket and/or pH can alter small molecule binding to proteins, if given information regarding the amino acids in the protein binding site.*
MS.26 *Identify a nucleotide from its molecular structure and be able to decompose the molecule into three main functional units: nitrogenous base, ribose, and phosphates.*

Lipids



Lipids are a diverse group of hydrophobic compounds that include molecules like fats, oils, waxes, phospholipids, and steroids. Most lipids are at their core hydrocarbons, molecules that include many nonpolar carbon-carbon or carbon-hydrogen bonds. The abundance of nonpolar functional groups give lipids a degree of hydrophobic (“water fearing”) character and most lipids have low solubility in water. Depending on their physical properties (encoded by their chemical structure), lipids can serve many functions in biological systems including energy storage, insulation, barrier formation, cellular signaling. The diversity of lipid molecules and their range of biological activities are perhaps surprisingly large to most new students of biology. Let's start by developing a core understanding of this class of biomolecules.

Fats and oils

A common fat molecule or **triglyceride**. These types of molecules are hydrophobic and, while they have many functions, are probably best known for their roles in body fat and plant oils. A triglyceride molecule derived from two types of molecular components—a polar "head" group and a nonpolar "tail" group. The "head" group of a triglyceride is derived from a single glycerol

molecule. Glycerol, a carbohydrate, is composed of three carbons, five hydrogens, and three hydroxyl (-OH) functional groups. The nonpolar **fatty acid** "tail" group comprises three hydrocarbons (a functional group composed of C-H bonds) that also have a polar carboxyl functional group (hence the term "fatty acid"—the carboxyl group is acidic at most biologically relevant pHs). The number of carbons in the fatty acid may range from 4–36; Most common are those containing 12–18 carbons.

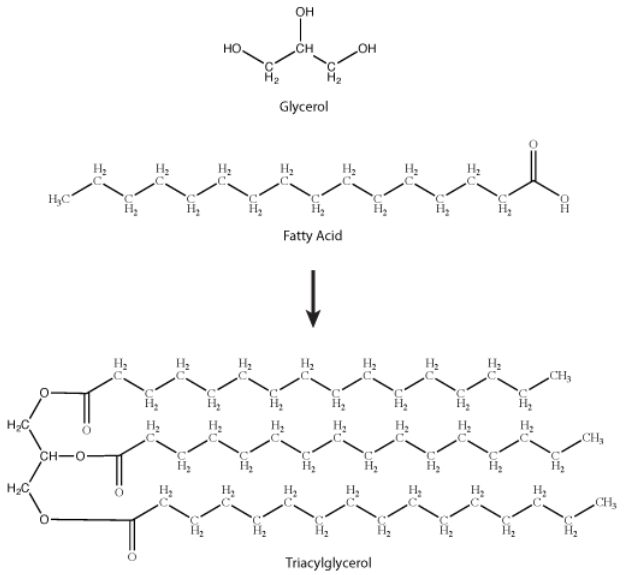


Figure 1. *Triacylglycerol is formed by the joining of three fatty acids to a glycerol backbone in a dehydration reaction. Three molecules of water are released in the process. Attribution: Marc T. Facciotti (own work).*

The models of the triglycerides shown above depict the relative positions of the atoms in the molecule. If you Google for images of triglycerides you will find some models that show the phospholipid tails in different positions from those depicted above. Using your intuition, can you give an opinion for which model you think is a more correct representation of real life? Why?

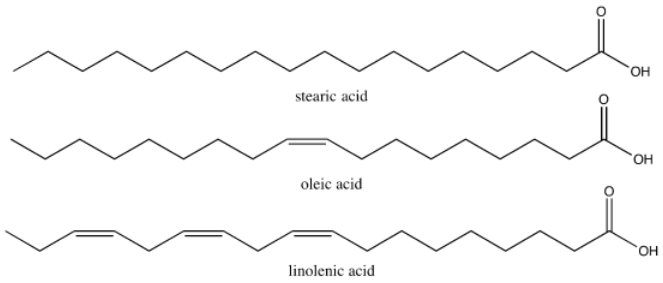


Figure 2. *Stearic acid is a common saturated fatty acid; oleic acid and linolenic acid are common unsaturated fatty acids. Attribution: Marc T. Facciotti (own work)*

Natural fats like butter, canola oil, etc., are composed mostly of triglycerides. The physical properties of these different fats vary depending on two factors:

1. The number of carbons in the hydrocarbon chains;
2. The number of desaturations, or double bonds, in the hydrocarbon chains.

The first factor influences how these molecules interact with each other and with water, while the second factor dramatically influences their shape. Introducing a double bond causes a "kink" in the otherwise relatively "straight" hydrocarbon, depicted in a slightly exaggerated was in Figure 3.

Based on what you can understand from this brief description, propose a rationale—in your own words—to explain why butter is solid at room temperature while vegetable oil is liquid.

Here is an additional piece of information that could help you answer question: butter has a greater percentage of longer and saturated hydrocarbons in its triglycerides than does vegetable oil.

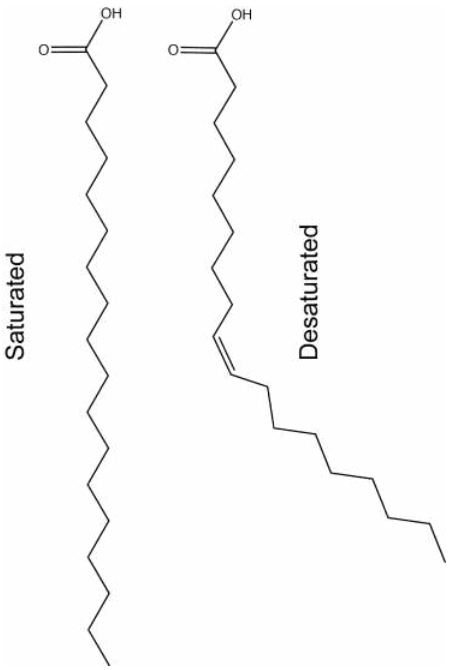


Figure 3. *The straight saturated fatty acid versus the "bent"/"kinked" unsaturated fatty acid. Attribution: Marc T. Facciotti (own work)*

Real-Life connection:

Ever thought about how lipids are important to vision? Read more here.

Sterols

Steroids are lipids with a fused ring structure. Although they do not resemble the other lipids discussed here, they are designated as lipids because they are also largely composed of carbons and hydrogens, are hydrophobic, and are insoluble in water. All steroids have four linked carbon rings. Many steroids also have the -OH functional group which puts them in the alcohol classification of sterols. Several steroids, like cholesterol, have a short tail. Cholesterol is the most common steroid. It is mainly synthesized in the liver and is the precursor to many steroid hormones such as testosterone. It is also the precursor to Vitamin D and of bile salts which help in the emulsification of fats and their subsequent absorption by cells. Although cholesterol is often spoken of in negative terms, it is necessary for the proper functioning of many animal cells, particularly in its role as a component of the plasma membrane where it is known to modulate membrane structure, organization, and fluidity.

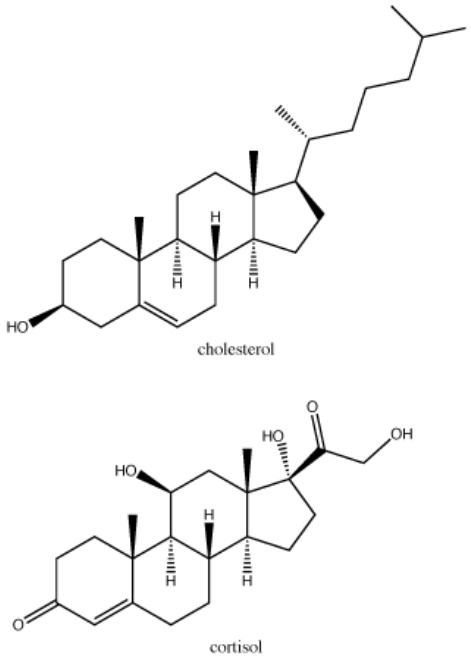
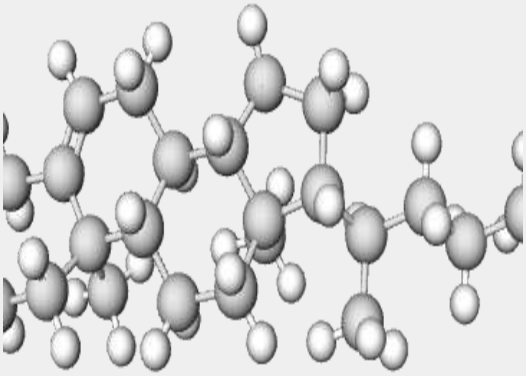
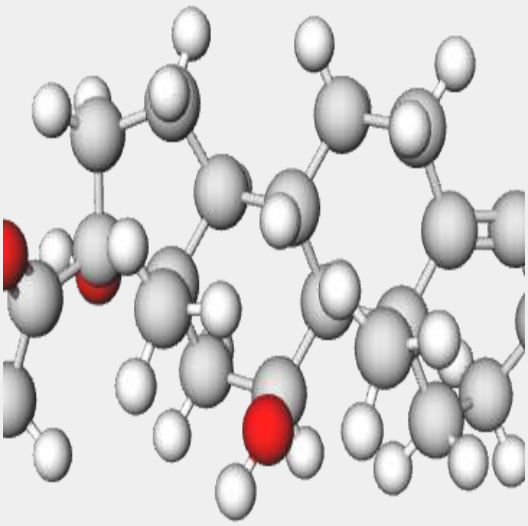


Figure 4. *Cholesterol is a modified lipid molecule that is synthesized by animal cells and is a key structural element in cellular membranes. Cortisol is a hormone (signaling molecule) that is often released in response to stress. Attribution: Marc T. Facciotti (own work)*

Cholesterol	Cortisol
	

Phospholipids

Phospholipids are major constituents of the cell membrane, the outermost layer of cells. Like fats, they are composed of fatty acid chains attached to a glycerol molecule. Unlike the triacylglycerols, phospholipids have two fatty acid tails and a phosphate group attached to the sugar. Phospholipid are therefore **amphipathic** molecules, meaning it they have a hydrophobic part and a hydrophilic part. The two fatty acid chains extending from the glycerol are hydrophobic and cannot interact with water, whereas the phosphate-containing head group is hydrophilic and interacts with water. Can you identify the functional groups on the phospholipid below that give each part of the phospholipid its properties?

Note

Note in Figure 5 that the phosphate group has an R group linked to one of the oxygen atoms. R is a variable commonly used in these types of diagrams to show some other atom or molecule bound at that position. That part of the molecule can be different in different phospholipids—and will impart some different chemistry to the whole molecule. At the moment, however, you are responsible for being able to recognize this type of molecule (no matter what the R group is) because of the common core elements—the glycerol backbone, the phosphate group, and the two hydrocarbon tails.

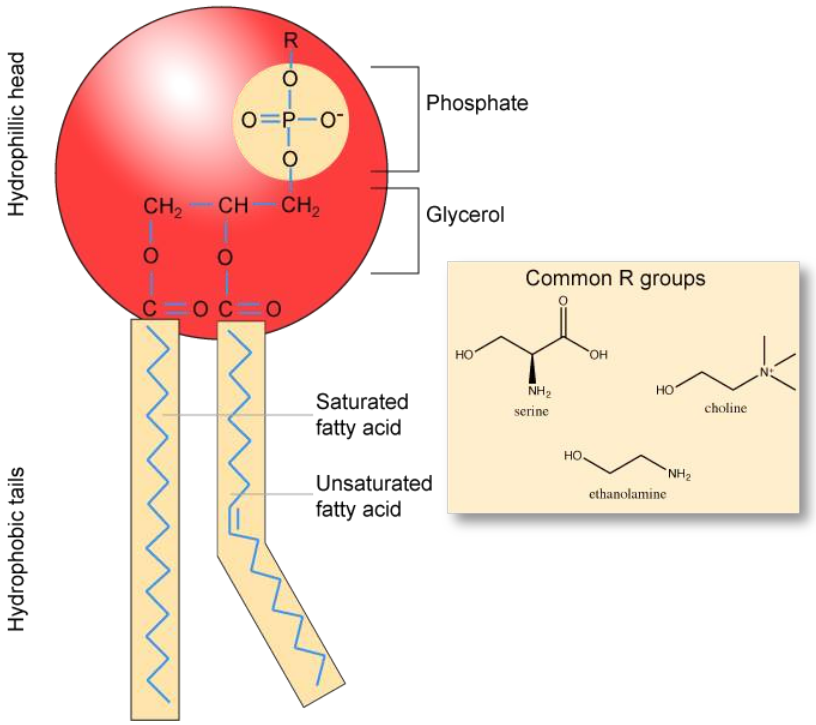
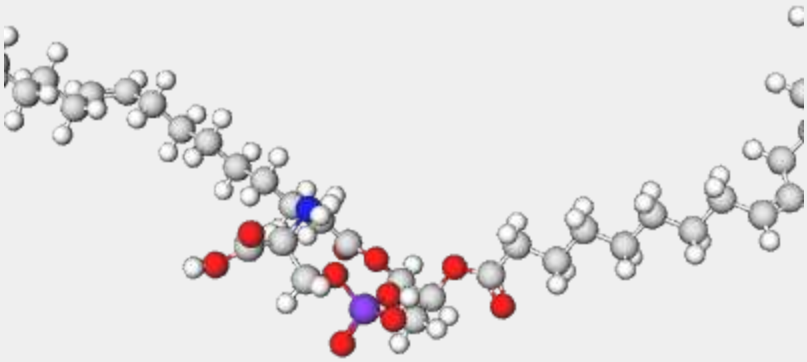


Figure 5. *A phospholipid is a molecule with two fatty acids and a modified phosphate group attached to a glycerol backbone. The phosphate may be modified by the addition of charged or polar chemical groups. Several chemical R groups may modify the phosphate. Choline, serine, and*

ethanolamine are shown here. These attach to the phosphate group at the position labeled R via their hydroxyl groups. Attribution: Marc T. Facciotti (own work).

1,2-Dioleoyl-sn-glycero-3-phospho-L-serine



In the presence of water, some phospholipids will spontaneously arrange themselves into a micelle (Figure 6). The lipids arrange such that their polar groups are on the outside of the micelle, and the nonpolar tails are on the inside. Under other conditions, a lipid bilayer can also form. This structure, only a few nanometers thick, is composed of two opposing layers of phospholipids such that all the hydrophobic tails align face-to-face in the center of the bilayer and are surrounded by the hydrophilic head groups. A phospholipid bilayer forms as the basic structure of most cell membranes and are responsible for the dynamic nature of the plasma membrane.

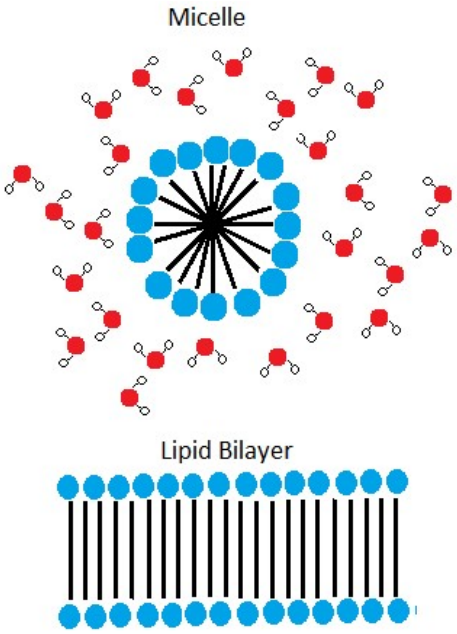


Figure 6. *In the presence of water, some phospholipids will spontaneously arrange themselves into a micelle. Source: Created by Erin Easlon (own work)*

As mentioned above, if you were to take some pure phospholipids and drop them into water that some phospholipid would spontaneously form into micelles. This sounds like a process that an Energy Story could describe. Go back to the Energy Story rubric and see if you can try to create an Energy Story for this process — I expect that the steps involving the description of energy might be difficult at this point (we'll come back to that later) but you should be able to do at least the first three steps.

We discuss the phospholipid membrane in a later module. It is important to remember the chemical properties associated with the functional groups in the phospholipid to understand the function of the cell membrane.

Carbohydrates



Carbohydrates are one of the four main classes of macromolecules that make up all cells and are an essential part of our diet; grains, fruits, and vegetables are all natural sources. While we may be most familiar with the role carbohydrates play in nutrition, they also have a variety of other essential functions in humans, animals, plants, and bacteria. In this section, we will discuss and review basic concepts of carbohydrate structure and nomenclature, and a variety of functions they play in cells.

Molecular structures

In their simplest form, **carbohydrates** can be represented by the stoichiometric formula $(CH_2O)_n$, where $_n$ is the number of carbons in the molecule. For simple carbohydrates, the ratio of carbon-to-hydrogen-to-oxygen in the molecule is 1:2:1. This formula also explains the origin of the term “carbohydrate”: the components are carbon (“carbo”) and the components of water (“hydrate”). Simple carbohydrates are classified into three subtypes: monosaccharides, disaccharides, and polysaccharides, which will be discussed below. While simple carbohydrates fall nicely into this 1:2:1 ratio, carbohydrates can also be structurally more complex. For example, many carbohydrates contain functional groups (remember them from our basic discussion about chemistry) besides the obvious hydroxyl. For example, carbohydrates can have phosphates or amino groups substituted at a variety of sites within the molecule. These functional groups can provide additional properties to the molecule and will alter its overall function. However, even with these types of substitutions, the basic overall structure of the carbohydrate is retained and easily identified.

Nomenclature

One issue with carbohydrate chemistry is the nomenclature. Here are a few quick and simple rules:

1. Simple carbohydrates, such as glucose, lactose, or dextrose, end with an "-ose."
2. Simple carbohydrates can be classified based on the number of carbon atoms in the molecule, as with triose (three carbons), pentose (five carbons), or hexose (six carbons).
3. Simple carbohydrates can be classified based on the functional group found in the molecule, i.e ketose (contains a ketone) or aldose (contains an aldehyde).
4. Polysaccharides are often organized by the number of sugar molecules in the chain, such as in a monosaccharide, disaccharide, or trisaccharide.

For a short video on carbohydrate classification, see the 10-minute Khan Academy video by clicking here.

Monosaccharides

Monosaccharides ("mono-" = one; "sacchar-" = sweet) are simple sugars; the most common is glucose. In monosaccharides, the number of carbons usually ranges from three to seven. If the sugar has an aldehyde group (the functional group with the structure R-CHO), it is known as an aldose; if it has a ketone group (the functional group with the structure RC(=O)R'), it is known as a ketose.

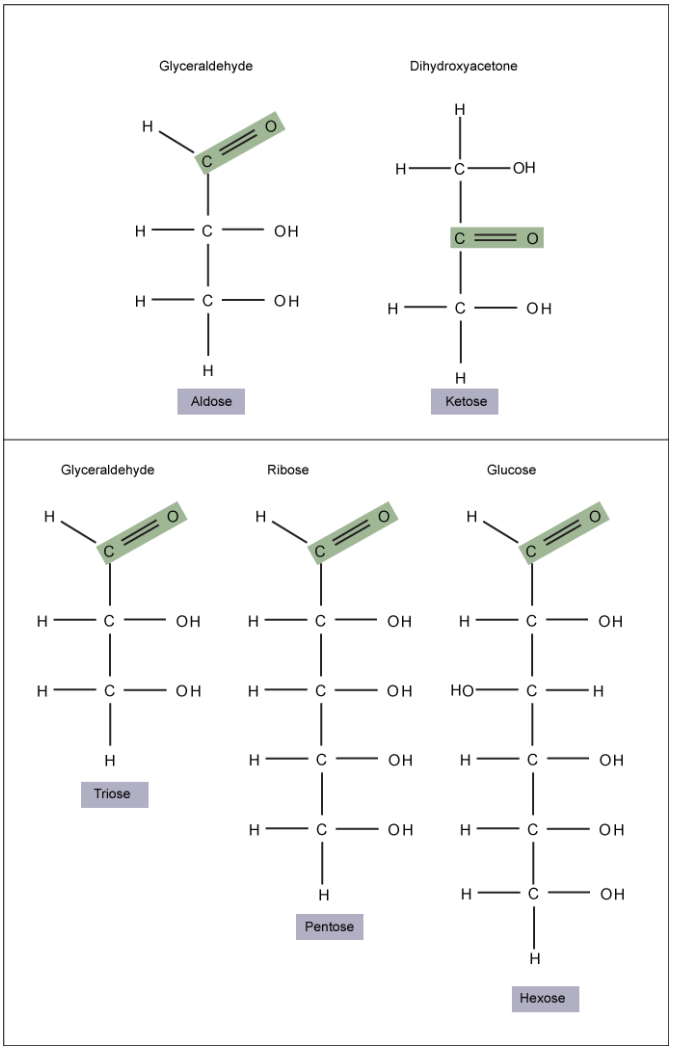


Figure 1. *Monosaccharides are classified based on the position of their carbonyl group and the number of carbons in the backbone. Aldoses have a carbonyl group (indicated in green) at the end of the carbon chain and ketoses have a carbonyl group in the middle of the carbon chain. Trioses, pentoses, and hexoses have three, five, and six carbons in their backbones, respectively. Attribution: Marc T. Facciotti (own work).*

Glucose versus galactose

Galactose (part of lactose, or milk sugar) and **glucose** (found in sucrose, glucose disaccharride) are other common monosaccharides. The chemical formula for glucose and galactose is $C_6H_{12}O_6$; both are hexoses, but the arrangements of the hydrogens and hydroxyl groups are different at position C_4. Because of this small difference, they differ structurally and chemically and are known as chemical isomers because of the different arrangement of functional groups around the asymmetric carbon; both of these monosaccharides have more than one asymmetric carbon (compare the structures in the figure below).

Fructose versus both glucose and galactose

A second comparison can be made when looking at glucose, galactose, and **fructose** (the second carbohydrate that with glucose makes up the disaccharide sucrose and is a common sugar found in fruit). All three are hexoses; however, there is a major structural difference between glucose and galactose versus fructose: the carbon that contains the **carbonyl** (C=O).

In glucose and galactose, the carbonyl group is on the C_1 carbon, forming an **aldehyde** group. In fructose, the carbonyl group is on the C_2 carbon, forming a **ketone** group. The former sugars are called **aldoses** based on the aldehyde group that is formed; the latter is designated as a **ketose** based on the ketone group. Again, this difference gives fructose different chemical and structural properties from those of the aldoses, glucose, and galactose, even though fructose, glucose, and galactose all have the same chemical composition: $C_6H_{12}O_6$. From the standpoint of most General Biology courses, the key takeaway is that, like many other molecules, the core building blocks of carbohydrates (i.e. C, OH, and H) can be reorganized in a number of ways to build chemically and functionally different molecules. In this examples, all three sugars have the same chemical forumla (i.e. the same building blocks), but because those building blocks have been organized a bit differently, each sugar is a unique kind of molecule.

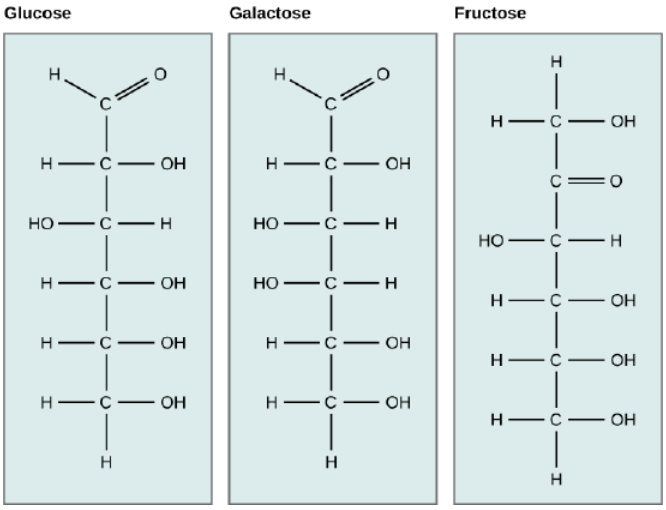


Figure 2. *Glucose, galactose, and fructose are all hexoses. They are structural isomers, meaning they have the same chemical formula ($C_6H_{12}O_6$) but a different arrangement of atoms.*

Linear versus ring form of the monosaccharides

Monosaccharides can exist as a linear chain or as ring-shaped molecules. In aqueous solutions, monosaccharides are usually found in ring form (Figure 3). Glucose in a ring form can have two different arrangements of the hydroxyl group (OH) around the anomeric carbon (C_1 that becomes asymmetric in the process of ring formation). If the hydroxyl group is below C_1 in the sugar, it is said to be in the alpha (α) position, and if it is above C_1 in the sugar, it is said to be in the beta (β) position.

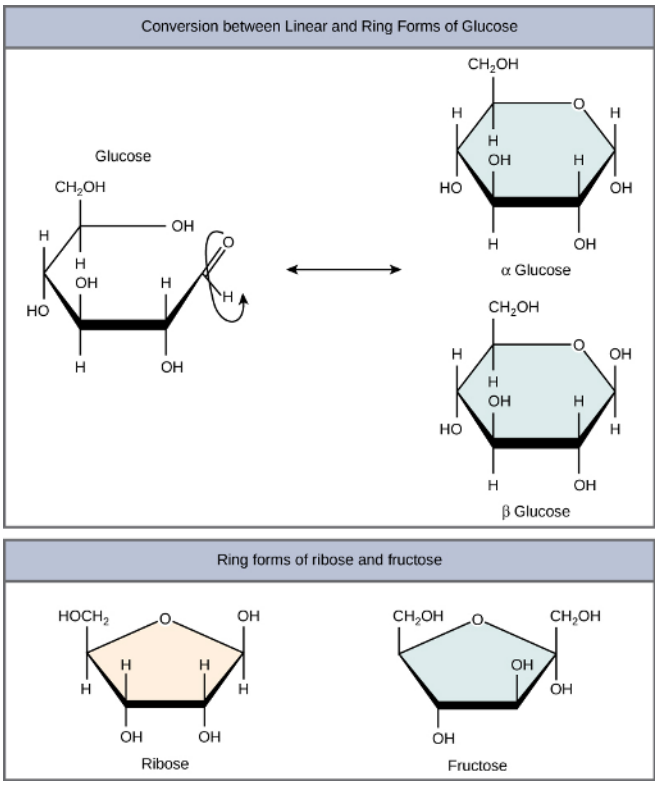


Figure 3. *Five- and six-carbon monosaccharides exist in equilibrium between linear and ring form. When the ring forms, the side chain it closes on is locked into an α or β position. Fructose and ribose also form rings, although they form five-membered rings as opposed to the six-membered ring of glucose.*

Disaccharides

Disaccharides ("di-" = two) form when two monosaccharides undergo a dehydration reaction (also known as a condensation reaction or dehydration synthesis). During this process, the hydroxyl group of one monosaccharide combines with the hydrogen of another monosaccharide, releasing a molecule of water and forming a covalent bond. A covalent bond formed between a carbohydrate molecule and another molecule (in this case, between two monosaccharides) is known as a **glycosidic** bond. Glycosidic bonds (also called glycosidic linkages) can be of the alpha or the beta type.

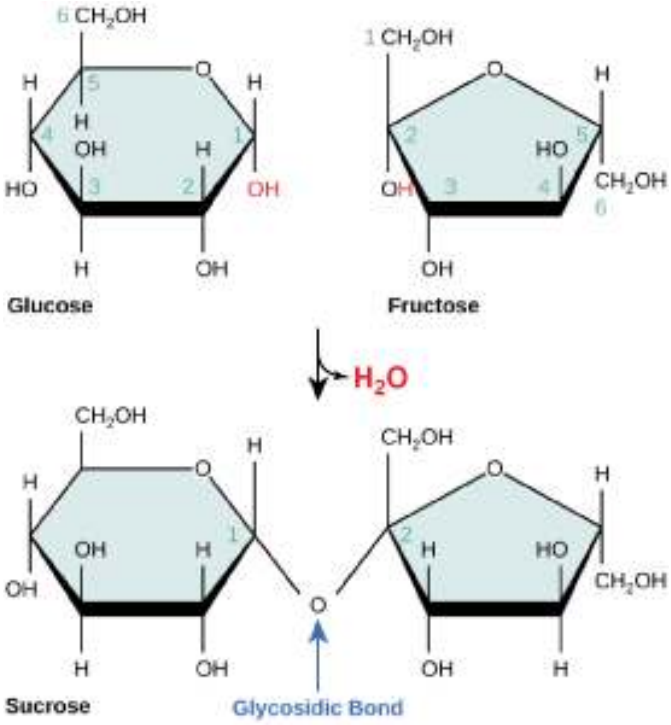


Figure 4. *Sucrose is formed when a monomer of glucose and a monomer of fructose are joined in a dehydration reaction to form a glycosidic bond. In the process, a water molecule is lost. By convention, the carbon atoms in a monosaccharide are numbered from the terminal carbon closest to the carbonyl group. In sucrose, a glycosidic linkage is formed between the C_1 carbon in glucose and the C_2 carbon in fructose.*

Common disaccharides include lactose, maltose, and sucrose (Figure 5). Lactose is a disaccharide consisting of the monomers glucose and galactose. It is found naturally in milk. Maltose, or malt/grain sugar, is a disaccharide formed by a dehydration reaction between two glucose molecules. The most common disaccharide is sucrose, or table sugar, which is composed of the monomers glucose and fructose.

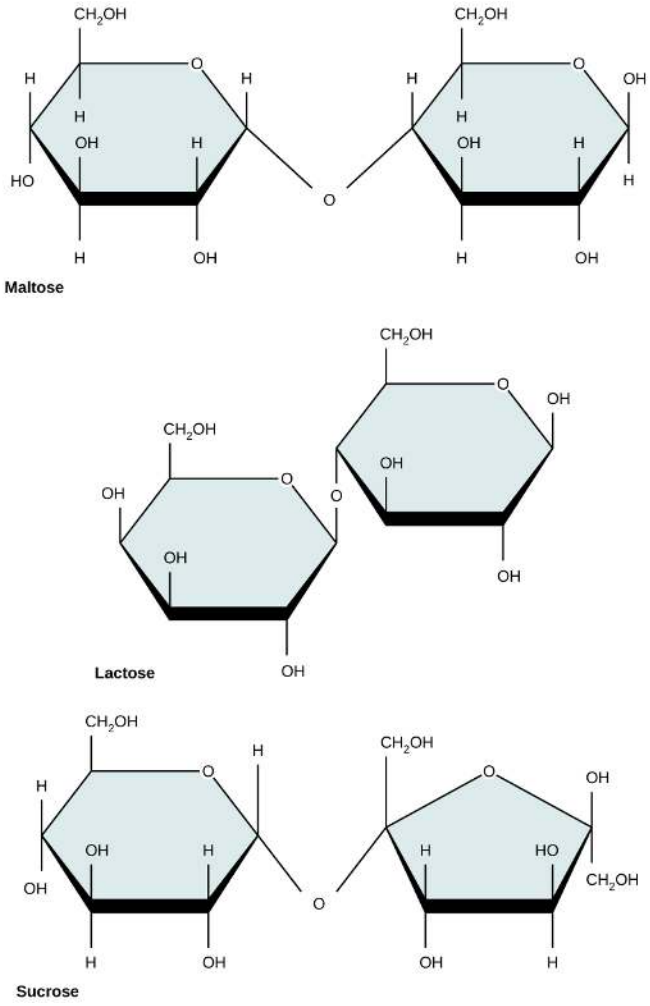
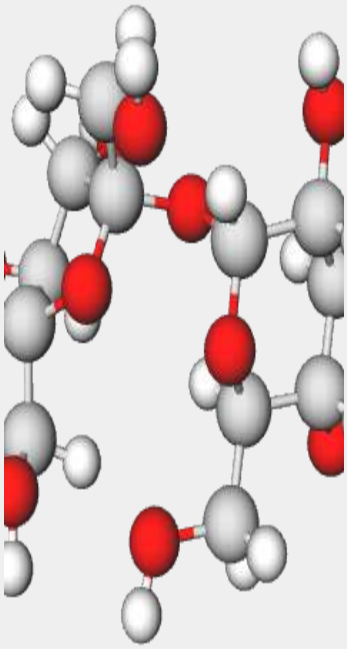
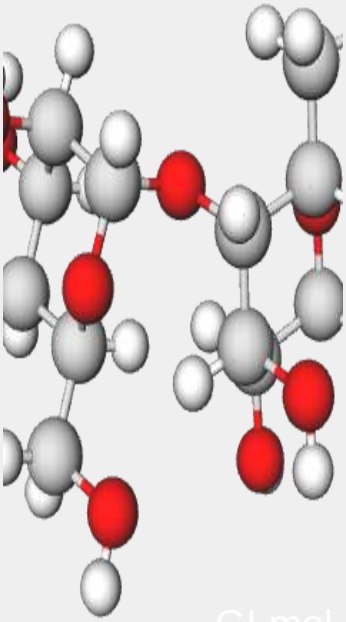
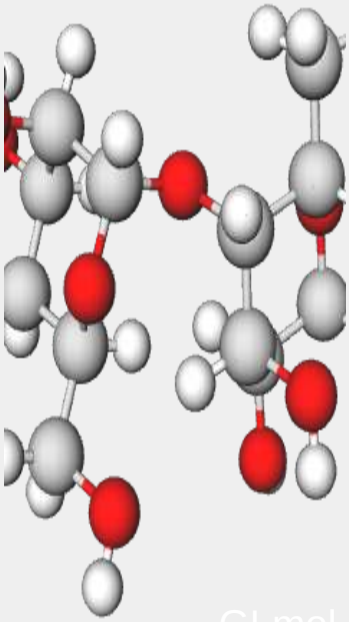


Figure 5. *Common disaccharides include maltose (grain sugar), lactose (milk sugar), and sucrose (table sugar).*

Sucrose	Lactose	Maltose
		

Polysaccharides

A long chain of monosaccharides linked by glycosidic bonds is known as a **polysaccharide** ("poly-" = many). The chain may be branched or unbranched, and it may contain different types of monosaccharides. The molecular weight may be 100,000 Daltons or more, depending on the number of monomers joined. Starch, glycogen, cellulose, and chitin are primary examples of polysaccharides.

Starch is the stored form of sugars in plants and is made up of a mixture of amylose and amylopectin; both are polymers of glucose. Plants are able to synthesize glucose. Excess glucose, the amount synthesized that is beyond the plant's immediate energy needs, is stored as starch in different plant parts, including roots and seeds. The starch in the seeds provides food for the embryo as it germinates and can also act as a source of food for humans and animals who may eat the seed. Starch that is consumed by humans is broken down by enzymes, such as salivary amylases, into smaller molecules, such as maltose and glucose.

Starch is made up of glucose monomers that are joined by 1-4 or 1-6 glycosidic bonds; the numbers 1-4 and 1-6 refer to the carbon number of the two residues that have joined to form the bond. As illustrated in Figure 6, amylose is starch formed by unbranched chains of glucose monomers (only 1-4 linkages), whereas amylopectin is a branched polysaccharide (1-6 linkages at the branch points).

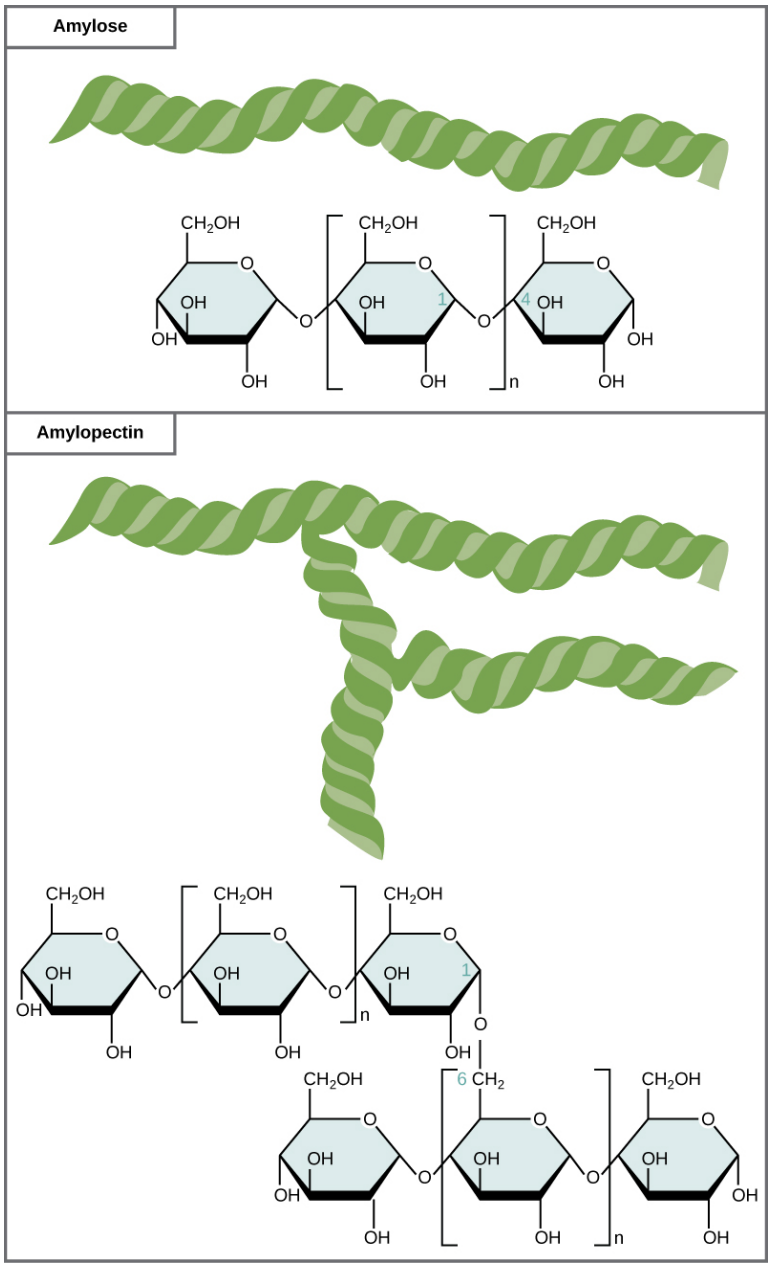
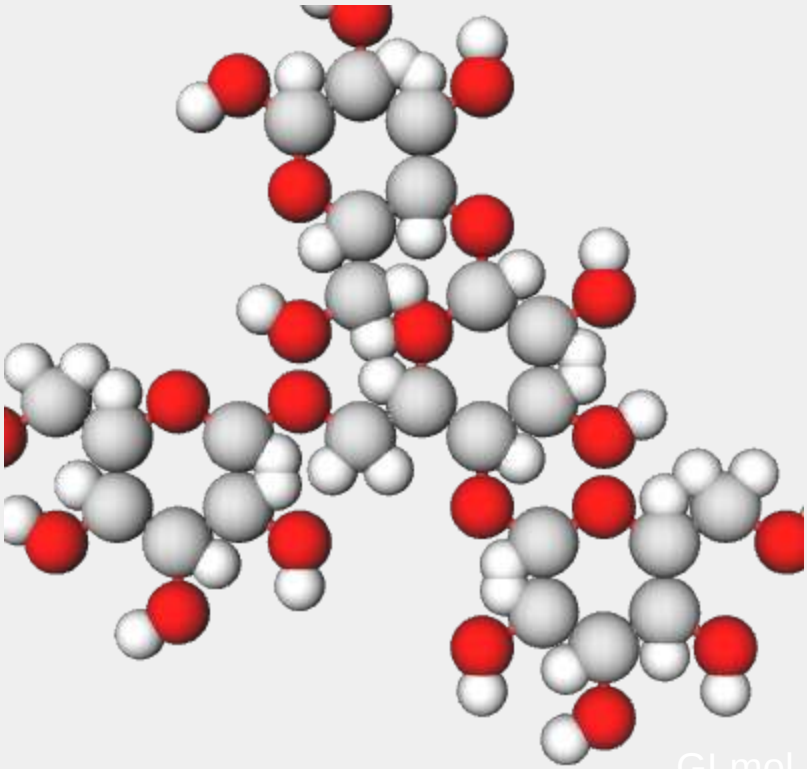


Figure 6. *Amylose and amylopectin are two different forms of starch. Amylose is composed of unbranched chains of glucose monomers connected by 1-4 glycosidic linkages. Amylopectin is composed of branched chains of glucose monomers connected by 1-4 and 1-6 glycosidic linkages. Because of the way the subunits are joined, the glucose chains have a helical structure. Glycogen (not shown) is similar in structure to amylopectin but more highly branched.*

Glycogen

Glycogen is a common stored form of glucose in humans and other vertebrates. Glycogen is the animal equivalent of starch and is a highly branched molecule usually stored in liver and muscle cells. Whenever blood glucose levels decrease, glycogen is broken down to release glucose in a process known as glycogenolysis.

Glycogen



Cellulose

Cellulose is the most abundant natural biopolymer. The cell wall of plants is mostly made of cellulose, which provides structural support to the cell. Wood and paper are mostly cellulosic in nature. Cellulose is made up of glucose monomers that are linked by β 1-4 glycosidic bonds.

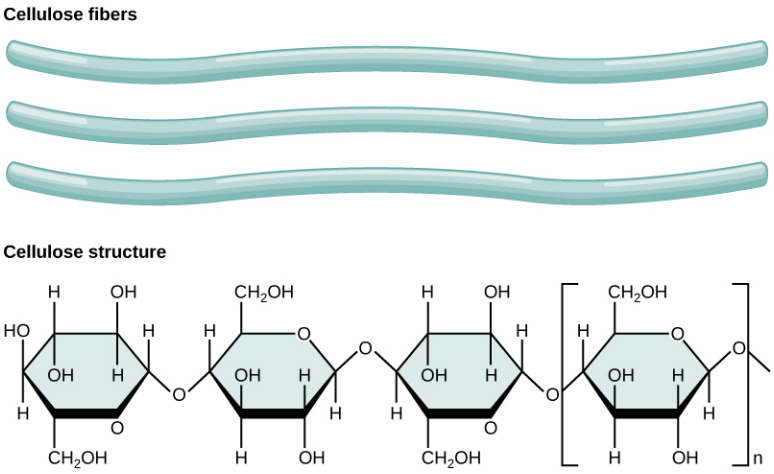


Figure 7. *In cellulose, glucose monomers are linked in unbranched chains by β 1-4 glycosidic linkages. Because of the way the glucose subunits are joined, every glucose monomer is flipped relative to the next one, resulting in a linear, fibrous structure.*

As shown in the figure above, every other glucose monomer in cellulose is flipped over, and the monomers are packed tightly as extended, long chains. This gives cellulose its rigidity and high tensile strength—which is so important to plant cells. While the β 1-4 linkage cannot be broken down by human digestive enzymes, herbivores such as cows, koalas, buffalos, and horses are able, with the help of the specialized flora in their stomach, to digest plant material that is rich in cellulose and use it as a food source. In these animals, certain species of bacteria and protists reside in the rumen (part of the digestive system of herbivores) and secrete the enzyme cellulase. The appendix of grazing animals also contains bacteria that digest cellulose, giving it an important role in the digestive systems of ruminants. Cellulases can break down cellulose into glucose monomers that can be used as an energy source by the animal. Termites are also able to break down cellulose because of the presence of other organisms in their bodies that secrete cellulases.

Interactions with carbohydrates

We have just discussed the various types and structures of carbohydrates found in biology. The next thing to address is how these compounds interact with other compounds. The answer to that is that it depends on the final structure of the carbohydrate. Because carbohydrates have many hydroxyl groups associated with the molecule, they are therefore excellent **H-bond donors** and **acceptors**. Monosaccharides can quickly and easily form H-bonds with water and are readily soluble. All of those H-bonds also make them quite "sticky". This is also true for many disaccharides and many short-chain polymers. Longer polymers may not be readily soluble.

Finally, the ability to form a variety of H-bonds allows polymers of carbohydrates or **polysaccharides** to form strong intramolecular and intermolocular bonds. In a polymer, because there are so many H-bonds, this can provide a lot of strength to the molecule or molecular complex, especially if the polymers interact. Just think of cellulose, a polymer of glucose, if you have any doubts.

Possible NB Discussion Point

Lipids and carbohydrates are not just classes of macromolecules that we discuss in BIS 2A but are also two of the essential macronutrients that we can obtain from eating various foods. Some popularized diet programs (e.g., Atkins, ketogenic) suggest limiting carbohydrates and/or fats. As you learn more about biomolecules and their roles in living systems, are you refining your perspective on foods and diets? What have you learned so far? Do you think there is anything missing in your understanding? Are you able to better understand and evaluate certain diets, such as the aforementioned?

Nucleic acids

There are two types of nucleic acids in biology: DNA and RNA. DNA carries the heritable genetic information of the cell and is composed of two antiparallel strands of nucleotides arranged in a helical structure. Each nucleotide subunit is composed of a pentose sugar (deoxyribose), a nitrogenous base, and a phosphate group. The two strands associate via hydrogen bonds between chemically complementary nitrogenous bases. Interactions known as "base stacking" interactions also help stabilize the double helix. By contrast to DNA, RNA can be either be single stranded, or double stranded. It too is composed of a pentose sugar (ribose), a nitrogenous base, and a phosphate group. RNA is a molecule of many tricks. It is involved in protein synthesis as a messenger, regulator, and catalyst of the process. RNA is also involved in various other cellular regulatory processes and helps to catalyze some key reactions (more on this later). With respect to RNA, in this course we are primarily interested in (a) knowing the basic molecular structure of RNA and what distinguishes it from DNA, (b) understanding the basic chemistry of RNA synthesis that occurs during a process called transcription, (c) appreciating the various roles that RNA can have in the cell, and (d) learning the major types of RNA that you will encounter most frequently (i.e. mRNA, rRNA, tRNA, miRNA etc.) and associating them with the processes they are involved with. In this module we focus primarily on the chemical structures of DNA and RNA and how they can be distinguished from one another.

Nucleotide structure

The two main types of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA and RNA are made up of monomers known as **nucleotides**. Individual nucleotides condense with one another to form a **nucleic acid** polymer. Each nucleotide is made up of three components: a nitrogenous base (for which there are five different types), a pentose sugar, and a phosphate group. These are depicted below. The main difference between these two types of nucleic acids is the presence or absence of a hydroxyl group at the C_2 position, also called the 2' position (*read "two prime"*), of the pentose (see Figure 1 legend and the section on the pentose sugar for more on carbon numbering). RNA has a hydroxyl functional group at that 2' position of the pentose sugar; the sugar is called ribose, hence the name ribonucleic acid. By contrast, DNA lacks the hydroxyl group at that position, hence the name, "deoxy" ribonucleic acid. DNA has a hydrogen atom at the 2' position.

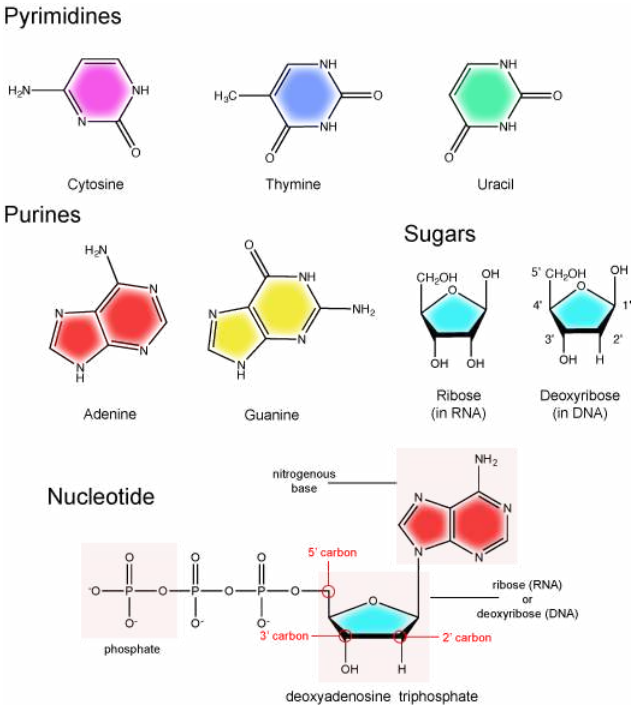


Figure 1. *A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and one or more phosphate groups. Carbons in the pentose are numbered 1' through 5' (the prime distinguishes these residues from those in the base, which are numbered without using a prime notation). The base is attached to the 1' position of the ribose, and the phosphate is attached to the 5' position. When a polynucleotide is formed, the 5' phosphate of the incoming nucleotide attaches to the 3' hydroxyl group at the end of the growing chain. Two types of pentose are found in nucleotides, deoxyribose (found in DNA) and ribose (found in RNA). Deoxyribose is similar in structure to ribose, but it has an -H instead of an -OH at the 2' position. Bases can be divided into two categories: purines and pyrimidines. Purines have a double ring structure, and pyrimidines have a single ring.*
Attribution: Marc T. Facciotti (original work)

The nitrogenous base

The nitrogenous bases of nucleotides are organic molecules and are so named because they contain carbon and nitrogen. They are bases because they contain an amino group that has the potential of binding an extra hydrogen, and thus acting as a base by decreasing the hydrogen ion concentration in the local environment. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G), cytosine (C), and thymine (T). By contrast, RNA contains adenine (A), guanine (G) cytosine (C), and uracil (U) instead of thymine (T).

Adenine and guanine are classified as **purines**. The primary distinguishing structural feature of a purine is double carbon-nitrogen ring. Cytosine, thymine, and uracil are classified as **pyrimidines**. These are structurally distinguished by a single carbon-nitrogen ring. You will be expected to recognize that each of these ring structures is decorated by functional groups that may be involved in a variety of chemistries and interactions.

Possible NB Discussion Point

Take a moment to review the five nitrogenous bases in Figure 1 above. Identify functional groups as described in class. For each functional group identified, describe what type(s) of chemistry you expect it to be involved in. Will the functional group act as a hydrogen bond donor, acceptor, or both?

The pentose sugar

The pentose sugar contains five carbon atoms. Each carbon atom of the sugar molecule are numbered as 1′, 2′, 3′, 4′, and 5′ (1′ is read as “one prime”). The two main functional groups that are attached to the sugar are often named in reference to the carbon to which they are bound. For example, the phosphate residue is attached to the 5′ carbon of the sugar and the hydroxyl group is attached to the 3′ carbon of the sugar. We will often use the carbon number to refer to functional groups on nucleotides so be very familiar with the structure of the pentose sugar.

The pentose sugar in DNA is called deoxyribose, and in RNA, the sugar is ribose. The difference between the sugars is the presence of the hydroxyl group on the 2' carbon of the ribose and its absence on the 2' carbon of the deoxyribose. You can, therefore, determine if you are looking at a DNA or RNA nucleotide by the presence or absence of the hydroxyl group on the 2' carbon atom —you will probably be asked to do so on numerous occasions, including exams.

The phosphate group

There can be anywhere between one and three phosphate groups bound to the 5' carbon of the sugar. When one phosphate is bound, the nucleotide is referred to as a Nucleotide MonoPhosphate (NMP). If two phosphates are bound the nucleotide is referred to as Nucleotide DiPhosphate (NDP). When three phosphates are bound to the nucleotide it is referred to as a Nucleotide TriPhosphate (NTP). The phosphoanhydride bonds between that link the phosphate groups to each other have specific chemical properties that make them good for various biological functions. The hydrolysis of the bonds between the phosphate groups is thermodynamically exergonic in biological conditions; Nature has evolved many mechanisms to couple this negative change in free energy to help drive many reactions in the cell. Figure 2 shows the structure of the nucleotide triphosphate Adenosine Triphosphate, ATP, which we will discuss in greater detail in other chapters.

Note: "high-energy" bonds

The term "high-energy bond" is used A LOT in biology. This term is, however, a verbal shortcut that can cause some confusion. The term refers to the amount of negative free energy associated with the hydrolysis of the bond in question. The water (or other equivalent reaction partner) is an important contributor to the energy calculus. In ATP, for instance, simply "breaking" a phosphoanhydride bond - say with imaginary molecular tweezers - by pulling off a phosphate would not be energetically favorable. We must therefore be careful not to say that breaking bonds in ATP is energetically favorable or that it "releases energy". Rather, we should be more specific, noting that the hydrolysis of the bond is energetically favorable. Some of this common misconception is tied to, in our opinion, the use of the term "high-energy bonds". While in General Biology we have tried to minimize the use of the vernacular "high energy" when referring to bonds, trying instead to describe biochemical reactions by using more specific terms, as students of biology you will no doubt encounter the potentially misleading - though admittedly useful - shortcut "high-energy bond" as you continue in your studies. So, keep the above in mind when you are reading or listening to various discussions in biology. Heck, use the term yourself. Just make sure that you really understand what it refers to.

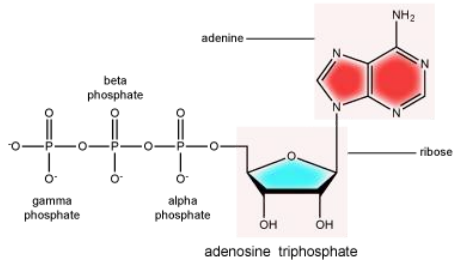
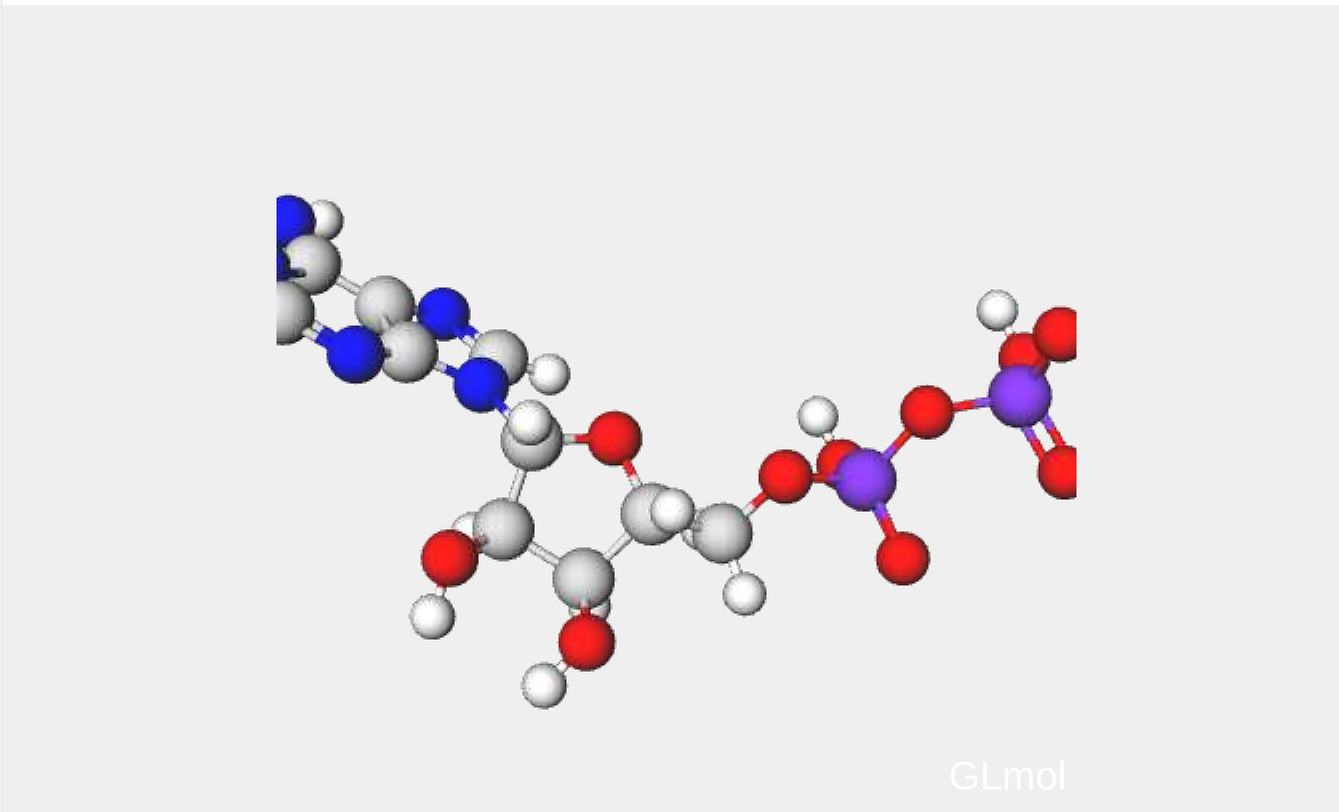


Figure 2. *ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate). Attribution: Marc T. Facciotti (original work)*

Adenosine triphosphate



Double helix structure of DNA

DNA has a double helix structure (shown below) created by two strands of covalently linked nucleotide subunits. The sugar and phosphate groups of each strand of nucleotides are positioned on the outside of the helix, forming the backbone of the DNA (highlighted by the orange ribbons in Figure 3). The two strands of the helix run in opposite directions, meaning that the 5′ carbon end of one strand will face the 3′ carbon end of its matching strand (See Figures 4 and 5). We referred to this orientation of the two strands as **antiparallel**. Note too that phosphate groups are depicted in Figure 3 as orange and red "sticks" protruding from the ribbon. The phosphates are negatively charged at physiological pHs and therefore give the backbone of the DNA a strong local negatively charged character. By contrast, the nitrogenous bases are stacked in the interior of the helix (these are depicted as green, blue, red, and white sticks in Figure 3). Pairs of nucleotides interact with one another through specific hydrogen bonds (shown in Figure 5). Each pair of separated from the next base pair in the ladder by 0.34 nm and this close stacking and planar orientation gives rise to energetically favorable base-stacking interactions. The specific chemistry associated with these interactions is typically beyond the content of a General Biology, but is described in more detail here for the curious or more advanced students. We

do expect, however, that students are aware that the stacking of the nitrogenous bases contributes to the stability of the double helix and defer to your upper-division genetics and organic chemistry instructors to fill in the chemical details.

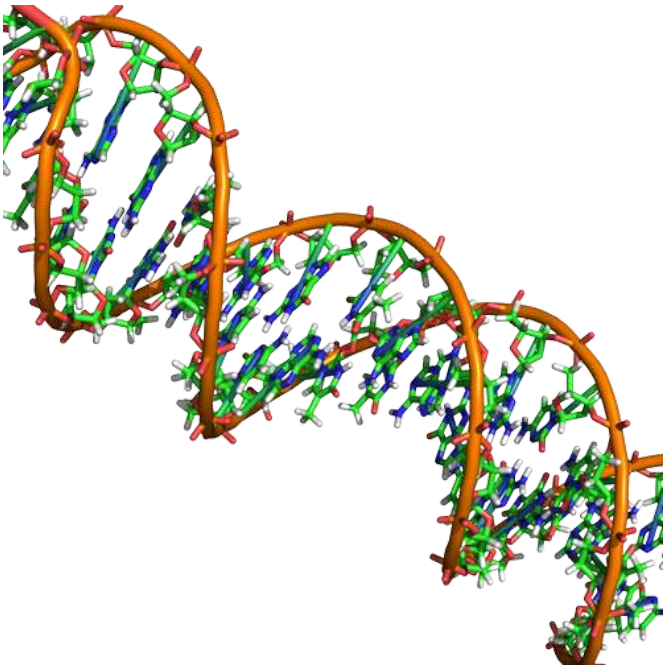


Figure 3. *Native DNA is an antiparallel double helix. The phosphate backbone (indicated by the curvy lines) is on the outside, and the bases are on the inside. Each base from one strand interacts via hydrogen bonding with a base from the opposing strand. Attribution: Marc T. Facciotti (original work)*

In a double helix, certain combinations of base pairing are chemically more favored than others based on the types and locations of functional groups on the nitrogenous bases of each nucleotide. In biology we find that:

Adenine (A) is chemically complementary with thymidine (T) (A pairs with T)

and

Guanine (G) is chemically complementary with cytosine (C) (G pairs with C).

We often refer to this pattern as "base complementarity" and say that the antiparallel strands are **complementary** to each other. For example, if the sequence of one strand is of DNA is 5'-AATTGGCC-3', the complementary strand would have the sequence 5'-GGCCAATT-3'.

We sometimes choose to represent complementary double-helical structures in text by stacking the complementary strands on top of on another as follows:

```
5' - GGCCAATTCCATACTAGGT - 3'
3' - CCGGTTAAGGTATGATCCA - 5'
```

Note that each strand has its 5' and 3' ends labeled and that if one were to walk along each strand starting from the 5' end to the 3' end that the direction of travel would be opposite the other for each strand; the strands are antiparallel. We commonly say things like "running 5-prime to 3-prime" or "synthesized 5-prime to 3-prime" to refer to the direction we are reading a sequence or the direction of synthesis. Start getting yourself accustomed to this nomenclature.

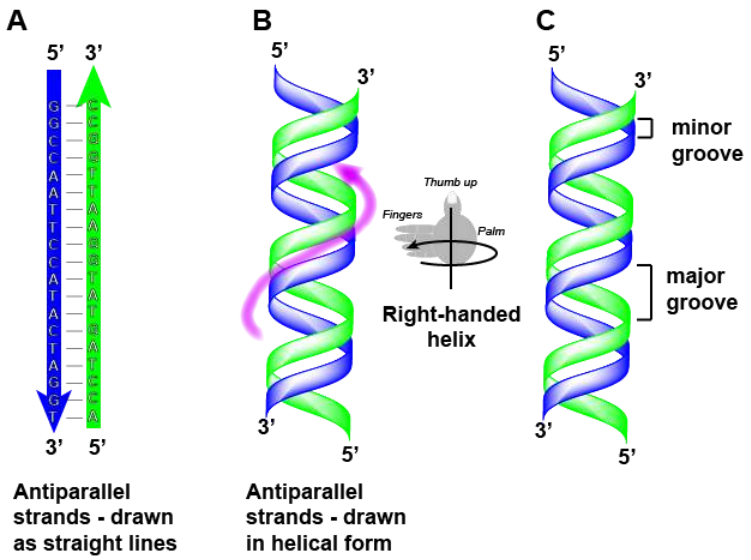


Figure 4. ***Panel A.*** *In a double-stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5′ to 3′ and the other 3′ to 5′. Here the strands are depicted as blue and green lines pointing in the 5′ to 3′ orientation. Complementary base pairing is depicted with a horizontal line between complementary bases.* ***Panel B.*** *The two antiparallel strands are depicted in double-helical form. Note that the orientation of the strands is still represented. Note that the helix is right-handed - the "curl" of the helix, depicted in purple, winds in the direction of the fingers of the hand if the right hand is used and the direction of the helix points towards the thumb.* ***Panel C.*** *This representation shows two structural features that arise from the assembly of the two strands called the major and minor grooves. These grooves can also be seen in Figure 3.*
Attribution: Marc T. Facciotti (original work)

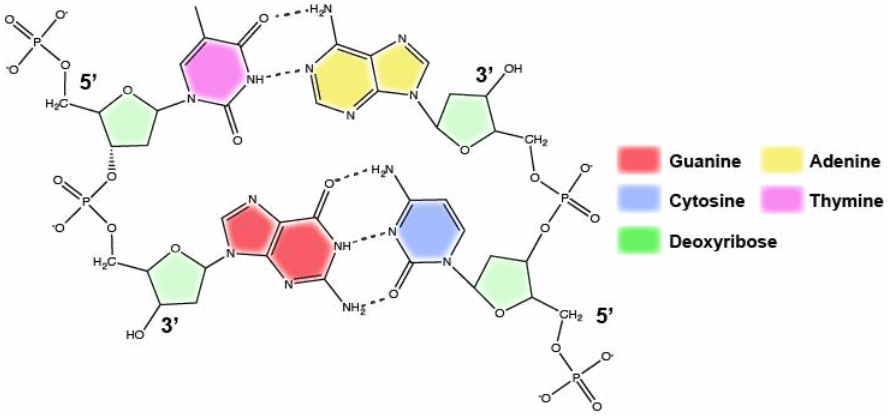


Figure 5. A zoomed-in molecular-level view of the antiparallel strands in DNA. *In a double-stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5′ to 3′ and the other 3′ to 5′. The phosphate backbone is located on the outside, and the bases are in the middle. Adenine forms hydrogen bonds (or base pairs) with thymine, and guanine base pairs with cytosine.*
Attribution: Marc T. Facciotti (original work)

Functions and roles of nucleotides and nucleic acids to look out for in General Biology

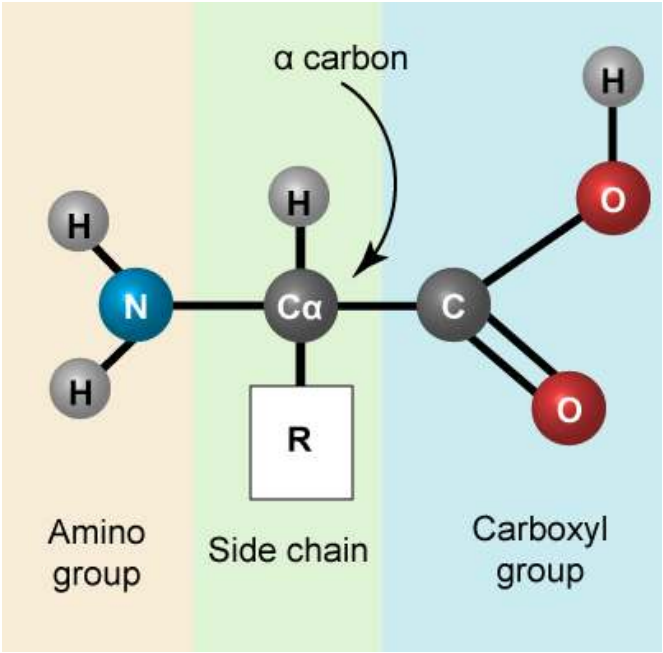
Besides their structural roles in DNA and RNA, nucleotides such as ATP and GTP also serve as mobile energy carriers for the cell. It surprises some students when they learn to appreciate that the ATP and GTP molecules we discuss in bioenergetics are the same as those involved in the formation of nucleic acids. We will cover this in more detail when we discuss DNA and RNA synthesis reactions. Nucleotides also play important roles as co-factors in many enzymatically catalyzed reactions.

Nucleic acids, RNA in particular, play a variety of roles in in cellular process besides being information storage molecules. Some roles that you should keep an eye out for as we progress through the course include: (a) **Riboprotein** complexes - RNA-Protein complexes in which the RNA serves both catalytic and structural roles. Examples of such complexes include, ribosomes (rRNA), RNases, splicesosome complexes, and telomerase. (b) Information storage and transfer roles. These roles include molecules like DNA, messenger RNA (mRNA), transfer RNA (tRNA). (c) Regulatory roles. Examples of these include various non-coding (ncRNA). Wikipedia has a comprehensive summary of the different known RNA molecules that we recommend browsing to get a better sense of the great functional diversity of these molecules.

Amino Acids



Amino acids are the monomers that make up proteins. Each amino acid has the same core structure, which comprises a central carbon atom, also known as the alpha (α) carbon, bonded to an amino group (NH2), a carboxyl group (COOH), and a hydrogen atom. Every amino acid also has another atom or group of atoms bonded to the alpha carbon known alternately as the R group, the variable group or the side-chain.



Amino acids have a central asymmetric carbon to which an amino group, a carboxyl group, a hydrogen atom, and a side chain (R group) are attached. Recall that one of the learning goals for this class is that you (a) be able to recognize, in a molecular diagram, the backbone of an amino acid and its side chain (R-group) and (b) that you be able to draw a generic amino acid. Make sure that you practice both. You should be able to recreate something like the figure above from memory (a good use of your sketchbook is to practice drawing this structure until you can do it with the crutch of a book or the internet).

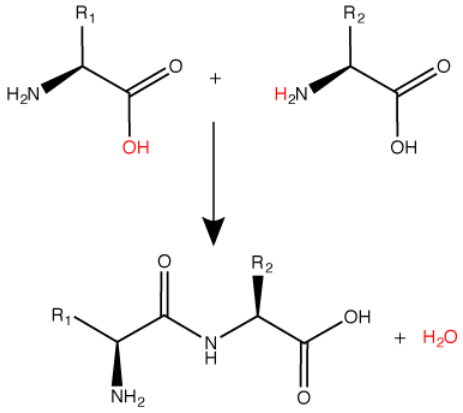
Attribution: Marc T. Facciotti (own work)

The Amino Acid Backbone

The name "amino acid" derives from the fact that all amino acids contain both an amino group and carboxyl-acid-group in their backbone. There are 20 common amino acids present in natural proteins and each of these contain the same backbone. The backbone, when ignoring the hydrogen atoms, comprises the pattern:

N-C-C

When looking at a chain of amino acids it is always helpful to first orient yourself by finding this backbone pattern starting from the N terminus (the amino end of the first amino acid) to the C terminus (the carboxylic acid end of the last amino acid).



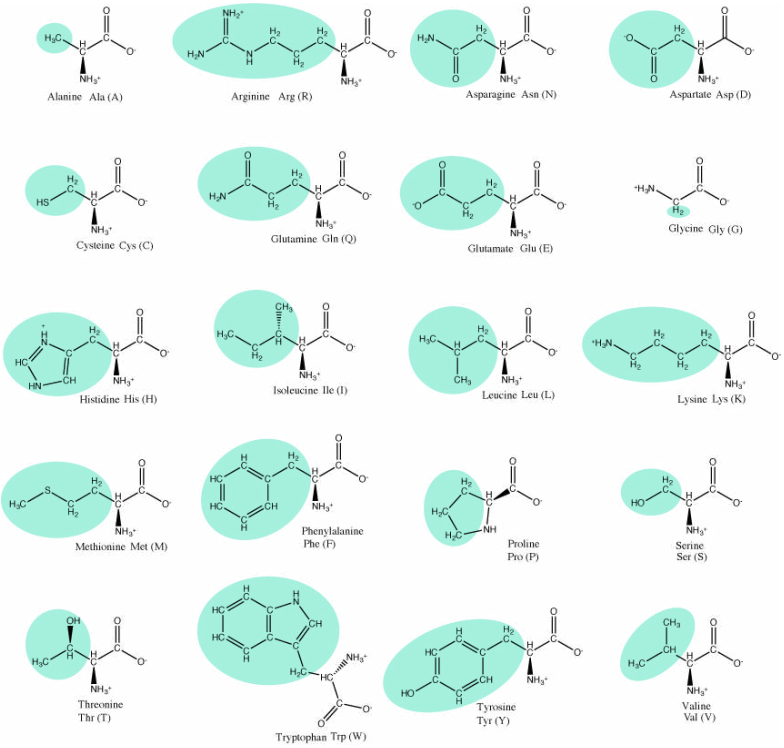
Peptide bond formation is a dehydration synthesis reaction. The carboxyl group of the first amino acid is linked to the amino group of the second incoming amino acid. In the process, a molecule of water is released and a peptide bond is formed.
Try finding the backbone in the dipeptide formed from this reaction. The pattern you are looking for is: N-C-C-N-C-C

Attribution: Marc T. Facciotti (own work)

The sequence and the number of amino acids ultimately determine the protein's shape, size, and function. Each amino acid is attached to another amino acid by a covalent bond, known as a **peptide bond**, which is formed by a dehydration synthesis (condensation) reaction. The carboxyl group of one amino acid and the amino group of the incoming amino acid combine, releasing a molecule of water and creating the peptide bond.

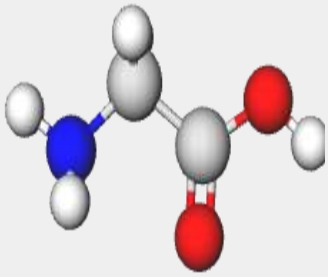
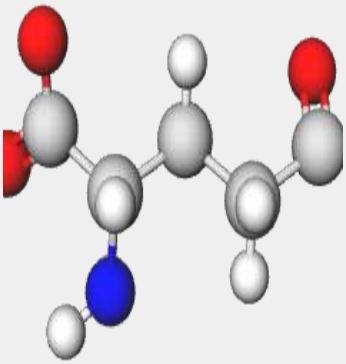
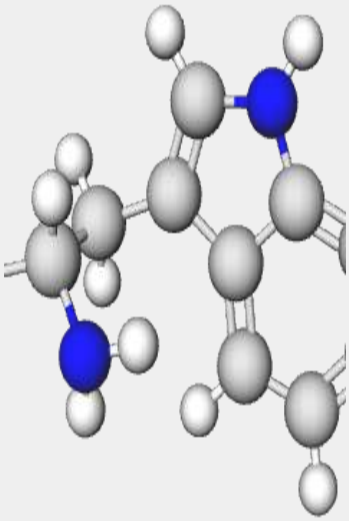
Amino Acid R group

The amino acid **R group** is a term that refers to the variable group on each amino acid. The amino acid backbone is identical on all amino acids, the R groups are different on all amino acids. For the structure of each amino acid refer to the figure below.



There are 20 common amino acids found in proteins, each with a different R group (variant group) that determines its chemical nature. R-groups are circled in teal. Charges are assigned assuming pH ~6.0. The full name, three letter abbreviation and single letter abbreviations are all shown.

Attribution: Marc T. Facciotti (own work)

Glycine	Glutamate	Tryptophan
		

Each variable group on an amino acid gives that amino acid specific chemical properties (acidic, basic, polar, or nonpolar). You should be familiar with most of the functional groups in the R groups by now. The chemical properties associated with the whole collection of individual functional groups gives each amino acid R group unique chemical potential.

For example, amino acids such as valine, methionine, and alanine are typically nonpolar or hydrophobic in nature, while amino acids such as serine and threonine are said to have polar character and possess hydrophilic side chains.

Proteins

Proteins are a class of biomolecules that perform an array of functions in biological systems. Some proteins serve as catalysts for specific biochemical reactions. Other proteins act as signaling molecules that allow cells to "talk" with one another. Proteins, like keratin in fingernails, can also act in a structural capacity. While the variety of functions for proteins is remarkably diverse, these functions are encoded by a linear assembly of amino acids, each connected to their neighbor via a peptide bond. The unique composition (types of amino acids and the number of each) and the order in which they are linked together determine the final three dimensional form that the protein will adopt and therefore, also the protein's biological "function". Many proteins can, in a cellular environment, spontaneously and often rapidly take on their final form in a process called **protein folding**. To watch a short (four minutes) introduction video on protein structure click here.

Protein structure

We can describe protein structures by four different levels of structural organization called primary, secondary, tertiary, and quaternary structures. These are briefly introduced in the sections that follow.

Primary structure

The unique sequence of amino acids in a polypeptide chain is its **primary structure** (Figure 1). The amino acids in this chain are linked to one another other via a series of peptide bonds. The chain of amino acids is often referred to as a **polypeptide** (multiple peptides).

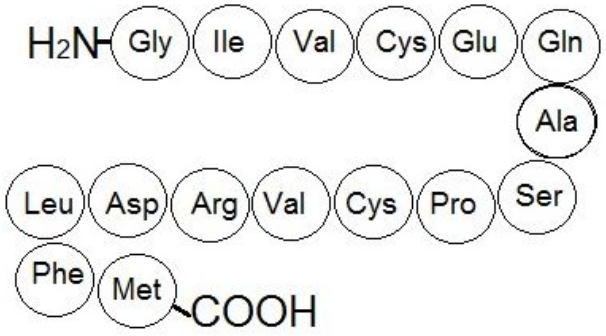


Figure 1. *The primary structure of a protein is depicted here as "beads on a string" with the N terminus and C terminus labeled. The order in which you would read this peptide chain would begin with the N terminus as Glycine, Isoleucine, etc., and end with Methionine. Source: Erin Easlon (own work)*

Due to the common backbone structure of amino acids, the resulting backbone of the protein has a repeating -N-Cα-C-N-Cα-C- pattern that can be readily identified in atomic resolution models of protein structures (Figure 2). Know that one of the learning goals for this class is for you to examine a model like the one below and to identify the backbone from the side chain atoms (e.g. create the purple trace and blue shading if there aren't any). This can be done by finding the -N-Cα-C-N-Cα-C- pattern. Another learning goal for this class is that you can create drawings that model the structure of a typical protein backbone and its side chains (aka. variable group, R group). This task can be greatly simplified if you remember to start your model by first creating the -N-Cα-C-N-Cα-C- pattern and then filling in the variable groups.

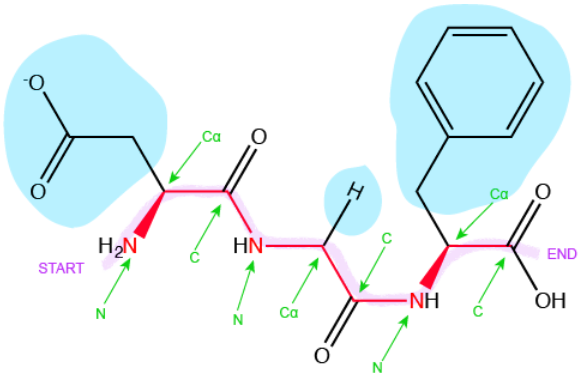


Figure 2. *A model of a short 3 amino acid long peptide. The backbone atoms are colored in red. The variable R groups are circled in light blue. A purple line traces the backbone from the N-terminus (start) to the C-terminus (end) of the protein. One can identify (in green) the repeated -N-Cα-C-N-Cα-C- ordered pattern by following the purple line from start to end and listing off the backbone atoms in the order that they are encountered. Attribution: Marc T. Facciotti (own work)*

Secondary structure

Because of the specific chemistry of the peptide bond the backbone between adjacent alpha-carbon atoms forms a highly planar structure (Figure 3). This means that all the atoms linked by the pink quadrilateral lie on the same plane. The polypeptide is therefore structurally constrained since very little rotation can happen around the peptide bond itself. Rather, rotations occur around the two bonds extending away from the alpha carbons. These structural constraints lead to two commonly observed patterns of structure that are associated with the organization of the backbone itself.

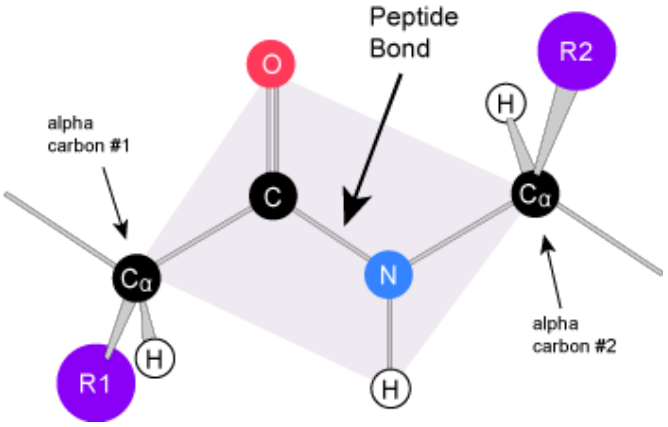


Figure 3. *The peptide bond between two amino acids is depicted. The shaded quadrilateral represents planar nature of this bond. Attribution: Marc T. Facciotti (own work)*

We call these patterns of backbone structure the **secondary structure** of the protein. The most common secondary structure patterns occurring via rotations of the bonds around each alpha-carbon, are the ***α*-helix, *β*-sheet** and **loop** structures. As the name suggests, the *α*-helix is characterized by a helical structure made by twisting the backbone. The *β*-sheet is actually the association between two or more structures called *β*-strands. If the orientation (N-terminus to C-terminus direction) of two associating *β*-strands are oriented in the same/parallel direction, the resulting *β*-sheet is called a **parallel *β*-sheet**. Meanwhile, if two associating *β*-strands are oriented in opposite/anti-parallel directions, the resulting *β*-sheet is called an **anti-parallel *β*-sheet**. The *α*-helix and *β*-sheet are both stabilized by hydrogen bonds that form between backbones atoms of amino acids near one another. The oxygen atom in the carbonyl group from one amino acid can form a hydrogen bond with a hydrogen atom bound to the nitrogen in the amino group of another amino acid. Loop structures refer to all secondary structure (e.g. backbone structure) that can not be identified as either *α*-helix or *β*-sheet.

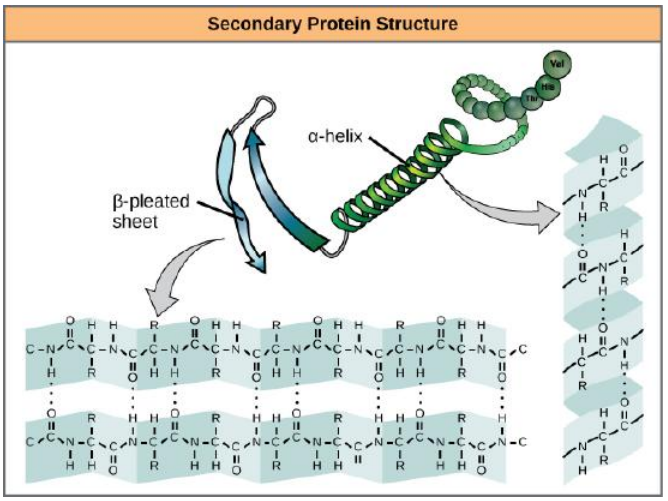


Figure 4. *The α-helix and β-sheet are secondary structures of proteins that are stabilized by hydrogen bonding between carbonyl and amino groups in the peptide backbone. Note how the hydrogen bonds in an alpha-helix occur between amino acids that are relatively close to one another (about 4 amino acids apart in the amino acid chain) while the interactions that occur in β-sheets may occur between amino acids that are much farther apart in the chain.*

Tertiary structure

The backbone and secondary structure elements will further fold into a unique and relatively stable three-dimensional structure called the **tertiary structure** of the protein. The tertiary structure is what we typically associate with the "functional" form of a protein. In Figure 6 two examples of tertiary structure are shown. In both structures, the protein is abstracted into a "cartoon" that depicts the polypeptide chain as a single continuous line or ribbon tracing the path between alpha carbons of amino acids linked to one another by peptide bonds - the ribbon traces the backbone of the protein (Figure 5).

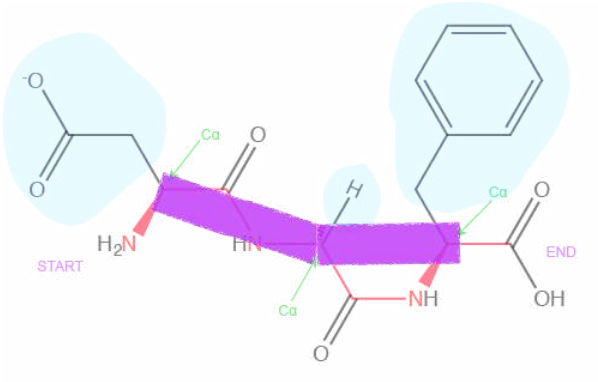


Figure 5. *How protein "cartoon" figures are drawn. Protein cartoons (like those shown in Figure 6) are perhaps the most common representation of three-dimensional protein structure. These cartoon models help us visualize the major features of a protein structure by tracing the path from one alpha-carbon to the next along the polypeptide backbone. This is depicted as a thick purple line. In a longer polypeptide this line would continue and join to the next alpha carbon until the end of the polypeptide was reached. While these models allow us to visualize the general structure of a protein, they leave out a lot of molecular-level detail.*

The ribbon created by joining alpha-carbons can be drawn as a simple continuous line or it can be enhanced by uniquely representing secondary structural elements. For instance, when an α-helix is identified, the helix is usually highlighted by accentuating/broadening the ribbon to make the helical structure stand out. When a β-strand is present, the ribbon is usually broadened and an arrow is typically added to the C-terminal end of each β-strand - the arrow helps to identify the orientation of the polypeptide and whether β-sheets are parallel or anti-parallel. The thin ribbon connecting α-helix and β-strand elements is used to represent the loops. Loops in proteins can be highly structured and play an important role in the protein's function. They should not be treated lightly or dismissed as unimportant because their name lacks a Greek letter.

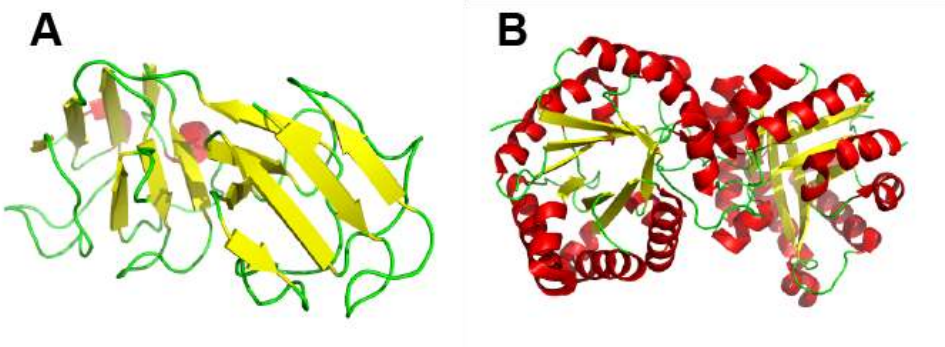
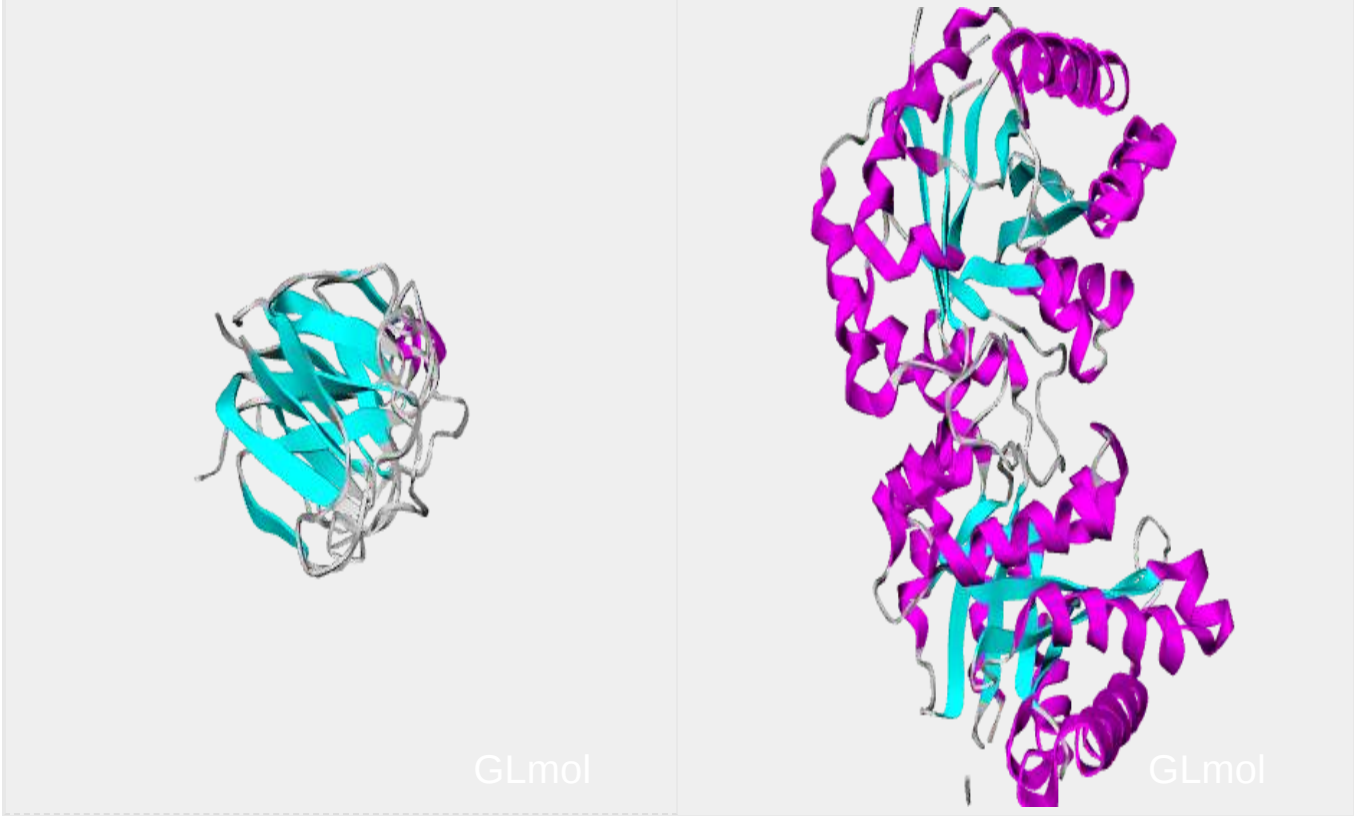


Figure 6. *Examples of tertiary structures of proteins. Secondary structure elements are colored as follows: β-sheet - yellow, α-helix - red; loop - green. In panel A the structure of protein gamma crystallin (PDBID 1a45) - a protein found in the vertebrate eye - is depicted. This protein is composed largely of β-sheet and loops. In panel B the structure of the protein triose phosphate isomerase (PDBID 1tim) - a protein found in the*

*glycolytic pathway - is composed of β-sheet, α-helix, and loops joining the secondary structural el*ements. Attribution: Marc T. Facciotti (own work)

Crystallin (PDBID 1a45)	Triose phosphate isomerase (PDBID 1tim)
	

The tertiary structure is the product of many types of chemical interactions among amino acid R groups, backbone atoms, ions in solution and water. These bonds include ionic, covalent, and hydrogen bonds and Van der Waals interactions. For example, ionic bonds may form between various ionizable side chains. It may, for instance, be energetically favorable for a negatively charged R group (e.g. an Aspartate) to interact with a positively charged R group (e.g. an Arginine). The resulting ionic interaction may then become part of the network of interactions that helps to stabilize the three dimensional fold of the protein. By contrast, R groups with like charges will likely be repelled by each other and be therefore unlikely to form a stable association thereby disfavoring a structure that would include that association. Likewise, hydrogen bonds may form between various R groups or between R groups and backbone atoms. These hydrogen bonds may also contribute to stabilizing the tertiary structure of the protein. In some cases covalent bonds may also form between amino acids. The most commonly observed covalent linkage between amino acids involves two cysteines and is termed a disulfide bond or disulfide linkage.

Finally, the association of the protein's functional groups with water also helps to drive chemical associations that help to stabilize the final protein structure. The interactions with water can include the formation of hydrogen bonds between polar functional groups on the protein and water molecules. Perhaps more importantly, however, is the drive for the protein to avoid placing too many hydrophobic functional groups in contact with water. The result of this desire to avoid interactions between water and hydrophobic functional groups means that the less polar side chains will often associate with one another away from water resulting in some energetically favorable Van der Waals interactions and the avoidance of energetic penalties associated with exposing the non-polar side chains to water. The energetic penalty is so high for "exposing" the non-polar side chains to water that burying these groups away from water is thought to be one of the primary energetic drivers of protein folding and stabilizing forces holding the protein together in its tertiary structure.

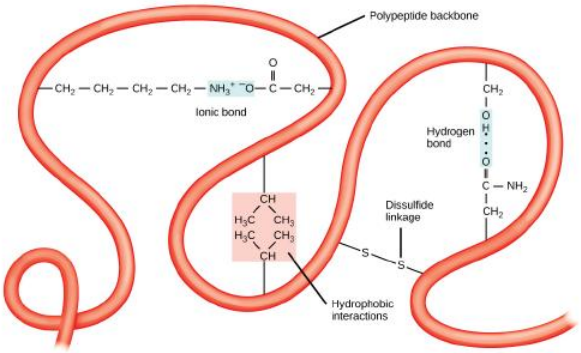


Figure 6. *The tertiary structure of proteins is determined by a variety of chemical interactions. These include hydrophobic interactions, ionic bonding, hydrogen bonding, and disulfide linkages. This image shows a flattened representation of a protein folded in tertiary structure. Without flattening, this protein would be a globular 3-D shape.*

Quaternary structure

In nature, the functional forms of some proteins are formed by the close association of several polypeptides. In such cases the individual polypeptides are also known as subunits. When the functional form of a protein requires the assembly of two or more subunits we call this level of structural organization the protein's **quaternary structure**. Yet again, combinations of ionic, hydrogen, and covalent bonds together with Van der Waals associations that occur through the "burial" of hydrophobic group at the interfaces between subunits help to stabilize the quaternary structures of proteins.

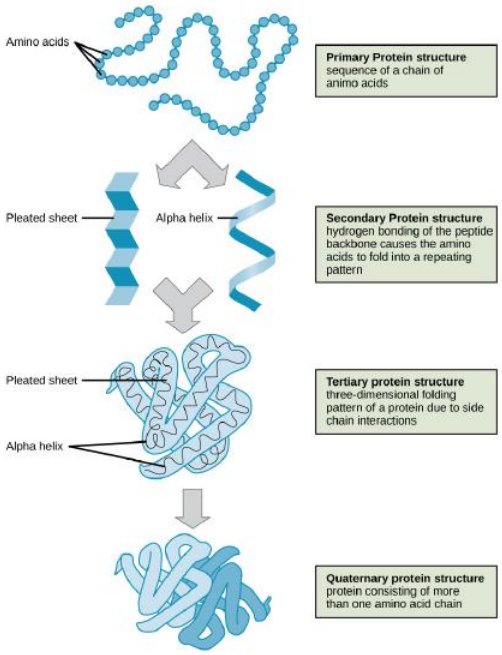


Figure 7. *The four levels of protein structure can be observed in these illustrations.*
Source: Modification of work by National Human Genome Research Institute

Possible NB Discussion Point

If the 1° structure of a protein encodes its 3° structure, how can you explain apparent contradictions that we find (1) proteins in Nature that have very similar 3° structures despite having less that 30% amino acid sequence identity (similar structures not similar sequences) and (2) while less frequent, other pairs of proteins that share higher

amino acid sequence identity but are not structurally similar (similar sequences with not similar structures)? What kinds of ideas do these observations simulate?

Denaturation

As was previously described, each protein has its own unique structure that is held together by various types of chemical interactions. If the protein is subject to changes in temperature, pH, or exposure to chemicals, that change the nature of or interfere with the associations between functional groups, the protein's secondary, tertiary and/or quaternary structures may change, even though the primary structure remains the same. This process is known as **denaturation**. While in the test tube denaturation is often reversible, in the cell the process can often be, for practical purposes, irreversible, leading to loss of function and the eventual recycling of the protein's amino acids. Resistance to environmental stresses that can lead to denaturation varies amongst the proteins found in nature. For instance, some proteins are remarkably resistant to high temperatures; for instance, bacteria that survive in hot springs have proteins that function at temperatures close to the boiling point of water. Some proteins are able to withstand the very acidic, low pH, environment of the stomach. Meanwhile some proteins are very sensitive to organic solvents while others can be found that are remarkably tolerant of these chemicals (the latter are prized for use in various industrial processes).

Finally, while many proteins can form their three dimensional structures completely on their own, in many cases proteins often receive assistance in the folding process from protein helpers known as **chaperones** (or chaperonins) that associate with their protein targets during the folding process. The chaperones are thought to act by minimizing the aggregation of polypeptides into non-functional forms - a process that can occur through the formation of non-ideal chemical associations.

Khan Academy link

Protein structure.

Reading_04 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_05

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_05

GC.34 *Interpret reaction coordinate diagrams showing either or both catalyzed and uncatalyzed reaction coordinates and identify respective activation energy barriers and relate these to the forward and reverse rates of reaction.*

GC.38 *Describe the concept of equilibrium in the context of reaction coordinate diagrams.*

GC.39 *Interpret reaction coordinate diagrams and associate changes in Gibbs enthalpy/energy and activation energy with relative rates of reactions, equilibrium conditions, and whether a reaction is endergonic or exergonic.*

GC.40 *Understand how to use the equation $\Delta G = \Delta H - T\Delta S$ and explain what each term represents.*

GC.41 *Interpret a biochemical transformation and predict whether or not the reaction is spontaneous by using a Gibbs enthalpy/energy reaction coordinate diagram.*

GC.46 *Develop and apply an "energy story" about a biological or biochemical reaction using the first and second laws of thermodynamics. Describe events in terms of energy, awareness of energy conservation, energy transfer, entropy, and then relate these to what is happening at the molecular level.*

GC.47 *Explain the first law of thermodynamics (conservation of energy).*

GC.48 *Explain the second law of thermodynamics (entropy is increasing) and how it relates to biological reactions.*

GC.49 *Describe the relationship between free energy and chemical equilibrium using the equation $\Delta G° = -RTlnKeq$, explicitly invoking appropriate "initial" and "final" states (as is done in an Energy Story).*

Energy

The concept of energy is a central to all sciences and helps us to understand why and how things change. Despite it's ubiquity, coming to a concrete understanding of what energy IS can be extremely challenging, if one can come to such a complete understanding at all. In biology - and in many other instances - it can be useful to describe energy as a property of a system. What does that mean, a property of a system? Simply put, we can imagine that things - or collection of things - that are made of atoms, molecules (aka "stuff") will have a certain amount of something we call energy. That energy may be associated with how the "stuff" is composed (e.g. chemical energy), whether the "stuff" is in motion (i.e. kinetic energy), the position of the "stuff" in a gravitational or electric field (i.e. potential energy) or other ways we might describe how energy is distributed in the system (e.g. internal energy).

We will think of energy as property that can be "stored" in a physical system and transferred or "moved" from one place or form to another (i.e. to different parts of the system; between kinetic and potential forms; etc). The idea is to reinforce the concept that energy maintains its identity when transferred. Unless we are concerned about the expansion of the universe (in biology we are not), we can rightfully assume that energy is conserved in any transformation. This idea is known as the First Law of Thermodynamics (described below) and tells us that energy does not get "made" or get "lost". We'll see below that in trying to explain "how things go" this property of energy conservation, and the associated energy accounting tracking where energy comes from and where it goes, can be very useful.

In the following sections, we will explore some basic concepts associated with common transformations in biology and chemistry: the solubility of various biomolecules, the making and breaking of chemical bonds, transferring electrons, transferring energy to and from light, and transferring energy as heat. In class, many of the discussions will happen in the context of the **Energy Story** rubric (described below), so when we consider a reaction of transformation, we will be interested in precisely defining the system in question and trying to account for all the various transfers of energy that occur within the system.

Thermodynamics

MCAT Connection

Thermodynamics is concerned with describing the changes in systems before and after a transformation. This usually involves discussing the dispersion of energy within and between the system and its surroundings.

The First Law of Thermodynamics

The first law of thermodynamics deals with the total amount of energy in a system. It is sometimes succinctly summarized by stating that the total amount of energy in the universe is constant. In other words, there always has been, and always will be, exactly the same amount of energy in the universe. While this is not strictly the case at cosmological scales, when considering events that happen on smaller scales (like what happens on earth when we talk about biological processes), the idea that the total energy of system and its environment are constant is valid.

According to the first law of thermodynamics, energy may be transferred/transformed, but it cannot be created or destroyed. Energy transfers take place around us all the time. Light bulbs transfer energy from electrical power stations into heat and photons of light. Gas stoves transfer energy stored in the bonds of chemical compounds into heat and light. (Heat, by the way, is the amount of energy transferred from one system to another because of a temperature difference.)

Plants perform one of the most biologically useful energy transfers on earth: they transfer energy in the photons of sunlight into the chemical bonds of organic molecules. In every one of these cases, energy is neither made nor destroyed, and we must try to account for *all* the energy when we examine some of these reactions.

The First Law and the Energy Story

The first law of thermodynamics is deceptively simple. Students often understand that energy cannot be created or destroyed. Yet, when telling an energy story for a process they often make the mistake of saying things like "energy is produced from the transfer of electrons from atom A to atom B." While most of us will understand the point the student is trying to make, the wrong words are being used. Energy is not made or produced; it is simply transferred. To be consistent with the first law, when telling an energy story, make sure you try to track explicitly all the places that ALL the energy in the system at the start of a process goes by the end of a process.

The Second Law of Thermodynamics

An important concept in physical systems is **entropy**. Entropy relates to how energy can be distributed or dispersed within a system. The Second Law of Thermodynamics states that entropy is always increasing in a system *and* its surroundings (that is, everything inside and outside the system combined).

This idea helps explain the directionality of natural phenomena. The notion is that the directionality comes from the tendency for energy in a system to move towards a state of maximal dispersion. The Second Law, therefore, implies that in any transformation, we should look for an overall increase in entropy (or dispersion of energy), somewhere. As dispersion of energy in a system or its surroundings increases, the ability of the energy to be directed towards work decreases.

Keep in mind: you will find many examples in which the entropy of a system decreases *locally*. However, according to the Second Law, the entropy of the entire universe can *never* decrease. This must mean that there is an equal or greater increase in entropy somewhere else in the surroundings (most likely in a closely connected system) that compensates for the local decrease.

We associate the four scenarios below with increasing entropy of the system. Try to think of specific examples for when:

a. the system gains energy;
b. a change of state occurs from solid to liquid to gas;
c. a mixing of substances occurs;
d. the number of particles increases during a reaction.

Possible NB Discussion Point

Justify or refute the following statement: *"Biological systems are an exception to the Second Law of Thermodynamics, since cells are known to arrange themselves into very highly organized structures (think: tissues, organs, etc.) rather than into a more disordered state."* Make sure to check out what your peers are saying -- do you agree or disagree with their position and/or their rationale?

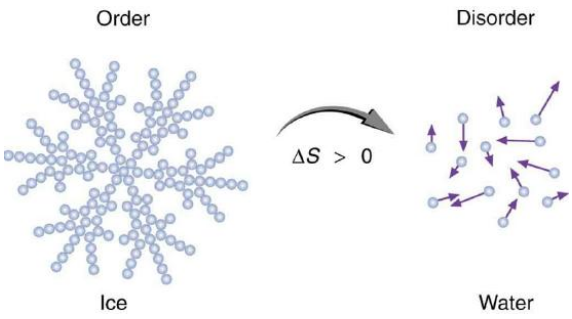


Figure 1. *An increase in disorder can happen in different ways. An ice cube melting on a hot sidewalk is one example. Here, ice is displayed as a snowflake, with organized, structured water molecules forming the snowflake. Over time, the snowflake will melt into a pool of disorganized, freely moving water molecules. It is common to describe entropy as a measure of order as a way to simplify the more concrete description relating entropy to the number of states in which energy can be dispersed in a system. While the idea of measuring order to define entropy has some flaws, it is sometimes a useful, if imperfect, proxy. (Source)*

If we consider the first and second laws together, we come to a useful conclusion. Whenever energy is transferred or redistributed within a system, entropy must increase. This increase in entropy is related to how "useful" the energy is to do work. Recall again that this energy becomes less and less available as entropy increases.

Keep in mind: you will find many examples in which the entropy of a system decreases *locally*. However, according to the Second Law, the entropy of the entire universe can *never* decrease. This must mean that there is an equal or greater increase in entropy somewhere else in the surroundings (most likely in a closely connected system) that compensates for the local decrease.

We conclude that while all the energy must be conserved, if the required change increases entropy, it means that some energy will become distributed in a way that makes it less useful for work. Most times, particularly in biology, some increase in entropy can be chalked up to a transfer of energy to heat in the environment.

Energy in chemical reactions

Chemical reactions involve a redistribution of energy within the reacting chemicals and with their environment. So, like it or not, we need to develop some models that can help us describe where energy is in a system (perhaps how it is "stored"/distributed) and how it can be moved around in a reaction. The models we develop will not be overly detailed in the sense that they would satisfy a hard-core chemist or physicist with their level of technical detail, but we expect that they should still be technically correct and not form incorrect mental models that will make it difficult to understand the "refinements" later.

In this respect, one of the key concepts to understand is that we will think about energy being transferred between parts of a system rather than referring thinking too much about it as being transformed. The distinction between "transfer" and "transform" is important because the latter gives the impression that energy is a property that exists in different forms, that it gets reshaped somehow. The common use of the term "transform" in relation to energy is understandable as different phenomena associated with the concept of energy physically "look" different to us. However, one potential problem with using the "transform" language is that it is sometimes difficult to reconcile with the idea that energy is being conserved (according to the first law of thermodynamics) if it is constantly changing form. How can the entity of energy be conserved if after a transformation it is no longer the same thing (e.g. transformed)? The second law of thermodynamics tells us that no transformation conserves all energy in a system. If energy is getting "transformed," how can it be conserved and still be consistent with the second law of thermodynamics?

So, instead, we will approach this issue by transferring and storing energy between different parts of a system and thus think about energy as a property that can get redistributed. That'll hopefully make the accounting of energy easier. Not that the energy "transfer" idea is consistent and compatible with terms like "potential energy" and "kinetic energy", as these are useful for describing how the energy is distributed between the motion of matter and the various fields (e.g. electric, gravitational, etc.) in a system.

CAUTION

If we are going to think about transferring energy from one part of a system to another, we also need to be careful about **NOT** treating energy like a substance that moves like a fluid or "thing." Rather, we need to appreciate energy simply as a property of a system that can be measured and reorganized but that is neither a "thing" nor something that is at one time in one form then later in another.

Since we will often be dealing with transformations of biomolecules, we can start by thinking about where energy can be found/stored in these systems. We'll start with a couple of ideas and add more to them later.

Let us propose that one place that energy can be stored is in the motion of matter. For brevity, we'll give the energy stored in motion a name: **kinetic energy**. Molecules in biology are in constant motion and therefore have a certain amount of kinetic energy (energy stored in motion) associated with them.

Let us also propose that there is a certain amount of energy stored in the biomolecules themselves and that the amount of energy stored in those molecules is associated with the types and numbers of atoms in the molecules and their organization (the number and types of bonds between them). The discussion of exactly where the energy is stored in the molecules is beyond the scope of this class, but we can approximate it by suggesting that a good proxy is in the bonds. Different types of bonds may be associated with storing different amounts of energy. In some contexts, this type of energy storage could be labeled **potential energy** or **chemical energy**. With this view, one of the things that happens during the making and breaking of bonds in a chemical reaction is that the energy is transferred about the system into different types of bonds. In the context of an Energy Story, one could theoretically count the amount of energy stored in the bonds and motion of the reactants and the energy stored in the bonds and energy of the products.

In some cases, you might find that when you add up the energy stored in the products and the energy stored in the reactants that these sums are not equal. If the energy in the reactants is greater than that in the products, where did this energy go? It had to get transferred to something else. Some will certainly have moved into other parts of the system, stored in the motion of other molecules (warming the environment) or perhaps in the energy associated with photons of light. One good, real-life example is the chemical reaction between wood and oxygen (reactants) and it's conversion to carbon dioxide and water (products). At the beginning, the energy in the system is largely in the molecular bonds of oxygen and the wood (reactants). There is still energy left in the carbon dioxide and water (products) but less than at the beginning. We all appreciate that some of that energy was transferred to the energy in light and heat. This reaction where energy is transferred to the environment is termed **exothermic**. By contrast, in some reactions, energy will transfer in from the environment. These reactions are **endothermic**.

The transfer of energy in or out of the reaction from the environment is NOT the only thing that determines whether a reaction will be spontaneous or not. We'll discuss that soon. For the moment, it is important to get comfortable with the idea that energy can be transferred among different components of a system during a reaction and that you should be able to envision tracking it.

Free Energy

If we want to describe transformations, it is useful to have a measure of (a) how much energy is in a system, (b) the dispersal of that energy within the system and (c) how these factors change between the start and end of a process. The concept of ***free energy***, often referred to as Gibbs energy or Gibbs enthalpy (abbreviated with the letter G), in some sense, does just that. We can define Gibbs energy in several interconvertible ways, but a useful one in the context of biology is the enthalpy (internal energy) of a system minus the entropy of the system scaled by the temperature. The difference in free energy when a process takes place is often reported in terms of the change (Δ) of enthalpy (internal energy) denoted H, minus the temperature scaled change (Δ) in entropy, denoted S. See the equation below.

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

We call reactions with ΔG < 0 **exergonic** and reactions with ΔG > 0 **endergonic**. Notice that these vocabulary terms look very similar to the terms exothermic and endothermic. While the pairs of terms sound similar, they have distinctly different meanings. The former refer to changes in ΔG while the latter refer to changes in ΔH.

We often interpret the Gibbs energy as the amount of energy available to do useful work. With a bit of hand waving, we can interpret this statement by invoking the idea presented in the section on entropy, which states the dispersion of energy (required by the Second Law) associated with a positive change in entropy somehow renders some energy that is transferred less useful to work. One can say that it reflects this in part in the $T\Delta S$ term of the Equation 1.

To provide a basis for fair comparisons of changes in Gibbs energy amongst different biological transformations or reactions, the free energy change of a reaction is measured under a set of common standard experimental conditions. The resulting standard free energy change of a chemical reaction is expressed as an amount of energy per mole of the reaction product (either in kilojoules or kilocalories, kJ/mol or kcal/mol; 1 kJ = 0.239 kcal), when measured at a standard pH, temperature, and pressure conditions. Standard pH, temperature, and pressure conditions are generally standardized at pH 7.0, 25 degrees Celsius, and 100 kilopascals (1 atm pressure), respectively. **It is important to note that cellular conditions differ from these standard conditions, and so actual ΔG inside a cell will differ considerably from those calculated under standard conditions.**

Endergonic and exergonic reactions

Any system of molecules that undergoes a physical transformation/reorganization (aka. reaction) will have an associated change in internal energy and entropy. If we examine a single isolated reaction, in which unique reactants convert into unique products the Gibbs free energy/enthalpy of the system will be dependent several factors, key among which are (a) the internal energy and entropy differences associated with the molecular rearrangements and (b) the degree to which the reaction is out-of-equilibrium.

If, for the sake of simplicity we begin by considering only the contribution of the molecular transformations in the system on ΔG, we conclude that reactions with ΔG < 0, the products of the reaction have less Gibbs energy than the reactants. Since ΔG is the difference between the enthalpy and temperature-scaled entropy changes in a reaction, a net negative ΔG can arise in through changes largely of enthalpy, entropy, or most often both. The left panel of Figure 1 below shows a common graphical representation of an **exergonic** reaction. This graph is called a reaction coordinate diagram. It plots Gibbs energy on the y-axis, and the x-axis in arbitrary units shows the progress of a reaction. With an exergonic reaction, the figure on the left shows two key things: (1) the difference between the free energy of the reactants and products is negative and (2) the progress of the reaction requires some input of free energy (shown as an energy "hill" or barrier). This graph does not tell us how the energy in the system redistributed, only that the difference between enthalpy and temperature-scaled entropy between reactants and products (e.g. ΔG) is negative. **Exergonic reactions** are said to occur spontaneously and are sometimes colloquially also referred to as "energetically favorable" reactions. Understanding which chemical reactions are spontaneous is useful for biologists who are trying to understand whether a reaction is likely to "go" or not.

It is important to note that the term "spontaneous"—in thermodynamics—implies nothing about how fast the reaction proceeds. The change in free energy only describes the difference between beginning and end states, NOT how fast that transition takes place. This is contrary to the everyday use of the term which usually carries the implicit understanding that something happens quickly. As an example, the oxidation/rusting of iron is a spontaneous reaction. However, an iron nail exposed to air does not rust instantly—it may take years.

A chemical reaction with a positive ΔG means that the products of the reaction have a higher free energy than the reactants (see the right panel of Figure 1). These chemical reactions are called **endergonic reactions**, and they are NOT spontaneous and are sometimes referred to as "energetically unfavorable" reactions. An endergonic reaction will not take place on its own without a transfer of energy into the reaction or an increase of entropy somewhere else.

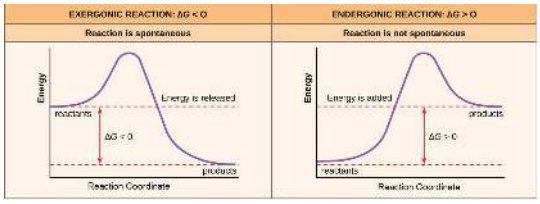


Figure 1. *Reaction coordinate diagrams of exergonic and endergonic reactions. Exergonic and endergonic reactions are characterized by changes in Gibbs energy. In the equilibrium state of an exergonic reaction, the Gibbs energy of the products is lower than that of the reactants. Meanwhile, the*

equilibrium state of an endergonic reaction in, the Gibbs energy of the products is higher than that of the reactants. Attribution: Marc T. Facciotti (own work)

The building of complex molecules, such as sugars, from simpler ones is an anabolic process and is endergonic. On the other hand, the catabolic process, such as the breaking down of sugar into simpler molecules, is exergonic. Like the example of rust above, while the breakdown of biomolecules is generally spontaneous, these reactions don't occur need to occur instantaneously (quickly). Remember, the terms endergonic and exergonic only refer to the difference in Gibbs energy between the products and reactants; they don't tell you about the rate of the reaction (how fast it happens). Rate will be discussed in later sections.

An important concept in the study of metabolism and energy is that of chemical equilibrium. Most chemical reactions are reversible. They can proceed in both directions, often transferring energy into their environment in one direction and transferring energy in from the environment in the other direction. The same is true for the chemical reactions involved in cell metabolism, such as the breaking down and building up of proteins into and from individual amino acids, respectively. Reactants within a closed system will undergo chemical reactions in both directions until a state of equilibrium is reached. **Equilibrium** in a chemical reaction is the state in which both reactants and products are present in concentrations that have no further tendency to change with time. Usually, this state results when the forward reaction proceeds at the same rate as the reverse reaction. NOTE THIS LAST STATEMENT! Equilibrium means that the relative concentrations of reactants and products are not changing in time, BUT it does NOT mean that there is no interconversion between substrates and products—it just means that when the reactant(s) are converted to product(s) that product(s) are converted to reactant(s) at an equal rate (see Figure 2). The state of equilibrium is also one of the lowest possible free energy states for the reaction and is a state of maximal entropy.

If a reaction is kept or started far out of equilibrium this state of the system also contributes to the overall Gibbs energy of a reaction. Either a rebalancing of substrate or product concentrations (by adding or removing substrate or product) or a positive change in free energy, typically by the transfer of energy from outside the reaction, can to move a reaction to an out-of-equilibrium state. Note that in a living cell, most chemical reactions do not reach a state of equilibrium—this would require that they reach their lowest free energy state, a state that is almost by definition incompatible with life. Energy is therefore required to keep biological reactions out of their equilibrium state. In this way, living organisms are in a constant, energy-requiring, uphill battle against equilibrium and entropy. This also means that the Gibbs energy of most biological reactions as they occur in the cell must also include a contribution from this out-of-equilibrium state. The Gibbs energy of these reactions, therefore, is often different from that reported under standard conditions.

You can use the simulation below on "Reversible Reactions" to develop an intuition about how ΔG relates to the equilibrium state of the reaction. Adjust the relative levels of of the gray platforms relative to one another (the difference in relative heights is analogous to ΔG as shown in reaction coordinate diagrams) and start a simulation. Watch for the reaction to reach equilibrium and take note of the number of A and B molecules at equilibrium for different configurations of the system.

Source: <https://phet.colorado.edu/en/simulat...ible-reactions>

Going a bit more into the relationship between chemical equilibrium and Gibbs energy

MCAT Connection

In a previous section, we began a description of chemical equilibrium in the context of forward and reverse rates. We presented three key ideas:

1. At equilibrium, the concentrations of reactants and products in a reversible reaction are not changing in time.
2. A reversible reaction at equilibrium is not static—reactants and products continue to interconvert at equilibrium, but the rates of the forward and reverse reactions are the same.
3. We were NOT going to fall into a common student trap of assuming that chemical equilibrium means that the concentrations of reactants and products are equal at equilibrium.

Here we extend our discussion and put the concept of equilibrium into the context of Gibbs energy, also reinforcing the Energy Story exercise of considering the "Before/Start" and "After/End" states of a reaction (including the inherent passage of time).

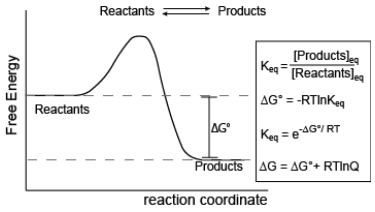


Figure 1. *Reaction coordinate diagram for a generic exergonic reversible reaction. Equations relating Gibbs energy and the equilibrium constant: R = 8.314 J mol^{-1} K^{-1} or 0.008314 kJ mol^{-1} K^{-1}; T is temperature in Kelvin. Attribution: Marc T. Facciotti (original work)*

The figure above shows a commonly cited relationship between ΔG° and Keq:

$$\Delta G^{o} = -RT \ln K_{eq}. \tag{2}$$

Here, G° indicates the Gibbs energy under standard conditions (e.g., 1 atmosphere of pressure, 298 K). This equation describes the change in Gibbs energy for reactants converting to products in a reaction that is at equilibrium. The value of ΔG° can therefore be thought of as being intrinsic to the reactants and products themselves. ΔG° is like a potential energy difference between reactants and products. With this concept as a basis, one can also consider a reaction where the "starting" state is somewhere out of equilibrium. In this case, there may be an additional “potential” associated with the out-of-equilibrium starting state. This “added” component contributes to the ΔG of a reaction and can be effectively added to the expression for Gibbs energy as follows:

$$\Delta G = \Delta G^{o} + RT \ln Q, \tag{3}$$

where Q is called the reaction quotient. From the standpoint of General Biology, we will use a simple (a bit incomplete but functional) definition for

$$Q = \frac{[Products]_{st}}{[Reactants]_{st}} \tag{4}$$

at a defined non-equilibrium condition, st. One can extend this idea and calculate the Gibbs energy difference between two non-equilibrium states, provided they are properly defined and thus compute Gibbs energy changes between specifically defined out-of-equilibrium states. This last point is often relevant in reactions found in biological systems as these reactions are often found in multi-step pathways that effectively keep individual reactions in an out-of-equilibrium state.

This takes us to a point of confusion for some. In many biology books, the discussion of equilibrium includes not only the discussion of forward and reverse reaction rates, but also a statement that ΔG = 0 at equilibrium. This can be confusing because these very discussions often follow discussions of non-zero ΔG° values in the context of equilibrium (ΔG° = -RTlnKeq). The nuance to point out is that ΔG° is referring to the Gibbs energy potential inherent in the chemical transformation between reactants and products alone. This is different from considering the progress of the reaction from an out-of-equilibrium state that is described by

$$\Delta G = \Delta G^{o} + RT \ln Q. \tag{5}$$

This expression can be expanded as follows:

$$\Delta G = -RT \ln K_{eq} + RT \ln Q \tag{6}$$

to bring the nuance into clearer focus. In this case note that as Q approaches Keq that the reaction ΔG becomes closer to zero, ultimately reaching zero when Q = Keq. This means that the Gibbs energy of the reaction (ΔG) reaches zero at equilibrium not that the potential difference between substrates and products (ΔG°) reaches zero.

The energy story

Overview of the energy story

Whether we know it, we tell stories that involve matter and energy every day. We just rarely use terminology associated with scientific discussions of matter and energy.

Example 1

The setup: a simple statement with implicit details

You tell your roommate a story about how you got to campus by saying, "I biked to campus today." In this simple statement are several assumptions that are instructive to unpack, even if they may not seem very critical to include explicitly in a casual conversation between friends over transportation choices.

An outsider's reinterpretation of the process

To illustrate this, imagine an external observer, such as an alien being watching the comings and goings of humans on Earth. Without the benefit of knowing much of the implied meanings and reasonable assumptions that are buried in our language, the alien's description of the morning cycling trip would differ from your own. What you described efficiently as "biking to campus" might be more specifically described by the alien as a change in location of a human body and its bicycle from one location (the apartment, termed position A) to a different location (the university, termed position B). The alien might be even more abstract and describe the bike trip as the movement of matter (the human body and its bike) between an initial state (at location A) to a final state (at location B). From the alien's standpoint, what you'd call "biking" might be more specifically described as the use of a two-wheeled tool that couples the transfer of energy from the electric fields in chemical compounds to the acceleration of the two-wheeled, tool-person combo that heats its environment. Finally, buried within the simple statement describing how we got to work is also the tacit understanding that the mass of the body and bike were conserved in the process (with some important caveats we’ll look at in future lectures) and that some energy was transferred around the system and environment to enable the movement of the body from position A to position B.

Details are important. What if you owned a fully electric bike, and the person you were talking with didn’t know that? What important details might this change about the “every day” story you told that the more detailed description would have cleared up? How would the alien’s story have changed? In what

scenarios might these changes be relevant?

As this simple story illustrates, irrespective of many factors, the act of creating a full description of a process includes some accounting of what happened to the matter, what happened to the energy, and almost always some description of a mechanism that describes how changes in matter and energy of a system were brought about.

We will make use of something we like to call the "Energy Story." You may be asked to tell an "energy story" in class, to practice telling energy stories on your lecture study guides, and to use the concept on your exams. In this section, we focus primarily on introducing the concept of an energy story and explaining how to tell one. It is worth noting that the term "energy story" is used almost exclusively and has a specific meaning in this class. This precise term will likely not appear in other courses (at least in the short term), or if it appears, is not likely to be used in the same manner.

You can think of “The Energy Story” as a systematic approach to creating a statement or story describing a biological process or event. Your BIS2A instructors have given this approach a short name (energy story) so that we can all associate it with the common exercise. That way, when the instructor asks the class to tell or construct an energy story, everyone knows what is meant.

Definition 1: energy story

An energy story is a narrative describing a process or event. The critical elements of this narrative are as follows:

1. Identify at least two states (e.g., start and end) in the process.
2. Identify and list the matter in the system and its state at the start and end of the process.
3. Describe the transformation of the matter that occurs during the process.
4. Account for the “location” of energy in the system at the start and end of the process.
5. Describe the transfer of energy that happens during the process.
6. Identify and describe mechanism(s) responsible for mediating the transformation of matter and transfer of energy.

A complete energy story will include a description of the initial reactants and their energetic states as well as a description of the final products and their energetic states after the process or reaction is completed.

Possible NB Discussion Point

We argue that the energy story can be used to communicate all the useful details that are required to describe nearly any process. Can you think of a process that cannot be adequately described by an energy story? If so, describe such a process.

Example 2: energy story example

Let us suppose that we are talking about the process of driving a car from "Point A" to "Point B" (see Figure 1).

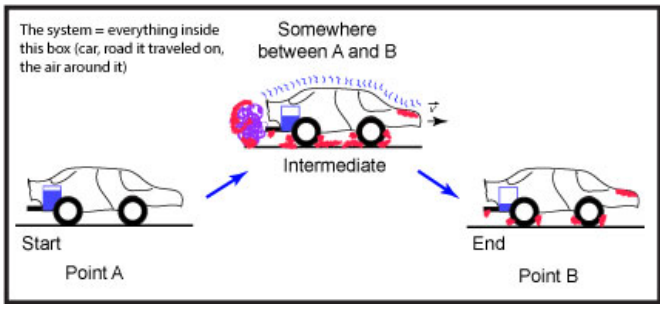


Figure 1: *This is a schematic of a car moving from a starting position, "Point A," to an end point, "Point B." The blue rectangle depicted in the back of the car represents the level of the gasoline; the purple, squiggly line near the exhaust pipe represents the exhaust; squiggly blue lines on top of the car represent sound vibrations; and the red shading represents areas that are hotter than at the start. Source: created by Marc T. Facciotti (own work)*

Let's step through the Energy Story rubric:

1. Identify at least two states (e.g., start and end) in the process.

In this example, we can easily identify two states. The first state is the nonmoving car at "Point A," the start of the trip. The second state, after the process is done, is the nonmoving car at "Point B."

2. Identify and list the matter in the system and its state at the start and end of the process.

In this case, we first note that the "system" includes everything in the figure—the car, the road, the air around the car, etc.

It is important to understand the we are going to apply the physical law of conservation of matter. That is, in any of the processes that we will discuss, matter is neither created or destroyed. It might change form, but one should be able to account for everything at the end of a process that was there at the beginning.

At the beginning of the process, the matter in the system consists of the following:

1. The car and all the stuff in it
2. The fuel in the car (a special thing in the car)
3. The air (including oxygen) around the car.
4. The road
5. The driver

At the end of the process, the matter in the system is distributed as follows:

1. The car and all the stuff in it is in a new place (let's assume, aside from the fuel and position, that nothing else changed).
2. There is less fuel in the car, and it too is in a new place.
3. The air has changed; it now has less molecular oxygen, more carbon dioxide, and more water vapor.
4. The road did not change (let's assume it didn't change—other than a few pebbles that moved around).
5. The driver did not change (let's assume she didn't change—though we'll see by the end of the term that she did, at least a little). However, the driver is now in a different place.

3. Describe the transformation of the matter that occurs during the process.

What happened to the matter in this process? Thanks to a lot of simplifying assumptions, we see that two big things happened. First, the car and its driver changed positions—they went from "Point A" to "Point B." Second, we note that some of the molecules in the fuel, which used to be in the car as a liquid, have changed forms and are now mostly in the form of carbon dioxide and water vapor (purple blob coming out of the tailpipe). Some of the oxygen molecules that used to be in the air are now also in a new place as part of the carbon dioxide and water that left the car.

4. Account for the “location” of energy in the system at the start and end of the process.

It is again important to understand that we are going to invoke the physical law of conservation of energy. That is, we stipulate that the energy in the system cannot be created or destroyed, and therefore, the energy that is in the system at the start of the process must still be there at the end of the process. It may have been redistributed, but you should be able to account for all the energy.

At the beginning of the process, the energy in the system is distributed as follows:

1. The energy is tied up in the associations between atoms that make up the matter of the car.
2. The energy is tied up in the associations between atoms that make up the fuel.
3. The energy is tied up in the associations between atoms that make up the air.
4. The energy is tied up in the associations between atoms that make up the road.
5. The energy is tied up in the associations between atoms that make up the driver.
6. For all the things above, we can also say that there is energy in the molecular motions of the atoms that make up the stuff.

At the end of the process, the energy in the system is distributed as follows:

1. The energy is tied up in the associations between atoms that make up the matter of the car.
2. The energy is tied up in the associations between atoms that make up the fuel.
3. The energy is tied up in the associations between atoms that make up the air.
4. The energy is tied up in the associations between atoms that make up the road.
5. The energy is tied up in the associations between atoms that make up the driver.
6. For all the things above, we can also say that there is energy in the molecular motions of the atoms that make up the stuff.

This is interesting in some sense, because the lists are about the same. We know that the amount of energy stored in the car has decreased, because there is less fuel. Something must have happened.

5. Describe the transfer of energy that happens during the process.

In this particular example, it is the transfer of energy among the components of the system that is most interesting. As we mentioned, there is less energy stored in the gas tank of the car at the end of the trip, because there is now less fuel. We also know intuitively (from real-life experience) that the transfer of energy from the fuel to something else was instrumental in moving the car from "Point A" to "Point B." So, where did this energy go? Remember, it didn't just disappear. It must have moved somewhere else in the system.

Well, we know that there is more carbon dioxide and water vapor in the system after the process. There is energy in the associations between those atoms (atoms that used to be in the fuel and air). So some of the energy that was in the fuel is now in the exhaust. Let's also draw from real-life experience again, and state that we know that parts of our car have gotten hot by the end of the trip (e.g., the engine, transmission, wheels/tires, exhaust, etc.). For the moment, we'll just use our intuition, and say that we understand that making something hot involves some transfer of energy. So we can reasonably postulate that some of the energy in the fuel went (directly or indirectly) into heating the car, parts of the road, and the exhaust—and thus the environment

around the car. An amount of energy also went into accelerating the car from zero velocity to whatever speed it traveled at, but most of that energy eventually became heat when the car came to a stop.

This is a bit of a hand-wavy explanation, and we'll learn how to do a better job throughout the quarter. The main point is that we should be able to add all the energy of the system at the beginning of the process (in all the places it is found) and at the end of the process (in all the places it is found), and those two values should be the same.

6. Identify and describe mechanism(s) responsible for mediating the transformation of matter and transfer of energy.

Finally, it is useful to try understanding how those transformations of matter and transfers of energy might have been facilitated. For the sake of brevity, we might just say that there was a complicated mechanical device (the engine) that helped facilitate the conversion of matter and transfer of energy about the system and coupled this to the change in position of the car. Someone interested in engines would, of course, give a more detailed explanation.

In this example, we made a bunch of simplifying assumptions to highlight the process and to focus on the transformation of the fuel. But that's fine. The more you understand about the processes, the finer details you can add. Note that you can use the Energy Story rubric for describing your understanding (or looking for holes in your understanding) of nearly any process (certainly in biology). In BIS2A, we'll use the Energy Story to get an understanding of processes as varied as biochemical reactions, DNA replication, the function of molecular motors, etc.

Important:

First: We will be working on many examples of the energy story throughout the course—do not feel that you need to have mastery over this topic today.

Second: While it is tempting to think all this is superfluous or not germane to your study of biology in BIS2A, let this serve as a reminder that your instructors (those creating the course midterm and final assessments) view it as core material. We will revisit this topic often throughout the course but need you to get familiar with some of the basic concepts now.

This is important material and an important skill to develop—do not put off studying it because it doesn't "look" like "biology" to you today. The academic term moves VERY quickly, and it will be difficult to catch up later if you don't give this some thought now.

Reading_05 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_06

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_06

GC.36 *Describe mechanisms used by enzymes to lower the activation energy and increase rates of reaction.*

GC.37 Interpret the influence of a regulatory molecule on an enzyme and propose a related functional hypothesis if given kinetic data.

MS.18 Hypothesize how binding of small molecules to one or more binding pockets can lead to changes in protein function (i.e. competitive inhibition and/or allostery).

MS.19 Describe in general terms the functional link between cofactors, coenzymes, and their associated proteins.

MS.17 Draw a rough sketch of an enzyme including its active site and other sites in the enzyme that might impact its function, such as an inhibitor binding site.

GC.42 Predict whether two reactions can be theoretically productively coupled by interpreting tables of standard Gibbs enthalpy (energy).

GC.50 Describe why an enzyme is critical for coupling an exergonic reaction to an endergonic reaction by applying the concepts from the chemistry and thermodynamics lectures.

MS. 14 Form hypotheses about how different types of changes at a molecular interface might influence binding (i.e. proteins-proteins, protein, DNA, protein-lipid, small molecule-lipid, small molecule-protein, etc.).This would include recognizing or inferring different types of chemical bonds from an analysis or knowledge of functional group interactions

Catalysts

For a chemical reaction to happen, the reactants must first find one another in space. Chemicals in solution don't "plan" these collisions; they happen at random. In fact, most times, it's even more complicated. Not only do the reactants need to run into one another, but they also need to come into contact in a specific orientation. If reactants are very dilute, the rate of the reaction will be slow—collisions will happen infrequently. Increasing the concentrations will increase the rate of productive collisions. Another way to change the rate of reaction is to increase the rate of collisions by increasing the rate at which the reactants explore the reaction space—by increasing the velocity of the molecules or their kinetic energy. This can be accomplished by transferring heat into the system or raising the temperature. Those two strategies are often suitable for increasing the rates of chemical reactions that happen in a tube. Use **this interactive PhET simulation** to develop a better intuition on the topic of reaction rates. In the cell, however, transferring heat may not be practical, as it may damage cellular components and lead to death. Cells sometimes use mechanisms to increase concentrations of reactants (we'll see some examples below), but this is rarely sufficient to drive reaction rates in a biologically relevant regime. This is where catalysts come in.

A **catalyst** is something that helps increase the rate of a chemical reaction undergoing no change itself. You can think of a catalyst as a chemical change agent.

The most important catalysts in biology are called **enzymes**. An **enzyme** is a protein catalyst. Other cellular catalysts include molecules called ribozymes. A **ribozyme** is a catalyst composed of a ribonucleic acid (RNA). Both will be discussed in more detail later in the course. Like all catalysts, enzymes work by lowering the level of energy that needs to be transferred into a chemical reaction to make it happen. A chemical reaction's **activation energy** is the "threshold" level of energy needed to start the reaction.

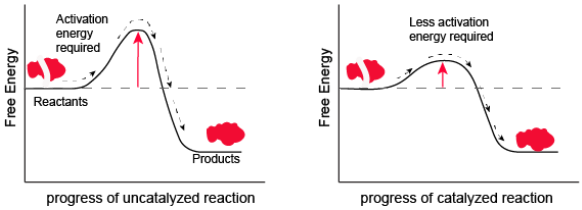


Figure 1. *Enzymes and other catalysts decrease the activation energy required to start a given chemical reaction. Without an enzyme (left), the energy input needed for a reaction to begin is high. With the help of an enzyme (right), la reaction needs less energy to begin. Attribution: Marc T. Facciotti (original work)*

In the figure above, What do you think the units are on the x-axis? Time would be one guess. However, if you compare the figures, it appears that the products are formed at the same time whether the activation energy barrier is high or low. Wasn't the point of this

figure to illustrate that reactions with high activation energy barriers are slower than those with low activation energy barriers? What's going on?

Enzymes Section Overview



Enzymes are biological catalysts that speed up chemical reactions by lowering the activation energy. Enzymes are proteins comprising one or more polypeptide chains. Enzymes have an active site that provides a unique chemical environment made up of certain amino acid R groups (residues). This unique environment is well suited to convert particular chemical reactants for that enzyme, called substrates, into unstable intermediates, called transition states. Enzymes and substrates are thought to bind with an induced fit which means that enzymes and substrates undergo slight conformational adjustments upon substrate contact, leading to binding. Enzymes bind to substrates and catalyze reactions in four different ways: bringing substrates together in an optimal orientation, compromising the bond structures of substrates so that bonds can be more easily broken, providing optimal environmental conditions for a reaction to occur, or taking part directly in their chemical reaction by forming transient covalent bonds with the substrates.

Enzyme action must be regulated so that, in a given cell at a given time, the desired reactions are being catalyzed and the undesired reactions are not. Enzymes are regulated by cellular conditions, such as temperature and pH. They are also regulated through their location within a cell, sometimes being compartmentalized so that they can only catalyze reactions under certain circumstances. Inhibition and activation of enzymes via other molecules are other important ways that enzymes are regulated. Inhibitors can act competitively, noncompetitively, or allosterically; noncompetitive inhibitors are usually allosteric. Activators can also enhance the function of enzymes allosterically. The most common method by which cells regulate the enzymes in metabolic pathways is through feedback inhibition. During feedback inhibition, the products of a metabolic pathway serve as inhibitors (usually allosteric) of one or more of the enzymes (usually the first committed enzyme of the pathway) involved in the pathway that produces them.

Enzymes

A substance that helps a chemical reaction to occur is a *catalyst*, and the special molecules that catalyze biochemical reactions are called *enzymes*. Almost all enzymes are proteins, made up of chains of amino acids, and they perform the critical task of lowering the activation energies of chemical reactions inside the cell. Enzymes do this by binding to the reactant molecules and holding them in such a way as to make the chemical bond-breaking and bond-forming processes take place more readily. It is important to remember that enzymes don't change the ΔG of a reaction. They don't change whether a reaction is exergonic (spontaneous) or endergonic (not spontaneous). This is because they don't change the free energy of the reactants or products. They only reduce the activation energy required to reach the transition state.

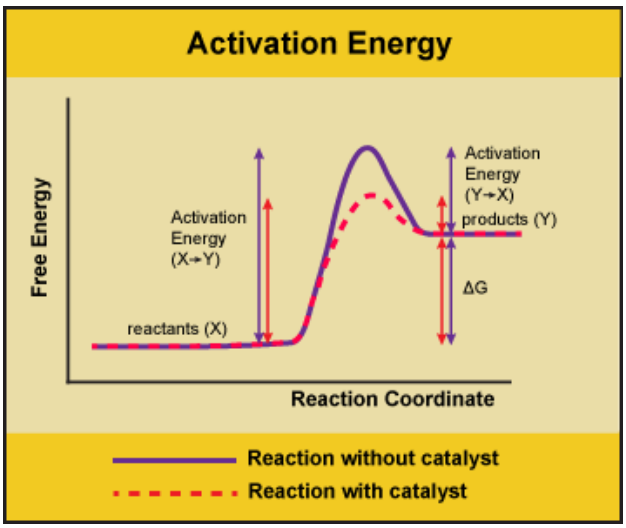


Figure 1. *Enzymes lower the activation energy of the reaction but do not change the free energy of the reaction. Here, the solid line in the graph shows the energy required for reactants to turn into products without a catalyst. The dotted line shows the energy required using a catalyst. This figure should say Gibbs Free Energy on the Y-axis and instead of noting ΔH should have ΔG. Attribution: Marc T. Facciotti (own work)*

Enzyme active site and substrate specificity

The chemical reactants to which an enzyme binds are the enzyme's *substrates*. There may be one or more substrates, depending on the particular chemical reaction. In some reactions, a single-reactant substrate is broken down into multiple products. In others, two substrates may come together to create one larger molecule. Two reactants might also enter a reaction, both become changed, and leave the reaction as two products. The location within the enzyme where the substrate binds is called the enzyme's *active site*. The active site is where the "action" happens, so to speak. Since enzymes are proteins, there is a unique combination of amino acid residues (also called side chains, or R groups) within the active site. Each amino acid side chain is characterized by different properties. Amino acids can be classified as large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of amino acids (their positions, sequences, structures, and properties) creates a very specific chemical environment within the active site. This specific environment is suited to bind, albeit briefly, to a specific chemical substrate (or substrates). Due to this jigsaw-puzzle-like match between an enzyme and its substrates (which adapts to find the best fit between the transition state and the active site), enzymes are known for their specificity. The "best fit" between an enzyme and its substrates results from their respective shapes and the chemical complementarity of the functional groups on each binding partner.

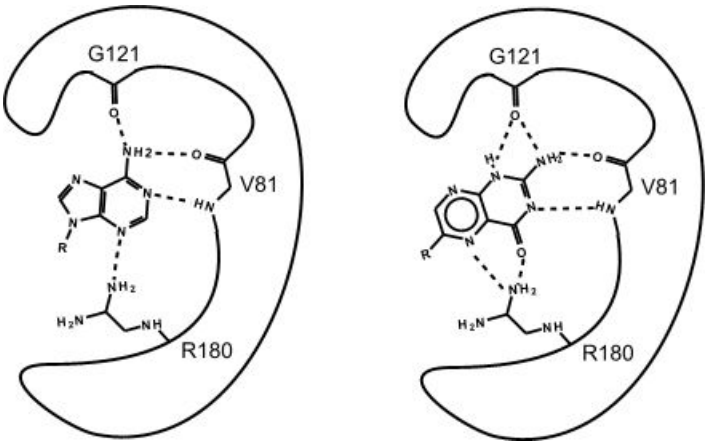


Figure 2. *This is an enzyme with two different substrates bound in the active site. The enzymes are represented as blobs, except for the active site, which shows the three R-groups of each of the three amino acids in the active site. These R groups are interacting with the substrates through hydrogen bonding (represented as dashed lines).*

At this point in the class, you should be familiar with all the types of bonds as well as the chemical characteristics of all the functional groups. For example, the R group of R180 in the enzyme depicted above is the amino acid Arginine (abbreviated as R) and has an R group that comprises several amino functional groups. Amino functional groups contain a nitrogen (N) and hydrogen (H) atoms. Nitrogen is more electronegative than hydrogen, so the covalent bond between N-H is a polar covalent bond. The hydrogen atoms in this bond will have a positive dipole moment, and the nitrogen atom will have a negative dipole moment. This allows amino groups to form hydrogen bonds with other polar compounds. Likewise, the backbone carbonyl oxygens of valine (V) 81 and glycine (G) 121 the backbone amino hydrogen of V81 are depicted engaged in hydrogen bonds with the small molecule substrate.

Exercise

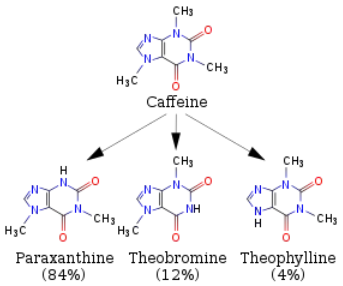
Look to see which atoms in Figure 2 (above) are involved in the hydrogen bonds between the amino acid R groups and the substrate. You will need to be able to identify these on your own; hydrogen bonds may not be drawn in for you on the test.

If you changed the pH of the solution that this enzyme is located in, would the enzyme still be able to form hydrogen bonds with the substrate?

Which substrate (the left or right one) do you think is more stable in the active site? Why? How?

Possible NB Discussion Point: How your body breaks down caffeine

When you drink coffee or other caffeinated beverages like some sodas, you are consuming a molecule called caffeine! Caffeine over time gets metabolized (broken down) via a set of very related "CYP (Cytochrome P450)" enzymes to yield the three products shown in the figure below (Source: Wikipedia). To simplify slightly, you can interpret one arrow to represent a reaction catalyzed by one of the related CYP enzymes to yield paraxanthine, theobromine, or theophylline ... all of which themselves get recognized by other enzymes that will further break them down, so on and so forth. Take a moment to examine the four structures below; the general structure might look vaguely familiar to you (think about what you've learned so far about the structure of nucleotides). Compare the reactant and the three products -- what are the noteworthy functional groups and properties of these molecules? How are the products similar to each other and caffeine? How are they different from one another? What do you predict to be the key features of the active sites for the enzymes that break down these four molecules? If you were to design an enzyme that would break down caffeine AND theophylline only, how would you design your active site?



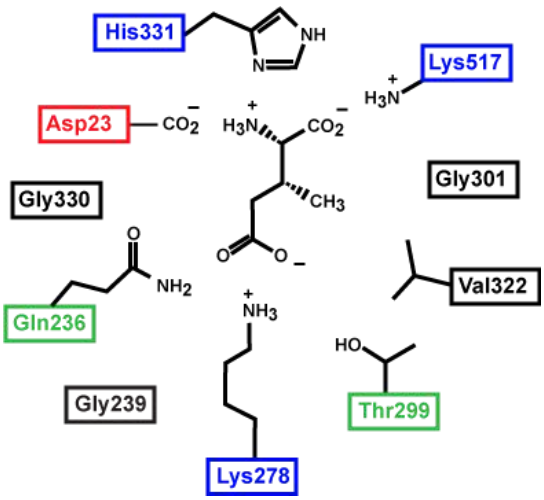


Figure 3. *This is a depiction of an enzyme active site. Only the amino acids in the active site are drawn. The substrate is sitting directly in the center.*
Source: created by Marc T. Facciotti (original work)

Exercise

First, identify the type of macromolecule in Figure 3. Second, draw in and label the appropriate interactions between the R groups and the substrate. Explain how these interactions might change if the pH of the solution changed.

Real-Life Connection

A new way of visualizing diseases in the body.

Structural instability of enzymes

The fact that active sites are so well suited to provide specific environmental conditions also means that they are subject to influences by the local environment. It is true that increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise. However, increasing or decreasing the temperature outside of an optimal range can affect chemical bonds within the active site in such a way that they are less well suited to bind substrates. High temperatures will eventually cause enzymes, like other biological molecules, to *denature*, a process that changes the natural properties of a substance. Likewise, the pH of the local environment can also affect enzyme function. Active site amino acid residues have their own acidic or basic properties that are optimal for catalysis. These residues are sensitive to changes in pH that can impair the way substrate molecules bind. Enzymes are suited to function best within a certain pH range, and, as with temperature, extreme pH values (acidic or basic) of the environment can cause enzymes to denature.

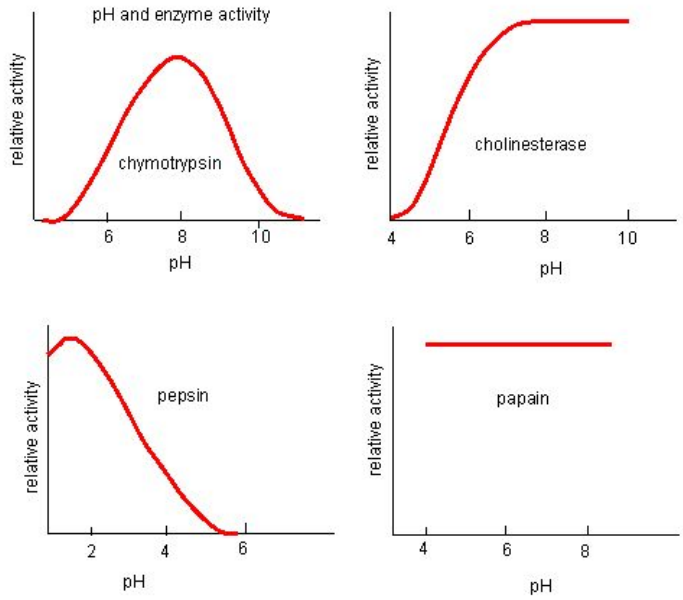


Figure 4. *Enzymes have an optimal pH. The pH at which the enzyme is most active will be the pH where the active site R groups are protonated/deprotonated such that the substrate can enter the active site and the initial step in the reaction can begin. Some enzymes require a very low pH (acidic) to be completely active. In the human body, these enzymes are most likely located in the lower stomach, or located in lysosomes (a cellular organelle used to digest large compounds inside the cell).*
Source: http://biowiki.ucdavis.edu/Biochemis..._pH_Inhibition

The process where enzymes denature usually starts with the unwinding of the tertiary structure through destabilization of the bonds holding the tertiary structure together. Hydrogen bonds, ionic bonds, and covalent bonds (disulfide bridges and peptide bonds) can all be disrupted by large changes in temperate and pH. Using the chart of enzyme activity and temperature below, make an energy story for the red enzyme. Explain what might be happening from 37 °C to 95 °C.

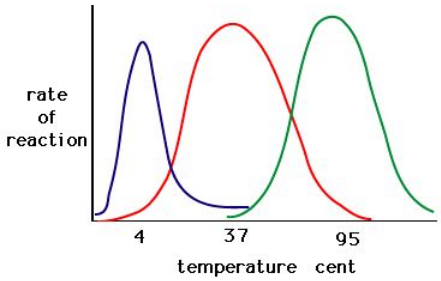


Figure 5. *Enzymes have an optimal temperature. The temperature at which the enzyme is most active will usually be the temperature where the structure of the enzyme is stable or uncompromised. Some enzymes require a specific temperature to remain active and not denature. Source: http://academic.brooklyn.cuny.edu/bi...ge/enz_act.htm*

Induced fit and enzyme function

For many years, scientists thought that enzyme-substrate binding took place in a simple "lock-and-key" fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a more refined view called *induced fit*. The induced-fit model expands upon the lock-and-key model by describing a more dynamic interaction between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme's structure that confirms a more productive binding arrangement between the enzyme and the transition state of the substrate. This energetically favorable binding maximizes the enzyme's ability to catalyze its reaction.

When an enzyme binds its substrate, an enzyme-substrate complex is formed. This complex lowers the activation energy of the reaction and promotes its rapid progression in one of many ways. On a basic level, enzymes promote chemical reactions that involve more than one substrate by bringing the substrates together in an optimal orientation. The appropriate region (atoms and bonds) of one molecule is juxtaposed to the appropriate region of the other molecule with which it must react. Another way in which enzymes promote the reaction of their substrates is by creating an energetically favorable environment within the active site for the reaction to occur. Certain chemical reactions might proceed best in a slightly acidic or nonpolar environment. The chemical properties that emerge from the particular arrangement of amino acid residues within an active site create the energetically favorable environment for an enzyme's specific substrates to react.

The activation energy required for many reactions includes the energy involved in slightly contorting chemical bonds so that they can more easily react. Enzymatic action can aid this process. The enzyme-substrate complex can lower the activation energy by contorting substrate molecules in such a way as to facilitate bond breaking. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. The amino acid residues can provide certain ions or chemical groups that actually form covalent bonds with substrate molecules as a necessary step of the reaction process. In these cases, it is important to remember that the enzyme will always return to its original state at the completion of the reaction. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme is done catalyzing a reaction, it releases its product(s).

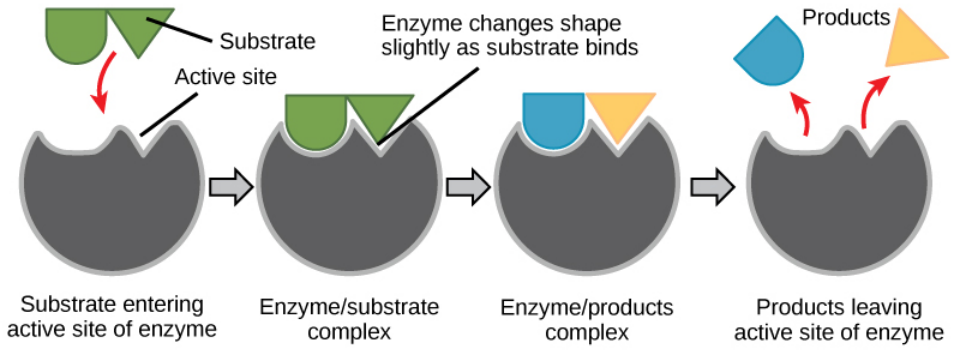


Figure 6. *According to the induced-fit model, both enzyme and substrate undergo dynamic conformational changes upon binding. The enzyme contorts the substrate into its transition state, thereby increasing the rate of the reaction.*

Creating an energy story for the reaction above

Using Figure 6, answer the questions posed in the energy story.

1. What are the reactants (define the start)? What are the products (define the end)?
2. How did the "stuff" in this system change between the beginning and end of the reaction?
3. How did the energy in this system change between the beginning and end of the reaction?
4. By what mechanism did this system transform? Imagine something. How might the enzyme have played a role?

Enzyme regulation

Why regulate enzymes?

Cellular needs and conditions vary from cell to cell and change within individual cells over time. The required enzymes and energetic demands of stomach cells are different from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive cell works much harder to process and break down nutrients during the time closely following a meal compared with many hours after a meal. As these cellular demands and conditions vary, so do the needed amounts and functionality of different enzymes.

Regulation of enzymes by molecules

Enzymes can be regulated in ways that either promote or reduce their activity. There are many kinds of molecules that inhibit or promote enzyme function, and various mechanisms exist for doing so. In some cases of enzyme inhibition, for example, an inhibitor molecule is similar enough to a substrate that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through *competitive inhibition*, because an inhibitor molecule competes with the substrate for active site binding. On the other hand, in noncompetitive inhibition, an inhibitor molecule binds to the enzyme in a location other than an active site and still manages to block substrate binding to the active site.

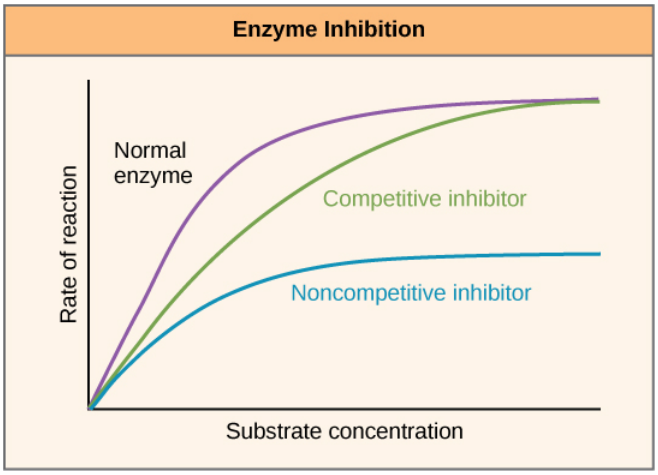


Figure 7. *Competitive and noncompetitive inhibition affect the rate of reaction differently. Competitive inhibitors affect the initial rate but do not affect the maximal rate, whereas noncompetitive inhibitors affect the maximal rate.*

Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the affinity of the enzyme for its substrate. This type of inhibition is called *allosteric inhibition*. Most allosterically regulated enzymes are made up of more than one polypeptide, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to an enzyme, all active sites on the protein subunits are changed slightly such that they bind their substrates with less efficiency. There are *allosteric activators* as well as inhibitors. Allosteric activators bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme's active site(s) for its substrate(s).

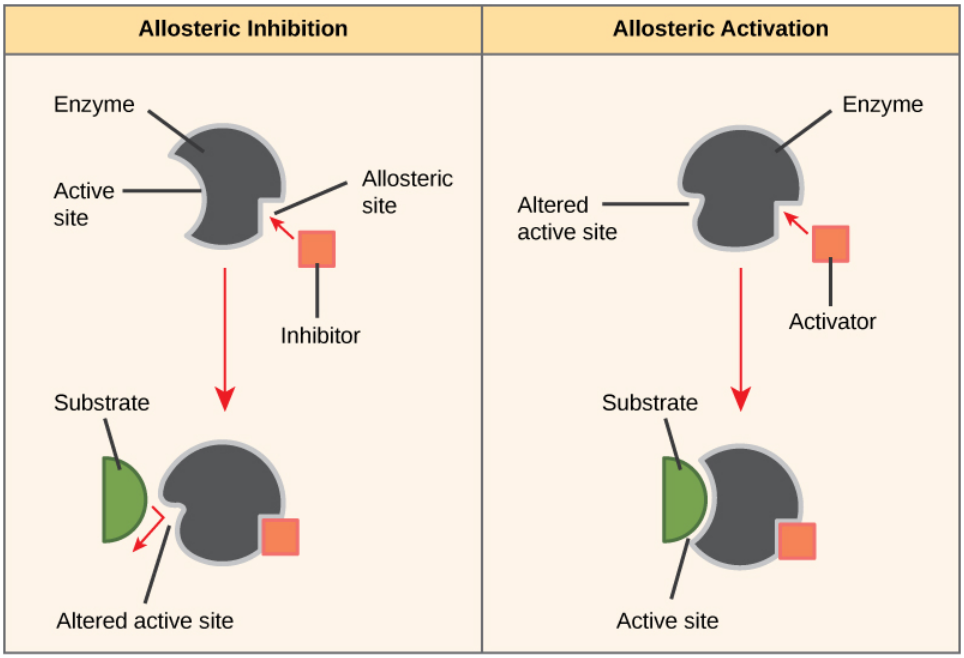


Figure 8. *Allosteric inhibitors modify the active site of the enzyme so that substrate binding is reduced or prevented. In contrast, allosteric activators modify the active site of the enzyme so that the affinity for the substrate increases.*

Video link

Check out this short (one-minute) video on competitive vs. noncompetitive enzymatic inhibition. Also, take a look at this video (1.2 minutes) on feedback inhibition.

Many enzymes don't work optimally, or even at all, unless bound to other specific non-protein helper molecules, either temporarily through ionic or hydrogen bonds or permanently through stronger covalent bonds. Two types of helper molecules are *cofactors* and *coenzymes*. Binding to these molecules promotes optimal conformation and function for their respective enzymes. Cofactors are inorganic ions such as iron(II) (Fe^{2+}) and magnesium(II) (Mg^{2+}). One example of an enzyme that requires a metal ion as a cofactor is the enzyme that builds DNA molecules, DNA polymerase, which requires a bound zinc(II) ion (Zn^{2+}) to function. Coenzymes are organic helper molecules, with a basic atomic structure made up of carbon and hydrogen, that are required for enzyme action. The most common sources of coenzymes are dietary vitamins. Some vitamins are precursors to coenzymes, and others act directly as coenzymes. Vitamin C is a coenzyme for multiple enzymes that take part in building the important connective tissue component, collagen. An important step in the breakdown of glucose to yield energy is catalysis by a multi-enzyme complex called pyruvate dehydrogenase. Pyruvate dehydrogenase is a complex of several enzymes that actually requires one cofactor (a magnesium ion) and five different organic coenzymes to catalyze its specific chemical reaction. Therefore, enzyme function is, in part, regulated by an abundance of various cofactors and coenzymes, which are supplied primarily by the diets of most organisms.

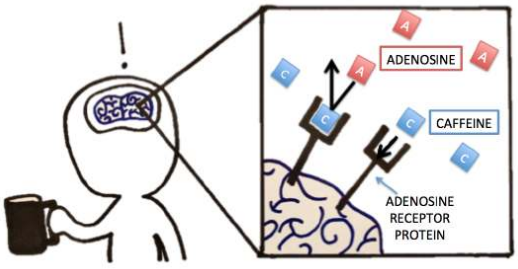
Enzyme compartmentalization

In eukaryotic cells, molecules such as enzymes are usually compartmentalized into different organelles. This allows for yet another level of regulation of enzyme activity. Enzymes required only for certain cellular processes can be housed separately along with their substrates, allowing for more efficient chemical reactions. Examples of this sort of enzyme regulation based on location and proximity include the enzymes involved in the latter stages of cellular respiration, which take place exclusively in the mitochondria, and the enzymes involved in the digestion of cellular debris and foreign materials, located within lysosomes.

Possible NB Discussion Point: Reversing the Effects of Caffeine

Previously, we discussed caffeine and its metabolism. Let's now think about caffeine's pharmacology (mode of action). Were you able to identify, compare, and contrast a molecule that caffeine had a similar structure to? Did you see its similarity to adenosine? If not, go take a look!

Because of caffeine's structural similarity to the molecule adenosine, it is actually able to bind an adenosine-specific receptor protein in the brain. However, because caffeine and adenosine are not structurally identical, caffeine will not "activate" the adenosine receptors upon binding as adenosine normally would. Why is this important? Normally, when adenosine binds to and activates its specific receptor protein in the brain, the physiological effect is increased drowsiness and muscle relaxation. It is believed that we accumulate adenosine over the waking hours of a day and that this leads to increased adenosine receptor activation. This increase in receptor activation then leads to getting tired at night. When caffeine is present, however, it can bind to the adenosine receptor protein, thereby blocking adenosine from binding/activating its receptor. The lack of adenosine action is thought to then lead to suppressed sleepiness and increased alertness. The inhibition seen with this receptor protein and caffeine is similar to some of the inhibition we see with enzymes. From the reading above, what type of inhibition would you classify this as? Follow up question: If you were hired by a company to design a solution to reverse the effect of caffeine, post-ingestion, what strategies would you try to test? Explain!



Check this out for more on why it makes sense to get some sleep.

Additional links

Khan Academy

The following links will take you to a series of videos on kinetics. The first link contains four videos on reaction rates, and the second link contains nine videos related to the relationship between reaction rates and concentration. These videos are supplemental and are provided to give you an outside resource to further explore enzyme kinetics.

- Introduction to enzyme kinetics
- Reaction mechanism

UCD ChemWiki

Allosteric regulation

Reading_06 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_07

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_07

CS.20 Create a labeled illustration depicting a typical bacterial or archaeal cell. The drawing should include some common features that distinguish bacterial and archaeal cells from eukaryotic cells.
CS.21 Describe some of the basic requirements for survival of both bacterial and archaeal cells and discuss some of the physical constraints that may limit the size of bacterial and archaeal cells.
CS.2 Discuss the various functional roles carried out by a biological membrane. Invoke the Design Challenge to describe the problems they solve and the costs and benefits associated with their construction and use.
CS.3 Describe the qualities and components of the plasma membrane.
MS.22 Describe how the chemical structures and properties of lipids contribute to their functions (e.g. formation of membrane structures, contribution to membrane fluidity, etc.).

CI.11 Describe the basic steps in the Design Challenge rubric and how it can be used in the context of learning biology.
CI.12 Create a statement of the key question/problem associated with a topic in the reading or in class using the Design Challenge rubric.
CI.10 Apply the Design Challenge rubric to dissect a complex problem/question into smaller, more manageable questions.

Introduction to bacterial and archaeal diversity



"Perhaps bacteria may tentatively be regarded as biochemical experiments; owing to their relatively small size and rapid growth, variations must arise much more frequently than in more differentiated forms of life, and they can in addition afford to occupy more precarious positions in natural economy than larger organisms with more exacting requirements." - Marjory Stephenson, in Bacterial Metabolism, (1930)

Prokaryotes are single-celled organisms with neither a membrane-bound nucleus nor other lipid membrane-bound organelles. They are composed of two phylogenetically distinct groups of organisms: **Bacteria** and **Archaea**. Although bacteria and archaea share many morphological, structural, and metabolic attributes, there are many differences between the organisms in these two clades. The most notable differences are in the chemical structure and compositions of membrane lipids, the chemical composition of the cell wall, and the makeup of the information processing machinery (e.g., replication, DNA repair, and transcription).

Bacterial and archaeal diversity

Bacteria and archaea were on Earth long before multicellular life appeared. They are ubiquitous and have highly diverse metabolic activities. This diversity allows different species within clades to inhabit every imaginable surface where there is sufficient moisture. For example, there is broad agreement that the typical human body hosts more microbial cells . Bacteria and archaea comprise most living things in all ecosystems. Certain bacterial and archaeal species can thrive in environments inhospitable for most other life. Bacteria and archaea, along with microbial eukaryotes, are also critical for recycling the **nutrients** essential for creating new biomolecules. They also drive the evolution of new ecosystems (natural or man-made).

Bacteria and archaea are adaptable: life in moderate and extreme environments

Some organisms have developed strategies that allow them to survive harsh conditions. Bacteria and archaea thrive in a vast array of environments: some grow in conditions that would seem very normal to us, whereas others are able to thrive and grow under conditions that would kill a plant or an animal. Almost all bacteria and archaea have some form of a cell wall, a protective structure that allows them to survive in both hyper- and hypo-osmotic conditions. Some soil bacteria are able to form endospores that resist heat and drought, thereby allowing the organism to survive until more favorable conditions recur. These adaptations, along with others, allow bacteria to be the most abundant life forms in all terrestrial and aquatic ecosystems.

Some bacteria and archaea are adapted to grow under extreme conditions and are called **extremophiles**, meaning “lovers of extremes.” Extremophiles have been found in all kinds of environments, such as in the depths of the oceans and the earth; in hot springs, the Artic, and the Antarctic; in very dry places; in harsh chemical environments; and in high-radiation environments, just to

mention a few. These organisms help to give us a better understanding of the diversity of life and open up the possibility of finding microbial species that may lead to the discovery of new therapeutic drugs or have industrial applications. Because they have specialized adaptations that allow them to live in extreme conditions, many extremophiles cannot survive in moderate environments. There are many different groups of extremophiles. They are categorized based on the conditions in which they grow best, and several habitats are extreme in multiple ways. For example, a soda lake is both salty and alkaline, so organisms that live in a soda lake must be both **alkaliphiles** and **halophiles**. Other extremophiles, like **radioresistant** organisms, do not prefer an extreme environment (in this case, one with high levels of radiation) but have adapted to survive in it.

Possible NB Discussion Point: How do extremophiles do it?

You just read that soil bacteria are able to survive through heat and droughts by forming dormant heat- and drought-resistant endospores. However, not all extremophiles form endospores as a means to survive their own harsh environmental conditions. Can you think of other strategies that other extremophiles might have developed? Choose a row in Table 1 (below) and try to brainstorm some creative survival mechanisms specific for that extremophile type!

***Table 1**. This table lists some extremophiles and their preferred conditions.*

Extremophile Type	*Conditions for Optimal Growth*
Acidophiles	pH 3 or below
Alkaliphiles	pH 9 or above
Thermophiles	Temperature of 60–80 °C (140–176 °F)
Hyperthermophiles	Temperature of 80–122 °C (176–250 °F)
Psychrophiles	Temperature of -15 °C (5 °F) or lower
Halophiles	Salt concentration of at least 0.2 M
Osmophiles	High sugar concentration

Cellular structure of bacteria and archaea

In this section, we will discuss the basic structural features of both bacteria and archaea. There are many structural, morphological, and physiological similarities between bacteria and archaea. As discussed in the previous section, these microbes inhabit many ecological niches and carry out a great diversity of biochemical and metabolic processes. Both bacteria and archaea lack a membrane-bound nucleus and membrane-bound organelles, which are hallmarks of eukaryotes.

While bacteria and archaea are separate domains, morphologically they share a number of structural features. As a result, they face similar problems, such as the transport of nutrients into the cell, the removal of waste material from the cell, and the need to respond to rapid local environmental changes. In this section, we will focus on how their common cell structure allows them to thrive in various environments and simultaneously puts constraints on them. One of the biggest constraints is related to cell size.

Although bacteria and archaea come in a variety of shapes, the most common three shapes are as follows: cocci (spherical), bacilli (rod-shaped), and spirilli (spiral-shaped) (figure below). Both bacteria and archaea are generally small compared to typical eukaryotes. For example, most bacteria tend to be on the order of 0.2 to 1.0 µm (micrometers) in diameter and 1-10 µm in length. However, there are exceptions. *Epulopiscium fishelsoni* is a bacillus-shaped bacterium that is typically 80 µm in diameter and 200-

600 µm long. *Thiomargarita namibiensis* is a spherical bacterium between 100 and 750 µm in diameter and is visible to the naked eye. For comparison, a typical human neutrophil is approximately 50 µm in diameter.

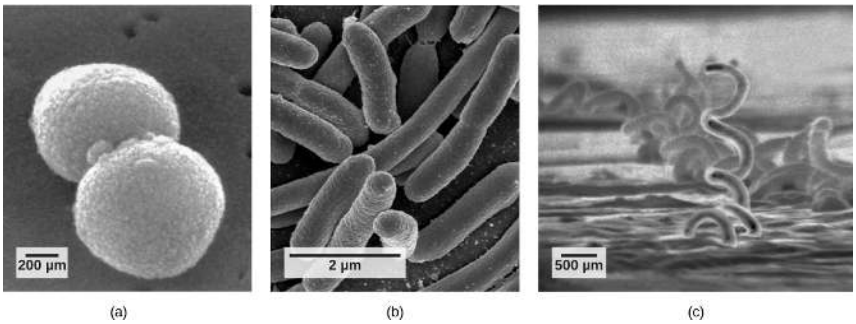


Figure 1. *This figure shows the three most common shapes of bacteria and archaea: (a) cocci (spherical), (b) bacilli (rod-shaped), and (c) spirilli (spiral-shaped).*

Possible NB Discussion Point: Why are bacteria and archaea so tiny anyway?

Why are bacteria and archaea typically so small? What are the constraints that are keeping them microscopic in size (i.e., what is preventing from getting bigger?)? How then exactly does the relatively giant *Thiomargarita namibiensis* (which has a cell volume that is three million times the volume of an average bacteria and is visible to the naked eye) and other larger bacteria overcome these constraints? Think of possible explanations or hypotheses that might answer these questions. We'll explore and develop an understanding of these questions in more detail below and in class.

The bacterial and archaeal cell: common structures

Introduction to the basic cell structure

Bacteria and archaea are unicellular organisms, which lack internal membrane-bound structures that are disconnected from the plasma membrane, a phospholipid membrane that defines the boundary between the inside and outside of the cell. In bacteria and archaea, the cytoplasmic membrane also contains all membrane-bound reactions, including those related to the electron transport chain, ATP synthase, and photosynthesis. By definition, these cells lack a nucleus. Instead, their genetic material is located in a self-defined area of the cell called the nucleoid. The bacterial and archaeal chromosome is often a single covalently closed circular double-stranded DNA molecule. However, some bacteria have linear chromosomes, and some bacteria and archaea have more than one chromosome or small non-essential circular replicating elements of DNA called plasmids. Besides the nucleoid, the next common feature is the cytoplasm (or cytosol), the "aqueous," jelly-like region encompassing the internal portion of the cell. The cytoplasm is where the soluble (non-membrane-associated) reactions occur and contains the ribosomes, the protein-RNA complex where proteins are synthesized. Finally, many bacteria and archaea also have cell walls, the rigid structural feature surrounding the plasma membrane that helps provide protection and constrain the cell shape. You should learn to create a simple sketch of a general bacterial or archaeal cell from memory.

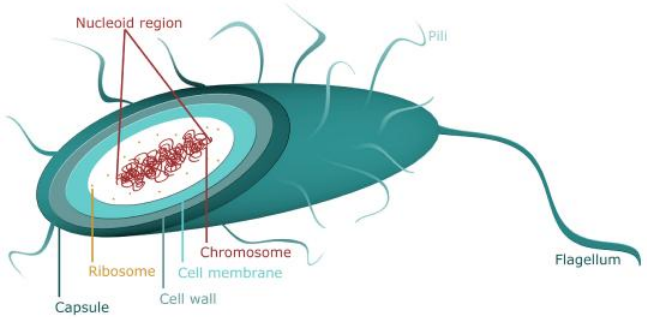


Figure 2. *The features of a typical prokaryotic cell are shown.*

Constraints on the bacterial and archaeal cell

One common, almost universal, feature of bacteria and archaea is that they are small, microscopic to be exact. Even the two examples given as exceptions, *Epulopiscium fishelsoni* and *Thiomargarita namibiensis*, still face the basic constraints all bacteria and archaea face; they simply found unique strategies around the problem. So what is the largest constraint when it comes to dealing with the size of bacteria and archaea? Think about what the cell must do to survive.

Some basic requirements

So what do cells have to do to survive? They need to transform energy into a usable form. This involves making ATP, maintaining an energized membrane, and maintaining productive $NAD^+/NADH_2$ ratios. Cells also need to be able to synthesize the appropriate macromolecules (proteins, lipids, polysaccharides, etc.) and other cellular structural components. To do this, they need to be able to either make the core, key precursors for more complex molecules or get them from the environment.

Diffusion and its importance to bacteria and archaea

Movement by diffusion is passive and proceeds down the concentration gradient. For compounds to move from the outside to the inside of the cell, the compound must be able to cross the phospholipid bilayer. If the concentration of a substance is lower inside the cell than outside and it has chemical properties that allow it to move across the cell membrane, that compound will energetically tend to move into the cell. While the "real" story is a bit more complex and will be discussed in more detail later, diffusion is one of the mechanisms bacteria and archaea use to aid in the transport of metabolites.

Diffusion can also be used to get rid of some waste materials. As waste products accumulate inside the cell, their concentration rises compared to that of the outside environment, and the waste product can leave the cell. Movement within the cell works the same way: compounds will move down their concentration gradient, away from where they are synthesized to places where their concentration is low and therefore may be needed. Diffusion is a random process—the ability of two different compounds or reactants for chemical reactions to interact becomes a meeting of chance. Therefore, in small, confined spaces, random interactions or collisions can occur more frequently than they can in large spaces.

The ability of a compound to diffuse depends on the viscosity of the solvent. For example, it is a lot easier for you to move around in air than in water (think about moving around underwater in a pool). Likewise, it is easier for you to swim in a pool of water than in a pool filled with peanut butter. If you put a drop of food coloring into a glass of water, it quickly diffuses until the entire glass has changed color. Now what do you think would happen if you put that same drop of food coloring into a glass of corn syrup (very viscous and sticky)? It will take a lot longer for the glass of corn syrup to change color.

The relevance of these examples is to note that the cytoplasm tends to be very viscous. It contains many proteins, metabolites, small molecules, etc. and has a viscosity more like corn syrup than water. So, diffusion in cells is slower and more limited than you might have originally expected. Therefore, if cells rely solely on diffusion to move compounds around, what do you think happens to the efficiency of these processes as cells increase in size and their internal volumes get bigger? Is there a potential problem to getting big that is related to the process of diffusion?

So how do cells get bigger?

As you've likely concluded from the discussion above, with cells that rely on diffusion to move things around the cell—like bacteria and archaea—size does matter. So how do you suppose *Epulopiscium fishelsoni* and *Thiomargarita namibiensis* got so big? Take a look at these links, and see what these bacteria look like morphologically and structurally: *Epulopiscium fishelsoni* and *Thiomargarita namibiensis*.

Based on what we have just discussed, in order for cells to get bigger, that is, for their volume to increase, intracellular transport must somehow become independent of diffusion. One of the great evolutionary leaps was the ability of cells (eukaryotic cells) to transport compounds and materials intracellularly, independent of diffusion. Compartmentalization also provided a way to localize processes to smaller organelles, which overcame another problem caused by the large size. Compartmentalization and the complex intracellular transport systems have allowed eukaryotic cells to become very large in comparison to the diffusion-limited bacterial and archaeal cells. We'll discuss specific solutions to these challenges in the following sections.

Membranes



Plasma membranes enclose and define the borders between the inside and the outside of cells. They are typically composed of dynamic bilayers of phospholipids into which various other lipid-soluble molecules and proteins have also been embedded. These bilayers are asymmetric—the outer leaf differing from the inner leaf in lipid composition and in the proteins and carbohydrates that are displayed to either the inside or outside of the cell. One major function of the outer cell membrane is to communicate the cell's unique identity to other cells. The proteins, lipids, and sugars displayed on the cell membrane allow for cells to be detected by and to interact with specific partners.

Various factors influence fluidity, permeability, and various other physical properties of the membrane. These include temperature, the configuration of the fatty acid tails (some are kinked by double bonds), sterols (i.e., cholesterol) embedded in the membrane, and the mosaic nature of the many proteins embedded within it. **The plasma membrane is "selectively permeable".** This means it allows only some substances through while excluding others. In addition, the plasma membrane must, sometimes, be flexible enough to allow certain cells, such as amoebae, to change shape and direction as they move through the environment, hunting smaller, single-celled organisms.

Amoebae Hunting Video

Cellular membranes

A subgoal in our "build-a-cell" design challenge is to create a boundary that separates the "inside" of the cell from the environment "outside". This boundary needs to serve multiple functions that include:

1. Act as a barrier by blocking some compounds from moving in and out of the cell.
2. Be selectively permeable in order to transport specific compounds into and out of the cell.
3. Receive, sense, and transmit signals from the environment to inside of the cell.
4. Project "self" to others by communicating identity to other nearby cells.





Figure 1. *The diameter of a typical balloon is 25cm and the thickness of the plastic of the balloon of around 0.25mm. This is a 1000X difference. A typical eukaryotic cell will have a cell diameter of about 50µm and a cell membrane thickness of 5nm. This is a 10,000X difference.*

Fluid mosaic model

The fluid mosaic model describes the dynamic movement of the numerous proteins, sugars, and lipids embedded in the cell's plasma membrane.

For some insight into the history of our understanding of the plasma membrane structure, click here.

It is sometimes useful to start our discussion by recalling the size of the cell membrane relative to the size of the entire cell cell. Plasma membranes range from 5 to 10 nm in thickness. For comparison, human red blood cells, visible via light microscopy, are approximately 8 µm wide, or approximately 1,000 times wider than a plasma membrane is thick. This means that the cellular barrier is very thin compared to the size of the volume it encloses. Despite this dramatic size differential, the cellular membrane must nevertheless still carry out its key barrier, transport and cellular recognition capacities and so must be a relatively "sophisticated" and dynamic structure.

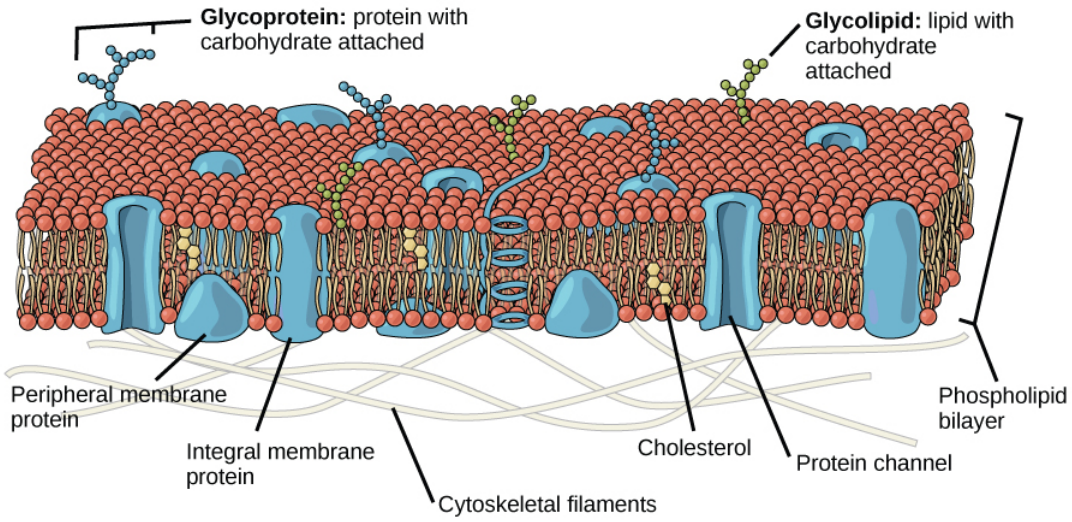


Figure 2. *The fluid mosaic model of the plasma membrane describes the plasma membrane as a fluid combination of phospholipids, cholesterol, and proteins. Carbohydrates attached to lipids (glycolipids) and to proteins (glycoproteins) extend from the outward-facing surface of the membrane.*

The principal components of a plasma membrane are **lipids** (phospholipids and cholesterol), **proteins**, and **carbohydrates**. Carbohydrates are present only on the exterior surface of the plasma membrane and are attached to proteins, forming *glycoproteins*, or to lipids, forming *glycolipids*. The proportions of proteins, lipids, and carbohydrates in the plasma membrane may vary with organism and cell type. In a typical human cell, proteins account for a massive 50 percent of the composition by mass, lipids (of all types) account for about 40 percent of the composition by mass, and carbohydrates account for the remaining 10 percent of the composition by mass. However, cellular functional specialization may cause these ratios of components to vary dramatically. For example, myelin, an outgrowth of the membrane of specialized cells, insulates the axons of the peripheral nerves, contains only 18 percent protein and 76 percent lipid. By contrast, the mitochondrial inner membrane contains 76 percent protein and only 24 percent lipid and the plasma membrane of human red blood cells is 30 percent lipid.

Phospholipids

Phospholipids are major constituents of the cell membrane. *Phospholipids* are made of a glycerol *backbone to which two* fatty acid tails and a phosphate group have been attached - one to each of each of the glycerol carbons atoms. The phospholipid is therefore an *amphipathic* molecule, meaning it has a hydrophobic part (fatty acid tails) and a hydrophilic part (phosphate head group).

Make sure to note in Figure 3 that the phosphate group has an R group linked to one of the oxygen atoms. R is a variable commonly used in these types of diagrams to indicate that some other atom or molecule is bound at that position. That part of the molecule can be different in different phospholipids—and will impart some different chemistry to the whole molecule. At the moment, however, you are responsible for being able to recognize this type of molecule (no matter what the R group is) because of the common core elements—the glycerol backbone, the phosphate group, and the two hydrocarbon tails.

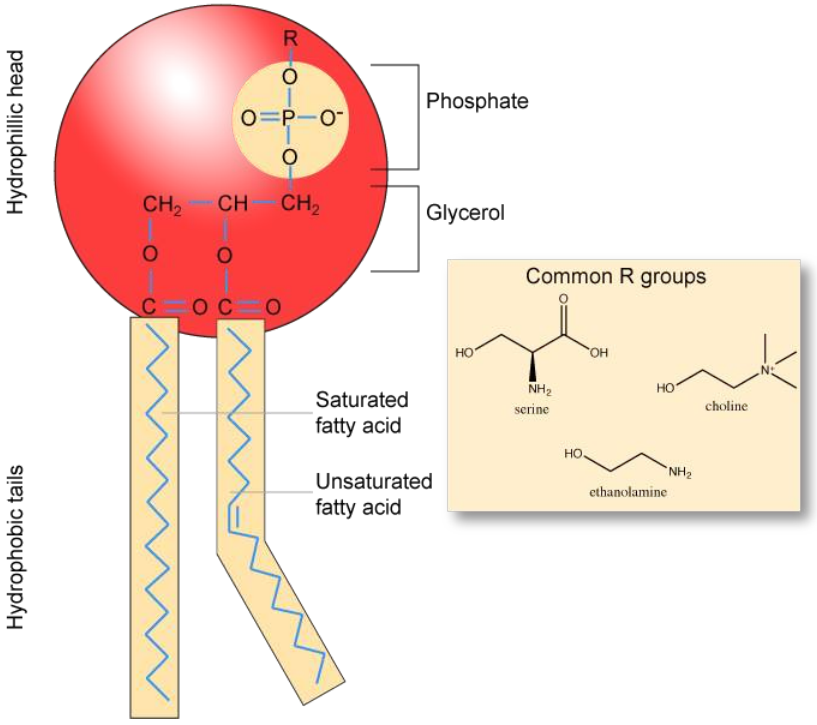


Figure 3. *A phospholipid is a molecule with two fatty acids and a modified phosphate group attached to a glycerol backbone. The phosphate may be modified by the addition of charged or polar chemical groups. Several chemical R groups may modify the phosphate. Choline, serine, and ethanolamine are shown here. These attach to the phosphate group at the position labeled R via their hydroxyl groups.*
Attribution: Marc T. Facciotti (own work)

When many phospholipids are together exposed to an aqueous environment, they can spontaneously arrange themselves into various structures including micelles and phospholipid bilayers. The latter is the basic structure of the cell membrane. In a phospholipid bilayer, the phospholipids associate with one another into two oppositely-facing sheets. In each sheet nonpolar parts of the phospholipids face inward towards one another, composing the internal part of the membrane, and polar head groups facing oppositely to both the aqueous extracellular and intracellular environments.

Possible NB Discussion Point

Earlier in the course, we discussed the Second Law of Thermodynamics, which states that the overall entropy of the universe is always increasing. Apply this law in the context of the formation of the lipid bilayer membrane. How is it possible that the lipids are able to spontaneously arrange themselves into such an organized structure instead of scatter into a

more disordered state? Or in other words -- if the second law holds true, then how exactly does the spontaneous lipid organization lead to increased entropy?

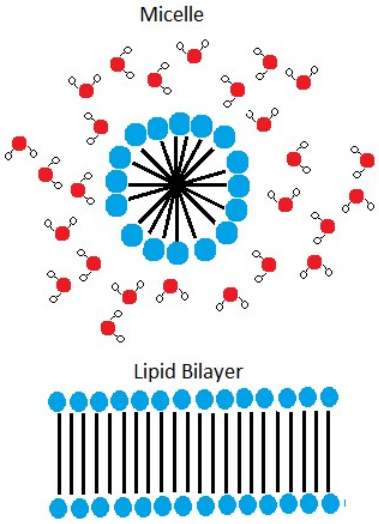


Figure 4. *In the presence of water, some phospholipids will spontaneously arrange themselves into a micelle. The lipids will be arranged such that their polar groups will be on the outside of the micelle, and the nonpolar tails will be on the inside. A lipid bilayer can also form, a two layered sheet only a few nanometers thick. The lipid bilayer consists of two layers of phospholipids organized in a way that all the hydrophobic tails align side by side in the center of the bilayer and are surrounded by the hydrophilic head groups. Source: Created by Erin Easlon (own work)*

Membrane proteins

Proteins make up the second major component of plasma membranes. *Integral membrane proteins*, as their name suggests, integrate completely into the membrane structure, and their hydrophobic membrane-spanning regions interact with the hydrophobic region of the phospholipid bilayer.

Some membrane proteins associate with only one half of the bilayer, while others stretch from one side of the membrane to the other, and are exposed to the environment on either side. Integral membrane proteins may have one or more transmembrane segments typically consisting of 20–25 amino acids. Within the transmembrane segments, hydrophobic amino acid variable groups arrange themselves to form a chemically complementary surface to the hydrophobic tails of the membrane lipids.

Peripheral proteins are found on only one side of the membrane, but never embed into the membrane. They can be on the intracellular side or the extracellular side, and weakly or temporarily associated with the membranes.

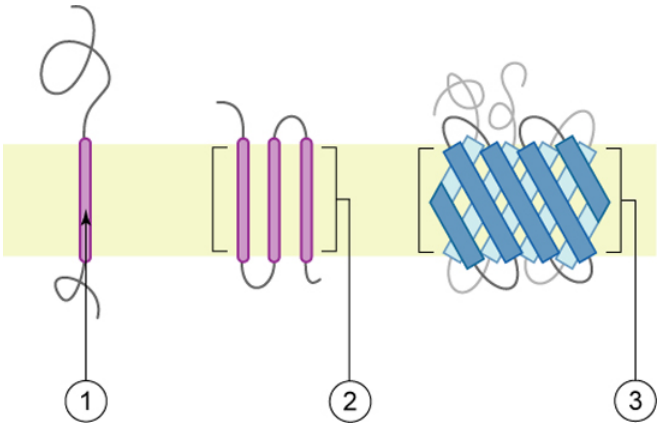


Figure 5. *Integral membranes proteins may have one or more α-helices (pink cylinders) that span the membrane (examples 1 and 2), or they may have β-sheets (blue rectangles) that span the membrane (example 3). (credit: "Foobar"/Wikimedia Commons)*

Carbohydrates

Carbohydrates are a third major component of plasma membranes. They are always found on the exterior surface of cells and are bound either to proteins (forming glycoproteins) or to lipids (forming glycolipids). These carbohydrate chains may consist of 2–60 monosaccharide units and can be either straight or branched. Along with peripheral proteins, carbohydrates form specialized sites on the cell surface that allow cells to recognize each other (one of the core functional requirements noted above.

Membrane fluidity

The integral proteins and lipids exist in the membrane as separate molecules and they "float" in the membrane, moving with respect to one another. The membrane is not like a balloon, however; because of the elastic properties of its plastic a balloon can easily grow and shrink its surface area without popping and while also maintaining the same rough circular shape. By contrast the plasma membrane is not able to withstand isotropic stretching or compression and can be easily popped when an imbalance of solute between inside and out causes water to rush in suddenly. A sudden loss of water will cause it to shrivel and wrinkle, dramatically changing the shape of the cell. it is fairly rigid and can burst if penetrated or if a cell takes in too much water and the membrane is stretched too far. However, because of its mosaic nature, a very fine needle can easily penetrate a plasma membrane without causing it to burst (the lipids flow around the needle point), and the membrane will self-seal when the needle is extracted.

Different organisms and cell types in multicellular organisms can tune fluidity of their membrane to be more compatible with specialized functions and/or in response to environmental factors. This tuning can be accomplished by adjusting the type and concentration of various components of the membrane, including the lipids, their degree of saturation, the lipids, their degree of saturation, the proteins, and other molecules like cholesterol. There are two other factors that help maintain this fluid characteristic. One factor is the nature of the phospholipids themselves. In their saturated form, the fatty acids in phospholipid tails are saturated with hydrogen atoms. There are no double bonds between adjacent carbon atoms, resulting in tails that are relatively straight. By contrast, unsaturated fatty acids do not have a full complement of hydrogen atoms on their fatty acid tails and therefore contain some double bonds between adjacent carbon atoms; a double bond results in a bend in the string of carbons of approximately 30 degrees.

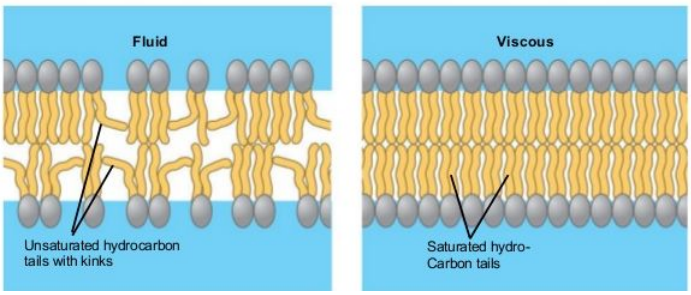


Figure 6. *Any given cell membrane will be composed of a combination of saturated and unsaturated phospholipids. The ratio of the two will influence the permeability and fluidity of the membrane. A membrane composed of completely saturated lipids will be dense and less fluid, and a membrane composed of completely unsaturated lipids will be very loose and very fluid.*

Saturated fatty acids, with straight tails, are compressed by decreasing temperatures, and they will press in on each other, making a dense and fairly rigid membrane. Conversely, when unsaturated fatty acids are compressed, the “kinked” tails elbow adjacent phospholipid molecules away, maintaining some space between the phospholipid molecules. This “elbow room” helps to maintain fluidity in the membrane at temperatures at which membranes with high concentrations of saturated fatty acid tails would “freeze” or solidify. The relative fluidity of the membrane is particularly important in a cold environment. Many organisms (fish are one example) are capable of adapting to cold environments by changing the proportion of unsaturated fatty acids in their membranes in response to the lowering of the temperature.

Cholesterol

Animal cells have cholesterol, an additional membrane constituent that assists in maintaining fluidity. Cholesterol, which lies right in between the phospholipids in the membrane, tends to dampen the effects of temperature on the membrane.Cholesterol both stiffens and increases membrane fluidity, depending on the temperature. Low temperatures cause phospholipids to pack together more tightly, creating a stiffer membrane. In this case, the cholesterol molecules serve to space the phospholipids apart and prevent the membrane from becoming totally rigid. Conversely, higher temperatures contribute to phospholipids moving farther apart from each other and therefore a more fluid membrane, but cholesterol molecules in the membrane take up space and prevent the complete dissociation of phospholipids.

Thus, cholesterol extends, in both directions, the range of temperature in which the membrane is appropriately fluid and consequently functional. Cholesterol also serves other functions, such as organizing clusters of transmembrane proteins into lipid rafts.

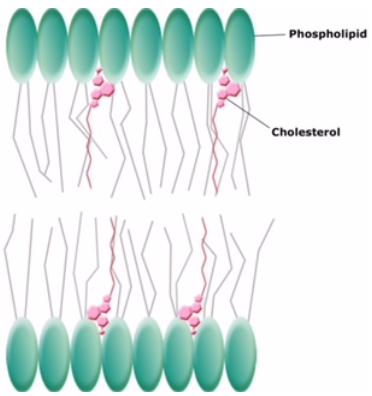


Figure 7. *Cholesterol fits between the phospholipid groups within the membrane.*

Review of the components of the membrane

The components and functions of the plasma membrane	
Component	Location
Phospholipid	Main fabric of the membrane
Cholesterol	Between phospholipids and between the two phospholipid layers of animal cells
Integral proteins (e.g., integrins)	Embedded within the phospholipid layer(s); may or may not penetrate through both layers
Peripheral proteins	On the inner or outer surface of the phospholipid bilayer; not embedded within the phospholipids
Carbohydrates (components of glycoproteins and glycolipids)	Generally attached to proteins on the outside membrane layer

Archaeal membranes

One major difference setting archaea apart from eukaryotes and bacteria is their membrane lipid composition. Though eukaryotes, bacteria, and archaea all use glycerol backbones in their membrane lipids, Archaea use long isoprenoid chains (20-40 carbons in length, derived from the five-carbon lipid isoprene) that are attached via *ether* linkages to glycerol, whereas eukaryotes and bacteria have fatty acids bonded to glycerol via *ester* linkages.

The polar head groups differ based on the genus or species of the Archaea and consist of mixtures of glyco groups (mainly disaccharides), and/or phospho groups primarily of phosphoglycerol, phosphoserine, phosphoethanolamine or phosphoinositol. The inherent stability and unique features of archaeal lipids have made them a useful biomarker for archaea within environmental samples, though approaches based on genetic markers are now more commonly used.

A second difference between bacterial and archaeal membranes that is associated with some archaea is the presence of *monolayer membranes*, as depicted below. Notice that the isoprenoid chain is attached to the glycerol backbones at both ends, forming a single molecule consisting of two polar head groups attached via two isoprenoid chains.

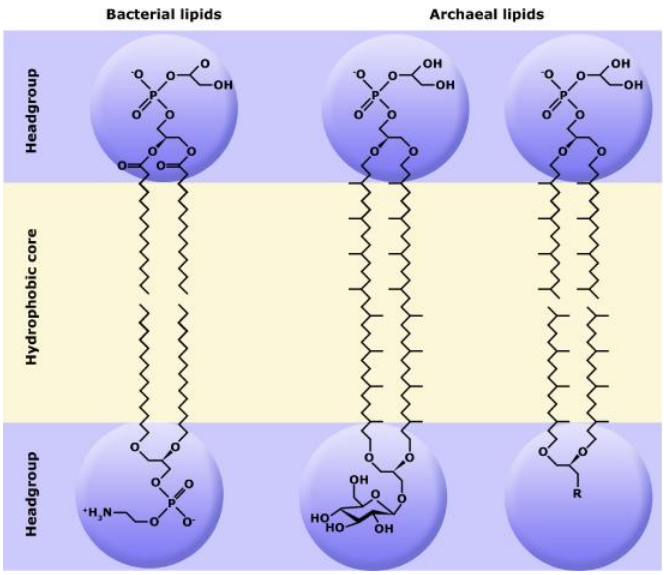


Figure 8. *The exterior surface of the archaeal plasma membrane is not identical to the interior surface of the same membrane.*

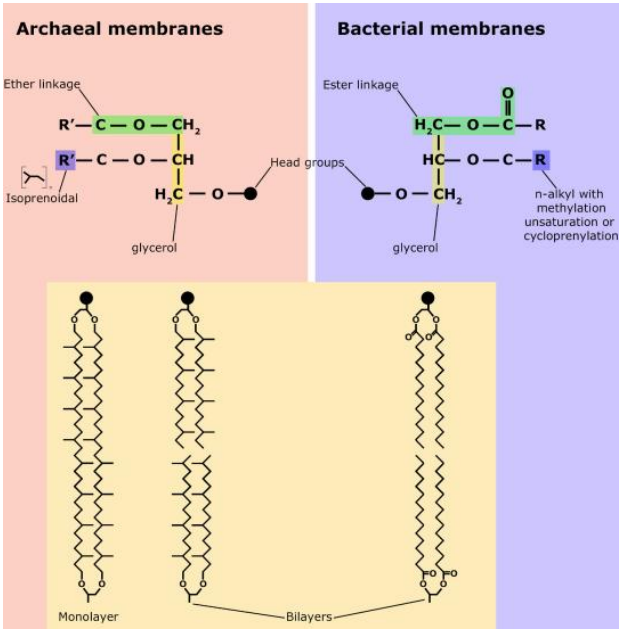


Figure 9. *Comparisons of different types of archaeal lipids and bacterial/eukaryotic lipids*

Reading_07 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_08

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_08

GC.42 Predict whether two reactions can be theoretically productively coupled by interpreting tables of standard Gibbs enthalpy (energy).
GC.50 Describe why an enzyme is critical for coupling an exergonic reaction to an endergonic reaction by applying the concepts from the chemistry and thermodynamics lectures.
CS.5 Discuss the advantages and trade-offs associated with simple diffusion, passive diffusion, facilitated diffusion, and active transport.
MS.16 Predict the properties of the R-groups on amino acids located in different locations within transmembrane transporters.
CS.8 Understand the role of the plasma membrane in maintaining chemical and electrical gradients.
CS.4 Define and correctly use the properties of molecules that can travel through the plasma membrane and contrast them to properties that render other molecules incapable of penetrating the membrane.
CS.1 Explain how molecules move in and out of cells.

Transport across the membrane



Design challenge problem and subproblems

General Problem: The cell membrane must simultaneously act as a barrier between "IN" and "OUT" and control specifically which substances enter and leave the cell and how quickly and efficiently they do so.

Subproblems: The chemical properties of molecules that must enter and leave the cell are highly variable. Some subproblems associated with this are: (a) Large and small molecules or collections of molecules must be able to pass across the membrane. (b) Both hydrophobic and hydrophilic substances must have access to transport. (c) Substances must be able to cross the membrane with and against concentration gradients. (d) Some molecules look very similar (e.g. Na^+ and K^+) but transport mechanisms must still be able to distinguish between them.

Energy story perspective

We can consider transport across a membrane from an energy story perspective; it is a process after all. For instance, at the beginning of the process a generic substance X may be on the inside or outside of the cell. At the end of the process, the substance will be on the opposite side from which it started.

e.g. $X_{(in)}$ ---> $X_{(out)}$,

where in and out refer to inside the cell and outside the cell, respectively.

At the beginning the matter in the system might be a very complicated collection of molecules inside and outside of the cell but with one molecule of X more inside the cell than out. At the end, there is one more molecule of X on the outside of the cell and one less on the inside. The energy in the system at the beginning is stored largely in the molecular structures and their motions and in electrical and chemical concentration imbalances across the cell membrane. The transport of X out of the cell will not change the energies of the molecular structures significantly but it will change the energy associated with the imbalance of concentration and or charge across the membrane. That is the transport will, like all other reactions, be exergonic or endergonic. Finally, some mechanism or sets of mechanisms of transport will need to be described.

Selective permeability



One of the great wonders of the cell membrane is its ability to regulate the concentration of substances inside the cell. These substances include: ions such as Ca^{2+}, Na^+, K^+, and Cl^-; nutrients including sugars, fatty acids, and amino acids; and waste products, particularly carbon dioxide (CO_2), which must leave the cell.

The membrane's lipid bilayer structure provides the first level of control. The phospholipids pack tightly, and the membrane has a hydrophobic interior. This structure alone creates what is known as a **selectively permeable** barrier, one that only allows substances meeting certain physical criteria to pass through it. In the case of the cell membrane, only relatively small,

nonpolar materials can move through the lipid bilayer at biologically relevant rates (remember, the lipid tails of the membrane are nonpolar).

Selective permeability of the cell membrane refers to its ability to differentiate between different molecules, only allowing some molecules through while blocking others. Some of this selective property stems from the intrinsic diffusion rates for different molecules across a membrane. A second factor affecting the relative rates of movement of various substances across a biological membrane is the activity of various protein-based membrane transporters, both passive and active, that will be discussed in more detail in subsequent sections. First, we take on the notion of intrinsic rates of diffusion across the membrane.

Relative permeability

That different substances might cross a biological membrane at different rates should be relatively intuitive. There are differences in the mosaic composition of membranes in biology and differences in the sizes, flexibility, and chemical properties of molecules so it stands to reason that the permeability rates vary. It is a complicated landscape. The permeability of a substance across a biological membrane can be measured experimentally and we can report the rate of movement across a membrane in what are known as membrane permeability coefficients.

Membrane permeability coefficients

Below, we plot a variety of compounds with respect to their membrane permeability coefficients (MPC) as measured against a simple biochemical approximation of a real biological membrane. The reported permeability coefficient for this system is the rate at which simple diffusion through a membrane occurs and is reported in units of centimeters per second (cm/s). The permeability coefficient is proportional to the partition coefficient and is inversely proportional to the membrane thickness.

It is important that you can read and interpret the diagram below. The larger the coefficient, the more permeable the membrane is to the solute. For example, hexanoic acid is very permeable, a MPC of 0.9; acetic acid, water, and ethanol have MPCs between 0.01 and 0.001, and they are less permeable than hexanoic acid. Whereas ions, such as sodium (Na^+), have an MPC of 10^{-12}, and cross the membrane at a comparatively slow rate.

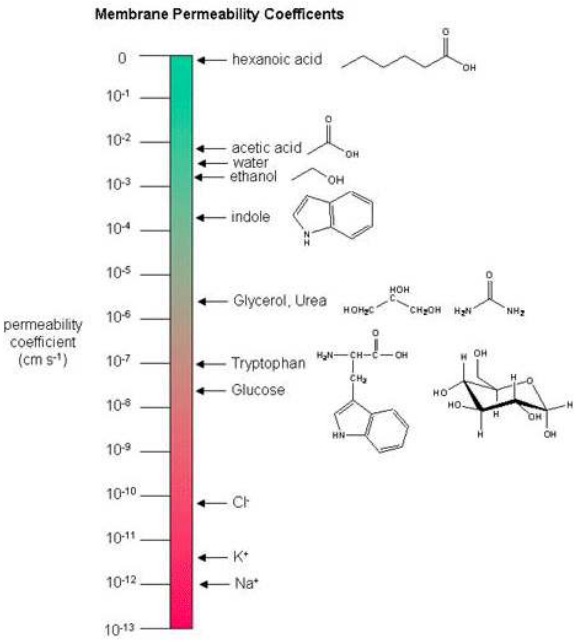


Figure 1. *Membrane permeability coefficient diagram. The diagram was taken from BioWiki and can be found at http://biowiki.ucdavis.edu/Biochemis...e_Permeability.*

While there are certain trends or chemical properties that can be roughly associated with a different compound permeability (small things go through "fast", big things "slowly", charged things not at all etc.), we caution against over-generalizing. The molecular

determinants of membrane permeability are complicated and involve many factors including: the specific composition of the membrane, temperature, ionic composition, hydration; the chemical properties of the solute; the potential chemical interactions between the solute in solution and in the membrane; the dielectric properties of materials; and the energy trade-offs associated with moving substances into and out of various environments. So, in this class, rather than try to apply "rules" and try to develop too many arbitrary "cut-offs", we will strive to develop a general sense of some properties that can influence permeability and leave the assignment of absolute permeability to experimentally reported rates. In addition, we will also try to minimize the use of vocabulary that depends on a frame of reference. For instance, saying that compound A diffuses "quickly" or "slowly" across a bilayer only means something if the terms "quickly" or "slow" are numerically defined or the biological context is understood.

Osmosis

Osmosis is the movement of water through a semipermeable membrane according to the concentration gradient of water across the membrane, which is inversely proportional to the concentration of solutes. While diffusion transports material across membranes and within cells, osmosis transports *only water* across a membrane and the membrane limits the diffusion of solutes in the water. The aquaporins that facilitate water movement play a large role in osmosis, most prominently in red blood cells and the membranes of kidney tubules.

Mechanism

Osmosis is a special case of diffusion. Water, like other substances, moves from an area of high concentration to one of low concentration. An obvious question is what makes water move at all? Imagine a beaker with a semipermeable membrane separating the two sides or halves. On both sides of the membrane the water level is the same, but there are different concentrations of a dissolved substance, or **solute**, that cannot cross the membrane (otherwise the concentrations on each side would be balanced by the solute crossing the membrane). If the volume of the solution on both sides of the membrane is the same, but the concentrations of solute are different, then there are different amounts of water, the solvent, on either side of the membrane.

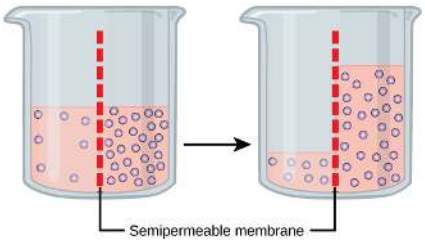


Figure 8. *In osmosis, water always moves from an area of higher water concentration to one of lower concentration. In the diagram shown, the solute cannot pass through the selectively permeable membrane, but the water can.*

To illustrate this, imagine two full glasses of water. One has a single teaspoon of sugar in it, whereas the second one contains one-quarter cup of sugar. If the total volume of the solutions in both cups is the same which cup contains more water? Because the large amount of sugar in the second cup takes up much more space than the teaspoon of sugar in the first cup, the first cup has more water in it.

Returning to the beaker example, recall that it has a mixture of solutes on either side of the membrane. A principle of diffusion is that the molecules move around and will spread evenly throughout the medium if they can. However, only the material capable of getting through the membrane will diffuse through it. In this example, the solute cannot diffuse through the membrane, but the water can. Water has a concentration gradient in this system. Thus, water will diffuse down its concentration gradient, crossing the membrane to the side where it is less concentrated. This diffusion of water through the membrane—osmosis—will continue until the concentration gradient of water goes to zero or until the hydrostatic pressure of the water balances the osmotic pressure. Osmosis proceeds constantly in living systems.

Tonicity

Tonicity describes how an extracellular solution can change the volume of a cell by affecting osmosis. A solution's tonicity often directly correlates with the osmolarity of the solution. **Osmolarity** describes the total solute concentration of the solution. A solution with low osmolarity has a greater number of water molecules relative to the number of solute particles; a solution with

high osmolarity has less water molecules with respect to solute particles. In a situation in which solutions of two different osmolarities are separated by a membrane permeable to water, though not to the solute, water will move from the side of the membrane with lower osmolarity (and more water) to the side with higher osmolarity (and less water). This effect makes sense if you remember that the solute cannot move across the membrane, and thus the only component in the system that can move—the water—moves along its own concentration gradient. An important distinction that concerns living systems is that osmolarity measures the number of particles (which may be molecules) in a solution. Therefore, a solution that is cloudy with cells may have a lower osmolarity than a solution that is clear if the second solution contains more dissolved molecules than there are cells.

Hypotonic solutions

Three terms—hypotonic, isotonic, and hypertonic—are used to relate the osmolarity of a cell to the osmolarity of the extracellular fluid that contains the cells. In a **hypotonic** situation, the extracellular fluid has lower osmolarity than the fluid inside the cell, and water enters the cell (in living systems, the point of reference is always the cytoplasm, so the prefix *hypo-* means that the extracellular fluid has a lower concentration of solutes, or a lower osmolarity, than the cell cytoplasm). It also means that the extracellular fluid has a higher concentration of water in the solution than does the cell. In this situation, water will follow its concentration gradient and enter the cell.

Hypertonic solutions

As for a **hypertonic** solution, the prefix *hyper-* refers to the extracellular fluid having a higher osmolarity than the cell's cytoplasm; therefore, the fluid contains less water than the cell does. Because the cell has a relatively higher concentration of water, water will leave the cell.

Isotonic solutions

In an **isotonic** solution, the extracellular fluid has the same osmolarity as the cell. If the osmolarity of the cell matches that of the extracellular fluid, there will be no net movement of water into or out of the cell, although water will still move in and out. Blood cells and plant cells in hypertonic, isotonic, and hypotonic solutions take on characteristic appearances.

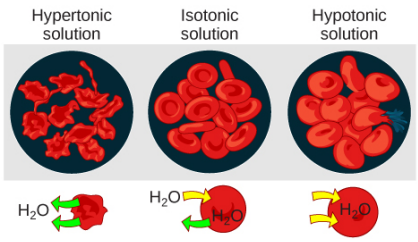


Figure 9. *Osmotic pressure changes the shape of red blood cells in hypertonic, isotonic, and hypotonic solutions. (credit: Mariana Ruiz Villareal)*

Possible NB Discussion Point

Of course there is such thing as drinking too little water... but is there such thing as drinking too much water? Discuss what you think happens when you drink an excessive amount of water -- what is happening at the level of the cell membrane? What is happening to the cell size? Is drinking too much water actually a health hazard? Predict what would happen if we changed the beverage to Gatorade instead of water.

Tonicity in living systems

In a hypotonic environment, water enters a cell, and the cell swells. In an isotonic condition, the relative concentrations of solute and solvent are equal on both sides of the membrane. There is no net water movement; therefore, there is no change in the size of

the cell. In a hypertonic solution, water leaves a cell and the cell shrinks. If either the hypo- or hyper- condition goes to excess, the cell's functions become compromised, and the cell may be destroyed.

A red blood cell will burst, or lyse, when it swells beyond the plasma membrane's capability to expand. Remember, the membrane resembles a mosaic, with discrete spaces between the molecules composing it. If the cell swells, and the spaces between the lipids and proteins become too large, and the cell will break apart.

In contrast, when excessive amounts of water leave a red blood cell, the cell shrinks. This has the effect of concentrating the solutes left in the cell, making the cytosol denser and interfering with diffusion within the cell. The cell's ability to function will be compromised and may also result in the cell's death.

Various living things have ways of controlling the effects of osmosis—a mechanism called osmoregulation. Some organisms, such as plants, fungi, bacteria, and some protists, have cell walls that surround the plasma membrane and prevent cell lysis in a hypotonic solution. The plasma membrane can only expand to the limit of the cell wall, so the cell will not lyse. In fact, the cytoplasm in plants is always slightly hypertonic to the cellular environment, and water will always enter a cell if water is available. This inflow of water produces turgor pressure, which stiffens the cell walls of the plant. In nonwoody plants, turgor pressure supports the plant. Conversely, if the plant is not watered, the extracellular fluid will become hypertonic, causing water to leave the cell. In this condition, the cell does not shrink because the cell wall is not flexible. However, the cell membrane detaches from the wall and constricts the cytoplasm. This is called **plasmolysis**. Plants lose turgor pressure in this condition and wilt.

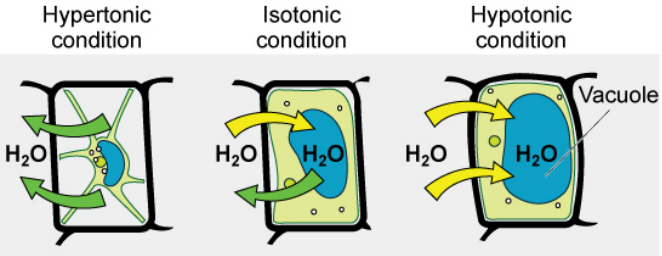


Figure 10. *The turgor pressure within a plant cell depends on the tonicity of the solution that it is bathed in. (credit: modification of work by Mariana Ruiz Villareal)*



Figure 11. *Without adequate water, the plant on the left has lost turgor pressure, visible in its wilting; the turgor pressure is restored by watering it (right). (credit: Victor M. Vicente Selvas)*

Tonicity is a concern for all living things. For example, paramecia and amoebas, which are protists that lack cell walls, have contractile vacuoles. This vesicle collects excess water from the cell and pumps it out, keeping the cell from bursting as it takes on water from its environment.

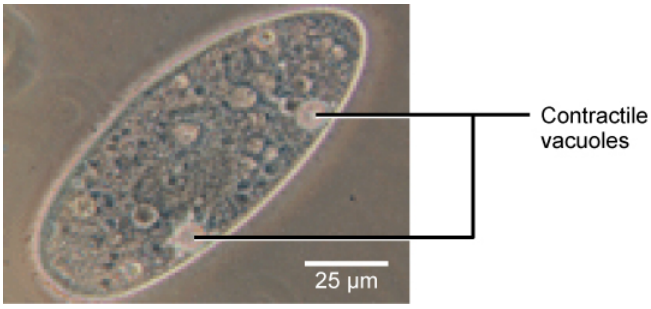


Figure 12. *A paramecium's contractile vacuole, here visualized using bright field light microscopy at 480x magnification, continuously pumps water out of the organism's body to keep it from bursting in a hypotonic medium. (credit: modification of work by NIH; scale-bar data from Matt Russell)*

Many marine invertebrates have internal salt levels matched to their environments, making them isotonic with the water in which they live. Fish, however, must spend approximately five percent of their metabolic energy maintaining osmotic homeostasis. Freshwater fish live in an environment that is hypotonic to their cells. These fish actively take in salt through their gills and excrete diluted urine to rid themselves of excess water. Saltwater fish live in the reverse environment, which is hypertonic to their cells, and they secrete salt through their gills and excrete highly concentrated urine.

In vertebrates, the kidneys regulate the amount of water in the body. Osmoreceptors are specialized cells in the brain that monitor the concentration of solutes in the blood. If the levels of solutes increase beyond a certain range, a hormone is released that retards water loss through the kidney and dilutes the blood to safer levels. Animals also have high concentrations of albumin, which is produced by the liver, in their blood. This protein is too large to pass easily through plasma membranes and is a major factor in controlling the osmotic pressures applied to tissues.

Energetics of transport



All substances that move through the membrane do so by one of two general methods, which are categorized based on whether the transport process is exergonic or endergonic. **Passive transport** is the exergonic movement of substances across the membrane. In contrast, **active transport** is the endergonic movement of substances across the membrane that is coupled to an exergonic reaction.

Passive Transport

Passive transport does not require the cell to expend energy. In passive transport, substances move from an area of higher concentration to an area of lower concentration, down the **concentration gradient** and energetically favorable. Depending on the chemical nature of the substance, different processes may be associated with passive transport.

Diffusion

Diffusion is a passive process of transport. A single substance tends to move from an area of high concentration to an area of low concentration until the concentration is equal across a space. You are familiar with diffusion of substances through the air. For example, think about someone opening a bottle of ammonia in a room filled with people. The ammonia gas is at its highest concentration in the bottle; its lowest concentration is at the edges of the room. The ammonia vapor will diffuse, or spread away, from the bottle, and gradually, more and more people will smell the ammonia as it spreads. Materials move within the cell's cytosol by diffusion, and certain materials move through the plasma membrane by diffusion.

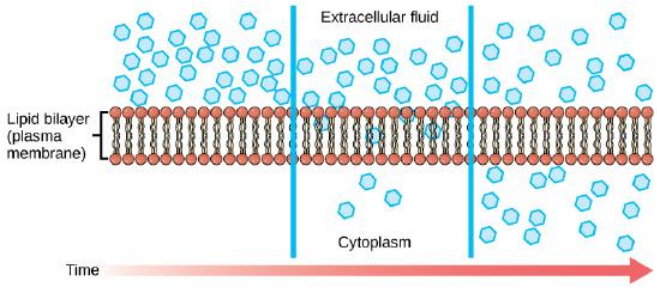


Figure 1. *Diffusion through a permeable membrane moves a substance from an area of high concentration (extracellular fluid, in this case) down its concentration gradient (into the cytoplasm). Each separate substance in a medium, such as the extracellular fluid, has its own concentration gradient, independent of the concentration gradients of other materials. In addition, each substance will diffuse according to that gradient. Within a system, there will be different rates of diffusion of the different substances in the medium (Attribution: Mariana Ruiz Villareal, modified)*

Factors That Affect Diffusion

If unconstrained, molecules will move through and explore space randomly at a rate that depends on their size, their shape, their environment, and their thermal energy. This type of movement underlies the diffusive movement of molecules through whatever medium they are in. The absence of a concentration gradient does not mean that this movement will stop, just that there may be no ***net*** movement of the number of molecules from one area to another, a condition known as **dynamic equilibrium**.

Factors influencing diffusion include:

- Extent of the concentration gradient: The greater the difference in concentration, the more rapid the diffusion. The closer the distribution of the material gets to equilibrium, the slower the rate of diffusion becomes.
- Shape, size and mass of the molecules diffusing: Large and heavier molecules move more slowly; therefore, they diffuse more slowly. The reverse is typically true for smaller, lighter molecules.
- Temperature: Higher temperatures increase the energy and therefore the movement of the molecules, increasing the rate of diffusion. Lower temperatures decrease the energy of the molecules, thus decreasing the rate of diffusion.
- Solvent density: As the density of a solvent increases, the rate of diffusion decreases. The molecules slow down because they have a more difficult time getting through the denser medium. If the medium is less dense, rates of diffusion increase. Since cells primarily use diffusion to move materials within the cytoplasm, any increase in the cytoplasm's density will decrease the rate at which materials move in the cytoplasm.
- Solubility: As discussed earlier, nonpolar or lipid-soluble materials pass through plasma membranes more easily than polar materials, allowing a faster rate of diffusion.
- Surface area and thickness of the plasma membrane: Increased surface area increases the rate of diffusion, whereas a thicker membrane reduces it.
- Distance traveled: The greater the distance that a substance must travel, the slower the rate of diffusion. This places an upper limitation on cell size. A large, spherical cell will die because nutrients or waste cannot reach or leave the center of the cell, respectively. Therefore, cells must either be small in size, as in the case of many prokaryotes, or be flattened, as with many single-celled eukaryotes.

Facilitated transport

In **facilitated transport**, also called facilitated diffusion, materials diffuse across the plasma membrane with the help of membrane proteins. A concentration gradient exists that allows these materials to diffuse into or out of the cell without expending cellular energy. In the case that the materials are ions or polar molecules, compounds that are repelled by the hydrophobic parts of the cell membrane, facilitated transport proteins help shield these materials from the repulsive force of the membrane, allowing them to diffuse into the cell.

Possible NB Discussion Point

Channels

The integral proteins involved in facilitated transport are collectively referred to as **transport proteins**, and they function as either channels for the material or carriers. In both cases, they are transmembrane proteins. Different channel proteins have different transport properties. Some have evolved to be have very high specificity for the substance that is being transported while others transport a variety of molecules sharing some common characteristic(s). The interior "passageway" of **channel proteins** have evolved to provide a low energetic barrier for transport of substances across the membrane through the complementary arrangement of amino acid functional groups (of both backbone and side-chains). Passage through the channel allows polar compounds to avoid the nonpolar central layer of the plasma membrane that would otherwise slow or prevent their entry into the cell. While at any one time significant amounts of water crosses the membrane both in and out the rate of individual water molecule transport may not be fast enough to adapt to changing environmental conditions. For such cases Nature has evolved a special class of membrane proteins called **aquaporins** that allow water to pass through the membrane at a very high rate.

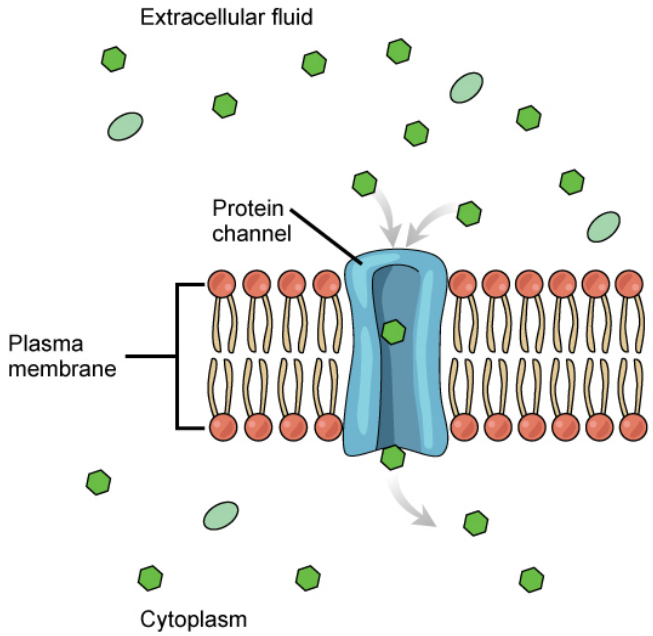


Figure 2: *Facilitated transport moves substances down their concentration gradients. They may cross the plasma membrane with the aid of channel proteins. (Attribution: Mariana Ruiz Villareal, modified.)*

Channel proteins are either open at all times or they are “gated.” The latter controls the opening of the channel. Various mechanisms may be involved in the gating mechanism. For instance, the attachment of a specific ion or small molecule to the channel protein may trigger opening. Changes in local membrane "stress" or changes in voltage across the membrane may also be triggers to open or close a channel.

Different organisms and tissues in multicellular species express different sets of channel proteins in their membranes depending on the environments they live in or specialized function they play in an organisms. This provides each type of cell with a unique membrane permeability profile that is evolved to complement its "needs" (note the anthropomorphism). For example, in some tissues, sodium and chloride ions pass freely through open channels, whereas in other tissues a gate must be opened to allow passage. This occurs in the kidney, where both forms of channels are found in different parts of the renal tubules. Cells involved in the transmission of electrical impulses, such as nerve and muscle cells, have gated channels for sodium, potassium, and calcium in

their membranes. Opening and closing of these channels changes the relative concentrations on opposing sides of the membrane of these ions, resulting a change in electrical potential across the membrane that lead to message propagation in the case of nerve cells or in muscle contraction in the case of muscle cells.

Carrier Proteins

Another type of protein embedded in the plasma membrane is a **carrier protein**. This aptly named protein binds a substance and, in doing so, triggers a change of its own shape, moving the bound molecule from the outside of the cell to its interior; depending on the gradient, the material may move in the opposite direction. Carrier proteins are typically specific for a single substance. This selectivity adds to the overall selectivity of the plasma membrane. The molecular-scale mechanism of function for these proteins remains poorly understood.

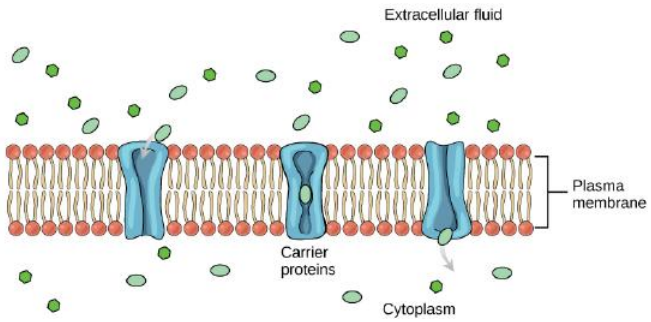


Figure 3: *Some substances are able to move down their concentration gradient across the plasma membrane with the aid of carrier proteins. Carrier proteins change shape as they move molecules across the membrane. (Attribution: Mariana Ruiz Villareal, modified.)*

Carrier proteins play an important role in the function of kidneys. Glucose, water, salts, ions, and amino acids needed by the body are filtered in one part of the kidney. This filtrate, which includes glucose, is then reabsorbed in another part of the kidney with the help of carrier proteins. Because there are only a finite number of carrier proteins for glucose, if more glucose is present in the filtrate than the proteins can handle, the excess is not reabsorbed and it is excreted from the body in the urine. In a diabetic individual, this is described as “spilling glucose into the urine.” A different group of carrier proteins called glucose transport proteins, or GLUTs, are involved in transporting glucose and other hexose sugars through plasma membranes within the body.

Channel and carrier proteins transport materials at different rates. Channel proteins transport much more quickly than do carrier proteins. Channel proteins facilitate diffusion at a rate of tens of millions of molecules per second, whereas carrier proteins work at a rate of a thousand to a million molecules per second.

Note: A note of appreciation

The rates of transport just discussed are astounding. Recall that these molecular catalysts are on the scale of 10s of nanometers (10^{-9} meters) and that they are composed of a self-folding string of 20 amino acids and the relatively small selection of chemical functional groups that they carry.

Active Transport



Active transport mechanisms require the use of the cell’s energy, usually in the form of adenosine triphosphate (ATP). If a substance must move into the cell against its concentration gradient—that is, if the concentration of the substance inside the cell is greater than its concentration in the extracellular fluid (and vice versa)—the cell must use energy to move the substance. Some active transport mechanisms move small-molecular weight materials, such as ions, through the membrane. Other mechanisms transport much larger molecules.

Moving Against a Gradient

To move substances against a concentration or electrochemical gradient, the cell must use energy. This energy is harvested from ATP generated through the cell's metabolism. Active transport mechanisms, collectively called **pumps**, work against electrochemical gradients. Small substances constantly pass through plasma membranes. Active transport maintains concentrations of ions and other substances needed by living cells in the face of these passive movements. Much of a cell's supply of metabolic energy may be spent maintaining these processes. (Most of a red blood cell's metabolic energy is used to maintain the imbalance between exterior and interior sodium and potassium levels required by the cell.) Because active transport mechanisms depend on a cell's metabolism for energy, they are sensitive to many metabolic poisons that interfere with the supply of ATP.

Two mechanisms exist for the transport of small-molecular weight material and small molecules. **Primary active transport** moves ions across a membrane and creates a difference in charge across that membrane, which is directly dependent on ATP. **Secondary active transport** describes the movement of material that is due to the electrochemical gradient established by primary active transport that does not directly require ATP.

Carrier Proteins for Active Transport

An important membrane adaption for active transport is the presence of specific carrier proteins or pumps to facilitate movement: there are three types of these proteins or **transporters**. A **uniporter** carries one specific ion or molecule. A **symporter** carries two different ions or molecules, both in the same direction. An **antiporter** also carries two different ions or molecules, but in different directions. All of these transporters can also transport small, uncharged organic molecules like glucose. These three types of carrier proteins are also found in facilitated diffusion, but they do not require ATP to work in that process. Some examples of pumps for active transport are Na^+-K^+ ATPase, which carries sodium and potassium ions, and H^+-K^+ ATPase, which carries hydrogen and potassium ions. Both of these are antiporter carrier proteins. Two other carrier proteins are Ca^{2+} ATPase and H^+ ATPase, which carry only calcium and only hydrogen ions, respectively. Both are pumps.

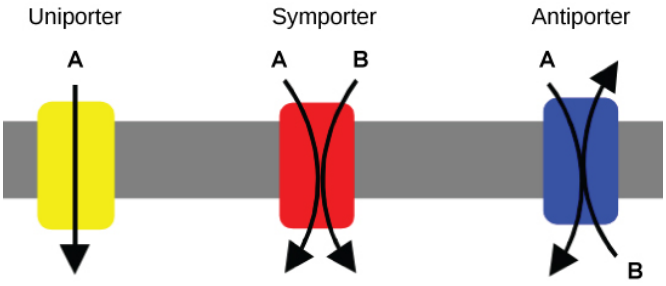


Figure 1: A uniporter carries one molecule or ion. A symporter carries two different molecules or ions, both in the same direction. An antiporter also carries two different molecules or ions, but in different directions. (credit: modification of work by "Lupask"/Wikimedia Commons)

Primary Active Transport

In primary active transport, the energy is derived directly from the breakdown of ATP. Often times, primary active transport such as that shown below which functions to transport sodium and potassium ions allows secondary active transport to occur (discussed in the section below). The second transport method is still considered active because it depends on the use of energy from the primary transport.

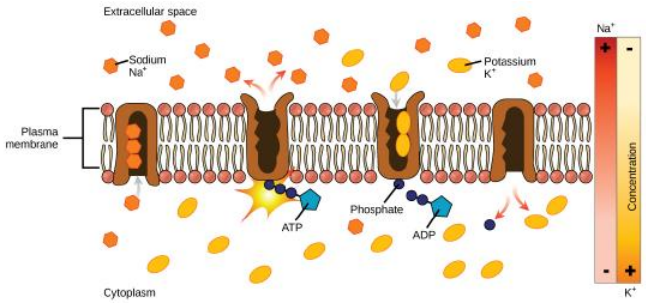


Figure 2: Primary active transport moves ions across a membrane, creating an electrochemical gradient (electrogenic transport). (credit: modification of work by Mariana Ruiz Villareal)

One of the most important pumps in animal cells is the sodium-potassium pump (Na^+-K^+ ATPase), which maintains the electrochemical gradient (and the correct concentrations of Na^+ and K^+) in living cells. The sodium-potassium pump moves K^+ into the cell while moving Na^+ out at the same time, at a ratio of three Na^+ for every two K^+ ions moved in. The Na^+-K^+ ATPase exists in two forms, depending on its orientation to the interior or exterior of the cell and its affinity for either sodium or potassium ions. The process consists of the following six steps.

1. With the enzyme oriented towards the interior of the cell, the carrier has a high affinity for sodium ions. Three ions bind to the protein.
2. ATP is hydrolyzed by the protein carrier and a low-energy phosphate group attaches to it.
3. As a result, the carrier changes shape and re-orients itself towards the exterior of the membrane. The protein's affinity for sodium decreases and the three sodium ions leave the carrier.
4. The shape change increases the carrier's affinity for potassium ions, and two such ions attach to the protein. Subsequently, the low-energy phosphate group detaches from the carrier.
5. With the phosphate group removed and potassium ions attached, the carrier protein repositions itself towards the interior of the cell.
6. The carrier protein, in its new configuration, has a decreased affinity for potassium, and the two ions are released into the cytoplasm. The protein now has a higher affinity for sodium ions, and the process starts again.

Several things have happened as a result of this process. At this point, there are more sodium ions outside of the cell than inside and more potassium ions inside than out. For every three ions of sodium that move out, two ions of potassium move in. This results in the interior being slightly more negative relative to the exterior. This difference in charge is important in creating the conditions necessary for the secondary process. The sodium-potassium pump is, therefore, an **electrogenic pump** (a pump that creates a charge imbalance), creating an electrical imbalance across the membrane and contributing to the membrane potential.

Link to Learning

Visit the site to see a simulation of active transport in a sodium-potassium ATPase.

Secondary Active Transport (Co-transport)

Secondary active transport brings sodium ions, and possibly other compounds, into the cell. As sodium ion concentrations build outside of the plasma membrane because of the action of the primary active transport process, an electrochemical gradient is created. If a channel protein exists and is open, the sodium ions will be pulled through the membrane. This movement is used to transport other substances that can attach themselves to the transport protein through the membrane. Many amino acids, as well as glucose, enter a cell this way. This secondary process is also used to store high-energy hydrogen ions in the mitochondria of plant and animal cells for the production of ATP. The potential energy that accumulates in the stored hydrogen ions is translated into kinetic energy as the ions surge through the channel protein ATP synthase, and that energy is used to convert ADP into ATP.

Tags recommended by the template: article:topic

Reading_08 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_09

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_09

CS.9 Solve basic mathematical problems involving the surface-area-to-volume relationship and interpret the numerical solutions in terms of functional challenges faced by the cell.
CS.10 Recognize examples of strategies for increasing membrane surface area and be able to explain what problems are solved by these solutions.
CS.21 Describe some of the basic requirements for survival of both bacterial and archaeal cells and discuss some of the physical constraints that may limit the size of bacterial and archaeal cells.
CS.11 Discuss some of the functional challenges a crowded cell interior poses, and construct hypotheses for how these challenges might be overcome.
CS.17 Identify and illustrate the components of subcellular infrastructure that are important for distinguishing between prokaryotic and eukaryotic cells, and different types of eukaryotic cells.
CS.18 Discuss the advantages of the subcellular compartmentalization of biochemical processes/reactions into a specific organelle and the evolutionary trade-offs of this localization on the evolution of cellular processes.
CS.19 Describe the functional roles and interrelationships of the cell membrane, nucleus, mitochondria, endoplasmic reticulum, Golgi apparatus, peroxisome, lysosome, vacuole, centrosomes, and chloroplast.
CS.12 Create an energy story describing the coupling of exergonic cellular reactions to the transport of molecules by motor proteins throughout the cell.
CS.13 Compare and contrast microfilaments and microtubules and describe the important roles they play in the cell.
CS.14 Compare and contrast the motor proteins: kinesins and dyneins.
CS.15 Discuss how the cytoskeleton and motor proteins function together to address various problems associated with transport within the cell.
CS.20 Create a labeled illustration depicting a typical bacterial or archaeal and eukaryotic cell. The drawing should include some common features that distinguish bacterial and archaeal cells from eukaryotic cells.

Eukaryotes



Living things fall into three large groups: Archaea, Bacteria, and Eukarya. The first two groups include non-nucleated cells, and the third contains all eukaryotes. A relatively sparse fossil record is available to help us discern what the first members of each of these lineages looked like, so it is possible that all the events that led up to the last common ancestor of extant eukaryotes will remain unknown. However, comparative biology of extant organisms and the limited fossil record provide some insights into the history of Eukarya.

The earliest fossils found appear to be bacteria, most likely cyanobacteria. They are about 3.5 billion years old and are recognizable because of their relatively complex structure and, for bacteria, relatively large cells. Most other bacteria and archaea have small cells, 1 or 2 μm in size, and would be difficult to pick out as fossils. Most living eukaryotes have cells measuring 10 μm or greater. Structures this size, which might be fossils, appear in the geological record about 2.1 billion years ago.

Characteristics of eukaryotes

Data from these fossils have led biologists to the conclusion that living eukaryotes are all descendants of a single common ancestor. Mapping the characteristics found in all major groups of eukaryotes reveals that the following characteristics must have been present in the last common ancestor, because these characteristics are present in at least some of the members of each major lineage.

1. Cells with nuclei surrounded by a nuclear envelope with nuclear pores. This is the single characteristic that is both necessary and sufficient to define an organism as a eukaryote. All extant eukaryotes have cells with nuclei.
2. Mitochondria. Some extant eukaryotes have very reduced remnants of mitochondria in their cells, whereas other members of their lineages have “typical” mitochondria.
3. A cytoskeleton containing the structural and motility components called actin microfilaments and microtubules. All extant eukaryotes have these cytoskeletal elements.

4. Flagella and cilia, organelles associated with cell motility. Some extant eukaryotes lack flagella and/or cilia, but they are descended from ancestors that possessed them.
5. Chromosomes, each consisting of a linear DNA molecule coiled around basic (alkaline) proteins called histones. The few eukaryotes with chromosomes lacking histones clearly evolved from ancestors that had them.
6. Mitosis, a process of nuclear division wherein replicated chromosomes are divided and separated using elements of the cytoskeleton. Mitosis is universally present in eukaryotes.
7. Sex, a process of genetic recombination unique to eukaryotes in which diploid nuclei at one stage of the life cycle undergo meiosis to yield haploid nuclei and subsequent karyogamy, a stage where two haploid nuclei fuse together to create a diploid zygote nucleus.
8. Members of all major lineages have cell walls, and it might be reasonable to conclude that the last common ancestor could make cell walls during some stage of its life cycle. However, not enough is known about eukaryotes' cell walls and their development to know how much homology exists among them. If the last common ancestor could make cell walls, it is clear that this ability must have been lost in many groups.

Endosymbiosis and the evolution of eukaryotes

In order to understand eukaryotic organisms fully, it is necessary to understand that all extant eukaryotes are descendants of a chimeric organism that was a composite of a host cell and the cell(s) of an alpha-proteobacterium that "took up residence" inside it. This major theme in the origin of eukaryotes is known as *endosymbiosis*, one cell engulfing another such that the engulfed cell survives and both cells benefit. Over many generations, a symbiotic relationship can result in two organisms that depend on each other so completely that neither could survive on its own. Endosymbiotic events likely contributed to the origin of the last common ancestor of today's eukaryotes and to later diversification in certain lineages of eukaryotes. Before explaining this further, it is necessary to consider metabolism in bacteria and archaea.

Bacterial and archaeal metabolism

Many important metabolic processes arose in bacteria and archaea, and some of these, such as nitrogen fixation, are never found in eukaryotes. The process of aerobic respiration is found in all major lineages of eukaryotes, and it is localized in the mitochondria. Aerobic respiration is also found in many lineages of bacteria and archaea, but it is not present in all of them, and many forms of evidence suggest that such anaerobic microbes never carried out aerobic respiration nor did their ancestors.

While today's atmosphere is about one-fifth molecular oxygen (O_2), geological evidence shows that it originally lacked O_2. Without oxygen, aerobic respiration would not be expected, and living things would have relied on fermentation instead. Around 3.5 billion years ago, some bacteria and archaea began using energy from sunlight to power anabolic processes that reduce carbon dioxide to form organic compounds. They evolved the ability to photosynthesize. Hydrogen, derived from various sources, was captured using light-powered reactions to reduce fixed carbon dioxide in the Calvin cycle. The group of Gram-negative bacteria that gave rise to cyanobacteria used water as the hydrogen source and released O_2 as a waste product.

Eventually, the amount of photosynthetic oxygen built up in some environments to levels that posed a risk to living organisms, since it can damage many organic compounds. Various metabolic processes evolved that protected organisms from oxygen; one of which, aerobic respiration, also generated high levels of ATP. It became widely present among microbes, including in a group we now call alpha-proteobacteria. Organisms that did not gain aerobic respiration had to remain in oxygen-free environments. Originally, oxygen-rich environments were likely localized around places where cyanobacteria were active, but by about 2 billion years ago, geological evidence shows that oxygen was building up to higher concentrations in the atmosphere. Oxygen levels similar to today's levels only arose within the last 700 million years.

Recall that the first fossils that we believe to be eukaryotes are about 2 billion years old, so they appeared as oxygen levels were increasing. Also, recall that all extant eukaryotes descended from an ancestor with mitochondria. These organelles were first observed by light microscopists in the late 1800s, where they appeared to be worm-shaped structures that seemed to move around in the cell. Some early observers suggested that they might be bacteria living inside host cells, but these hypotheses remained unknown or rejected in most scientific communities.

Endosymbiotic theory

As cell biology developed in the twentieth century, it became clear that mitochondria were the organelles responsible for producing ATP using aerobic respiration. In the 1960s, American biologist Lynn Margulis developed *endosymbiotic theory*, which states that eukaryotes may have been a product of one cell engulfing another (one living within another) and evolving until the separate cells were no longer recognizable as such. In 1967, Margulis introduced new work on the theory and substantiated her findings through microbiological evidence. Although Margulis' work initially was met with resistance, this once-revolutionary hypothesis is now widely (but not completely) accepted, with work progressing on uncovering the steps involved in this evolutionary process and the key players involved. Much remains to be discovered about the origins of the cells that now make up the cells in all living eukaryotes.

Broadly, it has become clear that many of our nuclear genes and the molecular machinery responsible for replication and expression appear closely related to those in Archaea. On the other hand, the metabolic organelles and genes responsible for many energy-harvesting processes had their origins in bacteria. Much remains to be clarified about how this relationship occurred; this continues to be an exciting field of discovery in biology. For instance, it is not known whether the endosymbiotic event that led to mitochondria occurred before or after the host cell had a nucleus. Such organisms would be among the extinct precursors of the last common ancestor of eukaryotes.

Mitochondria

One of the major features distinguishing bacteria and archaea from eukaryotes is mitochondria. Eukaryotic cells may contain anywhere from one to several thousand mitochondria, depending on the cell's level of energy consumption. Each mitochondrion measures 1 to 10 or greater micrometers in length and exists in the cell as an organelle that can be ovoid, worm-shaped, or intricately branched. Mitochondria arise from the division of existing mitochondria; they may fuse; and they may be moved around inside the cell by interactions with the cytoskeleton. However, mitochondria cannot survive outside the cell. As the atmosphere was oxygenated by photosynthesis, and as successful aerobic microbes evolved, evidence suggests that an ancestral cell with some membrane compartmentalization engulfed a free-living aerobic bacterium, specifically an alpha-proteobacterium, thereby giving the host cell the ability to use oxygen to release energy stored in nutrients. Alpha-proteobacteria are a large group of bacteria that includes species symbiotic with plants, disease organisms that can infect humans via ticks, and many free-living species that use light for energy. Several lines of evidence support that mitochondria are derived from this endosymbiotic event. Most mitochondria are shaped like alpha-proteobacteria and are surrounded by two membranes, which would result when one membrane-bound organism is engulfed into a vacuole by another membrane-bound organism. The mitochondrial inner membrane is extensive and involves substantial infoldings called cristae that resemble the textured, outer surface of alpha-proteobacteria. The matrix and inner membrane are rich with the enzymes necessary for aerobic respiration.

Eukaryotic Cell: Structure and Function

Introduction to eukaryotic cell structure

By definition, ***eukaryotic cells*** are cells that contain a membrane-bound nucleus, which is not present in bacterial or archaeal cells. Besides the nucleus, eukaryotic cells are characterized by many membrane-bound **organelles** such as the endoplasmic reticulum, Golgi apparatus, chloroplasts, mitochondria, and others.

In previous sections, we considered the Design Challenge of making cells larger than a small bacterium—more precisely, growing cells to sizes at which, in the eyes of natural selection, relying on diffusion of substances for transport through a highly viscous cytosol comes with inherent functional trade-offs that offset most selective benefits of getting larger. In the lectures and readings on bacterial cell structure, we discovered some morphological features of large bacteria that allow them to effectively overcome diffusion-limited size barriers (e.g., filling the cytoplasm with a large storage vacuole maintains a small volume for metabolic activity that remains compatible with diffusion-driven transport).

As we transition our focus to eukaryotic cells, we want you to approach the study by constantly returning to the Design Challenge. We will cover many subcellular structures unique to eukaryotes, and **you will be expected to know the names of these structures**

or organelles, to associate them with one or more "functions", and to identify them on a canonical cartoon representation of a eukaryotic cell. We will also ask you to think deeper about some functional and evolutionary costs and benefits **(trade-offs)** of both evolving eukaryotic cells and various eukaryotic organelles, and how a eukaryotic cell might coordinate the functions of different organelles.

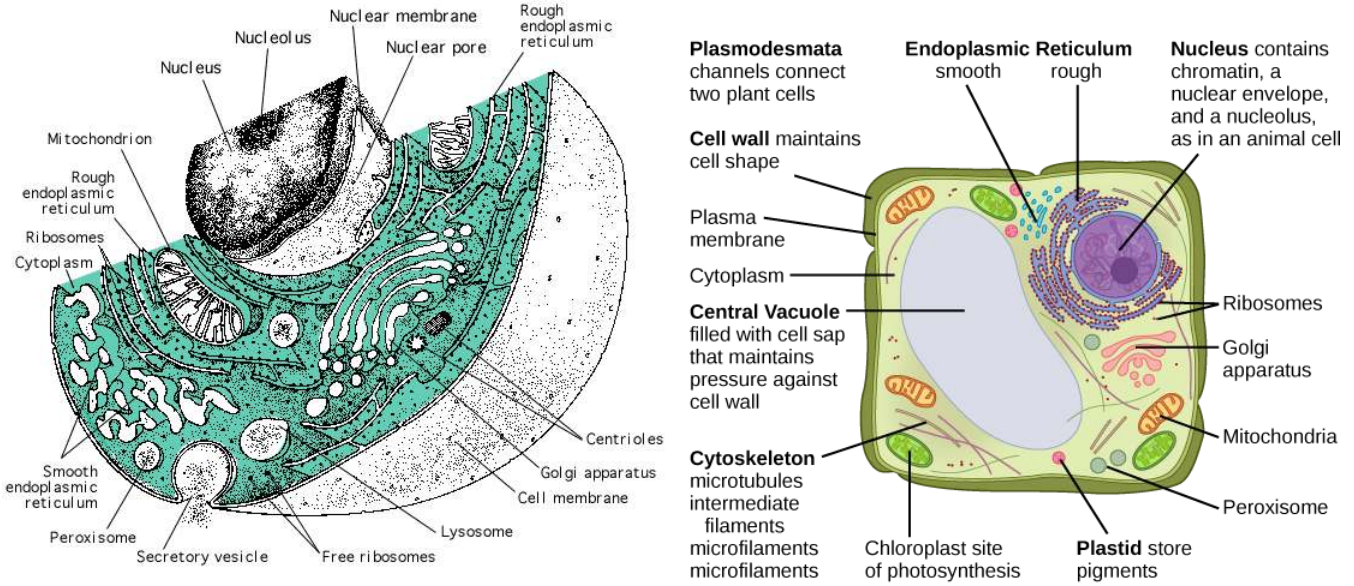


Figure 1. *These figures show the major organelles and other cell components of (a) a typical animal cell and (b) a typical eukaryotic plant cell. The plant cell has a cell wall, chloroplasts, plastids, and a central vacuole—structures not found in animal cells. Plant cells do not have lysosomes or centrosomes. Panel (a) from <https://upload.wikimedia.org/wikiped...dimensions.png>*

The plasma membrane

Like bacteria and archaea, eukaryotic cells have a **plasma membrane**, a phospholipid bilayer with embedded proteins that separates the internal contents of the cell from its surrounding environment. The plasma membrane controls the passage of organic molecules, ions, water, and oxygen into and out of the cell. Wastes (such as carbon dioxide and ammonia) also leave the cell by passing through the plasma membrane, usually with some help of protein transporters.

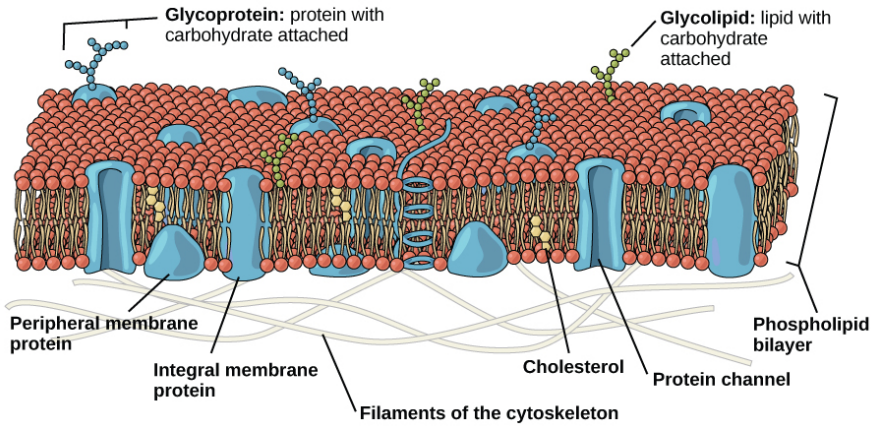


Figure 2. *The eukaryotic plasma membrane is a phospholipid bilayer with proteins and cholesterol embedded in it.*

As discussed in bacterial cell membranes, the plasma membranes of eukaryotic cells may also adopt unique structural conformations. For instance, in cells that specialize in absorption the plasma membrane often fold into fingerlike projections called microvilli (singular = microvillus) (see figure below). The "folding" of the membrane into microvilli increases the surface area for absorption while minimally impacting the cytosolic volume. We can find such cells lining the small intestine, the organ that absorbs nutrients from digested food.

An aside: People with celiac disease have an immune response to gluten, a protein found in wheat, barley, and rye. The immune response damages microvilli. As a result, afflicted individuals have an impaired ability to absorb nutrients. This can lead to malnutrition, cramping, and diarrhea.

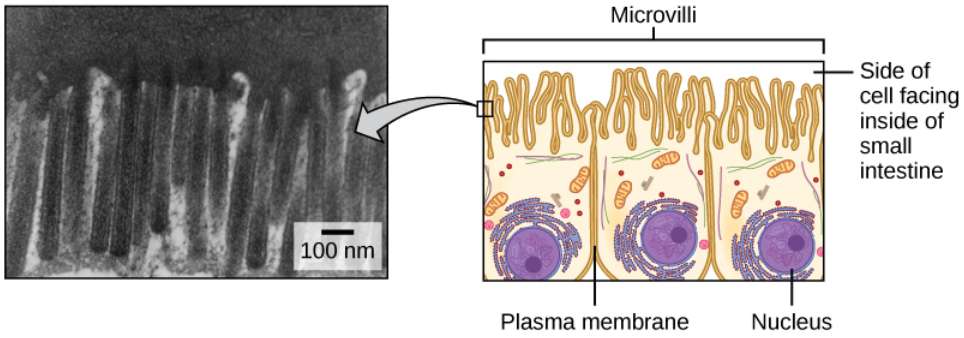


Figure 3. *Microvilli, shown here as they appear on cells lining the small intestine, increase the surface area available for absorption. These microvilli are only found on the area of the plasma membrane that faces the cavity from which substances will be absorbed. Credit: "micrograph", modification of work by Louisa Howard*

The cytoplasm

The **cytoplasm** refers to the entire region of a cell between the plasma membrane and the nuclear envelope. It is composed of organelles suspended in the gel-like **cytosol**, the cytoskeleton, and various chemicals. Even though the cytoplasm consists of 70 to 80 percent water, it nevertheless has a semisolid consistency. It is very crowded in there. Proteins, simple sugars, polysaccharides, amino acids, nucleic acids, fatty acids, ions, and many other water-soluble molecules are all competing for space and water.

The nucleus

Typically, the nucleus is the most prominent organelle in a cell (see figure below) when viewed through a microscope. The **nucleus** (plural = nuclei) houses the cell's DNA. Let's look at it in more detail.

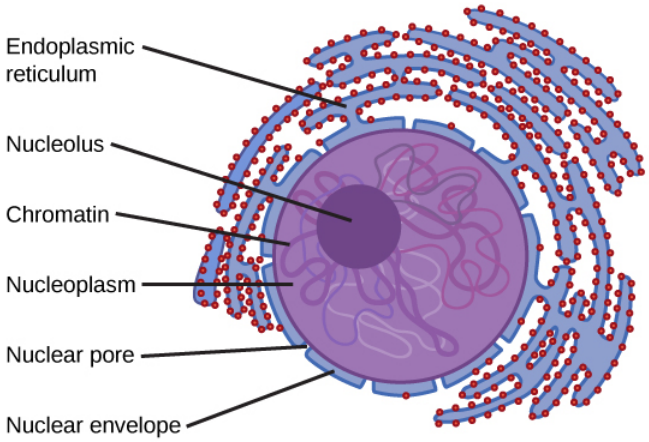


Figure 4. *The nucleus stores chromatin (DNA plus proteins) in a gel-like substance called the nucleoplasm. The nucleolus is a condensed region of chromatin where ribosome synthesis occurs. The boundary of the nucleus is called the nuclear envelope. It*

consists of two phospholipid bilayers: an outer membrane and an inner membrane. The nuclear membrane is continuous with the endoplasmic reticulum. Nuclear pores allow substances to enter and exit the nucleus.

The nuclear envelope

The **nuclear envelope** ***is a double phospholipid bilayer*** that constitutes the outermost boundary of the nucleus. The nuclear envelope is also punctuated with nuclear pores, which are protein-based channels that control the passage of ions, molecules, and RNA between the nucleoplasm and cytoplasm. The **nucleoplasm** is the name for the semisolid fluid inside the nucleus where we find chromatin and the **nucleolus**, a condensed region of chromatin where ribosome synthesis occurs.

Chromatin and chromosomes

Chromosomes are structures within the nucleus that are made up of long coils of DNA. Recall that in bacteria and archaea, DNA is typically organized into one or more circular chromosome(s). In eukaryotes, chromosomes are linear structures. Every eukaryotic species has a specific number of chromosomes in the nuclei of its cells. In humans, for example, the chromosome number is 23, while in fruit flies, it is 4.

Chromosomes are only clearly distinguishable from one another by visible light microscopy when the cell is preparing to divide and the DNA is tightly packed by proteins into easily distinguishable shapes. When the cell is in the growth and maintenance phases of its life cycle, numerous proteins are still associated with the nucleic acids, but the DNA strands more closely resemble an unwound, jumbled bunch of threads. The term **chromatin** is used to describe chromosomes (the protein-DNA complexes) when they are both condensed and decondensed.

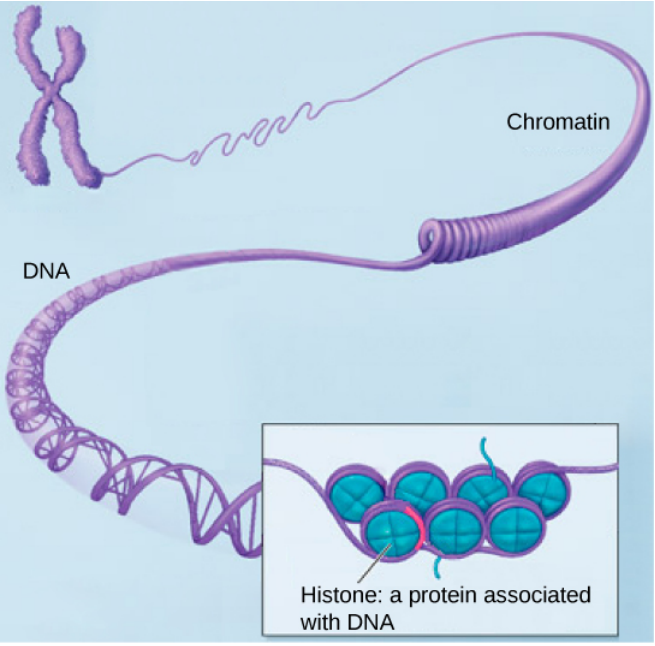


Figure 5. *(a) This image shows various levels of the organization of chromatin (DNA and protein). (b) This image shows paired chromosomes. Credit (b): modification of work by NIH; scale-bar data from Matt Russell*

The nucleolus

Some chromosomes have sections of DNA that encode ribosomal RNA. A darkly staining area within the nucleus called the **nucleolus** (plural = nucleoli) aggregates the ribosomal RNA with associated proteins to assemble the ribosomal subunits that are then transported out to the cytoplasm through the nuclear envelope pores.

Ribosomes

Ribosomes are the cellular structures responsible for the process of protein synthesis referred to as **translation**. When viewed through an electron microscope, ribosomes appear either as clusters (polyribosomes) or single, tiny dots that float freely in the cytoplasm. They may also appear to be attached to (1) the cytoplasmic side of the plasma membrane or (2) the cytoplasmic side of the endoplasmic reticulum and (3) the outer membrane of the nuclear envelope.

Electron microscopy has shown us that ribosomes, which are large complexes of protein and RNA, consist of two subunits, aptly called large and small (figure below). Ribosomes receive their "instructions" for protein synthesis from the nucleus, where the DNA is transcribed into messenger RNA (mRNA). The mRNA travels to the ribosomes, which translate the code provided by the sequence of the nitrogenous bases in the mRNA into a specific order of amino acids in a protein. This is covered in greater detail in the section covering the process of translation.

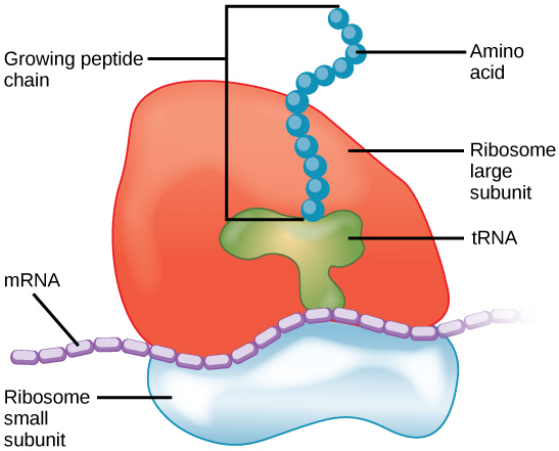


Figure 6. *Ribosomes are made up of a large subunit (top) and a small subunit (bottom). During protein synthesis, ribosomes assemble amino acids into proteins.*

Mitochondria

Mitochondria (singular = mitochondrion) are often called the “powerhouses” or “energy factories” of a cell because they are the primary site of metabolic respiration in eukaryotes. Depending on the species and the type of mitochondria found in those cells, the respiratory pathways may be anaerobic or aerobic. By definition, when respiration is **aerobic**, the terminal electron acceptor is **oxygen**; when respiration is anaerobic, a compound other than oxygen functions as the terminal electron acceptor. In either case, the result of these respiratory processes is the production of ATP via oxidative phosphorylation. Recall that ATP is a “high-energy” molecule the cell can use to power several different processes. Mitochondria also possess a small genome that encodes genes whose functions are typically restricted to the mitochondrion.

In some cases, the number of mitochondria per cell is adjustable depending, typically, on energy demand. For example, cells from tissues that are consistently more active (for instance those of athletes who consistently train need lots of ATP). To solve this problem, those cells often have significantly more mitochondria than cells that do not require such a large constant energy supply.

The structure of the mitochondria can vary significantly depending on the organism and the state of the cell cycle which one is observing. The typical textbook image, however, depicts mitochondria as oval-shaped organelles with a double inner and outer membrane (see figure below); learn to recognize this generic representation. Both the inner and outer membranes are phospholipid bilayers embedded with proteins that mediate transport across them and catalyze various other biochemical reactions. The inner membrane layer has folds called **cristae** that increase the surface area into which respiratory chain proteins can be embedded. The region within the cristae is called the mitochondrial **matrix** and contains enzymes of the TCA cycle. During respiration, protons are pumped by respiratory chain complexes from the matrix into a region known as the **intermembrane space** (which is between the inner and outer mitochondrial membranes).

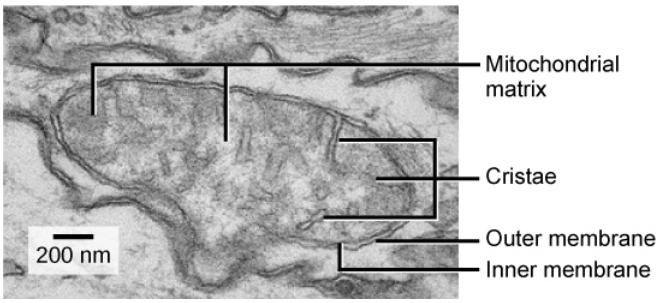


Figure 7. *This electron micrograph shows a mitochondrion as viewed with a transmission electron microscope. This organelle has an outer membrane and an inner membrane. The inner membrane contains folds, called cristae, which increase its surface area. The space between the two membranes is called the intermembrane space, and the space inside the inner membrane is called the mitochondrial matrix. ATP synthesis takes place on the inner membrane. Credit: modification of work by Matthew Britton; scale-bar data from Matt Russell*

There are many other organelles, but they all serve essential functions to the cell. Let's introduce a few more.

Peroxisomes

Peroxisomes are small, round organelles enclosed by single membranes. These organelles carry out chemical reactions known as **redox reactions** that oxidize and break down fatty acids and amino acids. They also help to detoxify many toxins that may enter the body. Many of these redox reactions release hydrogen peroxide, H_2O_2, which would be damaging to cells; however, when these reactions are confined to peroxisomes, enzymes safely break down the H_2O_2 into harmless oxygen and water. For example, alcohol is detoxified by peroxisomes in liver cells. Glyoxysomes, which are specialized peroxisomes in plants, are responsible for converting stored fats into sugars.

Lysosomes

Animal cells have another set of organelles not found in plant cells: **lysosomes**. Colloquially, the lysosomes are sometimes called the cell's "garbage disposal". Enzymes within the lysosomes aid the breakdown of proteins, polysaccharides, lipids, nucleic acids, and even "worn-out" organelles. These enzymes are active at a much lower pH than that of the cytoplasm. Therefore, the pH within lysosomes is more acidic than the pH of the cytoplasm. In plant cells, many of the same digestive processes take place in vacuoles.

Vesicles and vacuoles

Vesicles and *vacuoles* are membrane-bound sacs that function in storage and transport. Other than the fact that vacuoles are somewhat larger than vesicles, there is a very subtle distinction between them: the membranes of vesicles can fuse with either the plasma membrane or other membrane systems within the cell. Additionally, some agents such as enzymes within plant vacuoles break down macromolecules. The membrane of a vacuole does not fuse with the membranes of other cellular components.

The central vacuole

Vacuoles as essential components of plant cells. If you look at the cartoon figure of the plant cell, you will see that it depicts a large central vacuole that occupies most of the area of the cell. The **central vacuole** plays a key role in regulating the cell's concentration of water in changing environmental conditions.

Vacuole factoid: Have you ever noticed that if you forget to water a plant for a few days, it wilts? That's because as the water concentration in the soil becomes lower than the water concentration in the plant, water moves out of the central vacuoles and cytoplasm. As the central vacuole shrinks, it leaves the cell wall unsupported. This loss of support to the cell walls of plant cells results in the wilted appearance of the plant.

The central vacuole also supports the expansion of the cell. When the central vacuole holds more water, the cell gets larger without having to invest a lot of energy in synthesizing new cytoplasm.

The centrosome

The *centrosome* is the organelle where all microtubules originate in animal and yeast cells. It is also a microtubule-organizing center found near the nuclei of animal cells. It contains a pair of centrioles, two structures that lie perpendicular to each other (see figure below). Each centriole is a cylinder of nine triplets of microtubules.

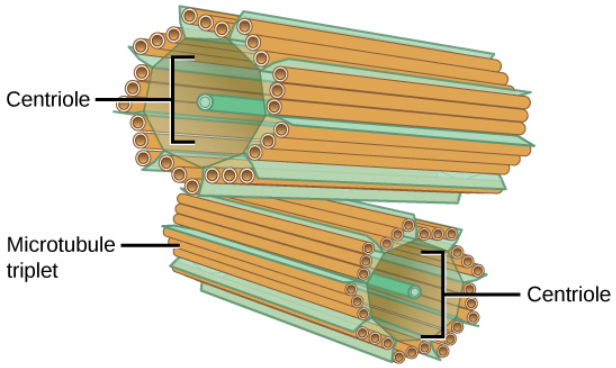


Figure 8. *The centrosome consists of two centrioles that lie at right angles to each other. Each centriole is a cylinder made up of nine triplets of microtubules. Nontubulin proteins (indicated by the green lines) hold the microtubule triplets together.*

The centrosome replicates itself before a cell divides, and the centrioles appear to have some role in pulling the duplicated chromosomes to opposite ends of the dividing cell.

The cell wall

If you examine the diagram above depicting plant and animal cells, you will see in the diagram of a plant cell a structure external to the plasma membrane called the **cell wall**. The cell wall is a rigid covering that protects the cell, provides structural support, and gives shape to the cell.

Fungal and protistan cells also have cell walls. While the chief component of bacterial cell walls is peptidoglycan, the major organic molecule in the plant cell wall is cellulose (see structure below), a polysaccharide made up of glucose subunits.

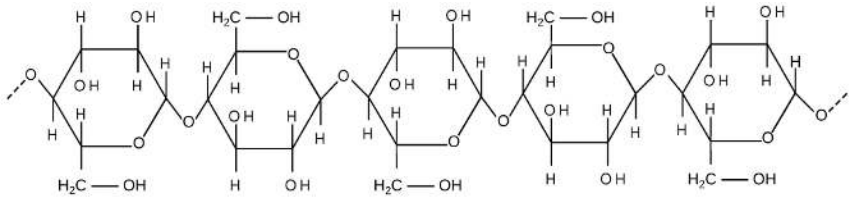


Figure 9. *Cellulose is a long chain of β-glucose molecules connected by a 1-4 linkage. The dashed lines at each end of the figure indicate a series of many more glucose units. The size of the page makes it impossible to portray an entire cellulose molecule.*

Chloroplasts

Chloroplasts are plant cell organelles that carry out photosynthesis. Like the mitochondria, chloroplasts have their own DNA and ribosomes, but chloroplasts have an entirely different function.

Like mitochondria, chloroplasts have outer and inner membranes, but within the space enclosed by a chloroplast's inner membrane is a set of interconnected and stacked fluid-filled membrane sacs called *thylakoids* (figure below). Each stack of thylakoids is called a *granum* (plural = grana). The fluid enclosed by the inner membrane that surrounds the grana is called the stroma.

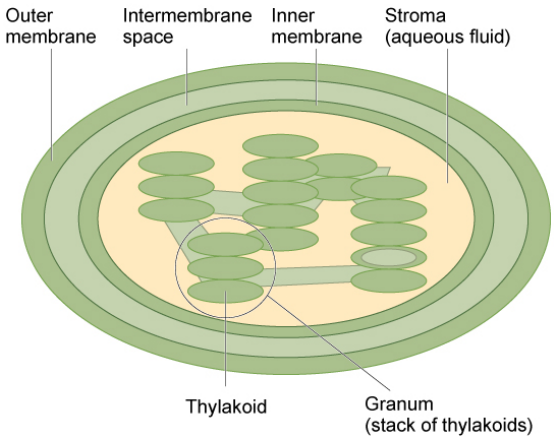
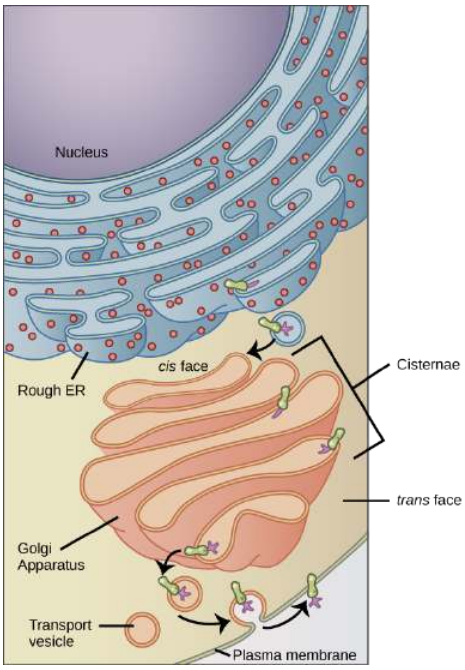


Figure 10. *The chloroplast has an outer membrane, an inner membrane, and membrane structures called thylakoids that are stacked into grana. The space inside the thylakoid membranes is called the thylakoid space. The light harvesting reactions take place in the thylakoid membranes, and the synthesis of sugar takes place in the fluid inside the inner membrane, which is called the stroma. Chloroplasts also have their own genome, which is contained on a single circular chromosome.*

The chloroplasts contain a green pigment called *chlorophyll*, which captures the light energy that drives the reactions of photosynthesis. Like plant cells, photosynthetic protists also have chloroplasts. Some bacteria perform photosynthesis, but their chlorophyll is not relegated to an organelle.

The Endomembrane System

The endomembrane system (endo = "within") is a group of membranes and organelles in eukaryotic cells that works together to change, package, and transport lipids and proteins. It includes the nuclear envelope, lysosomes, and vesicles, which we've already mentioned, and the endoplasmic reticulum and Golgi apparatus, which we will cover shortly. Although not technically *within* the cell, we include the plasma membrane in the endomembrane system because, as you will see, it interacts with the other endomembranous organelles. The endomembrane system does not include the membranes of either mitochondria or chloroplasts.



Membrane and secretory proteins are synthesized in the rough endoplasmic reticulum (RER). The RER also sometimes modifies proteins. In this illustration, a (green) integral membrane protein in the ER is modified by attachment of a (purple) carbohydrate.

Vesicles with the integral protein bud from the ER and fuse with the cis face of the Golgi apparatus. As the protein passes along the Golgi's cisternae, it is further modified by the addition of more carbohydrates. After its synthesis is complete, it exits as integral membrane protein of the vesicle that bud from the Golgi's trans face and when the vesicle fuses with the cell membrane the protein becomes integral portion of that cell membrane. (credit: modification of work by Magnus Manske)

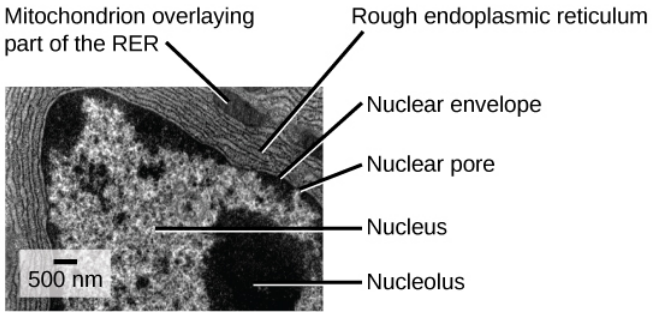
The Endoplasmic Reticulum

The *endoplasmic reticulum (ER)* (see figure above) is a series of interconnected membranous sacs and tubules that collectively modifies proteins and synthesizes lipids. However, these two functions are performed in separate areas of the ER: the rough ER and the smooth ER, respectively.

The hollow portion of the ER tubules is called the lumen or cisternal space. The membrane of the ER, which is a phospholipid bilayer embedded with proteins, is continuous with the nuclear envelope.

Rough ER

The *rough endoplasmic reticulum (RER)* is so named because the ribosomes attached to its cytoplasmic surface give it a studded appearance when viewed through an electron microscope (see figure below).



This transmission electron micrograph shows the rough endoplasmic reticulum and other organelles in a pancreatic cell. (credit: modification of work by Louisa Howard)

Ribosomes transfer their newly synthesized proteins into the lumen of the RER where they undergo structural modifications, such as folding or the acquisition of side chains. These modified proteins will be incorporated into cellular membranes—the membrane of the ER or those of other organelles—or secreted from the cell (such as protein hormones, enzymes). The RER also makes phospholipids for cellular membranes.

If the phospholipids or modified proteins are not destined to stay in the RER, they will reach their destinations via transport vesicles that bud from the RER's membrane.

Since the RER is engaged in modifying proteins (such as enzymes, for example) that will be secreted from the cell, you would be correct in assuming that the RER is abundant in cells that secrete proteins. This is the case with cells of the liver, for example.

Smooth ER

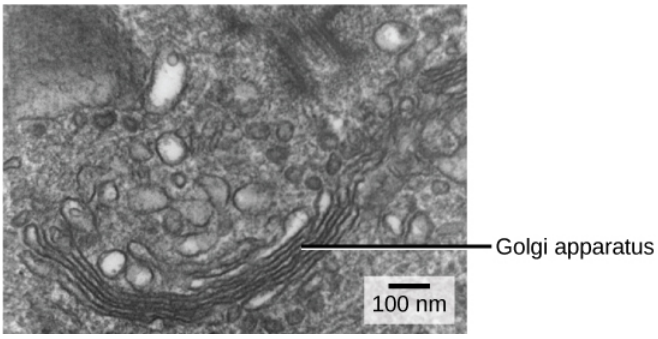
The *smooth endoplasmic reticulum (SER)* is continuous with the RER but has few or no ribosomes on its cytoplasmic surface. Functions of the SER include synthesis of carbohydrates, lipids, and steroid hormones; detoxification of medications and poisons; and storage of calcium ions.

In muscle cells, a specialized SER called the sarcoplasmic reticulum is responsible for storage of the calcium ions that are needed to trigger the coordinated contractions of the muscle cells.

The Golgi Apparatus

We have already mentioned that vesicles can bud from the ER and transport their contents elsewhere, but where do the vesicles go? Before reaching their final destination, the lipids or proteins within the transport vesicles still need to be sorted, packaged, and

tagged so that they wind up in the right place. Sorting, tagging, packaging, and distribution of lipids and proteins takes place in the *Golgi apparatus* (also called the Golgi body), a series of flattened membranes (see figure below).



The Golgi apparatus in this white blood cell is visible as a stack of semicircular, flattened rings in the lower portion of the image. Several vesicles can be seen near the Golgi apparatus. (credit: modification of work by Louisa Howard)

The receiving side of the Golgi apparatus is called the *cis* face. The opposite side is called the *trans* face. The transport vesicles that formed from the ER travel to the *cis* face, fuse with it, and empty their contents into the lumen of the Golgi apparatus. As the proteins and lipids travel through the Golgi, they undergo further modifications that allow them to be sorted. The most frequent modification is the addition of short chains of sugar molecules. These newly modified proteins and lipids are then tagged with phosphate groups or other small molecules so that they can be routed to their proper destinations.

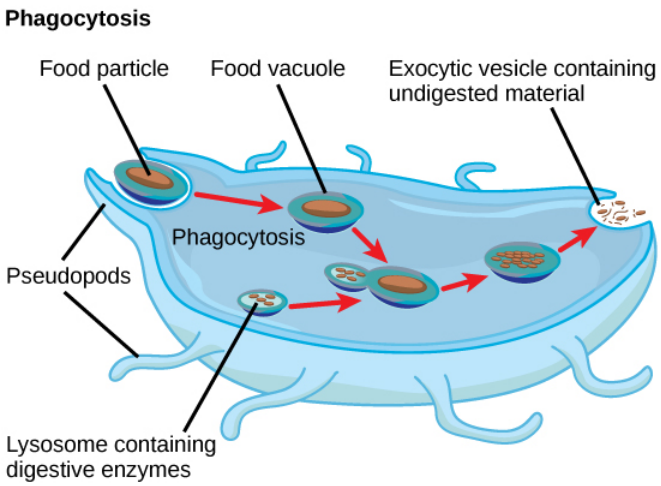
Finally, the modified and tagged proteins are packaged into secretory vesicles that bud from the *trans* face of the Golgi. While some of these vesicles deposit their contents into other parts of the cell where they will be used, other secretory vesicles fuse with the plasma membrane and release their contents outside the cell.

In another example of form following function, cells that engage in a great deal of secretory activity (such as cells of the salivary glands that secrete digestive enzymes or cells of the immune system that secrete antibodies) have an abundance of Golgi.

In plant cells, the Golgi apparatus has the additional role of synthesizing polysaccharides, some of which are incorporated into the cell wall and some of which are used in other parts of the cell.

Lysosomes

In addition to their role as the digestive component and organelle-recycling facility of animal cells, lysosomes are considered to be parts of the endomembrane system. Lysosomes also use their hydrolytic enzymes to destroy pathogens (disease-causing organisms) that might enter the cell. A good example of this occurs in a group of white blood cells called macrophages, which are part of your body's immune system. In a process known as phagocytosis or endocytosis, a section of the plasma membrane of the macrophage invaginates (folds in) and engulfs a pathogen. The invaginated section, with the pathogen inside, then pinches itself off from the plasma membrane and becomes a vesicle. The vesicle fuses with a lysosome. The lysosome's hydrolytic enzymes then destroy the pathogen (figure below).



A macrophage has engulfed (phagocytized) a potentially pathogenic bacterium and then fuses with a lysosomes within the cell to destroy the pathogen. Other organelles are present in the cell but for simplicity are not shown.

Summary of Endomembranes

The endomembrane system includes the nuclear envelope, lysosomes, vesicles, the ER, and Golgi apparatus, as well as the plasma membrane. These cellular components work together to modify, package, tag, and transport proteins and lipids that form the membranes.

The RER modifies proteins and synthesizes phospholipids used in cell membranes. The SER synthesizes carbohydrates, lipids, and steroid hormones; engages in the detoxification of medications and poisons; and stores calcium ions. Sorting, tagging, packaging, and distribution of lipids and proteins take place in the Golgi apparatus. Lysosomes are created by the budding of the membranes of the RER and Golgi. Lysosomes digest macromolecules, recycle worn-out organelles, and destroy pathogens.

Possible NB Discussion Point

If you were given a lab and the authority to pursue research in one of the organelles you just learned about, which one would you want to conduct research on? Why? What specific aspect of the organelle would you want to learn more about?

The Cytoskeleton

The cytoskeleton is a network of different protein fibers that provides many functions: it maintains or changes the shape of the cell; it secures some organelles in specific positions; it enables movement of cytoplasm and vesicles within the cell; and it enables the cell to move in response to stimuli. There are three types of fibers within the cytoskeleton: microfilaments, intermediate filaments, and microtubules. Some of the cytoskeletal fibers work in conjunction with molecular motors which move along the fibers within the cell to carry out a diverse set of functions. There are two main families of cytoskeletally-associated *molecular motors: dyneines* and *kinesins*.

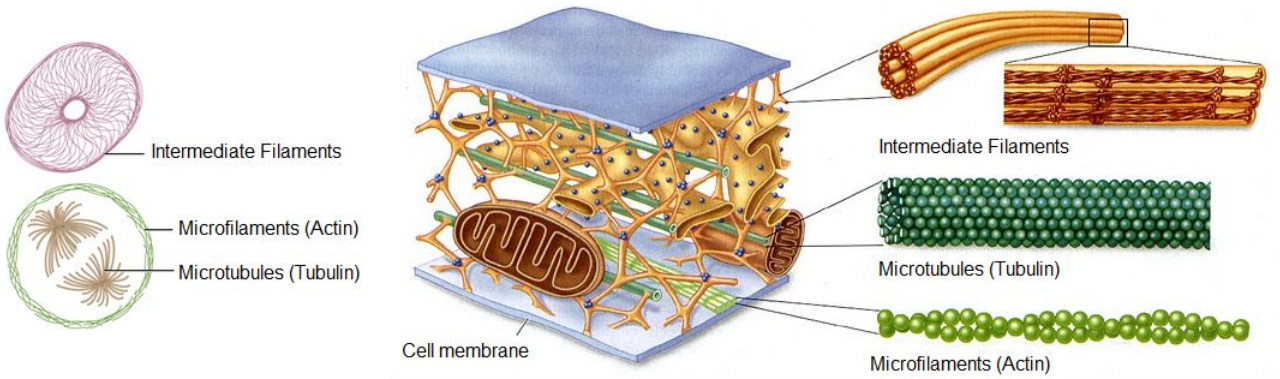


Figure 1. *Microfilaments thicken the cortex around the inner edge of a cell; like rubber bands, they resist tension. Microtubules are found in the interior of the cell where they maintain cell shape by resisting compressive forces. Intermediate filaments are found throughout the cell and hold organelles in place.*

Design challenge

Problem statement: Eukaryotic cells contain membrane-bound organelles that effectively separate materials, processes, and reactions from one another and from the cytoplasm. This in itself poses a problem for eukaryotes.

How can the cell purposely move and control the location of materials between these organelles? More specifically, how can a eukaryotic cell transport compounds from their place of origin (in most cases the cyotoplasm) to where they are needed (perhaps the nucleus, the mitochondria, or the cell surface)?

One possible solution is for the cell to create a network that can connect all the different parts of the cell together. This network could be used not only as a scaffold to hold components in place but also as a reference for direction. For example, we can use a map to determine the direction we need to travel and roads to connect and travel from home to campus. Likewise, an interconnecting network inside the cell can be used to direct and move compounds from one location to a final destination. Some of the required characteristics of this network are listed below. Can you add to this list?

Intracellular network

- The network needs to be extensive, and connect every area of the cell.
- The network needs to be flexible, able to change and adapt as the cell grows larger, divides into two cells, or physically moves from one environment to another.
- The network needs to be strong, able to hold up to mechanical pressure from inside the cell or from outside of the cell.
- The network needs to be composed of different fibers and each of these fibers needs to be for a specific connection in the cell. For example, certain fibers might be involved in holding organelles in place, and other fibers would be involved in connecting two different organelles.
- The fibers need to have directionality (or polarity), meaning they need to have a defined starting point and a defined end to help direct movement from one location to another.
- The fibers need to work with proteins that can convert chemical energy into kinetic energy, to actively transport compounds along the fibers.

Microfilaments

Actin

Microfilaments are cytoskeleton fibers composed of *actin* subunits. Actin is one of the most abundant proteins in eukaryotic cells and comprises 20% of total cellular protein by weight in muscle cells. The actin amino acid sequence is highly conserved in eukaryotic cells, meaning that the protein amino acid sequence, and therefore its final 3-D shape, has changed little over the course of evolution, maintaining more than 80% similarity between algae and humans.

Actin can be present as either a free monomer called G-actin (globular) or as part of a polymer microfilament called F-actin ("F" for filamentous). Actin must be bound to ATP in order to assemble into its filamentous form and maintain the structural integrity of the filament. The actin filament itself has structural polarity. This term "polarity", in reference to a cytoskeleton filament, does not mean what it did when we discussed polar functional groups earlier in this course. Polarity here refers to the fact that there are two distinct ends to the filament. These ends are called the "(-)" end and the "(+)" end. At the "(+)" end, actin subunits are adding onto the elongating filament and at the "(-)" end, actin subunits are disassembling or falling off of the filament. This process of assembly and disassembly is controlled by the ATP to ADP ratio in the cytoplasm.

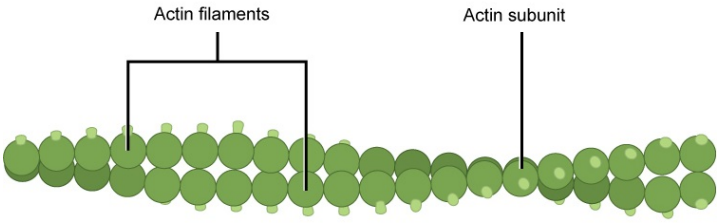


Figure 2. *Microfilaments are the narrowest of the three cytoskeleton fibers, with a diameter of about seven nm. Microfilaments are composed of actin subunits which form into two intertwined strands.*

Actin participates in many cellular processes, including muscle contraction, cell motility, cytokinesis during cell division, vesicle and organelle movement, and the maintenance of cell shape. Actin filaments serve as a track for the movement of a family of motor proteins called *myosins* discussed in more detail in a section below.

Link to learning:

To see an example of a white blood cell in action, click here and watch a short time-lapse video of the cell capturing two bacteria. It engulfs one and then moves on to the other.

Animations on actin filaments and how they work

- Actin filament assembly
- Muscle movement and the role of actin
- Sliding movement of actin filaments

Intermediate filaments

Intermediate filaments are made of several strands of fibrous proteins that are wound together. These elements of the cytoskeleton get their name from the fact that their diameter, eight to ten nm, is between those of the smaller microfilaments and the larger microtubules. The intermediate filaments are the most diverse group of cytoskeletal elements. Several types of fibrous proteins are found in the intermediate filaments. You are probably most familiar with keratin, the fibrous protein that strengthens your hair, nails, and the epidermis of the skin.



Figure 3. *Intermediate filaments consist of several intertwined strands of fibrous proteins.*

Intermediate filaments have no role in cell movement. Their function is purely structural. They bear tension, thus maintaining the shape of the cell, and anchor the nucleus and other organelles in place. The figure above shows how intermediate filaments create a cable-like supportive scaffolding inside the cell.

Microtubules

Microtubules are the largest component of the cytoskeleton and are found throughout the cytoplasm. These polymers are made up of globular protein subunits called *α-tubulin* and *β-tubulin*. Microtubules are found not only in eukaryotic cells but in some bacteria as well.

Both the α-tubulin and β-tubulin subunits bind to GTP. When bound to GTP, the formation of the microtubule can begin, this is called the nucleation event. As more GTP tubulin dimers assemble onto the filament, GTP is slowly hydrolyzed by β-tubulin to form GDP. Tubulin bound to GDP is less structurally robust and can lead to disassembly of the microtubule.

Much like the actin filaments discussed above, microtubules also have a distinct polarity that is critical for their biological function. Tubulin polymerizes end to end, with the β-subunits of one tubulin dimer contacting the α-subunits of the next dimer. These differences lead to different subunits being exposed on the two ends of the filament. The ends are designated the "(−)" and "(+)" ends. Unlike actin filaments, microtubules can elongate at both the "(+)" and "(-)" ends, but elongation is significantly more rapid at the "(+)" end.

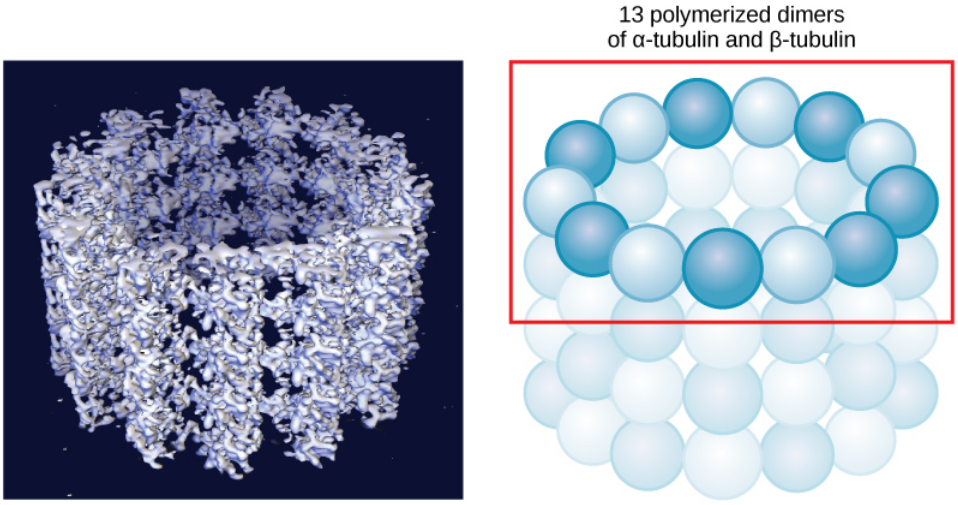


Figure 4. *Microtubules are hollow. Their walls consist of 13 polymerized dimers of α-tubulin and β-tubulin (right image). The left image shows the molecular structure of the tube.*

Microtubules help the cell resist compression, provide a track along which vesicles move through the cell, pull replicated chromosomes to opposite ends of a dividing cell, and are the structural elements of flagella, cilia, and centrioles (the latter are the two perpendicular bodies of the centrosome). In fact, in animal cells, the centrosome is the microtubule organizing center. In eukaryotic cells, flagella and cilia are quite different structurally from their counterparts in bacteria, discussed below.

Animations of the cytoskeleton

- Cytoskeleton
- Microtubules
- Another video on microtubules

Where did these fibers come from?

The cytoskeleton probably has its origins in bacterial and/or archaeal ancestry. There are ancient relatives to both actin and tubulin in bacterial systems. In bacteria, the MreB protein and the ParM protein are believed to be early ancestors to Actin. MreB functions in maintaining cell shape and ParM functions in plasmid (DNA) partitioning. The FtsZ protein in bacteria functions in cytokinesis, it is a GTPase, spontaneously forms filaments and is hypothesized to be an ancient form of tubulin. These findings support the hypothesis that the eukaryotic cytoskeleton has its origins in the bacterial world.

Flagella and cilia

Flagella (singular=flagellum) are long, hair-like structures that extend from the plasma membrane and are used to move an entire cell (for example, sperm, *Euglena*). When present, the cell has just one flagellum or a few flagella. *Cilia* are short, hair-like structures that are used to move entire cells (such as paramecia) or substances along the outer surface of the cell (for example, the cilia of cells lining the fallopian tubes that move the ovum toward the uterus, or cilia lining the cells of the respiratory tract that trap particulate matter and move it toward your nostrils.) When cilia are present, there can be many of them, extending along the entire surface of the plasma membrane.

Despite their differences in length and number, flagella and cilia share a common structural arrangement of microtubules called a "9+2 array." This is an appropriate name because a single flagellum or cilium is made of a ring of nine microtubule doublets, surrounding a single microtubule doublet in the center (Figure 5).

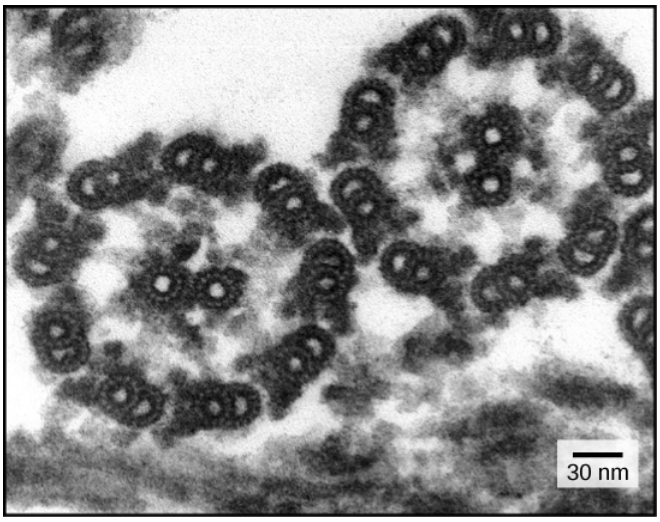


Figure 5. *This transmission electron micrograph of two flagella shows the "9+2 array" of microtubules: nine microtubule doublets surround a single microtubule doublet. (credit: modification of work by Dartmouth Electron Microscope Facility, Dartmouth College; scale-bar data from Matt Russell)*

For a video on flagellar and ciliar movement in *eukaryotes*, see the YouTube video: click here (you can skip the commericial).

Motor proteins

One function of the cytoskeleton is to move cellular components from one part of the cell to another. These cellular components are called "cargo" and are often stored within a vesicle for transport. You can think of the cytoskeleton as "railroad tracks" providing support and directionality inside of the cell.

Of course, if there are "railroad tracks" there needs to be an engine that can both move on the tracks and pull or push cargo along. In this case the engines are molecular motors that can move along the tracks in a specific direction. There are two families of *molecular motors* associated with the cytoskeleton; *dyneines* and *kinesins*. These motor proteins (train engines) and the cytoskeleton create a comprehensive network within the cell for moving vesicles (box cars) from one organelle to another or from one organelle to the cell surface.

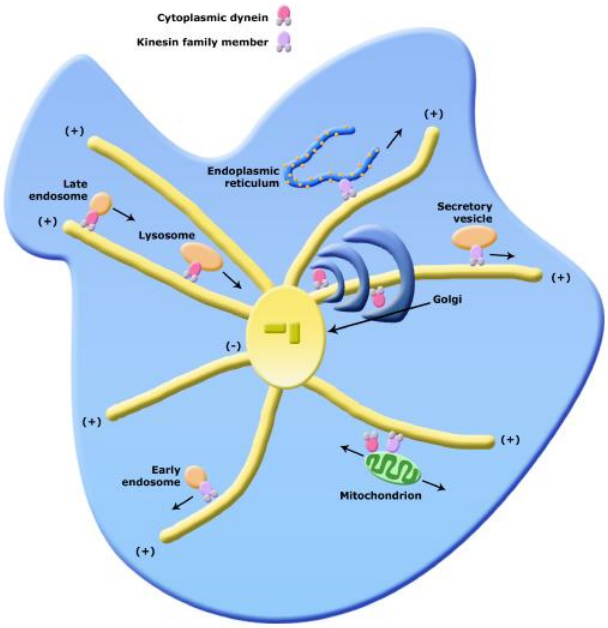


Figure 6. *Organelle transport via microtubules and kinesins and dynes. Note that the figure is conceptual and only intended to show directionality of movement of various organelles; it does not necessarily represent all of their forms faithfully.*

Cytoplasmic dyneins

Dynein is a protein complex that functions as a molecular motor. In cells, it converts the chemical energy from ATP hydrolysis into the mechanical energy of movement to 'walk' along the microtubule while carrying a vesicle. Dyneins bind to microtubules and move or "walk" from the plus "(+)" end of the cytoskeletal microtubule filament to the minus "(-)" end of the filament, which is usually oriented towards the cell center. Thus, they are often referred to as "minus end directed motors" and this vesicular transport is refereed to as *retrograde transport*. Cytoplasmic dynein moves processively along the microtubule, hydrolyzing ATP with each "step" it takes along the microtubule. During this process, one or the other of its "stalks" is always attached to the microtubule, allowing for the dynein motor (and its cargo) to "walk" a considerable distance along a microtubule without detaching.

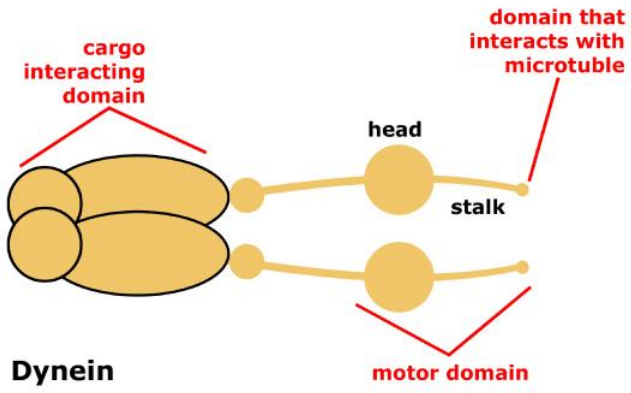


Figure 7. *Schematic of cytoplasmic dynein motor protein. Dyneins are protein complexes composed of many smaller polypeptide subunits. The overall structure of the dynien motors are relatively simple, consisting of two identical complexes each having a motor domain that interacts with the microtubule, a stalk, or stem region that connects the motor head to the cargo interacting domain.*

Cytoplamic dyneins are used in many different processes: they are involved in organelle movement such as the positioning of the Golgi complex and other organelles in the cell; they are used in the transport of cargo such as the movement of vesicles made by

the endoplasmic reticulum, endosomes, and lysosomes; and they are responsible for the movement of chromosomes during cell division. Axonemal dyneins are motor proteins used in the sliding of microtubules in the axonemes of cilia and flagella in eukaryotic cells.

Kinesins

Kinesins, like cytoplasmic dyneins are motor-protein complexes that "walk" along the microtubules and are involved in vesicle transport. Unlike cytoplasmic dyneins, the polarity of kinesin movement is from the "(-)" end of the microtubule to the "(+)" end with the hydrolysis of ATP. In most cells, this entails transporting cargo from the center of the cell towards the periphery (the opposite direction to dyneins). This form of transport is known as *anterograde* or *orthrograde transport*. Like cytoplasmic dyneins, kinesins are involved in a variety of cellular processes including vesicle movement and chromosome movement during cell division.

The structure of kinesins are similar to cytoplasmic dyneins and is diagrammed in Figure 8. Members of the kinesin superfamily vary in shape, but the overall structure is that of a heterotetramer whose motor subunits (heavy chains) form a protein dimer (molecule pair) that binds two light chains.

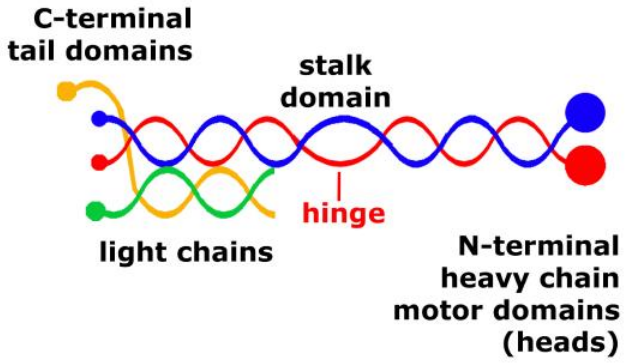


Figure 8. *Schematic of kinesin motor proteins. The heavy chains comprise a globular head (the motor domain) at the amino terminal end connected via a short, flexible neck linker to the stalk—a long, central α-helical coiled-coil domain—that ends in a carboxy terminal tail domain which associates with the light-chains. The stalks of two light chains intertwine to form a coiled-coil that directs dimerization of the two heavy chains. In most cases transported cargo binds to the kinesin light chains, but in some cases cargo binds to the C-terminal domains of the heavy chains.*

Possible NB Discussion Point

Explain the functions and importance of dyneins and kinesins as if you were trying to explain this to your 10 year old cousin. How are they similar and how are they different?

Animations of kinesin and dynein at work

- Animation of a cyplasmic dynein motor on a microtubule
- How dynein moves along a microtubule
- Mechanism of Kinesin moving on a microtubule
- Kinesin and dynein motors

How do the motors interact with cargo and the microtubules?

Cytoplasmic dyneins and kinesins interact with both cargo and microtubules in similar fashion. The light chains interact with receptors on the various cargo vesicles and the globular motor domains, specifically interact with the microtubules.

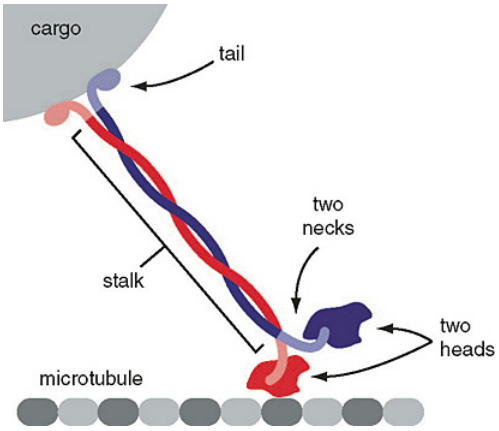


Figure 9. *Schematic of kinesin motor protein carrying a cargo vesicle along a microtubule filament.*

Reading_09 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_10

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_10

GC.28 Identify biological redox reactions involving common electron carriers.
GC.29 Given a redox reaction, identify the reducing agent, oxidizing agent, molecule that becomes oxidized, and the reduced species. Identify which species the electron(s) "starts" in, and to which species it "goes."
GC.30 Write a composite chemical equation when given two redox half-reactions.
GC.31 Calculate the ΔE_0' for a given redox reaction using the equation $\Delta E_0' = E_0'_{(oxidant)} - E_0'_{(reductant)}$
GC.44 Define and correctly use each variable and its role in the equation: $\Delta G_0' = -nF\Delta E_0'$.
GC.45 Convert between $\Delta G_0'$ and $\Delta E_0'$ for a given redox reaction using the equation $\Delta G_0' = -nF\Delta E_0'$.
GC.43 Qualitatively relate the difference in redox potentials with a corresponding delta of Gibbs enthalpy (energy).
GC.26 Predict whether a directional transfer of electrons between two chemical species is endergonic or exergonic by applying the concept of redox potential to provided data.
ME.2 Identify NAD^+ from its molecular structure and identify the functional group involved in its function as an oxidizing or reducing agent.
ME.1 Tell an energy story for a redox reaction that utilizes the electron carrier NAD+/NADH as the second substrate in the simple, generic reaction scheme: AH + NAD^+ -> A + NADH.

Metabolism in General Biology

Cellular metabolism represents roughly 1/3 of the curriculum for our class. While this may seem like a lot, we cover very little of what a classic course in metabolism covers, and a minuscule fraction of the metabolism that occurs on the planet. What we do cover, however, is very important foundational knowledge. You will learn about common chemical reactions that are associated with the transformation of life's molecular building blocks and about different core modes of energy transfer that you will encounter often in biology. The energy story and the design challenge rubrics introduced earlier will become increasingly important in these next few modules and beyond.

What have we learned? How will it relate to metabolism?

1. We have focused on the identification and chemical properties of common biological functional groups. As we dive into metabolism, this will help you be familiar with and sometimes even predict the chemical nature/reactivity of compounds you have never seen before.
2. We have practiced recognizing and classifying molecules into four major functional groups. This will help you as we discuss how to build and break down these molecules.
3. We have learned some basic thermodynamics. This gives us a common set of concepts with which to discuss whether a biochemical reaction or process is likely to occur, and if so in which direction and how fast. This will be critical as we consider some key reactions that take place in metabolism.
4. We have learned and practiced the energy story rubric. This will allow us to study new biochemical reactions and to discuss them with a common, consistent language and approach which also reinforces the lessons we learned about thermodynamics.

An overview of this section

- We will introduce an important concept called **reduction potential** and you will be given the opportunity to use a redox tower. There is also a discussion on redox chemistry in your discussion manual. Use both resources.
- We will introduce two major players in metabolism, ATP and NADH. We expect you to recognize their structures if shown on an exam.
- We will cover the metabolic pathway glycolysis. Keep in mind that we want you to look at any reaction and tell us an energy story of that reaction. You should not try to memorize these pathways (though it will help to remember some big picture things - we will stress these). Often we will give you the pathway as a figure on the exams. Glycolysis produces 2 ATP via a process called substrate level phosphorylation, 2 NADH and 2 pyruvate compounds.

- We will use the reactions of the TCA cycle to create multiple examples of energy stories. The TCA cycle will also produce more ATP, NADH and oxidize glucose into CO_2.
- We will look at an alternative pathway to that of the TCA cycle, fermentation. Here, for the first time, we will see NADH used as a reactant in a metabolic reaction.
- We will follow NADH to the end of its journey, as it donates its electrons to the electron transport chain (ETC). In this module, you will need to use a redox tower. The ETC produces a proton gradient. No ATP is directly generated in this process. However, the proton gradient is then used by the cell to run an enzyme called ATP synthase, which catalyzes the reaction ADP + Pi --> ATP. This method of ATP production (called oxidative phosphorylation) results in more ATP being produced than through substrate level phosphorylation.
- And finally, we will go through the process of photosynthesis.

Reduction-Oxidation Reactions (Electron Transfer)



Most of the **reduction/oxidation reactions (redox)** that we discuss here occur in metabolic pathways (connected sets of biochemical reactions) and can be described as electron transfers between two molecules. These electron transfers play two fundamental roles. First, they are involved in the building, breakdown and/or conversions of molecules into chemically different forms. Second, in all known forms of life these electron transfer reactions also play critical roles in energy transfer, either between a cell and its environment or within the cell itself. It is therefore important to develop, at the very least, an intuitive understanding and appreciation for redox/electron transfer reactions in biology.

Since redox reactions are generally important, most students of biology will also study this chemical process in chemistry courses. Unfortunately, biologists and chemists tend to approach teaching this common foundational topic in different ways (even if they're both teaching the same thing). That can be confusing for new students.

Chemists often introduce the concepts of oxidation and reduction using the concept of oxidation states. See this link for more information: <https://chem.libretexts.org/Bookshel...ation_Numbers)>. Chemists usually ask students to apply a set of rules (see link) to determine the oxidation states of individual atoms in the molecules involved in a chemical reaction. The chemistry formalism defines oxidation as an increase in oxidation state and reduction as a decrease in oxidation state.

However, biologists don't typically teach the topic in this way. Why? It's probably because most of the redox reactions encountered in biology involve a change in oxidation state that comes about through a transfer of electron(s). Consequently biologists, typically define reduction as a gain of electrons and oxidation as a loss of electrons. We note that the biological electron-exchange view of redox reactions is entirely consistent with the more general definition associated strictly with changes in oxidation states. The electron-exchange model does not, however, explain redox reactions that do not involve a transfer of electrons, which sometimes occur in the context of a chemistry class. The biologist's view of redox chemistry has the advantage (in the context of biology) of being relatively easy to create a mental picture for. There are no lists of rules to remember or much inspection of molecular structure involved in developing at least a basic conceptual picture of the topic. We simply imagine an exchange between two parties - one molecule handing off one or more electrons to a partner who accepts them.

Since this is a biology reading for a biology class we approach redox from the "gain/loss of electrons" conceptualization. If you have already taken a chemistry class and this topic seems to be presented a little different in your biology course, remember that at its core, you are learning the same thing. Biologists just adapted what you learned in chemistry to make more intuitive sense in the context of biology. If you haven't learned about redox, yet don't worry. If you can understand what we are trying to do here, when you cover this concept in chemistry class you will be a few steps ahead. You will just need to work to generalize your thinking a little bit.

Let's start with some generic reactions

Transferring electrons between two compounds results in one of these compounds losing an electron and one compound gaining an electron. For example, look at the figure below. If we use the energy story rubric to look at the overall reaction, we can compare the before and after characteristics of the reactants and products. What happens to the matter (stuff) before and after the reaction? **Compound A** starts as neutral and becomes positively charged. **Compound B** starts as neutral and becomes negatively charged.

Because electrons are negatively charged, we can explain this reaction with the movement of an electron from **Compound A** to **B**. That is consistent with the changes in charge. **Compound A** loses an electron (becoming positively charged), and we say that A has become oxidized. For biologists, **oxidation** is associated with the loss of electron(s). **B** gains the electron (becoming negatively charged), and we say that **B** has become reduced. **Reduction** is associated with the gain of electrons. We also know, since a reaction occurred (something happened), that energy must have been transferred and/or reorganized in this process and we'll consider this shortly.

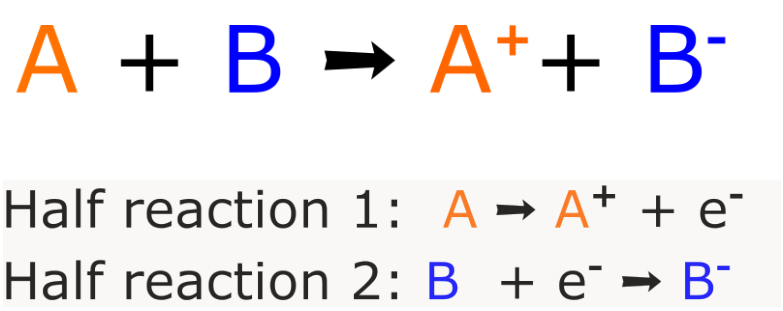


Figure 1. *Generic redox reaction with half-reactions Attribution: Mary O. Aina*

To reiterate: When an electron(s) is lost, or a molecule is **oxidized**, the electron(s) must then pass to another molecule. We say that the molecule gaining the electron becomes **reduced**. Together these paired electron gain-loss reactions are known as an **oxidation-reduction reaction** (also called a redox reaction).

This idea of paired half-reactions is critical to the biological concept of redox. Electrons don't drop out of the universe for "free" to reduce a molecule nor do they jump off a molecule into the ether. Donated electrons MUST come from a donor molecule and be transferred to some other acceptor molecule. For example, in the figure above the electron the reduces molecule B in half-reaction 2 must come from a donor - it just doesn't appear from nowhere! Likewise, the electron that leaves A in half-reaction 1 above must "land" on another molecule - it doesn't just disappear from the universe.

Therefore, **oxidation and reduction reactions must ALWAYS be paired.** We'll examine this idea in more detail below when we discuss the idea of "half-reactions".

- **A tip to help you remember:** The mnemonic LEO says GER (**L**ose **E**lectrons = **O**xidation and **G**ain **E**lectrons = **R**eduction) can help you remember the biological definitions of oxidation and reduction.



Figure 2. *A figure for the mnemonic "LEO the lion says GER." LEO: Loss of Electrons = Oxidation. GER: Gain of Electrons = Reduction. Attribution: Kamali Sripathi*

• **The vocabulary of redox can be confusing:** Students studying redox chemistry can often become confused by the vocabulary used to describe the reactions. Terms like oxidation/oxidant and reduction/reductant look and sound very similar but mean distinctly different things. An electron donor is also sometimes called a reductant because it is the compound that causes the reduction (gain of electrons) of another compound (the oxidant). In other words, the reductant is donating it's electrons to the oxidant which is gaining those electrons. Conversely, the electron acceptor is called the oxidant because it is the compound that is causing the oxidation (loss of electrons) of the other compound. Again, this simply means the oxidant is gaining electrons from the reductant who is donating those electrons. Confused yet?

Yet another way to think about definitions is to remember that describing a compound as reduced/oxidized is describing the **state** that *the compound itself* is in, whereas labeling a compound as a reductant/oxidant describes how the compound can act, to either reduce or oxidize *another compound*. Keep in mind that the term reductant is also synonymous with reducing agent and oxidant is also synonymous with oxidizing agent. The chemists who developed this vocabulary need to be brought up on charges of "willful thickheadedness" at science trial and then be forced to explain to the rest of us why they needed to be so deliberately obtuse.

The confusing language of redox: quick summary

1. A compound can be described as "reduced" - term used to describe the compound's **state.**

2. A compound can be a "reductant" - term used to describe a compound's **capability** (it can reduce something else). The synonymous term "reducing agent" can be used to describe the same capability (the term "agent" refers to the thing that can "do something" - in this case reduce another molecule).

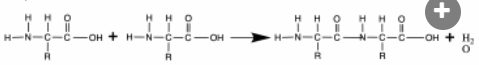
3. A compound can be an "oxidant" - term used to describe a compound's **capability** (it can oxidize something else). The synonymous term "oxidizing agent" can be used to describe the same capability (the term "agent" refers to the thing that can "do something" - in this case oxidize another molecule).

4. A compound can "become reduced" or "become oxidized"- term used to describe the **transition** to a new state.

Since all of these terms are used in biology, in General Biology we expect you to become familiar with this terminology. Try to learn it and use it as soon as possible - we will use the terms frequently and will not have the time to define terms each time.

Knowledge Check Quiz

A)



B) $C_{12}H_{11}O_{11}$ (Sucrose) + H_2O ⟶ $C_6H_{12}O_6$ (Glucose) + $C_6H_{12}O_6$ (Fructose)

C) NAD^+ + cytochrome c (red) ⟶ cytochrome c (ox) + NADH

Which of the following is a redox reaction?

- ○ B
- ○ A
- ○ C

Check

Reuse <> Embed

Knowledge Check Quiz

$$C4H6O4(succinate) + FAD+ \rightarrow C4H4O4(fumarate) + FADH2$$

A. C4H6O4 is in _____ form. C4H4O4 is in _____ form.
FADH2 is in its _____ form. FAD+ is in its _____ form.
B. FAD+ gets _____ and C4H6O4 gets _____.
C. The oxidizing agent is _____ and the reducing agent is _____

C4H6O4
reduced
reduced
reduced
FAD+
oxidized
oxidized
oxidized

Check

Reuse <> Embed

The Half Reaction

Here we introduce the concept of the half reaction. We can think each half reaction as a description of what happens to one of the two molecules (i.e. the donor or the acceptor) involved in a "full" redox reaction. A "full" redox reaction requires two half reactions. In the example below, half reaction #1 depicts the molecule **AH** losing two electrons and a proton and in the process becoming $\mathbf{A}^+$. This reaction depicts the oxidation of AH. Half reaction #2 depicts the molecule $\mathbf{B}^+$ gaining two electrons and a proton to become **BH**. This reaction depicts the reduction of $\mathbf{B}^+$. Each of these two half reactions is conceptual and neither can happen on its own. The electrons lost in half reaction #1 MUST go somewhere, they can't just disappear. Likewise, the electrons gained in half reaction #2 must come from something. They too just can not appear out of nowhere.

One can imagine that there might be different molecules that can serve as potential acceptors (the place for the electrons to go) for the electrons lost in half reaction #1. Likewise, there might be many potential reduced molecules that can serve as the electron donors (the source of electrons) for half reaction #2. In the example below, we show what happens (the reaction) when molecule **AH** is the donor of electrons for molecule $\mathbf{B}^+$. When we put the donor and acceptor half reactions together, we get a "full" redox reaction. In the figure below we call that reaction "Reaction #1". When this happens, we say that the two half reactions **coupled**. NOTE: In Chemistry class, because there are sometimes more than two reactions involved in a complete redox reaction, instead of the term "coupled reaction", you might encounter the more generic terms "simultaneous reactions" or "consecutive reactions". In all cases, these terms are meant to communicate that the reactions must happen together, either at the same time or in sequence.

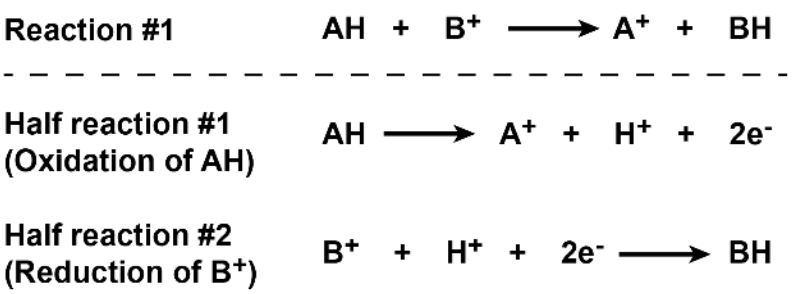


Figure 3. *Generic redox reaction where compound AH is being oxidized by compound B^+. Each half reaction represents a single species or compound to either lose or gain electrons (and a subsequent proton as shown in the figure above). In half reaction #1 AH loses a proton and 2 electrons: in the second half reaction, B^+ gains 2 electrons and a proton. In this example HA is oxidized to A^+ while B^+ is reduced to BH.*

Using this idea, we can theoretically couple and think about any two half reactions, one half reaction serving as the electron donor for the other half reaction that accepts the donated electrons. For instance, using the example above, we could consider coupling the reduction of $\mathbf{B}^+$ that happens in half reaction #2 with another half reaction describing the oxidation of the molecule NADH. In that case, the NADH would be the electron donor for $\mathbf{B}^+$. Likewise you could couple the oxidation of AH that happens in half reaction #1 with a half reaction describing the the reduction of hypothetical molecule Z^+. You can mix-and-match half reactions together as you please provided one half is describing the oxidation of a compound (it's donating electrons) and the reduction of another compound (it's accepting the donated electrons).

- **A note on how we write full reactions versus half reactions:** In the example above, when we write Reaction #1 as an equation, the 2 electrons - *in Chemistry class, these electrons might be referred to as "intermediates" because they do not appear in the overall redox reaction* - and the $\mathbf{H}^+$ that are explicitly described in the underlying half reactions, are not explicitly included in the text of the full reaction. In the reaction above you must infer that an exchange of electrons happens. This can be observed by trying to balance charges between each reactant and its corresponding product. Reactant **AH** becomes product $\mathbf{A}^+$. In this case, you can infer that some movement of electrons must have taken place. To balance the charges on this compound (make the sum of charges on each side of the equation equal) you need to add 2 electrons to the right side of the equation, one to account for the "+" charge on $\mathbf{A}^+$ and a second to go with the H^+ that was also lost. The other reactant $\mathbf{B}^+$ is converted to **BH**.

It must therefore gain 2 electrons to balance charges, one for $\mathbf{B}^+$ and a second for the additional H^+ that was added. Together this information leads you to conclude that the most likely thing to have happened is that two electrons were exchanged between **AH** and $\mathbf{B}^+$.

- This will also be the case for most redox reactions in biology. Fortunately, in most cases, either the context of the reaction, the presence of chemical groups often engaged in redox (e.g. metal ions), or the presence of commonly used electron carriers (e.g. NAD^+/NADH, $FAD^+/FADH_2$, ferredoxin, etc.) will alert you that the reaction is of class "redox". You will be expected to learn to recognize some of these common molecules.

Reduction Potential

By convention, we quantitatively characterize redox reactions using an measure called **reduction potentials.** The reduction potential attempts to quantitatively describe the “ability” of a compound or molecule to gain or lose electrons. The specific value of the reduction potential is determined experimentally, but for the purpose of this course we assume that the reader will accept that the values in provided tables are reasonably correct. We can anthropomorphize the reduction potential by saying that it is related to the strength with which a compound can “attract” or “pull” or “capture” electrons. Not surprisingly this is is related to but not identical to electronegativity.

What is this intrinsic property to attract electrons?

Different compounds, based on their structure and atomic composition have intrinsic and distinct "attractions" for electrons. This quality leads each molecule to have its own standard **reduction potential** or $\mathbf{E^{0'}}$. The reduction potential is a relative quantity (relative to some “**standard**” reaction). If a test compound has a stronger "attraction" to electrons than the standard (if the two competed, the test compound would "take" electrons from the standard compound), we say that the test compound has a positive reduction potential. The magnitude of the difference in E_0’ between any two compounds (including the standard) is proportional to how much more or less the compounds "want" electrons. The relative strength of the reduction potential is measured and reported in units of **Volts (V)** (sometimes written as electron volts or eV) or **milliVolts (mV)**. The reference compound in most redox towers is H_2.

Possible NB Discussion Point

Rephrase for yourself: How do you describe or think about the difference between the concept of electronegativity and red/ox potential?

Redox student misconception alert: The standard redox potential for a compound reports how strongly a substance wants to hold onto an electron in comparison to hydrogen. Since both redox potential and electronegativity are both discussed as measurements for how strongly something "wants" an electron, they are sometimes conflated or confused for one another. However, they are not the same. While the electronegativity of atoms in a molecule may influence its redox potential, it is not the only factor that does. You don't need to worry about how this works. For now, try to keep them as different and distinct ideas in your mind. The physical relationship between these two concepts is well beyond the scope of this general biology class.

REQUEST

As you go through the Redox section of the reading please use NB to also comment on parts of the reading that are hard to understand, confusing, and places that you think are well explained and clear. If you have suggested edits please let us know. Use the following emoji to tag these comments:



Table of Standard Reduction Potentials (aka. The Redox Tower)

All kinds of compounds can take part in redox reactions. Scientists have developed a graphical tool, a table of standard reduction potentials (aka. a redox tower), to tabulate redox half reactions based on their $E^{0'}$ values. This tool can help predict the direction of electron flow between potential electron donors and acceptors and how much free energy change might be expected to change in a specific reaction. By convention, all half reactions in the table are written in the direction of reduction for each compound listed.

In the biology context, the redox tower usually ranks a variety of common compounds (their half reactions) from most negative $E^{0'}$ (compounds that readily get rid of electrons), to the most positive $E^{0'}$ (compounds most likely to accept electrons). The tower below lists the number of electrons that are transferred in each reaction. For example, the reduction of NAD^+ to NADH involves two electrons, written in the table as $2e^-$.

---

oxidized form	reduced form	n (electrons)	$E^{0'}$ (volts)
PS1* (ox)	PS1* (red)	-	-1.20
Acetate + CO_2	pyruvate	2	-0.7
ferredoxin (ox) *version 1*	ferredoxin (red) *version 1*	1	-0.7
succinate + CO_2 + $2H^+$	a-ketoglutarate + H_2O	2	-0.67
PSII* (ox)	PSII* (red)	-	-0.67
P840* (ox)	PS840* (red)	-	-0.67
acetate	acetaldehyde	2	-0.6
glycerate-3-P	glyceraldehyde-3-P + H_2O	2	-0.55
O_2	O_2^-	1	-0.45
ferredoxin (ox) *version 2*	ferredoxin (red) *version 2*	1	-0.43
CO_2	glucose	24	-0.43
CO_2	formate	2	-0.42
2H+	H_2	2	-0.42 (at $[H^+] = 10^{-7}$; pH=7) Note: at $[H^+] = 1$; pH=0 the $E^{0'}$ for hydrogen is ZERO. You will see this in chemistry class.
α-ketoglutarate + CO_2 + $2H^+$	isocitrate	2	-0.38
acetoacetate	b-hydroxybutyrate	2	-0.35
Cystine	cysteine	2	-0.34
Pyruvate + CO_2	malate	2	-0.33
$NAD^+ + 2H^+$	NADH + H^+	2	-0.32
$NADP^+ + 2H^+$	NADPH + H^+	2	-0.32
Complex I FMN (enzyme bound)	$FMNH_2$	2	-0.3
Lipoic acid, (ox)	Lipoic acid, (red)	2	-0.29
1,3 bisphosphoglycerate +	glyceraldehyde-3-P	2	-0.29

oxidized form	reduced form	n (electrons)	$E^{0'}$ (volts)
$2H^+$	+ P_i		
Glutathione, (ox)	Glutathione, (red)	2	-0.23
FAD^+ (free) + $2H^+$	$FADH_2$	2	-0.22
Acetaldehyde + $2H^+$	ethanol	2	-0.2
Pyruvate + $2H^+$	lactate	2	-0.19
Oxalacetate + $2H^+$	malate	2	-0.17
α-ketoglutarate + NH_4+	glutamate	2	-0.14
FAD^+ + $2H^+$ (bound)	$FADH_2$ (bound)	2	0.003-0.09
Methylene blue, (ox)	Methylene blue, (red)	2	0.01
Fumarate + $2H^+$	succinate	2	0.03
CoQ (Ubiquinone - UQ + H^+)	UQH.	1	0.031
UQ + $2H^+$	UQH_2	2	0.06
Dehydroascorbic acid	ascorbic acid	2	0.06
Plastoquinone; (ox)	Plastoquinone; (red)	-	0.08
Ubiquinone; (ox)	Ubiquinone; (red)	2	0.1
Complex III Cytochrome b_2; Fe^{3+}	Cytochrome b_2; Fe^{2+}	1	0.12
Fe^{3+} (pH = 7)	Fe^{2+} (pH = 7)	1	0.20
Complex III Cytochrome c_1; Fe^{3+}	Cytochrome c_1; Fe^{2+}	1	0.22
Cytochrome c; Fe^{3+}	Cytochrome c; Fe^{2+}	1	0.25
Complex IV Cytochrome a; Fe^{3+}	Cytochrome a; Fe^{2+}	1	0.29
1/2 O_2 + H_2O	H_2O_2	2	0.3
$P840^{GS}$ (ox)	$PS840^{GS}$ (red)	-	0.33
Complex IV Cytochrome a_3; Fe^{3+}	Cytochrome a_3; Fe^{2+}	1	0.35

oxidized form	reduced form	n (electrons)	$E^{0'}$ (volts)
Ferricyanide	ferrocyanide	2	0.36
Cytochrome f; Fe^{3+}	Cytochrome f; Fe^{2+}	1	0.37
PSI^{GS} (ox)	PSI^{GS} (red)	.	0.37
Nitrate	nitrite	1	0.42
Fe^{3+} (pH = 2)	Fe^{2+} (pH = 2)	1	0.77
1/2 O_2 + $2H^+$	H_2O	2	0.816
$PSII^{GS}$ (ox)	$PSII^{GS}$ (red)	-	1.10

* Excited State, after absorbing a photon of light
GS Ground State, state prior to absorbing a photon of light
PS1: Oxygenic photosystem I
P840: Bacterial reaction center containing bacteriochlorophyll (anoxygenic)
PSII: Oxygenic photosystem II

Table 1. *Common redox tower used in Bis2A. By convention the tower half reactions are written with the oxidized form of the compound on the left and the reduced form on the right. Compounds that make good electron donors have highly negative reduction potentials. Compounds such as Glucose and Hydrogen gas are excellent electron donors. By contrast compounds that make excellent electron acceptors, such as Oxygen and Nitrite are excellent electron acceptors.*

Video on redox tower

For a short video on how to use the electron tower in redox problems click here or below. This video was made by Dr. Easlon for Bis2A students. (This is quite informative.)



What is the relationship between $\Delta E^{0'}$ and $\Delta G^{0'}$?

How do we know if any given redox reaction (the specific combination of two half reactions) is energetically spontaneous or not (exergonic or endergonic)? Moreover, how can we determine what the quantitative change in free energy is for a specific redox

reaction? The answer lies in the difference in the reduction potentials of the two compounds. The difference in the reduction potential for the reaction ($\Delta E^{0'}$), can be calculated by taking the difference between the $E^{0'}$ for the **oxidant** (the compound getting the electrons and causing the oxidation of the other compound) and the **reductant** (the compound losing the electrons). In our generic example below, AH is the reductant and B^+ is the oxidant. Electrons are moving from AH to B^+. Using the $E^{0'}$ of -0.32 for the reductant and +0.82 for the oxidant the total change in $E^{0'}$ or $\Delta E^{0'}$ is 1.14 eV.

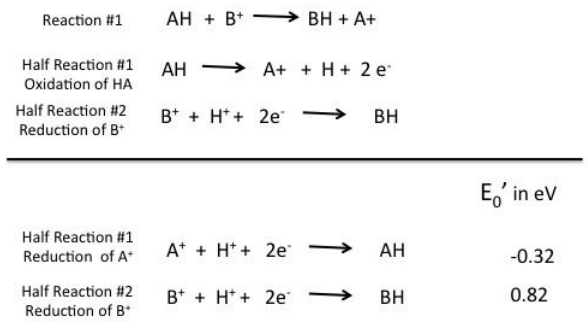


Figure 4. *Generic red/ox reaction with half reactions written with reduction potential ($E^{0'}$) of the two half reactions indicated.*

$\Delta E^{0'}$ between oxidant and reductant can tell us about the spontaneity of a proposed electron transfer. Intuitively, if electrons are proposed to move from a compound that "wants" electrons less to a compound that "wants" electrons more (i.e. a move from a compound with a lower $E^{0'}$ to a compound with a higher $E^{0'}$, the reaction will be energetically spontaneous). If the electrons are proposed to move from a compound that "wants" electrons more to a compound that "wants" electrons less (i.e. a move from a compound with a higher $E^{0'}$ to a compound with a lower $E^{0'}$, the reaction will be energetically non-spontaneous). Because of the way biological/biochemical redox tables are ordered (small $E^{0'}$ on top and larger $E^{0'}$ on the bottom) transfers of electrons from donors higher on the table to acceptors lower on the table will be spontaneous.

It is also possible to quantify the amount of free energy change associated with a specific redox reaction. The relationship is given by the following equation:

$$\Delta G^{\circ'} = -n\mathrm{F}\Delta \mathrm{E}^{\circ'}$$

Figure 5. *The relationship between free energy of a redox reaction to the difference in reduction potential between the reduced products of the reaction and oxidized reactant. Attribution: Marc T. Facciotti*

Where:

- n is the number of moles of electrons transferred
- F is the Faraday constant of 96.485 kJ/V. Sometimes it is given in units of kcal/V which is 23.062 kcal/V, which is the amount of energy (in kJ or kcal) released when one mole of electrons passes through a potential drop of 1 volt

Note that the signs of $\Delta E^{0'}$ and $\Delta G^{0'}$ are opposite one another. When $\Delta E^{0'}$ is positive, $\Delta G^{0'}$ will be negative. When $\Delta E^{0'}$ is negative, $\Delta G^{0'}$ will be positive.

Alternative Views of Some Common Confusing Issues in Basic Redox Chemistry for Biology

This reading tries to break down some of the more challenging topics that students encounter when studying redox chemistry in General Biology. This reading is not a substitute for your main reading but rather a complement to it that revisits some of the same topics through a different lens.

Finding $\Delta E_0'$

Students often struggle with finding the ΔE for a given redox reaction. One of the main barriers to developing this skill seems to be associated with developing a picture of the redox reaction itself. From the context of most biological redox reactions it is useful to imagine/picture a redox reaction as a simple exchange of electrons between two molecules, an electron donor and an electron acceptor that accepts electrons from the donor.

An analogy with kiwi fruit: To help build this mental picture we offer an analogy. Two people are standing next to one another. At the start, one person is holding a kiwi fruit in their hand and the second person's hands are empty. In this reaction, person 1 gives the kiwi to person 2. At the end of the reaction, person 2 is holding a kiwi and person 1 is not. We can write this exchange of kiwi fruit in the form of a chemical reaction:

person 1(kiwi) + person 2() <-> person 1() + person 2(kiwi).

start/initial state <-> *final/end state*

If we read this "reaction" from left to right, person 1 is a kiwi donor and person 2 is a kiwi acceptor. We can extend this analogy a little by proposing that person 1 and person 2 have different desire and ability to grab and hold kiwi fruit - we'll call that property "kiwi-potential". We can then propose a situation where person 1 and person 2 compete for a kiwi. Let's propose that person 2 has a higher "kiwi-potential" than person 1 - that is, person 2 has a stronger desire and ability to grab and hold kiwi than person 1.

If we set up a competition where person 1 starts with the kiwi and person 2 competes for it, we should expect that after some time the kiwi will be exchanged to person 2 and stay there most often. At the end of the reaction the kiwi will be with person 2. Due to the difference in "kiwi-potential" between person 1 and person 2, we can say that the spontaneous direction of kiwi flow is from person 1 to person 2. If we ever observed the kiwi flow from person 2 to person 1 we could probably conclude that person 1 required some extra help/energy to make that happen - flow from person 2 to person 1 would be non-spontaneous.

Let's call the "kiwi-potential", Kp. In our analogy, $Kp_{person\ 1} < Kp_{person\ 2}$. We can calculate ΔKp, the difference in Kp between the two people, and that will tell us something about how likely we can expect to see kiwi exchange hands between these two people. The bigger the difference in Kp the more likely the kiwi will move from the person who has a lower Kp to the person who has the higher Kp.

By definition, to calculate ΔKp we obtain the solution to $\Delta Kp = Kp_{final/end} - Kp_{initial/start}$. Since the kiwi is with person 2 at the end of the reaction and it starts with person 1 at the beginning of the reaction we would calculate $\Delta Kp = Kp_{person\ 2} - Kp_{person\ 1}$.

Doing it with electrons instead of kiwi fruit: To find $\Delta E^{0'}$ for a redox reaction we can translate this analogy to the molecular space. Instead of people, we have two molecules. Instead of a kiwi, we have electrons. Different molecules have different inherent abilities to grab and hold electrons and this can be measured by the value E. If two molecules exchange one or more electrons we can imagine that electrons will flow spontaneously from a molecule with lower $E^{0'}$ to one with a higher $E^{0'}$. We can write a familiar reaction with those substitutions.

molecule 1(electron) + molecule 2() <-> person 1() + molecule 2(electron).

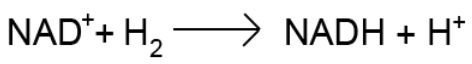
start/initial state <-> *final/end state*

To find $\Delta E^{0'}$, you solve for $\Delta E^{0'} = E^{0'}_{\text{-final/end}} - E^{0'}_{\text{-initial/start}}$. Alternatively, you can think of it as $\Delta E^{0'} = E^{0'}_{\text{-acceptor}} - E^{0'}_{\text{-donor}}$.

When evaluating a redox reaction for $E^{0'}$ you therefore need to:

1. First, find which of the reactants is the electron donor. The donor can also be associated with the initial state because it is the molecule that *initially* (before the start of the reaction) has the electron(s) to donate. This will always be one of the reactants and will be the molecule that gets oxidized (i.e. the molecule that loses electrons).
2. Second, find which of the reactants is the electron acceptor. This will also always be a reactant and will be the molecule that becomes reduced by the reaction (i.e. gains electrons). This molecule can be also associated with the final state since, in its reduced form, it is the molecule that has the electrons at the end of the reaction.

3. Third, calculate $\Delta E^{0'} = E^{0'}_{\text{-acceptor}} - E^{0'}_{\text{-donor}}$ or if you prefer, $\Delta E^{0'} = E^{0'}_{\text{-final/end}} - E^{0'}_{\text{-initial/start}}$.



In the example above, we can examine the reactants and determine that NAD^+ is the oxidized form of the electron carrier - it can, therefore, not be the donor. This means that H_2 must act as the electron donor in this reactant. During the reaction electrons flow onto NAD^+ from the donor H_2 creating the reduced product NADH and oxidized product H^+. To calculate $\Delta E^{0'}$ we say that at the start of the reaction the exchanged electrons are on the donor H_2. We say that at the end of the reaction the electrons are found on NADH. Calculating $\Delta E^{0'}$ requires us to evaluate the difference:

$E^{0'}_{\text{-acceptor}} - E^{0'}_{\text{-donor}}$

or equivalently,

$E^{0'}_{\text{-final/end}} - E^{0'}_{\text{-initial/start}}$.

Using a redox table to find $E^{0'}$ values for the start and end molecules shows us that NAD^+/NADH has an $E^{0'}$ of -0.30 while H^+/H_2 has an $E^{0'}$ of -0.42.

Therefore, $\Delta E^{0'} = (-0.30) - (-0.42) = 0.12$ V.

We can see intuitively that this reaction is spontaneous: electrons are flowing from a molecule that "wants" electrons less ($E^{0'}$ of $H^+/H_2 = -0.42$) to a molecule that wants them more ($E^{0'}$ of NAD^+/NADH = -0.30).

Reading Different Looking Redox Towers

Novice students of redox chemistry will all undoubtedly run across different ways of representing a redox tower. These different representations may look different but contain the same information. Without explanation, however, reading these tables - when they look different - can be confusing. We will compare and contrast different common forms of redox towers.

Redox Tower: Type 1

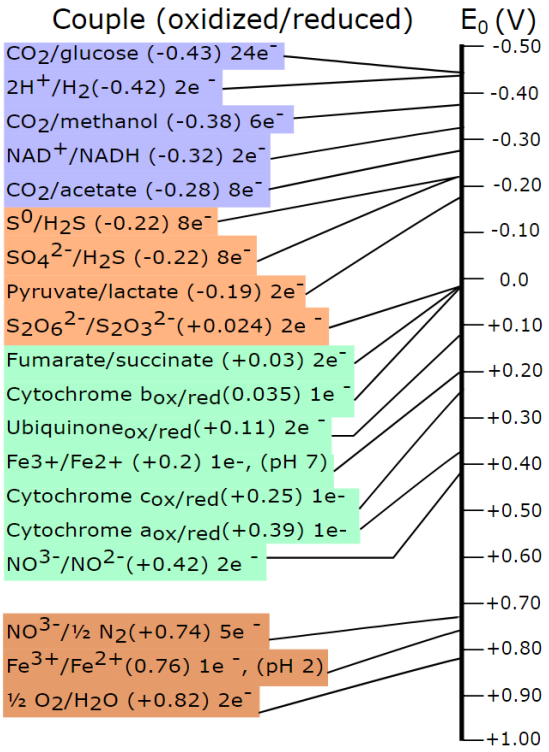


Figure 1. A *Generic redox tower with oxidized/reduced couple listed with its reduction potential ($E^{0'}$). Attribution: Caidon Iwuagwu*

In this type of redox tower, the oxidized and reduced forms of a molecule are separated by a slash. There is a line drawn from each half-reaction to its redox potential $E^{0'}$ reported on the vertical axis.

Redox Tower: Type 2

Electron Acceptor	Electron Donor	$E^{0'}$ (eV)
$CO_2 + 24e^- \rightarrow$	glucose	- 0.43
$2H^+ + 2e^- \rightarrow$	H_2	- 0.42
$CO_2 + 6e^- \rightarrow$	methanol	- 0.38
$NAD^+ + 2e^- \rightarrow$	NADH	- 0.32
$CO_2 + 8e^- \rightarrow$	acetate	- 0.28
$S_0 + 2e^- \rightarrow$	H_2S	- 0.28
$SO_4^{2-} + 8e^- \rightarrow$	H_2S	- 0.22
Pyruvate + $2e^- \rightarrow$	lactate	- 0.19
$S_4O_6^{2-} + 2e^- \rightarrow$	$S_2O_3^{2-}$	+ 0.024
Fumarate + $2e^- \rightarrow$	succinate	+ 0.03
Cytochrome b_{ox} + $1e^- \rightarrow$	Cytochrome b_{red}	+ 0.035
Ubiquinone$_{ox}$ + $2e^- \rightarrow$	Ubiquinone$_{red}$	+ 0.11
Fe^{3+} + 1e- $\rightarrow$ (pH 7)	Fe^{2+}	+ 0.2
Cytochrome c_{ox} + $1e^- \rightarrow$	Cytochrome c_{red}	+ 0.25
Cytochrome a_{ox} + $1e^- \rightarrow$	Cytochrome a_{red}	+ 0.39
NO_3^- + $2e^- \rightarrow$	NO_2^-	+ 0.42
NO_3^- + $5e^- \rightarrow$	1/2 N_2	+ 0.74
Fe^{3+} + $1e^- \rightarrow$ (pH 2)	Fe^{2+}	+ 0.77
1/2 O_2 + $2e^- \rightarrow$	H_2O	+ 0.82

In this type of redox tower, each row consists of a half-reaction. The oxidized form of a molecule is shown in the first column, the reduced form of the molecule is shown in the second column. Finally, the $E^{0'}$ value of the molecule is listed in the third column from the left. The number of electrons transferred to reduce the oxidized form of the molecule is shown in column 1. While the format of the table looks different from Type 1 tower, both contain the exact same information.

Redox Tower: Type 3

oxidized form	reduced form	n (electrons)	$E^{o'}$ (volts)
CO_2	glucose	24	-0.43
$2H^+$	H_2	2	-0.42 (at $[H^+] = 10^{-7}$; pH=7) Note: at $[H^+] = 1$; pH=0 the Eo' for hydrogen is ZERO. You will see this in chemistry class.
CO_2	methanol	6	-0.38
$NAD^+ + 2H^+$	$NADH + H^+$	2	-0.32
CO_2	acetate	8	-0.28
S^0	H_2S	2	-0.28
SO_4^{2-}	H_2S	8	-0.22
Pyruvate + $2H^+$	lactate	2	-0.19
$S_4O_6^{2-}$	$S_4O_6^{2-}$	2	0.024
Fumarate	succinate	2	0.03
Cytochrome b_{ox}	Cytochrome b_{red}	1	0.035
Ubiquinone; (ox)	Ubiquinone; (red)	2	0.1
Fe^{3+} (pH = 7)	Fe^{2+} (pH = 7)	1	0.20
Cytochrome c; Fe^{3+}	Cytochrome c; Fe^{2+}	1	0.25
Cytochrome a	Cytochrome a	1	0.39
Nitrate	nitrite	2	0.42
Nitrate	1/2 N_2	5	0.74
Fe^{3+} (pH = 2)	Fe^{2+} (pH = 2)	1	0.77
1/2 O_2 + $2H^+$	H_2O	2	0.816

In this redox tower, the oxidized form of a molecule is in the leftmost column, its reduced form is in the second column from the left, the number of electrons transferred is in the third column from the left, and the $E^{0'}$ is in the far right column.

Again, all of these towers contain the exact same information and are used in an identical manner.

Special note: If you have studied redox chemistry in a formal chemistry course, you might notice two key differences between the towers you use in a biology setting and those used by chemists.

1. In chemistry, the redox towers are flipped relative to those in biology: In chemistry, the molecules with the most positive $E^{0'}$ are listed starting at the top of the table and the compounds with the most negative $E^{0'}$ are listed at the bottom. In bioloigal redox tables molecules with the largest $E^{0'}$ are listed at the bottom while those with the smallest $E^{0'}$ are listed starting at the top. The biology

orientation has the advantage of making it easy to picture electrons spontaneously falling down the table from molecules that "want" the electrons less (lower $E^{0'}$) to molecules that "want" electrons more (higher $E^{0'}$).

2. In chemistry, the redox potential for hydrogen (H^+/H_2) is listed as 0. This is because (a) redox potentials for chemistry are measured under a set of non-biologically relevant standard conditions and (b) hydrogen is being used as the common standard redox potential against which all other redox potentials are measured. In biology, the redox potential for hydrogen (H^+/H_2) is listed as -0.42. This difference between the chemistry and biology tables comes about because the redox potential for (H^+/H_2) in biology is measured at a physiological pH of 7.0.

Familiarize yourself with how to read and interpret all three types of redox towers!

Chemistry and Biology Teach Redox Slightly Differently

For students who have been taught redox chemistry in a formal chemistry course, biological lessons in redox can sometimes seem like they're talking about something completely different. Not surprisingly, chemists tend to teach the most proper and universally applicable approach to evaluating redox reaction. This approach consists of using a set of rules to formally evaluate whether atoms in a molecule have undergone a change in oxidation state. Meanwhile, biologists tend to approach the discussion of redox reactions by thinking about electron transfers between molecules. It turns out that the approach biologists take is not as rigorous as how chemists approach redox reactions and can sometimes not identify bona fide redox reactions that wouldn't be missed using the chemist's approach. However, since the vast majority of biological redox reactions do involve a transfer of electron (and therefore change in redox states).

Let's look at a specific example to see the differences in approach.

The Chemistry Approach (oxidation numbers):

To evaluate/solve redox reactions in *CHEMISTRY*, we use the concept of oxidation states. The oxidation state of an element refers to how the electrons are shared between atoms in a chemical compound and they tell us about the movement of the electrons in the redox reaction. There are specific rules to assigning oxidation states, we will not be going over all of them since they will not be applicable to the redox reactions you will see in General Biology, but here are a few:

1. A single element has an oxidation state of 0
2. Fluorine ALWAYS has an oxidation state of -1
3. Hydrogen has an oxidation state of +1 with nonmetals and -1 with metals.

etc.

For more on calculating oxidation states see: <https://chem.libretexts.org/Bookshel...ation_Numbers)>

So, to find which elements are reduced/oxidized when given a redox reaction, you must track the change of the oxidation states between the reactants and the products. Here is an example:

In the unbalanced reaction $NO_3^- + FADH_2 \longrightarrow NO_2^- + FAD^+$

1. Using the rules, we observe that in NO_3^-, the oxidation number of Nitrogen is +5. In NO_2^-, the oxidation number of Nitrogen is +3. So because $+5 \longrightarrow +3$, N is reduced in this reaction.
2. We could conduct a similar calculation for key atoms on FAD^+ and $FADH_2$ to discover that $FADH_2$ is oxidized in the reaction.

The Biology/Biochemistry Approach (electron flow):

To evaluate/solve redox reactions in *biology/biochemistry*, we typically do not use or assign oxidation numbers to evaluate redox reactions. Rather we follow the exchange of electrons between molecules. Fortunately, biology tends to reuse a limited number of electron carriers and redox towers tell us which form of a compound is reduced or oxidized. As mentioned above, the basic approach to redox in biology is to define *oxidation as: the* loss of electrons. *Reduction* is defined as: *the gain of electrons.* Here is an example:

$NO_3^- + FADH_2 \longrightarrow NO_2^- + FAD^+$

Here we examine the reactants and immediately spot the common electron carrier $FADH_2$, the reduced form of the electron carrier. In the products we observe the oxidized form of the electron carrier FAD^+. We conclude that $FADH_2$ lost electrons (became oxidized) in the reaction. Since the electrons had to go somewhere they were likely accepted by NO_3^- which then became reduced to NO_2^-. In this case the biologist's model arrives at the same conclusion as the chemist's approach through a more intuitive approach that doesn't require memorizing numerous rules and how to apply them.

In our General Biology class, we take the biology/biochemistry approach to redox. You will not need to know how to calculate oxidation states in this course.

DISCLAIMER: DO NOT WORRY IF YOU HAVE NOT TAKEN CHEMISTRY YET. WE WILL NOT BE USING THE CHEMISTRY APPROACH WHEN IT COMES TO REDOX REACTIONS IN OUR CLASS. THE PURPOSE OF THIS IS JUST TO DISTINGUISH AND HOPEFULLY CLARIFY THE TWO APPROACHES FOR STUDENTS THAT MAY HAVE ALREADY TAKEN A CHEMISTRY COURSE!!

Reading_10 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_11

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_11

ME.16 Apply the concept of the “conservation of mass” to metabolism by describing the different forms mass takes as it enters and leaves the cell (e.g. input: reduced molecules like glucose, lipids, proteins, etc. & output: oxidized molecules like CO_2, H_2O etc.).

ME.15 Apply the concept of the "conservation of energy" to central metabolism. Follow energy from "sources" of electrons with relatively low reduction potentials to "sinks" with higher redox potentials, describe the major transfers of energy and how this energy is "stored" at each stage.

————-*Learning objectives above are for the full course, but particularly relevant to the discussion of metabolism and thus extend well beyond this lecture*————————-

MS.28 Identify ATP from its molecular structure and identify the core nucleotide functional units: nitrogenous base, ribose, and phosphates.

MS.27 Identify the “high-energy” bonds in ATP.

GC.71 Describe how the term "high energy" is commonly used in the context of molecular bonds and what is implied by its use. That is, answer: what makes a bond "high energy" and when is it appropriate to use the term?

ME.3 Create a thermodynamic argument for how ATP hydrolysis and/or phosphate group transfer can be coupled to drive endergonic reactions.

ME.4 Explain the process of substrate level phosphorylation (SLP) and identify the SLP reactions when given a collection of reactions, such as in a pathway.

ME.5 Explain the important contribution of water in determining the negative $\Delta G^{0'}$ of the hydrolysis of a phosphoanhydride bond in ATP.

ME.7 Create an "energy story" for glycolysis. The story should list the overall reactants and products, sources of energy, energy transfers, the types of reactions involved in energy transfer, and the mediators of the transformations of matter and transfers of energy.

ME.17 Explain the importance of the reaction catalyzed by glyceraldehyde-3-phosphate dehydrogenase in the harvesting of energy in glycolysis. Use this mechanism and figures to unite lessons from the course: enzymes and catalysts, energy coupling, redox, functional group chemistry, etc.

ME.20 Be able to interpret figures depicting the mechanism of glyceraldehyde-3-phosphate dehydrogenase and identify key steps in the reaction including the role of a catalytic histidine and the formation of a covalent thioester linkage (including its role in energy transfer).

ME.21 Create an energy story for the reaction catalyzed by glyceraldehyde-3-phosphate dehydrogenase, that discusses specifically the coupling of a redox reaction to a phosphate transfer.

ATP

An important chemical compound is adenosine triphospate (**ATP**). The main cellular role of ATP is as a “short-term” energy transfer device for the cell. The hydrolysis reactions that liberate one or more of ATP's phosphates are exergonic and many, many cellular proteins have evolved to interact with ATP in ways that help facilitate the transfer of energy from hydrolysis to myriad other cellular functions. In this way, ATP is often called the “energy currency” of the cell: it has reasonably fixed values of energy to transfer to or from itself and can exchange that energy between many potential donors and acceptors. We will see many examples of ATP "at work" in the cell, so be looking for them. As you see them, try to think of them as functional examples of Nature's uses for ATP that you could expect to see in another reaction or context.

ATP structure and function

At the heart of ATP is the nucleotide called adenosine monophosphate (AMP). Like the other nucleotides, AMP is composed of a nitrogenous base (an adenine molecule) bonded to a ribose molecule and a single phosphate group. The addition of a second phosphate group to this core molecule results in the formation of adenosine diphosphate (ADP); the addition of a third phosphate group forms adenosine triphosphate (ATP).

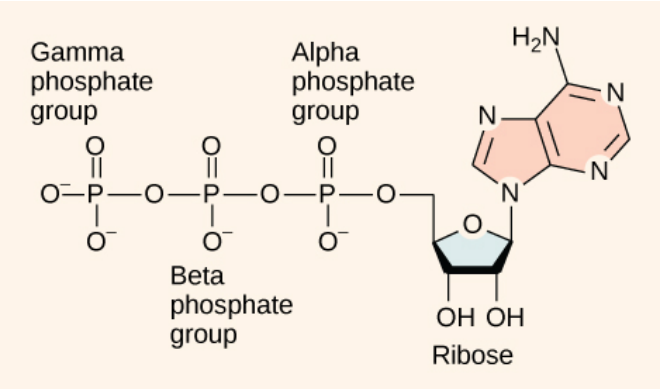


Figure 1. *ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate).*

The **phosphorylation** (or condensation of phosphate groups onto AMP) is an endergonic process. By contrast, the hydrolysis of one or two phosphate groups from ATP, a process called **dephosphorylation**, is exergonic. Why? Let's recall that the terms endergonic and exergonic refer to the sign on the difference in standard free energy of a reaction between the products and reactants, $\Delta G°'$. Here we are explicitly assigning a direction to the reaction, either toward phosphorylation or dephosphorylation of the nucleotide. In the phosphorylation reaction, the reactants are the nucleotide and an inorganic phosphate while the products are a phosphorylated nucleotide and WATER. In the dephosphorylation/hydrolysis reaction, the reactants are the phosphorylated nucleotide and WATER while the products are inorganic phosphate and the nucleotide minus one phosphate.

Since Gibbs free energy is a state function, it doesn't matter how the reaction happens; you just consider the beginning and ending states. As an example, let's examine the hydrolysis of ATP. The reactants ATP and water are characterized by their atomic makeup and the kinds of bonds between the constituent atoms. We can associate some free energy with each of the bonds and their possible configurations—likewise for the products. If we examine the reaction from the standpoint of the products and reactants and ask "how can we recombine atoms and bonds in the reactants to get the products?," we find that a phosphoanhydride bond between an oxygen and a phosphorus must be broken in the ATP, a bond between an oxygen and hydrogen must be broken in the water, a bond must be made between the OH (that came from the splitting of water) and the phosphorus (from the freed PO_3^{-2}), and a bond must be formed between the H (derived from the splitting of water) and the terminal oxygen on the phosphorylated nucleotide. It is the sum of energies associated with all of those bond rearrangements (including those directly associated with water) that makes this reaction exergonic. We could make a similar analysis with the reverse reaction.

Is there something special about the specific bonds involved in these molecules? Much is made in various texts about the types of bonds between the phosphates of ATP. Certainly, the properties of the bonds in ATP help define the molecule's free energy and reactivity. However, while it is appropriate to apply concepts like charge density and availability of resonance structures to this discussion, trotting these terms out as an "explanation" without a thorough understanding of how these factors influence the free energy of the reactants is a special kind of hand-waving that we shouldn't engage in. Numerous students of general biloogy have not taken any college chemistry and those who have are not likely to have discussed those terms in any meaningful way. So, explaining the process using the ideas above only gives a false sense of understanding, assigns some mystical quality to ATP and its "special" bonds (that don't exist), and distracts from the real point: the hydrolysis reaction is exergonic because of the properties of ATP and ALSO because of the chemical properties of water and those of the reaction products. For this class, it is sufficient to know that dedicated physical chemists are still studying the process of ATP hydrolysis in solution and in the context of proteins and that they are still trying to account for the key enthalpic and entropic components of the component free energies. We'll just need to accept a certain degree of mechanistic chemical ignorance and be content with a description of gross thermodynamic properties. The latter is perfectly sufficient to have deep discussions about the relevant biology.

"High-Energy" bonds

What about the term "high-energy bonds" that we so often hear associated with ATP? If there is nothing "special" about the bonds in ATP, why do we always hear the term "high-energy bonds" associated with the molecule? The answer is deceptively simple. In biology the term "high-energy bond" is used to describe an exergonic reaction involving the hydrolysis of the bond in question that

results in a "large," negative change in free energy. Remember that this change in free energy has not only to do with the bond in question but rather the sum of all bond rearrangements in the reaction. What constitutes a large change? It is a rather arbitrary assignment usually associated with an amount of energy associated with the types of anabolic reactions we typically observe in biology. If there is something special about the bonds in ATP, it is not uniquely tied to the free energy of hydrolysis, as there are plenty of other bonds whose hydrolysis results in greater negative differences in free energy.

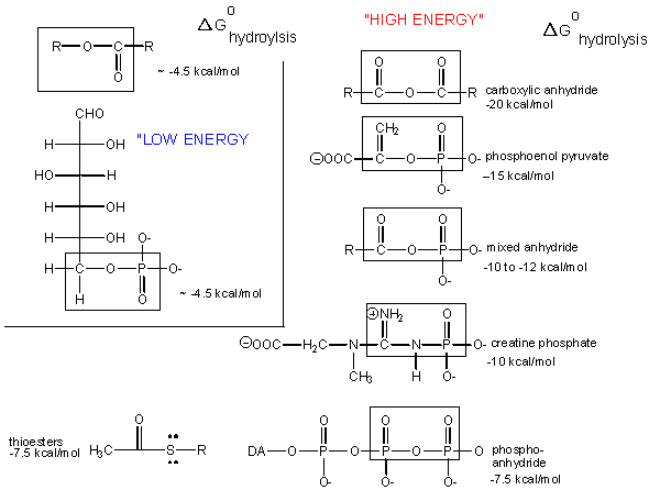


Figure 2. *The standard free energy of hydrolysis of different types of bonds can be compared to that of the hydrolysis of ATP. Source:* *http://bio.libretexts.org/Core/Biochemistry/Oxidation_and_Phosphorylation/ATP_and_Oxidative_Phosphorylation/Properties_of_ATP*

Compound	$\Delta G^{0\prime}$ kJ/mol
Phosphoenol pyruvate	-51.6
1,3-bis-Phosphoglycerate	-52.0
Acetylphosphate	-44.8
ATP	-31.8
ADP	-31.8
Acetyl-CoA	-31
AMP	-14.2
Glucose-6-P	-13.8

Table 1. *Table of common cellular phosphorylated molecules and their respective standard free energies of hydrolysis.*

External link discussing the energetics of coupling ATP hydrolysis to other reactions

http://bio.libretexts.org/Core/Biochemistry/Oxidation_and_Phosphorylation/ATP_and_Oxidative_Phosphorylation/Properties_of_ATP

Possible NB Discussion Point

You have just now read about two important molecules: NADH/NAD^+ and ATP. In what biological contexts/process do you expect to see NADH/NAD^+? What about ATP? Can you state what you know so far about

the relationship between NADH/NAD^+ and ATP? Take a moment to identify any gaps in comprehension you might have -- what questions are you left with after reading the text? Help your peers out with their questions/discussions to reinforce your own knowledge!

The cycling of ATP pools

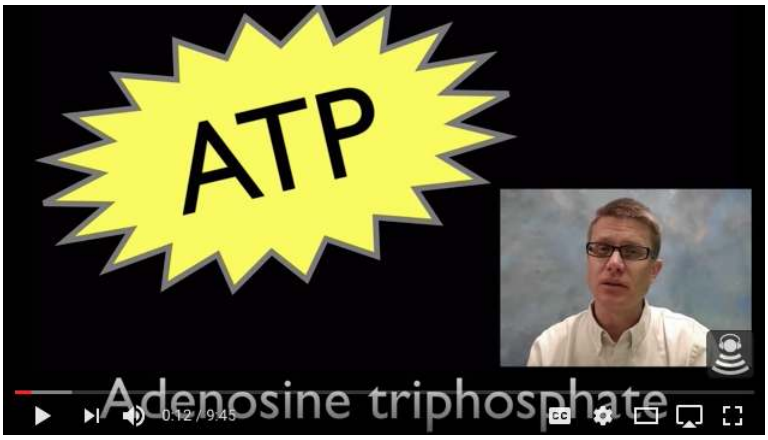
Estimates for the number of ATP molecules in a typical human cell range from ~$3x10^7$ (~$5x10^{-17}$ moles ATP/cell) in a white blood cell to $5x10^9$ (~$9x10^{-15}$ moles ATP/cell) in an active cancer cell. While these numbers might seem large, and already amazing, consider that it is estimated that this pool of ATP turns over (becomes ADP and then back to ATP) 1.5 x per minute. Extending this analysis yields the estimate that this daily turnover amounts to roughly the equivalent of one body weight of ATP getting turned over per day. That is, if no turnover/recycling of ATP happened, it would take one body weight worth of ATP for the human body to function, hence our previous characterization of ATP as a "short-term" energy transfer device for the cell.

While the pool of ATP/ADP may be recycled, some of the energy that is transferred in the many conversions between ATP, ADP, and other biomolecules is also transferred to the environment. In order to maintain cellular energy pools, energy must transfer in from the environment as well. Where does this energy come from? The answer depends a lot on where energy is available and what mechanisms Nature has evolved to transfer energy from the environment to molecular carriers like ATP. In nearly all cases, however, the mechanism of transfer has evolved to include some form of redox chemistry.

In this and the sections that follow we are concerned with learning some critical examples of energy transfer from the environment, key types of chemistry and biological reactions involved in this process, and key biological reactions and cellular components associated with energy flow between different parts of the living system. We focus first on reactions involved in the (re)generation of ATP in the cell (not those involved in the creation of the nucleotide per se but rather those associated with the transfer of phosphates onto AMP and ADP).

Video link

For another perspective - including places you'll see ATP in Bis2a, take a look at this video (10 minutes) by clicking here.



How do cells generate ATP?

A variety of mechanisms have emerged over the 3.25 billion years of evolution to create ATP from ADP and AMP. The majority of these mechanisms are modifications on two themes: direct synthesis of ATP or indirect synthesis of ATP with two basic mechanisms known respectively as **substrate level phosphorylation (SLP)** and **oxidative phosphorylation**. Both mechanisms rely on biochemical reactions that transfer energy from some energy source to ADP or AMP to synthesize ATP. These topics are substantive, so they will be discussed in detail in the next few modules.

Glycolysis: An overview



Organisms, whether unicellular or multicellular, need to find ways of getting at least two key things from their environment: (1) matter or raw materials for maintaining a cell and building new cells and (2) energy to help with the work of staying alive and reproducing. Energy and the raw materials may come from different places. For instance, organisms that primarily harvest energy from sunlight will get raw materials for building biomolecules from sources like CO_2. By contract, some organisms rely on red/ox reactions with small molecules and/or reduced metals for energy and get their raw materials for building biomolecules from compounds unconnected to the energy source. Meanwhile, some organisms (including ourselves), have evolved to get energy AND the raw materials for building and cellular maintenance from sometimes associated sources.

Glycolysis is the first **metabolic pathway** discussed in BIS2A; a metabolic pathway is a series of linked biochemical reactions. Because of its ubiquity in biology, we hypothesize that glycolysis was probably one of the earliest metabolic pathways to evolve (more on this later). Glycolysis is a ten-step metabolic pathway that is centered on the processing of glucose for both energy extraction from chemical fuel and for the processing of the carbons in glucose into various other biomolecules (some of which are key precursors of many much more complicated biomolecules). We will therefore examine our study of glycolysis using the precepts outlined in the energy challenge rubric that ask us to formally consider what happens to BOTH matter and energy in this multi-step process.

The energy story and design challenge of glycolysis

Our investigation of glycolysis is a good opportunity to examine a biological process using both the energy story and the design challenge rubrics and perspectives.

The design challenge rubric will try to get you to think actively, and broadly and specifically, about why we are studying this pathway—what is so important about it? What "problems" does the evolution of a glycolytic pathway allow life to solve or overcome? We will also want to think about alternate ways to solve the same problems and why they may or may not have evolved. Later, we will examine a hypothesis for how this pathway—and other linked pathways—may have evolved, and thinking about alternative strategies for satisfying various constraints will come in handy then.

We ask you to think about glycolysis through the lens of an energy story in which you examine the 10-step process as a set of matter and energy inputs and outputs, a process with a beginning and an end. By taking this approach you will learn not only about glycolysis, but also some skills required to read and interpret other biochemical pathways.

So what is glycolysis? Let's find out.

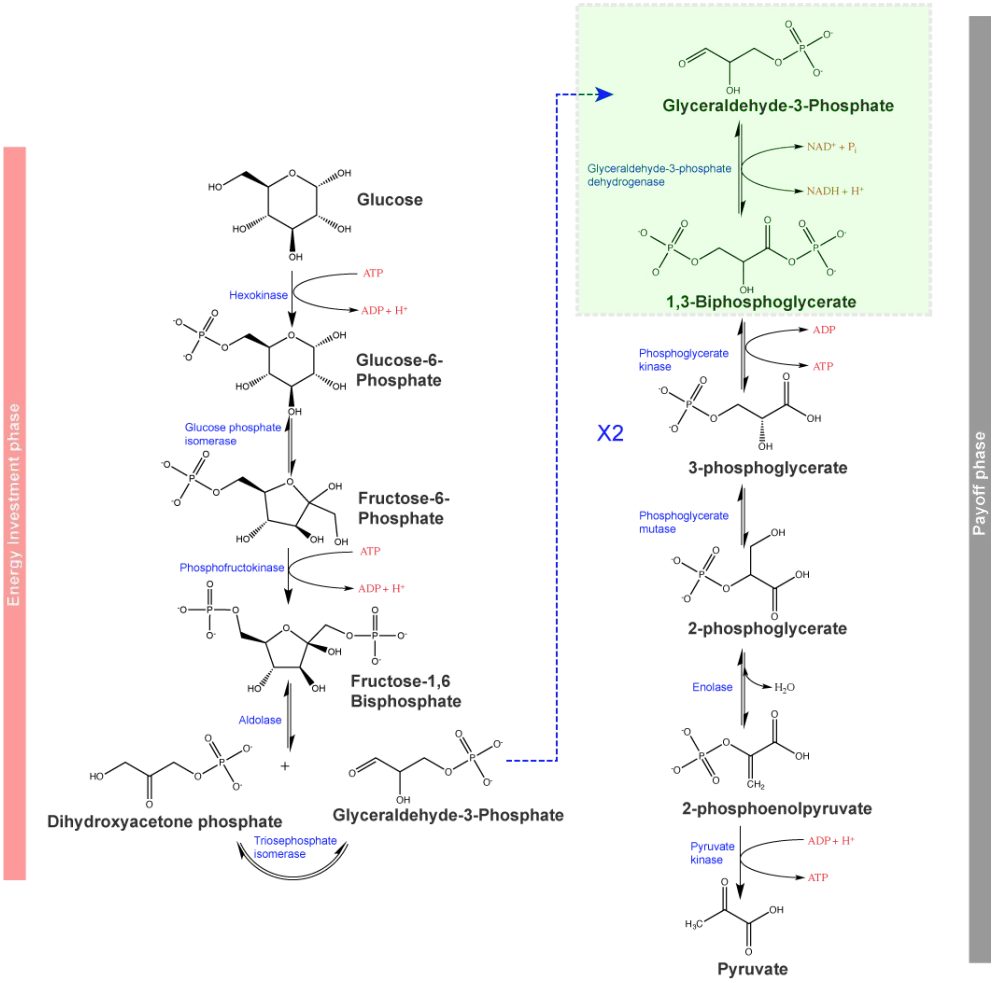


Figure 1. *The ten biochemical reactions of glycolysis are shown. Enzymes are labeled in blue. The structure of each sugar-derived compound is depicted as a molecular model; other reactants and products may be abbreviated (e.g., ATP, NAD+, etc.). The box surrounding the reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase indicates that this reaction is of special interest in the course. Attribution: Marc T. Facciotti* (original work)

Table 1. *This table shows glycolytic enzymes and measurements of the energy at standard state (ΔG°'/(kJ/mol)) compared with measurements taken from a living cell (ΔG/(kJ/mol)). Under conditions of constant temperature and pressure, (ΔG°'/(kJ/mol)), reactions will occur in the direction that leads to a decrease in the value of the Gibbs free energy. Cellular measurements of ΔG can be dramatically different from ΔG°' measurements because of cellular conditions, such as concentrations of relevant metabolites, etc. There are three large, negative ΔG drops in the cell in the process of glycolysis. We consider these reactions irreversible and are often subject to regulation*

Enzyme	Step	ΔG/(kJ/mol)	ΔG°'/(kJ/mol)
Hexokinase	1	-34	-16.7
Phosphoglucose isomerase	2	-2.9	1.67
Phosphofructokinase	3	-19	-14.2
Fructose-bisphosphate aldolase	4	-0.23	23.9
Triose phosphate isomerase	5	2.4	7.56
Glyceraldehyde 3-phosphate dehydrogenase	6	-1.29	6.30
Phosphoglycerate kinase	7	0.09	-18.9
Phosphoglycerate mutase	8	0.83	4.4
Enolase	9	1.1	1.8
Pyruvate kinase	10	-23.0	-31.7

Overall, the glycolytic pathway comprises 10 enzyme-catalyzed steps. The primary input into this pathway is a single molecule of glucose, though we discover that other molecules may enter this pathway at various steps. We will focus our attention on (1) consequences of the overall process, (2) several key reactions that highlight important types of biochemistry and biochemical principles we will want to carry forward to other contexts, and (3) alternative fates of the intermediates and products of this pathway.

Note for reference that glycolysis is an ***anaerobic*** process. There is no requirement for molecular oxygen in glycolysis - oxygen gas is not a reactant in any of the chemical reactions in glycolysis. Glycolysis occurs in the ***cytosol*** *or* ***cytoplasm*** of cells. For a short (three-minute) overview YouTube video of glycolysis, click here.

First half of glycolysis: energy investment phase

We typically refer the first few steps of glycolysis as an "energy investment phase" of the pathway. This, however, doesn't make much intuitive sense (in the framework of a design challenge; it's not clear what problem this energy investment solves) if one only looks at glycolysis as an "energy-producing" pathway and until these steps of glycolysis are put into a broader metabolic context. We'll try to build that story as we go, so for now just recall that we mentioned that some first steps are often associated with energy investment and ideas like "trapping" and "commitment" that are noted in the figure below.

Step 1 of glycolysis:

The first step in glycolysis, shown below in Figure 2, is glucose being catalyzed by hexokinase, an enzyme with broad specificity that catalyzes the phosphorylation of six-carbon sugars. Hexokinase catalyzes the phosphorylation of glucose, where glucose and ATP are substrates for the reaction, producing a molecule called glucose 6-phosphate and ADP as products.

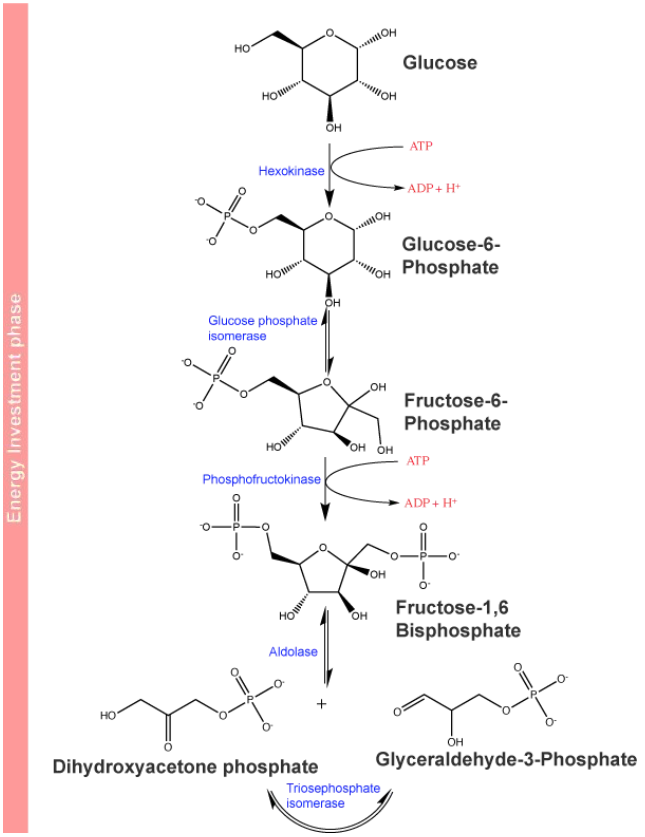


Figure 2. *The first half of glycolysis is called the energy investment phase. In this phase, the cell spends two ATPs into the reactions. Attribution: Marc T. Facciotti (original work)*

Note:

The paragraph above states that the enzyme hexokinase has "broad specificity." This means that it can catalyze reactions with different sugars, not just glucose. From a molecular perspective, can you explain why this might be the case? Does this challenge your conception of enzyme specificity? If you Google the term "enzyme promiscuity" (don't worry; it's safe for work), you will hopefully gain a broader appreciation for enzyme selectivity and activity.

The conversion of glucose to the negatively charged glucose 6-phosphate significantly reduces the likelihood that the phosphorylated glucose leaves the cell by diffusion across the hydrophobic interior of the plasma membrane. It also "marks" the glucose in a way that tags it for several possible fates (see Figure 3).

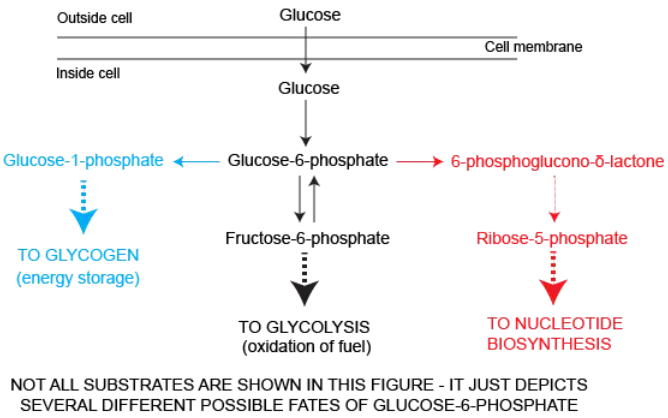


Figure 3. *Note that this figure shows that glucose 6-phosphate can, depending on cellular conditions, be directed to multiple fates. While it is a component of the glycolytic pathway, it is not only involved in glycolysis but also in the storage of energy as glycogen*

(colored in cyan) and in the building of various other molecules like nucleotides (colored in red). Source: Marc T. Facciotti (original work)

As Figure 3 shows, glycolysis is but one fate for glucose 6-phosphate (G6P). Depending on cellular conditions, G6P may be diverted to the biosynthesis of glycogen (for energy storage), or it may be diverted into the pentose phosphate pathway for the biosynthesis of various biomolecules, including nucleotides. This means that G6P, while involved in the glycolytic pathway, is not solely tagged for oxidation at this phase. Perhaps showing the broader context that this molecule is involved in (in addition to the rationale that tagging glucose with a phosphate decreases the likelihood that it will leave the cell) helps to explain the seemingly contradictory (if you only consider glycolysis as an "energy-producing" process) reason for transferring energy from ATP onto glucose if it is only to be oxidized later—that is, glucose is not only used by the cell for harvesting energy and several other metabolic pathways depend on the transfer of the phosphate group.

Step 2 of glycolysis:

In the second step of glycolysis, an ***isomerase*** catalyzes the conversion of glucose 6-phosphate into one of its isomers, fructose 6-phosphate. An *isomerase* is an enzyme that catalyzes the conversion of a molecule into one of its isomers.

Step 3 of glycolysis:

The third step of glycolysis is the phosphorylation of fructose 6-phosphate, catalyzed by the enzyme phosphofructokinase. A second ATP molecule donates a phosphate to fructose 6-phosphate, producing fructose 1,6-bisphosphate and ADP as products. In this pathway, phosphofructokinase is a rate-limiting enzyme, and its activity is tightly regulated. It is ***allosterically*** activated by AMP when the concentration of AMP is high and when it is moderately allosterically inhibited by ATP at the same site. Citrate, a compound we'll discuss soon, also acts as a negative ***allosteric*** regulator of this enzyme. In this way, phosphofructokinase monitors or senses molecular indicators of the energy status of the cells and can in response act as a switch that turns on or off the flow of the substrate through the rest of the metabolic pathway depending on whether there is “sufficient” ATP in the system. The conversion of fructose 6-phosphate into fructose 1,6-bisphosphate is sometimes referred to as a commitment step by the cell to the oxidation of the molecule in the rest of the glycolytic pathway by creating a substrate for and helping to energetically drive the next highly endergonic (under standard conditions) step of the pathway.

Step 4 of glycolysis:

In the fourth step in glycolysis, an enzyme, fructose-bisphosphate aldolase, cleaves 1,6-bisphosphate into two three-carbon isomers: dihydroxyacetone phosphate and glyceraldehyde 3-phosphate.

Second half: energy payoff phase

If viewed in the absence of other metabolic pathways, glycolysis has so far cost the cell two ATP molecules and produced two small, three-carbon sugar molecules: dihydroxyacetone phosphate (DAP) and glyceraldehyde 3-phosphate (G3P). When viewed in a broader context, this investment of energy to produce a variety of molecules that can be used in a variety of other pathways doesn't seem like such a bad investment.

Both DAP and G3P can proceed through the second half of glycolysis. We now examine these reactions.

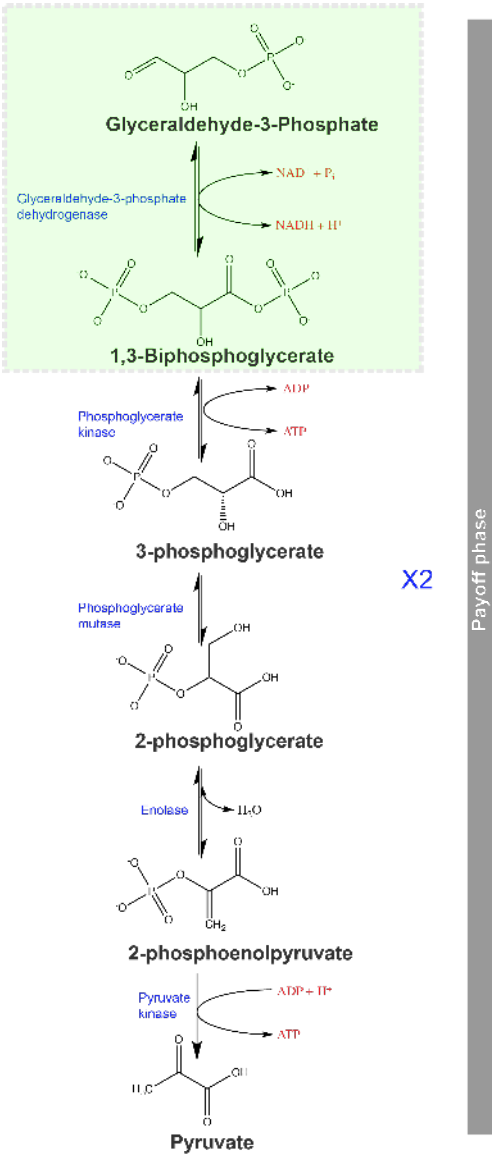


Figure 4. *The second half of glycolysis is called the energy payoff phase. In this phase, the cell gains two ATP and two NADH compounds. At the end of this phase, glucose has become partially oxidized to form pyruvate. Attribution: Marc T. Facciotti (original work).*

Step 5 of glycolysis:

In the fifth step of glycolysis, an isomerase transforms the dihydroxyacetone phosphate into its isomer, glyceraldehyde 3-phosphate. The six-carbon glucose has therefore now been converted into two phosphorylated three-carbon molecules of G3P.

Step 6 of glycolysis:

The sixth step is key and one from which we can now leverage our understanding of the several chemical reactions that we've studied so far. If you're energy focused, this is finally a step of glycolysis where some reduced sugar becomes oxidized. The reaction is catalyzed by the enzyme glyceraldehyde 3-phosphate dehydrogenase. This enzyme catalyzes a multi-step reaction between three substrates—glyceraldehyde 3-phosphate, the cofactor NAD^+, and inorganic phosphate (P_i)—and produces three products: 1,3-bisphosphoglycerate, NADH, and H^+. One can think of this reaction as two reactions: (1) an oxidation/reduction reaction and (2) a condensation reaction in which an inorganic phosphate is transferred onto a molecule. Here, the red/ox reaction, a transfer of electrons off G3P and onto NAD^+, is exergonic, and the phosphate transfer is endergonic. The net standard free energy change hovers around zero—more on this later. The enzyme here acts as a molecular **coupling** agent to couple the energetics of the

exergonic reaction to that of the endergonic reaction, thus driving both forward. This processes happens through a multi-step mechanism in the enzyme's active site and involves the chemical activity of a variety of functional groups.

It is important to note that this reaction depends upon the availability of the oxidized form of the electron carrier, NAD^+. If we consider that there is a limiting pool of NAD^+, we can then conclude that the reduced form of the carrier (NADH) must continuously oxidize back into NAD^+ to keep this step going. If NAD^+ is not available, the second half of glycolysis slows down or stops.

Possible NB Discussion Point

Can you write an energy story for Step 6 of glycolysis (the reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase)? When you discuss energy, just describe whether steps are exergonic or endergonic. As a group, try to build up an increasingly "expert" version that is complete, brief, and uses appropriate vocabulary. Amend one another's texts in a polite and constructive way.

Step 7 of glycolysis:

In the seventh step of glycolysis, catalyzed by phosphoglycerate kinase (an enzyme named for the reverse reaction), 1,3-bisphosphoglycerate transfers a phosphate to ADP, forming one molecule of ATP and a molecule of 3-phosphoglycerate. This reaction is exergonic and is also an example of substrate-level phosphorylation.

Step 8 of glycolysis:

In the eighth step, the remaining phosphate group in 3-phosphoglycerate moves from the third carbon to the second carbon, producing 2-phosphoglycerate (an isomer of 3-phosphoglycerate). The enzyme catalyzing this step is a mutase (isomerase).

Step 9 of glycolysis:

Enolase catalyzes the ninth step. This enzyme causes 2-phosphoglycerate to lose water from its structure; this is a dehydration reaction, resulting in the formation of a double bond that increases the potential energy in the remaining phosphate bond and produces phosphoenolpyruvate (PEP).

Step 10 of glycolysis:

The last step in glycolysis is catalyzed by the enzyme pyruvate kinase (the enzyme in this case is named for the reverse reaction of pyruvate’s conversion into PEP) and results in the production of a second ATP molecule by substrate-level phosphorylation and the compound pyruvic acid (or its salt form, pyruvate). Many enzymes in enzymatic pathways are named for the reverse reactions, since the enzyme can catalyze both forward and reverse reactions (these may have been described initially by the reverse reaction that takes place in vitro, under non-physiological conditions).

Outcomes of glycolysis

Here are a couple of things to consider:

One of the clear outcomes of glycolysis is the biosynthesis of compounds that can enter into a variety of metabolic pathways. Likewise, compounds coming from other metabolic pathways can feed into glycolysis at various points. So, this pathway can be part of a central exchange for carbon flux within the cell.

If glycolysis runs long enough, the constant oxidation of glucose with NAD^+ can leave the cell with a problem: how to regenerate NAD^+ from the two molecules of NADH produced. If the cell does not regenerate NAD^+, nearly all the cell's NAD+ will transform into NADH. So how do cells regenerate NAD^+?

Pyruvate is not completely oxidized; There is still some energy to extract. How might this happen? Also, what should the cell do with all of that NADH? Is there any energy there to extract?

Possible NB Discussion Point

To some, that glycolysis is such a complex, multi-step pathway may seem counter-intuitive: *"Why wouldn't evolution lead to a *simpler* way to extract energy from food since energy is an important requirement for life?"* Explain the necessity/advantage of having glucose get broken down in many steps.

Substrate-level phosphorylation (SLP)

The simplest route to synthesize ATP is substrate-level phosphorylation. ATP molecules are generated (that is, regenerated from ADP) because of a chemical reaction that occurs in catabolic pathways. A phosphate group is removed from an intermediate reactant in the pathway, and the free energy of the reaction is used to add the third phosphate to an available ADP molecule, producing ATP. This very direct method of phosphorylation is called *substrate-level phosphorylation (SLP)*. We can find SLP in a variety of catabolic reactions, most notably in two specific reactions in glycolysis (which we will discuss specifically later). What the reaction requires is a high-energy intermediate compound whose free energy of oxidation can drive the synthesis of ATP.

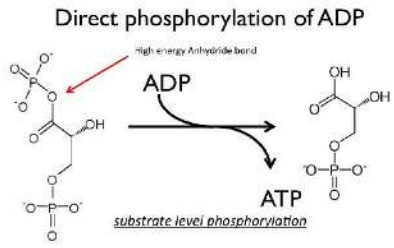


Figure 5. *Here is one example of substrate-level phosphorylation occurring in glycolysis. There is a direct transfer of a phosphate group from the carbon compound onto ADP to form ATP. Attribution: Marc T. Facciotti (own work)*

In this reaction, the reactants are a phosphorylated carbon compound called G3P (from step 6 of glycolysis) and an ADP molecule, and the products are 1,3-BPG and ATP. The transfer of the phosphate from G3P to ADP to form ATP in the active site of the enzyme is *substrate-level phosphorylation*. This occurs twice in glycolysis and once in the TCA cycle (for a subsequent reading).

Reading_11 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_12

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_12

ME.16 Apply the concept of the “conservation of mass” to metabolism by describing the different forms mass takes as it enters and leaves the cell (e.g. input: reduced molecules like glucose, lipids, proteins, etc. & output: oxidized molecules like CO_2, H_2O etc.).

ME.15 Apply the concept of the "conservation of energy" to central metabolism. Follow energy from "sources" of electrons with relatively low reduction potentials to "sinks" with higher redox potentials, describe the major transfers of energy and how this energy is "stored" at each stage.

————*-Learning objectives above are for the full course, but particularly relevant to the discussion of metabolism and thus extend well beyond this lecture*————-

ME.11 Create an energy story explaining the functional value of pyruvate oxidation and its importance in generating building blocks for biomolecules.

ME.8 Create an “energy story” explaining the functional value of pyruvate reduction (fermentation) to a cell and its importance in regenerating the NAD^+ pool.

ME.14 Given metabolic maps of glycolysis and TCA cycle, follow the flow of electrons from energy source to mobile carrier NADH and use a provided redox tower to quantitatively describe the energy transformations.

ME.12 Create an energy story for each reaction in the TCA cycle.

ME.10 Explain the central role of Acetyl CoA in carbon metabolism including its formation from various sources (the oxidative decarboxylation of pyruvate, oxidation of fatty acids, and/or oxidative degradation of some amino acids) and as a carbon source in both the TCA cycle and lipid synthesis.

ME.13 Describe and give examples of cyclic processes and how they are involved in energy harvesting and the inter-conversion of molecule types.

ME.18 Relate Coenzyme A's ability to form thioesters with the concepts of its function of carrying "activated" acyl groups.

ME.19 Identify Coenzyme A and its functionally important thiol group by inspection of a molecular model.

ME.9 Discuss at least two key fates for glucose other than its direct oxidation to pyruvate, highlighting the role of glycolytic intermediates in processes related to carbon flow (functions that are not associated with energy harvesting).

ME.91 Explain the role of glucose in the creation of medium- and long-term energy storage compounds.

Fermentation and regeneration of NAD+



Section summary

This section discusses the process of fermentation. Due to the heavy emphasis in this course on central carbon metabolism, the discussion of fermentation understandably focuses on the fermentation of pyruvate. Nevertheless, some of the core principles that we cover in this section apply equally well to the fermentation of many other small molecules.

The "purpose" of fermentation

The oxidation of a variety of small organic compounds is a process that is utilized by many organisms to garner energy for cellular maintenance and growth. The oxidation of glucose via glycolysis is one such pathway. Several key steps in the oxidation of glucose to pyruvate involve the reduction of the electron/energy shuttle NAD^+ to NADH. You were already asked to figure out what options the cell might reasonably have to reoxidize the NADH to NAD^+ in order to avoid consuming the available pools of NAD^+ and to thus avoid stopping glycolysis. Put differently, during glycolysis, cells can generate large amounts of NADH and slowly exhaust their supplies of NAD^+. If glycolysis is to continue, the cell must find a way to regenerate NAD^+, either by synthesis or by some form of recycling.

In the absence of any other process—that is, if we consider glycolysis alone—it is not immediately obvious what the cell might do. One choice is to try putting the electrons that were once stripped off of the glucose derivatives right back onto the downstream product, pyruvate, or one of its derivatives. We can generalize the process by describing it as the returning of electrons to the molecule that they were once removed, usually to restore pools of an oxidizing agent. This, in short, is fermentation. As we will

discuss in a different section, the process of respiration can also regenerate the pools of NAD^+ from NADH. Cells lacking respiratory chains or in conditions where using the respiratory chain is unfavorable may choose fermentation as an alternative mechanism for garnering energy from small molecules.

An example: lactic acid fermentation

An everyday example of a fermentation reaction is the reduction of pyruvate to lactate by the lactic acid fermentation reaction. This reaction should be familiar to you: it occurs in our muscles when we exert ourselves during exercise. When we exert ourselves, our muscles require large amounts of ATP to perform the work we are demanding of them. As the ATP is consumed, the muscle cells are unable to keep up with the demand for respiration, O_2 becomes limiting, and NADH accumulates. Cells need to get rid of the excess and regenerate NAD^+, so pyruvate serves as an electron acceptor, generating lactate and oxidizing NADH to NAD^+. Many bacteria use this pathway as a way to complete the NADH/NAD^+ cycle. You may be familiar with this process from products like sauerkraut and yogurt. The chemical reaction of lactic acid fermentation is the following:

Pyruvate + NADH ↔ lactic acid + NAD^+

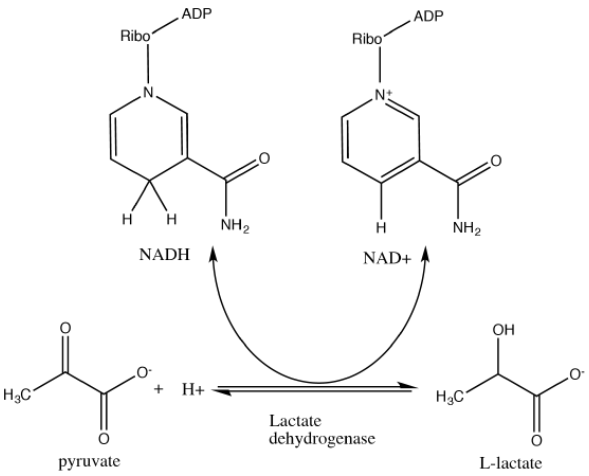


Figure 1. *Lactic acid fermentation converts pyruvate (a slightly oxidized carbon compound) to lactic acid. In the process, NADH is oxidized to form NAD^+. Attribution: Marc T. Facciotti (original work)*

Energy story for the fermentation of pyruvate to lactate

An example (if a bit lengthy) energy story for lactic acid fermentation is the following:

The reactants are pyruvate, NADH, and a proton. The products are lactate and NAD^+. The process of fermentation results in the reduction of pyruvate to form lactic acid and the oxidation of NADH to form NAD^+. Electrons from NADH and a proton are used to reduce pyruvate into lactate. If we examine a table of standard reduction potential, we see under standard conditions that a transfer of electrons from NADH to pyruvate to form lactate is exergonic and thus thermodynamically spontaneous. The reduction and oxidation steps of the reaction are coupled and catalyzed by the enzyme lactate dehydrogenase.

A second example: alcohol fermentation

Another familiar fermentation process is alcohol fermentation, which produces ethanol, an alcohol. The alcohol fermentation reaction is the following:

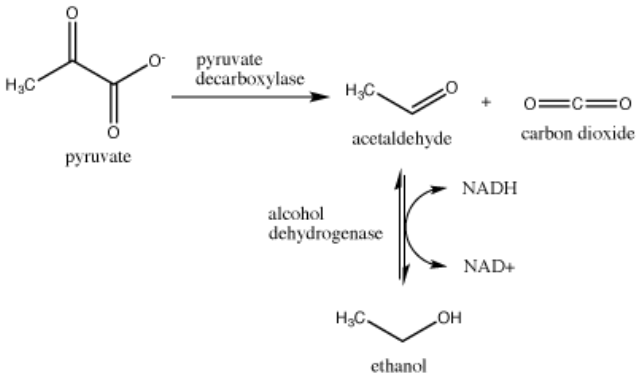


Figure 2. *Ethanol fermentation is a two-step process. Pyruvate (pyruvic acid) is first converted into carbon dioxide and acetaldehyde. The second step converts acetaldehyde to ethanol and oxidizes NADH to NAD$^+$. Attribution: Marc T. Facciotti (original work)*

In the first reaction, a carboxyl group is removed from pyruvic acid, releasing carbon dioxide as a gas (some of you may be familiar with this as a key component of various beverages). The second reaction removes electrons from NADH, forming NAD$^+$ and producing ethanol (another familiar compound—usually in the same beverage) from the acetaldehyde, which accepts the electrons.

Fermentation pathways are numerous

While the lactic acid fermentation and alcohol fermentation pathways described above are examples, there are many more reactions (too numerous to go over) that Nature has evolved to complete the NADH/NAD$^+$ cycle. It is important that you understand the general concepts behind these reactions. In general, cells try to maintain a balance or constant ratio between NADH and NAD$^+$; when this ratio becomes unbalanced, the cell compensates by modulating other reactions to compensate. The only requirement for a fermentation reaction is that it uses a small organic compound as an electron acceptor for NADH and regenerates NAD$^+$. Other familiar fermentation reactions include ethanol fermentation (as in beer and bread), propionic fermentation (it's what makes the holes in Swiss cheese), and malolactic fermentation (it's what gives Chardonnay its more mellow flavor—the more conversion of malate to lactate, the softer the wine). In Figure 3, you can see a large variety of fermentation reactions that various bacteria use to reoxidize NADH to NAD$^+$. All of these reactions start with pyruvate or a derivative of pyruvate metabolism, such as oxaloacetate or formate. Pyruvate is produced from the oxidation of sugars (glucose or ribose) or other small, reduced organic molecules. It should also be noted that other compounds can be used as fermentation substrates besides pyruvate and its derivatives. These include methane fermentation, sulfide fermentation, or the fermentation of nitrogenous compounds such as amino acids. You are not expected to memorize all of these pathways. You are, however, expected to recognize a pathway that returns electrons to products of the compounds that were originally oxidized to recycle the NAD$^+$/NADH pool and to associate that process with fermentation.

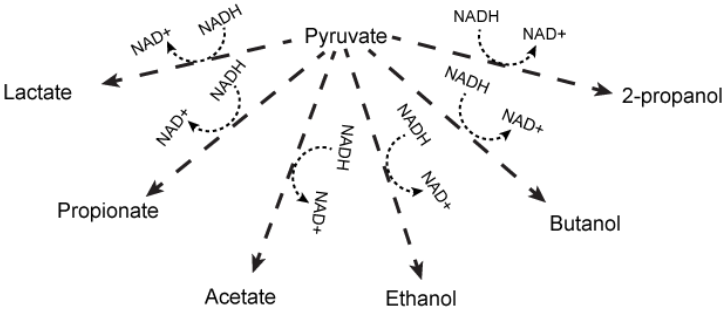


Figure 3. *This figure shows various fermentation pathways using pyruvate as the initial substrate. In the figure, pyruvate is reduced to a variety of products via different and sometimes multistep (dashed arrows represent possible multistep processes) reactions. All details are deliberately not shown. The key point is to appreciate that fermentation is a broad term not solely associated with the conversion of pyruvate to lactic acid or ethanol. Source: Marc T. Facciotti (original work)*

A note on the link between substrate-level phosphorylation and fermentation

Fermentation occurs in the absence of molecular oxygen (O_2). It is an anaerobic process. Notice there is no O_2 in any of the fermentation reactions shown above. Many of these reactions are quite ancient, hypothesized to be some of the first energy-generating metabolic reactions to evolve. This makes sense if we consider the following:

1. The early atmosphere was highly reduced, with little molecular oxygen readily available.
2. Small, highly reduced organic molecules were relatively available, arising from a variety of chemical reactions.
3. These types of reactions, pathways, and enzymes are found in many different types of organisms, including bacteria, archaea, and eukaryotes, suggesting these are very ancient reactions.
4. The process evolved long before O_2 was found in the environment.
5. The substrates, highly reduced, small organic molecules, like glucose, were readily available.
6. The end products of many fermentation reactions are small organic acids, produced by the oxidation of the initial substrate.
7. The process is coupled to substrate-level phosphorylation reactions. That is, small, reduced organic molecules are oxidized, and ATP is generated by first a red/ox reaction followed by the substrate-level phosphorylation.
8. This suggests that substrate-level phosphorylation and fermentation reactions coevolved.

Consequences of fermentation

Imagine a world where fermentation is the primary mode for extracting energy from small molecules. As populations thrive, they reproduce and consume the abundance of small, reduced organic molecules in the environment, producing acids. One consequence is the acidification (decrease in pH) of the environment, including the internal cellular environment. This can be disruptive, since changes in pH can have a profound influence on the function and interactions among various biomolecules. Therefore, mechanisms needed to evolve that could remove the various acids. Fortunately, in an environment rich in reduced compounds, substrate-level phosphorylation and fermentation can produce large quantities of ATP.

It is hypothesized that this scenario was the beginning of the evolution of the F_0F_1-ATPase, a molecular machine that hydrolyzes ATP and translocates protons across the membrane (we'll see this again in the next section). With the F_0F_1-ATPase, the ATP produced from fermentation could now allow for the cell to maintain pH homeostasis by coupling the free energy of hydrolysis of ATP to the transport of protons out of the cell. The downside is that cells are now pumping all of these protons into the environment, which will now start to acidify.

Oxidation of Pyruvate and the TCA Cycle



Overview of Pyruvate Metabolism and the TCA Cycle

Under appropriate conditions, pyruvate can be further oxidized. One of the most studied oxidation reactions involving pyruvate is a two-part reaction involving NAD^+ and a molecule called co-enzyme A, often abbreviated as "CoA". This reaction oxidizes pyruvate, leads to a loss of one carbon via decarboxylation, and creates a new molecule called acetyl-CoA. The resulting acetyl-CoA can enter several pathways for the biosynthesis of larger molecules or it can flow into another pathway of central metabolism called the Citric Acid Cycle, sometimes also called the Krebs Cycle, or Tricarboxylic Acid (TCA) Cycle. Here, the remaining two carbons in the acetyl group can either be further oxidized or serve again as precursors for the construction of various other molecules. We discuss these scenarios below.

The different fates of pyruvate and other end products of glycolysis

The glycolysis module left off with the end-products of glycolysis: 2 pyruvate molecules, 2 ATPs and 2 NADH molecules. This module and the module on fermentation explore what the cell can do with the pyruvate, ATP and NADH that were generated.

The fates of ATP and NADH

ATP can be used for or coupled to a variety of cellular functions including biosynthesis, transport, replication, etc. We will see many such examples throughout the course.

What to do with the NADH however, depends on the conditions under which the cell is growing. Sometimes, the cell will opt to recycle NADH rapidly back into NAD^+. This occurs through a process called fermentation. This process returnsthe electrons initially taken from the glucose derivatives to more downstream products through another red/ox transfer (described in more detail in the module on fermentation). Alternatively, NADH can recycle back into NAD^+ by donating electrons to something known as an electron transport chain (we cover this in the module on respiration and electron transport).

The fate of cellular pyruvate

- Pyruvate can be a terminal electron acceptor (either directly or indirectly) in fermentation reactions and we discuss this in the fermentation module.
- Pyruvate can be secreted from the cell as a waste product.
- Pyruvate can be further oxidized to extract more free energy from this fuel.
- Pyruvate can serve as a valuable intermediate compound linking some core carbon processing metabolic pathways

The further oxidation of pyruvate

In respiring bacteria and archaea, the pyruvate is further oxidized in the cytoplasm. In aerobically respiring eukaryotic cells, cells transport the pyruvate molecules produced at the end of glycolysis into mitochondria. These sites of cellular respiration house oxygen consuming electron transport chains (ETC in the module on respiration and electron transport). Organisms from all three domains of life share similar mechanisms to further oxidize the pyruvate to CO_2. First pyruvate is decarboxylated and covalently linked to **co-enzyme A** via a **thioester** linkage to form the molecule known as **acetyl-CoA**. While acetyl-CoA can feed into multiple other biochemical pathways, we now consider its role in feeding the circular pathway known as the **Tricarboxylic Acid Cycle**, also referred to as the **TCA cycle**, the **Citric Acid Cycle** or the **Krebs Cycle**. We detail this process below.

Conversion of Pyruvate into Acetyl-CoA

In a multi-step reaction catalyzed by the enzyme pyruvate dehydrogenase, pyruvate is oxidized by NAD^+, decarboxylated, and covalently linked to a molecule of co-enzyme A via a **thioester bond**. The release of the carbon dioxide is important here, this reaction often results in a **loss of mass from the cell** as the CO_2 will diffuse or be transported out of the cell and become a waste product. In addition, cells reduce one molecule of NAD^+ to NADH during this process per molecule of pyruvate oxidized. Remember: there are *two* pyruvate molecules produced at the end of glycolysis for every molecule of glucose metabolized; thus, if both pyruvate molecules are oxidized to acetyl-CoA two of the original six carbons will have converted to waste.

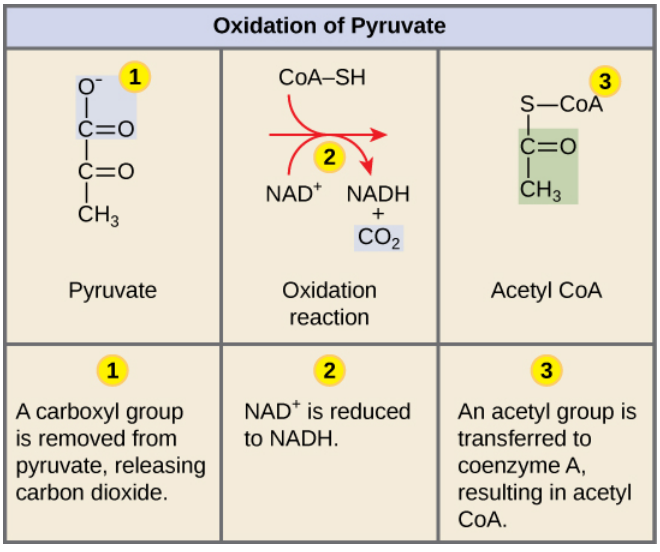


Figure 1. *Upon entering the mitochondrial matrix, a multi-enzyme complex converts pyruvate into acetyl CoA. In the process, carbon dioxide is released and one molecule of NADH is formed.*

In the presence of a suitable ***terminal electron acceptor,*** acetyl CoA delivers (exchanges a bond) its acetyl group to a four-carbon molecule, oxaloacetate, to form citrate (designated the first compound in the cycle). This cycle is called by different names: the **citric acid cycle** (for the first intermediate formed—citric acid, or citrate), the **TCA cycle** (since citric acid or citrate and isocitrate are tricarboxylic acids), and the **Krebs cycle**, after Hans Krebs, who first identified the steps in the pathway in the 1930s in pigeon flight muscles.

The Tricarboxcylic Acid (TCA) Cycle

In bacteria and archaea reactions in the TCA cycle typically happen in the cytosol. In eukaryotes, the TCA cycle takes place in the matrix of mitochondria. Almost all (but not all) of the enzymes of the TCA cycle are water soluble (not in the membrane), with the single exception of the enzyme succinate dehydrogenase, which is embedded in the inner membrane of the mitochondrion (in eukaryotes). Unlike glycolysis, the TCA cycle is a closed loop: the last part of the pathway regenerates the compound used in the first step. The eight steps of the cycle are a series of red/ox, dehydration, hydration, and decarboxylation reactions that produce two carbon dioxide molecules, one ATP, and reduced forms of NADH and $FADH_2$.

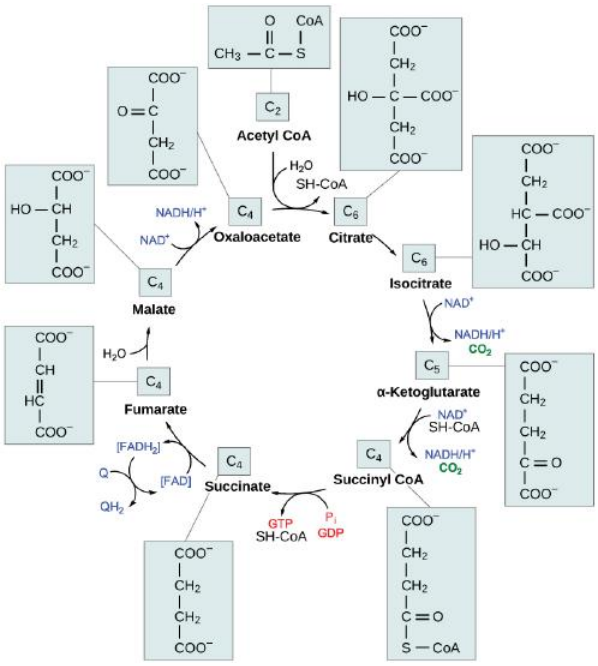


Figure 2. *In the TCA cycle, the acetyl group from acetyl CoA attaches to a four-carbon oxaloacetate molecule to form a six-carbon citrate molecule. Through a series of steps, citrate is oxidized, releasing two carbon dioxide molecules for each acetyl group fed into the cycle. In the process, three NAD^+ molecules are reduced to NADH, one FAD^+ molecule is reduced to $FADH_2$, and one ATP or GTP (depending on the cell type) is produced (by substrate-level phosphorylation). Because the final product of the TCA cycle is also the first reactant, the cycle runs continuously in the presence of sufficient reactants. Attribution: "Yikrazuul"/Wikimedia Commons (modified)*

NOTE:

We are explicitly referring to eukaryotes, bacteria and archaea when we discuss the location of the TCA cycle because many beginning students of biology only associate the TCA cycle with mitochondria. Yes, the TCA cycle occurs in the mitochondria of eukaryotic cells. However, this pathway is not exclusive to eukaryotes; it occurs in bacteria and archaea too!

Steps in the TCA Cycle

Step 1:

The first step of the cycle is a condensation reaction involving the two-carbon acetyl group of acetyl-CoA with one four-carbon molecule of oxaloacetate. The products of this reaction are the six-carbon molecule citrate and free co-enzyme A. This step is considered irreversible because it is so highly exergonic. ATP concentration controls the rate of this reaction through negative feedback. If ATP levels increase, the rate of this reaction decreases. If ATP is in short supply, the rate increases. If not already, the reason will become clear shortly.

Step 2:

In step two, citrate loses one water molecule and gains another as citrate converts into its isomer, isocitrate.

Step 3:

In step three, isocitrate is oxidized by NAD^+ and decarboxylated. Keep track of the carbons! This carbon now more than likely leaves the cell as waste and is no longer available for building new biomolecules. The oxidation of isocitrate therefore produces a five-carbon molecule, α-ketoglutarate, a molecule of CO_2 and NADH. This step is also regulated by negative feedback from ATP and NADH, and via positive feedback from ADP.

Step 4:

Step 4 is catalyzed by the enzyme succinate dehydrogenase. Here, α-ketoglutarate is further oxidized by NAD^+. This oxidation again leads to a decarboxylation and thus the loss of another carbon as waste. **So far two carbons have come into the cycle from acetyl-CoA and two have left as CO_2.** At this stage, there is no net gain of carbons assimilated from the glucose molecules that are oxidized to this stage of metabolism. Unlike the previous step however succinate dehydrogenase - like pyruvate dehydrogenase before it - couples the free energy of the exergonic red/ox and decarboxylation reaction to drive the formation of a thioester bond between the substrate co-enzyme A and succinate (what is left after the decarboxylation). Succinate dehydrogenase is regulated by feedback inhibition of ATP, succinyl-CoA, and NADH.

Possible NB Discussion Point

We have seen several steps in this and other pathways that are regulated by allosteric feedback mechanisms. Why is it so important to be able to regulate cellular processes in the context of metabolism? Is there something(s) common to these regulated steps in the TCA cycle? Why might these steps be good steps to regulate in particular?

Step 5:

In step five, a substrate level phosphorylation event occurs. Here an inorganic phosphate (P_i) is added to GDP or ADP to form GTP (an ATP equivalent for our purposes) or ATP. The energy that drives this substrate level phosphorylation event comes from the hydrolysis of the CoA molecule from succinyl~CoA to form succinate. Why is GTP or ATP produced? In animal cells there are two isoenzymes (different forms of an enzyme that carries out the same reaction), for this step, depending upon the type of animal tissue in which we find those cells. We find one isozyme in tissues that use large amounts of ATP, such as heart and skeletal muscle. This isozyme produces ATP. We find the second isozyme of the enzyme in tissues that have many anabolic pathways, such as liver. This isozyme produces GTP. GTP is energetically equivalent to ATP; However, its use is more restricted. In particular, the process of protein synthesis primarily uses GTP. Most bacterial systems produce GTP in this reaction.

Step 6:

Step six is another red/ox reactions in which succinate is oxidized by FAD^+ into fumarate. Two hydrogen atoms are transferred to FAD^+, producing $FADH_2$. The difference in reduction potential between the fumarate/succinate and NAD^+/NADH half reactions does not make NAD^+ a suitable reagent for oxidizing succinate with NAD^+ under cellular conditions. However, the difference in reduction potential with the $FAD^+/FADH_2$ half reaction is adequate to oxidize succinate and reduce FAD^+. Unlike NAD^+, FAD^+ remains attached to the enzyme and transfers electrons to the electron transport chain directly. A cell makes this process possible by

localizing the enzyme catalyzing this step inside the inner membrane of the mitochondrion or plasma membrane (depending on whether or not the organism in question is eukaryotic).

Step 7:

Water is added to fumarate during step seven, and malate is produced. The last step in the citric acid cycle regenerates oxaloacetate by oxidizing malate with NAD^+. Another molecule of NADH is produced in the process.

Summary

Note that this process (oxidation of pyruvate to Acetyl-CoA followed by one "turn" of the TCA cycle) completely oxidizes 1 molecule of pyruvate, a 3 carbon organic acid, to 3 molecules of CO_2. Overall, 4 molecules of NADH, 1 molecule of $FADH_2$, and 1 molecule of GTP (or ATP) are also produced. For respiring organisms this is a significant mode of energy extraction, since each molecule of NADH and FAD_2 can feed directly into the electron transport chain, and as we will soon see, the subsequent red/ox reactions that are driven by this process will indirectly power the synthesis of ATP. The discussion so far suggests that the TCA cycle is primarily an energy extracting pathway; evolved to extract or convert as much potential energy from organic molecules to a form that cells can use, ATP (or the equivalent) or an energized membrane. However - and let us not forget - the other important outcome of evolving this pathway is the ability to produce several precursors or substrate molecules necessary for various catabolic reactions (this pathway provides some early building blocks to make bigger molecules). As we will discuss below, there is a strong link between carbon metabolism and energy metabolism.

Connections to Carbon Flow

One hypothesis that we have explored in this reading and in class is the idea that "central metabolism" evolved to generate carbon precursors for catabolic reactions. Our hypothesis also states that as cells evolved, these reactions became linked into pathways: glycolysis and the TCA cycle, to maximize their effectiveness for the cell. We can postulate that a side benefit to evolving this metabolic pathway was the generation of NADH from the complete oxidation of glucose - we saw the beginning of this idea when we discussed fermentation. We have already discussed how glycolysis not only provides ATP from substrate level phosphorylation but also yields a net of 2 NADH molecules and 6 essential precursors: glucose-6-P, fructose-6-P, 3-phosphoglycerate, phosphoenolpyruvate, and pyruvate. While ATP can be used by the cell directly as an energy source, NADH poses a problem and must be recycled back into NAD^+, to keep the pathway in balance. As we see in the fermentation module, the most ancient way cells deal with this problem is to use fermentation reactions to regenerate NAD^+.

During the process of pyruvate oxidation via the TCA cycle, 4 additional essential precursors are formed: acetyl~CoA, α-ketoglutarate, oxaloacetate, and succinyl~CoA. Three molecules of CO_2 are lost, and this represents a net loss of mass for the cell. These precursors, however, are substrates for a variety of catabolic reactions including the production of amino acids, fatty acids, and various co-factors, such as heme. This means that the rate of reactions through the TCA cycle will be sensitive to the concentrations of each **metabolic intermediate** (more on the thermodynamics in class). A metabolic intermediate is a compound that is produced by one reaction (a product) and then acts as a substrate for the next reaction. This also means that metabolic intermediates, in particular the 4 essential precursors, can be removed for catabolic reactions, if there is a demand, changing the thermodynamics of the cycle.

Possible NB Discussion Point

Some of the TCA cycle intermediates, particularly glutamate and succinyl-CoA, are diverted away from the TCA cycle to other reactions. Unless something is done about it, this can leave the TCA cycle with too few intermediates to function effectively. Either through your own imagination or external study (hint: search "anapleurotic reactions" with a search engine) describe how this problem might be solved. What is required for this to happen? As a follow-up, you can also try to explain what you might need to run the TCA cycle backwards. Many organisms need to do this. What is required for this to happen and under what circumstances might it be helpful? Search "reductive Krebs Cycle" for leads.

Not all cells have a complete TCA cycle.

Since all cells require the ability of make these precursor molecules, one might expect that all organisms would have a fully complete TCA cycle. In fact, the cells of many organisms DO NOT have all the enzymes required to form a complete cycle - all cells, however, DO have the capability of making the 4 TCA cycle precursors noted in the previous paragraph. How can the cells make precursors and not have a full cycle? Remember that most of these reactions are freely reversible, so, if NAD^+ is required to for the oxidation of pyruvate or acetyl~CoA, then the reverse reactions would require NADH. We often refer to this process as the reductive TCA cycle. To drive these reactions in reverse (regarding the direction discussed above) requires energy, in this case carried by ATP and NADH. If you get ATP and NADH driving a pathway one direction, driving it in reverse will require ATP and NADH as "inputs". So, organisms that do not have a full cycle can still make the 4 key metabolic precursors by using previously extracted energy and electrons (ATP and NADH) to drive some key steps in reverse.

Additional Links

Here are some additional links to videos and pages that you may find useful.

Chemwiki Links

- Chemwiki TCA cycle - link down until key content corrections are made to the resource

Khan Academy Links

- Khan Academy TCA cycle - link down until key content corrections are made to the resource

Reading_12 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_13

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_13

ME.39 Create a conceptual diagram linking various modes of ATP synthesis with diverse uses of ATP in the cell that highlights the concepts of an "ATP pool" and the role of ATP as a short-term energy carrier.
ME.33 Explain the thermodynamic principle underlying the sequential ordering of electron carriers in a functional electron transport chain (ETC).
ME.34 Describe the molecular make-up of the components of typical electron transport chains. What are they made of, and how do you expect them to behave and interact with the molecules they encounter?
ME.32 Given a series of electron carriers and their associated reduction potentials, be able to generate from scratch a putatively functional electron transport chain (ETC).
ME.35 Define and correctly use the term "terminal electron acceptor" and create a list of at least three criteria that apply to most "good" terminal electron acceptors.
ME.42 Explain in detail how an electron transport chain (ETC) can harvest usable "cellular energy" by coupling a series of sequential exergonic redox transfers to the active transport of protons across the membrane from low to high concentration (endergonic work).
ME.37 Create an energy story describing the enzyme catalyzed active "pumping" of protons across the membrane (remember, track both mass and energy transfers).
ME.38 Compare and contrast the roles of protons and electrons in the electron transport chain (ETC).
ME.36 Justify the proton gradient derived from ETC activity as a critical component of energy transfer.
ME.40 Compare and contrast the three forms of ATP production - substrate-level phosphorylation, oxidative phosphorylation and photophosphorylation.
GC.51 Create a concept map explicitly linking all of the following course topics and the concept of redox chemistry: enzymes, thermodynamics, coupled reactions, protein chemistry, and electron transport.

Introduction to Respiration and Electron Transport Chains

General Overview and Points to Keep In Mind

In the next few modules, we learn about the process of respiration and the roles that electron transport chains play in this process. A definition of the word "respiration" that most people are familiar with is "the act of breathing". When we breathe, we bring air, including molecular oxygen, into our lungs from outside of the body. The oxygen then becomes reduced, and waste products, including the reduced oxygen in the form of water, are exhaled. More generically, some reactant comes into the organism and then gets reduced and leaves the body as a waste product.

This generic idea, in a nutshell, can be applied across biology. Note that oxygen need not always be the compound that is brought in, reduced, and dumped as waste. The compounds onto which the electrons that are "dumped" are known as "**terminal electron acceptors**." The molecules from which the electrons originate vary across biology (so far, we have only looked at one source - the reduced carbon-based molecule glucose).
In between the original electron source and the terminal electron acceptor are a series of biochemical reactions involving at least one redox reaction. These redox reactions harvest energy for the cell by coupling an exergonic redox reaction to an energy-requiring reaction in the cell. In respiration, a special set of enzymes carry out a linked series of redox reactions that ultimately transfer electrons to the terminal electron acceptor.

These "chains" of redox enzymes and electron carriers are called **electron transport chains** (ETC). In aerobically respiring eukaryotic cells the ETC is composed of four large, multi-protein complexes embedded in the inner mitochondrial membrane and two small diffusible electron carriers shuttling electrons between them. Electrons pass from enzyme to enzyme through a series of redox reactions. These reactions couple exergonic redox reactions to the endergonic transport of hydrogen ions across the inner mitochondrial membrane. This process contributes to the creation of a transmembrane electrochemical gradient. The electrons passing through the ETC gradually lose potential energy until the point the ETC deposits them on the terminal electron acceptor. The cell typically disposes of the reduced terminal electron as waste. When oxygen acts as the final electron acceptor, the free

energy difference of this multi-step redox process is ~ -60 kcal/mol when NADH donates electrons or ~ -45 kcal/mol when $FADH_2$ donates.

Note: Oxygen is not the only terminal electron acceptor in nature

Recall, that we use oxygen as an example of only one of many possible terminal electron acceptors that can be found in nature. The free energy differences associated with respiration in anaerobic organisms will be different.

In prior modules we discussed the general concept of redox reactions in biology and introduced the Electron Tower, a tool to help you understand redox chemistry and to estimate the direction and magnitude of potential energy differences for various redox couples. In later modules, substrate level phosphorylation and fermentation were discussed, and we saw that enzymes could directly couple exergonic redox reactions to the endergonic synthesis of ATP.

We hypothesize these processes to be one of the oldest forms of energy production used by cells. In this section, we discuss the next evolutionary advancement in cellular energy metabolism, oxidative phosphorylation. Foremost recall that oxidative phosphorylation *does not* imply the use of oxygen. Rather, the term oxidative phosphorylation is used because this process of ATP synthesis relies on redox reactions to generate an electrochemical *transmembrane potential* the cell can then use to do the work of ATP synthesis.

A Quick Overview of Principles Relevant to Electron Transport Chains

An ETC begins with the addition of electrons, donated from NADH, $FADH_2$ or other reduced compounds. These electrons move through a series of electron transporters, enzymes that are embedded in a membrane, or other carriers that undergo redox reactions. The free energy transferred from these exergonic redox reactions is often coupled to the endergonic movement of protons across a membrane. Since the membrane is an effective barrier to charged species, this pumping results in an unequal accumulation of protons on either side of the membrane. This "polarizes" or "charges" the membrane, with a net positive (protons) on one side of the membrane and a negative charge on the other side of the membrane. The separation of charge creates an ***electrical potential***. In addition, the accumulation of protons also causes a pH gradient known as a ***chemical potential*** across the membrane. Together these two gradients (electrical and chemical) are called an ***electro-chemical gradient***.

Review: The Electron Tower

Since redox chemistry is so central to the topic we begin with a quick review of the table of reduction potential - sometimes called the "redox tower" or "electron tower". You may hear your instructors use these terms interchangeably. As we discussed in previous modules, all kinds of compounds can take part in biological redox reactions. Making sense of all of this information and ranking potential redox pairs can be confusing. We have developed a tool to rate redox half reactions based on their reduction potentials or E_0' values. Whether a particular compound can act as an electron donor (reductant) or electron acceptor (oxidant) depends on what other compound it is interacting with. The redox tower ranks a variety of common compounds (their half reactions) from most negative E_0', compounds that readily get rid of electrons, to the most positive E_0', compounds most likely to accept electrons. The tower organizes these half reactions based on the ability of electrons to accept electrons. In addition, in many redox towers, each half reaction is written by convention with the oxidized form on the left followed by the reduced form to its right. The two forms may be either separated by a slash, for example, we write the half reaction for the reduction of NAD^+ to NADH: $NAD^+/NADH + 2e^-$, or by separate columns. We show an electron tower below.

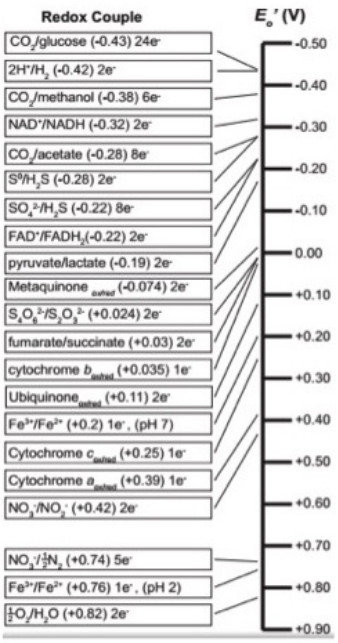


Figure 1. *A common biological "redox tower"*

Note

Use the redox tower above as a reference guide to orient you as to the reduction potential of the various compounds in the ETC. Redox reactions may be either exergonic or endergonic depending on the relative redox potentials of the donor and acceptor. Also remember there are many different ways of looking at this conceptually; this type of redox tower is just one way.

Note: **Language shortcuts reappear**

In the redox table above some entries seem to be written in unconventional ways. For instance Cytochrome $c_{ox/red}$. There only appears to be one form listed. Why? This is another example of language shortcuts (likely because someone was too lazy to write cytochrome twice) that can confuse - particularly to students. We could rewrite the notation above as Cytochrome c_{ox}/Cytochrome c_{red} to show that the cytochrome c protein can exist in either and oxidized state Cytochrome c_{ox} or reduced state Cytochrome c_{red}.

Review Redox Tower Video

For a short video on how to use the redox tower in redox problems click here. This video was made by Dr. Easlon for Bis2A students.



Using the redox tower: A tool to help understand electron transport chains

By convention, we write the tower half reactions with the oxidized form of the compound on the left and the reduced form on the right. Notice that compounds such as glucose and hydrogen gas are excellent electron donors and have very low reduction potentials $E^{0'}$. Compounds, such as oxygen and nitrite, whose half reactions have relatively high positive reduction potentials ($E^{0'}$) make we find good electron acceptors at the opposite end of the table.

Example: **Menaquinone**

Let's look at menaquinone$_{ox/red}$. This compound sits in the middle of the redox tower with a half-reaction $E^{0'}$ value of -0.074 eV. Menaquinone$_{ox}$ can spontaneously ($\Delta G<0$) accept electrons from reduced forms of compounds with lower half-reaction $E^{0'}$. Such transfers form menaquinone$_{red}$ and the oxidized form of the original electron donor. In the table above, examples of compounds that could act as electron donors to menaquinone include $FADH_2$, an $E^{0'}$ value of -0.22, or NADH, with an $E^{0'}$ value of -0.32 eV. Remember, the reduced forms are on the right-hand side of the redox pair.

Once menaquinone has been reduced, it can now spontaneously ($\Delta G<0$) donate electrons to any compound with a higher half-reaction $E^{0'}$ value. Electron acceptors include cytochrome b$_{ox}$ with an $E^{0'}$ value of 0.035 eV; or ubiquinone$_{ox}$ with an $E^{0'}$ of 0.11 eV. Remember that the oxidized forms lie on the left side of the half reaction.

Electron Transport Chains



An **electron transport chain**, or **ETC**, is composed of a group of protein complexes in and around a membrane that help energetically couple a series of exergonic/spontaneous red/ox reactions to the endergonic pumping of protons across the membrane to generate an electrochemical gradient. This electrochemical gradient creates a free energy potential that we call a **proton motive force** whose energetically "downhill" exergonic flow can later be coupled to a variety of cellular processes.

ETC overview

Step 1: Electrons enter the ETC from an electron donor, such as NADH or $FADH_2$, which are generated during a variety of catabolic reactions, including those associated glucose oxidation. Depending on the number and types of electron carriers of the ETC being used by an organism, electrons can enter at a variety of places in the electron transport chain. Entry of electrons at a specific "spot" in the ETC depends upon the respective reduction potentials of the electron donors and acceptors.

Step 2: After the first red/ox reaction, the initial electron donor will become oxidized and the electron acceptor will become reduced. The difference in red/ox potential between the electron acceptor and donor is related to ΔG by the relationship $\Delta G = -nF\Delta E$, where n = the number of electrons transferred and F = Faraday's constant. The larger a positive ΔE, the more exergonic the red/ox reaction is.

Step 3: If sufficient energy is transferred during an exergonic red/ox step, the electron carrier may couple this negative change in free energy to the endergonic process of transporting a proton from one side of the membrane to the other.

Step 4: After usually multiple red/ox transfers, the electron is delivered to a molecule known as the terminal electron acceptor. With humans, the terminal electron acceptor is oxygen. However, there are many, many, many other possible electron acceptors in nature; see below.

Note: NADH AND $FADH_2$ ARE NOT THE ONLY ELECTRON DONORS

Electrons entering the ETC do not have to come from NADH or $FADH_2$. Many other compounds can serve as electron donors; the only requirements are (1) that there is an enzyme that can oxidize the electron donor and then reduce another compound, and (2) that the $\Delta E_0'$ is positive (e.g., $\Delta G<0$). Even a small amount of free energy transfers can add up. For example, there are bacteria that use H_2 as an electron donor. This is not too difficult to believe because the half reaction $2H^+ + 2\ e^-/H_2$ has a reduction potential (E_0') of -0.42 V. If these electrons are eventually delivered to oxygen, then the $\Delta E_0'$ of the reaction is 1.24 V, which corresponds to a large negative ΔG ($-\Delta G$). Alternatively, there are some bacteria that can oxidize iron, Fe^{2+} at pH 7 to Fe^{3+} with a reduction potential (E_0') of + 0.2 V. These bacteria use oxygen as their terminal electron acceptor, and, in this case, the $\Delta E_0'$ of the reaction is approximately 0.62 V. This still produces a $-\Delta G$. The bottom line is that, depending on the electron donor and acceptor that the organism uses, a little or a lot of energy can be transferred and used by the cell per electrons donated to the electron transport chain.

What are the complexes of the ETC?

ETCs comprise a series (at least one) of membrane-associated red/ox proteins or (some are integral) protein complexes (complex = more than one protein arranged in a quaternary structure) that move electrons from a donor source, such as NADH, to a final terminal electron acceptor, such as oxygen. This specific donor/terminal acceptor pair is the primary one used in human mitochondria. Each electron transfer in the ETC requires a reduced substrate as an electron donor and an oxidized substrate as the electron acceptor. In most cases, the electron acceptor is a member of the enzyme complex itself. Once the complex is reduced, the complex can serve as an electron donor for the next reaction.

How do ETC complexes transfer electrons?

As previously mentioned, the ETC is composed of a series of protein complexes that undergo a series of linked red/ox reactions. These complexes are in fact multi-protein enzyme complexes referred to as **oxidoreductases** or simply, **reductases**. The one exception to this naming convention is the terminal complex in aerobic respiration that uses molecular oxygen as the terminal electron acceptor. That enzyme complex is referred to as an **oxidase**. Red/ox reactions in these complexes are typically carried out by a non-protein moiety called a **prosthetic group**. The prosthetic groups are directly involved in the red/ox reactions being catalyzed by their associated oxidoreductases. In general, these prosthetic groups can be divided into two general types: those that carry both electrons and protons and those that only carry electrons.

Note:

This use of prosthetic groups by members of ETC is true for all of the electron carriers with the exception of quinones, which are a class of lipids that can directly be reduced or oxidized by the oxidoreductases. Both the $Quinone_{(red)}$ and the $Quinone_{(ox)}$ forms of these lipids are soluble within the membrane and can move from complex to complex to shuttle electrons.

The electron and proton carriers

- **Flavoproteins (Fp)**, these proteins contain an organic prosthetic group called a **flavin**, which is the actual moiety that undergoes the oxidation/reduction reaction. $FADH_2$ is an example of an Fp.
- **Quinones** are a family of lipids, which means they are soluble within the membrane.
- We also note that we consider NADH and NADPH electron ($2e^-$) and proton ($2\ H^+$) carriers.

Electron carriers

- **Cytochromes** are proteins that contain a heme prosthetic group. The heme can carry a single electron.
- **Iron-Sulfur proteins** contain a nonheme iron-sulfur cluster that can carry an electron. We often abbreviate the prosthetic group as **Fe-S**

Aerobic versus anaerobic respiration

We humans use oxygen as the **terminal electron acceptor** for the ETCs in our cells. This is also the case for many of the organisms we intentionally and frequently interact with (e.g. our classmates, pets, food animals, etc). We breathe in oxygen; Our cells take it up and transport it into the mitochondria, where it becomes the final acceptor of electrons from our electron transport chains. We call the process where oxygen is the terminal electron acceptor **aerobic respiration**.

While we may use oxygen as the terminal electron acceptor for our respiratory chains, this is not the only mode of respiration on the planet. The more general processes of respiration evolved when oxygen was not a major component of the atmosphere. As a result, many organisms can use a variety of compounds, including nitrate (NO_3^-), nitrite (NO_2^-), even iron (Fe_3^+) as terminal electron acceptors. When oxygen is **NOT** the terminal electron acceptor, we refer the process to as **anaerobic respiration**. Therefore, respiration or oxidative phosphorylation does not require oxygen at all; It requires a compound with a high enough reduction potential to act as a terminal electron acceptor, accepting electrons from one complex within the ETC.

The ability of some organisms to vary their terminal electron acceptor provides metabolic flexibility and can ensure better survival if any given terminal acceptor is in limited supply. Think about this: in the absence of oxygen, we die; but other organisms can use a different terminal electron acceptor when conditions change to survive.

Possible NB Discussion Point

Nature has figured out how to use different molecules as terminal electron acceptors of ETCs. Yet humans seem limited to using only oxygen. Can you offer any hypotheses why humans have not evolved to use multiple different terminal electron acceptors? Why do you think it might be advantageous for an organism to use oxygen as a sole terminal electron acceptor?

A generic example: A simple, two-complex ETC

The figure below depicts a generic electron transport chain, composed of two integral membrane complexes; Complex $I_{(ox)}$ and Complex $II_{(ox)}$. A reduced electron donor, designated DH (such as NADH or $FADH_2$) reduces Complex $I_{(ox)}$, giving rise to the oxidized form D (such as NAD^+ or FAD^+). Simultaneously, a prosthetic group within Complex I is now reduced (accepts the electrons). In this example, the red/ox reaction is exergonic and the free energy difference is coupled by the enzymes in Complex I to the endergonic translocation of a proton from one side of the membrane to the other. The net result is that one surface of the membrane becomes more negatively charged, because of an excess of hydroxyl ions (OH^-), and the other side becomes positively charged because of an increase in protons on the other side. Complex $I_{(red)}$ can now reduce a mobile electron carrier Q, which will then move through the membrane and transfer the electron(s) to the prosthetic group of Complex $II_{(red)}$. Electrons pass from Complex I to Q then from Q to Complex II via thermodynamically spontaneous red/ox reactions, regenerating Complex $I_{(ox)}$, which can repeat the previous process. Complex $II_{(red)}$ then reduces A, the terminal electron acceptor to regenerate Complex $II_{(ox)}$ and create the reduced form of the terminal electron acceptor, AH. In this specific example, Complex II can also translocate a proton during the process. If A is molecular oxygen, AH represents water and the process would be considered being a model of an aerobic ETC. If A is nitrate, NO_3^-, then AH represents NO_2^- (nitrite) and this would be an example of an anaerobic ETC.

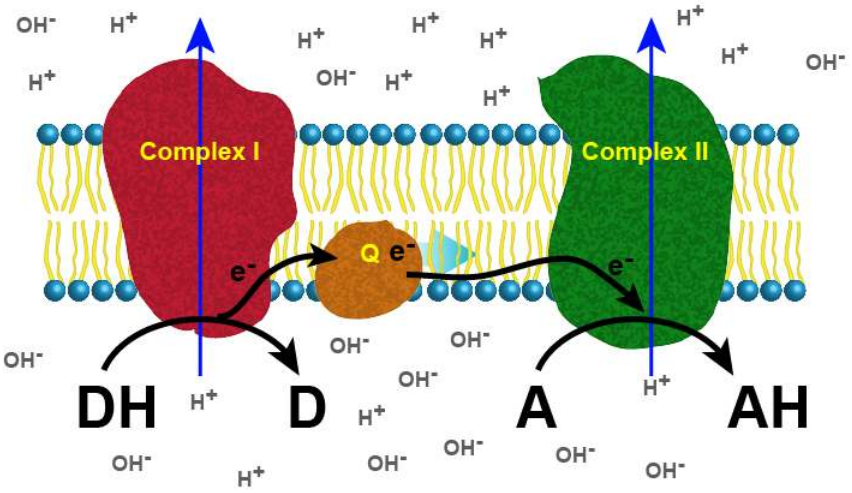


Figure 1. *Generic 2 complex electron transport chain. In the figure, DH is the electron donor (donor reduced), and D is the donor oxidized. A is the oxidized terminal electron acceptor, and AH is the final product, the reduced form of the acceptor. As DH is oxidized to D, protons are translocated across the membrane, leaving an excess of hydroxyl ions (negatively charged) on one side of the membrane and protons (positively charged) on the other side of the membrane. The same reaction occurs in Complex II as the terminal electron acceptor is reduced to AH. Attribution: Marc T. Facciotti (original work)*

Detailed look at aerobic respiration

The eukaryotic mitochondria have evolved a very efficient ETC. There are four complexes composed of proteins, labeled I through IV depicted in the figure below. The aggregation of these four complexes, together with associated mobile, accessory electron carriers, is called an electron transport chain. This electron transport chain is present in multiple copies in the inner mitochondrial membrane of eukaryotes.

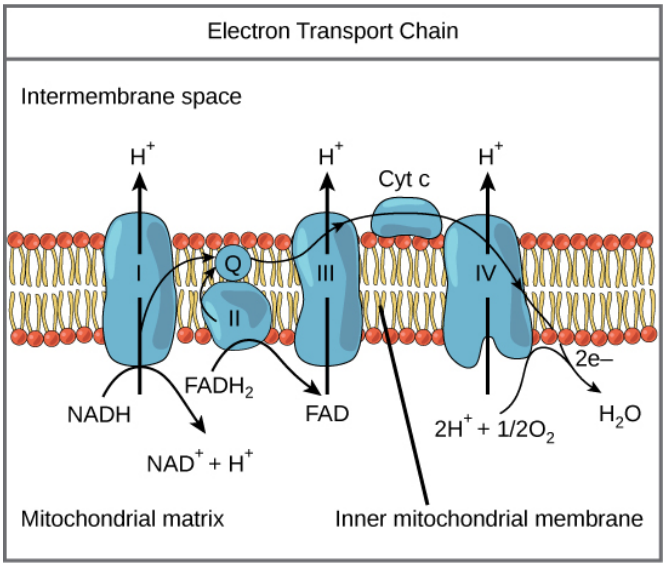


Figure 2. *The electron transport chain is a series of electron transporters embedded in the inner mitochondrial membrane that shuttles electrons from NADH and $FADH_2$ to molecular oxygen. In the process, protons are pumped from the mitochondrial matrix to the intermembrane space, and oxygen is reduced to form water.*

Complex I

To start, NADH delivers two electrons to the first protein complex. This complex, labeled I in Figure 2, includes flavin mononucleotide (FMN) and iron-sulfur (Fe-S)-containing proteins. FMN, which is derived from vitamin B_2, also called riboflavin, is one of several prosthetic groups or cofactors in the electron transport chain. Prosthetic groups are organic or inorganic, nonpeptide molecules bound to a protein that facilitate its function; prosthetic groups include coenzymes, which are the prosthetic groups of enzymes. We also call the enzyme in Complex I NADH dehydrogenase. This protein complex contains 45 individual polypeptide chains. Complex I can pump four hydrogen ions across the membrane from the matrix into the intermembrane space helping to generate and maintain a hydrogen ion gradient between the two compartments separated by the inner mitochondrial membrane.

Q and Complex II

Complex II directly receives $FADH_2$, which does not pass through Complex I. The compound connecting the first and second complexes to the third is ubiquinone (Q). The Q molecule is lipid soluble and freely moves through the hydrophobic core of the membrane. Once reduced, (QH_2), ubiquinone delivers its electrons to the next complex in the electron transport chain. Q receives the electrons derived from NADH from Complex I and the electrons derived from $FADH_2$ from Complex II, succinate dehydrogenase. Since these electrons bypass and thus do not energize the proton pump in the first complex, fewer ATP molecules are made from the $FADH_2$ electrons. As we will see in the following section, the number of ATP molecules ultimately obtained is directly proportional to the number of protons pumped across the inner mitochondrial membrane.

Complex III

The third complex is composed of cytochrome b, another Fe-S protein, Rieske center (2Fe-2S center), and cytochrome c proteins; we also call this complex cytochrome oxidoreductase. Cytochrome proteins have a prosthetic group of heme. The heme molecule is like the heme in hemoglobin, but it carries electrons, not oxygen. As a result, the iron ion at its core is reduced and oxidized as it passes the electrons, fluctuating between different oxidation states: Fe^{2+} (reduced) and Fe^{3+} (oxidized). The heme molecules in the cytochromes have slightly different characteristics because of the effects of the different proteins binding them, giving slightly different characteristics to each complex. Complex III pumps protons through the membrane and passes its electrons to cytochrome c for transport to the fourth complex of proteins and enzymes (cytochrome c is the acceptor of electrons from Q; however, whereas Q carries pairs of electrons, cytochrome c can accept only one at a time).

Complex IV

The fourth complex is composed of cytochrome proteins c, a, and a3. This complex contains two heme groups (one in each of the two Cytochromes, a, and a3) and three copper ions (a pair of CuA and one CuB in Cytochrome a3). The cytochromes hold an oxygen molecule tightly between the iron and copper ions until it completely reduces the oxygen. The reduced oxygen then picks up two hydrogen ions from the surrounding medium to make water (H_2O). The removal of the hydrogen ions from the system contributes to the ion gradient used in the process of chemiosmosis.

Chemiosmosis

In **chemiosmosis**, the free energy from the series of red/ox reactions just described is used to pump protons across the membrane. The uneven distribution of H^+ ions across the membrane establishes both concentration and electrical gradients (thus, an electrochemical gradient), owing to the proton's positive charge and their aggregation on one side of the membrane.

If the membrane were open to diffusion by protons, the ions would tend to diffuse back across into the matrix, driven by their electrochemical gradient. Ions, however, cannot diffuse through the nonpolar regions of phospholipid membranes without the aid of ion channels. Similarly, protons in the intermembrane space can only traverse the inner mitochondrial membrane through an integral membrane protein called ATP synthase (depicted below). This complex protein acts as a tiny generator, turned by transfer of energy mediated by protons moving down their electrochemical gradient. The movement of this molecular machine (enzyme)

serves to lower the activation energy of reaction and couples the exergonic transfer of energy associated with the movement of protons down their electrochemical gradient to the endergonic addition of a phosphate to ADP, forming ATP.

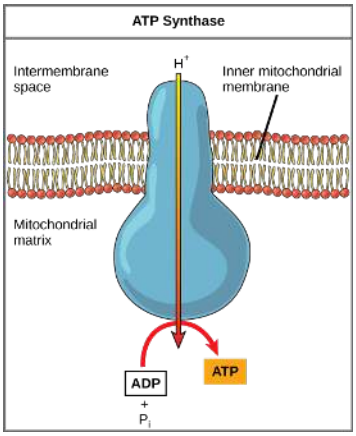


Figure 3. *ATP synthase is a complex, molecular machine that uses a proton (H^+) gradient to form ATP from ADP and inorganic phosphate (Pi). Credit: modification of work by Klaus Hoffmeier*

Possible NB Discussion Point

Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of the intermembrane space to increase or decrease? What effect would cyanide have on ATP synthesis? How would this affect the rates of reactions in glycolysis and the TCA cycle?

In healthy cells, chemiosmosis (depicted below) is used to generate 90 percent of the ATP made during aerobic glucose catabolism; it is also the method used in the light reactions of photosynthesis to harness the energy of sunlight in the process of photophosphorylation. Recall that the production of ATP using the process of chemiosmosis in mitochondria is called oxidative phosphorylation and that a similar process can occur in the membranes of bacterial and archaeal cells. The overall result of these reactions is the production of ATP from the energy of the electrons removed originally from a reduced organic molecule like glucose. In the aerobic example, these electrons ultimately reduce oxygen and create water.

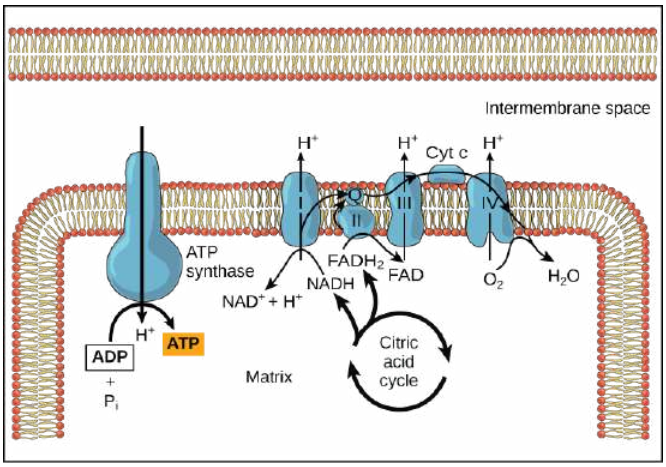


Figure 4. *In oxidative phosphorylation, the pH gradient formed by the electron transport chain is used by ATP synthase to form ATP in a Gram-bacteria.*

Helpful link: How ATP is made from ATP synthase



A Hypothesis for How ETCs May Have Evolved

A proposed link between SLP/fermentation and the evolution of ETCs:

In a previous discussion of energy metabolism, we explored substrate level phosphorylation (SLP) and fermentation reactions. While SLP and fermentation together are perfectly good ways to harvest energy, one byproduct of these reactions is the acidification of the cell. Early cells that used these modes of energy harvested therefore needed to co-evolve mechanisms that helped remove protons accumulated from SLP and fermentation from the cytosol (interior of the cell). One solution to the "proton problem" may have been the evolution of the F_0F_1-ATPase, a multi-subunit enzyme that translocated protons from the inside of the cell to the outside of the cell by hydrolyzing ATP (see the figure below). This arrangement works as long as small reduced organic molecules are abundant and freely available to generate ATP can through SLP that can "fuel" the business of the cell. However, as these biological processes continue, the small reduced organic molecules will become depleted. The resulting scarcity of fuel, therefore, puts a demand on cells to find alternative mechanisms to harness energy and/or to become more efficient.

In the scheme proposed above, one potential source of "wasted ATP" is its use in the removal of protons from the cell's cytosol; organisms that could find other mechanisms to expel accumulating protons while still preserving ATP could have a selective advantage. We hypothesize that this selective evolutionary pressure potentially led to the evolution of the first membrane-bound proteins that used red/ox reactions as their energy source (depicted in a second picture) to pump out the accumulating protons. Enzymes and enzyme complexes with these properties exist today in the as electron transport complexes like Complex I, the NADH dehydrogenase.

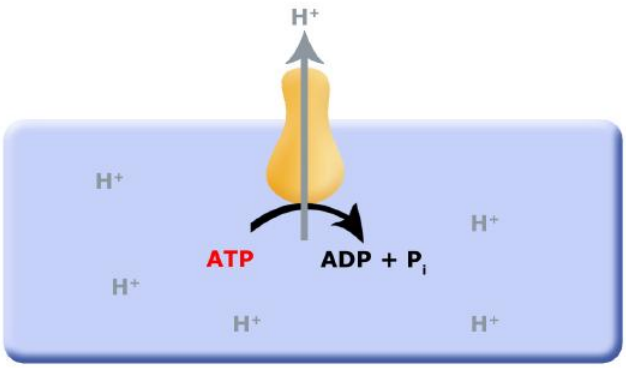


Figure 1. *Proposed evolution of an ATP dependent proton translocator*

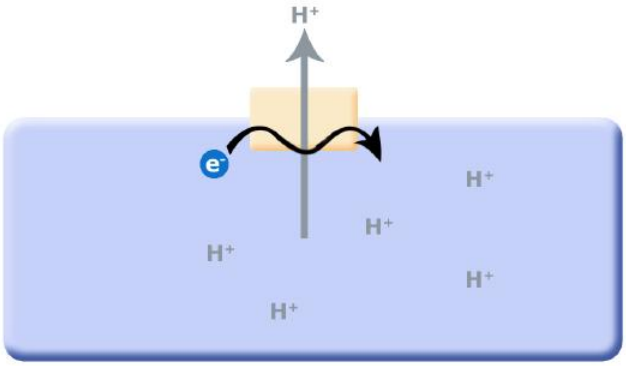


Figure 2. *As small reduced organic molecules become limited, organisms that can find alternative mechanisms to remove protons from the cytosol may have an a selective advantage. The evolution of a proton translocator that uses red/ox reactions rather than ATP hydrolysis could substitute for the ATPase.*

Continuing with this line of logic, if organisms evolved that could now use red/ox reactions to translocate protons across the membrane they would create an electrochemical gradient, separating both charge (positive on the outside and negative on the inside, creating an electrical potential) and pH (low pH outside, higher pH inside). With excess protons on the outside of the cell membrane, and the F_0F_1-ATPase no longer consuming ATP to translocate protons, we hypothesize that the electrochemical gradient could then power the F_0F_1-ATPase "backwards" — that is, to form or produce ATP by using the energy in the charge/pH gradients set up by the red/ox pumps (as depicted below). We call this arrangement an **electron transport chain (ETC)**.

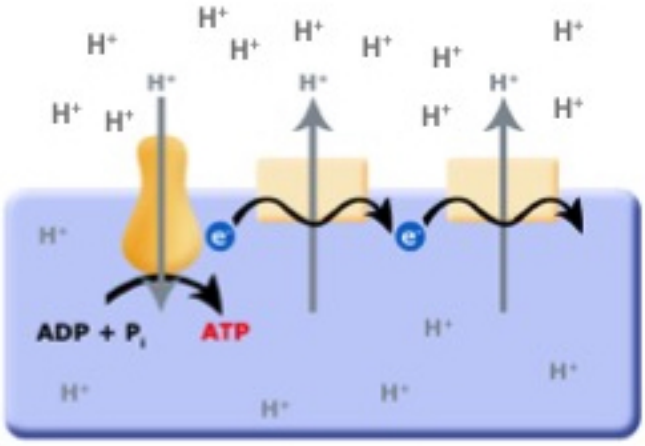


Figure 3. *The evolution of the ETC; the combination of the red/ox driven proton translocators coupled to the production of ATP by the F_0F_1-ATPase.*

Note: Extended reading on the evolution of electron transport chains

If you're interested in the the evolution of electron transport chains, check out this more in-depth discussion of the topic at NCBI.

Possible NB Discussion Point

Dinitrophenol (DNP) is a small chemical that serves to uncouple the flow of protons across the inner mitochondrial membrane to the ATP synthase, making the membrane leaky to protons. People used it until 1938 as a weight-loss drug. What effect would you expect DNP to have on the difference in pH across both sides of the inner mitochondrial membrane? Why do you think this might be an effective weight-loss drug? Why might it be dangerous? Can you think of any scenarios where it is non-harmful, or even beneficial, to uncouple proton flow with ATP synthase?

Reading_13 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_14

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_14

ME.6 Describe the difference between ΔG and ΔG°', why the former is necessary and useful to describe processes in the cell, and the significance of the former on the regulation of flow of metabolites in pathways.

ME.22 Describe the mechanism of allosteric regulation and provide examples of how this mechanism can be used to tune the flow of molecules in a metabolic pathway.

ME.23 In a given metabolic pathway, think about: why you might need regulation (ex: conservation of resources), what information will you use to regulate (ex: [metabolite]), where does it make sense to regulate (ex: branch points, essential steps with large -,ΔG), by what mechanism might you use to regulate this pathway (how), and what can/should do the regulating.

An introduction to metabolic regulation

Metabolic pathways

We can describe cellular metabolic pathways as sets of interconnected enzyme-catalyzed chemical reactions. Together these pathways form complex networks that collectively guide the flow of molecules and energy throughout a cell by breaking down, building, and rearranging the atoms and molecules of nature into their many different forms.

The collective activity of metabolic pathways allows organisms to eat food (e.g. atoms and molecules) from their environments and to convert that food at a molecular level into the specific forms of matter the organism needs to support life. Metabolic pathways also allow organisms to harvest energy from the environment and to move that energy to places in their cells that need energy to get work done.

Understanding how metabolic pathways help biological systems reorganize matter and energy is therefore fundamental to our understanding of biology from the smallest cellular scale to the functioning of connected global ecosystems and their interactions with the non-living environment.

Metabolic decision making, sensors and switches

Many living systems have complex metabolic pathways that enable them to make and process many different types of molecules. While these complex pathways likely provide an organism with selective advantage not all pathways need to be functioning at the same levels all of the time.

If, for instance, a unicellular organism is able to find and eat plenty of the amino acid tryptophan from its environment, the organism's metabolic pathways that synthesize tryptophan do not need to be highly active. Running metabolic pathways that don't need to be active wastes both the atoms that pass through the pathways and the energy used in the creation of molecules that aren't needed at a particular time. Moreover, since most living systems do not exist in resource-rich environments, atoms and energy must be carefully distributed to cellular functions that need them most, only when they are needed. Resources can then be converted into other forms when they are needed elsewhere in the cell.

To be competitive, the cell must be frugal with resources and thus control how and when different metabolic pathways are active. Controlling the flow of materials through metabolic pathways requires mechanisms for decision making - we use the anthropomorphic word "decision" to imply that the cell seemingly knows how to make choices, not to suggest that it has a brain. For example, the cell in the example above needs a way to know that tryptophan is in abundance and then to use that information to switch off the biosynthetic pathway for tryptophan biosynthesis. Concomitantly, the cell can choose to use the metabolic precursors for tryptophan biosynthesis for another purpose, activating enzymes in other pathways that can use those same precursor molecules. This kind of sensing and control is required to simultaneously coordinate the flow of biomolecules and energy throughout all of cellular metabolism. This means that there must be molecular sensors and switches for controlling metabolism distributed throughout the cell.

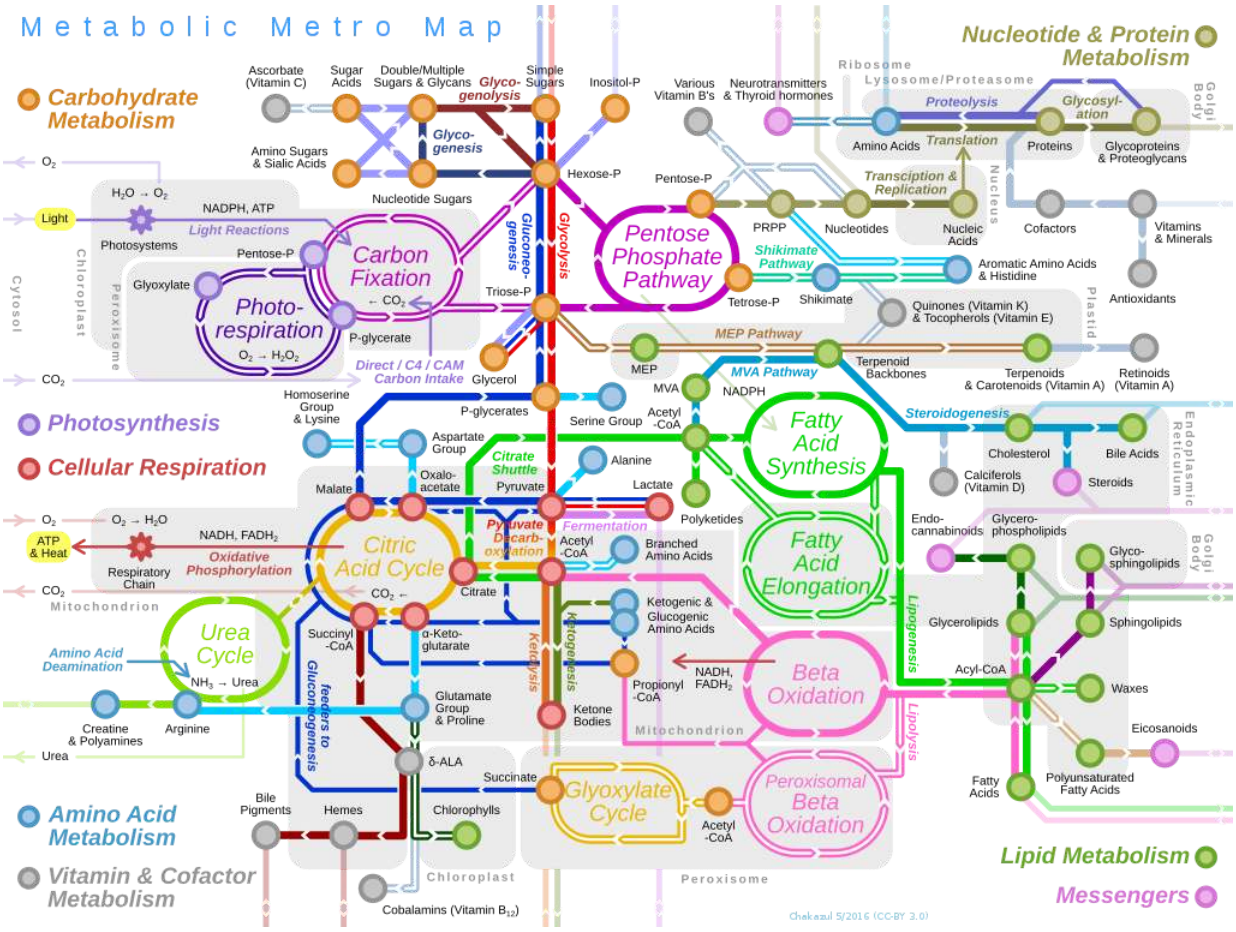


Figure 1. Overview of the major connections between core metabolic pathways in eukaryotes.

Attribution:Chakazul / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0) <https://upload.wikimedia.org/wikipedia/commons/6/6e/Metabolic_Metro_Map.svg>

Our daily lives are full of similar experiences and concepts that can inform how we think about metabolic control in biology. Our homes have all sorts of sensors and switches for managing energy. We have light switches in every room to control the flow of electricity and conversion of energy through light bulbs. Our stoves have switches on them to control energy use on each burner. Thermostats in the home sense temperature and use that information to adjust the activity of our air conditioners and/or heaters. Some of these switches are bistable, they turn things fully “on” or “off” (e.g. most light switches). Other switches are continuous (e.g. dial switches on a stove, the volume knob on the radio). They allow the user to control energy use along a gradient.

The hierarchical nature of regulatory systems

Another useful concept to draw from looking at how we manage energy in our homes is to note the hierarchical nature of the control systems we have engineered for our living spaces. Nearly all homes/apartments have circuit breakers that can help control the flow of power from an energy provider into the home. In a circuit breaker box one usually finds a master switch that can turn power on or off for the entire home. One can usually also find switches that control power to individual rooms or parts of the house (e.g. the kitchen, the bathrooms). Within each room, light switches mounted on the wall typically allow the resident to control energy flow to all or just some of the outlets in the wall. Finally, each device plugged into the wall has its own “on”/“off” switch. Many devices, like a radio, have additional switches on them that allow individualized energy use management. As we develop our understanding of regulation in biology, we will see that Nature has also evolved conceptually similar hierarchical control structures (though typically more interconnected) to help manage metabolism and various other cellular functions. Some signaling molecules target and regulate very specific control points - a single enzyme. Other molecular signals target many control

points simultaneously (many enzymes can be coordinately regulated). The interplay between specific control signals and broader control signals creates the ability to respond to big changes quickly and also to fine tune responses in the cell in a very detailed way.

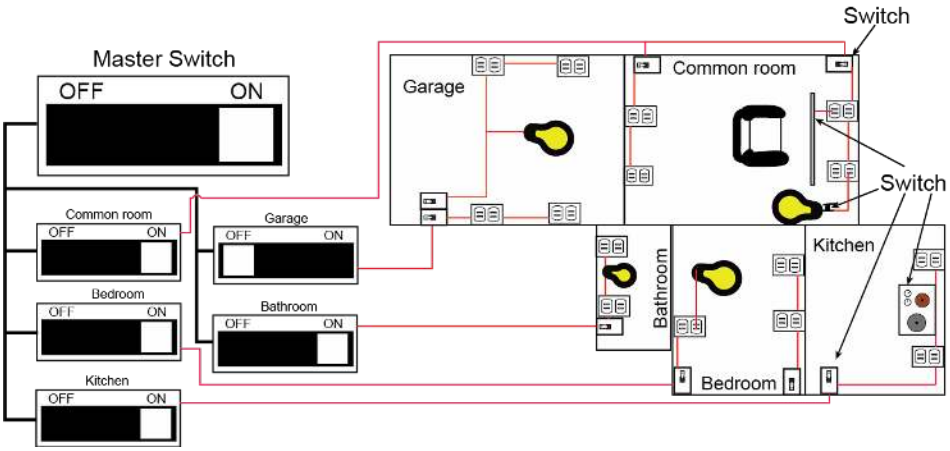


Figure 2. A schematic of a household wiring scheme showing hierarchical control of power to rooms, lights and devices. The master switch can turn off all power. Sub-switches can control power to individual rooms. Meanwhile, room-level switches can control power to parts of the room. Finally, switches on devices can control power at an individual device level. This serves as an analogy for some features of metabolic and cellular regulation.
Attribution: Original work - Marc T. Facciotti

The Who, What, When, Where, Why and How of Regulation

We structure the rest of this discussion of metabolic regulation by using a formalism often associated with journalism. We examine answers to the questions who, what, when, where, why and how <https://en.wikipedia.org/wiki/Five_Ws>.

Why regulate?

To some degree, we have made the argument about why regulation is necessary above. Largely, regulation in metabolism is involved in making decisions about how to best manage limited resources - the atoms that make up biomolecules, the biomolecules themselves, and the energy needed to build new molecules and/or carry out cellular functions. Good decision-making leads to more efficient resource utilization and by consequence we suspect higher evolutionary fitness.

Who to regulate?

When we consider the regulation of metabolism and metabolic pathways, the key control points are the enzymes that catalyze individual reactions. In the simplest terms, increasing the abundance of a specific enzyme and/or increasing its activity (how fast it catalyzes a reaction) can increase the flow of molecules through that step of a pathway. By contrast, decreasing the abundance of an enzyme and/or lowering its activity will reduce the amount of material flowing through that step in a pathway.

Where should regulation be exercised?

If enzymes and their relative abundance are good targets of regulation we can now ask whether some steps in metabolic pathways are more important to regulate than others. The answer to this question is yes and can be largely summarized by finding enzymes that belong to one or more of three classes: (a) enzymes that catalyze steps at branch points in a pathway; (b) enzymes that catalyze rate-limiting reactions; and (c) enzymes that catalyze so-called irreversible reactions (i.e. reaction with a large negative ΔG). We'll examine the logic behind each.

Enzymes that catalyze reactions at branch points: Branch points in metabolic reactions are points in a metabolic pathway in which a compound can be used as a substrate in two or more different biochemical reactions. The cell must "decide" which of the two or

more pathways to direct the biomolecule to. This is no different than the everyday experience we each time we encounter an intersection in the road we're traveling on - we must decide whether to continue straight or turn down a different path. We can't do both at the same time. Another everyday example might be in the construction of a water distribution pipeline in a garden with two or more planting beds. Depending on what is planted in each bed the gardener may wish to tune the flow of water to each bed differently and would do so by opening and closing valves at the branch points of the watering system.

Enzymes that catalyze rate-limiting steps in a pathway: A rate-limiting step in a metabolic pathway is defined as the reaction that determines how fast the overall pathway overall can convert input into output. This is almost always the slowest step in the pathway. Regulating this step by either slowing or speeding flow of metabolites through it can change the flow for the entire pathway. By analogy, you can think of a 5-lane highway that needs to constrict to 3 lanes over a bridge. While the highway may widen back up to 5 lanes after the bridge, the maximum rate of traffic flow before the bridge and after the bridge is limited by how many cars can get across the bridge at any one time. Narrowing or widening the bridge will have a major impact on traffic flow both before and after.

Enzymes that catalyze irreversible reactions: Reactions that have a large negative ΔG are often called irreversible reactions. These reactions are also often considered to be "commitment" steps in a biochemical pathway because it is difficult to reverse the reaction once it is done. Typically, different enzymes and external energy sources are required to run the reaction in the endergonic direction. We can also draw upon an analogy here to understand why these are important decision points by thinking about decisions in life that are hard to take back once they are made. Consider, for instance, the act of buying a used car or other equivalent item of value that you might need, or care to have. Few people would consider buying such an item sight unseen and without asking questions, doing some research, and maybe even getting a professional opinion on the condition of the car or other item. If the car/item is still functional and well-maintained the purchase could work out fantastically well. On the other hand, if the car/item has some defects or in need of repair, the purchase may not end well at all. The decision to buy the car/item must be carefully considered BEFORE the purchase. After you hand over the money there is typically no opportunity to changing your mind – the sale if final - and you will need to live with the consequences of that decision, good or bad. The same idea applies to reactions with large negative ΔG. Once the decision has been made to catalyze the reaction, that decision cannot be easily reversed. Therefore, the cell must evolve mechanisms that help make these decisions carefully on the basis of good information.

In many metabolic pathways, particularly those that are as interconnected to other pathways as glycolysis, the picture is - of course - more complicated as each these three types of control points are each used for regulation and it is too simplistic to simply state that there is a single point in the pathway that controls flow. Rather, it is the combination and interplay of regulatory sites and regulatory mechanisms that ultimately determines flow into, through, and out of the pathway.

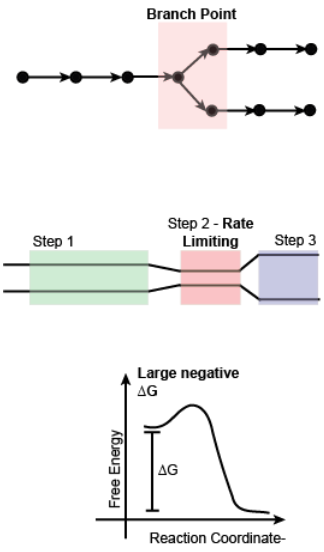


Figure 3. Schematic of three important types of steps in metabolic pathways that are key targets of regulation.
Attribution: Original work - Marc T. Facciotti

When to regulate?

A cell is always regulating the flow of molecules through its metabolic pathways. Understanding when to change the regulation of different metabolic pathways is the critical issue. When does the organism start to run glycolysis? When does it ramp up nucleotide biosynthesis and turn down amino acid biosynthesis? The short answer is that pathways are up-regulated when they are needed and down-regulated when they aren't. But how does the organism know when something is needed and something else isn't? The answer to this question is a bit more involved than what we can discuss in the time and space provided here. However, we can intuit that whatever these processes are, they must involve the sensing of environmental and/or cellular information. Something must be measured.

What to measure?

In the context of metabolic pathways, what information could be measured to help a cell make decisions about whether to up- or down-regulate flow through a metabolic pathway? Since metabolic pathways are in "the business" of consuming input molecules and converting them into some other product, it might be reasonable to expect that knowing whether there is enough product in the cell already (in which case the pathway doesn't need to run) or whether there is enough of the original substrate around to feed the pathway could be useful information to know for regulation. Other cellular information of relevance to metabolism that might be useful to know about when deciding on whether to up- or down-regulate a pathway is the general level of usable energy (e.g. the balance between levels of ATP, ADP, and AMP) and the availability of reduced and oxidized electron carriers (e.g. the balance between NADH and NAD^+). Clearly, in some cases the levels and balance between other molecules will also be important.

How can the cell use molecular information to make metabolic decisions?

Above we suggested some rationale about the regulation of metabolic pathways. We proposed that metabolic pathways must be regulated to manage cellular resources **(why)** and that this regulation should happen in response to changing cellular needs **(when)**. We propose that regulation happens when branch points, irreversible reactions, and rate-limiting steps are found **(where)**. Furthermore, we suggest that regulation happen by controlling the abundance and/or activity of the enzymes that catalyze reactions in a pathway **(who)**. Finally, we posit that knowing the abundance of pathway inputs and products as well as general indicators of cellular energy and redox stores **(what)** is generally useful for cellular decision-making.

This leaves us to answer the final question. **How?** How can a cell use abundance of cellular molecules to inform enzymes catalyzing key steps in metabolic pathways about whether they are needed more or less when cellular conditions change? The complete answer to this question is, like seemingly everything else in biology, multifaceted, involving different targets of regulation (e.g. the genes encoding an enzyme to the enzyme itself) and different mechanisms. Here we briefly discuss one mechanism that links back to an earlier lesson on proteins; allosteric regulation.

Allosteric regulation involves the binding of molecules to allosteric sites on an enzyme that are by definition not in the active site. This binding alters the protein structure and can cause in different instances both either up or down regulation of enzyme activity. Often, enzymes that are the subject of allosteric regulation can be influenced by more than one ligand. By binding different phosphorylation states of ATP (i.e. AMP, ADP, and ATP) at allosteric sites enzymes can measure the ratios of ATP to its other forms and thus assess the energy "status" of the cell.

Examples of metabolic regulation by allosteric binding

Reactions with large negative ΔG

An early intermediate step of glycolysis, the phosphorylation of fructose-6-phosphate by ATP to yield fructose-1,6-bisphosphate and ADP, has an approximate ΔG of -19kJ/mol under cellular conditions. This is one of three largest, by comparison to all other reactions in the pathway, drops in free energy in the glycolytic pathway. In mammalian systems, the enzyme catalyzing this reaction, phosphofructokinase, is both positively and negatively regulated by multiple small molecules. ATP, citrate, and phosphoenolpyruvate (PEP) have been found to bind allosteric sites on the enzyme and lead to a lowering of enzyme activity. AMP, ADP, and fructose-2,6-bisphosphate (a product created by another enzyme) each increase phosphofructokinase activity. A number

of other small molecules have also been found to influence enzyme activity in the test tube, but these are typically considered to play an insignificant role in the living system.

Given the central role of the glycolytic pathway in both energy harvesting and the creation of key precursors for other metabolic pathways, it is perhaps not surprising to have this enzyme's activity influenced by a measure of the energy status of the cell (the ATP/AMP ratio) and by molecules in connected pathways (PEP and citrate). It is perhaps easiest to understand that when the enzyme senses low energy levels in the cell, when the ATP/AMP ratio is low, that the enzyme committing sugar to enter the energy extraction phase of glycolysis should become more active and that the converse should be true when ATP is abundant.

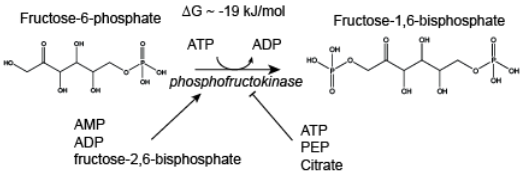


Figure 4. The enzyme phosphofructokinase is regulated allosterically both positively and negatively by numerous small molecules indicators of cellular energy state and of key concentrations of pathway intermediates. *Attribution: Original work - Marc T. Facciotti*

Branch Points

An example of branch point regulation can be found in the pathways leading to the biosynthesis of aromatic amino acids (i.e. Phenylalanine, Tyrosine, and Tryptophan). The synthesis of each of these three amino acids begins with the production of the compound Chorismate. This compound can then be taken down two independent pathways. The first leads to Phenylalanine and Tyrosine biosynthesis while the second leads to Tryptophan biosynthesis. The enzyme catalyzing the first step in the path towards Phenylalanine and Tyrosine, Chrosimate mutase, is negatively regulated by the two products of the pathway and activated by Tryptophan, the product of the second pathway. Meanwhile, the enzyme Anthranilate Synthase is negatively regulated by the final product of its pathway, Tryptophan. The feedback of end-product levels on the enzymes responsible for catalyzing the synthesis of these three amino acids allow the cell to decide how to best utilize the stock of Chorismate and make that decision at the one of the core branch points for its use.

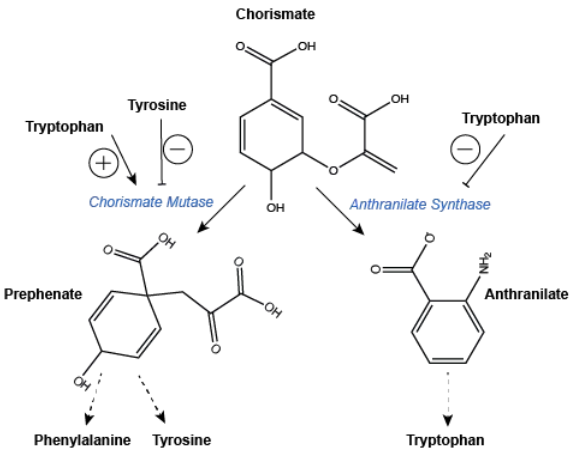


Figure 5: Regulation of branchpoint enzymes in the biosynthesis of aromatic amino acids. Dashed arrows indicate one or more steps are still required to create product of the pathway and these are not drawn explicitly. Note that this is also an example of pathway products feeding back to regulate upstream enzyme activity.
Attribution: Original work - Marc T. Facciotti

Reading_14 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_15

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_15

GC.52 Explain why organisms might use different pigments depending on their habitat.
GC.53 Understand that pigment's chemical structure helps to determine how it interacts with light.
ME.27 Describe the role of the photosystem in coupling the absorption of light to the creation of ATP and NADP(H).
ME.29 List the advantages and disadvantages of having two photosystems and the ability to use water as an electron donor.
ME.28 Use your understanding of respiratory electron transport chains to interpret the function(s) of photosynthetic electron transport chains and their individual components.
ME.30 Create a sketch model and an energy story for photosynthesis that includes the flow of electrons from a "low energy" source to an electron transporter (e.g. NADP), and finally to a reduced carbon compound.
ME.24 Express the functional links between carbon fixation and light-driven ATP and NADPH production.
ME.25 Create an energy story for the Calvin cycle.
ME.26 Explain the difference between cyclic and non-cyclic photophosphorylation. Be able to illustrate the flow of electrons in both processes, highlighting the electron source and sink in both cases.
ME.40 Compare and contrast the three forms of ATP production - substrate-level phosphorylation, oxidative phosphorylation and photophosphorylation.

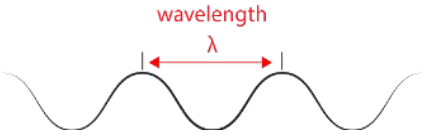
Light Energy and Pigments

Light Energy

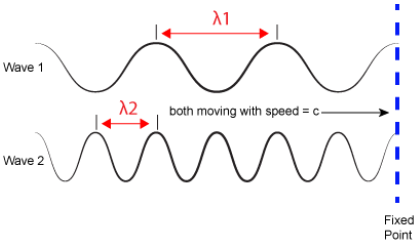
The sun emits an enormous amount of electromagnetic radiation (solar energy) that spans a broad swath of the **electromagnetic spectrum**, the range of all radiation frequencies. When solar radiation reaches Earth, a fraction of this energy interacts with and may transfer to the matter on the planet. This energy transfer results in a wide variety of different phenomena, from influencing weather patterns to driving myriad biological processes. In general biology, we are largely concerned with the latter, and below, we discuss some very basic concepts related to light and its interaction with biology.

First, however we need to refresh a few key properties of light:

1. Light in a vacuum travels at a constant speed of 299,792,458 m/s. We often abbreviate the speed of light with the variable "c".
2. Light has properties of waves. A specific "color" of light has a characteristic wavelength.



We refer to the distance between peaks in a wave as the wavelength and abbreviate it with the Greek letter lambda (λ). Attribution: Marc T. Facciotti (original work)



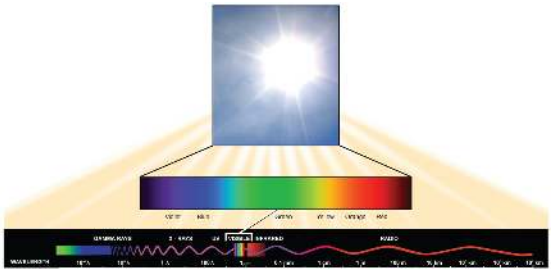
The inverse proportionality of frequency and wavelength. Wave 1 has a wavelength that is 2x that of wave 2 ($\lambda 1 > \lambda 2$). If the two waves are traveling at the same speed (c)—imagine that both of the whole lines that are dragged past the fixed vertical line at the same speed—then the number of times a wave peak passes a fixed point is greater for wave 2 than wave 1 (f2 > f1). Attribution: Marc T. Facciotti (original work)

---

3. Finally, each frequency (or wavelength) of light is associated with a specific energy. We'll call energy "E". The relationship between frequency and energy is:

$$E = h \times f \tag{1}$$

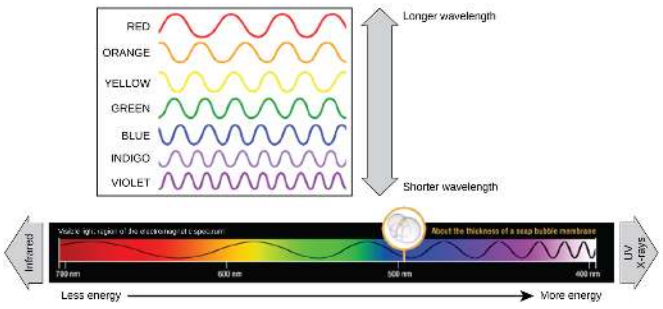
where h is a constant called the Planck constant (~6.626×10^{-34} Joule•second when frequency is expressed in cycles per second). Given the relationship between frequency and wavelength, you can also write $E = h*c/\lambda$. Therefore, the larger the frequency (or shorter the wavelength), the more energy is associated with a specific "color". Wave 2 in the figure above is associated with greater energy than wave 1.



The sun emits energy in the form of electromagnetic radiation. All electromagnetic radiation, including visible light, is characterized by its wavelength. The longer the wavelength, the less energy it carries. The shorter the wavelength, the more energy is associated with that band of the electromagnetic spectrum.

The Light We See

The visible light seen by humans as white light is composed of a rainbow of colors, each with a characteristic wavelength. Certain objects, such as a prism or a drop of water, disperse white light to reveal the colors to the human eye. In the visible spectrum, violet and blue light have shorter (higher energy) wavelengths while the orange and red light have a longer (lower energy) wavelengths.



The colors of visible light do not carry the same amount of energy. Violet has the shortest wavelength and therefore carries the most energy, whereas red has the longest wavelength and carries the least amount of energy. Credit: modification of work by NASA.

Absorption by Pigments

The interaction between light and biological systems occurs through several mechanisms, some of which you may learn about in upper division courses in cellular physiology or biophysical chemistry. In general biology, we concern ourselves mostly with the interaction between light and biological pigments. These interactions can start a variety of light-dependent biological processes that can be grossly grouped into two functional categories: cellular signaling and energy harvesting. Signaling molecules perceive changes in the environment (in this case, changes in light). An example of a signaling interaction might be the interaction between light and the pigments expressed in an eye. Light/pigment interactions that are involved in energy harvesting are used for—not surprisingly—capturing the energy in the light and transferring it to the cell to fuel biological processes. Photophosphorylation, which we will learn more about soon, is one example of an energy harvesting interaction.

Possible NB Discussion Point



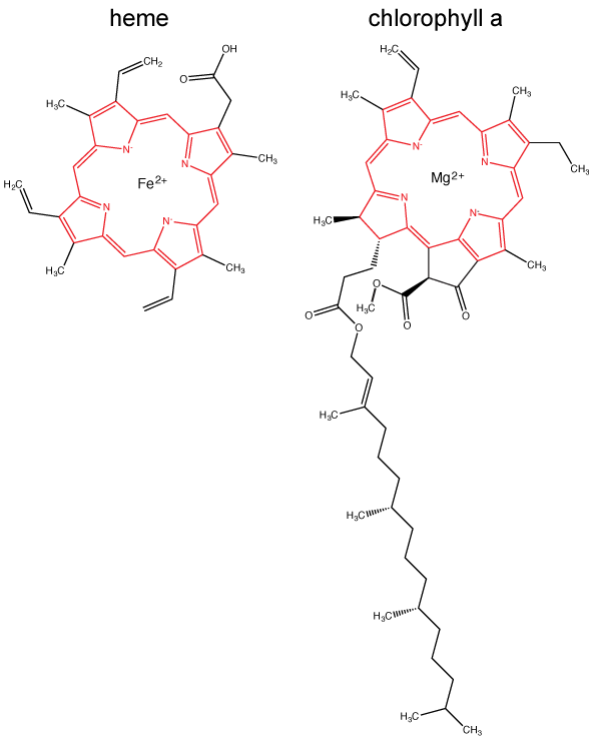
Photophosphorylation is a process involving an electron transport chain that allows organisms to harvest energy from light. Some of you may already be familiar with this process. Many of you are learning about this for the first time. Given your current knowledge base, offer your best explanation or hypothesis as to how light interacts with the ETC. You will have a chance to revisit this topic very soon.

At the center of the biological interactions with light are groups of molecules we call organic pigments. Whether in the human retina, chloroplast thylakoid, or microbial membrane, organic pigments often have specific ranges of energy or wavelengths that they can absorb. The sensitivity of these molecules for different wavelengths of light is due to their unique chemical makeups and structures. A range of the electromagnetic spectrum is given a couple of special names because of the sensitivity of some key biological pigments: The retinal pigment in our eyes, when coupled with an opsin sensor protein, “sees” (absorbs) light predominantly between the wavelengths between of 700 nm and 400 nm. Because this range defines the physical limits of the electromagnetic spectrum that we can actually see with our eyes, we refer to this wavelength range as the “visible range”. For similar reasons, as plants pigment molecules tend to absorb wavelengths of light mostly between 700 nm and 400 nm, plant physiologists refer to this range of wavelengths as "photosynthetically active radiation".

Three Key Types of Pigments Commonly Discussed in General Biology

Chlorophylls

Chlorophylls (including bacteriochlorophylls) are part of a large family of pigment molecules. There are five major chlorophyll pigments named: *a*, *b*, *c*, *d*, and *f*. Chlorophyll *a* is related to a class of more ancient molecules found in bacteria called **bacteriochlorophylls**. Chlorophylls are structurally characterized by ring-like porphyrin group that coordinates a metal ion. This ring structure is chemically related to the structure of heme compounds that also coordinate a metal and are involved in oxygen binding and/or transport in many organisms. We distinguish different chlorophylls from one another by different "decorations"/chemical groups on the porphyrin ring.



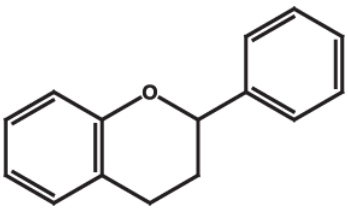
The structure of heme and chlorophyll a molecules. The common porphyrin ring is colored in red. Attribution: Marc T. Facciotti (original work)

Carotenoids

Carotenoids are the red/orange/yellow pigments found in nature. They are found in fruit—the red of tomato (lycopene), the yellow of corn seeds (zeaxanthin), or the orange of an orange peel (β-carotene)—which serve as biological "advertisements" to attract seed dispersers (animals or insects that may carry seeds elsewhere). In photosynthesis, carotenoids function as photosynthetic pigments. In addition, when a leaf is exposed to full sun, that surface is required to process an enormous amount of energy; If that energy is not managed properly, it can do significant damage. Therefore, many carotenoids help absorb excess energy in light and safely dissipate that energy as heat.

Flavonoids

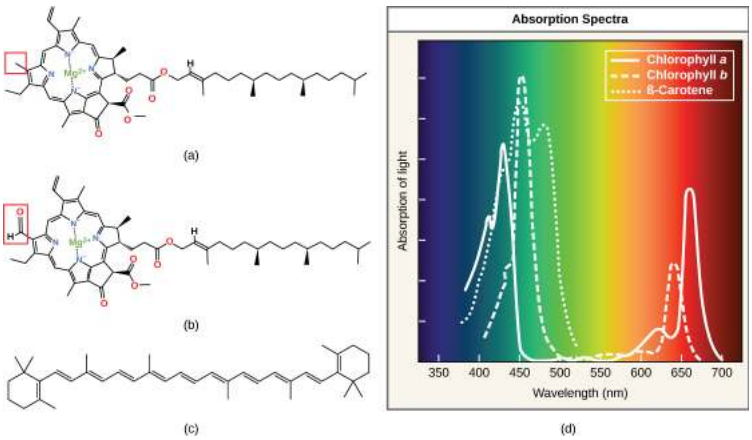
Flavonoids are a very broad class of compounds that are found in great diversity in plants. These molecules come in many forms but all share a common core structure shown below. The diversity of flavonoids comes from the many combinations of functional groups that can "decorate" the core flavone.



The core ring structure of flavans.

Each type of pigment can be identified by the specific pattern of wavelengths it absorbs from visible light. We define this characteristic as the pigment's **absorption spectrum**. The graph in the figure below shows the absorption spectra for chlorophyll *a*, chlorophyll *b*, and a type of carotenoid pigment called β-carotene (which absorbs blue and green light). Notice how each pigment

has a distinct set of peaks and troughs, revealing a highly specific pattern of absorption. These differences in absorbance are because of differences in chemical structure (some of these are highlighted in the figure). Chlorophyll *a* absorbs wavelengths from either end of the visible spectrum (blue and red), but not green. Because chlorophyll reflects green light and absorbs other wavelengths of light, things containing this pigment appear green. Carotenoids absorb in the short-wavelength blue region and reflect the longer yellow, red, and orange wavelengths.



(a) Chlorophyll a, (b) chlorophyll b, and (c) β-carotene are hydrophobic organic pigments found in the thylakoid membrane. Chlorophyll a and b, which are identical except for the part indicated in the red box, are responsible for the green color of leaves. Note how the small amount of difference in chemical composition between different chlorophylls leads to different absorption spectra. β-carotene is responsible for the orange color in carrots. Each pigment has a unique absorbance spectrum (d).

Importance of having multiple different pigments

Not all photosynthetic organisms have full access to sunlight. Some organisms grow underwater where light intensity and the number of wavelengths decrease and change, respectively, with depth. Other organisms grow in competition for light. For instance, plants on the rainforest floor must be able to absorb any bit of light that comes through because the taller trees absorb most of the sunlight and scatter the remaining solar radiation. To account for these variable light conditions, many photosynthetic organisms have a mixture of pigments whose expression they can tune to improve the organism's ability to absorb energy from a wider range of wavelengths than would be possible with one pigment alone.

Photophosphorylation

Photophosphorylation an overview

Photophosphorylation is the process of transferring the energy from light into chemicals, particularly ATP. The evolutionary roots of photophosphorylation are likely in the anaerobic world, between 3 billion and 1.5 billion years ago, when life was abundant in the absence of molecular oxygen. Photophosphorylation probably evolved relatively shortly after electron transport chains (ETC) and **anaerobic respiration** provided metabolic diversity. The first step of the process involves the absorption of a photon by a pigment molecule. Light energy transfers to the pigment and promotes electrons (e^-) into a higher quantum energy state—something biologists term an "excited state". Note the use of anthropomorphism here; the electrons are not "excited" in the classic sense and aren't suddenly hopping all over or celebrating their promotion. They are simply in a higher energy quantum state. In this state, the electrons are colloquially said to be "energized". While in the "excited" state, the pigment now has a much lower reduction potential and can donate the "excited" electrons to other carriers with greater reduction potentials. These electron acceptors may become donors to other molecules with greater reduction potentials and, in doing so, form an electron transport chain.

As electrons pass from one electron carrier to another via red/ox reactions, enzymes can couple these exergonic electron transfers

to the endergonic transport (or pumping) of protons across a membrane to create an electrochemical gradient. This electrochemical gradient generates a proton motive force (PMF). Enzymes can couple the exergonic drive of these protons to reach equilibrium to the endergonic production of ATP, via ATP synthase. As we will see in more detail, the electrons involved in this electron transport chain can have one of two fates: (1) they may return to their initial source in a process called cyclic photophosphorylation; or (2) they can transfer onto a close relative of NAD^+ called $NADP^+$. If electrons return to the original pigment in a cyclic process, the whole process can start over. If, however, the electron transfers onto $NADP^+$ to form NADPH (**shortcut note—we didn't explicitly mention any protons but assume that they are also involved**), the original pigment must regain an electron from somewhere else. This electron must come from a source with a smaller reduction potential than the oxidized pigment and depending on the system there are different sources, including H_2O, reduced sulfur compounds such as SH_2 and even elemental S^0.

What happens when a compound absorbs a photon of light?

When a compound absorbs a photon of light, the compound is said to leave its ground state and become "excited".

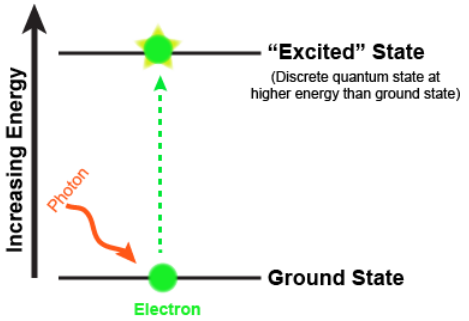


Figure 1. *A diagram depicting what happens to a molecule that absorbs a photon of light. Attribution: Marc T. Facciotti (original work)*

What are the fates of the "excited" electron? There are four possible outcomes, which are schematically diagrammed in the figure below. These options are:

1. The e^- can relax to a lower quantum state, transferring energy as heat.
2. The e^- can relax to a lower quantum state and transfer energy into a photon of light—a process known as fluorescence.
3. The energy can be transferred by resonance to a neighboring molecule as the e^- returns to a lower quantum state.
4. The energy can change the reduction potential such that the molecule can become an e^- donor. Linking this excited e^- donor to a proper e^- acceptor can lead to an exergonic electron transfer. The excited state can be involved in red/ox reactions.

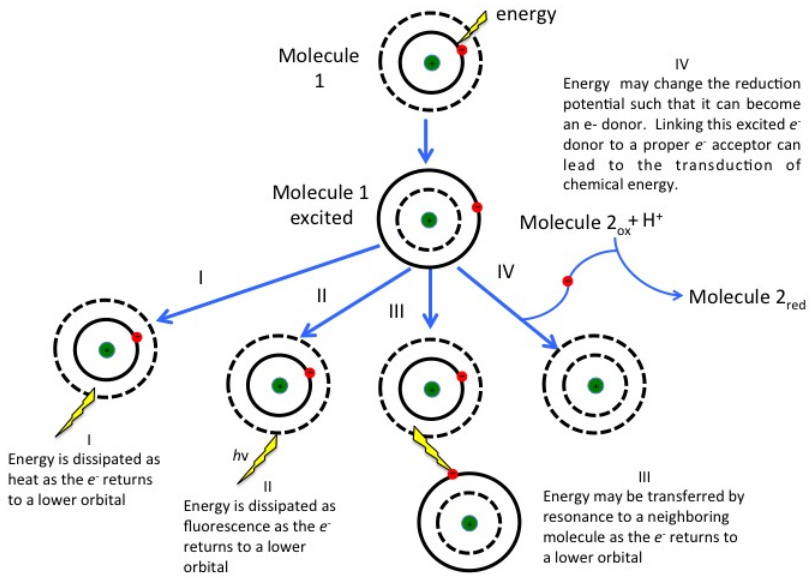


Figure 2. *What can happen to the energy absorbed by a molecule.*

As the excited electron decays back to its lower energy state, it can transfer its energy in a variety of ways. While many so-called antenna or auxiliary pigments absorb light energy and transfer it to something known as a reaction center (by mechanisms depicted in option III in Figure 2), it is what happens at the reaction center that we are most concerned with (option IV in the figure above). Here a chlorophyll or bacteriochlorophyll molecule absorbs a photon's energy, and an electron is excited. This energy transfer suffices to allow the reaction center to donate the electron in a red/ox reaction to a second molecule. This starts the electron transport reactions. The result is an oxidized reaction center that must now be reduced to start the process again. How this happens is the basis of electron flow in photophosphorylation and we describe this below.

Simple photophosphorylation systems: anoxygenic photophosphorylation

Early in the evolution of photophosphorylation, these reactions evolved in anaerobic environments where there was very little molecular oxygen available. Two sets of reactions evolved under these conditions, both directly from anaerobic respiratory chains as described previously. We know these as the **light reactions** because they require the activation of an electron (an "excited" electron) from the absorption of a photon of light by a reaction center pigment, such as bacteriochlorophyll. We classify the light reactions either as **cyclic** or as **noncyclic** photophosphorylation, depending upon the final state of the electron(s) removed from the reaction center pigments. If the electron(s) returns to the original pigment reaction center, such as bacteriochlorophyll, this is cyclic photophosphorylation; the electrons make a complete circuit. We diagram this in Figure 4. If the electron(s) are used to reduce $NADP^+$ to NADPH, the electron(s) are removed from the pathway and end up on NADPH; we refer to this process as noncyclic since the electrons are no longer part of the circuit. Here the reaction center must be re-reduced before the process can happen again. Therefore, an external electron source is required for noncylic photophosphorylation. In these systems reduced forms of Sulfur, such as H_2S, which can be used as an electron donor and is diagrammed in Figure 5. To help you better understand the similarities of photophosphorylation to respiration, we provided a red/ox tower that contains many commonly used compounds involved with photosphosphorylation.

oxidized form	reduced form	n (electrons)	Eo´ (volts)
PS1* (ox)	PS1* (red)	-	-1.20
ferredoxin (ox) version 1	ferredoxin (red) version 1	1	-0.7
PSII* (ox)	PSII* (red)	-	-0.67
P840* (ox)	PS840* (red)	-	-0.67
acetate	acetaldehyde	2	-0.6
CO_2	Glucose	24	-0.43
ferredoxin (ox) version 2	ferredoxin (red) version 2	1	-0.43
CO_2	formate	2	-0.42
$2H^+$	H_2	2	-0.42 (at $[H^+] = 10^{-7}$; pH=7)
$NAD^+ + 2H^+$	$NADH + H^+$	2	-0.32
$NADP^+ + 2H^+$	$NADPH + H^+$	2	-0.32
Complex I FMN (enzyme bound)	$FMNH_2$	2	-0.3
Lipoic acid, (ox)	Lipoic acid, (red)	2	-0.29
FAD^+ (free) + $2H^+$	$FADH_2$	2	-0.22
Pyruvate + $2H^+$	lactate	2	-0.19
$FAD^+ + 2H^+$ (bound)	$FADH_2$ (bound)	2	0.003-0.09
CoQ (Ubiquinone - UQ + H^+)	UQH**.**	1	0.031
$UQ + 2H^+$	UQH_2	2	0.06
Plastoquinone; (ox)	Plastoquinone; (red)	-	0.08
Ubiquinone; (ox)	Ubiquinone; (red)	2	0.1
Complex III Cytochrome b_2; Fe^{3+}	Cytochrome b_2; Fe^{2+}	1	0.12
Complex III Cytochrome c_1; Fe^{3+}	Cytochrome c_1; Fe^{2+}	1	0.22
Cytochrome c; Fe^{3+}	Cytochrome c; Fe^{2+}	1	0.25
Complex IV Cytochrome a; Fe^{3+}	Cytochrome a; Fe^{2+}	1	0.29
1/2 $O_2 + H_2O$	H_2O_2	2	0.3
$P840^{GS}$ (ox)	$PS840^{GS}$ (red)	-	0.33
Complex IV Cytochrome a_3; Fe^{3+}	Cytochrome a_3; Fe^{2+}	1	0.35

oxidized form	reduced form	n (electrons)	Eo´ (volts)
Ferricyanide	ferrocyanide	2	0.36
Cytochrome f; Fe^{3+}	Cytochrome f; Fe^{2+}	1	0.37
PSI^{GS} (ox)	PSI^{GS} (red)	.	0.37
Nitrate	nitrite	1	0.42
Fe^{3+}	Fe^{2+}	1	0.77
1/2 O_2 + $2H^+$	H_2O	2	0.816
$PSII^{GS}$ (ox)	$PSII^{GS}$ (red)	-	1.10

* Excited State, after absorbing a photon of light
GS Ground State, state prior to absorbing a photon of light
PS1: Oxygenic photosystem I
P840: Bacterial reaction center containing bacteriochlorophyll (anoxygenic)
PSII: Oxygenic photosystem II

Figure 3. *Electron tower that has a variety of common photophosphorylation components. PSI and PSII refer to Photosystems I and II of the oxygenic photophosphorylation pathways.*

Cyclic photophosphorylation

In cyclic photophosphorylation the bacteriochlorophyll$_{red}$ molecule absorbs enough light energy to energize and eject an electron to form bacteriochlorophyll$_{ox}$. The electron reduces a carrier molecule in the reaction center which in turn reduces a series of carriers via red/ox reactions. These carriers are the same carriers found in respiration. If the change in reduction potential from the various red/ox reactions are sufficiently large, H^+ protons can be translocated across a membrane. Eventually, the electron is used to reduce bacteriochlorophyll$_{ox}$ (making a complete loop) and the whole process can start again. This flow of electrons is cyclic and is therefore said to drive a processed called cyclic photophosphorylation. The electrons make a complete cycle: bacteriochlorophyll is the initial source of electrons and is the final electron acceptor. ATP is produced via the **F_1F_0 ATPase**. The schematic in Figure 4 demonstrates how cyclic electrons flow and thus how cyclic photophosphorylation works.

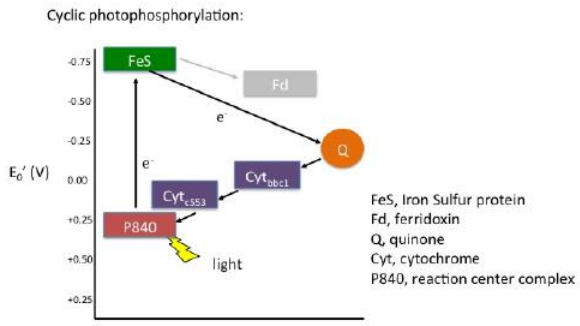


Figure 4. *Cyclic electron flow. The reaction center P840 absorbs light energy and becomes excited, denoted with an *. The excited electron is ejected and used to reduce an FeS protein leaving an oxidized reaction center. The electron its transferred to a quinone, then to a series of cytochromes, which in turn reduces the P840 reaction center. The process is cyclical. Note the gray array coming from the FeS protein going to a ferridoxin (Fd), also in gray. This represents an alternative pathway the electron can take and will be discussed below in noncyclic photophosphorylation. Note that the electron that initially leaves the P840 reaction center is not necessarily the same electron that eventually finds its way back to reduce the oxidized P840.*

Possible NB Discussion Point

The figure of cyclic photophosphorylation above depicts the flow of electrons in a respiratory chain. How does this process help generate ATP? Why might running the process in a cyclical fashion be advantageous for a cell?

Noncyclic photophosphorylation

In cyclic photophosphorylation, electrons cycle from bacteriochlorophyll (or chlorophyll) to a series of electron carriers and eventually back to bacteriochlorophyll (or chlorophyll); there is theoretically no net loss of electrons and they stay in the system. In noncyclic photophosphorylation, electrons leave from the photosystem and red/ox chain and eventually end up on NADPH. That means there needs to be a source of electrons, a source that has a smaller reduction potential than bacteriochlorophyll (or chlorophyll) that can donate electrons to bacteriochlorophyll$_{ox}$ to reduce it. From looking at the electron tower in Figure 3, you can see what compounds can reduce the oxidized form of bacteriochlorophyll. The second requirement is that, when bacteriochlorophyll becomes oxidized by ejecting its excited electron, it must reduce a carrier that has a greater reduction potential than NADP/NADPH (see the electron tower). Here, electrons can flow from energized bacteriochlorophyll to NADP forming NADPH and oxidized bacteriochlorophyll. The system loses electrons and they end up on NADPH; to complete the circuit, bacteriochlorophyll$_{ox}$ is reduced by an external electron donor such as H_2S or elemental S^0.

Noncyclic electron flow

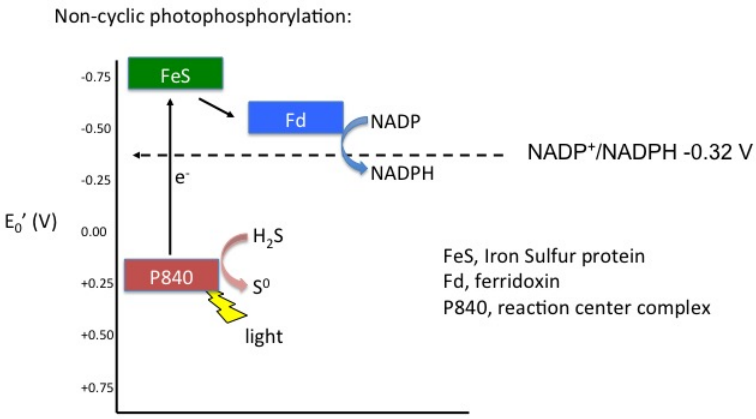


Figure 5. *Noncyclic electron flow. In this example, the P840 reaction center absorbs light energy and becomes energized; the emitted electron reduces a FeS protein and reduces ferridoxin. Reduced ferridoxin (Fd_{red}) can now reduce NADP to form NADPH. The electrons are now removed from the system, finding their way to NADPH. The electrons need to be replaced on P840, which requires an external electron donor. Here, H_2S serves as the electron donor.*

We note that for bacterial photophosphorylation pathways, for each electron donated from a reaction center [remember only one electron is actually donated to the reaction center (or chlorophyl molecule)], the resulting output from that electron transport chain is either the formation of NADPH (requires two electrons) or ATP can be made but NOT not both. The path the electrons take in the ETC can have one or two outcomes. This puts limits on the versatility of the bacterial anoxygenic photosynthetic systems. But what would happen if a process evolved that used both systems? More precicely, a cyclic and noncyclic photosynthetic pathway which could form both ATP and NADPH from a single input of electrons? A second limitation is that these bacterial systems require compounds such as reduced sulfur to act as electron donors to reduce the oxidized reaction centers, but they are not necessarily widely found compounds. What would happen if a chlorophyll $_{ox}$ molecule would have a reduction potential higher (more positive) than that of the molecular O_2/H_2O reaction? Answer: a planetary game changer.

Oxygenic photophosphorylation

Generation of NADPH and ATP

The overall function of light-dependent reactions is to transfer solar energy into chemical compounds, largely the molecules NADPH and ATP. This energy supports the light-independent reactions and fuels the assembly of sugar molecules. We depict the light-dependent reactions in Figures 6 and 7. Protein complexes and pigment molecules work together to produce NADPH and ATP.

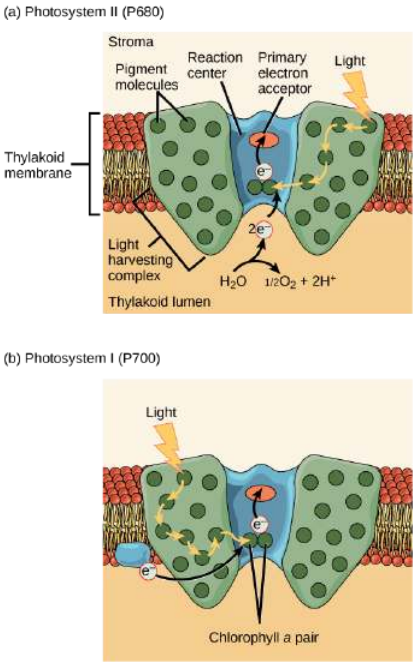


Figure 1. *A photosystem consists of a light-harvesting complex and a reaction center. Pigments in the light-harvesting complex pass light energy to two special chlorophyll* a *molecules in the reaction center. The light excites an electron from the chlorophyll* a *pair, which passes to the primary electron acceptor. The excited electron must then be replaced. In (a) photosystem II, the electron comes from the splitting of water, which releases oxygen as a waste product. In (b) photosystem I, the electron comes from the chloroplast electron transport chain discussed below.*

The actual step that transfers light energy into a biomolecule takes place in a multiprotein complex called a **photosystem**, two types of which are found embedded in the thylakoid membrane, **photosystem II** (PSII) and **photosystem I** (PSI). The two complexes differ based on what they oxidize (that is, the source of the low energy electron supply) and what they reduce (the place to which they deliver their energized electrons).

Both photosystems have the same basic structure; several **antenna proteins** to which the chlorophyll molecules bind surround the **reaction center** in which the photochemistry takes place. Each photosystem associates with the **light-harvesting complex**, which passes energy captured from sunlight to the reaction center; it comprises multiple antenna proteins that contain a mixture of 300–400 chlorophyll *a* and *b* molecules and other pigments like carotenoids. The absorption of a single **photon**—a distinct quantity or "packet" of light—by any of the chlorophylls pushes that molecule into an excited state. In short, the biological molecule has now captured light energy. However, the energy is not yet stored in any useful form. The captured energy transfers from chlorophyll to chlorophyll until, eventually (after about a millionth of a second), it arrives to the reaction center. Up to this point, only energy has been transferred between molecules, not electrons.

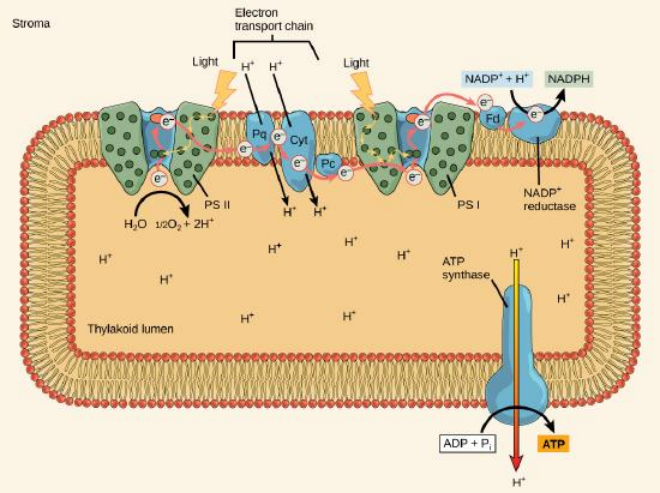


Figure 2. *In the photosystem II (PSII) reaction center, energy from sunlight is used to extract electrons from water. The electrons travel through the chloroplast electron transport chain to photosystem I (PSI), which reduces $NADP^+$ to NADPH. The electron transport chain moves protons across the thylakoid membrane into the lumen. At the same time, splitting of water adds protons to the lumen, and reduction of NADPH removes protons from the stroma. The net result is a low pH in the thylakoid lumen, and a high pH in the stroma. ATP synthase uses this electrochemical gradient to make ATP.*

The reaction center contains a pair of chlorophyll *a* molecules with a special property. Those two chlorophylls can undergo oxidation upon excitation; they can give up an electron in a process called a **photoactivation**. It is at this step in the reaction center, this step in photophosphorylation, that light energy transfers into an excited electron. All the subsequent steps involve a getting of that electron onto the energy carrier NADPH for delivery to the Calvin cycle where the electron can be deposited onto carbon for long-term storage in the form of a carbohydrate.

The Z-scheme

PSII and PSI are two major components of the photosynthetic **electron transport chain**, which also include the **cytochrome complex**. The reaction center of PSII (called **P680**) delivers its high-energy electrons, one at a time, to a **primary electron acceptor** called pheophytin (Ph), and then sequentially to two bound plastoquinones Q_A and Q_B. Electrons then transfer off of PSII onto a pool of mobile plastoquinones (Q pool) which then transfer the electrons to a protein complex called Cytochromeb_6f. The cytochrome complex uses the red/ox transfers to pump proton across the thylakoyd membrane establishing a proton-motive force that can be used for the synthesis of ATP. Electrons leaving the Cytochrome transfer to a copper-containing protein called plastocyanin (PC) which then transfers electrons to PSI (P700). P680's missing electron is replaced by extracting an electron from water; thus, water splits and PSII is re-reduced after every photoactivation step. Just for the sake of sharing some numbers: Splitting one H_2O molecule releases two electrons, two hydrogen atoms, and one atom of oxygen. The formation of one molecule of diatomic O_2 gas requires the splitting two water molecules. In plant tissue, mitochondria use about ten percent of that oxygen to support oxidative phosphorylation. The rest escapes to the atmosphere where it is used by aerobic organisms to support respiration.

As electrons move through the proteins that live between PSII and PSI, they take part in exergonic red/ox transfers. The free energy associated with the exergonic red/ox reaction is coupled to the endergonic transport of protons from the stromal side of the membrane to the thylakoid lumen by the cytochrome complex. Those hydrogen ions, plus the ones produced by splitting water, accumulate in the thylakoid lumen and create a proton motive force that will drive the synthesis of ATP in a later step. Since the electrons on PSI now have a greater reduction potential than when they started their trek (it is important to note that unexcited PSI has a greater red/ox potential than $NADP^+$/NADPH), they must be re-energized in PSI before getting deposited onto $NADP^+$. Therefore, to complete this process, another photon must be absorbed by the PSI antenna. That energy transfers to the PSI reaction center (called **P700**). P700 then oxidizes and sends an electron through several intermediate red/ox steps to $NADP^+$ to form

NADPH. Thus, PSII captures the energy in light and couples its transfer via red/ox reactions to the creation of a proton gradient. As already noted, the exergonic and controlled relaxation of this gradient can be coupled to the synthesis of ATP. PSI captures energy in light and couples that, through a series of red/ox reactions, to reduce $NADP^+$ into NADPH. The two photosystems work in concert, in part, to guarantee that the production of NADPH will be in the right proportion to the production of ATP. Other mechanisms exist to fine tune that ratio to match the chloroplast's constantly changing energy needs.

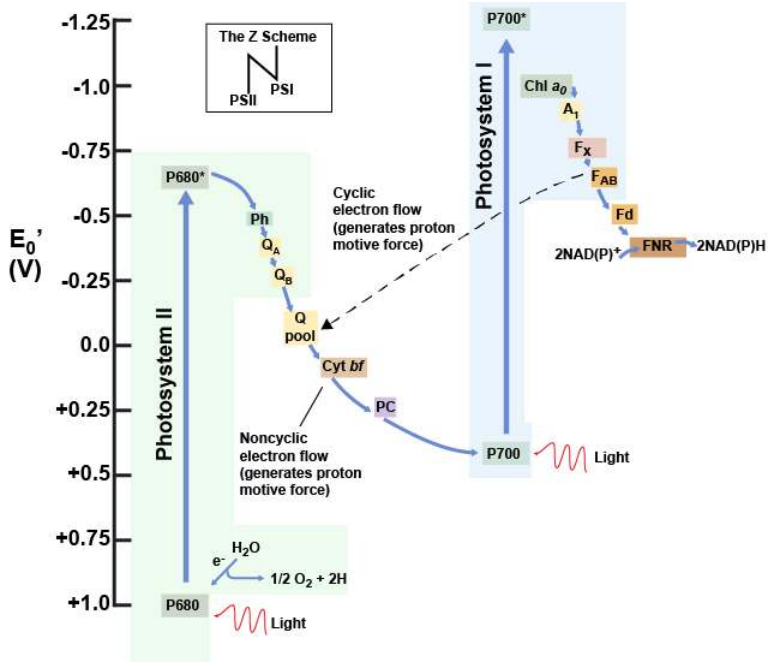


Figure 3. *A diagram depicting the flow of electrons and the red/ox potentials of their carriers in oxygenic photosynthetic systems expressing both photosystem I (boxed in blue) and photosystem II (boxed in green). Ph* = pheophytin; Q_A = bound plastoquinone, Q_B = more loosely associated plastoquinone; Q pool = mobile plastoquinone pool; Cyt *bf* = Cytochrome b_6f complex; PC = plastocyanin; Chl_{a0} = special chrolophyl; A_1 = vitamin K; F_x and F_{AB} = iron-sulfur centers; F_d = ferredoxin; FNR = ferredoxin-NADP reductase. Attribution: Marc T. Facciotti (own work)

Reading_15 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_16

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_16

ME.41 Properly connect the pentose phosphate pathway (PPP) to glycolysis and explain the PPP's role in supplying NADPH for cellular anabolic reactions and in the biosynthesis of nucleotides and aromatic amino acids (tryptophan, phenylalanine, tyrosine).

Introduction into the pentose phosphate pathway (PPP)



Discussions of metabolism in most introductory biology courses focus on glycolysis (oxidation of glucose to pyruvate) and the TCA cycle (oxidation of pyruvate to acetyl-CoA and the eventual complete oxidation to CO_2). While these are important and universal metabolic pathways, many courses leave out the pentose phosphate pathway (PPP), also known as the hexose monophosphate shunt. In this class, we consider the PPP important for two key reasons. First, it is the primary route for the formation of pentoses, the five-carbon sugar required for nucleotide biosynthesis and a variety of other essential cellular components. Second, redox reactions in the PPP generate NADPH, the main mobile electron donor used in anabolic (building) reactions.

A note from the instructor

Like the modules on glycolysis and the TCA cycle, we do not expect students to memorize specific compound names or details of molecular structures in the pathway. We provide those details in the reading so you can understand the transformations occurring in this pathway and refer back to them when needed. Rather than memorizing, focus instead on the mastering the assigned learning goals related to the PPP.

Oxidative pentose phosphate pathway: a.k.a., the hexose monophosphate shunt

We call glycolysis, the TCA cycle and the pentose phosphate pathway **central carbon metabolism**. These three pathways (along with the reaction that converts pyruvate to acetyl-CoA) contain all the chemical precursors required by cells for the biosynthesis of nearly all other biomolecules. The PPP produces **pentose phosphates** (five-carbon sugars), **eyrthrose-phosphate** (a four-carbon sugar), and **NADPH**. The pentose phosphates are key precursors for nucleotide biosynthesis while NADPH serves as the main mobile electron donor for anabolic (building) reactions. The PPP also produces **sedoheptulose-phosphate**, an essential seven-carbon sugar used in the building of Gram-negative bacteria's outer cell membranes.

Below is a diagram of the pathway. The pathway involves several redox reactions and multiple molecular rearrangement that interconvert molecules of 3-, 4-, 5-, 6-, and 7-carbons. The pathway begins with the oxidation of Glucose-6-phosphate (G6P), a key intermediate of glycolysis, by the enzyme glucose-6-phosphate dehydrogenase (G6PDH). This enzyme oxidizes G6P through the coupled reduction of the electron carrier NADP+ to make NADPH. Enzymes called **transaldolases** and **transketalases** are used to produce the intermediates within the pathway. The net result is oxidation and subsequent decarboxylation of glucose to form a pentose. The total reaction involves three glucose-6-phosphate (in green) molecules being oxidized to form three CO_2 molecules, one glyceraldehyde-phosphate (in red), and two hexose-phosphates (in red). In this cycle, the formed glyceradehyde-phosphate feeds into glycolysis and the two hexose-phosphates (e.g., glucose-phosphates) can recycle into the PPP or glycolysis.

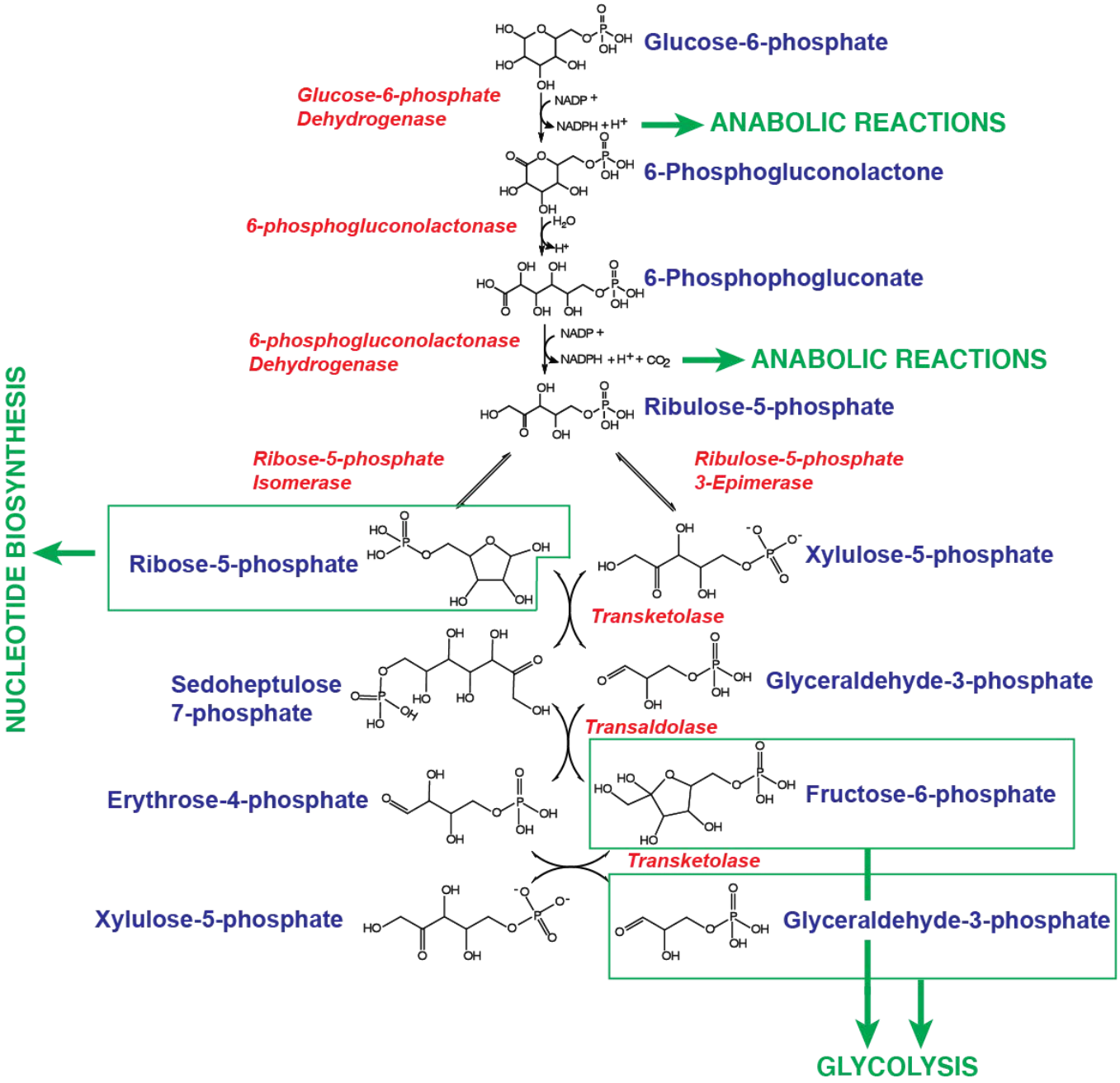


Figure 1. **The *Pentose Phosphate Pathway (PPP).*** The Pentose Phosphate Pathway diverts the glucose-6-phosphate from glycolysis, oxidizes the sugar to produce NADPH for anabolic reactions, creates ribose-5-phosphate for nucleotide biosynthesis, and other key 3, 4 and 7 carbon intermediates. Two of the products, fructose-6-phosphate and glyceraldehyde-3-phosphate reenter glycolysis. Compound names are blue while enzyme names are red italic font. *Attribution: Marc T. Facciotti*

As shown in the figure above, products of the pathway include glyceraldehyde-3-phosphate. This sugar can then be further oxidized via glycolysis. Fructose-6-phosphate that can reenter glycolysis and NADPH, a reductant for many biosynthetic (anabolic) reactions is also made. In addition, the pathway provides a variety of intermediate sugar-phosphates that the cell requires, such as pentose-phosphates (for nucleotides and some amino acids), erythrose-phosphate (for amino acids) and sedohepulose-phosphate (for gram-negative bacteria). The figure below illustrates the input-output relationship between PPP and the "top half" of glycolysis.

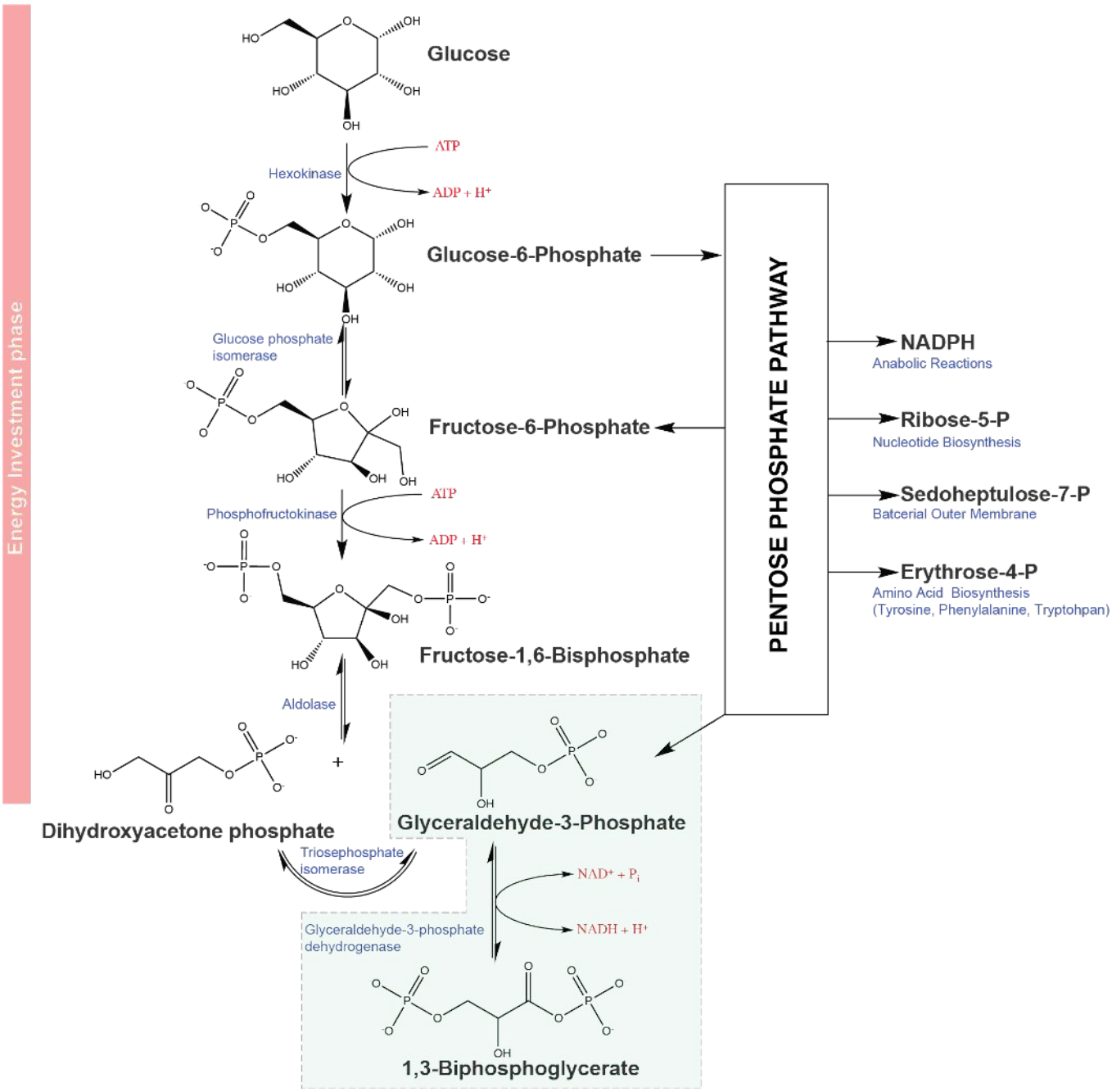


Figure 2. **The relationship between glycolysis and PPP.** The Pentose Phosphate Pathway is shown as a "shunt" (alternative metabolic pathway) for glucose-6-phosphate. *Attribution: Marc T. Facciotti*

This is a very short reading and should give you some space and time to start incorporating/integrating the lessons of this section on metabolism. Also since there is little room for commenting in NB on the content above, I suggest having some discussions with you classmates in NB covering some of the following questions:

1. Prof. Facciotti asked us to frame the discussion on metabolism around two key questions. How/where does an organism get its building blocks to make new molecules? How/where does an organism get its energy? How has/can/does this framing help you to shape how you approach your studies of metabolism so far? (Perhaps recall that you also had an exercise in discussion section with different types of metabolic modes that can be framed in this way).

2. The exchange of electrons between molecules is a universal (used by all living systems) process involved in energy "harvesting" and flow in an organism - from the simplest single-celled microbe to the most complex organism on the planet, all use electron transfer reactions to manage energy. How has this idea impacted your thinking about biology and how energy is gathered and flows? Is this a new perspective for you? A different perspective? What insight has it given you? How can you frame your studies of metabolism around this idea?

3. We watched a clip of a Veritasium movie in class where the host asks folks "what do we get from the sun?". Reflect back on that question. What is it that we get from the sun? How does that idea tie in to the flow of electrons we discussed earlier? What about the whole concept of what life "is"? What does this imply about what life "does" on the planet and how it may have managed to get more and more complex and abundant on Earth?

4. Any other notes you might want to discuss?

5. Another anchor point for added discussion.

Reading_16 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_17

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_17

MS.26 Identify a nucleotide from its molecular structure and be able to decompose the molecule into three main functional units: nitrogenous base, ribose, and phosphates.

GI.1 Create illustrations that serve as models of the three-dimensional structures of DNA. These should span several levels of detail and abstraction that include cartoon models of:

(a) three nucleotides and their phosphodiester bonds;

(b) two antiparallel strands showing hydrogen bonds between nucleotides in opposite strands; and

(c) a zoomed out representation of the polymer's helical structure.

GI.2 Name the nitrogenous bases that naturally form hydrogen bonds to one another in nucleic acids and identify instances of complementary base pairing ("base-pairing rules").

GI.38 Describe a genome as the complete collection of heritable nucleotides whose sequence can be increasingly quickly and inexpensively determined and annotated with rapidly evolving instrumentation (micro and nano) and computational technologies.

GI.13 Formulate and argument for why it is important to understand how to read and interpret the information encoded within genomes.

GI.15 Describe the diversity of genome sizes, genomic content, and genome organization that has been found in nature.

GI.8 Dissect the chemical reaction responsible for DNA synthesis and tell its energy story.

Nucleic acids



There are two types of nucleic acids in biology: DNA and RNA. DNA carries the heritable genetic information of the cell and is composed of two antiparallel strands of nucleotides arranged in a helical structure. Each nucleotide subunit is composed of a pentose sugar (deoxyribose), a nitrogenous base, and a phosphate group. The two strands associate via hydrogen bonds between chemically complementary nitrogenous bases. Interactions known as "base stacking" interactions also help stabilize the double helix. By contrast to DNA, RNA can be either be single stranded, or double stranded. It too is composed of a pentose sugar (ribose), a nitrogenous base, and a phosphate group. RNA is a molecule of many tricks. It is involved in protein synthesis as a messenger, regulator, and catalyst of the process. RNA is also involved in various other cellular regulatory processes and helps to catalyze some key reactions (more on this later). With respect to RNA, in this course we are primarily interested in (a) knowing the basic molecular structure of RNA and what distinguishes it from DNA, (b) understanding the basic chemistry of RNA synthesis that occurs during a process called transcription, (c) appreciating the various roles that RNA can have in the cell, and (d) learning the major types of RNA that you will encounter most frequently (i.e. mRNA, rRNA, tRNA, miRNA etc.) and associating them with the processes they are involved with. In this module we focus primarily on the chemical structures of DNA and RNA and how they can be distinguished from one another.

Nucleotide structure

The two main types of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA and RNA are made up of monomers known as **nucleotides**. Individual nucleotides condense with one another to form a **nucleic acid** polymer. Each nucleotide is made up of three components: a nitrogenous base (for which there are five different types), a pentose sugar, and a phosphate group. These are depicted below. The main difference between these two types of nucleic acids is the presence or absence of a hydroxyl group at the C_2 position, also called the 2' position (*read "two prime"*), of the pentose (see Figure 1 legend and the section on the pentose sugar for more on carbon numbering). RNA has a hydroxyl functional group at that 2' position of the pentose sugar; the sugar is called ribose, hence the name ribonucleic acid. By contrast, DNA lacks the hydroxyl group at that position, hence the name, "deoxy" ribonucleic acid. DNA has a hydrogen atom at the 2' position.

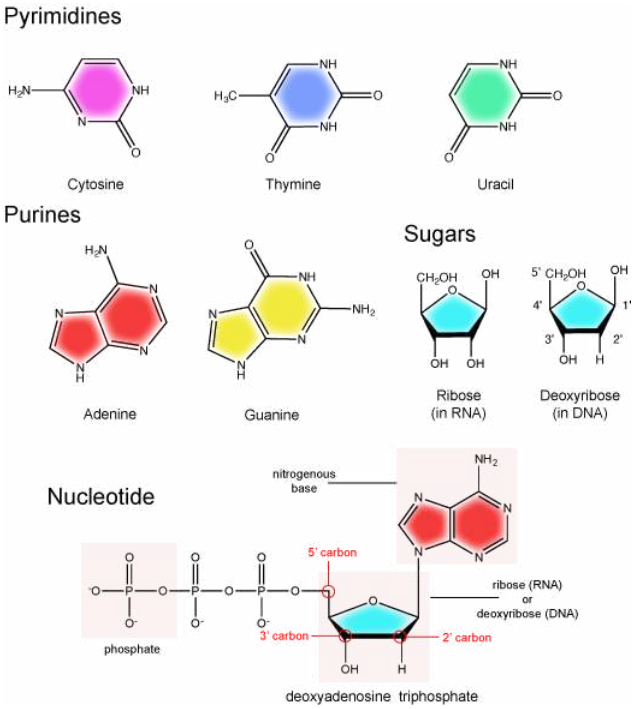


Figure 1. *A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and one or more phosphate groups. Carbons in the pentose are numbered 1' through 5' (the prime distinguishes these residues from those in the base, which are numbered without using a prime notation). The base is attached to the 1' position of the ribose, and the phosphate is attached to the 5' position. When a polynucleotide is formed, the 5' phosphate of the incoming nucleotide attaches to the 3' hydroxyl group at the end of the growing chain. Two types of pentose are found in nucleotides, deoxyribose (found in DNA) and ribose (found in RNA). Deoxyribose is similar in structure to ribose, but it has an -H instead of an -OH at the 2' position. Bases can be divided into two categories: purines and pyrimidines. Purines have a double ring structure, and pyrimidines have a single ring. Attribution: Marc T. Facciotti (original work)*

The nitrogenous base

The nitrogenous bases of nucleotides are organic molecules and are so named because they contain carbon and nitrogen. They are bases because they contain an amino group that has the potential of binding an extra hydrogen, and thus acting as a base by decreasing the hydrogen ion concentration in the local environment. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G), cytosine (C), and thymine (T). By contrast, RNA contains adenine (A), guanine (G) cytosine (C), and uracil (U) instead of thymine (T).

Adenine and guanine are classified as **purines**. The primary distinguishing structural feature of a purine is double carbon-nitrogen ring. Cytosine, thymine, and uracil are classified as **pyrimidines**. These are structurally distinguished by a single carbon-nitrogen ring. You will be expected to recognize that each of these ring structures is decorated by functional groups that may be involved in a variety of chemistries and interactions.

Possible NB Discussion Point

Take a moment to review the five nitrogenous bases in Figure 1 above. Identify functional groups as described in class. For each functional group identified, describe what type(s) of chemistry you expect it to be involved in. Will the functional group act as a hydrogen bond donor, acceptor, or both?

The pentose sugar

The pentose sugar contains five carbon atoms. Each carbon atom of the sugar molecule are numbered as 1′, 2′, 3′, 4′, and 5′ (1′ is read as “one prime”). The two main functional groups that are attached to the sugar are often named in reference to the carbon to which they are bound. For example, the phosphate residue is attached to the 5′ carbon of the sugar and the hydroxyl group is attached to the 3′ carbon of the sugar. We will often use the carbon number to refer to functional groups on nucleotides so be very familiar with the structure of the pentose sugar.

The pentose sugar in DNA is called deoxyribose, and in RNA, the sugar is ribose. The difference between the sugars is the presence of the hydroxyl group on the 2' carbon of the ribose and its absence on the 2' carbon of the deoxyribose. You can, therefore, determine if you are looking at a DNA or RNA nucleotide by the presence or absence of the hydroxyl group on the 2' carbon atom —you will probably be asked to do so on numerous occasions, including exams.

The phosphate group

There can be anywhere between one and three phosphate groups bound to the 5' carbon of the sugar. When one phosphate is bound, the nucleotide is referred to as a Nucleotide MonoPhosphate (NMP). If two phosphates are bound the nucleotide is referred to as Nucleotide DiPhosphate (NDP). When three phosphates are bound to the nucleotide it is referred to as a Nucleotide TriPhosphate (NTP). The phosphoanhydride bonds between that link the phosphate groups to each other have specific chemical properties that make them good for various biological functions. The hydrolysis of the bonds between the phosphate groups is thermodynamically exergonic in biological conditions; Nature has evolved many mechanisms to couple this negative change in free energy to help drive many reactions in the cell. Figure 2 shows the structure of the nucleotide triphosphate Adenosine Triphosphate, ATP, which we will discuss in greater detail in other chapters.

Note: "high-energy" bonds

The term "high-energy bond" is used A LOT in biology. This term is, however, a verbal shortcut that can cause some confusion. The term refers to the amount of negative free energy associated with the hydrolysis of the bond in question. The water (or other equivalent reaction partner) is an important contributor to the energy calculus. In ATP, for instance, simply "breaking" a phosphoanhydride bond - say with imaginary molecular tweezers - by pulling off a phosphate would not be energetically favorable. We must therefore be careful not to say that breaking bonds in ATP is energetically favorable or that it "releases energy". Rather, we should be more specific, noting that the hydrolysis of the bond is energetically favorable. Some of this common misconception is tied to, in our opinion, the use of the term "high-energy bonds". While in Bis2a we have tried to minimize the use of the vernacular "high energy" when referring to bonds, trying instead to describe biochemical reactions by using more specific terms, as students of biology you will no doubt encounter the potentially misleading - though admittedly useful - shortcut "high-energy bond" as you continue in your studies. So, keep the above in mind when you are reading or listening to various discussions in biology. Heck, use the term yourself. Just make sure that you really understand what it refers to.

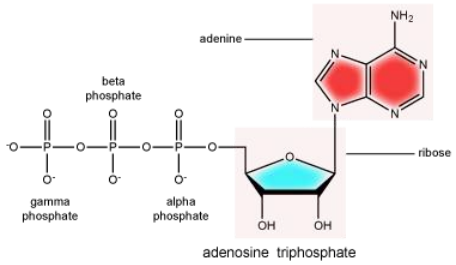
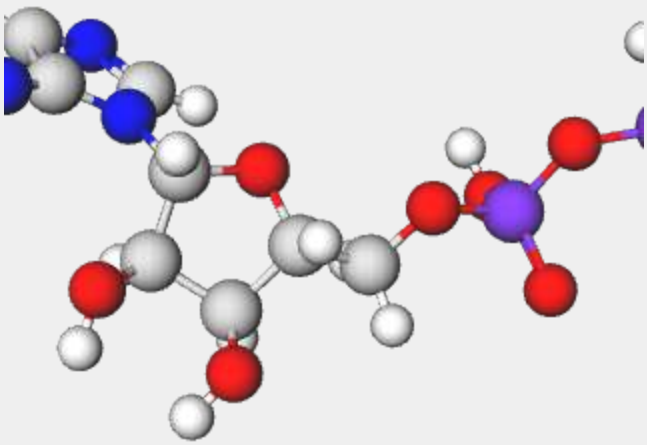


Figure 2. *ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate). Attribution: Marc T. Facciotti (original work)*

Adenosine triphosphate



Double helix structure of DNA

DNA has a double helix structure (shown below) created by two strands of covalently linked nucleotide subunits. The sugar and phosphate groups of each strand of nucleotides are positioned on the outside of the helix, forming the backbone of the DNA (highlighted by the orange ribbons in Figure 3). The two strands of the helix run in opposite directions, meaning that the 5′ carbon end of one strand will face the 3′ carbon end of its matching strand (See Figures 4 and 5). We referred to this orientation of the two strands as **antiparallel**. Note too that phosphate groups are depicted in Figure 3 as orange and red "sticks" protruding from the ribbon. The phosphates are negatively charged at physiological pHs and therefore give the backbone of the DNA a strong local negatively charged character. By contrast, the nitrogenous bases are stacked in the interior of the helix (these are depicted as green, blue, red, and white sticks in Figure 3). Pairs of nucleotides interact with one another through specific hydrogen bonds (shown in Figure 5). Each pair of separated from the next base pair in the ladder by 0.34 nm and this close stacking and planar orientation gives rise to energetically favorable base-stacking interactions. The specific chemistry associated with these interactions is beyond the content of Bis2a but is described in more detail here for the curious or more advanced students. We do expect, however, that students are aware that the stacking of the nitrogenous bases contributes to the stability of the double helix and defer to your upper-division genetics and organic chemistry instructors to fill in the chemical details.

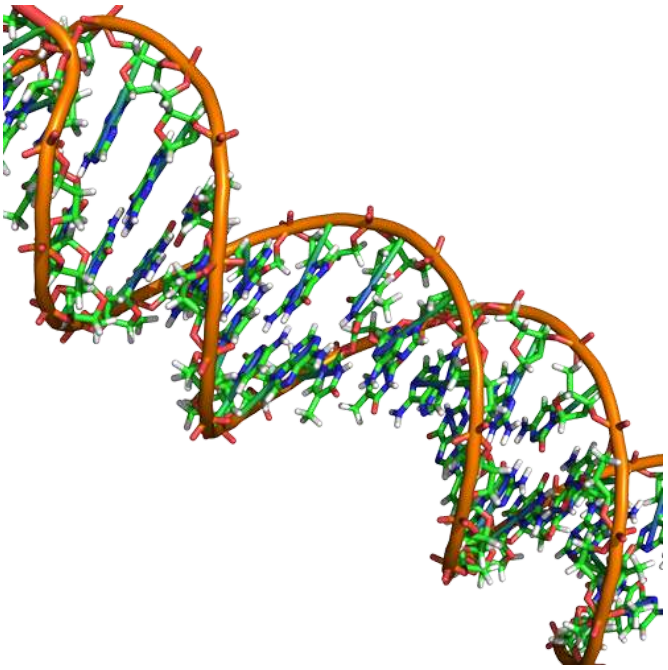


Figure 3. *Native DNA is an antiparallel double helix. The phosphate backbone (indicated by the curvy lines) is on the outside, and the bases are on the inside. Each base from one strand interacts via hydrogen bonding with a base from the opposing strand. Attribution: Marc T. Facciotti (original work)*

In a double helix, certain combinations of base pairing are chemically more favored than others based on the types and locations of functional groups on the nitrogenous bases of each nucleotide. In biology we find that:

Adenine (A) is chemically complementary with thymidine (T) (A pairs with T)

and

Guanine (G) is chemically complementary with cytosine (C) (G pairs with C).

We often refer to this pattern as "base complementarity" and say that the antiparallel strands are **complementary** to each other. For example, if the sequence of one strand is of DNA is 5'-AATTGGCC-3', the complementary strand would have the sequence 5'-GGCCAATT-3'.

We sometimes choose to represent complementary double-helical structures in text by stacking the complementary strands on top of on another as follows:

5' - GGCCAATTCCATACTAGGT - 3'

3' - CCGGTTAAGGTATGATCCA - 5'

Note that each strand has its 5' and 3' ends labeled and that if one were to walk along each strand starting from the 5' end to the 3' end that the direction of travel would be opposite the other for each strand; the strands are antiparallel. We commonly say things like "running 5-prime to 3-prime" or "synthesized 5-prime to 3-prime" to refer to the direction we are reading a sequence or the direction of synthesis. Start getting yourself accustomed to this nomenclature.

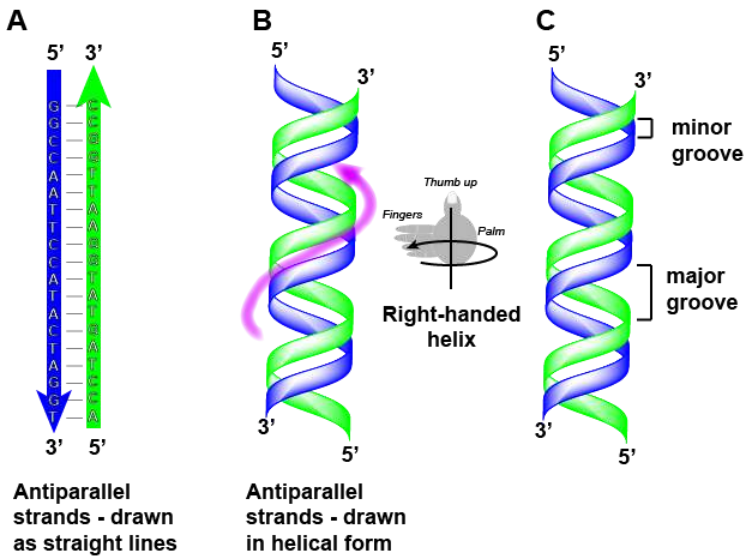


Figure 4. ***Panel A.*** *In a double-stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5′ to 3′ and the other 3′ to 5′. Here the strands are depicted as blue and green lines pointing in the 5′ to 3′ orientation. Complementary base pairing is depicted with a horizontal line between complementary bases.* ***Panel B.*** *The two antiparallel strands are depicted in double-helical form. Note that the orientation of the strands is still represented. Note that the helix is right-handed - the "curl" of the helix, depicted in purple, winds in the direction of the fingers of the hand if the right hand is used and the direction of the helix points towards the thumb.* ***Panel C.*** *This representation shows two structural features that arise from the assembly of the two strands called the major and minor grooves. These grooves can also be seen in Figure 3.*
Attribution: Marc T. Facciotti (original work)

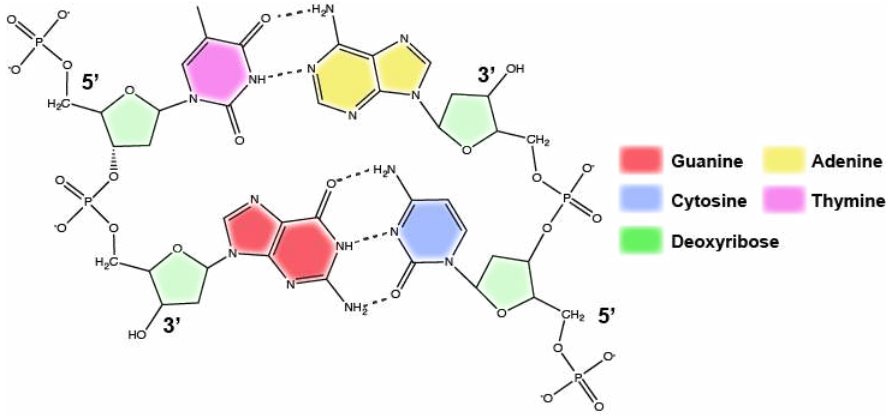


Figure 5. A zoomed-in molecular-level view of the antiparallel strands in DNA. *In a double-stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5′ to 3′ and the other 3′ to 5′. The phosphate backbone is located on the outside, and the bases are in the middle. Adenine forms hydrogen bonds (or base pairs) with thymine, and guanine base pairs with cytosine. Attribution: Marc T. Facciotti (original work)*

Functions and roles of nucleotides and nucleic acids to look out for in General Biology

Besides their structural roles in DNA and RNA, nucleotides such as ATP and GTP also serve as mobile energy carriers for the cell. It surprises some students when they learn to appreciate that the ATP and GTP molecules we discuss in bioenergetics are the same as those involved in the formation of nucleic acids. We will cover this in more detail when we discuss DNA and RNA synthesis reactions. Nucleotides also play important roles as co-factors in many enzymatically catalyzed reactions.

Nucleic acids, RNA in particular, play a variety of roles in in cellular process besides being information storage molecules. Some roles that you should keep an eye out for as we progress through the course include: (a) **Riboprotein** complexes - RNA-Protein complexes in which the RNA serves both catalytic and structural roles. Examples of such complexes include, ribosomes (rRNA), RNases, splicesosome complexes, and telomerase. (b) Information storage and transfer roles. These roles include molecules like DNA, messenger RNA (mRNA), transfer RNA (tRNA). (c) Regulatory roles. Examples of these include various non-coding (ncRNA). Wikipedia has a comprehensive summary of the different known RNA molecules that we recommend browsing to get a better sense of the great functional diversity of these molecules.

Genomes as organismal blueprints

A genome is an organism's complete collection of heritable information stored in DNA. Differences in information content help to explain the diversity of life we see all around us. Changes to the information encoded in the genome are the primary drivers of the phenotypic diversity we see (and some we can't) around us that are filtered by natural selection, and they are thus the drivers of evolution. This leads to questions. If every cell in a multicellular organism contains the same sequence of DNA, how can there be different cell types (e.g., how can a cell in a liver be so different from a cell in the brain if they both carry the same DNA)? How do we read the information? How do we interpret what we read? How do we understand how all the "parts" we identify in the genome functionally interrelate? How is all of this related to the expression of traits? How do changes in the genome lead to changes in traits?

Determining a genome sequence

The information encoded in genomes provides important data for understanding life, its functions, its diversity, and its evolution. Therefore, a reasonable place to begin studies in biology would be to read the information content encoded in the genome(s) in question. A good starting point is to determine the sequence of nucleotides (A, G, C, T) and their organization into one or more independently replicating units of DNA (e.g., think chromosomes and/or plasmids). For 30+ years after the discovery that DNA is the hereditary material, this was a daunting proposition. In the late 1980s, however, the advent of semi-automated tools for DNA sequencing were pioneered, and this began a revolution that has dramatically changed how we approach the study of life. Twenty years later, in the mid-2000s, we entered a period of accelerated technological progress in which advances in materials sciences (particularly, advances in our ability to make things on a tiny scale), optics, electrical and computer engineering, bioengineering, and computer sciences have all converged to bring us dramatic increases in our capacity to sequence DNA and correspondingly dramatic decreases in the cost of sequencing DNA. A famous example to illustrate this point is to compare the changes in cost to sequence the human genome. The first draft of the human genome took nearly 15 years and $3 billion dollars to complete. Today, 10's of human genomes can be sequenced in a single day on a single instrument at a cost of less than $1000 each (the cost and time continue to decrease). Today, companies like Illumina, Pacific Biosciences, Oxford Nanopore, and others offer competing technologies that are driving down the cost and increasing the volume, quality, speed, and portability of DNA sequencing.

One of the very exciting elements of the DNA sequencing revolution is that it has required and continues to require contributions from biologists, chemists, materials scientists, electrical engineers, mechanical engineers, computer scientists and programmers, mathematicians and statisticians, product developers, and many other technical experts. The potential applications and implications of unlocking barriers to DNA sequencing have also engaged investors, business people, product developers, entrepreneurs, ethicists, policy makers, and many others to pursue new opportunities and to think about how to best and most responsibly use this growing technology.

The technological advances in genome sequencing have resulted in a virtual flood of complete genome sequences being determined and deposited into publicly available databases. You can find many of them at the National Center for Biotechnology Information. The number of available, completely sequenced genomes numbers in the tens of thousands—over 2,000 eukaryotic genomes, over 600 archaeal genomes, and nearly 12,000 bacterial genomes at the time of this writing. Tens of thousands of additional genome sequencing projects are in progress. With this many genome sequences available—or soon to be available—we can start asking many questions about what we see in these genomes. What patterns are common to all genomes? How many genes are encoded in genomes? How are these organized? How many types of features can we find? What do the features we find do? How different are

the genomes from one another? Is there evidence that can tell us how genomes evolve? Let's briefly examine a few of these questions.

Diversity of genomes

Diversity of sizes, number of genes, and chromosomes

Let's start by examining the range of genome sizes. In the table below, we see a sampling of genomes from the database. We can see that the genomes of free-living organisms range tremendously in size. The smallest known genome is encoded by 580,000 base pairs while the largest is 150 billion base pairs—for reference, recall that the human genome is 3.2 billion base pairs. That's a huge range of sizes. Similar disparities in the number of genes also exist.

Table 1. *This table shows some genome data for various organisms. 2n = diploid number. Attribution: Marc T. Facciotti (own work—reproduced from http://book.bionumbers.org/how-big-are-genomes/)*

Organism	Genome size (base pairs)	Protein coding genes	Number of chromosomes
model organisms			
model bacteria *E. coli*	4.6 Mpb	4,300	1
budding yeast *S. cerevisiae*	12 Mbp	6,600	16
fission yeast *S. pombe*	13 Mbp	4,800	3
amoeba *D. discoideum*	34 Mbp	13,300	6
nematode *C. elegans*	100 Mbp	20,000	12 (2n)
fruit fly *D. melanogaster*	140 Mbp	14,000	8 (2n)
model plant *A. thaliana*	140 Mbp	27,000	10 (2n)
moss *P. patens*	510 Mbp	28,000	27
mouse *M. musculus*	2.8 Gbp	20,000	40 (2n)
human *H. sapiens*	3.2 Gbp	21,000	46 (2n)
viruses			
hepatitis D virus (smallest known RNA virus)	1.7 Kb	1	ssRNA
HIV-1	9.7 kbp	9	2 ssRNA (2n)
influenza A	15 kbp	11	8 ssRNA
bacteriophage λ	49 kbp	66	1 dsDNA
Pandoravirus salinus (largest known viral genome)	2.8 Mbp	2500	1 dsDNA
organelles			
mitochondria - *H. sapiens*	16.8 kbp	13 (+22 tRNA +2 rRNA)	1
mitochondria - *S. cerevisiae*	86 kbp	8	1
chloroplast - *A. thaliana*	150 kbp	100	1
bacteria			
C. ruddy (smallest genome of an endosymbiont bacteria)	160 kbp	182	1
M. genitalium (smallest genome of a free living bacteria)	580 kbp	470	1
H. pylori	1.7 Mbp	1,600	1
Cyanobacteria S. elongatus	2.7 Mbp	3,000	1
methicillin-resistant *S. aureus* (MRSA)	2.9 Mbp	2,700	1
B. subtilis	4.3 Mbp	4,100	1
S. cellulosum (largest known bacterial genome)	13 Mbp	9,400	1
archaea			
Nanoarchaeum equitans (smallest parasitic archaeal genome)	490 kbp	550	1
Thermoplasma acidophilum (flourishes at pH < 1)	1.6 Mbp	1,500	1
Methanocaldococcus jannashii (from ocean bottom hydrothermal vents; pressure > 200 atm)	1.7 Mbp	1,700	1
Pyrococcus furious (optimal temp 100°C)	1.9 Mbp	2,000	1
Haloarcula marismortui (flourishes at 4.5M salt)	4.3 Mbp	4,242	4 (+5 other replicons)
eukaryotes - multicellular			
pufferfish *Fugu rubripes* (smallest known vertebrate genome)	400 Mbp	19,000	22
poplar *P. trichocarpa* (first tree genome sequenced)	500 Mbp	46,000	19
corn *Z. mays*	2.3 Gbp	33,000	20 (2n)
dog *C. familiaris*	2.4 Gbp	19,000	40
chimpanzee *P. troglodytes*	3.3 Gbp	19,000	48 (2n)
wheat *T. aestivum* (hexaploid)	16.8 Gbp	95,000	42 (2n = 6x)
marbled lungfish *P. aethiopicus* (largest known animal genome)	130 Gbp	unknown	34
herb plant *Paris japonica* (largest known genome)	150 Gbp	unknown	40 (2n)

Adapted from http://book.bionumbers.org/how-big-are-genomes/
kbp = 1000's of base pairs; Mbp = millions of base pairs; Gbp = billions of base pairs

Examining Table 1 also reveals that some organisms carry with them more than one chromosome. Some genomes are also *polyploid*, meaning that they maintain multiple copies of similar but not identical (*homologous*) copies of each chromosome. A diploid organism carries in its genome two homologous copies (usually one from Mom and one from Dad) of each chromosome. Humans are diploid. Our somatic cells carry 2 homologous copies of 23 chromosomes. We received 23 copies of individual

chromosomes from our mother and 23 copies from our father, for a total of 46. Some plants have a higher ploidy. For example, we term a plant with four homologous copies of each chromosome *tetraploid*. We term an organism with a single copy of each chromosome *haploid*.

Structure of genomes

Table 1 also provides clues to other points of interest. For instance, if we compare the pufferfish genome to the chimpanzee genome, we note that they encode roughly the same number of genes (19,000), but they do so on dramatically differently sized genomes—400 million base pairs versus 3.3 billion base pairs, respectively. That implies that the pufferfish genome must have much less space between its genes than what we might expect to find in the chimpanzee genome. This is the case, and the difference in gene density is not unique to these two genomes. If we look at Figure 1, which attempts to represent a 50-kb part of the human genome, we notice that besides the protein-coding regions (indicated in red and pink) that many other so-called "features" can be read from the genome. Many of these elements contain highly repetitive sequences.

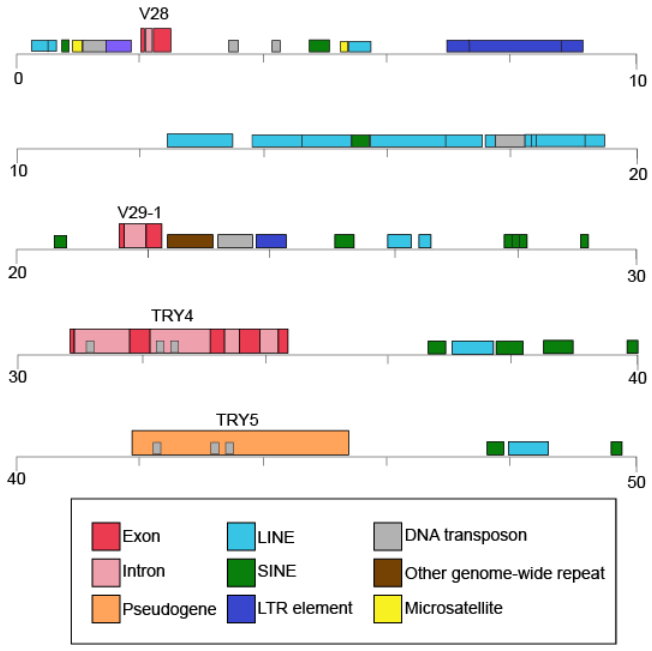


Figure 1. *This figure shows a 50-kb segment of the human β T-cell receptor locus on chromosome 7. This figure depicts a small region of the human genome and the types of "features" that can be read and decoded in the genome, including, but also besides, protein-coding sequences. Red and pink correspond to regions that encode proteins. Other colors represent different types of genomic elements. Attribution: Marc T. Facciotti (own work—reproduced from www.ncbi.nlm.nih.gov/books/NBK21134/)*

If we now look at what fraction of the whole human genome each of these types of elements makes up (see Figure 2), we see that protein-coding genes only make up 48 million of the 3.2 billion bases of the haploid genome.

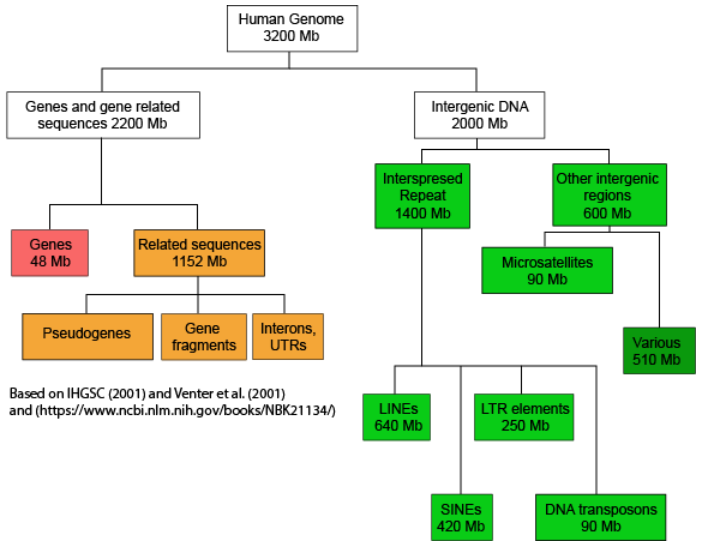


Figure 2. *This graph depicts how the many base pairs of DNA in the human haploid genome are distributed between various identifiable features. Note that only a small fraction of the genome is associated directly with protein-coding regions. Attribution: Marc T. Facciotti (own work—reproduced from sources noted in figure)*

When we examine the frequency of repeat regions versus protein-coding regions in different species, we note large differences in protein-coding versus non-coding regions.

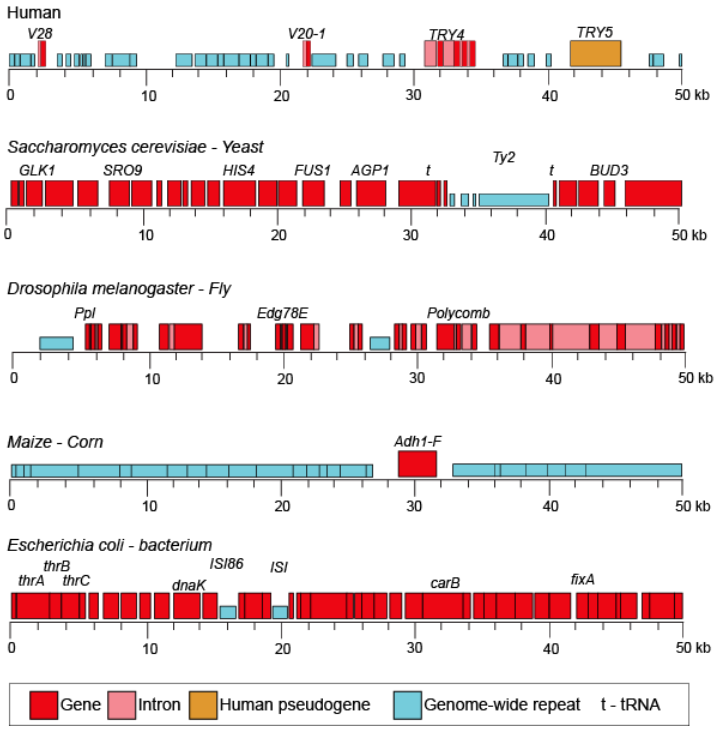


Figure 3. This figure shows *50-kb segments of different genomes, illustrating the highly variable frequency of repeat versus protein-coding elements in different species. Attribution: Marc T. Facciotti (own work—reproduced from www.ncbi.nlm.nih.gov/books/NBK21134/)*

Possible NB Discussion Point

Propose a hypothesis for why you think some genomes might have more or fewer noncoding sequences.

Dynamics of genome structure

Genomes change over time, and many events can change their sequence.

1. **Mutations** are accumulated during DNA replication or through environmental exposure to chemical mutagens or radiation. These changes typically occur at the level of single nucleotides.
2. **Genome rearrangements** describe a class of large-scale changes that can occur, and they include: (a) deletions—where segments of the chromosome are lost; (b) duplication—where regions of the chromosome are inadvertently duplicated; (c) insertions—the insertion of genetic material (note that sometimes this is acquired from viruses or the environment, and deletion/insertion pairs may happen across chromosomes); (d) inversions—where regions of the genome are flipped within the same chromosome; and (e) translocations—where segments of the chromosome are translocated (moved elsewhere in the chromosome).

These changes happen at different rates, and some are facilitated by the activity of enzyme catalysts (e.g., transposases).

Possible NB Discussion Point

How might mutations and genome rearrangements complicate studying/analyzing genomes? Conversely, can you think of interesting questions we can ask by comparing variation between genomes that occur because of events like mutation and genome rearrangements?

The study of genomes

Comparative genomics

One of the most common things to do with a collection of genome sequences is to compare the sequences of multiple genomes to one another. In general terms, these types of activities fall under the umbrella of a field called *comparative genomics*.

Comparing the genomes of people who suffer from an inheritable disease to the genomes of people who are not afflicted can help us uncover the genetic basis for the malady. Comparing the gene content, order, and sequence of related microbes can help us find the genetic basis of why some microbes cause disease while their close cousins are virtually harmless. We can compare genomes to understand how a new species may have evolved. There are many possible analyses! The basis of these analyses is similar: look for differences across multiple genomes and try to associate those differences with different traits or behaviors in those organisms.

Last, some people are comparing genome sequences to understand the evolutionary history of the organisms. Typically, these types of comparisons result in a graph known as a phylogenetic tree, which is a graphical model of the evolutionary relationship between the various species being compared. This field, not surprisingly, is called *phylogenomics*.

Metagenomics: Who is living somewhere and what are they doing?

Besides studying the genomes of individual species, the increasingly powerful DNA-sequencing technologies are making it possible to sequence simultaneously the genomes of environmental samples inhabited by many species. This field is called *metagenomics*. These studies are typically focused on trying to understand what microbial species inhabit different environments. There is great interest in using DNA sequencing to study the populations of microbes in the gut and to watch how the population changes in response to different diets, to see if there is any association between the abundance of different microbes and various

diseases, or to look for pathogens. People are using DNA sequencing of environmental metagenomic samples to explore which microbes inhabit different environments on Earth (from the deep sea, to soil, to air, to hypersaline ponds, to cat feces, to some common surfaces we touch every day).

Besides discovering "who lives where," the sequencing of microbial populations in different environments can also reveal what protein-coding genes are present in an environment. This can give investigators clues into what metabolic activities might occur in that environment. Besides providing important information about what kind of chemistry might happen in a specific environment, the catalog of genes that is accumulated can also serve as an important resource for the discovery of novel enzymes for applications in biotechnology.

Reading_17 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_18

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_18

GI.9 Create a dynamic model of DNA replication for the leading strand, including key sites on the DNA, reactants, products, enzymes, and the depiction of energy requirements for each step. Use the Design Challenge rubric to identify problems/challenges that must be overcome to replicate DNA and describe how the key parts in your figure help to resolve them.
GI.10 Repeat GI.9 for the lagging strand, include the design challenge associated with replicating the lagging strand template, and the solutions afforded by Okazaki fragments and DNA ligase.
GI.11 Describe the challenges of replicating the ends of linear chromosomes. Contrast these challenges with prokaryotic chromosome replication.
GI.12 Create a depiction of the telomerase-mediated telomere-lengthening mechanism that explicitly describes the role of structural RNA in the enzyme complex.
GI.14 Explain how DNA replication is a source of genetic variation, and describe the cellular mechanisms that can change the frequency of accumulating mutation and the trade-offs associated with fixing errors.

The DNA Double Helix and its Replication

The Problem

In this module, we discuss the replication of DNA—one of the key requirements for a living system to regenerate and create the next generation. Let us first briefly consider the problem through a literary analogy.

The human genome comprises roughly 6.5 billion base pairs of DNA if one considers the full diploid genome (i.e., if you count the DNA inherited from both parents). Six point five billion looks like this: 6,500,000,000. That's a large number. To get a better idea of what that number means, imagine that our DNA is a set of written instructions for constructing one of us. By analogy, we can then compare it to another written document. For this example, we begin by considering Tolstoy's *War and Peace*, a novel many people are familiar with for its voluminous nature. Data from Wikipedia estimates that *War and Peace* contains about 560,000 words. A second written work many are familiar with are the seven volumes of J.K. Rowling's *Harry Potter*. This work checks in at ~1,080,000 words (Referenced Statistics on Wikipedia). If we assume that the length of the average English word is five characters, the two literary works are 2.8 million and 5.4 million characters, respectively. Therefore, even all seven volumes of "Harry Potter" have over 1000x fewer characters than our own genomes. The number of characters in these novels are, however, much closer to the number of nucleotides in a typical bacterial genome.

Now imagine for a moment developing a machine or mechanical process (not an electronic process) that reads and copies these books. Or imagine yourself copying these texts. How fast could you do it? How many mistakes are you likely to make? Do you expect there to be a trade-off between the speed at which you can copy and the accuracy? What type of resources does this process need? How much energy does the process require? Now imagine copying something 1000x larger! Oh, and just for good measure, your imaginary mechanical device needs to do its work on text that is ~25Å wide (i.e., 0.0000000025 meters wide). By comparison, a typical ten point font is ~0.00025 meters wide, about 100,000x larger than the width of a DNA base pair.

With that in mind, it is worth noting that a human cell can take about 24 hours to divide (DNA replication must therefore be a little faster). A healthy *E. coli* cell may take only 20 minutes to divide (including replicating its ~4.5 million base pair genome). Both the human and bacterium do this while typically making few enough mistakes that the subsequent generation remains viable and recognizable. That should seem rather amazing! Now consider that we estimate the human body comprises ~10 trillion cells (10,000,000,000,000) and that it may have between two and ten times that number of microbial residents. That's a lot of cell division to consider.

Design challenge

If the cell is to replicate—its ultimate goal—a copy of the DNA must be created. So one clear problem statement/question is "how can the cell effectively copy its DNA?" Given the analogy above, here are some relevant sub-questions: What are the chemical and physical properties that enable DNA to be copied? With what fidelity must the organism copy its genome? What speed must it be copied at? Where does the energy come from for this task and how much is necessary? Where do the "raw materials" come from? How do the molecular machines involved in this process couple the assembly of raw materials and the energy required to build a new DNA molecule together? The list could, of course, go on.

In the following discussion and in the lecture, we examine how the process of DNA replication is accomplished while keeping in mind some driving questions. As you go through the reading and lecture materials, try to be constantly aware of these and other questions associated with this process. Use these questions as guideposts for organizing your thoughts and try to find matches between the "facts" that you think we might expect you to know and the driving questions.

The DNA double helix



To build some extra context, we also need a little of empirically determined knowledge. Perhaps one of the best-known and popular features of the hereditary form of the DNA molecule is that it has a double helical tertiary structure. Our appreciation of the double-helical structure of DNA dates to the 1950s. For more on this story, see the short film here. <https://www.biointeractive.org/class...s/double-helix>.

Models of the structure of DNA revealed that molecule comprises two strands of covalently linked nucleotides that are twisted around each other to form a right-handed helix. In each strand, nucleotides are covalently joined to two other nucleotides (except at the very ends of a linear strand) via phosphodiester bonds that link the sugars via the 5' and 3' hydroxyl groups (panel b in Figure 1). Recall that the labels 5' and 3' refer to the carbons on the sugar molecule. These sugars and phosphate chains form a contiguous set of covalent links that are often referred to as the "backbone" of the structure. In a linear molecule, each strand has two free ends. We call one free end the 5' end because the unlinked functional group that is typically involved in joining nucleotides is the phosphate linked to the 5' carbon. We call the other end of the strand the 3' end because the unlinked functional group that is typically involved in joining nucleotides is the hydroxyl group linked to the 3' carbon of the sugar. Since the two ends of the strand are not symmetrical, this makes it easy to designate a direction one the strand—one can, for instance, say that they are reading from the 5' end to 3' end to indicate that they are "walking" along the strand starting at the 5' end and moving towards the 3' end. This direction (5' to 3') is the convention used by most biologists. One can read in the opposite direction (3' to 5') provided we make the direction explicit. We find the two strands of covalently linked nucleotides to be **anti-parallel** to one another in the double-helix; that is, the orientation/direction of one strand is opposite to that of the other strand (panel b in Figure 1). The backbone is structurally on the "outside" of the double helix, creating a band of negative charges on the surface. The nitrogenous bases of each of the antiparallel strands stack on the inside of the structure and oppose one another in a way that allows hydrogen bonds between unique purine/pyrimidine pairs (A pairing with T and G pairing with C) to form. We call these specific base pairings **complementary base pairs**. Thus, we refer to the paired strands of a double helix as **complementary strands**.

Complementary strands carry redundant information. Because of the strict chemical pairing, if you know the sequence of one strand, you obligatorily know the strand of its complement. Take, for example, the sequence 5′- C A T A T G G G A T G - 3′. Note how the sequence is annotated with the orientation (indicated by 5' and 3' labels). The complement of this sequence—written according to the 5' to 3' convention is: 5′- C A T C C C A T A T G - 3′. If you aren't convinced, write these two sequences out across from one another in your notes, writing them as antiparallel strands. Note that the twisting of the two complementary strands around each other results in the formation of structural features called the major and minor grooves that will become more important when we discuss the binding of proteins to DNA (panel c in Figure 1).

Most general biology instructors will expect you to recognize key structural features depicted in the figure below and that you will be able to create a basic figure of the structure of DNA yourself.

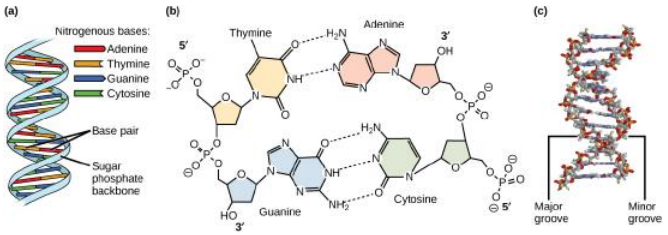


Figure 1. *DNA has (a) a double helix structure and (b) phosphodiester bonds. The (c) major and minor grooves are binding sites for DNA binding proteins during processes such as transcription (the creation of RNA from a DNA template) and replication.*

Possible NB Discussion Point

Take a moment to review the nitrogenous bases in Figure 1. Identify functional groups as described in class. For each functional group identified, describe what type of chemistry you expect it to be involved in. Try to identify whether the functional group can act as either a hydrogen bond donor, acceptor, or both?

At around the same time, three hypotheses for the modes of DNA replication were being considered. The models for replication were known as: the conservative model, the semi-conservative model, and the dispersive model.

1. Conservative: The conservative model of replication postulated that each whole double-stranded molecule could act as a template for the synthesis of a new double-stranded molecule. If one were to put a chemical tag on the template DNA molecule after replication, none of that tag would be found on the new copy.
2. Semi-conservative: This hypothesis stipulated that each individual strand of a DNA molecule could serve as a template for a new strand to which it would now associate with. In this case, if a chemical label were placed on a double-stranded DNA molecule, one strand on each of the copies would keep the label.
3. Dispersive: This model proposed that a copied double helix would be a piecewise combination of continuous segments of "old" and "new" strands. If a chemical label were placed on a DNA molecule that were copied using a dispersive mechanism, one would find discrete segments of the resulting copy that were labeled on both strands separated by completely unlabeled parts.

Meselson and Stahl resolved the issue in 1958 when they reported results of a now famous experiment (describe on Wikipedia) which showed that DNA replication is semi-conservative (Figure 2), where each strand is used as a template for the creation of the new strand. To learn more about this experiment, watch The Meselson-Stahl Experiment.

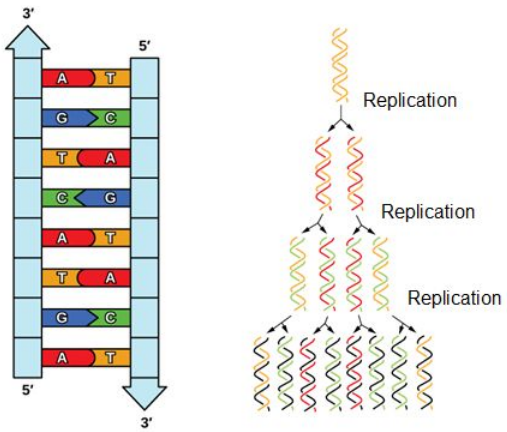


Figure 2. *DNA has an antiparallel double helix structure, the nucleotide bases are hydrogen bonded together, and each strand complements the other. DNA is replicated in a semi-conservative manner, each strand is used as the template for the newly made strand.*

DNA replication

Having established some basic structural features and the need for a semi-conservative mechanism, it is important to understand some of what we know about the process and to think about what questions one might want to answer if they are to better understand what is going on.

Since DNA replication is a process, we can invoke the energy story rubric to think about it. Recall that the energy story rubric is there to help us think systematically about processes (how things go from A to B). In this case the process in question is the act of starting with one double-stranded DNA molecule and ending up with two double-stranded molecules. So, we will ask a variety of questions: What does the system look like at the beginning (matter and energy) of replication? How are matter and energy transferred in the system, and what catalyzes the transfers? What does the system look like at the end of the process? We can also ask questions regarding specific events that MUST happen during the process. For instance, since DNA is a long molecule and it is sometimes circular, we can ask basic questions like, where does the process of replication start? Where does it end? We can also ask practical questions about the process like, what happens when a double-stranded structure is unwound?

We consider some of these key questions in the text and in class and encourage you to do the same.

Requirements for DNA replication

Let's start by listing some basic functional requirements for DNA replication that we can infer just by thinking about the process that must happen and/or be required for the replication to happen. So, what do we need?

- We know that DNA is composed of nucleotides. If we want to create a new strand, we will need a source of nucleotides.
- We can infer that building a new strand of DNA will require an energy source—we should try to find this.
- We can infer that that there must be a process for finding a place to start replication.
- We can infer that there will be one or more enzymes that help catalyze the process of replication.
- We can also infer that since this is a biochemical process, that it will make some mistakes.

Nucleotide structure review

Recall some basic structural features of the nucleotide building blocks of DNA. The nucleotides start off as nucleotide triphosphates. Nucleotides are composed of a nitrogenous base, deoxyribose (five-carbon sugar), and a phosphate group. We name the nucleotide according to its nitrogenous base, purines such as adenine (A) and guanine (G), or pyrimidines such as cytosine (C) and thymine (T). Recall the structures below. Note that the nucleotide Adenosine triphosphate (ATP) is a precursor of the deoxyribonucleotide (dATP) which is incorporated into DNA.

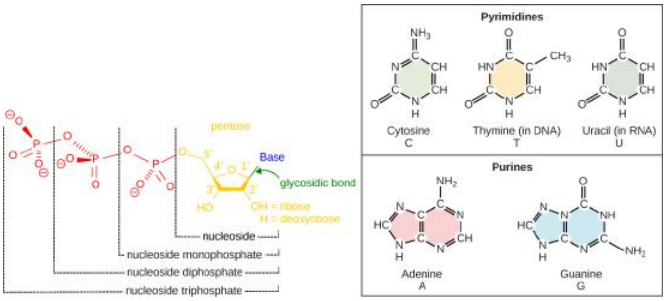


Figure 3. *Each nucleotide is made up of a sugar (ribose or deoxyribose depending on whether it builds RNA or DNA, respectively), a phosphate group, and a nitrogenous base. The purines have a double ring structure with a six-membered ring fused to a five-membered ring. Pyrimidines are smaller in size; they have a single six-membered ring structure. The carbon atoms of the five-*

carbon sugar are numbered 1', 2', 3', 4', and 5' (1' is read as "one prime"). The phosphate residue is attached to the hydroxyl group of the 5' carbon of one sugar of one nucleotide and the hydroxyl group of the 3' carbon of the sugar of the next nucleotide, thereby forming a 5'-3' phosphodiester bond.

Initiation of replication

Where along the DNA does the replication machinery start DNA replication?

With millions, if not billions, of nucleotides to copy how does the DNA polymerase know where to start? This process turns out not to be random. There are specific nucleotide sequences called **origins of replication** along the DNA at which replication begins. Once this site is identified, however, there is a problem. The DNA double helix is held together by base stacking interactions and hydrogen bonds. If each strand must be read and copied individually, there must be some mechanism responsible for helping to dissociate the two strands from one another. Energetically, this is an endergonic process. Where does the energy come from, and how is this reaction catalyzed? Basic reasoning should lead to the hypothesis that a protein catalyst is likely involved, and that this enzyme either creates new bonds that are energetically more favorable (exergonic) than the bonds it breaks AND/OR it can couple the use of an external energy source to help dissociate the strands.

It turns out that the details of this process and the proteins involved differ depending on the specific organism in question, and many of the molecular level details are not completely understood. There are, however, some common features in the replication of eukaryotes, bacteria, and archaea, and one of these features is that the process involves multiple different types of proteins in replicating DNA. First, proteins called "initiators" can bind DNA at or very near origins of replication. The interaction of the initiator proteins with the DNA helps to destabilize the double helix and also help to recruit other proteins, including an enzyme called a DNA **helicase** to the DNA. Here the energy required to destabilize the DNA double helix seems to come from the formation of new associations between DNA and the initiator proteins and the proteins themselves. The DNA helicase is a multi-subunit protein important in the process of replication because it couples the exergonic hydrolysis of ATP to the unwinding of the DNA double helix. Additional proteins must be recruited to the initiation complex (the collection of proteins involved in initiating transcription). These include, but are not limited to, additional enzymes called **primase** and **DNA polymerase**. While the initiators depart soon after the initiation of replication, the rest of the proteins work in concert to execute the process of DNA replication. This complex of enzymes function at Y-shaped structures in the DNA called **replication forks** (Figure 4). For any replication event, two replication forks can form at each origin of replication, extending in both directions. Multiple origins of replication can be found on eukaryotic chromosomes and some archaea, while the genome of the bacterium, *E. coli*, seems to encode one origin of replication.

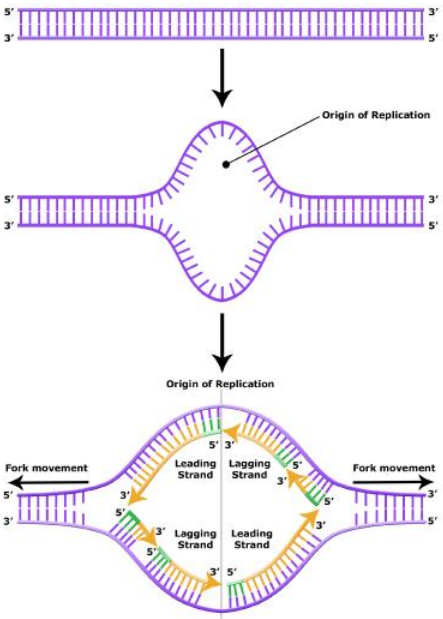


Figure 4. *At the origin of replication, a replication bubble forms. The replication bubble is composed of two replication forks, each traveling in opposite directions along the DNA. It is understood that the replication forks include all of the enzymes required for replication to occur—they are just not drawn explicitly in the figure to provide room to illustrate the relationships between the template and new DNA strands.*
Attribution: BIS2A team original image.

Elongation of replication

The melting open of the DNA double helix and the assembling the DNA replication complex is just the first step in the process of replication. Now the process of creating a new strand actually needs to get started. Here, we encounter additional challenges. The first obvious issue is that of determining which of the two strands should get copied at any replication fork (i.e., Which strand will serve as a template for semi-conservative synthesis? Are both strands equally viable alternatives?). There is also the problem of getting the process of the new strand synthesis started. Can the DNA polymerase start the new strand on its own? We will discuss later the answer to the latter question and some rationale and consequences. The key idea to note at this point is that it has been experimentally determined that DNA polymerase can NOT start strand synthesis on its own. Rather, DNA polymerase requires a short stretch of double-stranded structure followed by a single-stranded template. The enzyme **primase** creates a short oligonucleotide polymer of RNA (not DNA) called a **primer** (these are depicted by short green lines in the figures above and below). DNA polymerase uses the primer to nucleate and grow a new strand.

During the process of strand elongation, the **DNA polymerase** polymerizes a new covalently linked strand of DNA nucleotides (in bacteria this specific enzyme may be called DNA polymerase III; in eukaryotes, polymerase nomenclature is more complex and the roles of several polymerase proteins are not completely understood). It turns out that one strand is favored over the other to serve as a template. DNA polymerase will "read" the template strand from 3' to 5' and synthesize a new strand in the 5' to 3' direction. Hypotheses to explain this universal observation usually center on the energetics associated with the addition of a new nucleotide and arguments associated with DNA repair that we will describe shortly. Let us, therefore, briefly consider the reaction involving the addition of a single nucleotide. The primer provides an important 3' hydroxyl on which to begin synthesis. The next deoxyribonucleotide triphosphate enters the binding site of the DNA polymerase and, as shown in Figure 5 below, is oriented by

the polymerase such that a hydrolysis of the 5' triphosphate can occur. This reaction releases pyrophosphate and couples the exergonic hydrolysis of the phosphoanhydride to the synthesis of a phosphodiester bond between the 5' phosphate of the incoming nucleotide and the 3' hydroxyl group of the primer. This process repeats until deoxyribonucleotide triphosphates run out or the replication complex falls off of the DNA. In effect, DNA polymerase adds the phosphate group (5') from the incoming nucleotide to the existing hydroxyl group (3') of the previously added nucleotide.

Correct base pairing, or selection of correct nucleotide to add at each step, is accomplished by structural constraints felt by the DNA polymerase and the energetically favorable hydrogen bonds formed between complementary nucleotides. The process is energetically driven by the hydrolysis of the incoming 5' triphosphate and the energetically favorable interactions formed by the inter-nucleotide interactions in the growing double helix (base stacking and complementary base pairing hydrogen bonds). Note that the energetics of nucleotide addition do not technically prevent a strand growing in the 3' to 5' direction. The key difference in this “backwards” synthesis scheme is that the energy “source” for synthesis would need to come from a nucleotide already incorporated into the growing strand rather than the new incoming nucleotide (which this might be an important selective disadvantage is discussed briefly). After elongation has started a different DNA polymerase (in bacteria we usually call this enzyme DNA Polymerase I) comes in to remove the RNA primer and to synthesize the remaining bit of missing DNA.

As discussed in more detail in class, the movement of the replication fork induces winding of the DNA in both directions of replication. Another ATP consuming enzyme called **topoisomerase** helps to relieve this stress.

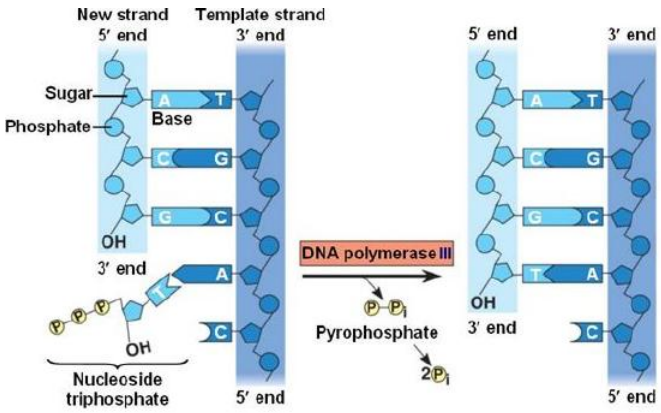


Figure 5. *DNA polymerase catalyzes the addition of the 5' phosphate group from an incoming nucleotide to the 3' hydroxyl group of the previous nucleotide. This process creates a phosphodiester bond between the nucleotides while hydrolyzing the phosphoanhydride bond in the nucleotide.*
Source: http://bio1151.nicerweb.com/Locked/m...h16/elong.html

Leading and lagging strand

The discussion above about strand elongation describes the process of new strand synthesis if that strand is synthesized in the same direction as the replication fork is or appears to be moving along the DNA. This strand can be synthesized continuously and is called the **leading strand**. However, both strands of the original DNA double helix must be copied. Since the DNA polymerase can only synthesize DNA in a 5' to 3' direction, the polymerization of the strand opposite of the leading strand must occur in the opposite direction that helicase, or front of the replication fork, is traveling. This strand is called the **lagging strand,** and due to geometric constraints, must be synthesized through a series of RNA priming and DNA synthesis events into short segments called **Okazaki fragments**. As noted, the initiation of synthesis of each Okazaki fragment requires a primase to synthesize an RNA primer, and each of these RNA primers must be removed and replaced with DNA nucleotides by a different DNA polymerase. The covalent bonds between each of the Okazaki fragment can not be made by the DNA polymerase and must therefore be formed by

yet another enzyme called **DNA ligase**. The geometry of lagging strand synthesis is difficult to visualize and will be covered in class.

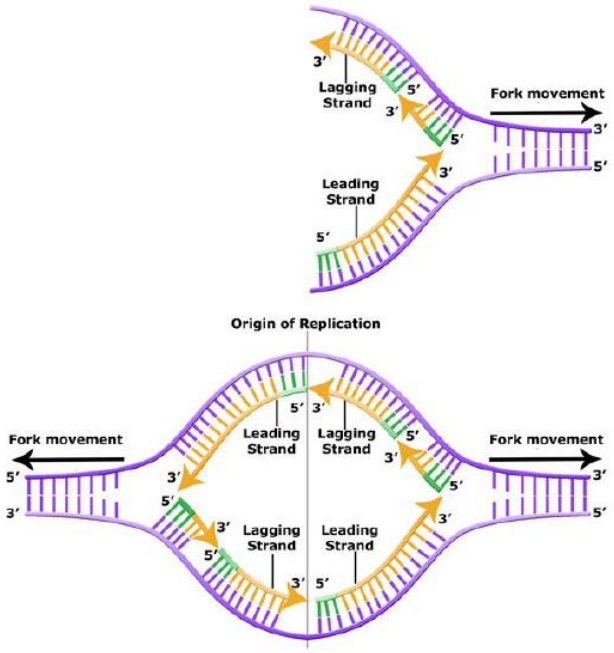


Figure 6. *The lagging strand is created in multiple segments. A replication fork shows the leading and lagging strand. A replication bubble shows the leading and lagging strands. BIS2A Team original image.*

Termination of replication

Telomeres and telomerase

The ends of replication in circular bacterial chromosomes pose few practical problems. However, the ends of linear eukaryotic chromosomes pose a specific problem for DNA replication. Because DNA polymerase can add nucleotides in only one direction (5' to 3'), the leading strand allows for continuous synthesis until the end of the chromosome is reached; however, as the replication complex arrives at the end of the lagging strand there is no place for the primase to "land" and synthesize an RNA primer so that the synthesis of the missing lagging strand DNA fragment at the end of the chromosome can be initiated by the DNA polymerase. Without some mechanism to help fill this gap, this chromosomal end will remain unpaired and will be lost to nucleases. Over time, and several rounds of replication, this would result in the ends of linear chromosomes getting progressively shorter, ultimately compromising the ability of the organism to survive. These ends of the linear chromosomes are known as **telomeres,** and nearly all eukaryotic species have evolved repetitive sequences that do not code for a specific gene. As a result, these "non-coding" telomeres act as replication buffers and are shortened with each round of DNA replication instead of critical genes. For example, in humans, a six base-pair sequence, TTAGGG, repeats 100 to 1000 times at the end of most chromosomes. Besides acting as a potential buffer, the discovery of the enzyme **telomerase** helped in understanding how chromosome ends are maintained. Telomerase is an enzyme composed of protein and RNA. Telomerase attaches to the end of the chromosome by complementary base pairing between the RNA component of telomerase and the DNA template. Telomerase uses the RNA as a complementary strand for the short elongation of its complement. The cell repeats this process many times. Once the lagging strand template is sufficiently elongated by telomerase, primase will create a primer followed by DNA polymerase which can now add nucleotides complementary to the ends of the chromosomes. Thus, the ends of the chromosomes are replicated.

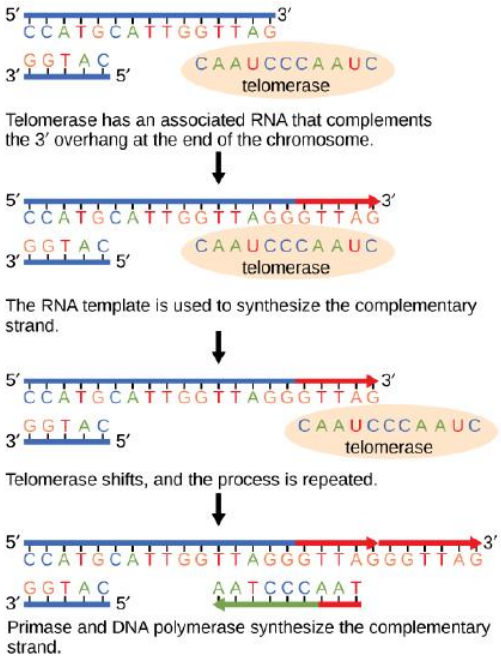


Figure 7. *The ends of linear chromosomes are maintained by the action of the telomerase enzyme.*

Telomerase is not active in adult somatic cells. Adult somatic cells that undergo cell division continue to have their telomeres shortened. This means that telomere shortening is associated with aging. In 2010, scientists found that telomerase can reverse some age-related conditions in mice, and this may have potential in regenerative medicine.[1] Telomerase-deficient mice were used in these studies; these mice have tissue atrophy, stem-cell depletion, organ system failure, and impaired tissue injury responses. Telomerase reactivation in these mice caused the extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved functioning of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

Possible NB Discussion Point

Imagine that researchers have invented a drug for humans that upregulates telomerase activity. Would you be excited or would you be skeptical about being part of a clinical trial? Do you trust this drug? Can you think of any negative consequences to activating telomerase to a level that it would not “naturally” be at? Assuming scientists proved without doubt that their drug has no negative side effects -- what are some of the ethical questions you would consider?

Differences in DNA replication rates between bacteria and eukaryotes

DNA replication has been extremely well studied in bacteria, primarily because of the small size of the genome and the large number of variants available. *E. coli* has 4.6 million base pairs in a single circular chromosome, and all of it gets replicated in approximately 42 minutes, starting from a single origin of replication and proceeding around the chromosome in both directions. This means that the polymerase adds approximately 1000 nucleotides per second to a growing strand. The process is much more rapid than in eukaryotes. Table 1 summarizes the differences between bacterial and eukaryotic replications.

Table 1. *Differences between bacterial and eukaryotic replication.*

Differences between prokaryotic and eukaryotic replication		
Property	Bacteria	Eukaryotes
Origin of replication	Single	Multiple
Rate of polymerization per polymerase	1000 nucleotides/s	50 to 100 nucleotides/s
Chromosome structure	Circular	Linear
Telomerase	Not present	Present

Link to external resources

Click through a tutorial on DNA replication.

Replication design challenge: proofreading



When the cell replicates its DNA, it does so in response to environmental signals that tell the cell it is time to divide. The ideal goal of DNA replication is to produce two identical copies of the double-stranded DNA template and to do it in an amount of time that does not pose an unduly high evolutionarily selective cost. This is a daunting task when you consider that there are ~6,500,000,000 base pairs in the human genome and ~4,500,000 base pairs in the genome of a typical *E. coli* strain and that Nature has determined that the cells must replicate within 24 hours and 20 minutes, respectively. In either case, many individual biochemical reactions need to take place.

While ideally replication would happen with perfect fidelity, DNA replication, like all other biochemical processes, is imperfect—bases may be left out, extra bases may be added, or bases may be added that do not properly base-pair. In many organisms, many of the mistakes that occur during DNA replication are corrected promptly by DNA polymerase itself via a mechanism known as *proofreading*. In *proofreading*, the DNA polymerase "reads" each newly added base via sensing the presence or absence of small structural anomalies before adding the next base to the growing strand. In doing so, a correction can be made.

If the polymerase detects that a newly added base has been paired correctly with the base in the template strand, it adds the next nucleotide. If, however, a wrong nucleotide is added to the growing polymer, the misshaped double helix will cause the DNA polymerase to stall, and it will eject the newly made strand from the polymerizing site. Consequently, the DNA strand will enter an exonuclease site. In this site, DNA polymerase can cleave off the last several nucleotides that were added to the polymer. Once the polymerase removes the incorrect nucleotides, the DNA strand can return to the polymerizing site and new nucleotides will be added again. This proofreading capability comes with some trade-offs: using an error-correcting/more accurate polymerase requires time (the trade-off is speed of replication) and energy (always an important cost to consider). The slower you go, the more accurate you can be. Going too slow, however, may keep you from replicating as fast as your competition, so figuring out the balance is key.

We call errors that are not corrected by proofreading *mutations*.

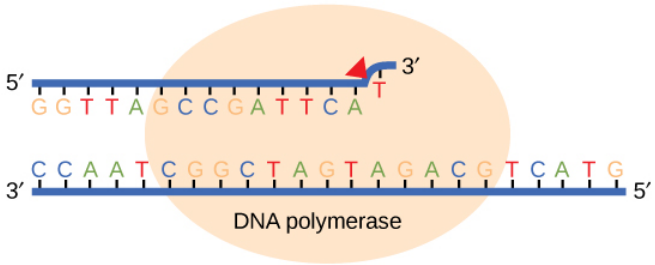


Figure 1. *Proofreading by DNA polymerase corrects errors during replication.*

Replication mistakes and DNA repair

Although DNA replication is typically a highly accurate process and proofreading DNA polymerases helps to keep the error rate low, mistakes still occur. Besides errors of replication, environmental damage may also occur to the DNA. Such uncorrected errors of replication or environmental DNA damage may lead to serious consequences. Therefore, Nature has evolved several mechanisms for repairing damaged or incorrectly synthesized DNA.

Mismatch repair

Errors not corrected during replication may instead get corrected after replication completes; we know this kind of repair as *mismatch repairs*. Specific enzymes recognize the incorrectly added nucleotide and excise it, replacing it with the correct base. But, how do mismatch repair enzymes recognize which of the two bases is the incorrect one?

In *E. coli*, after replication, the nitrogenous base adenine gains a methyl group; this means that directly after replication, the parental DNA strand will have methyl groups, whereas the newly synthesized strand lacks them. Thus, mismatch repair enzymes can scan the DNA and remove the wrongly incorporated bases from the newly synthesized, non-methylated strand by using the methylated strand as the "correct" template from which to incorporate a new nucleotide. In eukaryotes, the mechanism is not as well understood, but we believe it to involve recognition of unsealed nicks in the new strand, and a short-term, continuing association of some replication proteins with the new daughter strand after replication has completed.

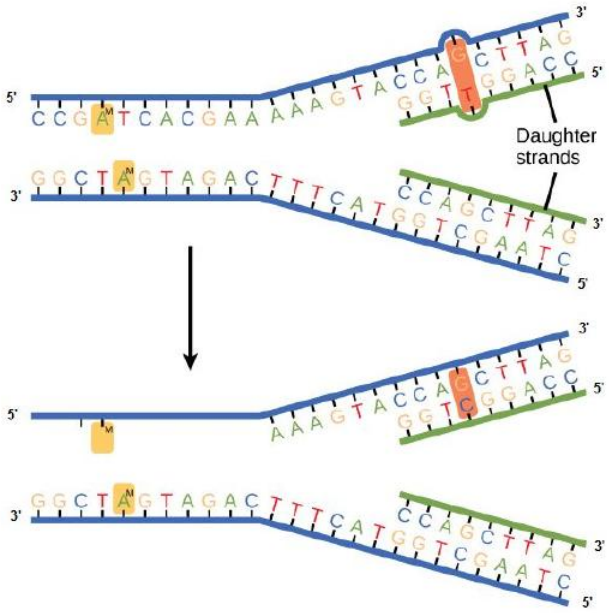


Figure 2. *In mismatch repair, the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base.*

Nucleotide excision repair

Nucleotide excision repair enzymes replace incorrect bases by making a cut on both the 3' and 5' ends of the incorrect base. The entire segment of DNA is removed and replaced with correctly paired nucleotides by the action of a DNA polymerase. Once the bases are filled in, it seal the remaining gap with a phosphodiester linkage catalyzed by the enzyme DNA ligase. This repair mechanism is often employed when UV exposure causes the formation of pyrimidine dimers.

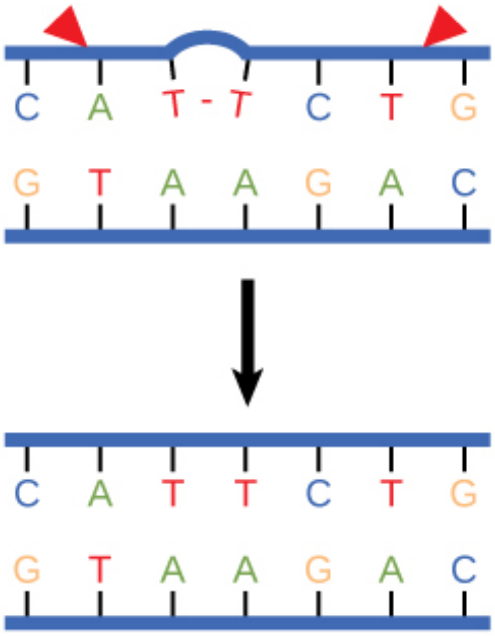


Figure 3. *Nucleotide excision repairs thymine dimers. When exposed to UV, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced.*

Consequences of errors in replication, transcription, and translation

Something key to think about:

Cells have evolved a variety of ways to make sure DNA errors are both detected and corrected. We have already discussed several of them. But why did so many mechanisms evolve? From proofreading by the various DNA-dependent DNA polymerases, to the complex repair systems. Such mechanisms did not evolve for errors in transcription or translation. If you are familiar with the processes of transcription and/or translation, think about what the consequences would be of an error in transcription. Would such an error affect the offspring? Would it be lethal to the cell? What about errors in translation? Ask the same questions about the process of translation. What would happen if the wrong amino acid is accidentally put into the growing polypeptide during translation? How do these contrast with DNA replication? If you are not familiar with transcription or translation, don't fret. We'll learn those soon and return to this question again.

Reading_18 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_19

Learning Objectives Associated with SS1_2025_Bis2a_Facciotti_Reading_19

GI.40 Describe the key components of the central dogma and the relationships between them, including the processes of replication, transcription, translation, reverse transcription and the possible fates of the various possible products (i.e. mRNA, tRNA, rRNA, enzymes, structural proteins, replicated chromosomes etc.).
GI.22 Create illustrations that model the structures of RNA. These should span several levels of detail and abstraction that include cartoon models of:
(a) three nucleotides and their phosphodiester bonds;
(b) a zoomed-out representation of an mRNA and a tRNA.
GI.3 Given an abstracted model of a genomic region, correctly interpret representations of genes, promoters, and other functional elements that are represented as geometric objects, abstract representations of DNA sequences, on the linear genome coordinate.
GI.17 Create a model for a basic transcriptional unit that includes promoters, transcriptional regulatory sites for transcription factor binding, ribosome binding site, coding region, stop codon and transcriptional terminator.
GI.18 Use the model of a transcriptional unit to discuss the roles of each of the structural elements of a transcriptional unit. Identify those that are transcribed, those that may be translated and those that serve other roles.
GI.21 Create a dynamic model (verbal, drawing, etc.) of the process of transcription that includes the reactants, products, enzymes, the sites on the DNA required for transcription to take place, and how these components interact with the DNA template at different times during the process. Put GI.17 and GI.18 into action!

The flow of genetic information



In bacteria, archaea, and eukaryotes, the primary role of DNA is to store heritable information that encodes the instruction set required for creating the organism in question. While we have gotten much better at quickly reading the chemical composition (the sequence of nucleotides in a genome and some chemical modifications that are made to it), we still don't know how to decode reliably all the information within and all the mechanisms by which it is read and ultimately expressed.

There are, however, some core principles and mechanisms associated with the reading and expression of the genetic code whose basic steps are understood and that need to be part of the conceptual toolkit for all biologists. Two of these processes are transcription and translation, which are the coping of parts of the genetic code written in DNA into molecules of the related polymer RNA and the reading and encoding of the RNA code into proteins, respectively.

In BIS2A, we focus on developing an understanding of the process of transcription (recall that an Energy Story is a rubric for describing a process) and its role in the expression of genetic information. We motivate our discussion of transcription by focusing on functional problems (bringing in parts of our problem solving/design challenge rubric) that must be solved the process to take place. We then describe how the process is used by Nature to create a variety of functional RNA molecules (that may have various structural, catalytic or regulatory roles) including so called messenger RNA (mRNA) molecules that carry the information required to synthesize proteins. Likewise, we focus on challenges and questions associated with the process of translation, the process by which the ribosomes synthesize proteins.

We often depict the basic flow of genetic information in biological systems in a scheme known as "the central dogma" (see figure below). This scheme states that information encoded in DNA flows into RNA via transcription and ultimately translated to proteins. Processes like reverse transcription (the creation of DNA from and RNA template) and replication also represent mechanisms for propagating information in different forms. This scheme, however, says nothing per se about how information is encoded or about the mechanisms by which regulatory signals move between the various layers of molecule types depicted in the model. Therefore, while the scheme below is a nearly required part of the lexicon of any biologist, perhaps left over from old tradition, students

should also know mechanisms of information flow are more complex (we'll learn about some as we go, and that "the central dogma" only represents some core pathways).

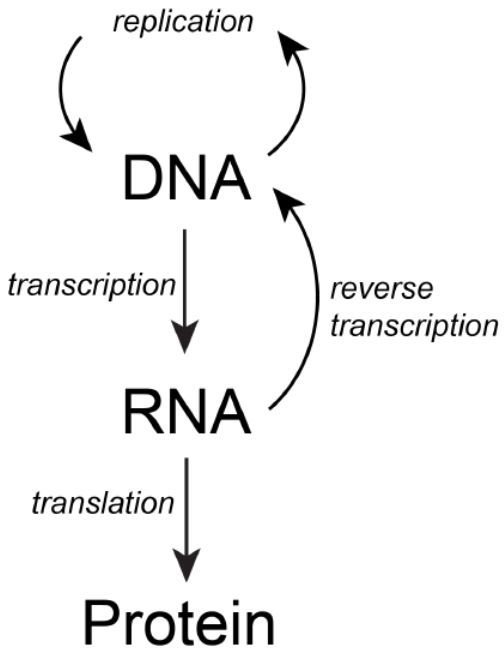


Figure 1. *The flow of genetic information. Attribution: Marc T. Facciotti (original work)*

Genotype to phenotype

An important concept in the following sections is the relationship between genetic information, the **genotype**, and the result of expressing it, the **phenotype**. These two terms and the mechanisms that link the two will be discussed repeatedly over the next few weeks—start becoming proficient with using this vocabulary.

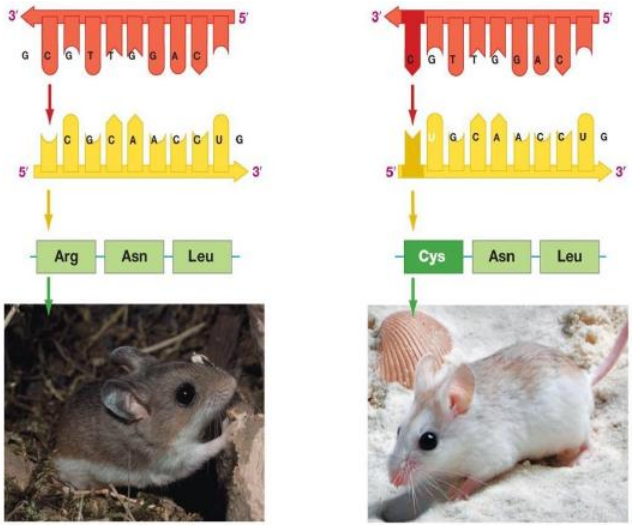


Figure 2. *Information in DNA that will be expressed by transcription is stored in the sequence of individual nucleotides read in the 5' to 3' direction. Conversion of the information from DNA into RNA (a process called transcription) expresses that information*

into a temporary copy that can be functional as is (e.g. tRNA, rRNA), or a message encoding the information required to build a protein (e.g. mRNA). Cells use the mRNA as the template for the creation proteins via translation. Here, we show two different DNA sequences. The differences in each DNA sequence result in the production of two different mRNAs, followed by the synthesis of two different proteins. Ultimately, these different proteins create two different coat colors in the mice.

Genotype refers to the information stored in the DNA of the organism, the sequence of the nucleotides, and the compilation of its genes. **Phenotype** refers to any physical characteristic that you can measure, such as height, weight, amount of ATP produced, ability to metabolize lactose, and response to environmental stimuli. Differences in genotype, even slight, can lead to different phenotypes subject to natural selection. The figure above depicts this idea. Also note that, while we talk classically about the genotype and phenotype relationship in the context of multicellular organisms, this nomenclature and the underlying concepts apply to all organisms, even single-celled organisms like bacteria and archaea.

Genes

What is a gene? A **gene** is a segment of DNA in an organism's genome that encodes a functional RNA (such as rRNA, tRNA, etc.) or protein product (enzymes, tubulin, etc.). A generic gene contains elements encoding regulatory regions and a region encoding a transcribed unit.

Genes can gain **mutations**—defined as changes in the composition and or sequence of the nucleotides—either in the coding or regulatory regions. These mutations can lead to several outcomes: (1) nothing measurable happens as a result; (2) the gene is no longer expressed; or (3) the expression or behavior of the gene product(s) are different. In a population of organisms sharing the same gene different variants of the gene are known as **alleles**. Different alleles can lead to differences in phenotypes of individuals and contribute to the diversity in biology under selective pressure.

Start learning these vocabulary terms and associated concepts. You will then be somewhat familiar with them when we dive into them in more detail over the next lectures.

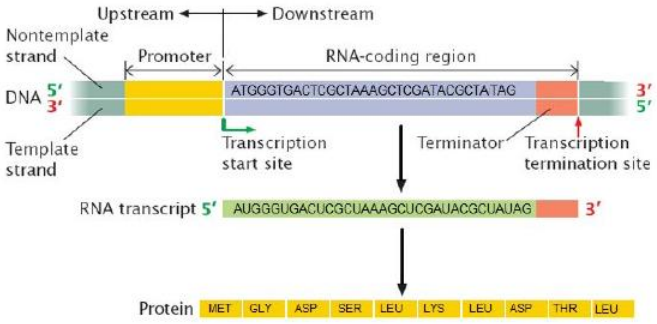


Figure 3. *A gene consists of a coding region for an RNA or protein product accompanied by its regulatory regions. The coding region is transcribed into RNA which is then translated into protein.*

Transcription from DNA to RNA

Section summary



Bacteria, archaea, and eukaryotes must all transcribe genes from their genomes. While the cellular location may be different (eukaryotes perform transcription in the nucleus; bacteria and archaea perform transcription in the cytoplasm), the mechanisms by which organisms from each of these clades carry out this process are fundamentally the same and can be characterized by three stages: initiation, elongation, and termination.

A short overview of transcription

Transcription is the process of creating an RNA copy of a segment of DNA. Since this is a process, we want to apply the Energy Story rubric to develop a functional understanding of transcription. What does the system of molecules look like before the start of the transcription? What does it look like at the end? What transformations of matter and transfers of energy happen during the transcription, and what catalyzes the process? We also want to think about the process from a Design Challenge standpoint. If the biological task is to create a copy of DNA in the chemical language of RNA, what challenges can we reasonably hypothesize or expect, given our knowledge about other nucleotide polymer processes, must be overcome? Is there evidence that Nature solved these problems in different ways? What seem to be the criteria for the success of transcription? You get the idea.

Listing some basic requirements for transcription

Let us first consider the tasks at hand by using some of our foundational knowledge and imagining what might need to happen during transcription if the goal is to make an RNA copy of a piece of one strand of a double-stranded DNA molecule. We'll see that using some basic logic allows us to infer many of the important questions and things that we need to know in order to describe the process properly.

Let us imagine that we want to design a nanomachine/nanobot that would conduct transcription. We can use some Design Challenge thinking to identify problems and subproblems that need to be solved by our little robot.

- Where should the machine start? Along the millions to billions of base pairs, where should the machine be directed?
- Where should the machine stop?
- If we have start and stop sites, we will need ways of encoding that information so that our machine(s) can read this information—how will that be accomplished?
- How many RNA copies of the DNA will we need to make?
- How fast do the RNA copies need to be made?
- How accurately do the copies need to be made?
- How much energy will the process take and where is the energy going to come from?

These are only some core questions. One can dig deeper if they wish. However, these are already good enough for us to get a good feel for this process. Notice, too, that many of these questions are remarkably similar to those we inferred might be necessary to understand DNA replication.

The building blocks of transcription

The building blocks of RNA

Recall from our discussion on the structure of nucleotides that the building blocks of RNA are very similar to those in DNA. In RNA, the building blocks comprise nucleotide triphosphates that are composed of a ribose sugar, a nitrogenous base, and three phosphate groups. The key differences between the building blocks of DNA and those of RNA are that RNA molecules are composed of nucleotides with ribose sugars (as opposed to deoxyribose sugars) and use uridine, a uracil containing nucleotide (as opposed to thymidine in DNA). Note below that uracil and thymine are structurally very similar—the uracil is just lacking a methyl (CH_3) functional group compared to thymine.

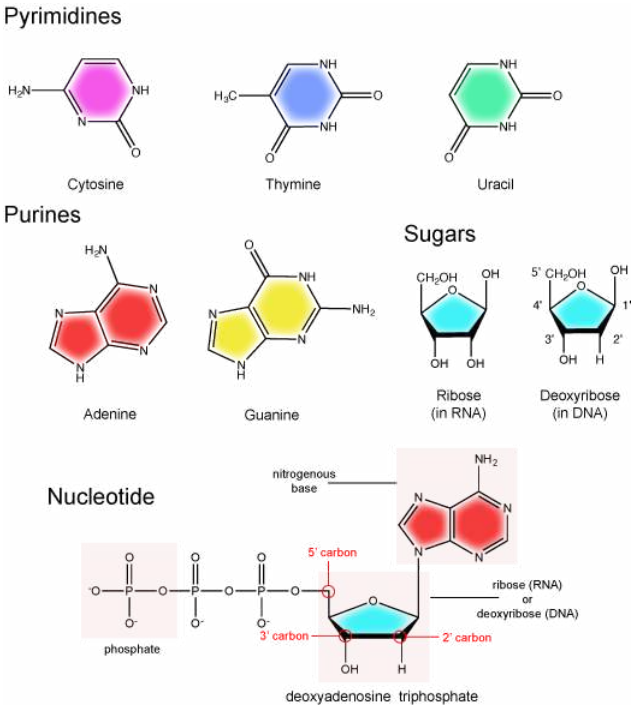


Figure 1. *The basic chemical components of nucleotides. Attribution: Marc T. Facciotti (original work)*

Transcription initiation

Promoters

Proteins responsible for creating an RNA copy of a specific piece of DNA (transcription) must first be able to recognize the beginning of the element to be copied. A **promoter** is a DNA sequence onto which various proteins, collectively known as the transcription machinery, bind and start transcription. In most cases, promoters exist upstream (5' to the coding region) of the genes they regulate. The specific sequence of a promoter is very important because it determines whether the cell transcribes the corresponding coding portion of the gene all the time, sometimes, or infrequently. Although promoters vary among species, a few elements of similar sequences are sometimes conserved. At the -10 and -35 regions upstream of the initiation site, there are two promoter **consensus** sequences, or regions similar across many promoters and across various species. Some promoters will have a sequence very similar to the consensus sequence (the sequence containing the most common sequence elements), and others will look very different. These sequence variations affect the strength to which the transcriptional machinery can bind to the promoter to start transcription. This helps to control the number of transcripts that are made and how often they get made.

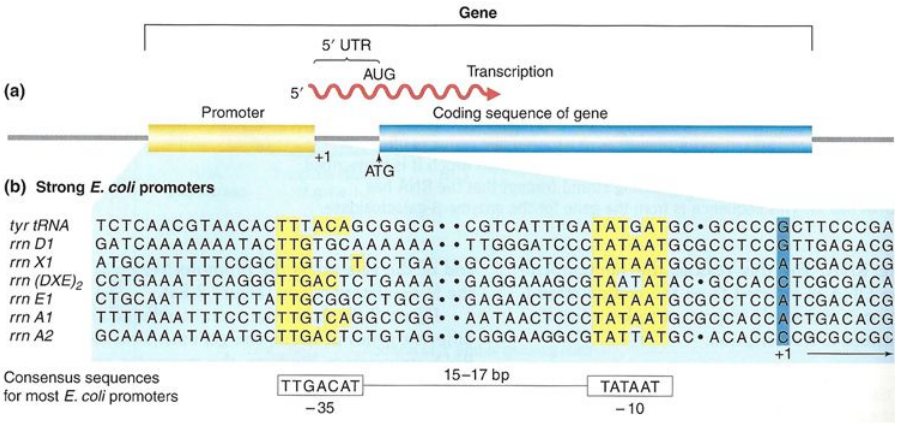


Figure 2. *(a) A general diagram of a gene. The gene includes the promoter sequence, an untranslated region (UTR), and the coding sequence. (b) A list of several strong E. coli promoter sequences. The -35 box and -10 box are highly conserved sequences throughout the strong promoter list. Weaker promoters will have more base pair differences when compared to these sequences. Source: http://www.discoveryandinnovation.co...lecture12.html*

Bacterial vs. eukaryotic promoters

In bacterial cells, the -10 consensus sequence, called the -10 region, is AT rich, often TATAAT. The -35 sequence, TTGACA, is recognized and bound by the protein *σ*. Once this protein-DNA interaction is made, the subunits of the core RNA polymerase bind to the site. Because of the relatively lower stability of AT associations, the AT-rich -10 region facilitates unwinding of the DNA template, and several phosphodiester bonds are made.

Eukaryotic promoters are much larger and more complex than prokaryotic promoters, but both have an AT-rich region—in eukaryotes, it is typically called a TATA box. For example, in the mouse thymidine kinase gene, the TATA box is located at approximately -30. For this gene, the exact TATA box sequence is TATAAAA, as read in the 5' to 3' direction on the nontemplate strand. This sequence is not identical to the *E. coli* -10 region, but both share the quality of being AT-rich element.

Instead of a single bacterial polymerase, the genomes of most eukaryotes encode three different RNA polymerases, each made up of ten protein subunits or more. Each eukaryotic polymerase also requires a distinct set of proteins known as **transcription factors** to recruit it to a promoter. In addition, an army of other transcription factors, proteins known as enhancers, and silencers help to regulate the synthesis of RNA from each promoter. Enhancers and silencers affect the efficiency of transcription but are not necessary for the initiation of transcription or its procession. Basal transcription factors are crucial in the formation of a **preinitiation complex** on the DNA template that subsequently recruits RNA polymerase for transcription initiation.

The initiation of transcription begins with the binding of RNA polymerase to the **promoter**. Transcription requires the DNA double helix to partially unwind such that one strand can be used as the template for RNA synthesis. The region of unwinding is called a **transcription bubble**.

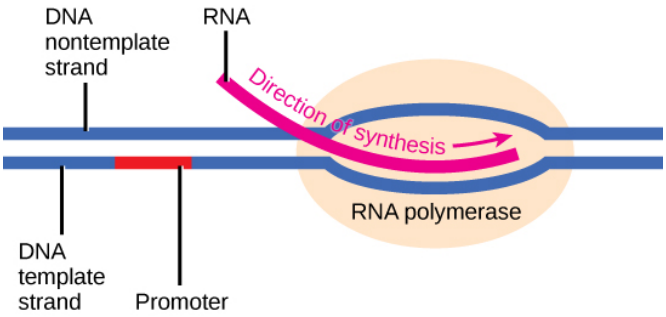


Figure 3. *During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds then rewinds the DNA as it is read.*

Elongation

Transcription always proceeds from the **template strand**, one of the two strands of the double-stranded DNA. The RNA product is complementary to the template strand and is almost identical to the nontemplate strand, called the **coding strand**, with the exception that RNA contains a uracil (U) in place of the thymine (T) found in DNA. During elongation, an enzyme called **RNA polymerase** proceeds along the DNA template, adding nucleotides by base pairing with the DNA template in a manner similar to DNA replication, with the difference being an RNA strand that is synthesized does not remain bound to the DNA template. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it. Note that the direction of synthesis is identical to that of synthesis in DNA—5' to 3'.

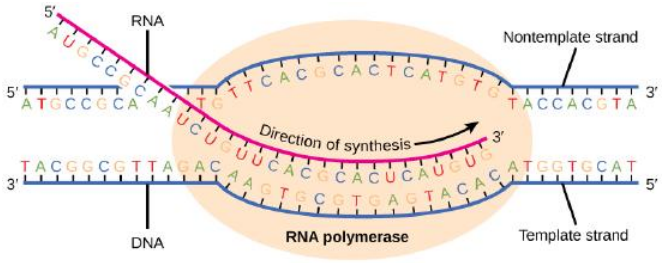


Figure 4. *During elongation, RNA polymerase tracks along the DNA template, synthesizing mRNA in the 5' to 3' direction, unwinding and then rewinding the DNA as it is read.*

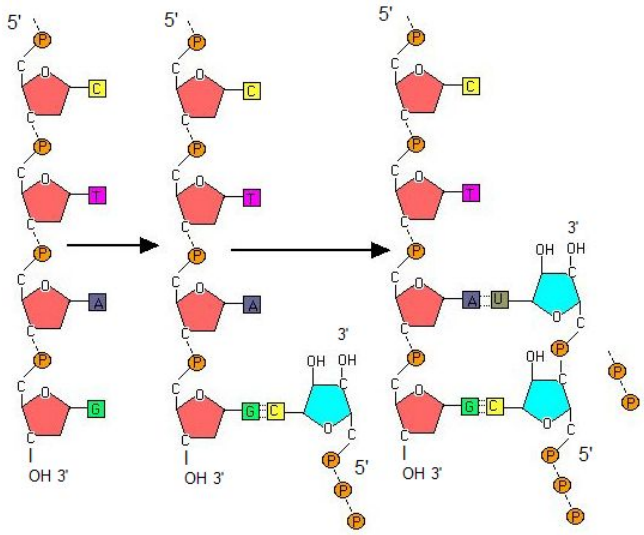


Figure 5. *The addition of nucleotides during the process of transcription is very similar to nucleotide addition in DNA replication. The RNA is polymerized from 5' to 3', and with each addition of a nucleotide, a phosphoanhidride bond is hydrolized by the enzyme, resulting in a longer polymer and the release of two inorganic phosphates. Source: http://utminers.utep.edu/rwebb/html/...longation.html*

Bacterial vs. eukaryotic elongation

In bacteria, elongation begins with the release of the σ subunit from the polymerase. The dissociation of σ allows the core enzyme to proceed along the DNA template, synthesizing mRNA in the 5' to 3' direction at a rate of approximately 40 nucleotides per second. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it. The base pairing between DNA and RNA is not stable enough to maintain the stability of the mRNA synthesis components. Instead, the RNA polymerase acts as a stable linker between the DNA template and the nascent RNA strands to ensure that elongation is not interrupted prematurely.

In eukaryotes, following the formation of the preinitiation complex, the polymerase is released from the other transcription factors, and elongation is allowed to proceed as it does in prokaryotes with the polymerase synthesizing pre-mRNA in the 5' to 3' direction. As discussed previously, RNA polymerase II transcribes the major share of eukaryotic genes, so this section will focus on how this polymerase accomplishes elongation and termination.

Possible NB Discussion Point

Compare and contrast the energy story for DNA replication initiation + elongation to the energy story for transcription initiation + elongation.

Termination

In bacteria

Once a gene is transcribed, the bacterial polymerase needs to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals. One is a protein-based, and the other is RNA-based. **Rho-dependent termination** is controlled by the rho protein, which tracks along the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

Rho-independent termination is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in CG nucleotides. The mRNA folds back on itself, and the complementary CG nucleotides bind together. The result is a stable **hairpin** that causes the polymerase to stall as soon as it begins to transcribe a region rich in AT nucleotides. The complementary UA region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

In eukaryotes

The termination of transcription is different for the different polymerases. Unlike in prokaryotes, elongation by RNA polymerase II in eukaryotes takes place 1,000–2,000 nucleotides beyond the end of the gene being transcribed. This pre-mRNA tail is subsequently removed by cleavage during mRNA processing. On the other hand, RNA polymerases I and III require termination signals. Genes transcribed by RNA polymerase I contain a specific 18-nucleotide sequence that is recognized by a termination protein. The process of termination in RNA polymerase III involves a mRNA hairpin similar to rho-independent termination of transcription in prokaryotes.

In archaea

Termination of transcription in the archaea is far less studied than in the other two domains of life and is still not well understood. While the functional details are likely to resemble mechanisms that have been seen in the other domains of life, the details are beyond the scope of this course.

Cellular location

In bacteria and archaea

In bacteria and archaea, transcription occurs in the cytoplasm, where the DNA is located. Because the location of the DNA, and thus the process of transcription, are not physically segregated from the rest of the cell, translation often starts before transcription has finished. This means that mRNA in bacteria and archaea is used as the template for a protein before it produces the entire mRNA. The lack of spatial segregation also means that there is very little temporal segregation for these processes. Figure 6 shows the processes of transcription and translation occurring simultaneously.

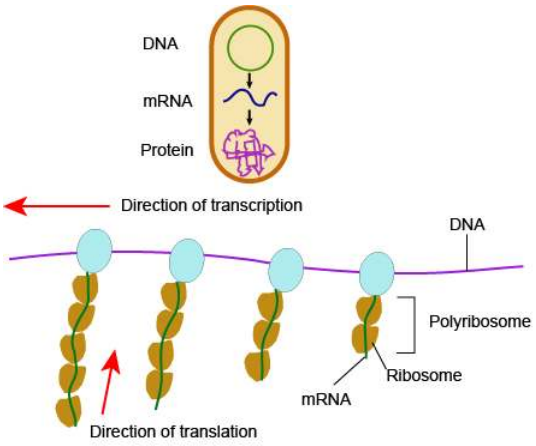


Figure 6. *The addition of nucleotides during the process of transcription is very similar to nucleotide addition in DNA replication. Source: Marc T. Facciotti (own work)*

In eukaryotes....

In eukaryotes, the process of transcription is physically segregated from the rest of the cell, sequestered inside of the nucleus. This results in two things: the mRNA is completed before translation can start, and there is time to "adjust" or "edit" the mRNA before translation starts. The physical separation of these processes gives eukaryotes a chance to alter the mRNA in such a way as to extend the lifespan of the mRNA or even alter the protein product that will be produced from the mRNA.

mRNA processing

5' G-cap and 3' poly-A tail

When an eukaryotic gene is transcribed, the cell processes the primary transcript in the nucleus in several ways. Eukaryotic cells modify mRNAs at the 3' end by the addition of a poly-A tail. This run of A residue is added by an enzyme that does not use genomic DNA as a template. The mRNAs have a chemical modification of the 5' end, called a 5'-cap. Data suggests that these modifications both help to increase the lifespan of the mRNA (prevent its premature degradation in the cytoplasm) and to help the mRNA start translation.

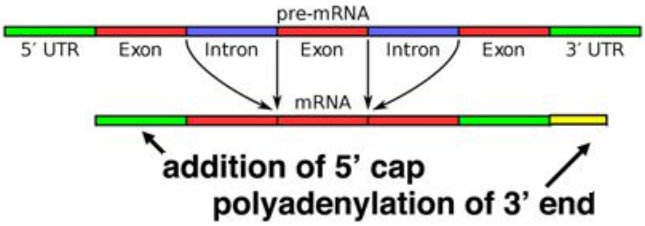


Figure 7. *pre-mRNAs are processed in a series of steps. Introns are removed, a 5' cap and poly-A tail are added. Source: http://www.discoveryandinnovation.co...lecture12.html*

Possible NB Discussion Point

Transcriptomics is a branch of “-omics” that involves studying an organism or population’s transcriptome or, the complete set of all RNA molecules. What kind of information can you obtain from studying the transcriptome(s)? Can you think of any cool scientific questions that a transcriptomic analysis might help resolve? What are some constraints to transcriptomic approaches one might keep in mind when conducting analyses?

Alternative splicing

Splicing occurs on most eukaryotic mRNAs in which introns are removed from the mRNA sequence and exons are ligated together. This can create a much shorter mRNA than initially transcribed. Splicing allows cells to mix and match which exons are incorporated into the final mRNA product. As shown in the figure below, this can lead to multiple proteins being coded for by a single gene.

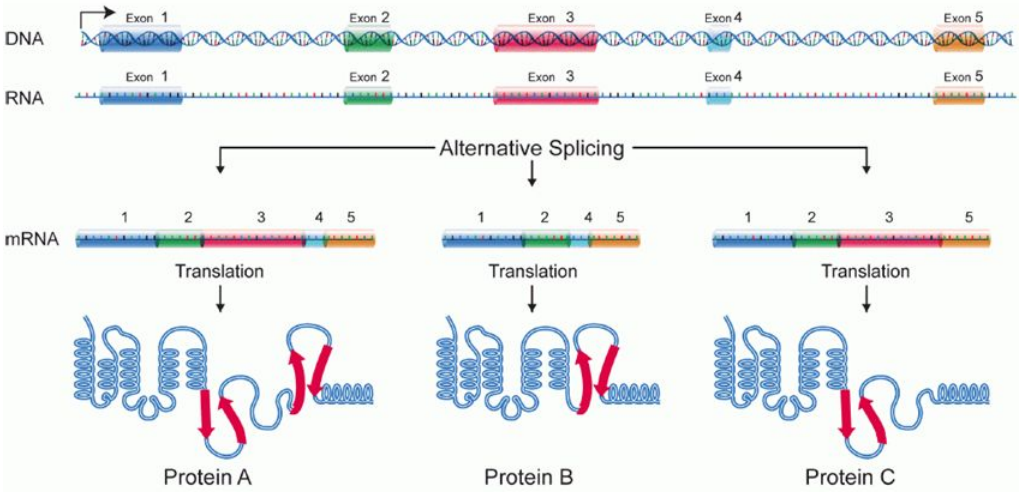


Figure 8. *The information stored in the DNA is finite. In some cases, organisms can mix and match this information to create different end products. In eukaryotes, alternative splicing allows for the creation of different mRNA products, which in turn are used in translation to create different protein sequences. This ultimately leads to the production of different protein shapes, and thus different protein functions.*
Source: http://www.discoveryandinnovation.co...lecture12.html

Reading_19 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_20

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_20

GI.48 Compare the similarities and differences between prokaryotic and eukaryotic transcriptional regulatory components (the parts on the DNA and the proteins involved) and the mechanisms that are used for regulating transcription in each case.
GI.39 Explain the key differences between bacterial and eukaryotic gene expression. Be sure to include in your explanation the role compartmentalization plays in these differences.
GI.23 Describe how eukaryotic gene transcripts are processed before translation and the significance of this for generating functional diversity.
GI.19 Differentiate and convert between coding/noncoding, template/non-template strands.
GI.20 Define, explain and correctly use the function and structure of an open reading frame (ORF).
GI.24 Use a codon table to "translate" an RNA sequence and possible variants of the sequence into a protein sequence.
GI.25 Explain what is meant by the genetic code being degenerate.
GI.26 Diagram the process of translation. The diagram should include reactants (including the mRNA template and the tRNAs), the products, enzymes, and the sites on the mRNA template required for translation to take place.
GI.30 Create a dynamic model of the chemical reaction responsible for protein synthesis and tell its energy story. Then compare this story to the one you prepared for the chemical reactions of DNA synthesis and RNA synthesis.
GI.27 Draw a rough sketch of an aminoacyl tRNA synthetase, including its active site and other sites in the enzyme that bind to the reactants.
GI.28 Describe the steps in the chemical reaction responsible for tRNA charging and lay out its energy story.

Protein Synthesis



Introduction

The process of **translation** in biology is the decoding an mRNA message into a polypeptide product. Put another way, a message written in the chemical language of nucleotides is "translated" into the chemical language of amino acids. Amino acids are linearly strung together via covalent bonds (called peptide bonds) between amino and carboxyl termini of adjacent amino acids. The decoding and "linking" process is catalyzed by a ribonucleoprotein complex called the **ribosomes** and can result in chains of amino acids of lengths ranging from tens to over 1,000.

The resulting proteins are so important to the cell that their synthesis consumes more of a cell's energy than any other metabolic process. Like DNA replication and transcription, translation is a complex molecular process that we can approach using both the Energy Story and Design Challenge rubrics. Describing the overall process, or steps, requires the accounting of the matter and energy before the process and after the process and a description of how that matter is transformed and energy transferred during the process. From a Design Challenge standpoint, we can - even before digging any further into what is or is not understood about translation - try to infer some basic questions that we will need to answer regarding this process.

Let us start by considering the basic problem. We have a strand of RNA (called mRNA) and a bunch of amino acids and we need to design a machine that will somehow:

(a) decode the chemical language of nucleotides into the language of amino acids,
(b) join amino acids in a very specific manner,
(c) complete this process with reasonable accuracy, and
(d) do this at a reasonable speed. Reasonable is defined by natural selection.

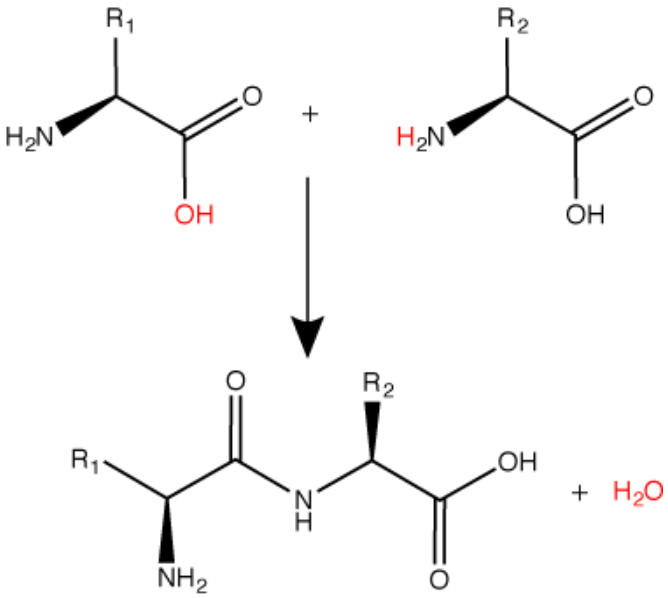
As before, we can identify subproblems.

(a) How does our molecular machine determine where and when to work?

(b) How does the molecular machine coordinate decoding and bond formations?
(c) Where does the energy for this process come from and how much?
(d) How does the machine know where to stop?

Other questions and functional problems/challenges will arise as we dig deeper.

The point, as always, is that even knowing no specifics about translation we can use our imaginations, curiosity and common sense to imagine some requirements for the process that we will need to learn more about. Understanding these questions as the context for what follows is key.



A peptide bond links the carboxyl end of one amino acid with the amino end of another, expelling one water molecule. The R_1 and R_2 designation refer to the side chain of the two amino acids. Attribution: Marc T. Facciotti (original work).

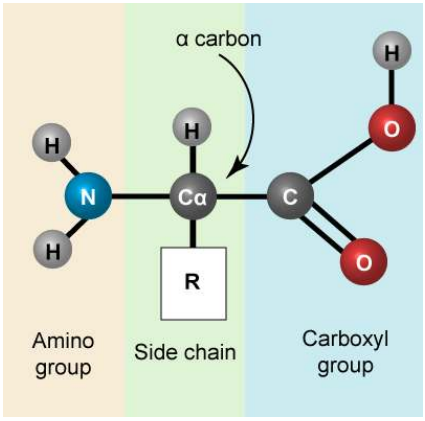
Protein Synthesis Machinery

The components that go into the process

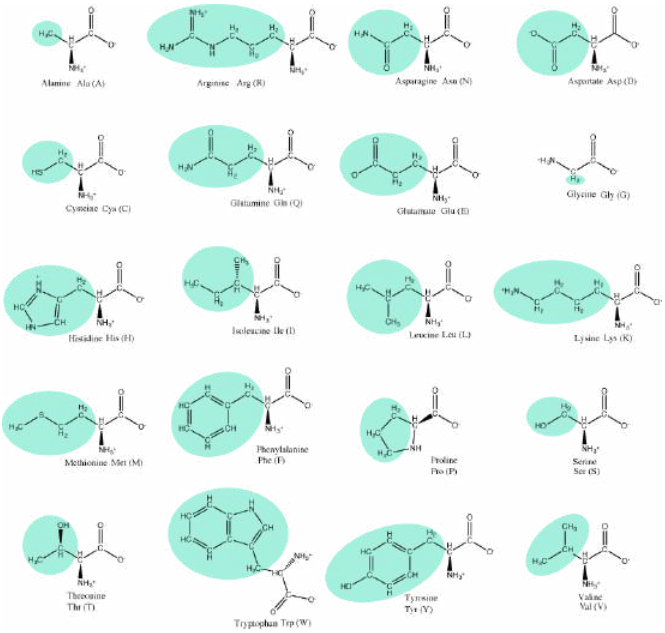
Many molecules and macromolecules contribute to the process of translation. While the exact composition of "the players" in the process may vary from species to species - for instance, ribosomes may comprise different numbers of *rRNAs* (ribosomal RNAs) and polypeptides depending on the organism - the general functions of the protein synthesis machinery are comparable from bacteria to human cells. We focus on these similarities. At a minimum, translation requires an **mRNA template**, **amino acids**, **ribosomes**, **tRNAs**, an energy source, and various additional accessory enzymes and small molecules.

Reminder: Amino acids

Let us recall that the basic structure of amino acids consists of a backbone composed of an amino group, a central carbon (called the α-carbon), and a carboxyl group. Attached to the α-carbon is a variable group that helps determine some chemical properties and reactivity of the amino acid.



A generic amino acid. *Attribution: Marc T. Facciotti (own work)*



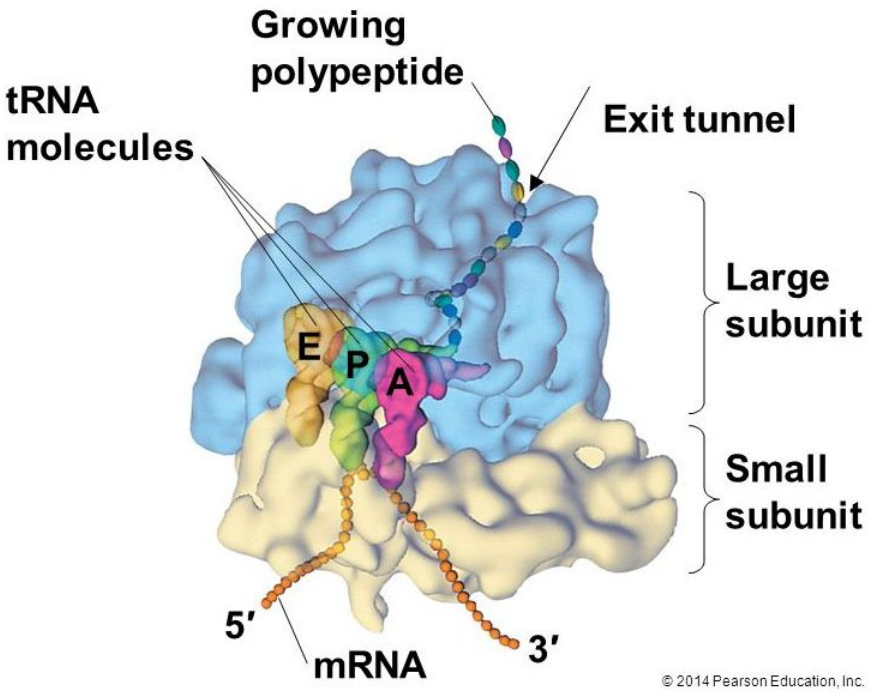
The 20 common amino acids. *Attribution: Marc T. Facciotti (own work)*

Ribosomes

A **ribosome** is a complex macromolecule composed of structural and catalytic rRNAs, and many distinct polypeptides. As we try thinking about energy accounting in the cell, we note that the ribosomes themselves do not come for "free". Even before an mRNA is translated, a cell must invest energy to build each of its ribosomes. In *E. coli*, there are between 10,000 and 70,000 ribosomes present in each cell.

Ribosomes exist in the cytoplasm in bacteria and archaea and in the cytoplasm and on the rough endoplasmic reticulum in eukaryotes. Mitochondria and chloroplasts also have their own ribosomes in the matrix and stroma, which look more similar to bacterial ribosomes (and have similar drug sensitivities), than the ribosomes just outside their outer membranes in the cytoplasm. Ribosomes dissociate into large and small subunits when they are not synthesizing proteins and reassociate during the initiation of translation. In *E. coli*, we describe the small subunit as 30S, and the large subunit as 50S. Mammalian ribosomes have a small 40S subunit and a large 60S subunit. The small subunit binds the mRNA template, whereas the large subunit sequentially binds tRNAs. Many ribosomes can simultaneously translate an individual mRNA molecule, each ribosome synthesizing protein in the same

direction: reading the mRNA from 5' to 3' and synthesizing the polypeptide from the N terminus to the C terminus. The complete mRNA/poly-ribosome structure is called a **polysome**.

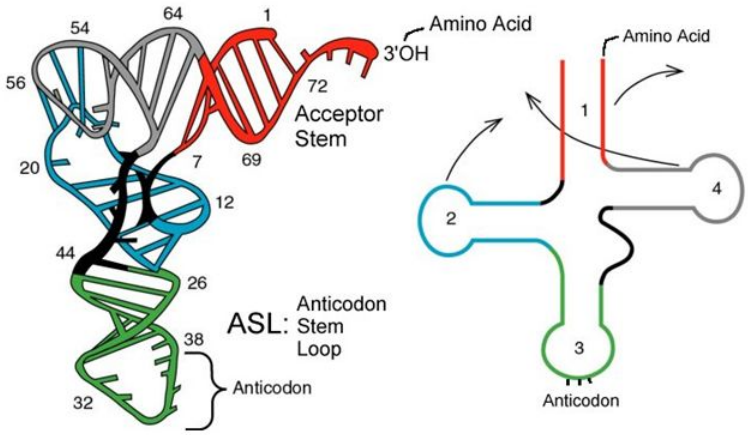


The protein synthesis machinery includes the large and small subunits of the ribosome, mRNA, and tRNA.
Source: http://bio1151.nicerweb.com/Locked/m.../ribosome.html

tRNAs

tRNAs are structural RNA molecules that were transcribed from genes. Depending on the species, 40 to 60 types of tRNAs exist in the cytoplasm. Serving as adaptors, specific tRNAs bind to sequences on the mRNA template and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually “translate” the language of RNA into the language of proteins.

Of the 64 possible mRNA **codons**—or triplet combinations of A, U, G, and C, three specify the termination of protein synthesis and 61 specify the addition of amino acids to the polypeptide chain. Of these 61, one codon (AUG) also encodes the initiation of translation. Each tRNA **anticodon** can base pair with one of the mRNA codons and add an amino acid or terminate translation, according to the genetic code. For instance, if the sequence CUA occurred on an mRNA template in the proper reading frame, it would bind a tRNA expressing the complementary sequence, GAU, which would be linked to the amino acid leucine.



The folded secondary structure of a tRNA. The anticodon loop and amino acid acceptor stem are indicated.
Source: http://mol-biol4masters.masters.grkr...ansfer_RNA.htm

---

Aminoacyl tRNA Synthetases

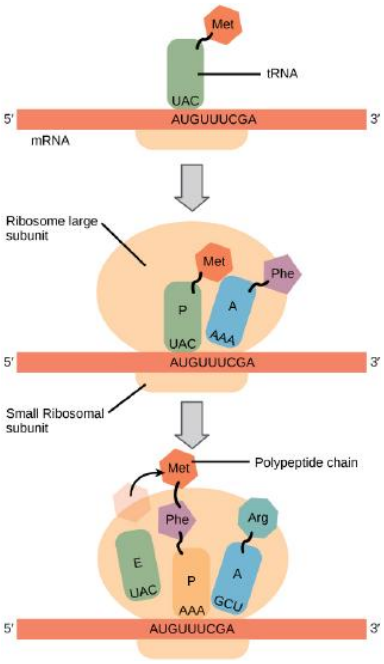
The process of pre-tRNA synthesis by RNA polymerase III only creates the RNA portion of the adaptor molecule. The corresponding amino acid must be added later, once the tRNA is processed and exported to the cytoplasm. Through the process of tRNA "charging," each tRNA molecule is linked to its correct amino acid by a group of enzymes called **aminoacyl tRNA synthetases**. At least one type of aminoacyl tRNA synthetase exists for each of the 20 amino acids; the exact number of aminoacyl tRNA synthetases varies by species. These enzymes first bind and hydrolyze ATP to catalyze a high-energy bond between an amino acid and adenosine monophosphate (AMP); they expel a pyrophosphate molecule in this reaction. The activated amino acid is then transferred to the tRNA, and AMP is released.

The Mechanism of Protein Synthesis

Like in transcription, we can divide protein synthesis into three phases: initiation, elongation, and termination. The process of translation is similar in bacteria, archaea and eukaryotes.

Translation Initiation

Generally, protein synthesis begins with the formation of an initiation complex. The small ribosomal subunit will bind to the mRNA at the **ribosomal binding site**. Soon after, the methionine-tRNA will bind to the AUG start codon (through complementary binding with its anticodon). This complex is then joined by large ribosomal subunit. This initiation complex then recruits the second tRNA and thus translation begins.



Translation begins when a tRNA anticodon recognizes a codon on the mRNA. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Bacterial vs Eukaryotic initiation

In *E. coli* mRNA, a sequence upstream of the first AUG codon, called the **Shine-Dalgarno sequence** (AGGAGG), interacts with a rRNA molecule. This interaction anchors the 30S ribosomal subunit at the correct location on the mRNA template. Stop for a moment to appreciate the repetition of a mechanism you've encountered before. Here, getting a protein complex to associate - in proper register - with a nucleic acid polymer is accomplished by aligning two antiparallel strands of complementary nucleotides with one another. We also saw this in the function of telomerase.

Instead of binding at the Shine-Dalgarno sequence, the eukaryotic initiation complex recognizes the 7-methylguanosine cap at the 5' end of the mRNA. A cap-binding protein (CBP) assists the movement of the ribosome to the 5' cap. Once at the cap, the initiation complex tracks along the mRNA in the 5' to 3' direction, searching for the AUG start codon. Many eukaryotic mRNAs are translated from the first AUG, but this is not always the case. According to **Kozak's rules**, the nucleotides around the AUG indicate whether it is the correct start codon. Kozak's rules state that the following consensus sequence must appear around the AUG of vertebrate genes: 5'-gccRccAUGG-3'. The R (for purine) indicates a site that can be A or G, but cannot be C or U. Essentially, the closer the sequence is to this consensus, the higher the efficiency of translation.

Possible NB Discussion Point

Compare and contrast the initiation of translation with that of transcription — in what ways are these processes similar and in what ways do they differ?

Translation Elongation

During translation elongation, the mRNA template provides specificity. As the ribosome moves along the mRNA, each mRNA codon comes into 'view', and specific binding with the corresponding charged tRNA anticodon is ensured. If mRNA were not present in the elongation complex, the ribosome would bind tRNAs nonspecifically. Note again the use of base pairing between two antiparallel strands of complementary nucleotides to bring and keep our molecular machine in register and in this case also to accomplish the job of "translating" between the language of nucleotides and amino acids.

The large ribosomal subunit comprises three compartments: the A site binds incoming charged tRNAs (tRNAs with their attached specific amino acids), the P site binds charged tRNAs carrying amino acids that have formed bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA, and the E site which releases dissociated tRNAs so they can be recharged with another free amino acid.

Elongation proceeds with charged tRNAs entering the A site and then shifting to the P site followed by the E site with each single-codon “step” of the ribosome. Ribosomal steps are induced by conformational changes that advance the ribosome by three bases in the 3' direction. The energy for each step of the ribosome is donated by an elongation factor that hydrolyzes GTP. Peptide bonds form between the amino group of the amino acid attached to the A-site tRNA and the carboxyl group of the amino acid attached to the P-site tRNA. The formation of each peptide bond is catalyzed by **peptidyl transferase**, an RNA-based enzyme that is integrated into the 50S ribosomal subunit. The energy for each peptide bond formation is derived from GTP hydrolysis, which is catalyzed by a separate elongation factor. The amino acid bound to the P-site tRNA is also linked to the growing polypeptide chain. As the ribosome steps across the mRNA, the former P-site tRNA enters the E site, detaches from the amino acid, and is expelled. The ribosome moves along the mRNA, one codon at a time, catalyzing each process that occurs in the three sites. With each step, a charged tRNA enters the complex, the polypeptide becomes one amino acid longer, and an uncharged tRNA departs. Amazingly, this process occurs rapidly in the cell, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino acid polypeptide could be translated in just 10 seconds.

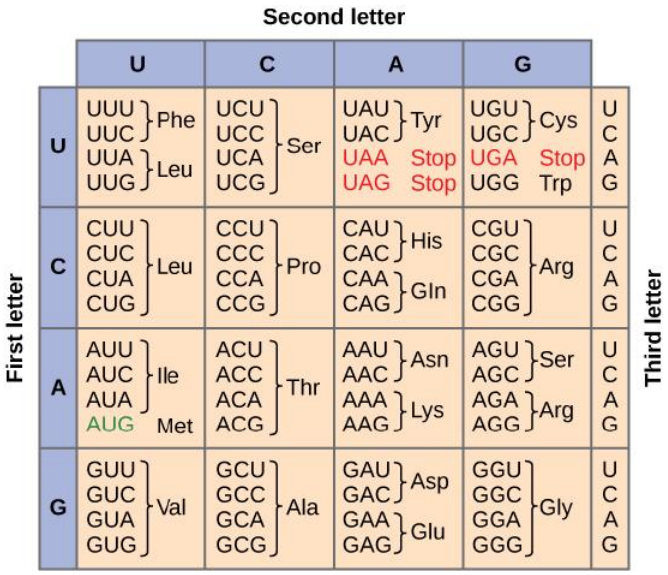
Possible NB Discussion Point

Tetracycline is an antibiotic on the World Health Organization’s List of Essential Medicines. It mitigates infections by blocking the A site on the bacterial ribosome. Another antibiotic, chloramphenicol, blocks peptidyl transfer. Describe the immediate and long-term effects of these two antibiotics. What other strategies can you think of to battle infection specifically at the level of translation?

The Genetic Code

To summarize what we know to this point, the cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template converts nucleotide-based genetic information into a protein product. Protein sequences consist of 20 commonly occurring amino acids; therefore, we can say that the protein alphabet consists of 20 letters. We define each amino acid by a three-nucleotide sequence called the triplet **codon**. The relationship between a nucleotide codon and its corresponding amino acid is called the **genetic code**. Given the different numbers of “letters” in the mRNA and protein “alphabets,” means that there are a total of 64 (4 × 4 × 4) possible codons; therefore, an amino acid (20 total) must be encoded for by more than one codon.

Three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **stop codons**. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the **start codon** to initiate translation. The reading frame for translation is set by the AUG start codon near the 5' end of the mRNA. The genetic code is universal. With a few exceptions, virtually all species use the same genetic code for protein synthesis, which is powerful evidence that all life on Earth shares a common origin.



This figure shows the genetic code for translating each nucleotide triplet, or codon, in mRNA into an amino acid or a termination signal in a nascent protein. (credit: modification of work by NIH)

Redundant (aka degenerate), not Ambiguous

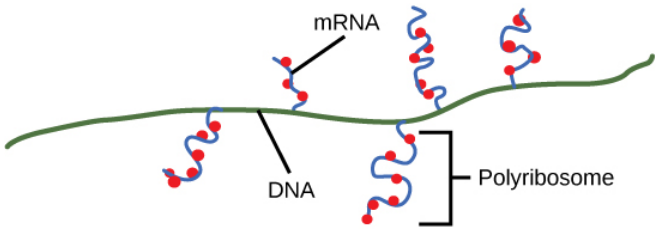
The information in the genetic code is redundant, also known as the degeneracy of the genetic code. Multiple codons code for the same amino acid. For example, using the chart above, you can find 4 different codons that code for Valine, likewise, there are two codons that code for Leucine, etc. But the code is not ambiguous, meaning, that if you were given a codon you would know definitively which amino acid it is coding for, a codon will only code for a specific amino acid. For example, GUU will always code for Valine, and AUG will always code for Methionine. This is important, you will be asked to translate an mRNA into a protein using a codon chart like the one shown above.

Translation Termination

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is encountered. When the ribosome encounters the stop codon no tRNA enters the A site. Instead, a protein known as a **release factor** binds to the complex. This interaction destabilizes the translation machinery, causing the release of the polypeptide and the dissociation of the ribosome subunits from the mRNA. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

Coupling between Transcription and Translation

As discussed previously, bacteria and archaea need not transport their RNA transcripts between a membrane bound nucleous and the cytoplasm. The RNA polymerase is therefore transcribing RNA directly into the cytoplasm. Here ribosomes can bind to the RNA and begin the process of translation, sometimes while transcription is still occurring. The coupling of these two processes, and even mRNA degradation, is facilitated not only because transcription and translation happen in the same compartment but also because both of the processes happen in the same direction - synthesis of the RNA transcript happens in the 5' to 3' direction and translation reads the transcript in the 5' to 3' direction. This "coupling" of transcription with translation occurs in both bacteria and archaea and is, in fact, essential for proper gene expression sometimes.



Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

Section Summary

mRNA is used to synthesize proteins by the process of translation. The genetic code is the correspondence between the three-nucleotide mRNA codon and an amino acid. The genetic code is "translated" by the tRNA molecules, which associate a specific codon with a specific amino acid. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three stop codons. This means that more than one codon corresponds to an amino acid. Almost every species on the planet uses the same genetic code.

The players in translation include the mRNA template, ribosomes, tRNAs, and various enzymatic factors. The small ribosomal subunit binds to the mRNA template. Translation begins at the starting AUG on the mRNA. The formation of bonds occurs between sequential amino acids specified by the mRNA template according to the genetic code. The ribosome accepts charged tRNAs, and as it steps along the mRNA, it catalyzes bonding between the new amino acid and the end of the growing polypeptide. The entire mRNA is translated in three-nucleotide "steps" of the ribosome. When a stop codon is encountered, a release factor binds and dissociates the components and frees the new protein.

Reading_20 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_21

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_21

GI.29 Discuss how the primary structure of a protein influences its target destination in the cell for both eukaryotic and bacterial organisms.

GI.XX Describe the steps a protein with a signal sequence follow to reach the part of the secretory pathway which is its final localization.

GI.XX.2 Distinguish between co-translational and post-translational translocation and classify subcellular compartments accordingly

Post-translational Protein Modification

After translation individual amino acids may be chemically modified. These modifications add chemical variation and new properties that are rooted in the chemistries of the functional groups that are being added. Common modifications include phosphate groups, methyl, acetate, and amide groups. Some proteins, typically targeted to membranes, will be lipidated - a lipid will be added. Other proteins will be glycosylated - a sugar will be added. Another common post-translational modification is cleavage or linking of parts of the protein itself. Signal-peptides may be cleaved, parts may be excised from the middle of the protein, or new covalent linkages may be made between cysteine or other amino acid side chains. Enzymes will catalyze nearly all modifications and all change modifications the functional behavior of the protein.

Post-Translational Modifications (PTMs) overview, Protein Activity, and Longevity

The addition or removal of these added chemical groups to/from proteins can regulate their activity, add new functions, or help regulate how long they exist in the cell. Sometimes these chemical modifications can also help to regulate where a protein is found in the cell—for example, in the nucleus, the cytoplasm, or attached to the plasma membrane.

Post-Translational modifications (PTMs) can be triggered in response to external stimuli such as stress, the lack of nutrients, heat, or ultraviolet light exposure. In addition to regulating the function of the proteins themselves, if these changes occur on specific proteins they can alter epigenetic accessibility (in the case of histone modification), transcription (transcription factors), mRNA stability (RNA binding proteins), or translation (eIF-2) thus feeding back and regulating various parts of the process of gene expression. In the case of modification to regulatory proteins, this can be an efficient way for the cell to rapidly change the levels of specific proteins in response to the environment by regulating various steps in the process.

The human genome contains approximately 20,000 to 25,000 genes. When analyzing the transcriptome, it becomes apparent that the genome becomes amplified by the wide array of splice variants that can occur during the processing of transcripts. It is estimated that the transcriptome contains roughly 100,000 transcripts. This is amplified again within the proteome that contains over 1,000,000 unique proteins. One of the main routes of further proteome expansion is through PTMs of proteins. PTMs are present in both eukaryotes and prokaryotes, but it is estimated that PTMs are more common in eukaryotic cells, in which about 5% of the genome is dedicated to enzymes that carry out posttranslational modifications of proteins.

Protein PTM results from the enzymatic or nonenzymatic attachment of specific chemical groups to amino acid side chains. Such modifications occur either following protein translation or concomitant with translation. PTM influences both protein structure and physiological and cellular functions. Examples of enzymatic PTMs include phosphorylation, glycosylation, acetylation, methylation, sumoylation, and ubiquitination. Note that many of these modifications are not made in isolation. It is common for proteins to have several different types of modifications and that these modifications can differ depending on the tissue type and environmental circumstances present.

PTMs can exert their effect through a many different structural changes, including opening and closing the active site, changing the conformation and electrostatic properties of binding sites, changing the flexibility of the chain including increasing or decreasing intrinsically disordered regions, altering protein:protein interactions, etc. These effecs can also arise from the binding of small molecules.

Protein Phosphorylation

One of the most common posttranslational modifications, protein phosphorylation, is the reversible addition of a phosphoryl group from adenosine triphosphate (ATP) to amino acid side chains such as serine, threonine, and tyrosine residues. This modification causes conformational changes that either (1) affect the catalytic activity to activate or inactivate the protein and/or cause the tendency of a protein to misfold and aggregate or (2) recruit other proteins to bind; both responses result in altered protein function and cell signaling. Phosphorylated proteins have critical and well-known functions in diverse cellular processes across eukaryotes, but phosphorylation also occurs in prokaryotic cells. In humans, about one-third of proteins are estimated to be substrates for phosphorylation. Indeed, phosphorylated proteins are now identified and characterized by high-throughput phosphoproteomics studies. The reversibility of protein phosphorylation is attributed to the actions of ***kinases*** and ***phosphatases***, which phosphorylate and dephosphorylate substrates, respectively. The temporal and spatial balance of kinase and phosphatase concentrations within a cell mediates the size of its phosphoproteome.

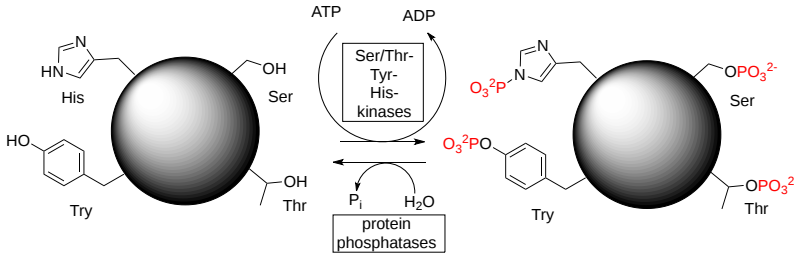


Figure 1: Amino acid phosphorylation in proteins 6.6.11

Protein Acetylation

The simplest form of acetylation is protein N-Acetylation. This occurs at the amino terminus amine and the ε-amino group of the lysine side chains through the action of acetylases. The acetylation of lysine side chains can be reversed through the actions of deacetylases (similar to the the combined actions of protein kinases and phosphatases). Interestingly, 80–90% of eukaryotic proteins are acetylated, yet the underlying biological significance remains unclear.

Although first described in histones, acetylation is also observed in cytoplasmic proteins. Acetylated proteins can also be modulated by the cross-talk with other posttranslational modifications, including phosphorylation, ubiquitination, and methylation. Therefore, acetylation may contribute to cell biology beyond transcriptional regulation.

Protein Glycosylation

Protein glycosylation involves the addition of a diverse set of sugar moieties to the protein core. Glycosylation has significant implications for protein folding, conformation, distribution, stability, and activity. Glycosylated proteins can have additions of simple monosaccharides. For example, many nuclear transcription factors are modified in this way. Alternatively, some proteins are modified with highly complex branched polysaccharides, such as those seen on cell surface protein receptors.

More than half of all mammalian proteins are believed to be glycosylated. However, glycoprotein functions, at both molecular and cellular levels, remain unclear. While proteins exhibit improved stability and trafficking after glycosylation *in vivo*, glycan structures can alter protein functions or activities. These structures often result from the activities of glycan-processing enzymes working within a cell at any given time. However, the structures are sometimes protein-specific, depending on protein trafficking properties and interactions with other cellular factors.

Protein Ubiquitination and Sumoylation

The addition of an ubiquitin group has another function - it marks that protein for degradation. Ubiquitin is a small molecule (an 8 kDa polypeptide) that acts like a flag indicating that the tagged proteins should be targeted to an organelle called the proteasome. This organelle is a large multi-protein complex that functions to cleave proteins into smaller pieces that can then be recycled. Ubiquitination (the addition of a ubiquitin tag), therefore helps to control gene expression by altering the functional lifetime of the protein product.

The addition of one ubiquitin can lead to the addition of more ubiquitins to form ubiquitin chains on the target protein. If the protein target is monoubiquinated, the activity of the target protein can be modified. For example monoubiquitination of histone proteins promotes the release the DNA enabling transcription. However, if multiple ubiquitins are added to the target protein forming a polyubiquitinated protein, the protein is tagged for degradation by the 26S proteasome.

Similarly, protein sumoylation is a reversible posttranslational modification whereby a **s**mall **u**biquitin-like **mo**difier (**SUMO**) protein is covalently attached to target. Like ubiquitin, SUMO is conjugated to the lysine side chains of target proteins, and it is removed by SUMO-specific isopeptidases. Sumoylation controls many aspects of nuclear function. However, recent research has started to unveil a determinant role of protein sumoylation in many extranuclear neuronal processes and potentially in a wide range of neuropathological conditions.

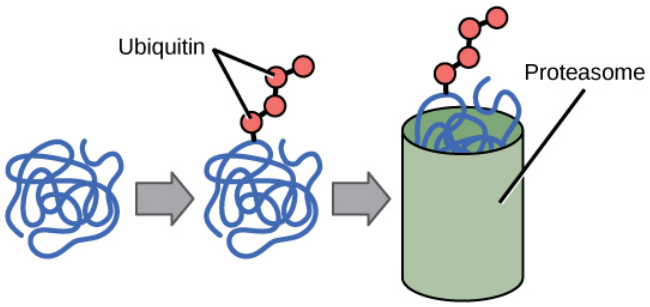


Figure 2: *Proteins with ubiquitin tags are marked for degradation within the proteasome.*

Protein Sorting

In the context of a protein synthesis Design Challenge we can also raise the question/problem of how proteins get to where they are supposed to go. We know that some proteins are destined for the plasma membrane, others in eukaryotic cells need to be directed to various organelles, some proteins, like hormones or nutrient scavenging proteins, are intended to be secreted by cells while others may need to be directed to parts of the cytosol to serve structural roles. How does this happen?

Since we have uncovered various mechanisms, the details of this process are not easily summarized in a brief paragraph or two. However, we can mention some key elements common to all mechanisms. First, is the need for a specific "tag" that can provide some molecular information about where the protein of interest is destined for. This tag usually takes the form of a short string of amino acids - a so-called signal peptide - that can encode information about where the protein should end up. The second required component of the protein sorting machinery must be a system to read and sort the proteins. In bacterial and archaeal systems this usually consists of proteins that can identify the signal peptide during translation, bind to it, and direct the synthesis of the nascent protein to the plasma membrane. In eukaryotic systems, the sorting is by necessity more complex, and involves a rather elaborate set of mechanisms of signal recognition, protein modification, and trafficking of vesicles between organelles or the membrane. These biochemical steps are initiated in the endoplasmic reticulum and further "refined" in the Golgi apparatus where proteins are modified and packaged into vesicles bound for various parts of the cell.

Some specific mechanisms may be discussed by your instructor in class. The key for all students it so appreciate the problem and to have a general idea of the high-level requirements that cells have adopted to solve them.

Reading_21 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_22

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_22

GI.45 Using a mental model for the central dogma, propose a rationale for the need to regulate each step, including biomolecule degradation.
GI.42 Describe the roles of both positive and negative transcriptional regulators in the control of gene expression.
GI.43 Given information regarding the allosteric regulation of a DNA binding protein, predict what effect changing the concentration of the allosteric regulator would have on the binding of a transcription factor to a regulatory element.
GI.41 Draw models that help explain how the allosteric binding of small molecules to both positively and negatively acting transcription factors can in both cases explain their ability to either "turn up" or "turn down" transcription in a small-molecule concentration-dependent manner.
GI.40 Describe the key components of the central dogma and the relationships between them, including the processes of replication, transcription, translation, reverse transcription and the possible fates of the various possible products (i.e. mRNA, tRNA, rRNA, enzymes, structural proteins, replicated chromosomes etc.).

Introduction to gene regulation

Regulation is all about decision making. Gene regulation is, therefore, all about understanding how cells decide about which genes to turn on, turn off or tune up or tune down. In the following section, we discuss some fundamental mechanisms and principles used by cells to regulate gene expression in response to changes in cellular or external factors. This biology is important for understanding how cells adjust changing environments, including how some cells, in multicellular organisms, become specialized for certain functions (e.g. tissues).

Since the subject of regulation is both a deep and broad topic of study in biology, in Bis2a we don't cover every detail - there are way too many. Rather, as we have done for all other topics, we try to focus on (a) outlining some core logical constructs and questions that you must have when you approach ANY scenario involving regulation, (b) learning some common vocabulary and ubiquitous mechanisms and (c) examining a few concrete examples that illustrate the points made in a and b.

Gene Expression

Introduction

All cells control when and how much each one of its genes is expressed. This simple statement - one that could be derived from observing cellular behavior - brings up many questions that we can decompose using our Design Challenge rubric.

Trying to define "gene expression"

The first thing we need to do, however, is to define what it means when we say that a gene is "expressed". If the gene encodes a protein, one might reasonably propose that "expression" of a gene means how much functional protein the cell makes. But what if the gene does not encode a protein but some functional RNA. Then, in this case, "expressed might mean how much of the functional RNA the cell makes. Yet another person might reasonably suggest that "expression" just refers to the initial step in creating a copy of the genomic information. By that definition, one might want to count how many full-length transcripts are being made. Is it the number of end products encoded by the genomic information or is it the number of reads of the information important to describe "expression" properly. Unfortunately, in practice, we often find that the definition depends on the context of the discussion. Keep that in mind. For the sake of making sure that we are talking about the same thing, in Bis2A we'll try to use the term "expression" primarily to describe the creation of the final functional product(s). Depending on the specific case, the final product may be a protein or RNA species.

The design challenge of regulating gene expression

To drive this discussion from a design challenge perspective, we can formally stipulate that the "big problem" we are interested in understating is that of regulating protein abundance in a cell. Problem: The cell must regulate the abundance of each functional protein. We can then start by posing subproblems:

Let's take a moment, though, first to reload a few ideas. The process of gene expression requires multiple steps depending on what the fate of the final product will be. With structural and regulatory RNAs (i.e. tRNA, rRNA, snRNA, etc.) the process requires that a gene be transcribed and that any needed post-transcriptional processing takes place. With a protein-coding gene, the transcript must also be translated into protein and, if required, modifications to the protein must also be made. Both transcription and translation are multi-step processes, and most of those sub-steps are also potential sites of control.

Some subproblems might therefore be:

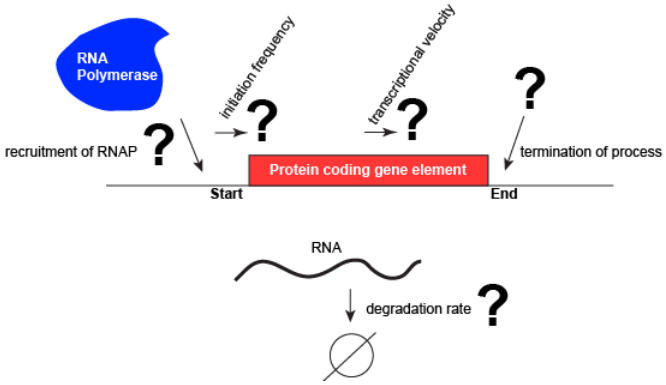
1. It is reasonable to postulate that there must be some mechanism(s) to regulate the first step of this multi-step process, the initiation of transcription (just getting things started). So, we could state, "we need a mechanism to regulate the initiation of transcription." We could also turn this into a question and ask, "how can the initiation of transcription be accomplished"?
2. We can use similar thinking to state, "we need a mechanism for regulating the end of transcription" or to ask "how is transcription terminated?"
3. Using this convention we can state, "we need to regulate the initiation of translation and the stop of translation".
4. We've talked only about the synthesis of protein and RNA. It is also reasonable to state, "we need a mechanism to regulate the degradation of RNA and protein."

Focusing on transcription

In this course we begin by focusing primarily on examining the first couple of problems/questions, regulating transcription initiation and termination - from genomic information to a functional RNA, either ready as is (e.g. with a functional RNA) or ready for translation. This allows us to examine some fundamental concepts regarding the regulation of gene expression and to examine a few real examples of those concepts in action.

Subproblems for transcription and the activity or RNA polymerase

Let us consider a protein-coding gene and work through some logic. We start by imagining a simple case, where a protein-coding gene is encoded by a single contiguous stretch of DNA. We know that to transcribe this gene an RNA polymerase will need to be recruited to the start of the coding region. The RNA polymerase is not "smart" *per se*. There needs to be some mechanism, *based on chemical logic*, to help recruit the RNA polymerase to the start of the protein-coding gene. Likewise, if this process is to be regulated, there needs to be some mechanism, or mechanisms to dictate when an RNA polymerase should be recruited to the start of a gene, ***when*** it should not, and/or ***if*** it is recruited to the DNA whether it should begin transcription and ***how many times*** this process should happen. Note, that the previous sentence, has several distinct subproblems/questions (e.g. when is the polymerase recruited?; if recruited, should it start transcription?; if it starts transcription, how many times should this process repeat?). We can also reasonably infer, that there will need to be some mechanisms to "instruct" (more anthropomorphisms) the polymerase to stop transcription. Finally, since the role of transcription is to create RNA copies of the genome segments, we should also consider problems/questions related to other factors that influence the abundance of RNA, like mechanisms of degradation. There must be some mechanisms, and these mechanisms will probably involve regulating this process.



A schematic showing a protein-coding gene and some questions or problems that we need to ask ourselves or problems we need to know solutions for if we are to understand how regulation of the transcriptional portion of the gene's expression is regulated. Attribution: Marc T. Facciotti (own work)

Activation and Repression of Transcription

Some basics

Let us consider a protein-coding gene and work through some logic. We start by imagining a simple case, where a protein-coding gene is encoded by a single contiguous stretch of DNA. We know that to transcribe this gene an RNA polymerase will need to be recruited to the start of the coding region. The RNA polymerase is not "smart" *per se*. There needs to be some mechanism, based on chemical logic, to help recruit the RNA polymerase to the start of the protein-coding gene. Likewise, if this process is to be regulated, there needs to be some mechanism, or mechanisms to dictate when an RNA polymerase should be recruited to the start of a gene, when it should not, and/or if it is recruited to the DNA, whether it should begin transcription and how many times this process should happen. Note, that the previous sentence, has several distinct subproblems/questions (e.g. when is the polymerase recruited?, if recruited should it start transcription?, if it starts transcription, how many times should this process repeat?). We can also reasonably infer, that there will need to be some mechanisms to "instruct" (more anthropomorphisms) the polymerase to stop transcription.

Recruiting RNA polymerase to specific sites

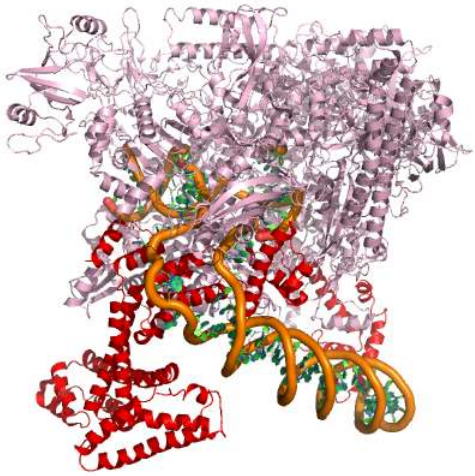
To initiate transcription, the RNA polymerase must be recruited to a segment of DNA near the start of a region of DNA encoding a functional transcript. The function of the RNA polymerase as described so far, however, is not to bind specific sequences but to move along any segment of DNA. Recruiting the polymerase to a specific site therefore seems contradictory to its usual behavior. Explaining this contradiction requires us to invoke something new. Either transcription can start anywhere and just those events that lead to a full productive transcript do anything useful or something other than the RNA polymerase itself helps to recruit the enzyme to the beginning of a gene. The latter, we now take for granted, is the case.

The recruitment of the RNA polymerase is mediated by proteins called general transcription factors. In bacteria, they have a special name: sigma factors. In archaea they are called TATA binding protein and transcription factor IIB. In eukaryotes, relatives of the archaeal proteins function with many others to recruit the RNA polymerase. The general transcription factors have at least two basic functions: (1) They can chemically recognize a specific sequence of DNA and (2) they can bind the RNA polymerase. Together these two functions of general transcription factors solve the problem of recruiting an enzyme that is otherwise not capable of binding a specific DNA site. In some texts, the general transcription factors (and particularly the sigma factor varieties) are said to be part of the RNA polymerase. While they are part of the complex when they help to target the RNA polymerase they do not continue with the RNA polymerase after it starts transcription.

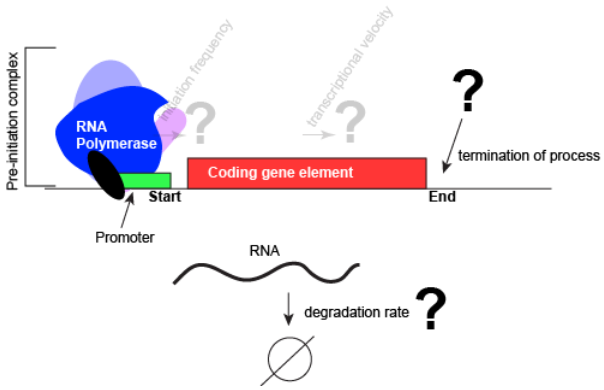
The DNA site that recruits an RNA polymerase has a special name. We call it a *promoter*. While the DNA sequences of different promoters need not be exactly the same, different promoter sequences typically have some special chemical properties in common.

Obviously, one property is that they can associate with an RNA polymerase. In addition, the promoter usually has a DNA sequence that facilitates the dissociation of the double stranded DNA such that the polymerase can begin reading and transcription the coding region. (Note: technically we could have broken down the properties of the promoter into design challenge subproblems. Here we skipped it, but you should still be able to step backwards and create the problem statements and or relevant questions once you find out about promoters).

In nearly all cases, but particularly in eukaryotic systems the complex of proteins that assembles with the RNA polymerase at promoters (typically called the pre-initiation complex) can number in the tens of proteins. Each of these other proteins has a specific function, but this is far too much detail to dive into for Bis2A.



A model of the E. coli pre-initiation complex. The sigma factor is colored red. The DNA is depicted as orange tubes and opposing ring structures. The rest of the pre-initiation complex is colored pink. Note that the DNA has regions of the double helix and an open structure inside the PIC. Attribution: Structure derived from PDB coordinates (4YLN) Marc T. Facciotti (own work)



An abstract model of a generic transcriptional unit shown above with the addition of a promoter and PIC. Questions noted earlier that will probably not be covered in Bis2a have been grayed out. Attribution: Marc T. Facciotti (own work)

States of a regulated promoter

Since promoters recruit an RNA polymerase these sites and the assembly of the pre-initiation complex are obvious sites for regulating the first steps of gene expression. At the level of transcription initiation, we often classify promoter into one of three classes. We call the first *constitutive*. Constitutive promoters are generally not regulated very strongly. Their base state is "on". When the constitutive transcription from a promoter is very high (relative to most other promoters), we will colloquially call that promoter a "strong constitutive" promoter. If the amount of transcription from a constitutive promoter is low (relative to most other

promoters) we call that promoter a "weak constitutive" promoter.

A second way to classify promoters by the use of the term *activated* or equivalently *induced*. These interchangeable terms are used to describe promoters sensitive to some external stimulus and respond to said stimulus by increasing transcription. Activated promoters have a base state exhibits little to no transcription. Transcription is then "activated" in response to a stimulus - the stimulus turns the promoter "on".

Finally, the third term used to classify promoters is by the use of the term *repressed*. These promoters also respond to stimuli but do so by decreasing transcription. The base state for these promoters is some basal level of transcription, and the stimulus acts to turn down or repress transcription. Transcription is "repressed" in response to a stimulus - the stimulus turns the promoter "off".

The examples given above assumed that a single stimulus acts to regulate promoters. While this is the simplest case, many promoters may integrate different information and may be alternately activated by some stimuli and repressed by other stimuli.

Possible NB Discussion Point

In Bis2A, you have learned about how important it is for the cell to be able to regulate its biology, with several examples of regulatory mechanisms. Given this, can you explain why it is advantageous to still have constitutive promoters that are generally not regulated very strongly? What kinds of genes would you expect to have a strong constitutive promoter? What about a weak constitutive promoter?

Transcription factors help to regulate the behavior of a promoter

How are promoters sensitive to external stimuli? Mechanistically, in both activation and repression require regulatory proteins to change the transcriptional output of the gene being observed. The proteins responsible for helping to regulate expression are generally called *transcription factors*. The specific DNA sequences bound by transcription factors are often called operators and most times the operators are very close to the promoter sequences.

Here's where the nomenclature gets potentially confusing - particularly when comparing across bacterial and eukaryotic research. In bacterial research, if the transcription factor acts by binding DNA and the RNA polymerase in a way that increases transcription, then it is typically called an *activator*. If, by contrast, the transcription factor acts by binding DNA to repress or decrease transcription of the gene then it is called a *repressor*.

Why are the classifications of activator and repressor potentially problematic? These terms describe the proteins with idealized single functions. While this may be true with some transcription factors, in reality other transcription factors may act to activate gene expression in some conditions while repressing in other conditions. Some transcription factors will act to modulate expression either up or down depending on context rather than shutting transcription "off" or turning it completely "on". To circumvent some of this confusion, some of your instructors prefer to avoid using the terms activator and repressor and instead prefer to discuss the activity of transcription various transcription factors as either a positive or a negative influence on gene expression in specific cases. If we use these terms, you might hear your instructor saying that the transcription factor in question ACTS LIKE/AS a repressor or that it ACTS LIKE/AS an activator, taking care not to call it simply an activator or repressor. It is more likely, however, that you will hear them say that a transcription factor is acting to positively or negatively influence transcription.

CAUTION: Depending on your instructor, you may cover a few real biological examples of positive and negative regulatory mechanisms. These specific examples will use the common names of the transcription factors - since the examples will typically be drawn from the bacterial literature, the names of the transcription factors may include the terms "repressor" or "activator". These names are relics of when they were first discovered. Try to spend more time examining the logic of how the system works than trying to commit to memory any special properties of that specific protein to all other cases with the same label. Just because we

label a protein as a repressor does not mean that it only acts as a negative regulator in all cases or that it interacts with external signals in the same was as the example.

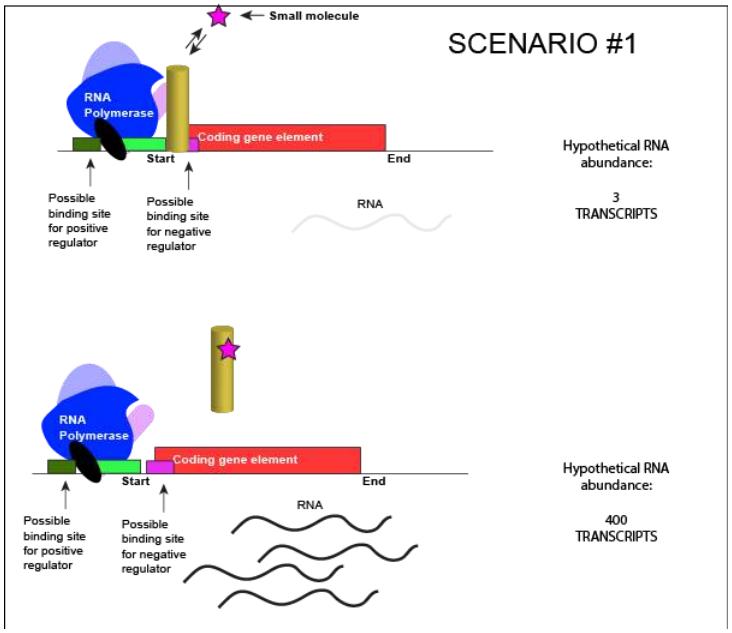
Possible NB Discussion Point

What types of interactions do you think happen between the amino acids of the transcription factor and the double helix of the DNA? How do transcription factors recognize their binding site on the DNA?

Allosteric Modulators of Regulatory Proteins

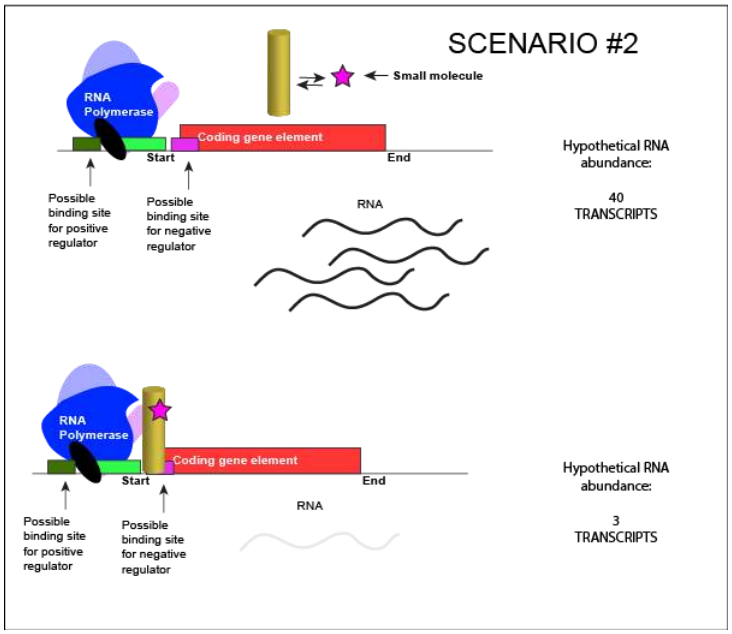
The activity of many proteins, including regulatory proteins and various transcription factors, can be allosterically modulated various factors, including by the relative abundance of small molecules in the cell. We often refer these small molecules as *inducers* or *co-repressors*or *co-activators* and are often metabolites, such as lactose or tryptophan or small regulatory molecules, such as cAMP or GTP. These interactions allow the TF to respond to environmental conditions and to modulate its function accordingly. It helps to decide whether to transcribe a gene depending on the abundance of the environmental signal.

Let us imagine a negative transcriptional regulator. In the most simple case we've considered so far, transcription of a gene with a binding site for this transcription factor would be low when the TF is present and high when the TF is absent. We can now add a small molecule to this model. Here the small molecule can bind the negative transcriptional regulator through sets of complementary hydrogen and ionic bonds. In this first example we will consider the case where the binding of the small molecule to the TF induces a conformational change to the TF that severely reduces its ability to bind DNA. If this is the case, the negative regulator - once bound by its small molecule - would release from the DNA. This would thereby relieve the negative influence and lead to increased transcription. This regulatory logic might be appropriate to have evolved in the following scenario: a small molecule food-stuff is typically absent from the environment. Therefore, genes encoding enzymes that will degrade/use that food should be kept "off" most of the time to preserve the cellular energy that their synthesis would use. A negative regulator could accomplish this. When the food-stuff appears in the environment, it would be appropriate for the enzymes responsible for its processing to be expressed. The food-stuff could then act by binding to the negative regulator, changing the TF's conformation, causing its release from the DNA and turning on transcription of the processing enzymes.



An abstract model of a generic transcriptional unit regulated by a negative regulator whose activity is modulated by a small molecule (depicted by a star). Here, the binding of the small molecule causes the TF to release from the DNA. Attribution: Marc T. Facciotti (own work)

We can consider a second model for how a negatively acting TF might interact with a small molecule. Here, the TF alone cannot bind its regulatory site to the DNA. However, when a small molecule binds to the TF a conformational change occurs that reorients DNA binding amino-acids into the "correct" orientation for DNA binding. The TF-small molecule complex now binds to the DNA and acts to influence transcription negatively.



An abstract model of a generic transcriptional unit regulated by a negative regulator whose activity is modulated by a small molecule (depicted by a star). Here, binding of the small molecule causes the TF to bind to the DNA. Attribution: Marc T. Facciotti (own work)

Note how the activity of the TF can be modulated in distinctly different ways by a small molecule. Depending on the protein, the binding of this external signal can either cause binding of the TF-small molecule complex to DNA OR binding of the small molecule can cause the release of the TF-small molecule complex from the DNA. We can work up the same kinds of examples for a positive regulator.

In both cases proposed above, the binding of a small molecule to a TF will depend on how strongly the TF interacts with the small molecule. This will depend on the types and spatial orientation of the protein's chemical functional groups, their protonation states (if applicable), and the complementary functional groups on the small molecule. It should not be surprising, therefore, to learn that the binding of the small molecule to the TF will depend on various factors, including but not limited to the relative concentrations of small-molecule and TF and pH.

Is it positive or negative regulation?

Resolving a common point of confusion

It is not uncommon for many Bis2a students to be slightly confused about how to determine if a transcription factor is acting as a positive or negative regulator. This confusion often comes after a discussion of the modes that the stimulus (i.e. small molecule) can influence the activity of a transcription factor. This is not too surprising. In the examples above, the binding of an effector molecule to a transcription factor could have one of two different effects: (1) binding of the effector molecule could induce a DNA-bound transcription factor to release from its binding site, derepressing a promoter, and "turning on" gene expression. (2) binding of the effector molecule to the transcription factor could induce the TF to bind to its DNA binding site, repressing transcription, and "turning off" gene expression. In the first case, it might appear that the TF is acting to regulate expression positively, while in the second example it might appear that the TF acts negatively.

However, in both examples above, the TF is acting as a negative regulator. To determine this, we look at what happens when the TF binds DNA (whether a small molecule is bound to the TF or not). In both cases, binding of the TF to DNA represses transcription. The TF is therefore acting as a negative regulator. We can conduct a similar analysis fr positively acting TFs.

Note that sometimes a TF may act as a positive regulator at one promoter and negative regulator at a different promoter so describing the behavior of the TF on a per case basis is often important (reading too much from the name it has been assigned can mislead sometimes). Other TF proteins can act alternately as both positive or negative regulators of the same promoter depending on conditions. Again, describing the behavior of the TF specifically for each case is advised.

A genetic test for positive or negative regulatory function of a TF

How does one determine if a regulatory protein functions in a positive or negative way? A simple genetic test is to ask "what happens to expression if the regulatory protein is absent?" This can be accomplished by removing the coding gene for the transcription factor from the genome. If a transcription factor acts positively, then its presence is required to activate transcription. In its absence, there is no regulatory protein, therefore no activation, and the outcome is lower transcription levels of a target gene. The opposite is true for a transcription factor acting negatively. In its absence expression should be increased, because the gene keeping expression low is no longer around.

Termination of Transcription and RNA degradation

Termination of transcription in bacteria

Once a gene is transcribed, the bacterial polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals. One is protein-based and the other is RNA-based. *Rho-dependent termination* is controlled by the rho protein, which tracks along behind the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

Rho-independent termination is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in C–G nucleotides. The mRNA folds back on itself, and the complementary C–G nucleotides bind together. The result is a stable *hairpin* that causes the polymerase to stall as soon as it transcribes a region rich in A–T nucleotides. The complementary U–A region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

Termination of transcription in eukaryotes

The termination of transcription is different for the different polymerases. Unlike in prokaryotes, elongation by RNA polymerase II in eukaryotes takes place 1,000–2,000 nucleotides beyond the end of the gene being transcribed. This pre-mRNA tail is subsequently removed by cleavage during mRNA processing. On the other hand, RNA polymerases I and III require termination signals. Genes transcribed by RNA polymerase I contain a specific 18-nucleotide sequence that is recognized by a termination protein. The process of termination in RNA polymerase III involves an mRNA hairpin similar to rho-independent termination of transcription in prokaryotes.

Termination of transcription in archaea

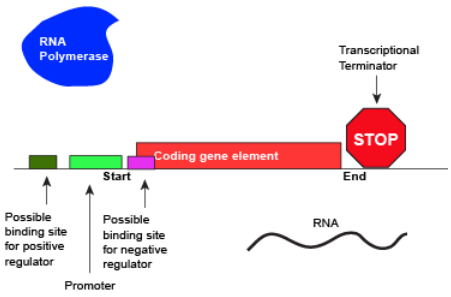
Termination of transcription in the archaea is far less studied than in the other two domains of life and is still not well understood. While the functional details are likely to resemble mechanisms that have been seen in the other domains of life the details are beyond the scope of this course.

Degradation of RNA

The lifetime of different RNA species in the cell can vary dramatically, from seconds to hours. The mean lifetime of mRNA can also vary dramatically depending on the organism. For instance, the median lifetime for mRNA in *E. coli* is ~5 minutes. The half-life of mRNA in yeast is ~20 minutes and 600 minutes for human cells. Some degradation is "targeted". Some transcripts include a short sequence that targets them for RNA degrading enzymes, speeding the degradation rate. It doesn't take too much imagination to infer that this process might also be evolutionarily tuned for different genes. Realizing that degradation - and the tuning of degradation - can also be a factor in controlling the expression of a gene is sufficient for Bis2a.

Tuning termination

Like all other biological processes, the termination of transcription is not perfect. Sometimes, the RNA polymerase can read through terminator sequences and transcribe adjacent genes. This is particularly true in bacterial and archaeal genomes where the density of coding genes is high. This transcriptional read-through can have various effects but the most common two outcomes are: (1) If the adjacent genes are encoded on the same DNA strand as the actively transcribed gene, the adjacent genes may also be transcribed. (2) If the adjacent genes are transcribed on the opposite strand of the actively transcribed genes, RNA polymerase read-through by interfere with polymerases actively transcribing the neighboring gene. Not surprisingly, biology has sometimes evolved mechanisms that both act to minimize the influence of read-through and to take advantage of it. Therefore, the "strength" of a terminator - and its effectiveness of terminating transcription in a particular direction - are "tuned" by Nature and "used" (note the anthropomorphism in quotes) to regulate the expression of genes.



An abstract model of a full basic transcriptional unit and the various "parts" encoded on the DNA that may influence the expression of the gene. We expect students in Bis2A to be able to recreate a similar conceptual figure from memory. Attribution: Marc T. Facciotti (own work)

Summary of gene regulation

In the preceding text we have examined several ways to solve some design challenges associated with regulating the amount of transcript that is created for a single coding region of the genome. We have looked in abstract terms at some processes responsible for controlling the initiation of transcription, how these may be made sensitive to environmental factors, and very briefly at the processes that terminate transcription and handle the active degradation of RNA. We have also discussed how Nature can quantitatively tune each of these processes to be "stronger" or "weaker" and that this tuning of "strength" (e.g. promoter strength, degradation rates, etc.) Influence the behavior of the overall process in potentially functionally important ways.

Reading_22 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_23

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_23

ME.22 Describe the mechanism of allosteric regulation and provide examples of how this mechanism can be used to tune the flow of molecules in a metabolic pathway.
GI.39 Explain the key differences between bacterial and eukaryotic gene expression. Be sure to include in your explanation the role compartmentalization plays in these differences.
GI.44 Given some specific constraints on a regulatory circuit, predict the behavior of the circuit under the possible influence of environmental or genetic perturbations (changes).
GI.46 Be able to rank the different levels of regulation discussed in class by energy efficiency for the cell and create statements that illustrate the logic used in the rankings.
GI.47 Using a regulatory network, accurately interpret relationships between genes/proteins in the network and predict potential targets of regulation by different possible types of signaling molecules.
GI.51 Describe how environmental information can shape transcriptional and translational output in ways that lead to different phenotypes and cellular specialization.
GI.52 Explain and describe situations when single-celled organisms would need to control gene expression and compare and contrast that with the needs of multicellular organisms.
GI.6 Describe how genomic DNA is packaged both in prokaryotes and eukaryotes and the contributions of basic proteins, such as histones and histone-like proteins, to this packaging.
GI.7 Given a molecular representation of DNA-protein interaction, analyze and interpret the diagram to make sensible predictions about which bonds provide specificity to the interaction, and which make energetically favorable contributions to the association.

Examples of Bacterial Gene Regulation

This section describes two examples of transcriptional regulation in bacteria. These are presented as illustrative examples. Use these examples to learn some basic principles about mechanisms of transcriptional regulation. Be on the lookout in class, in discussion, and in the study-guides for extensions of these ideas and use these to explain the regulatory mechanisms used for regulating other genes.

Gene Regulation Examples in *E. coli*

The DNA of bacteria and archaea are usually organized into one or more circular chromosomes in the cytoplasm. The dense aggregate of DNA that can be seen in electron micrographs is called the nucleoid. In bacteria and archaea, genes, whose expression needs to be tightly coordinated (e.g. genes encoding proteins involved in the same biochemical pathway) are often grouped closely together in the genome. When the expression of multiple genes is controlled by the same promoter and a single transcript is produced these expression units are called *operons*. For example, in the bacterium *Escherichia coli all the genes needed to utilize lactose* are encoded next to one another in the genome. We call this arrangement the lactose (or *lac*) operon. In many bacteria and archaea nearly 50% of all genes are encoded into operons of two or more genes.

The Role of the Promoter

The first level of control of gene expression is at the promoter itself. Some promoters recruit RNA polymerase and turn those DNA-protein binding events into transcripts more efficiently than other promoters. This intrinsic property of a promoter, it's ability to produce transcript at a particular rate, is referred to as promoter strength. The stronger the promoter, the more RNA is made in any given time period. Promoter strength can be "tuned" by Nature in very small or very large steps by changing the nucleotide sequence the promoter (e.g. mutating the promoter). This results in families of promoters with different strengths that can be used to control the maximum rate of gene expression for certain genes.

UC Davis Undergraduate Connection:

A group of UC Davis students interested in synthetic biology used this idea to create synthetic promoter libraries for engineering microbes as part of their design project for the 2011 iGEM competition.

Example #1: Trp Operon

Logic for regulating tryptophan biosynthesis

E. coli, like all organisms, needs to either synthesize or consume amino acids to survive. The amino acid *tryptophan* is one such amino acid. *E. coli* can either import tryptophan from the environment (eating what it can scavenge from the world around it) or synthesize tryptophan *de novo* using enzymes encoded by five genes. These five genes reside next to each other in the *E. coli* genome in what we call the *tryptophan (trp) operon* (Figure below). If tryptophan is present in the environment, then *E. coli* need not synthesize it and the switch controlling the activation of the genes in the *trp* operon switches off. However, when environmental tryptophan availability is low, the switch controlling the operon is turned on, transcription is initiated, the genes are expressed, and the organism synthesizes tryptophan. See the figure and paragraphs below for a mechanistic explanation.

Organization of the *trp* operon

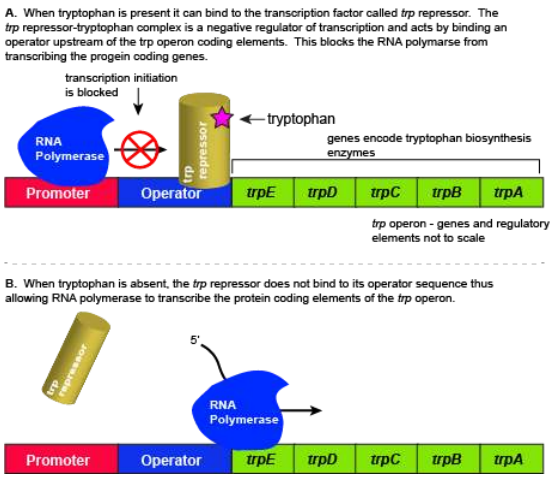
The five genes encoding tryptophan biosynthesis enzymes are arranged sequentially on the chromosome and are under the control of a single promoter - i.e. natural selection organized these genes into an operon. Just before the coding region is the *transcriptional start site*. This is, as the name implies, the location where the RNA polymerase starts a new transcript. The promoter sequence is further upstream of the transcriptional start site.

A DNA sequence called an "operator" is also encoded between the promoter and the first *trp* coding gene. This *operator* is the DNA sequence to which the transcription factor protein will bind.

A few more details regarding TF binding sites

We should note that the use of the term "operator" is limited to just a few regulatory systems and almost always refers to the binding site for a negatively acting transcription factor. Conceptually what you need to remember is that there are sites on the DNA that interact with regulatory proteins, allowing them to perform their appropriate function (e.g. repress or activate transcription). This theme will repeat universally across biology whether the "operator" term is used.

While the specific examples you will be show depict TF binding sites in their known locations, these locations are not universal to all systems. Transcription factor binding sites can vary in location relative to the promoter. There are some patterns (e.g. positive regulators are often upstream of the promoter and negative regulators bind downstream), but these generalizations are not true for all cases. Again, the key thing to remember is that transcription factors (both positive and negatively acting) have binding sites with which they interact to help regulate the initiation of transcription by RNA polymerase.



The five genes that required to synthesize tryptophan in E. coli group next to each other in the trp operon. When tryptophan is plentiful, two tryptophan molecules bind to the transcription factor and allow the TF-tryptophan complex to bind at the operator sequence. This physically blocks the RNA polymerase from transcribing the tryptophan biosynthesis genes. When tryptophan is absent, the transcription factor does not bind to the operator and the genes are transcribed. Attribution: Marc T. Facciotti (own work)

Regulation of the *trp* operon

When tryptophan is present in the cell: two tryptophan molecules bind to the *trp* repressor protein. When tryptophan binds to the transcription factor it causes a conformational change in the protein which now allows the TF-tryptophan complex to bind to the *trp* operator sequence. Binding of the tryptophan–repressor complex at the operator physically prevents the RNA polymerase from binding and transcribing the downstream genes. When tryptophan is not present in the cell, the transcription factor does not bind to the operator; therefore, the transcription proceeds, the tryptophan utilization genes are transcribed and translated, and tryptophan is thus synthesized.

Since the transcription factor actively binds to the operator to keep the genes turned off, the *trp* operon is said to be "negatively regulated" and the proteins that bind to the operator to silence *trp* expression are *negative regulators*.

Possible NB Discussion Point

Suppose nature took a different approach to regulating the trp operon. Propose a method for regulating the expression of the *trp* operon with a positive regulator instead of a negative regulator. Describe how this might work. (**Hint:** we ask this kind of question on exams)

External link

Watch this video to learn more about the *trp* operon.

Example #2: The *lac* operon

Rationale for studying the *lac* operon

In this example, we examine the regulation of genes encoding proteins whose physiological role is to import and assimilate the disaccharide lactose, the *lac* operon. The story of regulating *lac* operon is a common example used in many introductory biology classes to illustrate basic principles of inducible gene regulation. We describe this example second because it is, in our estimation, more complicated than the previous example involving the activity of a single negatively acting transcription factor. By contrast, the regulation of the *lac* operon is a wonderful example of how the coordinated activity of both positive and negative regulators around the same promoter can integrate multiple different sources of cellular information to regulate the expression of genes.

As you go through this example, keep in mind the last point. For many Bis2a instructors, it is more important for you to learn the lac operon story and guiding principles than it is for you to memorize the logic table presented below. When this is the case, the instructor will usually let you know. These instructors often deliberately do NOT include exam questions about the lac operon. Rather, they will test you on whether you understood the fundamental principles underlying the regulatory mechanisms that you study using the lac operon example. If it's not clear what the instructor wants, ask.

The utilization of lactose

Lactose is a disaccharide composed of the hexoses glucose and galactose. We commonly encounter lactose in milk and some milk products. As one can imagine, the disaccharide can be an important food-stuff for microbes that can use its two hexoses. *E. coli* can use multiple different sugars as energy and carbon sources, including lactose and the *lac operon* is a structure that encodes the genes necessary to acquire and process lactose from the local environment. *E. coli*, however, does not frequently encounter lactose,

and therefore the genes of the *lac* operon must typically be repressed (i.e. "turned off") when lactose is absent. Driving transcription of these genes when lactose is absent would waste precious cellular energy. By contrast, when lactose is present, it would make logical sense for the genes responsible for using the sugar to be expressed (i.e. "turned on"). So far, the story is very similar to that of the tryptophan operon described above.

However, there is a catch. Experiments conducted in the 1950's by Jacob and Monod showed that *E. coli* prefers to use all the glucose present in the environment before it uses lactose. This means that the mechanism used to decide whether or not to express the lactose utilization genes must be able to integrate two types of information (1) the concentration of glucose and (2) the concentration of lactose. While this could theoretically be accomplished in multiple ways, we will examine how the *lac* operon accomplishes this by using multiple transcription factors.

The transcriptional regulators of the *lac* operon

The *lac* repressor - a direct sensor of lactose

As noted, the *lac* operon normally has very low to no transcriptional output in the absence of lactose. This is because of two factors: (1) the constitutive promoter strength for the operon is relatively low and (2) the constant presence of the LacI repressor protein negatively influences transcription. This protein binds to the operator site near the promoter and blocks RNA polymerase from transcribing the *lac* operon genes. By contrast, if lactose is present, lactose will bind to the LacI protein, inducing a conformational change that prevents LacI-lactose complex from binding to its binding sites. Therefore, when lactose is present, the negative regulatory LacI is not bound to its binding site and transcription of lactose using genes can proceed.

CAP protein - an indirect sensor of glucose

In *E. coli*, when glucose levels drop, the small molecule *cyclic AMP (cAMP)* accumulates in the cell. cAMP is a common signaling molecule that is involved in glucose and energy metabolism in many organisms. When glucose levels decline in the cell, the increasing concentrations of cAMP allow this compound to bind to the positive transcriptional regulator called *catabolite activator protein (CAP)* - also referred to as CRP. cAMP-CAP complex has many sites throughout the *E. coli* genome and many of these sites are located near the promoters of many operons that control the processing of various sugars.

In the *lac* operon, the cAMP-CAP binding site is located upstream of the promoter. Binding of cAMP-CAP to the DNA helps to recruit and keep RNA polymerase to the promoter. The increased occupancy of RNA polymerase to its promoter, in turn, results in increased transcriptional output. Here, the CAP protein is acting as a positive regulator.

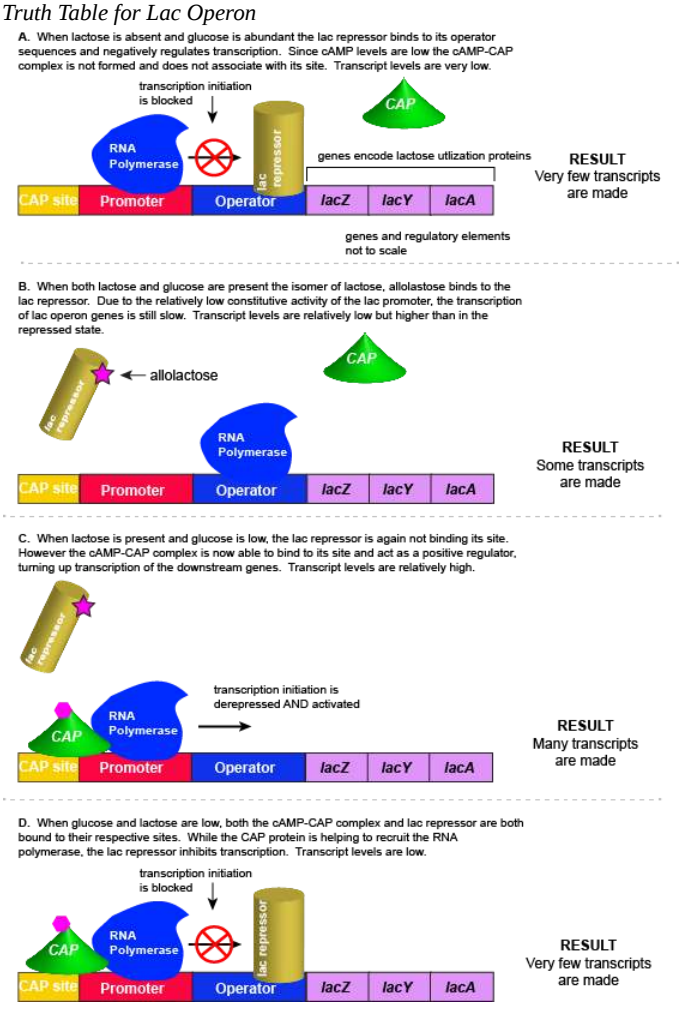
Note that the CAP-cAMP complex can, in other operons, also act as a negative regulator depending upon where the binding site for CAP-cAMP complex is located relative to the RNA polymerase binding site.

Putting it all together: Inducing expression of the lac operon

For the *lac* operon to be activated, two conditions must be met. First, the level of glucose must be very low or non-existent. Second, lactose must be present. Only when glucose is absent and lactose is present, will the *lac* operon be transcribed. When this condition is achieved the LacI-lactose complex dissociates the negative regulator from near the promoter, freeing the RNA polymerase to transcribe the operon's genes. High cAMP (indirectly indicative of low glucose) levels trigger the formation of the CAP-cAMP complex. This TF-inducer pair now bind near the promoter and act to positively recruit the RNA polymerase. This added positive influence boosts transcriptional output and lactose can be efficiently utilized. The mechanistic output of other combinations of binary glucose and lactose conditions are described in the table below and in the figure that follows.

Signals that Induce or Repress Transcription of the *lac* Operon				
Glucose	CAP binds	Lactose	Repressor binds	Transcription
+	-	-	+	No
+	-	+	-	Some
-	+	-	+	No
-	+	+	-	Yes

Truth Table for Lac Operon

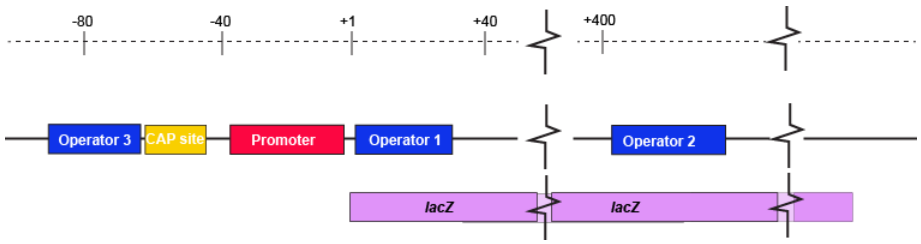


Transcription of the lac operon is carefully regulated so that its expression only occurs when glucose is limited and lactose is present to serve as an alternative fuel source.
Attribution: Marc T. Facciotti (own work)

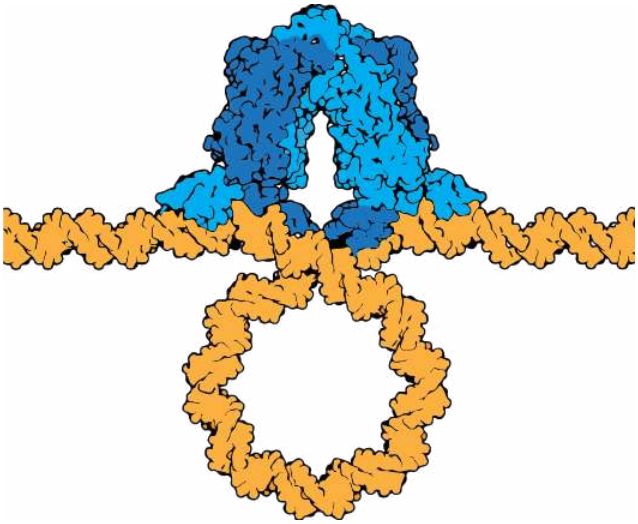
A more nuanced view of lac repressor function

The description of the lac repressor's function correctly describes the logic of the control mechanism used around the lac promoter. However, the molecular description of binding sites is a bit overly simplified. In reality, the lac repressor has three similar, but not identical, binding sites called Operator 1, Operator 2, and Operator 3. Operator 1 is very close to the transcript start site (denoted

+1). Operator 2 is located about +400nt into the coding region of the LacZ protein. Operator 3 is located about -80nt before the transcript start site (just "outside" of the CAP binding site).



The lac operon regulatory region depicting the promoter, three lac operators, and CAP binding site. The coding region for the Lac Z protein is also shown relative to the operator sequences. Note that two of the operators are in the protein coding region - there are multiple different types of information simultaneously encoded in the DNA. Attribution: Marc T. Facciotti (own work)



The lac repressor tetramer (blue) depicted binding two operators on a strand of looped DNA (orange).
Attribution: Marc T. Facciotti (own work) - Adapted from Goodsell (https://pdb101.rcsb.org/motm/39)

Eukaryotic Gene Regulation



Regulation overview

As previously noted, regulation is all about decision making. Gene regulation, as a general topic, relates to deciding about the functional expression of genetic material. Whether the final product is an RNA species or a protein, the production of the final expressed product requires processes that take multiple steps. We have spent some time discussing some of these steps (i.e. transcription and translation) and some mechanisms that nature uses for sensing cellular and environmental information to regulate the initiation of transcription.

When we discussed the concept of strong and weak promoters, we introduced the idea that regulating the amount (number of molecules) of a transcript produced from a promoter in some unit of time might also be important for function. This should not be entirely surprising. For a protein-coding gene, the more transcript produced, the greater potential there is to make more protein. This might be important when making a lot of a particular enzyme is key for survival. In other cases, the cell needs only a little of a specific protein and making too much would be a waste of cellular resources. Here, the cell may prefer low levels of

transcription. Promoters of differing strengths can accommodate these varying needs. Regarding transcript number, we also briefly mentioned that synthesis is not the only way to regulate abundance. Degradation processes are also important to consider.

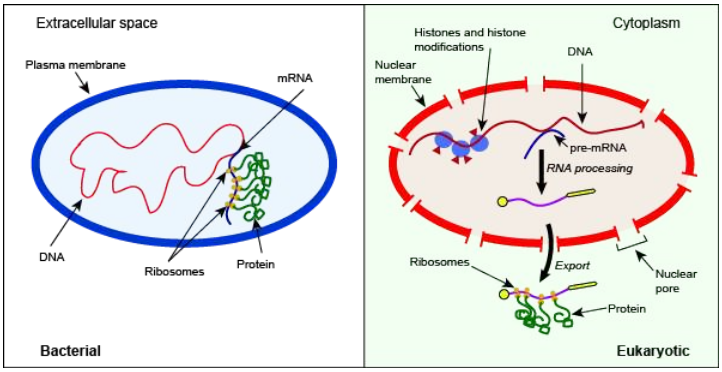
In this section, we add to these themes by focusing on eukaryotic regulatory processes. Specifically, we examine - and sometimes re-examine - the multiple steps required to express genetic material in eukaryotic organisms in the context of regulation. We want you not only to think about the processes but also to recognize that each step in the process of expression is also an opportunity to fine tune not only the abundance of a transcript or protein but also its functional state, form (or variant), and/or stability. Each of these additional factors may also be vitally important to consider for influencing the abundance of conditionally specific functional variants.

Structural differences between bacterial and eukaryotic cells influencing gene regulation

The defining hallmark of the eukaryotic cell is the nucleus, a double membrane that encloses the cell's hereditary material. In order to efficiently fits the organism's DNA into the confined space of the nucleus, the DNA is first packaged and organized by protein into a structure called chromatin. This packaging of the nuclear material reduces access to specific parts of the chromatin. Some elements of the DNA are so tightly packed that the transcriptional machinery cannot access regulatory sites like promoters. This means that one of the first sites of transcriptional regulation in eukaryotes must be the control access to the DNA itself. Chromatin proteins can be subject to enzymatic modification that can influence whether they bind tightly (limited transcriptional access) or more loosely (greater transcriptional access) to a segment of DNA . This process of modification - whichever direction is considered first - is reversible. Therefore, DNA can be dynamically sequestered and made available when the "time is right".

The regulation of gene expression in eukaryotes also involves some of the same additional fundamental mechanisms discussed in the module on bacterial regulation (i.e. the use of strong or weak promoters, transcription factors, terminators etc.) but the actual number of proteins involved is typically much greater in eukaryotes than bacteria or archaea.

The post-transcriptional enzymatic processing of RNA that occurs in the nucleus and the export of the mature mRNA to the cytosol are two additional difference between bacterial and eukaryotic gene regulation. We will consider this level of regulation in more detail below.



Depiction of some key differences between the processes of bacterial and eukaryotic gene expression. Note in this case the presence of histone and histone modifiers, the splicing of pre-mRNA, and the export of the mature RNA from the nucleus as key differentiators between the bacterial and eukaryotic systems. Attribution: Marc T. Facciotti (own work)

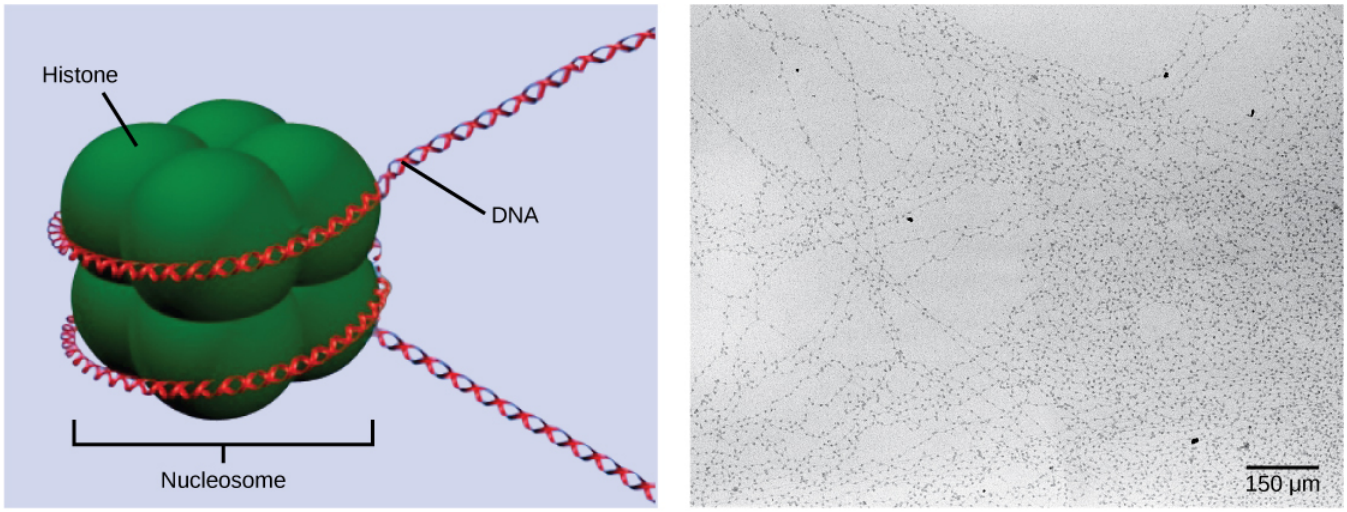
DNA Packing and Epigenetic Markers

The DNA in eukaryotic cells is precisely wound, folded, and compacted into chromosomes so it will fit into the nucleus. The nucleus also organizes the DNA so key proteins can easily access specific segments of the chromosomes as needed. Areas of the chromosomes that are more tightly compacted will be harder for proteins to bind and therefore lead to reduced gene expression of

genes encoded in those areas. Loosely compacted regions of the genome will be easier for proteins to access, thus increasing the likelihood that a gene will be transcribed. Discussed here are how cells regulate the density of DNA compaction.

DNA packing

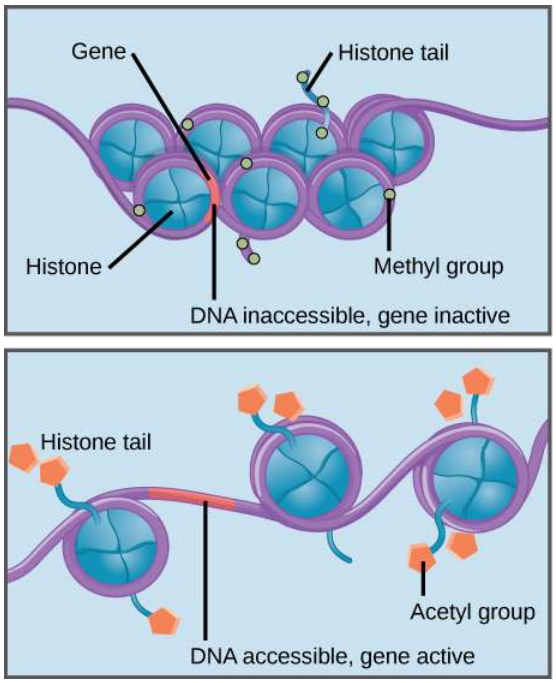
The first level of organization, or packing, is the winding of DNA strands around *histone* proteins. Histones package and order DNA into structural units called *nucleosomes*, which can control the access of proteins to specific DNA regions. Under the electron microscope, this winding of DNA around histone proteins to form nucleosomes looks like small beads on a string. These beads (nucleosome complexes) can move along the string (DNA) to alter which areas of the DNA are accessible to transcriptional machinery. While nucleosomes can move to open the chromosome structure to expose a segment of DNA, they do so in a very controlled manner.



DNA folds around histone proteins to create (a) nucleosome complexes. These nucleosomes control the access of proteins to the underlying DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (credit "micrograph": modification of work by Chris Woodcock)

Histone Modification

How the histone proteins move depends on chemical signals found on both the histone proteins and on the DNA. These chemical signals are chemical tags added to histone proteins and the DNA that tell the histones if a chromosomal region should be "open" or "closed". The figure below depicts modifications to histone proteins and DNA. These tags are not permanent, but may be added or removed as needed. They are chemical modifications (phosphate, methyl, or acetyl groups) that attach to specific amino acids in the histone proteins or to the nucleotides of the DNA. The tags do not alter the DNA base sequence, but they do alter how tightly wound the DNA is around the histone proteins. DNA is a negatively charged molecule; therefore, changes in the histone's charge will change how tightly wound the DNA molecule will be. When unmodified, the histone proteins have a large positive charge; by adding chemical modifications like acetyl groups, the charge becomes less positive.



Methylation of DNA and histones causes nucleosomes to pack tightly together. Transcription factors cannot bind the DNA, and genes are not expressed.

Histone acetylation results in loose packing of nucleosomes. Transcription factors can bind the DNA and genes are expressed.

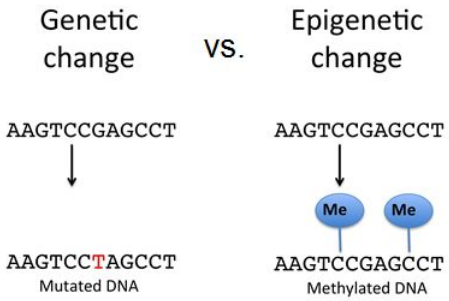
Nucleosomes can slide along DNA. When nucleosomes are spaced closely together (top), transcription factors cannot bind and gene expression is turned off. When the nucleosomes are spaced far apart (bottom), the DNA is exposed. Transcription factors can bind, allowing gene expression to occur. Modifications to the histones and DNA affect nucleosome spacing.

Possible NB Discussion Point

In the later maturation phase of sperm cells, histones (containing high numbers of lysine amino acids) are replaced by protamines, which are small, nuclear proteins that are very rich in arginine amino acids. This process is said to be essential for sperm head condensation and DNA stabilization. Based on this information, what comparisons can you draw between protamines and histones? Why is it significant that there are high numbers of lysine and arginine in histones and protamines? For what reasons do you think protamines replace histones in sperm but not other cells?

DNA Modification

The DNA molecule itself can also be modified. This occurs within very specific regions called CpG islands. These are stretches with a high frequency of cytosine and guanine dinucleotide DNA pairs (CG) often found in the promoter regions of genes. When this configuration exists, the cytosine member of the pair can be methylated (a methyl group is added). This modification changes how the DNA interacts with proteins, including the histone proteins that control access to the region. Highly methylated (hypermethylated) DNA regions with deacetylated histones are tightly coiled and transcriptionally inactive.



Epigenetic changes do not result in permanent changes in the DNA sequence. Epigenetic changes alter the chromatin structure (protein-DNA complex) to allow or deny access to transcribe genes. DNA modification such as methylation on cytosine nucleotides can either recruit repressor proteins that block RNA polymerase's access to transcribe a gene or they can aid in compacting the DNA to block all protein access to that area of the genome. These changes are reversible whereas mutations are not, however, epigenetic changes to the chromosome can also be inherited.
Source: modified from https://researcherblogski.wordpress....r/dudiwarsito/

Regulation of gene expression through chromatin remodeling is called epigenetic regulation. Epigenetic means “around genetics.” The changes that occur to the histone proteins and DNA do not alter the nucleotide sequence and are not permanent. Instead, these changes are temporary (although they often persist through multiple rounds of cell division and can be inherited) and alter the chromosomal structure (open or closed) as needed.

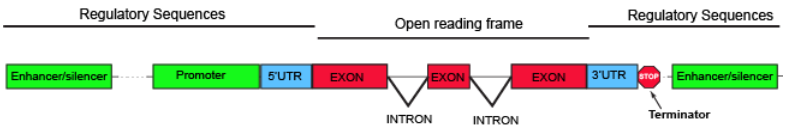
External link

View this video that describes how epigenetic regulation controls gene expression.

Eukaryotic gene structure and RNA processing

Eukaryotic gene structure

Many eukaryotic genes, particularly those encoding protein products, are encoded on the genome *discontinuously*. The coding region is broken into pieces by intervening non-coding gene elements. We term the coding regions *exons* while the intervening non-coding elements are termed *introns*. The figure below depicts a generic eukaryotic gene.



The parts of a typical discontinuous eukaryotic gene. Attribution: Marc T. Facciotti (own work)

Parts of a generic eukaryotic gene include familiar elements like a promoter and terminator. Between those two elements, the region encoding all the elements of the gene that have the potential to be translated (they have no stop codons), like in bacterial systems, is called the open reading frame (ORF). Enhancer and/or silencer elements are regions of the DNA that recruit regulatory proteins. These can be relatively close to the promoter, like in bacterial systems, or thousands of nucleotides away. Also present in many bacterial transcripts, 5' and 3' untranslated regions (UTRs) also exist. These regions of the gene encode segments of the transcript, which, as their names imply, are not translated and sit 5' and 3', respectively, to the ORF. The UTRs typically encode some regulatory elements critical for regulating transcription or steps of gene expression that occur post-transcriptionally.

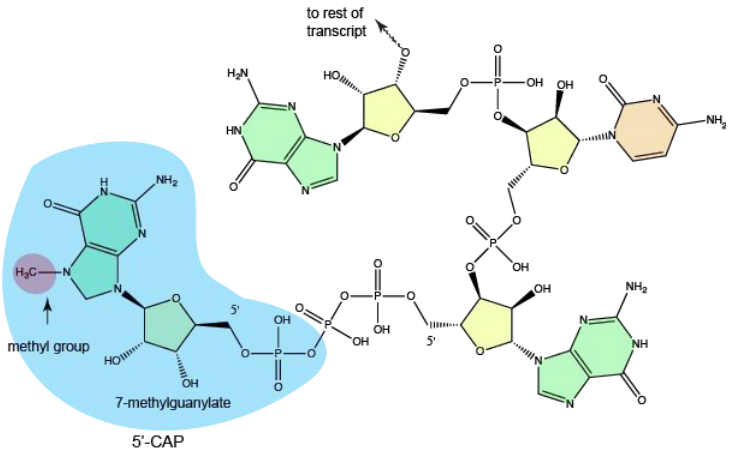
The RNA species resulting from the transcription of these genes are also discontinuous and must therefore be processed before exiting the nucleus to be translated or used in the cytosol as mature RNAs. In eukaryotic systems this includes RNA splicing, 5' capping, 3' end cleavage and polyadenylation. This series of steps is a complex molecular process that must occur within the closed

confines of the nucleus. Each one of these steps provides an opportunity for regulating the abundance of exported transcripts and the functional forms that these transcripts will take. While these would be topics for more advanced courses, think about how to frame some of the following topics as subproblems of the Design Challenge of genetic regulation. If nothing else, begin to appreciate the highly orchestrated molecular dance that must occur to express a gene and how this is a stunning bit of evolutionary engineering.

5' capping

Like in bacterial systems, eukaryotic systems must assemble a pre-initiation complex at and around the promoter sequence to start transcription. The complexes that assemble in eukaryotes serve many of the same function as those in bacterial systems, but they are significantly more complex, involving many more regulatory proteins. This added complexity allows for greater regulation and for the assembly of proteins with functions that occur predominantly in eukaryotic systems. One of these additional functions is the "capping" of nascent transcripts.

In eukaryotic protein-coding genes, the RNA that is first produced is called the pre-mRNA. The "pre" prefix signifies that this is not the full mature mRNA that will be translated and that it first requires some processing. The modification known as 5'-capping occurs after the pre-mRNA is about 20-30 nucleotides long. At this point the pre-RNA typically receives its first post-transcriptional modification, a 5'-cap. The "cap" is a chemical modification - a 7-methylguanosine - whose addition to the 5' end of the transcript is enzymatically catalyzed by multiple enzymes called the capping enzyme complex (CEC) a group of multiple enzymes that carry out sequential steps involved in adding the 5'-cap. The CEC binds to the RNA polymerase very early in transcription and carries out a modification of the 5' triphosphate, the subsequent transfer of at GTP to this end (connecting the two nucleotides using a unique 5'-to-5' linkage), the methylation of the newly transferred guanine, and in some transcripts the additional modifications to the first few nucleotides. This 5'-cap appears to function by protecting the emerging transcript from degradation and is quickly bound by RNA binding proteins known as the cap-binding complex (CBC). There is some evidence that this modification and the proteins bound to it play a role in targeting the transcript for export from the nucleus. Protecting the nascent RNA from degradation is not only important for conserving the energy invested in creating the transcript but is clearly involved in regulating the abundance of fully-functional transcript that is produced. Moreover, the role of the 5'-cap in guiding the transcript for export will directly help to regulate not only the amount of transcript that is made but, perhaps more importantly, the amount of transcript that is exported to the cytoplasm that has the potential to be translated.



The structure of a typical 7-methylguanylate cap. Attribution: Marc T. Facciotti (own work)

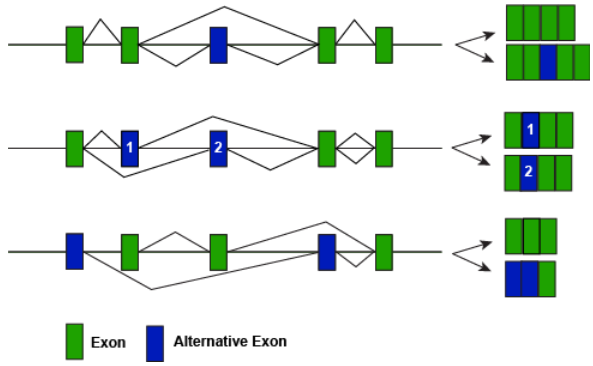
Transcript splicing

Cells must process nascent transcripts into mature RNAs by joining exons and removing the intervening introns. They accomplish this using a multicomponent complex of RNA and proteins called the spliceosome. The spliceosome complex assembles on the nascent transcript and most times the decisions about which introns to combine into a mature transcript are made at this point. How these decisions are made is still not completely understood but involves the recognition of specific DNA sequences at the splice sites by RNA and protein species and several catalytic events. It is interesting to note that the catalytic portion of the spliceosome is

made of RNA rather than protein. Recall that the ribosome is another example of a RNA-protein complex where the RNA serves as the primary catalytic component. The selection of which splice variant to make is a form of regulating gene expression. In this case rather than influencing abundance of a transcript, alternative splicing allows the cell to decide about which form of a transcript it makes.

The alternative splice forms of genes that result in protein products of related structure but of varying function are known as *isoforms*. The creation of isoforms is common in eukaryotic systems and is known to be important in different stages of development in multicellular organisms and in defining the functions of different cell types. By encoding multiple possible gene products from a single gene whose transcription initiation is encoded from a single transcriptional regulatory site (by making the decision of which end-product to produce post-transcriptionally) obviates the need to create and maintain independent copies of each gene in different parts of the genome and evolving independent regulatory sites. Therefore, the ability to form multiple isoforms from a single coding region is though to be evolutionarily advantageous because it enables some efficiency in DNA coding, minimizes transcriptional regulatory complexity, and may lower the energy burden of maintaining more DNA and protecting it from mutation. Some examples of possible outcomes of alternative splicing can include: the generation of enzyme variants with differential substrate affinity or catalytic rates; signal sequences that target proteins to various sub-cellular compartments can be changed; entirely new functions, via the swapping of protein domains can be created. These are just a few examples.

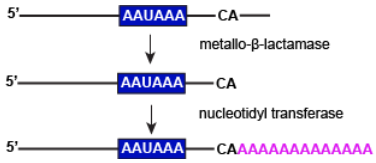
One additional interesting outcome of alternative splicing is the introduction of stop codons that can, through a mechanism that seems to require translation, lead to the targeted decay of the transcript. This means that, besides the control of transcription initiation and 5'-capping, we can also consider alternative splicing as one of the regulatory mechanisms that may influence transcript abundance. The effects of alternative splicing are therefore potentially broad - from complete loss of function to novel and diversified function to regulatory effects.



A figure depicting some different modes of alternative splicing illustrating how different splice variants can lead to different protein forms.
Attribution: Marc T. Facciotti (own work)

3' end cleavage and polyadenylation

One final modification is made to nascent pre-mRNAs before they leave the nucleus - the cleavage of the 3' end and its polyadenylation. This two step process is catalyzed by two different enzymes (as depicted below) and may decorate the 3' end of transcripts with up to nearly 200 nucleotides. This modification enhances the stability of the transcript. Generally, the more As in the polyA tag the longer lifetime that transcript has. The polyA tag also seems to play a role in the export of the transcript from the nucleus. Therefore, the 3' polyA tag plays a role in gene expression by regulating functional transcript abundance and how much is exported from the nucleus for translation.



A two step process is involved in modifying the 3' ends of transcripts prior to nuclear exports. These include cutting transcripts just downstream of a conserved sequence (AAUAAA) and transferring adenylate groups. Both processes are enzymatically catalyzed. Attribution: Marc T. Facciotti (own work)

---

microRNAs

RNA Stability and microRNAs

Besides the modifications of the pre-RNA described above and the associated proteins that bind to the nascent and transcripts, there are other factors that can influence the stability of the RNA in the cell. One example are elements called microRNAs. The microRNAs, or miRNAs, are short RNA molecules that are only 21–24 nucleotides in length. The miRNAs are transcribed in the nucleus as longer pre-miRNAs. These pre-miRNAs are subsequently chopped into mature miRNAs by a protein called dicer. These mature miRNAs recognize a specific sequence of a target RNA through complementary base pairing. miRNAs, however, also associate with a ribonucleoprotein complex called the RNA-induced silencing complex (RISC). RISC binds a target mRNA, along with the miRNA, to degrade the target mRNA. Together, miRNAs and the RISC complex rapidly destroy the RNA molecule. As one might expect, the transcription of pre-miRNAs and their subsequent processing is also tightly regulated.

Nuclear export

Nuclear export

Fully processed, mature transcripts, must be exported through the nucleus. Not surprisingly this process involves the coordination of a mature RNA species to which are bound many accessory proteins - some of which have been intimately involved in the modifications discussed above - and a protein complex called the *nuclear pore complex (NPC)*. Transport through the NPC allows flow of proteins and RNA species to move in both directions and is mediated by a number of proteins. This process can be used to selectively regulate the transport of various transcripts depending on which proteins associate with the transcript in question. This means that not all transcripts are treated equally by the NPC - depending on modification state and the proteins that have associated with a specific species of RNA it can be moved either more or less efficiently across the nuclear membrane. Since the rate of movement across the pore will influence the abundance of mature transcript that is exported into the cytosol for translation export control is another example of a step in the process of gene regulation that can be modulated. In addition, recent research has implicated interactions between the NPC and transcription factors in the regulation of transcription initiation, likely through some mechanism whereby the transcription factors tether themselves to the nuclear pores. This last example demonstrates how interconnected the regulation of gene expression is across the multiple steps of this complex process.

We know many additional details of the processes described above to some level of detail, but many more questions remain to be answered. For the sake of Bis2a it is sufficient to form a model of the steps that occur in the production of a mature transcript in eukaryotic organisms. We have painted a picture with very broad strokes, trying to present a scene that reflect what happens generally in all eukaryotes. Besides learning the key differentiating features of eukaryotic gene regulation, we would also like for Bis2a students to think of each of these steps as an opportunity for Nature to regulate gene expression in some way and to rationalize how deficiencies or changes in these pathways - potentially introduced through mutation - might influence gene expression.

While we did not explicitly bring up the Design Challenge or Energy Story here these formalisms are equally adept at helping you to make some sense of what is being described. We encourage you to try making an Energy Story for various processes. We also encourage you to use the Design Challenge rubric to reexamine the stories above: identify problems that need solving; hypothesize

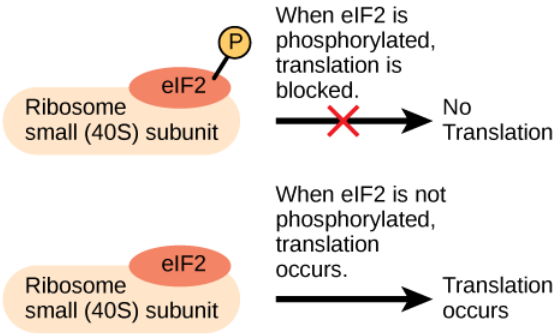
potential solutions and criteria for success. Use there formalisms to dig deeper and ask new questions/identify new problems or things that you don't know about the processes is what experts do. Chances are that doing this suggested exercise will lead you to identify a direction of research that someone has already pursued (you'll feel pretty smart about that!). Alternatively, you may raise some brand new question that no one has thought of yet.

Control of Protein Abundance

After a mRNA has been transported to the cytoplasm, it is translated into protein. Control of this process is largely dependent on the RNA molecule. As previously discussed, the stability of the RNA will have a large impact on its translation into a protein. As the stability changes, the amount of time that it is available for translation also changes.

The initiation complex and translation rate

Like transcription, translation is controlled by proteins complexes of proteins and nucleic acids that must associate to initiate the process. In translation, one of the first complexes that must assembles to start the process is referred to as the initiation complex. The first protein to bind to the mRNA that helps initiate translation is called eukaryotic initiation factor-2 (eIF-2). Activity of the eIF-2 protein is controlled by multiple factors. The first is whether or not it is bound to a molecule of GTP. When the eIF-2 is bound to GTP it is considered to be in an active form. The eIF-2 protein bound to GTP can bind to the small 40S ribosomal subunit. When bound, the eIF-2/40S ribosome complex, bringing with it the mRNA to be translated, also recruits the methionine initiator tRNA associates. At this point, when the initiator complex is assembled, the GTP is hydrolyzed into GDP creating an "inactive form of eIF-2 that is released, along with the inorganic phosphate, from the complex. This step, in turn, allows the large 60S ribosomal subunit to bind and to begin translating the RNA. The binding of eIF-2 to the RNA further controlled by protein phosphorylation. When eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP thus inhibiting the initiation complex from forming - translation is therefore inhibited (see the figure below). In the dephosphorylated state eIF-2 can bind GTP and allow the assembly of the translation initiation complex as described above. The ability of the cell therefore to tune the assembly of the translation invitation complex via a reversible chemical modification (phosphorylation) to a regulatory protein is another example of how Nature has taken advantage of even this seemingly simple step to tuned gene expression.



An increase in phosphorylation levels of eIF-2 has been observed in patients with neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's. What impact do you think this might have on protein synthesis?

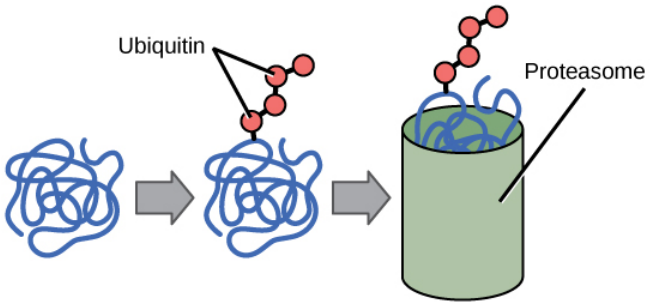
Chemical Modifications, Protein Activity, and Longevity

Not to be outdone by nucleic acids, proteins can also be chemically modified with the addition of groups including methyl, phosphate, acetyl, and ubiquitin groups. The addition or removal of these groups from proteins can regulate their activity or the length of time they exist in the cell. Sometimes these modifications can regulate where a protein is found in the cell—for example, in the nucleus, the cytoplasm, or attached to the plasma membrane.

Chemical modifications can occur in response to external stimuli such as stress, the lack of nutrients, heat, or ultraviolet light exposure. In addition to regulating the function of the proteins themselves, if these changes occur on specific proteins they can alter epigenetic accessibility (in the case of histone modification), transcription (transcription factors), mRNA stability (RNA binding

proteins), or translation (eIF-2) thus feeding back and regulating various parts of the process of gene expression. In the case of modification to regulatory proteins, this can be an efficient way for the cell to rapidly change the levels of specific proteins in response to the environment by regulating various steps in the process.

The addition of an ubiquitin group has another function - it marks that protein for degradation. Ubiquitin is a small molecule that acts like a flag indicating that the tagged proteins should be targeted to an organelle called the proteasome. This organelle is a large multi-protein complex that functions to cleave proteins into smaller pieces that can then be recycled. Ubiquitination (the addition of a ubiquitin tag), therefore helps to control gene expression by altering the functional lifetime of the protein product.



Proteins with ubiquitin tags are marked for degradation within the proteasome.

In conclusion, we see that gene regulation is complex and that it can be modulated at each step in the process of expressing a functional gene product. Moreover, the regulatory elements that happen at each step can act to influence other regulatory steps both earlier and later in the process of gene expression (i.e. the process of chemically altering a transcription factor can influence the regulation of its own transcription many steps earlier in the process). These complex sets of interactions form what are known as gene regulatory networks. Understanding the structure and dynamics of these networks is critical for understanding how different cells function, the basis for numerous diseases, developmental processes, and how cells make decisions about how to react to the many factors that are in constant flux both inside and outside.

Reading_23 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_24

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_24

GI.31 Define, describe and correctly use the terms: point mutation, deletion, insertion, nonsense, frameshift, null, loss-of-function and gain-of-function. Be able to describe their possible impact(s) on protein integrity.
GI.33 Use a codon table and your knowledge of protein structure and function to make predictions of how changes at the DNA level might influence protein structure and function.
GI.34 Explain the possible different mechanisms by which mutations can cause changes in phenotype. Include mutations to both protein-coding regions and non protein-coding regions in your discussion.
GI.35 Explain the potential influence of mutations on the specificity and affinity of protein-DNA interactions and the potential impact of these mutations on gene expression.
GI.53 Define and correctly use the terms "haploid", "diploid", and "polyploid" and describe some of the costs and benefits of polyploidy.
GI.54 Explain how multiple genes may interact to influence phenotype and apply this concept to interpret or propose a molecular mechanism that explains the phenotypes.
GI.55 Explain how multiple alleles for a single gene, when encoded in a genome, can result in multiple distinguishable phenotypes. Propose or interpret a molecular mechanism that explains these multiple phenotypes.
GI.56 Define and be able to correctly use the terms heterozygous, homozygous, mutant, wild type, dominant, recessive, allele, gene, loci, and chromosomes.
GI.57 Predict a phenotype(s) if given a genotype (for each haploid, diploid and polyploid) and related molecular mechanisms, and vice versa.

Mutations



Errors occurring during DNA replication are not the only way by which mutations can arise in DNA. Mutations, variations in the nucleotide sequence of a genome, can also occur because of physical damage to DNA. Such mutations may be of two types: induced or spontaneous. **Induced mutations** are those that result from an exposure to chemicals, UV rays, x-rays, or some other environmental agent. **Spontaneous mutations** occur with no exposure to any environmental agent; they result from spontaneous biochemical reactions taking place within the cell.

Mutations may have a wide range of effects. Some mutations create no observable impact; we call these **silent mutations**. **Point mutations** are those mutations that affect a single base pair. The most common nucleotide mutations are substitutions, in which one base is replaced by another. These can be of two types, either transitions or transversions. **Transition substitution** refers to a purine or pyrimidine being replaced by a base of the same kind; for example, a purine such as adenine may be replaced by the purine guanine. **Transversion substitution** refers to a purine being replaced by a pyrimidine, or vice versa; for example, cytosine, a pyrimidine, is replaced by adenine, a purine. Mutations can also be the result of the addition of a nucleotide, known as an insertion, or the removal of a base, also known as deletion. Sometimes a piece of DNA from one chromosome may get translocated to another chromosome or to another region of the same chromosome; this is known as translocation.

As we will visit later, when a mutation occurs in a protein coding region, it may have several effects. Transition or transversion mutants may lead to no change in the protein sequence (known as **silent mutations**), change the amino acid sequence (known as **missense mutations**), or create what is known as a stop codon (known as a **nonsense mutation**). Insertions and deletions in protein coding sequences lead to **frameshift mutations**. Missense mutations that lead to **conservative changes** results in the substitution of similar but not identical amino acids. For example, the acidic amino acid glutamate being substituted for the acidic amino acid aspartate would be considered conservative. We do not expect these types of missense mutations to be as severe as a **non-conservative** amino acid change; such as a glutamate substituted for a valine. Drawing from our understanding of functional group chemistry, we can correctly infer that this type of substitution may lead to severe functional consequences, depending upon location of the mutation.

Note: Vocabulary Watch

Note that the preceding paragraph had a lot of potentially new vocabulary - it would be a good idea to learn these terms.

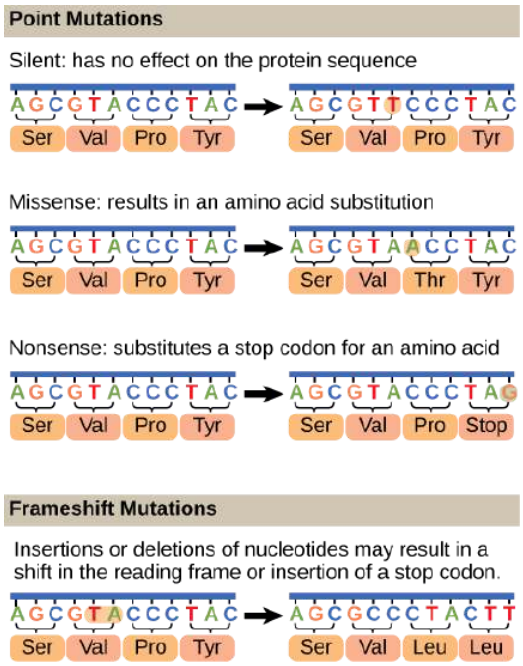


Figure 1. *Mutations can lead to changes in the protein sequence encoded by the DNA.*

Mutations: Some nomenclature and considerations

Mutation

Etymologically, the term mutation means a change or alteration. In genetics, a mutation is a change in the genetic material - DNA sequence - of an organism. By extension, a mutant is the organism in which a mutation has occurred. But what is the change compared to? The answer to this question is that it depends. We can make the comparison with the direct progenitor (cell or organism) or to patterns seen in a population of the organism in question. It mostly depends on the specific context of the discussion. Since genetic studies often look at a population (or key subpopulations) of individuals, we begin by describing the term "wild-type".

Wild Type vs Mutant

What do we mean by "wild type"? Since the definition can depend on context, this concept is not entirely straightforward. Here are a few examples of definitions you may run into:

Possible meanings of "wild-type"

1. An organism having an appearance characteristic of the species in a natural breeding population (i.e. a cheetah's spots and tear-like dark streaks that extend from the eyes to the mouth).
2. The form or forms of a gene most commonly occurring in nature in a given species.
3. A phenotype, genotype, or gene that predominates in a natural population of organisms or strain of organisms in contrast to that of natural or laboratory mutant forms.
4. The normal, as opposed to the mutant gene or allele.

The common thread to all of the definitions listed above is based on the "norm" for a set of characteristics with respect to a specific trait compared to the overall population. In the "Pre-DNA sequencing Age" species were classified based on common phenotypes (what they looked like, where they lived, how they behaved, etc.). A "norm" was established for the species in question. For

example, Crows display a common set of characteristics, they are large, black birds that live in specific regions, eat certain types of food and behave in a certain characteristic way. If we see one, we know its a crow based on these characteristics. If we saw one with a white head, we would think that either it is a different bird (not a crow) or a mutant, a crow that has some alteration from the norm or wild type.

In this class we take what is common about those varying definitions and adopt the idea that "wild type" is simply a reference standard against which we can compare members of a population.

Possible NB Discussion Point

If you were assigning wild type traits to describe a dog, what would they be? What is the difference between a mutant trait and variation of a trait in a population of dogs? Is there a wild type for a dog that we could use as a standard? How would we begin to think about this concept with respect to dogs?



Figure 2. *Mutations can lead to changes in the protein sequence encoded by the DNA that then impact the outward appearance of the organism.*
(Source)

Mutations are simply changes from the "wild type", reference or parental sequence for an organism. While the term "mutation" has colloquially negative connotations, we must remember that change is neither inherently "bad". Mutations (changes in sequences) should not primarily be thought of as "bad" or "good", but rather simply as changes and a source of genetic and phenotypic diversity on which evolution by natural selection can occur. Natural selection ultimately determines the long-term fate of mutations. If the mutation confers a selective advantage to the organism, the mutation will be selected and may eventually become very common in the population. Conversely, if the mutation is deleterious, natural selection will ensure that the mutation will be lost from the population. If the mutation is neutral, that is it neither provides a selective advantage or disadvantage, then it may

persist in the population. Different forms of a gene, including those associated with "wild type" and respective mutants, in a population are termed **alleles**.

Consequences of Mutations

For an individual, the consequence of mutations may mean little or it may mean life or death. Some deleterious mutations are **null** or **knock-out** mutations which result in a loss of function of the gene product. These mutations can arise by a deletion of the either the entire gene, a portion of the gene, or by a point mutation in a critical region of the gene that renders the gene product non-functional. These types of mutations are also referred to as **loss-of-function** mutations. Alternatively, mutations may lead to a modification of an existing function (i.e. the mutation may change the catalytic efficiency of an enzyme, a change in substrate specificity, or a change in structure). In rare cases a mutation may create a new or enhanced function for a gene product; this is often referred to as a **gain-of-function** mutation. Lastly, mutations may occur in non-coding regions of DNA. These mutations can have a variety of outcomes including altered regulation of gene expression, changes in replication rates or structural properties of DNA and other non-protein associated factors.

Mutations and cancer

Mutations can affect either somatic cells or germ cells. Sometimes mutations occur in DNA repair genes, in effect compromising the cell's ability to fix other mutations that may arise. If, as a result of mutations in DNA repair genes, many mutations accumulate in a somatic cell, they may lead to problems such as the uncontrolled cell division observed in cancer. Cancers, including forms of pancreatic cancer, colon cancer, and colorectal cancer have been associated with mutations like these in DNA repair genes. If, by contrast, a mutation in DNA repair occurs in germ cells (sex cells), the mutation will be passed on to the next generation, as in the case of diseases like hemophilia and xeroderma pigmentosa. In the case of xeroderma pigmentoas individuals with compromised DNA repair processes become very sensitive to UV radiation. In severe cases these individuals may get severe sun burns with just minutes of exposure to the sun. Nearly half of all children with this condition develop their first skin cancers by age 10.

Consequences of errors in replication, transcription and translation

Something key to think about:

Cells have evolved a variety of ways to make sure DNA errors are both detected and corrected, rom proof reading by the various DNA-dependent DNA polymerases, to more complex repair systems. Why did so many mechanisms evolve to repair errors in DNA? By contrast, similar proof-reading mechanisms did NOT evolve for errors in transcription or translation. Why might this be? What would be the consequences of an error in transcription? Would such an error effect the offspring? Would it be lethal to the cell? What about translation? Ask the same questions about the process of translation. What would happen if the wrong amino acid was accidentally put into the growing polypeptide during the translation of a protein? Contrast this with DNA replication.

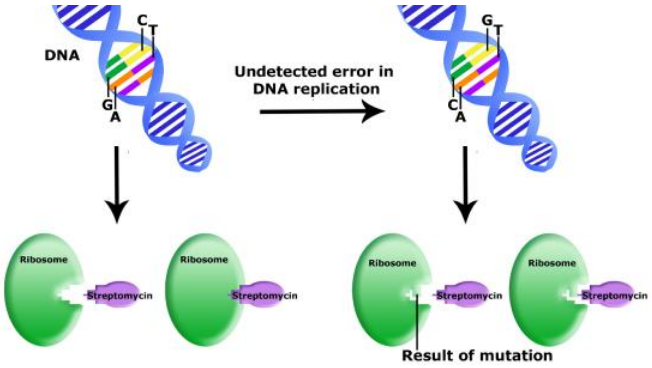
Mutations as instruments of change

Mutations are how populations can adapt to changing environmental pressures

Mutations are randomly created in the genome of every organism, and this in turn creates genetic diversity and a plethora of different alleles per gene per organism in every population on the planet. If mutations did not occur, and chromosomes were replicated and transmitted with 100% fidelity, how would cells and organisms adapt? Whether evolution keeps mutations in a population depends on whether the mutation provides selective advantage, poses some selective cost or is at the very least, not harmful. Indeed, mutations that appear neutral may persist in the population for many generations and only be meaningful when a population is challenged with a new environmental challenge. At this point, the apparently previously neutral mutations may provide a selective advantage.

Example: Antibiotic resistance

The bacterium *E. coli* is sensitive to an antibiotic called streptomycin, which inhibits protein synthesis by binding to the ribosome. The ribosomal protein L12 can be mutated such that streptomycin no longer binds to the ribosome and inhibits protein synthesis. Wild type and L12 mutants grow equally well and the mutation appears to be neutral in the absence of the antibiotic. In the presence of the antibiotic wild type cells die and L12 mutants survive. This example shows how genetic diversity is important for the population to survive. If mutations did not randomly occur, when the population is challenged by an environmental event, such as the exposure to streptomycin, the entire population would die. For most populations this becomes a numbers game. If the mutation rate is 10^{-6} then a population of 10^{7} cells would have 10 mutants; a population of 10^{8} would have 100 mutants, etc.



Uncorrected errors in DNA replication lead to mutation. In this example, an uncorrected error was passed onto a bacterial daughter cell. This error is in a gene that encodes for part of the ribosome. The mutation results in a different final 3D structure of the ribosome protein. While the wildtype ribosome can bind to streptomycin (an antibiotic that will kill the bacterial cell by inhibiting the ribosome function) the mutant ribosome cannot bind to streptomycin. This bacteria is now resistant to streptomycin. Source: Bis2A Team original image

An example: Lactate dehydrogenase

Lactate Dehydrogenase (LDH), the enzyme that catalyzes the reduction of pyruvate into lactic acid in fermentation, while virtually every organism has this activity, the corresponding enzyme and therefore gene differs immensely between humans and bacteria. The proteins are clearly related, they perform the same basic function but have a variety of differences, from substrate binding affinities and reaction rates to optimal salt and pH requirements. Each of these attributes have been evolutionarily tuned for each specific organism through multiple rounds of mutation and selection.

Possible NB Discussion Point

We can use comparative DNA sequence analysis to generate hypotheses about the evolutionary relationships between three or more organisms. One way to accomplish this is to compare the DNA or protein sequences of proteins found in each of the organisms we wish to compare. Let us, for example, imagine that we were to compare the sequences of lactate dehydrogenase (LDH) from three different organisms. The schematic below depicts the primary structures of LDH proteins from Organisms A, B, and C. The letters in the center of the proteins' line diagram represent amino acids at a unique position and the proposed differences in each sequences (Attribution: Marc T. Facciotti [original work]). **Question***:* Is Organism C more closely related to Organism A or B? The simplest explanation is that Organism A is the earliest form, a mutation occurred giving rise to Organism B. Over time a second mutation arose in the B lineage to give rise to the enzyme found in Organism C. This is the simplest explanation, however we cannot rule out other possibilities. Can you think of other ways the different forms of the LDH enzyme arose these three organisms?

H_2N—H—L—COOH Organism A

H_2N—A—L—COOH Organism B

H_2N—A—V—COOH Organism C

Real-life Application:

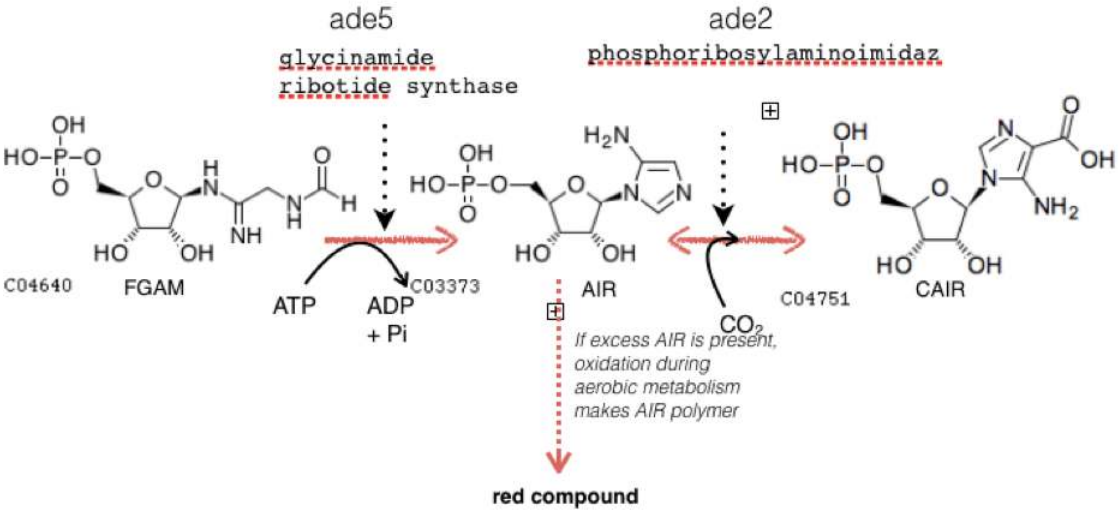
As we have seen in the "Mutations and Mutants" module, changing even one nucleotide can have major effects on the translated product. Read more about an undergraduate's work on point mutations and GMOs here.

POST PRACTICE GUIDE

General Practice

3. Why: Reading a figure that depicts steps in a metabolic pathway is a key skill that you need in future courses on cellular metabolism and anywhere else you might encounter metabolic pathways. Here we use and practice that skill - and thrown in a little wrinkle of thinking about the genes that encode the enzymes in the pathway while linking our hard-earned lessons on metabolism, enzymes, learning to read pathways, getting comfortable with molecular structures and the funny names of enzymes and compounds etc. to phenotype.

How to practice: Look at the figure below: This is the same figure we discussed in class with a couple of additions. Create and "energy story" for these reactions.



This exercise helps you practice learning objectives: This is "reloading" learning objectives from our section on metabolism that expect you to read a metabolic pathway. Here we add one more layer, which is to recognize the gene name that encodes the enzyme catalyzing a step in the pathway (in this figure ade5 and ade2) and to add this into your energy story.

4. Why: One of the big pictures objectives we instructors have for you is to start being able to tie together concepts and ideas from across the course. In this case we want you to be able to talk about the molecular (in this case can be gene and enzyme level) details that can explain a simple phenotype of an organism.

How to practice: Use the figure above to explain the molecular basis for the phenotype associated with a haploid genotype in which the *ade2* gene has been deleted.

This exercise helps you practice learning objectives: ***GI.57*** *Predict a phenotype(s) if given a genotype (for each haploid, diploid and polyploid) and related molecular mechanisms, and vice versa.*

5. Why: Mutations are the engine of organismal diversity required for natural selection. Here we want you to link the idea of mutation to different phenotype and to start linking the idea of a change in the genome and how that can mechanistically cause differences in phenotype. We want you to practice the ideas above but now get you to think about how mutations factor in to the story of the genotype-phenotype relationship.

How to practice: Explain the possible outcomes of mutations (nucleotide changes) in the region of DNA encoding the *ade2* gene. What are the possible effects on the protein encoded by the gene? Can you propose a hypothetical scenario that might explain HOW a change in the *ade2* gene leads to an amino-acid-level change in the protein that is responsible for a "novel" phenotype. (suggestion: think about the previous questions probing molecular interactions with the ribosome).

This exercise helps you practice learning objectives: ***GI.34*** *Explain the possible different mechanisms by which mutations can cause changes in phenotype. Include mutations to both protein-coding regions and non protein-coding regions in your discussion.;* ***GI.57*** *Predict a phenotype(s) if given a genotype (for each haploid, diploid and polyploid) and related molecular mechanisms, and vice versa.;* ***GI.54*** *Explain how multiple genes may interact to influence phenotype and apply this concept to interpret or propose a molecular mechanism that explains the phenotypes.;* ***GI.55*** *Explain how multiple alleles for a single gene, when encoded in a genome, can result in multiple distinguishable phenotypes. Propose or interpret a molecular mechanism that explains these multiple phenotypes.*

6. Why: What happens to phenotype when you can have multiple copies of a gene in a genome and those copies are different versions (mutants/alleles) of one another? In Bis2a we want to you start thinking about this with the very simplest starting assumptions (which you will see in later classes do not hold true in all cases). This exercise is basically asking you to take the concepts/skills you worked on in exercises 4 and 5 and to combine them in a scenario where multiple variants of a gene can co-exist in the same cell.

How to practice: Imagine the following partially hypothetical scenario: A typical sized yeast cell (haploid or diploid) requires a minimum of 100 units of enzyme activity from the protein encoded by the *ade2* gene in order to keep a white phenotype when the purine biosynthetic pathway is operating. The wild-type *ade2* allele produces 100 units of activity. A haploid yeast expresses an allele of *ade2* (we'll call it *ade2-a*) that encodes an enzyme that provides 50 units of activity. Create figures/cartoons to help you keep track of what is going on!!

Predict the phenotype and explain the molecular details of:

(a) a haploid cell expressing a single copy of *ade2-a*

(b) a haploid cell expressing a single copy of the *ade2* gene

(c) a heterozygous diploid cell expressing a single copy of *ade2-a* allele and a deletion of the wild-type *ade2* allele gene on the homologous chromosome

(d) a heterozygous diploid cell expressing a single copy of the *ade2* allele and and a deletion of the wild-type *ade2* allele gene on the homologous chromosome

(e) a homozygous diploid cell expressing two copies of the *ade2-a* allele

This exercise helps you practice learning objectives: ***GI.56*** *Define and be able to correctly use the terms heterozygous, homozygous, mutant, wild type, dominant, recessive, allele, gene, loci, and chromosomes.;* ***GI.53*** *Define and correctly use the terms "haploid", "diploid", and "polyploid" and describe some of the costs and benefits of polyploidy.;* ***GI.57*** *Predict a phenotype(s) if given a genotype (for each haploid, diploid and polyploid) and related molecular mechanisms, and vice versa.;*

GI.54 *Explain how multiple genes may interact to influence phenotype and apply this concept to interpret or propose a molecular mechanism that explains the phenotypes.*

7. Why: Here we want you to think about the same concepts as in exercise 6 but with a different example.

How to practice: The allele pool in a population can be large. Using the same logic as in exercise 6 propose a reasonable hypothesis for the phenotype of the following:

(a) haploid cells encoding an *ade-z* allele (the *ade-z* allele encodes and enzyme that produces 20 units of activity).

(b) haploid cells encoding an *ade-x* allele (the *ade-x* allele encodes and enzyme that produces 40 units of activity).

(c) a heterozygous diploid expressing the *ade-z* and *ade-x* alleles.

(d) a homozygous dipoid expressing the *ade-z* allele

(e) a homozygous dipoid expressing the *ade-x* allel

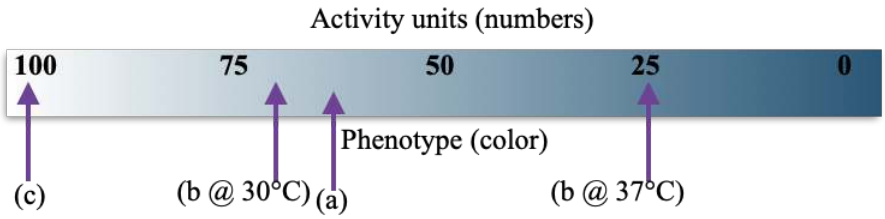
This exercise helps you practice learning objectives: ***GI.56*** *Define and be able to correctly use the terms heterozygous, homozygous, mutant, wild type, dominant, recessive, allele, gene, loci, and chromosomes.;* ***GI.53*** *Define and correctly use the terms "haploid", "diploid", and "polyploid" and describe some of the costs and benefits of polyploidy.;* ***GI.57*** *Predict a phenotype(s) if given a genotype (for each haploid, diploid and polyploid) and related molecular mechanisms, and vice versa.;* ***GI.54*** *Explain how multiple genes may interact to influence phenotype and apply this concept to interpret or propose a molecular mechanism that explains the phenotypes.*

8. Why: This is related to the questions in class related to the alleles of the *blue6* gene. More practice on genotype-to-phenotype relationship from a different angle. Same skills/different look.

How to practice: Given the following allele pool, propose a reasonable hypothesis for the following scenarios. You may consider various factors including ploidy.

Allele	Activity at 30°C
blue6-1	25
blue6-2	45
blue6-3	100
blue6-4	65
blue6-5	80
blue6-6	20

Propose hypotheses that explain phenotype (a). Propose one or more hypotheses that explain the observations about the two phenotypes for (b). Propose hypotheses that explain phenotype (c).

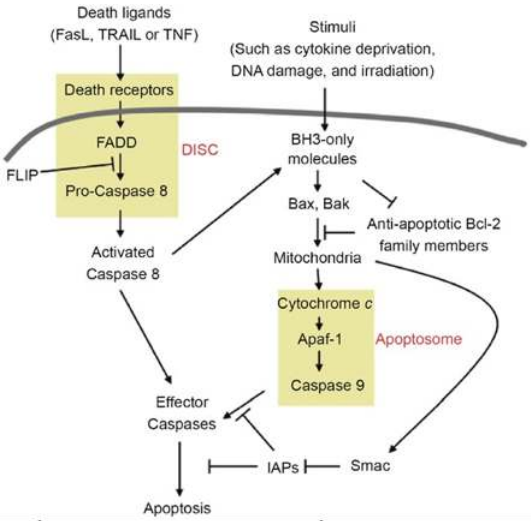


The Blue6 and the Ade2 gene examples have been idealized a little bit for the convenience of talking about interactions between different alleles for the same locus. What are some key assumptions that we make when we treat the total activity of multiple alleles as simply additive?

This exercise helps you practice learning objectives: ***GI.56*** *Define and be able to correctly use the terms heterozygous, homozygous, mutant, wild type, dominant, recessive, allele, gene, loci, and chromosomes.;* ***GI.53*** *Define and correctly use the terms "haploid", "diploid", and "polyploid" and describe some of the costs and benefits of polyploidy.;* ***GI.57*** *Predict a phenotype(s) if given a genotype (for each haploid, diploid and polyploid) and related molecular mechanisms, and vice versa.;* ***GI.54*** *Explain how multiple genes may interact to influence phenotype and apply this concept to interpret or propose a molecular mechanism that explains the phenotypes.*

9. Why: As we have seen in several cases many processes in biological systems are interconnected in various ways and all of these interconnected processes are regulated either positively or negatively at various points. The interconnections between functions, pathways, compounds, you name it in the cell are often drawn in network diagram form where positive and negative regulation is depicted by special symbols. Here we ask you to practice reading these kinds of figures.

How to practice: There are common symbols in regulatory pathways to indicate positive and negative regulatory circuits. Identify some of the positive and negative circuits in the image below. Then, select a specific protein and pretend that a mutation has been generated in the gene encoding it. What would be the impact of the mutation on the phenotype (apoptosis)?



(http://www.nature.com/cr/journal/v17/n9/images/cr200752f1.jpg).

This exercise helps you practice learning objectives: ***GI.47*** *Using a regulatory network, accurately interpret relationships between genes/proteins in the network and predict potential targets of regulation by different possible types of signaling molecules.;* ***GI.44*** *Given some specific constraints on a regulatory circuit, predict the behavior of the circuit under the possible influence of environmental or genetic perturbations (changes).*

10. Why: Translation is a core process in all cells. We practiced some skills related to this process in previous study guides but want to keep exercising the skills associated with the reading of DNA sequences and their translation. Keeping it fresh on the mind!

How to practice: Consider the following hypothetical sequence:

5'CGCTACGTCTTACGCTGGAGCTCTCATGGATCGGTTCGGTAGGGCTCGATCACATCGCTAGCCAT 3'

3'GCGATGCAGAATGCGACCTCGAGAGTACCTAGCCAAGCCATCCCGAGCTAGTGTAGCGATCGGTA 5'

a) How many potential open reading frames (pairs of start and stop codons) are present in this DNA fragment (you may need to look up a codon table on-line)?

b) Look at only the top strand. Identify the location(s) of the start codons and the stop codons. Which frame could encode a potential protein?

c) Look at only the bottom strand. Identify the location(s) of the start codons and the stop codons. Which frame could encode a potential protein? (Remember that mRNAs are read by the ribosome 5' to 3.')

d) Take the ORFs identified in (b) and (c) and use this information to obtain a possible protein sequence. Which ORF makes the longest protein?

This exercise helps you practice learning objectives: This is an exercise to help you keep practicing learning objectives from the translation section.

Practice Exam Questions

Question Q24.1.

Q24.1 In the reaction depicted in the figure below the most likely phenotype of a Δ*ade5* genotype expressed in a haploid yeast is:

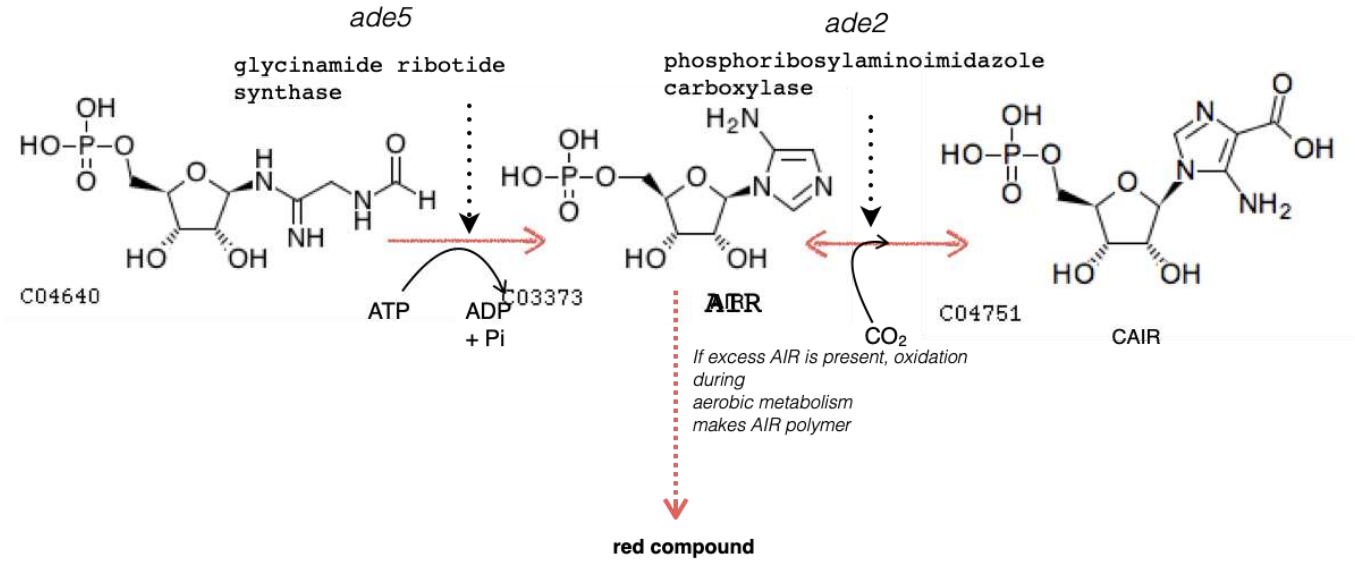
A. cells turn red

B. cells will be cream colored

C. cells will be pink

D. cells will turn blue

E. cells will turn purple



Question Q24.2.

Q24.2. The figure to the right diagrams a metabolic pathway that might controls the color of flowers in certain plants: The genotype of a diploid wild-type flower is AA BB CC DD, and its phenotype is light purple. Alleles A, B, C, D each encoded enzymes that catalyze the eponymous reactions in the Figure. a, b, c, and d are loss-of-function alleles. What will the mostly likely phenotype of the AA Bb cc DD double mutant?

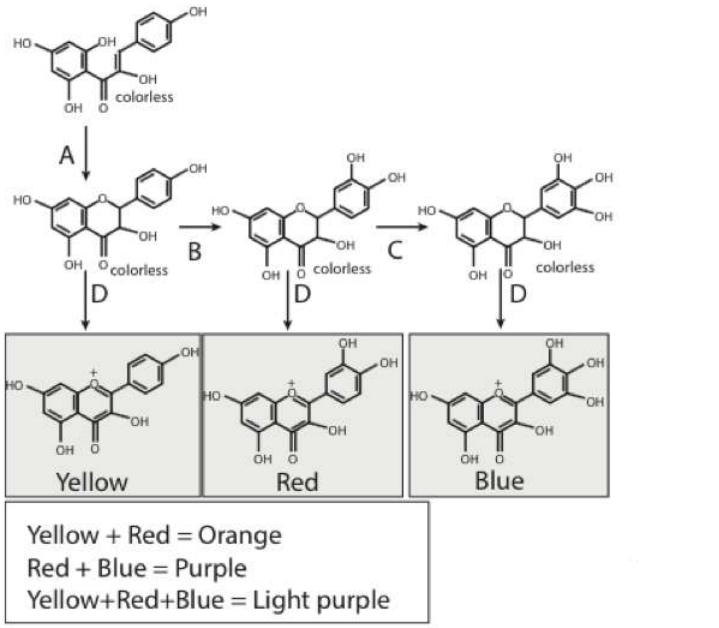
A. Purple

B. White

C. Orange

D. Red

E. Yellow



Question Q24.3. Phenotype/genotype with alleles and chromosomal location.

Q24.3. Within a group of mutants (genetically related organisms with slightly different genotypes) with the same overt phenotype, DNA sequencing studies mapping the location of the mutations determined that mutations in different organisms were on different chromosomes. This indicates that

A. the same gene governs all the steps in a particular biological pathway.

B. different genes can govern different individual steps in the same biological pathway.

C. different genes govern the same step in a particular biological pathway.

D. all biological pathways are governed by different genes.

E. genes do not govern steps in biological pathways.

Question Q24.4. A bit of practice reading and interpreting networks.

Q24.4 In the network below we see the functional relationships network of transcription factors, the genes encoding them, environmental factors and the genes responsible for a complex bacterial phenotype called heterocyst formation (*hgl, hep, and nif)*. Which interpretation of the figure is most reasonable?

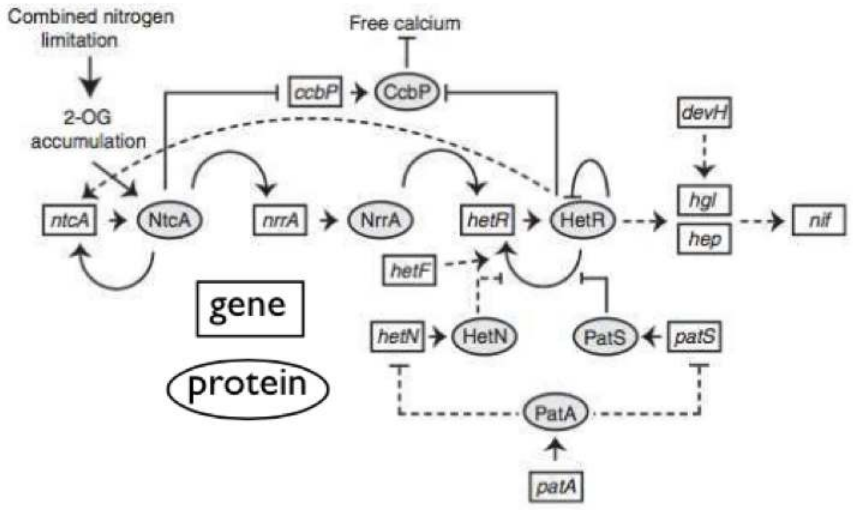
A. Calcium negatively regulates CcbP.

B. The expression of gene *patS* leads to an increase in the expression of *hgl.*

C. Nitrogen limitation “turns-on” *hgl* expression.

D.The NtcA transcription factor can act both negatively and positively.

E. C & D



Question Q24.5. Relatively easy question if you understand the definition of allele and diploid genomes.

Q24.5. Within the human population, 200 different alleles exist for a specific locus. Each person carries at most _____ alleles for this locus.

A. 1

B. 2

C. 100

D. 200

E. 400

GLOSSARY

mutation that results from exposure to chemicals or environmental agents

variation in the nucleotide sequence of a genome

type of repair mechanism in which mismatched bases are removed after replication

type of DNA repair mechanism in which the wrong base, along with a few nucleotides upstream or downstream, are removed

function of DNA pol in which it reads the newly added base before adding the next one

mutation that affects a single base

mutation that is not expressed

mutation that takes place in the cells as a result of chemical reactions taking place naturally without exposure to any external agent

when a purine is replaced with a purine or a pyrimidine is replaced with another pyrimidine

when a purine is replaced by a pyrimidine or a pyrimidine is replaced by a purine

induced mutation:

mutation:
mismatch repair:
nucleotide excision repair:
proofreading:
point mutation:
silent mutation:
spontaneous mutation:
transition substitution:
transversion substitution:

Reading_24 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_25

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_25

CD.1 Identify the signals responsible for entering the different phases of the cell cycle, what happens if these are disrupted, and how different modes of regulation could be used.
CD.2 List the requirements for cell division and how that relates to the different phases of the cell cycle.
CD.3 Compare and contrast the sequence of events that need to occur during mitosis versus meiosis and why they are necessary: include the roles of microtubules, motor proteins, centrosomes, and the level of DNA condensation.
CD.4 Compare and contrast the behaviors of sister chromatids, chromosomes, and homologous chromosomes in the context of mitosis versus meiosis.
CD.41 Correctly use the terms "chromosome", "homologous chromosome", "chromatid" and "sister chromatid" to describe their common representations in figures of cell division.

Introduction to cell division

An evolutionary goal of all living systems is to reproduce. Since the basic unit of life is a cell, and we know - thanks at least in part to Francesco Reid - that life begets new life - this means that there must be a process by which to create new cells from parental cells. We also know intuitively that multicellular organisms must somehow increase their number of cells during their growth by creating copies of existing cells. The process by which one cell creates one or more new cells, for both single and multi-celled organisms, requires a parental cell to divide and is called **cell division**.

From the standpoint of the Design Challenge framework we can stipulate that the big problem of cell division is to make a copy of a cell. If a condition for success requires that the daughter cells be viable then a number of subproblems can be defined:

1. The cell must replicate its DNA so that at least two cells have a functional copy after cell division is complete - we have discussed this process already.
2. The cell must make sufficient copies of the rest of the cellular content so that daughter cells are viable or it must find a way to ensure that the copied DNA (even without a full replica of cellular content) is viable.
3. The cell must divide the replicated cell content and DNA between at least two independently bounded compartments.
4. To ensure success, the process must be happen in an evolutionarily competitive time and be accomplished with an evolutionary selection-friendly amount of biochemical resources.

While it is not a strict requirement that this process happen in a coordinated manner, Nature has selected for systems in which all of the steps in the process happen in a highly coordinated way. This helps cells meet requirement number 4 in the list above. The coordinated process and the mechanisms of control are generally referred to as the **cell cycle**. This term can be used to describe the coordinate process used by any cell that is undergoing cell division. When we observe Nature we find that it has evolved two major modes of reproduction: sexual and asexual. Within each of these modes of reproduction we find several major modes of cell division that occur frequently across all domains of life. We consider three of these modes: binary fission (used primarily by single celled bacteria and archaea), mitosis (used often by eukaryotes in processes of cell division NOT associated with sexual reproduction) and meiosis (a process of cell division tightly linked to sexual reproduction). We discuss these processes in the sections that follow.



Figure. First-year UC Davis students enrolled in a hands-on first-year seminar course examine an agar plate of on which they have "painted" designs with an engineered strain of *Escherschia coli.* The designs only become apparent after the bacteria multiply through the process of binary fission. Opportunities for getting involved in hands-on research abound on the UC Davis campus - make sure to set time aside to get involved before you graduate. (Photo: by undergraduate student Daniel Oberbauer - 2017)

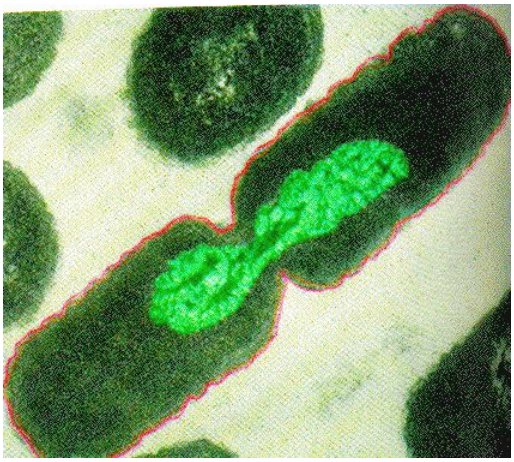
Cell division in the bacteria and archaea



Bacteria and Archaea

Like all other life forms, bacteria and archaea have one key evolutionary driver: to make more of themselves. Typically, bacterial and archaeal cells grow, duplicate all major cellular constituents, like DNA, ribosomes, etc., distribute this content and then divide into two nearly identical daughter cells. This process is called **binary fission** and is shown mid-process in the figure below. While some bacterial species are known to use several alternative reproductive strategies including making multiple offspring or budding - and all alternative mechanisms still meet the requirements for cell division stipulated above - binary fission is the most commonly laboratory-observed mechanisms for cell division the bacteria and archaea so we limit our discussion to this mechanisms alone.

(Aside: Those who want to read more about alternatives to binary fission in bacteria should check this link out.)



Binary Fission in bacteria starts with DNA replication at the replication origin attached to the cell wall, near the midpoint of the cell. New replication forks can form before the first cell division ends; this phenomenon allows an extremely rapid rate of reproduction. Source: http://biology.kenyon.edu/courses/bi...01/week01.html

---

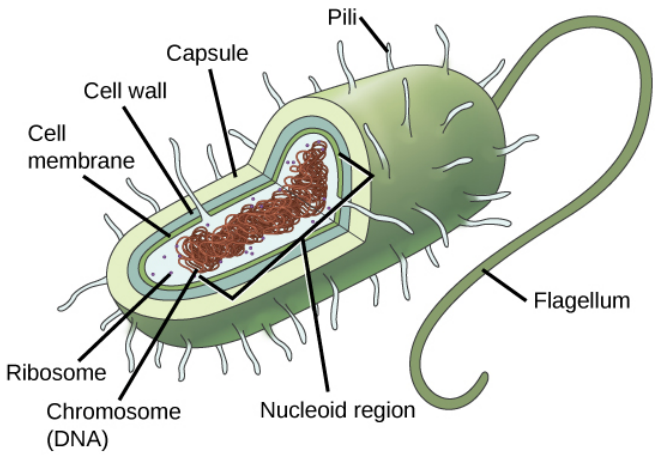
Binary Fission

The process of binary fission is the most commonly observed mechanism for cell division in bacteria and archaea (at least the culturable ones studied in the laboratory). The following is a description of a process that happens in some rod-shaped bacteria:

Since we must consider the replication of DNA one structural feature of relevance to DNA replication in the bacteria and archaea is that their genetic material is not enclosed in a nucleus, but instead occupies a specific location, the nucleoid, within the cell. Moreover, the DNA of the nucleoid is associated with numerous proteins that aid in compacting the DNA into a smaller, organized structure. Another organizational feature to note is that the bacterial chromosome is typically attached to the plasma membrane at about the midpoint of the cell. The starting point of replication, the **origin**, is close to this attachment site. Recall also that replication of the DNA is bidirectional, with replication forks moving away from the origin on both strands of the loop simultaneously. Due to the structural arrangement of the DNA at the midpoint this means that as the new double strands are formed, each origin point moves away from the cell wall attachment toward the opposite ends of the cell.

This process of DNA replication is typically occurring at the same time as a growth in the physical dimensions of the cell. Therefore, as the cell elongates, the growing membrane aids in the transport of the chromosomes towards the two opposite poles of the cells. After the chromosomes have cleared the midpoint of the elongated cell, cytoplasmic separation begins.

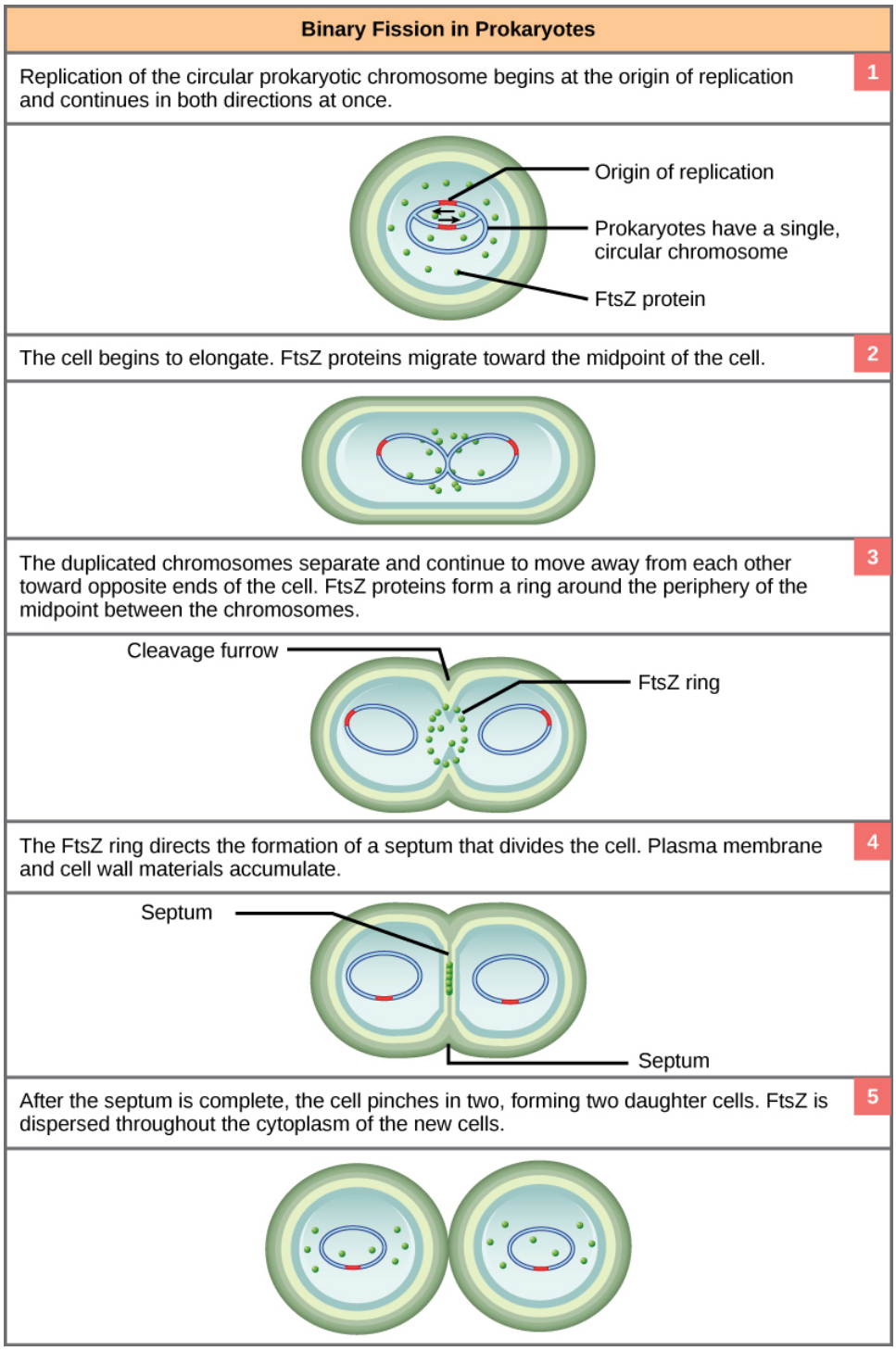
The formation of a ring composed of repeating units of a protein called **FtsZ** (a cytoskeletal protein) directs the formation of a partition between the two new nucleoids. Formation of the FtsZ ring triggers the accumulation of other proteins that work together to recruit new membrane and cell wall materials to the site. Gradually, a **septum** is formed between the nucleoids, extending from the periphery toward the center of the cell. When the new cell walls are in place, the daughter cells separate.



Prokaryotes, including bacteria and archaea, have a single, circular chromosome located in a central region called the nucleoid.

Possible discussion

How does attaching the replicating chromosome to the cell membrane aid in dividing the two chromosomes after replication is complete?



These images show the steps of binary fission in prokaryotes. (credit: modification of work by "Mcstrother"/Wikimedia Commons)

Control of these processes

Not surprisingly, the process of binary fission is strictly controlled in most bacteria and archaea. Somewhat surprisingly, however, while some key molecular players are know, much remains to be discovered and understood about how decisions are made to coordinate the activities.

Eukaryotic Cell Cycle and Mitosis

The cell cycle is an orderly sequence of events used by biological systems to coordinate cell division. In eukaryotes, asexual cell division proceeds via a cell cycle that includes multiple spatially and temporally coordinated events. These include a long preparatory period called **interphase**, and a **mitotic** phase called M phase. Interphase is often further divided into distinguishable subphases called G_1, S, and G_2 phases. Mitosis is the stage in which replicated DNA is distributed to daughter cells and is itself often subdivided into five distinguishable stages: **prophase**, **prometaphase**, **metaphase**, **anaphase**, and **telophase**. Mitosis is often accompanied by a process called **cytokinesis**, during which the cytoplasmic components of the daughter cells are separated either by an actin ring (animal cells) or by cell plate formation (plant cells). The passage through these phases are controlled by checkpoints. There are three major checkpoints in the cell cycle: one near the end of G_1, a second at the G_2–M transition, and the third during metaphase. These regulatory checks ensure that the processes required to move successfully on to the next phase of the cell cycle have been fully completed and that sufficient resources exist to move on to the next phase of cell division.

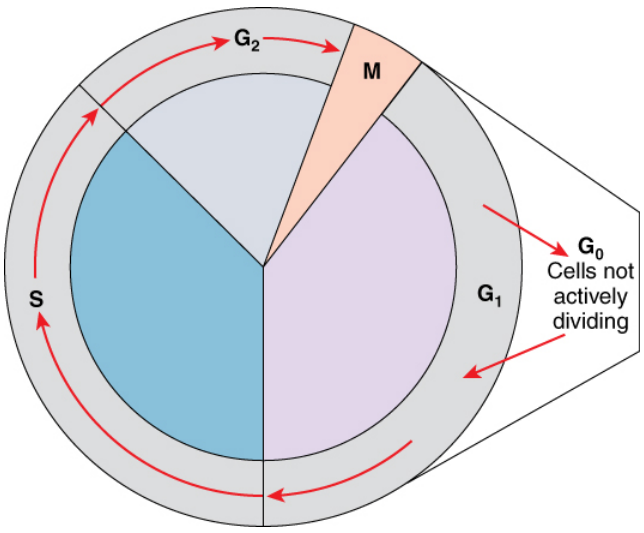
Cell Cycle

In asexually reproducing eukaryotic cells, one "turn" of the cell cycle consists of two general phases: **interphase**, followed by **mitosis** and **cytokinesis**. Interphase is the period of the cell cycle during which the cell may either be living and not dividing or in which it is preparing itself to divide. Most of the cells in a fully-developed multicellular organism typically live in interphase. **Mitosis** is the point in the cell cycle associated with division or distribution of replicated genetic material to two daughter cells. During mitosis the cell nucleus breaks down and two new, fully functional, nuclei form. **Cytokinesis** is the process that divides the cytoplasm into two distinctive cells.

Interphase

G1 Phase

The first stage of interphase is called the **G1 phase**, or first gap, because little change is visible. However, during the G_1 stage, the cell is quite active at the biochemical level. The cell is accumulating the building blocks of chromosomal DNA and the associated proteins, and accumulating enough energy reserves to complete replicating each chromosome in the nucleus.



A cell moves through a series of phases in an orderly manner. During interphase, G_1 involves cell growth and protein synthesis, the S phase involves DNA replication and the replication of the centrosome, and G_2 involves further growth and protein synthesis. The mitotic phase follows interphase. Mitosis is nuclear division during which duplicated chromosomes are segregated and distributed

into daughter nuclei. Usually the cell will divide after mitosis in a process called cytokinesis in which the cytoplasm is divided and two daughter cells are formed.

S Phase

Throughout interphase, nuclear DNA remains in a semi-condensed chromatin configuration. In *S phase* (synthesis phase), DNA replication results in the formation of two identical copies of each chromosome—**sister chromatids**—that are firmly attached at the centromere region. At the end of this stage, each chromosome has been replicated.

In cells using the organelles called **centrosomes**, these structures are often duplicated during S phase. Centrosomes consists of a pair of rod-like **centrioles** composed of tubulin and other proteins that sit at right angles to one another other. The two resulting centrosomes will give rise to the **mitotic spindle**, the apparatus that orchestrates the movement of chromosomes later during mitosis.

G2 Phase

During the **G2 phase**, or second gap, the cell replenishes its energy stores and synthesizes the proteins necessary for chromosome manipulation. Some cell organelles are duplicated, and the cytoskeleton is dismantled to provide resources for the mitotic spindle. There may be additional cell growth during G_2. The final preparations for the mitotic phase must be completed before the cell is able to enter the first stage of mitosis.

G0 Phase

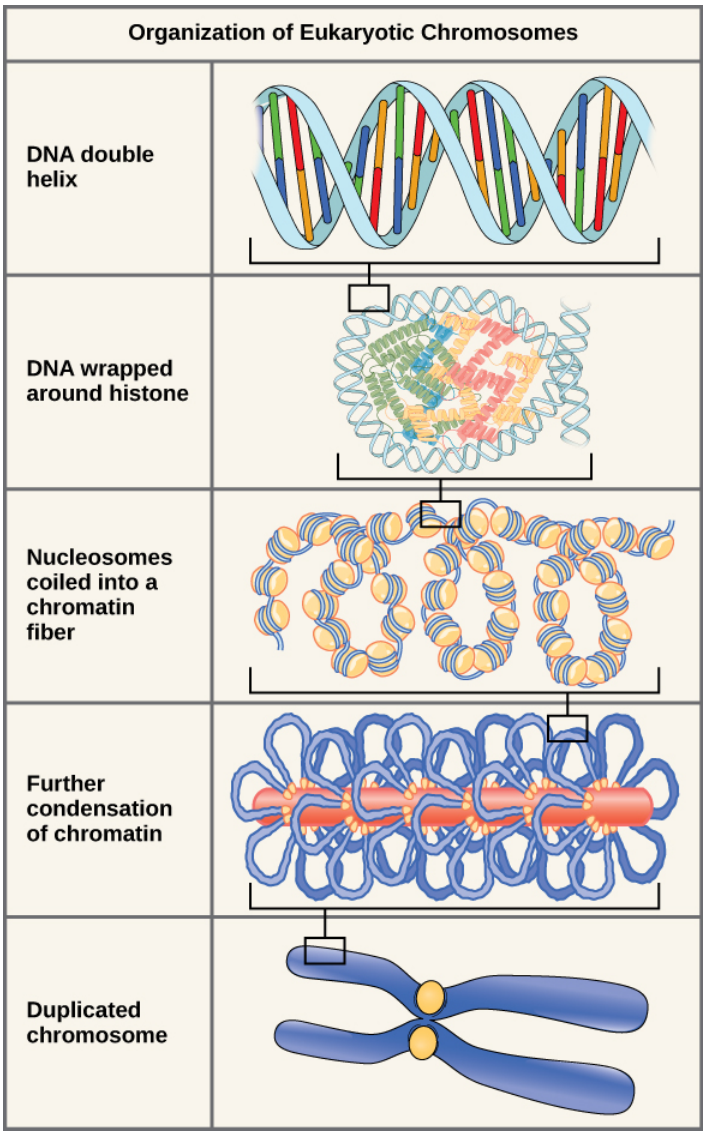
Not all cells adhere to the classic cell-cycle pattern in which a newly formed daughter cell immediately enters interphase, closely followed by the mitotic phase. Cells in the **G0 phase** are not actively preparing to divide. The cell is in a quiescent (inactive) stage, having exited the cell cycle. Some cells enter G_0 temporarily until an external signal triggers the onset of G_1. Other cells that never or rarely divide, such as mature cardiac muscle and nerve cells, remain in G_0 permanently.

A Quick Aside: Structure of Chromosomes During the Cell Cycle

If we lay out the DNA from all 46 chromosomes end to end, it would measure approximately two meters; however, its diameter would be only 2 nm. Considering that the size of a typical human cell is about 10 µm (100,000 cells lined up to equal one meter), DNA must pack tightly to fit in the cell's nucleus. At the same time, it must also be readily accessible for the genes to be expressed. During some stages of the cell cycle, the long strands of DNA are condensed into compact chromosomes. There are several ways that chromosomes are compacted.

Suggested discussion

When should we expect to see highly condensed DNA in the cell (which phases of the cell cycle)? When would the DNA remain un-compacted (during which phases of the cell cycle)?



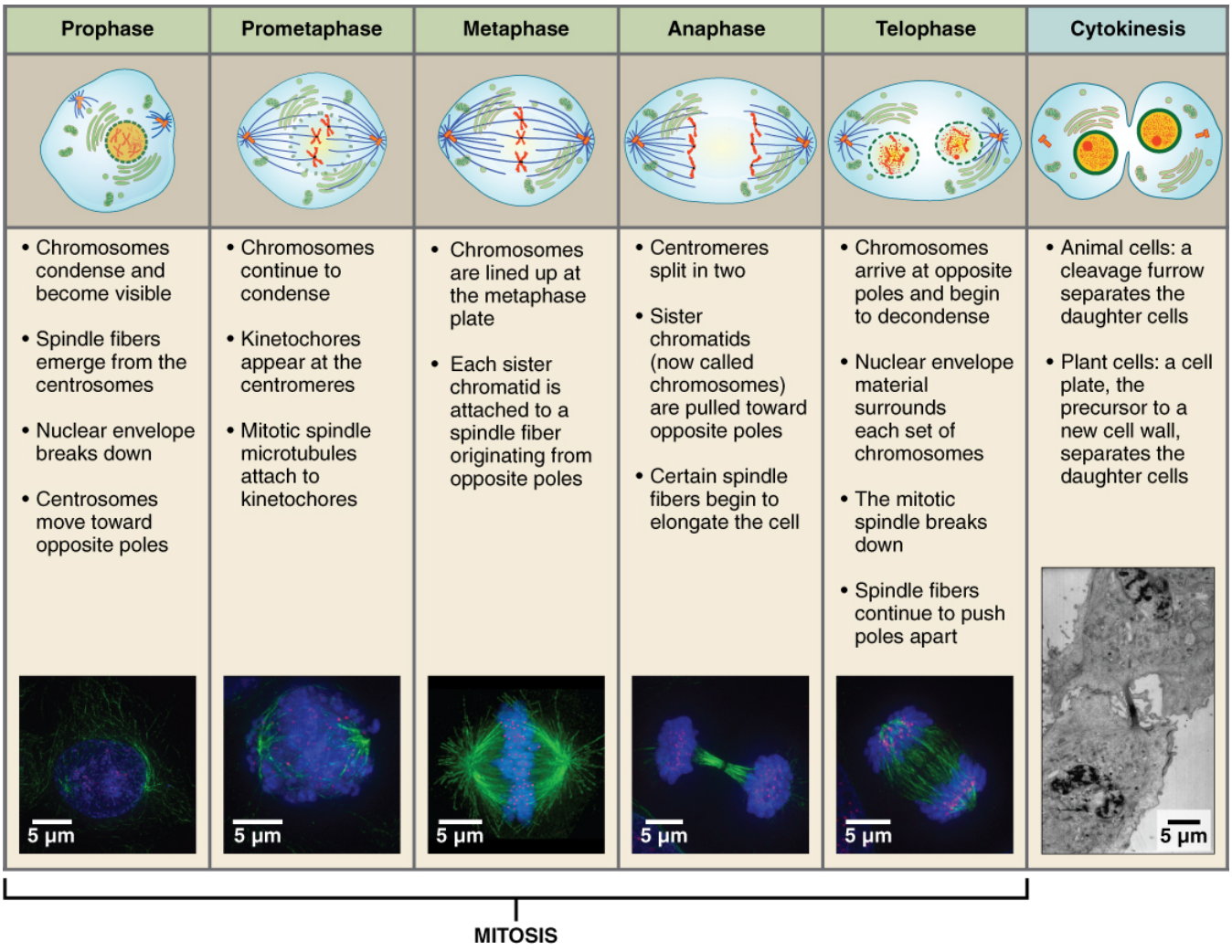
Double-stranded DNA wraps around histone proteins to form nucleosomes that appear like "beads on a string." The nucleosomes are coiled into a 30-nm chromatin fiber. When a cell undergoes mitosis, the chromosomes condense even further.

Mitosis and Cytokinesis

During the **mitotic phase**, a cell undergoes two major processes. First, it completes mitosis, during which the contents of the nucleus are equitably pulled apart and distributed between its two halves. **Cytokinesis** then occurs, dividing the cytoplasm and cell body into two new cells.

Note

The major phases of Mitosis are visually distinct from one another and were originally characterized by what could be seen by viewing dividing cells under a microscope. Some instructors may ask you be able to distinguish each phase be looking at images of cells or more commonly by inspection of cartoon depiction of mitosis. If your instructor is not explicit about this point, remember to ask whether this will be expected of you.



The stages of cell division oversee the separation of identical genetic material into two new nuclei, followed by the division of the cytoplasm. Animal cell mitosis is divided into five stages—prophase, prometaphase, metaphase, anaphase, and telophase—visualized here by light microscopy with fluorescence. Mitosis is usually accompanied by cytokinesis, shown here by a transmission electron microscope. (credit "diagrams": modification of work by Mariana Ruiz Villareal; credit "mitosis micrographs": modification of work by Roy van Heesbeen; credit "cytokinesis micrograph": modification of work by the Wadsworth Center, NY State Department of Health; donated to the Wikimedia foundation; scale-bar data from Matt Russell)

Prophase

Prophase is the first phase of mitosis, during which the loosely packed chromatin coils and condenses into visible chromosomes. During prophase, each chromosome becomes visible with its identical partner (**sister chromatid**) attached, forming the familiar X-shape of sister chromatids. The nucleolus disappears early during this phase, and the nuclear envelope also disintegrates.

A major occurrence during prophase concerns a very important structure that contains the origin site for microtubule growth. Cellular structures called centrioles that serve as origin points from which microtubules extend. These tiny structures also play a very important role during mitosis. A **centrosome** is a pair of centrioles together. The cell contains two centrosomes side-by-side, which begin to move apart during prophase. As the centrosomes migrate to two different sides of the cell, microtubules begin to extend from each like long fingers from two hands extending toward each other. The **mitotic spindle** is the structure composed of the centrosomes and their emerging microtubules.

Near the end of prophase there is an invasion of the nuclear area by microtubules from the mitotic spindle. The nuclear membrane has disintegrated, and the microtubules attach themselves to the centromeres that adjoin pairs of sister chromatids. The **kinetochore** is a protein structure on the centromere that is the point of attachment between the mitotic spindle and the sister chromatids. This stage is referred to as late prophase or "prometaphase" to indicate the transition between prophase and metaphase.

Metaphase

Metaphase is the second stage of mitosis. During this stage, the sister chromatids, with their attached microtubules, line up along a linear plane in the middle of the cell. A metaphase plate forms between the centrosomes that are now located at either end of the cell. The **metaphase plate** is the name for the plane through the center of the spindle on which the sister chromatids are positioned. The microtubules are now poised to pull apart the sister chromatids and bring one from each pair to each side of the cell.

Anaphase

Anaphase is the third stage of mitosis. Anaphase takes place over a few minutes, when the pairs of sister chromatids are separated from one another, forming individual chromosomes once again. These chromosomes are pulled to opposite ends of the cell by their kinetochores, as the microtubules shorten. Each end of the cell receives one partner from each pair of sister chromatids, ensuring that the two new daughter cells will contain identical genetic material.

Telophase

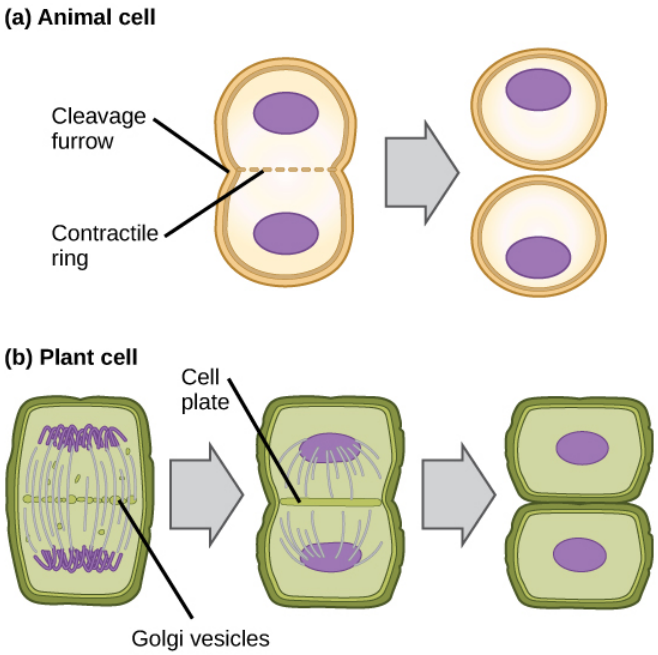
Telophase is the final stage of mitosis. Telophase is characterized by the formation of two new daughter nuclei at either end of the dividing cell. These newly formed nuclei surround the genetic material, which uncoils such that the chromosomes return to loosely packed chromatin. Nucleoli also reappear within the new nuclei, and the mitotic spindle breaks apart, each new cell receiving its own complement of DNA, organelles, membranes, and centrioles. At this point, the cell is already beginning to split in half as cytokinesis begins.

Cytokinesis

Cytokinesis is the second part of the mitotic phase during which cell division is completed by the physical separation of the cytoplasmic components into two daughter cells. Although the stages of mitosis are similar for most eukaryotes, the process of cytokinesis is quite different for eukaryotes that have cell walls, such as plant cells.

In cells such as animal cells that lack cell walls, cytokinesis begins following the onset of anaphase. A contractile ring composed of actin filaments forms just inside the plasma membrane at the former metaphase plate. The actin filaments pull the equator of the cell inward, forming a fissure. This fissure, or "crack," is called the **cleavage furrow**. The furrow deepens as the actin ring contracts, and eventually the membrane and cell are cleaved in two (see the figure below).

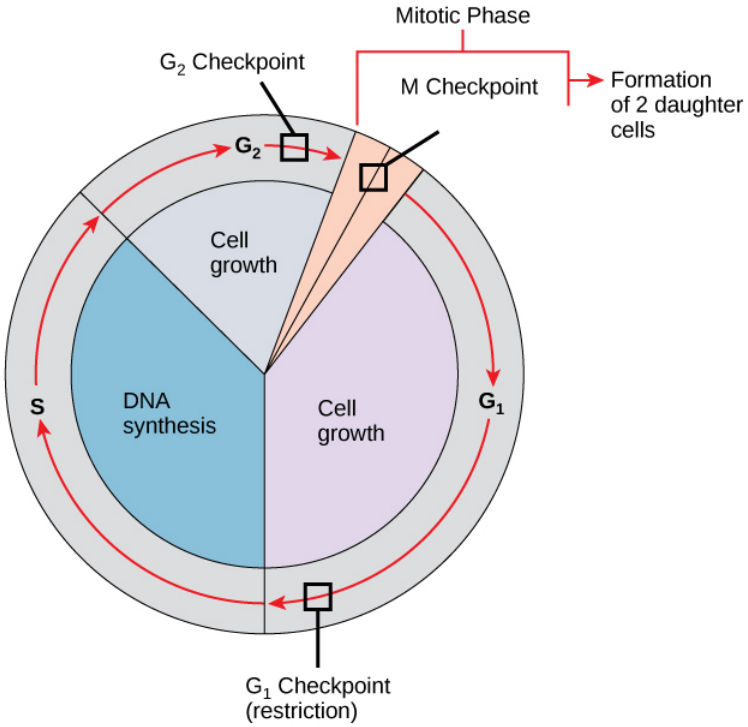
In plant cells, a cleavage furrow is not possible because of the rigid cell walls surrounding the plasma membrane. A new cell wall must form between the daughter cells. During interphase, the Golgi apparatus accumulates enzymes, structural proteins, and glucose molecules prior to breaking up into vesicles and dispersing throughout the dividing cell. During telophase, these Golgi vesicles move on microtubules to collect at the metaphase plate. There, the vesicles fuse from the center toward the cell walls; this structure is called a **cell plate**. As more vesicles fuse, the cell plate enlarges until it merges with the cell wall at the periphery of the cell. Enzymes use the glucose that has accumulated between the membrane layers to build a new cell wall of cellulose. The Golgi membranes become the plasma membrane on either side of the new cell wall (see panel b in the figure below).



In part (a), a cleavage furrow forms at the former metaphase plate in the animal cell. The plasma membrane is drawn in by a ring of actin fibers contracting just inside the membrane. The cleavage furrow deepens until the cells are pinched in two. In part (b), Golgi vesicles coalesce at the former metaphase plate in a plant cell. The vesicles fuse and form the cell plate. The cell plate grows from the center toward the cell walls. New cell walls are made from the vesicle contents.

Cell Cycle Check Points

It is essential that daughter cells be nearly exact duplicates of the parent cell. Mistakes in the duplication or distribution of the chromosomes lead to mutations that may pass forward to every new cell produced from the abnormal cell. To prevent a compromised cell from continuing to divide, there are internal control mechanisms that operate at three main **cell cycle checkpoints** at which the cell cycle can be stopped until conditions are favorable. These checkpoints occur near the end of G_1, at the G_2–M transition, and during metaphase (see figure below).



The cell cycle is controlled at three checkpoints. Integrity of the DNA is assessed at the G_1 checkpoint. Proper chromosome duplication is assessed at the G_2 checkpoint. Attachment of each kinetochore to a spindle fiber is assessed at the M checkpoint.

G1 Checkpoint

The G_1 checkpoint determines whether all conditions are favorable for cell division to proceed into S phase where DNA replication occurs. The G_1 checkpoint, also called the restriction point, is the point at which the cell irreversibly commits to the cell-division process. Besides adequate reserves and cell size, there is a check for damage to the genomic DNA at the G_1 checkpoint. A cell that does not meet all the requirements will not be released into the S phase.

G2 Checkpoint

The G_2 checkpoint bars the entry to the mitotic phase if certain conditions are not met. As in the G_1 checkpoint, cell size and protein reserves are assessed. However, the most important role of the G_2 checkpoint is to ensure that all of the chromosomes have been replicated and that the replicated DNA is not damaged.

M Checkpoint

The M checkpoint occurs near the end of the metaphase stage of mitosis. The M checkpoint is also known as the spindle checkpoint because it determines if all the sister chromatids are correctly attached to the spindle microtubules. Because the separation of the sister chromatids during anaphase is an irreversible step, the cycle will not proceed until the kinetochores of each pair of sister chromatids are firmly anchored to spindle fibers arising from opposite poles of the cell.

Note

Watch what occurs at the G_1, G_2, and M checkpoints by visiting this animation of the cell cycle.

When the Cell Cycle gets out of Control

Most people understand that cancer or tumors are caused by abnormal cells that multiply continuously. If the abnormal cells continue to divide unstopped, they can damage the tissues around them, spread to other parts of the body, and eventually result in death. In healthy cells, the tight regulation mechanisms of the cell cycle prevent this from happening, while failures of cell cycle control can cause unwanted and excessive cell division. Failures of control may be caused by inherited genetic abnormalities that

compromise the function of certain “stop” and “go” signals. Environmental insult that damages DNA can also cause dysfunction in those signals. Often, a combination of both genetic predisposition and environmental factors lead to cancer.

The process of a cell escaping its normal control system and becoming cancerous may actually happen throughout the body quite frequently. Fortunately, certain cells of the immune system are capable of recognizing cells that have become cancerous and destroying them. However, in certain cases the cancerous cells remain undetected and continue to proliferate. If the resulting tumor does not pose a threat to surrounding tissues, it is said to be benign and can usually be easily removed. If capable of damage, the tumor is considered malignant and the patient is diagnosed with cancer.

Homeostatic Imbalances: Cancer Arises from Homeostatic Imbalances

Cancer is an extremely complex condition, capable of arising from a wide variety of genetic and environmental causes. Typically, mutations or aberrations in a cell’s DNA that compromise normal cell cycle control systems lead to cancerous tumors. Cell cycle control is an example of a homeostatic mechanism that maintains proper cell function and health. While progressing through the phases of the cell cycle, a large variety of intracellular molecules provide stop and go signals to regulate movement forward to the next phase. These signals are maintained in an intricate balance so that the cell only proceeds to the next phase when it is ready. This homeostatic control of the cell cycle can be thought of like a car’s cruise control. Cruise control will continually apply just the right amount of acceleration to maintain a desired speed, unless the driver hits the brakes, in which case the car will slow down. Similarly, the cell includes molecular messengers, such as cyclins, that push the cell forward in its cycle.

In addition to cyclins, a class of proteins that are encoded by genes called proto-oncogenes provide important signals that regulate the cell cycle and move it forward. Examples of proto-oncogene products include cell-surface receptors for growth factors, or cell-signaling molecules, two classes of molecules that can promote DNA replication and cell division. In contrast, a second class of genes known as tumor suppressor genes sends stop signals during a cell cycle. For example, certain protein products of tumor suppressor genes signal potential problems with the DNA and thus stop the cell from dividing, while other proteins signal the cell to die if it is damaged beyond repair. Some tumor suppressor proteins also signal a sufficient surrounding cellular density, which indicates that the cell need not presently divide. The latter function is uniquely important in preventing tumor growth: normal cells exhibit a phenomenon called “contact inhibition;” thus, extensive cellular contact with neighboring cells causes a signal that stops further cell division.

These two contrasting classes of genes, proto-oncogenes and tumor suppressor genes, are like the accelerator and brake pedal of the cell’s own “cruise control system,” respectively. Under normal conditions, these stop and go signals are maintained in a homeostatic balance. Generally speaking, there are two ways that the cell’s cruise control can lose control: a malfunctioning (overactive) accelerator, or a malfunctioning (underactive) brake. When compromised through a mutation, or otherwise altered, proto-oncogenes can be converted to oncogenes, which produce oncoproteins that push a cell forward in its cycle and stimulate cell division even when it is undesirable to do so. For example, a cell that should be programmed to self-destruct (a process called apoptosis) due to extensive DNA damage might instead be triggered to proliferate by an oncoprotein. On the other hand, a dysfunctional tumor suppressor gene may fail to provide the cell with a necessary stop signal, also resulting in unwanted cell division and proliferation.

A delicate homeostatic balance between the many proto-oncogenes and tumor suppressor genes delicately controls the cell cycle and ensures that only healthy cells replicate. Therefore, a disruption of this homeostatic balance can cause aberrant cell division and cancerous growths.

FOOTNOTES

1. Leigh Van Valen, “A new evolutionary law,” Evolutionary Theory 1 (1973): 1–30.
2. Adam S. Wilkins and Robin Holliday, “The Evolution of Meiosis from Mitosis,” *Genetics* 181 (2009): 3–12.
3. Marilee A. Ramesh, Shehre-Banoo Malik and John M. Logsdon, Jr, “A Phylogenetic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis,” *Current Biology* 15 (2005):185–91.

Reading_25 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Reading_26

Learning Objectives Associated with SS1_2025_Bis2A_Facciotti_Reading_26

CD.3 Compare and contrast the sequence of events that need to occur during mitosis versus meiosis and why they are necessary: include the roles of microtubules, motor proteins, centrosomes, and the level of DNA condensation.
CD.4 Compare and contrast the behaviors of sister chromatids, chromosomes, and homologous chromosomes in the context of mitosis versus meiosis.
CD.41 Correctly use the terms "chromosome", "homologous chromosome", "chromatid" and "sister chromatid" to describe their common representations in figures of cell division.
CD.5 Create and discuss a picture that illustrates the importance of crossing over and chromatid exchange during meiosis I and explain what happens if these crossover events do not occur.
CD.17 Compare and contrast sexual and asexual modes of reproduction, explicitly referencing costs and benefits of each mode in the context of natural selection in various scenarios.
CD.14 Describe the basic principles of Mendelian inheritance (segregation and independent assortment) and apply these principles to the inheritance of nuclear genes to interpret the results of sexual crosses.

Sexual Reproduction



Sexual reproduction was an early evolutionary innovation after the appearance of eukaryotic cells. That most eukaryotes reproduce sexually is evidence of its evolutionary success. In many animals, it is the only mode of reproduction. And yet, scientists recognize some real disadvantages to sexual reproduction. On the surface, offspring that are genetically identical to the parent may appear to be more advantageous. If the parent organism is successfully occupying a habitat, offspring with the same traits would be similarly successful. There is also the obvious benefit to an organism that can produce offspring by asexual budding, fragmentation, or asexual eggs. These methods of reproduction do not require another organism of the opposite sex. There is no need to spend energy finding or attracting a mate. That energy can be spent on producing more offspring. Some organisms that lead a solitary lifestyle have retained the ability to reproduce asexually. In addition, in asexual populations every individual is capable of reproduction. In contrast, the males in sexual populations (half the population) are not producing offspring themselves. Because of this, an asexual population can in theory grow twice as fast as a sexual population. This means that in competition, the asexual population would have the advantage. These advantages to asexual reproduction, which are also disadvantages to sexual reproduction, should mean that the number of species with asexual reproduction should be more common.

However, multicellular organisms that exclusively depend on asexual reproduction are rare.

So why is sexual reproduction so common?

This is one of the important questions in biology and has been the focus of much research from the latter half of the twentieth century until now. A likely explanation is that the variation that sexual reproduction creates among offspring is very important to the survival and reproduction of those offspring. The only source of genetic variation in asexual organisms is mutation. In sexually reproducing organisms, mutations are continually reshuffled between generations when parents combine their unique genomes, and the genes are mixed into different combinations by the process of **meiosis**.

The Red Queen Hypothesis

There is no question that sexual reproduction provides evolutionary advantages to organisms that employ this mechanism to produce offspring. The problematic question is why, even in the face of seemingly stable conditions, sexual reproduction persists when it is more difficult and produces fewer offspring for individual organisms? Variation is the outcome of sexual reproduction, but why is ongoing variation necessary? Enter the Red Queen hypothesis, first proposed by Leigh Van Valen in 1973.[1] The concept was named in reference to the Red Queen's race in Lewis Carroll's book, *Through the Looking-Glass*, in which the Red Queen says one must run at full speed just to stay where one is.

All species coevolve with other organisms. For example, predators coevolve with their prey, and parasites coevolve with their hosts. A remarkable example of coevolution between predators and their prey is the unique coadaptation of night flying bats and their moth prey. Bats find their prey by emitting high-pitched clicks, but moths have evolved simple ears to hear these clicks so they can avoid the bats. The moths have also adapted behaviors, such as flying away from the bat when they first hear it, or dropping suddenly to the ground when the bat is upon them. Bats have evolved "quiet" clicks in an attempt to evade the moth's hearing. Some moths have evolved the ability to respond to the bats' clicks with their own clicks as a strategy to confuse the bats echolocation abilities.

Each tiny advantage gained by favorable variation gives a species an edge over close competitors, predators, parasites, or even prey. The only method that will allow a coevolving species to keep its own share of the resources is also to continually improve its ability to survive and produce offspring. As one species gains an advantage, other species must also develop an advantage or they will be outcompeted. No single species progresses too far ahead because genetic variation among progeny of sexual reproduction provides all species with a mechanism to produce adapted individuals. Species whose individuals cannot keep up become extinct. The Red Queen's catchphrase was, "It takes all the running you can do to stay in the same place." This is an apt description of coevolution between competing species.

Meiosis



Sexual reproduction requires **fertilization**, the union of two cells from two individual organisms. If those two cells each contain one set of chromosomes, then the resulting cell contains two sets of chromosomes. **Haploid** cells contain one set of chromosomes, **diploid** cells contain two sets of chromosomes. The number of sets of chromosomes in a cell is called its **ploidy** level. If the reproductive cycle is to continue, then the diploid cell must somehow reduce its number of chromosome sets before fertilization can occur again, or there will be a continual doubling in the number of chromosome sets in every generation. So, in addition to fertilization, sexual reproduction includes a nuclear division that reduces the number of chromosome sets.

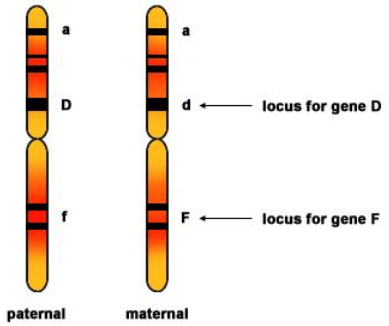
The nuclear division that forms haploid cells, which is called **meiosis**, is related to mitosis. In mitosis, both the parent and the daughter nuclei are at the same ploidy level—diploid for most plants and animals. Meiosis employs many of the same mechanisms as mitosis. However, the starting nucleus is always diploid and the nuclei that result at the end of a meiotic cell division are haploid. To achieve this reduction in chromosome number, meiosis consists of one round of chromosome duplication and two rounds of nuclear division. Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds of division, the major process and the stages are designated with a "I" or a "II." Thus, **meiosis I** is the first round of meiotic division and consists of prophase I, prometaphase I, and so on. **Meiosis II**, in which the second round of meiotic division takes place, includes prophase II, prometaphase II, and so on.

Meiosis I

Meiosis is preceded by an interphase consisting of the G_1, S, and G_2 phases, which are nearly identical to the phases preceding mitosis.

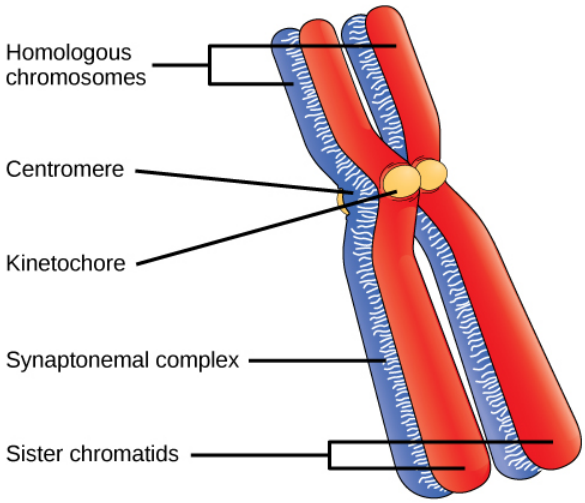
Prophase I

Early in prophase I, before the chromosomes can be seen clearly microscopically, **homologous chromosomes** are attached at their tips to the nuclear envelope by proteins. Homologous chromosomes are similar but not identical chromosomes. For example, chromosome 12 from your mother and chromosome 12 from your father will both be present inside each of your cells. Each chromosome 12 contains the same genes, usually in the same locations, however, each gene can be a different allele. Gene A on chromosome 12 from your mother may be allele R' and gene A on chromosome 12 from your father may be allele r. In species such as humans, even though the X and Y sex chromosomes are not homologous (most of their genes differ), they have a small region of homology that allows the X and Y chromosomes to pair up during prophase I. A partial synaptonemal complex develops only between the regions of homology. It will be very important to understand what homologous chromosomes are when following the process of meiosis.



Two homologous chromsomes are shown prior to DNA replication. Each chromosome has three genes with their locus marked. Homologous chromosomes contain the same genes but are not identical. They each can contain different alleles of each gene.
Source: http://mrphome.net/mrp/Homologous_Chromosome.html

As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. The **synaptonemal complex**, a lattice of proteins between the homologous chromosomes, first forms at specific locations and then spreads to cover the entire length of the chromosomes. The tight pairing of the homologous chromosomes is called **synapsis**. In synapsis, the genes on the chromatids of the homologous chromosomes are aligned precisely with each other. The synaptonemal complex supports the exchange of chromosomal segments between non-sister homologous chromatids, a process called crossing over. Crossing over can be observed visually after the exchange as **chiasmata** (singular = chiasma) (see figure below).

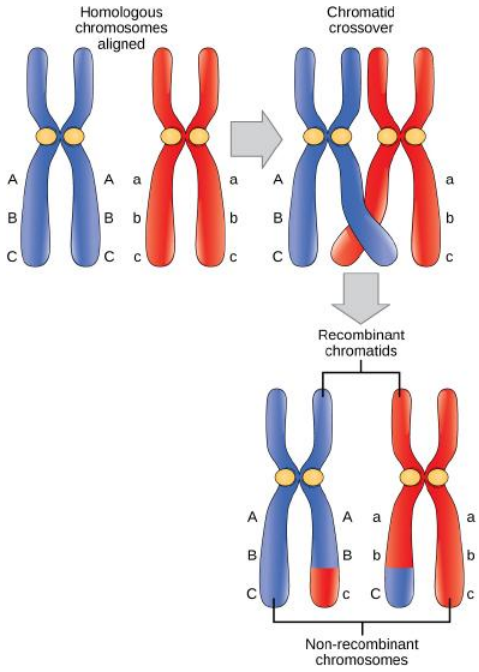


Early in prophase I, homologous chromosomes come together to form a synapse. The chromosomes are bound tightly together and in perfect alignment by a protein lattice called a synaptonemal complex and by cohesin proteins at the centromere.

Located at intervals along the synaptonemal complex are large protein assemblies called **recombination nodules**. These assemblies mark the points of later chiasmata and mediate the multi-step process of **crossover**—or genetic recombination—between the non-sister chromatids. Near the recombination nodule on each chromatid, the double-stranded DNA is cleaved, the cut ends are modified, and a new connection is made between the non-sister chromatids. As prophase progresses, the synaptonemal complex begins to break down and the chromosomes begin to condense. When the synaptonemal complex is gone, the homologous chromosomes remain attached to each other at the centromere and at chiasmata. The chiasmata remain until anaphase I. The number of chiasmata varies according to the species and the length of the chromosome. There must be at least one chiasma per chromosome for proper separation of homologous chromosomes during meiosis I, but there may be as many as 25. Following crossover, the synaptonemal complex breaks down and the cohesin connection between homologous pairs is also removed. At the

end of prophase I, the pairs are held together only at the chiasmata (figure below) and are called **tetrads** because the four sister chromatids of each pair of homologous chromosomes are now visible.

The crossover events are the first source of genetic variation in the nuclei produced by meiosis. A single crossover event between homologous non-sister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. Now, when that sister chromatid is moved into a gamete cell it will carry some DNA from one parent of the individual and some DNA from the other parent. The sister recombinant chromatid has a combination of maternal and paternal genes that did not exist before the crossover. Multiple crossovers in an arm of the chromosome have the same effect, exchanging segments of DNA to create recombinant chromosomes.



Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes.

Possible NB Discussion Point

What are the major differences between Prophase I of Meiosis and Prophase of Mitosis? Why are these distinctions so significant?

Prometaphase I

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. Kinetochore proteins are multiprotein complexes that bind the centromeres of a chromosome to the microtubules of the mitotic spindle. Microtubules grow from centrosomes placed at opposite poles of the cell. The microtubules move toward the middle of the cell and attach to one of the two fused homologous chromosomes. The microtubules attach at each chromosomes' kinetochores. With each member of the homologous pair attached to opposite poles of the cell, in the next phase, the microtubules can pull the homologous pair apart. A spindle fiber that has attached to a kinetochore is called a kinetochore microtubule. At the end of prometaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome facing each pole. The homologous chromosomes are still held together at chiasmata. In addition, the nuclear membrane has broken down entirely.

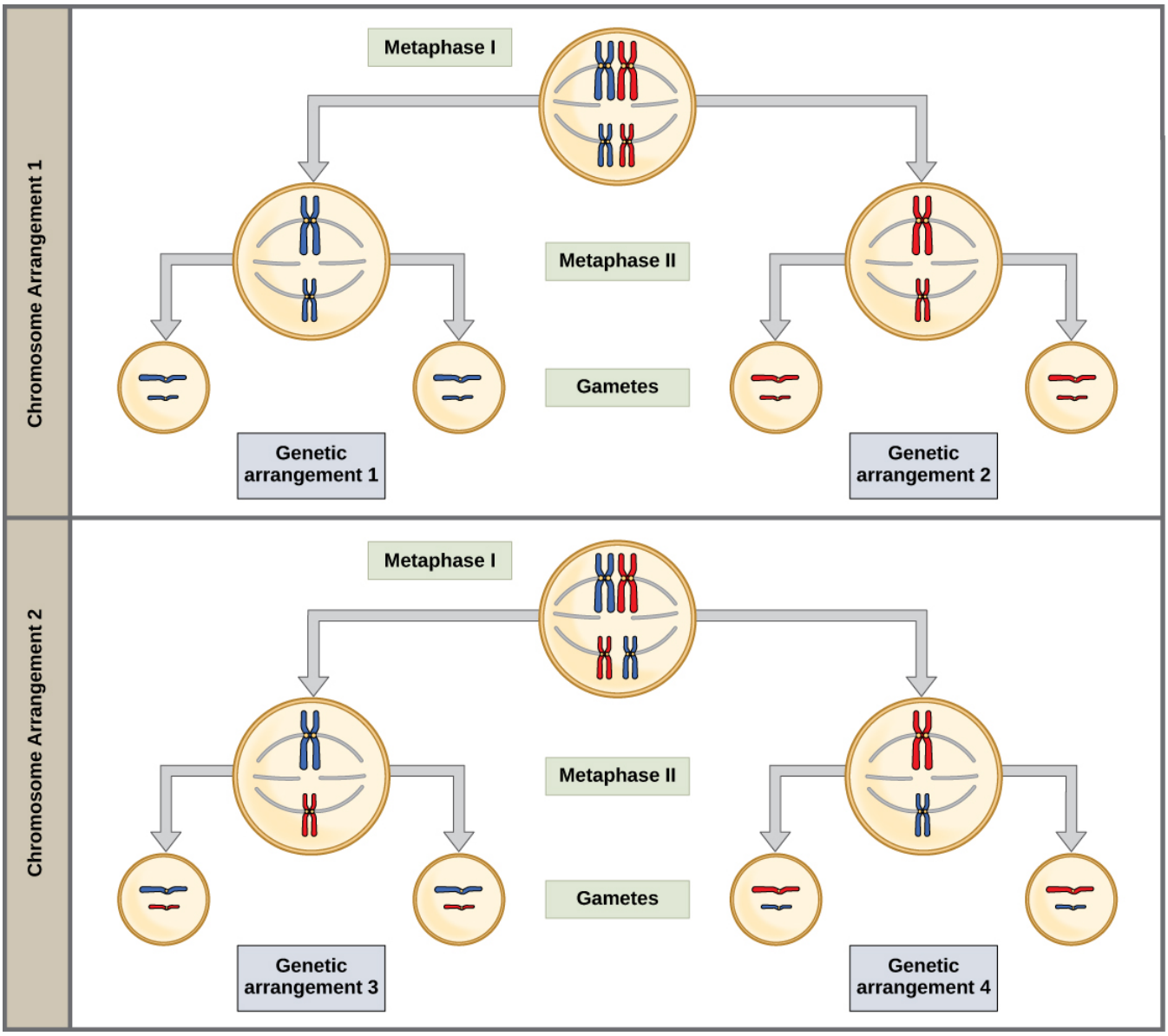
Metaphase I

During metaphase I, the homologous chromosomes are arranged in the center of the cell with the kinetochores facing opposite poles. The homologous pairs orient themselves randomly at the equator. For example, if the two homologous members of chromosome 1 are labeled a and b, then the chromosomes could line up a-b, or b-a. This is important in determining the genes carried by a gamete, as each will only receive one of the two homologous chromosomes. This is called **Independent Assortment**. Recall that homologous chromosomes are not identical, they contain slight differences in their genetic information, causing each gamete to have a unique genetic makeup.

This randomness is the physical basis for the creation of the second form of genetic variation in offspring. Consider that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets, one from each parent. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the egg. Every cell of the multicellular offspring has copies of the original two sets of homologous chromosomes. In prophase I of meiosis, the homologous chromosomes form the tetrads. In metaphase I, these pairs line up at the midway point between the two poles of the cell to form the metaphase plate. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Any maternally inherited chromosome may face either pole. Any paternally inherited chromosome may also face either pole. The orientation of each tetrad is independent of the orientation of the other 22 tetrads.

This event—the random (or independent) assortment of homologous chromosomes at the metaphase plate—is the second mechanism that introduces variation into the gametes or spores. In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations is dependent on the number of chromosomes making up a set. There are two possibilities for orientation at the metaphase plate; the possible number of alignments therefore equals $2n$, where n is the number of chromosomes per set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possible genetically-distinct gametes. This number does not include the variability that was previously created in the sister chromatids by crossover. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition (see figure below).

To summarize the genetic consequences of meiosis I, the maternal and paternal genes are recombined by crossover events that occur between each homologous pair during prophase I. In addition, the random assortment of tetrads on the metaphase plate produces a unique combination of maternal and paternal chromosomes that will make their way into the gametes.



Random, independent assortment during metaphase I can be demonstrated by considering a cell with a set of two chromosomes (n = 2). In this case, there are two possible arrangements at the equatorial plane in metaphase I. The total possible number of different gametes is 2n, where n equals the number of chromosomes in a set. In this example, there are four possible genetic combinations for the gametes. With n = 23 in human cells, there are over 8 million possible combinations of paternal and maternal chromosomes.

Anaphase I

In anaphase I, the microtubules pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. The chiasmata are broken in anaphase I as the microtubules attached to the fused kinetochores pull the homologous chromosomes apart.

Telophase I and Cytokinesis

In telophase, the separated chromosomes arrive at opposite poles. The remainder of the typical telophase events may or may not occur, depending on the species. In some organisms, the chromosomes decondense and nuclear envelopes form around the

chromatids in telophase I. In other organisms, cytokinesis—the physical separation of the cytoplasmic components into two daughter cells—occurs without reformation of the nuclei. In nearly all species of animals and some fungi, cytokinesis separates the cell contents via a cleavage furrow (constriction of the actin ring that leads to cytoplasmic division). In plants, a cell plate is formed during cell cytokinesis by Golgi vesicles fusing at the metaphase plate. This cell plate will ultimately lead to the formation of cell walls that separate the two daughter cells.

Two haploid cells are the end result of the first meiotic division. The cells are haploid because at each pole, there is just one of each pair of the homologous chromosomes. Therefore, only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, even though each homolog still consists of two sister chromatids. Recall that sister chromatids are merely duplicates of one of the two homologous chromosomes (except for changes that occurred during crossing over). In meiosis II, these two sister chromatids will separate, creating four haploid daughter cells.

Meiosis II

In some species, cells enter a brief interphase, or **interkinesis**, before entering meiosis II. Interkinesis lacks an S phase, so chromosomes are not duplicated. The two cells produced in meiosis I go through the events of meiosis II in synchrony. During meiosis II, the sister chromatids within the two daughter cells separate, forming four new haploid gametes. The mechanics of meiosis II is similar to mitosis, except that each dividing cell has only one set of homologous chromosomes. Therefore, each cell has half the number of sister chromatids to separate out as a diploid cell undergoing mitosis.

Prophase II

If the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The centrosomes that were duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed.

Prometaphase II

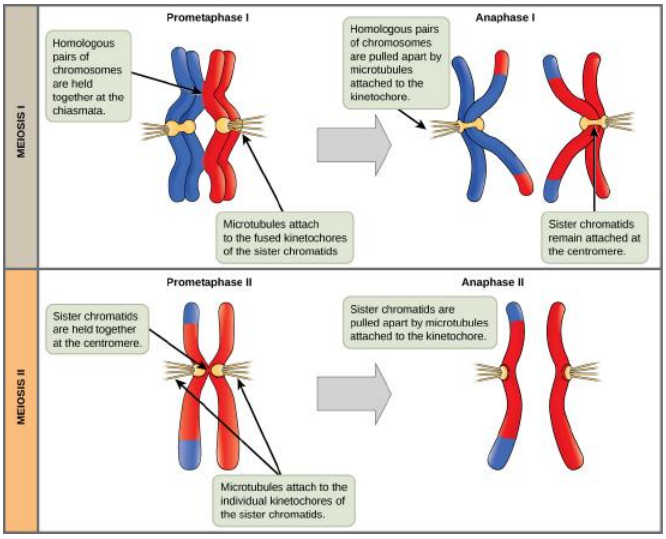
The nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches to microtubules from opposite poles.

Metaphase II

The sister chromatids are maximally condensed and aligned at the equator of the cell.

Anaphase II

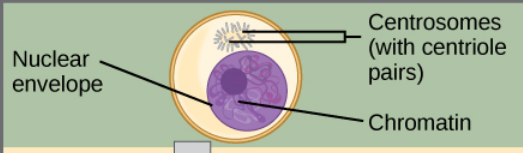
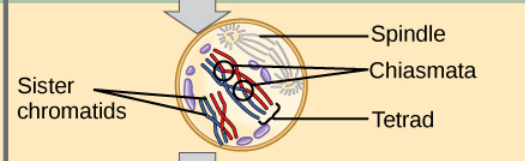
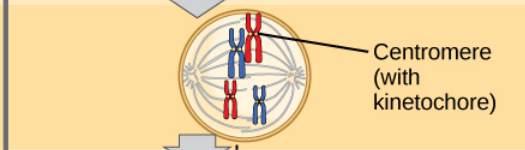
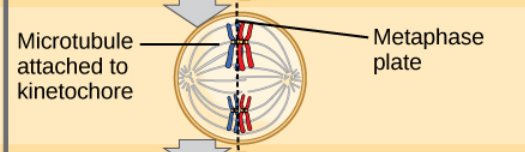
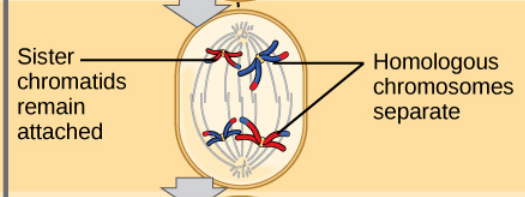
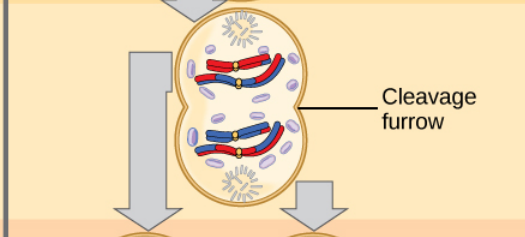
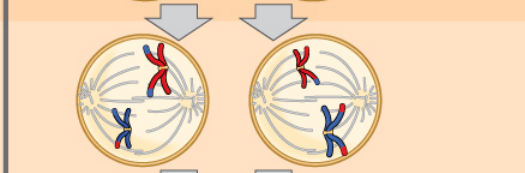
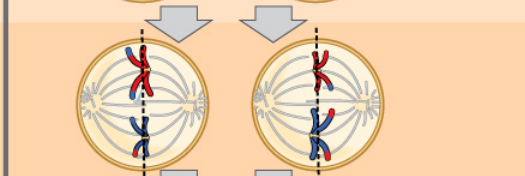
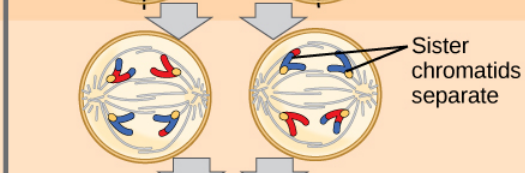
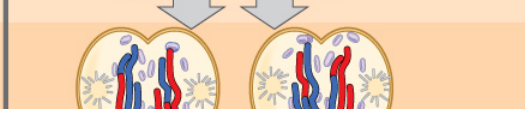
The sister chromatids are pulled apart by the kinetochore microtubules and move toward opposite poles. Non-kinetochore microtubules elongate the cell.

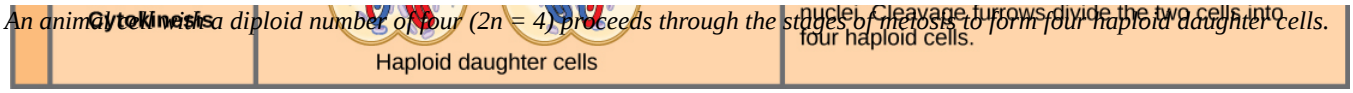


The process of chromosome alignment differs between meiosis I and meiosis II. In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes, and the homologous chromosomes are arranged at the midpoint of the cell in metaphase I. In anaphase I, the homologous chromosomes are separated. In prometaphase II, microtubules attach to the kinetochores of sister chromatids, and the sister chromatids are arranged at the midpoint of the cells in metaphase II. In anaphase II, the sister chromatids are separated.

Telophase II and Cytokinesis

The chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. Cytokinesis separates the two cells into four unique haploid cells. At this point, the newly formed nuclei are both haploid. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and because of the recombining of maternal and paternal segments of chromosomes (with their sets of genes) that occurs during crossover. The entire process of meiosis is outlined in the figure below.

	Stage	Event	Outcome
INTERPHASE	S phase		Chromosomes are duplicated during interphase. The resulting sister chromatids are held together at the centromere. The centrosomes are also duplicated.
MEIOSIS I	Prophase I		Chromosomes condense, and the nuclear envelope fragments. Homologous chromosomes bind firmly together along their length, forming a tetrad. Chiasmata form between non-sister chromatids. Crossing over occurs at the chiasmata. Spindle fibers emerge from the centrosomes.
	Prometaphase I		Homologous chromosomes are attached to spindle microtubules at the fused kinetochore shared by the sister chromatids. Chromosomes continue to condense, and the nuclear envelope completely disappears.
	Metaphase I		Homologous chromosomes randomly assemble at the metaphase plate, where they have been maneuvered into place by the microtubules.
	Anaphase I		Spindle microtubules pull the homologous chromosomes apart. The sister chromatids are still attached at the centromere.
	Telophase I and Cytokinesis		Sister chromatids arrive at the poles of the cell and begin to decondense. A nuclear envelope forms around each nucleus and the cytoplasm is divided by a cleavage furrow. The result is two haploid cells. Each cell contains one duplicated copy of each homologous chromosome pair.
MEIOSIS II	Prophase II		Sister chromatids condense. A new spindle begins to form. The nuclear envelope starts to fragment.
	Prometaphase II		The nuclear envelope disappears, and the spindle fibers engage the individual kinetochores on the sister chromatids.
	Metaphase II		Sister chromatids line up at the metaphase plate.
	Anaphase II		Sister chromatids are pulled apart by the shortening of the kinetochore microtubules. Non-kinetochore microtubules lengthen the cell.
	Telophase II and		Chromosomes arrive at the poles of the cell and decondense. Nuclear envelopes surround the four



An animal cell with a diploid number of four (2n = 4) proceeds through the stages of meiosis to form four haploid daughter cells.

Possible NB Discussion Point

Have you ever enjoyed the convenience of a seedless fruit? If you've eaten the modern day banana, then you have consumed a triploid fruit. While the wild fruit is diploid and can sexually reproduce, seedless bananas arise from mutations, planned hybridizations, and can propagate asexually. Explain why triploid organisms are incapable of successfully undergoing meiosis. Can you think of any benefits to being triploid instead of diploid?

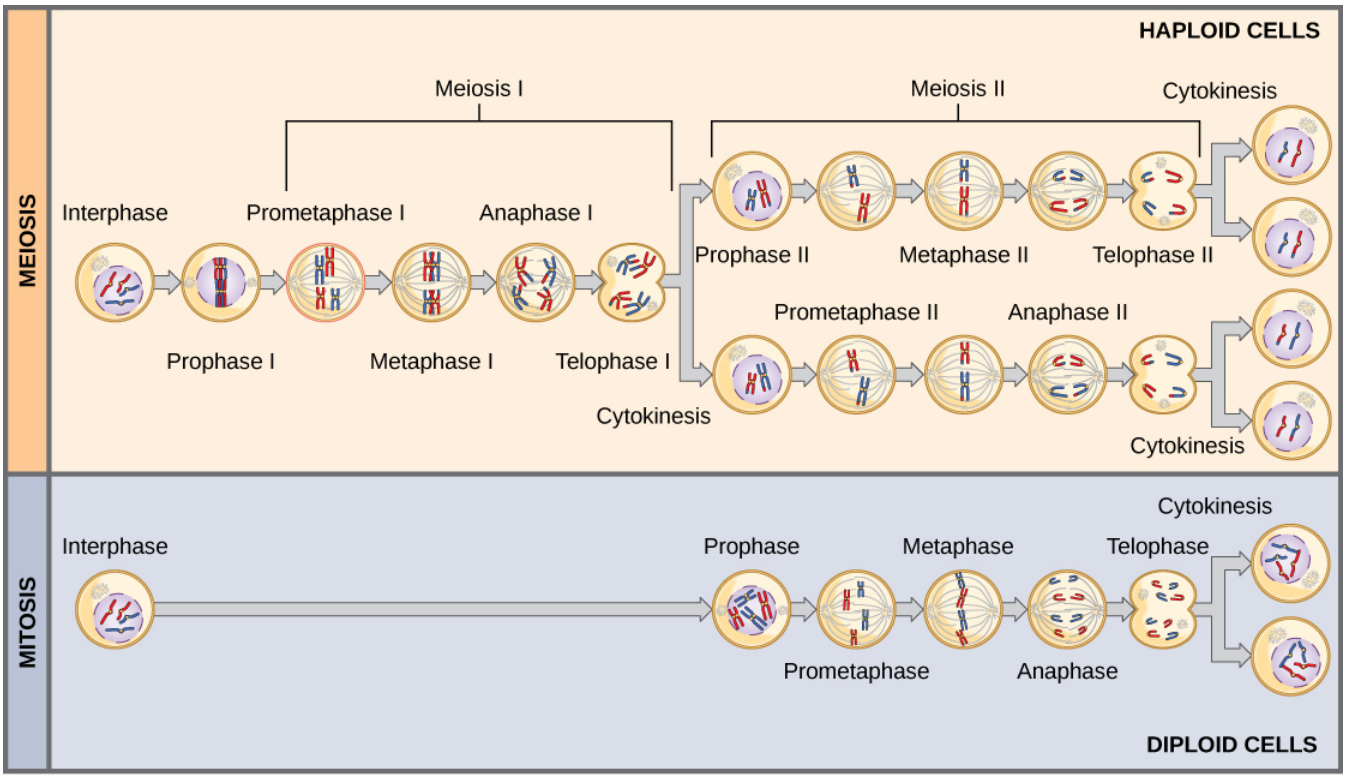
Comparing Mitosis and Meiosis

Mitosis and meiosis are both forms of division of the nucleus in eukaryotic cells. They share some similarities, but also exhibit distinct differences that lead to very different outcomes. Mitosis is a single nuclear division that results in two nuclei that are usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original nucleus. They have the same number of sets of chromosomes, one set in the case of haploid cells and two sets in the case of diploid cells. In most plants and all animal species, it is typically diploid cells that undergo mitosis to form new diploid cells. In contrast, meiosis consists of two nuclear divisions resulting in four nuclei that are usually partitioned into four new cells. The nuclei resulting from meiosis are not genetically identical and they contain one chromosome set only. This is half the number of chromosome sets in the original cell, which is diploid.

The main differences between mitosis and meiosis occur in meiosis I, which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs become associated with each other, are bound together with the synaptonemal complex, develop chiasmata and undergo crossover between sister chromatids, and line up along the metaphase plate in tetrads with kinetochore fibers from opposite spindle poles attached to each kinetochore of a homolog in a tetrad. All of these events occur only in meiosis I.

When the chiasmata resolve and the tetrad is broken up with the homologs moving to one pole or another, the ploidy level—the number of sets of chromosomes in each future nucleus—has been reduced from two to one. For this reason, meiosis I is referred to as a **reduction division**. There is no such reduction in ploidy level during mitosis.

Meiosis II is much more analogous to a mitotic division. In this case, the duplicated chromosomes (only one set of them) line up on the metaphase plate with divided kinetochores attached to kinetochore fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid—now referred to as a chromosome—is pulled to one pole while the other sister chromatid is pulled to the other pole. If it were not for the fact that there had been crossover, the two products of each individual meiosis II division would be identical (like in mitosis). Instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because although there are fewer copies of the genome in the resulting cells, there is still one set of chromosomes, as there was at the end of meiosis I.



						OUTCOME
PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Meiosis and mitosis are both preceded by one round of DNA replication; however, meiosis includes two nuclear divisions. The four daughter cells resulting from meiosis are haploid and genetically distinct. The daughter cells resulting from mitosis are diploid and identical to the parent cell.

The Mystery of the Evolution of Meiosis

Some characteristics of organisms are so widespread and fundamental that it is sometimes difficult to remember that they evolved like other simpler traits. Meiosis is such an extraordinarily complex series of cellular events that biologists have had trouble hypothesizing and testing how it may have evolved. Although meiosis is inextricably entwined with sexual reproduction and its advantages and disadvantages, it is important to separate the questions of the evolution of meiosis and the evolution of sex, because early meiosis may have been advantageous for different reasons than it is now. Thinking outside the box and imagining what the early benefits from meiosis might have been is one approach to uncovering how it may have evolved.

Meiosis and mitosis share obvious cellular processes and it makes sense that meiosis evolved from mitosis. The difficulty lies in the clear differences between meiosis I and mitosis. Adam Wilkins and Robin Holliday[2] summarized the unique events that needed to occur for the evolution of meiosis from mitosis. These steps are homologous chromosome pairing, crossover exchanges, sister

chromatids remaining attached during anaphase, and suppression of DNA replication in interphase. They argue that the first step is the hardest and most important, and that understanding how it evolved would make the evolutionary process clearer. They suggest genetic experiments that might shed light on the evolution of synapsis.

There are other approaches to understanding the evolution of meiosis in progress. Different forms of meiosis exist in single-celled protists. Some appear to be simpler or more "primitive" forms of meiosis. Comparing the meiotic divisions of different protists may shed light on the evolution of meiosis. Marilee Ramesh and colleagues [3] compared the genes involved in meiosis in protists to understand when and where meiosis might have evolved. Although research is still ongoing, recent scholarship into meiosis in protists suggests that some aspects of meiosis may have evolved later than others. This kind of genetic comparison can tell us what aspects of meiosis are the oldest and what cellular processes they may have borrowed from in earlier cells.

Link to Learning

Click through the steps of this interactive animation to compare the meiotic process of cell division to that of mitosis: How Cells Divide.

FOOTNOTES

1. Leigh Van Valen, "A new evolutionary law," Evolutionary Theory 1 (1973): 1–30.
2. Adam S. Wilkins and Robin Holliday, "The Evolution of Meiosis from Mitosis," *Genetics* 181 (2009): 3–12.
3. Marilee A. Ramesh, Shehre-Banoo Malik and John M. Logsdon, Jr, "A Phylogenetic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis," *Current Biology* 15 (2005):185–91.

Reading_26 is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Index

D

dire

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: SS1_2025_BIS2a_Facciotti

Webpages: 37

All licenses found:

- Undeclared: 100% (37 pages)

By Page

- SS1_2025_BIS2a_Facciotti - *Undeclared*
  - Front Matter - *Undeclared*
    - TitlePage - *Undeclared*
    - InfoPage - *Undeclared*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Readings - *Undeclared*
    - Reading_01 - *Undeclared*
    - Reading_02 - *Undeclared*
    - Reading_03 - *Undeclared*
    - Reading_04 - *Undeclared*
    - Reading_05 - *Undeclared*
    - Reading_06 - *Undeclared*
    - Reading_07 - *Undeclared*
    - Reading_08 - *Undeclared*
    - Reading_09 - *Undeclared*
    - Reading_10 - *Undeclared*
    - Reading_11 - *Undeclared*
    - Reading_12 - *Undeclared*
    - Reading_13 - *Undeclared*
    - Reading_14 - *Undeclared*
    - Reading_15 - *Undeclared*
    - Reading_16 - *Undeclared*
    - Reading_17 - *Undeclared*
    - Reading_18 - *Undeclared*
    - Reading_19 - *Undeclared*
    - Reading_20 - *Undeclared*
    - Reading_21 - *Undeclared*
    - Reading_22 - *Undeclared*
    - Reading_23 - *Undeclared*
    - Reading_24 - *Undeclared*
    - Reading_25 - *Undeclared*
    - Reading_26 - *Undeclared*
  - Back Matter - *Undeclared*
    - Index - *Undeclared*
    - Glossary - *Undeclared*
    - Detailed Licensing - *Undeclared*