Introduction to Molecular and Cell Biology

INTRODUCTION TO MOLECULAR AND CELL BIOLOGY

For use in RWU BIO103

KATHERINE MATTAINI

Introduction to Molecular and Cell Biology by Katherine R. Mattaini is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted.

CONTENTS

Preface ix

Part I. Part 1. What is life?

Chapter 1. The Study of Life 3

Part II. Part 2. What molecules make up all living things?

Chapter 2. The Chemical Context of Life 41
Chapter 3. Amino Acids & Proteins 95
Chapter 4. Carbohydrates 119
Chapter 5. Nucleotides & Nucleic Acids 145
Chapter 6. Lipids 162

Part III. Part 3. What are the main features and categories of cells?

Chapter 7. Introduction to Cells 181
Chapter 8. Membrane Transport 245
Chapter 9. Cell Communication 302

Part IV. Part 4. How do living things acquire and use energy?

Chapter 10: Introduction to Metabolism - Enzymes and Energy 341
Chapter 11. Cellular Respiration 385
Chapter 12. Photosynthesis 434

Part V. Part 5. How are heritable traits determined and passed on?

Chapter 13. The Cell Cycle & Mitosis 477
Chapter 14. DNA Replication 523

Chapter 15. Meiosis & Sexual Reproduction 561

1. Chapter 16. The Central Dogma: Genes to Traits 608
2. Chapter 17. Regulation of Gene Expression 656

Chapter 18. Mendelian Genetics 684

PREFACE

This book was modified from *Biology* at OpenStax.org. It was initially modified by instructors at Front Range Community College (FRCC, Colorado) in June 2019, and further modified by Katherine R. Mattaini in July 2020. Major changes made included reordering the chapters, replacing some images, and some rewording.

The information page for *Biology 2e*, including the link to the first edition, can be found here: https://openstax.org/details/books/biology-2e.

The original book was licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Senior Contributing Authors of the original book:

Yael Avissar, Rhode Island College

Jung Choi, Georgia Institute of Technology

Jean DeSaix, University of North Carolina at Chapel Hill

Vladimir Jurukovski, Suffolk County Community College

Robert Wise, University of Wisconsin, Oshkosh

Connie Rye, East Mississippi Community College

Contributing Authors of the Revised Book for Front Range Community College:

Ann Riedl, Front Range Community College

Christopher Wrobel, Front Range Community College

Maria Carolina Pilonieta, Front Range Community College

Richard Fulghum, Front Range Community College

PART I

PART 1. WHAT IS LIFE?

CHAPTER 1. THE STUDY OF LIFE

Figure 1.1 Behold one of the more stunningly detailed images of the Earth yet created. This Blue Marble Earth montage, created from photographs taken by the VIIRS instrument on board the Suomi NPP satellite, shows many stunning details of our home planet. (Credit: NASA)

Chapter Outline

- 1.1 The Study of Biology
- 1.2 Themes and Concepts of Biology

Introduction

Viewed from space, Earth offers no clues about the diversity of life forms that reside there. The first forms of life on Earth are thought to have been microorganisms that existed for billions of years in the ocean before plants and animals appeared. The mammals, birds, and flowers so familiar to us are all relatively recent, originating 130 to 200 million years ago. Humans have inhabited this planet for only the last 2.5 million years, and only in the last 200,000 years have humans started looking like we do today.

1.1 | The Study of Biology

Learning Objectives

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

- Describe the science of biology.
- Summarize the steps of the scientific method.

What is biology? In simple terms, **biology** is the study of living organisms and their interactions with one another and their environment. This is a very broad definition because the scope of biology is vast. Biologists may study anything from the microscopic or submicroscopic view of a cell to ecosystems and the whole living planet (**Figure 1.1**). Listening to the daily news, you will quickly realize how many aspects of biology are discussed every day. For example, recent news topics include *Escherichia coli* (**Figure 1.2**) outbreaks in spinach and *Salmonella* contamination in peanut butter. Other subjects include efforts toward finding cures for diseases such as AIDS, Alzheimer disease, and cancer. On a global scale, many researchers are committed to finding ways to

protect the planet, solve environmental issues, and reduce the effects of climate change. All of these diverse endeavors are related to different facets of the discipline of biology.

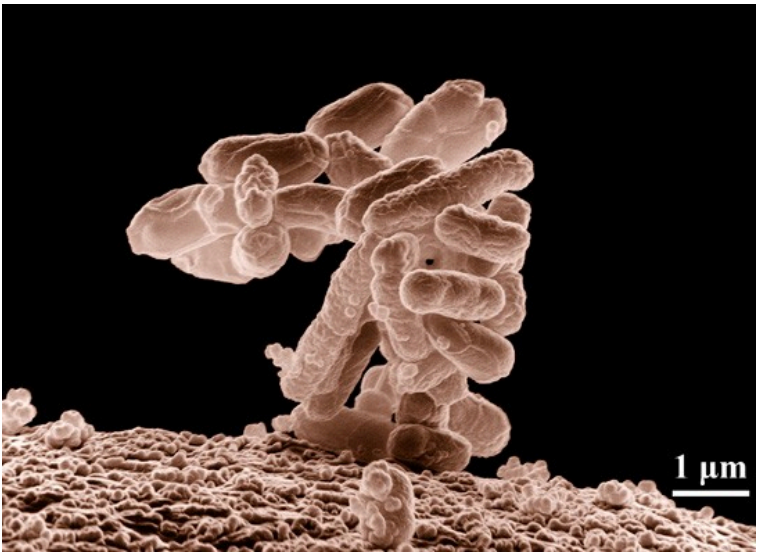


Figure 1.2 *Escherichia coli* (*E. coli*) bacteria, seen in this scanning electron micrograph, are normal residents of our digestive tracts that aid in the absorption of vitamin K and other nutrients. However, virulent strains are sometimes responsible for disease outbreaks. (Credit: Eric Erbe, digital colorization by Christopher Pooley, both of USDA, ARS, EMU)

1.1.1 The Process of Science

Biology is a science, but what exactly is science? What does the study of biology share with other scientific disciplines?

Science (from the Latin *scientia*, meaning "knowledge") can be defined as the process of acquiring knowledge about general truths or the operation of general laws, especially when acquired and tested by the scientific method. It becomes clear from this definition that the application of the scientific method plays a major role in science. The **scientific method** is a method of research with defined steps that include experiments and careful observation.

The steps of the scientific method will be examined in detail later, but one of the most important aspects of this method is the testing of hypotheses by means of repeatable experiments. A **hypothesis** is a suggested explanation for an event, which can be tested. Although using the scientific method is inherent to science, it is inadequate in determining what science is. This is because it is relatively easy to apply the scientific method to disciplines such as physics and chemistry, but when it comes to disciplines like archaeology and geology, the scientific method becomes less applicable as it becomes more difficult to repeat experiments.

These areas of study are still sciences, however. Consider archeology—even though one cannot perform repeatable experiments, hypotheses may still be supported. For instance, an archeologist can hypothesize that an ancient culture existed based on finding a piece of pottery. Further hypotheses could be made about various characteristics of this culture, and these hypotheses may be found to be supported or false from other findings. Over time, as hypotheses continue to be supported,

they may contribute to the formulation of a theory. A **theory** is a tested and confirmed explanation that fits all of the observations or phenomena in a given field of study.

Not surprisingly, the natural science of biology has many branches or subdisciplines. Cell biologists study cell structure and function, while biologists who study anatomy investigate the structure of an entire organism. Those biologists studying physiology focus on the internal functioning of an organism. Some areas of biology focus on only particular types of living things. For example, botanists explore plants, while zoologists specialize in animals.

1.1.2 The Scientific Method

Biologists study the living world by posing questions about it and seeking verifiable responses. This approach is common to other sciences as well and is often referred to as the **scientific method**. The scientific method was used even in ancient times, but it was first documented by England's Sir Francis Bacon (1561–1626), who set up inductive methods for scientific inquiry. The scientific method can be applied to almost all fields of study as a logical, rational problem-solving method.

The scientific process typically starts with an observation that leads to a question. Let's think about a simple problem that starts with an observation and apply the scientific method to solve the problem. One Monday morning, a student arrives

at class and quickly discovers that the classroom is too warm. That is an observation that also describes a problem: the classroom is too warm. The student then asks a question: "Why is the classroom so warm?"

Proposing a Hypothesis

Recall that a hypothesis is a suggested explanation that can be tested. To solve a problem, several hypotheses may be proposed. For example, one hypothesis might be, "The classroom is warm because no one turned on the air conditioning." But there could be other responses to the question, and therefore other hypotheses may be proposed. A second hypothesis might be, "The classroom is warm because there is a power failure, and so the air conditioning doesn't work."

Once a hypothesis has been formulated, the student can make a prediction. A prediction is similar to a hypothesis but it typically has the format "If . . . then . . . ." For example, the prediction for the first hypothesis might be, "*If* the student turns on the air conditioning, *then* the classroom will no longer be too warm."

Testing a Hypothesis

A valid hypothesis must be testable. It should also be **falsifiable**, meaning that it can be disproven by experimental results. Importantly, science does not claim to "prove"

anything because scientific understandings are always subject to modification with further information. This step—openness to disproving ideas—is what distinguishes sciences from non-sciences. The presence of the supernatural, for instance, is neither testable nor falsifiable.

To test a hypothesis, a researcher will conduct one or more experiments designed to eliminate one or more of the hypotheses. Each experiment will have one or more variables and one or more controls. A **variable** is any part of the experiment that can vary or change during the experiment. **Control variables** are variables that are held constant in all parts of the experiment. The **experimental variable** is the one that is changed. Many experiments are designed to test whether one variable, the **dependent variable**, is influenced by another variable, the **independent variable**. The researcher manipulates, or investigates several cases of, the independent variable and observes changes in the dependent variable. For example, if a scientist hypothesizes that coat color in rabbits is influenced by temperature, she may observe rabbits at different temperatures (the independent variable) and observe the coat color (the dependent variable) in each case.

To test the first hypothesis in the above example, the student would find out if the air conditioning is on. If the air conditioning is turned on but does not work, this hypothesis should be rejected. To test the second hypothesis, the student could check if the lights in the classroom are functional. If

so, there is no power failure and this hypothesis should be rejected. Be aware that rejecting one hypothesis does not determine whether or not the other hypotheses can be accepted; it simply eliminates one hypothesis that is not valid. Using the scientific method, the hypotheses that are inconsistent with experimental data are rejected (**Figure 1.3**).

The scientific method may seem too rigid and structured. It is important to keep in mind that, although scientists often follow this sequence, there is flexibility. Sometimes an experiment leads to conclusions that favor a change in approach; often, an experiment brings entirely new scientific questions to the puzzle. Many times, science does not operate in a linear fashion; instead, scientists continually draw inferences and make generalizations, finding patterns as their research proceeds. Scientific reasoning is more complex than the scientific method alone suggests. Notice, too, that the scientific method can be applied to solving problems that are not necessarily scientific in nature.

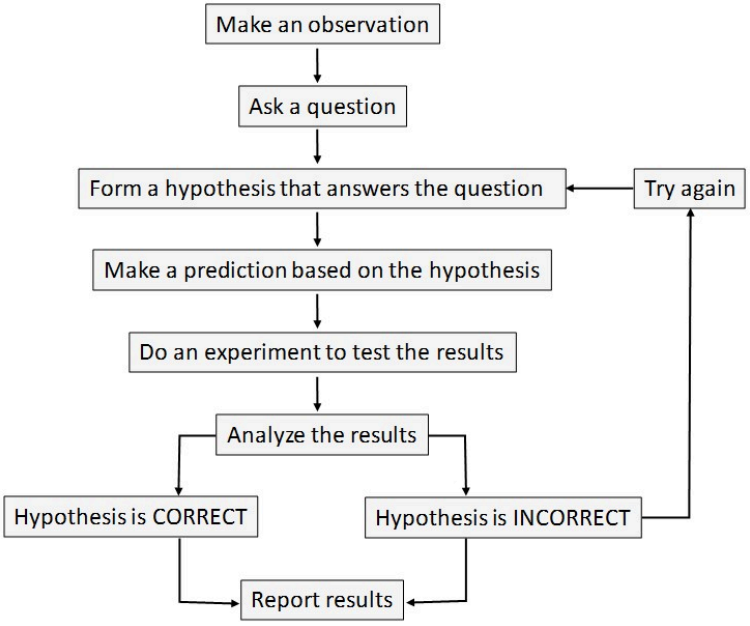


Figure 1.3 The scientific method consists of a series of well-defined steps. If a hypothesis is not supported by experimental data, a new hypothesis can be proposed.

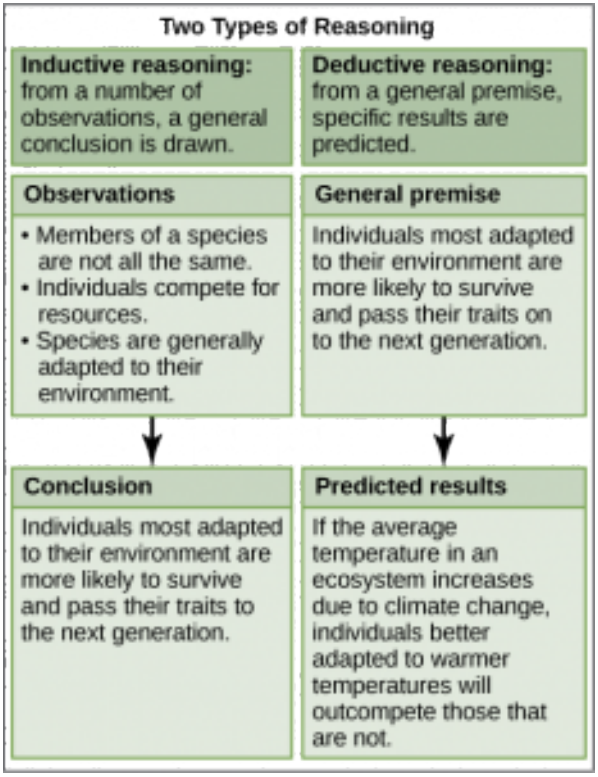


Figure 1.4 Scientists use two types of reasoning, inductive and deductive reasoning, to advance scientific knowledge. As is the case in this example, the conclusion from inductive reasoning can often become the premise for deductive reasoning.

Concept Check

Decide if each of the following is an example of inductive or deductive reasoning.

- All flying birds and insects have wings. Birds and insects flap their wings as they move through the air. Therefore, wings enable flight.
- Insects generally survive mild winters better than harsh ones. Therefore, insect pests will become more problematic if global temperatures increase.
- Chromosomes, which are made of DNA, pass genetic information from parent to offspring during cell division. Therefore, DNA is the genetic material.
- Animals as diverse as humans, insects, and wolves all exhibit social behavior. Therefore, social behavior must have an evolutionary advantage.

Concept Check

In the example below, the scientific method is used to solve an everyday problem.

- Order the scientific method steps (numbered items).
- Match these steps with the processes of solving the problem (lettered items).
- Based on the results of the experiment, is the hypotheses correct? If it is incorrect, propose some alternate hypotheses.

1. Experiment	a. There is something wrong with the electrical outlet.
2. Prediction	b. If something is wrong with the electrical outlet, my coffeemaker also won't work when plugged into it.
3. Question	c. My toaster doesn't toast my bread.
4. Observation	d. I plug my coffeemaker into the outlet.
5. Result	e. My coffeemaker works.
6. Hypothesis	f. Why doesn't my toaster work?

1.1.3 Reporting Scientific Work

Whether scientific research is basic science or applied science, scientists must share their findings in order for other researchers to expand and build upon their discoveries. Collaboration with other scientists—when planning, conducting, and analyzing results—are all important for scientific research. For this reason, important aspects of a scientist's work are communicating with and disseminating results to peers.

Scientists can share results by presenting them at a scientific

meeting or conference, but this approach can reach only the select few who are present. Instead, most scientists present their results in peer-reviewed manuscripts that are published in scientific journals. **Peer-reviewed manuscripts** are scientific papers that are reviewed by a scientist's colleagues, or peers. These colleagues are qualified individuals, often experts in the same research area, who judge whether or not the work is suitable for publication. The process of peer review helps to ensure that the research described in a scientific paper or grant proposal is original, significant, logical, and thorough. Grant proposals, which are requests for research funding, are also subject to peer review. Scientists publish their work so other scientists can reproduce their experiments under similar or different conditions to expand on the findings. The experimental results must be consistent with the findings of other scientists.

A scientific paper is very different from creative writing. Although creativity is required to design experiments, there are fixed guidelines when it comes to presenting scientific results. Scientific writing must be brief, concise, and accurate. Most scientific papers consist of the following:

Abstract: a concise summary of the results of the study

Introduction: background information about what is known in the field as well as the rationale of the work

Materials and Methods: complete and accurate

description of the substances used, and the methods and techniques used by the researchers to gather data. The description should be thorough enough to allow another researcher to repeat the experiment and obtain similar results. This section will also include information on how measurements were made and what types of calculations and statistical analyses were used.

Results and/or Discussion: description of the findings, usually by means of tables or graphs. The researcher will interpret the results, describe how variables may be related, and attempt to explain the observations.

Conclusion: summary of the importance of the experimental findings.

References: It is indispensable to conduct an extensive literature search to put the results in the context of previously published scientific research. Therefore, proper citations are included in this section.

Review articles do not present original scientific findings, or primary literature; instead, they summarize and comment on findings that were published as primary literature and typically include extensive reference sections.

1.2 | Themes and Concepts of

Biology

Learning Objectives

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

- Identify and describe the properties of life.
- Describe the levels of organization among living things.
- Summarize the three unifying theories of biology.
- Name and briefly characterize the three domains of life on Earth.

Biology is the science that studies life, but what exactly is life? This may sound like a silly question with an obvious response, but it is not always easy to define life. For example, a branch of biology called virology studies viruses, which exhibit some of the characteristics of living entities but lack others. It turns out that although viruses can attack living organisms, cause diseases, and even reproduce, they do not meet all of the criteria that most biologists use to define life.

From its earliest beginnings, biology has wrestled with three questions: What are the shared properties that make something "alive"? And once we know something is alive, how do we find meaningful levels of organization in its structure? And, finally, when faced with the remarkable diversity of life, how do we organize the different kinds of organisms so that we can better understand them? As new organisms are discovered every day, biologists continue to seek answers to these and other questions.

1.2.1. Properties of Life

All living organisms share several key characteristics or functions: order, sensitivity or response to the environment, reproduction, growth and development, regulation, homeostasis, energy processing, adaptation, and evolution. When viewed together, these nine characteristics serve to define life.

Order

Organisms are highly organized, coordinated structures that consist of one or more cells. Even very simple, single-celled organisms are remarkably complex: inside each cell, atoms make up molecules; these in turn make up cell organelles and other cellular inclusions. In multicellular organisms (**Figure 1.5**), similar cells form tissues. Tissues, in turn, collaborate

to create organs (body structures with a distinct function). Organs work together to form organ systems.

Figure 1.5 A toad represents a highly organized structure consisting of cells, tissues, organs, and organ systems. (Credit: "Ivengo"/Wikimedia Commons)

Sensitivity or Response to Stimuli

Organisms respond to diverse stimuli. For example, plants can bend toward a source of light, climb on fences and walls, or respond to touch (**Figure 1.6**). Even tiny bacteria can move toward or away from chemicals (a process called *chemotaxis*) or light (*phototaxis*). Movement toward a stimulus is considered a positive response, while movement away from a stimulus is considered a negative response.

Figure 1.6 The leaves of this sensitive plant (*Mimosa pudica*) will instantly droop and fold when touched. After a few minutes, the plant returns to normal. (Credit: Alex Lomas)

Reproduction

Single-celled organisms reproduce by first duplicating their DNA, and then dividing it equally as the cell prepares to divide to form two new cells. Multicellular organisms often produce specialized reproductive cells that will form new individuals. When reproduction occurs, genes containing DNA are passed along to an organism's offspring. These genes ensure that the offspring will belong to the same species and will have similar characteristics (**Figure 1.7**).

Figure 1.7 Although no two are identical, these kittens have inherited genes from both parents and share many of the same characteristics. (Credit: Rocky Mountain Feline Rescue)

Growth and Development

Organisms grow and develop following specific instructions coded for by their genes. These genes provide instructions that will direct cellular growth and development, ensuring that a species' young will grow up to exhibit many of the same characteristics as its parents (**Figure 1.7**).

Regulation and Homeostasis

Even the smallest organisms are complex and require multiple regulatory mechanisms to coordinate internal functions, respond to stimuli, and cope with environmental stresses. Two examples of internal functions regulated in an organism are nutrient transport and blood flow. Organs (groups of tissues

working together) perform specific functions, such as carrying oxygen throughout the body, removing wastes, delivering nutrients to every cell, and cooling the body.

In order to function properly, cells need to have appropriate conditions such as proper temperature, pH, and appropriate concentration of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to maintain internal conditions within a narrow range almost constantly, despite environmental changes, through **homeostasis** (literally, "steady state")—the ability of an organism to maintain constant internal conditions. For example, an organism needs to regulate body temperature through a process known as thermoregulation. Organisms that live in cold climates have body structures, including fur, feathers, blubber, and fat, that help them withstand low temperatures and conserve body heat. Structures that aid in this type of insulation In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.

Figure 1.8 Polar bears (*Ursus maritimus*) and other mammals living in ice-covered regions maintain their body temperature by generating heat and reducing heat loss through thick fur and a dense layer of fat under their skin. (Credit: "longhorndave"/Flickr)

Energy Processing

All organisms use a source of energy for their metabolic activities. Some organisms capture energy from the sun and convert it into chemical energy in food; others use chemical energy in molecules they take in as food (**Figure 1.9**).

Figure 1.9 The California condor (*Gymnogyps californianus*) uses chemical energy derived from food to power flight. California condors are an endangered species; this bird has a wing tag that helps biologists identify the individual. (Credit: Pacific Southwest Region U.S. Fish and Wildlife Service)

Adaptation and Evolution

All organisms have DNA as the genetic material that allows parents to pass traits to their offspring. Due to the

changeability of DNA, variability is introduced into populations as they reproduce. This causes species to change over time, or evolve. As variability increases, new species come into being (**Figure 1.10**). The mechanisms of this process will be discussed later.

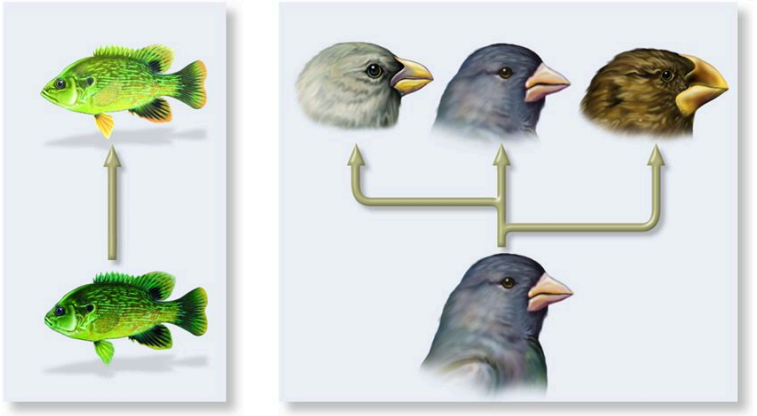


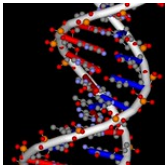
Figure 1.10 A single ancestral species of bird may be the progenitor for multiple species. This process is known as cladogenesis.

1.2.2 Levels of Organization of Living Things

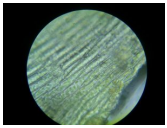
Living things are highly organized and structured, following a hierarchy that can be examined on a scale from small to large. The **atom** is the smallest and most fundamental unit of matter. It consists of a nucleus surrounded by electrons.

Atoms form molecules. A **molecule** is a chemical structure consisting of at least two atoms held together by one or more chemical bonds. Many molecules that are biologically important are **macromolecules**, large molecules that are typically formed from repeating units. An example of a macromolecule is deoxyribonucleic acid (DNA) (**Table 1.1**), which contains the instructions for the structure and functioning of all living organisms.

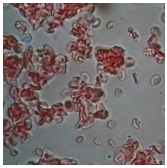
Table 1.1 The biological levels of organization of living things from small to large. From a single organelle to the entire biosphere, living organisms are parts of a highly structured hierarchy. (Credit "macromolecules": "Geoff Hutchison"/Flickr; credit "organelles": "Przemyslawtarka95"/Wikimedia Commons; Credit "cells": "Steve Begin"/Flickr; Credit "tissues": "Circa24"/Wikimedia Commons; credit "organs": anatomical drawing by Leonardo da Vinci; credit "organisms": "dmitry.kaglik"/Flickr; credit "ecosystems": "USFWS – Pacific Region"/Flickr; credit "biosphere": GSFC/NASA Goddard.)



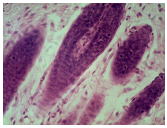
Macromolecules – Ex: DNA double helix



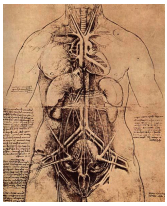
Organelles – Ex: chloroplasts in plant cells



Cells – Ex: human red blood cells



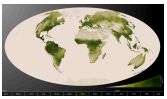
Tissues – Ex: hair follicles in mammalian skin



Organs and organ systems – Ex: "The Principal Organs and Vascular and Urino-Genital Systems of a Woman"

Organisms, populations and communities – Ex: pine trees in a forest community

Ecosystems – Ex: coral reef ecosystem



Biosphere – The sum of all ecosystems on Earth

Some cells contain small structures that exist within cells surrounded by membranes; these are called **organelles**. All living things are made of cells; the **cell** itself is the smallest fundamental unit of structure and function in living organisms. Some organisms consist of a single cell and others are multicellular.

In larger organisms, cells combine to make **tissues**, which are groups of similar cells carrying out similar or related functions. **Organs** are collections of tissues grouped together performing a common function. An **organ system** is a higher level of organization that consists of functionally related organs. Mammals have many organ systems, including the circulatory system and the digestive system. **Organisms** are individual living entities. For example, each tree in a forest is an organism. Some organisms are composed of a single cell.

All the individuals of a species living within a specific area

are collectively called a **population**. For example, a forest may include many pine trees. All of these pine trees represent the population of pine trees in this forest. Different populations may live in the same specific area. For example, the forest with the pine trees includes populations of flowering plants and also insects and microbial populations. A **community** is the sum of populations inhabiting a particular area. For instance, all of the trees, flowers, insects, and other populations in a forest form the forest's community. The forest itself is an ecosystem. An **ecosystem** consists of all the living things in a particular area together with the abiotic, non-living parts of that environment such as nitrogen in the soil or rain water. At the highest level of organization (**Table 1.1**), the **biosphere** is the collection of all ecosystems, and it represents the zones of life on Earth. It includes land, water, and the atmosphere.

Concept Check

Which of the following statements is false?

- Tissues exist within organs which exist within organ systems.
- Communities exist within populations which

exist within ecosystems.
- Organelles exist within cells which exist within tissues.
- Communities exist within ecosystems which exist in the biosphere.

1.2.3 The Three Unifying Theories of Biology

Scientists use the word theory differently than non-scientists do. In science, a **theory** is an explanation that encompasses all of the known information in a field of study. Although theories can continue to be modified, they are almost never replaced or dismissed entirely. Three of the most important theories in biology are cell theory, the chromosomal theory of inheritance, and the theory of evolution by natural selection. The details of these theories will be covered in later chapters, or in the case of evolution by natural selection, in other courses. The basic principles of these theories are described briefly below.

Cell Theory

Cell theory states that all living things are composed of cells,

that cells are the basic unit of life, and that cells come from pre-existing cells. All cells have the same basic structure of a gel-like cytoplasm enclosed by a double layer of lipid molecules called the plasma membrane. In addition, all cells contain DNA as the genetic material.

How did the first cells come into existence? Most scientists think that conditions on early Earth were ideal for the formation of living cells. The early seas were warm, contained a large number of carbon-based molecules, and had a lot of energy input from lightening and volcanos. Sometime around 3.5 billion years ago, carbon-based molecules became surrounded by lipid molecules and evolved the ability to reproduce. This was the first cell. Other scientists postulate that life may have arrived on earth from space on an asteroid or meteor. Since conditions on earth no longer support the formation of life, cells can only come from pre-existing cells at this time on Earth.

The Chromosomal Theory of Inheritance

The **chromosomal theory of inheritance** states that chromosomes carry the genetic material. **Chromosomes** are made of a single very long molecule of DNA wrapped around proteins that serve to pack it tightly into a cell (**Figure 1.11**). Segments of the chromosome called **genes** code for traits, or characteristics. Chromosomes are passed on from parents to offspring, which is the basis of **inheritance**.

Figure 1.11 Polytene chromosomes from the salivary glands of nonbiting midges larvae. (Credit: Doc. RNDr. Josef Reischig, CSc)

The Theory of Evolution by Natural Selection

Species change over time, or **evolve**. In 1869, Charles Darwin and Alfred Russel Wallace proposed a mechanism, called **natural selection**, for this observed evolution. They proposed that since not all of the members of a population can survive, the ones who are most fit will survive and reproduce at a higher rate and will therefore be more likely to pass on their traits. Therefore, over time, the characteristics that are present in a population will change. As populations change, new species come into existence.

For natural selection to work, there must first be **variation** among individuals in a population. The variation arises through changes in the DNA, which leads to changes in the traits of individuals. Secondly, there must be a **selective**

pressure, such as a shortage of food, a predator, a mating preference, etc. For example, if the selective pressure is a predator, individuals who are best able to avoid being eaten will survive and reproduce at a higher rate. Individuals may have different types of **fitness**: some may be able to move faster, others may be able to hide better, while still others may be able to convince the predator that they are poisonous. Regardless, some individuals get eaten and some survive. The final important part is that the survivors must be able to reproduce. Only the individuals who get to pass on their DNA to their offspring are truly fit.

It is important to understand that evolution does not work on individuals, but on populations. Since the most fit individuals have more offspring, their traits become more prevalent in the population over time.

Natural selection is very similar to breeding of domestic animals. Breeders choose animals or plants with favorable traits and allow them to breed. By continuing to choose the desired traits over many generations, breeders developed breeds or strains with strikingly different characteristics. Darwin called this process **artificial selection**. He reasoned that a similar process could take place with nature, rather than people, deciding who got to reproduce.

1.2.4 The Diversity of Life

The fact that biology has such a broad scope has to do with

the tremendous diversity of life on Earth. The source of this diversity is evolution. Evolutionary biologists study the evolution of living things in everything from the microscopic world to ecosystems. The evolution of various life forms on Earth can be summarized in a phylogenetic tree (**Figure 1.12**).

Cells are classified as prokaryotic or eukaryotic. **Prokaryotes** are single-celled or colonial organisms that do not have membrane-bound nuclei; in contrast, the cells of **eukaryotes** do have membrane-bound organelles and a membrane-bound nucleus. Eukaryotes can be either unicellular or multicellular. For most of the history of biology, scientists believed that the main division in the **Tree of Life** (the phylogenetic tree showing all organisms on Earth) was between prokaryotes and eukaryotes.

However, in 1977, American microbiologist Carl Woese, refuted that hypothesis using data obtained from sequencing ribosomal RNA genes. He found that the most fundamental division in life on Earth is between *three* domains: Bacteria, Archaea, and Eukarya (eukaryotes). Although bacteria and archaea are both prokaryotes, archaea are more closely related to eukaryotes than they are to bacteria.

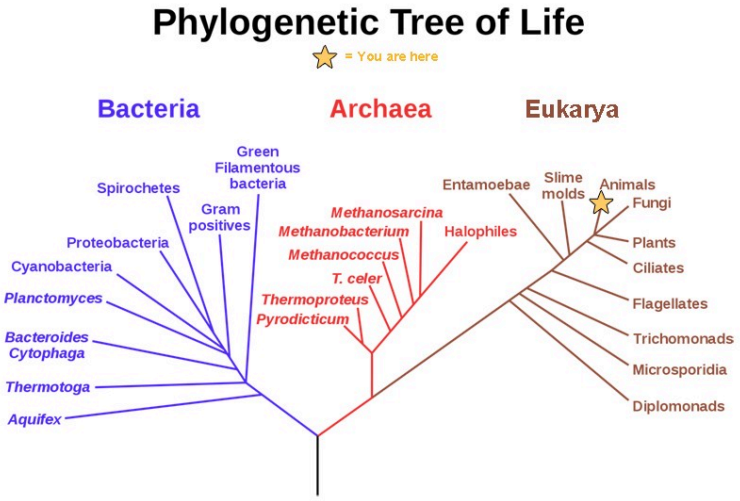


Figure 1.12 Phylogenetic tree constructed by microbiologist Carl Woese. The tree shows the separation of living organisms into three domains: Bacteria, Archaea, and Eukarya. Bacteria and Archaea are prokaryotes, single-celled organisms lacking intracellular organelles. (Credit: Eric Gaba; NASA Astrobiology Institute)

PART II

PART 2. WHAT MOLECULES MAKE UP ALL LIVING THINGS?

CHAPTER 2. THE CHEMICAL CONTEXT OF LIFE

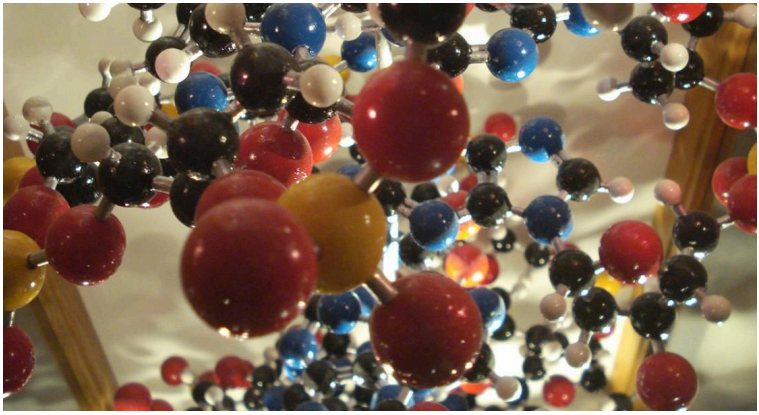


Figure 2.1 Atoms are the building blocks of all the molecules found in the universe—air, soil, water, rocks . . . and also the cells of all living organisms. In this model of an organic molecule, the atoms of carbon (black), hydrogen (white), nitrogen (blue), oxygen (red), and sulfur (yellow) are shown in proportional atomic size. The silver rods indicate chemical bonds. (Credit: modification of work by Christian Guthier)

Chapter Outline

- 2.1 Atoms
- 2.2 Chemical Bonding & Intermolecular Forces
- 2.3 Water, Acids & Bases
- 2.4 The Energy of Life
- 2.5 Carbon

Introduction

Elements in various combinations comprise all matter, including living things. Some of the most abundant elements in living organisms include carbon, hydrogen, nitrogen, oxygen, sulfur, and phosphorus. These form the nucleic acids, proteins, carbohydrates, and lipids that are the fundamental components of living matter. Biologists must understand these important building blocks and the unique structures of the atoms that make up molecules, allowing for the formation of cells, tissues, organ systems, and entire organisms.

All biological processes follow the laws of physics and chemistry; so in order to understand how biological systems

work, it is important to understand the underlying physics and chemistry. For example, the flow of blood within the circulatory system follows the laws of physics that regulate fluid flow. The breakdown of the large, complex molecules of food into smaller molecules—and the conversion of these to release energy to be stored in adenosine triphosphate (ATP)—is a series of chemical reactions that follow chemical laws. The properties of water and the formation of hydrogen bonds are key to understanding living processes. Recognizing the properties of acids and bases is important, for example, to our understanding of the digestive process. Therefore, the fundamentals of physics and chemistry are important for gaining insight into biological processes.

2.1 | Atoms

Learning Objectives

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

- Define matter and elements.
- Describe the interrelationship between

protons, neutrons, and electrons.
- Use atomic number to determine electron configuration.

At its most fundamental level, life is made up of **matter**: any substance that occupies space and has mass. **Elements** are unique forms of matter with specific chemical and physical properties that cannot be broken down into smaller substances by ordinary chemical reactions. There are 118 elements, but only 92 occur naturally. The remaining elements are synthesized in laboratories and are unstable.

Each element is designated by its chemical symbol, which is a single capital letter or, when the first letter is already used for another element, a combination of two letters. Some elements follow the English term for the element, such as C for carbon and Ca for calcium. Other elements' chemical symbols derive from their Latin names; for example, the symbol for sodium is Na, referring to *natrium*, the Latin word for sodium.

The four most common elements in all living organisms are oxygen (O), carbon (C), hydrogen (H), and nitrogen (N). These elements comprise 96% of living organisms. In the non-living world, elements are found in different proportions, and some elements common to living organisms

are relatively rare on the earth as a whole, as shown in **Table 2.1**. In spite of their differences in abundance, all elements and the chemical reactions between them obey the same chemical and physical laws regardless of whether they are a part of the living or non-living world.

Table 2.1. Percentage of elements in living organisms vs. the non-living world.

Element	Life (Humans)	Atmosphere	Earth's Crust
Oxygen (O)	65%	21%	46%
Carbon (C)	18%	trace	trace
Hydrogen (H)	10%	trace	0.1%
Nitrogen (N)	3%	78%	trace

2.1.1 The Structure of the Atom

An **atom** is the smallest unit of matter that retains all of the chemical properties of an element. For example, one gold atom has all of the properties of gold, such as that it is a solid metal at room temperature. Gold atoms cannot be broken down into anything smaller while still retaining the properties of gold.

An atom is composed of two regions: the **nucleus**, which is in the center of the atom and contains protons and neutrons, and the outer region of the atom, which holds its

electrons in orbit around the nucleus (**Figure 2.2**). Atoms contain **subatomic particles**, the largest of which are protons, electrons, and neutrons.

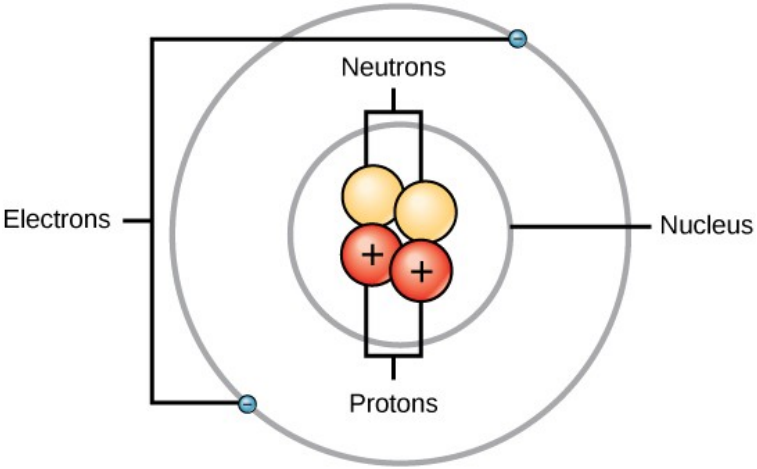


Figure 2.2 Elements, such as helium, depicted here, are made up of atoms. Atoms are made up of protons and neutrons located within the nucleus, with electrons in orbitals surrounding the nucleus.

Protons are positively charged, **electrons** are negatively charged, and **neutrons** are uncharged (**Table 2.2**). Each electron has a negative charge equal to the positive charge of a proton. In uncharged, neutral atoms, the number of electrons orbiting the nucleus is equal to the number of protons inside the nucleus. In these atoms, the positive and negative charges cancel each other out, leading to an atom with no net charge.

Protons and neutrons have approximately the same mass, about 1.67×10^{-24} grams. Scientists arbitrarily define this amount of mass as one **atomic mass unit** (**amu**) or one **Dalton** (**Da**) (**Table 2.2**). Electrons are much smaller in mass than protons, weighing only 9.11×10^{-28} grams, or about 1/1800 of an atomic mass unit. Hence, they do not contribute much to an element's overall atomic mass. Therefore, when considering atomic mass, it is customary to ignore the mass of any electrons and calculate the atom's mass based on the number of protons and neutrons alone.

Accounting for the sizes of protons, neutrons, and electrons, most of the volume of an atom—greater than 99 percent—is, in fact, empty space. With all this empty space, one might ask why so-called solid objects do not just pass through one another. The reason they do not is that the electrons that surround all atoms are negatively charged and negative charges repel each other.

Table 2.2. Properties of subatomic particles.

	Charge	Mass (amu)	Location
Proton	+1	1	nucleus
Neutron	0	1	nucleus
Electron	−1	0	orbitals

2.1.2 Atomic Number, Mass

Number, Isotopes, and Atomic Weight

Atoms of each element contain a characteristic number of protons. The number of protons determines an element's **atomic number** and is used to distinguish one element from another.

The number of neutrons in the atoms of a given element is variable. For example, the element carbon (C) has atomic number 6. Therefore, all neutral carbon atoms have 6 protons and 6 electrons. However, some carbon atoms have 6 neutrons, some have 7 neutrons, and some have 8 neutrons. Together, the number of protons plus the number of neutrons determines an atom's **mass number**. Note that the small contribution of mass from electrons is disregarded when calculating the mass number.

Isotopes are atoms that have the same number of protons but a different number of neutrons. Carbon atoms that have 6 neutrons have a mass number of 12 amu, and are referred to as Carbon-12 or ^{12}C. Carbon atoms with 7 neutrons have a mass number of 13 amu, and are referred to as Carbon-13 or ^{13}C. Carbon atoms that have 8 neutrons have a mass number of 14 amu, and are referred to as Carbon-14 or ^{14}C. These represent three naturally occurring isotopes of carbon (**Figure 2.3**).

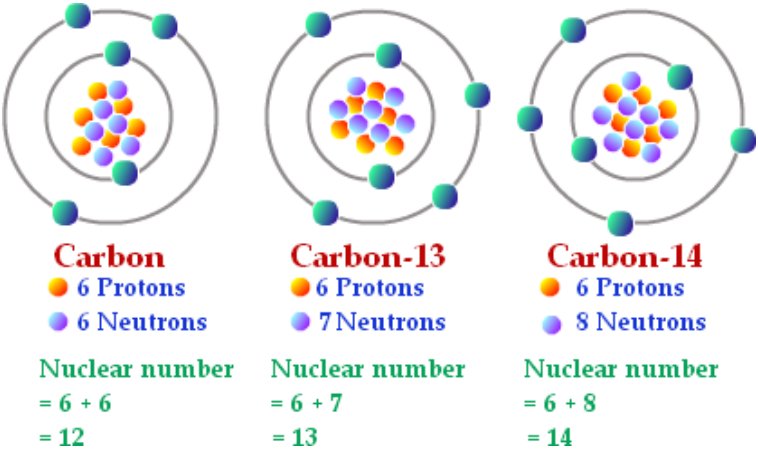


Figure 2.3 Elements may have more than one isotope. For example, the three isotopes of carbon are shown here. They all have 6 protons and electrons, but the number of neutrons varies.

Since an element's isotopes have different mass numbers, scientists also determine the **atomic weight**, which is the calculated mean of the mass number for the naturally occurring isotopes of an element on earth. Often, the resulting number is not a whole number. For example, the atomic mass of chlorine (Cl) is 35.45 because chlorine is composed of several isotopes, some (the majority) with atomic mass 35 (17 protons and 18 neutrons) and some with atomic mass 37 (17 protons and 20 neutrons). The atomic mass of carbon is 12.011 because the great majority of carbon on earth is Carbon-12.

Some isotopes may emit neutrons, protons, and electrons in order to become more stable. These are radioactive

isotopes, or **radioisotopes**. Radioactive decay describes the loss of energy and/or mass that occurs when an unstable atom's nucleus releases radiation. Carbon-14 is an example of a radioisotope (**Figure 2.4**).

evolution CONNECTION

Carbon Dating Carbon-14 (^{14}C) is a naturally occurring radioisotope. In a living organism, the relative amount of ^{14}C is equal to the concentration of ^{14}C in the atmosphere. When an organism dies, the ratio between ^{14}C and ^{12}C will decrease as ^{14}C decays.After approximately 5,730 years, half of the starting concentration of ^{14}C decayed. The time it takes for half of the original concentration of an isotope to decay to its more stable form is called its half-life. The number of half-lives since an object such as bones or wood was alive can be determined by comparing the ratio of the ^{14}C concentration in the object to the amount of ^{14}C detected in the atmosphere. The age of the material can be calculated with accuracy if it is not much older than about 50,000 years (Figure 2.4).

Figure 2.4 The age of carbon-containing remains less than about 50,000 years old, such as this pygmy mammoth, can be determined using carbon dating. (Credit: Bill Faulkner, NPS)

2.1.3 The Periodic Table

The different elements are organized and displayed in the **periodic table**. Devised by Russian chemist Dmitri Mendeleev (1834–1907) in 1869, the table groups elements that share certain chemical properties. The properties of elements are responsible for their physical state at room temperature: they may be gases, solids, or liquids. Elements

also have specific **chemical reactivity**, the ability to chemically bond with each other.

In the periodic table, shown in **Figure 2.5**, 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. Each square on the table gives the name, chemical symbol, atomic weight, and atomic mass for one element. For example, the first square contains hydrogen, its symbol (H), its atomic number of (1), and its atomic mass (1.01).

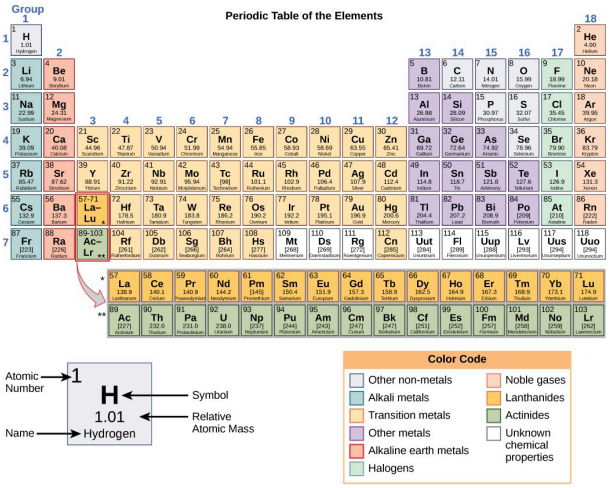


Figure 2.5 The periodic table shows the atomic mass and atomic number of each element. The atomic number appears above the symbol for the element and the approximate atomic mass appears below it.

2.1.4 Electron Shells and the Bohr Model

An early model of the atom was developed in 1913 by Danish scientist Niels Bohr (1885–1962). The Bohr model shows

the atom with a central nucleus and the electrons in circular **electron shells** at specific distances from the nucleus. The closer an energy shell is the nucleus, the lower the energy of the electrons that occupy that shell. The first (1n) electron shell can hold two electrons, while the second (2n) and third (3n) shells can hold eight electrons each (**Figure 2.6**).

Electron Orbitals

Although useful to explain the reactivity and chemical bonding of certain elements, the Bohr model of the atom does not accurately reflect how electrons are spatially distributed surrounding the nucleus. They do not circle the nucleus like the earth orbits the sun, but are found in three-dimensional **electron orbitals**. Mathematical equations can predict within a certain level of probability where an electron might be at any given time. The area where an electron is most likely to be found is called its orbital. Each electron orbital can hold only two electrons.

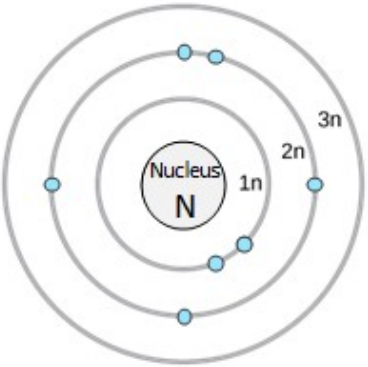


Figure 2.6 In the Bohr model, developed by Niels Bohrs in 1913, electrons exist in orbitals within electron shells. An electron normally exists in the electron shell with the lowest available energy, which is the one closest to the nucleus.

Electrons fill orbitals in a consistent order: they first fill

the orbitals closest to the nucleus, then they continue to fill orbitals of increasing energy further from the nucleus. If there are multiple orbitals of equal energy, one electron will be added to each orbital before a second electron is added to any of the orbitals. For example, the second energy level in the nitrogen atom shown here has one electron in each of three orbitals and two electrons in the fourth orbital (**Figure 2.6**).

The innermost shell has a single orbital, for a maximum of two electrons but the next two electron shells can each have four orbitals, for a maximum of eight electrons. The **octet rule** states that, with the exception of the innermost shell, atoms are most stable when they have eight electrons in their **valence shell**, the outermost electron shell. Examples of some neutral atoms and their electron configurations are shown in **Figure 2.7**.

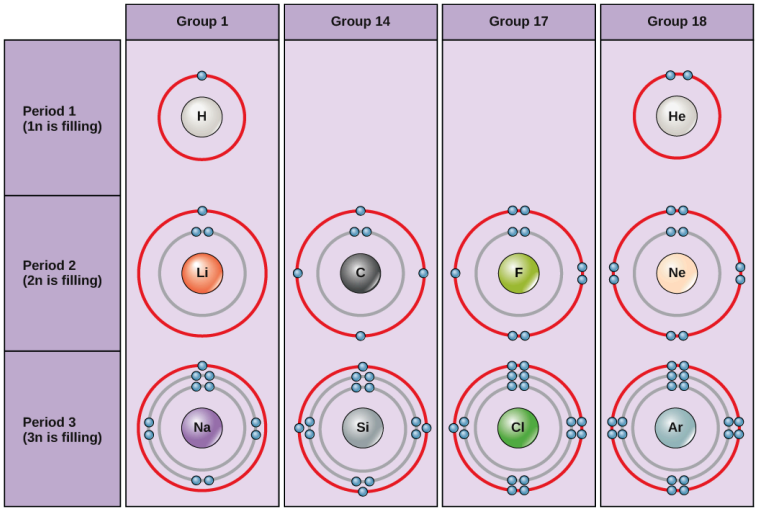


Figure 2.7 Bohr diagrams indicate how many electrons fill each shell. Group 18 elements (e.g., helium, neon, and argon) have a full valence shell. A full valence shell is the most stable electron configuration. Elements in other groups, with partially filled valence shells, gain or lose electrons to achieve a stable configuration.

Concept Check

Draw Bohr's diagrams for oxygen and magnesium atoms.

- How many electrons do oxygen atoms need to gain in order to achieve a stable electron configuration?
- How many electrons do magnesium atoms need to lose to achieve a stable configuration?

The periodic table is arranged in columns and rows based on the number of electrons and where these electrons are located. Note that elements in the far right column of the periodic table (**Figure 2.5**) all have filled valence shells. These atoms are highly stable, making it unnecessary for them to share electrons with other atoms. They are therefore non-reactive and are called **inert gases** (or **noble gases**).

In general, atoms with 4-7 electrons in their valence shell will either gain electrons to become negatively charged ions, or will share electrons with other atoms to form covalently bonded molecules. Atoms with 1-3 electrons in their valence shell will tend to donate their electrons to other atoms until they have a full outer shell. As a result of losing negatively charged electrons, they become positively charged **ions**.

2.2 | Chemical Bonding & Intermolecular Forces

Learning Objectives

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

- Compare the ways in which electrons can be donated or shared between atoms.
- Explain the ways in which naturally occurring elements combine to create molecules.
- Identify intermolecular forces that hold molecules together.

2.2.1 Chemical Reactions and Molecules

The octet rule drives the chemical behavior of atoms. Atoms will chemically react and bond to each other form **molecules**, which are simply two or more atoms chemically bonded together. A **compound** is a type of molecule that contains

two or more *different* types of atoms. In short, atoms form **chemical bonds** with other atoms, thereby obtaining the electrons they need to attain a stable electron configuration. The familiar water molecule, H2O, consists of two hydrogen atoms and one oxygen atom bonded together (**Figure 2.8**). Atoms can form molecules by donating, accepting, or sharing electrons to fill their outer shells.

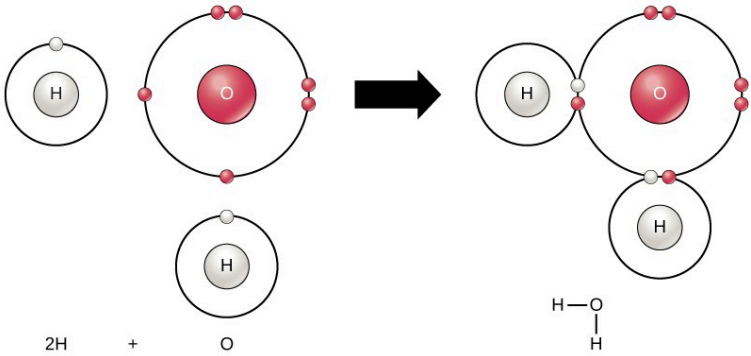


Figure 2.8 Two or more atoms may bond with each other to form a molecule. When two hydrogens and an oxygen share electrons via covalent bonds, a water molecule is formed.

Chemical reactions occur when two or more atoms bond together to form molecules or when bonded atoms are broken apart. The substances used in the beginning of a chemical reaction are called the **reactants** and the substances found at the end of the reaction are known as the **products.** An arrow is typically drawn between the reactants and products to indicate the direction of the chemical reaction.

Most chemical reactions can go in either direction. For the creation of the water molecule shown above, the chemical equation would be:

$$2\ H_2 + O_2 \rightarrow 2\ H_2O$$

This is an example of a **balanced chemical equation**, wherein the number of atoms of each element is the same on each side of the equation.

2.2.2. Ions and Ionic Bonds

Some atoms are more stable when they gain or lose an electrons and form ions. This fills their outermost electron shell and makes them more stable. Because the number of electrons does not equal the number of protons, each ion has a net charge. **Cations** are positive ions that are formed by losing electrons. **Anions** are negative ions that are formed by gaining electrons.

Certain ions, such as sodium, potassium, and calcium, are referred to in physiology as **electrolytes**. These ions are necessary for nerve impulse conduction, muscle contractions and water balance. Many sports drinks and dietary supplements provide these ions to replace those lost from the body via sweating during exercise.

Movement of electrons from one atom or molecule to another is referred to as **electron transfer** or as a **redox**

reaction. As **Figure 2.9** illustrates, sodium (Na) only has one electron in its outer electron shell. If sodium loses an electron, it now has 11 protons and only 10 electrons, making it a sodium cation with an overall charge of +1. Chlorine (Cl) has seven electrons in its outer shell. If it gains an electron, it now has 17 protons and 18 electrons, making it a chloride anion with an overall charge of -1. Both ions now satisfy the octet rule and have complete outermost shells.

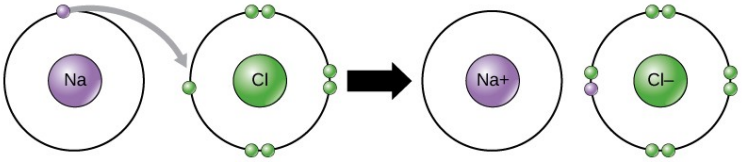


Figure 2.9 In the formation of an ionic compound, an electron is transferred from one atom to another, forming two oppositely charged ions, which are then attracted to each other.

An **ionic bond** is the electrical attraction that forms between ions with opposite charges. For example, positively charged sodium ions and negatively charged chloride ions bond together to make crystals of sodium chloride, or table salt, creating a crystalline molecule with zero net charge.

2.2.3 Covalent Bonds

Another way the octet rule can be satisfied is by the sharing

of electrons between atoms to form **covalent bonds**. One, two, or three pairs of electrons may be shared, making single, double, and triple bonds, respectively.

The formation of water molecules provides an example of covalent bonding (**Figure 2.8**). To completely fill the outer shell of oxygen, which has six electrons in its outer shell, two electrons (one from each hydrogen atom) are needed. The electrons are shared between the two elements to fill the outer shell of each, making both elements more stable.

Polar Covalent Bonds

Although atoms share electrons in covalent bonds, they do not always share the electrons equally. Atoms have different **electronegativities**, or attraction for electrons (**Figure 2.10**). When a covalent bond is formed between two atoms with different electronegativities, the shared electrons will spend more time around the nucleus of the atom with the higher electronegativity and less time around the nucleus with lower electronegativity. Since electrons are negatively charged, the atom that gets more time with the electron acquires a slightly negative charge ($\delta-$). The atom with

H 2.1						
Li 1.0	Be 1.6	B 2.0	C 2.5	N 3.0	O 3.5	F 4.0
Na 0.9	Mg 1.2	Al 1.5	Si 1.8	P 2.1	S 2.5	Cl 3.0

Figure 2.10 Electronegativity values for selected elements.

lower electronegativity gets less time with the electron and acquires a slightly positive charge ($\delta+$).

The type of covalent bond that forms between two atoms with different electronegativities is called a **polar covalent bond**. Molecules with polar covalent bonds are called **polar molecules**, due to the separation of charges across the molecule. For example, water is a polar molecule, since oxygen has an electronegativity of 3.5 and hydrogen has an electronegativity of 2.1. The oxygen atom in a water molecule attracts the shared electrons more and acquires a partial negative charge, while the hydrogen atoms attract the shared electrons less and acquire a partial positive charge (**Figure 2.11**). Many of the important properties of water result from its polarity.

Nonpolar Covalent Bonds

Nonpolar covalent bonds form between two atoms of the same element or between different elements that share electrons equally. For example, molecular oxygen (O_2) is nonpolar because the electrons will be equally distributed between the two oxygen atoms.

Another example of a nonpolar covalent bond is methane (CH_4), also shown in **Figure 2.11**. Carbon and hydrogen have similar electronegativity values. Therefore, these elements share electrons equally, creating a nonpolar covalent molecule. Some molecules are nonpolar due to

symmetry, as seen in the carbon dioxide molecule in **Figure 2.11**.

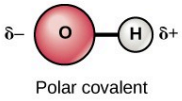
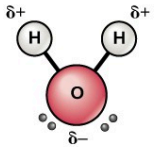
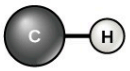
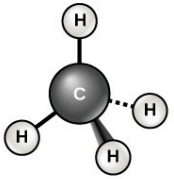
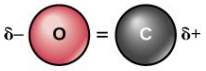
	Bond type	Molecular shape	Molecular type
Water	 Polar covalent	 Bent	Polar
Methane	 Nonpolar covalent	 Tetrahedral	Nonpolar
Carbon dioxide	 Polar covalent	Linear	Nonpolar

Figure 2.11 Whether a molecule is polar or nonpolar depends both on electronegativity values and molecular shape. Both water and carbon dioxide have polar covalent bonds, but carbon dioxide is linear, so the partial charges on the molecule cancel each other out.

2.2.4 Hydrogen Bonds and Van Der Waals Interactions

As described above, covalent and ionic bonds occur

between atoms to form molecules. Other types of interactions occur between molecules and are therefore called intermolecular forces. Two examples of weak attractions that occur frequently between molecules are hydrogen bonds and van der Waals interactions. Without these two types of attractions, life as we know it would not exist.

Hydrogen bonds are weak interactions between two polar molecules or between partially charged parts of molecules. The $\delta+$ of the hydrogen from one molecule is attracted to the $\delta-$ charge on the more electronegative atoms (usually oxygen or nitrogen) of another molecule. Hydrogen bonds can also occur between different parts of the same molecule. Individual hydrogen bonds are weak and easily broken; however, they occur in very large numbers in water and in organic polymers, forming very strong cumulative interactions. Hydrogen bonds between water molecules provide many of the critical, life-sustaining properties of water and also stabilize the structures of proteins and DNA.

Like hydrogen bonds, **van der Waals interactions** are weak attractions or interactions between molecules. Van der Waals attractions can occur between any two or more molecules and are dependent on slight fluctuations of the electron densities, around an atom. For these attractions to happen, the molecules need to be very close to one another. Although weaker than hydrogen bonds, van der Waals

interactions are also additive and can be quite strong in great numbers.

Pharmaceutical Chemist

Pharmaceutical chemists are responsible for developing new drugs and for trying to determine the mode of action of drugs. Drugs can be found in nature or can be synthesized in the laboratory. In many cases, potential drugs found in nature are changed chemically in the laboratory to make them safer and more effective.

After the initial discovery or synthesis of a drug, the chemist develops the drug, perhaps by chemically altering it, testing to see if it is toxic, and designing methods for large-scale production. Next, the process of getting the drug approved for human use by the Food and Drug Administration

(FDA) begins. This involves a series of large-scale experiments using human subjects to make sure the drug is safe and effective. Approval often takes several years and requires the participation of physicians and chemists.

An example of a drug that was originally discovered in a living organism is Paclitaxel (Taxol), an anti-cancer drug used to treat breast cancer. This drug was discovered in the bark of the pacific yew tree. Another example is aspirin, which was originally isolated from willow tree bark. Both of these drugs are now produced synthetically. Finding drugs often means testing hundreds of samples of plants, fungi, and other forms of life to see if any biologically active compounds are found within them. Sometimes, traditional medicine can give modern medicine clues to where an active compound can be found. For example, the use of willow bark to make medicine has been known for thousands of years, dating back to ancient Egypt. It was not until the late 1800s,

however, that the aspirin molecule, known as acetylsalicylic acid, was purified and marketed for human use.

Occasionally, drugs developed for one use are found to have unforeseen effects that allow these drugs to be used in other, unrelated ways. For example, the drug minoxidil (Rogaine) was originally developed to treat high blood pressure. When tested on humans, it was noticed that individuals taking the drug would grow new hair. Eventually the drug was marketed to men and women with baldness to restore lost hair.

The career of the pharmaceutical chemist may involve detective work, experimentation, and drug development, all with the goal of making human beings healthier.

2.3 | Water, Acids & Bases

Learning Objectives

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

- Describe the properties of water that are critical to maintaining life.
- Explain why water is an excellent solvent.
- Provide examples of water's cohesive and adhesive properties.
- Discuss the role of acids, bases, and buffers in homeostasis.

Why do scientists spend time looking for water on other planets? It is because water is essential to life as we know it. Water is one of the more abundant molecules and the one most critical to life on Earth. Approximately 60–70 percent of the human body is made up of water. Without it, life as we know it simply would not exist.

The polarity of the water molecule and its resulting hydrogen bonding make water a unique substance with

special properties that 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 watery contents of the cell's cytoplasm. Special properties of water include its high heat capacity and heat of vaporization, its ability to dissolve polar molecules, its cohesive and adhesive properties, and its dissociation into ions that leads to the generation of pH. Understanding these characteristics of water helps to elucidate its importance in maintaining life.

2.3.1 The Properties of Water

The Polarity of Water

One of water's important properties is that it is composed of polar molecules. While there is no net charge to a water molecule, the slight positive charges on the hydrogen atoms and the slight negative charges on the oxygen atoms contribute to water's properties of attraction.

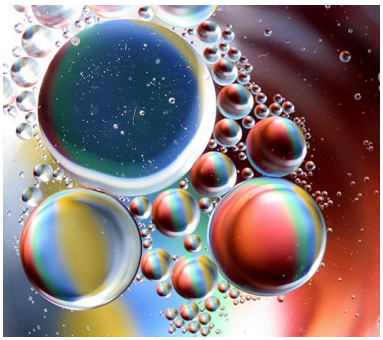


Figure 2.12 Oil and water do not mix. Since oil is nonpolar, it does not dissolve in water but forms droplets instead. (Credit: Gautam Dogra).

As a result of their polarity, water molecules form hydrogen bonds with each other. Water also attracts, or is attracted to, other polar molecules and ions. A polar substance that interacts readily with or dissolves in water is referred to as **hydrophilic** (hydro- = "water"; -philic = "loving"). In contrast, non-polar molecules, such as oils and fats, do not interact well with water, and separate from it rather than dissolving in it (**Figure 2.12**). These nonpolar compounds are called **hydrophobic** (hydro- = "water"; -phobic = "fearing").

Water's States: Gas, Liquid, and Solid

Its many hydrogen bonds cause water to have some unique chemical characteristics compared to other liquids. Since living things have a high water content, understanding these chemical features is key to understanding life. In liquid water, hydrogen bonds constantly form and break as the water molecules slide past each other. The bonds break due to the motion (kinetic energy) of the water molecules due to the heat contained in the system. As water is boiled, the higher kinetic energy of the water molecules causes the hydrogen bonds to break completely and allows water molecules to escape into the air as gas (steam or water vapor).

On the other hand, when water freezes, the water molecules form a crystalline structure that makes ice less dense than liquid water. Water is less dense as a solid because as water freezes the water molecules are pushed farther apart

by the hydrogen bonds. In most other substances, molecules pack more tightly as they freeze, making the solid more dense than the liquid. The lower density of ice causes it to float on liquid water (**Figure 2.13**). In lakes and ponds, ice will form on the surface of the water creating an insulating barrier that protects the living organisms in the pond from freezing. Without this layer of insulating ice, organisms living in the pond would freeze in the solid block of ice and could not survive.

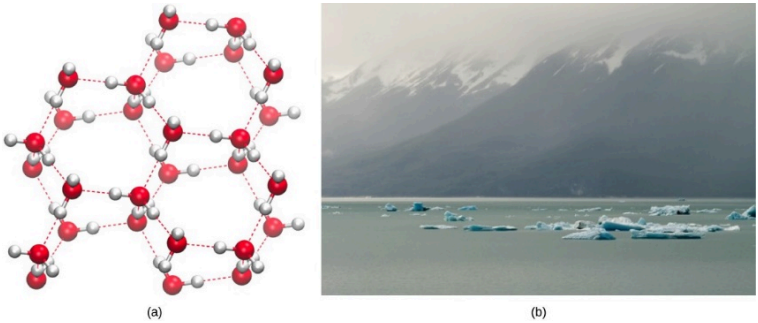


Figure 2.13 The (a) lattice structure of ice due to hydrogen bonding makes it less dense than liquid water, enabling it to (b) float on water. (Credit a: modification of work by Jane Whitney, image created using Visual Molecular Dynamics (VMD) software; b: modification of work by Carlos Ponte)

Water's High Heat Capacity

Water has the highest **specific heat capacity** of any liquid. Specific heat is defined as the amount of heat one gram of a substance must absorb or lose to change its temperature by one degree Celsius. Because of all of the hydrogen bonds

between water molecules, it takes water a long time to heat and long time to cool. In fact, the specific heat capacity of water is about five times more than that of sand. This explains why the land cools faster than the sea. Due to its high heat capacity, water is used by warm blooded animals to help maintain an even temperature.

Water's Heat of Vaporization

Water also has a high **heat of vaporization**, the amount of energy required to change one gram of a liquid substance to a gas. A considerable amount of heat energy is required to accomplish this change in water. As liquid water heats up, hydrogen bonding makes it difficult to separate the liquid water molecules from each other, in order for it to enter its gaseous phase (steam). As a result, water requires much more heat to boil than most other liquids. Eventually, as water reaches its boiling point, the heat is able to break the hydrogen bonds between the water molecules, and the kinetic energy (motion) between the water molecules allows them to escape from the liquid as a gas.

Even when below its boiling point, individual water molecules acquire enough energy from other water molecules that some can escape, in a process known as **evaporation**. Since the evaporation of water requires heat energy, it cools the environment where the evaporation is taking place. In many living organisms, including humans, the evaporation of

sweat allows the organism to cool so that homeostasis of body temperature can be maintained.

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, water is referred to as a **solvent**, a substance capable of dissolving other molecules. The substance that is dissolved in a liquid is called a **solute**. The mixture of a solute dissolved in a solvent is called a **solution**. If the solvent is water, the solution is called an **aqueous solution**.

Due to their charge, polar molecules and ions form hydrogen bonds with water and become surrounded by water molecules. The resulting **sphere of hydration**, or hydration shell, serves to keep the particles separated in the water. When ionic compounds are added to water, the individual ions react with the water molecules and their ionic bonds are disrupted in the process of **dissociation**. For example, when crystals of table salt (NaCl, or sodium chloride) are added to water, 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 (**Figure 2.14**).

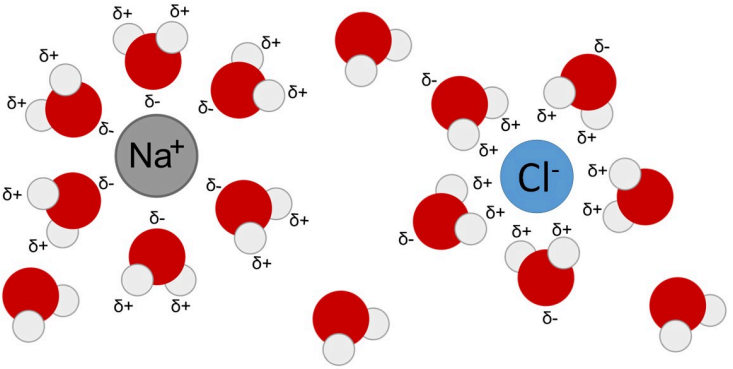


Figure 2.14 When table salt (NaCl) is mixed in water, spheres of hydration are formed around the ions.

Water's Cohesive and Adhesive Properties

Have you ever filled a glass of water to the very top and then slowly added a few more drops? Before it overflows, the water forms a dome-like shape above the rim of the glass. This water can stay above the glass because of the property of **cohesion**. In cohesion, water molecules are attracted to each other, keeping the molecules together at the water-air interface.

Cohesion allows for the development of **surface tension**, the capacity of a substance to withstand being ruptured when placed under tension or stress. This is also why water forms droplets when placed on a dry surface rather than being flattened out by gravity. It is even possible to "float" a

needle on top of a glass of water or for a water strider to stay afloat on the surface layer of water (**Figure 2.15**).

Figure 2.15 a. The weight of the needle pulls the surface of the water down while surface tension is pulling it up, suspending it on the surface of the water and keeping it from sinking. Notice the indentation in the water around the needle. b. Water's cohesive property allows this water strider to stay afloat. (credit: a. Cory Zanker; b. Tim Vickers)

Hydrogen bonding also leads to water's property of **adhesion**, or the attraction of water molecules to other molecules. Adhesion is observed when water "climbs" up the tube placed in a glass of water: notice that the water appears to be higher on the sides of the tube than in the middle. This is because the water molecules are attracted to the charged glass walls of the capillary and therefore adhere to it. This type of adhesion is called **capillary action** (**Figure 2.16**).

Cohesion and adhesion are important for the transport of water from the roots to the leaves in plants. As water molecules are evaporated from the surface of leaves, they tend to stay connected to water molecules below them, creating a "pull" up the water column. Ultimately, water is pulled into the roots, allowing the plant to receive the dissolved minerals they require from the soil.

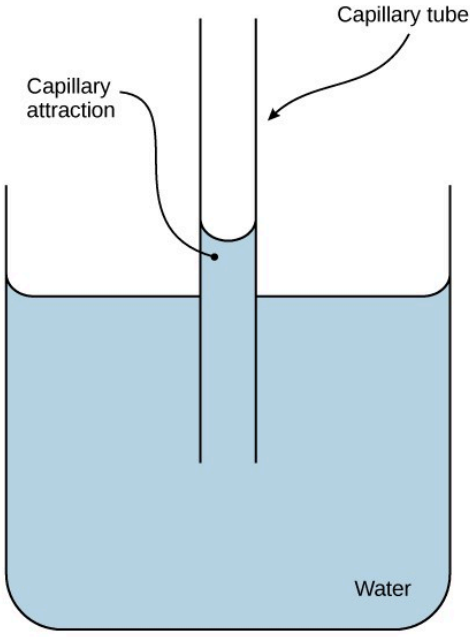


Figure 2.16 Capillary action in a glass tube is caused by the adhesion of water molecules to the charges on the glass. (credit: modification of work by Pearson-Scott Foresman, donated to the Wikimedia Foundation)

2.3.2 Acids and Bases

pH is a measure of the concentration of hydrogen ions $[H^+]$in a solution. Hydrogen ions are spontaneously generated in pure water by the dissociation (ionization) of a small percentage of water molecules into equal numbers of hydrogen ions (H^+) and hydroxide ions (OH^-) ions. While the hydroxide ions are kept in solution by hydrogen bonding

with other water molecules, the hydrogen ions are immediately attracted to water molecules, forming hydronium ions (H_30^+). Still, by convention, scientists refer to hydrogen ions and their concentration as if they were free in this state in liquid water.

$$H_2O \leftrightarrow H^+ + OH^-$$
$$H^+ + H_2O \leftrightarrow H_3O+$$

The concentration of hydrogen ions dissociating from pure water is 1×10^{-7} moles H^+ ions per liter of water. One **mole** of a substance is equal to 6.02×10^{23} particles of the substance. pH is calculated as the negative of the base 10 logarithm of the H^+ concentration. The $\log_{10}$ of 1×10^{-7} is -7.0, and the negative of this number yields a pH of 7.0, which is also known as neutral pH. The pH inside cells (6.8) and the pH of human blood (7.4) are both very close to neutral. Extremes in pH in either direction are usually considered inhospitable to life.

An **acid** is a substance that increases the H^+ concentration in a solution, usually by having one of its hydrogen atoms dissociate. A **base** provides either OH^- or other negatively charged ions that combine with hydrogen ions, reducing their concentration in the solution and thereby raising the pH. In cases where the base releases hydroxide ions, these ions bind to free hydrogen ions, generating new water molecules.

The stronger the acid, the more readily it donates H^+. For example, hydrochloric acid (HCl) completely dissociates into

hydrogen and chloride ions and is highly acidic, whereas the acids in tomato juice or vinegar do not completely dissociate and are considered weak acids. Conversely, strong bases are those substances that readily donate OH^- or take up hydrogen ions. Sodium hydroxide (NaOH) and many household cleaners are highly alkaline and give up OH^- rapidly when placed in water, the

reby raising the pH. An example of a weak basic solution is seawater, which has a pH near 8.0, close enough to neutral pH that marine organisms adapted to this saline environment are able to thrive in it.

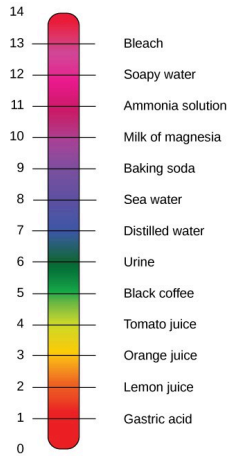


Figure 2.17 The pH scale measures the concentration of hydrogen ions (H+) in a solution. (credit: modification of work by Edward Stevens)

The **pH scale** ranges from 0 to 14 (**Figure 2.17**). Anything below 7.0 is acidic, and anything above 7.0 is basic, or alkaline. Since the pH scale is a negative logarithmic scale, a ten-fold change in $[H^+]$ results in a change of one in pH in the opposite direction. For example, increasing the $[H^+]$ from 1×10^{-4} to 1×10^{-3} decreases the pH from pH 4 to pH 3.

So how can organisms that require a near-neutral pH ingest acidic and basic substances (a human drinking orange juice,

for example) and survive? Buffers are the key. **Buffers** readily absorb excess H^+ or OH^-, keeping the pH of the body in homeostasis. For example, the buffer maintaining the pH of human blood is a mixture of carbonic acid (H_2CO_3), bicarbonate ion (HCO^-), and carbon dioxide (CO_2). When bicarbonate ions combine with free hydrogen ions and become carbonic acid, hydrogen ions are removed, moderating pH increases. Excess carbonic acid can be converted to carbon dioxide gas and exhaled through the lungs. This prevents too many free hydrogen ions from building up in the blood and dangerously reducing the blood's pH. Conversely, if too much OH– is introduced into the system, carbonic acid will combine with it to create bicarbonate, lowering the pH. Without this buffer system, the body's pH would fluctuate enough to put survival in jeopardy (**Figure 2.18**).

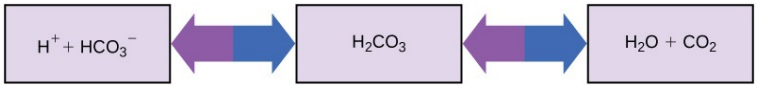


Figure 2.18 This diagram shows the body's buffering of blood pH levels. The blue arrows show the raising of pH as more CO2 is made. The purple arrows indicate the lowering of pH as more bicarbonate is created.

Other examples of buffers are antacids used to combat excess stomach acid. Many of these over-the-counter medications work in the same way as blood buffers, usually with at least

one ion capable of absorbing hydrogen and moderating pH, bringing relief to those that suffer "heartburn" after eating.

2.4 | Energy and Life

Learning Objectives

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

- Explain the concept of metabolism.
- Explain the difference between potential and kinetic energy.
- Explain how the first two laws of thermodynamics relate to living organisms.

2.4.1 Metabolism

Virtually every task performed by living organisms requires energy. In fact, the living cells of every organism constantly use energy. Just as energy is required to both build and demolish a building, energy is required for both the synthesis

and breakdown of molecules. Other cellular process that require energy include transport of signaling molecules, such as hormones and neurotransmitters; ingesting and breaking down pathogens, such as bacteria and viruses; importing nutrients and exporting waste; and many others.

The cellular processes listed above require a steady supply of energy. From where, and in what form, does this energy come? How do living cells obtain energy, and how do they use it? This section will discuss different forms of energy and the physical laws that govern energy transfer.

Cellular processes such as building and breaking down complex molecules occur through series of chemical reactions. All of the chemical reactions that take place inside cells, including those that use energy and those that release energy, are the cell's **metabolism**. Chemical reactions that require energy to synthesize complex molecules from simpler ones are called **anabolic** reactions, and chemical reactions that release energy as complex molecules are broken down are called **catabolic** reactions.

2.4.2. Potential vs. Kinetic Energy

Energy is defined as the capacity to do work. When an object is in motion, there is energy associated with that object because moving objects are capable of enacting a change, or doing work. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects.

However, a wrecking ball that is not in motion is incapable of performing work. Energy associated with objects in motion is called **kinetic energy**. A speeding bullet, a walking person, the rapid movement of molecules in the air, and electromagnetic radiation all have kinetic energy.

Now what if that same motionless wrecking ball is lifted two stories above a car with a crane? If the suspended wrecking ball is unmoving, is there energy associated with it? The answer is yes. The suspended wrecking ball has energy that results from the fact that there is the *potential* for the wrecking ball to do work. This type of energy is called **potential energy**. Another example of potential energy is the energy of water held behind a dam (**Figure 2.19**).

Figure 2.19 Water behind a dam has potential energy. Moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (Credit "dam": modification of work by "Pascal"/Flickr; credit "waterfall": modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of

matter (such as a wrecking ball being held up), but also with the structure of matter. A spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut.

Living cells rely heavily on structural potential energy. On a chemical level, the bonds that hold the atoms of molecules together have potential energy. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, because these bonds can release energy when broken. This type of potential energy is called **chemical energy** (**Figure 2.20**). Chemical energy provides cells with energy by breaking the molecular bonds within fuel molecules.

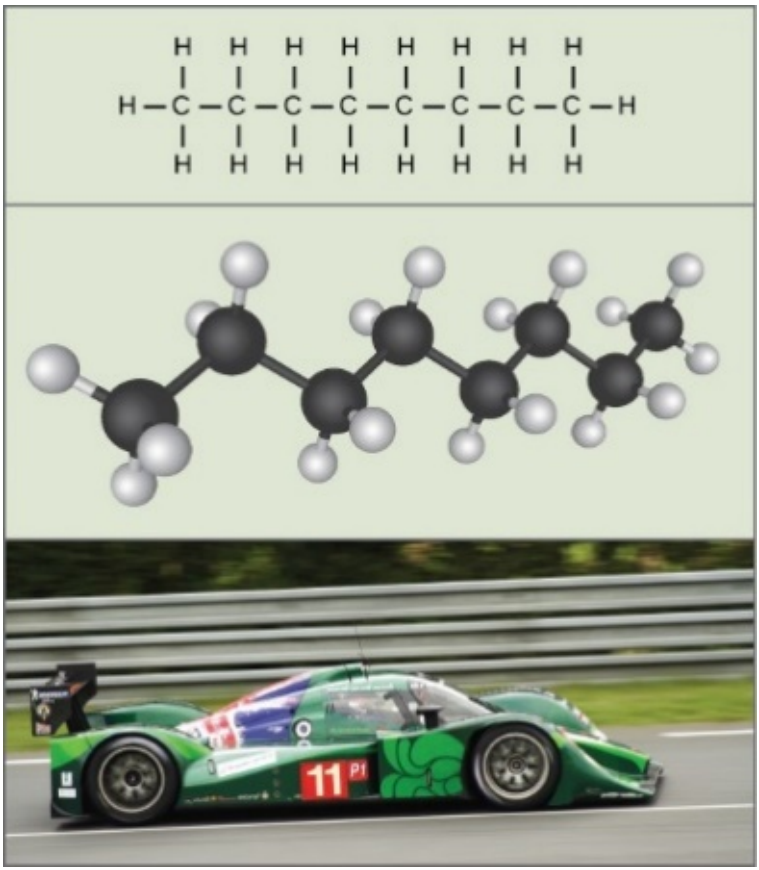


Figure 2.20 The molecules in gasoline (octane, the chemical formula shown) contain chemical energy within the chemical bonds. This energy is transformed into kinetic energy that allows a car to race on a racetrack. (credit "car": modification of work by Russell Trow)

2.4.3 The Laws of Thermodynamics

Thermodynamics refers to the study of energy and energy transfer. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe. The first and second laws of thermodynamics are relevant to biological systems and how they convert and exchange energy with their surroundings.

The First Law of Thermodynamics

The first law of thermodynamics states that the total amount of energy in the universe is constant. In other words, energy cannot be created or destroyed. However, energy may be transferred from one form to another. Transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants convert energy of sunlight into chemical energy stored within organic molecules. Some examples of energy transformations are shown in **Figure 2.21**.

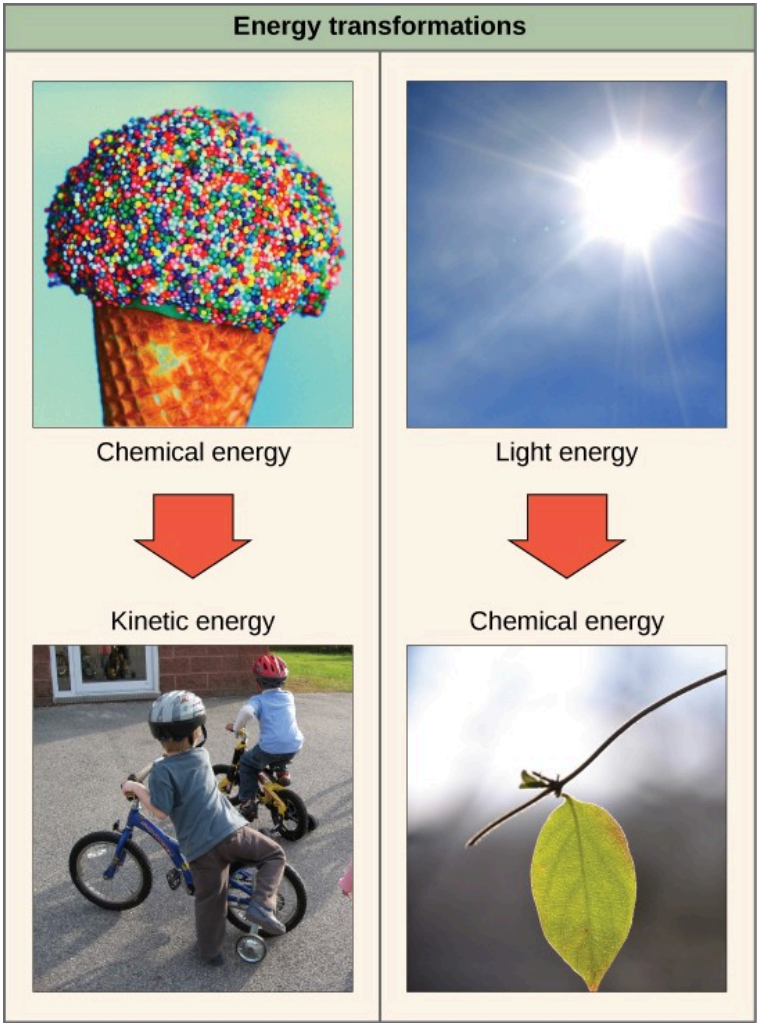


Figure 2.21 Shown are two examples of energy being transferred from one system to another and transformed from one form to another. Humans can convert the chemical energy in food, like this ice cream cone, into kinetic energy (the energy of movement to ride a bicycle). Plants can convert electromagnetic radiation (light energy) from the sun into chemical energy. (Credit "ice cream": modification of

work by D. Sharon Pruitt; credit "kids on bikes": modification of work by Michelle Riggen-Ransom; credit "leaf": modification of work by Cory Zanker)

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work. Living cells have evolved to meet this challenge very well. Chemical energy stored within organic molecules such as sugars and fats is transformed through a series of cellular chemical reactions into energy within molecules of ATP. Energy in ATP molecules is easily accessible to do work.

The Second Law of Thermodynamics

The second law of thermodynamics states that the disorder, or **entropy**, in the universe is always increasing. None of the energy transfers and transformations in the universe is completely efficient. In every energy transfer, some amount of energy is lost in a form that is unusable. The more energy that is lost by a system to its surroundings, the less ordered and more random the system is.

In most cases, the energy is lost in the form of heat energy. Thermodynamically, **heat energy** is the energy transferred from one system to another that is not doing work. For example, when an airplane flies through the air, some of the energy of the flying plane is lost as heat energy due to friction with the surrounding air. This friction actually heats the air

by temporarily increasing the speed of air molecules. Likewise, during cellular metabolic reactions, some energy is lost as heat energy. (This is good for warm-blooded creatures like us, because heat energy helps to maintain our body temperature.) In another example, as molecules at a high concentration in one place diffuse and spread out, entropy increases (**Figure 2.22**).

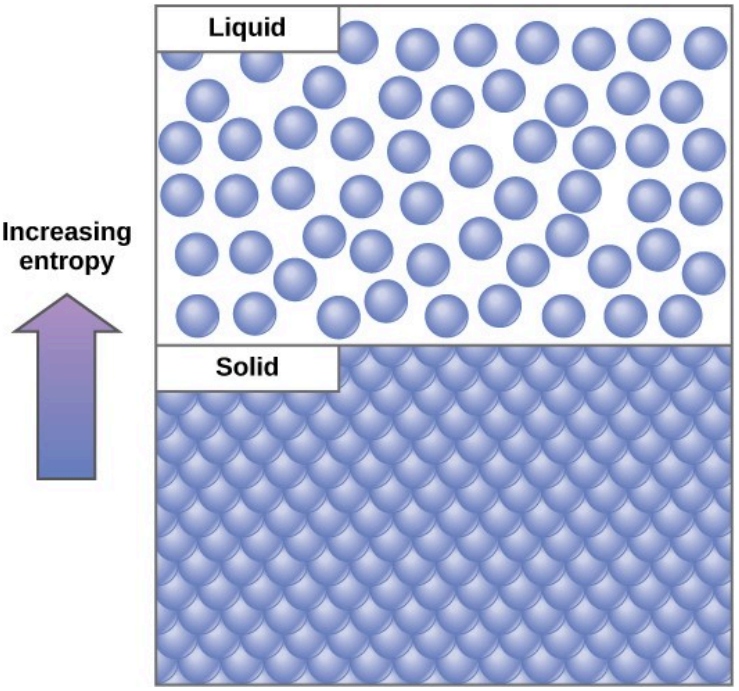


Figure 2.22 Entropy is a measure of randomness or disorder in a system. Gases have higher entropy than liquids, and liquids have higher entropy than solids.

Since living things are highly ordered, they require a constant input of energy. Essentially, living things are in a continuous uphill battle against this constant increase in universal entropy.

2.5 | Carbon

Learning Objectives

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

- Explain why carbon is important for life.
- Describe the role of functional groups in biological molecules.

Cells are made of many complex **organic** (carbon-containing) **molecules**, such as proteins and carbohydrates, which are especially important for life. The fundamental component of all of these macromolecules is carbon. Carbon atoms can form covalent bonds with up to four different atoms, making them ideal to form the "backbone" of macromolecules.

2.5.1 Hydrocarbons

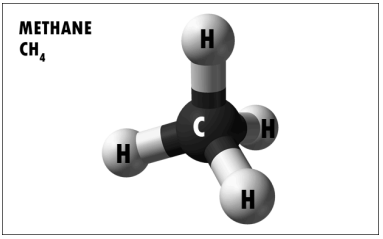


Figure 2.23 Methane, CH4, is the simplest hydrocarbon.

Hydrocarbons are organic molecules consisting entirely of carbon and hydrogen. We often use hydrocarbons as fuels—like the propane in a gas grill or the butane in a lighter. The covalent bonds between the atoms in hydrocarbons store a great amount of energy, which is released when these molecules are burned (oxidized). Methane, an excellent fuel, is the simplest hydrocarbon molecule (**Figure 2.23**). Hydrocarbons may exist as linear carbon chains, carbon rings, or combinations of both. Furthermore, individual carbon-to- carbon bonds may be single, double, or triple covalent bonds.

2.5.2 Functional Groups

Functional groups are groups of atoms that occur commonly within molecules and confer specific chemical properties to those molecules. They are found attached to the carbon "backbone" of macromolecules. Each of the four types of macromolecules—proteins, lipids, carbohydrates, and nucleic acids—has its own characteristic set of

functional groups that contributes greatly to its differing chemical properties and its function in living organisms.

Some of the important functional groups in biological molecules are shown in **Figure 2.24**. Functional groups are often classified as polar or non-polar, since that determines whether they are hydrophobic or hydrophilic. For example, non-polar methyl groups are hydrophobic and polar hydroxyl groups are hydrophilic. Functional groups can also be classified as acidic or basic, depending on whether they release or accept H^+ in solution. Releasing H^+ results in a negatively charged functional group, such as phosphate groups or carboxyl groups. Accepting H^+ results in a positively charged functional group, such as amino groups.

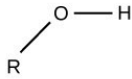
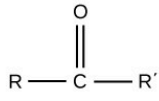
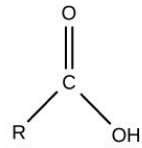
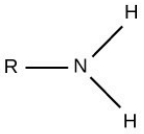
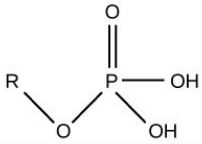
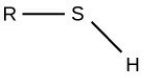
Functional Group	Structure	Properties
Hydroxyl		Polar
Methyl		Nonpolar
Carbonyl		Polar
Carboxyl		Charged, ionizes to release H^+. Since carboxyl groups can release H^+ ions into solution, they are considered acidic.
Amino		Charged, accepts H^+ to form NH_3^+. Since amino groups can remove H^+ from solution, they are considered basic.
Phosphate		Charged, ionizes to release H^+. Since phosphate groups can release H^+ ions into solution, they are considered acidic.
Sulfhydryl		Polar

Figure 2.24 The functional groups shown here are found in many different biological molecules. Properties of the functional groups affect the properties of the macromolecules that they are part of. For example, since DNA contains phosphate groups, it is both acidic and negatively charged.

CHAPTER 3. AMINO ACIDS & PROTEINS

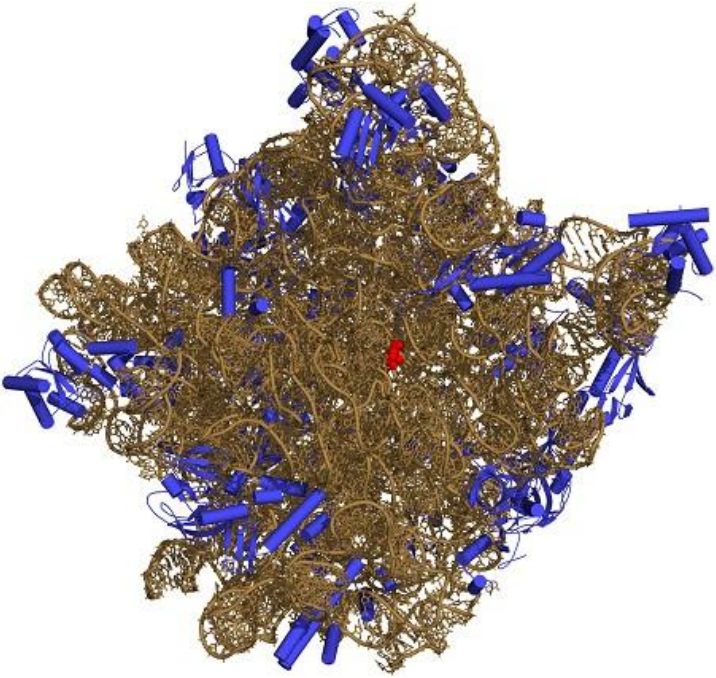


Figure 4.1 Atomic structure of the large subunit of a ribosome from *Haloarcula marismortui*. Ribosomes are large composites of nucleic acid (RNA) and proteins. In this figure, proteins are colored in blue and RNA in ochre. (Credit: by Yikrazuul. Data were taken from PDB 3CC2, rendered with PyMOL.)

Chapter Outline

- 3.1 Biological Macromolecules
- 3.2 Types and Functions of Proteins
- 3.3 Amino Acids
- 3.4 Protein Structure

Introduction

We will now begin our tour of the four major types of macromolecules found in living organisms. The first type of molecule, proteins, are molecular machines that do the work of cells. They have a huge variety of structure and function. But before we delve into how protein structure relates to protein function, we first have to discuss macromolecules.

3.1 | Biological Macromolecules

Learning Objectives

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

- Name the four major classes of biological macromolecules.
- Understand the synthesis of macromolecules.
- Describe dehydration synthesis and hydrolysis reactions.

Biological macromolecules are large molecules, necessary for life, that are built from smaller organic molecules. There are four major classes of biological macromolecules: carbohydrates, lipids, proteins, and nucleic acids. Each is an important cell component and performs a wide array of functions. Combined, these molecules make up the majority of a cell's dry mass (recall that water makes up the majority of its complete mass). Biological macromolecules are organic,

meaning they contain carbon. In addition, they may contain hydrogen, oxygen, nitrogen, and additional minor elements.

3.1.1 Dehydration Synthesis Reactions

Most macromolecules are made from single subunits, or building blocks, called **monomers**. The monomers combine with each other using covalent bonds to form larger molecules known as **polymers**. In doing so, monomers release water molecules as byproducts. This type of reaction is known as **dehydration synthesis** (also known as **condensation**), which means "to make while losing water."

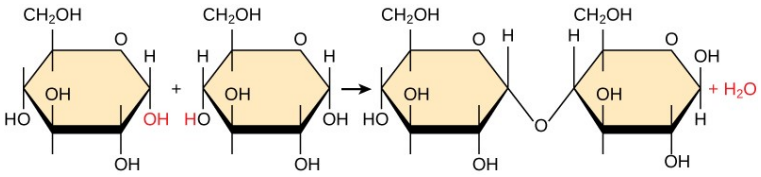


Figure 3.2 In the dehydration synthesis reaction shown above, two molecules of glucose are linked together to form the disaccharide maltose. In the process, a water molecule is formed.

In a dehydration synthesis reaction, the hydrogen of one monomer combines with the hydroxyl group of another monomer, releasing a molecule of water (**Figure 3.2**). 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.

3.1.2 Hydrolysis Reactions

Polymers are broken down into monomers in a process known as **hydrolysis**, which means "to split water." (**Figure 3.3**). During these reactions, the polymer is broken into two components: one part gains a hydrogen atom (H+) and the other gains a hydroxyl molecule (OH–) from a split water molecule.

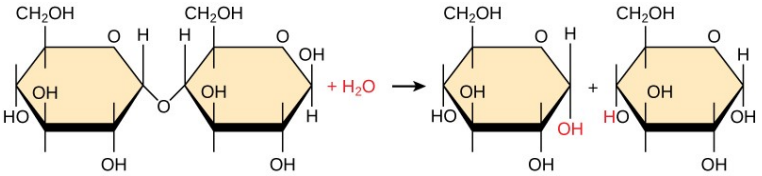


Figure 3.3 In the hydrolysis reaction shown here, the disaccharide maltose is broken down to form two glucose monomers with the addition of a water molecule. Note that this reaction is the reverse of the synthesis reaction shown in Figure 3.2.

Dehydration and hydrolysis reactions are catalyzed, or "sped up," by specific enzymes; dehydration reactions involve the formation of new bonds, requiring energy, while hydrolysis reactions break bonds and release energy. These reactions are similar for most macromolecules, but each monomer and polymer reaction is specific for its class. For example, in our bodies, food is hydrolyzed, or broken down, into smaller molecules by catalytic enzymes in the digestive system. This allows for easy absorption of nutrients by cells in the intestine. Each macromolecule is broken down by a specific enzyme. For instance, carbohydrates are broken down by amylase, sucrase, lactase, or maltase. Proteins are broken down by the enzymes pepsin and peptidase, and by hydrochloric acid. Lipids are broken down by lipases. Breakdown of these macromolecules provides energy for cellular activities.

3.2 | Types and Functions of Proteins

Learning Objectives

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

- Describe the functions proteins perform in the cell and in tissues.

Proteins are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules. They are all, however, polymers of amino acids, arranged in a linear sequence. Proteins may be structural, regulatory, contractile, or protective; they may serve in transport, storage, or membranes; or they may be toxins or enzymes. Each cell in a living system may contain thousands of proteins, each with a unique function. Their structures, like their functions, vary greatly.

Structural proteins make up the cytoskeleton inside cells and the scaffold outside of cells. They include the keratin of our skin and the collagen of our connective tissue. Contractile proteins include actin and myosin, which allow muscles to contract. Antibodies that help mount an immune response are proteins, as is hemoglobin, which transports oxygen in our blood. Cell membranes contain many proteins, including receptors, channels, and pumps, and many of the signaling molecules that bind to receptors, such as hormones, are proteins. Enzymes are proteins that catalyze biochemical

reactions. The primary types and functions of proteins are listed in **Table 3.1**.

Table 3.1 Protein types and functions.

Type	Examples	Functions
Digestive Enzymes	Amylase, lipase, pepsin, trypsin	Help in digestion of food by catabolizing nutrients into monomeric units
Transport	Hemoglobin, albumin	Carry substances in the blood or lymph throughout the body
Structural	Actin, tubulin, keratin	Construct different structures, like the cytoskeleton
Hormones	Insulin, thyroxine	Coordinate the activity of different body systems
Defense	Immunoglobulins	Protect the body from foreign pathogens
Contractile	Actin, myosin	Effect muscle contraction
Storage	Legume storage proteins, egg white (albumin)	Provide nourishment in early development of the embryo and the seedling

Reflecting their diverse functions, proteins have very diverse shapes and sizes. Some proteins, such as hemoglobin, are globular in shape whereas others, such as collagen, are fibrous (**Figure 3.4**). The shape of each protein is critical to its function, and this shape is maintained by many different types

of chemical bonds. Changes in temperature, pH, and exposure to chemicals may lead to permanent changes in the shape of the protein, leading to loss of function, known as **denaturation**.

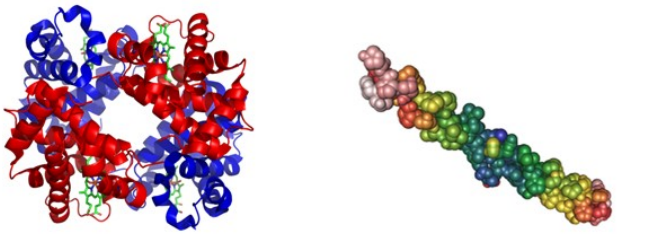


Figure 3.4 The diverse structures of proteins is responsible for their diverse functions. (Credit: A. Hemoglobin protein by Zephyris at the English language Wikipedia, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=2300973 B. Collagen protein by Nevit Dilmen [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], via Wikimedia Commons.)

3.3 | Amino Acids

Learning Objectives

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

- Discuss the relationship between amino acids and proteins.
- Describe the structure of an amino acid.
- Understand the peptide bond.

Amino acids are the monomers that make up proteins. All proteins are made up of different arrangements of the same 20 amino acids. Each amino acid has the same fundamental structure, which consists of a central carbon atom bonded to an amino group (NH2), a carboxyl group (COOH), a hydrogen atom, and a variable **"R" group** (**Figure** 3.5). The name "amino acid" is derived from the presence of the amino group and the acidic carboxyl group.

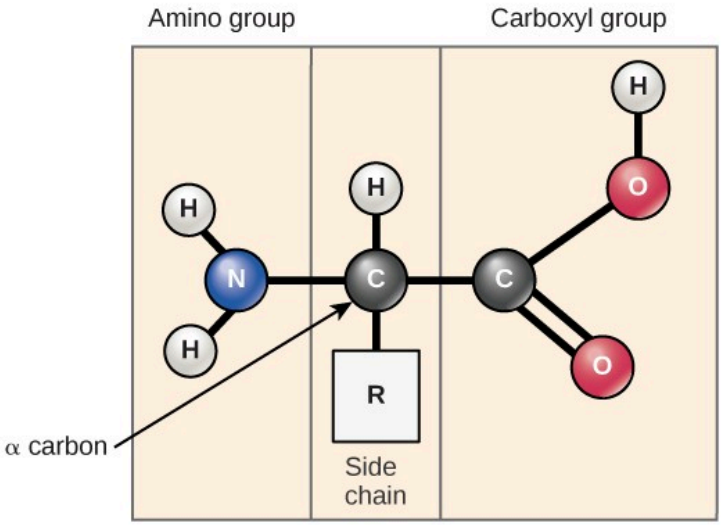


Figure 3.5 Amino acids have a central asymmetric carbon to which an amino group, a carboxyl group, a hydrogen atom, and a side chain (R group) are attached.

The same 20 common amino acids are present in proteins from all species of life. Ten of these are considered essential amino acids in humans because the human body cannot produce them and they must be obtained from the diet. Each amino acid differs only in the R group (or side chain). The chemical nature of the R group determines the nature of the amino acid (that is, whether it is acidic, basic, polar, or nonpolar). For example, amino acids such as valine, methionine, and alanine are nonpolar or hydrophobic in nature, while amino acids such as serine, threonine, and cysteine are polar and have hydrophilic side chains. The side chains of lysine and arginine are positively charged, while the

side chains of aspartate and glutamate are negatively charged. (**Figure 3.6**).

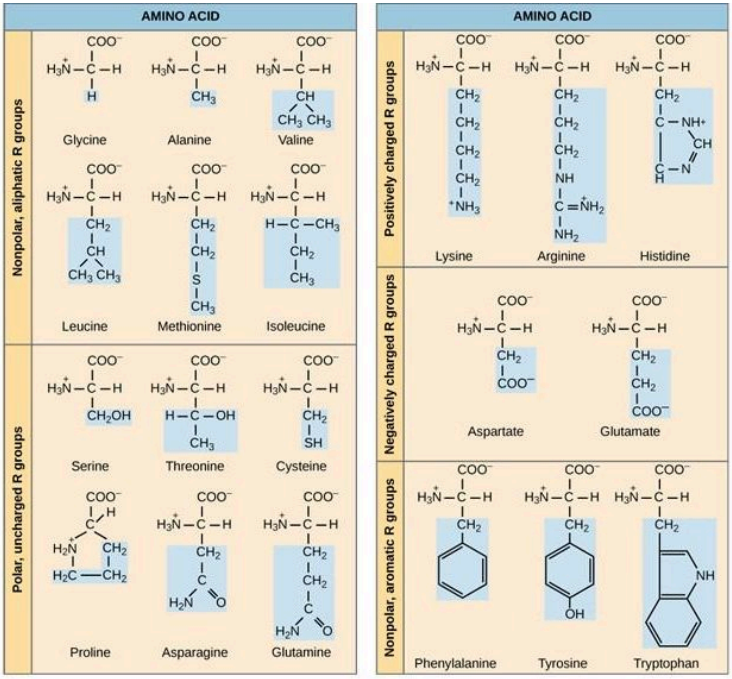


Figure 3.6 There are 20 amino acids commonly found in proteins, each with a different R group (side chain) that determines its chemical nature.

Amino acids are linked together into linear chains called **polypeptides**. While the terms polypeptide and protein are sometimes used interchangeably, a polypeptide is technically a polymer of amino acids, whereas the term protein is a polypeptide or polypeptides that have a distinct shape and

a unique function. The sequence and the number of amino acids determine the protein's shape, size, and function. After protein synthesis (translation), most proteins are modified. Only after these modifications is the protein completely functional.

Each polypeptide has an N terminal, with a free amino group, and a C terminal, with a free carboxyl group. Amino acids are attached to other amino acids by covalent bonds, known as **peptide bonds**, which are formed by dehydration synthesis reactions. The carboxyl group of one amino acid and the amino group of the incoming amino acid combine, releasing a molecule of water and forming a peptide bond (**Figure 3.7**). Additional amino acids are always added to the C terminus until the chain is complete. The resulting polypeptide chain has a **peptide backbone** (or carbon-nitrogen backbone), that is identical for all proteins, with the variable R groups extending off of the backbone. Polypeptides differ from each other only in the order of the R groups.

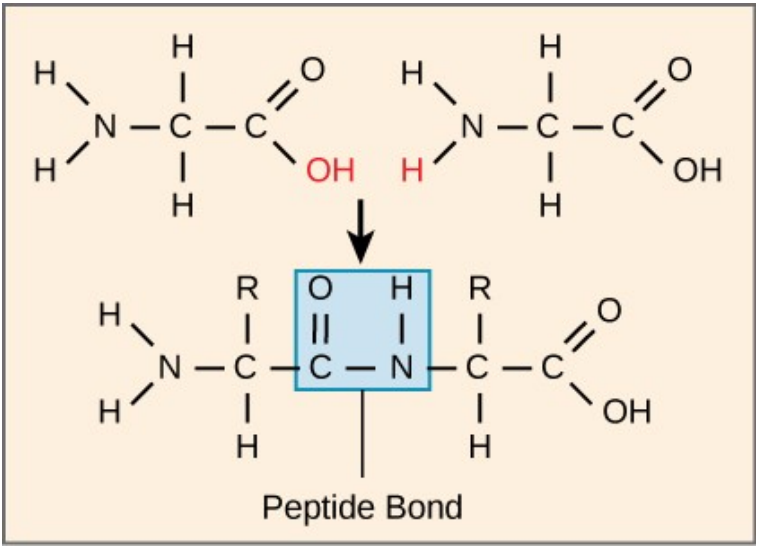


Figure 3.7 Peptide bond formation is a dehydration synthesis reaction. The carboxyl group of one amino acid is linked to the amino group of the incoming amino acid. In the process, a molecule of water is released.

evolution CONNECTION

The Evolutionary Significance of Cytochrome c Cytochrome c is a protein that is important for cellular respiration. Sequencing the cytochrome c protein from different species has shown that there is a considerable amount of

amino acid sequence homology among different species. Evolutionary kinship can be assessed by measuring the similarities or differences among various species' DNA or protein sequences.Human cytochrome c contains 104 amino acids. 37 of these amino acids appear in the same position in samples of cytochrome c from a variety of different species. Human and chimpanzee cytochrome c is identical and human and rhesus monkey cytochrome c differ in only one amino acid. Surprisingly, human and yeast cytochrome c also differ in just one amino acid.

3.4 | Protein Structure

Learning Objectives

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

- Explain the four levels of protein organization.
- Describe the ways in which protein structure and function are linked.

The structure of a protein is critical to its function. For example, an enzyme can bind to a specific substrate at a site known as the active site. If this active site is altered because of local changes or changes in overall protein structure, the enzyme may be unable to bind to the substrate. To understand how the protein gets its final shape or conformation, we need to understand the four levels of protein structure: primary, secondary, tertiary, and quaternary.

3.4.1 Primary Structure

The unique sequence of amino acids in a polypeptide chain is its **primary structure**. For example, the pancreatic hormone insulin has two polypeptide chains, A and B, which are linked together by disulfide bonds. The primary structure of each chain is indicated by three-letter abbreviations that represent the names and order of the amino acids. The N terminal amino acid of the A chain is glycine, whereas the C terminal amino acid is asparagine (**Figure 3.8**). The sequences of amino acids in the A and B chains are unique to insulin.

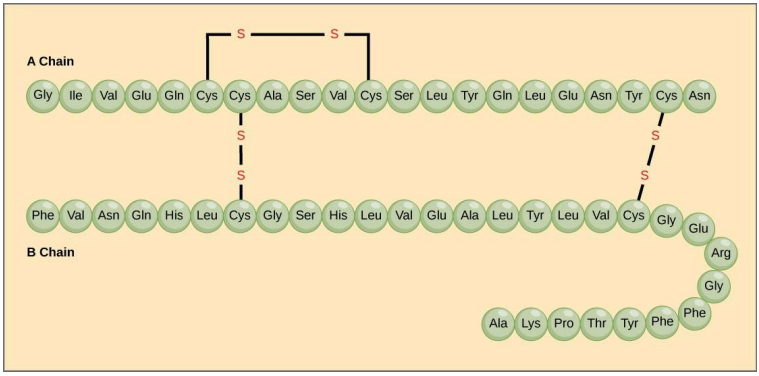
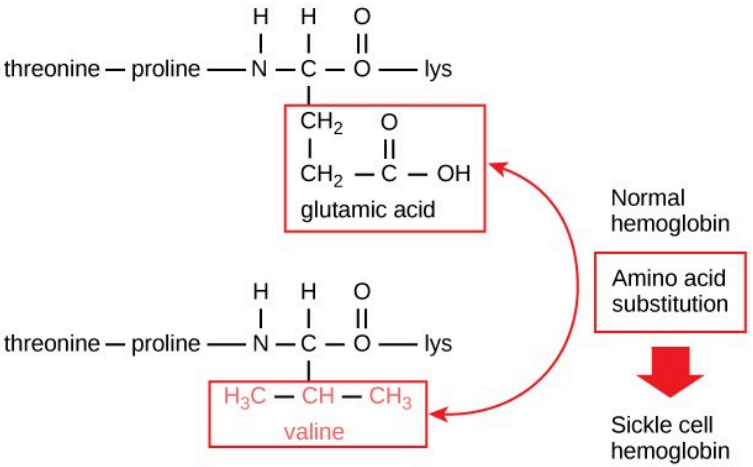


Figure 3.8 Bovine serum insulin is a protein hormone made of two peptide chains, A (21 amino acids long) and B (30 amino acids long).

The amino acid cysteine (cys) has a sulfhydryl (SH) group as a side chain. Two sulfhydryl groups can react in the presence of oxygen to form a disulfide (S-S) bond. Two disulfide bonds connect the A and B chains together, and a third helps the A chain fold into the correct shape.

The unique primary sequence for every protein is determined by the gene that encodes the protein. Even a small change in a gene can lead to a different amino acid being added to the growing polypeptide chain. For example, in the human genetic disease sickle cell anemia, the hemoglobin β chain (a small portion of which is shown in **Figure 3.9A**) has a single amino acid substitution (valine for glutamic acid). This change of one amino acid in the chain causes hemoglobin molecules to form long fibers that distort red blood cells into a crescent or

"sickle" shape, which clogs arteries and leads to serious health problems such as breathlessness, dizziness, headaches, and abdominal pain (**Figure 3.9B**).



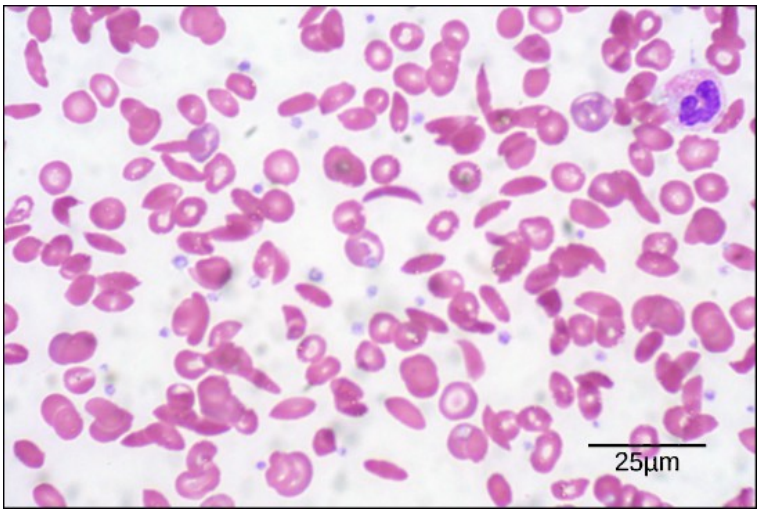


Figure 3.9 A. A single amino acid substitution in the beta chain of hemoglobin leads to sickle cell anemia. B. In this blood smear, visualized at 535x magnification using bright field microscopy, sickle cells are crescent shaped, while normal cells are disc-shaped. (Credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

3.4.2 Secondary Structure

The local folding of the polypeptide in some regions gives rise to the **secondary structure** of the protein. The most common are the ***α*-helix** and ***β*-pleated sheet** structures (**Figure 3.10**). Both structures are formed by hydrogen bonds forming between parts of the peptide backbone of the polypeptide. Specifically, the oxygen atom in the carbonyl group in one amino acid interacts with another amino acid that is four amino acids farther along the chain.

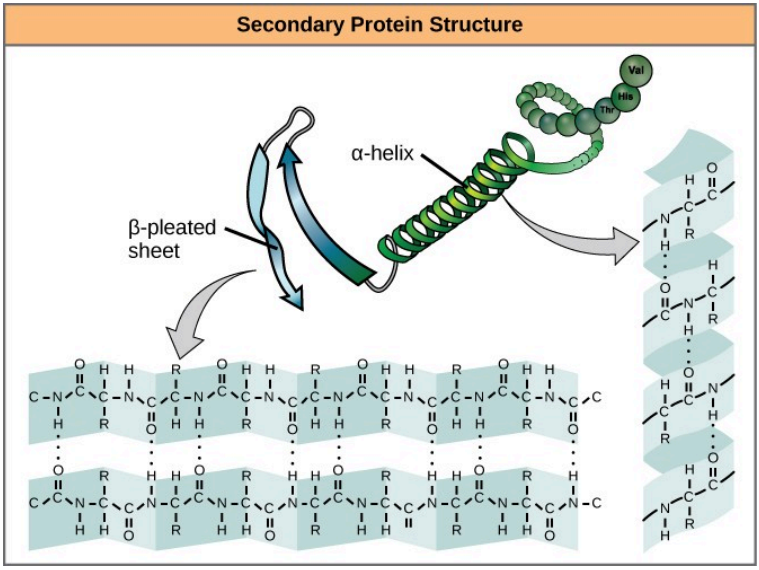


Figure 3.10 The α-helix and β-pleated sheet are secondary structures of proteins that form because of hydrogen bonding between carbonyl and amino groups in the peptide backbone.

3.4.3 Tertiary Structure

The unique three-dimensional structure of a polypeptide is its **tertiary structure** (**Figure 3.11**). This structure is primarily due to interactions among R groups. For example, R groups with like charges are repelled by each other and those with unlike charges are attracted to each other via ionic bonds. When protein folding takes place in a watery environment, such as that found inside cells, the hydrophobic R groups of

nonpolar amino acids lay in the interior of the protein, while the hydrophilic R groups face out. Hydrophobic R groups also interact with each other through van der Waals forces.Interaction between cysteine side chains forms disulfide linkages, which are the only covalent bond formed during protein folding. All of these interactions determine the final three-dimensional shape of the protein. When a protein loses its three-dimensional shape, it may no longer be functional.

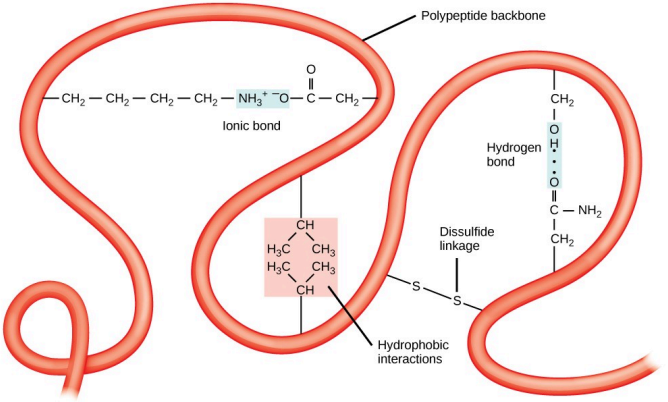


Figure 3.11 The tertiary structure of proteins is determined by a variety of chemical interactions, including hydrophobic interactions, ionic bonding, hydrogen bonding and disulfide linkages.

3.4.4 Quaternary Structure

In nature, some proteins are formed from several separate polypeptides, known as **subunits**. The interaction of these

subunits forms the **quaternary structure** of a protein. Weak interactions between the subunits help to stabilize the overall structure. For example, silk is a fibrous protein that results from hydrogen bonding between different chains.

The four levels of protein structure (primary, secondary, tertiary, and quaternary) are illustrated in **Figure 3.1**.

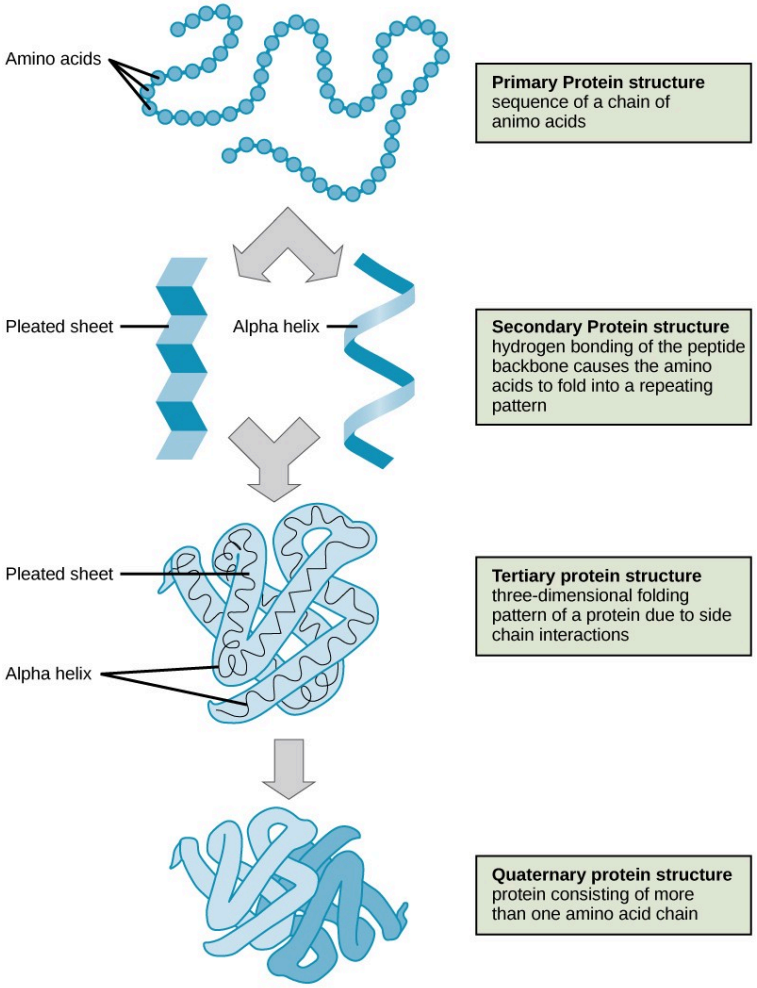


Figure **3.12** The four levels of protein structure can be observed in these illustrations. (Credit: modification of work by National Human Genome Research Institute)

3.4.5 Denaturation and Protein Folding

If a protein is subject to changes in temperature, pH, or exposure to chemicals, it may lose its shape, a process called **denaturation**. The primary structure of the protein is not changed by denaturation but some or all of the folding is lost. Denaturation is often reversible, allowing the protein to resume its function. Sometimes denaturation is irreversible, leading to permanent loss of function. One example of irreversible protein denaturation is cooking an egg white. Different proteins denature under different conditions. For example, bacteria from hot springs have proteins that function at temperatures close to boiling. Although stomach acid denatures proteins as part of digestion, the digestive enzymes of the stomach retain their activity under these conditions.

Correct folding of proteins is critical to their function. Although some proteins fold automatically, recent research has discovered that some proteins receive assistance in folding from protein helpers known as **chaperones**.

CHAPTER 4. CARBOHYDRATES

Image 4.1 Foods such as grains, vegetables, fruits, beans, and sugary sweets are all rich in carbohydrates. (Credit: “gigantfotos”/Unsplash.)

Chapter Outline

- Monosaccharides
- Disaccharides
- Polysaccharides
- The Functions and Benefits of Carbohydrates

Introduction

Most people are familiar with **carbohydrates**, one type of macromolecule, especially when it comes to what we eat. To lose weight, some individuals adhere to "low-carb" diets. Athletes, in contrast, often "carb-load" before important competitions to ensure that they have enough energy to compete at a high level. Carbohydrates are, in fact, an essential part of our diet; grains, fruits, and vegetables are all natural sources of carbohydrates. Carbohydrates provide energy to the body, particularly through **glucose**, a simple sugar that is a component of **starch** and an ingredient in many staple foods. Carbohydrates also have other important functions in humans, animals, and plants.

Carbohydrates can be represented by the chemical formula $(CH_2O)_n$, where n is the number of carbons in the molecule. In other words, the ratio of carbon to hydrogen to oxygen is 1:2:1 in carbohydrate molecules. This formula also explains the origin of the term "carbohydrate": the components are carbon

("carbo") and the components of water (hence, "hydrate"). Carbohydrates are classified into three subtypes: monosaccharides, disaccharides, and polysaccharides.

4.1 | Monosaccharides

Learning Objectives

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

- Describe the chemical nature of monosaccharides and ways in which individual monosaccharides can differ from one another.
- Name some common monosaccharides.

Monosaccharides (mono- = "one"; sacchar- = "sweet") are **simple sugars**, the most common of which is **glucose**. In monosaccharides, the number of carbons usually ranges from three to seven. Most monosaccharide names end with the suffix "-ose." If the sugar has an aldehyde group (the functional group with the structure R-CHO), it is known as an aldose,

and if it has a ketone group (the functional group with the structure RC(=O)R'), it is known as a ketose. Depending on the number of carbons in the sugar, they also may be known as trioses (three carbons), pentoses (five carbons), and or hexoses (six carbons). See **Figure 4.2** for an illustration of some monosaccharides.

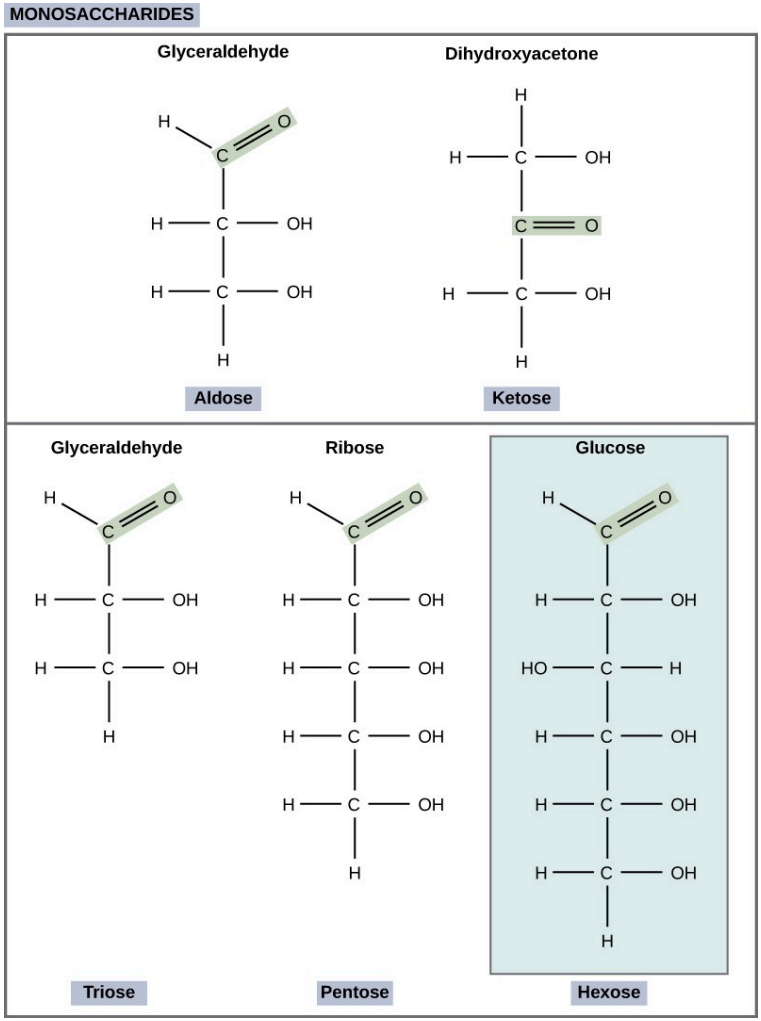


Figure 4.2 Monosaccharides are classified based on the position of their carbonyl group and the number of carbons in the backbone. Aldoses have a carbonyl group (indicated in green) at the end of the carbon chain, and ketoses have a carbonyl group in the middle of the carbon chain. Trioses, pentoses, and hexoses have three, five, and six carbon backbones, respectively.

The chemical formula for glucose is $C_6H_{12}O_6$. In humans, glucose is an important source of energy. During cellular respiration, energy is released from glucose, and that energy is used to help make adenosine triphosphate (ATP). Plants synthesize glucose using carbon dioxide and water, and glucose in turn is used for energy requirements for the plant. Excess glucose is often stored as starch that is catabolized (the breakdown of larger molecules by cells) by humans and other animals that feed on plants.

Galactose (part of lactose, or milk sugar) and fructose (found in sucrose, in fruit) are other common monosaccharides. Although glucose, galactose, and fructose all have the same chemical formula ($C_6H_{12}O_6$), they differ structurally and chemically (and are known as isomers) because of the different arrangement of functional groups around the asymmetric carbon; all of these monosaccharides have more than one asymmetric carbon (**Figure 4.3**).

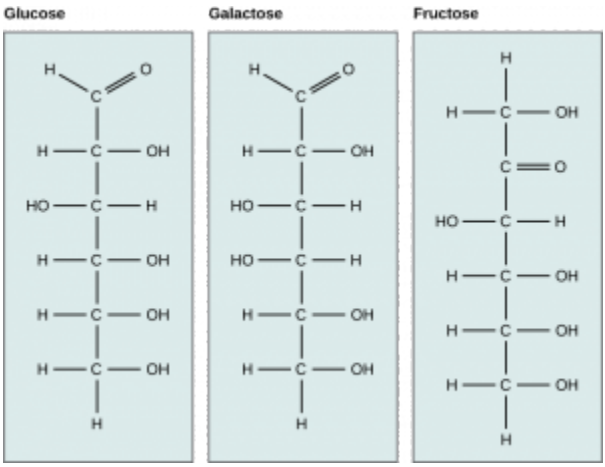
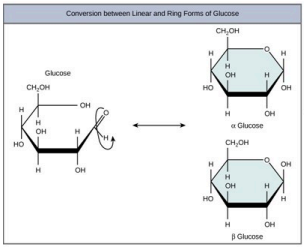


Figure 4.3 Glucose, galactose, and fructose are all hexoses. They are structural isomers, meaning they have the same chemical formula ($C_6H_{12}O_6$) but atoms bonded to each other in a different arrangement.

Glucose, galactose, and fructose are isomeric monosaccharides (hexoses), meaning they have the same chemical formula but have slightly different structures. Glucose and galactose are aldoses, and fructose is a ketose.

Monosaccharides can exist as a linear chain or as ring-shaped molecules; in aqueous solutions they are usually found in ring forms (**Figure 4.4**). Glucose in a ring form can have two different arrangements of the hydroxyl group (OH) around the anomeric carbon (carbon 1 that becomes asymmetric in the process of ring formation). If the hydroxyl group is below carbon number 1 in the sugar, it is said to be in the alpha (α)

position, and if it is above the plane, it is said to be in the beta (β) position.



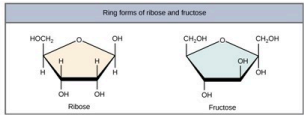


Figure 4.4 Five and six carbon monosaccharides exist in equilibrium between linear and ring forms. When the ring forms, the side chain it closes on is locked into an **α** or **β** position. Fructose and ribose also form rings, although they form five-membered rings as opposed to the six-mem

bered ring of glucose.

Concept Check

Concept Check What kind of sugars are glucose, galactose, and fructose: aldose or ketose?

4.2 | Disaccharides

Learning Objectives

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

- Describe the chemical nature of disaccharides.
- Name some common disaccharides and their constituent monosaccharides.
- Name the type of bond that holds monosaccharides together & identify it in a disaccharide structure.

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 or glycosidic linkage (Figure 4.5). Glycosidic bonds can be of the alpha or beta types.

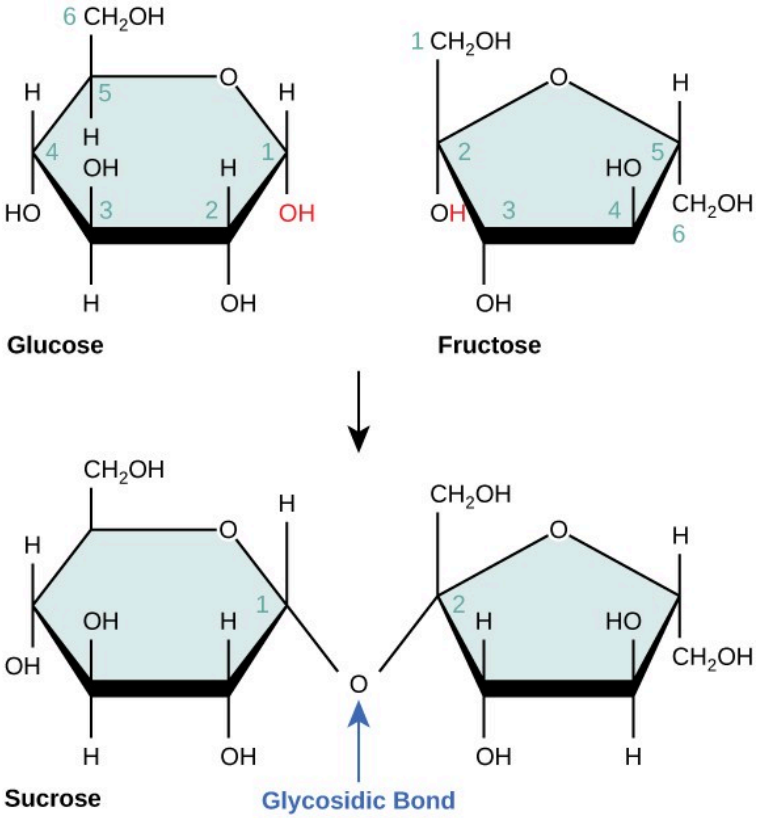


Figure 4.5 Sucrose is formed when a glucose and a monomer of fructose are joined in a dehydration reaction to form a glycosidic bond. In the process, a water molecule is lost. By convention, the carbon atoms in a monosaccharide are numbered from the terminal carbon closest to the carbonyl group. In sucrose, a glycosidic linkage is formed between carbon 1 in glucose and carbon 2 in fructose.

Common disaccharides include lactose, maltose, and sucrose (**Figure 4.6**). Lactose is a disaccharide consisting of the monomers glucose and galactose. It is found naturally in milk.

Maltose, or malt 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.

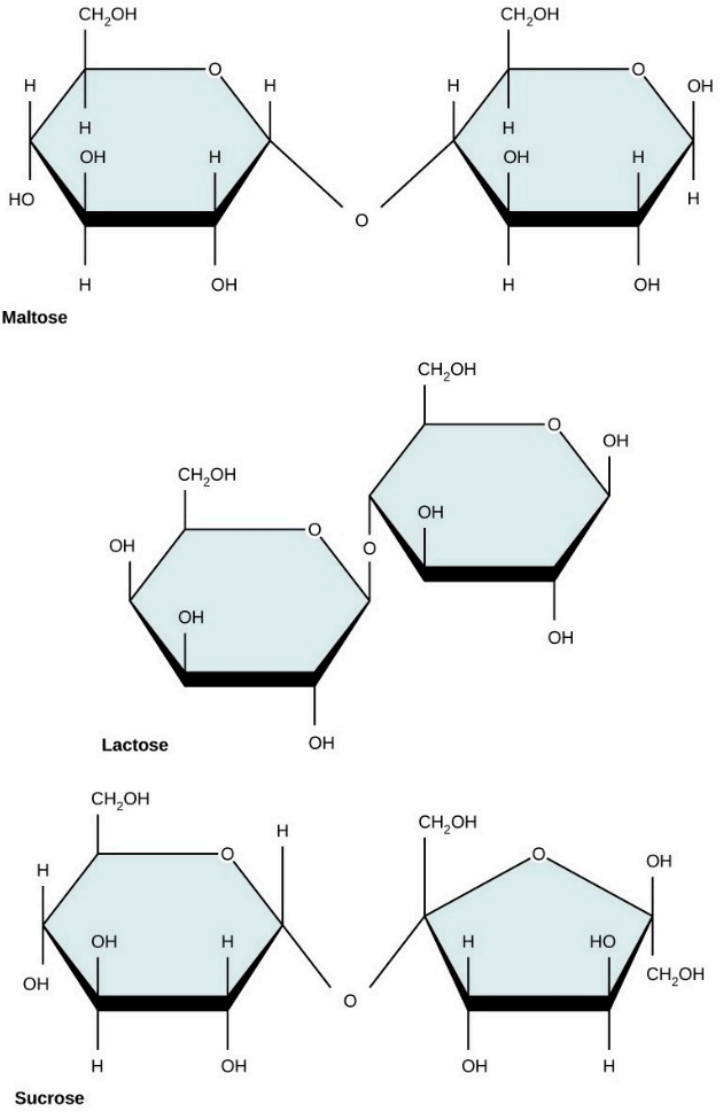


Figure 4.6 Common disaccharides include maltose (grain sugar), lactose (milk sugar), and sucrose (table sugar).

4.3 | Polysaccharides

Learning Objectives

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

- Describe the chemical nature of polysaccharides.
- Name some common polysaccharides and their functions.

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. Starch, glycogen, cellulose, chitin, and peptidoglycans 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 polymers of glucose). Plants are able to synthesize glucose, and the excess glucose, 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. The starch that is consumed by humans is broken down by enzymes, such as salivary amylases, into smaller molecules, such as maltose and glucose. The cells can then absorb the glucose.

Starch is made up of glucose monomers that are joined by α 1-4 or α 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 4.7**, amylose is starch formed by unbranched chains of glucose monomers (only α 1-4 linkages), whereas amylopectin is a branched polysaccharide (α 1-6 linkages at the branch points).

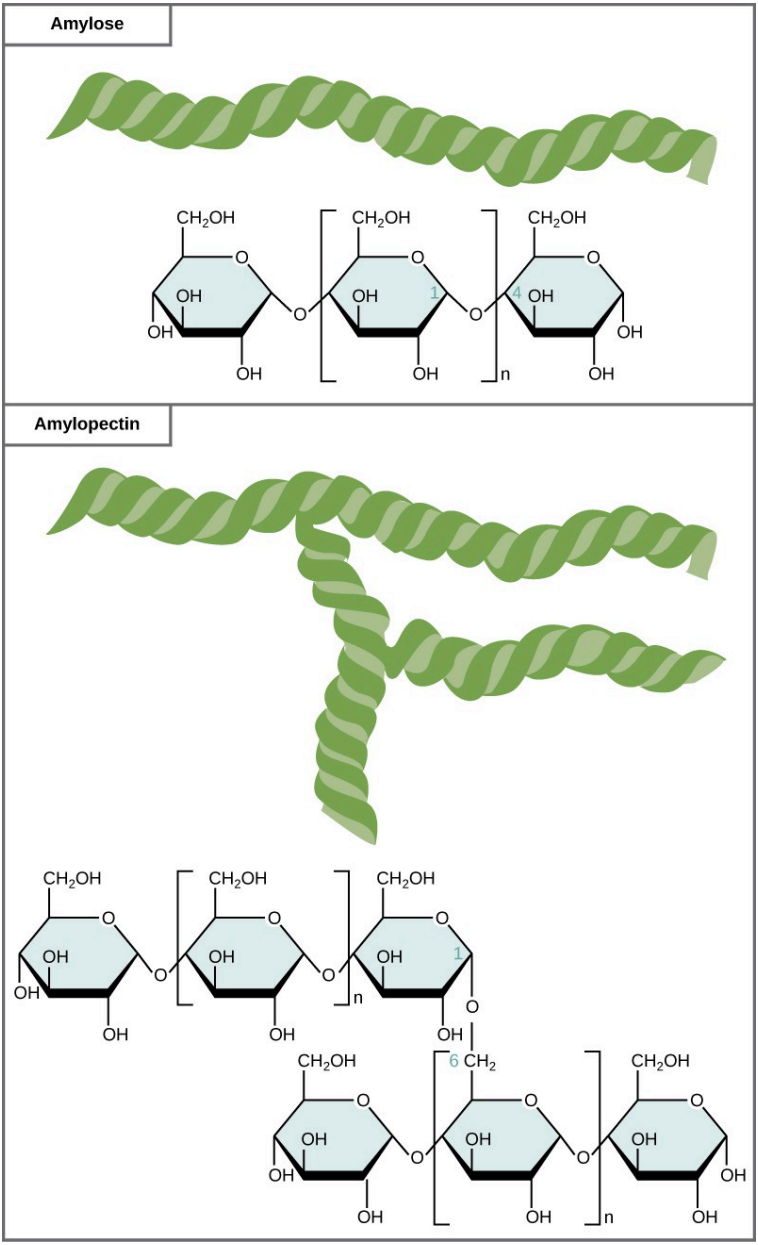


Figure 4.7 Amylose and amylopectin are two different

forms of starch. Amylose is composed of unbranched chains of glucose monomers connected by α 1,4 glycosidic linkages. Amylopectin is composed of branched chains of glucose monomers connected by α 1,4 and α 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 is the storage form of glucose in humans and other vertebrates and is made up of monomers of glucose. 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.

Cellulose is the most abundant natural biopolymer. The cell wall of plants is mostly made of cellulose; this 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 4.8**).

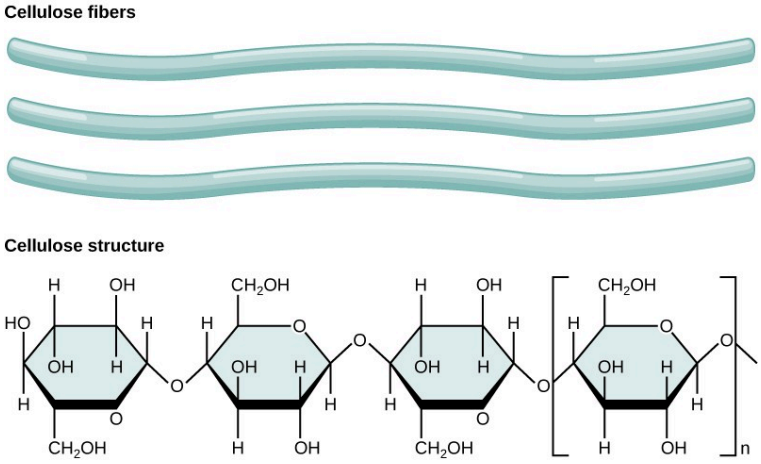


Figure 4.8 In cellulose, glucose monomers are linked in unbranched chains by β 1-4 glycosidic linkages. Because of the way the glucose subunits are joined, every glucose monomer is flipped relative to the next one resulting in a linear, fibrous structure.

As shown in **Figure 4.8**, 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.

Chitin forms the outer skeleton, called the exoskeleton, of arthropods (insects, crustaceans, and others) (**Figure 4.9**). Chitin is a nitrogen-containing polysaccharide that is made of repeating units of N-acetyl-β-d-glucosamine, a modified sugar. Chitin is also a major component of fungal cell walls.

Figure 4.9 Insects have a hard outer exoskeleton made of chitin, a type of polysaccharide. (credit: Louise Docker)

Peptidogylcans are the primary component of the cell walls of bacteria. Like chitin, peptidoglycans have nitrogen-containing sugars linked into long chains. Unlike chitin, peptidogylcans have short chains of amino acids attached to the sugars. These amino acids form peptide bonds between peptidoglycan molecules, linking strands together.

Many antibiotics, such as penicillin and its derivitives, target bacterial peptidoglycans. A common staining protocol, called Gram staining, can distinguish between different types of bacterial cell walls and help health care workers determine

which antibiotic should be used. Gram postitive bacteria have thicker cell walls that retain more of the purple Gram stain.

4.4 | Functions and Benefits of Carbohydrates

4.4.1 Carbohydrate functions

As described above, carbohydrates serve a variety of functions in cells. For example, disaccharides, starch, and glycogen serve as energy storage molecules, since they are composed of monosaccharides. Plants, algae, and some bacteria make monosaccharides using energy from the sun, in a process called **photosynthesis**. Photosynthesis essentially converts energy from the sun into chemical energy in the bonds of glucose. Glucose is broken down by **cellular respiration** or other pathways to convert its energy into ATP, the primary fuel molecule for cells.

Other carbohydrates, including cellulose, chitin, and peptidoglycans, are structural molecules. They are formed from long chains of monosaccharides that are cross-linked by hydrogen bonds or peptide bonds into tough, fibrous sheets.

Carbohydrates can also be used to build other complex molecules. For example, in the next chapter you will learn that the nucleic acid DNA contains a sugar called deoxyribose.

Sugars can also be used to make some of the amino acids that cells use to build proteins.

An important class of carbohydrates is found on the outer surface of many cells. These carbohydrates serve as identity markers to help cells of multicellular organisms recognize and communicate with each other (**Figure 4.10**).

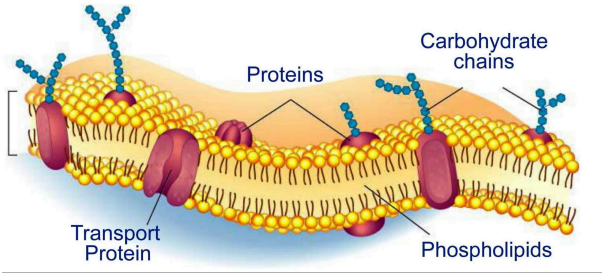


Figure 4.10 Cell membranes are made of a phospholipid bilayer (yellow) with proteins embedded in it (red). Attached to some of the proteins, on the outer surface of the cell, are carbohydrate chains (blue). These help cells recognize and communicate with each other. (Credit: image prepared by Maria Hrmova.)

4.4.2 Carbohydrate Benefits

Are carbohydrates good for you? People who wish to lose weight are often told that carbohydrates are bad for them and should be avoided. Some diets completely forbid carbohydrate consumption, claiming that a low-carbohydrate diet helps people to lose weight faster. However, carbohydrates have been

an important part of the human diet for thousands of years; artifacts from ancient civilizations show the presence of wheat, rice, and corn in our ancestors' storage areas.

Carbohydrates should be supplemented with proteins, vitamins, and fats for a well-balanced diet. Carbohydrates contain soluble and insoluble elements. The insoluble part, known as fiber, is mostly cellulose. Fiber has many uses; it promotes regular bowel movement by adding bulk, and it regulates the rate of consumption of blood glucose. Fiber also helps to remove excess cholesterol from the body by binding cholesterol in the small intestine, preventing it from entering the bloodstream, and causing it to exit the body via the feces. Fiber-rich diets also reduce the occurrence of colon cancer. In addition, eating a meal containing whole grains and vegetables gives a feeling of fullness. As an immediate source of energy, glucose is broken down during the process of cellular respiration, which produces ATP, the energy currency of the cell. Without the consumption of carbohydrates, the availability of "instant energy" is reduced. A low-calorie diet that is rich in whole grains, fruits, vegetables, and lean meat, together with plenty of exercise and plenty of water, is the most sensible way to lose weight.

Registered Dietitian

Obesity is a worldwide health concern, and many diseases such as diabetes and heart disease are becoming more prevalent because of obesity. This is one of the reasons why registered dietitians are increasingly sought after for advice. Registered dietitians help plan nutrition programs for individuals in various settings. They often work with patients in health care facilities, designing nutrition plans to treat and prevent diseases. For example, dietitians may teach a patient with diabetes how to manage blood sugar levels by eating the correct types and amounts of carbohydrates. Dietitians may also work in nursing homes, schools, and private practices.

To become a registered dietitian, one needs to earn at least a bachelor's degree in dietetics, nutrition, food technology, or a related field. In addition, registered dietitians must complete a

supervised internship program and pass a national exam. Those who pursue careers in dietetics take courses in nutrition, chemistry, biochemistry, biology, microbiology, and human physiology. Dietitians must become experts in the chemistry and physiology (biological functions) of food (proteins, carbohydrates, and fats).

CHAPTER 5. NUCLEOTIDES & NUCLEIC ACIDS

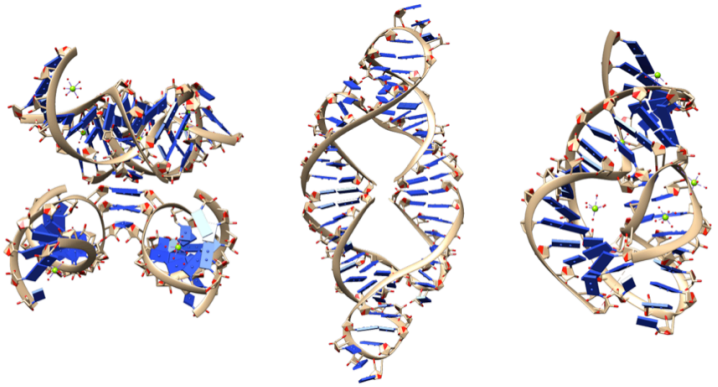


Figure 5.1 The DNA double helix is the most recognizable nucleic acid structure, but these are ribozymes. Ribozymes are ribonucleic acid molecules that can catalyze chemical reactions, like protein enzymes do. (Credit: "Lucasharr"/Wikimedia Commons)

Chapter Outline

- 5.1 Nucleotides and the Phosphodiester Bond
- 5.2 Deoxyribonucleic acid: DNA
- 5.3 Ribonucleic acid: RNA

Introduction

Nucleic acids are macromolecules made up of monomers called nucleotides. They are the most important macromolecules for the continuity of life. They carry the genetic information of a cell and instructions for the functioning of the cell. Nucleic acids are information molecules that serve as blueprints for the proteins that are made by cells. They are also the hereditary material in cells, as reproducing cells pass the blueprints on to their offspring.

The two main types of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material found in all living organisms. It is found in the nucleus of eukaryotes and in the chloroplasts and mitochondria. In prokaryotes, the DNA is not enclosed in a nucleus.

The entire genetic content of a cell is known as its **genome**. In eukaryotic cells, DNA forms a complex with histone proteins to form chromatin, the substance of eukaryotic

chromosomes. A chromosome may contain tens of thousands of genes. Many genes contain the information to make protein products; other genes code for RNA products. DNA controls all of the cellular activities by turning the genes "on" or "off."

The other type of nucleic acid, RNA, is mostly involved in protein synthesis. DNA molecules use an intermediary, called messenger RNA (mRNA), to communicate with the rest of the cell. Other types of RNA, such as rRNA, tRNA, and microRNA, are involved in protein synthesis and its regulation.

5.1 | Nucleotides and the Phosphodiester Bond

Learning Objectives

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

- Identify the three components of nucleotide structure.
- Recognize how nucleotides and nucleic acids

are related.
- Name the type of bond that holds nucleotides together & identify it in a nucleic acid structure.

DNA and RNA are made up of monomers known as **nucleotides**. The nucleotides combine with each other to form a nucleic acid, DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group (**Figure 5.2**). Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to one or more phosphate groups.

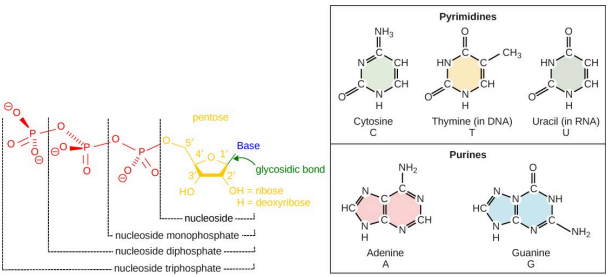


Figure 5.2 A nucleotide is made up of three components: a nitrogenous base, a pentose sugar, and one or more phosphate groups. The sugar is deoxyribose in DNA and ribose in RNA.

The nitrogenous bases are organic molecules that contain nitrogen. They are bases because they contain an amino group that has the potential of binding an extra hydrogen. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G) cytosine (C), and thymine (T). Each nucleotide in RNA contains one of four possible nitrogenous bases: adenine (A), guanine (G) cytosine (C), and uracil (U). Adenine and guanine are classified as purines and have two carbon-nitrogen rings. Cytosine, thymine, and uracil

are classified as pyrimidines, which have a single carbon-nitrogen ring (**Figure 5.2**).

The carbon atoms of the pentose sugar molecule in each nucleotide are numbered as 1′, 2′, 3′, 4′, and 5′ (1′ is read as "one prime"). The nitrogenous base is attached to the 1′ carbon and the phosphate group is attached to the hydroxyl group of the 5′ carbon. In RNA, the pentose sugar is ribose, which has a hydroxyl group attached to the 2′ carbon. In DNA, the pentose sugar is deoxyribose, which has a hydrogen atoms attached to the 2′ carbon. The "deoxy" in the name of DNA refers to the missing oxygen atom at the 2′ carbon (**Figure 5.2**).

Nucleic acids are long, linear chains of nucleotides. **Phosphodiester linkages** are covalent bonds between the 3′ carbon of one nucleotide and the 5′ phosphate group of another. They form by dehydration synthesis reactions **(Figure 5.3).** Nucleic acids have directionality: the first nucleotide in the chain has a free phosphate group at the 5′ end of the molecule. The last nucleotide added has a free 3′ hydroxy group at the 3′ end of the molecule. Nucleotides are always added on to the 3′ end.

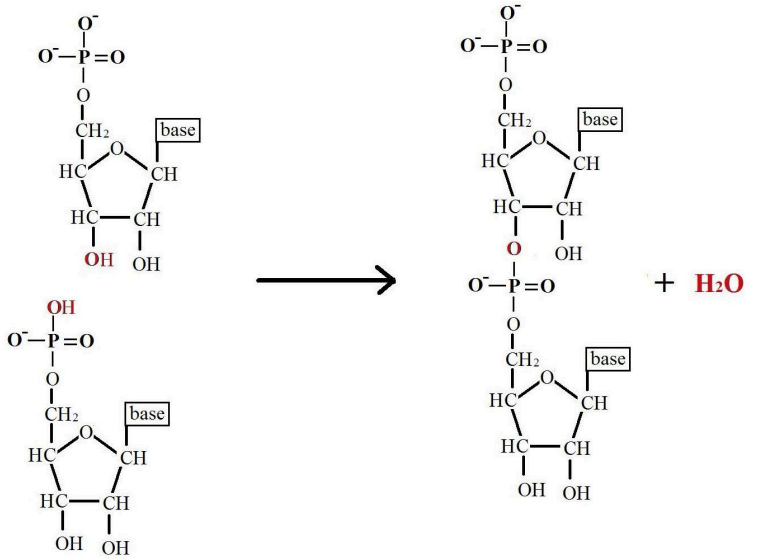


Figure 5.3 Covalent bonds that form between nucleotides are called phosphodiester linkages. They are formed by dehydration synthesis reactions when a hydroxyl group is removed from the 3' carbon of one nucleotide and a hydrogen atom is removed from the hydroxyl group attached to the 5' carbon of another nucleotide. Water is produced and a phosphodiester bond is formed.

5.2 | Deoxyribonucleic Acid: DNA

Learning Objectives

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

- Describe the structure and role of DNA.
- Discuss the similarities and differences between eukaryotic and prokaryotic DNA.

5.2.1 The Double Helix

In the 1950s, Francis Crick and James Watson worked together to determine the structure of DNA at the University of Cambridge, England. Other scientists like Linus Pauling and Maurice Wilkins were also actively exploring this field. Pauling had discovered the secondary structure of proteins using X-ray crystallography. In Wilkins' lab, researcher Rosalind Franklin was using X-ray diffraction methods to understand the structure of DNA. Watson and Crick were able to piece together the puzzle of the DNA molecule on the basis of Franklin's data because Crick had also studied X- ray diffraction (**Figure 5.4**). In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine. Unfortunately, by then Franklin had died, and Nobel prizes are not awarded posthumously.

(a) (b)

Figure 5.4 The work of pioneering scientists (a) James Watson, Francis Crick, and Maclyn McCarty led to our present day understanding of DNA. Scientist RosalindFranklin discovered the X-ray diffraction pattern of DNA, which helped to elucidate its double helix structure. (Credit: modification of work by Marjorie McCarty, Public Library of Science)

Watson and Crick correctly proposed that DNA is made up of two strands that are twisted around each other to form a right-handed helix. Two strands of nucleotides are held together by hydrogen bonds that form between pairs of nitrogenous bases. The sugar and phosphate "backbone" forms the outside of the helix. The nitrogenous bases are stacked in the interior, like the steps of a ladder. The two strands are anti-parallel in nature; that is, the 3′ end of one strand faces the 5′ end of the other strand (**Figure 5.5**).

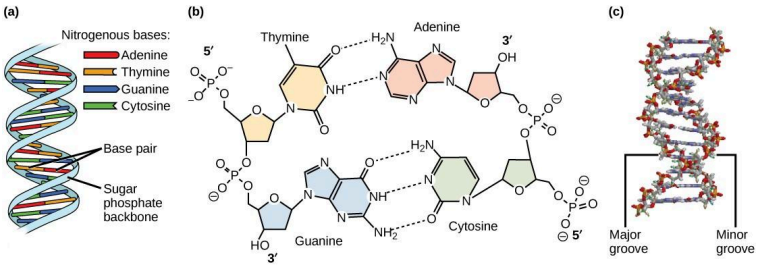


Figure 5.5 DNA has (a) a double helix structure and (b) hydrogen bonds. The major and minor grooves are binding sites for DNA binding proteins during processes such as transcription (the copying of RNA from DNA) and replication.

Only certain types of base pairing occur. A can only pair with T, and G can only pair with C, as shown in **Figure 5.5**. This is known as the base complementary rule. In other words, the DNA strands are complementary to each other. If the sequence of one strand is 5′-AATTGGCC-3′, the complementary strand would have the sequence 3′-TTAACCGG-5′. The fact that the two strands of a DNA molecule are complementary allows DNA to replicate. During DNA replication, each strand is copied, resulting in a daughter DNA double helix containing one parental DNA strand and a newly synthesized strand. The base pairs are stabilized by hydrogen bonds; adenine and thymine form two hydrogen bonds and cytosine and guanine form three hydrogen bonds.

Concept Check

A mutation occurs, and cytosine is replaced with adenine. What impact do you think this will have on the DNA structure?

5.2.2. DNA Packaging in Cells

When comparing prokaryotic cells to eukaryotic cells, prokaryotes are much simpler than eukaryotes in many of their features (**Figure 5.6**). Most prokaryotes contain a single, circular chromosome that is found in an area of the cytoplasm called the nucleoid.

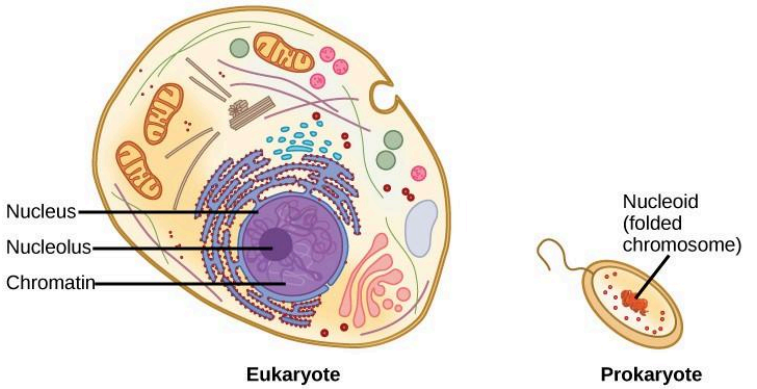


Figure 5.6 A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid.

The size of the genome in one of the most well-studied prokaryotes, *E. coli*, is 4.6 million base pairs (approximately 1.1 mm, if cut and stretched out). So how does this fit inside a small bacterial cell? The DNA is twisted by what is known as supercoiling. Supercoiling means that DNA is either under-wound (less than one turn of the helix per 10 base pairs) or over-wound (more than 1 turn per 10 base pairs) from its normal relaxed state. Some proteins are known to be involved in the supercoiling; other proteins and enzymes, such as DNA gyrase, help in maintaining the supercoiled structure.

Eukaryotes, whose chromosomes each consist of a linear DNA molecule, employ a different type of packing strategy to fit their DNA inside the nucleus (**Figure 5.7**). At the most basic level, DNA is wrapped around proteins known as **histones** to form structures called **nucleosomes**. The histones

are evolutionarily conserved proteins that are rich in basic amino acids and form an octamer. The DNA (which is negatively charged because of the phosphate groups) is wrapped tightly around the histone core. This nucleosome is linked to the next one with the help of a linker DNA. This is also known as the "beads on a string" structure. This is further compacted into a 30 nm fiber, which is the diameter of the structure. At the metaphase stage, the chromosomes are at their most compact, are approximately 700 nm in width, and are found in association with scaffold proteins.

In interphase, eukaryotic chromosomes have two distinct regions that can be distinguished by staining. The tightly packaged region is known as heterochromatin, and the less dense region is known as euchromatin. Heterochromatin usually contains genes that are not expressed (not actively transcribed to make a product), and is found in the regions of the centromere and telomeres. The euchromatin usually contains genes that are transcribed, with DNA packaged around nucleosomes but not further compacted.

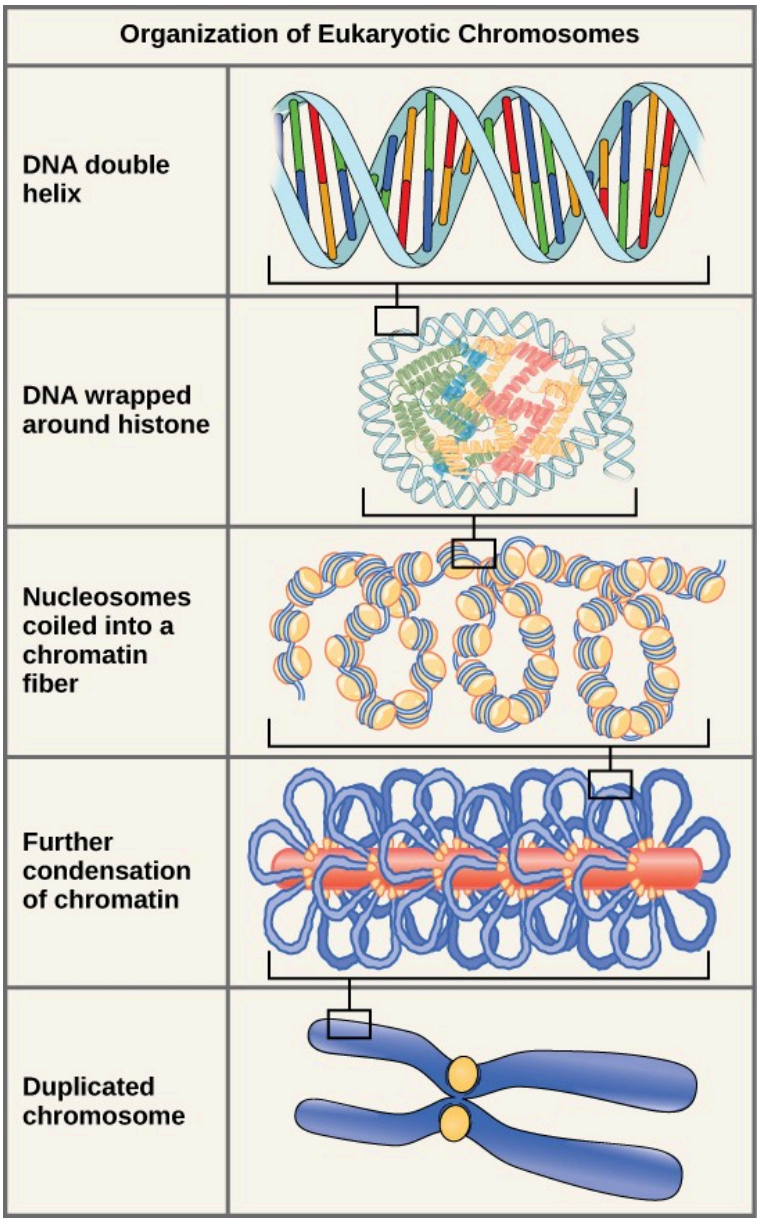


Figure 5.7 These figures illustrate the compaction of the eukaryotic chromosome.

5.3 | Ribonucleic Acid: RNA

Learning Objectives

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

- Explain the structure and roles of RNA.
- Compare and contrast the two types of nucleic acids.

Ribonucleic acid, or RNA, is mainly involved in protein synthesis. Like DNA, RNA is made of nucleotides linked by phosphodiester bonds. However, the nucleotides in RNA contain ribose sugar instead of deoxyribose and the nitrogenous base uracil (U) instead of thymine (T). Unlike DNA, RNA is usually single-stranded. However, most RNAs show internal base pairing between complementary sequences, creating a three-dimensional structure essential for their function.

There are four major types of RNA: **messenger RNA (mRNA)**, **ribosomal RNA** (**rRNA**), **transfer RNA** (**tRNA**), and **microRNA** (**miRNA**). mRNA carries a copy

of the genetic code from DNA. If a cell requires a certain protein to be synthesized, the gene is turned "on" and the corresponding messenger RNA is synthesized. The RNA sequence is complementary to the sequence of the DNA (except U replaces T). If the DNA strand has a sequence 5′-AATTGCGC-3′, the sequence of the complementary RNA is 3′-UUAACGCG-5′. The mRNA then interacts with ribosomes and other cellular machinery so that a protein can be made from the coded message. The mRNA is read in sets of three bases known as codons. Each codon codes for a single amino acid.

Thus, information flow in an organism goes from DNA to mRNA to protein. DNA dictates the sequence of mRNA in a process known as **transcription**, and RNA dictates the structure of protein in a process known as **translation**. This is known as the **Central Dogma of Molecular Biology**.

rRNA is a major constituent of ribosomes, to which the mRNA binds to make a protein product. tRNA carries the correct amino acid to the site of protein synthesis. miRNAs play a role in the regulation of gene expression. **Table 4.2** summarizes features of DNA and RNA.

Table 4.2 Features of DNA and RNA

	DNA	**RNA**
Function	Carries genetic information	Involved in protein synthesis and regulation of gene expression
Location	Remains in the nucleus	Leaves the nucleus
Structure	Double helix	Usually single-stranded
Sugar	Deoxyribose	Ribose
Pyrimidines	Cytosine, thymine	Cytosine, uracil
Purines	Adenine, guanine	Adenine, guanine

CHAPTER 6. LIPIDS

Figure 6.1 Lipids are abundant in oils and solid fats, like butter.

Chapter Outline

- 6.1 Fatty Acids
- 6.2 Triacylglycerols
- 6.3 Phospholipids
- 6.4 Steroids

Introduction

Lipids include a diverse group of compounds that are largely nonpolar in nature. This is because they are hydrocarbons that include mostly nonpolar carbon–carbon or carbon–hydrogen bonds. Non-polar molecules are hydrophobic, or insoluble in water. Lipids perform many different functions in a cell. Cells store energy for long-term use in the form of fats. Lipids also provide insulation from the environment for plants and animals (**Figure 6.2**). They help keep aquatic birds and mammals dry. Lipids are also the building blocks of many hormones and are an important constituent of all cellular membranes. Lipids include fats, oils, waxes, phospholipids, and steroids.

Figure 6.2 Hydrophobic lipids in the fur of aquatic mammals, such as this river otter, protect them from the elements.(Credit: Ken Bosma)

5.1 | Fatty Acids

Learning Objectives

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

- Label the structural components of a fatty acid.
- Explain why saturated and *trans* unsaturated fatty acids are solids at room temperature and *cis* unsaturated fatty acids are liquids.

Fatty acids are building blocks in several other types of lipids. They have a long chain of hydrocarbons to which a carboxyl (carboxylic acid) group is attached, hence the name "fatty acid." The number of carbons in the fatty acid may range from 4 to 36; most common are those containing 12–18 carbons. Some fatty acids have common names that specify their origin. For example, palmitic acid, is derived from the palm tree. Arachidic acid is derived from *Arachis hypogea,* the scientific name for peanuts.

Fatty acids may be **saturated** or **unsaturated**. In a fatty acid chain, if there are only single bonds between neighboring carbons, the fatty acid is said to be saturated. Saturated fatty acids are saturated with hydrogen; in other words, the number of hydrogen atoms attached to the carbon skeleton is maximized. Stearic acid is an example of a saturated fatty acid (**Figure 6.3**). When the hydrocarbon chain contains a double

bond, the fatty acid is said to be unsaturated. Oleic acid is an example of an unsaturated fatty acid (**Figure 6.4**).

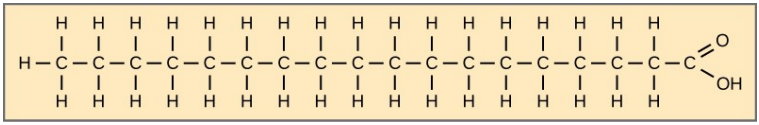


Figure 6.3 Stearic acid is a common saturated fatty acid.

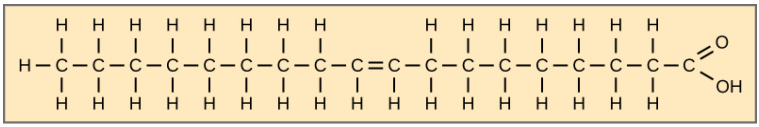


Figure 6.4 Oleic acid is a common unsaturated fatty acid.

If there is one double bond in the molecule, then it is known as a monounsaturated fat (e.g., olive oil), and if there is more than one double bond, then it is known as a polyunsaturated fat (e.g., canola oil).

The double bonds found in unsaturated fatty acids can be in either a *cis* or *trans* configuration. In the *cis* ("same") configuration, both hydrogen atoms are on the same side of the carbon chain, which results in a bend in the chain. As a result, the fatty acids cannot pack tightly together and therefore form a liquid, or oil, at room temperature. On the other hand, fatty acids in the *trans* ("across") configuration,

the hydrogen atoms are on opposite sides of the carbon chain. Like saturated fatty acids, *trans* unsaturated fatty acid chains are straight and pack tightly together, and therefore form a solid, or fat, at room temperature (**Figure 6.5).**

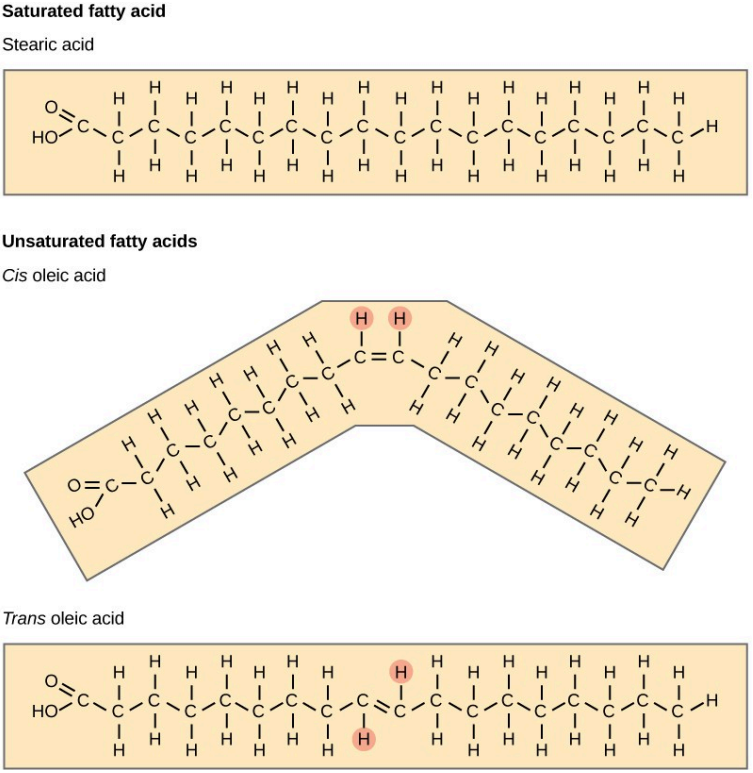


Figure 6.5 Saturated fatty acids have hydrocarbon chains connected by single bonds only. Unsaturated fatty acids have one or more double bonds. Each double bond may be in a *cis* or *trans* configuration. In the *cis* configuration, both hydrogens are on the same side of the hydrocarbon chain. In the *trans* configuration, the hydrogens are on opposite sides. A *cis* double bond causes a kink in the chain.

Trans Fats

In the food industry, oils are artificially hydrogenated to make them semi-solid and of a consistency desirable for many processed food products. Simply speaking, hydrogen gas is bubbled through oils to solidify them. During this hydrogenation process, double bonds of the *cis*– conformation in the hydrocarbon chain may be converted to double bonds in the *trans*– conformation.

Margarine, some types of peanut butter, and shortening are examples of artificially hydrogenated trans fats. Recent studies have shown that an increase in trans fats in the human diet may lead to an increase in levels of low-density lipoproteins (LDL), or "bad" cholesterol, which in turn may lead to plaque deposition in the arteries, resulting in heart disease. Many fast food restaurants have recently banned the use of trans fats, and food labels are required to display the trans fat content.

Omega Fatty Acids

Essential fatty acids are fatty acids required but not synthesized by the human body. Consequently, they must be supplied by the diet. **Omega-3** and **omega-6** fatty acids are the only two essential fatty acids known for humans. Omega-3 fatty acids are so named because the third carbon from the end of the hydrocarbon chain is connected to its neighboring carbon by a double bond (**Figure 6.6**).

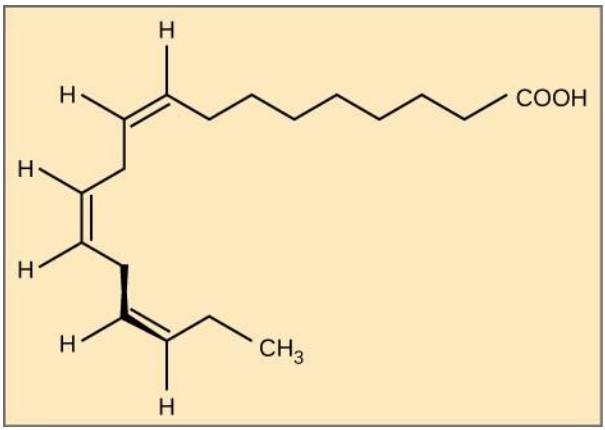


Figure 6.6 Alpha-linolenic acid is an example of an omega-3 fatty acid. It has three *cis* double bonds and, as a result, a curved shape. For clarity, the carbons are not shown. Each singly bonded carbon has two hydrogens associated with it, also not shown.

Omega-3 fatty acids include alpha-linoleic acid (ALA), eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA), all of which are polyunsaturated. Salmon, trout, and tuna are good sources of omega-3 fatty acids. Research indicates that omega-3 fatty acids reduce the risk of sudden death from heart attacks, reduce triglycerides in the blood, lower blood pressure, and prevent thrombosis by inhibiting blood clotting. They also reduce inflammation, and may help reduce the risk of some cancers in animals.

5.2 | Triacylglycerols

Learning Objectives

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

- Label the structural components of triacylglycerols and state their biological function.

A fat molecule, also called **triacylglycerols** or **triglycerides,** consists of two main components—glycerol and fatty acids. Glycerol is an organic compound with three carbons and three hydroxyl (OH) groups. In a fat molecule, the fatty acids are attached to each of the three hydroxyl groups of the glycerol molecule, forming ester bonds (**Figure 6.7**). Each ester bond forms by a dehydration synthesis reaction, forming a water molecule.

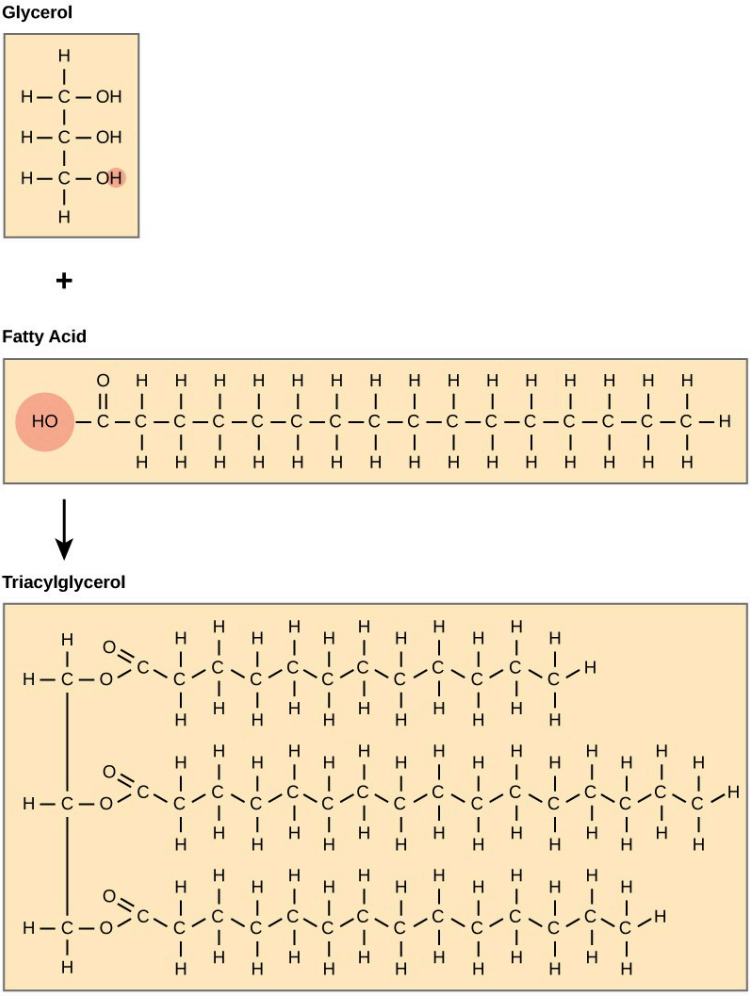


Figure 6.7 Triacylglycerol is formed by the joining of three fatty acids to a glycerol backbone in a dehydration reaction. Three molecules of water are released in the process.

Mammals store fats in specialized cells called adipocytes, where globules of fat occupy most of the cell's volume. In plants,

fat or oil is often stored in seeds and is used as a source of energy during seedling development. Unsaturated fats or oils are usually of plant or fish origin. Olive oil, corn oil, canola oil, and cod liver oil are examples of unsaturated fats.

Saturated fats usually occur in animals, and are found in meat and milk (butter). In humans, unsaturated fats help to lower blood cholesterol levels whereas saturated fats contribute of plaque formation in the arteries.

Like carbohydrates, fats have received a lot of bad publicity. It is true that eating an excess of fried foods and other "fatty" foods leads to weight gain. However, fats do have important functions. Many vitamins are fat soluble, and fats serve as a long-term storage form of fatty acids: a source of energy. They also provide insulation for the body. Therefore, "healthy" fats in moderate amounts should be consumed on a regular basis.

5.3 | Phospholipids

Learning Objectives

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

- Label the structural components of a phospholipid.
- Explain how the structure of phospholipids leads to their biological function in membranes.

Phospholipids are major constituents of the plasma membrane, the outermost layer of animal cells. Like fats, they are composed of fatty acid chains attached to a glycerol backbone. Instead of three fatty acids attached as in triglycerides, however, there are two fatty acids and a modified phosphate group (**Figure 6.8**).

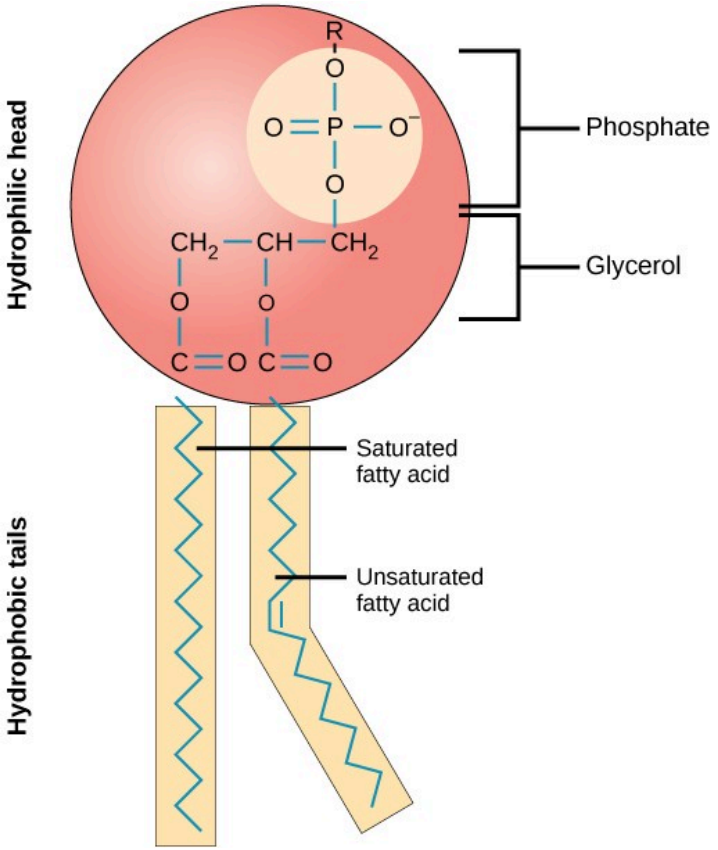


Figure 6.8 A phospholipid is a molecule with two fatty acids and a modified phosphate group attached to a glycerol backbone. The phosphate may be modified by the addition of charged or polar chemical groups.

A phospholipid is an **amphipathic** molecule, meaning it has a hydrophobic and a hydrophilic part. The "head" is composed of the hydrophilic phosphate group, and the "tail" contains the hydrophobic fatty acids. In a membrane, a bilayer of

phospholipids forms the matrix of the structure, the fatty acid tails of phospholipids face inside, away from water, whereas the phosphate group faces the outside, aqueous side (**Figure 6.9**).

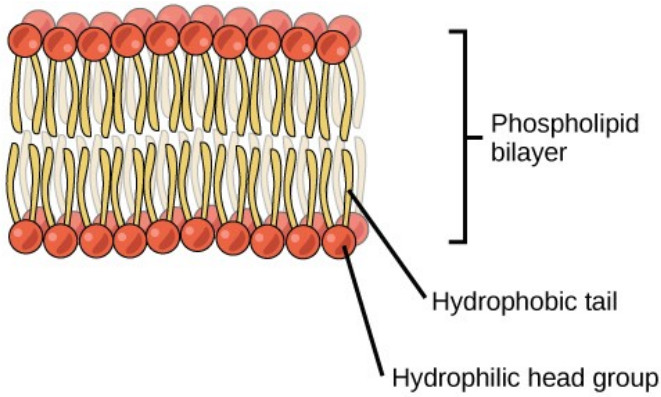


Figure 6.9 The phospholipid bilayer is the major component of all cellular membranes. The hydrophilic head groups of the phospholipids face the aqueous solution. The hydrophobic tails are sequestered in the middle of the bilayer.

Phospholipids are responsible for the dynamic nature of the plasma membrane. If a drop of phospholipids is placed in water, it spontaneously forms a structure known as a **micelle**, where the hydrophilic phosphate heads face the outside and the fatty acids face the interior.

5.4 | Steroids

Learning Objectives

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

- Name the four major types of lipids.
- Recognize the basic structure of a steroid and name some functions of steroids.

Unlike the phospholipids and fats discussed earlier, **steroids** have a fused ring structure. Although they do not resemble the other lipids, they are grouped with them because they are also hydrophobic and insoluble in water. All steroids have four linked carbon rings and several of them, like cholesterol, have a short tail (**Figure 6.10**). Many steroids also have the –OH functional group, which puts them in the alcohol classification (sterols).

Cholesterol is the most common steroid. Cholesterol is mainly synthesized in the liver and is the precursor to many steroid **hormones** such as testosterone and estradiol, which are secreted by the gonads and endocrine glands. It is also the

precursor to Vitamin D. Cholesterol is also the precursor 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 by lay people, it is necessary for proper functioning of the body. It is a component of the plasma membrane of animal cells and is found within the phospholipid bilayer.

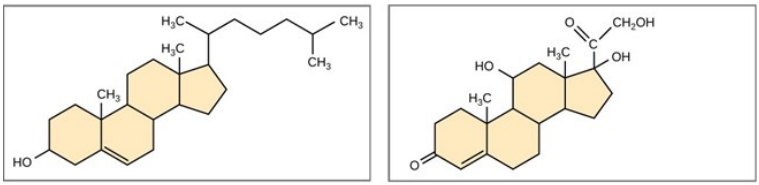


Figure 6.10 Steroids such as cholesterol and cortisol are composed of four fused hydrocarbon rings.

PART III

PART 3. WHAT ARE THE MAIN FEATURES AND CATEGORIES OF CELLS?

CHAPTER 7. INTRODUCTION TO CELLS

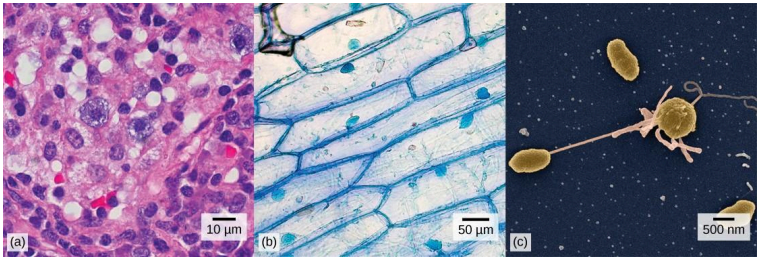


Figure 7.1 (a) Nasal sinus cells (viewed with a light microscope), (b) onion cells (viewed with a light microscope), and (c) *Vibrio tasmaniensis* bacterial cells (seen through a scanning electron microscope) are from very different organisms, yet all share certain characteristics of basic cell structure. (Credit a: modification of work by Ed Uthman, MD; credit b: modification of work by Umberto Salvagnin; credit c: modification of work by Anthony D'Onofrio, William H. Fowle, Eric J. Stewart, and Kim Lewis of the Lewis Lab at Northeastern University; scale-bar data from Matt Russell.)

Chapter Outline

Type your examples here.

- 7.1 Studying Cells
- 7.2 Prokaryotic Cells
- 7.3 Eukaryotic Cells
- 7.4 The Endomembrane System
- 7.5 The Cytoskeleton
- 7.6 Connections Between Cells

Introduction

Close your eyes and picture a brick wall. What is the basic building block of that wall? A single brick, of course. Like a brick wall, your body is composed of basic building blocks, and the building blocks of your body are cells.

Your body has many kinds of cells, each specialized for a specific purpose. Just as a home is made from a variety of building materials, the human body is constructed from many cell types. For example, epithelial cells protect the surface of the body and cover the organs and body cavities within. Bone cells help to support and protect the body. Cells of the immune system fight invading bacteria. Additionally, blood and blood cells carry nutrients and oxygen throughout the body while removing carbon dioxide. Each of these cell types plays a vital

role during the growth, development, and day-to-day maintenance of the body. In spite of their enormous variety, however, cells from all organisms—even ones as diverse as bacteria, onion, and human—share certain fundamental characteristics.

7.1 | Studying Cells

Learning Objectives

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

- Describe the role of cells in organisms.
- Compare and contrast light microscopy and electron microscopy.
- Summarize cell theory.

A cell is the smallest unit of a living thing. A living thing, whether made of one cell (like bacteria) or many cells (like a human), is called an **organism**. Thus, cells are the basic building blocks of all organisms.

Several cells of one kind that interconnect with each other

and perform a shared function form tissues, several tissues combine to form an organ (your stomach, heart, or brain), and several organs make up an organ system (such as the digestive system, circulatory system, or nervous system). Several systems that function together form an organism (like a human being). Here, we will examine the structure and function of cells.

There are many types of cells, all grouped into one of two broad categories: **prokaryotic** and **eukaryotic**. For example, both animal and plant cells are classified as eukaryotic cells, whereas bacterial cells are classified as prokaryotic. Before discussing the criteria for determining whether a cell is prokaryotic or eukaryotic, let's first examine how biologists study cells.

7.1.1 Microscopy

Most individual cells cannot be seen with the naked eye, so scientists use microscopes (micro- = "small"; -scope = "to look at") to study them. A **microscope** is an instrument that magnifies an object. Most photographs of cells are taken with a microscope, and these images can also be called micrographs.

Light Microscopes

To give you a sense of cell size, a typical human red blood cell is about eight millionths of a meter, or eight micrometers (abbreviated as 8 μm) in diameter; the head of a pin of is about

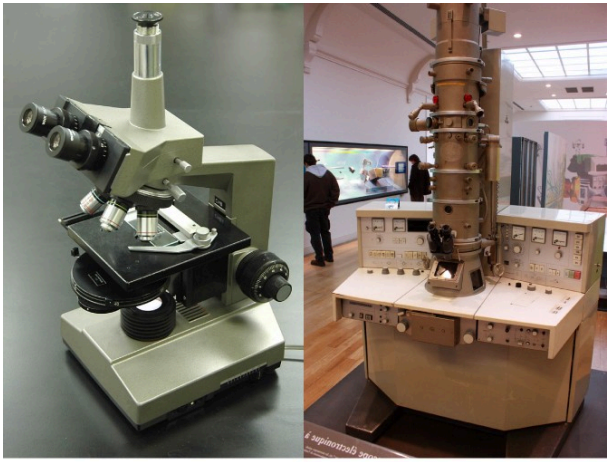
two thousandths of a meter (2 mm) in diameter. That means about 250 red blood cells could fit across the head of a pin.

Most commonly used microscopes are classified as **light microscopes** (**Figure 7.2a**). Visible light passes and is bent through the lens system to enable the user to see the specimen. Light microscopes are advantageous for viewing living organisms, but since individual cells are generally transparent, their components are not distinguishable unless they are colored with special stains..

The optics of a light microscope's lenses change the orientation of the image that the user sees. A specimen that is right-side up and facing right on the microscope slide will appear upside-down and facing left when viewed through a microscope, and vice versa. Similarly, if the slide is moved left while looking through the microscope, it will appear to move right, and if moved down, it will seem to move up. This occurs because microscopes use two sets of lenses to magnify the image. Because of the manner by which light travels through the lenses, this system of two lenses produces an inverted image (dissecting microscopes, work in a similar manner, but include an additional magnification system that makes the final image appear to be upright).

Light microscopes commonly used in the undergraduate college laboratory magnify up to approximately 400 times. Two parameters that are important in microscopy are magnification and resolving power. Magnification is the process of enlarging an object in appearance. Resolving power

is the ability of a microscope to distinguish two adjacent structures as separate: the higher the resolution, the better the clarity and detail of the image. When oil immersion lenses are used for the study of small objects, magnification is usually increased to 1,000 times.



(a) (b)

Figure 7.2 (a) Most light microscopes used in a college biology lab can magnify cells up to approximately 400 times and have a resolution of about 200 nanometers. (b) Electron microscopes provide a much higher magnification, 100,000x, and a have a resolution of 50 picomete

rs. (Credit a: modification of work by "GcG"/Wikimedia Commons; credit b: modification of work by Evan Bench)

Electron Microscopes

In contrast to light microscopes, **electron microscopes** (**Figure 7.2b**) use a beam of electrons instead of a beam of light. Not only does this allow for higher magnification and, thus, more detail (**Figure 7.3**), it also provides higher resolving power. The method used to prepare the specimen for viewing with an electron microscope kills the specimen. Electrons have short wavelengths (shorter than photons) that move best in a vacuum, so living cells cannot be viewed with an electron microscope.

In a **scanning electron microscope**, a beam of electrons moves back and forth across a cell's surface, creating details of cell surface characteristics. In a **transmission electron microscope**, the electron beam penetrates the cell and provides details of a cell's internal structures. As you might imagine, electron microscopes are significantly more bulky and expensive than light microscopes.

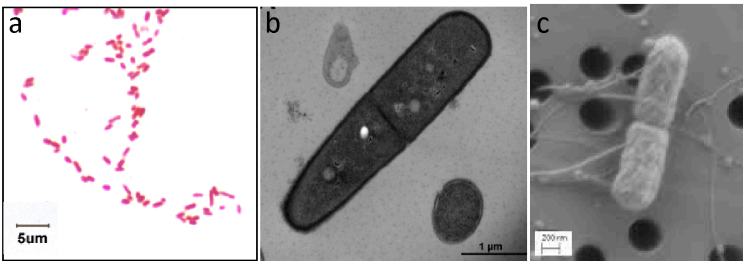


Figure 7.3 Salmonella enteritidis bacteria viewed with (a) light microscope (b) transmission electron microscope and (c) scanning electron microscope

career CONNECTION

Cytotechnologist

A Pap smear is a common medical test in which a small sample of cells are taken from the uterine cervix of a patient and sent to a medical lab. A professional called a cytotechnologist then stains the cells and examines them for any changes that could indicate cervical cancer or a microbial infection (**Figure 7.4**).

Cytotechnologists (cyto- = "cell") study cells via microscopic examinations and other laboratory tests. They are trained to determine which

cellular changes are within normal limits and which are abnormal. Their focus is not limited to cervical cells; they study cellular specimens that come from all organs. When they notice abnormalities, they consult a pathologist, who is a medical doctor who can make a clinical diagnosis. Cytotechnologists play a vital role in saving people's lives. When abnormalities are discovered early, a patient's treatment can begin sooner, which usually increases the chances of a successful outcome.

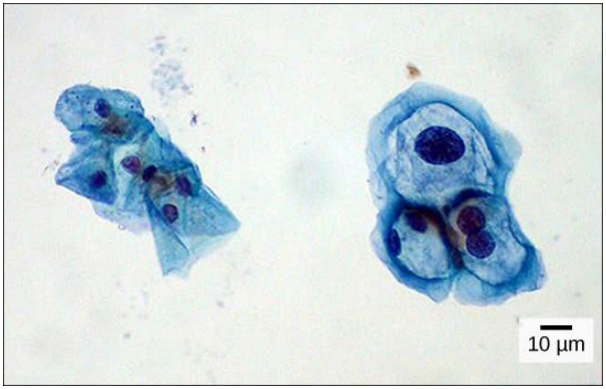


Figure 7.4 These uterine cervix cells, viewed through a light microscope, were obtained from a Pap smear. Normal cells are on the left. The cells on the right are infected with human papillomavirus (HPV). Notice that the infected cells are larger and two cells each have two nuclei. (Credit: modification of work by Ed Uthman, MD; scale-bar data from Matt Russell.)

7.1.2 Cell Theory

The microscopes we use today are far more complex than those used in the 1600s by Antony van Leeuwenhoek, a Dutch shopkeeper who had great skill in crafting lenses. Despite the limitations of his now-ancient lenses, van Leeuwenhoek observed the movements of protista (a type of single-celled

organism) and sperm, which he collectively termed "animalcules."

In a 1665 publication called Micrographia, Robert Hooke coined the term "cell" for the box-like structures he observed when viewing cork tissue through a lens. In the 1670s, van Leeuwenhoek discovered bacteria and protozoa. Later advances in lenses, microscope construction, and staining techniques enabled other scientists to see some components inside cells. By the late 1830s, botanist Matthias Schleiden and zoologist Theodor Schwann were studying tissues and proposed the unified cell theory, which states that all living things are composed of one or more cells, the cell is the basic unit of life, and new cells arise from existing cells.

7.2 | Prokaryotic Cells

Learning Objectives

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

- Name examples of prokaryotic and eukaryotic organisms.

- Compare and contrast prokaryotic cells and eukaryotic cells.
- Describe the relative sizes of different kinds of cells.
- Explain why cells must be small.

Cells fall into one of two broad categories: prokaryotic and eukaryotic. Only the single-celled organisms of the domains Bacteria and Archaea are classified as prokaryotes. Cells of animals, plants, fungi, and protists are all eukaryotes, and are made up of eukaryotic cells.

7.2.1. Components of Prokaryotic Cells

All cells share four common components: 1) a **plasma membrane**, an outer covering that separates the cell's interior from its surrounding environment; 2) **cytoplasm**, consisting of a jelly-like cytosol within the cell in which other cellular components are found; 3) DNA, the genetic material of the cell; and 4) **ribosomes**, which synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

A **prokaryote** is a simple, single-celled (**unicellular**) organism that lacks a nucleus, or other highly organized

membrane-bound organelle (pro- = "before"; -kary- = "kernel," which refers to the nucleus). Prokaryotic DNA is found in a central part of the cell called the **nucleoid** (**Figure 7.5**).

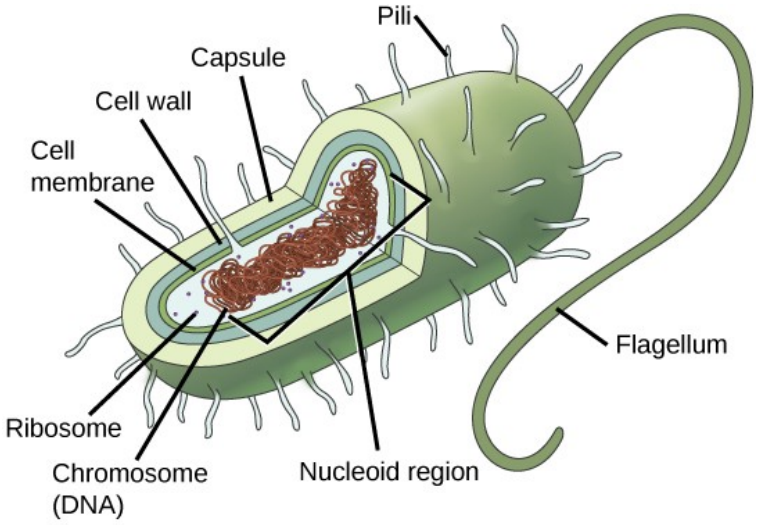


Figure 7.5 This figure shows the generalized structure of a prokaryotic cell. All prokaryotes have chromosomal DNA localized in a nucleoid, ribosomes, a cell membrane, and a cell wall. The other structures shown are present in some, but not all, bacteria.

The prokaryotes, which include bacteria and archaea, are mostly single-celled organisms that, by definition, lack membrane-bound nuclei and other organelles. A bacterial chromosome is a covalently closed circle that, unlike eukaryotic chromosomes, is not organized around histone proteins. The central region of the cell in which prokaryotic

DNA resides is called the nucleoid. In addition, prokaryotes often have abundant **plasmids**, which are shorter circular DNA molecules that may only contain one or a few genes. Plasmids can be transferred independently of the bacterial chromosome during cell division and often carry traits such as antibiotic resistance.

Most prokaryotes have a peptidoglycan cell wall and many have a polysaccharide capsule (**Figure 7.5**). The cell wall acts as an extra layer of protection, helps the cell maintain its shape, and prevents dehydration. The capsule enables the cell to attach to surfaces in its environment. Some prokaryotes have flagella, pili, or fimbriae. Flagella are used for locomotion. Pili are used to exchange genetic material during a type of reproduction called conjugation. Fimbriae are used by bacteria to attach to a host cell or other surface.

Microbiologist

The most effective action anyone can take to prevent the spread of contagious illnesses is to wash his or her hands. Why? Because microbes (organisms so tiny that they can only be seen

with microscopes) are ubiquitous. They live on doorknobs, money, your hands, and many other surfaces. If someone sneezes into his hand and touches a doorknob, and afterwards you touch that same doorknob, the microbes from the sneezer's mucus are now on your hands. If you touch your hands to your mouth, nose, or eyes, those microbes can enter your body and could make you sick.

However, not all microbes (also called microorganisms) cause disease; most are actually beneficial. You have microbes in your gut that make vitamin K. Other microorganisms are used to ferment beer and wine.

Microbiologists are scientists who study microbes. Microbiologists can pursue a number of careers. Not only do they work in the food industry, they are also employed in the veterinary and medical fields. They can work in the pharmaceutical sector, serving key roles in research and development by identifying new sources of antibiotics that could be used to treat bacterial infections.

Environmental microbiologists may look for new

ways to use specially selected or genetically engineered microbes for the removal of pollutants from soil or groundwater, as well as hazardous elements from contaminated sites. These uses of microbes are called bioremediation technologies. Microbiologists can also work in the field of bioinformatics, providing specialized knowledge and insight for the design, development, and specificity of computer models of, for example, bacterial epidemics.

7.2.2 Cell Size

At 0.1 to 5.0 μm in diameter, prokaryotic cells are significantly smaller than most eukaryotic cells, which have diameters ranging from 10 to 100 μm (**Figure 7.6**). The small size of prokaryotes allows ions and organic molecules that enter them to quickly diffuse to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly diffuse out. This is not the case in eukaryotic cells, which have developed different structural adaptations to enhance intracellular transport.

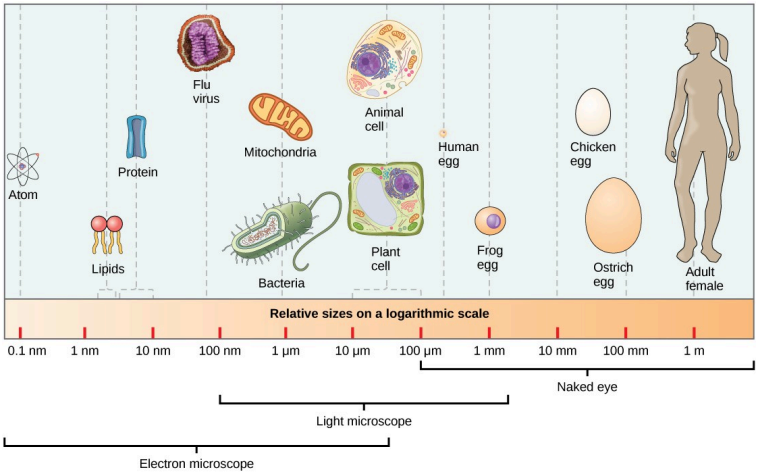


Figure 7.6 This figure shows relative sizes of structures on a logarithmic scale (recall that each unit of increase in a logarithmic scale represents a 10-fold increase in the quantity being measured).

Small size is necessary for all cells, whether prokaryotic or eukaryotic. Not all cells are spherical in shape, but most tend to approximate a sphere. The formula for the surface area of a sphere is $4\pi r^2$, while the formula for its volume is $4\pi r^3/3$. Thus, as the radius of a cell increases, its surface area increases as the square of its radius, but its volume increases as the cube of its radius. Therefore, as a cell increases in size, its surface area-to-volume ratio decreases. This same principle would apply if the cell had the shape of a cube (**Figure 7.7**). If the cell grows too large, the plasma membrane will not have sufficient surface area to support the rate of diffusion required for the increased volume. In other words, as a cell grows, it becomes

less efficient. One way to become more efficient is to divide; another way is to develop organelles that perform specific tasks. These adaptations lead to the development of more sophisticated cells called eukaryotic cells.

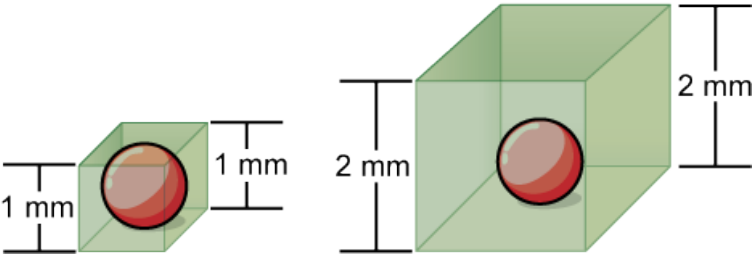


Figure 7.7 Notice that as a cell increases in size, its surface area-to-volume ratio decreases. When there is insufficient surface area to support a cell's increasing volume, a cell will either divide or die.

Concept Check

Prokaryotic cells are much smaller than eukaryotic cells.

- What advantages might small cell size confer on a cell?

- What advantages might large cell size have?

7.3 | Eukaryotic Cells

Learning Objectives

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

- Describe the structure of eukaryotic cells.
- Compare animal cells with plant cells.
- State the role of the plasma membrane.
- Summarize the functions of the major cell organelles.

Unlike prokaryotic cells, eukaryotic cells have a membrane-bound nucleus, numerous membrane-bound **organelles** that have specialized functions, and several, rod-shaped chromosomes. Because a eukaryotic cell's nucleus is

surrounded by a membrane, the word eukaryote means "true nucleus."

Eukaryotic cells have a more complex structure than prokaryotic cells. Organelles ("little organs") allow different functions to be compartmentalized in different areas of the cell. Before turning to organelles, let's first examine two important components of the cell: the plasma membrane and the cytoplasm.

7.3.1 The Plasma Membrane

Like prokaryotes, eukaryotic cells have a plasma membrane (**Figure 7.8**), consisting of 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.

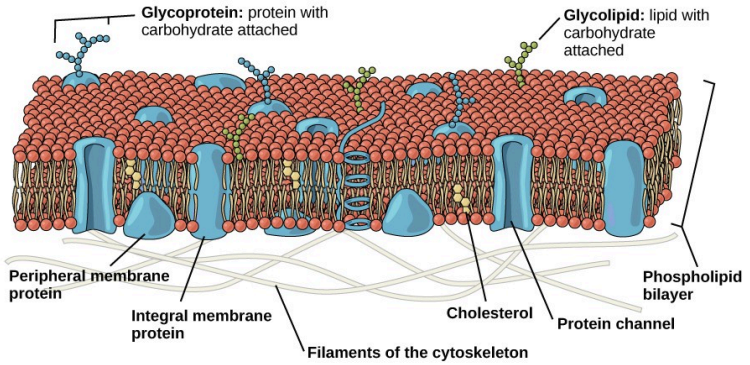


Figure 7.8 The eukaryotic plasma membrane is a phospholipid bilayer with proteins and cholesterol embedded in it.

The plasma membranes of cells that specialize in absorption are folded into fingerlike projections called microvilli (**Figure 7.9**). Such cells are typically found lining the small intestine, the organ that absorbs nutrients from digested food. People with celiac disease have an immune response to gluten, which is a protein found in wheat, barley, and rye. The immune response damages microvilli, preventing absorption of nutrients and leading to malnutrition, cramping, and diarrhea. Patients with celiac disease must follow a gluten-free diet.

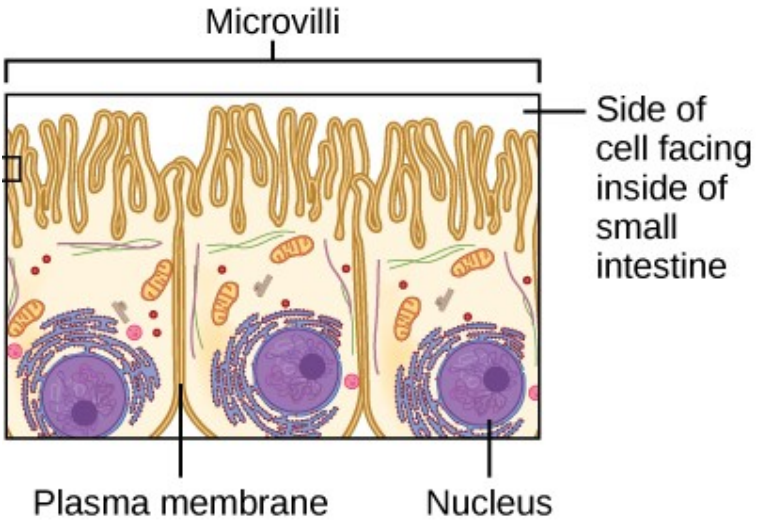


Figure 7.9 Microvilli increase the surface area for absorption. (credit "micrograph": modification of work by Louisa Howard)

7.3.2 The Cytoplasm

The **cytoplasm** is the entire region of a cell between the plasma membrane and the nuclear envelope. It is made up of organelles suspended in the gel-like **cytosol**, the cytoskeleton, and various chemicals (**Figure 7.10**). Even though the cytoplasm consists of 70 to 80 percent water, it has a semi-solid consistency, which comes from the proteins within it. Glucose and other simple sugars, polysaccharides, amino acids, nucleic acids, fatty acids, and ions are also found in the cytoplasm. Many metabolic reactions take place in the cytoplasm.

7.3.3 Organelles

The cytoplasm of eukaryotic cells is highly compartmentalized into structures called organelles. Like organs in your body, each type of organelle has a specialized function. The organelles and other cellular structures found in typical animal (**Figure 7.10a**) and plant (**Figure 7.10b**) cells are diagrammed below.

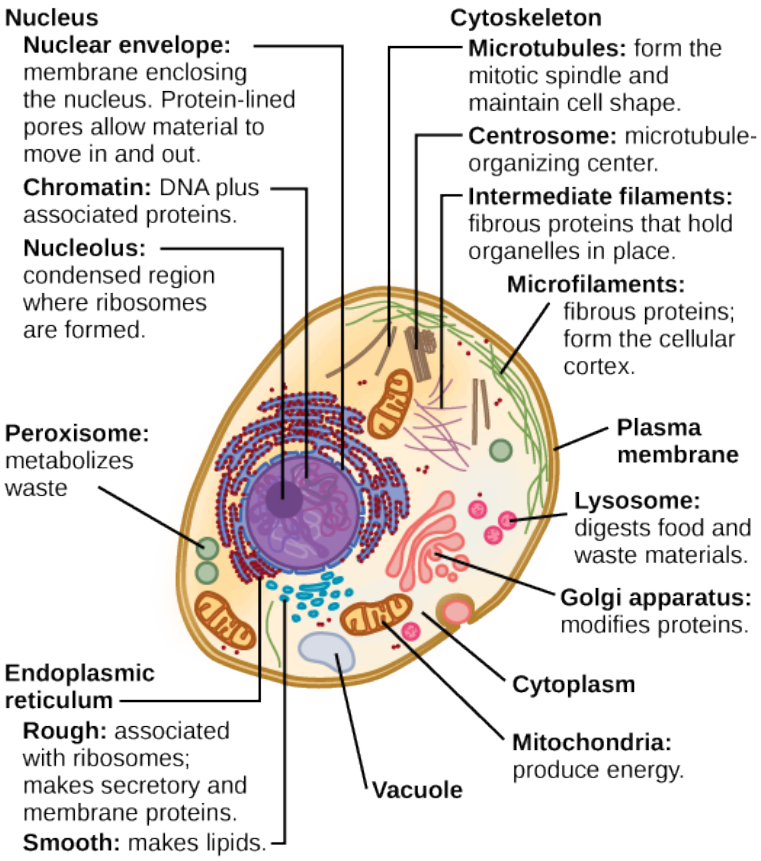


Figure 7.10a A typical animal cell, with the major organelles and other cell components

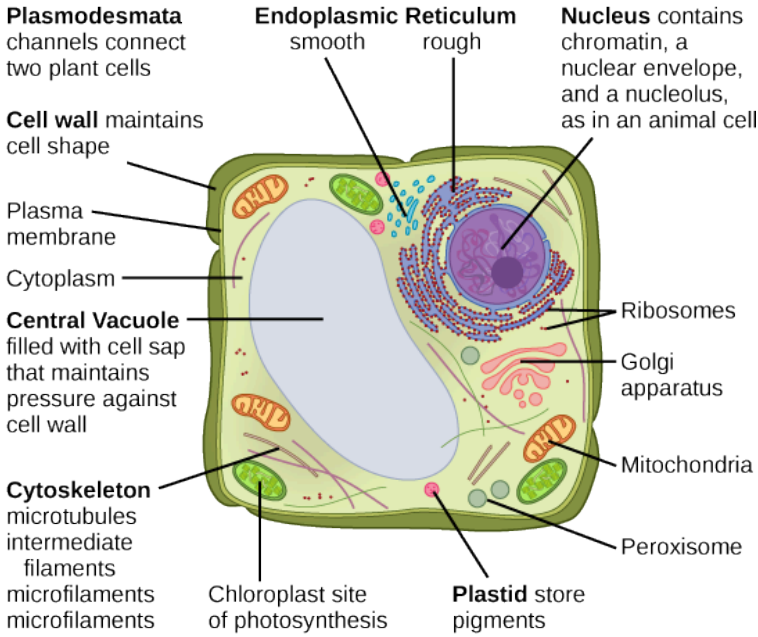


Figure 7.10b A typical plant cell. Plant cells have a cell wall, chloroplasts, plastids, and a central vacuole—structures not found in animal cells. Plant cells do not have lysosomes or centrosomes.

The Nucleus

Typically, the nucleus is the most prominent organelle in a cell. The **nucleus** (plural = nuclei) houses the cell's DNA and directs the synthesis of ribosomes and proteins. Let's look at it in more detail (**Figure 7.11**).

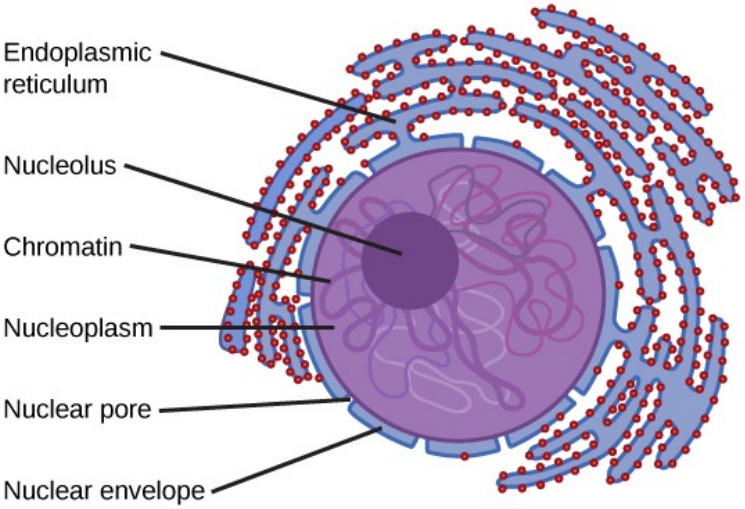


Figure 7.11 The nucleus stores chromatin (DNA plus proteins) in a gel-like substance called the nucleoplasm. The nucleolus is a condensed region of chromatin where ribosome synthesis occurs. The nucleus is enclosed by a double phospholipid bilayer called the nuclear envelope. Nuclear pores allow substances to enter and exit the nucleus.

The Nuclear Envelope

The **nuclear envelope** is a double-membrane structure that constitutes the outermost portion of the nucleus (**Figure 7.11**). Both the inner and outer membranes of the nuclear envelope are phospholipid bilayers. The nuclear envelope is punctuated with pores that control the passage of ions, molecules, and RNA between the nucleoplasm and cytoplasm. The nucleoplasm is the semi-solid fluid inside the nucleus.

Chromatin and Chromosomes

Chromosomes are structures within the nucleus that are made up of DNA, the hereditary material. You may remember that in prokaryotes, DNA is organized into a single circular chromosome. In eukaryotes, chromosomes are linear structures. Every eukaryotic species has a specific number of chromosomes in the nuclei of its body's cells. For example human cells have 46 chromosomes. Chromosomes are only visible and distinguishable from one another when the cell is getting ready to divide. When the cell is in the growth and maintenance phases of its life cycle, proteins are attached to chromosomes, and they resemble an unwound, jumbled bunch of threads. These unwound protein-chromosome complexes are called **chromatin** (**Figure 7.12**).

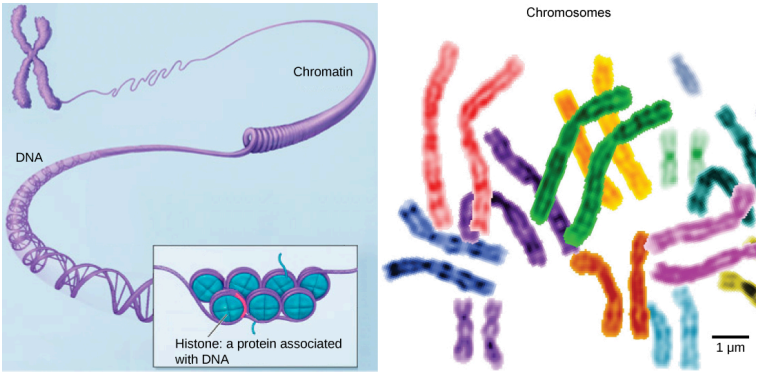


Figure 7.12 (a) Chromatin is composed of DNA and protein. (b) Paired eukaryotic chromosomes. (credit: modification of work by NIH; scale-bar data from Matt Russell)

The Nucleolus

We already know that the nucleus directs the synthesis of ribosomes, but how does it do this? Some chromosomes have sections of DNA that encode ribosomal RNA (rRNA). The darkly staining area within the nucleus called the **nucleolus** (plural = nucleoli) aggregates the ribosomal RNA with associated proteins to assemble ribosomal subunits that are then transported out through the pores in the nuclear envelope to the cytoplasm.

Ribosomes

Ribosomes are the cellular structures responsible for protein synthesis. They are large complexes of protein and rRNA, consisting of a large and a small subunit (**Figure 7.13**). When viewed through an electron microscope, ribosomes appear either as clusters or as single, tiny dots that float freely in the cytoplasm. They may be attached to the inside of the plasma membrane or the outside of the endoplasmic reticulum or nuclear envelope (**Figure 7.11**). Ribosomes bind to mRNA and translate the code provided by the sequence of the nitrogenous bases in the mRNA into a specific order of amino acids in a protein.

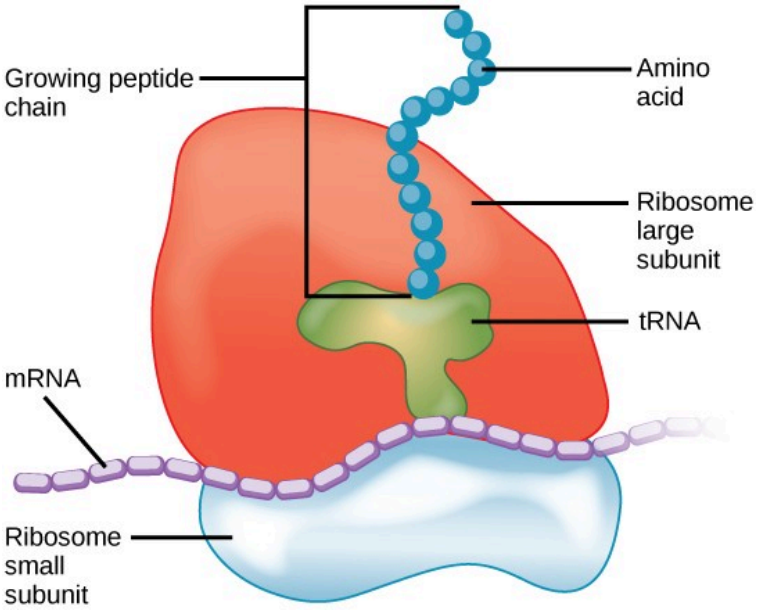


Figure 7.13 Ribosomes are made up of a large subunit (top) and a small subunit (bottom). During protein synthesis, ribosomes assemble amino acids into proteins.

Because proteins synthesis is an essential function of all cells, ribosomes are found in practically every cell. Ribosomes are particularly abundant in cells that synthesize large amounts of protein. For example, the pancreas is responsible for creating several digestive enzymes and the cells that produce these enzymes contain many ribosomes.

Mitochondria

Mitochondria (singular = mitochondrion) are often called

the "powerhouses" or "energy factories" of a cell because they are responsible for making adenosine triphosphate (ATP), the cell's main energy-carrying molecule. ATP is made using the chemical energy found in glucose and other nutrients by the process of cellular respiration. In mitochondria, this process uses oxygen and produces carbon dioxide as a waste product. In fact, the carbon dioxide that you exhale with every breath comes from the breakdown of nutrients you eat inside mitochondria.

Mitochondria are oval-shaped, double membrane organelles (**Figure 7.14**) that have their own ribosomes and DNA. Each membrane is a phospholipid bilayer embedded with proteins. The inner membrane has folds called cristae that increase surface area. The space between the two membranes is called the intermembrane space and the space inside the inner membrane is called the matrix. ATP synthesis takes place on the inner membrane.

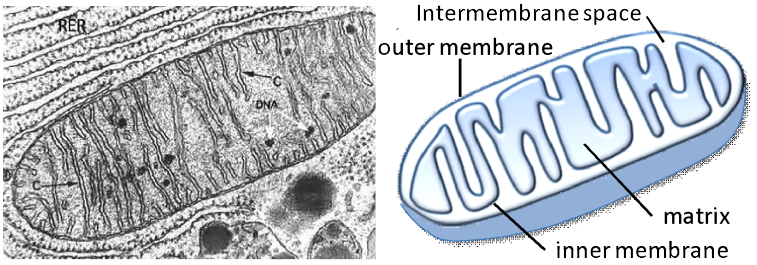


Figure 7.14 At left a transmission electron micrograph and at right a labeled diagram of a mitochondrion.

Peroxisomes

Peroxisomes are small, spherical organelles enclosed by single membranes. They carry out oxidation reactions that break down fatty acids and amino acids. They also detoxify many poisons that may enter the body. (Many of these oxidation 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 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.

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. Some types of cells have contractile vacuoles, which contract to expel water from the cell.

7.,3.4 Animal Cells versus Plant Cells

At this point, you know that each eukaryotic cell has a plasma membrane, cytoplasm, a nucleus, ribosomes, mitochondria, peroxisomes, and in some, vacuoles, but there are some striking differences between animal and plant cells. While both animal and plant cells have microtubule organizing centers (MTOCs), animal cells also have centrioles associated with the MTOC: a complex called the centrosome. Animal cells each have a centrosome and lysosomes, whereas plant cells do not. Plant cells have a cell wall, chloroplasts and other specialized plastids, and a large central vacuole, whereas animal cells do not.

The Centrosome

The **centrosome** is a microtubule-organizing center from which all microtubules in the cell originate. It is found near the nucleus of animal cells. Each centrosome contains a pair of **centrioles**, which lie perpendicular to each other (**Figure 7.15**). Each centriole is a cylinder of nine triplets of microtubules.

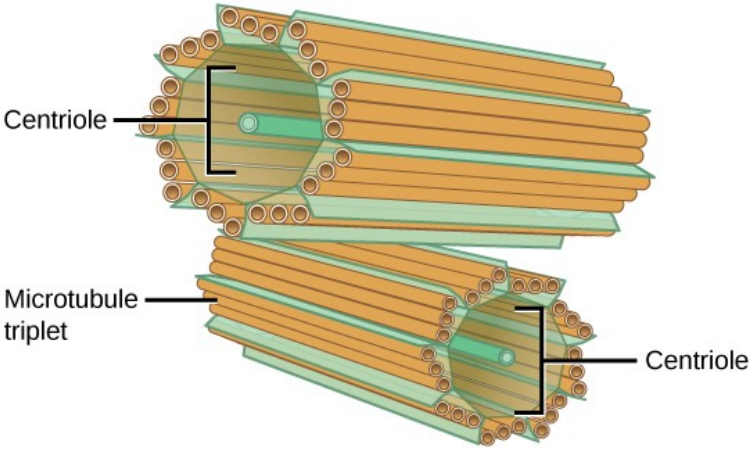


Figure 7.15 The centrosome consists of two centrioles that lie at right angles to each other. Each centriole is a cylinder made up of nine triplets of microtubules. Other proteins (indicated by the green lines) hold the microtubule triplets together.

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. However, the exact function of the centrioles in cell division isn't clear, because cells that have had the centrosome removed can still divide, and plant cells, which lack centrosomes, are capable of cell division.

Lysosomes

Animal cells have another set of organelles not found in plant cells: lysosomes. **Lysosomes** are the cell's "garbage disposal."

Enzymes inside 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. Proton pumps in the lysosome membrane pump protons in to maintain the low pH inside the lysosome. In plant cells, digestive processes take place in vacuoles.

The Cell Wall

If you examine **Figure 7.10b**, the diagram of a plant cell, you will see 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. While the chief component of prokaryotic cell walls is peptidoglycan, the major organic molecule in the plant cell wall is cellulose (**Figure 7.16**), a polysaccharide made up of glucose units. When you bite into a raw vegetable, like celery, it crunches because you are tearing the rigid cell walls of the celery cells with your teeth. Fungus and some protist cells also have cell walls, which are made of other structural molecules.

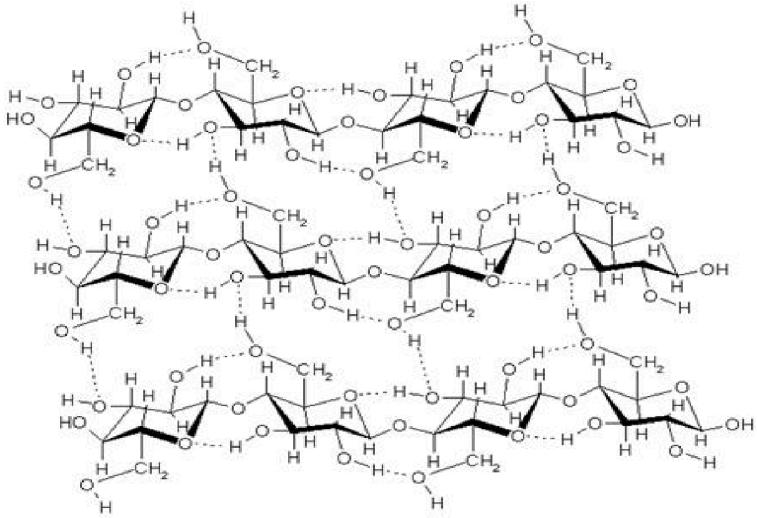


Figure 7.16 Cellulose is made of long chains of β-glucose molecules connected by a 1-4 linkage. These chains are crosslinked to form strong sheets that form plant cell walls.

Chloroplasts

Like mitochondria, chloroplasts have their own DNA and ribosomes, but chloroplasts have an entirely different function. **Chloroplasts** are plant cell organelles that carry out photosynthesis. Photosynthesis is the series of reactions that use carbon dioxide, water, and light energy to make glucose and oxygen. This is a major difference between plants and animals; plants (autotrophs) are able to make their own food, like sugars, while animals (heterotrophs) must ingest their food.

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 7.17**). 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. The fluid inside the thylakoids is called the thylakoid space. Light energy is harvested in the thylakoid membranes and sugar is made in the stroma.

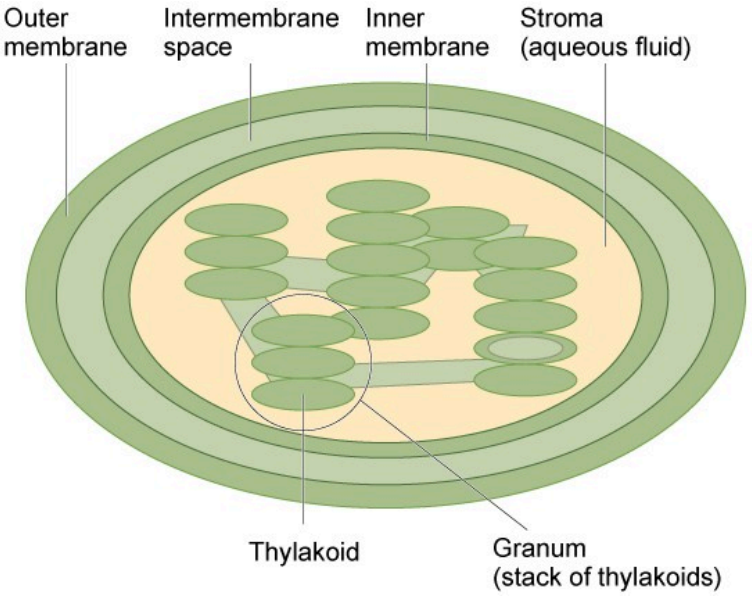


Figure 7.17 Chloroplasts have an outer membrane, an inner membrane, and thylakoids that are stacked into grana.

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 photosynthetic pigments are not relegated to an organelle.

The Central Vacuole

Plant cells each have a large **central vacuole** that occupies most of the area of the cell (**Figure 7.8b**). The central vacuole plays a key role in regulating the cell's concentration of water in changing environmental conditions. If you forget to water a plant for a few days, it wilts because when environmental water levels are low, water moves out of the central vacuoles and cytoplasm. As the central vacuole shrinks, it leaves the cell wall unsupported, resulting in wilting. 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.

evolution CONNECTION

Endosymbiosis

We have mentioned that both mitochondria and

chloroplasts contain DNA and ribosomes. Have you wondered why? Strong evidence points to endosymbiosis as the explanation.

Symbiosis is a relationship in which organisms from two separate species depend on each other for their survival. Endosymbiosis (endo- = "within") is a mutually beneficial relationship in which one organism lives inside the other. Endosymbiotic relationships abound in nature. We have already mentioned that microbes that produce vitamin K live inside the human gut. This relationship is beneficial for us because we are unable to synthesize vitamin K. It is also beneficial for the microbes because they are protected from other organisms and from drying out, and they receive abundant food from the environment of the large intestine.

Scientists have long noticed that bacteria, mitochondria, and chloroplasts are similar in size. We also know that bacteria have DNA and ribosomes, just as mitochondria and chloroplasts do. Scientists believe that host cells and bacteria formed an endosymbiotic relationship when the host cells ingested both aerobic and autotrophic

bacteria (cyanobacteria) but did not destroy them. Through many millions of years of evolution, these ingested bacteria became more specialized in their functions, with the aerobic bacteria becoming mitochondria and the autotrophic bacteria becoming chloroplasts.

7.4 | The Endomembrane System

Learning Objectives

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

- List the components of the endomembrane system.
- Recognize the relationship between the endomembrane system and its functions.

The **endomembrane system** (endo = "within") is a group of membranes and organelles (**Figure 7.18**) in eukaryotic cells that works together to modify, package, and transport lipids and proteins. It includes the nuclear envelope, lysosomes, and vesicles, which we've already mentioned, as well as the endoplasmic reticulum and Golgi apparatus. Although not technically *within* the cell, the plasma membrane is sometimes included in the endomembrane system because it interacts with the other endomembranous organelles.

7.4.1 The Endoplasmic Reticulum

The **endoplasmic reticulum (ER)** 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.

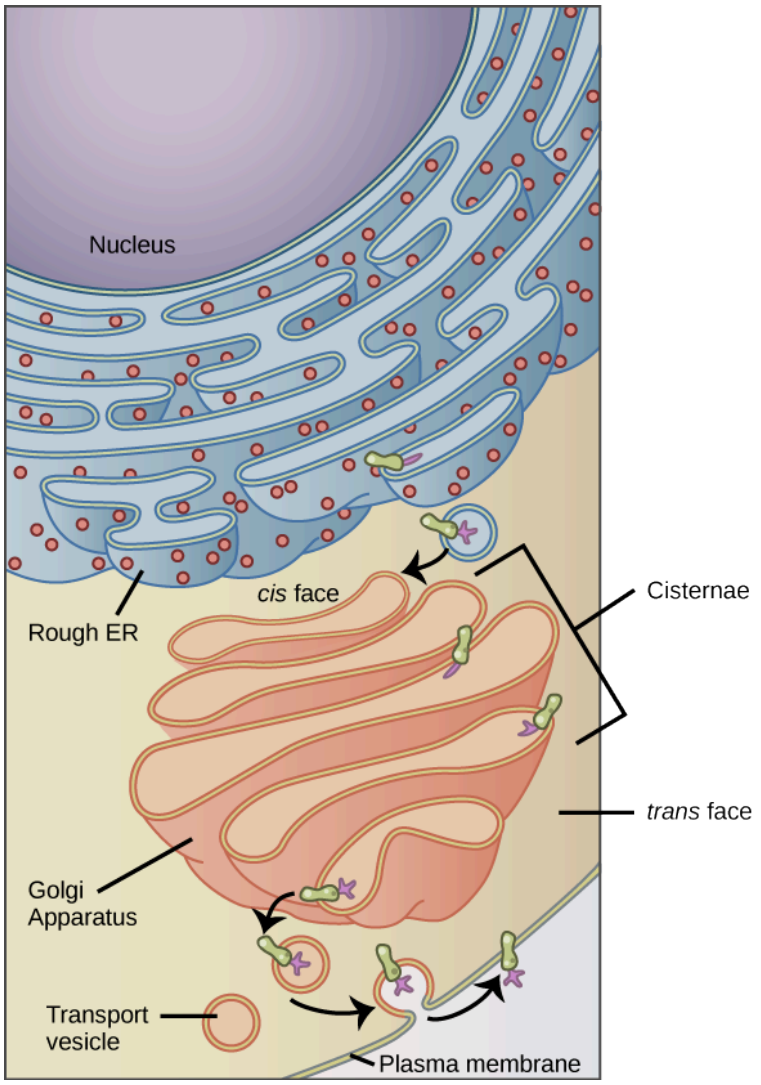


Figure 7.18 Membrane and secretory proteins are synthesized on bound ribosomes into the rough ER lumen. They then travel to the Golgi apparatus by vesicle. In the Golgi, they are modified and sorted before being sent to their final destination. (Credit: modification of work by Magnus Manske)

Rough ER

The **rough endoplasmic** reticulum **(RER)** is so named because ribosomes attached to its cytoplasmic surface give it a studded appearance when viewed through an electron microscope (**Figure 7.19**). These **bound ribosomes** are in the process of translating proteins directly into the lumen of the RER.

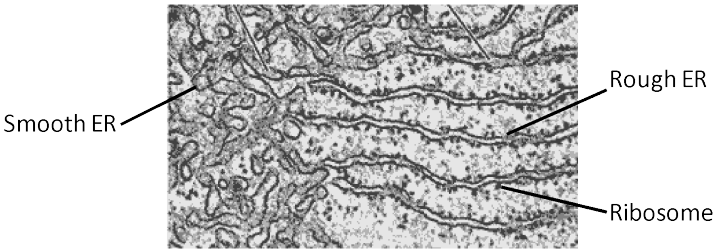


Figure 7.19 This transmission electron micrograph shows the rough and smooth endoplasmic reticulum.

Three types of proteins are made on bound ribosomes: proteins that will end up secreted from the cell: proteins that will end up inserted into the plasma membrane of the cell, and proteins that will end up inside organelles in the cell. What these three classes of proteins have in common is that they will not end up in the cytosol of the cell. From the moment that they start to be synthesized, these proteins are sequestered inside membrane-bound compartments and will never touch the cytosol.

Once the newly synthesized proteins are deposited into the lumen of the RER, they undergo structural modifications, such as folding or addition of side chains. If the modified proteins are not destined to stay in the RER, they will travel by small, membrane-bound transport vesicles that bud from the RER's membrane to the Golgi apparatus (**Figure 7.18**).

Since the RER is engaged in modifying proteins that will be secreted from the cell, the RER is abundant in cells that secrete proteins, such as liver cells, for example.

Smooth ER

The **smooth endoplasmic reticulum (SER)** is continuous with the RER but has few or no ribosomes on its surface (**Figure 7.19**). Functions of the SER include synthesis of carbohydrates, lipids, phospholipids, and steroid hormones and detoxification of medications and poisons. The SER takes on different functions depending on the needs of the cell. For example, in muscle cells, a specialized SER called the sarcoplasmic reticulum stores calcium ions that are needed to trigger the coordinated contractions of the muscle cells. Cells that make a lot of lipids have a large amount of SER. For example, Leydig cells in mammalian testes produce steroid hormones such as testosterone and therefore have abundant SER.

7.4.2 The Golgi Apparatus

Vesicles containing proteins and lipids bud from the ER and transport their contents to the **Golgi apparatus (GA).** In the GA, proteins and lipids are sorted, packaged, and tagged so that they wind up in the right place. For this reason, the GA is sometimes called the post office of the cell. The GA is a series of flattened membranes, each forming a separate compartment The receiving side of the Golgi apparatus is called the *cis* face. The opposite side is called the *trans* face (**Figure 7.20**).

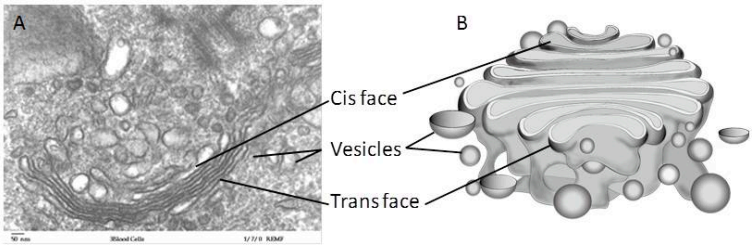


Figure 7.20 The Golgi apparatus in this white blood cell is visible as a stack of semicircular, flattened compartments. Several vesicles can be seen near the Golgi apparatus. (credit: A. modification of work by Louisa Howard. B. DBCLS 統合TV [CC BY 4.0 http://creativecommons.org/licenses/by/4.0)

Transport vesicles from the ER fuse with the *cis* face and empty their contents into the lumen of the GA. The proteins and lipids travel from compartment to compartment by vesicle. As they travel through the GA, they undergo further modifications. The most frequent modification is the addition of short chains of sugar molecules. They may also be 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 vesicles that bud from the *trans* face of the Golgi. Some of these vesicles deposit their contents into other organelles of the cell. Other vesicles fuse with the plasma membrane and their contents end up outside the cell or inserted into the plasma membrane.

Cells that engage in a great deal of secretory activity, such as cells of the salivary glands that secrete digestive enzymes, have an abundance of GA. 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.

Cardiologist

Heart disease is the leading cause of death in the United States. This is primarily due to sedentary lifestyles and poor diets. Heart failure is just one of many disabling heart conditions. Heart failure occurs when the heart cannot pump with

sufficient force to transport oxygenated blood to all the vital organs.

Heart failure occurs when the endoplasmic reticula of cardiac muscle cells do not function properly. As a result, an insufficient number of calcium ions are available to trigger a sufficient contractile force.

Cardiologists (cardi- = "heart") are doctors who specialize in treating heart diseases. Cardiologists can diagnose heart failure via physical examination, results from an electrocardiogram (a test that measures the electrical activity of the heart), a chest X-ray to see whether the heart is enlarged, and other tests. If heart failure is diagnosed, the cardiologist will typically prescribe appropriate medications and recommend a reduction in table salt intake and a supervised exercise program.

7.4.3 Lysosomes

Lysosomes are also part of the endomembrane system. Lysosomes are small, spherical compartments that function as the digestive and organelle-recycling facility of animal cells,

They contain hydrolytic enzymes that digest non-functioning organelles, macromolecules, and pathogens (disease-causing organisms). The enzymes in lysosomes are called acid hydrolases because they catalyze hydrolysis reaction and their optimal pH is around 4.5. Lysosomes contain proton pumps in their membranes that pump hydrogen ions into the lumen, thereby lowering the pH. This elegant mechanism allows the enzymes to be safely made and transported through the endomembrane system in an inactive state. They become active only when they encounter the acidic environment inside the lysosome.

Lysosomes also use their hydrolytic enzymes to destroy pathogens that enter the cell. A good example of this occurs in a type of white blood cells called macrophages, which are part of your body's immune system. In a process known as phagocytosis, macrophages engulf pathogens into a vesicle. The vesicle containing the pathogen fuses with a lysosome, and the lysosome's enzymes destroy the pathogen (**Figure 7.21**).

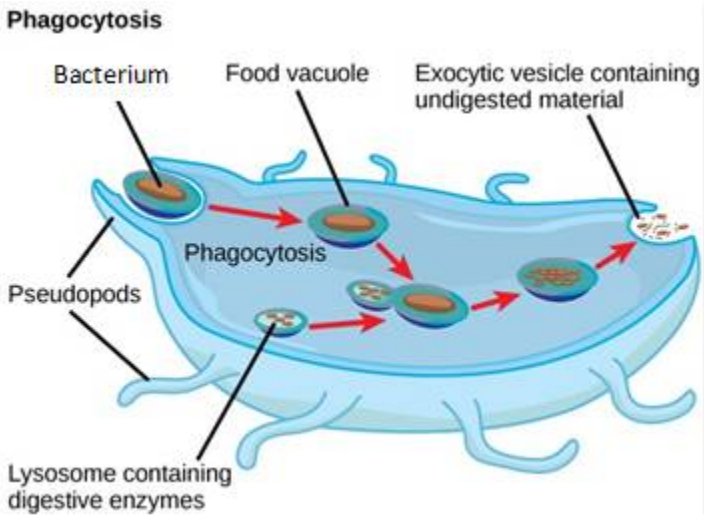


Figure 7.21 A macrophage has engulfed (phagocytized) a potentially pathogenic bacterium and then fuses with a lysosomes within the cell to destroy the pathogen. Other organelles are present in the cell but for simplicity are not shown.

7.5 | The Cytoskeleton

Learning Objectives

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

- Describe the cytoskeleton.
- Compare the roles of microfilaments, intermediate filaments, and microtubules.
- Compare and contrast cilia and flagella.
- Summarize the differences among the components of prokaryotic cells, animal cells, and plant cells.

If you were to remove all the organelles from a cell, would the plasma membrane and the cytoplasm be the only components left? No. Within the cytoplasm, there would still be ions and organic molecules, plus a network of protein fibers that help maintain the shape of the cell, secure some organelles in specific positions, allow cytoplasm and vesicles to move within the cell, and enable cells within multicellular organisms to move. Collectively, this network of protein fibers is known as the **cytoskeleton**. There are three types of fibers within the cytoskeleton: microfilaments, intermediate filaments, and microtubules (**Figure 7.22**). Here, we will examine each.

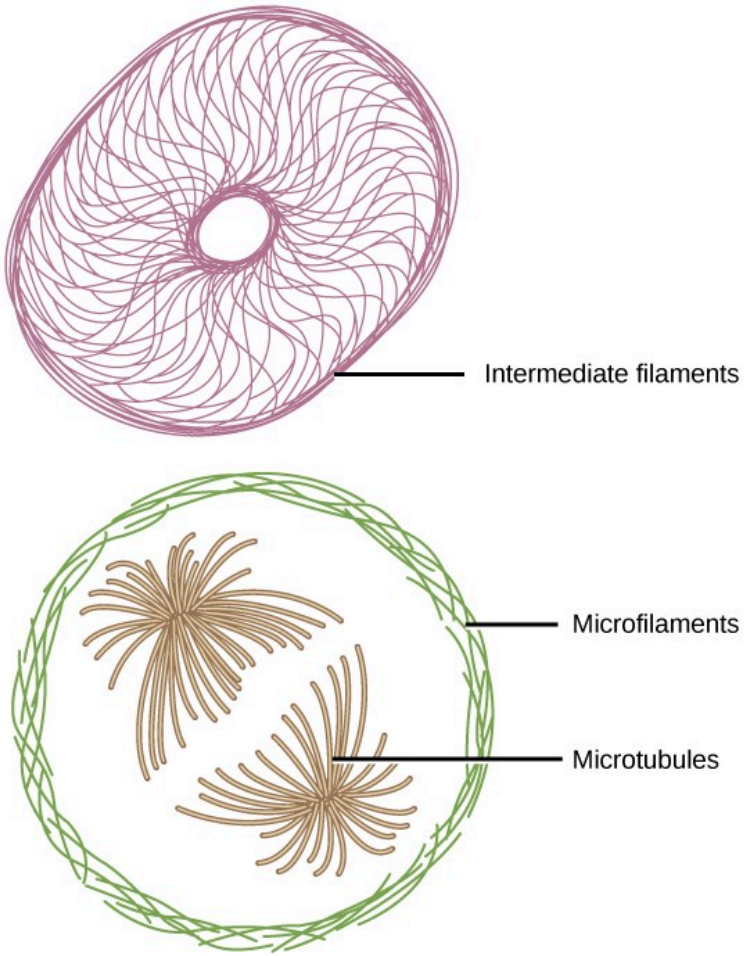


Figure 7.22 Microfilaments thicken the cortex around the inner edge of a cell; like rubber bands, they resist tension. Microtubules are found in the interior of the cell where they maintain cell shape by resisting compressive forces. Intermediate filaments are found throughout the cell and hold organelles in place.

7.5.1 Microfilaments

Of the three types of protein fibers in the cytoskeleton, **microfilaments** are the thinnest. They function in cellular movement and are made of two intertwined strands of a globular protein called **actin** (**Figure 7.23**). For this reason, microfilaments are also known as actin filaments.

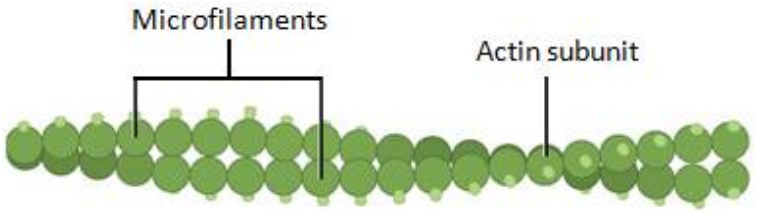


Figure 7.23 Microfilaments are made of two intertwined strands of actin protein.

Microfilaments provide some rigidity and shape to cells. They can depolymerize (disassemble) and reform quickly, thus enabling a cell to change its shape and move. Microfilaments are also involved in cellular movement, such as cell division in animal cells and cytoplasmic streaming, which is the circular movement of the cell cytoplasm, in plant cells. Actin also helps muscle cells contract.

7.5.2 Intermediate Filaments

Intermediate filaments are made of several strands of fibrous proteins that are wound together (**Figure 7.24**). These elements of the cytoskeleton get their name from the fact that their diameter is between those of microfilaments and microtubules. The function of intermediate filaments is purely structural. They bear tension, thus maintaining the shape of the cell, and create a supportive scaffolding to anchor the nucleus and other organelles in place.

Figure 7.24 Intermediate filaments consist of several intertwined strands of fibrous proteins.

Intermediate filaments are the most diverse group of cytoskeletal elements. Several types of fibrous proteins are found in the intermediate filaments, including keratin, the fibrous protein that strengthens your hair, nails, and skin.

7.5.3 Microtubules

As their name implies, microtubules are small hollow tubes. The walls of the microtubule are made of dimers of α-tubulin

and β-tubulin, two globular proteins (**Figure 7.25**). **Microtubules** are the widest components of the cytoskeleton. They help the cell resist compression, provide a track along which vesicles move through the cell, and pull replicated chromosomes to opposite ends of a dividing cell. Like microfilaments, microtubules can dissolve and reform quickly.

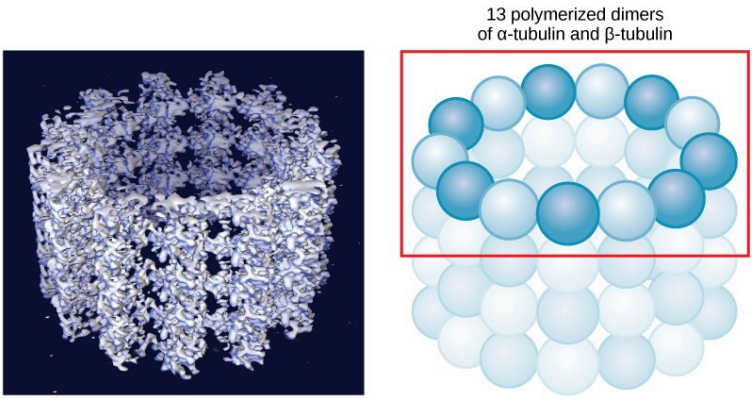


Figure 7.25 Microtubules are hollow. Their walls consist of 13 polymerized dimers of α-tubulin and β-tubulin (right image). The left image shows the molecular structure of the tube.

Microtubules are also the structural elements of flagella, cilia, and centrioles. In fact, in animal cells, the centrioles are the microtubule-organizing center. In eukaryotic cells, flagella and cilia are quite different structurally from their counterparts in prokaryotes, as discussed below.

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. When present, a cell has just one flagellum or a few flagella. **Cilia** (singular = cilium) are short, hair-like structures that are used to move entire cells or to move substances along the outer surface of the cell. For example, the cells lining the ovarian tubes have cilia that move the ovum toward the uterus, and the cells lining the respiratory tract have cilia that move mucus toward your nostrils. When cilia are present, they extend along the entire surface of the plasma membrane.

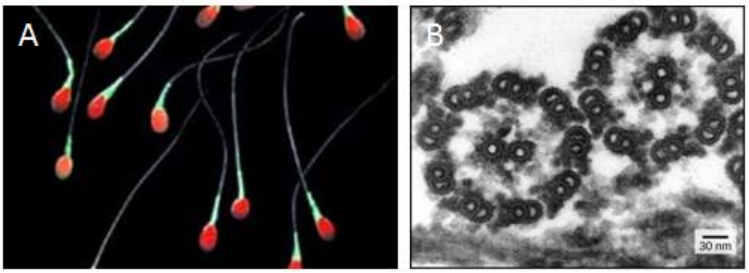


Figure 7.26 A. Light micrograph of human sperm, showing their long flagella B. Transmission electron micrograph of two flagella shows the 9 + 2 array of microtubules. (credit: modification of work by Dartmouth Electron Microscope Facility, Dartmouth College; scale-bar data from Matt Russell)

Despite their differences in length and number, flagella and

cilia share a common structural arrangement of microtubules called a "9 + 2 array." A single flagellum or cilium is made of a ring of nine microtubule pairs surrounding a single microtubule pair in the center (**Figure 7.26**).

Table 7.1 Components of prokaryotic and eukaryotic cells.

Component	Function
Plasma membrane	Separates cell from external environment; controls passage of organic molecules, ions, water, oxygen, and wastes in and out of cell
Cytoplasm	Provides turgor pressure to plant cells as fluid inside the central vacuole; site of many metabolic reactions; medium in which organelles are found
Nucleus	Cell organelle that houses DNA and directs synthesis of ribosomes and proteins
Nucleolus	Darkened area within the nucleus where ribosomal subunits are synthesized.
Ribosomes	Protein Synthesis
Mitochondria	ATP production/cellular respiration
Peroxisomes	Oxidizes and thus breaks down fatty acids and amino acids, and detoxifies poisons
Vesicles and Vacuoles	Storage and transport; digestive function in plant cells
Centrosome	Unspecified role in cell division in animal cells; source of microtubules in animal cells

Lysosome	Digestion of macromolecules; recycling of worn-out organelles
Cell Wall	Protection, structural support and maintenance of cell shape
Chloroplasts	Photosynthesis
Endoplasmic Reticulum	Modifies proteins and synthesizes lipids
Golgi Apparatus	Modifies, sorts, tags, packages, and distributes lipids and proteins
Cytoskeleton	Maintains cell's shape, secures organelles in specific positions, allows cytoplasm and vesicles to move within cell, and enables unicellular organisms to move independently
Flagella	Cellular locomotion
Cilia	Cellular locomotion, movement of particles along extracellular surface of plasma membrane, and filtration

7.6 | Connections Between Cells

Learning Objectives

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

- Describe the extracellular matrix.
- List examples of the ways that plant cells and animal cells communicate with adjacent cells.
- Summarize the roles of tight junctions, desmosomes, gap junctions, and plasmodesmata.

You already know that a group of similar cells working together is called a tissue. As you might expect, if cells are to work together, they must communicate with each other, just as you need to communicate with others if you work on a group project. Let's take a look at how cells communicate with each other.

7.6.1 Extracellular Matrix of

Animal Cells

Most animal cells release materials into the extracellular space. The primary components of these materials are proteins, and the most abundant protein is collagen. Collagen fibers are interwoven with carbohydrate-containing protein molecules called proteoglycans. Collectively, these materials are called the extracellular matrix (**Figure 7.27**). Not only does the extracellular matrix hold the cells together to form a tissue, but it also allows the cells within the tissue to communicate with each other. How can this happen?

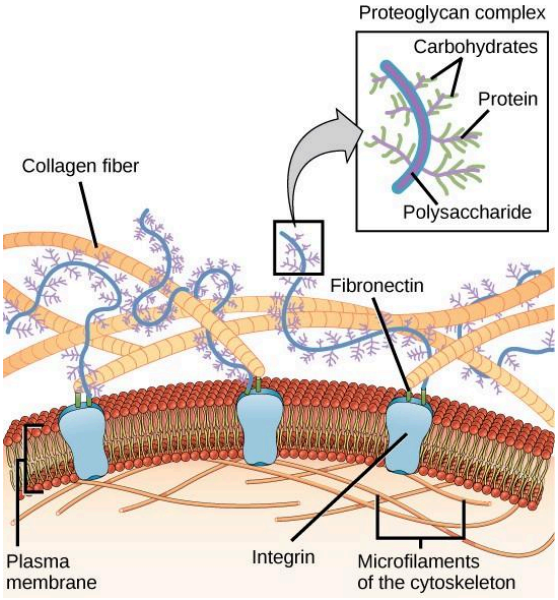


Figure 7.27 The extracellular matrix consists of a network of proteins and carbohydrates.

Cells have protein receptors on the extracellular surfaces of their plasma membranes. When a molecule within the matrix binds to the receptor, it changes the molecular structure of the receptor. The receptor, in turn, changes the conformation of the microfilaments positioned just inside the plasma membrane. These conformational changes induce chemical signals inside the cell that reach the nucleus and turn "on" or "off" the transcription of specific sections of DNA, which affects the production of associated proteins, thus changing the activities within the cell.

Blood clotting provides an example of the role of the extracellular matrix in cell communication. When the cells lining a blood vessel are damaged, they display a protein receptor called tissue factor. When tissue factor binds with another factor in the extracellular matrix, it causes platelets to adhere to the wall of the damaged blood vessel, stimulates the adjacent smooth muscle cells in the blood vessel to contract (thus constricting the blood vessel), and initiates a series of steps that stimulate the platelets to produce clotting factors.

7.6.2 Intercellular Junctions

Adjacent cells can also communicate with each directly through intercellular junctions. Plasmodesmata are junctions between plant cells; tight junctions, gap junctions, and desmosomes occur between animal cells.

Plasmodesmata

In general, plasma membranes of neighboring plant cells cannot touch one another because they are separated by the cell wall that surrounds each cell. Plasmodesmata (singular = plasmodesma) are numerous channels that pass between cell walls of adjacent plant cells, connect their cytoplasm, and enable materials to be transported from cell to cell (**Figure 7.28**). This allows plants to transfer water and other soil nutrients from its roots, through its stems, and to its leaves.

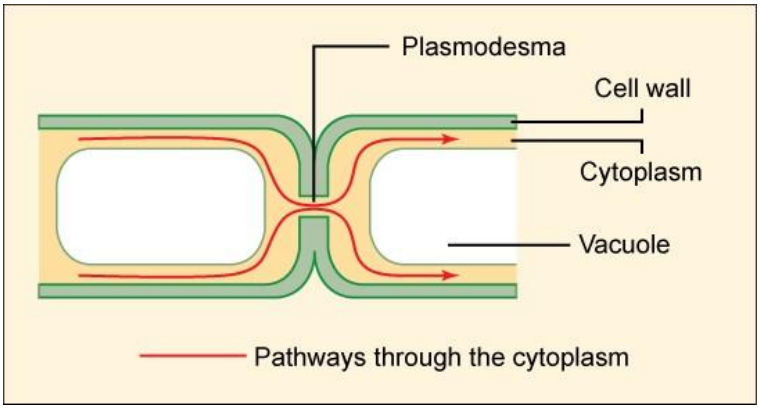


Figure 7.28 A plasmodesma is a channel between the cell walls of two adjacent plant cells. Plasmodesmata allow materials to pass from the cytoplasm of one plant cell to the cytoplasm of an adjacent cell.

Tight Junctions

A tight junction is a watertight seal between two adjacent

animal cells (**Figure 7.29A**). The cells are held tightly against each other by proteins, preventing materials from leaking between the cells. Tight junctions are typically found in tissues that line internal organs and cavities. For example, the tight junctions of the epithelial cells lining your urinary bladder prevent urine from leaking out into the extracellular space. Tight junctions in the gut normally prevent food from leaking between cells into your body. However, these junctions can be regulated to allow some substances to pass through.

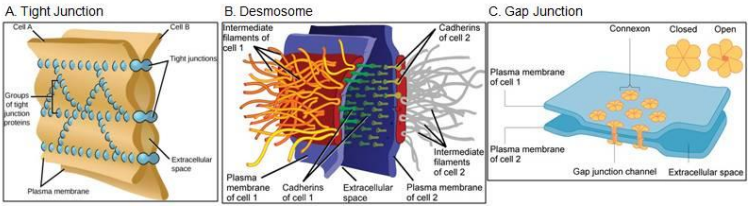


Figure 7.29 A. Tight junctions form watertight connections between adjacent animal cells. (Credit: modification of work by Mariana Ruiz Villareal) B. A desmosome forms a very strong spot weld between cells. (Credit: modification of work by Mariana Ruiz Villareal) C. A gap junction is a protein-lined pore that allows water and small molecules to pass between adjacent animal cells. (Credit: modification of work by Mariana Ruiz Villareal)

Desmosomes

Also found only in animal cells are desmosomes, which act like spot welds between adjacent epithelial cells (**Figure 7.29B**). Proteins in the plasma membrane connect to intermediate

filaments to create desmosomes. Effectively, the cytoskeletons of the two cells are linked together so they cannot easily be pulled apart. Skin is an example of a tissue with numerous desmosomes. This allows you to pull on your skin without it ripping apart.

Gap Junctions

Gap junctions in animal cells are similar to plant cell plasmodesmata. They are channels between adjacent cells that allow for the transport of ions, nutrients, and other substances that enable cells to communicate (**Figure 7.29C**). Structurally, however, gap junctions and plasmodesmata differ.

Gap junctions develop when a set of six proteins (called connexins) in the plasma membrane arrange themselves in a ring, called a connexon. When the pores of connexons in adjacent cells align, they form a channel between the two cells. Gap junctions are particularly important in cardiac muscle. The electrical signal for the muscle to contract is passed efficiently through gap junctions, allowing the heart muscle cells to contract in unison.

CHAPTER 8. MEMBRANE TRANSPORT

Figure 8.1 Despite its seeming hustle and bustle, Grand Central Station functions with a high level of organization: People and objects move from one location to another, they cross or are contained within certain boundaries, and they provide a constant flow as part of larger activity. Analogously, a plasma membrane's functions involve movement within the cell and across boundaries in the process of intracellular and intercellular activities. (Credit: modification of work by Randy Le'Moine)

Chapter Outline

- 8.1 Membrane Components and Structure
- 8.2 Passive Transport
- 8.3 Active Transport
- 8.4 Bulk Transport

Introduction

The plasma membrane, which is also called the cell membrane, has many functions, but the most basic one is to define the borders of the cell and keep the cell functional. The plasma membrane is selectively permeable. This means that the membrane allows some materials to freely enter or leave the cell, while other materials cannot move freely, but require the use of a specialized structure, and occasionally, even energy investment for crossing.

8.1 | Membrane Components and Structure

Learning Objectives

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

- Understand the fluid mosaic model of cell membranes.
- Describe the functions of phospholipids, proteins, and carbohydrates in membranes.
- Discuss membrane fluidity.

A cell's plasma membrane defines the cell, outlines its borders, and determines the nature of its interaction with its environment (see Table 8.1 for a summary). Cells exclude some substances, take in others, and excrete still others, all in controlled quantities. The plasma membrane must be very flexible to allow certain cells, such as red blood cells and white blood cells, to change shape as they pass through narrow capillaries. These are the more obvious functions of a plasma membrane. In addition, the surface of the plasma membrane

carries markers that allow cells to recognize one another, which is vital for tissue and organ formation during early development, and which later plays a role in the "self" versus "non-self" distinction of the immune response.

Among the most sophisticated functions of the plasma membrane is the ability to transmit signals by means of complex, integral proteins known as receptors. These proteins act both as receivers of extracellular inputs and as activators of intracellular processes. These membrane receptors provide extracellular attachment sites for effectors like hormones and growth factors, and they activate intracellular response cascades when their effectors are bound. Occasionally, receptors are hijacked by viruses (HIV, human immunodeficiency virus, is one example) that use them to gain entry into cells, and at times, the genes encoding receptors become mutated, causing the process of signal transduction to malfunction with disastrous consequences.

8.1.1 Fluid Mosaic Model

The existence of the plasma membrane was identified in the 1890s, and its chemical components were identified in 1915. The principal components identified at that time were lipids and proteins. The first widely accepted model of the plasma membrane's structure was proposed in 1935 by Hugh Davson and James Danielli; it was based on the "railroad track" appearance of the plasma membrane in early electron

micrographs. They theorized that the structure of the plasma membrane resembles a sandwich, with protein being analogous to the bread, and lipids being analogous to the filling. In the 1950s, advances in microscopy, notably transmission electron microscopy (TEM), allowed researchers to see that the core of the plasma membrane consisted of a double, rather than a single, layer. A new model that better explains both the microscopic observations and the function of that plasma membrane was proposed by S.J. Singer and Garth L. Nicolson in 1972.

The explanation proposed by Singer and Nicolson is called the **fluid mosaic model**. The model has evolved somewhat over time, but it still best accounts for the structure and functions of the plasma membrane as we now understand them. The fluid mosaic model describes the structure of the plasma membrane as a mosaic of components—including phospholipids, cholesterol, proteins, and carbohydrates—that gives the membrane a fluid character. 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. The membrane does look a bit like a sandwich (**Figure 8.2**).

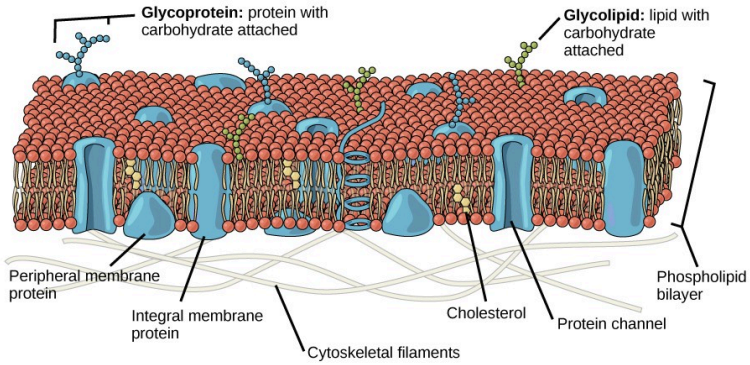


Figure 8.2 The fluid mosaic model of the plasma membrane describes the plasma membrane as a fluid combination of phospholipids, cholesterol, and proteins. Carbohydrates attached to lipids (glycolipids) and to proteins (glycoproteins) extend from the outward-facing surface of the membrane.

The principal components of a plasma membrane are lipids, proteins, and carbohydrates. The lipids include phospholipids and cholesterol Proteins either float in the bilayer or are attached to one side or the other of it. Carbohydrate chains are attached to the proteins and lipids on the outside surface of the membrane. The proportions of proteins, lipids, and carbohydrates in the plasma membrane vary with cell type, but for a typical human cell, protein accounts for about 50 percent of the composition by mass, lipids account for about 40 percent of the composition by mass, with the remaining 10 percent of the composition by mass being carbohydrates.

Phospholipids

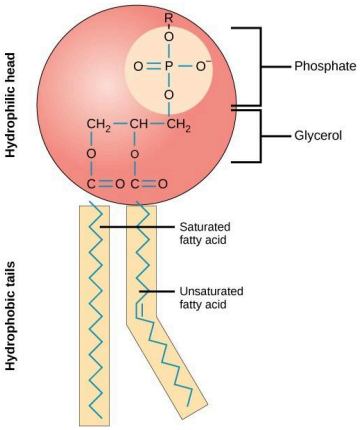
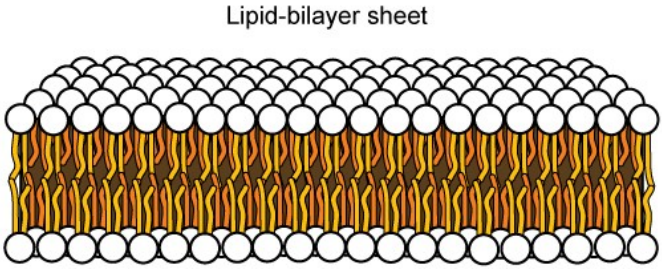


Figure 8.3 This phospholipid molecule is composed of a hydrophilic phosphate group head and two hydrophobic fatty acid tails.

The main fabric of the membrane is composed of **amphiphilic** phospholipid molecules. Recall from chapter 4 that a phospholipid is a molecule consisting of glycerol, two fatty acids, and a phosphate-linked head group (**Figure 8.3**) . The **hydrophilic** "head" of these molecules are in contact with the aqueous fluid both inside and outside the cell. The **hydrophobic** "tails" face each other in the inside of the bilayer. Therefore, phospholipids form an excellent two-layer cell membrane that separates fluid within the cell from the fluid outside of the cell (**Figure 8.2**).

The amphipathic nature of phospholipids is vital to the structure of a plasma membrane because, in water, phospholipids automatically become arranged with their hydrophobic tails facing each other and their hydrophilic heads facing out. In this way, they form a lipid bilayer—a barrier composed of a double layer of phospholipids that separates the water and other materials on one side of the

barrier from the water and other materials on the other side (**Figure 8.4 top**). In fact, phospholipids heated in an aqueous solution tend to spontaneously form small spheres or droplets called micelles, with their hydrophilic heads forming the exterior and their hydrophobic tails on the inside (**Figure 8.4 bottom**).



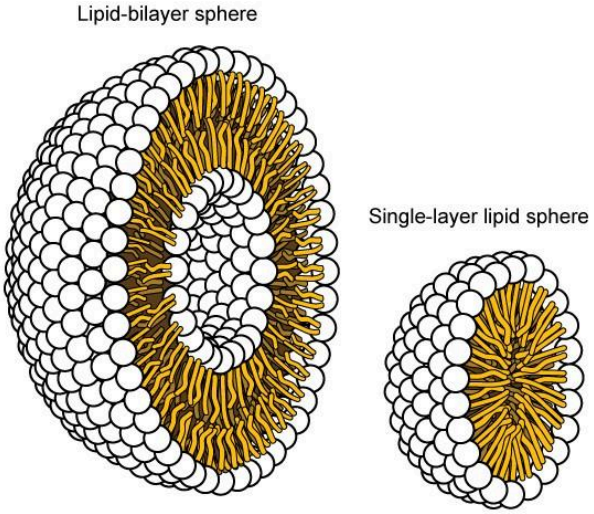


Figure 8.4 In an aqueous solution, phospholipids tend to arrange themselves with their polar heads facing outward and their hydrophobic tails facing inward. At high concentrations, they form a bilayer, such as the plasma membrane of cells (top). At lower concentrations, they form micelles (bottom). (Credit: modification of work by Mariana Ruiz Villareal)

Proteins

Proteins make up the second major component of plasma membranes. **Integral proteins** are, as their name suggests, integrated completely into the membrane structure, and their hydrophobic membrane-spanning regions interact with the hydrophobic region of the the phospholipid bilayer (**Figure 8.2**). Single-pass integral membrane proteins usually have a

hydrophobic transmembrane segment that consists of 20–25 amino acids. Some span only part of the membrane—associating with a single layer—while others stretch from one side of the membrane to the other, and are exposed on either side. Since they cross the membrane, these are often called **transmembrane proteins**.

Some complex integral proteins are composed of up to 12 segments, which are extensively folded and embedded in the membrane (**Figure 8.5**). This type of protein has a hydrophilic region or regions, and several hydrophobic regions. This arrangement of regions of the protein tends to orient the protein alongside the phospholipids, with the hydrophobic region of the protein adjacent to the tails of the phospholipids and the hydrophilic region or regions of the protein protruding from the membrane and in contact with the cytosol or extracellular fluid.

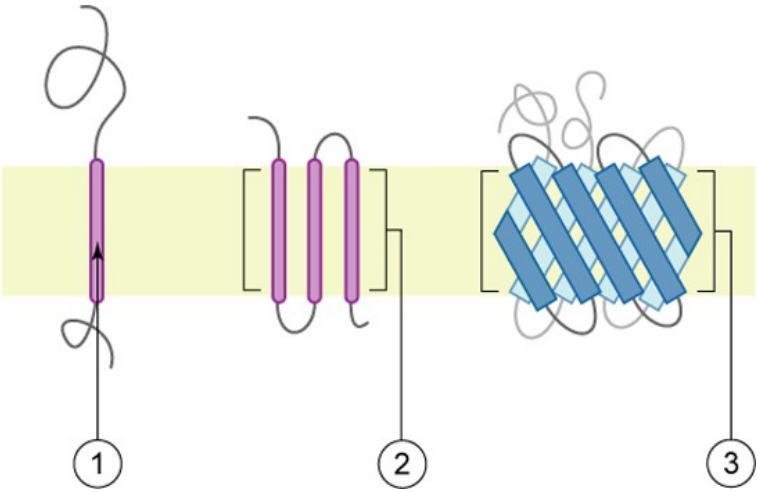


Figure 8.5 Integral membranes proteins may have one or more alpha-helices that span the membrane (examples 1 and 2), or they may have beta-sheets that span the membrane (example 3). (credit: "Foobar"/Wikimedia Commons)

Peripheral proteins are found on the exterior and interior surfaces of membranes, attached either to integral proteins or to phospholipids. Peripheral proteins, along with integral proteins, may serve as enzymes, as structural attachments for the fibers of the cytoskeleton, or as part of the cell's recognition sites. These are sometimes referred to as "cell-specific" proteins. The body recognizes its own proteins and attacks foreign proteins associated with invasive pathogens.

Carbohydrates

Carbohydrates are the 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**) (**Figure 8.2**). 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. These sites have unique patterns that allow the cell to be recognized, much the way that the facial features unique to each person allow him or her to be recognized. This recognition function is very important to cells, as it allows the immune system to differentiate between body cells (called "self") and foreign cells or tissues (called "non-self"). Similar types of glycoproteins and glycolipids are found on the surfaces of viruses and may change frequently, preventing immune cells from recognizing and attacking them.

These carbohydrates on the exterior surface of the cell—the carbohydrate components of both glycoproteins and glycolipids—are collectively referred to as the glycocalyx (meaning "sugar coating"). The glycocalyx is highly hydrophilic and attracts large amounts of water to the surface of the cell. This aids in the interaction of the cell with its watery environment and in the cell's ability to obtain substances dissolved in the water. As discussed above, the

glycocalyx is also important for cell identification, self/non-self determination, and embryonic development, and is used in cell-cell attachments to form tissues.

evolution CONNECTION

How Viruses Infect Specific Organs

Glycoprotein and glycolipid patterns on the surfaces of cells give many viruses an opportunity for infection. HIV and hepatitis viruses infect only specific organs or cells in the human body. HIV is able to penetrate the plasma membranes of a subtype of lymphocytes called T-helper cells, as well as some monocytes and central nervous system cells. The hepatitis virus attacks liver cells.

These viruses are able to invade these cells, because the cells have binding sites on their surfaces that are specific to and compatible with certain viruses (**Figure 8.6**). Other recognition sites on the virus's surface interact with the human immune system, prompting the body to produce antibodies.

Antibodies are made in response to the antigens or proteins associated with invasive pathogens, or in response to foreign cells, such as might occur with an organ transplant. These same sites serve as places for antibodies to attach and either destroy or inhibit the activity of the virus. Unfortunately, these recognition sites on HIV change at a rapid rate because of mutations, making the production of an effective vaccine against the virus very difficult, as the virus evolves and adapts. A person infected with HIV will quickly develop different populations, or variants, of the virus that are distinguished by differences in these recognition sites. This rapid change of surface markers decreases the effectiveness of the person's immune system in attacking the virus, because the antibodies will not recognize the new variations of the surface patterns. In the case of HIV, the problem is compounded by the fact that the virus specifically infects and destroys cells involved in the immune response, further incapacitating the host.

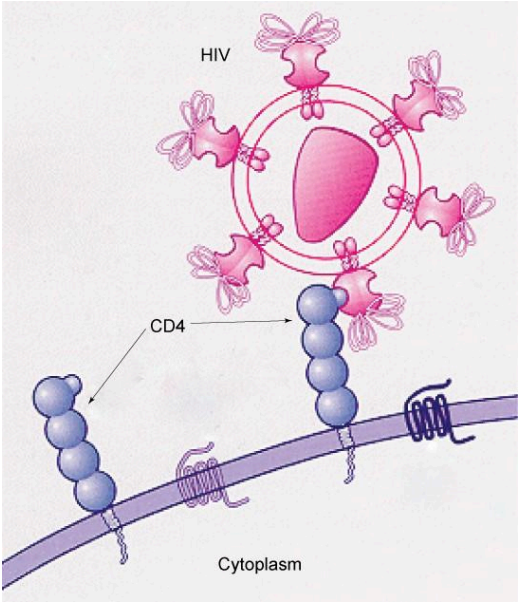


Figure 8.6 HIV binds to the CD4 receptor, a glycoprotein on the surfaces of T cells. (Credit: modification of work by NIH, NIAID)

8.1.2 Membrane Fluidity

The mosaic characteristic of the membrane, described in the fluid mosaic model, helps to illustrate its nature. The integral proteins and lipids exist in the membrane as separate but loosely attached molecules. These resemble the separate,

multicolored tiles of a mosaic picture, and they float, moving somewhat with respect to one another. The membrane is not like a balloon, however, that can expand and contract; rather, it is fairly rigid and can burst if penetrated or if a cell takes in too much water. However, because of its mosaic nature, a very fine needle can easily penetrate a plasma membrane without causing it to burst, and the membrane will flow and self-seal when the needle is extracted.

The mosaic characteristics of the membrane explain some but not all of its fluidity. 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 bound hydrogen atoms. There are no double bonds between adjacent carbon atoms. This results in tails that are relatively straight. In contrast, unsaturated fatty acids do not contain a maximal number of hydrogen atoms, but they do 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 8.3**).

Thus, if saturated fatty acids, with their straight tails, are compressed by decreasing temperatures, they press in on each other, making a dense and fairly rigid membrane. If unsaturated fatty acids are compressed, the "kinks" in their 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 saturated fatty acid tails in their phospholipids would "freeze" or solidify. The relative fluidity of the membrane is particularly important in a cold environment. A cold environment tends to compress membranes composed largely of saturated fatty acids, making them less fluid and more susceptible to rupturing. 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.

Animals have an additional membrane constituent that assists in maintaining fluidity. Cholesterol, which lies alongside the phospholipids in the membrane, tends to dampen the effects of temperature on the membrane. Thus, this lipid functions as a buffer, preventing lower temperatures from inhibiting fluidity and preventing increased temperatures from increasing fluidity too much. 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.

Table 8.1 The components and functions of the plasma membrane.

Component	Location
Phospholipid	Main fabric of the membrane
Cholesterol	Attached between phospholipids and between the two phospholipid layers
Integral proteins (for example, 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

Immunologist

The variations in peripheral proteins and

carbohydrates that affect a cell's recognition sites are of prime interest in immunology. These changes are taken into consideration in vaccine development. Many infectious diseases, such as smallpox, polio, diphtheria, and tetanus, were conquered by the use of vaccines.

dImmunologists are the physicians and scientists who research and develop vaccines, as well as treat and study allergies or other immune problems. Some immunologists study and treat autoimmune problems (diseases in which a person's immune system attacks his or her own cells or tissues, such as lupus) and immunodeficiencies, whether acquired (such as acquired immunodeficiency syndrome, or AIDS) or hereditary (such as severe combined immunodeficiency, or SCID). Immunologists are called in to help treat organ transplantation patients, who must have their immune systems suppressed so that their bodies will not reject a transplanted organ. Some immunologists work to understand natural immunity and the effects of a person's environment on it. Others work on questions about how the immune system affects diseases such as cancer. In the past, the

importance of having a healthy immune system in preventing cancer was not at all understood.

To work as an immunologist, a PhD or MD is required. In addition, immunologists undertake at least 2–3 years of training in an accredited program and must pass an examination given by the American Board of Allergy and Immunology. Immunologists must possess knowledge of the functions of the human body as they relate to issues beyond immunization, and knowledge of pharmacology and medical technology, such as medications, therapies, test materials, and surgical procedures.

8.2 | Passive Transport

Learning Objectives

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

- Explain why and how passive transport occurs.
- Understand the processes of osmosis and diffusion.
- Define tonicity and describe its relevance to passive transport.

Plasma membranes must allow certain substances to enter and leave a cell, and prevent some harmful materials from entering and some essential materials from leaving. In other words, plasma membranes are **selectively permeable**—they allow some substances to pass through, but not others. If they were to lose this selectivity, the cell would no longer be able to sustain itself, and it would be destroyed. Some cells require larger amounts of specific substances than do other cells; they must have a way of obtaining these materials from extracellular fluids. This may happen passively, as certain materials move back and forth, or the cell may have special mechanisms that facilitate transport. Some materials are so important to a cell that it spends some of its energy, hydrolyzing adenosine triphosphate (ATP), to obtain these materials. All cells spend the majority of their energy to maintain an imbalance of

sodium and potassium ions between the interior and exterior of the cell.

The most direct forms of membrane transport are passive. **Passive transport** is a naturally occurring phenomenon and does not require the cell to exert any of its energy to accomplish the movement. In passive transport, substances move from an area of higher concentration to an area of lower concentration. A physical space in which there is a range of concentrations of a single substance is said to have a **concentration gradient**.

8.2.1 Selective Permeability

Plasma membranes are asymmetric: the interior of the membrane is not identical to the exterior of the membrane. In fact, there is a considerable difference between the array of phospholipids and proteins between the two leaflets that form a membrane. On the interior of the membrane, some proteins serve to anchor the membrane to fibers of the cytoskeleton. There are peripheral proteins on the exterior of the membrane that bind elements of the extracellular matrix. Carbohydrates,

attached to lipids or proteins, are also found on the exterior surface of the plasma membrane. These carbohydrate complexes help the cell bind substances that the cell needs in the extracellular fluid. This adds considerably to the selective nature of plasma membranes (**Figure 8.7**).

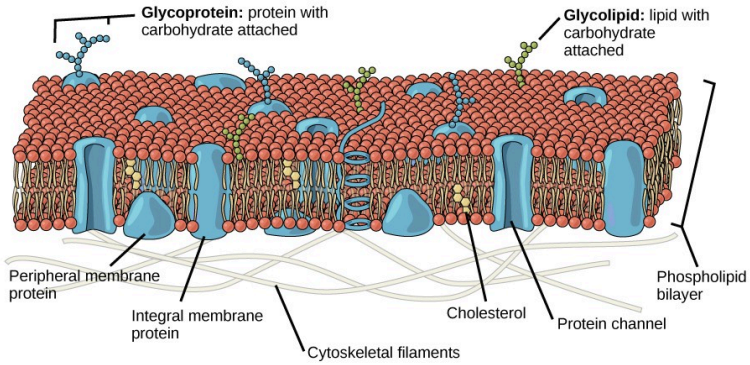


Figure 8.7 The exterior surface of the plasma membrane is not identical to the interior surface of the same membrane.

Recall that plasma membranes are amphipathic: They have hydrophilic and hydrophobic regions. This characteristic helps the movement of some materials through the membrane and hinders the movement of others. Lipid-soluble material with a low molecular weight can easily slip through the hydrophobic lipid core of the membrane. Substances such as the fat- soluble vitamins A, D, E, and K readily pass through the plasma membranes in the digestive tract and other tissues. Fat-soluble drugs and hormones also gain easy entry into cells and are readily transported into the body's tissues and organs. Molecules of oxygen and carbon dioxide have no charge and so pass through membranes by simple diffusion.

Polar substances present problems for the membrane. While some polar molecules connect easily with the outside of a cell, they cannot readily pass through the lipid core of the plasma membrane. Additionally, while small ions could easily slip

through the spaces in the mosaic of the membrane, their charge prevents them from doing so. Ions such as sodium, potassium, calcium, and chloride must have special means of penetrating plasma membranes. Larger polar molecules, such as simple sugars and amino acids also need help with transport across plasma membranes.

8.2.2 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 8.8**). Diffusion expends no energy. On the contrary, concentration gradients are a form of potential energy, dissipated as the gradient is eliminated.

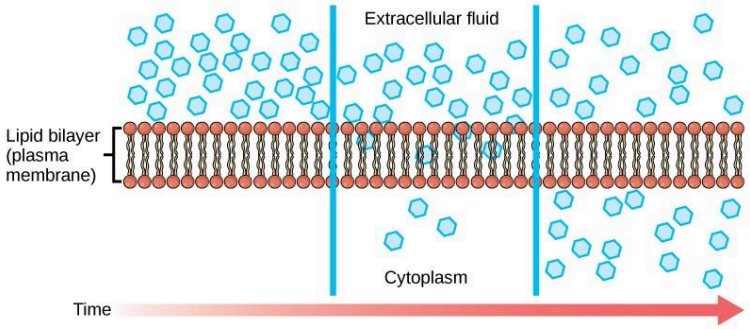


Figure 8.8 Diffusion through a permeable membrane moves a substance from an area of high concentration down its concentration gradient. (Credit: modification of work by Mariana Ruiz Villareal)

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.

Factors That Affect Diffusion

Molecules move constantly in a random manner, at a rate that depends on their mass, their environment, and the amount of thermal energy they possess, which in turn is a function of temperature. This movement accounts for the diffusion of molecules through whatever medium in which they are localized. A substance will tend to move into any space available to it until it is evenly distributed throughout it. After

a substance has diffused completely through a space, removing its concentration gradient, molecules will still move around in the space, but there will be no net movement of the number of molecules from one area to another. This lack of a concentration gradient in which there is no net movement of a substance is known as **dynamicequilibrium**. While diffusion will go forward in the presence of a concentration gradient of a substance, several factors affect the rate of diffusion.

"Steepness" 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.

Mass of the molecules diffusing: Heavier molecules move more slowly; therefore, they diffuse more slowly.

Temperature: Higher temperatures increase the energy and therefore the movement of the molecules, increasing 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, diffusion increases. Because cells primarily use diffusion to move materials within the cytoplasm, any increase in the cytoplasm's density will inhibit the movement of the materials. An example of this is a

person experiencing dehydration. As the body's cells lose water, the rate of diffusion decreases in the cytoplasm, and the cells' functions deteriorate. Neurons tend to be very sensitive to this effect. Dehydration frequently leads to unconsciousness and possibly coma because of the decrease in diffusion rate within the cells.

8.2.3 Facilitated diffusion

In **facilitated diffusion**, materials diffuse across the plasma membrane with the help of membrane proteins. A concentration gradient exists that would allow these materials to diffuse into the cell without expending cellular energy. However, these materials are ions or polar molecules that are repelled by the hydrophobic parts of the cell membrane. Facilitated diffusion proteins shield these materials from the repulsive force of the membrane, allowing them to diffuse into the cell. These proteins are called **transportproteins** and can be channels or carrier proteins.

Channels

Channelproteins are transmembrane proteins that fold in such as way as to form a channel or pore through the membrane. Each channel is specific for one particular substance. Channel proteins have hydrophilic domains exposed to the intracellular and extracellular fluids. In

addition, they have a hydrophilic channel through their core that provides a hydrated opening through the membrane layers (**Figure8.9**). 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. **Aquaporins** are channel proteins that allow water to pass through the membrane at a very high rate.

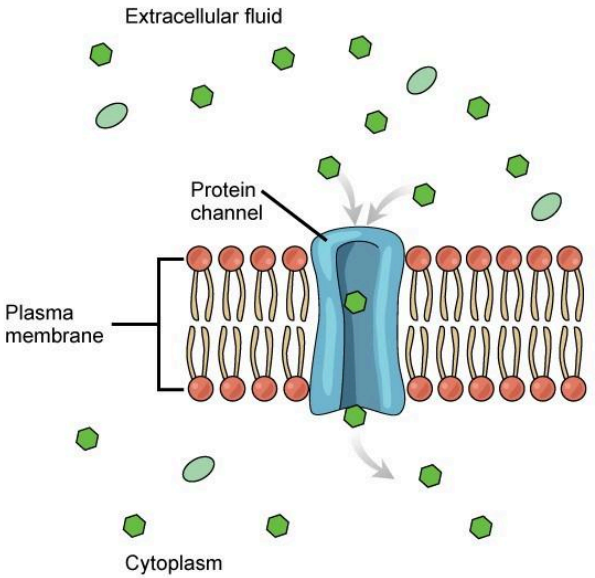


Figure 8.9 Facilitated transport moves substances down their concentration gradients. They may cross the plasma membrane with the aid of channel proteins. (Credit: modification of work by Mariana Ruiz Villareal)

Some channel proteins are always open but many are "gated,"

meaning that they can be opened and closed. If a channel is ligand-gated, the attachment of a particular molecule to the channel protein may cause it to open. Other channels are voltage-gated, requiring a change in voltage across the membrane to open them. Cells involved in the transmission of electrical impulses, such as nerve and muscle cells, have voltage-gated ion channels in their membranes.

Carrier Proteins

Another type of transmembrane transporter protein is a **carrierprotein**. Like channels, carrier proteins are usually specific for particular molecules. A carrier proteins binds a substance and, in doing so, triggers a change of its own shape, moving the bound molecule across the membrane (**Figure8.10**). Carrier proteins are used to transport molecules that are too large to pass through channels, such as amino acids and glucose.

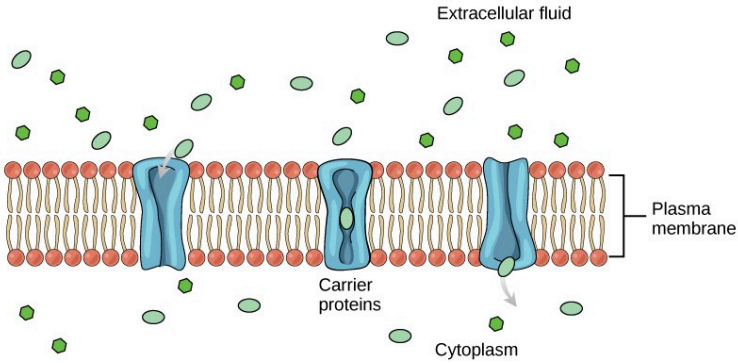


Figure 8.10 Some substances are able to move down their concentration gradient across the plasma membrane with the aid of carrier proteins. Carrier proteins change shape as they move molecules across the membrane. (Credit: modification of work by Mariana Ruiz Villareal)

There are a finite number of each type of carrier proteins in any membrane. This can cause problems in transporting enough of the material for the cell to function properly. When all of the proteins are bound to their ligands, they are saturated and the rate of transport is at its maximum. Increasing the concentration gradient at this point will not result in an increased rate of transport.

An example of this process occurs in the kidney. Glucose, water, salts, ions, and amino acids needed by the body are filtered out of the blood in one part of the kidney. This filtrate, which includes glucose, is then reabsorbed in another part of the kidney. Because there are only a finite number of carrier proteins for glucose, if more glucose is present than the proteins can handle, the excess is not transported 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 into cells within the body. The hormone insulin, increases the number of GLUTs on cells, causing them to take glucose from the blood when its levels are high. It is this process that is compromised in diabetic individuals.

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

8.2.4 Osmosis

Osmosis is the diffusion of water across a semipermeable membrane. Since it is diffusion, it depends on the concentration gradient, or the amount of water on each side of the membrane. The amount of water in a solute is inversely proportional to the concentration of solutes. In other words, the higher the concentration of water, the lower the concentration of solutes, and vice versa. Water can move readily across most membranes, due in part to the presence of aquaporins; however, the membrane limits the diffusion of solutes in the water.

Mechanism of Osmosis

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 (**Figure8.11**). 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.

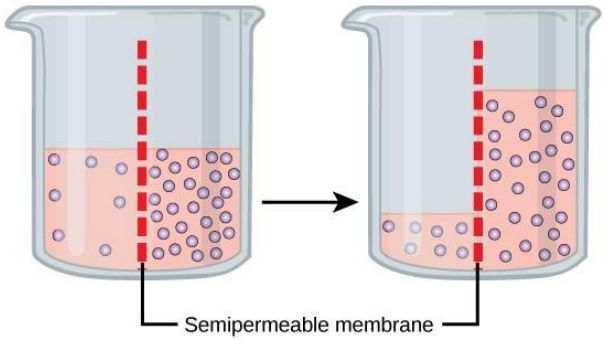


Figure 8.11 In osmosis, water always moves from an area of higher water concentration to one of lower concentration. In the diagram shown, the solute cannot pass through the selectively permeable membrane, but the water can.

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.

8.2.5 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 fewer 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.

Three terms—hypotonic, isotonic, and hypertonic—are

used to relate the osmolarity of a cell to the osmolarity of the extracellular fluid. In living systems, the point of reference is always the cytoplasm, so the prefix hypo- ("lower") means that the extracellular fluid has a lower concentration of solutes, or a lower osmolarity, than the cell cytoplasm. Blood cells and plant cells in hypertonic, isotonic, and hypotonic solutions take on characteristic appearances (**Figure 8.12**).

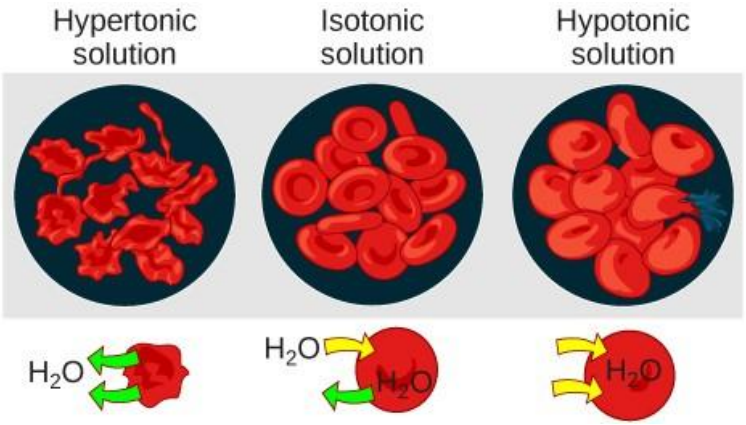


Figure 8.12 Osmotic pressure changes the shape of red blood cells in hypertonic, isotonic, and hypotonic solutions. (Credit: Mariana Ruiz Villareal)

Hypotonic Solutions

In a **hypotonic** situation, the extracellular fluid has lower osmolarity than the fluid inside the cell. The extracellular fluid has a higher concentration of water than does the cell and water will move down its concentration gradient and enter the cell.

Hypertonic Solutions

In a **hypertonic** solution (hyper- = "more"), the extracellular fluid has a higher osmolarity than the cell's cytoplasm. The fluid contains less water than the cell does, so water will leave the cell.

Isotonic Solutions

In an **isotonic** solution, the extracellular fluid has the same osmolarity as the cell. There is no net movement of water into or out of the cell (although water will still move in and out).

Concept Check

A doctor injects a patient with what the doctor thinks is an isotonic saline solution. The patient dies, and an autopsy reveals that many red blood cells have been destroyed. Do you think the solution the doctor injected was really isotonic?

8.2.6 Tonicity in Living Systems

A red blood cell will burst, or lyse, when it swells beyond the plasma membrane's capability to expand. In contrast, when excessive amounts of water leave a red blood cell, the cell shrinks, or crenates. Crenation 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 it may die. (**Figure 8.12**).

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 cells from lysing. 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 (**Figure 8.13**). In nonwoody plants, turgor pressure supports the plant. If the plant is not watered, the extracellular fluid will become hypertonic, causing water to leave the cell. In this condition, the cell membrane detaches from the cell wall and constricts the cytoplasm. This process, called **plasmolysis**, causes plants to lose turgor pressure (**Figure 8.14**).

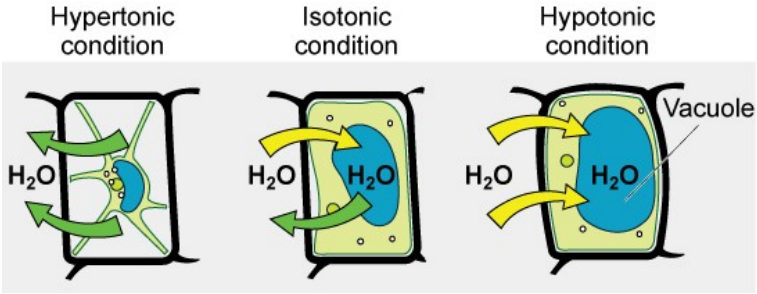


Figure 8.13 The turgor pressure within a plant cell depends on the tonicity of the solution that it is bathed in. (Credit: modification of work by Mariana Ruiz Villareal)

Figure 8.14 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 lysing as it takes on water from its environment (**Figure 8.15**).

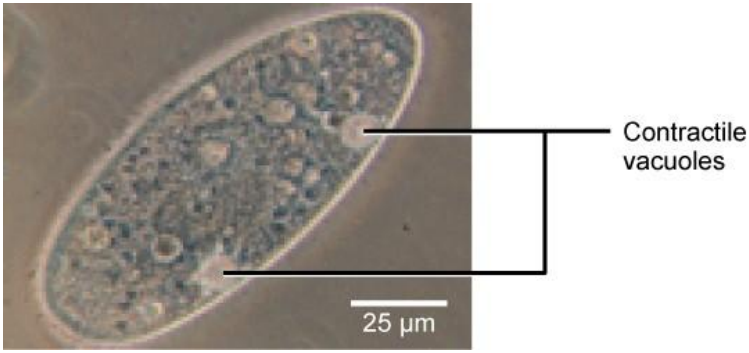


Figure 8.15 A paramecium's contractile vacuole, here visualized using bright field light microscopy at 480x magnification, continuously pumps water out of the organism's body to keep it from bursting in a hypotonic medium. (Credit: modification of work by NIH; scale-bar data from Matt Russell)

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.

8.3 | Active Transport

By the end of this section, you will be able to:Understand how electrochemical gradients affect ionsDistinguish between primary active transport and secondary 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.

8.3.1 Electrochemical Gradient

We have discussed simple concentration gradients—different concentrations of a substance across a space or a membrane—but in living systems, gradients are more complex. Because ions move into and out of cells and because cells contain proteins that do not move across the membrane and are mostly negatively charged, there is also an electrical gradient, a difference of charge, across the plasma membrane.

The interior of living cells is electrically negative with respect to the extracellular fluid surrounding them. At the same time, cells have a lower concentration of (Na+) than does the extracellular fluid. Therefore, both the concentration

gradient and the electrical gradient tend to drive Na+ into the cell. Conversely, cells have a higher concentration of K+ than the extracellular fluid does. Therefore, the concentration gradient tends to drive K+ out of the cell, while the electrical gradient tends to drive it inside the cell. The combined gradient of concentration and electrical charge that affects an ion is called its **electrochemical gradient** (**Figure 8.16**).

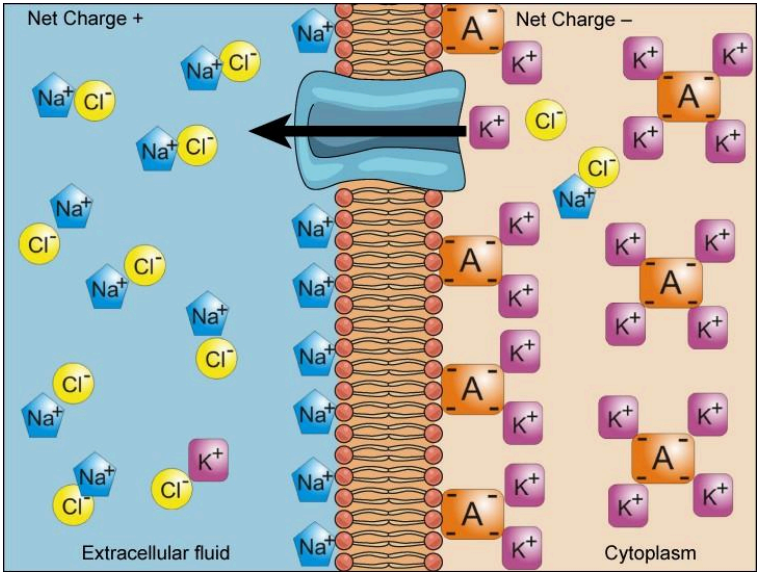


Figure 8.16 Electrochemical gradients arise from the combined effects of concentration gradients and electrical gradients. (Credit: "Synaptitude"/Wikimedia Commons)

Concept Check

Injection of a potassium solution into a person's blood is lethal; this is used in capital punishment and euthanasia. Why do you think a potassium solution injection is lethal?

Moving Against a Gradient

To move substances against a concentration or electrochemical gradient, the cell must use energy, usually in the form of ATP. Active transport proteins, 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.

Proteins for Active Transport

The specific proteins that facilitate active transport are called **transporters**. There are three types of transporters (**Figure 8.17**). A **uniporter** carries one specific ion or molecule. A **symporter** carries two different ions or molecules, both in the same direction. An **antiporter** carries two different ions or molecules in different directions. All of these transporters can transport small, uncharged organic molecules such as glucose.

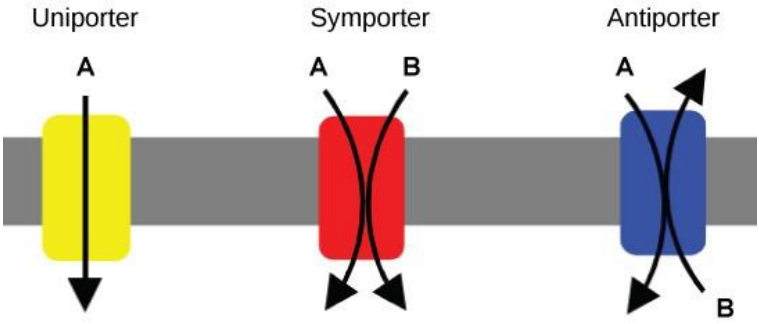


Figure 8.17 A uniporter carries one molecule or ion. A symporter carries two different molecules or ions, both in the same direction. An antiporter also carries two different molecules or ions, but in different directions. (credit: modification of work by "Lupask"/Wikimedia Commons)

Two mechanisms exist for the transport of small-molecular weight material and small molecules. **Primary active transport** is directly dependent on ATP. **Secondary active transport** does not directly require ATP, because it uses electrochemical gradients established by primary active transport for fuel. Primary active transport must occur first to in order to allow secondary active transport to occur. Although it does not use ATP, secondary active transport is still considered active because it requires energy.

8.3.2 Primary Active Transport

One of the most important pumps in animals 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 two K+ into the cell while moving three Na+ out of the cell (**Figure 8.18**).

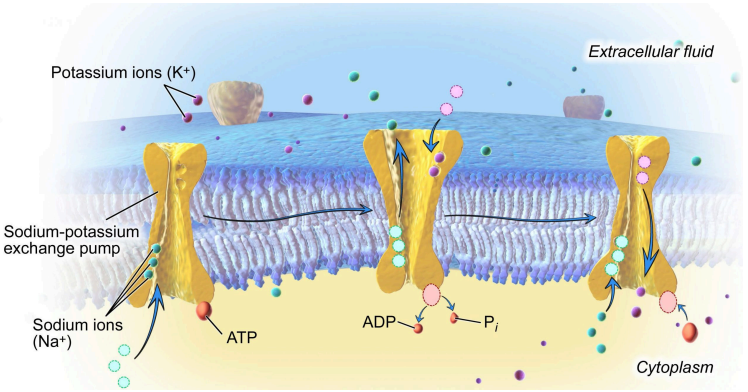


Figure 8.18 Primary active transport moves ions across a membrane, creating an electrochemical gradient (electrogenic transport). (Credit: Blausen.com staff. "Blausen gallery 2014". Wikiversity Journal of Medicine.)

The sodium-potassium pump works in the following six steps:

1. Three sodium ions bind to the protein.
2. ATP is hydrolyzed by the protein carrier and a low-energy phosphate group attaches to it.
3. The carrier changes shape and opens towards the exterior of the membrane. The three sodium ions are released.
4. Two potassium ions attach to the protein, causing

the low-energy phosphate group to detach.
5. The carrier protein changes shape so that is open towards the interior of the cell.
6. The two potassium ions are released into the cytoplasm and the process begins again.

Several things have happened as a result of this process. First, there are now more sodium ions outside of the cell than inside and more potassium ions inside than out. Second, since three sodium ions moved out for each two potassium ions that moved in, the interior is slightly more negative relative to the exterior. This difference in charge is important in creating the conditions necessary for secondary active transport. 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.

The sodium-potassium pump (Na+/K+ pump) is one example of energy coupling. Each cycle of the Na+/K+ pump moves three sodium out of the cell and brings two potassium into the cell. For each cycle, one ATP is hydrolyzed and its free phosphate group is transferred to the pump protein. This process of a phosphate group binding to a molecule is called **phosphorylation**. Phosphorylation of the pump protein causes it to change shape, moving ions across the membrane. ATP performs cellular work using this basic form of energy coupling through phosphorylation. Here, the **exergonic**

(energy-releasing) process of ATP breakdown "pays for" the **endergonic** (energy-requiring) process of moving ions against their concentration gradients.

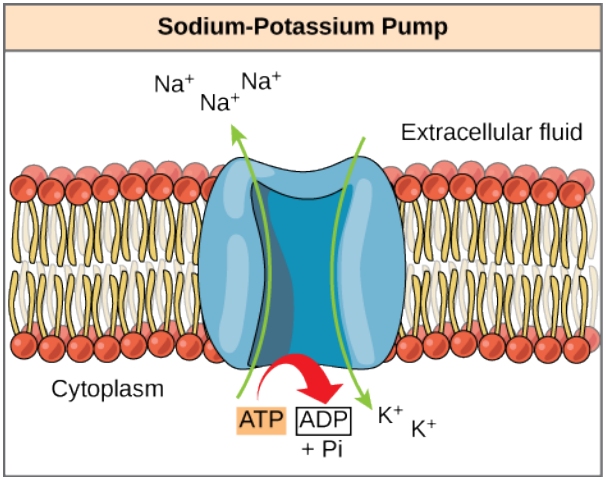


Figure 8.19 The sodium-potassium pump is an example of energy coupling. The energy derived from exergonic ATP hydrolysis is used to pump sodium and potassium ions across the cell membrane.

8.3.3 Secondary Active Transport (Co-transport)

Secondary active transport moves a solute against its concentration gradient, an endergonic process, by moving another solute down its concentration gradient, an exergonic process. For instance, as sodium ion concentrations build outside of the plasma membrane because of the action of the

sodium-potassium pump, an electrochemical gradient is created. If a channel protein exists and is open, the sodium ions will be pulled through the membrane, down their concentration gradient. This exergonic movement is used to transport other substances that can attach themselves to the transport protein through the membrane (**Figure 8.20**). Many amino acids, as well as glucose, enter a cell this way.

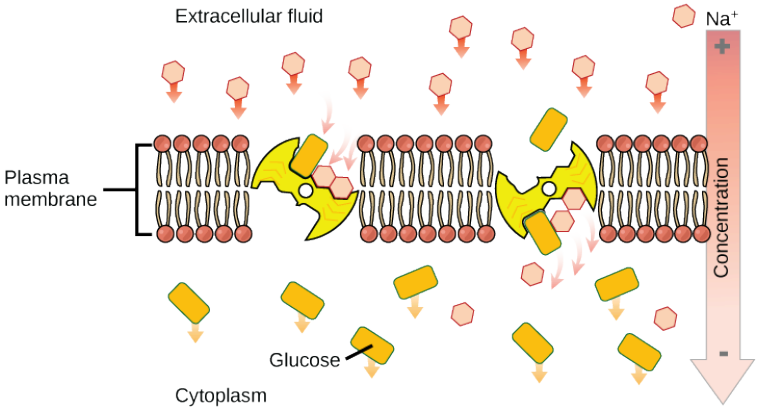


Figure 8.20 An electrochemical gradient, created by primary active transport, can move other substances against their concentration gradients, a process called co-transport or secondary active transport. (Credit: modification of work by Mariana Ruiz Villareal)

8.4 | Bulk Transport

Learning Objectives

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

- Describe endocytosis, including phagocytosis, pinocytosis, and receptor-mediated endocytosis.
- Understand the process of exocytosis.

In addition to moving small ions and molecules through the membrane, cells also need to remove and take in larger molecules and particles (see **Table 8.2** for examples). Some cells are even capable of engulfing entire unicellular microorganisms. You might have correctly hypothesized that the uptake and release of large particles by the cell requires energy. A large particle, however, cannot pass through the membrane, even with energy supplied by the cell.

8.4.1 Endocytosis

Endocytosis is a type of active transport that moves particles, such as large molecules, parts of cells, and even whole cells, into a cell. There are different variations of endocytosis, but all share a common characteristic: The plasma membrane of the cell invaginates, forming a pocket around the target particle. The pocket pinches off, resulting in the particle being contained in a newly created intracellular vesicle formed from the plasma membrane. The three types of endocytosis are phagocytosis, pinocytosis, and receptor-mediated endocytosis.

Phagocytosis

Phagocytosis ("cell eating") is the process by which large particles, such as other cells or relatively large particles, are taken in by a cell. For example, when microorganisms invade the human body, a type of white blood cell called a neutrophil will "eat" the invaders through phagocytosis, surrounding and engulfing the microorganism, which is then destroyed by lysosomes inside the neutrophil (**Figure 8.21**).

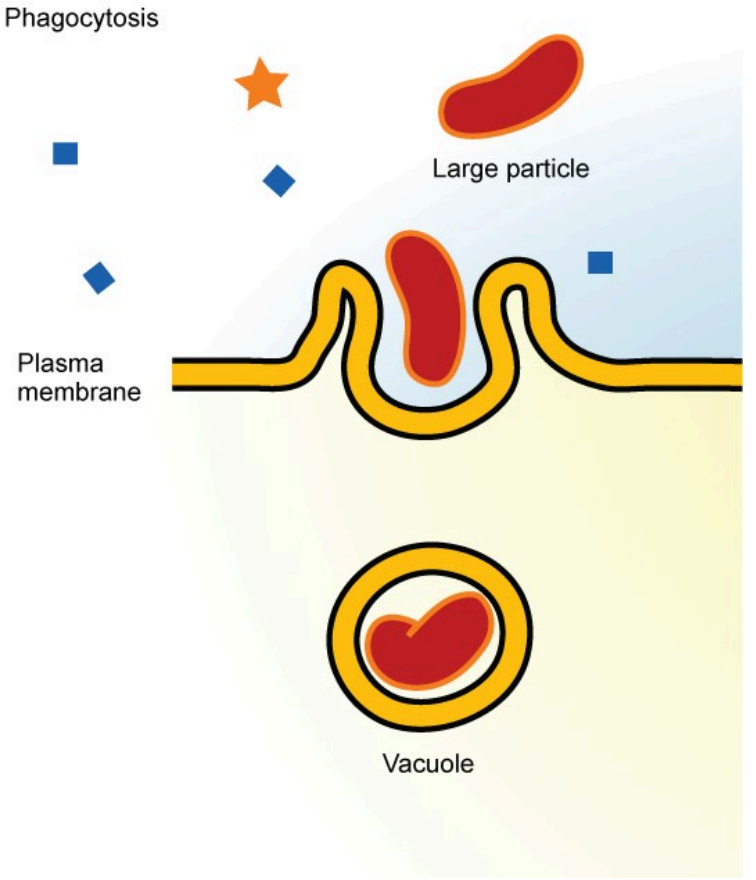


Figure 8.21 In phagocytosis, the cell membrane surrounds the particle and engulfs it. (Credit: Mariana Ruiz Villareal)

In preparation for phagocytosis, a portion of the inward-facing surface of the plasma membrane becomes coated with a protein called **clathrin**, which stabilizes this section of the membrane. The coated portion of the membrane then extends from the body of the cell and surrounds the particle, eventually

enclosing it. Once the vesicle containing the particle is enclosed within the cell, the clathrin disengages from the membrane and the vesicle merges with a lysosome for the breakdown of the material in the newly formed compartment. When accessible nutrients from the degradation of the vesicular contents have been extracted, the newly formed endosome merges with the plasma membrane and releases its contents into the extracellular fluid. The endosomal membrane again becomes part of the plasma membrane.

Pinocytosis

Through **pinocytosis** ("cell drinking"), cells take in molecules, including water, which the cell needs from the extracellular fluid. Pinocytosis results in a much smaller vesicle than does phagocytosis, and the vesicle does not need to merge with a lysosome (**Figure 8.22**).

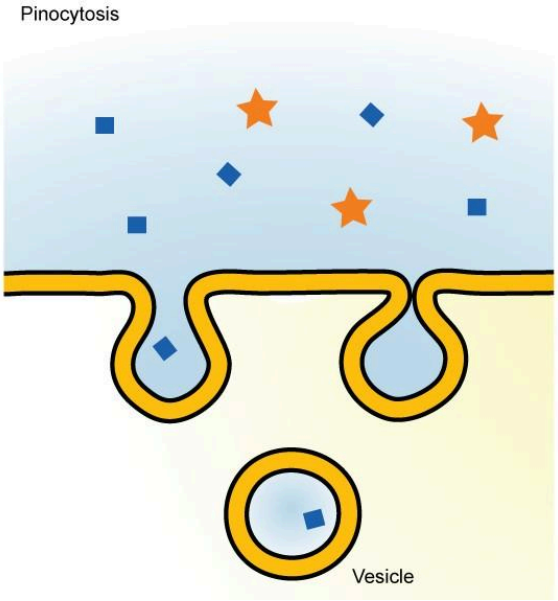


Figure 8.22 In pinocytosis, the cell membrane invaginates, surrounds a small volume of fluid, and pinches off. (Credit: Mariana Ruiz Villareal)

Receptor-mediated Endocytosis

Receptor-mediated endocytosis is a targeted variation of endocytosis that employs receptor proteins in the plasma membrane that have a specific binding affinity for certain substances (**Figure 8.23**).

Receptor-mediated endocytosis, as in phagocytosis, uses clathrin protein attached to the cytoplasmic side of the plasma membrane. Some human diseases are caused by the failure of receptor-mediated endocytosis. For example, the form of

cholesterol termed low-density lipoprotein or LDL (also referred to as "bad" cholesterol) is removed from the blood by receptor-mediated endocytosis. In the human genetic disease familial hypercholesterolemia, the LDL receptors are defective or missing entirely. People with this condition have life-threatening levels of cholesterol in their blood, because their cells cannot clear LDL particles from their blood.

Although receptor-mediated endocytosis is designed to bring specific substances that are normally found in the extracellular fluid into the cell, other substances may gain entry into the cell at the same site. Flu viruses, diphtheria, and cholera toxin all have sites that cross-react with normal receptor-binding sites and gain entry into cells.

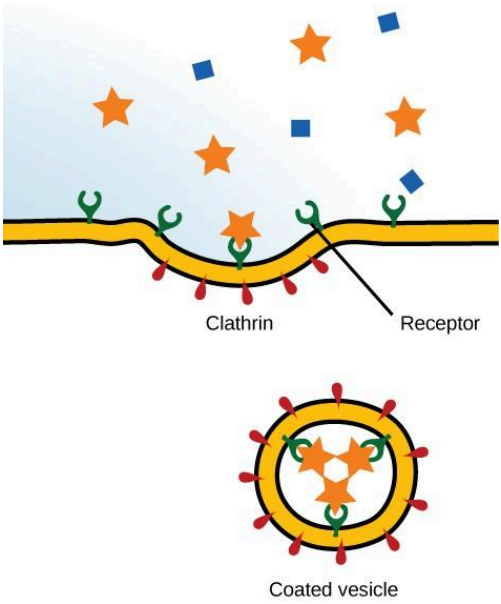


Figure 8.23 In receptor-mediated endocytosis, uptake of substances by the cell is targeted to a single type of substance that binds to the receptor on the external surface of the cell membrane. (Credit: modification of work by Mariana Ruiz Villareal)

8.4.2 Exocytosis

The reverse process of moving material into a cell is the process of **exocytosis**. The purpose of exocytosis is to expel material from the cell into the extracellular fluid. Waste material is enveloped in vesicle, which fuses with the interior of the plasma membrane, expelling the waste material into the

extracellular space (**Figure 8.24**). Cells also use exocytosis to secrete proteins such as hormones, neurotransmitters, or parts of the extracellular matrix.

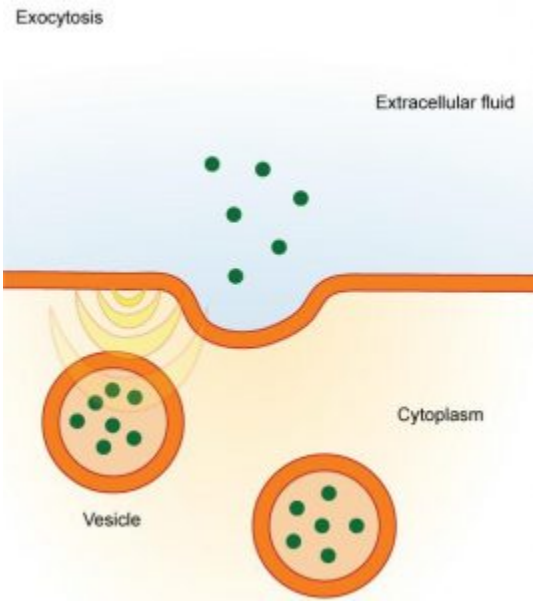


Figure 8.24 In exocytosis, vesicles containing substances fuse with the plasma membrane. The contents are then released to the exterior of the cell. (Credit: modification of work by Mariana Ruiz Villareal)

Table 8.2 Methods of transport, energy requirements, and types of material transported

Transport Method	Active / Passive	Material Transported
Diffusion	Passive	Small-molecular weight material
Osmosis	Passive	Water
Facilitated transport/ diffusion	Passive	Sodium, potassium, calcium, glucose
Primary active transport	Active	Sodium, potassium, calcium
Secondary active transport	Active	Amino acids, lactose
Phagocytosis	Active	Large macromolecules, whole cells, or cellular structures
Pinocytosis and potocytosis	Active	Small molecules (liquids/ water)
Receptor-mediated endocytosis	Active	Large quantities of macromolecules

CHAPTER 9. CELL COMMUNICATION

Figure 9.1 Have you ever become separated from a friend while in a crowd? If so, you know the challenge of searching for someone when surrounded by thousands of other people. If you and your friend have cell phones, your chances of finding each other are good. A cell phone's ability to send and receive messages makes it an ideal communication device. (Credit: modification of work by Vincent and Bella Productions)

Chapter Outline

Type your examples here.

- 9.1 Signaling Molecules and Cellular Receptors
- 9.2 Propagation of the Signal
- 9.3 Response to the Signal

Introduction

Imagine what life would be like if you and the people around you could not communicate. You would not be able to express your wishes to others, nor could you ask questions to find out more about your environment. Social organization is dependent on communication between the individuals that comprise that society.

As with people, it is vital for individual cells to be able to interact with their environment and with each other. This is true whether a cell is growing by itself in a pond or is one of many cells that form a larger organism. In order to properly respond to external stimuli, cells have developed complex mechanisms of communication so that they can receive a

message, transfer the information across the plasma membrane, and then produce changes within the cell in response to the message.

In multicellular organisms, cells constantly send and receive chemical messages to coordinate the actions of other organs, tissues, and cells. The ability to send messages quickly and efficiently enables cells to coordinate and fine-tune their functions. While the necessity for cellular communication in larger organisms seems obvious, even single-celled organisms communicate with each other. Yeast cells signal each other to aid mating. Some forms of bacteria coordinate their actions in order to form large complexes called biofilms or to organize the production of toxins to remove competing organisms. The ability of cells to communicate through chemical signals originated in single cells and was essential for the evolution of multicellular organisms. Efficient, error-free communication is vital for all life.

9.1 | Signaling Molecules and Cellular Receptors

Learning Objectives

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

- Describe four types of signaling found in multicellular organisms.
- Compare internal receptors with cell-surface receptors.
- Recognize the relationship between a ligand's chemistry and its mechanism of action.

There are two kinds of communication in the world of living cells. Communication between cells is called **intercellular signaling**, and communication within a cell is called **intracellular signaling**. An easy way to remember the distinction is by understanding that the prefix inter- means "between" (an interstate highway crosses between states) and intra- means "inside" (an IV means intravenous or "inside the vein").

Chemical signals are released by a **signaling cell** and received by a **target cell**. Target cells have proteins called **receptors**, which bind to signaling molecules and cause a response. Signaling molecules that bind to receptors are called **ligands**. Ligands and receptors are specific for each other; a receptor will typically bind only to its specific ligand. However, there are different types of signaling.

9.1.1 Forms of Signaling

There are four categories of chemical signaling found in multicellular organisms: **autocrine** signaling, **paracrine** signaling, **endocrine** signaling, and **direct** signaling across gap junctions (**Figure 9.2**). The main difference between the different categories of signaling is the distance that the signal travels to reach the target cell.

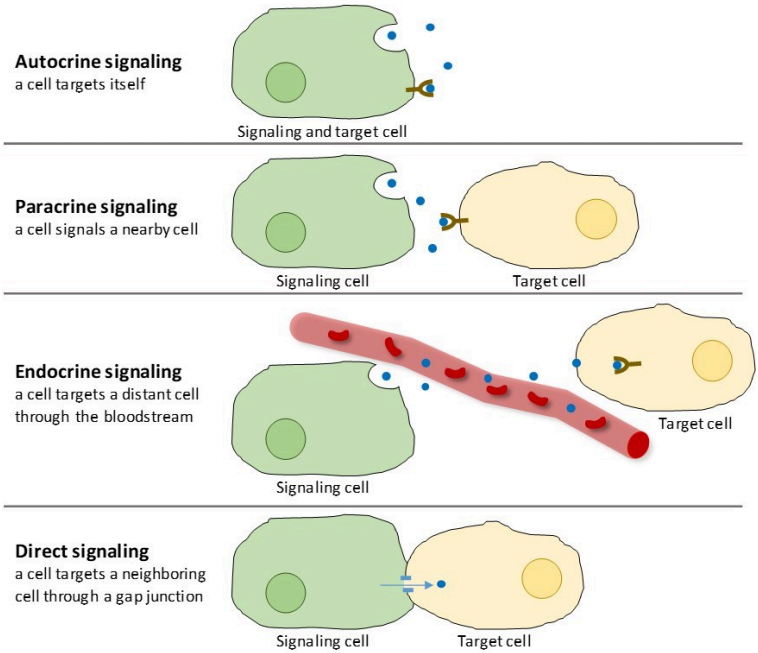


Figure 9.2 In chemical signaling, a cell may target itself, a nearby cell, a distant cell, or a cell that it is connected to by a gap junction.

Paracrine Signaling

Signals that act locally between cells that are close together are called **paracrine** signals. Paracrine signals move by diffusion through the extracellular matrix (**Figure 9.2**). These types of signals usually elicit quick responses that last only a short amount of time. In order to keep the response localized, paracrine ligands are usually quickly degraded by enzymes or removed by neighboring cells. Removing the signals reestablishes the concentration gradient for the signal molecule, allowing them to quickly diffuse through the intracellular space if released again.

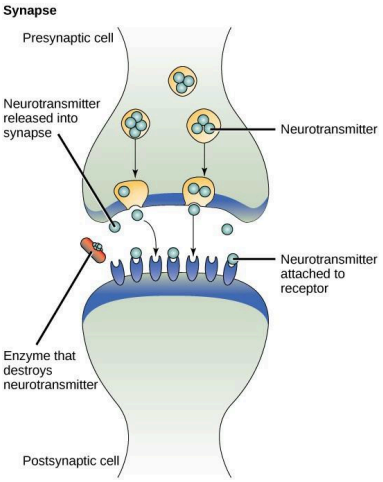


Figure 9.3 The distance between two nerve cells is very small and allows for rapid diffusion of the neurotransmitter. Enzymes in the synaptic cleft degrade some types of neurotransmitters to terminate the signal.

One example of paracrine signaling is the transfer of signals between nerve cells. The tiny space between nerve cells where signal transmission occurs is called a synapse. Signals are propagated along nerve cells by fast-moving electrical impulses. When these impulses reach the end of one nerve cell, chemical ligands called **neurotransmitters** are released into the synapse by the presynaptic cell (the cell emitting the signal). The neurotransmitters diffuse across the synapse (**Figure 9.3**). The small distance between nerve cells allows the signal to travel quickly, which enables an immediate response, such as, "take your hand off the stove!" When the neurotransmitter binds the receptor on the surface of the postsynaptic cell, the next electrical impulse is launched. The neurotransmitters are degraded quickly or are reabsorbed by the presynaptic cell so that the recipient

nerve cell can recover quickly and be prepared to respond rapidly to the next synaptic signal.

Autocrine Signaling

When a cell responds to its own signaling molecule, it is called **autocrine** signaling (auto = "self"). Autocrine signaling often occurs with other types of signaling. For example, when a paracrine signal is released, the signaling cell may respond to the signal along with its neighbors (**Figure 9.2**).

Autocrine signaling often occurs during early development of an organism to ensure that cells develop into the correct tissues. Autocrine signaling also regulates pain sensation and inflammatory responses. Further, if a cell is infected with a virus, the cell can signal itself to undergo programmed cell death, killing the virus in the process.

Endocrine Signaling

Signals from distant cells are called **endocrine signals**, and they originate from **endocrine cells**. (In the body, many endocrine cells are located in endocrine glands, such as the thyroid gland, the hypothalamus, and the pituitary gland.) These types of signals usually produce a slower response but have a longer-lasting effect. The ligands released in endocrine signaling are called **hormones**, signaling molecules that are

produced in one part of the body but affect other body regions some distance away (**Figure 9.2**).

Hormones travel the large distances between endocrine cells and their target cells via the bloodstream, which is a relatively slow way to move throughout the body. Because of their form of transport, hormones get diluted and are present in low concentrations when they act on their target cells. This is different from paracrine signaling, in which local concentrations of signaling molecules can be very high.

Direct Signaling

Gap junctions in animals and plasmodesmata in plants are connections between the plasma membranes of neighboring cells. These water-filled channels allow small signaling molecules to diffuse between the two cells. Small molecules, such as calcium ions (Ca2+), are able to move between cells, but large molecules like proteins and DNA cannot fit through the channels. The specificity of the channels ensures that the cells remain independent but can quickly and easily transmit signals. Direct signaling allows a group of cells to coordinate their response to a signal that only one of them may have received. In plants, plasmodesmata are ubiquitous, making the entire plant into a giant communication network.

9.1.2 Types of Receptors

Receptors are protein molecules in the target cell or on its surface that bind to ligands. There are two types of receptors, internal receptors and cell-surface receptors.

Internal receptors

Internal receptors, also known as intracellular or cytoplasmic receptors, are found in the cytoplasm of target cells and respond to hydrophobic ligand molecules that are able to travel across the plasma membrane. Once inside the cell, many of these molecules bind to proteins that act as regulators of mRNA synthesis (transcription) to mediate gene expression.

Gene expression is the cellular process of transforming the information in a cell's DNA into a sequence of amino acids, which ultimately forms a protein. When the ligand binds to the internal receptor, a conformational change is triggered that exposes a DNA-binding site on the receptor protein. The ligand-receptor complex moves into the nucleus, then binds to specific regulatory regions of the chromosomal DNA and promotes the initiation of transcription (**Figure 9.4**). Transcription is the process of copying the information in a cell's DNA into a special form of RNA called messenger RNA (mRNA); the cell uses information in the mRNA to link specific amino acids in the correct order, producing a protein.

Thus, when a ligand binds to an internal receptor, it can directly influence gene expression in the target cell.

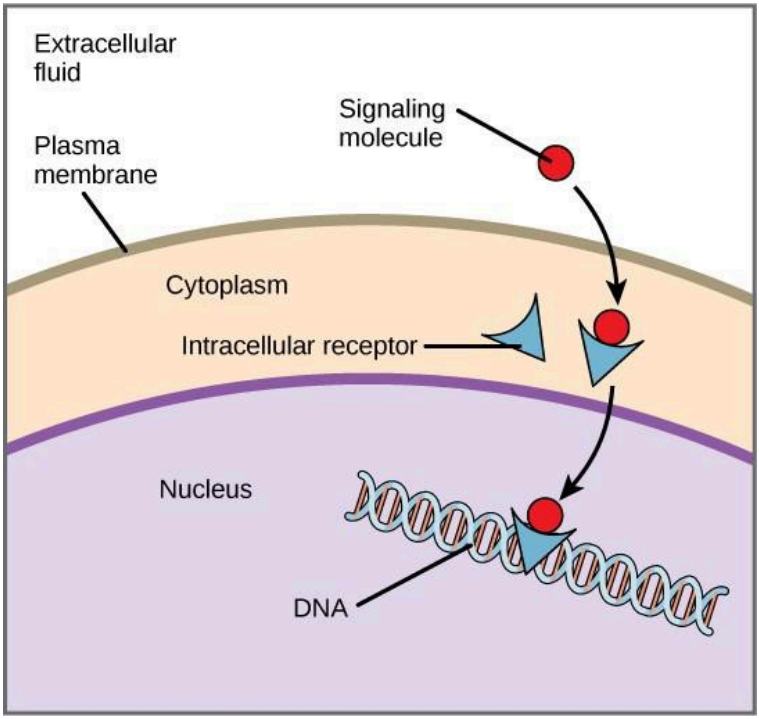


Figure 9.4 Hydrophobic signaling molecules typically diffuse across the plasma membrane and interact with intracellular receptors in the cytoplasm. Many intracellular receptors are transcription factors that interact with DNA in the nucleus and regulate gene expression.

Cell-Surface Receptors

Cell-surface receptors, also known as transmembrane receptors, are integral proteins that bind to external signaling

molecules. These receptors span the plasma membrane and perform **signal transduction**, in which an extracellular signal is converted into an intercellular signal. (**Figure 9.5**). Because cell-surface receptor proteins are fundamental to normal cell functioning, it should come as no surprise that a malfunction in any one of these proteins could have severe consequences. Errors in the protein structures of certain receptor molecules have been shown to play a role in hypertension (high blood pressure), asthma, heart disease, and cancer.

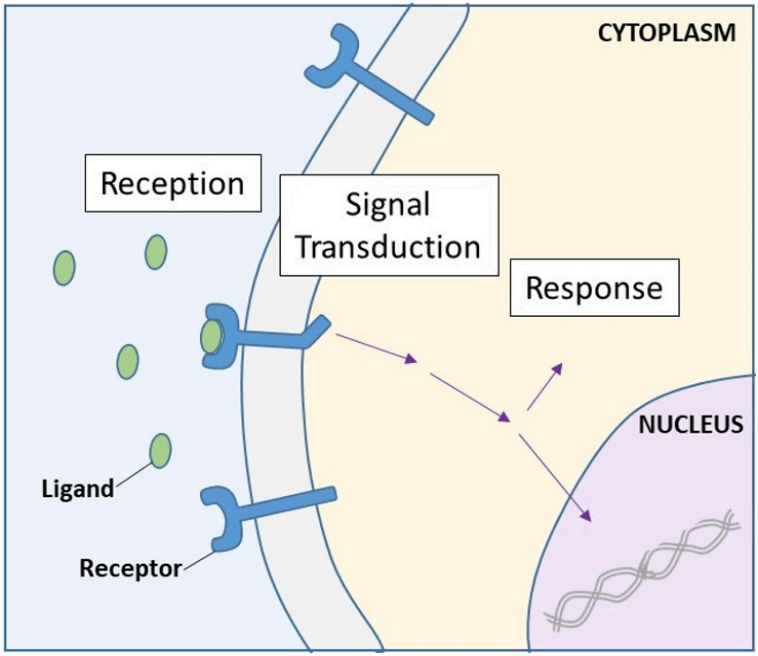


Figure 9.5 Hydrophilic signaling molecules typically work by binding to the extracellular portion of a receptor protein. The signal is then transduced across the membrane.

Each cell-surface receptor has three main components: an external ligand-binding domain, or **extracellular domain**; a hydrophobic membrane-spanning region; and an intracellular domain. Cell-surface receptors are involved in most of the signaling in multicellular organisms. There are three general categories of cell-surface receptors: enzyme-linked receptors, ion channel-linked receptors, and G-protein-linked receptors.

Enzyme-linked receptors are cell-surface receptors with intracellular domains that are associated with an enzyme. In some cases, the intracellular domain of the receptor itself is an enzyme. Other enzyme-linked receptors have a small intracellular domain that interacts directly with an enzyme. Enzyme-linked receptors normally have large extracellular and intracellular domains, but the membrane-spanning region consists of a single alpha-helix in the peptide strand.

When a ligand binds to the extracellular domain of an enzyme-linked receptor, a signal is transferred through the membrane, activating the enzyme. Activation of the enzyme sets off a chain of events within the cell that eventually leads to a response.

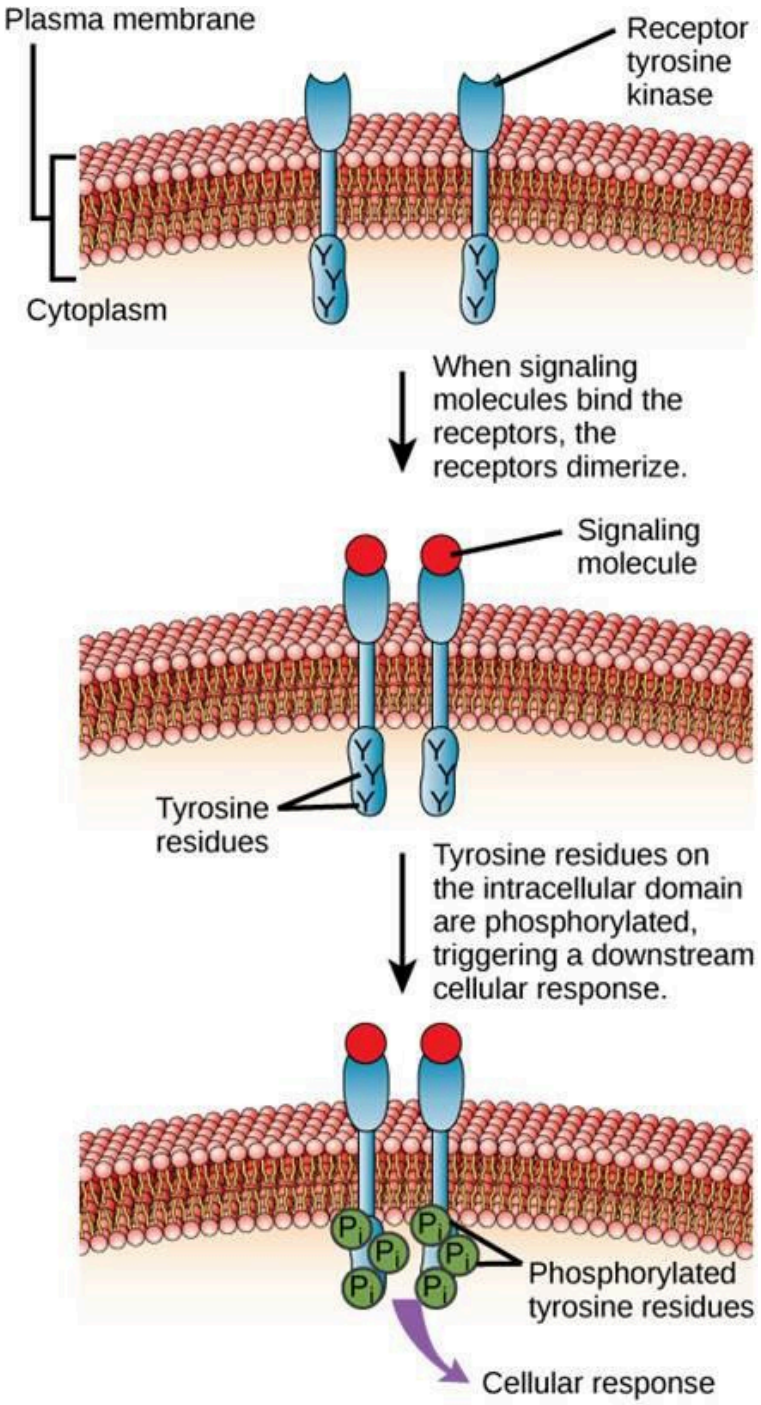

Plasma membrane
Receptor tyrosine kinase
Cytoplasm
When signaling molecules bind the receptors, the receptors dimerize.
Signaling molecule
Tyrosine residues
Tyrosine residues on the intracellular domain are phosphorylated, triggering a downstream cellular response.
Phosphorylated tyrosine residues
Cellular response

Figure 9.6 A receptor tyrosine kinase is an enzyme-linked receptor with a single transmembrane region, and extracellular and intracellular domains. Binding of a signaling molecule to the extracellular domain causes the receptor to dimerize. Tyrosine residues on the intracellular domain are then auto-phosphorylated, triggering a downstream cellular response. The signal is terminated by a phosphatase that removes the phosphates from the phosphotyrosine residues.

One example of an enzyme-linked receptor is the tyrosine kinase receptor (**Figure 9.6**). A **kinase** is an enzyme that transfers phosphate groups from ATP to another protein. The tyrosine kinase receptor transfers phosphate groups to tyrosine molecules. First, signaling molecules bind to the extracellular domain of two nearby tyrosine kinase receptors. The two neighboring receptors then bond together, or dimerize. Phosphates are then added to tyrosine residues on the intracellular domain of the receptors (phosphorylation). The phosphorylated residues can then transmit the signal to the next messenger within the cytoplasm.

Epidermal growth factor receptors are an example of receptor tyrosine kinases that follows this mode of signaling. Defects in ErbB signaling in this family can lead to neuromuscular diseases such as multiple sclerosis and Alzheimer's disease.

Ion channel-linked receptors bind to a ligand and open a channel through the membrane that allows specific ions to pass through. This type of cell-surface receptor has an extensive membrane-spanning region with hydrophobic amino acids.

Conversely, the amino acids that line the inside of the channel are hydrophilic to allow for the passage of ions. When a ligand binds to the extracellular region of the channel, there is a conformational change in the protein's structure that allows ions such as sodium, calcium, magnesium, or hydrogen to pass through (**Figure 9.7**).

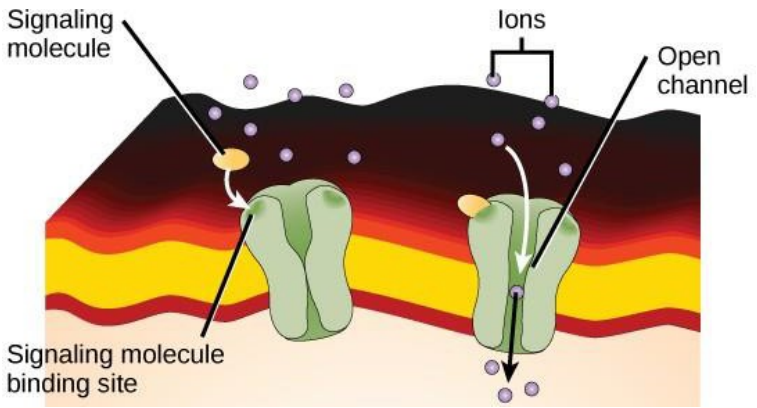


Figure 9.7 Ion channel-linked receptors open and allow ions to enter a cell. An example of an ion channel-linked receptor is found on neurons. When neurotransmitters bind to these receptors, a conformational change allows sodium ions to flow across the cell membrane, causing a change in the membrane potential.

G-protein-linked receptors bind to a ligand and activate an associated G-protein. The activated G- protein then interacts with a nearby membrane protein, which may be an ion channel or an enzyme (**Figure 9.8**). All G-protein-linked receptors have seven transmembrane domains, but each

receptor has a specific extracellular domain and G-protein-binding site.

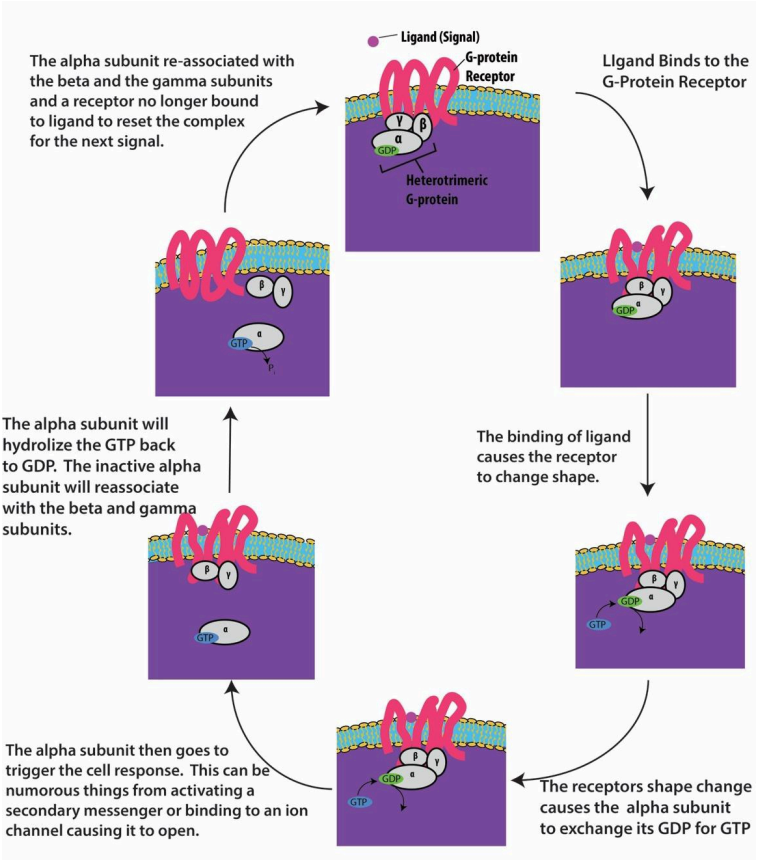
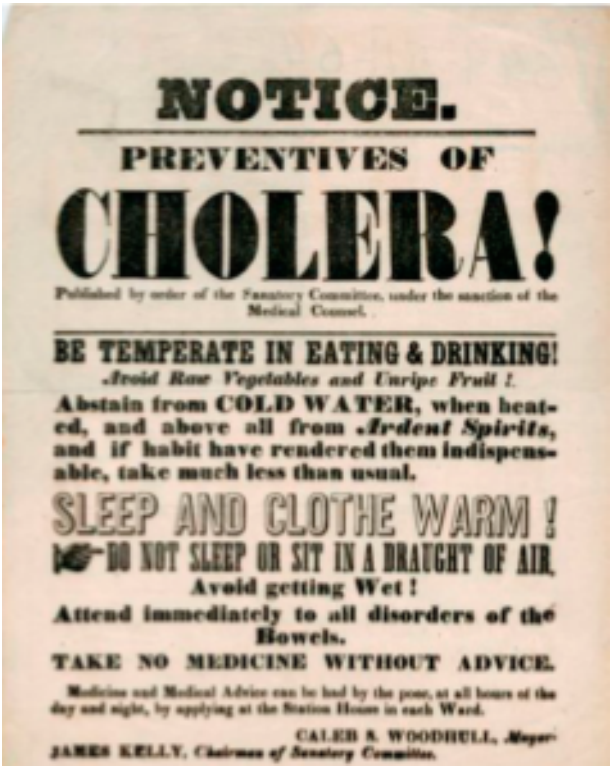


Figure 9.8 Some G proteins have three subunits: α, β, and γ. When a signaling molecule binds to a G- protein receptor, a GDP molecule associated with the α subunit is exchanged for GTP. The β and γ subunits dissociate from the α subunit, and a cellular response is triggered. Hydrolysis of GTP to GDP terminates the signal. Image by Chris Wrobel.

Cell signaling using G-protein-linked receptors occurs as a cycle. Once the ligand binds to the receptor, the resultant shape change activates the G-protein, which releases GDP and picks up GTP. The subunits of the G-protein then split into α and $\beta\gamma$ subunits. One or both of these G-protein fragments may be able to activate other proteins in the cell. After a while, the GTP on the active α subunit of the G-protein is hydrolyzed to GDP and the $\beta\gamma$ subunit is deactivated. The subunits re-associate to form the inactive G-protein and the cycle begins again (**Figure 9.8**).

G-protein linked receptors are used in many physiological processes including those for vision transduction, taste, and regulation of immune system and inflammation.



NOTICE.

PREVENTIVES OF

CHOLERA!

Published by order of the Sanitary Committee, under the sanction of the Medical Council.

BE TEMPERATE IN EATING & DRINKING!

Avoid Raw Vegetables and Unripe Fruit!

Abstain from COLD WATER, when heated, and above all from *Ardent Spirits*, and if habit have rendered them indispensable, take much less than usual.

SLEEP AND CLOTHE WARM!

DO NOT SLEEP OR SIT IN A DRAUGHT OF AIR.

Avoid getting Wet!

Attend immediately to all disorders of the Bowels.

TAKE NO MEDICINE WITHOUT ADVICE.

Medicine and Medical Advice can be had by the poor, at all hours of the day and night, by applying at the Station House in each Ward.

CALEB S. WOODHULL, *Mayor.*

JAMES KELLY, *Chairman of Sanitary Committee.*

Figure 9.9 Transmitted primarily through contaminated drinking water, cholera is a major cause of death in the developing world and in areas where natural disasters interrupt the availability of clean water. The cholera bacterium, Vibrio cholerae, creates a toxin that modifies G-protein-mediated cell signaling pathways in the intestines. Modern sanitation eliminates the threat of cholera outbreaks, such as the one that swept through New York City in 1866. This

poster shows how transmission of this disease was not understood at that time. (Credit: New York City Sanitary Commission)

Concept check

HER2 is a receptor tyrosine kinase. In 30 percent of human breast cancers, HER2 is permanently activated, resulting in unregulated cell division. Lapatinib, a drug used to treat breast cancer, inhibits the process by which the receptor phosphorylates itself, thus reducing tumor growth by 50 percent. Which of the following steps would be inhibited by Lapatinib?

- Signaling molecule binding, dimerization, and the downstream cellular response
- Dimerization, and the downstream cellular response

- The downstream cellular response
- Phosphatase activity, dimerization, and the downstream cellular response

9.1.3 Signaling Molecules

Produced by signaling cells, **ligands** are chemical signals that travel to target cells and cause a response. The types of molecules that serve as ligands are incredibly varied and range from small proteins to small ions. Ligands are categorized as either small hydrophobic ligands, which can cross plasma membranes, or water-soluble ligands, which cannot.

Small Hydrophobic Ligands

Small hydrophobic ligands, also called lipid-soluble ligands, can directly diffuse through the plasma membrane and interact with internal receptors. Important members of this class of ligands are the steroid hormones. Steroids are lipids that have a hydrocarbon skeleton with four fused rings; different steroids have different functional groups attached to the carbon skeleton. Steroid hormones include the female sex hormone estradiol, which is a type of estrogen; the male sex hormone testosterone; and cholesterol, which is an important

structural component of biological membranes and a precursor of steroid hormones (**Figure 9.10**). Other hydrophobic hormones include thyroid hormones and vitamin D. In order to be soluble in blood, hydrophobic ligands must bind to carrier proteins while they are being transported through the bloodstream.

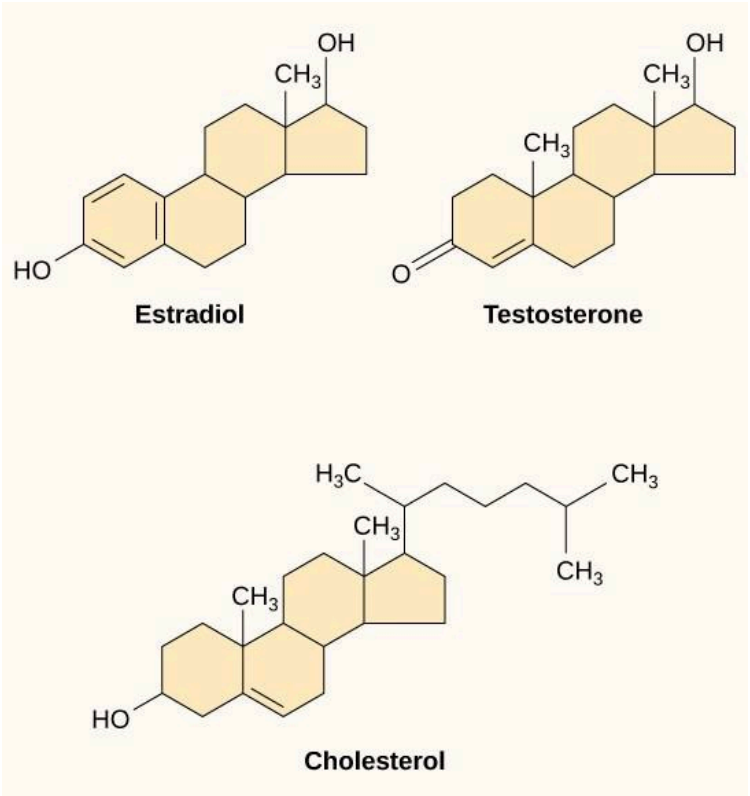


Figure 9.10 Steroid hormones are similar in structure to their precursor, cholesterol. Since they are small and hydrophobic, they can diffuse across plasma membranes and interact with internal receptors.

Water-Soluble Ligands

Since water-soluble ligands are polar, they cannot pass through the plasma membrane unaided. Sometimes they are too large to pass through the membrane at all. Instead, most water-soluble ligands bind to the extracellular domain of cell-surface receptors (see **Figure 9.5**). This group of ligands is quite diverse and includes small molecules, peptides, and proteins.

9.2 | Propagation of the Signal

Learning Objectives

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

- Explain how the binding of a ligand initiates signal transduction throughout a cell.
- Recognize that intracellular signals are transmitted by the role of phosphorylation or second messengers.

Once a water-soluble ligand binds to its receptor, the signal is

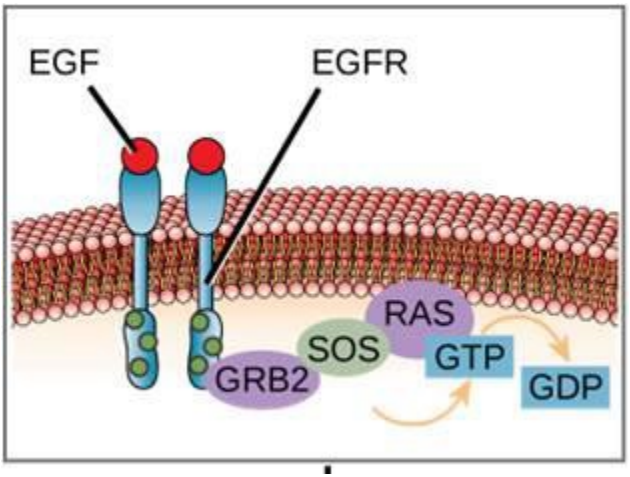
transmitted through the membrane and into the cytoplasm. Continuation of a signal in this manner is called signal transduction (**Figure 9.5**). Signal transduction only occurs with cell-surface receptors since internal receptors are able to enter the cell.

When a ligand binds to its receptor, conformational changes occur that affect the receptor's intracellular domain. These conformational changes lead to activation of the intracellular domain or its associated proteins. In some cases, binding of the ligand causes **dimerization** of the receptor, which means that two receptors bind to each other to form a stable complex called a dimer. A **dimer** is a chemical compound formed when two molecules (often identical) join together. The binding of the receptors in this manner enables their intracellular domains to come into close contact and activate each other.

9.2.1 Signaling Pathways and Signal Amplification

Although signaling molecules are often found at very low concentrations, they may produce profound effects. After the ligand binds to the cell-surface receptor, the activation of the receptor's intracellular components sets off a chain of events that is called a **signaling pathway** or a signaling cascade. In a signaling pathway, second messengers, enzymes, and/or activated proteins activate other proteins or messengers

(**Figure 9.11**). Each member of the pathway can activate thousands of the next member of the pathway in a process called **signal amplification**. Since the signal is amplified at each step, a very large response can be generated from a single receptor binding a ligand.



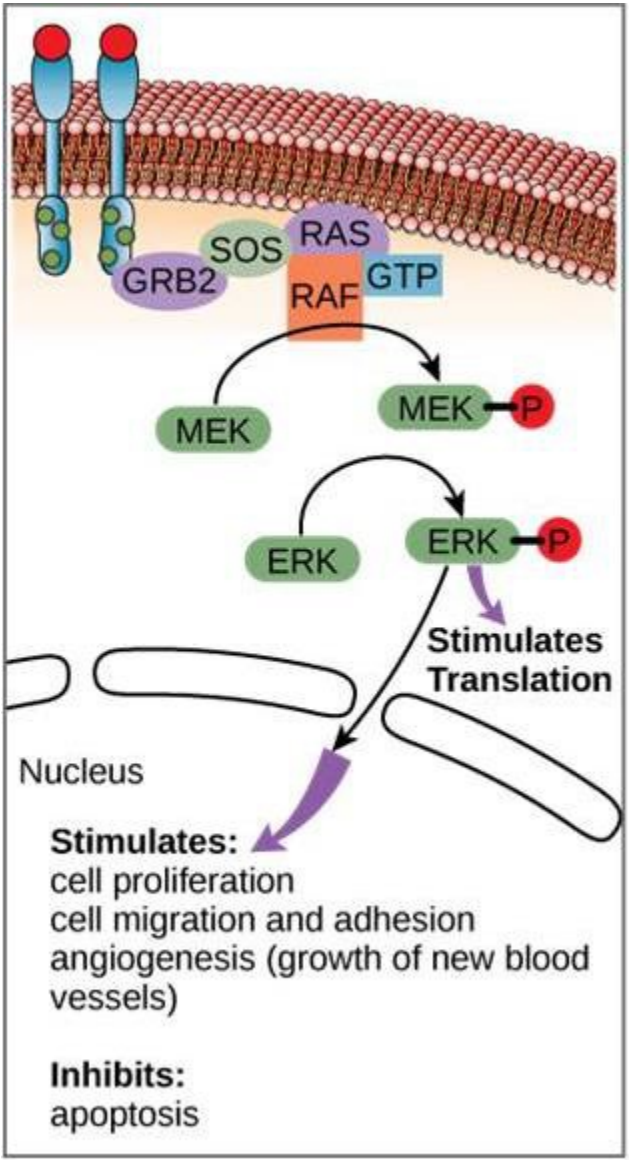


Figure 9.11 The epidermal growth factor receptor, EGFR, is a receptor tyrosine kinase. **Top.** When EGF binds to its receptor, two proteins activate

RAS, a small G-protein. **Bottom.** RAS activates RAF, which phosphorylates MEK, which phosphorylates ERK. Activated ERK enters the nucleus and triggers a cell response.

An example of a signaling pathway is shown is **Figure 9.11**. Epidermal growth factor (EGF) is a signaling molecule that is involved in the regulation of cell growth, wound healing, and tissue repair. The receptor for EGF (EGFR) is a tyrosine kinase. An activated kinase phosphorylates and activates many downstream molecules. When EGF binds to EGFR, a cascade of downstream phosphorylation events signals the cell to grow and divide. If EGFR is activated at inappropriate times, uncontrolled cell growth (cancer) may occur.

Concept Check

In certain cancers, the GTPase activity of the RAS G-protein is inhibited. This means that the RAS protein can no longer hydrolyze GTP into GDP. What effect would this have on downstream cellular events?

9.2.2 Methods of Intracellular

Signaling

The induction of a signaling pathway depends on the modification of a cellular component by an enzyme. There are numerous enzymatic modifications that can occur to activate the next component of the pathway. The following are some of the more common events in intracellular signaling.

Phosphorylation

One of the most common chemical modifications that occurs in signaling pathways is the addition of a phosphate group to a molecule in a process called phosphorylation. The phosphate can be added to a nucleotide such as GMP to form GDP or GTP. Phosphates are also often added to serine, threonine, and tyrosine residues of proteins, where they replace the hydroxyl group of the amino acid (**Figure 9.12**). The transfer of the phosphate is catalyzed by an enzyme called a **kinase**. Phosphorylation may activate or inactivate enzymes, and the reversal of phosphorylation, dephosphorylation, will reverse the effect.

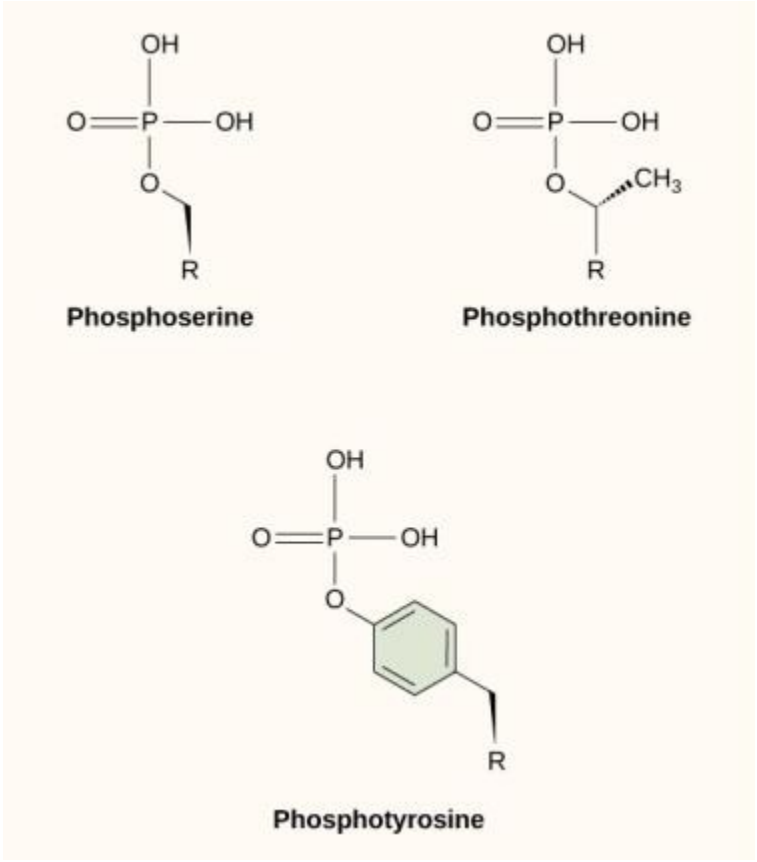


Figure 9.12 In protein phosphorylation, a phosphate group (PO_4^{-3}) is added to residues of the amino acids serine, threonine, or tyrosine. The phosphate group is added by a kinase. ATP is often used as the substrate to add the phosphate group to these amino acid. The phosphate group often results in a shape change in the protein that can activate or turns off the function of the protein.

Second Messengers

Second messengers are small molecules that propagate a signal after it has been initiated by the binding of the signaling molecule to the receptor. These molecules help to spread a signal through the cytoplasm by altering the behavior of certain cellular proteins. A second messenger utilized by many different cell types is cyclic AMP (cAMP). Cyclic AMP is synthesized by the enzyme adenylyl cyclase from ATP (**Figure 9.13**). The main role of cAMP in cells is to bind to and activate an enzyme called cAMP-dependent kinase (A-kinase). A-kinase regulates many vital metabolic pathways: It phosphorylates serine and threonine residues of its target proteins, activating them in the process. A-kinase is found in many different types of cells, and the target proteins in each kind of cell are different. Another secondary messenger is Ca2+which can be released to flood the cell.

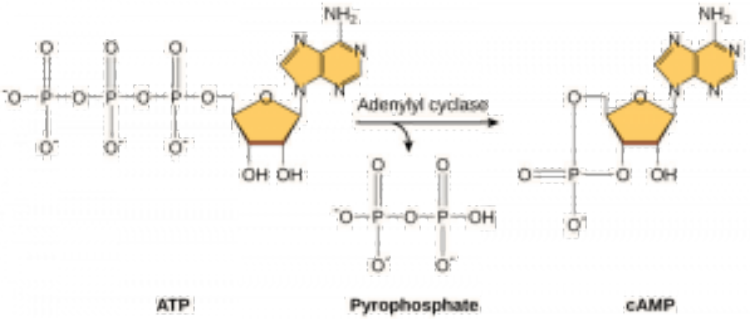


Figure 9.13 Formation of cyclic AMP (cAMP). cAMP serves as a second messenger in many cell types. Termination of the signal occurs when an enzyme called phosphodiesterase converts cAMP into AMP.

Different cells respond differently to cAMP. In **Figure 9.14**, the alpha subunit from a G-protein receptor is shown activating two different types of signaling. In the first image, cAMP is produced by the enzyme adenylate cyclase when activated by the alpha subunit. cAMP then activates other proteins that affect gene transcription. In the second image, the alpha subunit from the G-protein triggers a cascade that releases Ca2+ from the smooth endoplasmic reticulum. In this case Ca2+ is the secondary messenger that causes the cellular response.

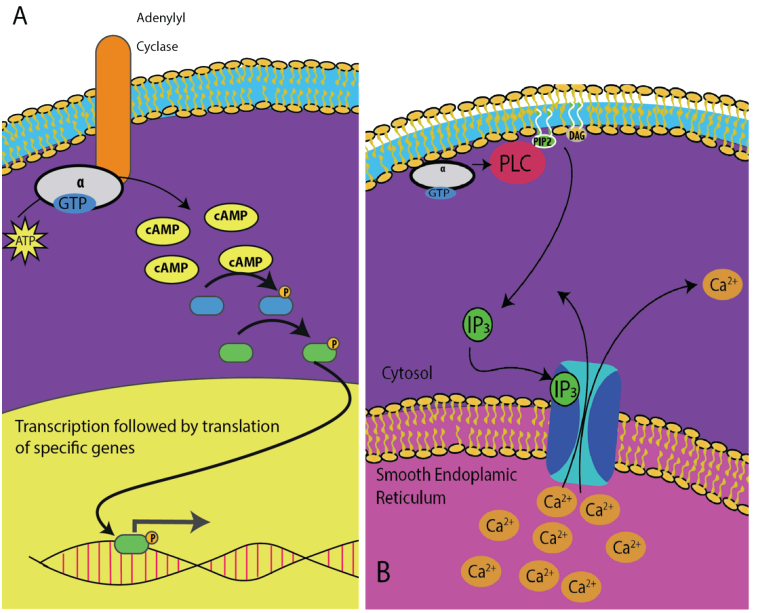


Figure 9.14 The G-protein α subunit causes different responses. **A.** Activated α subunit associates with adenylyl cyclase to produce cAMP, triggering a phosphorylation cascade that ultimately alters gene expression. **B.** Activated α subunit activates phospholipase C, ultimately leading to a flood of calcium ions. (Credit: Image by Chris Wrobel.)

9.3 | Response to the Signal

Learning Objectives

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

- Recognize that signaling pathways direct protein expression, cellular metabolism, and cell growth.
- Recognize the role of apoptosis in the development and maintenance of a healthy organism.

Using signal transduction pathways, receptors in the plasma membrane produce a variety of effects on the cell. Inside the cell, ligands bind to their internal receptors, allowing them to directly affect the cell's DNA and protein-producing machinery. The results of signaling pathways are extremely varied and depend on the type of cell involved as well as the external and internal conditions. A small sampling of responses is described below.

9.3.1 Responses to the Signaling Pathway

Gene Expression

Some signal transduction pathways regulate the transcription of RNA. Others regulate the translation of proteins.

Increase in Cellular Metabolism

The activation of β-adrenergic receptors in muscle cells by adrenaline leads to an increase in cyclic AMP inside the cell. Adrenaline is a hormone produced by the adrenal gland that readies the body for short-term emergencies.

Cell Growth

Cell signaling pathways also play a major role in cell division. Cells do not normally divide unless they are stimulated by signals from other cells. The ligands that promote cell growth are called **growth factors**. Most growth factors bind to cell-surface receptors that are linked to tyrosine kinases.

Cell Death

When a cell is damaged, superfluous, or potentially dangerous to an organism, a cell can initiate a mechanism to trigger programmed cell death, or **apoptosis**. Apoptosis allows a cell

to die in a controlled manner that prevents the release of potentially damaging molecules from inside the cell. However, in some cases, such as a viral infection or uncontrolled cell division due to cancer, the cell's normal checks and balances fail. External signaling can also initiate apoptosis. For example, most normal animal cells have receptors that interact with the extracellular matrix, a network of glycoproteins that provides structural support for animal cells. The binding of cellular receptors to the extracellular matrix initiates a signaling cascade within the cell. However, if the cell moves away from the extracellular matrix, the signaling ceases, and the cell undergoes apoptosis. This system helps prevent cells from traveling through the body and proliferating out of control, as happens with tumor cells that metastasize.

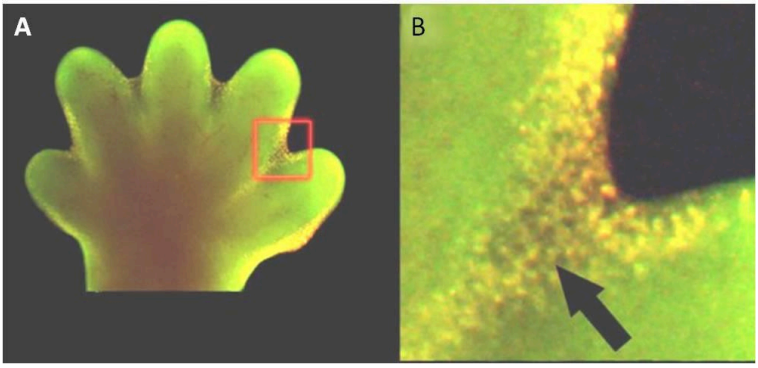


Figure 9.15 A. Apoptosis in a developing mouse paw. **B.** An enlarged view of the boxed area, showing cells dying in the webbing between the digits. (Credit: Amitabha Bandyopadhyay et. al., 2006 PLOS Genetics doi: 10.1371/journal.pgen.0020216)

Apoptosis is also essential for normal embryological development. In vertebrates, for example, early stages of development include the formation of web-like tissue between individual fingers and toes **(Figure 9.15)**. During the course of normal development, these unneeded cells must be eliminated, enabling fully separated fingers and toes to form. A cell signaling mechanism triggers apoptosis, which destroys the cells between the developing digits.

9.3.2 Termination of Signaling Pathways

The aberrant signaling often seen in tumor cells is proof that the termination of a signal at the appropriate time can be just as important as the initiation of a signal. One method of stopping a specific signal is to degrade the ligand or remove it so that it can no longer access its receptor. One reason that hydrophobic hormones like estrogen and testosterone trigger long-lasting events is because they bind to carrier proteins. These proteins allow the insoluble molecules to be soluble in blood, but they also protect the hormones from degradation by circulating enzymes.

Inside the cell, many different enzymes reverse the cellular modifications that result from signaling cascades. For example, **phosphatases** are enzymes that remove the phosphate group attached to proteins by kinases in a process called dephosphorylation. cAMP is degraded into AMP by

phosphodiesterase, thereby terminating its signal. Similarly, the release of calcium stores is reversed by Ca2+ pumps that are located in cell membranes.

PART IV

PART 4. HOW DO LIVING THINGS ACQUIRE AND USE ENERGY?

CHAPTER 10: INTRODUCTION TO METABOLISM - ENZYMES AND ENERGY

Figure 10.1 Breaking down organic material through oxidation releases energy. The energy can be released as flame, or inside cells, it can be released and harnessed in a controlled manner, using enzymes. (Credit: Photo by Sergiu Vălenaș on Unsplash.)

Chapter Outline

- 10.1 Energy Changes During Chemical Reactions
- 10.2 Enzymes
- 10.3 Energy Transfer in Living Systems: Adenosine Triphosphate (ATP)
- 10.4 Energy Transfer in Living Systems: Electron Transfer
- 10.5 Metabolic Pathways

Introduction

In Part 4 of this text, we will address how living things manage their energy needs. Many **autotrophs** (organisms able to form biological macromolecules from simple inorganic compounds) are photosynthetic, and obtain energy from light. Both autotrophs and **heterotrophs** (organisms that have to eat complex organic substances to live) can break down biological macromolecules to release useable energy. Consider the oxidation of one molecule of glucose:

$$C_6H_{12}O_6 + 6\,O_2 \rightarrow 6\,CO_2 + 6\,H_2O$$

If this reaction occurs in a single step, combustion occurs, with energy released as the light and heat of a flame. Obviously, this single-step reaction is not compatible with life. Fortunately, evolution has resulted in a slower way to release energy from glucose and harness that energy in a safe form for the cell's use: ATP.

Meeting a cell's energy needs is one of the functions of **metabolism**. Technically speaking, metabolism is the sum of all reactions occurring in a living organism. More explicitly, metabolism deals with how an organism manages its material and energy resources, including:

- Breaking down nutrients
- Generating & storing energy
- Synthesizing cellular building blocks (macromolecules)
- Eliminating waste & potentially harmful substances

In this course, we focus on the first two roles of metabolism, and only a portion of those. If you take an advanced biochemistry course, you will delve further into the pathways that allow organisms to function. To understand the basics of metabolism, you need an understanding of how living things manage energy and how enzymes allow them to do so.

10.1 | Energy Changes During

Chemical Reactions

Learning Objectives

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

- Identify endergonic and exergonic reactions and describe the change in Gibbs free energy.
- Define activation energy and explain why even exergonic reactions have an activation energy.

10.1.1 Gibbs Free Energy

It is critical for scientists to be able to quantify and express the energy changes associated with chemical reactions, specifically so that the energy released from one reaction be compared to that of another reaction. A measurement of **Gibbs free energy** is used to quantitate these energy transfers. Gibbs free energy (abbreviated with the letter G) was named after Josiah Willard Gibbs, the scientist who developed the measurement.

Recall that according to the second law of thermodynamics,

all energy transfers involve the loss of some amount of energy in an unusable form such as heat, resulting in entropy. Gibbs free energy specifically refers to the energy associated with a chemical reaction that is available after entropy is accounted for. In other words, Gibbs free energy is usable energy, or energy that is available to do work.

Every chemical reaction involves a change in free energy, called delta G (ΔG). The change in free energy can be calculated for any system that undergoes a change, such as a chemical reaction. To calculate ΔG, start with the total energy change of the system, which is called **enthalpy**, and is denoted as ΔH. Then subtract the amount of energy that is lost to entropy (denoted as ΔS). The formula for calculating ΔG is as follows, where the symbol T refers to absolute temperature in Kelvin (degrees Celsius + 273):

$$\Delta G = \Delta H - T\Delta S$$

The *standard* free energy change of a chemical reaction is expressed either in kilojoules per mole (kJ/mol) or kilocalories per mole (kcal/mol) under standard pH, temperature, and pressure conditions. In biological systems, standard conditions are typically pH 7.0, 25 degrees Celsius, and 1 atm pressure. However, cellular conditions can vary considerably from these standard conditions, and so *actual* ΔG values for biological reactions may differ.

10.1.2 Endergonic Reactions and Exergonic Reactions

If energy is released during a chemical reaction, then the resulting value from the above equation for ΔG will be a negative number. A negative ΔG means that the products of the reaction have less free energy than the reactants, because they gave off some free energy during the reaction. Reactions that have a negative ΔG and consequently release free energy are called **exergonic reactions**. (Think: *ex*ergonic means energy is *ex*iting the system.) These reactions are also referred to as spontaneous reactions, because they can occur without the addition of energy into the system (**Figure 10.2**). Understanding which chemical reactions are spontaneous and release free energy is extremely useful for biologists, because these reactions can be harnessed to perform work inside the cell. An important distinction must be drawn between the term spontaneous and the idea of a chemical reaction that occurs immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. The rusting of iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

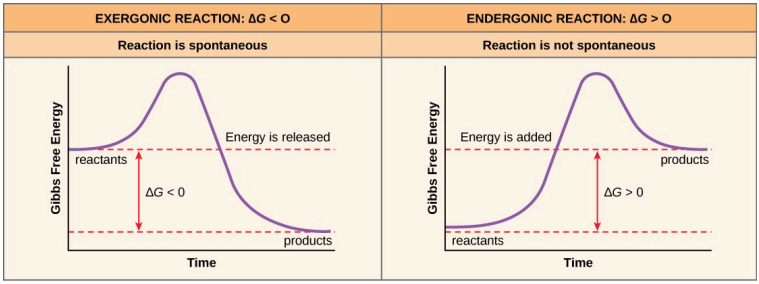


Figure 10.2 Exergonic and endergonic reactions result in changes in Gibbs free energy. Exergonic reactions release energy; endergonic reactions require energy to proceed.

If a chemical reaction requires an input of energy rather than releasing energy, then the ΔG for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, the products of these reactions can be thought of as energy-storing molecules. These chemical reactions are called **endergonic reactions**, and they are non-spontaneous. An endergonic reaction will not take place on its own with the addition of energy (**Figure 10.2**).

Let's consider the example of the synthesis and breakdown of the food molecule, glucose. In metabolism, the building of complex molecules, such as sugars, from simpler ones is called an **anabolic** process and requires energy. Therefore, the chemical reactions involved in anabolic processes are endergonic reactions. On the other hand, the **catabolic** process of breaking sugar down into simpler molecules releases energy in a series of exergonic reactions. Figure **10.3** shows some other examples of endergonic and exergonic reactions.

Figure 10.3 Shown are some examples of endergonic processes (ones that require energy) and exergonic processes (ones that release energy). These include (a) a compost pile decomposing, (b) a chick hatching from a fertilized egg, (c) sand art being destroyed, and (d) a ball rolling down a hill. (Credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by "Athlex"/Flickr; credit d: modification of work by Harry Malsch) Look at each of the processes shown, and decide if it is endergonic or exergonic.

Another 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, releasing energy into their environment in one direction, and absorbing it 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. This state of equilibrium is one of the lowest possible free energy and a state of maximal entropy.

10.1.3 Activation Energy

Even exergonic reactions require a small amount of energy input to get going before they can proceed with their energy-releasing steps. These reactions have a net release of energy, but still require some energy in the beginning. This small amount of energy input necessary for all chemical reactions to occur is called the **activation energy**.

Why would an energy-releasing, negative ΔG reaction actually require some energy to proceed? The reason lies in the steps that take place during a chemical reaction. During chemical reactions, certain chemical bonds are broken and new ones are formed. For example, when a glucose molecule is broken down, bonds between the carbon atoms of the molecule are broken. Since these are energy-storing bonds, they release energy when broken. However, to get them into a state that allows the bonds to break, the molecule must be somewhat contorted. A small energy input is required to achieve this contorted state. This contorted state is called the

transition state, and it is a high-energy, unstable state. For this reason, reactant molecules don't last long in their transition state, but very quickly proceed to the next steps of the chemical reaction.

Free energy diagrams (also called reaction coordinate diagrams) illustrate the energy profiles for a given reaction (**Figure 10.4**). Whether the reaction is exergonic or endergonic determines whether the products of the reaction will exist at a lower or higher energy state than the reactants. However, regardless of this measure, the transition state will always exists at a higher energy state than the reactants.

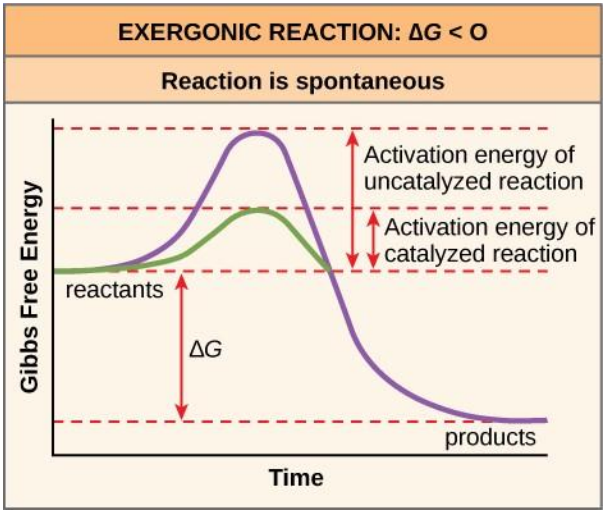


Figure 10.4 Activation energy is the energy required for a reaction to proceed, and it is lower if the reaction is catalyzed. The horizontal axis of this diagram describes the sequence of events in time. If no activation energy were required to break down sucrose (table sugar), would you be able to store it in a sugar bowl?

Where does the activation energy required by chemical reactants come from? The source of the activation energy needed to push reactions forward is typically heat energy from the surroundings. Heat energy speeds up the motion of molecules, increasing the frequency and force with which they collide; it also moves atoms and bonds within the molecule slightly, helping them reach their transition state. For this reason, heating up a system will cause chemical reactants within that system to react more frequently. Once reactants

have absorbed enough heat energy from their surroundings to reach the transition state, the reaction will proceed.

The activation energy of a particular reaction determines the rate at which it will proceed. The higher the activation energy, the slower the chemical reaction will be. The example of iron rusting illustrates an inherently slow reaction. This reaction occurs slowly over time because of its high activation energy. Additionally, the burning of many fuels, which is strongly exergonic, will take place at a negligible rate unless their activation energy is overcome by sufficient heat from a spark. Once they begin to burn, however, the chemical reactions release enough heat to continue the burning process, supplying the activation energy for surrounding fuel molecules.

Like these reactions outside of cells, the activation energy for most cellular reactions is too high for heat energy to overcome at efficient rates. In other words, in order for important cellular reactions to occur at appreciable rates (number of reactions per unit time), their activation energies must be lowered (**Figure 10.4**). As we are about to discuss, enzymes function by lowering the activation energy and increasing the reaction rate of reactions. This is a very good thing as far as living cells are concerned. Important macromolecules, such as proteins, DNA, and RNA, store considerable energy, and their breakdown is exergonic. If cellular temperatures alone provided enough heat energy for

these exergonic reactions to overcome their activation barriers, the essential components of a cell would disintegrate.

10.2 | Enzymes

Learning Objectives

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

- Explain how enzymes function as molecular catalysts.
- Discuss enzyme regulation by various factors.

A substance that speeds up a chemical reaction without being changed by that reaction is a **catalyst**. Proteins that catalyze biochemical reactions are called **enzymes**. An example of an enzyme is salivary amylase, which hydrolyzes its substrate amylose, a component of starch.

Enzymes 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. Since all enzymes increase the rate of reaction, they are considered to be organic catalysts (**Figure 10.4**). Note that although enzymes can speed up reactions, they can never change an endergonic reaction to an exergonic one. Whether a reaction is endergonic or exergonic depends only on the chemical makeup on the reactant and product molecules and cannot be changed by a catalyst.

10.2.1 Enzyme Active Site and Substrate Specificity

The chemical reactants to which an enzyme binds are called 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.

The location within the enzyme where the substrate binds is called the enzyme's **active site**. Since enzymes are proteins, there is a unique combination of amino acid side chains, which creates a very specific chemical environment within the active site. This specific environment is suited to bind to a specific substrate (or substrates). Due to this jigsaw puzzle-like match between an enzyme and its substrates, enzymes are known for their specificity. The "best fit" results from the shape and the amino acid functional group's attraction to the substrate.

There is a specifically matched enzyme for each substrate and, thus, for each chemical reaction.

Since enzymes are proteins, their shape is sensitive to variations in temperature and pH. 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 to denature. Likewise, pH can also affect enzyme function. Enzymes are suited to function best within a certain pH range, and extreme pH values can cause enzymes to denature.

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** (**Figure 10.5**). 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 an ideal binding arrangement between the enzyme and the transition state of the substrate. This ideal binding maximizes the enzyme's ability to catalyze its reaction.

Enzymes will always return to their original state once the

reaction is complete. One of the hallmark properties of enzymes is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme is finished catalyzing a reaction, it releases its product(s).

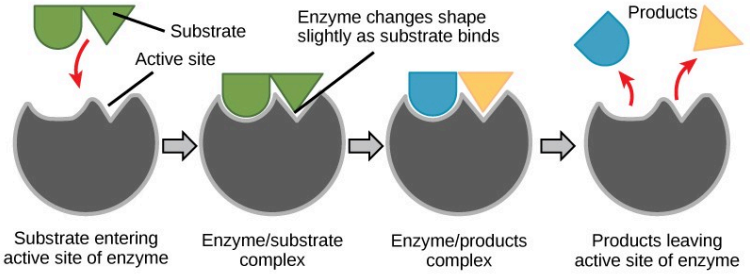


Figure 10.5 According to the induced-fit model, both enzyme and substrate undergo dynamic conformational changes upon binding. The enzyme contorts the substrate into its transition state, thereby increasing the rate of the reaction.

10.2.2 Control of Metabolism Through Enzyme Regulation

It would seem ideal to have a scenario in which all of the enzymes encoded in an organism's genome existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. In reality, this is far from the case. A variety of mechanisms ensure that this does not happen. 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 that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so do the amounts and functionality of different enzymes.

The relative amounts and functioning of the variety of enzymes within a cell ultimately determine which reactions will proceed and at which rates. This determination is tightly controlled. In certain cellular environments, enzyme activity is partly controlled by environmental factors, like pH and temperature. There are other mechanisms through which cells control the activity of enzymes and determine the rates at which various biochemical reactions will occur.

Regulation of Enzymes by Small Molecules

There are many different kinds of molecules that inhibit or promote enzyme function. In some cases, an inhibitor molecule binds to the active site of an enzyme and physically blocks the substrate from binding. This is called **competitive inhibition** because an inhibitor molecule competes with the substrate for active site binding (**Figure 10.6**).

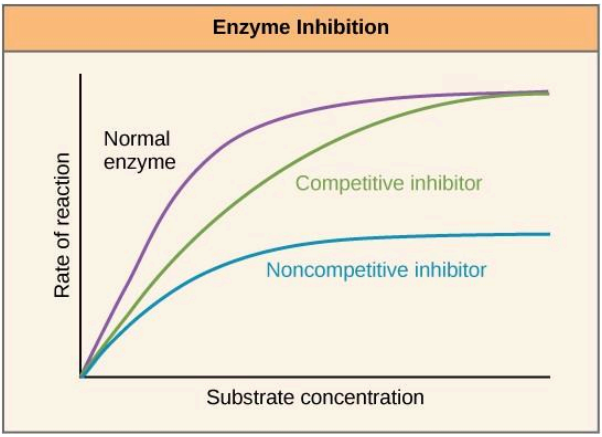


Figure 10.6 Competitive and noncompetitive inhibition affect the rate of reaction differently. Competitive inhibitors affect the initial rate but do not affect the maximal rate, whereas noncompetitive inhibitors affect the maximal rate.

In other cases, an inhibitor molecule binds to the enzyme in a location other than the active site, called the allosteric site, and causes a shape change in the enzyme that blocks substrate binding to the active site. This type of inhibition is called **allosteric inhibition** (**Figure 10.7**). There are also allosteric activators, which bind to an allosteric site and induce a conformational change that increases the affinity of the enzyme for its substrate(s).

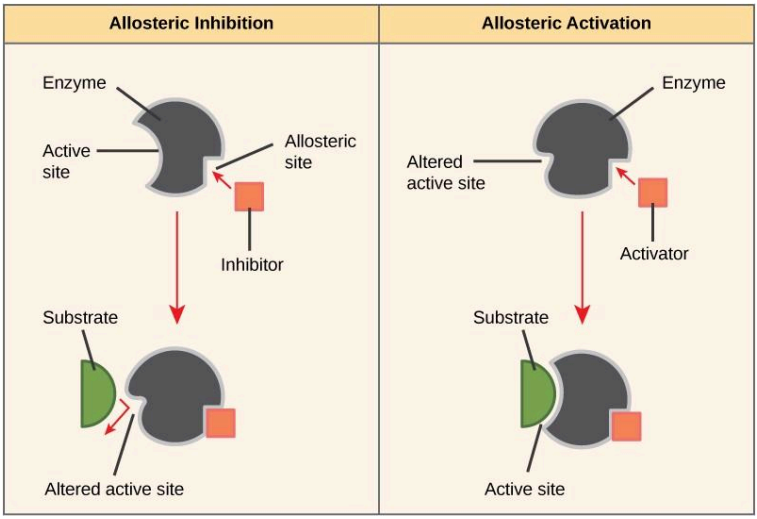


Figure 10.7 Allosteric inhibitors modify the active site of the enzyme so that substrate binding is reduced or prevented. In contrast, allosteric activators modify the active site of the enzyme so that the affinity for the substrate increases.

Figure 10.8 Have you ever wondered how pharmaceutical drugs are developed? (credit: Deborah Austin)

Drug Discovery by Looking for Inhibitors of Key Enzymes in Specific Pathways

Enzymes are key components of metabolic pathways. Understanding how enzymes work and how they can be regulated is a key principle behind the development of many of the pharmaceutical drugs (**Figure 10.8**) on the market today. Biologists working in this field

collaborate with other scientists, usually chemists, to design drugs.

Consider statins for example—which is the name given to the class of drugs that reduces cholesterol levels. These compounds are essentially inhibitors of the enzyme HMG-CoA reductase. HMG-CoA reductase is the enzyme that synthesizes cholesterol from lipids in the body. By inhibiting this enzyme, the levels of cholesterol synthesized in the body can be reduced. Similarly, acetaminophen, popularly marketed under the brand name Tylenol, is an inhibitor of the enzyme cyclooxygenase. While it is effective in providing relief from fever and inflammation (pain), its mechanism of action is still not completely understood.

How are drugs developed? One of the first challenges in drug development is identifying the specific molecule that the drug is intended to target. In the case of statins, HMG-CoA reductase is the drug target. Drug targets are identified through painstaking research in the laboratory. Identifying the target alone is not sufficient; scientists also need to know how the target acts

inside the cell and which reactions go awry in the case of disease. Once the target and the pathway are identified, then the actual process of drug design begins. During this stage, chemists and biologists work together to design and synthesize molecules that can either block or activate a particular reaction. However, this is only the beginning: both if and when a drug prototype is successful in performing its function, then it must undergo many tests from in vitro experiments to clinical trials before it can get FDA approval to be on the market.

Many enzymes don't work optimally, or even at all, unless bound to other specific non-protein helper molecules. Two types of helper molecules are **cofactors** and **coenzymes**. When these molecules bind to their enzymes, they promote optimal conformation and function. Cofactors are inorganic ions such as iron (Fe++) and magnesium (Mg++). Coenzymes are organic helper molecules. The most common sources of coenzymes are dietary vitamins (**Figure 10.9**). For example, vitamin C is a coenzyme for multiple enzymes that take part in building the important connective tissue component, collagen. 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.

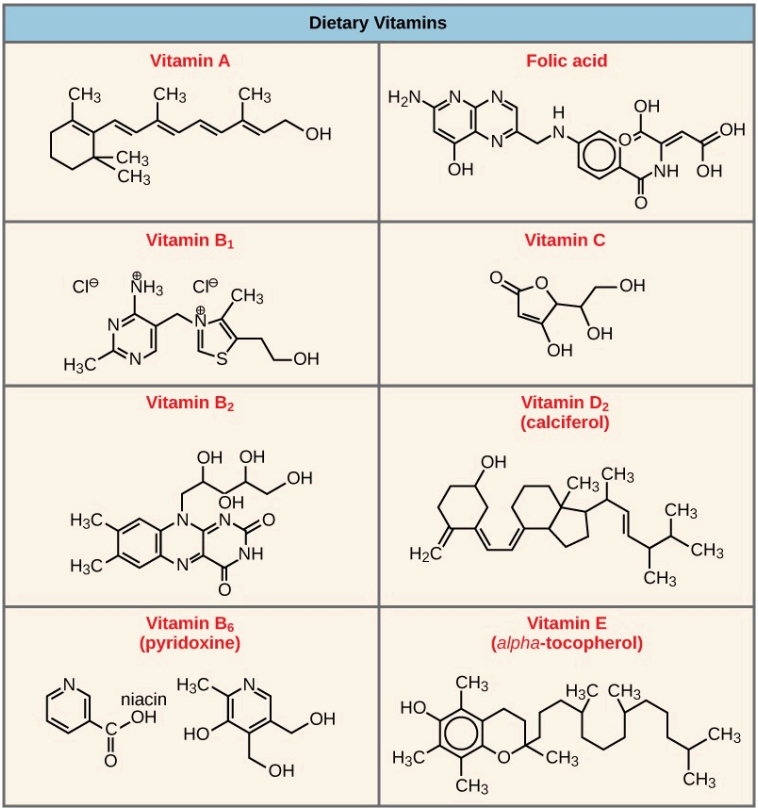


Figure 10.9 Vitamins are important coenzymes or precursors of coenzymes, and are required for enzymes to function properly. Multivitamin capsules usually contain mixtures of all the vitamins at different percentages.

Regulation of Enzymes by Post-translational Modification

Some amino acid side chains in enzymes can be chemically modified after the protein itself is synthesized (translated from an mRNA). These changes are collectively called **post-translational modifications (PTMs)** (**Table 10.1**).

Table 10.1 Some common post-translational modifications.

Name of modification	Chemical group added	Structure	Ami aci modi
Phosphorylation	Phosphate	PO_3^{4-}	Seri threo tyros
Glycosylation	Mono-, di-, and polysaccharides	Various	Aspara (N-lin serin threo (O-lin
Ubiquitination	Ubiquitin	8-kDa polypeptide of 76 amino acids	Lysi
Nitrosylation	Nitric oxide	NO	Cyst
Methylation	Methyl	$-CH_3$	Ma contai either N
Acetylation	Acetyl	$-COCH_3$	Lysi
Lipidation	Lipid	Many, but most well-known are prenyl and fatty acyl groups and glycosylphosphatidylinositols (GPI)	Vario inclu N- an term

Phosphorylation is the most well-studied PTM, where a phosphate group is added to a serine, threonine, or tyrosine residue by an enzyme known as a **kinase**. That phosphate group can then be removed by another enzyme known as a **phosphatase.**

How a post-translational modification affects a protein's activity depends on a) the protein being modified, b) the PTM

being added, and c) the site at which the PTM is added. Enzyme X might be activated by phosphorylation at serine 101, but deactivated by phosphorylation at serine 152. In this scenario, two different kinases would catalyze the addition of the phosphate group to the two different serine residues, and the kinases would likely respond to different cellular states.

Regulation of Enzymes by Changes in Gene Expression

As we will discuss in detail later in this text, all cells carry the same DNA blueprint, but individual cells express only a subset of those genes at a time. For instance, a cell growing in a low-serine environment might synthesize more of the enzyme molecules involved in the serine synthesis pathway than a cell growing in a serine-rich environment.

Regulation of Enzymes by 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.

Table 10.2 Summary of enzyme regulation methods

Method	Speed	Reversible?
Small molecules	Fast	Yes
Post-translational modifications	Fast	Yes
Gene expression	Slow	Yes, but on a longer timescale
Compartmentalization	N/A	No

10.3 | Energy Transfer in Living Systems: Adenosine Triphosphate (ATP)

Learning Objectives

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

- Explain the role of ATP as the cellular energy currency.
- Describe how energy is released through hydrolysis of ATP.

10.3.1 ATP Structure & Function

As discussed above, even exergonic reactions require a small amount of activation energy in order to proceed. However, consider endergonic reactions, which require much more energy input, because their products have more free energy than their reactants. Within the cell, where does energy to power such reactions come from? The answer frequently lies with an energy-supplying molecule called **adenosine triphosphate**, or **ATP**. ATP is a small, relatively simple molecule (**Figure 10.10**), but, due to its chemical structure, it contains the potential for a quick burst of energy that can be harnessed to perform cellular work. ATP is the primary energy currency of cells and is used to power the majority of energy-requiring cellular reactions.

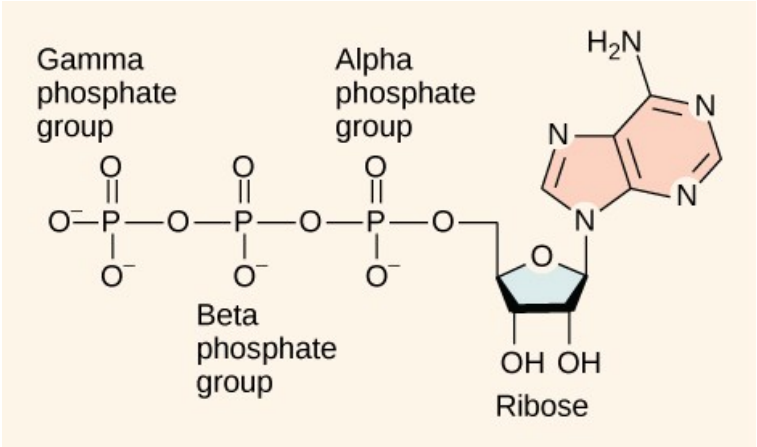


Figure 10.10 ATP is the primary energy currency of the cell. It has a nucleotide with ribose sugar, adenine, and three phosphate groups.

Adenosine triphosphate is comprised of a ribonucleotide containing adenine and three phosphate groups (**Figure 10.10**). When the bond that links the beta and gamma phosphate groups is broken, the reaction releases sufficient energy to power a variety of cellular reactions and processes.

The reason that this reaction releases so much energy is because the products of such bond breaking—adenosine diphosphate (ADP) and one inorganic phosphate group (P_i)—have considerably lower free energy than the reactants: ATP and a water molecule. Because this reaction takes place with the use of a water molecule, it is considered a hydrolysis reaction. In other words, ATP is hydrolyzed into ADP in the following reaction:

$$ATP + H_2O \rightarrow ADP + P_i + \text{free energy}$$

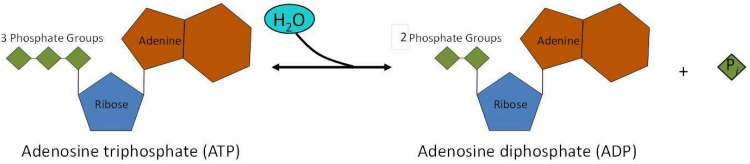


Figure 10.11 ATP is hydrolyzed into ADP and inorganic phosphate, releasing energy that can be used to do work in the cell.

The same reaction is shown in **Figure 10.11**. Like most chemical reactions, the hydrolysis of ATP to ADP is reversible. The reverse reaction regenerates ATP + H_2O from ADP + P_i.

ATP is a highly unstable molecule. Unless quickly used to perform work, ATP spontaneously dissociates into ADP + Pi, and the free energy released during this process is lost as heat. Cells use a process called **energy coupling** to harness the energy released by ATP hydrolysis to perform work inside the cell. In this process, an exergonic reaction, such as hydrolyzing ATP, is used to drive an endergonic reaction.

ATP is constantly being used and regenerated in a process known as the **ATP cycle** (**Figure 10.12**). ADP phosphorylation, an energy-requiring process, is coupled to a sufficiently energy-releasing process, nutrient breakdown, to allow ATP regeneration to occur. Then, ATP hydrolysis is

coupled to cellular endergonic processes such as anabolism, active transport, and cell division, providing the required energy.

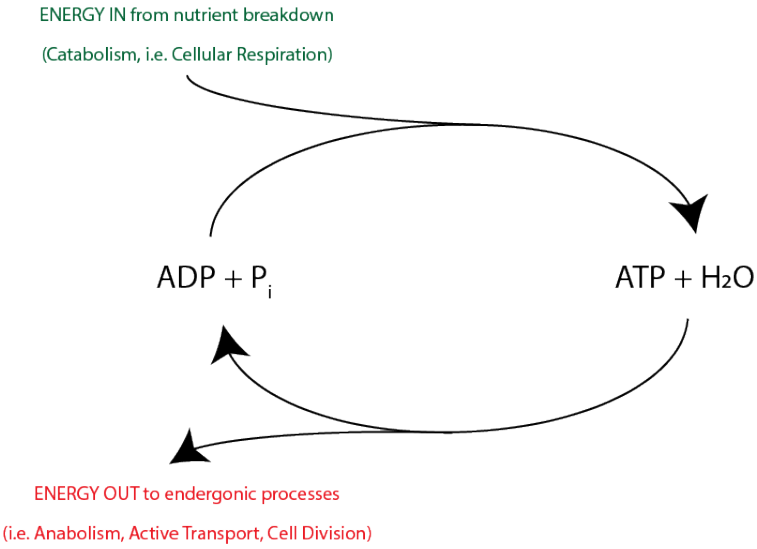


Figure 10.12 The ATP cycle. (Credit: Katherine Mattaini.)

10.3.2 ATP Generation

Substrate-Level Phosphorylation

During the breakdown of glucose, a few ATP molecules are regenerated from ADP as a direct result of the chemical reactions that occur in the 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 (**Figure 10.13**). This very direct method of phosphorylation is called **substrate-level phosphorylation**.

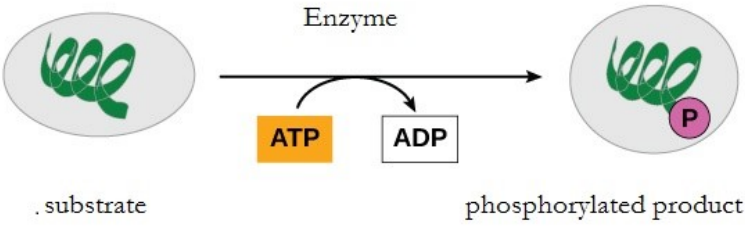


Figure 10.13 During phosphorylation reactions, a phosphate group is transferred from ATP to a substrate, resulting in a phosphorylated product.

Oxidative Phosphorylation

However, most of the ATP generated during glucose catabolism is derived from a much more complex process, **chemiosmosis**, which takes place in the **mitochondria** (**Figure 10.14**) of eukaryotic cells and in the plasma membrane of prokaryotic cells. The production of ATP by chemiosmosis in the mitochondria is called **oxidative phosphorylation** because it requires oxygen.

Chemiosmosis is also used in the light reactions of photosynthesis (in chloroplasts) to harness the energy of sunlight. This process is called **photophosphorylation** because of the involvement of light energy.

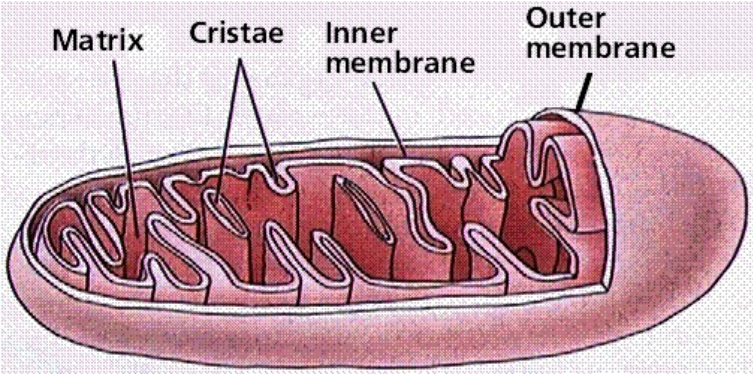


Figure 10.14 In eukaryotes, oxidative phosphorylation takes place in mitochondria. In prokaryotes, this process takes place in the plasma membrane.

10.4 | Energy Transfer in Living Systems: Electron Transfer

Learning Objectives

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

- Discuss the importance of electrons in the

transfer of energy in living systems.
- Identify reduction and oxidation reactions.
- Be familiar with common electron carriers.

10.4.1 The Carbon Cycle

The metabolism of sugar (a simple carbohydrate) is a classic example of the many cellular processes that use and produce energy. Living things consume sugar as a major energy source, because sugar molecules have a great deal of energy stored within their bonds. The breakdown of glucose into carbon dioxide and water is described by the equation:

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O + \text{energy}$$

Carbohydrates have their origins in photosynthesizing organisms, such as plants. During photosynthesis, plants use the energy of sunlight to convert carbon dioxide gas (CO_2) into sugar molecules, such as glucose ($C_6H_{12}O_6$). The synthesis of glucose is described by the reverse of the previous equation:

$$6CO_2 + 6H_2O + \text{energy} \rightarrow C_6H_{12}O_6 + 6O_2$$

Because photosynthesis involves synthesizing a larger, energy-storing molecule (glucose), it requires an input of energy to proceed. First, energy from sunlight is transformed into chemical energy in the form of ATP. The stored energy

in ATP is then used to build one molecule of glucose from six molecules of CO_2.

Both ATP and glucose store energy in their chemical bonds. However, ATP is very unstable and cannot be stored for more than a few seconds. Therefore, photosynthetic cells use the ATP to make glucose, which is more stable and can be used to make disaccharides or polysaccharides for long-term energy storage.

When cells need to use this energy to do work, sugar molecules are broken down through a complex series of chemical reactions in order to harvest the energy stored in their chemical bonds. The energy harvested is used to make ATP molecules, which can be used to power many chemical reactions and processes in the cell. If ATP is the dollar bill of the cell's energy currency, carbohydrates synthesized from glucose can be thought of as one of the organism's savings accounts.

Together, photosynthesis and cellular respiration make up the **carbon cycle,** so named because carbon cycles from CO_2 to glucose. Perhaps a better name for this cycle would be the "energy cycle". The point of photosynthesis is to use the sun's energy to make ATP, which is then used to make glucose (which stores the energy). The point of cellular respiration is to use the energy from glucose to make ATP, which is used to drive work in the cell. Photosynthetic organisms, such as plants, algae, and some bacteria, can do both parts of the carbon cycle. Non-photosynthetic organisms can only do

half. The next two chapters will describe the two halves of the carbon cycle in detail.

Figure 10.15 Plants use energy from sunlight to make sugar and other organic molecules. Both plants and animals use cellular respiration to derive energy from the organic molecules originally produced by plants. (Credit "acorn": modification of work by Noel Reynolds; credit "squirrel": modification of work by Dawn Huczek)

10.4.2 Redox Reactions and Energy

Free energy is defined as the ability to do work. Energy exists in different forms, including electrical energy, light energy, chemical energy, and heat energy. In order to appreciate the way energy flows into and out of biological systems, it is important to understand more about the different types of energy that exist in the physical world.

Energy production within a cell involves many coordinated biochemical pathways. Most of these pathways are combinations of oxidation and reduction reactions. Oxidation and reduction occur in tandem. An oxidation reaction strips an electron from an atom in a compound, and then adds this electron to another compound in a reduction reaction. Because oxidation and reduction usually occur together, these pairs of reactions are called reduction-oxidation, or **redox reactions**. A mnemonic to help remember which is which is OIL RIG, which stands for **O**xidation **I**s **L**oss of an electron and **R**eduction **I**s **G**ain of an electron.

The removal of an electron from a molecule, oxidizing it, results in a decrease in potential energy in the oxidized compound. The transfer of electrons between molecules is important because most of the chemical energy stored in molecules is in the form of high-energy electrons. The transfer of energy in the form of electrons allows the cell to transfer and use energy in an incremental fashion—in small packages rather than in a single, destructive burst.

10.4.3 Electron Carriers

In living systems, a small class of compounds functions as electron shuttles: They bind and carry high-energy electrons between compounds in pathways. The principal electron carriers we will consider are derived from the B vitamin group

and are derivatives of nucleotides. These compounds can be easily reduced (by accepting electrons) or oxidized (by losing electrons). Nicotinamide Adenine Dinucleotide (NAD^+) (**Figure 10.16**) is derived from vitamin B3 (niacin). NAD^+ is the oxidized form of the molecule; NADH is the reduced form of the molecule after it has accepted two electrons and a proton (the equivalent of a hydrogen atom plus an extra electron).

NAD^+ can accept electrons from an organic molecule according to the general equation:

R-H	+	**NAD+**	**NADH**	+	**R**
Reducing Agent		Oxidizing Agent	Reduced Form		Oxidized Form

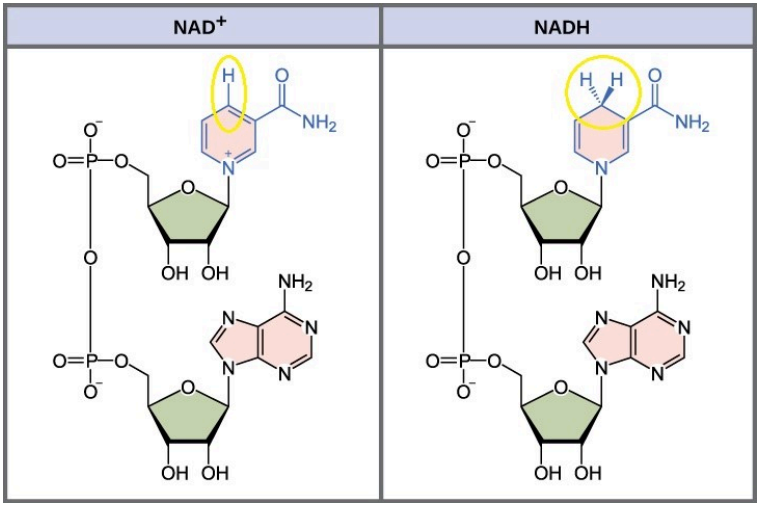


Figure 10.16 The oxidized form of the electron carrier (NAD+) is shown on the left and the reduced form (NADH) is shown on the right. The nitrogenous base in NADH has one more hydrogen ion and two more electrons than in NAD+. Note the yellow highlighted circle around the change in the molecule.

Similarly, flavin adenine dinucleotide (FAD) is derived from vitamin B2 (riboflavin). Its reduced form is $FADH_2$. $NADP^+$ is a variant of NAD^+ that has an extra phosphate group. Both NAD^+ and FAD are used in energy extraction from sugars, and $NADP^+$ plays an important role in photosynthesis.

10.5 Metabolic Pathways

Learning Objectives

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

- Define a metabolic pathway and identify whether one is catabolic or anabolic.
- Discuss feedback inhibition in metabolic pathways.

A metabolic pathway is a series of sequential biochemical reactions that convert a substrate molecule or molecules, step-by-step, through a series of metabolic intermediates, eventually yielding a final product or products. Pathways that build more complex molecules out of smaller ones are called **anabolic** pathways. Pathways that break more complex molecules into smaller ones are called **catabolic** pathways. Anabolic pathways require energy, while catabolic pathways produce energy. Consequently, metabolism is composed of building (anabolism) and degradation (catabolism) (**Figure 10.17**).

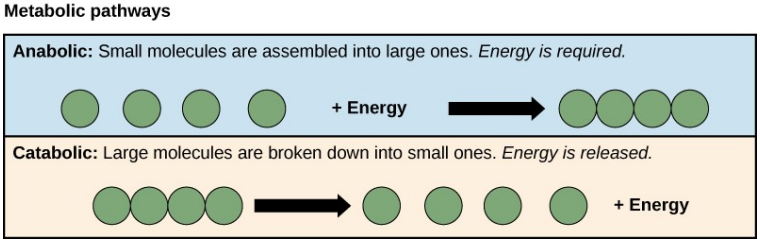


Figure 10.17 Anabolic pathways are those that require energy to synthesize larger molecules. Catabolic pathways are those that generate energy by breaking down larger molecules. Both types of pathways are required for maintaining the cell's energy balance.

10.4.1 Evolution of Metabolic Pathways

There is more to the complexity of metabolism than understanding the metabolic pathways alone. Metabolic complexity varies from organism to organism. Photosynthesis is the primary pathway in which photosynthetic organisms like plants harvest the sun's energy and convert it into carbohydrates. (The majority of global synthesis is done by planktonic algae.) The by-product of photosynthesis is oxygen, required by some cells to carry out cellular respiration. During cellular respiration, oxygen aids in the catabolic breakdown of carbon compounds, like carbohydrates. Among the products of this catabolism are CO_2 and ATP. In addition,

some eukaryotes perform catabolic processes without oxygen (fermentation); that is, they perform anaerobic metabolism.

Organisms probably evolved anaerobic metabolism to survive, since living organisms came into existence about 3.8 billion years ago, when the atmosphere lacked oxygen. Despite the differences between organisms and the complexity of metabolism, researchers have found that all branches of life share some of the same metabolic pathways, suggesting that all organisms evolved from the same ancient common ancestor (**Figure 10.18**). Evidence indicates that over time, the pathways diverged, adding specialized enzymes to allow organisms to better adapt to their environment, thus increasing their chance to survive. However, the underlying principle remains that all organisms must harvest energy from their environment and convert it to ATP to carry out cellular functions.

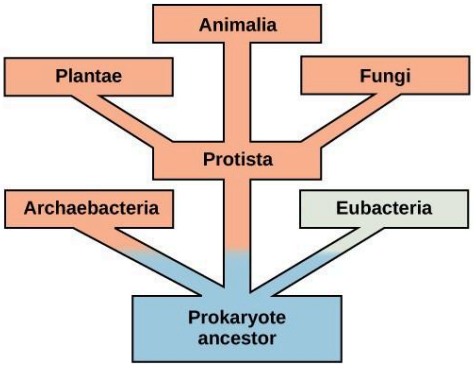


Figure 10.18 This tree shows the evolution of the various branches of life. The vertical dimension is time. Early life forms, in blue, used anaerobic metabolism to obtain energy from their surroundings.

10.4.2 Feedback Inhibition in Metabolic Pathways

Molecules can regulate enzyme function in many ways. Perhaps the most relevant sources of enzyme regulatory molecules are the products of the cellular metabolic reactions themselves. In a most efficient and elegant way, cells have evolved to use the products of their own reactions for feedback inhibition of enzyme activity. **Feedback inhibition** involves the use of a reaction product to regulate its own further production (**Figure 10.19**). The cell responds to the abundance of specific products by slowing down production. Such reaction products may inhibit the enzymes that catalyzed their production through the mechanisms described above.

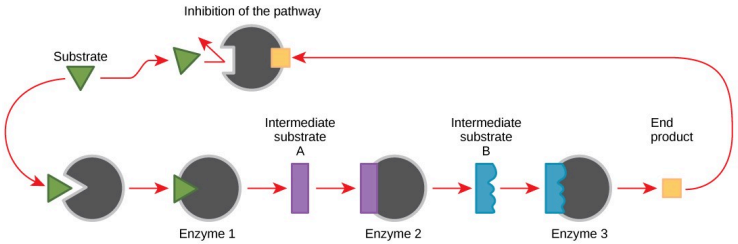


Figure 10.19 Metabolic pathways are a series of reactions catalyzed by multiple enzymes. Feedback inhibition, where the end product of the pathway inhibits an upstream step, is an important regulatory mechanism in cells.

For example, ATP is an allosteric regulator of some of the enzymes involved in the catabolic breakdown of sugar, the process that produces ATP. In this way, when ATP is abundant, the cell can prevent its further production. Remember that ATP is an unstable molecule that can spontaneously dissociate into ADP. If too much ATP were present in a cell, much of it would go to waste. On the other hand, ADP serves as an allosteric activator for some of the same enzymes that are inhibited by ATP. Thus, when relative levels of ADP are high compared to ATP, the cell is triggered to produce more ATP.

CHAPTER 11. CELLULAR RESPIRATION

Figure 11.1 This geothermal energy plant transforms thermal energy from deep in the ground into electrical energy, which can be easily used. (Credit: modification of work by the U.S. Department of Defense.)

Chapter Outline

- 11.1 Overview of Cellular Respiration

- 11.2 Glycolysis
- 11.3 Pyruvate Oxidation and the Citric Acid Cycle
- 11.4 Oxidative Phosphorylation
- 11.5 Metabolism Without Oxygen
- 11.6 Metabolism of Other Nutrients
- 11.7 Regulation of Cellular Respiration

Introduction

The electrical energy plant in **Figure 11.1** converts energy from one form to another form that can be more easily used. This type of generating plant starts with underground thermal energy (heat) and transforms it into electrical energy that will be transported to homes and factories. Like a generating plant, plants and animals must also take in energy from the environment and convert it into a form that their cells can use.

Energy enters an organism's body in one form and is converted into another form. In plants and other photosynthetic producers, the process of photosynthesis takes in light energy and converts it into chemical energy in the form of glucose. This energy is stored in the chemical bonds

of the glucose. Next, a series of metabolic pathways, called **cellular respiration**, extracts the energy from the bonds in glucose and converts it into a form that all living things can use: ATP.

11.1 | Overview of Cellular Respiration

Learning Objectives

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

- Discuss how electron movement and energy are related in cellular respiration.
- Discuss how breathing is related to cellular respiration.
- Name the four phases of cellular respiration.

As discussed in the previous chapter, oxidation of (removing electrons from) energy-storing molecules like glucose releases energy that can be used to do cellular work. Cellular respirations involves a series of electron transfers from a high

energy state in glucose to a low energy state, as part of water. The energy released in this process is used to make ATP, both through substrate-level phosphorylation and oxidative phosphorylation. As electrons are progressively removed from the glucose carbons, glucose is ultimately broken down into 6 molecules of CO_2.

The full equation for cellular respiration is listed below.

$$C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O + \text{energy}$$

As you can see, oxygen is required for cellular respiration. Without oxygen to act as the final electron acceptor, glucose cannot be fully broken down to CO_2. We breathe air and extract oxygen from it in order to break down glucose (and other nutrients) and produce ATP. When we exhale, we release the CO_2 that is the byproduct of glucose breakdown.

Cellular respiration occurs in four phases, that will be discussed in detail below.

Phase 1: Glycolysis

Phase 2: Pyruvate oxidation

Phase 3: The citric acid cycle

Phase 4: Oxidative phosphorylation

11.2 | Glycolysis

Learning Objectives

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

- Describe the overall result in terms of molecules produced in the breakdown of glucose by glycolysis.
- State the output of glycolysis in terms of ATP molecules and NADH molecules produced.

You have read that nearly all of the energy used by living cells resides in the bonds of the monosaccharide glucose. **Glycolysis** is the first step in the breakdown of glucose to extract energy for cellular metabolism. Nearly all living organisms carry out glycolysis as part of their metabolism. The process does not use oxygen and is therefore anaerobic. Glycolysis takes place in the cytoplasm of both prokaryotic and eukaryotic cells.

Before glycolysis can begin, glucose must be present in the cell. Glucose may be obtained in many ways, including

facilitated diffusion across the plasma membrane through carrier proteins, secondary active transport using sodium ion gradients set up by the sodium-potassium pump, breakdown of disaccharides, such as sucrose, to form glucose and other monosaccharides, and breakdown of storage polymers, such as glycogen or starch, to yield glucose.

11.2.1 Simplified Overview of Glycolysis

Glycolysis begins with a single six-carbon glucose molecule and ends with two molecules of a three-carbon sugar called pyruvate. Glycolysis consists of three stages: investment, cleavage, and harvest (**Figure 11.2).**

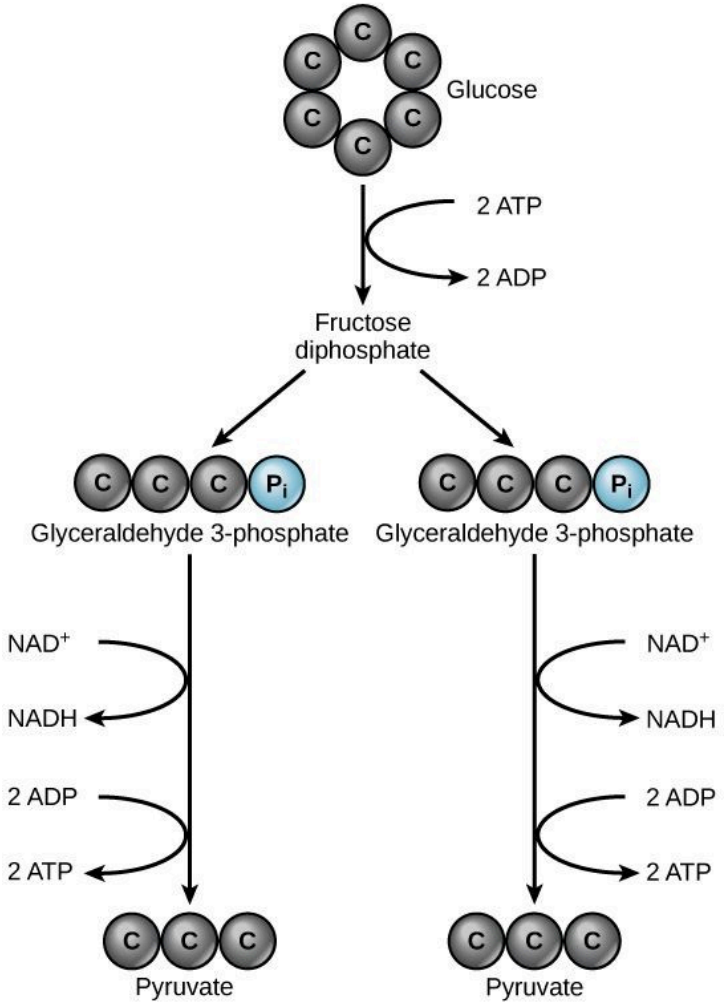


Figure 11.2 Simplified glycolysis showing investment, cleavage and harvest.

During investment, two molecules of ATP phosphorylate the six-carbon sugar. During cleavage, the six-carbon molecule is

split into two three-carbon molecules. During harvest, energy is captured from the two three carbon molecules and stored in the form of ATP and NADH.

11.2.2 Details of Glycolysis

Glycolysis is a ten-step metabolic pathway. A more detailed look at the steps of glycolysis is shown below.

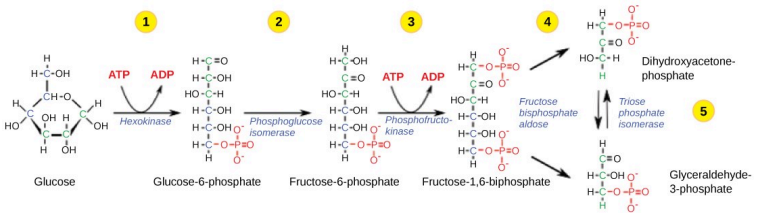


Figure 11.3 The first half of glycolysis invests two ATP molecules in the phosphorylation of glucose, to form fructose-1,6 biphosphate which is then split during cleavage into two three-carbon molecules.

Table 11.1 Steps in the first half of glycolysis.

Step 1	Investment #1	The enzyme hexokinase catalyzes the phosphorylation of glucose, using ATP as the source of phosphate, producing glucose-6-phosphate. This reaction prevents the phosphorylated glucose molecule from leaving the cell because it can no longer cross the phospholipid bilayer of the plasma membrane through any transport protein.
Step 2	Isomerize	The enzyme phosphogluco-isomerase converts glucose-6-phosphate into one of its isomers, fructose-6-phosphate. (Remember that isomers have the same chemical formula but have different chemical structures.)
Step 3	Investment #2	The enzyme phosphofructokinase catalyzes the phosphorylation of fructose-6-phosphate, using a second ATP, producing fructose-1,6-bisphosphate.
Step 4	Cleavage	The enzyme aldolase cleaves 1,6-bisphosphate into two three-carbon isomers: dihydroxyacetone-phosphate (DHAP) and glyceraldehyde-3-phosphate (3PGA)
Step 5	Isomerize	The enzyme triose-phosphate isomerase transforms DHAP into its isomer, 3PGA. Thus, the pathway will continue with two molecules of 3PGA.

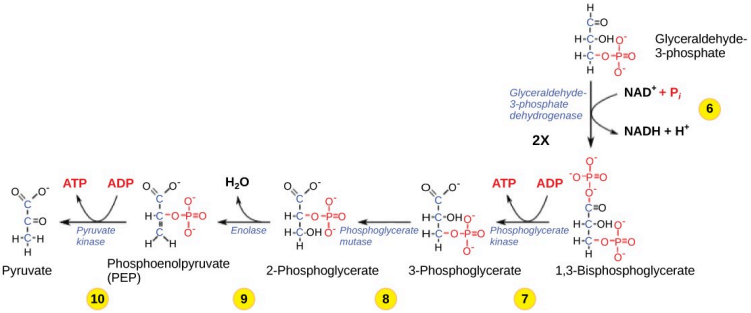


Figure 11.4. The second half of glycolysis involves phosphorylation without ATP investment (step 6) and produces two NADH and four ATP molecules per glucose. Therefore, there is a net profit of 2 ATP and 2 NADH during glycolysis. From this point in the glycolytic pathway each step must be performed twice to account for the two molecules of the 3-phosphoglyceraldehyde produced in step 5. Note the 2X in Figure 11.4.

Table 11.2 Steps in the second half of glycolysis.

Step 6	Produce NADH	The enzyme glyceraldehyde-3-phosphate dehydrogenase catalyzes the oxidation of glyceraldehyde-3-phosphate by transferring high-energy electrons to NAD+ to produce NADH. The sugar is then phosphorylated by the addition of a second phosphate group, producing 1,3-bisphosphoglycerate. Note that the second phosphate group does not require another ATP molecule.
Step 7	Produce ATP	The enzyme phosphoglycerate kinase catalyzes the transfer of a high-energy phosphate from 1,3-bisphosphoglycerate to ADP, forming one molecule of ATP. This is an example of substrate-level phosphorylation. The aldehyde group on the 1,3-bisphosphoglycerate is oxidized to a carboxyl group, and 3phosphoglycerate is formed.
Step 8	Isomerize	A mutase enzyme moves the remaining phosphate group in 3-phosphoglycerate from the third carbon to the second carbon, producing 2-phosphoglycerate (an isomer of 3-phosphoglycerate).
Step 9	Lose water	The enzyme enolase causes 2-phosphoglycerate to lose water (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	Produce ATP	The enzyme pyruvate kinase causes the production of a second ATP molecule by substrate-level phosphorylation and the compound pyruvic acid (or its ionized form, pyruvate).

11.2.3 Outcomes of Glycolysis

Glycolysis starts with glucose and ends with two **pyruvate** molecules, a total of four ATP molecules and two molecules of NADH. Since two ATP molecules were invested in the first stage of the pathway, the cell has a net gain of 2 ATP molecules and 2 NADH molecules. If the cell cannot catabolize the pyruvate molecules further, it will harvest only two ATP molecules from one molecule of glucose.

QUICK REVIEW OF GLYCOLYSIS:

- Splits a six-carbon glucose molecule into 2 three-carbon molecules of pyruvate, using 10 enzyme-catalyzed reactions
- Yields a net gain of 2 ATP and 2 NADH
- Takes place in the cytoplasm
- Can occur without oxygen; is an anaerobic process

11.3 | Pyruvate Oxidation and the Citric Acid Cycle

Learning Objectives

Type your learning objectives here.

- Describe the overall result in terms of molecules produced in the breakdown of pyruvate by pyruvate oxidation and the citric acid cycle.
- State the output of pyruvate oxidation and the citric acid cycle in terms of ATP molecules and NADH molecules produced.

11.3.1 Pyruvate Oxidation

If oxygen is available, aerobic respiration will proceed. In eukaryotic cells, the pyruvate molecules produced at the end of glycolysis are transported into the mitochondria. Once inside the mitochondria, one CO_2 will be removed from pyruvate, leaving an acetyl group that will be picked up and activated by a carrier compound called **coenzyme A (CoA)**. The resulting compound is called **acetyl CoA**. CoA is made from vitamin B5 (pantothenic acid). During the formation of acetyl CoA from pyruvate, 2 electrons are also transferred to NAD^+, forming NADH. Acetyl CoA can be used in a variety of ways by the cell, but its major function is to deliver the acetyl group derived from pyruvate to the citric acid cycle,

which is the next pathway in glucose catabolism. (**Figure 11.5**). Since there are two pyruvate molecules formed from each glucose molecule, this step occurs twice. Therefore, these are the first two of the six carbons from the original glucose molecule to be removed as CO_2.

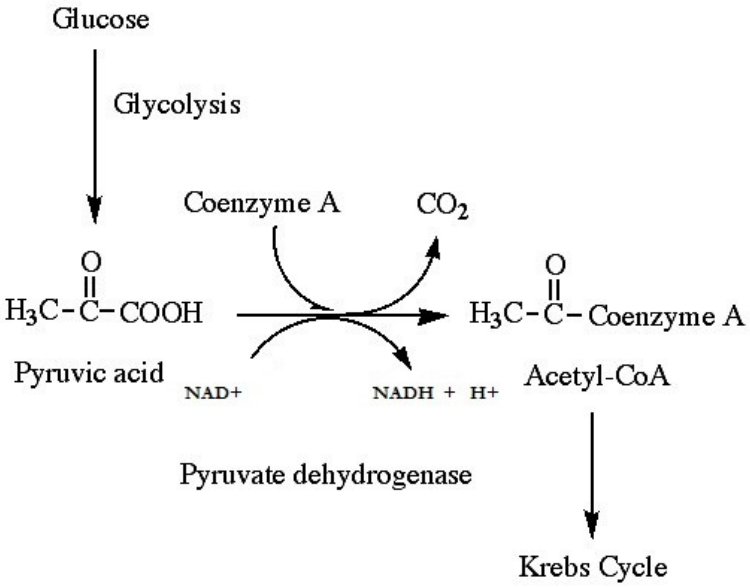


Figure 11.5 Upon entering the mitochondrial matrix, a multi-enzyme complex converts pyruvate into acetyl CoA. This complex is collectively called pyruvate dehydrogenase. In the process, carbon dioxide is released and one molecule of NADH is formed. Step 1: Split off a CO_2. The enzyme pyruvate dehydrogenase removes a carboxyl group from pyruvate, releasing a molecule of carbon dioxide into the surrounding medium. The resulting two-carbon hydroxyethyl group is then oxidized to an acetyl group, and two electrons are picked up by NAD+, forming NADH. Step 2: Attach to CoA. The acetyl group is transferred to Coenzyme A, producing a molecule of acetyl CoA. Acetyl CoA is ready to enter the citric acid cycle.

QUICK REVIEW OF PYRUVATE OXIDATION

- One carbon atom per pyruvate molecule is released

as CO_2, or two total CO_2 molecules per glucose. 2 acetyl CoA molecules per glucose remain.
- Yields a net gain of 2 NADH per glucose
- Takes place in the mitochondrial matrix
- Only occurs in the presence of oxygen; is an aerobic process

11.3.2 The Citric Acid Cycle

In the presence of oxygen, acetyl CoA delivers its acetyl group to a four-carbon molecule called **oxaloacetate**, to form a six-carbon molecule called **citric acid**. The next pathway will harvest the remainder of the extractable energy from what began as a glucose molecule. This single pathway is called the **citric acid cycle** (also called the TCA cycle or the Krebs cycle).

Like pyruvate oxidation, the citric acid cycle takes place in the matrix of mitochondria. Unlike glycolysis, the citric acid 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 redox, dehydration, hydration, and decarboxylation reactions that produce two carbon dioxide molecules from each two-carbon acetyl group, one GTP directly (which then forms ATP), and reduced forms of NADH and $FADH_2$ (**Figure 11.6**). This is considered an aerobic pathway because the NADH and $FADH_2$ produced must transfer their electrons to the next pathway in the

system, which will use oxygen. If this transfer does not occur, the oxidation steps of the citric acid cycle also do not occur. Note that the citric acid cycle produces very little ATP directly and does not directly consume oxygen.

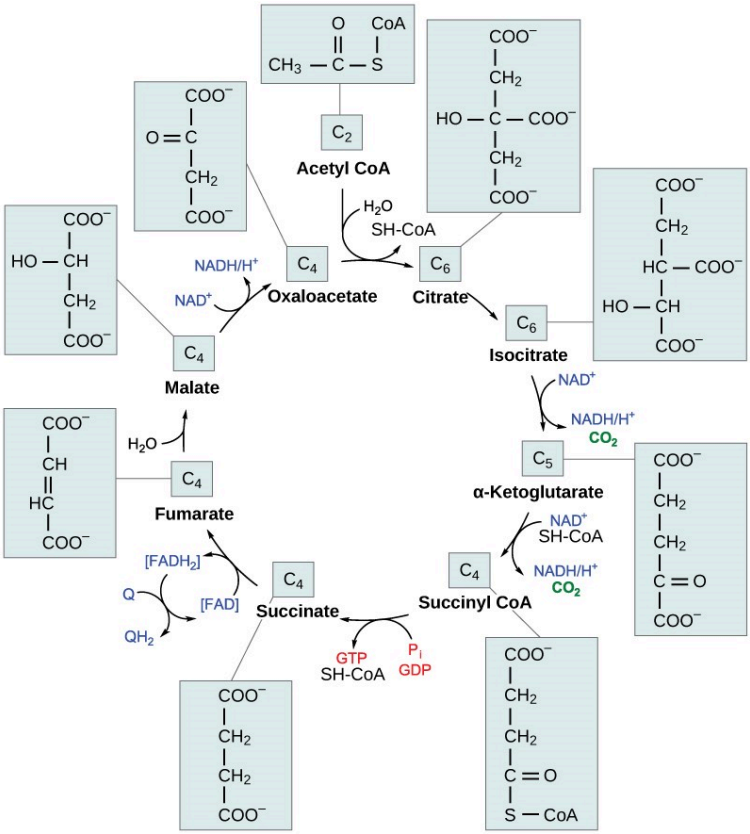


Figure 11.6 In the citric acid cycle, the acetyl group from acetyl CoA is attached to a four-carbon oxaloacetate to form a six-carbon citric acid. Citrate is oxidized, releasing two CO_2 for each acetyl group fed into the cycle. Three NAD+ and one FADH+ are reduced, and one ATP is produced. Because the final product of the citric acid cycle is also the first reactant, the cycle runs continuously in the presence of sufficient reactants. (Credit: modification of work by "Yikrazuul"/Wikimedia Commons)

Table 11.3 Steps of the citric acid cycle.

Step 1	Formation of citric acid (Top of **Figure 11.6**)	A two-carbon acetyl group is combined with a four-carbon oxaloacetate to form a six-carbon citric acid. This step is irreversible because it is highly exergonic.
Step 2	Isomerization	Citric acid is isomerized to isocitrate.
Step 3	CO_2 produced	Isocitrate is oxidized, producing a five-carbon α-ketoglutarate and CO2. NAD+ is reduced to NADH.
Step 4	CO_2 produced	α-ketoglutarate and CoA form succinyl-CoA and CO2. NAD+ is reduced to NADH. The bond between the succinate group and CoA hold a great deal of energy that will later be used to produce GTP or ATP in the next step.
Step 5	ATP or GTP produced	The bond between succinate and CoA is broken, which provides energy for the substrate-level phosphorylation of GDP to form GTP, which is quickly converted to ATP. Plants and bacteria form ATP instead. This is the only substrate-level phosphorylation in the citric acid cycle.
Step 6	$FADH_2$ produced	Succinate is converted to fumarate; two hydrogen atoms are transferred to FAD, producing $FADH_2$.
Step 7	Hydration	Fumarate is converted to malate.
Step 8	Oxaloacetate is regenerated	Malate is oxidized to form oxaloacetate and NAD+ is reduced to NADH. Since the end product re-creates one of the starting products, this pathway is a cycle.

By the end of the citric acid cycle, all six carbon atoms from

the original glucose molecule have been released as CO_2. Each turn of the cycle forms three NADH molecules and one $FADH_2$ molecule. These carriers will be used in the last pathway of aerobic respiration to produce ATP. One GTP or ATP is also made by substrate-level phosphorylation in each cycle. Since the original glucose molecule produces two acetyl CoA, two turns of the cycle occur for each glucose molecule.

QUICK REVIEW OF THE CITRIC ACID CYCLE

- The cycle begins with the two acetyl-CoA from pyruvate oxidation. By the end of the cycle, all of these carbons are released as CO_2.
- Two GTP or ATP per glucose are made by substrate-level phosphorylation
- Six NAD+ are reduced to NADH and two FAD are reduced to $FADH_2$ per glucose
- Takes place in the mitochondrial matrix
- Only occurs in the presence of oxygen; is an aerobic process

11.4 | Oxidative Phosphorylation

Learning Objectives

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

- Describe how electrons move through the electron transport chain.
- Explain how a proton (H^+) gradient is established and maintained by the electron transport chain.
- Explain how the electron transport chain and chemiosmosis together comprise oxidative phosphorylation.

You have just read about three pathways in glucose catabolism – glycolysis, pyruvate oxidation, and the citric acid cycle. Two of these pathways generate ATP a small amount of ATP by substrate-level phosphorylation. However, most of the ATP generated during the aerobic catabolism of glucose is not generated directly from these pathways. Rather, it is derived during a process called **oxidative phosphorylation,** which is made up of two parts: electron transport and chemiosmosis.

11.4.1 The Electron Transport Chain

The **electron transport chain (ETC)** is a series of proteins

embedded into the inner membrane of mitochondria and the plasma membrane of prokaryotes. The ETC consists of four protein complexes (labeled I through IV in **Figure 11.7**) and two mobile accessory electron carriers. Electron transport is a series of redox reactions that resemble a relay race or bucket brigade in that electrons are passed rapidly from one component to the next.

The NADH and $FADH_2$ that were produced during the previous phases of glucose catabolism provide the electrons to fuel the ETC. As these electrons are transported down the ETC, the electrons move from higher to lower energy states, and exergonic process. The energy released is used to shuttle hydrogen ions (H^+) from the mitochondrial matrix, across the inner membrane, into the intermembrane space. This produces an H^+ gradient, and endergonic process. The energy stored in the electrochemical gradient is then used to make ATP during chemiosmosis.

It is important to note that electrons from NADH enter the ETC at Complex I and then proceed to Complexes III and IV. However, electrons from $FADH_2$ enter the ETC at Complex II before proceeding to Complexes III and IV (**Figure 11.8**). Overall, NADH contains sufficient energy to transport three hydrogen ions into the intermembrane space, and FADH2 contains sufficient energy to transport two hydrogen ions into the intermembrane space. In this way, the hydrogen ion gradient is established and maintained between the two

compartments separated by the inner mitochondrial membrane.

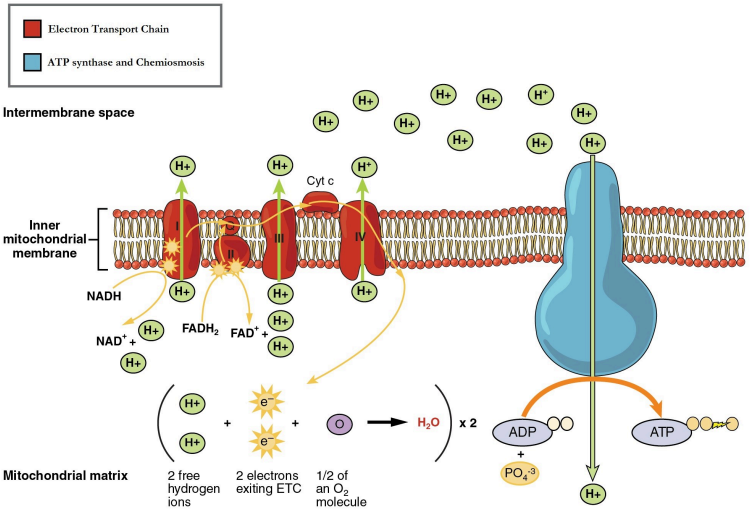


Figure 11.7 The electron transport chain is a series of electron transporters embedded in the inner mitochondrial membrane that shuttles electrons from NADH and $FADH_2$ to molecular oxygen. In the process, protons are pumped from the mitochondrial matrix to the intermembrane space, and oxygen is reduced to form water. During chemiosmosis, hydrogen ions (H^+) return to the matrix and activate ATP synthase in a process of oxidatively phosphorylation.

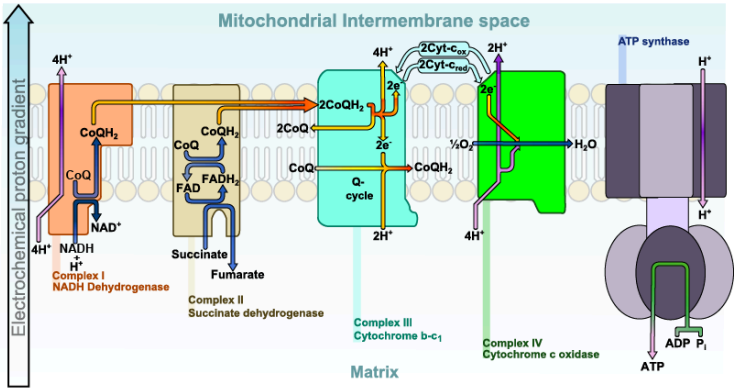


Figure 11.8 A detailed look at the movement of electrons through the electron transport chain. (Credit: "T-Fork"/Wikimedia Commons.)

Table 11.4 The movement of electrons through the mitochondrial electron transport chain.

Complex I	Integral membrane protein complex	Complex I is composed of flavin mononucleotide (FMN) and an iron-sulfur (Fe-S)-containing protein. The enzyme in complex I is NADH dehydrogenase. Two electrons are transferred from NADH to Complex I, providing the energy for it to transport a hydrogen ion from the matrix into the intermembrane space. Electrons that enter the ETC at Complex I bypass Complex II and are transferred to Coenzyme Q.
Complex II	Integral membrane protein complex	Complex II directly receives electrons from the oxidation of $FADH_2$ to FAD, using the enzyme succinate dehydrogenase (from the citric acid cycle) and does not pass through complex I. This enzyme and $FADH_2$ form a small complex that delivers electrons directly to the electron transport chain, bypassing the first complex. 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.
Coenzyme Q	Coenzyme embedded in the inner mitochondria membrane	Ubiquinone (coenzyme Q) receives electrons from complex I and complex II and delivers them to Complex III.

Complex III	Integral membrane protein complex	Complex III is composed of cytochrome b and c proteins (this complex is also called cytochrome oxidoreductase). Cytochrome proteins have a heme group, which is similar to the heme in hemoglobin, except that it carries electrons rather than oxygen. Complex III uses the energy of the electrons obtained from Q to transport another hydrogen ion from the matrix to the intermembrane space.
Cytochrome c	Small protein associated with the outside of the inner mitochondrial membrane	The electrons are then transported to cytochrome c, which transfers them to Complex IV.
Complex IV	Integral membrane protein complex	Complex IV is composed of cytochrome proteins c, a, and a3. This complex contains two heme groups and three copper ions. Complex IV uses the energy of the electrons obtained from complex III to pump another hydrogen ion to the intermembrane space. Complex IV also aids in the splitting of molecular oxygen (O2), so that it picks up two electrons (plus two hydrogen ions from the surrounding medium) to make water. The removal of the hydrogen ions from the system contributes to the ion gradient used in the process of chemiosmosis.

11.4.2 Chemiosmosis

As the ETC works, the concentration gradient of H^+ ions across the membrane establishes an **electrochemical gradient**. The hydrogen ions' positive charge comprises the electrical part of the gradient, and their aggregation on one side of the membrane comprises the chemical part of the gradient. Driven by their electrochemical gradient, the hydrogen ions tend to move back across the inner membrane into the matrix. They pass through an integral membrane protein called **ATP synthase** (**Figure 11.9**). This complex protein acts as a tiny turbine, spun by the force of the hydrogen ions diffusing through it. The turning of this molecular machine creates mechanical energy, which facilitates the addition of a phosphate to ADP, forming ATP.

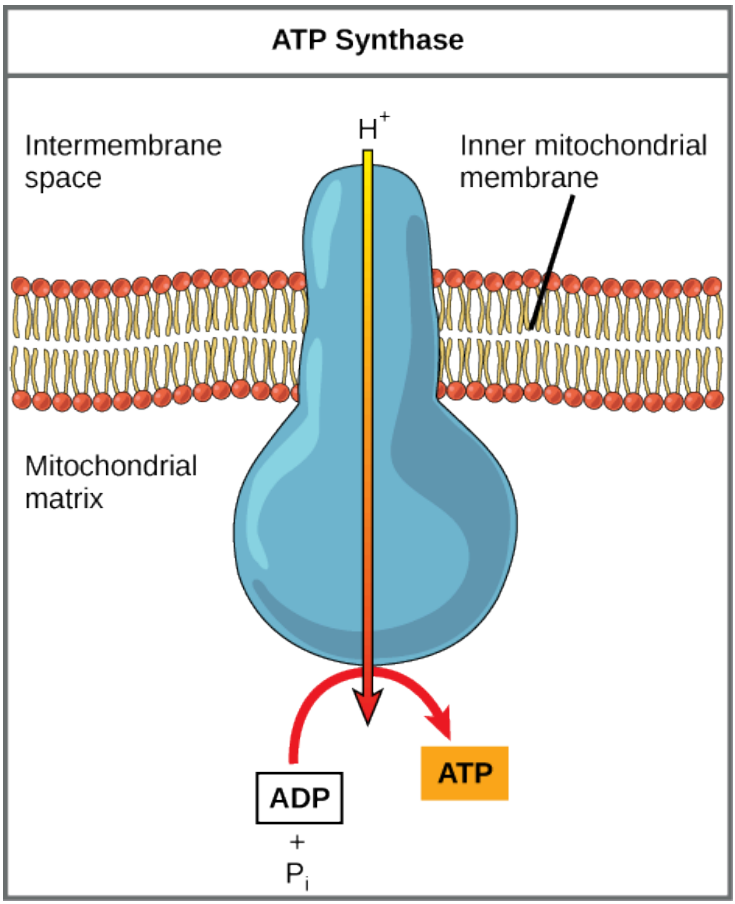


Figure 11.9 ATP synthase is a complex, molecular machine that uses a proton (H^+) gradient to form ATP from ADP and inorganic phosphate (P_i). (Credit: modification of work by Klaus Hoffmeier)

Chemiosmosis (**Figure 11.7**) is used to generate 90% of the ATP made during aerobic glucose catabolism. The overall result of these reactions is the production of ATP from the

energy of the electrons that were originally part of a glucose molecule.

Recall that the production of ATP using the process of chemiosmosis in mitochondria is called oxidative phosphorylation. The ETC is the only part of glucose metabolism that directly uses atmospheric oxygen. Oxygen continuously diffuses into plant and bacterial cells; in animals, it enters the body through the respiratory and circulatory systems, from which it diffuses into cells. In mitochondria, oxygen gas is the terminal electron acceptor for the ETC. If oxygen is not present, the ETC becomes backed up and eventually shuts down. With no ETC to accept electrons from NADH and $FADH_2$, these electron carriers remain reduced and are not available to accept additional electrons during pyruvate oxidation and the citric acid cycle. Therefore, these pathways also shut down. Most cells die in the absence of oxygen for this reason.

Concept Check

Dinitrophenol (DNP) is an uncoupler that makes the inner mitochondrial membrane leaky to protons. It was used until 1938 as a weight-loss drug. What

effect would you expect DNP to have on the change in pH across the inner mitochondrial membrane? Why do you think this might be an effective weight-loss drug? DNP was removed from the FDA's approved list in 1938 after people died from high fevers. Why do you think DNP would cause people to overheat?

Concept Check

The poison cyanide inhibits cytochrome c oxidase, a component of the ETC. 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?

11.4.3 ATP Yield

The number of ATP molecules generated from the

catabolism of glucose varies. For example, the number of hydrogen ions that the electron transport chain complexes can pump through the membrane varies between species. Overall, in living systems, these pathways of glucose catabolism extract about 34% of the energy gained in glucose. The maximum yield of ATP by oxidative phosphorylation under ideal conditions is approximately 34 ATP per glucose molecule. Since four ATP are made by substrate-level phosphorylation, the total maximum yield is 38 ATP per glucose molecule.

11.5 | Metabolism without Oxygen

Learning Objectives

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

- Discuss why fermentation is necessary in the absence of oxygen.
- Compare the ATP yield of aerobic and anaerobic metabolism.

- Describe the type of fermentation that readily occurs in animal cells and the conditions that initiate that fermentation.

In aerobic respiration, ATP is produced by the electron transport chain using the energy of high-energy electrons carried by NADH or $FADH_2$, and the final electron acceptor is an oxygen molecule. However, in some organisms and under some conditions, ATP can be produced in the absence of oxygen.

11.5.1 Anaerobic Cellular Respiration

In some organisms, molecules other than oxygen are used as the final electron acceptor. If an inorganic molecule is used as the final electron acceptor, the process is called **anaerobic cellular respiration**. Certain prokaryotes use anaerobic respiration to produce ATP. For example, the group of Archaea called methanogens uses carbon dioxide to oxidize NADH, producing methane. These microorganisms are found in soil and in the digestive tracts of ruminants, such as cows and sheep. Similarly, sulfur bacteria use sulfate to

regenerate NAD+ from NADH, producing hydrogen sulfide (**Figure 11.10**).

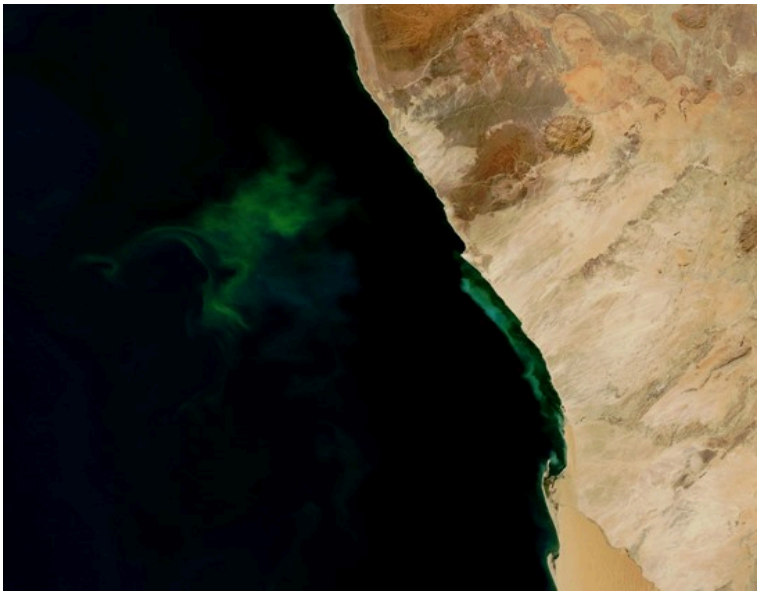


Figure 11.10 The green color seen in these coastal waters is from an eruption of hydrogen sulfide-producing bacteria. These anaerobic, sulfate-reducing bacteria release hydrogen sulfide gas. (Credit: modification of work by NASA/Jeff Schmaltz, MODIS Land Rapid Response Team at NASA GSFC, Visible Earth Catalog of NASA images)

11.5.2 Lactic Acid Fermentation

If neither aerobic nor anaerobic respiration occur, NADH accumulates in cells. If a cell does not have oxidized electron carriers, it cannot break glucose down because there is nowhere to put the electrons. In other words, NADH must

be re-oxidized to NAD^+ for reuse as an electron carrier for glycolysis to continue. Some living systems use an organic molecule as the final electron acceptor, in a process referred to as **fermentation**.

The fermentation method used by animals and certain bacteria, like those that make yogurt, is **lactic acid fermentation** (**Figure 11.11**). This type of fermentation is used routinely in mammalian red blood cells and in skeletal muscle cells that have an insufficient oxygen supply to allow aerobic respiration to continue (that is, in muscles used to the point of fatigue). The enzyme used in this reaction is lactate dehydrogenase (LDH). The reaction can proceed in either direction, but the reaction from left to right is inhibited by acidic conditions. Lactic acid accumulation was once believed to cause muscle stiffness, fatigue, and soreness, although more-recent research disputes this hypothesis. In muscles, lactic acid accumulation must be removed by the blood circulation and the lactate brought to the liver for further metabolism. Once the lactic acid has been removed from the muscle and circulated to the liver, it can be reconverted into pyruvic acid and further catabolized for energy.

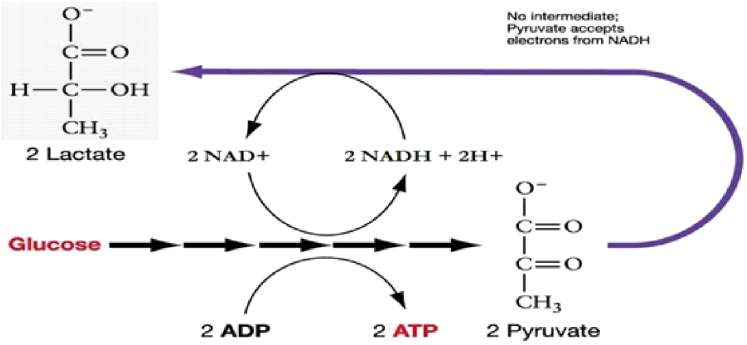


Figure 11.11 Lactic acid fermentation occurs in muscle cells that have run out of oxygen. In this pathway, electrons from NADH are transferred to pyruvate to make lactic acid. This regenerates NAD+ so that glycolysis can continue to make two ATP per glucose.

11.5.3 Alcohol Fermentation

Another familiar fermentation process is the two-step process of **alcohol fermentation** that produces the alcohol ethanol (**Figure 11.12**). The first step releases CO_2 gas and acetaldehyde. The second step oxidizes NADH to NAD+ and reduces acetaldehyde to ethanol.

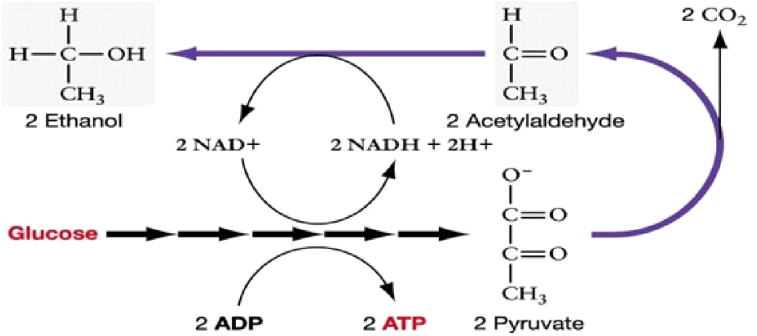


Figure 11.12 Alcohol fermentation is a common process that many organisms use in the absence of oxygen.

The first reaction is catalyzed by pyruvate decarboxylase with a coenzyme derived from vitamin B1 (thiamine). A carboxyl group is removed from pyruvate, releasing CO_2. The loss of carbon dioxide reduces the size of the molecule by one carbon, making acetaldehyde. The second reaction is catalyzed by alcohol dehydrogenase to oxidize NADH to NAD+ and reduce acetaldehyde to ethanol. The fermentation of pyruvic acid by yeast produces the ethanol found in alcoholic beverages. Ethanol is toxic to yeast, limiting the alcohol levels of fermented beverages. The tolerance of yeast is variable, ranging from about 5 percent to 21 percent, depending on the yeast strain and environmental conditions.

11.5.4 Other Types of Fermentation

Other fermentation methods occur in some types of bacteria. Many prokaryotes are "facultatively" anaerobic, meaning that they can switch between aerobic respiration and fermentation, depending on the availability of oxygen. Certain prokaryotes, like Clostridia, are obligate anaerobes, which live and grow in the absence of molecular oxygen. Oxygen is poisonous to these microorganisms and kills them on exposure. It should be noted that all forms of fermentation, except lactic acid fermentation, produce gas. The production of particular types of gas is used as an indicator of the fermentation of specific carbohydrates, which plays a role in the laboratory identification of the bacteria.

Regardless of the type of fermentation, the purpose is always the same: to ensure an adequate supply of NAD^+ so that glycolysis can continue and ATP can be harvested from the breakdown of glucose. Note that fermentation does not capture any energy; in fact, it wastes energy by returning the electrons held by NADH to pyruvate. The only energy produced in fermenting cells is the two ATP created during glycolysis. However, without fermentation to produce NAD^+, glycolysis could not continue and the cell would die.

11.6 | Metabolism of Other

Nutrients

Learning Objectives

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

- Discuss the ways in which carbohydrate metabolic pathways, glycolysis, and the citric acid cycle interrelate with protein and lipid metabolic pathways.
- Explain why metabolic pathways are not considered closed systems.

You have learned about the catabolism of glucose, which provides energy to living cells. However, living things consume more than glucose for food. How does a turkey sandwich end up as ATP in your cells? This happens because all of the catabolic pathways for carbohydrates, proteins, and lipids eventually connect into glycolysis and the citric acid cycle pathways (**Figure 11.13**). Metabolic pathways should not be thought about in a linear sense with a definite start and finish. It is best to think of them as being

interconnected—that is, substances enter from other pathways, and intermediates leave for other pathways. These pathways are not closed systems. Many of the substrates, intermediates, and products in a particular pathway are reactants in other pathways.

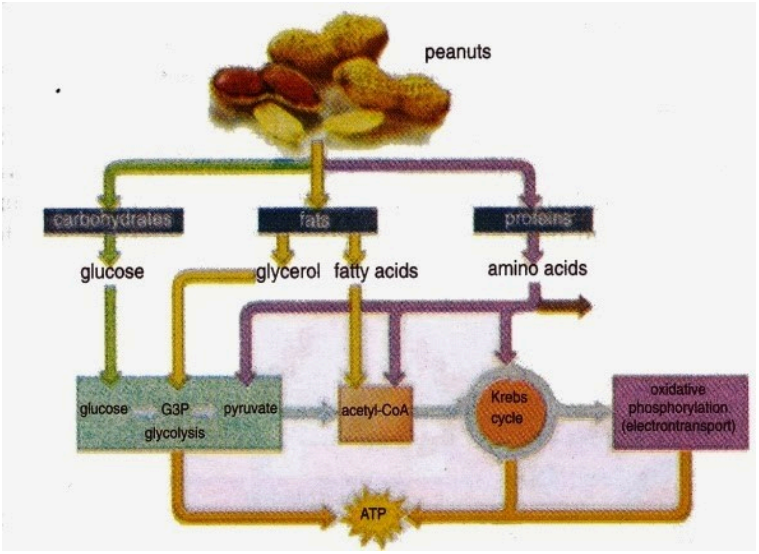


Figure 11.13 Peanuts are a source of protein, fat, and carbohydrates, which are absorbed by our intestinal tract. As these complex molecules are catabolized, the resulting less complex molecules may then enter the glycolysis, form acetyl-CoA, or enter the citric acid cycle as intermediates. The products then proceed to the electron transport chain and chemiosmosis within the mitochondria to yield ATP.

11.6.1 Metabolism of Other

Sugars

Glycogen, a polymer of glucose, is an energy storage molecule in animals. When there is adequate ATP present, excess glucose is stored as glycogen in liver and muscle cells. If blood sugar levels drop, glycogen will be hydrolyzed into glucose monomers (G1P) and converted into G6P, which enters glycolysis. The presence of glycogen as a source of glucose allows ATP to be produced for a longer time during exercise.

Sucrose is a disaccharide with a molecule of glucose and a molecule of fructose bonded together with a glycosidic linkage. Fructose is one of three dietary monosaccharides, along with glucose and galactose, which are absorbed directly into the bloodstream during digestion. The catabolism of both fructose and galactose produces the same number of ATP molecules as glucose.

11.6.2 Metabolism of Proteins

Proteins are hydrolyzed by a variety of enzymes in cells. Most of the time, the amino acids are recycled to synthesize new proteins. If there are excess amino acids, however, or if the body is in a state of starvation, some amino acids will be shunted into the pathways of glucose catabolism (**Figure 11.14**). Each amino acid must have its amino group removed prior to entry into these pathways. The amino group is converted into ammonia. In mammals, the liver synthesizes

urea from two ammonia molecules and a carbon dioxide molecule. Thus, urea is the principal waste product in mammals produced from the nitrogen originating in amino acids, and it leaves the body in urine.

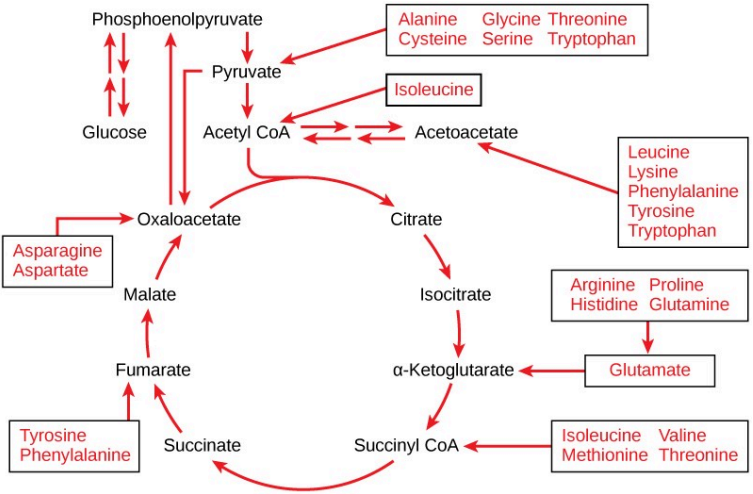


Figure 11.14 The carbon skeletons of certain amino acids (indicated in boxes) derived from proteins can feed into the citric acid cycle. (Credit: modification of work by Mikael Häggström)

11.6.3 Metabolism of Lipids

The lipids that are connected to the glucose pathways are cholesterol and triglycerides. Cholesterol is a lipid that contributes to cell membrane flexibility and is a precursor of steroid hormones. The synthesis of cholesterol starts with

acetyl groups and proceeds in only one direction. The process cannot be reversed.

Triglycerides are a form of long-term energy storage in animals. Triglycerides are made of glycerol and three fatty acids. Animals can make most of the fatty acids they need. Triglycerides can be both made and broken down through parts of the glucose catabolism pathways. Glycerol can be phosphorylated to glycerol-3-phosphate, which continues through glycolysis. Fatty acids are catabolized in a process called beta-oxidation that takes place in the matrix of the mitochondria and converts their fatty acid chains into two carbon units of acetyl groups. The acetyl groups are picked up by CoA to form acetyl CoA that proceeds into the citric acid cycle.

11.7 | Regulation of Cellular Respiration

Learning Objectives

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

- Describe how feedback inhibition would affect the production of an intermediate or product in a pathway.
- Identify the mechanism that controls the rate of the transport of electrons through the electron transport chain.

Cellular respiration must be regulated in order to provide balanced amounts of ATP. The cell also must generate a number of intermediate compounds that are used in the anabolism and catabolism of macromolecules. Without controls, metabolic reactions would quickly come to a standstill as the forward and backward reactions reach a state of equilibrium. At this point, resources would be used inappropriately. A cell does not need the maximum amount of ATP that it can make all the time. At times, the cell needs to shunt some of the intermediates to pathways for amino acid, protein, glycogen, lipid, and nucleic acid production. In short, the cell needs to control its metabolism.

11.7.1 Regulatory Mechanisms

A variety of mechanisms is used to control cellular respiration. Some type of control exists at each stage of glucose

metabolism. Access of glucose to the cell can be regulated using the **GLUT** (GLUcose Transport) **proteins** that transport glucose (**Figure 11.15**). Different forms of the GLUT protein control passage of glucose into the cells of specific tissues.

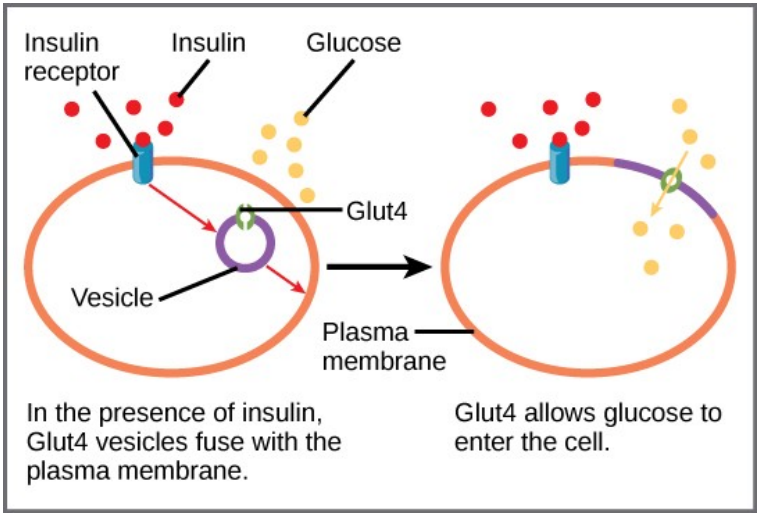


Figure 11.16 GLUT4 is a glucose transporter that is stored in vesicles. A cascade of events that occurs upon insulin binding to a receptor in the plasma membrane causes GLUT4-containing vesicles to fuse with the plasma membrane so that glucose may be transported into the cell.

Some reactions are controlled by having two different enzymes—one each for the two directions of a reversible reaction. Reactions that are catalyzed by only one enzyme can go to equilibrium, stalling the reaction. In contrast, if

two different enzymes (each specific for a given direction) are necessary for a reversible reaction, the opportunity to control the rate of the reaction increases, and equilibrium is not reached. A number of enzymes involved in each of the pathways—usually the first enzyme of the pathway—are controlled by allosteric regulation. The molecules most commonly used in this capacity are the nucleotides ATP, ADP, AMP, NAD+, and NADH. These allosteric regulators may increase or decrease enzyme activity, depending on the prevailing conditions.

Regulation of Glycolysis

Step 1: Hexokinase

The control of glycolysis begins with the first enzyme in the pathway, hexokinase. This enzyme catalyzes the phosphorylation of glucose, which helps to prepare the compound for cleavage in a later step. The presence of the negatively charged phosphate in the molecule also prevents the sugar from leaving the cell. When hexokinase is inhibited, glucose diffuses out of the cell and does not become a substrate for the respiration pathways in that tissue. The product of the hexokinase reaction is glucose-6-phosphate, which accumulates when a later enzyme, phosphofructokinase, is inhibited.

Step 3: Phosphofructokinase

Phosphofructokinase is the main enzyme controlled in glycolysis. High levels of ATP, citrate, or a lower pH decrease the enzyme's activity. An increase in citrate concentration can occur because of a blockage in the citric acid cycle.

Step 10: Pyruvate kinase

The last step in glycolysis is catalyzed by pyruvate kinase. The pyruvate produced can proceed to be catabolized or converted into the amino acid alanine. If no more energy is needed and alanine is in adequate supply, the enzyme is inhibited. The enzyme's activity is increased when fructose-1,6-bisphosphate levels increase. (Recall that fructose-1,6-bisphosphate is an intermediate in the first half of glycolysis.) The regulation of pyruvate kinase involves phosphorylation by a kinase (pyruvate kinase kinase), resulting in a less-active enzyme. Dephosphorylation by a phosphatase reactivates it. Pyruvate kinase is also regulated by ATP (a negative allosteric effect).

If more energy is needed, more pyruvate will be converted into acetyl CoA through the action of pyruvate dehydrogenase. If either acetyl groups or NADH accumulate, there is less need for the reaction and the rate decreases. Pyruvate dehydrogenase is also regulated by phosphorylation: A kinase phosphorylates it to form an inactive enzyme, and a phosphatase reactivates it. The kinase and the phosphatase are also regulated.

Regulation of the Citric Acid Cycle

The citric acid cycle is controlled through the enzymes that catalyze the reactions that make the first two molecules of NADH. These enzymes are isocitrate dehydrogenase and α-ketoglutarate dehydrogenase. When adequate ATP and NADH levels are available, the rates of these reactions decrease. When more ATP is needed, as reflected in rising ADP levels, the rate increases. α-Ketoglutarate dehydrogenase will also be affected by the levels of succinyl CoA—a subsequent intermediate in the cycle—causing a decrease in activity. A decrease in the rate of operation of the pathway at this point is not necessarily negative, as the increased levels of the α-ketoglutarate not used by the citric acid cycle can be used by the cell for amino acid (glutamate) synthesis.

Regulation of the Electron Transport Chain

Specific enzymes of the electron transport chain are unaffected by feedback inhibition, but the rate of electron transport through the pathway is affected by the levels of ADP and ATP. Greater ATP consumption by a cell is indicated by a buildup of ADP. As ATP usage decreases, the concentration of ADP decreases and ATP begins to build up in the cell. This change is the relative concentration of ADP to ATP triggers the cell to slow down the electron transport chain.

Table 11.5 Summary of feedback controls in cellular respiration.

Pathway	Enzyme	Elevated levels of effector
Glycolysis	hexokinase	glucose-6-phosphate
	phosphofructokinase	low-energy charge (ATP, AMP), fructose-6-phosphate via fructose-2,6-bisphosphate
	phosphofructokinase	high-energy charge (ATP, AMP), citra pH
	pyruvate kinase	fructose-1,6-bisphosphate
	pyruvate kinase	high-energy charge (ATP, AMP), alan
Pyruvate oxidation	pyruvate dehydrogenase	ADP, pyruvate
	pyruvate dehydrogenase	acetyl CoA, ATP, NADH
Citric acid cycle	isocitrate dehydrogenase	ADP
	isocitrate dehydrogenase	ATP, NADH

	α-ketoglutarate dehydrogenase	Calcium ions, ADP
	α-ketoglutarate dehydrogenase	ATP, NADH, succinyl CoA
Oxidative phosphorylation		ADP
		ATP

CHAPTER 12. PHOTOSYNTHESIS

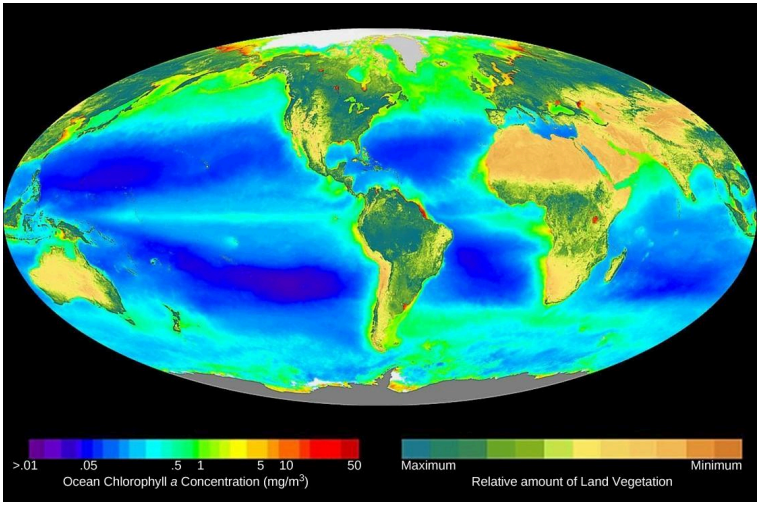


Figure 12.1 This world map shows Earth's distribution of photosynthesis as seen via chlorophyll a concentrations. On land, this is evident via terrestrial plants, and in oceanic zones, via phytoplankton. (Credit: modification of work by SeaWiFS Project, NASA/Goddard Space Flight Center and ORBIMAGE)

Chapter Outline

- 12.1 Overview of Photosynthesis
- 12.2 The Light-Dependent Reactions of Photosynthesis
- 12.3 Using Light Energy to Make Organic Molecules

Introduction

The processes in all organisms—from bacteria to humans—require energy. To get this energy, many organisms access stored energy by eating, that is, by ingesting other organisms. But where does the stored energy in food originate? All of this energy can be traced back to photosynthesis.

12.1 | Overview of Photosynthesis

Learning Objectives

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

- Explain the relevance of photosynthesis to other living things.
- Describe the main structures involved in photosynthesis.
- Identify the substrates and products of photosynthesis.
- Summarize the process of photosynthesis.

Photosynthesis is essential to all life on earth. It is the only biological process that can capture energy that originates in outer space (sunlight) and convert it into chemical compounds (carbohydrates) that every organism uses to power its metabolism. In brief, the energy of sunlight is used to energize electrons, which are then stored in the covalent bonds of sugar molecules. The energy extracted today by the burning of coal and petroleum products represents sunlight energy captured and stored by photosynthesis almost 200 million years ago.

Figure 12.2 Photoautotrophs including (a) plants, (b) algae, and (c) cyanobacteria synthesize their organic compounds via photosynthesis using sunlight as an energy source. Cyanobacteria and planktonic algae can grow over enormous areas in water, at times completely covering the surface. (Credit a: modification of work by Steve Hillebrand, U.S. Fish and Wildlife Service; b: modification of work by "eutrophication&hypoxia"/Flickr; c: modification of work by NASA

Plants, algae, and a group of bacteria called cyanobacteria are the only organisms capable of performing photosynthesis (**Figure 12.2**). Because they use light to manufacture their own food, they are called **autotrophs** ("self-feeders"). Other organisms, such as animals, fungi, and most other bacteria, are termed **heterotrophs** ("other feeders"), because they must rely on the sugars produced by photosynthetic organisms for their energy needs.

The importance of photosynthesis is not just that it can capture sunlight's energy. A lizard sunning itself on a cold day can use the sun's energy to warm up. Photosynthesis is vital because it evolved as a way to store the energy in solar radiation as high-energy electrons in the carbon-carbon bonds of carbohydrate molecules. Those carbohydrates are the energy

source that heterotrophs use to power the synthesis of ATP via cellular respiration. Therefore, photosynthesis powers Earth's ecosystems. When a top predator, such as a wolf, preys on a deer, the wolf is at the end of an energy path that went from nuclear reactions on the surface of the sun, to light, to photosynthesis, to vegetation, to deer, and finally to wolf.

12.1.1 Main Structures and Summary of Photosynthesis

Photosynthesis is a multi-step process that requires sunlight, carbon dioxide, and water as substrates (**Figure 12.3**). After the process is complete, it releases oxygen and produces glyceraldehyde-3-phosphate (GA3P), simple carbohydrate molecules (which are high in energy) that can subsequently be converted into glucose, sucrose, or any of dozens of other sugar molecules. These sugar molecules contain energy and the energized carbon that all living things need to survive.

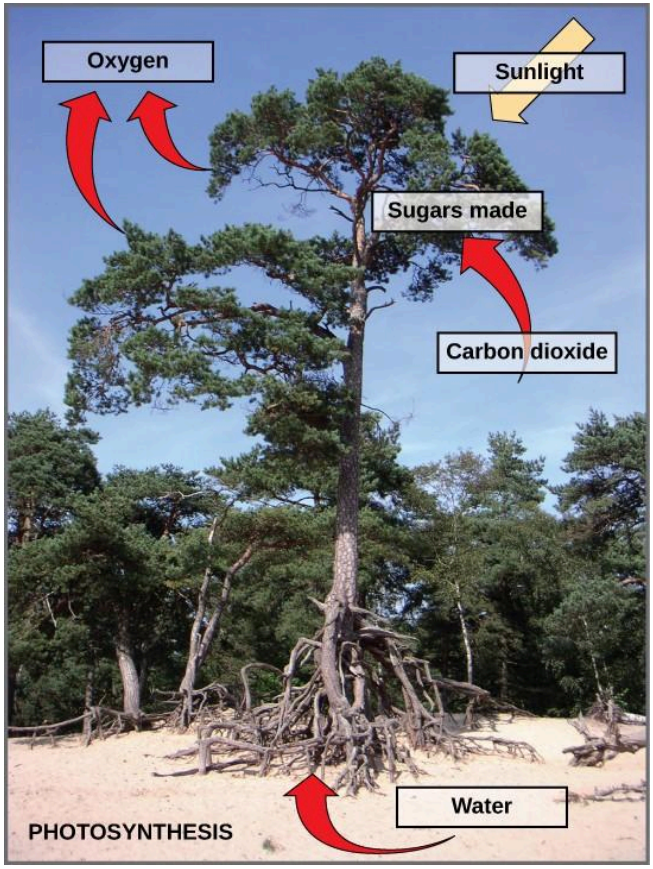


Figure 12.3 Photosynthesis uses solar energy, carbon dioxide, and water to produce energy-storing carbohydrates. Oxygen is generated as a waste product of photosynthesis.

The following is the chemical equation for photosynthesis:

Carbon dioxide	+	Water	SUNLIGHT →	Sugar	+	Oxygen
$6CO_2$		$6H_2O$		$C_6H_{12}O_6$		$6O_2$

Although the equation looks simple, the many steps that take place during photosynthesis are actually quite complex. Before learning the details of how photoautotrophs use sunlight to synthesize food, it is important to become familiar with the structures involved.

In plants, photosynthesis generally takes place in leaves, which consist of several layers of cells. Photosynthesis occurs in a middle layer called the **mesophyll**. The gas exchange of carbon dioxide and oxygen occurs through small, regulated openings called **stomata** (singular: stoma), which also play roles in the regulation of gas exchange and water balance. The stomata are typically located on the underside of the leaf, which helps to minimize water loss. Each stoma is flanked by guard cells that regulate the opening and closing of the stomata by swelling or shrinking in response to osmotic changes (**Figure 12.4**).

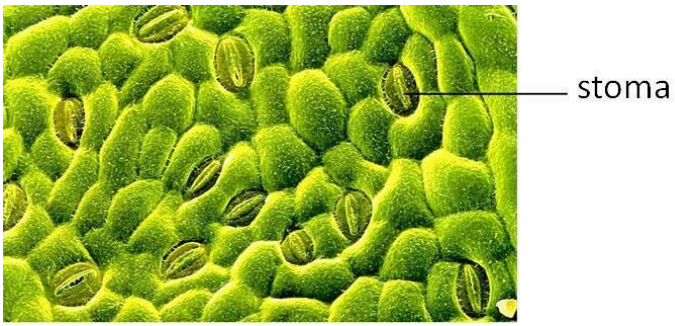


Figure 12.4 Stomata on the bottom of a rose leaf. Each stoma is formed by guard cells that swell to close the stoma. Water, oxygen, and carbon dioxide pass through open stomata.

Concept Check

On a hot, dry day, plants close their stomata to conserve water. What impact will this have on photosynthesis?

In all autotrophic eukaryotes, photosynthesis takes place inside an organelle called a **chloroplast**. For plants, chloroplast-containing cells exist in the mesophyll. Chloroplasts have a double membrane envelope composed of

an outer membrane and an inner membrane. Within the chloroplast are stacked, disc-shaped structures called **thylakoids**. Embedded in the thylakoid membrane is chlorophyll, a pigment (molecule that absorbs light) responsible for the initial interaction between light and plant material, and numerous proteins that make up the electron transport chain. The thylakoid membrane encloses an internal space called the **thylakoid lumen**. As shown in **Figure 12.5**, a stack of thylakoids is called a **granum**, and the liquid-filled space surrounding the granum is called **stroma** (not to be confused with stoma).

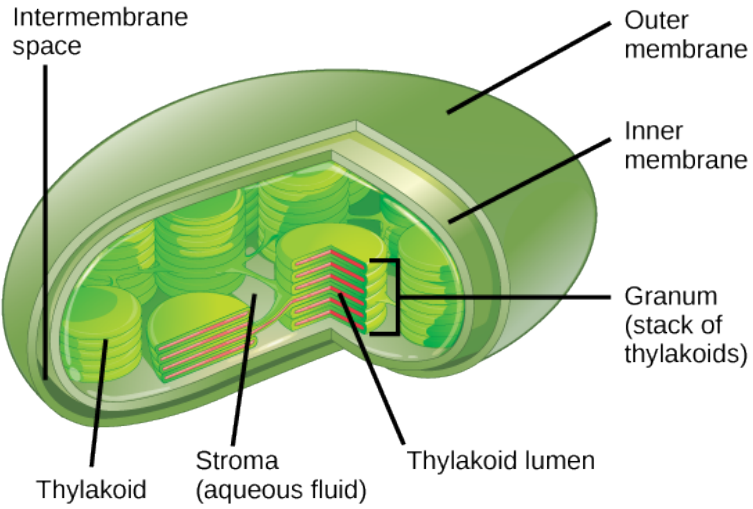


Figure 12.5 Photosynthesis takes place in chloroplasts, which have an outer membrane and an inner membrane. Stacks of thylakoids called grana form a third membrane layer.

12.1.2 The Two Halves of Photosynthesis

Photosynthesis takes place in two sequential stages: the light-dependent reactions and the light independent-reactions. In the **light-dependent reactions**, energy from sunlight is absorbed by chlorophyll and that energy is converted into stored chemical energy. In the **light-independent reactions**, the chemical energy harvested during the light-dependent reactions drives the assembly of sugar molecules from carbon dioxide.

Although the light-independent reactions do not use light directly as a reactant, they require the products of the light-dependent reactions. The light-dependent reactions utilize molecules called energy carriers to temporarily store energy. The energy carriers that move energy from light-dependent reactions to light-independent reactions can be thought of as "full" because they are rich in energy. After the energy is released, the "empty" energy carriers return to the light-dependent reaction to obtain more energy. **Figure 12.6** illustrates the components inside the chloroplast where the light-dependent and light-independent reactions take place.

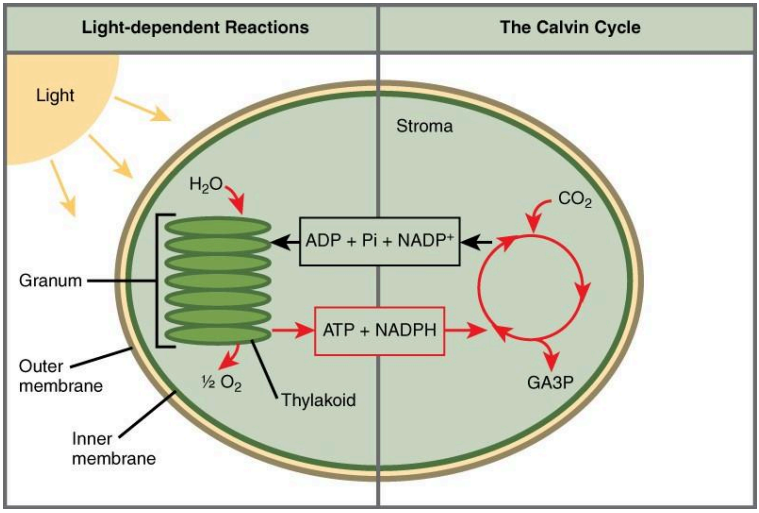


Figure 12.6 Photosynthesis takes place in two stages: light dependent reactions and the Calvin cycle. Light- dependent reactions, which take place in the thylakoid membrane, use light energy to make ATP and NADPH. The Calvin cycle, which takes place in the stroma, uses energy derived from these compounds to make GA3P from CO_2.

EVERYDAY CONNECTION

Photosynthesis at the Grocery Store

Major grocery stores in the United States are organized into departments, such as

dairy, meats, produce, bread, cereals, and so forth. Each aisle (**Figure 12.7**) contains hundreds, if not thousands, of different products for customers to buy and consume.

Although there is a large variety, each item links back to photosynthesis. Meats and dairy link, because the animals were fed plant-based foods. The breads, cereals, and pastas come largely from starchy grains, which are the seeds of photosynthesis-dependent plants. What about desserts and drinks? All of these products contain sugar—sucrose is a plant product, a disaccharide, a carbohydrate molecule, which is built directly from photosynthesis. Many items are less obviously derived from plants: For instance, paper goods are generally plant products, and many plastics are derived from algae. Virtually every spice and flavoring in the spice aisle was produced by a plant as a leaf, root, bark, flower, fruit, or stem. Ultimately,

photosynthesis connects to every meal and every food a person consumes.

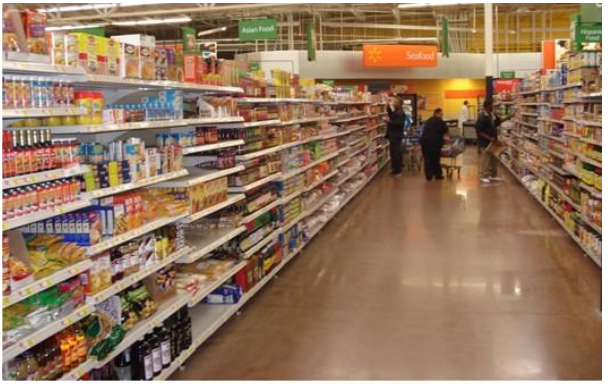


Figure 12.7 Foods that humans consume originate from photosynthesis. (Credit: Associação Brasileira de Supermercados)

12.2 | The Light-Dependent Reactions of Photosynthesis

Learning Objectives

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

- Explain how plants absorb energy from sunlight.
- Describe short and long wavelengths of light.
- Describe how and where photosynthesis takes place within a plant.

How can light be used to make food? Like all other forms of kinetic energy, light can travel, change form, and be harnessed to do work. In photosynthesis, light energy is converted into chemical energy, which photoautotrophs use to build carbohydrate molecules (**Figure 12.8**). However, autotrophs only use a few specific components of sunlight.

Figure 12.8 Photoautotrophs can capture light energy from the sun, converting it into the chemical energy used to build food molecules. (Credit: Gerry Atwell)

12.2.1 What Is Light Energy?

The sun emits an enormous amount of electromagnetic radiation, or solar energy. Solar energy is composed of tiny, mass-less packets of energy called **photons**. Humans can see only the tiny fraction of this energy that is capable of stimulating photoreceptor cells in our retinas. This portion is referred to as "visible light."

The manner in which photons travel is described as waves.

Scientists can determine the amount of energy in a particular photon by measuring its **wavelength**, the distance between consecutive points of a wave. A single wave is measured from two consecutive points, such as from crest to crest or from trough to trough (**Figure 12.9**).

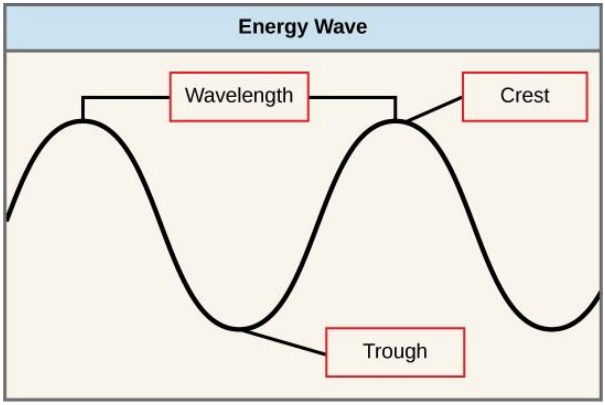


Figure 12.9 The wavelength of a single wave is the distance between two consecutive points of similar position (two crests or two troughs) along the wave.

Visible light is only one of many types of electromagnetic radiation emitted from the sun. Scientists describe the various types of radiant energy from the sun on an electromagnetic spectrum. The **electromagnetic spectrum** is the range of all possible frequencies of radiation, including X-rays, ultraviolet (uv) rays, microwaves, and radio waves (**Figure 12.10**).

Each type of electromagnetic radiation has photons that

travel at a particular wavelength. Since the height of the waves is the same, the wavelength relates to how quickly the photons are vibrating up and down as they travel a given distance. Therefore, the longer the wavelength, the less energy is carried by the photons. Photons traveling in short, tight waves carry the most energy. The higher-energy waves can penetrate tissues and damage cells and DNA, explaining why both X- rays and UV rays can be harmful to living organisms.

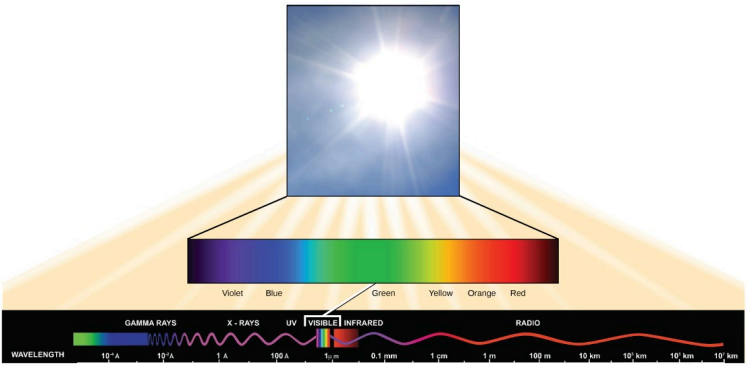


Figure 12.10 The sun emits energy in the form of electromagnetic radiation. This radiation exists at different wavelengths, each of which has its own characteristic energy. All electromagnetic radiation, including visible light, is characterized by its wavelength.

12.2.2 Absorption of Light

Pigments are molecules that absorb a portion of visible light. Interestingly, the portion of visible light that is NOT absorbed

by a pigment is reflected to our eyes. Therefore, the color we perceive represents the wavelengths of visible light that are NOT absorbed by a particular pigment. For example, pigments in a red-skinned apple absorb yellow, green and blue light and reflect red light (**Figure 12.11**).

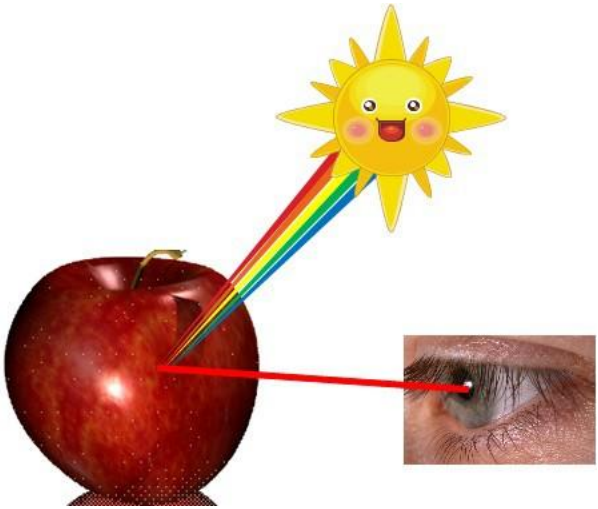


Figure 12.11 Pigments absorb some photons of visible light and reflect the others. The photons that are reflected are the ones that we "see", when they hit our retina.

Organic pigments, such as those used for photosynthesis, have a narrow range of energy levels that they can absorb. In fact, they can only absorb the exact amount of energy needed to raise an electron to an excited state. Recall that electrons occupy discrete energy levels in atoms. They can only absorb a

photon if it has the exact amount of energy to raise it by one or two energy levels. Photons with different amounts of energy cannot be absorbed (**Figure 12.12**) Therefore, pigments in our retinas can only "see" (absorb) wavelenghts of 700 nm to 400 nm , which is therefore called visible light. Plant pigment molecules also only absorb light in the wavelength range of 700 nm to 400 nm.

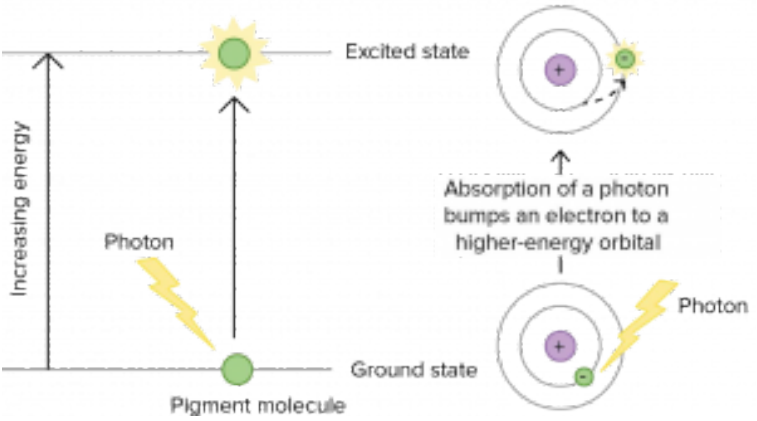


Figure 12.12 Pigments can only absorb photons with the amount of energy needed to bump an electron to a higher energy orbital. Other photons are reflected. (Credit: image modified from "Bis2A 06.3 Photophosphorylation: the light reactions of photosynthesis: Figure 7," by Mitch Singer.)

The visible light portion of the electromagnetic spectrum shows a rainbow of colors, with violet and blue having shorter wavelengths, and therefore higher energy. At the red end of the

spectrum, the wavelengths are longer and have lower energy (**Figure 12.13**).

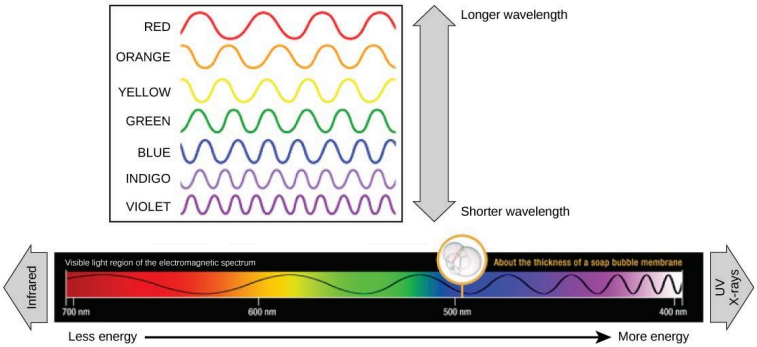


Figure 12.13 The colors of visible light do not carry the same amount of energy. Violet has the shortest wavelength and therefore carries the most energy, whereas red has the longest wavelength and carries the least amount of energy.(credit: modification of work by NASA)

Understanding Pigments

Different kinds of pigments exist, and each has evolved to absorb only certain wavelengths of visible light. Chlorophylls and carotenoids are the two major classes of photosynthetic pigments found in plants and algae. There are five major types of chlorophylls: *a*, *b*, *c* and *d* and a related molecule found in prokaryotes called bacteriochlorophyll. **Chlorophyll *a*** and **chlorophyll *b*** are found in higher plant chloroplasts.

With dozens of different forms, carotenoids are a much larger group of pigments. The carotenoids found in

fruit—such as the red of tomato (lycopene), the yellow of corn seeds (zeaxanthin), or the orange of an orange peel (β-carotene)—are used to attract seed dispersers. **Carotenoids** also serve as photosynthetic pigments that reside in the thylakoid membrane, absorb excess energy that could damage the plant, and safely dissipate that energy as heat.

Each type of pigment can be identified by the specific pattern of visible light it absorbs, which can be visualized on an **absorption spectrum**. **Figure 12.14** shows the absorption spectra for chlorophyll *a*, chlorophyll *b*, and β-carotene. Notice how each pigment has a distinct set of peaks and troughs, revealing a highly specific pattern of absorption. Chlorophyll *a* absorbs wavelengths from either end of the visible spectrum (blue and red), but not green. Because green is reflected or transmitted, chlorophyll appears green. Carotenoids absorb in the short-wavelength blue region, and reflect the longer yellow, red, and orange wavelengths.

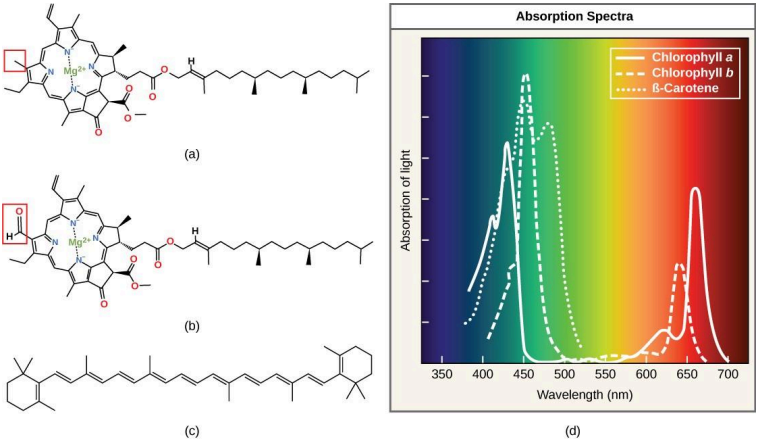


Figure 12.14 (a) Chlorophyll a, (b) chlorophyll b, and (c) β-carotene are organic pigments found in the thylakoid membrane of plants. Chlorophyll a and b, which are identical except for the part indicated in the red box, are responsible for the green color of leaves. β- carotene is responsible for the orange color in carrots. Each pigment has (d) a unique absorbance spectrum.

Many photosynthetic organisms have a mixture of pigments, which allows them to absorb energy from a wider range of wavelengths. For example, 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 rest (**Figure 12.15**).

Figure 12.15 Plants that commonly grow in the shade have adapted to low levels of light by changing the relative concentrations of their chlorophyll pigments. (Credit: Jason Hollinger)

12.2.3 How Light-Dependent Reactions Work

Photosystems

The overall function of light-dependent reactions is to convert solar energy into chemical energy in the form of NADPH and ATP. This chemical energy will fuel the assembly of sugar molecules during the light-independent reactions.

Light energy is converted into chemical energy in **photosystems**. Photosystems are large multiprotein complexes that consist of hundreds of pigment molecules

embedded in a protein matrix **(Figure 12.16)**. Two special pigment molecules are attached to the **reaction center**, while the rest are part of the **antennae complex**.

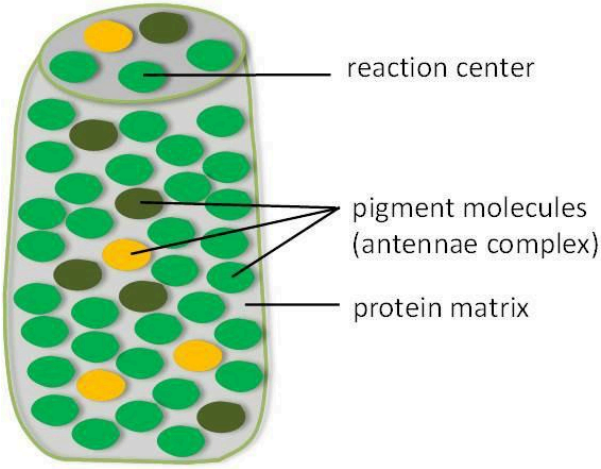


Figure 12.16 A photosystem consists of an antennae complex and a reaction center. Pigments in the light-harvesting complex pass light energy to two special chlorophyll a molecules in the reaction center. The light excites an electron from the chlorophyll a pair, which passes to the primary electron acceptor.

When a photon is absorbed by a pigment molecule in the antennae complex, an electron in that molecule is pushed into an excited state. At this point, the light energy has been captured by a biological molecule but is not yet stored in any useful form. The energy is transferred from chlorophyll to chlorophyll until eventually (after about a millionth of a

second), it is delivered to the reaction center. Up to this point, only energy has been transferred between molecules, not electrons.

The reaction center contains a pair of chlorophyll *a* molecules with a special property. Those two chlorophylls can undergo oxidation upon excitation; they can actually give up an electron. The excited electron is transferred to an electron acceptor and then to an electron transport chain, where its energy is used 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 actually give up an electron. The excited electron is transferred to an electron acceptor and then to an electron transport chain (ETC), where its energy is used to make ATP in much the same way as in the mitochondria (described below). Since energy from sunlight is used to add phosphate groups to ATP, the process is called **photophosphorylation**.

Since the reaction center chlorophyll *a* molecules give up excited electrons, their electrons must be constantly replaced with low energy electrons so that photosynthesis can continue. Two methods have evolved to replace the reaction center electrons. Photosynthetic bacteria use a single photosystem and a process called cyclic photophosphorylation. Plants and algae use two photosystems and a process called non-cyclic photophosphorylation.

Cyclic Photophosphorylation

In the process of **cyclic photophosphorylation**, bacteria reuse the same electron over and over again. When the reaction center pigment is excited, it transfers an excited electron to an ETC. As the electron is passed down the ETC, its energy is used to set up a proton gradient across folds of the plasma membrane. Protons flow back across the membrane through ATP synthase, turning its rotor, and forcing phosphate groups onto ADP to make ATP. At the bottom of the ETC, most of the electron's energy has been spent. The low energy electron is returned to the reaction center, where it can be re-excited by solar energy (**Figure 12.17a**).

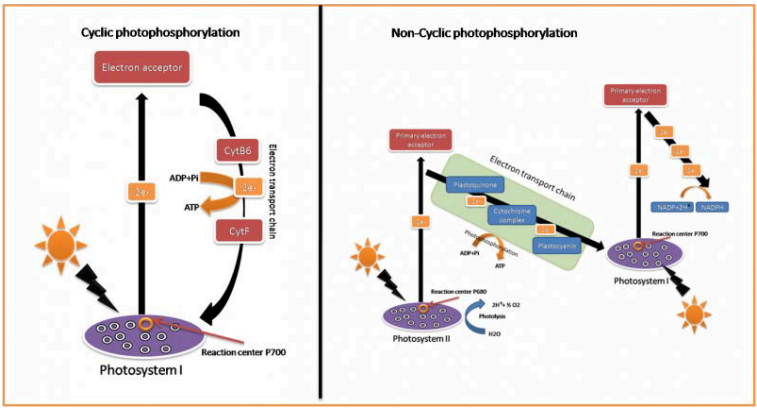


Figure 12.17 Two methods of returning a low energy electron to the reaction center chlorophyll are: (a) cyclic photophosphorylation, during which the same electron is returned after its energy is spent, and (b) non-cyclic photophosphorylation, during which electrons from water replace the electrons from photosystem II, while spent electrons from photosystem II replace the electrons from photosystem I. (Credit: www.biopicture.loxblog.com)

Non-Cyclic Photophosphorylation

In the process of non-**cyclic photophosphorylation**, plants and algae use two photosystems and the same electrons do not return to be reused (**Figure 12.17b**). The two types of photosystems, called photosystem II (PSII) and photosystem I (PSI), are found embedded in the thylakoid membrane of chloroplasts. The two complexes differ on the basis of 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) (**Figure 12.18**).

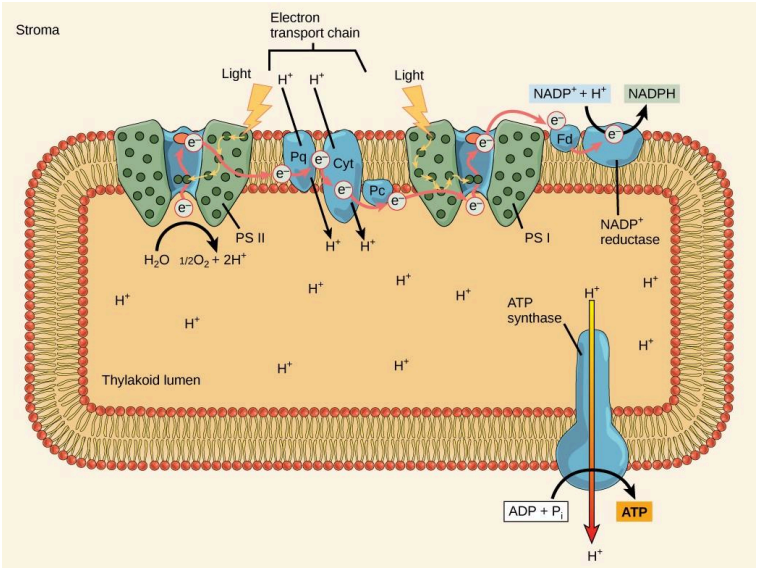


Figure 12.18 In the photosystem II reaction center, energy from sunlight is used to extract electrons from water. The electrons travel through the chloroplast electron transport chain (ETC) to photosystem I, which reduces $NADP^+$ to NADPH. The ETC moves protons from the stroma into the thylakoid lumen. ATP synthase uses the resulting electrochemical gradient to make ATP.

The reaction center of PSII (called **P680**) delivers its high-energy electrons, one at the time, to an electron acceptor and then through the ETC to PSI. P680's missing electron is replaced by extracting a low-energy electron from water. Splitting one H_2O molecule releases two electrons, two hydrogen atoms, and one atom of oxygen. Splitting two molecules is required to form one molecule of diatomic O_2

gas. About 10 percent of the oxygen is used by mitochondria in the leaf to support oxidative phosphorylation. The remainder escapes to the atmosphere where it is used by aerobic organisms to support cellular respiration.

As electrons move through the ETC that resides between PSII and PSI, they lose energy. That energy is used to move hydrogen ions from the stromal side of the membrane to the thylakoid lumen. Those hydrogen ions, plus the ones produced by splitting water, accumulate in the thylakoid lumen and will be used synthesize ATP in a later step. Because the electrons have lost energy prior to their arrival at PSI, they must be re-energized by photons absorbed by the PSI antenna complex. That energy is relayed to the PSI reaction center (called **P700**). P700 is oxidized and sends a high-energy electron to NADP+ to form NADPH. Thus, PSII captures the energy to create proton gradients to make ATP, and PSI captures the energy to reduce NADP+ into NADPH. The two photosystems work in concert, in part, to guarantee that the production of NADPH will roughly equal the production of ATP. Other mechanisms exist to fine tune that ratio to exactly match the chloroplast's constantly changing energy needs.

Generating an Energy Carrier: ATP

As in the intermembrane space of the mitochondria during cellular respiration, the buildup of hydrogen ions inside the thylakoid lumen creates a concentration gradient. The

exergonic movement of hydrogen ions from high concentration (in the thylakoid lumen) to low concentration (in the stroma) is harnessed to create ATP, just as in the mitochondrial electron transport chain of cellular respiration.

To release the energy stored in the electrochemical gradient, hydrogen ions will rush through any opening, similar to water jetting through a hole in a dam. In the thylakoid, that opening is a passage through a specialized protein channel called ATP synthase. The energy released by the hydrogen ion stream allows ATP synthase to attach a third phosphate group to ADP, which forms a molecule of ATP (**Figure 12.18**). The flow of hydrogen ions through ATP synthase is called chemiosmosis because the ions move from an area of high to an area of low concentration through a semi-permeable structure.

Concept Check

What is the initial source of electrons for the chloroplast electron transport chain?

- water
- oxygen

- carbon dioxide
- NADPH

12.3 | Using Light Energy to Make Organic Molecules

Learning Objectives

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

- Describe the Calvin cycle.
- Define carbon fixation.
- Explain how photosynthesis works in the energy cycle of all living organisms

After the energy from the sun is converted into chemical energy and temporarily stored in ATP and NADPH molecules, the cell has the fuel needed to build carbohydrate

molecules for long-term energy storage. The products of the light-dependent reactions, ATP and NADPH, have lifespans in the range of millionths of seconds, whereas the products of the light-independent reactions (carbohydrates and other forms of reduced carbon) can survive for hundreds of millions of years. The carbohydrate molecules made will have a backbone of carbon atoms. Where does the carbon come from? It comes from carbon dioxide, the gas that is a waste product of respiration in microbes, fungi, plants, and animals.

12.3.1 Light-Independent Reactions: The Calvin Cycle

In plants, carbon dioxide (CO_2) enters leaves through stomata and diffuses until it reaches the mesophyll cells. Once in the mesophyll cells, CO2 diffuses into the stroma of the chloroplast—the site of the **light-independent reactions**. These reactions are sometimes called the **Calvin cycle** after the scientist who discovered them (**Figure 12.19**). The light-independent reactions of the Calvin cycle can be organized into three stages: fixation, reduction, and regeneration.

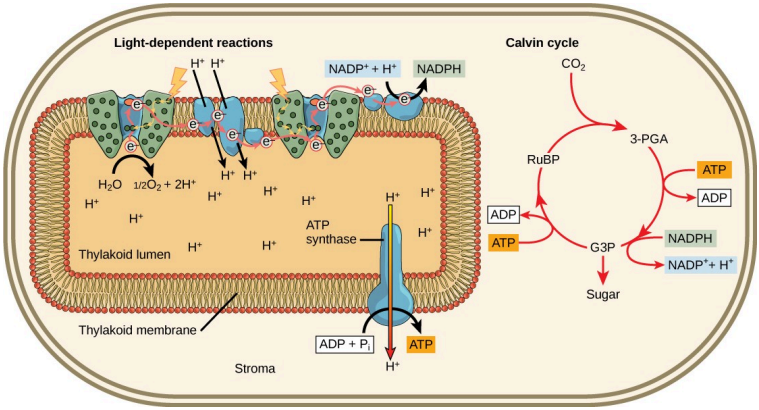


Figure 12.19 Light-dependent reactions use energy from the sun to make ATP and NADPH. These energy-carrying molecules are then used to drive the light-independent reactions, during which carbon is fixed and sugars are made.

Stage 1: Fixation

In the stroma, in addition to CO_2, two other components are present to initiate the light-independent reactions: an enzyme called RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase), and three molecules of RuBP (ribulose bisphosphate). RuBP has five atoms of carbon, flanked by two phosphates (**Figure 12.20**).

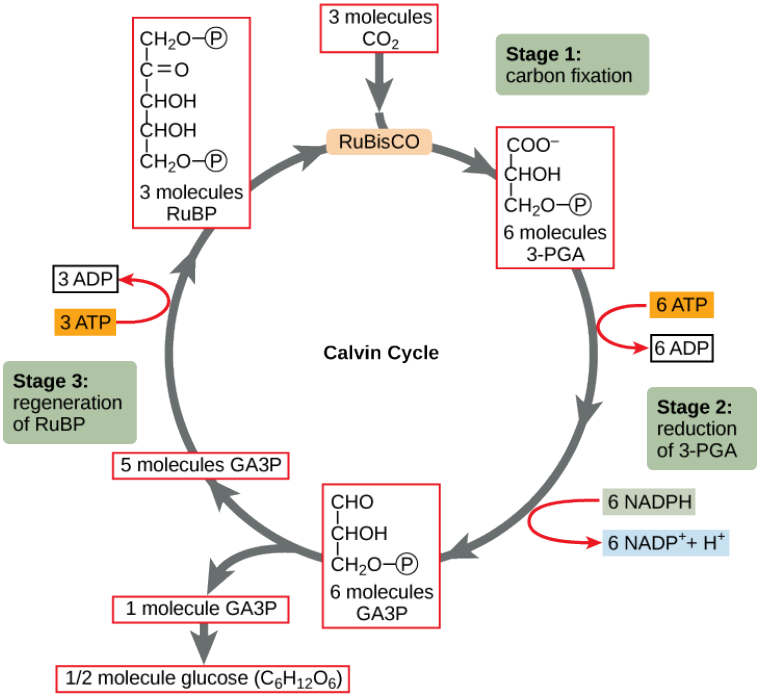


Figure 12.20 The Calvin cycle has three stages. (1) The enzyme RuBisCo fixes carbon dioxide to RuBP and produces two 3-PGA. (2) 3-PGA is reduced to GA3P using electrons from NADPH. (3) RuBP is regenerated so that the cycle can continue. The cycle must be completed three times to produce a single three-carbon GA3P molecule, and six times to produce a six-carbon glucose molecule.

RuBisCO catalyzes a reaction between CO_2 and RuBP. For each CO2 molecule that reacts with one RuBP, two molecules of another compound (3-PGA) form. 3-PGA has three carbons and one phosphate. Each turn of the cycle involves only one RuBP and one carbon dioxide and forms two molecules of 3-PGA. Note that the number of carbon atoms

remains the same, as the atoms form new bonds. This process is called **carbon fixation**, because CO_2 is "fixed" from an inorganic form into organic molecules.

Stage 2: Reduction

ATP and NADPH are used to convert the six molecules of 3-PGA into six molecules of a chemical called glyceraldehyde 3-phosphate (GA3P). This is a reduction reaction because it involves the gain of electrons by 3-PGA. Six molecules each of ATP and NADPH are used.

Stage 3: Regeneration

Interestingly, at this point, only one of the GA3P molecules leaves the Calvin cycle and is sent to the cytoplasm to contribute to the formation of other compounds needed by the plant. Because the GA3P exported from the chloroplast has three carbon atoms, it takes three "turns" of the Calvin cycle to fix enough net carbon to export one GA3P. Since each turn makes two GA3Ps, three turns make six GA3Ps. The remaining five GA3P molecules remain in the cycle and are used to regenerate RuBP, which enables the system to prepare for more CO_2 to be fixed. Three more molecules of ATP are used in these regeneration reactions.

Concept Check

Which of the following statements is true for photosynthesis?

- Oxygen, carbon dioxide, ATP, and NADPH are reactants.
- GA3P and water are products.
- Chlorophyll, water, and carbon dioxide are reactants.
- GA3P and oxygen are products.
- Water, carbon dioxide, ATP, and NADPH are reactants.
- RuBP and oxygen are products.
- In photosynthesis, water and carbon dioxide are reactants. GA3P and oxygen are products.

Photosynthesis

Photosynthesis in desert plants has evolved adaptations that conserve water. In the harsh dry heat, every drop of water must be used to survive. Because stomata must open to allow for the uptake of CO2, water escapes from the leaf during active photosynthesis. Desert plants have evolved processes to conserve water and deal with harsh conditions. A more efficient use of CO2 allows plants to adapt to living with less water. Some plants such as cacti (Figure 12.21) can prepare materials for photosynthesis during the night by a temporary carbon fixation/storage process, because opening the stomata at this time conserves water due to cooler temperatures. In addition, cacti have evolved the ability to carry out low levels of photosynthesis without opening stomata at all, an extreme mechanism to face extremely dry periods.

Figure 12.21 The harsh conditions of the desert have led plants like these cacti to evolve variations of the light-independent reactions of photosynthesis. These variations increase the efficiency of water usage, helping to conserve water and energy. (Credit: Piotr Wojtkowski)

12.3.2 The Energy Cycle

All living things access energy by breaking down carbohydrate molecules. In nature, there is no such thing as waste. Every single atom of matter and energy is conserved, recycling over and over infinitely. Substances change form or move from one type of molecule to another, but their constituent atoms never disappear (**Figure 12.22**).

CO_2 is no more a form of waste than oxygen is wasteful to photosynthesis. Both are byproducts of reactions that move on to other reactions. Photosynthesis absorbs light energy to build carbohydrates in chloroplasts, and aerobic cellular respiration releases energy by using oxygen to metabolize carbohydrates in the cytoplasm and mitochondria. Both processes use electron transport chains to capture the energy necessary to drive other reactions. These two powerhouse processes, photosynthesis and cellular respiration, function in biological, cyclical harmony to allow organisms to access life-sustaining energy that originates millions of miles away in a burning star humans call the sun.

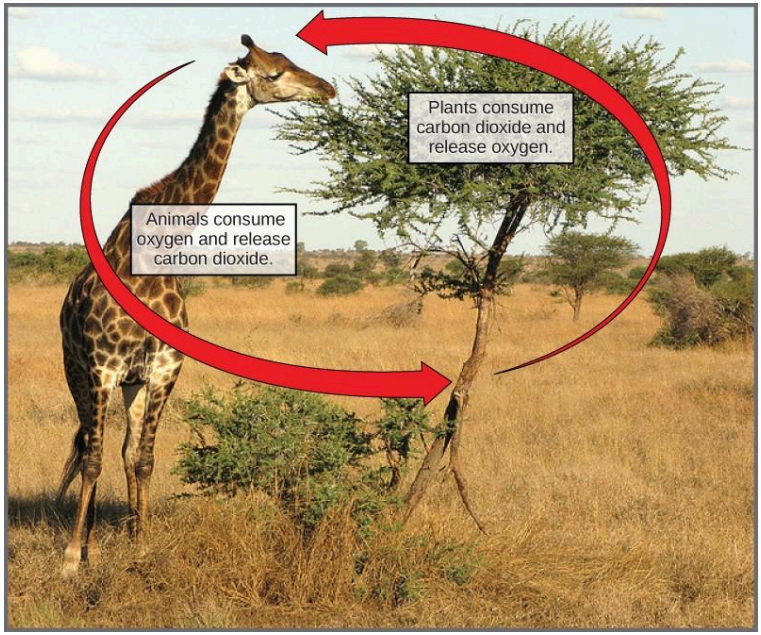


Figure 12.22 Photosynthesis consumes carbon dioxide and produces oxygen. Aerobic respiration consumes oxygen and produces carbon dioxide. These two processes play an important role in the carbon cycle. (Credit: modification of work by Stuart Bassil)

PART V

PART 5. HOW ARE HERITABLE TRAITS DETERMINED AND PASSED ON?

CHAPTER 13. THE CELL CYCLE & MITOSIS

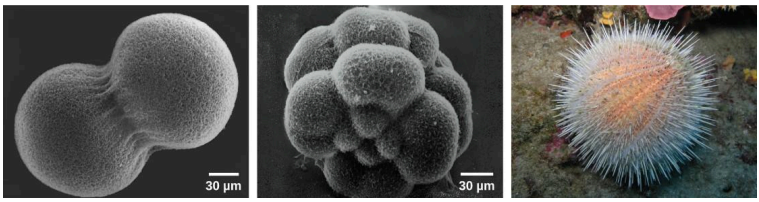


Figure 13.1 A sea urchin begins life as a single cell that (a) divides to form two cells, visible by scanning electron microscopy. After four rounds of cell division, (b) there are 16 cells, as seen in this SEM image. After many rounds of cell division, the individual develops into a complex, multicellular organism, as seen in this (c) mature sea urchin. (Credit a: modification of work by Evelyn Spiegel, Louisa Howard; credit b: modification of work by Evelyn Spiegel, Louisa Howard; credit c: modification of work by Marco Busdraghi; scale-bar data from Matt Russell)

Chapter Outline

- 13.1 DNA Organization and the Cell Cycle
- 13.2 The Cell Cycle
- 13.3 Control of the Cell Cycle
- 13.4 Cancer and the Cell Cycle
- 13.5 Prokaryotic Cell Division

Introduction

A human, as well as every sexually reproducing organism, begins life as a fertilized egg or zygote. Trillions of cell divisions subsequently occur in a controlled manner to produce a complex, multicellular human. In other words, that original single cell is the ancestor of every other cell in the body. Once a being is fully grown, cell reproduction is still necessary to repair or regenerate tissues. For example, new blood and skin cells are constantly being produced. All multicellular organisms use cell division for growth, maintenance, and repair of cells and tissues. Cell division is tightly regulated, and the occasional failure of regulation can have life-threatening consequences. Single-celled organisms use cell division as their method of reproduction.

13.1 | DNA Organization and

the Cell Cycle

Learning Objectives

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

- Describe the structure of prokaryotic and eukaryotic genomes.
- Distinguish between chromosomes, genes, and traits.
- Describe the mechanisms of chromosome compaction.

The continuity of life from one cell to another has its foundation in the reproduction of cells by way of the cell cycle. The **cell cycle** is an orderly sequence of events that describes the stages of a cell's life from the division of a single parent cell to the production of two genetically identical new daughter cells. The mechanisms involved in the cell cycle are highly regulated.

13.1.1 Genomic DNA

Before discussing the steps a cell must undertake to replicate, we need a deeper understanding of the structure and function of a cell's genetic information. A cell's DNA, packaged as double-stranded DNA molecules, is called its **genome**.

In prokaryotes, the genome is composed of a single circular double-stranded DNA molecule. (**Figure 13.2**). The region in the cell containing this genetic material is called a **nucleoid**. Some prokaryotes also have smaller loops of non- essential DNA called **plasmids.** Bacteria can exchange these plasmids with other bacteria, sometimes receiving beneficial new genes that the recipient can add to their chromosomal DNA. Antibiotic resistance is one trait that often spreads through a bacterial colony through plasmid exchange.

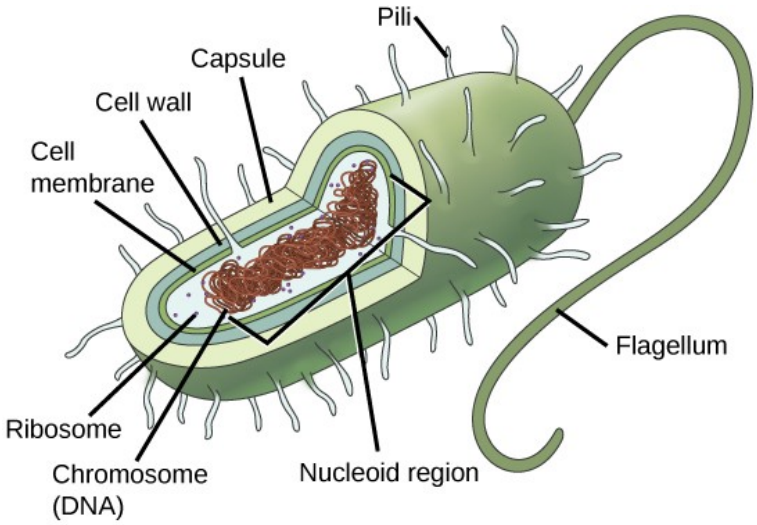


Figure 13.2 Prokaryotic cells, including all of the Domain Bacteria and all of the Domain Archaea, have a single, circular chromosome located in a central region called the nucleoid.

In eukaryotic cells, the genome consists of several double-stranded linear DNA molecules (**Figure 13.3**). Each species has a characteristic number of chromosomes in the nuclei of its cells. Human body cells have 46 chromosomes, while human **gametes** (sperm or eggs) have 23 chromosomes each. A typical body cell, or **somatic cell**, contains two matched sets of chromosomes, a configuration known as **diploid**. The letter *n* is used to represent a single set of chromosomes; therefore, a diploid organism is designated 2*n*. Human cells that contain one set of chromosomes are called gametes, or

sex cells; these are eggs and sperm, and are designated *1n*, or **haploid**.

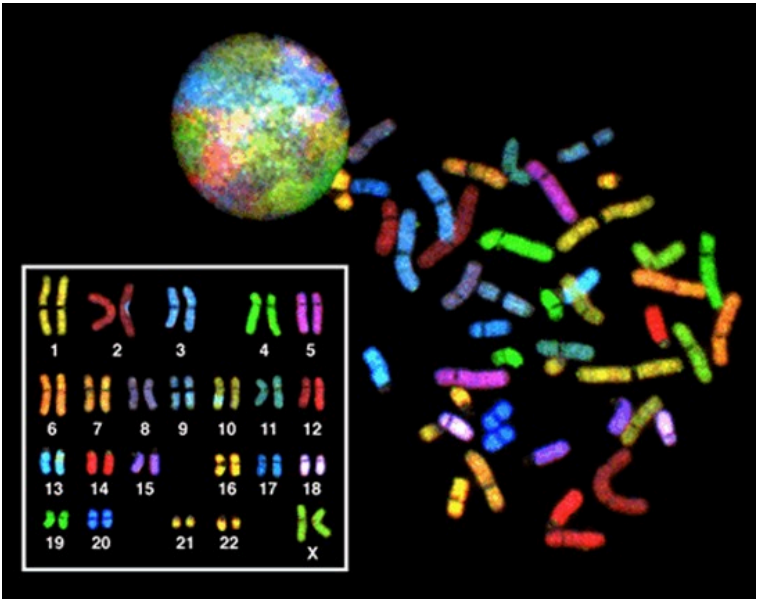


Figure 13.3 A karyotype of human chromosomes, showing their distinct sizes and banding patterns. In this image, the chromosomes were exposed to fluorescent stains to highlight chromosomes in different colors. (credit: National Human Genome Project/NIH)

Matched pairs of chromosomes in a diploid organism are called **homologous** ("same knowledge") **chromosomes**. Homologous chromosomes are the same length and have specific nucleotide segments called genes in exactly the same location, or locus. Genes are the functional units of chromosomes and determine specific characteristics by coding

for specific proteins. Traits are the variations of those characteristics. For example, hair color is a characteristic with traits that are blonde, brown, or black. Each copy of a homologous pair of chromosomes originates from a different parent; therefore, the genes themselves are not identical. The variation of individuals within a species is due to the specific combination of the genes inherited from both parents. Even a slightly altered sequence of nucleotides within a gene can result in an alternative trait.

For example, there are three possible gene sequences on the human chromosome that code for blood type: sequence A, sequence B, and sequence O. Because all diploid human cells have only two copies of the chromosome that determines blood type, the blood type trait is determined by which two versions of the marker gene are inherited. It is possible to have two copies of the same gene sequence on both homologous chromosomes, with one on each (for example, AA, BB, or OO), or two different sequences, such as AO, BO, or AB. Minor variations of traits, such as blood type, eye color, and handedness all contribute to the natural variation found within a species. If the entire DNA sequence from any pair of human homologous chromosomes is compared, the difference is less than one percent. The sex chromosomes, X and Y, are the single exception to the rule of homologous chromosome uniformity: Other than a small amount of homology that is necessary to accurately produce gametes, the genes found on the X and Y chromosomes are different

13.1.2 Eukaryotic Chromosomal Structure and Compaction

If the DNA from all 46 chromosomes in a human cell nucleus was laid out 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 be tightly packaged 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 a number of ways that chromosomes are compacted.

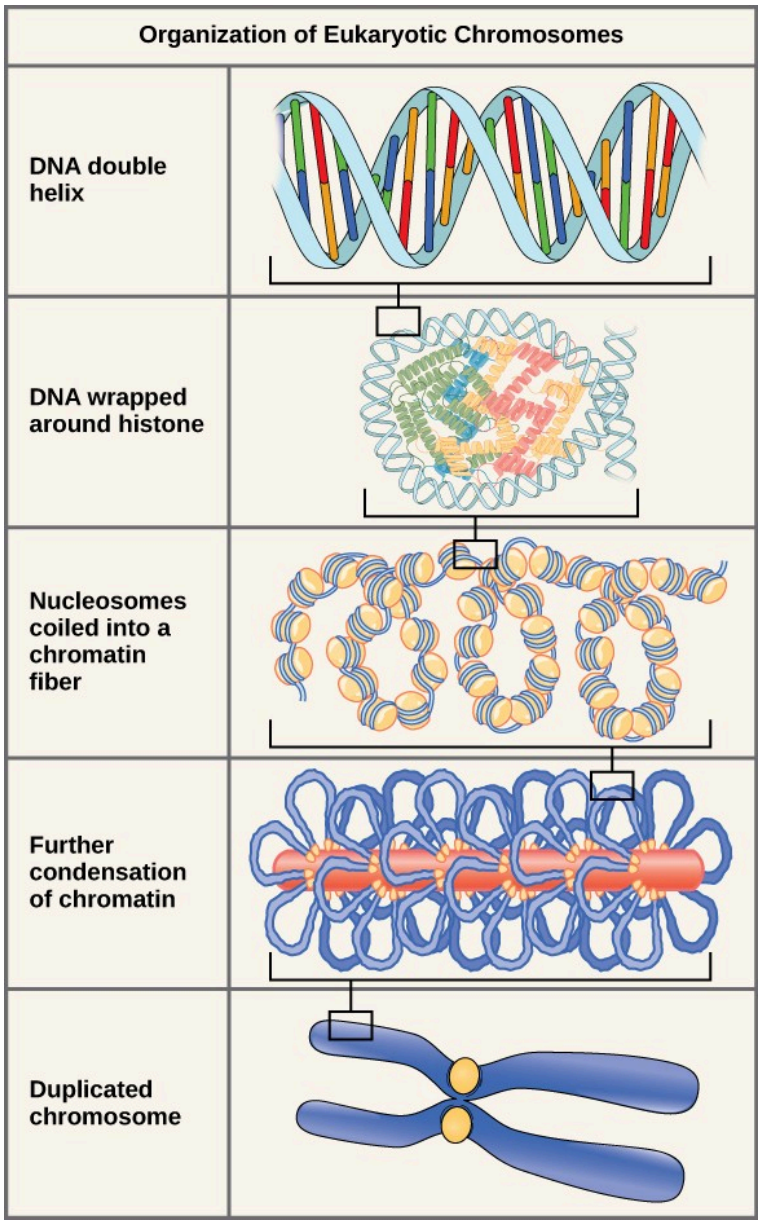


Figure 13.4 Double-stranded DNA wraps around histone proteins to form nucleosomes that have the appearance of

"beads on a string." The nucleosomes are coiled into a 30-nm chromatin fiber. When a cell undergoes mitosis, the chromosomes condense even further.

In the first level of compaction, short stretches of the DNA double helix wrap around a core of eight **histone proteins** at regular intervals along the entire length of the chromosome (**Figure 13.4**). The DNA-histone complex is called **chromatin**. The beadlike, histone DNA complex is called a **nucleosome**, and DNA connecting the nucleosomes is called linker DNA. A DNA molecule in this form is about seven times shorter than the double helix without the histones, and the beads are about 10 nm in diameter, in contrast with the 2-nm diameter of a DNA double helix. The next level of compaction occurs as the nucleosomes and the linker DNA between them are coiled into a 30-nm chromatin fiber. This coiling further shortens the chromosome so that it is about 50 times shorter. In the third level of packing, a variety of fibrous proteins is used to pack the chromatin. These fibrous proteins also ensure that each chromosome in a non-dividing cell occupies a particular area of the nucleus that does not overlap with that of any other chromosome (see the top image in **Figure 13.4**).

DNA replicates in the S phase of interphase. After replication, the chromosomes are composed of two linked sister **chromatids**. When fully compact, the pairs of identically packed chromosomes are bound to each other by

cohesin proteins. The connection between the sister chromatids is closest in a region called the **centromere**. The conjoined sister chromatids, with a diameter of about 1 μm, are visible under a light microscope. The centromeric region is highly condensed and thus will appear as a constricted area.

13.2 | The Cell Cycle

Learning Objectives

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

- Describe the three stages of interphase.
- Discuss the behavior of chromosomes during mitosis.
- Explain how the cytoplasmic content is divided during cytokinesis.
- Define the quiescent G_0 phase.

The cell cycle is an ordered series of events involving cell growth and cell division that produces two new daughter cells. Somatic cells on the path to cell division proceed through a

series of precisely timed and carefully regulated stages of growth, DNA replication, and division that produces two genetically identical cells. In other words, a typical 2n somatic cell will divide into two 2n somatic cells that are genetically identical. This is a form of asexual reproduction.

As a quick overview of the cell cycle, we see two major phases: **interphase** and the **M phase** (**Figure 13.5**). During interphase, the cell undergoes three distinct periods: **G1, S, and G2**. During G1, S, and G2, the cell grows, DNA is replicated, and the cell grows some more. During the M phase, the cell undergoes two distinct periods: **mitosis** (also called karyokinesis) and **cytokinesis,** the division of the cytoplasm.

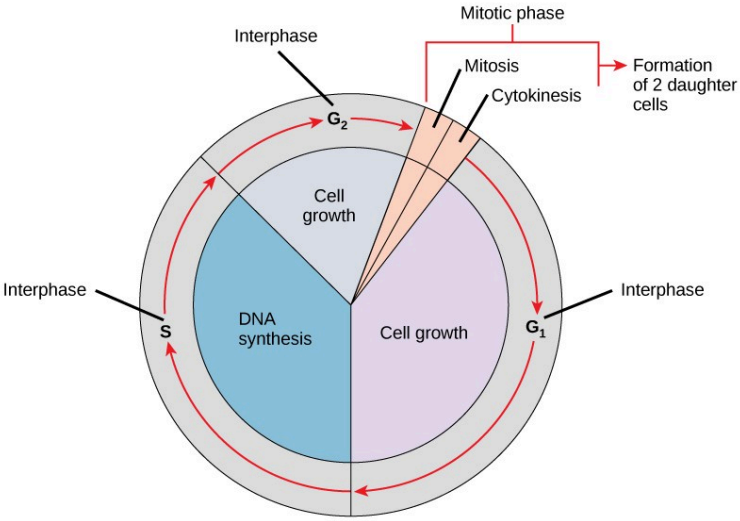


Figure 13.5 The cell cycle consists of interphase and the mitotic phase. Interphase consists of G1, S, and G2. The mitotic phase consists of mitosis and cytokinesis.

13.2.1 Interphase

During interphase, the cell undergoes normal growth processes while also preparing for cell division. In order for a cell to move from interphase into the mitotic phase, many internal and external conditions must be met. The three aspects or stages of interphase are called G1, S, and G2.

G1 Phase (First Gap)

The first stage of interphase is called the **G1 phase** (first gap) because, from a microscopic aspect, little change is visible. However, during the G1 stage, the cell is quite active at the biochemical level. The cell is accumulating the building blocks of chromosomal DNA and the associated proteins as well as accumulating sufficient energy reserves to complete the task of replicating each chromosome in the nucleus.

S Phase (Synthesis of DNA)

Throughout interphase, nuclear DNA remains in a semi-condensed chromatin configuration. In the **S phase**, DNA replication can proceed through the mechanisms that result in the formation of identical pairs of DNA molecules—sister chromatids—that are firmly attached to the centromeric region. The centrosome is duplicated during the S phase. The two centrosomes will give rise to the **mitotic spindle**, the apparatus that orchestrates the movement of chromosomes

during mitosis. At the center of each animal cell, the centrosomes of animal cells are associated with a pair of rod-like objects, the **centrioles**, which are at right angles to each other. Centrioles help organize cell division. Centrioles are not present in the centrosomes of other eukaryotic species, such as plants and most fungi.

G2 Phase (Second Gap)

In the **G2 phase**, the cell replenishes its energy stores and synthesizes proteins necessary for chromosome manipulation. Some cell organelles are duplicated, and the cytoskeleton is dismantled to provide resources for the mitotic phase. There may be additional cell growth during G2. The final preparations for the mitotic phase must be completed before the cell is able to enter the first stage of mitosis.

13.2.2 The Mitotic Phase

The mitotic phase is a multistep process during which the duplicated chromosomes are aligned, separated, and moved into two new, identical daughter cells. M phase is divided into mitosis and cytokinesis.

Mitosis

Mitosis is divided into a series of phases—Prophase,

Prometaphase, Metaphase, Anaphase, and Telophase—that result in the division of the cell nucleus (**Figure 13.6**).

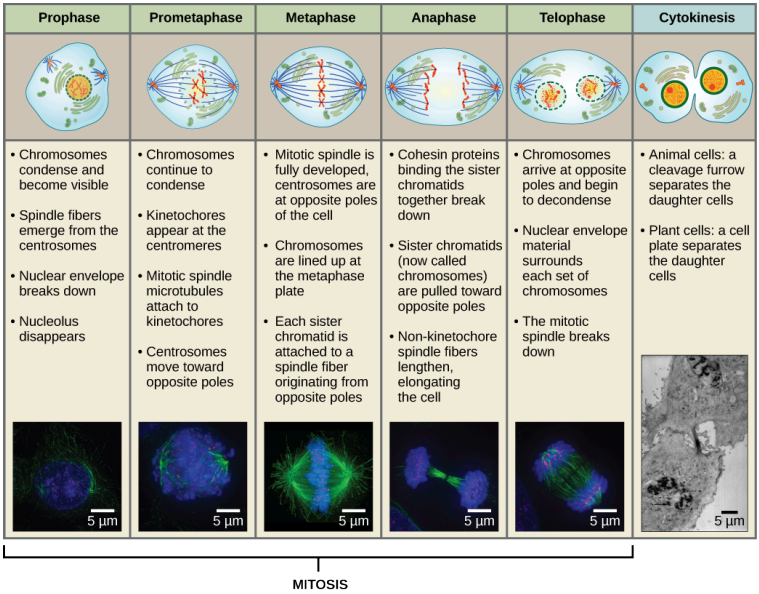


Figure 13.6 Mitosis is divided into five stages—prophase, prometaphase, metaphase, anaphase, and telophase. The pictures at the bottom were taken by fluorescence microscopy (hence, the black background) of cells stained by fluorescent dyes: blue indicates DNA (chromosomes) and green fluorescence indicates microtubules (spindle apparatus). (Credit "mitosis drawings": modification of work by Mariana Ruiz Villareal; credit "micrographs": modification of work by Roy van Heesbeen; credit "cytokinesis micrograph": Wadsworth Center/New York State Department of Health; scale-bar data from Matt Russell)

During **Prophase**, the nuclear envelope starts to dissociate

into small vesicles and the membranous organelles (such as the Golgi complex or Golgi apparatus, and endoplasmic reticulum), fragment and disperse toward the periphery of the cell. The nucleolus disappears (disperses). The centrosomes begin to migrate to opposite poles of the cell. Microtubules that will form the mitotic spindle extend between the centrosomes, pushing them farther apart as the microtubule fibers lengthen. The sister chromatids begin to coil more tightly with the aid of **condensin** proteins and become visible under a light microscope.

During **Prometaphase**, many processes that were begun in prophase continue to advance. The remnants of the nuclear envelope fragment. The mitotic spindle continues to develop as more microtubules assemble and stretch across the length of the former nuclear area. Chromosomes become more condensed and discrete. Each sister chromatid develops a protein structure called a **kinetochore** in the centromeric region (**Figure 13.7**). The proteins of the kinetochore attract and bind mitotic spindle microtubules. As the spindle microtubules extend from the centrosomes, some of these microtubules come into contact with and firmly bind to the kinetochores. Once a mitotic fiber attaches to a chromosome, the chromosome will be oriented until the kinetochores of sister chromatids face the opposite poles. Eventually, all the sister chromatids will be attached via their kinetochores to microtubules from opposing poles. Spindle microtubules that do not engage the chromosomes are called polar microtubules.

These microtubules overlap each other midway between the two poles and contribute to cell elongation. Astral microtubules are located near the poles, aid in spindle orientation, and are required for the regulation of mitosis.

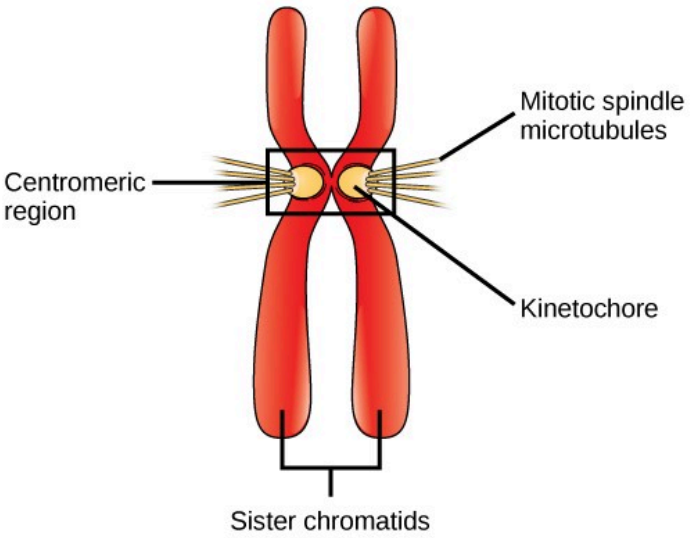


Figure 13.7 During prometaphase, mitotic spindle microtubules from opposite poles attach to each sister chromatid at the kinetochore. In anaphase, the connection between the sister chromatids breaks down, and the microtubules pull the chromosomes toward opposite poles.

During **Metaphase**, all the chromosomes are aligned in a plane called the **metaphase plate**, or the equatorial plane, midway between the two poles of the cell. The sister chromatids are still tightly attached to each other by cohesin

proteins. At this time, the chromosomes are maximally condensed and are at their most visible under a microscope.

During **Anaphase**, the cohesin proteins degrade, and the sister chromatids separate at the centromere. Each chromatid, now called a chromosome, is pulled rapidly toward the centrosome to which its microtubule is attached. The cell becomes visibly elongated (oval shaped) as the polar microtubules slide against each other at the metaphase plate where they overlap.

During **Telophase**, the chromosomes reach the opposite poles, the nuclear envelopes reforms around the chromosomes. The mitotic spindles are depolymerized into tubulin monomers that will be used to assemble cytoskeletal components for each daughter cell. The enclosed chromosomes begin to decondense (unravel), relaxing into a chromatin configuration, and nucleosomes appear within the nuclear area. The cell will appear as a cell with two distinct nuclei. At this stage, mitosis is complete and division of the cytoplasm will follow during cytokinesis.

Cytokinesis

Cytokinesis, or "cell motion," is the second main stage of the mitotic phase during which cell division is completed via the physical separation of the cytoplasmic components into two daughter cells. Division is not complete until the cell components have been apportioned and completely separated into the 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 actually begins at the midpoint 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 constriction. This constriction or fissure is called the **cleavage furrow**. The furrow deepens as the actin ring contracts, and eventually the membrane is cleaved in two (**Figure 13.8**).

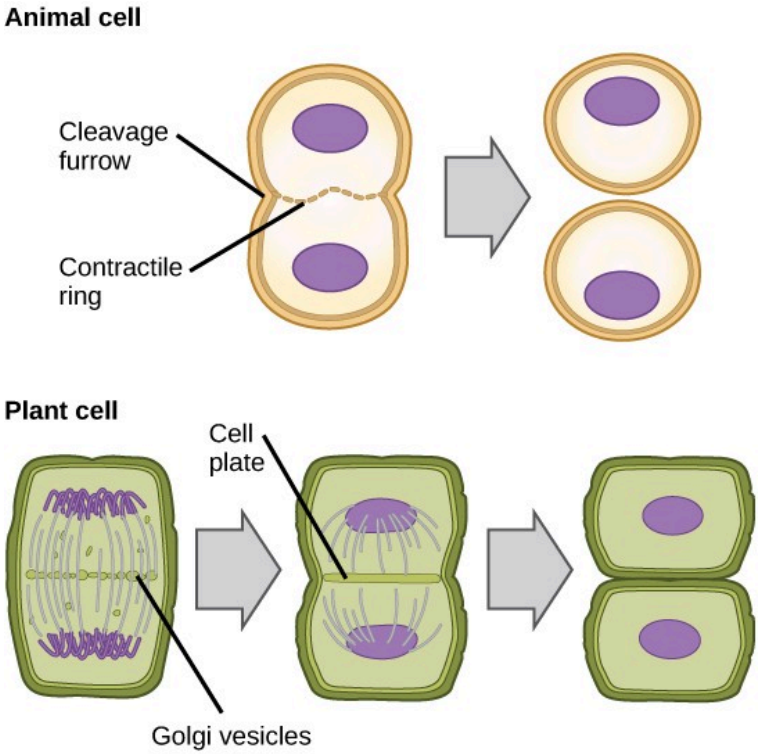


Figure 13.8 In animal cell cytokinesis, a ring of actin filaments forms at the metaphase plate. The ring contracts, forming a cleavage furrow, which divides the cell in two. In plant cells, Golgi vesicles coalesce at the former metaphase plate, forming a phragmoplast. A cell plate formed by the fusion of the vesicles of the phragmoplast grows from the center toward the cell walls, and the membranes of the vesicles fuse to form a plasma membrane that divides the cell in two.

In plant cells, 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 into vesicles and dispersing throughout the dividing cell. During telophase, these Golgi vesicles are transported on microtubules to form a phragmoplast (a vesicular structure) at the metaphase plate. There, the vesicles fuse and coalesce 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 walls at the periphery of the cell. Enzymes use the glucose that has accumulated between the membrane layers to build a new cell wall made of cellulose. Remember that cellulose is a structural polymer of glucose. The Golgi membranes become parts of the plasma membrane on either side of the new cell wall (**Figure 13.8**).

13.2.3 G_0 Phase (A variation of the cell cycle)

Not all cells adhere to the classic cell cycle pattern in which a newly formed daughter cell immediately enters the preparatory phases of interphase, closely followed by the mitotic phase. Cells in **G_0 phase** ("G zero") are not actively preparing to divide. The cell is in a **quiescent** (inactive) stage that occurs when cells exit the cell cycle. Some cells enter G_0 temporarily until an external signal triggers the onset of G1. Other cells that never or rarely divide, such as mature cardiac muscle and nerve cells, remain in G_0 permanently. Although cells in G_0 are inactive in the sense that they are not actively

preparing for and undergoing cell division, they can be very active in other ways. Many cells in the adult body for instance, are permanently in G_0 while fulfilling their specialized functions.

13.3 | Control of the Cell Cycle

Learning Objectives

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

- Understand how the cell cycle is controlled by mechanisms both internal and external to the cell.
- Explain how the three internal control checkpoints occur at the end of G1, at the G2/M transition, and during metaphase.
- Describe the molecules that control the cell cycle through positive and negative regulation.

The length of the cell cycle is highly variable, even within the

cells of a single organism. In humans, the frequency of cell turnover ranges from a few hours in early embryonic development, to an average of two to five days for epithelial cells. Some cells, such as cortical neurons or cardiac muscle cells spend all of their time in G_0 and never divide. In early embryos of fruit flies, the cell cycle is completed in about eight minutes. The timing of events in the cell cycle is controlled by mechanisms that are both internal and external to the cell.

There is also variation in the time that a cell spends in each phase of the cell cycle. When fast-dividing mammalian cells are grown in culture (outside the body under optimal growing conditions), the length of the cycle is about 24 hours. In rapidly dividing human cells with a 24-hour cell cycle, the G1 phase lasts approximately 9 hours, the S phase lasts 10 hours, the G2 phase lasts about 4.5 hours, and the M phase lasts approximately 0.5 hours.

13.3.1 Regulation of the Cell Cycle by External Events

Both the initiation and inhibition of cell division are triggered by events external to the cell when it is about to begin the replication process. An event may be as simple as the death of a nearby cell or as sweeping as the release of growth- promoting hormones, such as human growth hormone (HGH). A lack of HGH can inhibit cell division, resulting in dwarfism, whereas too much HGH can result in gigantism. Crowding of cells

can also inhibit cell division. Another factor that can initiate cell division is the size of the cell; as a cell grows, it becomes inefficient due to its decreasing surface-to-volume ratio. The solution to this problem is to divide.

Whatever the source of the message, the cell receives the signal, and a series of events within the cell allows it to proceed into interphase. Moving forward from this initiation point, every parameter required during each cell cycle phase must be met or the cycle cannot progress.

13.3.2 Regulation of the Cell Cycle at Internal Checkpoints

It is essential that the daughter cells that have been produced be exact duplicates of the parent cell. Mistakes in the duplication or distribution of the chromosomes lead to mutations that may be passed forward to every new cell produced from an abnormal cell. To prevent a compromised cell from continuing to divide, there are internal control mechanisms that operate at three main **cell cycle checkpoints**. A checkpoint is one of several points in the eukaryotic cell cycle at which the progression of a cell to the next stage in the cycle can be halted until conditions are favorable. These checkpoints occur near the end of G1, at the G2/M transition, and during metaphase (**Figure 13.10**).

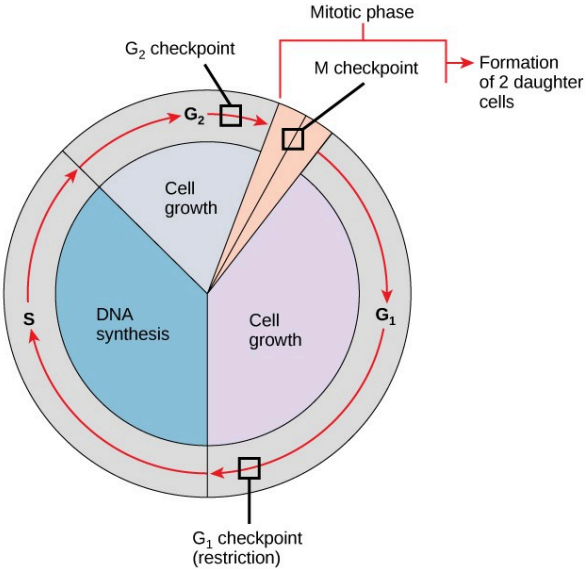


Figure 13.10 The cell cycle is controlled at three checkpoints. The integrity of the DNA is assessed at the G1 checkpoint. Proper chromosome duplication is assessed at the G2 checkpoint. Attachment of each kinetochore to a spindle fiber is assessed at the M checkpoint.

The G1 Checkpoint

The G1 checkpoint determines whether all conditions are favorable for cell division to proceed. The G1 checkpoint, also called the restriction point (in yeast), is a point at which the cell irreversibly commits to the cell division process. External influences, such as growth factors, play a large role in carrying the cell past the G1 checkpoint. In addition to adequate reserves and cell size, there is a check for genomic DNA damage at the G1 checkpoint. A cell that does not meet all

the requirements will not be allowed to progress into the S phase. The cell can halt the cycle and attempt to remedy the problematic condition, or the cell can advance into G_0 and await further signals when conditions improve.

The G2 Checkpoint

The G2 checkpoint bars entry into the mitotic phase if certain conditions are not met. As at the G1 checkpoint, cell size and protein reserves are assessed. However, the most important role of the G2 checkpoint is to ensure that all of the chromosomes have been replicated and that the replicated DNA is not damaged. If the checkpoint mechanisms detect problems with the DNA, the cell cycle is halted, and the cell attempts to either complete DNA replication or repair the damaged DNA.

The 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 whether all of 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 at least two spindle fibers arising from opposite poles of the cell.

13.3.3 Regulator Molecules of the Cell Cycle

In addition to the internally controlled checkpoints, there are two groups of intracellular molecules that regulate the cell cycle. These regulatory molecules either promote progress of the cell to the next phase (positive regulation) or halt the cycle (negative regulation). Regulator molecules may act individually, or they can influence the activity or production of other regulatory proteins. Therefore, the failure of a single regulator may have almost no effect on the cell cycle, especially if more than one mechanism controls the same event. Conversely, the effect of a deficient or non-functioning regulator can be wide-ranging and possibly fatal to the cell if multiple processes are affected.

Positive Regulation of the Cell Cycle

Two groups of proteins, called **cyclins** and **cyclin-dependent kinases** (Cdks), are responsible for the progress of the cell through the various checkpoints. The levels of the four cyclin proteins fluctuate throughout the cell cycle in a predictable pattern (**Figure 13.11**). Increases in the concentration of cyclin proteins are triggered by both external and internal signals. After the cell moves to the next stage of the cell cycle, the cyclins that were active in the previous stage are degraded.

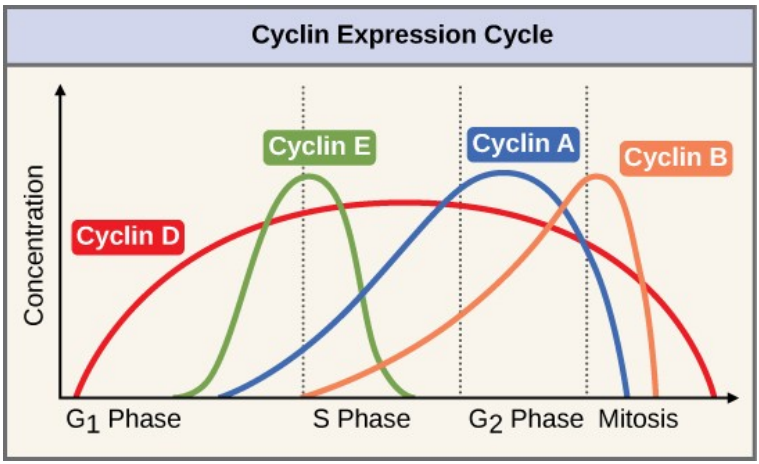


Figure 13.11 The concentrations of cyclin proteins change throughout the cell cycle. There is a direct correlation between cyclin accumulation and the three major cell cycle checkpoints. Also note the sharp decline of cyclin levels following each checkpoint (the transition between phases of the cell cycle), as cyclin is degraded by cytoplasmic enzymes. (Credit: modification of work by "WikiMiMa"/Wikimedia Commons.)

Cyclins regulate the cell cycle only when they are tightly bound to Cdks. To be fully active, the Cdk/cyclin complex must also be phosphorylated in specific locations. Like all kinases, Cdks are enzymes (kinases) that phosphorylate other proteins. Phosphorylation activates the protein by changing its shape. The proteins phosphorylated by Cdks are involved in advancing the cell to the next phase. (**Figure 13.12**). The levels of Cdk proteins are relatively stable throughout the cell cycle; however, the concentrations of cyclin fluctuate and determine when Cdk/cyclin complexes form. The different

cyclins and Cdks bind at specific points in the cell cycle and thus regulate different checkpoints.

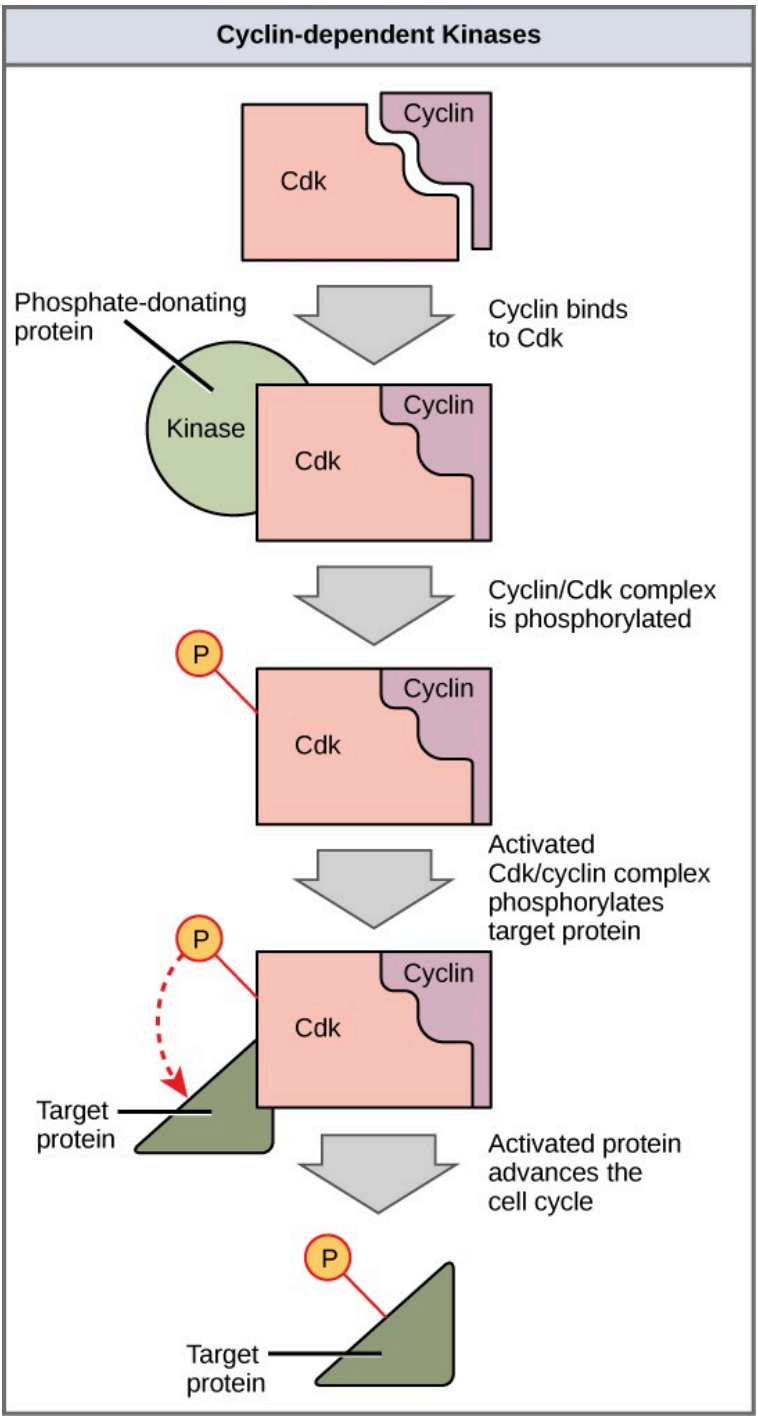

Cyclin-dependent Kinases
Cyclin
Cdk
Phosphate-donating protein
Cyclin binds to Cdk
Kinase
Cyclin
Cdk
Cyclin/Cdk complex is phosphorylated
P
Cyclin
Cdk
Activated Cdk/cyclin complex phosphorylates target protein
P
Cyclin
Cdk
Target protein
Activated protein advances the cell cycle
P
Target protein

Figure 13.12 Cyclin-dependent kinases (Cdks) are protein kinases that, when fully activated, can phosphorylate and thus activate other proteins that advance the cell cycle past a checkpoint. To become fully activated, a Cdk must bind to a cyclin protein and then be phosphorylated by another kinase.

Since the cyclic fluctuations of cyclin levels are based on the timing of the cell cycle and not on specific events. Regulation of the cell cycle usually occurs by either the Cdk molecules alone or the Cdk/cyclin complexes. Without a specific concentration of fully activated cyclin/Cdk complexes, the cell cycle cannot proceed through the checkpoints.

Although the cyclins are the main regulatory molecules that determine the forward momentum of the cell cycle, there are several other mechanisms that fine-tune the progress of the cycle with negative, rather than positive, effects. These mechanisms essentially block the progression of the cell cycle until problematic conditions are resolved. Molecules that prevent the full activation of Cdks are called Cdk inhibitors. Many of these inhibitor molecules directly or indirectly monitor a particular cell cycle event. The block placed on Cdks by inhibitor molecules will not be removed until the specific event that the inhibitor monitors is completed.

Negative Regulation of the Cell Cycle

The second group of cell cycle regulatory molecules are negative regulators. Negative regulators halt the cell cycle.

Remember that in positive regulation, active molecules cause the cycle to progress.

The best understood negative regulatory molecules are **retinoblastoma protein (Rb)**, **p53**, and **p21**. Retinoblastoma proteins are a group of tumor-suppressor proteins common in many cells. The 53 and 21 designations refer to the functional molecular masses of the proteins (p) in kilodaltons. Much of what is known about cell cycle regulation comes from research conducted with cells that have lost regulatory control. All three of these regulatory proteins were discovered to be damaged or non-functional in cells that had begun to replicate uncontrollably (became cancerous). In each case, the main cause of the unchecked progress through the cell cycle was a faulty copy of the regulatory protein.

Rb, p53, and p21 act primarily at the G1 checkpoint. Regulator protein **p53** is a multi-functional protein that has a major impact on the commitment of a cell to division because it acts when there is damaged DNA in cells that are undergoing the preparatory processes during G1. If damaged DNA is detected, p53 halts the cell cycle and recruits enzymes to repair the DNA. If the DNA cannot be repaired, p53 can trigger **apoptosis**, or cell suicide, to prevent the duplication of damaged chromosomes. As p53 levels rise, the production of p21 is triggered. Regulator protein **p21** enforces the halt in the cycle dictated by p53 by binding to and inhibiting the activity of the Cdk/cyclin complexes. As a cell is exposed to more

stress, higher levels of p53 and p21 will accumulate, making it less likely that the cell will move into the S phase.

Rb exerts its regulatory influence on other positive regulator proteins. Chiefly, Rb monitors cell size. In the active, dephosphorylated state, Rb binds to proteins called transcription factors, most commonly, E2F (**Figure 13.13**). Transcription factors "turn on" specific genes, allowing the production of proteins encoded by that gene. When Rb is bound to E2F, production of proteins necessary for the G1/S transition is blocked. As the cell increases in size, Rb is slowly phosphorylated until it becomes inactivated. Rb releases E2F, which can now turn on the gene that produces the transition protein, and this particular block is removed. For the cell to move past each of the checkpoints, all positive regulators must be "turned on," and all negative regulators must be "turned off."

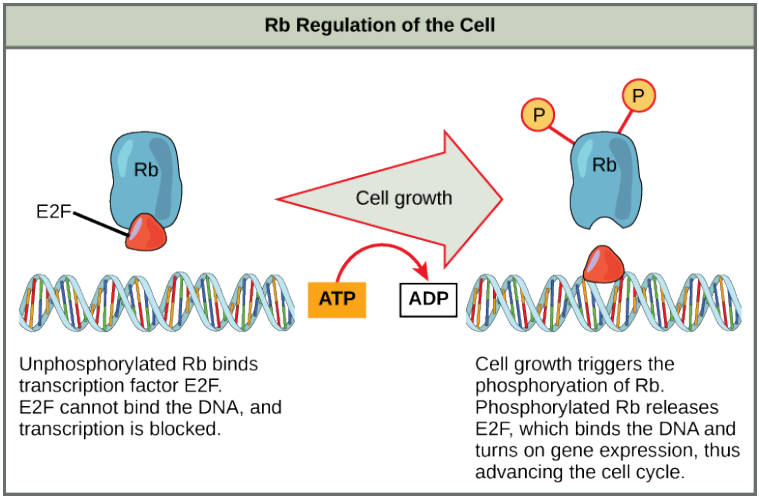


Figure 13.13 Rb halts the cell cycle and releases its hold in response to cell growth.

13.4 | Cancer and the Cell Cycle

Learning Objectives

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

- Describe how cancer is caused by

uncontrolled cell growth.

- Understand how proto-oncogenes are normal cell genes that, when mutated, become oncogenes.
- Describe how tumor suppressors function.
- Explain how mutant tumor suppressors cause cancer.

Cancer comprises many different diseases caused by a common mechanism: uncontrolled cell growth. Despite the overlapping levels of cell cycle control, errors do occur. One of the critical processes monitored by the cell cycle checkpoint surveillance mechanism is the proper replication of DNA during the S phase. Even when all of the cell cycle controls are fully functional, a small percentage of replication errors (mutations) will be passed on to the daughter cells. When changes to a DNA nucleotide sequence occur within a coding portion of a gene and are not corrected, a gene mutation results.

All cancers start when a gene mutation gives rise to a faulty protein that plays a key role in cell reproduction. The change in the cell that results from the malformed protein may be minor: perhaps a slight delay in the binding of Cdk to cyclin or an Rb protein that detaches from its target DNA while still

phosphorylated. Even minor mistakes may allow subsequent mistakes to occur more readily. Over and over, small uncorrected errors are passed from the parent cell to the daughter cells and amplified as each generation produces more non-functional proteins from uncorrected DNA damage. Eventually, the pace of the cell cycle speeds up as the effectiveness of the control and repair mechanisms decreases. Uncontrolled growth of the mutated cells outpaces the growth of normal cells in the area, and a tumor can result.

13.4.1 Proto-oncogenes

The genes that code for the positive cell cycle regulators are called **proto-oncogenes**. Proto-oncogenes are normal genes that, when mutated in certain ways, become **oncogenes**, genes that cause a cell to become cancerous. Consider what might happen to the cell cycle in a cell with a recently acquired oncogene. In most instances, the alteration of the DNA sequence will result in a less functional (or non-functional) protein. The result is detrimental to the cell and will likely prevent the cell from completing the cell cycle; however, the organism is not harmed because the mutation will not be carried forward. If a cell cannot reproduce, the mutation is not propagated and the damage is minimal. Occasionally, a gene mutation causes a change that increases the activity of a positive regulator. For example, a mutation that allows Cdk to be activated without being partnered with cyclin could push

the cell cycle past a checkpoint before all of the required conditions are met. If the resulting daughter cells are too damaged to undergo further cell divisions, the mutation would not be propagated and no harm would come to the organism. If the atypical daughter cells are able to undergo further cell divisions, subsequent generations of cells will probably accumulate even more mutations, some possible in additional genes that regulate the cell cycle.

The Cdk gene in the above example is only one of many genes that are considered proto-oncogenes. In addition to the cell cycle regulatory proteins, any protein that influences the cycle can be altered in such a way as to override cell cycle checkpoints. An oncogene is any gene that, when altered, leads to an increase in the rate of cell cycle progression.

13.4.2 Tumor Suppressor Genes

Like proto-oncogenes, many of the negative cell cycle regulatory proteins were discovered in cells that had become cancerous. **Tumor suppressor genes** are segments of DNA that code for negative regulator proteins, the type of regulators that, when activated, can prevent the cell from undergoing uncontrolled division. The collective function of the best-understood tumor suppressor gene proteins, Rb, p53, and p21, is to put up a roadblock to cell cycle progression until certain events are completed. A cell that carries a mutated form of a negative regulator might not be able to halt the cell cycle if

there is a problem. Tumor suppressors are similar to brakes in a vehicle: Malfunctioning brakes can contribute to a car crash.

Mutated p53 genes have been identified in more than one-half of all human tumor cells. This discovery is not surprising in light of the multiple roles that the p53 protein plays at the G1 checkpoint. A cell with a faulty p53 may fail to detect errors present in the genomic DNA (**Figure 13.14**). Even if a partially functional p53 does identify the mutations, it may no longer be able to signal the necessary DNA repair enzymes. Either way, damaged DNA will remain uncorrected. At this point, a functional p53 will deem the cell unsalvageable and trigger programmed cell death (apoptosis). The damaged version of p53 found in cancer cells cannot trigger apoptosis.

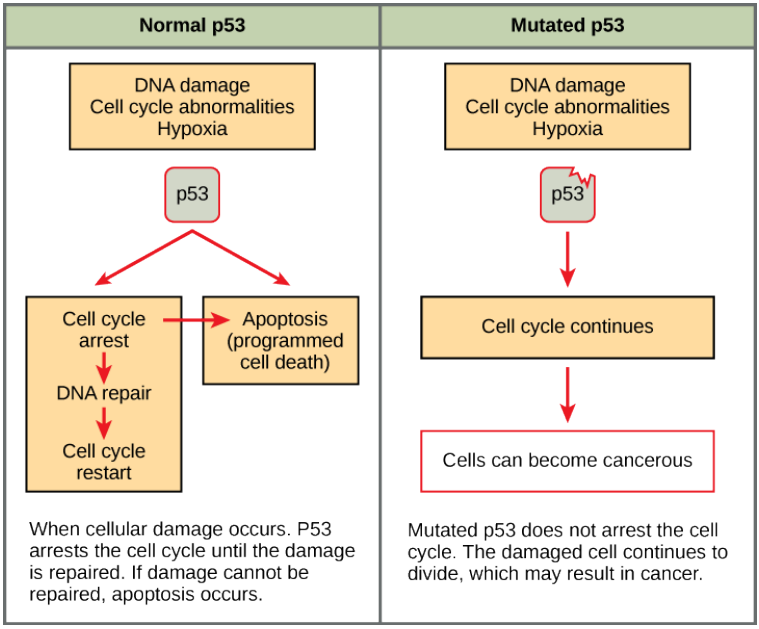


Figure 13.14 The role of normal p53 is to monitor DNA and the supply of oxygen. If damage is detected, p53 triggers repair mechanisms. If repairs are unsuccessful, p53 signals apoptosis. A cell with an abnormal p53 protein cannot repair damaged DNA and thus cannot signal apoptosis. Cells with abnormal p53 often become cancerous. (Credit: modification of work by Thierry Soussi)

The loss of p53 function has other repercussions for the cell cycle. Mutated p53 might lose its ability to trigger p21 production. Without adequate levels of p21, there is no effective block on Cdk activation. Essentially, without a fully functional p53, the G1 checkpoint is severely compromised and the cell proceeds directly from G1 to S regardless of internal and external conditions. At the completion of this

shortened cell cycle, two daughter cells are produced that have inherited the mutated p53 gene. Given the non-optimal conditions under which the parent cell reproduced, it is likely that the daughter cells will have acquired other mutations in addition to the faulty tumor suppressor gene. Cells such as these daughter cells quickly accumulate both oncogenes and non-functional tumor suppressor genes. Again, the result is tumor growth.

13.5 | Prokaryotic Cell Division

Learning Objectives

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

- Describe the process of binary fission in prokaryotes.
- Explain how FtsZ and tubulin proteins are examples of homology.

Prokaryotes propagate by **binary fission**. For unicellular organisms, cell division is the only method to produce new

individuals. In both prokaryotic and eukaryotic cells, the outcome of cell reproduction is a pair of daughter cells that are genetically identical to the parent cell. In unicellular organisms, daughter cells are individuals. To achieve the outcome of cloned offspring, certain steps are essential. The genomic DNA must be replicated and then allocated into the daughter cells; the cytoplasmic contents must also be divided to give both new cells the machinery to sustain life. In bacterial cells, the genome consists of a single, circular DNA chromosome; therefore, the process of cell division is greatly simplified. Karyokinesis is unnecessary because there is no nucleus and thus no need to direct one copy of the multiple chromosomes into each daughter cell.

13.5.1 Binary Fission

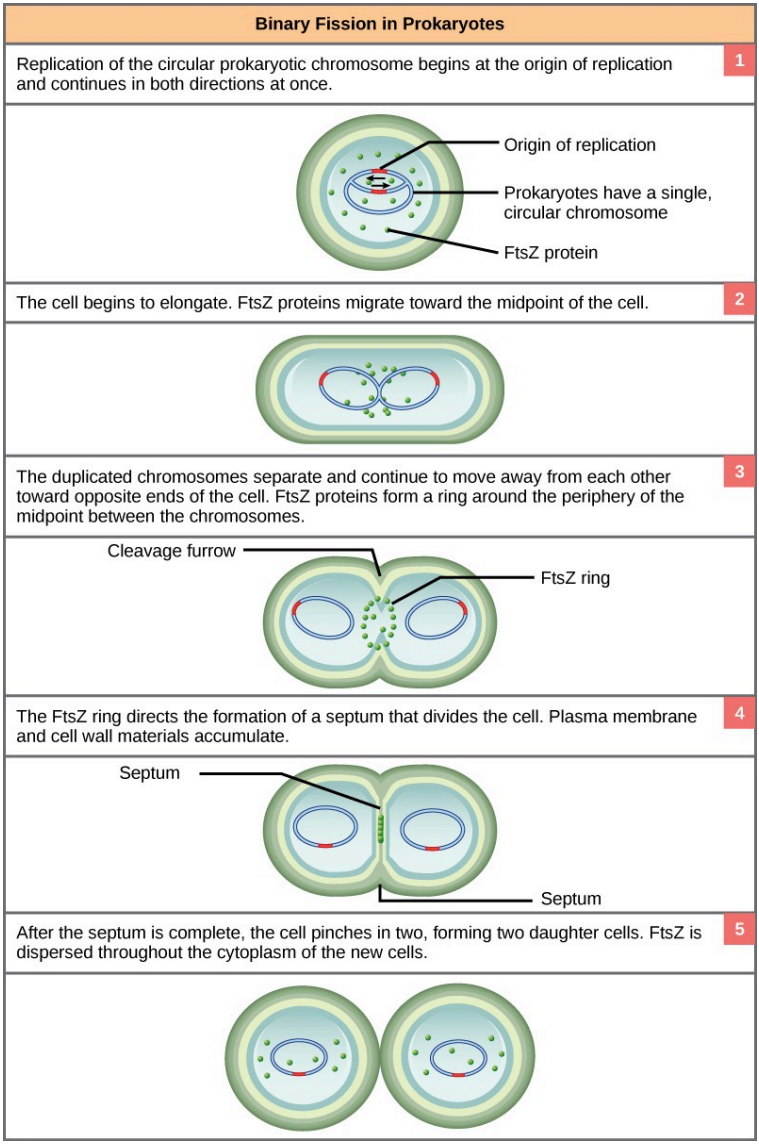


Figure 13.15 These images show the steps of binary fission in prokaryotes. (Credit: modification of work by "Mcstrother"/ Wikimedia Commons)

Due to the relative simplicity of the prokaryotes, the cell division process, called binary fission, is a less complicated and much more rapid process than cell division in eukaryotes. The single, circular DNA chromosome of bacteria is not enclosed in a nucleus, but instead occupies a specific location, the nucleoid, within the cell (**Figure 13.2**). Although the DNA of the nucleoid is associated with proteins that aid in packaging the molecule into a compact size, there are no histone proteins and thus no nucleosomes in prokaryotes. The packing proteins of bacteria are related to the cohesin and condensin proteins involved in the chromosome compaction of eukaryotes.

The bacterial chromosome is attached to the plasma membrane at about the midpoint of the cell. The starting point of replication, the **origin**, is close to the binding site of the chromosome to the plasma membrane (**Figure 13.15**). Replication of the DNA is bidirectional, moving away from the origin on both strands of the loop simultaneously. As the new double strands are formed, each origin point moves away from the plasma membrane attachment toward the opposite ends of the cell. As the cell elongates, the growing membrane aids in the transport of the chromosomes. 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** directs the partition, cell wall, between the 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. A **septum** is formed between the nucleoids, extending gradually from the periphery toward the center of the cell. When the new cell walls are in place, the daughter cells separate.

evolution CONNECTION

Mitotic Spindle Apparatus

The precise timing and formation of the mitotic spindle is critical to the success of eukaryotic cell division. Prokaryotic cells, on the other hand, do not undergo karyokinesis and therefore have no need for a mitotic spindle. The FtsZ protein that plays such a vital role in prokaryotic cytokinesis is structurally and functionally very similar to tubulin, the building block of microtubules that make up the mitotic spindle fibers that are necessary for eukaryotes. FtsZ proteins can form filaments, rings, and other three-dimensional structures that resemble the way tubulin forms microtubules, centrioles, and various cytoskeletal components. In addition, both FtsZ and tubulin employ the same energy sourse, GTP (guanosine

triphosphate), to rapidly assemble and disassemble complex structures.

FtsZ and tubulin are homologous structures derived from common evolutionary origins. In this example, FtsZ is the ancestor protein to tubulin (a modern protein). While both proteins are found in extant organisms, tubulin function has evolved and diversified tremendously since evolving from its FtsZ prokaryotic origin.

CHAPTER 14. DNA REPLICATION

Figure 14.1 Dolly the sheep was the first large mammal to be cloned.

Chapter Outline

- 14.1 Historical Basis of Modern Understanding
- 14.2 Overview of DNA Replication
- 14.3 DNA Replication in Prokaryotes
- 14.4 DNA Replication in Eukaryotes
- 14.5 DNA Repair

Introduction

The three letters "DNA" have now become synonymous with crime solving, paternity testing, human identification, and genetic testing. DNA can be retrieved from hair, blood, or saliva. Each person's DNA is unique, and it is possible to detect differences between individuals within a species on the basis of these unique features. DNA analysis has many practical applications beyond forensics. In humans, DNA testing is applied to numerous uses: tracing genealogy, identifying pathogens, archeological research, tracing disease outbreaks, and studying human migration patterns. In the medical field, DNA is used in diagnostics, new vaccine development, and cancer therapy. It is now possible to determine predisposition to diseases by looking at genes.

Each human cell has 23 pairs of chromosomes: one set of chromosomes is inherited from the mother and the other set is inherited from the father. There is also a mitochondrial genome, inherited exclusively from the mother, which can be involved in inherited genetic disorders. On each chromosome, there are thousands of genes, sequences of DNA that code for a functional product, that are responsible for determining the genotype and phenotype of the individual. The human haploid genome contains 3 billion base pairs and has between 20,000 and 25,000 functional genes.

In order for DNA to serve its role as the genetic material, all organisms must be able to faithfully copy the entire genome.

This process, **DNA replication**, is the precursor to all forms of cell division.

14.1 | Historical Basis of Modern Understanding

Learning Objectives

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

- Explain transformation of DNA.
- Describe the key experiments that helped identify that DNA is the genetic material.
- State and explain Chargaff's rules

Modern understandings of DNA have evolved from the discovery of nucleic acids to the development of the double-helix model. In the 1860s, Friedrich Miescher (**Figure 14.2**), a physician by profession, was the first person to isolate phosphate- rich chemicals from white blood cells or leukocytes. He named these chemicals (which would

eventually be known as RNA and DNA) nuclein because they were isolated from the nuclei of the cells.

Figure 14.2 Friedrich Miescher (1844–1895) discovered nucleic acids.

A half century later, British bacteriologist Frederick Griffith was perhaps the first person to show that hereditary information could be transferred from one cell to another "horizontally," rather than by descent. In 1928, he reported the first demonstration of bacterial **transformation**, a process in which external DNA is taken up by a cell, thereby changing morphology and physiology. He was working with *Streptococcus pneumoniae,* the bacterium that causes pneumonia. Griffith worked with two strains, rough (R) and smooth (S). The R strain is non-pathogenic (does not cause disease) and is called rough because its outer surface is a cell wall and lacks a capsule; as a result, the cell surface appears uneven under the microscope. The S strain is pathogenic (disease-causing) and has a capsule outside its cell wall. As a result, it has a smooth appearance under the microscope. Griffith

injected the live R strain into mice and they survived. In another experiment, when he injected mice with the heat-killed S strain, they also survived. In a third set of experiments, a mixture of live R strain and heat-killed S strain were injected into mice, and—to his surprise—the mice died. Upon isolating the live bacteria from the dead mouse, only the S strain of bacteria was recovered. When this isolated S strain was injected into fresh mice, the mice died. Griffith concluded that something had passed from the heat-killed S strain into the live R strain and transformed it into the pathogenic S strain, and he called this the transforming principle (**Figure 11.3**). These experiments are now famously known as Griffith's transformation experiments.

Mouse injected with heat-killed virulant S strain lives.

Mouse injected with both heat-killed S strain and live non-virulant R strain dies.

Figure 14.3 Two strains of S.pneumoniae were used in Griffith's transformation experiments. The R strain is non-pathogenic. The S strain is pathogenic and causes death. When Griffith injected a mouse with the heat-killed S strain and a live R strain, the mouse died. The S strain was recovered from the dead mouse. Thus, Griffith concluded that something had passed from the heat-killed S strain to the R strain, transforming the R strain into S strain in the process. (credit "living mouse": modification of work by NIH; credit "dead mouse": modification of work by Sarah Marriage)

Scientists Oswald Avery, Colin MacLeod, and Maclyn McCarty (1944) were interested in exploring this transforming principle further. They isolated the S strain from the dead mice and isolated the proteins and nucleic acids, namely RNA and DNA, as these were possible candidates for the molecule of heredity. They conducted a systematic elimination study. They used enzymes that specifically degraded each component and then used each mixture separately to transform the R strain. They found that when DNA was degraded, the resulting mixture was no longer able to transform the bacteria, whereas all of the

other combinations were able to transform the bacteria. This led them to conclude that DNA was the transforming principle.

Experiments conducted by Martha Chase and Alfred Hershey in 1952 provided confirmatory evidence that DNA was the genetic material and not proteins. Chase and Hershey were studying a bacteriophage, which is a virus that infects bacteria. Viruses typically have a simple structure: a protein coat, called the capsid, and a nucleic acid core that contains the genetic material, either DNA or RNA. The bacteriophage infects the host bacterial cell by attaching to its surface, and then it injects its nucleic acids inside the cell. The phage DNA makes multiple copies of itself using the host machinery, and eventually the host cell bursts, releasing a large number of bacteriophages. Hershey and Chase labeled one batch of phage with radioactive sulfur, ^{35}S, to label the protein coat. Another batch of phage were labeled with radioactive phosphorus, ^{32}P. Because phosphorous is found in DNA, but not protein, the DNA and not the protein would be tagged with radioactive phosphorus.

Each batch of phage was allowed to infect the cells separately. After infection, the phage bacterial suspension was put in a blender, which caused the phage coat to be detached from the host cell. The phage and bacterial suspension was spun down in a centrifuge. The heavier bacterial cells settled down and formed a pellet, whereas

the lighter phage particles stayed in the supernatant (the liquid above the pellet). In the tube that contained phage labeled with ^{35}S, the supernatant contained the radioactively labeled phage, whereas no radioactivity was detected in the pellet. In the tube that contained the phage labeled with ^{32}P, the radioactivity was detected in the pellet that contained the heavier bacterial cells, and no radioactivity was detected in the supernatant. Hershey and Chase concluded that it was the phage DNA that was injected into the cell and carried information to produce more phage particles, thus providing evidence that DNA was the genetic material and not proteins (**Figure 14.4**).

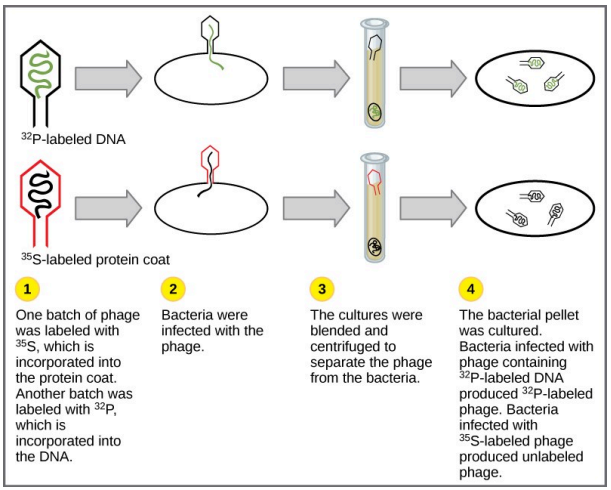


Figure 14.4 In Hershey and Chase's experiments, bacteria were infected with phage radiolabeled with either 35S, which labels protein, or 32P, which labels DNA. Only 32P entered the bacterial cells, indicating that DNA is the genetic material.

Around this same time, Austrian biochemist Erwin Chargaff examined the content of DNA in different

species and found that the amounts of adenine, thymine, guanine, and cytosine were not found in equal quantities, and that it varied from species to species, but not between individuals of the same species. He found that the amount of adenine equals the amount of thymine, and the amount of cytosine equals the amount of guanine, or A = T and G = C. These are also known as Chargaff's rules. This finding proved immensely useful when Watson and Crick were getting ready to propose their DNA double helix model, discussed in Chapter 5.

14.2 | Overview of DNA Replication

Learning Objectives

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

- Explain how the structure of DNA reveals the replication process.
- Describe the Meselson and Stahl experiments.

The elucidation of the structure of the double helix provided a hint as to how DNA divides and makes copies of itself. This model suggests that the two strands of the double helix separate during replication, and each strand serves as a template from which the new complementary strand is copied. What was not clear was how the replication took place. There were three models suggested (**Figure 14.5**): conservative, semi-conservative, and dispersive.

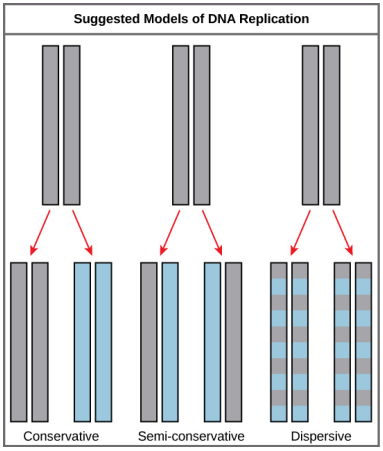


Figure 14.5 The three suggested models of DNA replication. Grey indicates the original DNA strands, and blue indicates newly synthesized DNA.

In conservative replication, the parental DNA remains together, and the newly formed daughter strands are together. The semi-conservative method suggests that each of the two parental DNA strands act as a template for new DNA to be synthesized; after replication, each double-stranded DNA includes one parental or "old" strand and one "new" strand. In the dispersive model, both copies of DNA have double-stranded segments of parental DNA and newly synthesized DNA interspersed. Meselson and Stahl were interested in

understanding how DNA replicates. They grew *E. coli* for several generations in a medium containing a "heavy" isotope of nitrogen (^{15}N) that gets incorporated into nitrogenous bases, and eventually into the DNA (**Figure 14.6**).

The *E. coli* culture was then shifted into medium containing ^{14}N and allowed to grow for one generation. The cells were harvested and the DNA was isolated. The DNA was centrifuged at high speeds in an ultracentrifuge. Some cells were allowed to grow for one more life cycle in ^{14}N and spun again. During the density graditent centrifugation, the DNA is loaded into a gradient (typically a salt such as cesium chloride or sucrose) and spun at high speeds of 50,000 to 60,000 rpm. Under these circumstances, the DNA will form a band according to its density in the gradient. DNA grown in ^{15}N will band at a higher density position than that grown in ^{14}N. Meselson and Stahl noted that after one generation of growth in ^{14}N after they had been shifted from ^{15}N, the single band observed was intermediate in position in between DNA of cells grown exclusively in ^{15}N and ^{14}N. This suggested either a semi-conservative or dispersive mode of replication. The DNA harvested from cells grown for two generations in ^{14}N formed two bands: one DNA band was at the intermediate position, between ^{15}N and ^{14}N and the other corresponded to the band of ^{14}N DNA. These results could only be explained if DNA replicates in a semi-conservative manner. Therefore, the other two modes were ruled out.

During DNA replication, each of the two strands that make up the double helix serves as a template from which new strands are copied. The new strand will be complementary to the parental or "old" strand. When two daughter DNA copies are formed, they have the same sequence and are divided equally into the two daughter cells.

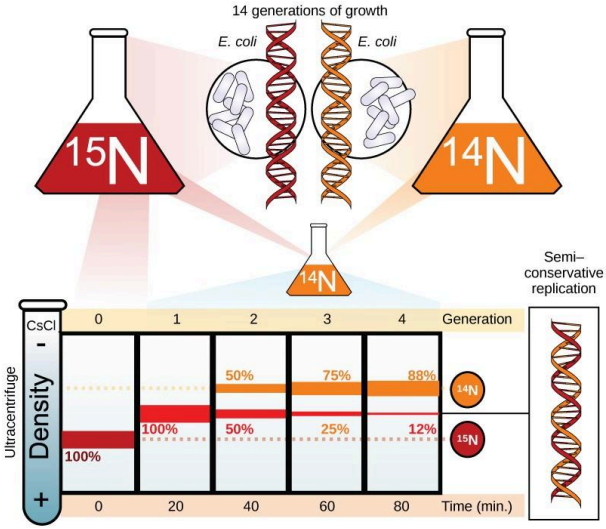


Figure 14.6 Meselson and Stahl experimented with E. coli grown first in heavy nitrogen (^{15}N) then in ^{14}N. DNA grown in ^{15}N (red band) is heavier than DNA grown in ^{14}N (orange band), and sediments to a lower level in cesium chloride solution in an ultracentrifuge. When DNA

14.3 | DNA Replication in Prokaryotes

grown in ^{15}N is switched to media containing ^{14}N, after one round of cell division the DNA sediments halfway between the ^{15}N and ^{14}N levels, indicating that it now contains fifty percent ^{14}N. In subsequent cell divisions, an increasing amount of DNA contains ^{14}N only. This data supports the semi-con

Learning Objectives

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

- Explain the process of DNA replication in prokaryotes.
- Discuss the role of different enzymes and proteins in supporting this process.

DNA replication has been extremely well studied in prokaryotes primarily because of the small size of the genome and the mutants that are 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 circle in both directions. This means that approximately 1000 nucleotides are added per

servative replicatio n model. (Credit: modificati on of work by Mariana Ruiz Villareal)

second. The process is quite rapid and occurs without many mistakes.

DNA replication employs a large number of proteins and enzymes, each of which plays a critical role during the process. One of the key players is the enzyme DNA polymerase, also known as DNA pol, which adds nucleotides one by one to the growing DNA chain that are complementary to the template strand. The addition of nucleotides requires energy; this energy is obtained from the nucleotides that have three phosphates attached to them, similar to ATP which has three phosphate groups attached. When the bond between the phosphates is broken, the energy released is used to form the phosphodiester bond between the incoming nucleotide and the growing chain. In prokaryotes, three main types of polymerases are known: DNA pol I, DNA pol II, and DNA pol III. It is now known that DNA pol III is the enzyme required for DNA synthesis; DNA pol I and DNA pol II are primarily required for repair.

How does the replication machinery know where to begin? It turns out that there are specific nucleotide sequences called origins of replication where replication begins. In E. coli, which has a single origin of replication on its one chromosome (as do most prokaryotes), it is approximately 245 base pairs long and is rich in AT sequences. The origin of replication is recognized by certain proteins that bind to this site. An enzyme called **helicase** unwinds the DNA by breaking the

hydrogen bonds between the nitrogenous base pairs. ATP hydrolysis is required for this process. As the DNA opens up, Y-shaped structures called **replication forks** are formed. Two replication forks are formed at the origin of replication and these get extended bi- directionally as replication proceeds. **Single-strand binding proteins** coat the single strands of DNA near the replication fork to prevent the single-stranded DNA from winding back into a double helix. DNA polymerase is able to add nucleotides only in the 5′ to 3′ direction (a new DNA strand can be only extended in this direction). It also requires a free 3′- OH group to which it can add nucleotides by forming a phosphodiester bond between the 3′-OH end and the 5′ phosphate of the next nucleotide. This essentially means that it cannot add nucleotides if a free 3′-OH group is not available. Then how does it add the first nucleotide? The problem is solved with the help of a primer that provides the free 3′-OH end. Another enzyme, RNA **primase**, synthesizes an RNA **primer** that is about five to ten nucleotides long and complementary to the DNA. Because this sequence primes the DNA synthesis, it is appropriately called the primer. DNA polymerase can now extend this RNA primer, adding nucleotides one by one that are complementary to the template strand (**Figure 14.7**).

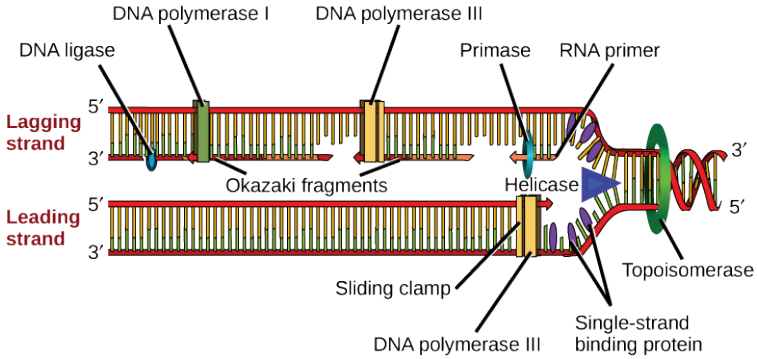


Figure 14.7 A replication fork is formed when helicase separates the DNA strands at the origin of replication. The DNA tends to become more highly coiled ahead of the replication fork. Topoisomerase breaks and reforms DNA's phosphate backbone ahead of the replication fork, thereby relieving the pressure that results from this supercoiling. Single-strand binding proteins bind to the single-stranded DNA to prevent the helix from re-forming. Primase synthesizes an RNA primer. DNA polymerase III uses this primer to synthesize the daughter DNA strand. On the leading strand, DNA is synthesized continuously, whereas on the lagging strand, DNA is synthesized in short stretches called Okazaki fragments. DNA polymerase I replaces the RNA primer with DNA. DNA ligase seals the gaps between the Okazaki fragments, joining the fragments into a single DNA molecule. (Credit: modification of work by Mariana Ruiz Villareal)

The replication fork moves at the rate of 1000 nucleotides per second. DNA polymerase can only extend in the 5′ to 3′ direction, which poses a slight problem at the replication fork. As we know, the DNA double helix is anti-parallel; that is, one strand is in the 5′ to 3′ direction and the other is oriented in

the 3′ to 5′ direction. One strand, which is complementary to the 3′ to 5′ parental DNA strand, is synthesized continuously towards the replication fork because the polymerase can add nucleotides in this direction. This continuously synthesized strand is known as the **leading strand**. The other strand, complementary to the 5′ to 3′ parental DNA, is extended away from the replication fork, in small fragments known as **Okazaki fragments**, each requiring a primer to start the synthesis. Okazaki fragments are named after the Japanese scientist who first discovered them. The strand with the Okazaki fragments is known as the **lagging strand**.

Concept check

You isolate a cell strain in which the joining together of Okazaki fragments is impaired and suspect that a mutation has occurred in an enzyme found at the replication fork. Which enzyme is most likely to be mutated?

The leading strand can be extended by one primer alone, whereas the lagging strand needs a new primer for each of the short Okazaki fragments. The overall direction of the lagging

strand will be 3′ to 5′, and that of the leading strand 5′ to 3′. A protein called the **sliding clamp** holds the DNA polymerase in place as it continues to add nucleotides. The sliding clamp is a ring-shaped protein that binds to the DNA and holds the polymerase in place. **Topoisomerase** prevents the over-winding of the DNA double helix ahead of the replication fork as the DNA is opening up; it does so by causing temporary nicks in the DNA helix and then resealing it. As synthesis proceeds, the RNA primers are replaced by DNA. The primers are removed by the exonuclease activity of DNA pol I, and the gaps are filled in by deoxyribonucleotides. The nicks that remain between the newly synthesized DNA (that replaced the RNA primer) and the previously synthesized DNA are sealed by the enzyme DNA **ligase** that catalyzes the formation of phosphodiester linkage between the 3′-OH end of one nucleotide and the 5′ phosphate end of the other fragment.

Once the chromosome has been completely replicated, the two DNA copies move into two different cells during cell division. The process of DNA replication can be summarized as follows:

1. DNA unwinds at the origin of replication.
2. Helicase opens up the DNA-forming replication forks; these are extended bidirectionally.
3. Single-strand binding proteins coat the DNA around the replication fork to prevent rewinding of the DNA.
4. Topoisomerase binds at the region ahead of the

replication fork to prevent supercoiling.

5. Primase synthesizes RNA primers complementary to the DNA strand.
6. DNA polymerase starts adding nucleotides to the 3′-OH end of the primer.
7. Elongation of both the lagging and the leading strand continues.
8. RNA primers are removed by exonuclease activity.
9. Gaps are filled by DNA pol by adding dNTPs.
10. The gap between the two DNA fragments is sealed by DNA ligase, which helps in the formation of phosphodiester bonds.

Table 11.1 Prokaryotic DNA replication: enzymes and their functions.

Enzyme/ Protein	**Specific Function**
DNA pol I	Exonuclease activity removes RNA primer and replaces with newly synthesized DNA
DNA pol II	Repair function
DNA pol III	Main enzyme that adds nucleotides in the 5′-3′ direction
Helicase	Opens the DNA helix by breaking hydrogen bonds between the nitrogenous bases
Ligase	Seals the gaps between the Okazaki fragments to create one continuous DNA strand
Primase	Synthesizes RNA primers needed to start replication
Sliding Clamp	Helps to hold the DNA polymerase in place when nucleotides are being added
Topoisomerase	Helps relieve the stress on DNA when unwinding by causing breaks and then resealing the DNA
Single-strand binding proteins (SSB)	Binds to single-stranded DNA to avoid DNA rewinding back.

14.4 | DNA Replication in Eukaryotes

Learning Objectives

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

- Discuss the similarities and differences between DNA replication in eukaryotes and prokaryotes.
- State the role of telomerase in DNA replication.

14.4.1 Prokaryote vs. Eukaryote Replication

Eukaryotic genomes are much more complex and larger in size than prokaryotic genomes. The human genome has three billion base pairs per haploid set of chromosomes, and 6 billion base pairs are replicated during the S phase of the cell cycle. There are multiple origins of replication on the

eukaryotic chromosome; humans can have up to 100,000 origins of replication. The rate of replication is approximately 100 nucleotides per second, much slower than prokaryotic replication. In yeast, which is a eukaryote, special sequences known as Autonomously Replicating Sequences (ARS) are found on the chromosomes. These are equivalent to the origin of replication in E. coli.

The number of DNA polymerases in eukaryotes is much more than prokaryotes: 14 are known, of which five are known to have major roles during replication and have been well studied. They are known as pol α, pol β, pol γ, pol δ, and pol ε.

The essential steps of replication are the same as in prokaryotes. Before replication can start, the DNA has to be made available as template. Eukaryotic DNA is bound to basic proteins known as histones to form structures called nucleosomes. The chromatin (the complex between DNA and proteins) may undergo some chemical modifications, so that the DNA may be able to slide off the proteins or be accessible to the enzymes of the DNA replication machinery. At the origin of replication, a pre-replication complex is made with other initiator proteins. Other proteins are then recruited to start the replication process (**Table 11.2**).

A helicase using the energy from ATP hydrolysis opens up the DNA helix. Replication forks are formed at each replication origin as the DNA unwinds. The opening of the double helix causes over-winding, or supercoiling, in the DNA ahead of the replication fork. These are resolved with the

action of topoisomerases. Primers are formed by the enzyme primase, and using the primer, DNA pol can start synthesis. While the leading strand is continuously synthesized by the enzyme pol δ, the lagging strand is synthesized by pol ε. A sliding clamp protein known as PCNA (Proliferating Cell Nuclear Antigen) holds the DNA pol in place so that it does not slide off the DNA. RNase H removes the RNA primer, which is then replaced with DNA nucleotides. The Okazaki fragments in the lagging strand are joined together after the replacement of the RNA primers with DNA. The gaps that remain are sealed by DNA ligase, which forms the phosphodiester bond.

Table 11.2 Differences between Prokaryotic and Eukaryotic Replication

Property	Prokaryotes	Eukaryotes
Origin of replication	Single	Multiple
Rate of replication	1000 nucleotides/s	50 to 100 nucleotides/s
DNA polymerase types	5	14
Telomerase	Not present	Present
RNA primer removal	DNA pol I	RNase H
Strand elongation	DNA pol III	Pol δ, pol ε
Sliding clamp	Sliding clamp	PCNA

14.4.2 Telomere Replication

Unlike prokaryotic chromosomes, eukaryotic chromosomes are linear. As you've learned, the enzyme DNA pol can add nucleotides only in the 5′ to 3′ direction. In the leading strand, synthesis continues until the end of the chromosome is reached. On the lagging strand, DNA is synthesized in short stretches, each of which is initiated by a separate primer. When the replication fork reaches the end of the linear chromosome, there is no place for a primer to be made for the DNA fragment to be copied at the end of the chromosome. These ends thus remain unpaired, and over time these ends may get progressively shorter as cells continue to divide.

The ends of the linear chromosomes are known as **telomeres**, which have repetitive sequences that code for no particular gene. In a way, these telomeres protect the genes from getting deleted as cells continue to divide. In humans, a six base pair sequence, TTAGGG, is repeated 100 to 1000 times. The discovery of the enzyme **telomerase** (**Figure 14.8**) helped in the understanding of how chromosome ends are maintained. The telomerase enzyme contains a catalytic part and a built-in RNA template. It attaches to the end of the chromosome, and complementary bases to the RNA template are added on the 3′ end of the DNA strand. Once the 3′ end of the lagging strand template is sufficiently elongated, DNA polymerase can add the nucleotides complementary to the

ends of the chromosomes. Thus, the ends of the chromosomes are replicated.

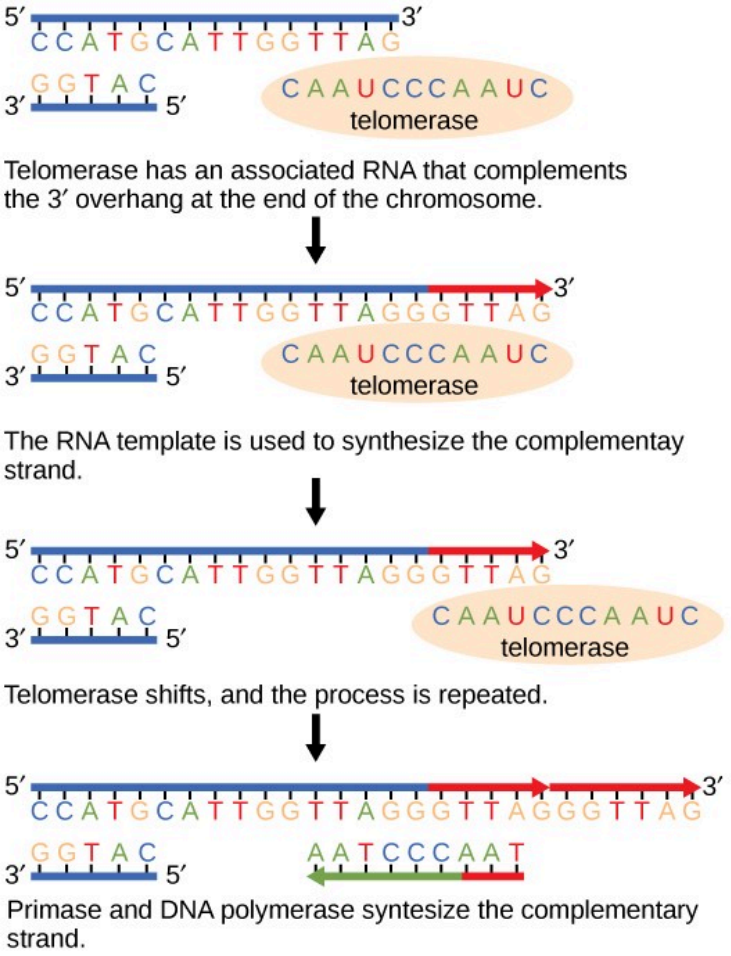


Figure 14.8 The ends of linear chromosomes are maintained by the action of the telomerase enzyme.

Figure 14.9 Elizabeth Blackburn, 2009 Nobel Laureate, is the scientist who discovered how telomerase works. (Credit: US Embassy Sweden)

Telomerase is typically active in germ cells and adult stem cells. It is not active in adult somatic cells. For her discovery of telomerase and its action, Elizabeth Blackburn (**Figure 14.9**) received the Nobel Prize for Medicine and Physiology in 2009.

Telomerase and Aging

Cells that undergo cell division continue to have their telomeres shortened because most somatic cells do not make telomerase. This essentially means that telomere shortening is associated with aging. With the advent of modern medicine, preventative health care, and healthier lifestyles, the human life span has increased, and there is an increasing demand for people to look younger and have a better quality of life as they grow older.

In 2010, scientists found that telomerase can reverse some age-related conditions in mice. 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 extension of telomeres, reduced DNA damage, reversed neurodegeneration, and improved the function of the testes, spleen, and intestines. Thus, telomere reactivation may have potential for treating age-related diseases in humans.

Cancer is characterized by uncontrolled cell division of abnormal cells. The cells accumulate mutations, proliferate uncontrollably, and can migrate to different parts of the body through a process called metastasis. Scientists have observed that cancerous cells have considerably shortened telomeres and that telomerase is active in these cells. Interestingly, only after the telomeres were shortened in the cancer cells did the telomerase become active. If the action of telomerase in these cells can be inhibited by drugs during cancer therapy, then the cancerous cells could potentially be stopped from further division.

1. Jaskelioff et al., "Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice," Nature 469 (2011): 102-7.

14.5 | DNA Repair

Learning Objectives

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

- Discuss the different types of mutations in DNA.
- Explain DNA repair mechanisms.

DNA replication is a highly accurate process, but mistakes can occasionally occur, such as a DNA polymerase inserting a wrong base. Uncorrected mistakes may sometimes lead to serious consequences, such as cancer. Repair mechanisms correct the mistakes. In rare cases, mistakes are not corrected, leading to mutations; in other cases, repair enzymes are themselves mutated or defective.

Most of the mistakes during DNA replication are promptly corrected by DNA polymerase by proofreading the base that has been just added (**Figure 14.10**). In **proofreading**, the DNA pol reads the newly added base before adding the next one, so a correction can be made. The polymerase checks

whether the newly added base has paired correctly with the base in the template strand. If it is the right base, the next nucleotide is added. If an incorrect base has been added, the enzyme makes a cut at the phosphodiester bond and releases the wrong nucleotide. This is performed by the exonuclease action of DNA pol III. Once the incorrect nucleotide has been removed, a new one will be added again.

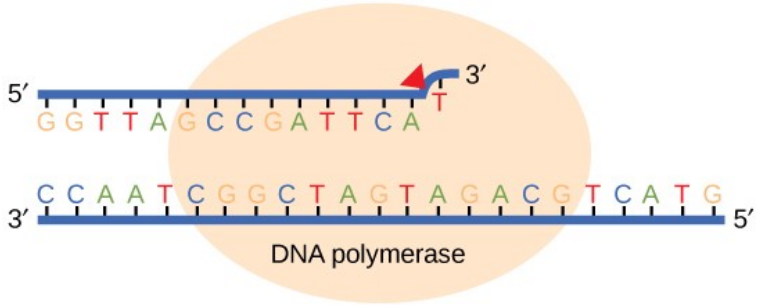


Figure 14.10 Proofreading by DNA polymerase corrects errors during replication.

Some errors are not corrected during replication, but are instead corrected after replication is completed; this type of repair is known as **mismatch repair** (**Figure 14.11**). The enzymes recognize the incorrectly added nucleotide and excise it; this is then replaced by the correct base. If this remains uncorrected, it may lead to more permanent damage. How do mismatch repair enzymes recognize which of the two bases is the incorrect one? In *E. coli*, after replication, the nitrogenous base adenine acquires a methyl

group; the parental DNA strand will have methyl groups, whereas the newly synthesized strand lacks them. Thus, DNA polymerase is able to remove the wrongly incorporated bases from the newly synthesized, non-methylated strand. In eukaryotes, the mechanism is not very well understood, but it is believed to involve recognition of unsealed nicks in the new strand, as well as a short-term continuing association of some of the replication proteins with the new daughter strand after replication has completed.

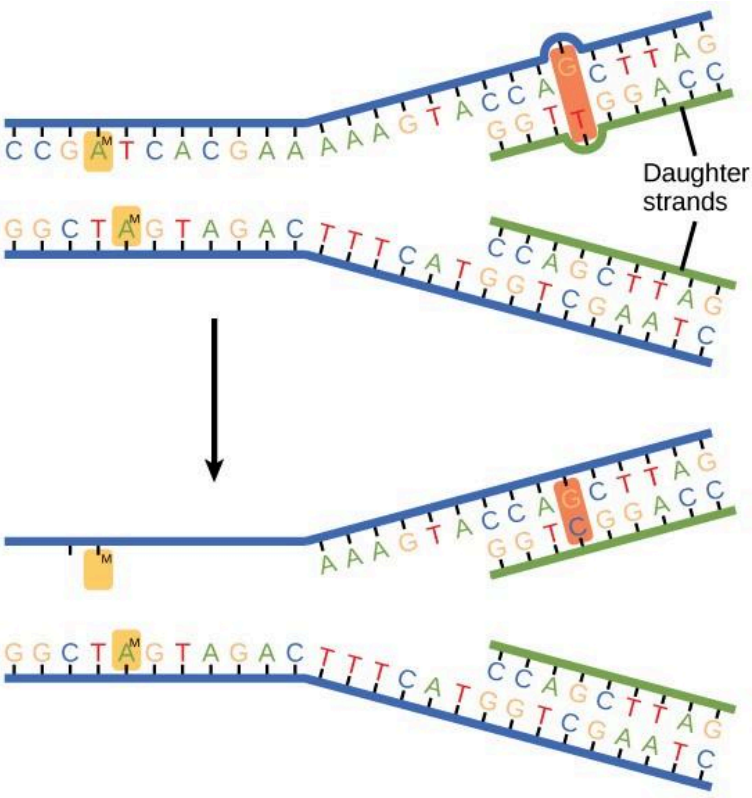


Figure 14.11 In mismatch repair, the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base.

In another type of repair mechanism, **nucleotide excision repair**, enzymes replace incorrect bases by making a cut on both the 3′ and 5′ ends of the incorrect base (**Figure 14.12**). The segment of DNA is removed and replaced with the correctly paired nucleotides by the action of DNA pol.

Once the bases are filled in, the remaining gap is sealed with a phosphodiester linkage catalyzed by DNA ligase. This repair mechanism is often employed when UV exposure causes the formation of pyrimidine dimers.

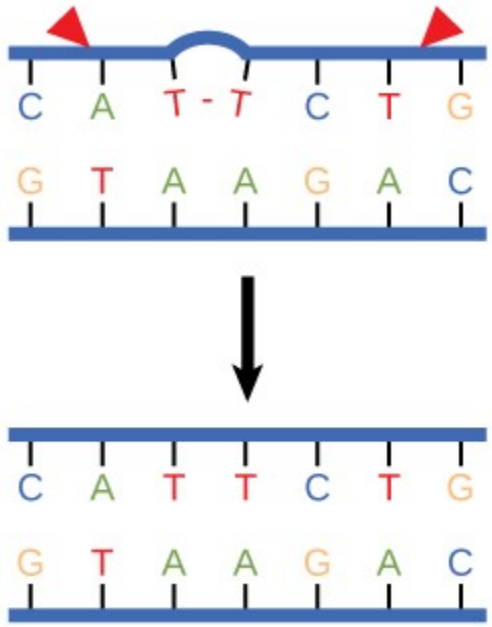


Figure 14.12 Nucleotide excision repairs thymine dimers. When exposed to UV, thymines lying adjacent to each other can form thymine dimers. In normal cells, they are excised and replaced.

A well-studied example of mistakes not being corrected is seen in people suffering from xeroderma pigmentosa (**Figure 14.13**). Affected individuals have skin that is highly sensitive to UV rays from the sun. When individuals are exposed to

UV, pyrimidine dimers, especially those of thymine, are formed; people with xeroderma pigmentosa are not able to repair the damage. These are not repaired because of a defect in the nucleotide excision repair enzymes, whereas in normal individuals, the thymine dimers are excised and the defect is corrected. The thymine dimers distort the structure of the DNA double helix, and this may cause problems during DNA replication. People with xeroderma pigmentosa may have a higher risk of contracting skin cancer than those who dont have the condition.

Figure 14.13 Xeroderma pigmentosa is a condition in which thymine dimerization from exposure to UV is not repaired. Exposure to sunlight results in skin lesions. (Credit: James Halpern et al.)

Errors during DNA replication are not the only reason why mutations arise in DNA. **Mutations**, variations in the nucleotide sequence of a genome, can also occur because of 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 without any exposure to any environmental agent; they are a result of natural reactions taking place within the body.

Mutations may have a wide range of effects. Some mutations are not expressed; these are known as **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 base, 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 also known as **translocation**. Some of these mutation types are shown in **Figure 14.14.**

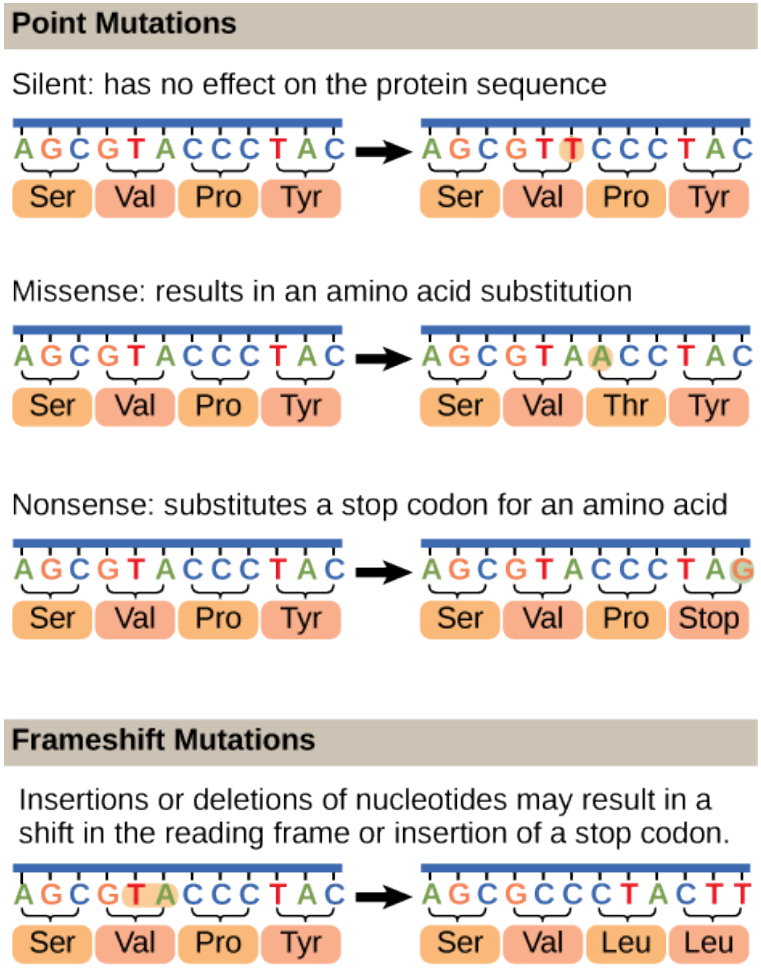


Figure 14.14 Mutations can lead to changes in the protein sequence encoded by the DNA.

Mutations in repair genes have been known to cause cancer. Many mutated repair genes have been implicated in certain

forms of pancreatic cancer, colon cancer, and colorectal cancer. Mutations can affect either somatic cells or germ cells. If many mutations accumulate in a somatic cell, they may lead to problems such as the uncontrolled cell division observed in cancer. If a mutation takes place in germ cells, the mutation will be passed on to the next generation, as in the case of hemophilia and xeroderma pigmentosa.

CHAPTER 15. MEIOSIS & SEXUAL REPRODUCTION

Figure 15.1 Each of us, like these other large multicellular organisms, begins life as a fertilized egg. After trillions of cell divisions, each of us develops into a complex, multicellular organism. (Credit a: modification of work by Frank Wouters; credit b: modification of work by Ken Cole, USGS; credit c: modification of work by Martin Pettitt)

Chapter Outline

- 15.1 The Process of Meiosis
- 15.2 Disorders in Chromosome Number
- 15.3 Sexual Reproduction

Introduction

The ability to reproduce in kind is a basic characteristic of all living things. In kind means that the offspring of any organism closely resemble their parents. Hippopotamuses give birth to hippopotamus calves, Joshua trees produce seeds from which Joshua tree seedlings emerge, and adult flamingos lay eggs that hatch into flamingo chicks. In kind does not generally mean exactly the same. Whereas many unicellular organisms and a few multicellular organisms can produce genetically identical clones of themselves through cell division, many single-celled organisms and most multicellular organisms reproduce sexually. **Sexual reproduction** is the production by parents of two haploid cells and the fusion of two haploid cells to form a single, unique diploid cell. In most plants and animals, through tens of rounds of mitotic cell division, this diploid cell will develop into an adult organism. Haploid cells that are part of the sexual reproductive cycle are produced by a type of cell division called **meiosis**. The vast majority of eukaryotic

organisms employ some form of meiosis and fertilization to reproduce. The evolutionary success of sexual reproduction may result from the genetic variation that it introduces into a species' genome.

15.1 | The Process of Meiosis

Learning Objectives

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

- Describe the behavior of chromosomes during meiosis.
- Describe cellular events during meiosis.
- Explain the differences between meiosis and mitosis.
- Explain the mechanisms within meiosis that generate genetic variation among the products of 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. Cells containing one set of chromosomes are called **haploid**. Cells containing two sets of chromosomes are called **diploid**. If the reproductive cycle is to continue, then diploid cells must somehow reduce the number of chromosome sets before fertilization can occur again, or the number of chromosome will double in every generation. So, sexual reproduction requires a nuclear division that reduces the number of chromosome sets.

Most animals and plants are diploid. In each **somatic** (non-reproductive) cell of the organism, the nucleus contains two copies of each chromosome. The two copies of a chromosome are called **homologous chromosomes**, or homologs. Homologous means "similar"; homologous chromosomes are not identical, since they come from different parents. However, homologous chromosomes are matched pairs that contain the same genes in identical locations along their length. Diploid organisms inherit one copy of each homologous chromosome from each parent. For example, your body cells each contain one set of 23 chromosomes from your mother and one set of 23 chromosomes from your mother. Chromosome 1 from your mother and chromosome 1 from your father are homologous to each other.

Haploid cells are found only within structures that give rise to reproductive cells, such as gametes or spores. **Spores** are haploid cells that can produce a haploid organism or can fuse with another spore to form a diploid cell. All animals and most

plants produce eggs and sperm, or **gametes**. Some plants and all fungi produce spores.

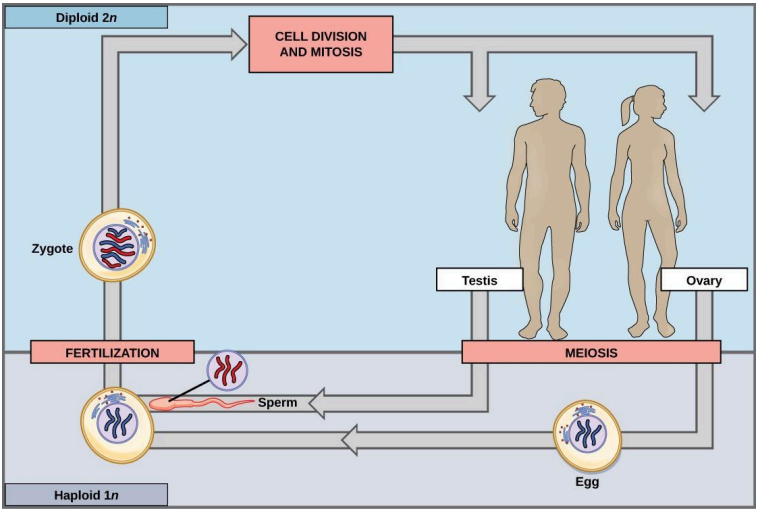


Figure 15.2 In animals, sexually reproducing adults form haploid gametes from diploid germ cells. Fusion of the gametes gives rise to a fertilized egg cell, or zygote. The zygote will undergo multiple rounds of mitosis to produce a multicellular offspring. The germ cells are generated early in the development of the zygote.

Meiosis is the nuclear division that forms haploid cells. Meiosis employs many of the same mechanisms as mitosis. However, whereas mitosis produces two genetically identical diploid daughter cells, meiosis produces four genetically different haploid daughter cells. To achieve this reduction in chromosome number, meiosis consists of 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 "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.

15.1.1 Meiosis I

Meiosis is preceded by an interphase consisting of the G1, S, and G2 phases, which are nearly identical to the phases preceding mitosis. The G1 phase, which is also called the first gap phase, is the first phase of the interphase and is focused on cell growth. The S phase is the second phase of interphase, during which the DNA of the chromosomes is replicated. Finally, the G2 phase, also called the second gap phase, is the third and final phase of interphase; in this phase, the cell undergoes the final preparations for meiosis.

During DNA duplication in the S phase, each chromosome is replicated to produce two identical copies, called sister chromatids, that are held together at the centromere by **cohesin** proteins. Cohesin holds the chromatids together until anaphase II. The centrosomes, which are the structures that organize the microtubules of the meiotic spindle, also

replicate. This prepares the cell to enter prophase I, the first meiotic phase.

Prophase I

Early in prophase I, before the chromosomes can be seen clearly microscopically, the homologous chromosomes are attached at their tips to the nuclear envelope by proteins. As the nuclear envelope begins to break down, the proteins associated with homologous chromosomes bring the pair close to each other. Recall that in mitosis homologous chromosomes do not pair with each other. In mitosis, homologous chromosomes line up end-to-end, held together by a lattice of proteins called the **synaptonemal complex**. This 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. (**Figure 15.3**).

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.

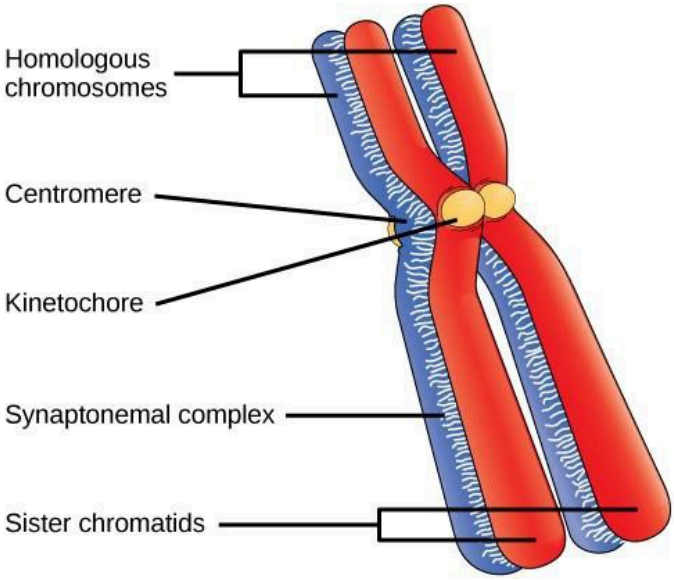


Figure 15.3 Early in prophase I, homologous chromosomes come together to form a synapse. The chromosomes are bound tightly together and in perfect alignment by a protein lattice called a synaptonemal complex and by cohesin proteins at the centromere.

The synaptonemal complex allows the exchange of chromosomal segments between homologous chromatids, a process called **crossing over** or **crossover**. Crossing over can be observed visually under a microscope as **chiasmata** (singular = chiasma) (**Figure 15.4**). 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.

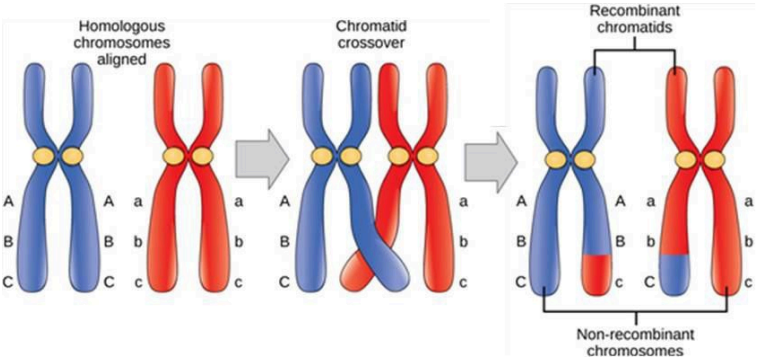


Figure 15.4 Crossover occurs between non-sister chromatids of homologous chromosomes. The result is an exchange of genetic material between homologous chromosomes.

As prophase I progresses, the synaptonemal complex begins to break down and the chromosomes begin to condense. At the end of prophase I, the pairs are held together only at the chiasmata and are called **tetrads** because the four sister chromatids of each pair of homologous chromosomes are now visible.

Crossover is the first source of genetic variation produced by meiosis. Crossover between a pair of homologous chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. 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. This combination of maternal and paternal genes did not exist before the crossover.

Prometaphase I

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. The microtubules move toward the middle of the cell and attach to one of the two fused homologous chromosomes. 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. 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 tetrads.

This is important in determining the genes carried by a gamete, as each will only receive one of the two homologous chromosomes.

The random orientation of homologous chromosomes creates the second form of genetic variation in offspring. Recall that homologous chromosomes contain slight differences in their genetic information. In each cell that undergoes meiosis, the arrangement of the tetrads is different. Since humans have 23 chromosome pairs, there are over eight million 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 (**Figure 15.5**).

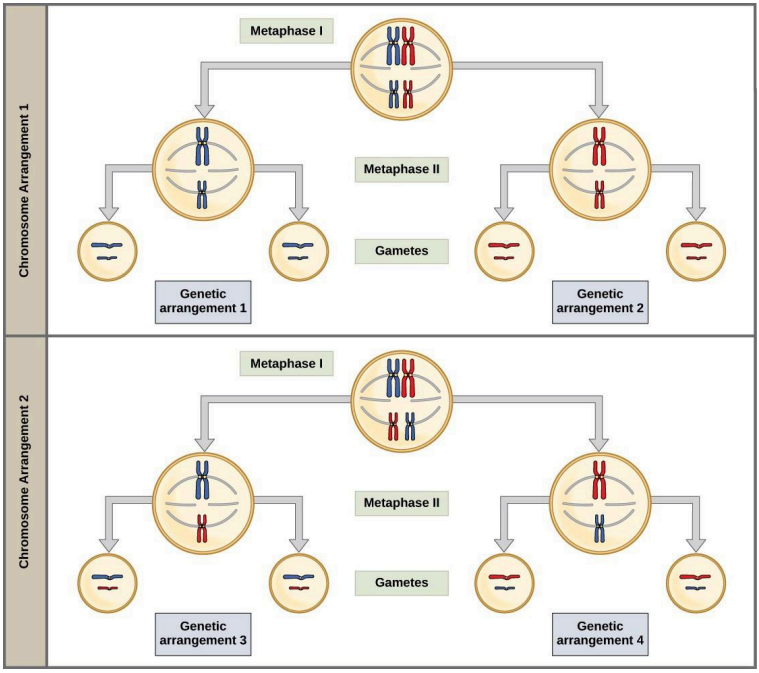


Figure 15.5 Random, independent assortment during metaphase I can be demonstrated in a cell with a set of two chromosomes (n = 2). The total possible number of different gametes is 2n, where n equals the number of chromosomes in a set. In this example, there are four possible genetic combinations.

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 (**Figure 15.6**).

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. However, 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.

15.1.2 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 and Prometaphase 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. 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. Note that during meiosis I, the homologous chromosomes are pulled apart, while in meiosis II, sister chromatids are pulled apart (**Figure 15.6**).

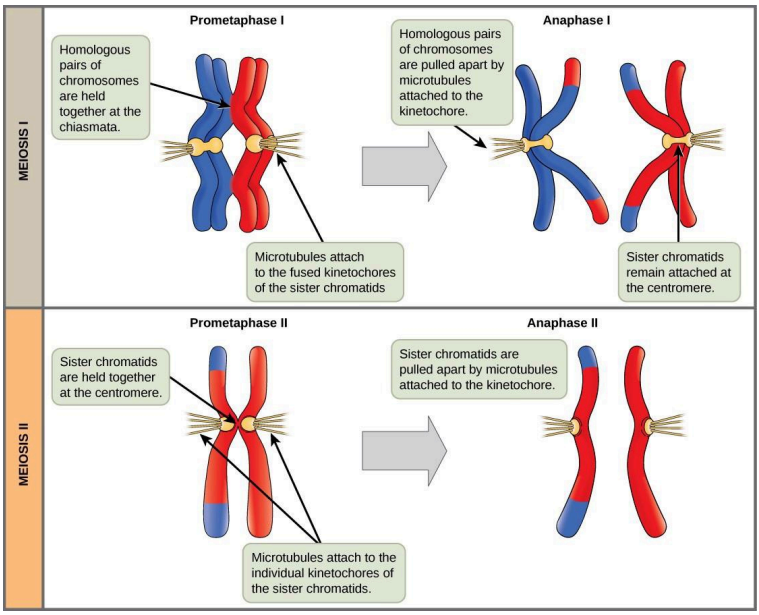


Figure 15.6 In anaphase I, homologous chromosomes are separated. In anaphase II, sister chromatids are separated.

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 **Figure 15.7**.

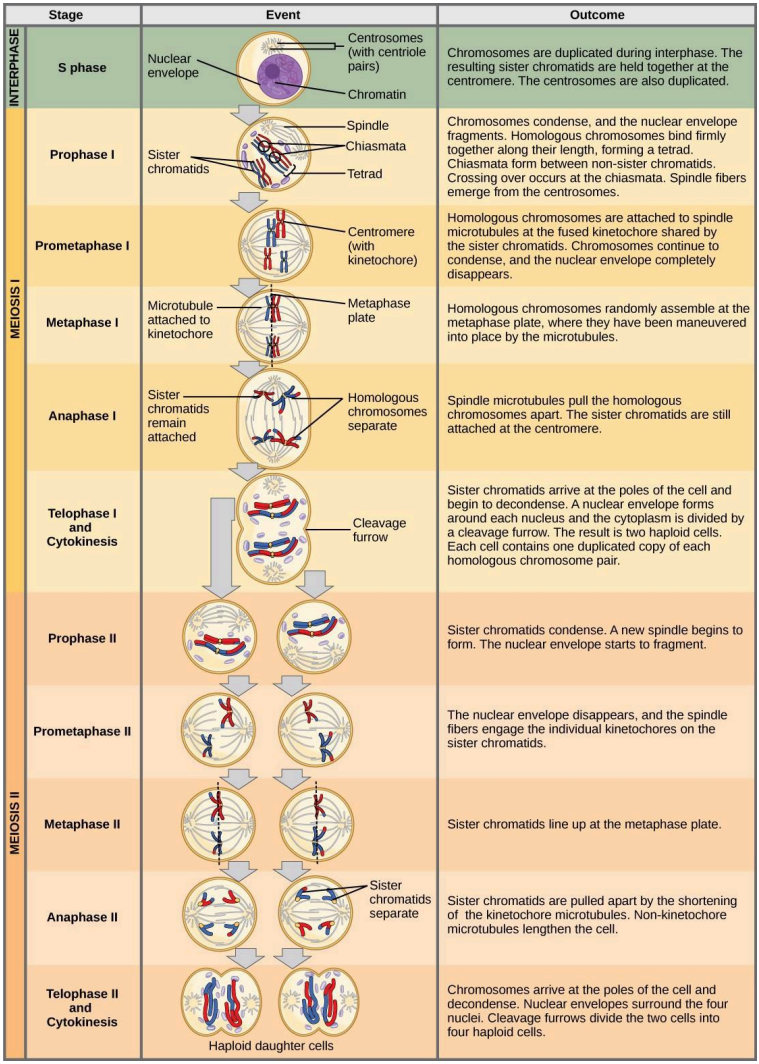


Figure 15.7 An animal cell with a diploid number of four (2n = 4) proceeds through the stages of meiosis to form four haploid daughter cells.

15.1.3 Comparing Meiosis and

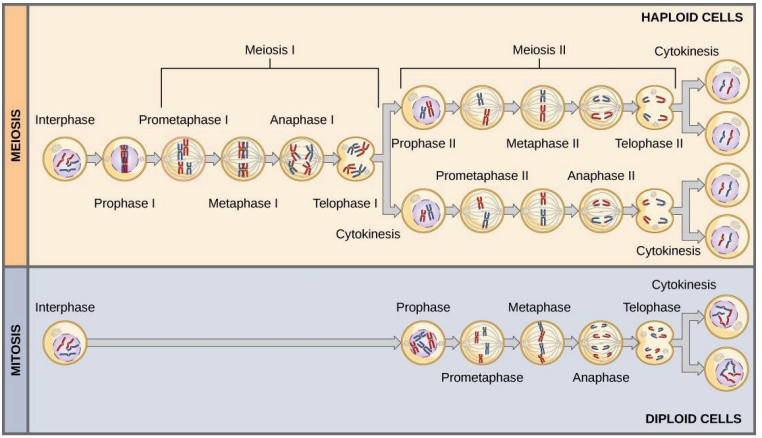
Mitosis

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 (**Figure 15.8**). 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. In most plants and all animal species, diploid cells typically 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 diploid cell.

The main differences between mitosis and meiosis occur in meiosis I. In meiosis I, homologous chromosome pairs are bound together, undergo crossover, and line up randomly along the metaphase plate. When the homologs separate, the ploidy level is 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 similar to mitosis. Duplicated chromosomes (only one set of them) line up on the metaphase plate and sister chromatids are separated to opposite poles.

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 (**Figure 15.8**).



						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

Figure 15.8 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.

15.2 | Disorders in Chromosome Number

Learning Objectives

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

- Describe how a karyotype is made and used.
- Explain how nondisjunction occurs during meiosis.
- Describe common types of aneuploidy in humans.
- Explain why X chromosome nondisjunction causes less serious effects.
- Describe common chromosome structural rearrangements

15.2.1 Identification of Chromosomes

A **karyotype** is the number and appearance of chromosomes,

and includes their length, banding pattern, and centromere position. To obtain a view of an individual's karyotype, cytologists photograph the chromosomes and then cut and paste each chromosome into a chart, or karyogram (**Figure 15.9**).

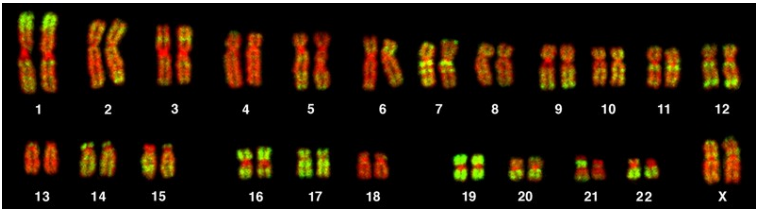


Figure 15.9 This karyotype is of a female human. Notice that homologous chromosomes are the same size, and have the same centromere positions and banding patterns. A human male would have an XY chromosome pair instead of the XX pair shown (Credit: Andreas Blozer et al.).

career CONNECTION

Geneticists Use Karyograms to Identify Chromosomal Aberrations

Although Mendel is referred to as the "father of modern genetics," he performed his experiments with none of the tools that the geneticists of

today routinely employ. One such powerful cytological technique is karyotyping, a method in which traits characterized by chromosomal abnormalities can be identified from a single cell. To observe an individual's karyotype, a person's cells (such as white blood cells) are first collected from a blood sample or other tissue. In the laboratory, the isolated cells are stimulated to begin actively dividing. A chemical called colchicine is then applied to cells to arrest condensed chromosomes in metaphase. Cells are then made to swell using a hypotonic solution so the chromosomes spread apart. Finally, the sample is preserved in a fixative and applied to a slide.

The geneticist then stains chromosomes with one of several dyes to better visualize the distinct and reproducible banding patterns of each chromosome pair. Following staining, the chromosomes are viewed using bright-field microscopy. A common stain choice is the Giemsa stain. Giemsa staining results in approximately 400–800 bands (of tightly coiled DNA and condensed proteins) arranged along all of the 23 chromosome pairs; an experienced

geneticist can identify each band. Chromosomes are further identified on the basis of size and centromere location. The geneticist obtains a digital image, identifies each chromosome, and manually arranges the chromosomes into a pattern (**Figure 15.8**).

At its most basic, the karyogram may reveal genetic abnormalities in which an individual has too many or too few chromosomes per cell. Examples of this are Down Syndrome, which is identified by a third copy of chromosome 21, and Turner Syndrome, which is characterized by the presence of only one X chromosome in women instead of the normal two. Geneticists can also identify large deletions or insertions of DNA. For instance, Jacobsen Syndrome—which involves distinctive facial features as well as heart and bleeding defects—is identified by a deletion on chromosome 11. Finally, the karyotype can pinpoint **translocations**, which occur when a segment of genetic material breaks from one chromosome and reattaches to another chromosome or to a different part of the same chromosome. Translocations are implicated in

certain cancers, including chronic myelogenous leukemia.

During Mendel's lifetime, inheritance was an abstract concept that could only be inferred by performing crosses and observing the traits expressed by offspring. By observing a karyogram, today's geneticists can actually visualize the chromosomal composition of an individual to confirm or predict genetic abnormalities in offspring, even before birth.

In a given species, chromosomes can be identified by their number, size, centromere position, and banding pattern. In a human karyotype, **autosomes** (all of the non–sex chromosomes) are organized from largest (chromosome 1) to smallest (chromosome 22). The X and Y chromosomes are not autosomes. The chromosome "arms" projecting from either end of the centromere may be designated as short or long, depending on their relative lengths. The short arm is abbreviated p (for "petite"), whereas the long arm is abbreviated q (because it follows "p" alphabetically). Each arm is further subdivided and denoted by a number. Using this naming system, locations on chromosomes can be described consistently in the scientific literature.

15.2.2 Nondisjunction during Meiosis Leads to Aneuploidy

Of all of the chromosomal disorders, abnormalities in chromosome number are the most obviously identifiable. Disorders of chromosome number include the duplication or loss of entire chromosomes, as well as changes in the number of complete sets of chromosomes. They are caused by **nondisjunction**, which occurs when pairs of homologous chromosomes or sister chromatids fail to separate during meiosis. Misaligned or incomplete synapsis, or a dysfunction of the spindle apparatus that facilitates chromosome migration, can cause nondisjunction. The risk of nondisjunction increases with the age of the parents.

Nondisjunction can occur during either meiosis I or II (**Figure 15.10**). If homologous chromosomes fail to separate during meiosis I, the result is two gametes that lack that particular chromosome and two gametes with two copies of the chromosome. If sister chromatids fail to separate during meiosis II, the result is one gamete that lacks that chromosome, two normal gametes with one copy of the chromosome, and one gamete with two copies of the chromosome.

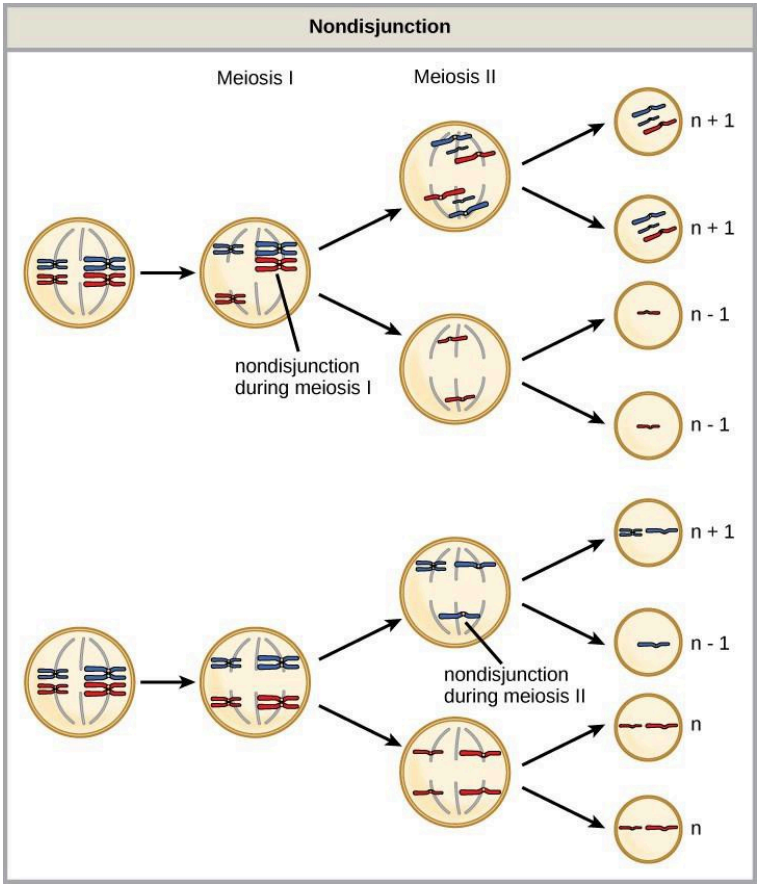


Figure 15.10 Nondisjunction occurs when homologous chromosomes or sister chromatids fail to separate during meiosis, resulting in an abnormal chromosome number. Nondisjunction may occur during meiosis I or meiosis II.

Aneuploidy

An individual with the appropriate number of chromosomes for their species is called **euploid**; in humans, euploidy

corresponds to 22 pairs of autosomes and one pair of sex chromosomes. An individual with an error in chromosome number is described as **aneuploid**.

If a gamete that lacks a chromosome due to nondisjunction during meiosis is fertilized, the resulting individual will have only one copy of that chromosome, a condition referred to as **monosomy**. If a gamete with an extra copy of a chromosome is fertilized, the resulting individual will have three copies of that chromosome, a condition referred to as **trisomy.**

Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they lack essential genes. Most autosomal trisomies also fail to develop to birth; however, trisomy of some of the smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years. Trisomic individuals suffer from an excess in gene dose, as they may synthesize too much of the gene products encoded by that chromosome. This extra dose of specific genes can lead to a number of functional challenges and often precludes development. The most common trisomy among viable births is that of chromosome 21, which corresponds to Down Syndrome. Individuals with this inherited disorder are characterized by short stature and stunted digits, facial distinctions that include a broad skull and large tongue, and significant developmental delays. The incidence of Down syndrome is correlated with maternal age; older women are

more likely to become pregnant with fetuses carrying the trisomy 21 genotype (**Figure 15.11**).

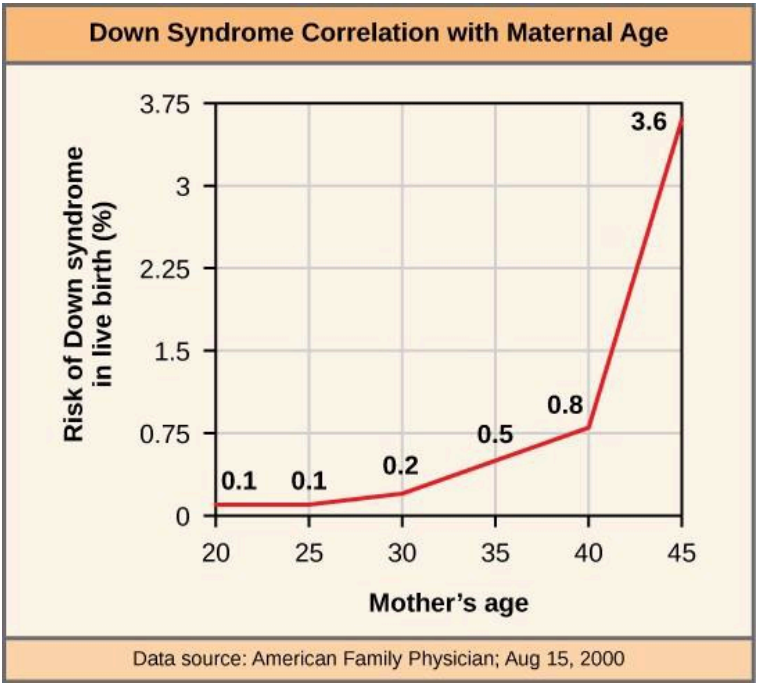


Figure 15.11 The incidence of having a fetus with trisomy 21 increases dramatically with maternal age.

Polyploidy

An individual with more than the correct number of chromosome sets is called **polyploid**. For instance, fertilization of an abnormal diploid egg with a normal haploid sperm would yield a triploid zygote. Polyploid animals are extremely

rare and are usually sterile because meiosis cannot proceed normally. Rarely, polyploid animals can reproduce asexually. In contrast, polyploidy is very common in the plant kingdom, and polyploid plants tend to be larger and more robust than euploids of their species (**Figure 15.12**).

Figure 15.12 As with many polyploid plants, this triploid orange daylily (Hemerocallis fulva) is particularly large and robust, and grows flowers with triple the number of petals of its diploid counterparts (Credit: Steve Karg).

15.2.3 Sex Chromosome Nondisjunction in Humans

In contrast to autosomal trisomies and monosomies, humans

can often function normally with different numbers of the X chromosome. In part, this occurs because of a process called **X inactivation**. Early in development, when female mammalian embryos consist of just a few thousand cells (relative to trillions in the newborn), one X chromosome in each cell inactivates by tightly condensing into a dormant structure called a Barr body. The chance that an X chromosome (maternally or paternally derived) is inactivated in each cell is random, but once the inactivation occurs, all cells derived from that one will have the same inactive X chromosome. By this process, females compensate for their double genetic dose of X chromosome. In so-called "tortoiseshell" cats, embryonic X inactivation is observed as color variegation (**Figure 12.13**). Females that are heterozygous for an X-linked coat color gene will express one of two different coat colors over different regions of their body, corresponding to whichever X chromosome is inactivated in the embryonic cell progenitor of that region.

Figure 15.13 In cats, the gene for coat color is located on the X chromosome. In the embryonic development of female cats, one of the two X chromosomes is randomly inactivated in each cell, resulting in a tortoiseshell pattern if the cat has two different alleles for coat color. Normal male cats, having only one X chromosome, never exhibit a tortoiseshell coat color. (Credit: Michael Bodega)

X Chromosome Aneuploidy in Humans

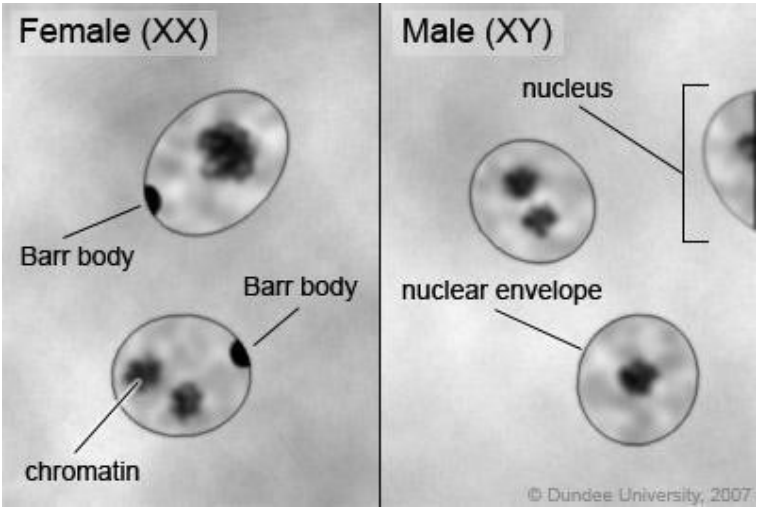


Figure 15.14 In human females, one X chromosome is inactivated, forming a Barr body. Since males only have one X chromosome, they do not have Barr bodies.

An individual carrying an abnormal number of X chromosomes will inactivate all but one X chromosome in each of her cells (**Figure 15.14**). However, even inactivated X chromosomes continue to express a few genes, and X chromosomes must reactivate for the proper maturation of female ovaries. As a result, X- chromosomal abnormalities are typically associated with mild mental and physical defects, as well as sterility.

Several errors in sex chromosome number have been characterized. Individuals with three X chromosomes (XXX) are phenotypically female but express developmental delays and reduced fertility. The XXY genotype, corresponding to

one type of Klinefelter syndrome, corresponds to phenotypically male individuals with small testes, enlarged breasts, and reduced body hair. More complex types of Klinefelter syndrome exist in which the individual has as many as five X chromosomes. In all types, every X chromosome except one undergoes inactivation to compensate for the excess genetic dosage. This can be seen as several Barr bodies in each cell nucleus. Turner syndrome, characterized as an X0 genotype (i.e., only a single sex chromosome), corresponds to a phenotypically female individual with short stature, webbed skin in the neck region, hearing and cardiac impairments, and sterility.

Concept check

Do women with Turner syndrome have Barr bodies?

15.2.4 Duplications and Deletions

In addition to the loss or gain of an entire chromosome, a chromosomal segment may be duplicated or lost. Duplications and deletions often produce offspring that survive but exhibit

physical and mental abnormalities. Duplicated chromosomal segments may fuse to existing chromosomes or may be free in the nucleus. Cri-du-chat (from the French for "cry of the cat") is a syndrome associated with nervous system abnormalities and identifiable physical features that result from a deletion of most of the small arm of chromosome 5 (**Figure 15.15**). Infants with this genotype emit a characteristic high-pitched cry on which the disorder's name is based.

Figure 15.15 This individual with cri-du-chat syndrome is shown at two, four, nine, and 12 years of age. (Credit: Paola Cerruti Mainardi)

15.2.5 Chromosomal Structural Rearrangements

Cytologists have characterized numerous structural

rearrangements in chromosomes, but chromosome inversions and translocations are the most common.

Chromosome Inversions

A **chromosome inversion** is the detachment, 180° rotation, and reinsertion of part of a chromosome. Unless they disrupt a gene sequence, inversions only change the orientation of genes and are likely to have mild effects. However, altered gene orientation can result in functional changes because regulators of gene expression could be moved out of position with respect to their targets, causing aberrant levels of gene products.

An inversion can be pericentric and include the centromere, or paracentric and occur outside of the centromere (**Figure 15.16**). A pericentric inversion that is asymmetric about the centromere can change the relative lengths of the chromosome arms, making these inversions easily identifiable.

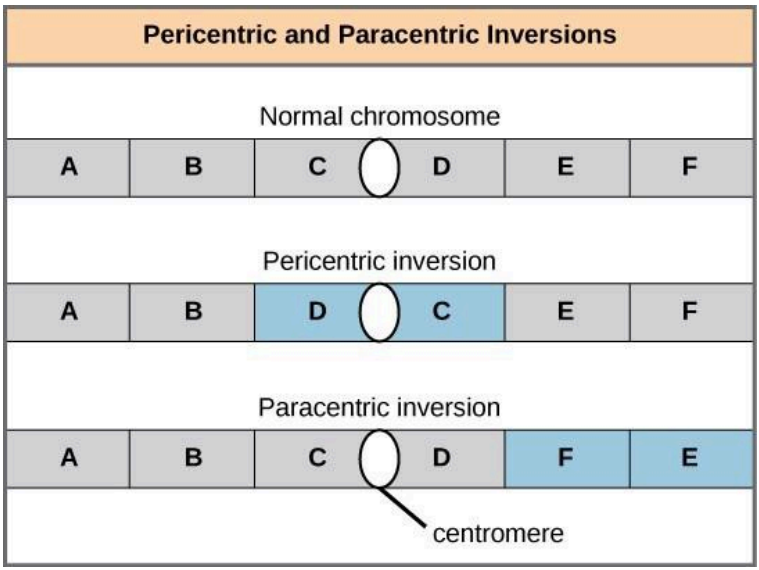


Figure 15.16 Pericentric inversions include the centromere, and paracentric inversions do not.

Translocations

A **translocation** occurs when a segment of a chromosome dissociates and reattaches to a different chromosome. Translocations can be benign or have devastating effects depending on how the positions of genes are altered with respect to regulatory sequences. Notably, specific translocations have been associated with several cancers and with schizophrenia. Reciprocal translocations result from the exchange of chromosome segments between two nonhomologous chromosomes such that there is no gain or loss of genetic information (**Figure 15.17**).

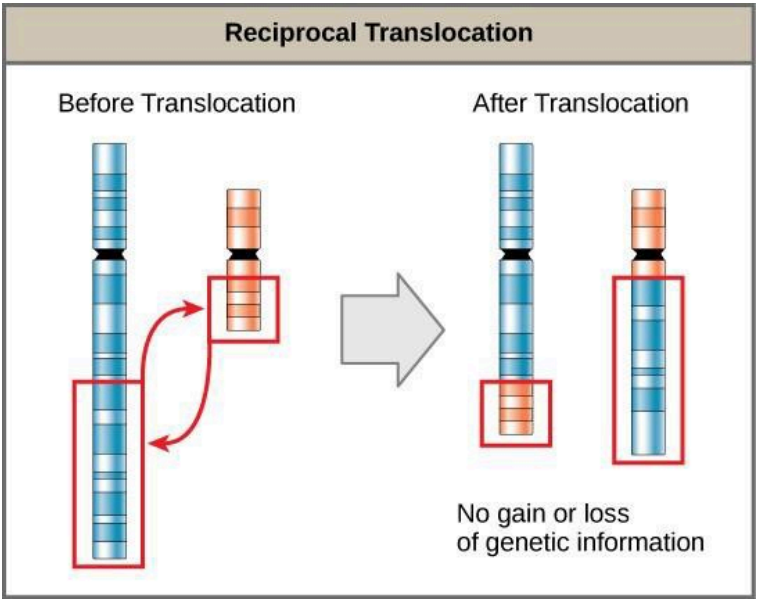


Figure 15.17 A reciprocal translocation occurs when a segment of DNA is transferred from one chromosome to another, nonhomologous chromosome. (Credit: modification of work by National Human Genome Research/USA)

evolution CONNECTION

The Chromosome 18 Inversion

Not all structural rearrangements of chromosomes produce noviable, impaired, or

infertile individuals. In rare instances, such a change can result in the evolution of a new species. In fact, a pericentric inversion in chromosome 18 appears to have contributed to the evolution of humans. This inversion is not present in our closest genetic relatives, the chimpanzees. Humans and chimpanzees differ by pericentric inversions on several chromosomes and by the fusion of two separate chromosomes in chimpanzees that correspond to chromosome two in humans.

The pericentric chromosome 18 inversion if believed to have occurred in early humans following their divergence from a common ancestor with chimpanzees approximately five million years ago. Researchers characterizing this inversion have suggested that approximately 19,000 nucleotide bases were duplicated on 18p, and the duplicated region inverted and reinserted on chromosome 18 of an ancestral human.

A comparison of human and chimpanzee genes in the region of this inversion indicates that two genes – *ROCK1* and *USP14* – that are adjacent on

chimpanzee chromosome 17 (which corresponds to human chromosome 18) are more distantly positioned on human chromosome 18. This suggests that one of the inversion breakpoints occurred between these two genes. Interestingly, humans and chimpanzees express *USP14* at distinct levels in specific cell types, including cortical cells and fibroblasts. Perhaps the chromosome 18 inversion in ancestral humans repositioned specific genes and reset their expression levels in a useful way. Because both *ROCK1* and *USP14* encode cellular enzymes, a change in their expression could alter cellular function. It is not known how this inversion contributed to hominid evolution, but it appears to be a significant factor in the divergence of humans from other primates.[1]

1. Violaine Goidts et al., "Segment duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates," Human Genetics. 115 (2004): 116-122.

15.3 | Sexual Reproduction

Learning Objectives

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

- Explain that meiosis and sexual reproduction are evolved traits.
- Identify variation among offspring as a potential evolutionary advantage to sexual reproduction.
- Describe the three different life-cycle types among sexual multicellular organisms and their commonalities.

Sexual reproduction was an early evolutionary innovation after the appearance of eukaryotic cells. It appears to have been very successful because most eukaryotes are able to reproduce sexually, and in many animals, it is the only mode of reproduction. And yet, scientists recognize some real disadvantages to sexual reproduction.

On the surface, creating offspring that are genetic clones of the parent appears to be a better system. If the parent organism

is successfully occupying a habitat, offspring with the same traits would be similarly successful. There is also the obvious benefit of not requiring another organism of the opposite sex. Indeed, 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. Since males in sexual populations d o not p r o d u c e offspring, in theory an asexual population could grow twice as fast.

However, multicellular organisms that exclusively depend on asexual reproduction are exceedingly rare. Why is sexuality (and meiosis) so common? This is one of the important unanswered questions in biology and has been the focus of much research beginning in the latter half of the twentieth century. There are several possible explanations, one of which is that the variation that sexual reproduction creates among offspring is very important to the survival and reproduction of the population. The only source of variation in asexual organisms is mutation. This is the ultimate source of variation in sexual organisms, but in addition, those different mutations are continually reshuffled from one generation to the next when different parents combine their unique genomes and the genes are mixed into different combinations by crossovers during prophase I and random assortment at metaphase I.

15.3.1 Life Cycles of Sexually Reproducing Organisms

Fertilization and meiosis alternate in sexual **life cycles**. There are three main categories of life cycles in multicellular organisms: diploid-dominant, haploid-dominant, and alternation of generation.

In **diploid-dominant** organisms, including most animals, the multicellular diploid stage is the most obvious life stage. In animals, the only haploid cells are the gametes. Early in the development of the embryo, specialized diploid cells, called **germ cells**, are produced within the gonads, such as the testes and ovaries. Germ cells are capable of mitosis to perpetuate the cell line and meiosis to produce gametes. Once the haploid gametes are formed, they lose the ability to divide again. There is no multicellular haploid life stage. Fertilization occurs with the fusion of two gametes, usually from different individuals, restoring the diploid state (**Figure 15.2**).

In **haploid-dominant** organisms, including fungi and some algae, the multicellular haploid stage is the most obvious life stage. The haploid cells that make up the tissues of the dominant multicellular stage are formed by mitosis. During sexual reproduction, specialized haploid cells from two individuals, designated the (+) and (−) mating types, join to form a diploid zygote. The zygote immediately undergoes meiosis to form four haploid cells called spores. Although haploid like the "parents," these spores contain a new genetic

combination from two parents. The spores can remain dormant for various time periods. Eventually, when conditions are conducive, the spores form multicellular haploid structures by many rounds of mitosis (**Figure 15.18**).

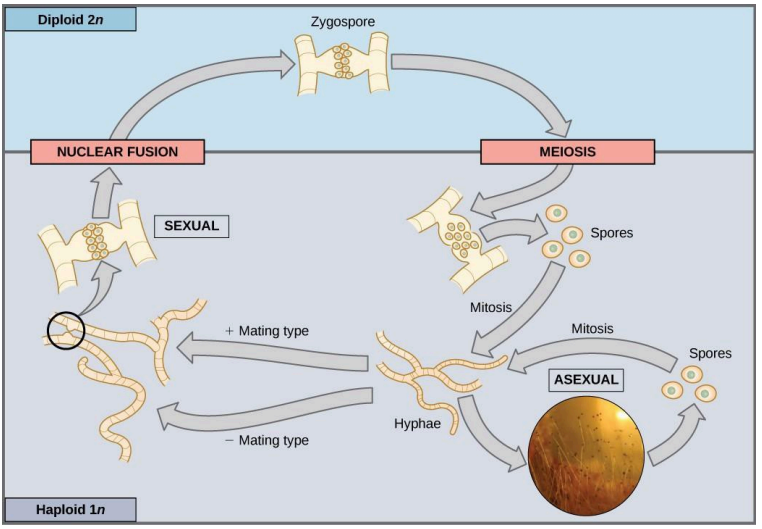


Figure 15.18 Fungi, such as black bread mold (*Rhizopus nigricans*), have haploid-dominant life cycles. (Credit "zygomycota" micrograph: modification of work by "Fanaberka"/Wikimedia Commons.)

Organisms that show **alternation of generations**, including plants and some algae, have both haploid and diploid multicellular organisms as part of their life cycle. The haploid multicellular plants are called **gametophytes**, because they produce gametes from specialized cells. Meiosis is not directly involved in the production of gametes in this case, because

the organism that produces the gametes is already a haploid. Fertilization between the gametes forms a diploid zygote. The zygote will undergo many rounds of mitosis and give rise to a diploid multicellular plant called a **sporophyte**. Specialized cells of the sporophyte will undergo meiosis and produce haploid spores. The spores will subsequently develop into the gametophytes (**Figure 15.19**).

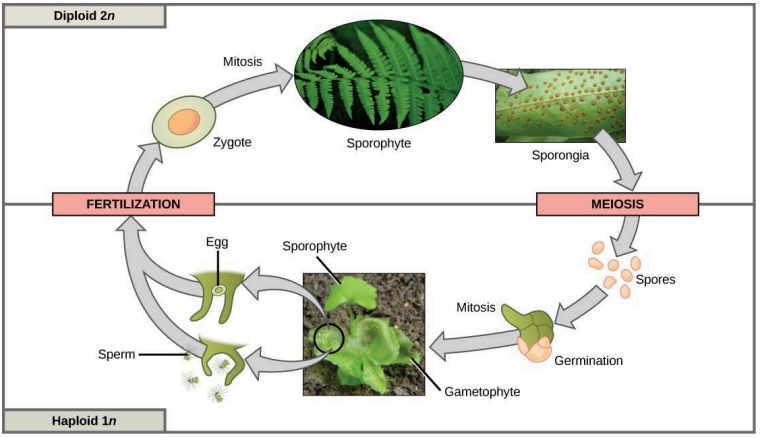


Figure 15.19 Plants have a life cycle that alternates between a multicellular haploid organism and a multicellular diploid organism. (Credit "fern": modification of work by Cory Zanker; credit "sporangia": modification of work by "Obsidian Soul"/Wikimedia Commons; credit "gametophyte and sporophyte": modification of work by "Vlmastra"/ Wikimedia Commons.)

In summary, sexual reproduction takes many forms in multicellular organisms. However, at some point in each type of life cycle, meiosis produces haploid cells that will fuse with

the haploid cell of another organism. The mechanisms of variation—crossover, random assortment of homologous chromosomes, and random fertilization—are present in all versions of sexual reproduction. The fact that nearly every multicellular organism on Earth employs sexual reproduction is strong evidence for the benefits of producing offspring with unique gene combinations, though there are other possible benefits as well.

evolution CONNECTION

The Red Queen Hypothesis

It is not in dispute that sexual reproduction provides evolutionary advantages to organisms that employ this mechanism to produce offspring. But why, even in the face of fairly stable conditions, does sexual reproduction persist when it is more difficult and costly for individual organisms? Variation is the outcome of sexual reproduction, but why are ongoing variations necessary? Enter the Red Queen hypothesis, first proposed by Leigh Van Valen in 1973. The concept was named in reference to the Red Queen's race inLewis Carroll's book, Through the Looking-Glass.

All species co-evolve with other organisms; for example predators evolve with their prey, and parasites evolve with their hosts. 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 co-evolving species to maintain its own share of the resources is to also continually improve its fitness. As one species gains an advantage, this increases selection on the other species; they must also develop an advantage or they will be outcompeted. No single species progresses too far ahead because genetic variation among the progeny of sexual reproduction provides all species with a mechanism to improve rapidly. Species that 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 co-evolution between competing species.

1.

CHAPTER 16. THE CENTRAL DOGMA: GENES TO TRAITS

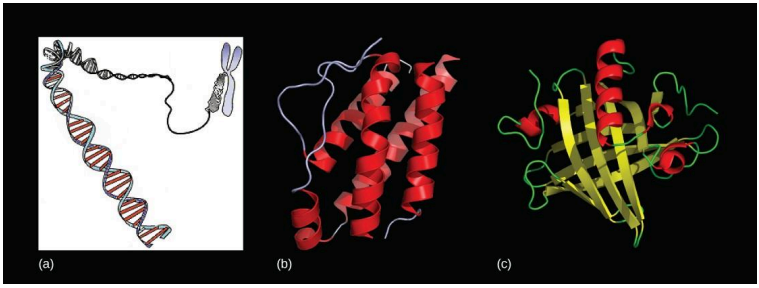


Figure 16.1 Genes, which are carried on (a) chromosomes, are instructions for making the RNA and protein molecules that are necessary for all of processes of life. The (b) interleukin-2 protein and (c) alpha-2u-globulin protein are just two examples of human proteins that are encoded by genes. (Credit "chromosome: National Human Genome Research Institute; credit "interleukin-2": Ramin Herati/ Created from PDB 1M47 and rendered with Pymol; credit "alpha-2u-globulin": Darren Logan/rendered with AISMIG)

Chapter Outline

- 16.1 The Genetic Code
- 16.2 Transcription
- 16.3 RNA Processing in Eukaryotes
- 16.4 Translation

Introduction

Proteins are responsible for orchestrating nearly every cell function. Therefore, proteins also influence most of the traits displayed by multicellular organisms. So, how do cells know how to make the hundreds or thousands of different proteins that they need?

DNA contained in cells is a blueprint for making proteins (**Figure 16.1**). Each chromosome is a single molecule of DNA wound tightly around proteins. Segments of the DNA molecule, called genes, code for the order of amino acids in each protein. Both genes and the proteins they encode are absolutely essential to life as we know it.

16.1 | The Genetic Code

Learning Objectives

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

- Explain the Central Dogma of molecular biology.
- Describe the genetic code.
- Explain how the nucleotide sequence of a gene prescribes the amino acid sequence of a protein.
- Identify various types of point mutations and chromosomal mutations.

DNA contains the genes that serve as the blueprint for making proteins. Each protein has a unique sequence of amino acids. Since different amino acids have different chemistries (such as acidic vs. basic, or polar vs. nonpolar), the order of amino acids in a protein determines how the protein folds, thereby determining the shape and function

of the protein. The order of amino acids in a protein is coded for by the gene for that protein.

16.1 The Central Dogma: DNA Encodes mRNA and mRNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the **Central Dogma of molecular biology** (**Figure 16.2**). When a cell needs a particular protein, the gene that codes for that protein is activated and a single-stranded mRNA copy is made of the gene, in a process called **transcription**. The code copied into the mRNA is then used to determine the order of amino acids in the protein, in a process called **translation**. The copying of DNA to RNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every nucleotide read in the DNA strand. The translation to protein is a bit more complex because three mRNA nucleotides correspond to one amino acid in the polypeptide sequence (**Figure 16.2**).

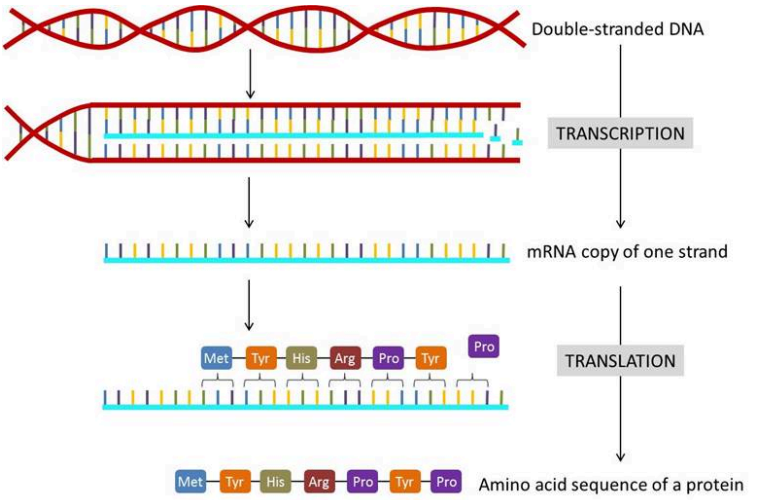


Figure 16.2 The central dogma of molecular biology. Segments of DNA, called genes, are transcribed into mRNA copies. mRNA is then "read" in three-nucleotide codons to specify the order of amino acids in a protein.

The Genetic Code Is Universal and Redundant

How does the order of nucleotides in an mRNA specify the order of amino acids in a protein? mRNA is "read" in three nucleotide segments called **codons**. Since RNA has four nucleotides (A, C, U, and G), there are 64 (4^3) possible combinations of three nucleotides (**Figure 16.3**). 61 of these codons code for one of the 20 common amino acids. The other three are called **stop codons** or **nonsense codons** because they do not code for an amino acid.

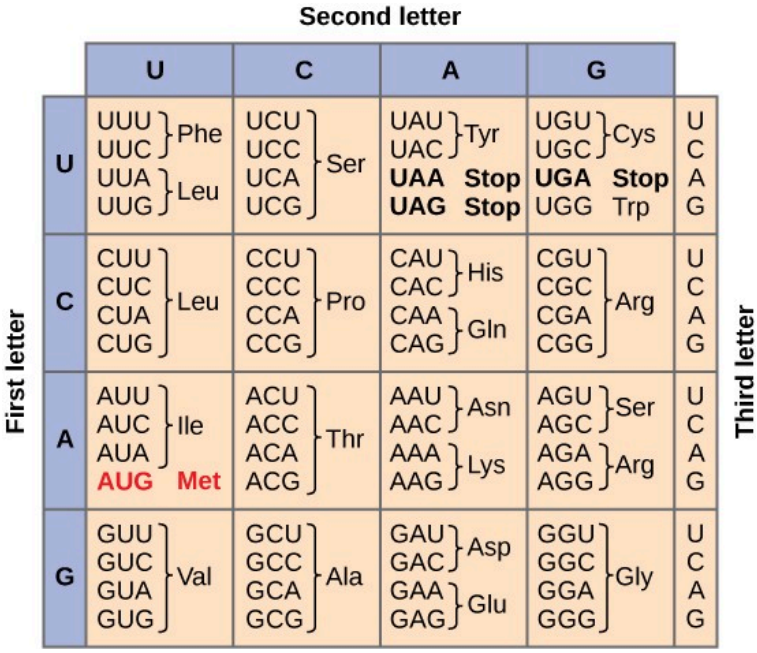


Figure 16.3 The genetic code allows cells to translate each nucleotide triplet in mRNA into an amino acid or a termination signal in a protein. (credit: modification of work by NIH)

Scientists painstakingly solved the genetic code by translating synthetic mRNAs in vitro and sequencing the proteins they specified (**Figure 16.4**). Once all of the codons were known, they discovered some important features of the code.

The genetic code is **universal**. With a few exceptions, virtually all species use the same genetic code for protein

synthesis. Conservation of codons means that a purified mRNA encoding the globin protein in horses could be transferred to a tulip cell, and the tulip would synthesize horse globin. That there is only one genetic code is powerful evidence that all of life on Earth shares a common origin

Since there are more nucleotide triplets than there are amino acids, the genetic code is **redundant**. In other words, a given amino acid can be encoded by more than one nucleotide triplet. Redundancy reduces the negative impact of random mutations. Codons that specify the same amino acid typically only differ by one nucleotide, usually the third one. For example, ACU, ACC, ACA and ACG all code for the amino acid threonine. In addition, amino acids with chemically similar side chains are encoded by similar codons. For example, UGU and UGC code for the amino acid cysteine, while AGU and AGC code for the amino acid serine. Cysteine and serine both have polar side chains that are very similar in size and other properties. Thus, the redundancy of the genetic code ensures that a single- nucleotide substitution mutation might specify either the same amino acid or a similar amino acid, preventing the protein from being rendered completely nonfunctional.

While 61 of the 64 codons specify the addition of a specific amino acid to a polypeptide chain, the remaining three codons terminate protein synthesis and release the

polypeptide from the translation machinery. These triplets are called **nonsense codons**, or **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.

Elucidating the Genetic Code

Given the different numbers of "letters" in the mRNA and protein "alphabets," scientists theorized that combinations of nucleotides corresponded to single amino acids. Nucleotide doublets would not be sufficient to specify every amino acid because there are only 16 possible two-nucleotide combinations (4^2). In contrast, there are 64 possible nucleotide triplets (4^3). The fact that amino acids were encoded by nucleotide triplets was confirmed experimentally by Francis Crick and Sydney Brenner. They inserted one, two, or three nucleotides into the gene of a virus. When one or two nucleotides were inserted, the protein was not made. When three nucleotides were inserted, the protein was synthesized and functional. This demonstrated that three nucleotides specify each amino acid.The nucleotide triplets that code for amino acids are called **codons**. The insertion of one or two nucleotides completely changed the triplet r**eading frame**, thereby altering the message for every subsequent amino acid (**Figure 16.4**). Though insertion of

three nucleotides caused an extra amino acid to be inserted during translation, the integrity of the rest of the protein was maintained.

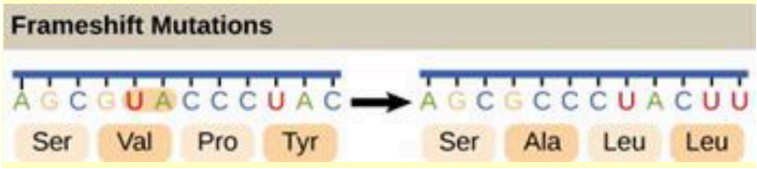


Figure 16.4 The deletion of two nucleotides shifts the reading frame of an mRNA and changes the entire protein message, creating a nonfunctional protein or terminating protein synthesis altogether.

Exceptions to the Central Dogma

Many genes code for RNA molecules that do not function as mRNAs and are therefore not translated into proteins. Some RNAs, called rRNA, form parts of the ribosomes. Others form transfer RNAs, or tRNA, which help with translation. Still others can regulate which genes are expressed.

Another exception to the central dogma is in some cases, information flows backwards as is seen in certain viruses called retroviruses.

These viruses have genes made up of RNA and when retroviruses infect a cell, the virus has to synthesize a DNA version of the RNA genes using a specialized viral polymerase called reverse transcriptase. The human immunodeficiency virus (HIV), which causes AIDS, is a retrovirus and many of the prescribed drugs used for AIDS patients target the HIV reverse transcriptase.

16.2 Mutations

In the living cell, DNA undergoes frequent chemical changes, especially as it is being replicated. These changes, result from incorrect nucleotides being inserted into the coding region of a gene. Most of these changes are quickly repaired. Those that are not repaired result in a **mutation**: a heritable change in the DNA.

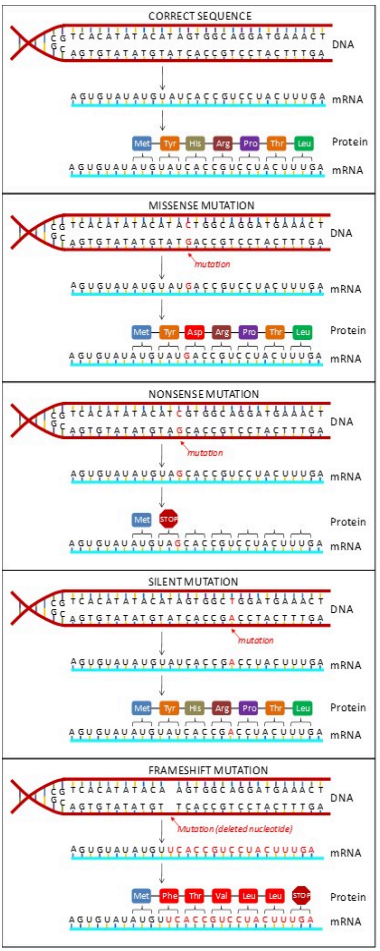


Figure 16.5 Examples of point mutations and their effects on mRNA and protein sequences. Incorrect nucleotides and amino acids are shown in red.

Point Mutations

If a single base is changed in the DNA sequence, say from an A to a G, it results in a **point mutation**. A point mutation can have drastic consequences at the protein level depending on which base was substituted. Different types of point

mutations are named based on their effect on the resulting protein. (**Figure 16.5**)

Missense Mutation

A **missense mutation** causes a different amino acid to be inserted in the protein. Even a single incorrect amino acid can have dire effects on an organism. For example, sickle-cell disease in humans results from the replacement of A by T at the 17th nucleotide of a hemoglobin gene, thereby substituting valine (GTG) for glutamic acid (GAG).

Nonsense Mutation

A **nonsense mutation** changes a codon that specified an amino acid to a stop codon. Translation of the resulting mRNA will therefore stop prematurely. The earlier in the gene that this mutation occurs, the shorter the protein will be and the more likely that it will be unable to function.

Silent Mutation

A **silent mutation** results in changing a codon to another codon that encodes the same amino acid. This is possible because the genetic code is redundant. Since this type of mutation does not change the protein product, it cannot be detected unless the gene is sequenced.

Frameshift Mutation

A **frameshift mutation** results when extra base pairs are added (insertion) or removed (deletion) from a gene. The resulting shift in the reading frame can have devastating consequences (**Figure 16.4**). Frameshifts can alter every subsequent codon and may also create stop codons.

Chromosome Mutations

Chromosomal mutations are large-scale mutations resulting in changes at the chromosome level instead of at the gene level. For example, a **deletion** results in a large section of a chromosome, usually involving several to hundreds of genes, being lost. A **duplication** is the opposite situation: a large section of a chromosome is repeated. In an **inversion**, a piece of a chromosome breaks off, flips, and rejoins the chromosome. A **translocation** results when a large segment of one chromosome breaks off and attaches to a different chromosome (**Figure 16.6**).

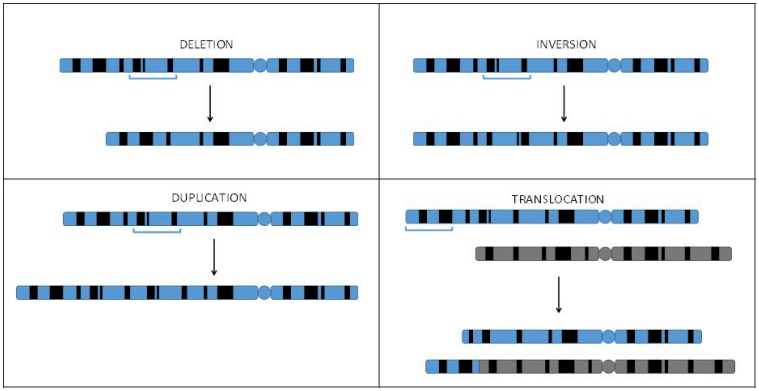


Figure 16.6 Examples of chromosomal mutations. A deletion removes a segment of a chromosome, while a duplication inserts an extra copy of a segment. An inversion flips a segment of a chromosome around. A translocation occurs when a segment from one chromosome becomes attached to a different chromosome.

Since they involve large numbers of genes being removed or repeated, deletions and duplications often cause serious problems for the individual who has them. Conversely, inversions and translocations are often not problematic. Notice that in both cases, no genetic material is lost (**Figure 5.6**). The only problems that may arise are if the breakpoints are in the middle of genes. However, both inversions and translocations cause problems during reproduction and may lead to infertility.

Results of Mutations

Regardless of type, all mutations can be divided into three categories:

Beneficial mutations increase the fitness (its ability to survive and reproduce) of an organism.

Neutral mutations do not have an effect on fitness. Silent mutations fall in this category.

Deleterious mutations lower fitness in an organism.

16.2 | Transcription

Learning Objectives

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

- Describe the different steps in transcription.
- Discuss the role of promoters in transcription.
- Describe the role of RNA polymerases in transcription.

Transcription is the process of making an RNA copy of a

gene (**Figure 16.2**). In eukaryotic cells, DNA is housed in the nucleus and transcription takes place in the nucleus. Since prokaryotic cells do not have a nucleus, transcription takes place in the cytoplasm in these cells.

Transcription is mediated by enzymes called **RNA polymerases**. RNA polymerases synthesize RNA in a 5′ ➞ 3′ direction. One of the strands of DNA, called the **template strand**, is used as a template to construct a complementary copy of the DNA. The other strand of DNA is nearly identical to the RNA copy, and is therefore called the **coding strand**. Since RNA has the nucleotide uracil (U) rather than thymine (T), the RNA copy has the same sequence as the coding strand except with U instead of T. For example, if the template strand of DNA has the sequence: ATCAGT, the coding strand will have the sequence TAGTCA and the RNA will have the sequence UAGUCA.

The first nucleotide pair in the DNA double helix, from which the first RNA nucleotide is transcribed is called the +1 site, or the **initiation site**. Nucleotides preceding the initiation site are given negative numbers and are designated **upstream**. Conversely, nucleotides following the initiation site are given positive numbers and are called **downstream**.

Transcription takes place in three stages: initiation, elongation, and termination.

16.2.1 Initiation of Transcription

Promoters

The first step in transcription is to identify where to begin copying the DNA strand. Each gene has a specific sequence of DNA, called the **promotor**, which specifies where to begin transcription and denotes which DNA strand to copy. The promoter is the site where the transcription machinery binds and initiates transcription. In most cases, promoters exist upstream of the genes they regulate.

The specific sequence of a promoter is very important because it determines whether the corresponding gene is transcribed all the time, some of the time, or infrequently. Prokaryotic promoters have regions at -10 and -35 upstream of the initiation site, where the promoter binds (**Figure 16.7**). The -10 sequence is TATAAT, and is therefore called the TATA box. The -35 sequence is TTGACA. A subunit of RNA polymerase called **sigma** (σ) binds to the -35 sequence. Once this interaction is made, the rest of the RNA polymerase enzyme binds to the promoter. Eukaryotic promoters are much larger and more complex than prokaryotic promoters, but they also have a TATA box. (**Figure 16.8**).

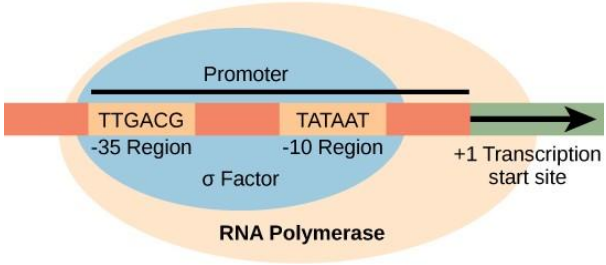


Figure 16.7 The σ subunit of prokaryotic RNA polymerase recognizes sequences in the promoter region upstream of the transcription start sight. The σ subunit dissociates from the polymerase after transcription has been initiated.

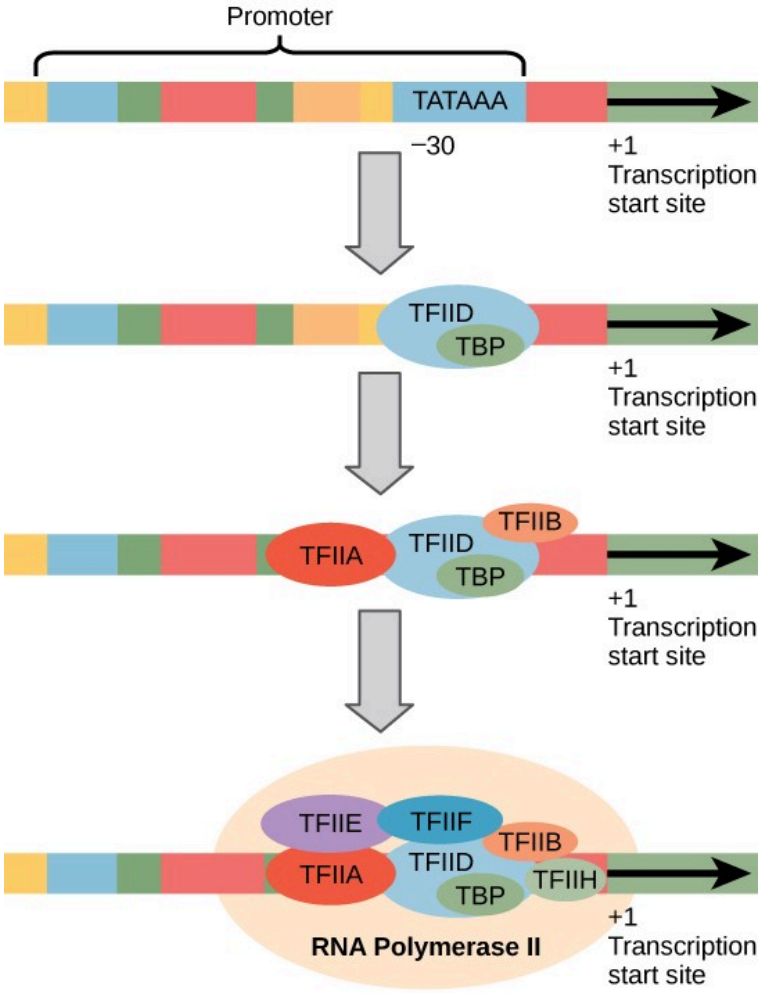


Figure 16.8 A generalized eukaryotic promoter is shown. Transcription factors recognize and bind to the promoter. RNA polymerase then binds and forms the initiation complex.

Initiation Complex

In prokaryotes, σ controls when RNA polymerase binds to a promoter and begins to transcribe a gene. Eukaryotes rely on an army of **transcription factors** and other regulatory proteins to help regulate the frequency with which RNA is synthesized from a gene. Transcription factors bind to the DNA template strand and subsequently recruit RNA polymerase for transcription initiation. Transcription factors are essential to ensure that the cell transcribes precisely the RNAs that it needs.

The complex of transcription factors and RNA polymerase bound to a promoter is called an **initiation complex (Figure 16.8)**. Once RNA polymerase is bound to the promoter, the DNA double helix of a gene must be unwound to make room for RNA synthesis. The region of unwinding is called a **transcription bubble (Figure 16.9)**.

16.2.2 Transcription Elongation

During elongation, RNA polymerase proceeds 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 enzyme and rewound behind it (**Figure 16.9**).

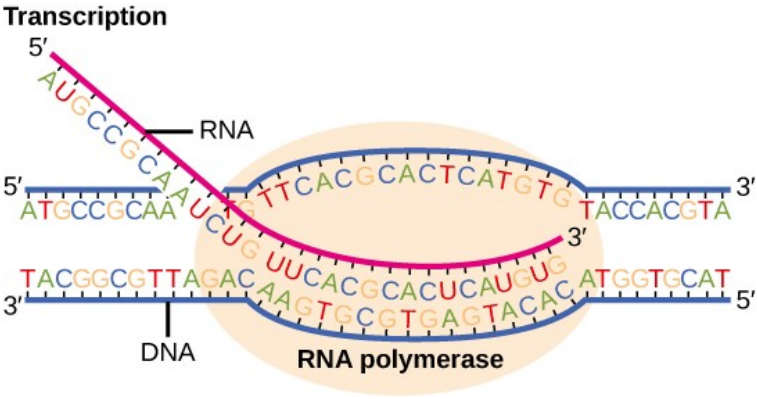


Figure 16.9 During elongation, RNA polymerase unwinds the DNA, tracks along the DNA template strand, synthesizes mRNA in the 5' to 3' direction, and rewinds the DNA.

16.2.3 Transcription Termination

Once a gene is transcribed, the polymerase needs to dissociate from the DNA template and liberate the newly made mRNA. In prokaryotes, 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 break away and liberate the new mRNA transcript.

In eukaryotes, termination of transcription is different for the three different RNA polymerases. RNA Polymerase II transcribes 1,000–2,000 extra nucleotides beyond the end of the gene. This tail is subsequently removed during mRNA processing. 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.

Upon termination, the process of transcription is complete. By the time termination occurs, the prokaryotic transcript would already have been used to begin synthesis of numerous copies of the encoded protein because these processes can occur concurrently (**Figure 16.10**). The unification of transcription and translation is possible because there is no nucleus in the prokaryotic cell. In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.

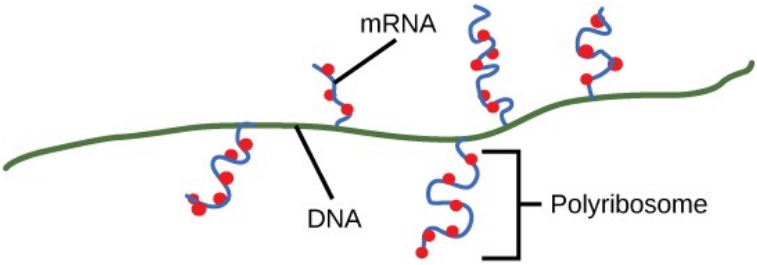


Figure 16.10 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

Concept Check

A scientist splices a eukaryotic promoter in front of a bacterial gene and inserts the gene in a bacterial chromosome. Would you expect the bacteria to transcribe the gene?

RNA Polymerases

Prokaryotes use the same RNA polymerase to transcribe all of their genes. In *E. coli*, the polymerase is composed of five subunits, two of which are identical. Four of these subunits, denoted α, α, β, and β' comprise the polymerase **core enzyme**. Each subunit has a unique role; the two α-subunits assemble the polymerase on the DNA; the β subunit binds to the ribonucleoside triphosphate that will become part of the new mRNA molecule; and the β' binds the DNA template strand. The fifth subunit, σ, is involved only in transcription initiation. Without σ, the core enzyme would transcribe from random sites and would produce mRNA molecules that specified protein gibberish. The polymerase comprised of all five subunits is called the **holoenzyme**.

Eukaryotes employ three different RNA polymerases that each transcribe a different type of gene. Each RNA

polymerase is made of 10 or more subunits. Each requires a distinct set of transcription factors to bring it to the DNA template.

RNA polymerase I transcribes most ribosomal RNA (rRNA), which becomes part of ribosomes. The rRNA molecules are considered structural RNAs because they are not translated into protein. The rRNAs are components of the ribosome and are essential to the process of translation.

RNA polymerase II synthesizes all protein-coding nuclear pre-mRNAs. Eukaryotic pre-mRNAs undergo extensive processing after transcription to become mature mRNAs. RNA polymerase II is responsible for transcribing the overwhelming majority of eukaryotic genes.

RNA polymerase III transcribes a variety of structural RNAs that includes one rRNA, transfer RNAs (tRNAs), and small nuclear RNAs (snRNAs**)**. tRNAs have a

critical role in translation; they bring amino acids to the growing polypeptide chain. snRNAs have a variety of functions, including "splicing" pre-mRNAs and regulating transcription factors.

A scientist characterizing a new gene can determine which polymerase transcribes it by testing whether the gene is expressed in the presence of a particular mushroom poison, α-amanitin. Interestingly, α-amanitin affects the three polymerases very differently. RNA polymerase I is completely insensitive to α-amanitin. In contrast, RNA polymerase II is extremely sensitive to α-amanitin, and RNA polymerase III is moderately sensitive. Knowing the transcribing polymerase can clue a researcher into the general function of the gene being studied.

16.3 | RNA Processing in Eukaryotes

Learning Objectives

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

- Describe the different steps in RNA processing.
- Understand the significance of exons, introns, and splicing.
- Explain how tRNAs and rRNAs are processed.

After transcription, eukaryotic pre-mRNAs must undergo several processing steps before they can be translated. Eukaryotic (and prokaryotic) tRNAs and rRNAs also undergo processing before they can function as components in the protein synthesis machinery.

16.3.1 mRNA Processing

The eukaryotic pre-mRNA undergoes extensive processing before it is ready to be translated. The additional steps involved in eukaryotic mRNA maturation create a molecule that is much more stable than a prokaryotic mRNA. Eukaryotic

mRNAs last for several hours, whereas the typical *E. coli* mRNA lasts no more than five seconds.

Pre-mRNAs are first coated in RNA-stabilizing proteins that protect the pre-mRNA from degradation while it is processed and exported out of the nucleus. The three most important steps of pre-mRNA processing are the addition of stabilizing and signaling factors at the 5′ and 3′ ends of the molecule, and the removal of intervening sequences that do not specify the appropriate amino acids.

5′ Capping

While the pre-mRNA is still being synthesized, a **7-methylguanosine cap** is added to the 5′ end of the growing transcript. This "**5' cap**" protects the mRNA from degradation. In addition, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

3′ Poly-A Tail

Once elongation is complete, the pre-mRNA is cleaved by an endonuclease between an AAUAAA consensus sequence and a GU-rich sequence, leaving the AAUAAA sequence on the pre-mRNA. An enzyme called poly-A polymerase then adds a string of approximately 200 A residues, called the **poly-A tail**. This modification further protects the pre-mRNA from degradation.

Pre-mRNA Splicing

Eukaryotic genes are composed of **exons**, which correspond to protein-coding sequences (ex-on signifies that they are expressed), and intervening sequences called **introns** (intron denotes their intervening role). Intron sequences in mRNA do not encode functional proteins and are removed from the pre-mRNA during processing.

All of a pre-mRNA's introns must be completely and precisely removed before protein synthesis begins. If the process errs by even a single nucleotide, the reading frame of the rejoined exons would shift, and the resulting protein would be dysfunctional (**Figure 16.5**). The process of removing introns and reconnecting exons is called splicing (**Figure 16.11**). Introns are removed and degraded while the pre-mRNA is still in the nucleus. Splicing occurs by a sequence-specific mechanism that ensures introns will be removed and exons rejoined with accuracy and precision. The splicing of pre-mRNAs is conducted by complexes of proteins and RNA molecules called spliceosomes. Note that more than 70 individual introns can be present, and each has to undergo the process of splicing – in addition to 5′ capping and the addition of a poly-A tail – just to generate a single, translatable mRNA molecule (**Figure 16.12**).

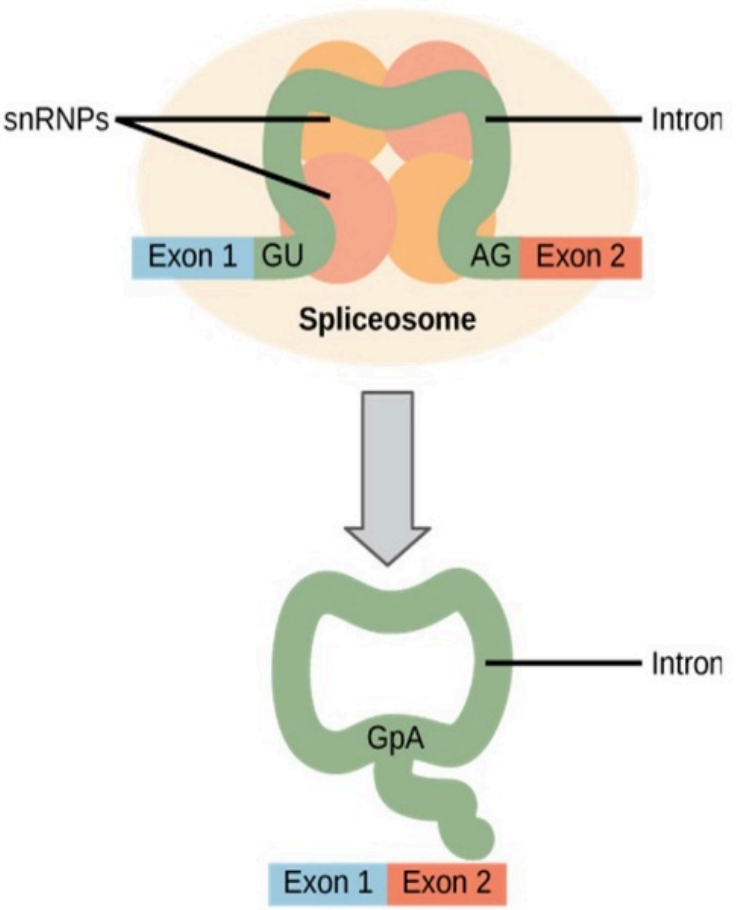


Figure 16.11 Pre-mRNA splicing involves the precise removal of introns from the primary RNA transcript. The splicing process is catalyzed by protein complexes called spliceosomes that are composed of proteins and RNA molecules called snRNAs. Spliceosomes recognize sequences at the 5' and 3' end of the intron.

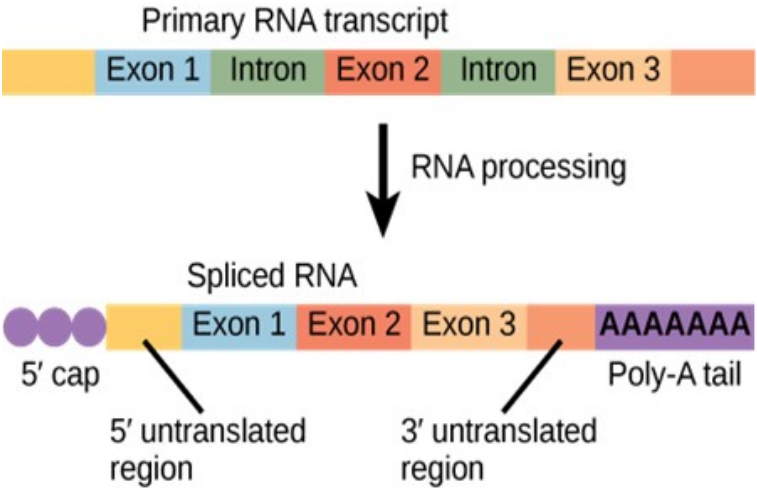


Figure 16.12 Pre-mRNA splicing involves the precise removal of introns from the primary RNA transcript. The splicing process is catalyzed by protein complexes called spliceosomes that are composed of proteins and RNA molecules called snRNAs. Spliceosomes recognize sequences at the 5' and 3' end of the intron.

Concept Check

Errors in splicing are implicated in cancers and other human diseases. What kinds of mutations might lead to splicing errors? Think of different possible outcomes if splicing errors occur.

An Unexpected Discovery

The discovery of introns came as a surprise to researchers in the 1970s who expected that pre-mRNAs would specify protein sequences without further processing, as they had observed in prokaryotes. The genes of higher eukaryotes very often contain one or more introns. These regions may correspond to regulatory sequences; however, the biological significance of having many introns or having very long introns in a gene is unclear. It is possible that introns slow down gene expression because it takes longer to transcribe pre-mRNAs with lots of introns. Alternatively, introns may be nonfunctional sequence remnants left over from the fusion of ancient genes throughout evolution. This is supported by the fact that separate exons often encode separate protein subunits or domains. For the most part, the sequences of introns can be mutated without ultimately affecting the protein product.

16.3.2 Processing of tRNAs and

rRNAs

tRNAs and rRNAs are structural molecules that have roles in protein synthesis. However, these RNAs are not translated. Pre-rRNAs are transcribed, processed, and assembled into ribosomes, in a specialized region of the nucleus called the **nucleolus**. Ribosomes are described in the next section. Pre-tRNAs are transcribed and processed in the nucleus and then released into the cytoplasm. Mature tRNAs take on a three-dimensional structure through hydrogen bonding between nucleotides within the tRNA. The result is an amino acid binding site at one end of the tRNA and an **anticodon** at the other end (**Figure 16.14**). The anticodon is a three-nucleotide sequence in a tRNA that base pairs with a complementary mRNA codon.

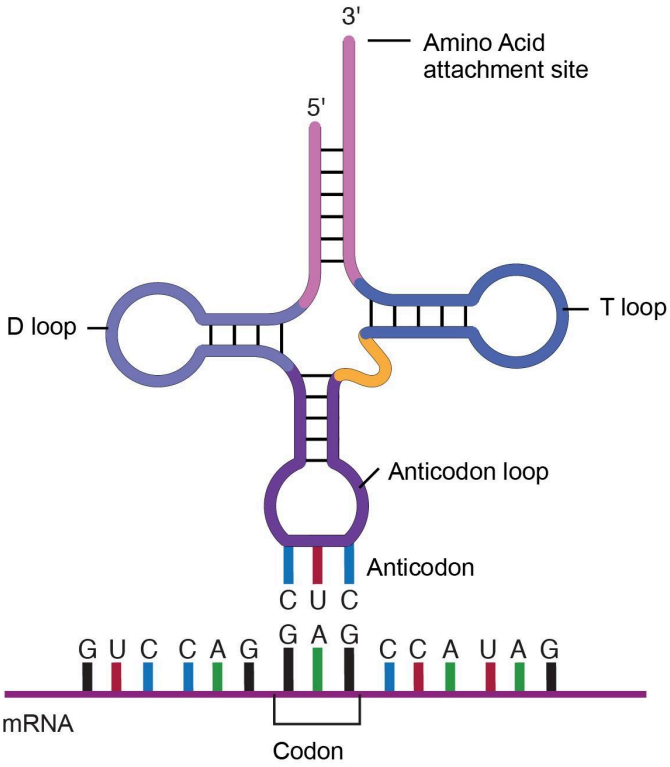


Figure 16.14 Each tRNA has two important areas: a trinucleotide region called the anticodon and a region for attaching a specific amino acid. During translation, tRNAs form base pairs with complementary sequences on the mRNA, ensuring that the appropriate amino acid is inserted into the protein. Credit: Darryl Leja, NHGRI (www.genome.gov).

16.4 | Ribosomes and Protein Synthesis

Learning Objectives

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

- Describe the different steps in protein synthesis.
- Discuss the role of ribosomes in protein synthesis.

The synthesis of proteins consumes more of a cell's energy than any other metabolic process. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform virtually every function of a cell. The process of **translation** involves the decoding of an mRNA message to form a polypeptide. Fifty to more than 1000 amino acids are covalently strung together by dehydration synthesis reactions, forming peptide bonds. (**Figure 16.15**). This reaction is catalyzed by ribosomes.

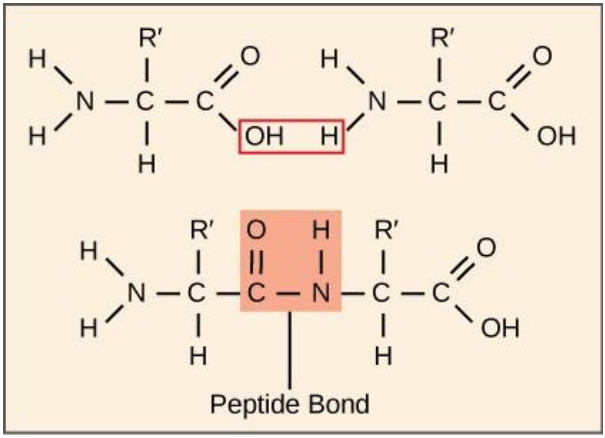


Figure 16.15 A peptide bond links the carboxyl end of one amino acid with the amino end of another, expelling one water molecule. For simplicity in this image, only the functional groups involved in the peptide bond are shown. The R and R' designations refer to the rest of each amino acid structure.

16.4.1 The Protein Synthesis Machinery

In addition to the mRNA template, many molecules and macromolecules contribute to the process of translation. The composition of each component may vary across species; for instance, ribosomes may consist of different numbers of rRNAs and proteins. However, the general structures and functions of the protein synthesis machinery are comparable

from bacteria to human cells. Translation requires the input of an mRNA template, ribosomes, tRNAs, and various enzymes.

Ribosomes

Before an mRNA is translated, a cell must build ribosomes. In a single prokaryotic E. coli cell, there are between 10,000 and 70,000 ribosomes present at any given time. A ribosome is a complex macromolecule composed of rRNAs and many distinct polypeptides (**Figure 16.16**). In eukaryotes, the nucleolus is specialized for the synthesis and assembly of ribosomes.

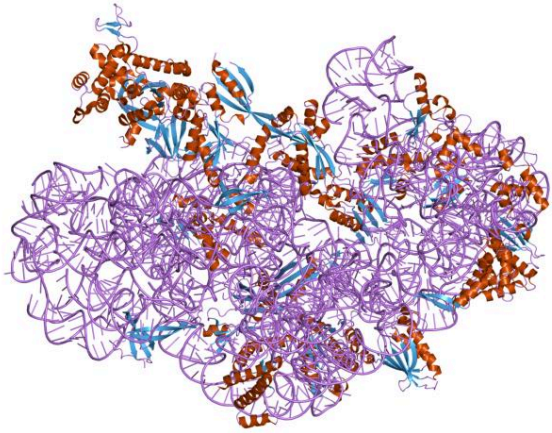


Figure 16.16 A ribosome is made of four rRNA molecules (shown in purple) in a complex with over 70 proteins. (By Jawahar Swaminathan and MSD staff at the European Bioinformatics Institute. Public Domain. https://commons.wikimedia.org/w/index.php?curid=5843147).

Ribosomes dissociate into large and small subunits when they are not synthesizing proteins and reassociate during the initiation of translation. The small subunit is responsible for binding the mRNA template, whereas the large subunit binds tRNAs. Each mRNA molecule is simultaneously translated by many ribosomes, all 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** (**Figure 16.17**).

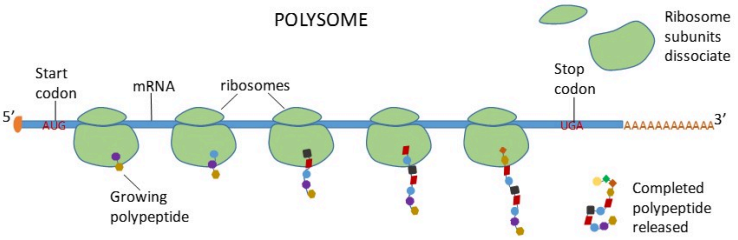


Figure 16.17 Many ribosomes can translate an mRNA at the same time in a 5' to 3' direction. The resulting complex of many ribosomes simultaneously making many copies of a protein is called a polysome.

Ribosomes are found either in the cytoplasm or attached to the rough endoplasmic reticulum. Mitochondria and chloroplasts also have their own ribosomes, which look and act more similar to prokaryotic ribosomes than to the nuclear ribosomes in the same cell.

tRNAs

tRNAs are structural RNA molecules that 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.

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 the codon GAG

on an mRNA template would bind a tRNA that had the complementary anticodon CUC and was linked to the amino acid glutamic acid (**Figure 16.14**).

Aminoacyl tRNA Synthetases

tRNA molecules are linked to their correct amino acids by a group of enzymes called **aminoacyl tRNA synthetases.** Adding the amino acid to a tRNA is called "**charging**" the tRNA. At least one type of aminoacyl tRNA synthetase exists for each of the 20 amino acids.

16.4.2 The Mechanism of Protein Synthesis

As with mRNA synthesis, protein synthesis can be divided into three phases: initiation, elongation, and termination.

Initiation of Translation

Protein synthesis begins with the formation of an initiation complex. This complex involves the small ribosomal subunit, the mRNA template, initiation factors, and a special **initiator tRNA**, called tRNAMet. The initiator tRNA has anticodon UAG, which interacts with the start codon AUG and is charged with the amino acid methionine. Methionine is therefore the first amino acid of every polypeptide chain.

First, the mRNA binds to the small ribosomal subunit. Initiation factors help the small subunit bind and scan along until it identifies the AUG start codon. Next, the anticodon of the initiator tRNA hydrogen bonds to the start codon. Finally, the large subunit of the ribosome binds, lining up so the initiator tRNA is in the P site. This step completes initiation of translation (**Figure 16.18**).

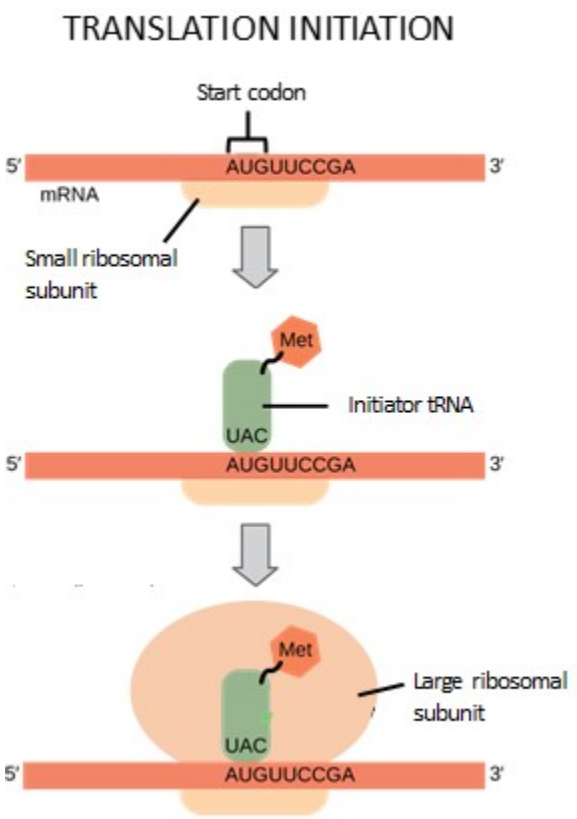


Figure 16.18 Initiation of translation begins with the small ribosomal subunit binding to the mRNA and scanning along until it reaches the start codon. Next, the initiator tRNA hydrogen bonds to the start codon. Finally, the large ribosomal subunit binds.

Elongation and Termination of Translation

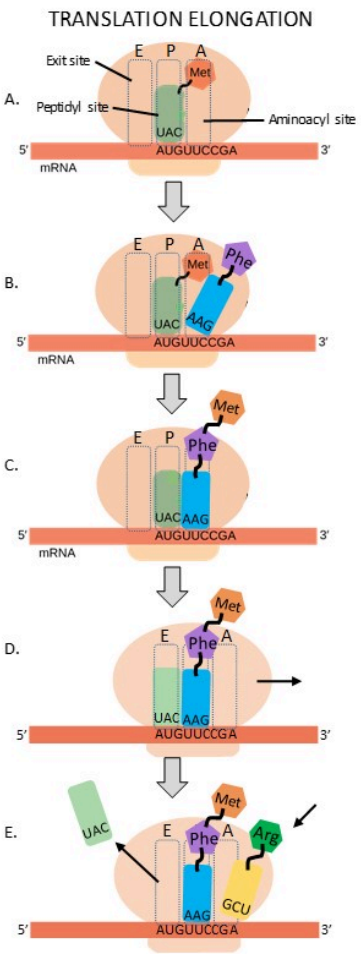


Figure 16.19 After initiation of translation (A), the second tRNA binds to the codon in the A site of the ribosome (B). The methionine is transferred to the phenylalanine, and a peptide bond is made (C). The ribosome slides one codon

The large ribosomal subunit contains three tRNA binding sites: the A (aminoacyl) site, the P (peptidyl) site, and the E (exit) site (**Figure 16.19A**). During translation elongation, the mRNA template provides specificity. As the ribosome moves along the mRNA, each mRNA codon comes into register, and specific binding with the corresponding charged tRNA anticodon is ensured.

The first step in elongation occurs when the second codon is recognized by the appropriate charged tRNA. As the anticodon of this tRNA hydrogen bonds to the second codon, it enters the A site

downstream (D). The initiator tRNA exits the E site, while a third tRNA enters the A site (E).

of the ribosome (**Figure 16.19B**).

Next, the two amino acids are covalently bonded together with a peptide bond. The growing polypeptide chain remains attached to the tRNA in the A site (**Figure 16.19C**).

The ribosome now slides along the mRNA by a distance of one codon, in a process called **translocation**. Since the tRNAs are hydrogen bonded to the mRNA, they remain behind as the ribosome moves. The result of translocation is that the initiator tRNA is now in the E site of the ribosome, the second tRNA is in the P site, and the A site is empty (**Figure 16.19D**).

After translocation, the initiator tRNA exits from the E site. The third tRNA recognizes the third codon and hydrogen bonds to it, entering the A site of the ribosome (**Figure 16.19E**).

Steps C, D and E repeat until the stop codon is reached. Each cycle of these three steps adds one amino acid to the growing polypeptide. Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino acid protein can be translated in just 10 seconds!

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is encountered. Upon aligning with the A site, the stop codon is recognized by a release factor.

The P-site amino acid detaches from its tRNA, and the newly made polypeptide is released. The small and large ribosomal subunits dissociate from the mRNA and from each other and find another mRNA to translate (**Figure 16.20**). After many ribosomes have translated it, the mRNA is degraded so that its nucleotides can be reused.

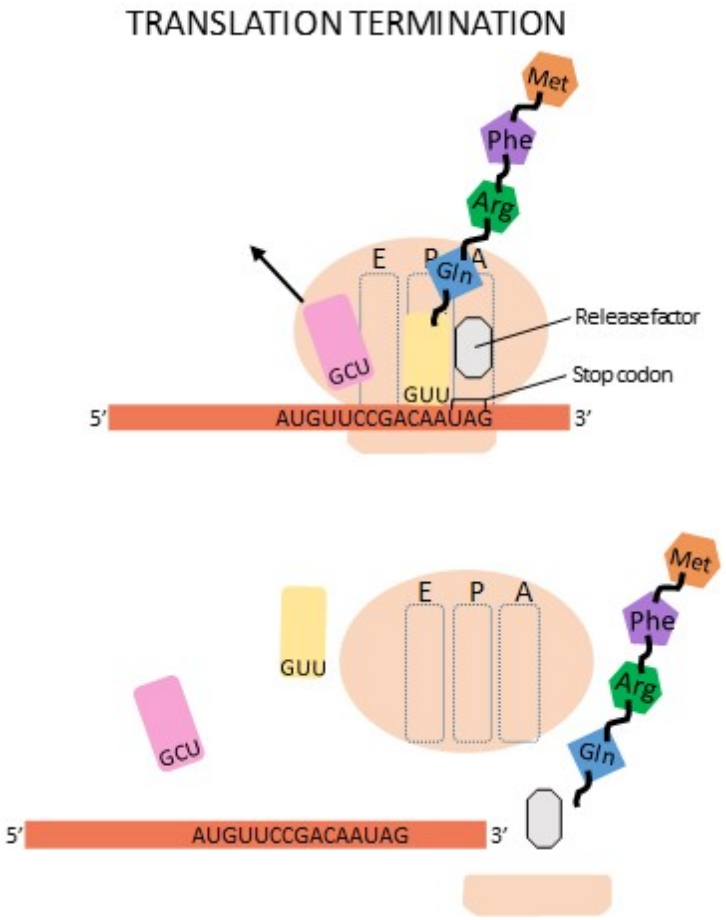


Figure 16.20 When the stop codon is reached, a release factor binds in the A site of the ribosome. The protein, ribosomal subunits, mRNA, and tRNAs dissociate and are reused to translate another mRNA.

16.4.3 Protein Folding, Modification, and Targeting

During and after translation, individual amino acids may be chemically modified, signal sequences may be added, and the new protein "folds" into a distinct three-dimensional structure as a result of intramolecular interactions. A **signal sequence** is a short tail of amino acids that directs a protein to a specific cellular compartment. These sequences at the amino end or the carboxyl end of the protein can be thought of as the protein's "train ticket" to its ultimate destination. Other cellular factors recognize each signal sequence and help transport the protein from the cytoplasm to its correct compartment. For instance, a specific sequence at the amino terminus will direct a protein to the mitochondria or chloroplasts (in plants). Once the protein reaches its cellular destination, the signal sequence is usually clipped off.

Many proteins fold spontaneously, but some proteins require helper molecules, called **chaperones**, to help them fold correctly. Even if a protein is properly specified by its corresponding mRNA, it could take on a completely dysfunctional shape if abnormal temperature or pH conditions prevent it from folding correctly.

2.

CHAPTER 17. REGULATION OF GENE EXPRESSION

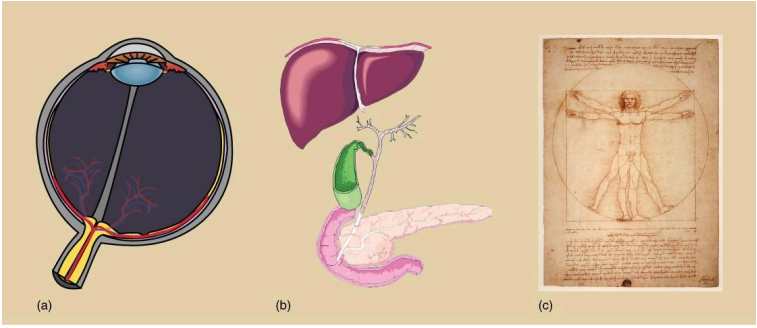


Figure 17.1 The genetic content of each somatic cell in an organism is the same, but not all genes are expressed in every cell. The control of which genes are expressed dictates whether a cell is (a) an eye cell or (b) a liver cell. It is the differential gene expression patterns that arise in different cells that give rise to (c) a complete organism.

Chapter Outline

- 17.1 Overview of Regulation of Gene Expression
- 17.2 Prokaryotic Gene Regulation
- 17.3 Eukaryotic Gene Regulation

Introduction

Each somatic cell in the body generally contains the same DNA. (A few exceptions include red blood cells, which contain no DNA in their mature state, and some immune system cells that rearrange their DNA while producing antibodies.) In general, the genes that determine whether you have green eyes or brown hair, or how fast you metabolize food are the same in eye cells and liver cells, even though these organs function quite differently. If each cell has the same DNA, how is it that cells differ in their structure and function? Why do cells in the eye differ so dramatically from cells in the liver?

Although each cell in your body contains the same DNA sequences, each cell does not turn on, or express, the same set of genes. In fact, only a small subset of proteins are made by any one cell. In other words, in any given cell, not all genes encoded in the DNA are transcribed into mRNA or translated

into protein. Cells in the eye make a certain subset of proteins, and liver cells make a different subset of proteins. In addition, at different times, liver cells may make different subsets of liver proteins. The expression of specific genes is a highly regulated process with many levels and stages of control. This complexity ensures expression of each protein in the proper cells at the proper time.

17.1 | Overview of Regulation of Gene Expression

Learning Objectives

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

- Discuss why every cell does not express all of its genes.
- Describe some major differences between prokaryotic and eukaryotic gene regulation.

For a cell to function properly, necessary proteins must be synthesized at the proper time. All cells control or regulate

the synthesis of proteins from information encoded in their DNA. The process of "turning on" a gene to produce mRNA and protein is called **gene expression**. Whether in a simple unicellular organism or a complex multi-cellular organism, each cell controls when its genes are expressed, how much of the protein is made, and when it is time to stop making that protein because it is no longer needed.

The regulation of gene expression conserves energy and space. It is more energy efficient to turn on the genes only when they are required. In addition, only expressing a subset of genes in each cell saves space because DNA must be unwound from its tightly coiled structure to transcribe and translate the DNA. Cells would have to be enormous if every protein were expressed in every cell all the time. The control of gene expression is extremely complex. Malfunctions in this process are detrimental to the cell and can lead to the development of many diseases, including cancer.

17.1.1 Prokaryotic versus Eukaryotic Gene Expression

Since prokaryotic organisms are single-celled organisms that lack a cell nucleus, their DNA floats freely in the cell's cytoplasm. When a particular protein is needed, the gene that codes for it is transcribed in mRNA, which is simultaneously translated into protein. When the protein is no longer needed, transcription stops. As a result, the primary method to control

how much of each protein is expressed in a prokaryotic cell is the regulation of transcription.

Eukaryotic cells, in contrast, have intracellular organelles that add to their complexity. In eukaryotic cells, the DNA is contained inside the cell's nucleus, where it is transcribed into mRNA. The newly synthesized mRNA is then modified and transported out of the nucleus into the cytoplasm, where ribosomes translate the mRNA into protein. The processes of transcription and translation are physically separated by the nuclear membrane; transcription occurs only within the nucleus, and translation occurs only in the cytoplasm. The regulation of gene expression in eukaryotes can occur at all stages of the process (**Figure 17.2**).

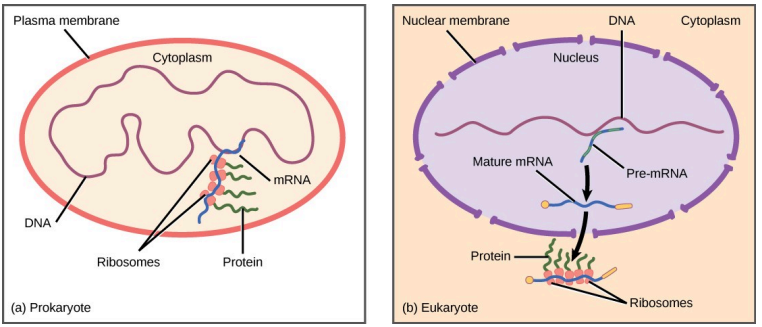


Figure 17.2 Prokaryotic transcription and translation occur simultaneously in the cytoplasm, and regulation occurs at the level of transcription. In eukaryotes, transcription and translation are physically separated, and gene expression is regulated at many different levels.

Some of the differences in the regulation of gene expression between prokaryotes and eukaryotes are summarized in **Table 17.1**.

Table 17.1 Differences in prokaryotic and eukaryotic gene regulation.

Prokaryotic organisms	Eukaryotic organisms
Lack nucleus	Contain nucleus
DNA is found in the cytoplasm	DNA is in the nucleus
Transcription and translation occur almost simultaneously	Transcription occurs in the nucleus prior to translation, which occurs in the cytoplasm.
Gene expression is regulated primarily at the transcriptional level	Gene expression is regulated at many levels: epigenetic, transcriptional, nuclear shuttling, post-transcriptional, translational, and post-translational

17.2 | Prokaryotic Gene Regulation

Learning Objectives

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

- Describe the steps involved in prokaryotic gene regulation.
- Explain the roles of activators, inducers, and repressors in gene regulation

The DNA of prokaryotes is organized into a circular chromosome that resides in the cell's cytoplasm. Proteins that are needed for a specific function, or that are involved in the same biochemical pathway, are often encoded together in blocks called **operons**. For example, all five of the genes needed to make the amino acid tryptophan in the bacterium *E. coli* are located next to each other in the *trp* operon. The genes in an operon are transcribed into a single mRNA molecule. This allows the genes to be controlled as a unit: either all are expressed, or none is expressed. Each operon needs only one regulatory region, including a **promoter**, where RNA polymerase binds, and an **operator**, where other regulatory proteins bind.

In prokaryotic cells, there are three types of regulatory molecules that can affect the expression of operons. **Activators** are proteins that increase the transcription of a gene. **Repressors** are proteins that suppress transcription of a gene. Finally, **inducers** are molecules that bind to repressors

and inactivate them. Below are two examples of how these molecules regulate different operons.

17.2.1 The *trp* Operon: A Repressor Operon

Like all cells, bacteria need amino acids to survive. Tryptophan is one amino acid that the bacterium *E. coli* can either ingest from the environment or synthesize. When E. coli needs to synthesize tryptophan, it must express a set of five proteins that are encoded by five genes. These five genes are located next to each other in the **tryptophan (*trp*) operon** (**Figure 17.3**).

When tryptophan is present in the environment, *E. coli* does not need to synthesize it, and the *trp* operon is switched off. However, when tryptophan availability is low, the *trp* operon is turned on so that the genes are transcribed, the proteins are made, and tryptophan can be synthesized.

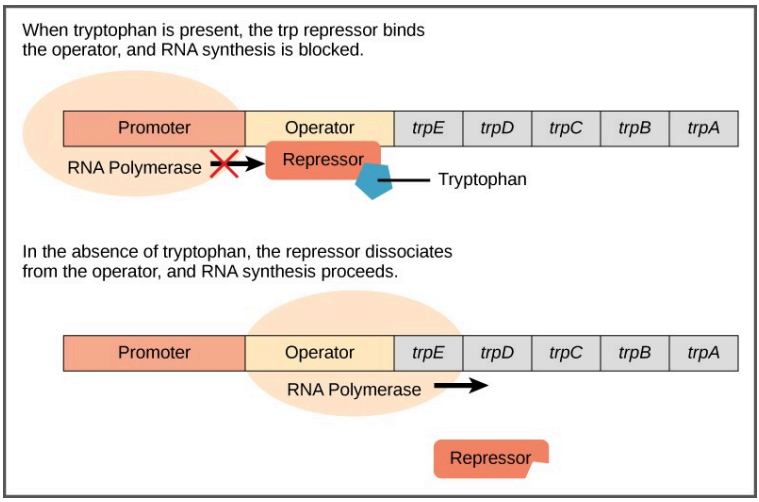


Figure 17.3 The five genes that are needed to synthesize tryptophan in E. coli are located next to each other in the trp operon.

A DNA sequence called the operator is located between the promoter and the first *trp* gene. The operator contains the DNA code to which the repressor protein can bind. The repressor protein is regulated by levels of tryptophan in the cell.

When tryptophan is present in the cell, two tryptophan molecules bind to the *trp* repressor. This causes the repressor to change shape and bind to the *trp* operator. Binding of the tryptophan–repressor complex at the operator physically blocks the RNA polymerase from binding, and transcribing the downstream genes. Thus, when the cell has enough tryptophan, it is preventing from making more.

When tryptophan is not present in the cell, the repressor

has no tryptophan to bind to it. The repressor is not activated and it does not bind to the operator. Therefore, RNA polymerase can transcribe the operon and make the enzymes to synthesize tryptophan. Thus, when the cell does not have enough tryptophan, it synthesizes it.

17.2.2 The *lac* Operon: An Inducer Operon

The *lac* operon in *E. coli* has more complex regulation, involving both a repressor and an activator. *E. coli* uses glucose for food, but is able to use other sugars, such as lactose, when glucose concentrations are low. Three proteins are needed to break down lactose; they are encoded by the three genes of the *lac* operon.

When lactose is not present, the proteins to digest lactose are not needed. Therefore, a repressor binds to the operator and prevents RNA polymerase from transcribing the operon.

When lactose is present, lactose binds to the repressor and removes it from the operator. RNA polymerase is now free to transcribe the genes necessary to digest lactose (**Figure 17.4**)

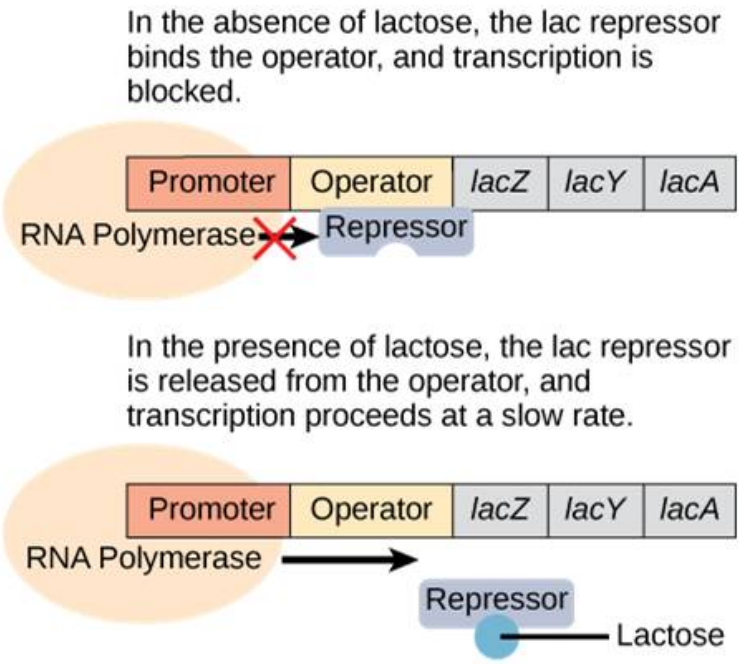


Figure 17.4 Transcription of the lac operon only occurs when lactose is present. Lactose binds to the repressor and removes it from the operator.

However, the story is more complex than this. Since *E. coli* prefers to use glucose for food, the *lac* operon is only expressed at low levels even when the repressor is removed. But what happens when ONLY lactose is present? Now the bacterium needs to ramp up production of the lactose-digesting proteins. It does so by using an activator protein called catabolite activator protein (CAP).

When glucose levels drop, cyclic AMP (cAMP) begins to accumulate in the cell. cAMP binds to CAP and the complex

binds to the *lac* operon promoter (**Figure 17.5**). This increases the binding ability of RNA polymerase to the promoter and ramps up transcription of the genes.

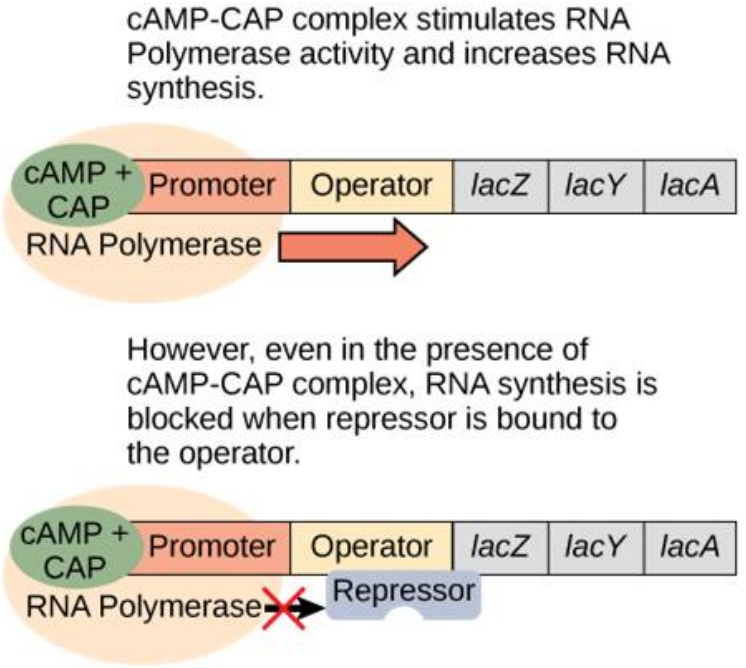


Figure 17.5 When there is no glucose, the CAP activator increases transcription of the lac operon. However, if no lactose is present, the operon is not activated.

In summary, for the *lac* operon to be fully 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 maximally. This makes sense for the cell, because it would be energetically wasteful to create the proteins to process lactose if glucose was plentiful or lactose was not available (**Table 17.2**).

Table 17.2 Summary of signals that induce or repress transcription of the *lac* operon.

Lactose present?	Repressor bound?	Glucose present?	CAP bound?	Transcription of *lac* operon?
-	Yes	+	No	No
-	Yes	-	Yes	No
+	No	+	No	Some
+	No	-	Yes	Lots

15.3 Eukaryotic Gene Regulation

Learning Objectives

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

- Explain the process of epigenetic gene regulation in eukaryotic cells.
- Explain the process of transcriptional gene regulation in eukaryotic cells.
- Explain the process of post-transcriptional gene regulation in eukaryotic cells.
- Explain the process of translational gene regulation in eukaryotic cells.
- Explain the process of post-transcriptional gene regulation in eukaryotic cells.

In eukaryotes, control of gene expression is more complex and can happen at many different levels. Eukaryotic genes are not organized into operons, so each gene must be regulated independently. In addition, eukaryotic cells have many more genes than prokaryotic cells. Regulation of gene expression can happen at any of the stages as DNA is transcribed into mRNA and mRNA is translated into protein. For convenience, regulation is divided into five levels: epigenetic, transcriptional, post-transcriptional, translational, and post-translational (**Figure 17.6**).

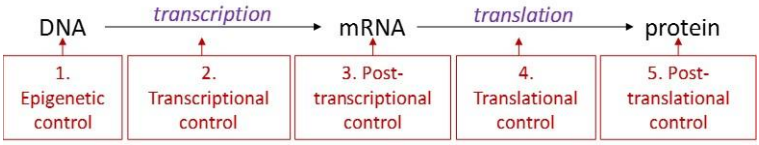


Figure 17.6 Regulation of gene expression in eukaryotes can occur at five different levels. Here, the Central Dogma is diagrammed with arrows showing where each type of eukaryotic regulation of gene expression interrupts it.

17.3.1 Epigenetic Control fo Gene Expression

The first level of control of gene expression is **epigenetic** ("around genetics") regulation. Epigenetics is a relatively new, but growing, field of biology.

Epigenetic control involves changes to genes that do not alter the nucleotide sequence of the DNA and are not permanent. Instead, these changes alter the chromosomal structure so that genes can be turned on or off. This level of control occurs through heritable chemical modifications of the DNA and/or chromosomal proteins.

One example of chemical modifications of DNA is the addition of methyl groups to the DNA, in a process called methylation, In general, methylation suppresses transcription. Interestingly, methylation patterns can be passed on as cells divide. Thus, parents may be able to pass on the tendency

of a gene to be expressed in their offspring. Other heritable chemical modifications of DNA may also occur.

Modification of Histone Proteins is an Example of Epigenetic Control

The best-studied example of epigenetic regulation is modification of histone proteins. Histones are chromosomal proteins that tightly wind DNA so that it fits into the nucleus of a cell. The human genome, for example, consists of over three billion nucleotide pairs. An average chromosome contains 130 million nucleotide pairs, and each body cell contains 46 chromosomes. If stretched out linearly, an average human chromosome would be over four centimeters long. In order to fit all of this DNA into the nucleus of a microscopic cell, the DNA must be tightly wound around proteins. It is also organized so that specific segments can be accessed as needed by a specific cell type (**Figure 17.7**).

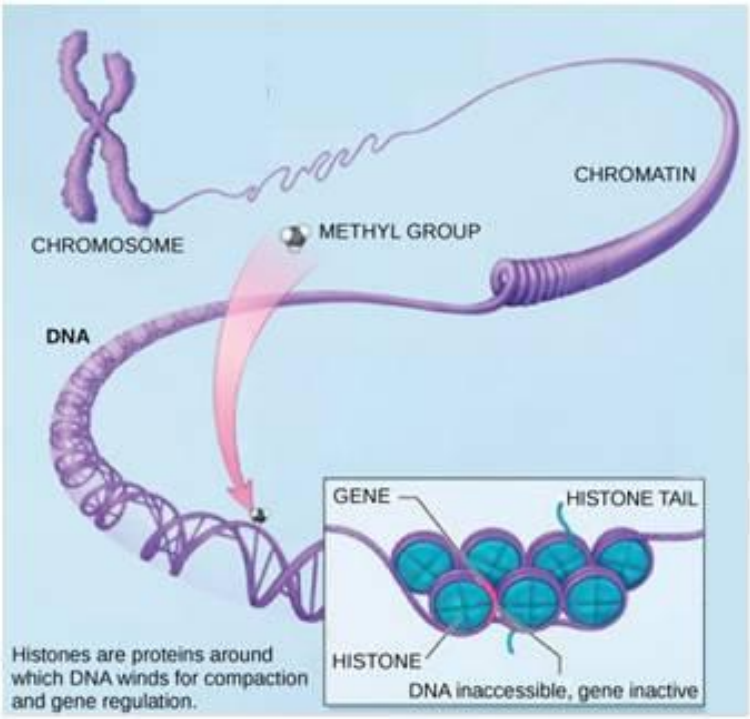


Figure 17.7 In each chromosome, DNA is wound around histone proteins to pack it into the nucleus of a cell. (Credit: modification of work by NIH.)

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 nucleosome complexes, which can control the access of proteins to the DNA regions (**Figure 17.8a**). Under the electron microscope, this winding of DNA around histone proteins to form nucleosomes looks like small beads on a string (**Figure 17.8b**). These beads

(histone proteins) can move along the string (DNA) and change the structure of the molecule.

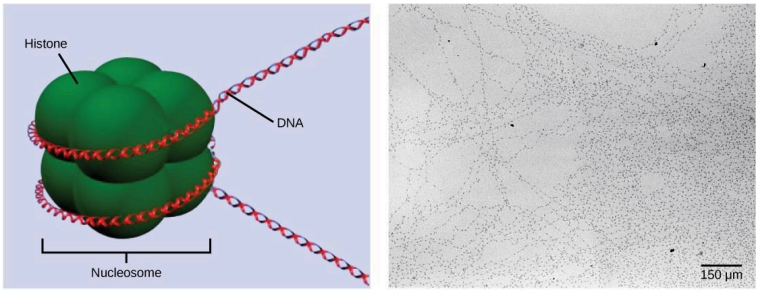


Figure 17.8 DNA is wrapped around histones to create nucleosomes (a), which control the access of proteins to DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (Credit "micrograph": modification of work by Chris Woodcock.)

If a gene is to be transcribed, the nucleosomes surrounding that region of DNA can slide down the DNA to open that specific chromosomal region and allow access for RNA polymerase and other proteins, called transcription factors, to bind to the promoter region and initiate transcription. If a gene is to remain turned off, or silenced, the histone proteins and DNA have different modifications that signal a closed chromosomal configuration. In this closed configuration, the RNA polymerase and transcription factors do not have access to the DNA and transcription cannot occur (**Figure 17.9**).

How the histone proteins move is dependent on signals found on the histone proteins. These signals are "tags" – in

the form of phosphate, methyl, or acetyl groups – that open or close a chromosomal region (**Figure 17.9**). These tags are not permanent, but may be added or removed as needed. Since DNA negatively charged, changes in the charge of the histone 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.

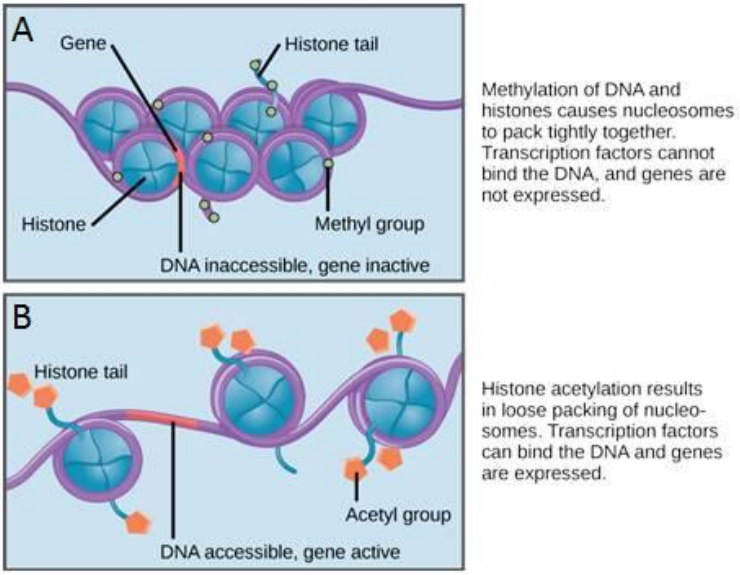


Figure 17.9 Nucleosomes can slide along DNA. (A) When nucleosomes are spaced closely together, transcription factors cannot bind and gene expression is turned off. (B) When nucleosomes are spaced far apart, transcription factors can bind, allowing gene expression to occur.

17.3.2 Transcriptional Control of Gene Expression

Transcriptional regulation is control of whether or not an mRNA is transcribed from a gene in a particular cell. Like prokaryotic cells, the transcription of genes in eukaryotes requires an RNA polymerase to bind to a promoter to initiate transcription. In eukaryotes, RNA polymerase requires other proteins, or **transcription factors**, to facilitate transcription initiation. Transcription factors are proteins that bind to the promoter sequence and other regulatory sequences to control the transcription of the target gene. RNA polymerase by itself cannot initiate transcription in eukaryotic cells. Transcription factors must bind to the promoter region first and recruit RNA polymerase to the site for transcription to begin.

The Promoter and Transcription Factors

In eukaryotic genes, the promoter region is immediately upstream of the coding sequence. This region can range from a few to hundreds of nucleotides long. The length of the promoter is gene-specific and can differ dramatically between genes. The longer the promoter, the more available space for proteins to bind. Consequently, the level of control of gene expression can differ quite dramatically between genes. The purpose of the promoter is to bind transcription factors that control the initiation of transcription (**Figure 17.10, top**).

Within the promoter region, just upstream of the transcriptional start site, resides the TATA box. This box is simply a repeat of thymine and adenine dinucleotides (literally, TATA repeats). Transcription factors bind to the TATA box, assembling an initiation complex. Once this complex is assembled, RNA polymerase binds to its upstream sequence and becomes phosphorylated. This releases part of the protein from the DNA, activates the transcription initiation complex, and places RNA polymerase in the correct orientation to begin transcription (**Figure 17.10, top**).

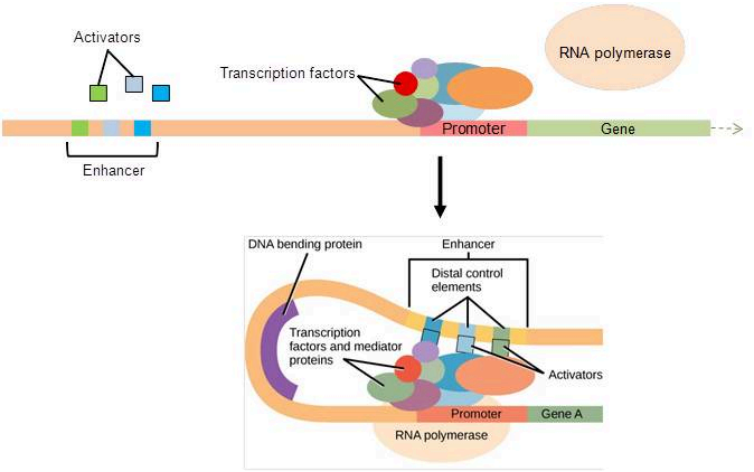


Figure 17.10 Top. Each gene has a promoter upstream of the coding sequence. The promoter binds to transcription factors and helps RNA polymerase to bind and start transcription. Bottom. Many genes also have upstream enhancers. Enhancers bind activators, bend around, and help RNA polymerase start transcription.

Enhancers and Repressors

In some eukaryotic genes, there are regions that help increase transcription. These regions, called **enhancers**, are not necessarily close to the genes; they can be located thousands of nucleotides away. They can be found upstream, within the coding region, or downstream of a gene. Enhancers are binding sites for activators. When an enhancer is far away from a gene, the DNA folds such that the enhancer is brought into proximity with the promoter, allowing interaction between the activators and the transcription initiation complex (**Figure 17.10, bottom**).

Like prokaryotic cells, eukaryotic cells also have mechanisms to prevent transcription. Transcriptional **repressors** can bind to promoter or enhancer regions and block transcription. Both activators and repressors respond to external stimuli to determine which genes need to be expressed.

17.3.3 Post-transcriptional Control of Gene Expression

Post-transcriptional regulation occurs after the mRNA is transcribed but before translation begins. This regulation can occur at the level of mRNA processing, transport from the nucleus to the cytoplasm, or binding to ribosomes.

Alternative RNA splicing

Recall from chapter 5 that in eukaryotic cells the RNA primary transcript often contains introns, which are removed prior to translation.

Alternative RNA splicing is a mechanism that allows different combinations of introns, and sometimes exons, to be removed from the primary transcript (**Figure 17.11**). This allows different protein products to be produced from one gene. Alternative splicing can act as a mechanism of gene regulation. Differential splicing is used to produce different protein products in different cells or at different times within the same cell. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes; up to 70 percent of genes in humans are expressed as multiple proteins through alternative splicing.

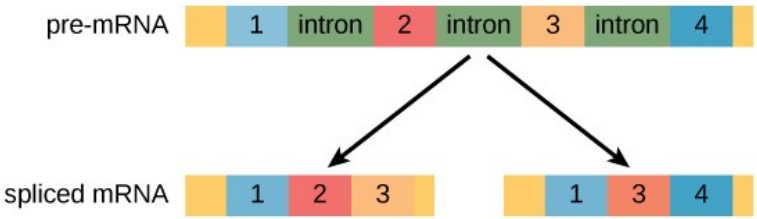


Figure 17.11 Before a RNA can be translated, introns must be removed by splicing. Pre-mRNA can be alternatively spliced to create different proteins.

evolution CONNECTION

Evolution of Alternative Splicing

How could alternative splicing evolve? Introns have a beginning and ending recognition sequence; it is easy to imagine the failure of the splicing mechanism to identify the end of an intron and instead find the end of the next intron, thus removing two introns and the intervening exon. In fact, there are mechanisms in place to prevent such intron skipping, but mutations are likely to lead to their failure. Such "mistakes" would more than likely produce a nonfunctional protein. Indeed, the cause of many genetic diseases is alternative splicing rather than mutations in a sequence. However, alternative splicing would create a protein variant without the loss of the original protein, opening up possibilities for adaptation of the new variant to new functions. Gene duplication has played an important role in the evolution of new functions in a similar way by providing genes that may evolve without eliminating the original, functional protein.

Control of RNA Stability

Another type of post-transcriptional control involves the stability of the mRNA in the cytoplasm. The longer an mRNA exists in the cytoplasm, the more time it has to be translated, and the more protein is made. Many factors contribute to mRNA stability, including the length of its poly-A tail.

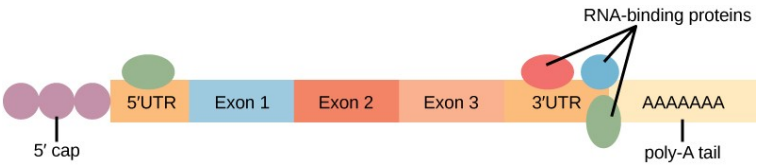


Figure 17.12 The protein-coding region of mRNA is flanked by 5' and 3' untranslated regions (UTRs). RNA-binding proteins at the 5' or 3' UTR influence the stability of the RNA molecule.

Proteins, called RNA-binding proteins (RBPs) can bind to the regions of the RNA just upstream or downstream of the protein-coding region. These regions in the RNA that are not translated into protein are called the **untranslated regions**, or UTRs. The region just before the protein-coding region is called the **5′ UTR**, whereas the region after the coding region is called the **3′ UTR** (**Figure** 17**.12**). The binding of RBPs to these regions can increase or decrease the stability of an RNA molecule, depending on the specific RBP that binds.

microRNAs, or **miRNAs**, can also bind to the RNA

molecule. miRNAs are short (21–24 nucleotides) RNA molecules that are made in the nucleus as longer pre-miRNAs and then chopped into mature miRNAs by a protein called **dicer**. miRNAs bind to mRNA along with a ribonucleoprotein complex called the **RNA-induced silencing complex (RISC).** The RISC-miRNA complex rapidly degrades the target mRNA.

17.3.4 Translational Control of Gene Expression

After an mRNA has been transported to the cytoplasm, it is translated into proteins. Control of this process is largely dependent on the mRNA molecule. As previously discussed, the stability of the mRNA will have a large impact on its translation into a protein. Translation can also be regulated at the level of binding of the mRNA to the ribosome. Once the mRNA bound to the ribosome, the speed and level of translation can still be controlled. An example of translational control occurs in proteins that are destined to end up in an organelle called the endoplasmic reticulum (ER). The first few amino acids of these proteins are a tag called a signal sequence. As soon as these amino acids are translated, a signal recognition particle (SRP) binds to the signal sequence and stops translation while the mRNA-ribosome complex is shuttled to the ER. Once they arrive, the SRP is removed and translation resumes.

17.3.5 Post-translational Control of Gene Expression

The final level of control of gene expression in eukaryotes is **post-translational regulation**. This type of control involves modifying the protein after it is made, in such as way as to affect its activity. One example of post-translational regulation is enzyme inhibition. When an enzyme is no longer needed, it is inhibited by a competitive or allosteric inhibitor, which prevents it from binding to its substrate. The inhibition is reversible, so that the enzyme can be reactivated later. This is more efficient than degrading the enzyme when it is not needed and then making more when it is needed again.

The activity and/or stability of proteins can also be regulated by adding functional groups, such as methyl, phosphate, or acetyl groups. 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.

The addition of an **ubiquitin** group to a protein marks that protein for degradation. Ubiquitin acts like a flag indicating that the protein's lifespan is complete. Tagged proteins are moved to a **proteasome**, an organelle that degrades proteins (**Figure 17.13**). One way to control gene expression, therefore, is to alter the longevity of the protein.

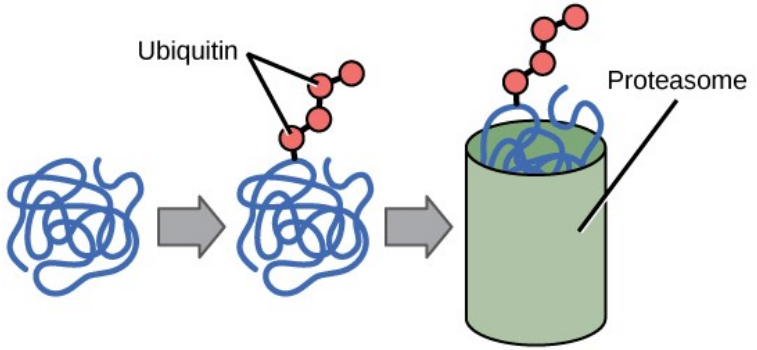


Figure 17.13 Proteins with ubiquitin tags are marked for degradation within the proteasome.

CHAPTER 18. MENDELIAN GENETICS

Figure 18.1 Mendel experimented with garden peas to uncover the fundamentals of genetics. (Credit: modification of work by Jerry Kirkhart)

Chapter Outline

- 18.1 Mendel's Experiments

- 18.2 Mendel's Principles of Inheritance
- 18.3 Exceptions to Mendel's Principles of Inheritance

Introduction

Genetics is the study of heredity. Johann Gregor Mendel (1822–1884) set the framework for genetics long before chromosomes or genes had been identified, at a time when meiosis was not well understood (**Figure 18.2**). Mendel selected a simple biological system and conducted methodical, quantitative analyses using large sample sizes. Because of Mendel's work, the fundamental principles of heredity were revealed. We now know that genes, carried on chromosomes, are the basic functional units of heredity with the capability to be

Figure 18.2 Johann Gregor Mendel is considered to be the father of genetics.

replicated, expressed, or mutated. Today, the postulates put forth by Mendel form the basis of classical, or Mendelian, genetics. Not all genes are transmitted from parents to offspring according to Mendelian genetics, but Mendel's experiments serve as an excellent starting point for thinking about inheritance.

18.1 | Mendel's Experiments

Learning Objectives

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

- Describe the scientific reasons for the success of Mendel's experimental work.
- Describe the expected outcomes of monohybrid crosses involving dominant and recessive alleles.

Johann Gregor Mendel (1822–1884) was a lifelong learner, teacher, scientist, and man of faith. As a young adult, he joined the Augustinian Abbey of St. Thomas in Brno in what is now

the Czech Republic. Supported by the monastery, he taught physics, botany, and natural science courses at the secondary and university levels. In 1856, he began a decade-long research pursuit involving inheritance patterns in honeybees and plants, ultimately settling on pea plants as his primary **model system**. In 1865, Mendel presented the results of his experiments with nearly 30,000 pea plants to the local Natural History Society. He demonstrated that traits are transmitted faithfully from parents to offspring independently of other traits and in dominant and recessive patterns. In 1866, he published his work, *Experiments in Plant Hybridization*[1] in the proceedings of the Natural History Society of Brünn.

Mendel's work went virtually unnoticed by the scientific community that believed, incorrectly, that the process of inheritance involved a blending of parental traits that produced an intermediate physical appearance in offspring; this hypothetical process appeared to be correct because of what we know now as continuous variation. **Continuous variation** results when many genes work together to determine a characteristic, such as human height or eye color.

1. Johann Gregor Mendel, Versuche über Pflanzenhybriden Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865 Abhandlungen, 3–47. [for English translation see http://www.mendelweb.org/Mendel.plain.html]

Offspring appear to be a "blend" of their parents' traits when we look at characteristics that exhibit continuous variation.

Mendel worked with traits that were inherited in distinct classes, such as violet versus white flowers. These traits display **discontinuous variation**. Mendel's choice of these kinds of traits allowed him to see that the traits were not blended in the offspring, nor were they absorbed, but rather that they kept their distinctness and could be passed on. In 1868, Mendel became abbot of the monastery and exchanged his scientific pursuits for his pastoral duties. He was not recognized for his extraordinary scientific contributions during his lifetime. In fact, it was not until 1900 that his work was rediscovered, reproduced, and revitalized by scientists on the brink of discovering the chromosomal basis of heredity.

18.1.1 Mendel's Model System

Mendel's seminal work was accomplished using the garden pea, *Pisum sativum*, to study inheritance. This species naturally self-fertilizes, such that pollen encounters ova within individual flowers. The flower petals remain sealed tightly until after pollination, preventing pollination from other plants. The result is highly inbred, or "true-breeding," pea plants. These are plants that always produce offspring that look like the parent. By experimenting with true-breeding pea plants, Mendel avoided the appearance of unexpected traits in offspring that might occur if the plants were not true breeding.

The garden pea also grows to maturity within one season, meaning that several generations could be evaluated over a relatively short time. Finally, large quantities of garden peas could be cultivated simultaneously, allowing Mendel to conclude that his results did not come about simply by chance.

18.1.2 Mendelian Crosses

Mendel performed **hybridizations**, which involve mating two true-breeding individuals that have different traits. In the pea, this is done by manually transferring pollen from one pea plant to the stigma of another pea plant. In plants, pollen carries the male gametes (sperm) to the stigma, a sticky organ that traps pollen and allows the sperm to move down the pistil to the female gametes (ova) below. To prevent the pea plant that was receiving pollen from self-fertilizing and confounding his results, Mendel painstakingly removed all of the pollen-producing anthers from the plant's flowers before they had a chance to mature.

Plants used in first-generation crosses were called **P**, or parental generation, plants (**Figure 18.3**). Mendel collected the seeds that resulted from each cross and grew them the following season. These offspring were called the **F1**, or the first filial (*filial* = offspring, daughter or son), generation. Once Mendel examined the characteristics in the F1 generation of plants, he allowed them to self-fertilize. He then

collected and grew the seeds from the F1 plants to produce the **F2**, or second filial, generation. Mendel's experiments extended beyond the F2 generation to the F3 and F4 generations, and so on, but it was the ratio of characteristics in the P–F1–F2 generations that were the most intriguing and became the basis for Mendel's principles.

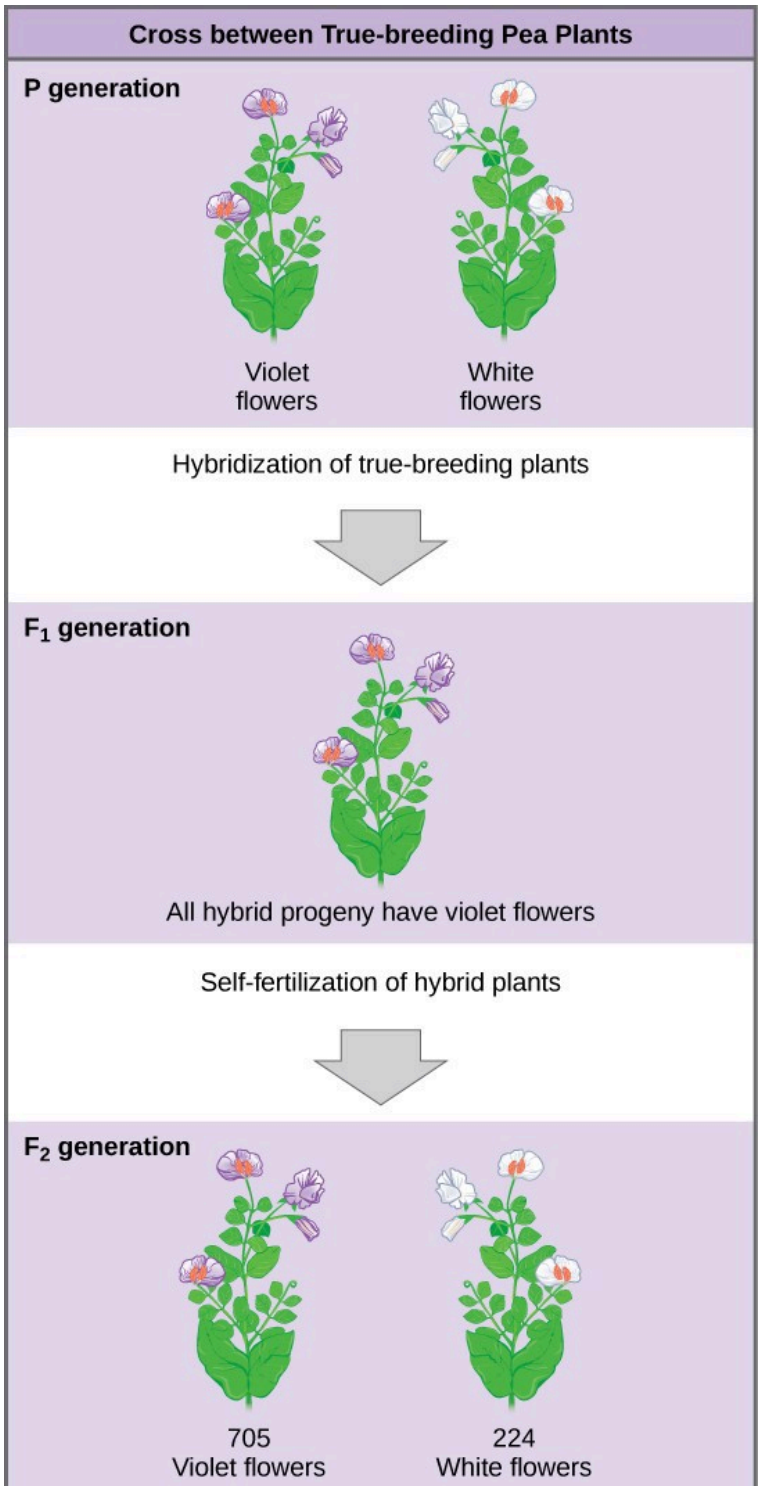

Cross between True-breeding Pea Plants
P generation
Violet flowers
White flowers
Hybridization of true-breeding plants
F1 generation
All hybrid progeny have violet flowers
Self-fertilization of hybrid plants
F2 generation
705
Violet flowers
224
White flowers

Figure 18.3 In one of his experiments on inheritance patterns, Mendel crossed plants that were true-breeding for violet flower color with plants true-breeding for white flower color (the P generation). The resulting hybrids in the F1 generation all had violet flowers. In the F2 generation, approximately three quarters of the plants had violet flowers, and one quarter had white flowers.

18.1.3 Garden Pea Characteristics Revealed the Basics of Heredity

In his 1865 publication, Mendel reported the results of his crosses involving seven different characteristics, each with two contrasting traits. A **trait** is defined as a variation in the physical appearance of a heritable characteristic. The characteristics include: tall vs. short plant height, wrinkled vs. round seeds, green vs. yellow seeds, violet vs. white flowers, etc. (**Table 18.1**). To fully examine each characteristic, Mendel generated large numbers of F1 and F2 plants, reporting results from 19,959 F2 plants alone.

As an example, let us look at Mendel's results for the flower color trait. First, Mendel confirmed that he had plants that bred true for white or violet flower color. Regardless of how many generations Mendel examined, all self-crossed offspring of parents with white flowers had white flowers, and all self-crossed offspring of parents with violet flowers had violet flowers. In addition, Mendel confirmed that, other than flower color, the pea plants were physically identical.

Once these validations were complete, Mendel applied pollen from a plant with violet flowers to the stigma of a plant with white flowers. After gathering and sowing the seeds that resulted from this cross, Mendel found that 100 percent of the F1 hybrid generation had violet flowers. Conventional wisdom at that time would have predicted the hybrid flowers to be pale violet or for hybrid plants to have equal numbers of white and violet flowers. In other words, the contrasting parental traits were expected to blend in the offspring. Instead, Mendel's results demonstrated that the white flower trait in the F1 generation had completely disappeared.

Importantly, Mendel did not stop his experimentation there. He allowed the F1 plants to self-fertilize and found that, of F2- generation plants, 705 had violet flowers and 224 had white flowers. This was a ratio of 3.15 violet flowers per one white flower, or approximately 3:1. When Mendel transferred pollen from a plant with violet flowers to the stigma of a plant with white flowers and vice versa, he obtained about the same ratio regardless of which parent, male or female, contributed which trait. This is called a **reciprocal cross**—a paired cross in which the respective traits of the male and female in one cross become the respective traits of the female and male in the other cross. For the other six characteristics Mendel examined, the F1 and F2 generations behaved in the same way as they had for flower color. One of the two traits would disappear completely from the F1 generation only to reappear in the F2 generation at a ratio of approximately 3:1 (**Table 18.1**).

Table 18.1 The Results of Mendel's Garden Pea Hybridizations

Characteristic	Contrasting P0 Traits	F1 Offspring Traits	F2 C
Flower color	Violet vs. white	100 percent violet	705 v 22
Flower position	Axial vs. terminal	100 percent axial	651 a 20
Plant height	Tall vs. dwarf	100 percent tall	787 t 27
Seed texture	Round vs. wrinkled	100 percent round	5,474 1,8
Seed color	Yellow vs. green	100 percent yellow	6,022 2,0
Pea pod texture	Inflated vs. constricted	100 percent inflated	882 i 29
Pea pod color	Green vs. yellow	100 percent green	428 g 15

18.2 | Mendel's Principles of Inheritance

Learning Objectives

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

- Describe the three principles of inheritance.
- Explain the relationship between phenotype and genotype.
- Develop a Punnett square to calculate the expected proportions of genotypes and phenotypes in a monohybrid cross.
- Explain the purpose and methods of a test cross.
- Draw and interpret a pedigree.

Mendel generalized the results of his pea-plant experiments into three principles that describe the basis of inheritance in diploid organisms. They are: the principle of segregation, the principle of dominance, and the principle of independent

assortment. Together, these principles summarize the basics of classical, or Mendelian, genetics.

18.2.1 The Principle of Segregation

Since the white flower trait reappeared in the F2 generation, Mendel saw that the traits remained separate (not blended) in the plants of the F1 generation. This led to the **principle of segregation**, which states that individuals have two copies of each trait, and that each parent transmits one of its two copies to its offspring.

We now know that the traits that are passed on are a result of genes that are inherited on chromosomes during meiosis and fertilization. The fact that the genetic factors proposed by Mendel were carried on chromosomes was proposed in 1902 by Walter and Sutton and Theodor Boveri (**Figure 18.4**) as the **Chromosomal Theory of Inheritance**.

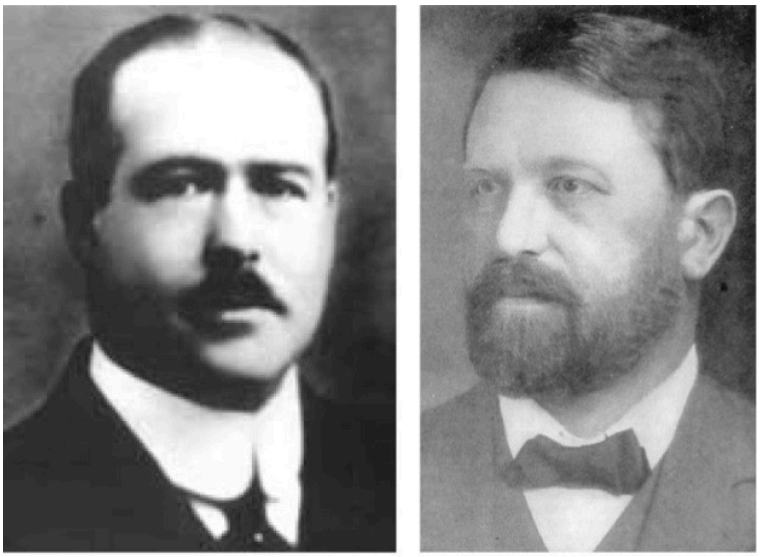


Figure 18.4 (a) Walter Sutton and (b) Theodor Boveri are credited with developing the Chromosomal Theory of Inheritance, which states that chromosomes carry the unit of heredity (genes).

Different versions of genes are called **alleles**. Diploid organisms that have two identical alleles of a gene on their two homologous chromosomes are **homozygous** for that trait. Diploid organisms that have two different alleles of a gene on their two homologous chromosomes are **heterozygous** for that trait.

The physical basis of the principle of segregation is the first division of meiosis, in which the homologous chromosomes with their different versions of each gene are segregated into daughter nuclei. Since each gamete receives only one homolog

of each chromosome, it follows that they receive only one allele for each trait. At fertilization, the zygote receives one of each homologous chromosome, and one of each allele, from each parent.

18.2.2 The Principle of Dominance

Upon compiling his results for many thousands of plants, Mendel concluded that the characteristics could be divided into dominant and recessive traits. **Dominant traits** are those that are expressed in a hybridization. **Recessive traits** become latent, or disappear, in the offspring of a hybridization but reappear in the progeny of the hybrid offspring. Thus, the violet-flower trait is dominant and the white-flower trait is recessive.

Figure 18.5 The child in the photo expresses albinism, a recessive trait.

The **principle of dominance** states that in a heterozygote, only the dominant allele will be expressed. The recessive allele will remain "latent" but will be transmitted to offspring by the same manner in which the dominant allele is transmitted. The recessive trait will only be expressed by offspring that have two copies of this allele (**Figure 18.5**). Individuals with a dominant trait could have either two dominant versions of the trait or one dominant and one recessive version of the trait. Individuals with a recessive trait have two recessive alleles.

In Mendel's experiments, the principle of dominance explains why the F1 heterozygous offspring were identical to one of the parents, rather than expressing both alleles. For a gene that is expressed in a dominant and recessive pattern, homozygous dominant and heterozygous organisms will look identical. The recessive allele will only be observed in homozygous recessive individuals. Some examples of human dominant and recessive traits are shown in **Table 18.2**.

Table 18.2 Examples of dominant and recessive traits in humans.

Dominant Traits	Recessive Traits
Achondroplasia	Albinism
Brachydactyly	Cystic fibrosis
Huntington's disease	Duchenne muscular dystrophy
Marfan syndrome	Galactosemia
Neurofibromatosis	Phenylketonuria
Widow's peak	Sickle-cell anemia
Wooly hair	Tay-Sachs disease

The principles of segregation and dominance could be deduced by simple crosses that follow only one genetic trait. These crosses are called **monohybrid crosses**. Before we discuss the principle of independent assortment, let's look at some tools and terminology used for monohybrid crosses.

18.2.3 Phenotypes and Genotypes

Several conventions exist for referring to genes and alleles. For the purposes of this chapter, we will abbreviate genes using the first letter of the gene's corresponding dominant trait. For

example, green is the dominant trait for pea pod color, so the pod-color gene would be abbreviated as *G* (note that it is customary to italicize gene designations). Furthermore, we will use uppercase and lowercase letters to represent dominant and recessive alleles, respectively. Therefore, we would refer to the genotype of a homozygous dominant pea plant with green pods as *GG*, a homozygous recessive pea plant with yellow pods as *gg*, and a heterozygous pea plant with green pods as *Gg*.

The two alleles for each given gene in a diploid organism may be expressed and interact to produce physical characteristics. The observable traits expressed by an organism are referred to as its **phenotype**. An organism's underlying genetic makeup, which alleles it has, is called its **genotype**. Mendel's hybridization experiments demonstrate the difference between phenotype and genotype. When true-breeding plants in which one parent had yellow pods and one had green pods were cross-fertilized, all of the F1 hybrid offspring had green pods. Although the hybrid offspring had the same phenotype as the true-breeding parent with green pods, we know that the genotype of the parent was homozygous dominant (*GG*), while the genotype of the F1 offspring was heterozygous (*Gg*). We know this since the yellow pod allele reappeared in some of the F2 offspring (*gg*).

18.2.4 Using Punnett Squares for Monohybrid Crosses

Punnett squares, devised by the British geneticist Reginald Punnett, can be used to predict the possible outcomes of a genetic cross or mating and their expected frequencies. To demonstrate a monohybrid cross, consider the case of true-breeding pea plants with yellow versus green pea seeds. The dominant seed color is yellow; therefore, the parental genotypes were *YY* for the plants with yellow seeds and *yy* for the plants with green seeds, respectively. To prepare a Punnett square, all possible combinations of the parental alleles are listed along the top (for one parent) and side (for the other parent) of a grid, representing their meiotic segregation into haploid gametes. Then the combinations of egg and sperm are made in the boxes in the table to show which alleles are combining. Each box then represents the diploid genotype of a zygote, or fertilized egg, that could result from this mating. Because each possibility is equally likely, genotypic ratios can be determined from a Punnett square. If the pattern of inheritance (dominant or recessive) is known, the phenotypic ratios can be inferred as well. For a monohybrid cross of two true-breeding parents, each parent contributes one type of allele. In this case, only one genotype is possible. All offspring are *Yy* and have yellow seeds (**Figure 18.6**).

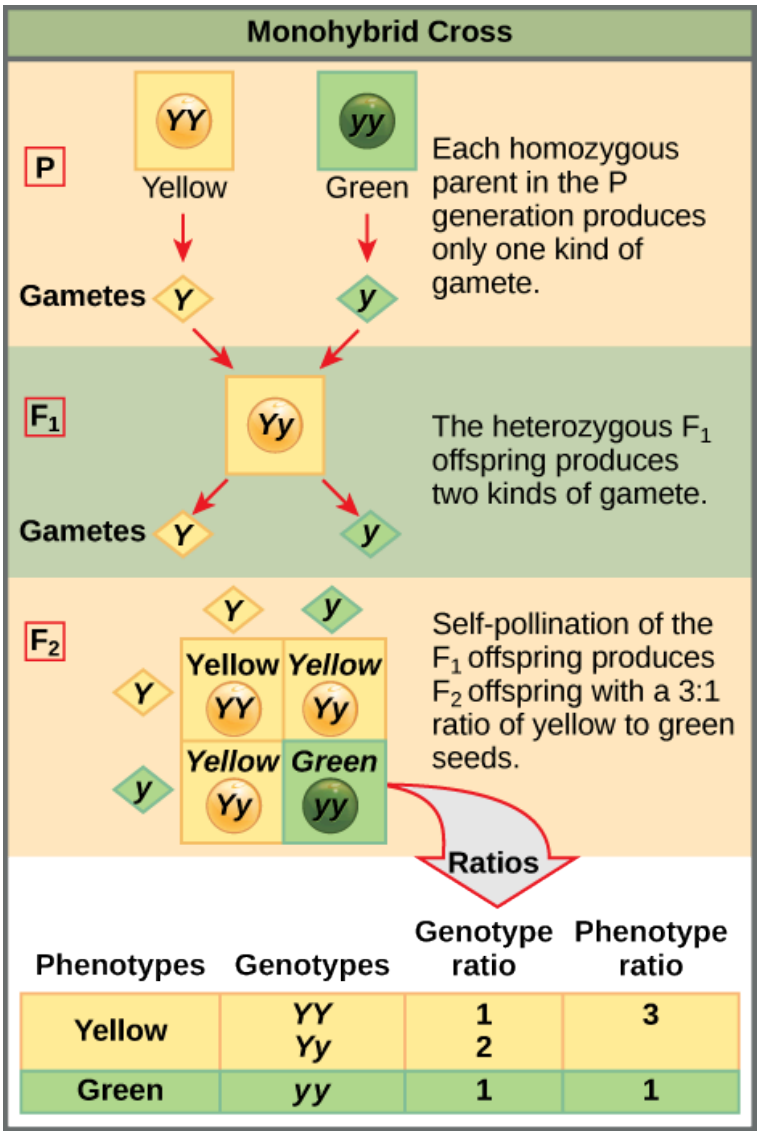


Figure 18.6 In the P generation, pea plants that are true-breeding for the dominant yellow phenotype are crossed with plants with the recessive green phenotype. This cross produces F1 heterozygotes with a yellow phenotype.

Punnett square analysis can be used to predict the genotypes of the F2 generation.

A self-cross of one of the *Yy* heterozygous offspring can be represented in a 2 × 2 Punnett square because each parent can donate one of two different alleles. Therefore, the offspring can potentially have one of four allele combinations: *YY*, *Yy*, *yY*, or *yy* (**Figure 18.6**). Notice that there are two ways to obtain the *Yy* genotype: a *Y* from the egg and a *y* from the sperm, or a *y* from the egg and a *Y* from the sperm. Both of these possibilities must be counted. Recall that Mendel's pea- plant characteristics behaved in the same way in reciprocal crosses. Therefore, the two possible heterozygous combinations produce offspring that are genotypically and phenotypically identical despite their dominant and recessive alleles deriving from different parents.

Because fertilization is a random event, we expect each combination to be equally likely and for the offspring to exhibit a ratio of *YY*:*Yy*:*yy* genotypes of 1:2:1 (**Figure 18.6**). Furthermore, because the *YY* and *Yy* offspring have yellow seeds and are phenotypically identical, we expect the offspring to exhibit a phenotypic ratio of 3 yellow:1 green. Indeed, working with large sample sizes, Mendel observed approximately this ratio in every F2 generation resulting from crosses for individual traits.

Using a Test Cross to Determine Genotype

Beyond predicting the offspring of a cross between known homozygous or heterozygous parents, Mendel also developed a way to determine whether an organism that expressed a dominant trait was a heterozygote or a homozygote. Called the **test cross**, this technique is still used by plant and animal breeders. In a test cross, an organism with the dominant phenotype is crossed with an organism that is homozygous recessive for the same characteristic. If the dominant-expressing organism is a homozygote, then all F1 offspring will be heterozygotes expressing the dominant trait. Alternatively, if the dominant expressing organism is a heterozygote, the F1 offspring will exhibit a 1:1 ratio of heterozygotes and recessive homozygotes (**Figure 18.7**). The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.

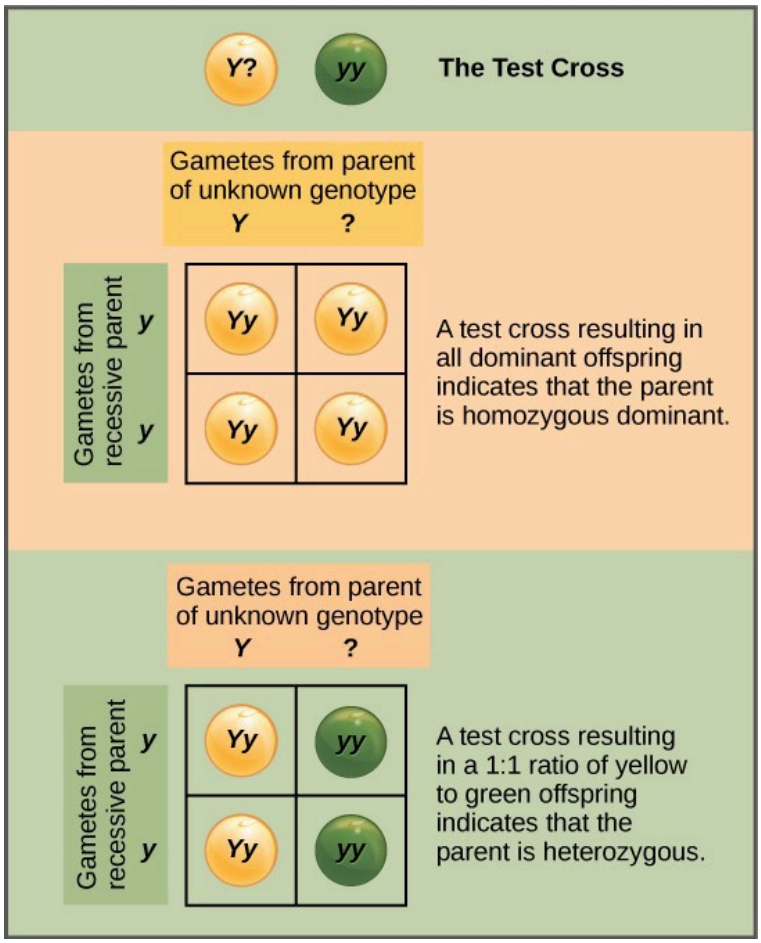


Figure 18.7 A test cross can be performed to determine whether an organism expression a dominant trait is a homozygote or a heterozygote.

Concept Check

In pea plants, round peas (R) are dominant to wrinkled peas (r). You do a test cross between a pea plant with wrinkled peas (genotype rr) and a plant of unknown genotype that has round peas. You end up with three plants, all which have round peas.

- From this data, can you tell if the round pea parent plant is homozygous dominant or heterozygous?
- If the round pea parent plant is heterozygous, what is the probability that a random sample of 3 progeny peas will all be round?

18.2.5 Using Pedigrees to Study Inheritance Patterns

Many human diseases are inherited genetically. A healthy person in a family in which some members suffer from a recessive genetic disorder may want to know if he or she has the disease-causing gene and what risk exists of passing the

disorder on to his or her offspring. Of course, doing a test cross in humans is unethical and impractical. Instead, geneticists use **pedigree analysis** to study the inheritance pattern of human genetic diseases.

Each row of a pedigree represents one generation of the family. Women are represented by circles; males by squares. People who had children together are connected with a horizontal line and their children are connected to this line with a vertical line. See **Figure 18.8** for an example of a pedigree for a human genetic disease.

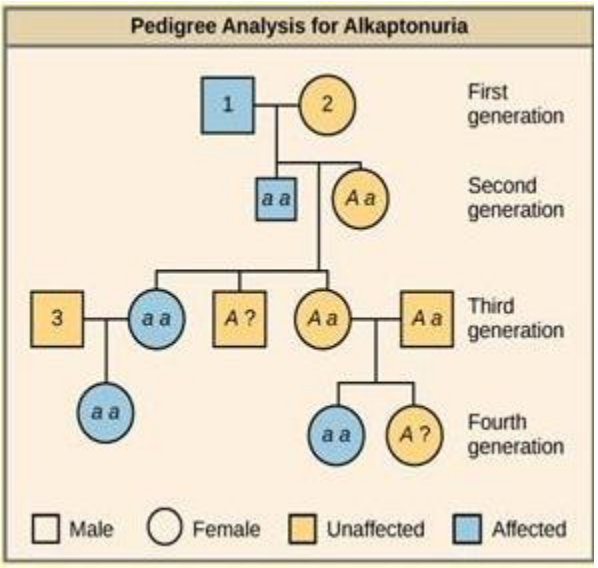


Figure 18.8 Pedigree of a human family with the recessive genetic disease alkaptonuria.

People with the recessive genetic disease alkaptonuria cannot properly metabolize two amino acids, phenylalanine and tyrosine. Affected individuals may have darkened skin and brown urine, and may suffer joint damage and other complications.

In this pedigree, individuals with the disorder are indicated in blue and have the genotype *aa*. Unaffected individuals are indicated in yellow

and have the genotype *AA* or *Aa*. Note that it is often possible to determine a person's genotype from the genotype of their offspring. For example, if neither parent has the disorder but their child does, both parents must be heterozygous. Two individuals on the pedigree have an unaffected phenotype but unknown genotype. Because they do not have the disorder, they must have at least one normal allele, so their genotype gets the "*A?*" designation.

Concept Check

What are the genotypes of the individuals labeled 1, 2, and 3?

18.2.6 Principle of Independent Assortment

Mendel's **principle of independent assortment** states that genes do not influence each other with regard to the sorting of alleles into gametes, and every possible combination of alleles for every gene is equally likely to occur. The independent assortment of genes can be illustrated by a **dihybrid** cross, a cross between two true-breeding parents that express different traits for two characteristics. Consider the characteristics of seed color and seed texture for two pea plants, one that has green, wrinkled seeds (*yyrr*) and another that has yellow, round seeds (*YYRR*). Because each parent is homozygous, the principle of segregation indicates that the gametes for the green/wrinkled plant all are *yr*, and the gametes for the yellow/round plant are all *YR*. Therefore, the F1 generation of offspring all are *YyRr* (**Figure 18.9**).

For the F2 generation, the principle of segregation requires that each gamete receive either an *R* allele or an *r* allele along with either a *Y* allele or a *y* allele. The principle of independent assortment states that a gamete into which an *r* allele sorted would be equally likely to contain either a *Y* allele or a *y* allele. Thus, there are four equally likely gametes that can be formed when the *YyRr* heterozygote is self-crossed, as follows: *YR*, *Yr*, *yR*, and *yr*. Arranging these gametes along the top and left of a 4 × 4 Punnett square gives us 16 equally likely genotypic combinations. From these genotypes, we infer a phenotypic

ratio of 9 round/yellow:3 round/green:3 wrinkled/yellow:1 wrinkled/green (**Figure 18.9**).

The physical basis for the principle of independent assortment also lies in meiosis I, in which the different homologous pairs line up in random orientations. Each gamete can contain any combination of paternal and maternal chromosomes (and therefore the genes on them) because the orientation of tetrads on the metaphase plane is random.

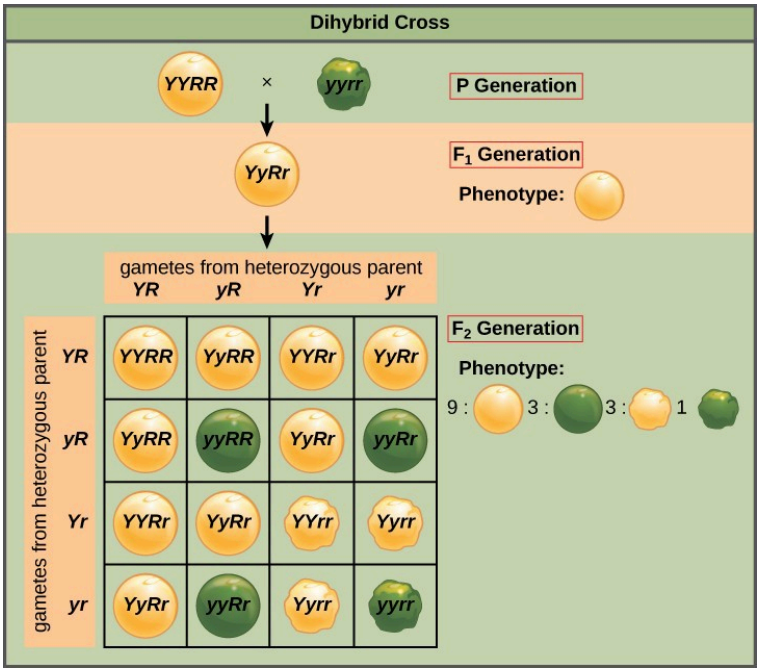


Figure 18.9 In a dihybrid cross, two traits are followed in a single cross. Here, both seed color and seed smoothness are followed into the F2 generation.

Testing the Hypothesis of Independent Assortment

To better appreciate the amount of labor and ingenuity that went into Mendel's experiments, proceed through one of Mendel's dihybrid crosses.

Question: What will be the offspring of a dihybrid cross?

Background: Consider that you have access to a large garden in which you can cultivate thousands of pea plants. There are several true-breeding plants with the following pairs of traits: tall plants with inflated pods, and dwarf plants with constricted pods. Before the plants have matured, you remove the pollen-producing organs from the tall/inflated plants in your crosses to prevent self-fertilization. When the plants mature, they are manually crossed by transferring pollen from the dwarf/constricted plants to the stigmata of the tall/inflated plants.

Hypothesis: Both trait pairs will sort independently according to Mendelian principles. When the true-breeding parents are crossed, all of the F1 offspring are tall and have inflated pods, which indicates that the tall (T) and inflated (I) traits are dominant over the dwarf (t) and constricted (i) traits, respectively. A self-cross of the F1 heterozygotes results in 2,000 F2 progeny.

Test the hypothesis: You cross the dwarf and tall plants and then self-cross the offspring. For best results,

this is repeated with hundreds or even thousands of pea plants. What special precautions should be taken in the crosses and in growing the plants?

If these traits sort independently, the ratios of tall:dwarf and inflated:constricted will each be 3:1. Each member of the F1 generation therefore has a genotype of *TtIi*. **Figure 18.10** shows a cross between two *TtIi* individuals. There are 16 possible offspring genotypes. The offspring proportions: tall/inflated:tall/constricted:dwarf/inflated:dwarf/constricted show a 9:3:3:1 ratio. Notice from the grid that when considering the tall/dwarf and inflated/constricted trait pairs in isolation, they are each inherited in 3:1 ratios.

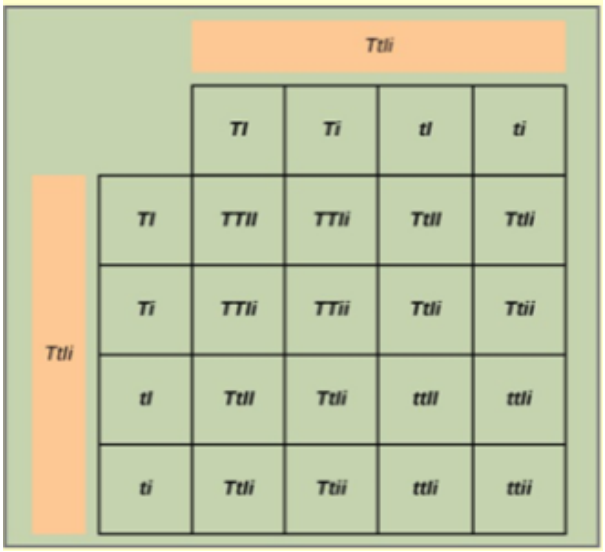


Figure 18.10 This figure shows all possible combinations of offspring resulting from a dihybrid cross of pea plants that are heterozygous for the tall/dwarf and inflated/constricted alleles.

Analyze your data: You observe the following plant phenotypes in the F2 generation: 2706 tall/inflated, 930 tall/constricted, 888 dwarf/inflated, and 300 dwarf/constricted. Reduce these findings to a ratio and determine if they are consistent with Mendelian principles.

Form a conclusion: Were the results close to the expected 9:3:3:1 phenotypic ratio? Do the results support the prediction? What might be observed if far fewer plants were used, given that alleles segregate

randomly into gametes? Try to imagine growing that many pea plants, and consider the potential for experimental error. For instance, what would happen if it was extremely windy one day?

18.3 | Exceptions to Mendel's Principles of Inheritance

Learning Objectives

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

- Identify non-Mendelian inheritance patterns such as incomplete dominance, codominance, and sex linkage.
- Describe genetic linkage.
- Describe how chromosome maps are created.
- Explain the phenotypic outcomes of epistatic effects between genes.

Although Mendel's principles still apply to some situations,

many situations exist in which they do not apply. These "exceptions" to Mendelian genetics are discussed below.

18.3.1 Alternatives to Dominance and Recessiveness

Since Mendel's experiments with pea plants, other researchers have found that the principle of dominance does not always hold true. Instead, several different patterns of inheritance have been found to exist.

Incomplete Dominance

Figure 18.11 These pink flowers of a heterozygote snapdragon result from incomplete dominance. (Credit: "storebukkebruse"/Flickr)

Mendel's results, that traits are inherited as dominant and recessive pairs, contradicted the view at that time that offspring exhibited a blend of their parents' traits. However, the heterozygote phenotype occasionally does appear to be intermediate between the two parents. For example, in the snapdragon, *Antirrhinum majus* (**Figure 18.11**), a cross between a homozygous parent with white flowers (*CWCW*) and a homozygous parent with red flowers (*CRCR*) will produce offspring with pink flowers (*CRCW*). (Note that different genotypic abbreviations are used for Mendelian extensions to distinguish these patterns from simple dominance and recessiveness.) This pattern of inheritance is described as **incomplete dominance**, denoting the expression of two contrasting alleles such that the individual displays an intermediate phenotype. The allele for red flowers is incompletely dominant over the allele for white

flowers. However, the results of a heterozygote self-cross can still be predicted, just as with Mendelian dominant and recessive crosses. In this case, the genotypic ratio would be 1 *CRCR*:2 *CRCW*:1 *CWCW*, and the phenotypic ratio would be 1:2:1 for red:pink:white.

Codominance

A variation on incomplete dominance is **codominance**, in which both alleles for the same characteristic are simultaneously expressed in the heterozygote. An example of codominance is the MN blood groups of humans. The M and N alleles are expressed in the form of an M or N antigen present on the surface of red blood cells. Homozygotes (*LMLM* and *LNLN*) express either the M or the N allele, and heterozygotes (*LMLN*) express both alleles equally. In a self-cross between heterozygotes expressing a codominant trait, the three possible offspring genotypes are phenotypically distinct. However, the 1:2:1 genotypic ratio characteristic of a Mendelian monohybrid cross still applies.

Multiple Alleles

Mendel implied that only two alleles, one dominant and one recessive, could exist for a given gene. We now know that this is an oversimplification. Although individual humans (and all diploid organisms) can only have two alleles for a given gene, multiple alleles may exist at the population level such that

many combinations of two alleles are observed. Note that when many alleles exist for the same gene, the convention is to denote the most common phenotype or genotype among wild animals as the **wild type** (often abbreviated "+"); this is considered the standard or norm. All other phenotypes or genotypes are considered **variants** of this standard, meaning that they deviate from the wild type. The variant may be recessive or dominant to the wild-type allele.

An example of multiple alleles is coat color in rabbits (**Figure 18.12**). Here, four alleles exist for the *c* gene. The wild-type version, C^+C^+, is expressed as brown fur. The chinchilla phenotype, $c^{ch}c^{ch}$, is expressed as black-tipped white fur. The Himalayan phenotype, c^hc^h, has black fur on the extremities and white fur elsewhere. Finally, the albino, or "colorless" phenotype, *cc*, is expressed as white fur. In cases of multiple alleles, dominance hierarchies can exist. In this case, the wild- type allele is dominant over all the others, chinchilla is incompletely dominant over Himalayan and albino, and Himalayan is dominant over albino. This hierarchy, or allelic series, was revealed by observing the phenotypes of each possible heterozygote offspring.

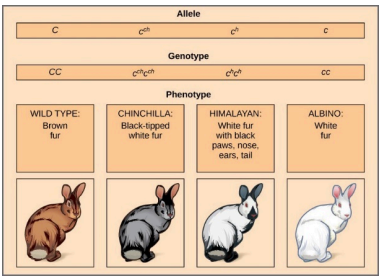


Figure 18.12 Four different alleles exist for the rabbit coat color (C) gene.

An example of multiple allelism in humans pertains to ABO blood type. A person's blood type (e.g., type A or type O) is caused by different combinations of three alleles: IA, IB, and IO. A person with type A blood could have either IAIA or IAIO genotype. A person with type B blood could have IBIB or IBIO genotype. A person with type O blood must have the IOIO genotype. Note that type AB blood is an example of codominance (IAIB).

The complete dominance of a wild-type phenotype over all other mutants often occurs as an effect of "dosage" of a specific gene product, such that the wild-type allele supplies the correct amount of gene product whereas the mutant alleles cannot. For rabbit fur color, the wild-type allele may supply a given dosage of fur pigment, whereas the mutants supply a lesser dosage or none at all.

evolution CONNECTION

Multiple Alleles Confer Drug Resistance in the Malaria Parasite

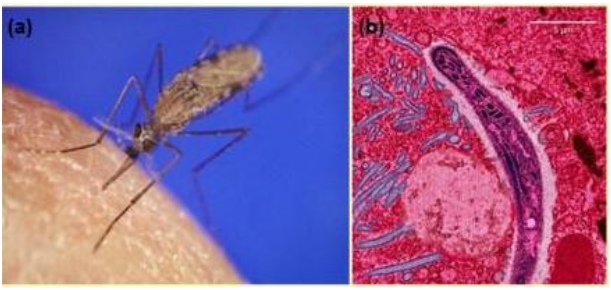


Figure 18.13 (a) The mosquito *Anopheles gambiae* transmits the malaria-causing parasite to humans. (Credit: James D. Gathany) (b) The malaria parasite, *Plasmodium falciparum*, visualized by false-color transmission electron microscopy. (Credit: Ute Frevert; false color by Margaret Shear; scale-bar data from Matt Russell)

Malaria is a parasitic disease that is transmitted to humans by infected female *Anopheles gambiae* mosquitos (**Figure 18.13a**). It is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* is the most deadly causative agent of malaria (**Figure 18.13b**). When promptly and correctly treated, *P. falciparum* malaria has a mortality rate of 0.1%. However, in some parts of the world, the parasite has evolved resistance to commonly used malaria treatments, so the most

effective malarial treatments can vary by geographic region.

In Southeast Asia, Africa, and South America, *P. falciparum* has developed resistance to the anti-malarial drugs chloroquine, mefloquine, and sulfadoxine-pyrimethamine. *P. falciparum*, which is haploid during the life stage in which it infects humans, has evolved multiple drug-resistant mutant alleles of the *dhps* gene. Varying degrees of sulfadoxine resistance are associated with each of these alleles. Being haploid, *P. falciparum* needs only one drug-resistant allele to express this trait.

In Southeast Asia, different sulfadoxine-resistant alleles of the *dhps* gene are localized to different geographic regions. This is a common evolutionary phenomenon that occurs because drug-resistant mutants arise in a population and interbreed with other *P. falciparum* isolates in close proximity. Sulfadoxine-resistant parasites cause considerable human hardship in regions where this drug is widely used as an over-the-counter malaria remedy. As is common with pathogen that multiply to large numbers within

an infection cycle, *P. falciparum* evolves relatively rapidly (over a decade or so) in response to the selective pressure of commonly used anti-malarial drugs. For this reason, scientists must constantly work to develop new drugs or drug combinations to combat the worldwide malaria burden.

Environmental Effects

Interestingly, the Himalayan phenotype in rabbits is the result of an allele that produces a temperature-sensitive gene product that only produces pigment in the cooler extremities of the rabbit's body. In this case, the protein product of the gene does not fold correctly at high temperatures. A similar gene gives Siamese cats their distinctive coloration.

Temperature-sensitive proteins are also at work in arctic foxes and rabbits, which are white in the winter and darker colored during the summer. In these cases, the protein product of the gene does not fold correctly at colder temperatures. The mutation that caused this coloration was advantageous to these species, so they persisted in the populations.

18.3.2 X-Linked Traits are an Exception to the Principle of Segregation

Figure 18.14 In Drosophila, several genes determine eye color. The genes for white and vermilion eye colors are located on the X chromosome. Others are located on the autosomes. Clockwise from top right are brown, cinnabar, sepia, vermilion, white, and red. Red eye color is wild-type and is dominant to white eye color.

In humans, as well as in many other animals and some plants, the sex of the individual is determined by sex chromosomes. The sex chromosomes are one pair of non-homologous chromosomes. Until now, we have only considered inheritance patterns among non-sex chromosomes, or autosomes. In addition to 22 homologous pairs of autosomes, human females have a homologous pair of X chromosomes, whereas human males have an XY chromosome pair. Although the Y chromosome contains a small region of similarity to the X chromosome so that they can pair during meiosis, the Y chromosome is much shorter and contains many fewer genes. When a gene is present on the X chromosome, it is said to be **X-linked**.

Eye color in *Drosophila* was one of the first X-linked traits

to be identified. Like humans, *Drosophila* males are XY and females are XX. In flies, the wild-type eye color is red (X^W) which is dominant to white eye color (X^w) (**Figure 18.14)**. Females can be X^WX^W, X^WX^w or X^wX^w. However, *Drosophila* males lack a second allele copy on the Y chromosome, so their genotype can only be X^WY or X^wY. Males are said to be **hemizygous**, because they have only one allele for any X- linked characteristic. Hemizygosity makes the descriptions of dominance and recessiveness irrelevant for XY males.

In an X-linked cross, the genotypes of F1 and F2 offspring depend on whether the recessive trait was expressed by the male or the female in the P generation. When the P male expresses the white-eye phenotype and the female is homozygous red-eyed, all members of the F1 generation exhibit red eyes (**Figure 18.15**). The F1 females are heterozygous (X^WX^w), and the males are all X^WY, since they received their X chromosome from the homozygous dominant P female and their Y chromosome from the P male. A cross between a X^WX^w female and an X^WY male would produce only red-eyed females and both red- and white-eyed males. A cross between a homozygous white-eyed female and a male with red eyes would produce only heterozygous red-eyed females and only white-eyed males.

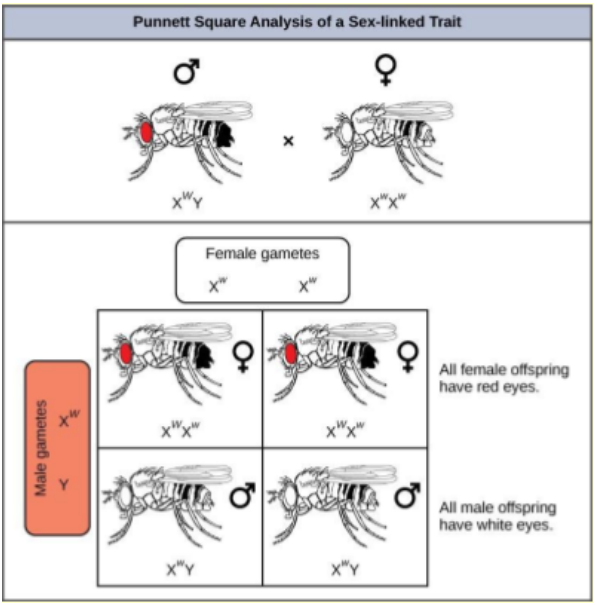


Figure 18.15 Punnett square analysis is used to determine the ratio of offspring from a cross between a red-eyed male fruit fly and a white-eyed female fruit fly.

Concept Check

What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

In some groups of organisms with sex chromosomes, the gender with the non-homologous sex chromosomes is the female rather than the male. This is the case for all birds. In this case, sex-linked traits will be more likely to appear in the female, in which they are hemizygous.

Human Sex-linked Disorders

Sex-linkage studies in Morgan's laboratory provided the fundamentals for understanding X-linked recessive disorders in humans, which included red-green color blindness, Types A and B hemophilia, and muscular dystrophy. Because human males need to inherit only one recessive mutant X allele to be affected, X-linked disorders are disproportionately observed in males. Females must inherit recessive X-linked alleles from both of their parents in order to express the trait. When they inherit one recessive X-linked mutant allele and one dominant X-linked wild-type allele, they are carriers of the trait and are typically unaffected. Carrier females can manifest mild forms of the trait due to the inactivation of the dominant allele located on one of the X chromosomes. However, female carriers can contribute the trait to their sons, resulting in the son exhibiting the trait, or they can contribute the recessive allele to their daughters, resulting in the daughters being carriers of the trait (**Figure 18.16**). Although some Y-linked recessive disorders exist, typically they are associated with infertility in males and are therefore not transmitted to subsequent generations.

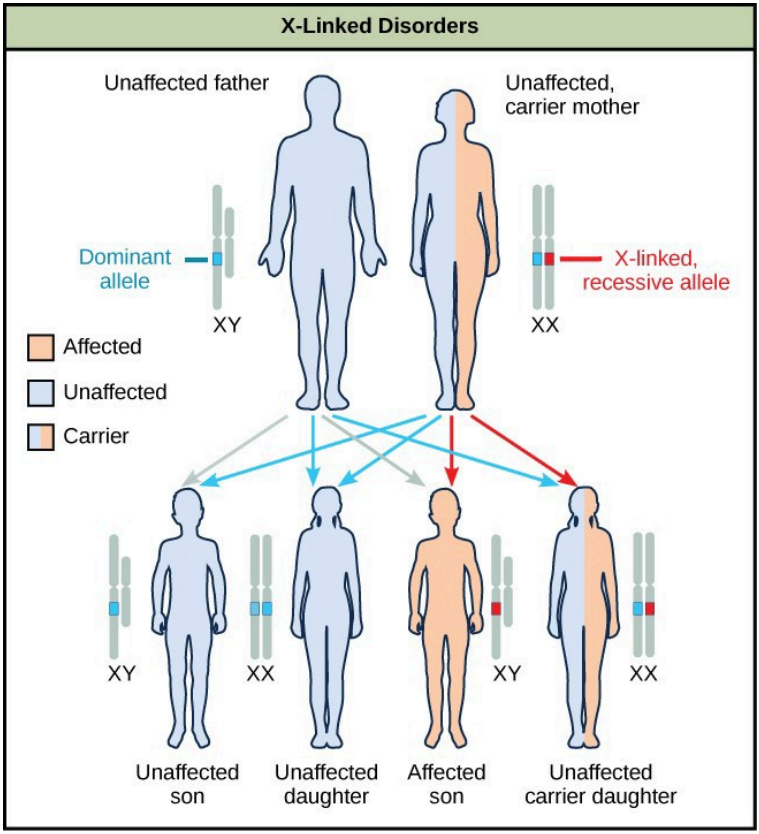


Figure 18.16 The son of a woman who is a carrier of a recessive X-linked disorder will have a 50 percent chance of being affected. A daughter will not be affected, but she will have a 50 percent chance of being a carrier like her mother.

18.3.3 Lethal Alleles are Apparent Exceptions to the Principle of Segregation

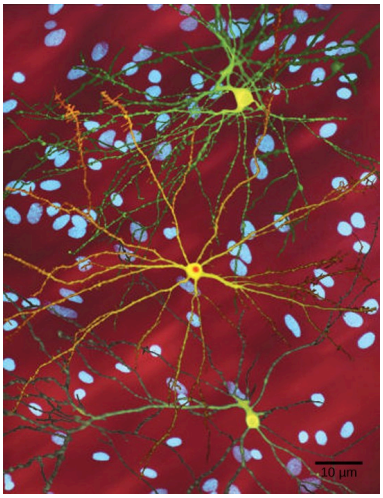


Figure 18.17 The neuron in the center of this micrograph (yellow) has nuclear inclusions characteristic of Huntington disease (orange area in the center of the neuron). Huntington disease occurs when an abnormal dominant allele for the Huntington gene is present. (Credit: Dr. Steven Finkbeiner, Gladstone Institute of Neurological Disease, The Taube- Koret Center for Huntington Disease Research, and the University of California San Francisco/ Wikimedia)

A large proportion of genes in an individual's genome are essential for survival. Occasionally, a nonfunctional allele for an essential gene can arise by mutation and be transmitted in a population through heterozygous carriers. The wild-type allele functions at a capacity sufficient to sustain life and is therefore considered to be dominant over the nonfunctional allele. If two heterozygous parents mate, one quarter of their offspring will be homozygous recessive. Because the gene is essential, these individuals will die. This will cause the genotypic ratio among surviving offspring to be 2:1 rather than 3:1. This inheritance pattern is referred to as **recessive lethal**.

The **dominant lethal** inheritance pattern is one in which

an allele is lethal both in the homozygote and the heterozygote. Dominant lethal alleles are very rare because, as you might expect, the allele only lasts one generation and is not transmitted. However, dominant lethal alleles might not be expressed until adulthood. The allele may be unknowingly passed on, resulting in a delayed death in both generations. An example of this in humans is Huntington disease, in which the nervous system gradually wastes away (**Figure 18.17**). People who are heterozygous for the dominant Huntington allele (*Hh*) will inevitably develop the fatal disease. However, the onset of Huntington disease may not occur until age 40, at which point the afflicted persons may have already passed the allele to 50 percent of their offspring.

18.3.4 Linked Genes Violate the Principle of Independent Assortment

Although all of Mendel's pea characteristics behaved according to the principle of independent assortment, we now know that some allele combinations are not inherited independently of each other. Genes that are located on different chromosomes will always sort independently. However, each chromosome contains hundreds or thousands of genes, organized linearly on chromosomes like beads on a string. Genes that are on the same chromosome are **linked**

and are therefore likely to be inherited together. When homologs separate during meiosis I, entire chromosomes segregate into separate daughter cells, carrying all of their linked genes with them.

However, because of crossover, it is possible for two genes on the same chromosome to behave independently, or as if they are not linked. To understand this, let's consider the biological basis of gene linkage and recombination.

Homologous chromosomes possess the same genes in the same order. However, since each homolog came from a different parent, the alleles may differ on homologous chromosome pairs. Prior to meiosis I, homologous chromosomes replicate and synapse so that genes on the homologs align with each other. At this stage, segments of homologous chromosomes cross over and exchange segments of genetic material (**Figure 18.18**). Because the genes are aligned, the gene order is not altered. Instead, the result of recombination is that maternal and paternal alleles are combined onto the same chromosome. Across a given chromosome, several recombination events may occur, causing extensive shuffling of alleles.

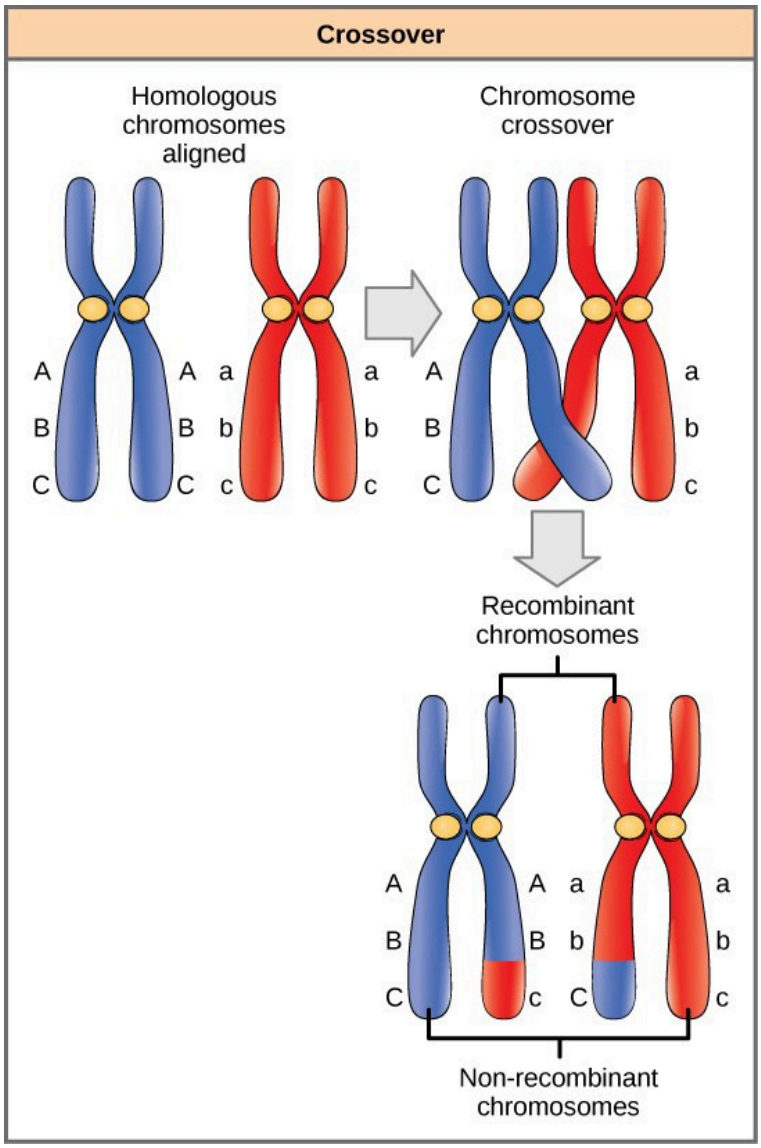


Figure 18.18 The process of crossover, or recombination, occurs when two homologous chromosomes align during meiosis and exchange a segment of genetic material. Here, the alleles for gene C were exchanged. The result is two

recombinant and two non-recombinant chromosomes.

When two genes are located in close proximity on the same chromosome, their alleles are more likely to be transmitted through meiosis together. To exemplify this, imagine a dihybrid cross involving flower color and plant height in which the genes are next to each other on the chromosome. If the homologous chromosome from one parent has alleles for tall plants and red flowers, and the homolog from the other parent has alleles for short plants and yellow flowers, then when the gametes are formed, the tall and red alleles will go together into a gamete and the short and yellow alleles will go into other gametes. These are called the **parental genotypes** because they have been inherited intact from the parents of the individual producing gametes. Since the genes were close together on the same chromosomes, the chance of a crossover event happening between them is slim. Therefore, there will be no gametes with tall and yellow alleles and no gametes with short and red alleles. If you create the Punnett square with these gametes, you will see that the classical Mendelian prediction of a 9:3:3:1 outcome of a dihybrid cross would not apply

As the distance between two genes increases, the probability of crossovers between them increases, and the genes behave more as if they are on separate chromosomes. The further apart two linked genes are on a chromosome, the more progeny with **nonparental genotypes** will appear.

Genetic Linkage and Distances

Geneticists have used the proportion of nonparental gametes as a measure of how far apart genes are on a chromosome. Using this information, they have constructed elaborate maps of genes on chromosomes. Briefly, the more crossover that occurs between two linked genes, the further apart they are on the chromosome. The frequency of crossover is measured by counting the number of offspring that have nonparental genotypes. By using recombination frequency to predict genetic distance, the relative order of genes on chromosome 2 could be inferred.

18.3.5 Epistasis is an Exception to the Principle of Independent Assortment

Mendel's studies in pea plants implied that every characteristic was distinctly and completely controlled by a single gene. In fact, single observable characteristics are almost always under the influence of multiple genes (each with two or more alleles) acting in unison. For example, at least eight genes contribute to eye color in humans.

Genes may function in complementary or synergistic fashions, such that two or more genes need to be expressed simultaneously to affect a phenotype. Genes may also oppose each other. In **epistasis**, the interaction between genes is

antagonistic, such that one gene masks or interferes with the expression of another. Often the biochemical basis of epistasis is a gene pathway in which the expression of one gene is dependent on the function of a gene that precedes or follows it in the pathway.

An example of epistasis is pigmentation in mice. The wild-type coat color, agouti (*AA*), is dominant to solid-colored fur (*aa*). However, a separate gene (*C*) is necessary for pigment production. A mouse with a recessive *c* allele at this locus is unable to produce pigment and is albino regardless of the allele present at locus *A*. Therefore, the genotypes *AAcc*, *Aacc*, and *aacc* all produce an albino phenotype. A cross between heterozygotes for both genes (*AaCc* x *AaCc*) would generate offspring with a phenotypic ratio of 9 agouti:3 solid color:4 albino (**Figure 18.19**). In this case, the *C* gene is epistatic to the *A* gene.

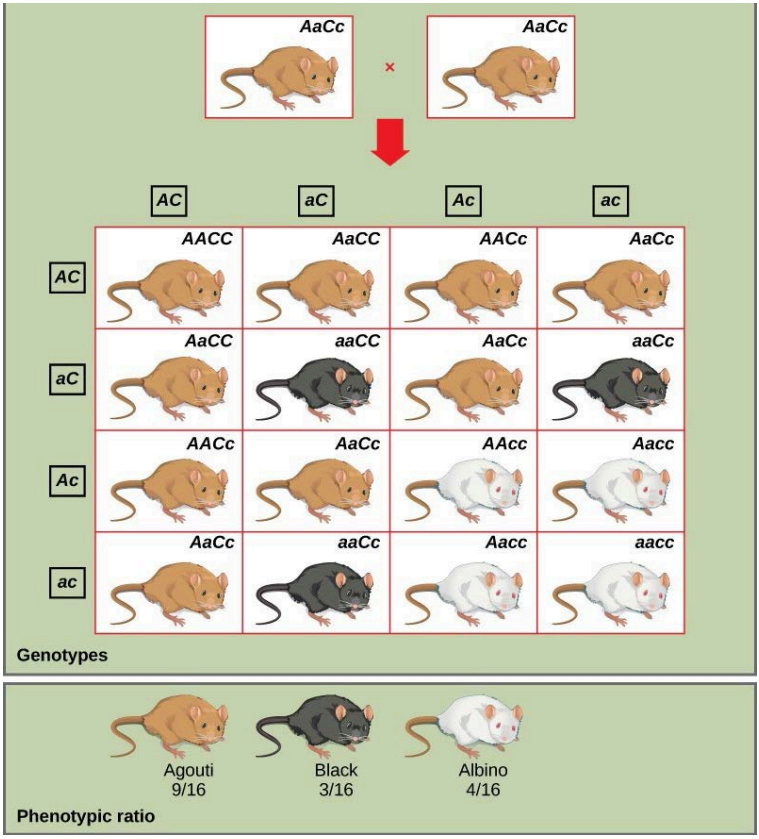


Figure 18.19 In mice, the mottled agouti coat color (A) is dominant to a solid coloration, such as black or gray. A gene at a separate locus (C) is responsible for pigment production. The recessive c allele does not produce pigment, and a mouse with the homozygous recessive cc genotype is albino regardless of the allele present at the A locus. Thus, the C gene is epistatic to the A gene.

Epistasis can also occur when a dominant allele masks expression at a separate gene. Fruit color in summer squash is expressed in this way. Homozygous recessive expression of

the *W* gene (*ww*) coupled with homozygous dominant or heterozygous expression of the *Y* gene (*YY* or *Yy*) generates yellow fruit, and the *wwyy* genotype produces green fruit. However, if a dominant copy of the *W* gene is present in the homozygous or heterozygous form, the summer squash will produce white fruit regardless of the *Y* alleles. A cross between white heterozygotes for both genes (*WwYy* × *WwYy*) would produce offspring with a phenotypic ratio of 12 white:3 yellow:1 green.

Finally, epistasis can be reciprocal such that either gene, when present in the dominant (or recessive) form, expresses the same phenotype. In the shepherd's purse plant (*Capsella bursa-pastoris*), the characteristic of seed shape is controlled by two genes in a dominant epistatic relationship. When the genes *A* and *B* are both homozygous recessive (*aabb*), the seeds are ovoid. If the dominant allele for either of these genes is present, the result is triangular seeds. That is, every possible genotype other than *aabb* results in triangular seeds, and a cross between heterozygotes for both genes (*AaBb* x *AaBb*) would yield offspring with a phenotypic ratio of 15 triangular:1 ovoid.

As you work through genetics problems, keep in mind that any single characteristic that results in a phenotypic ratio that totals 16 is typical of a two-gene interaction. Recall the phenotypic inheritance pattern for Mendel's dihybrid cross, which considered two non-interacting genes—9:3:3:1. Similarly, we would expect interacting gene pairs to also

exhibit ratios expressed as 16 parts. Note that we are assuming the interacting genes are not linked; they are still assorting independently into gametes.