BI 112 - Cell Biology for Health Occupations

OER Textbook

Content updated 9/20/2019

Print version generated 10/18/2019

1. Introduction and Organizing Principles

Overview of Anatomy and Physiology

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

- Compare and contrast anatomy and physiology, including their specializations and methods of study
- Discuss the fundamental relationship between anatomy and physiology

Human anatomy is the scientific study of the body's structures. Some of these structures are very small and can only be observed and analyzed with the assistance of a microscope. Other larger structures can readily be seen, manipulated, measured, and weighed. The word "anatomy" comes from a Greek root that means "to cut apart." Human anatomy was first studied by observing the exterior of the body and observing the wounds of soldiers and other injuries. Later, physicians were allowed to dissect bodies of the dead to augment their knowledge. When a body is dissected, its structures are cut apart in order to observe their physical attributes and their relationships to one another. Dissection is still used in medical schools, anatomy courses, and in pathology labs. In order to observe structures in living people, however, a number of imaging techniques have been developed. These techniques allow clinicians to visualize structures inside the living body such as a cancerous tumor or a fractured bone.

Like most scientific disciplines, anatomy has areas of specialization. Gross anatomy is the study of the larger structures of the body, those visible without the aid of magnification (Figure 1.2a). Macro- means "large," thus, gross anatomy is also referred to as macroscopic anatomy. In contrast, micro- means "small," and microscopic anatomy is the study of structures that can be observed only with the use of a microscope or other magnification devices (Figure 1.2b). Microscopic anatomy includes cytology, the study of cells and histology, the study of tissues. As the technology of microscopes has advanced, anatomists have been able to observe smaller and smaller structures of the body, from slices of large structures like the heart, to the three-dimensional structures of large molecules in the body.

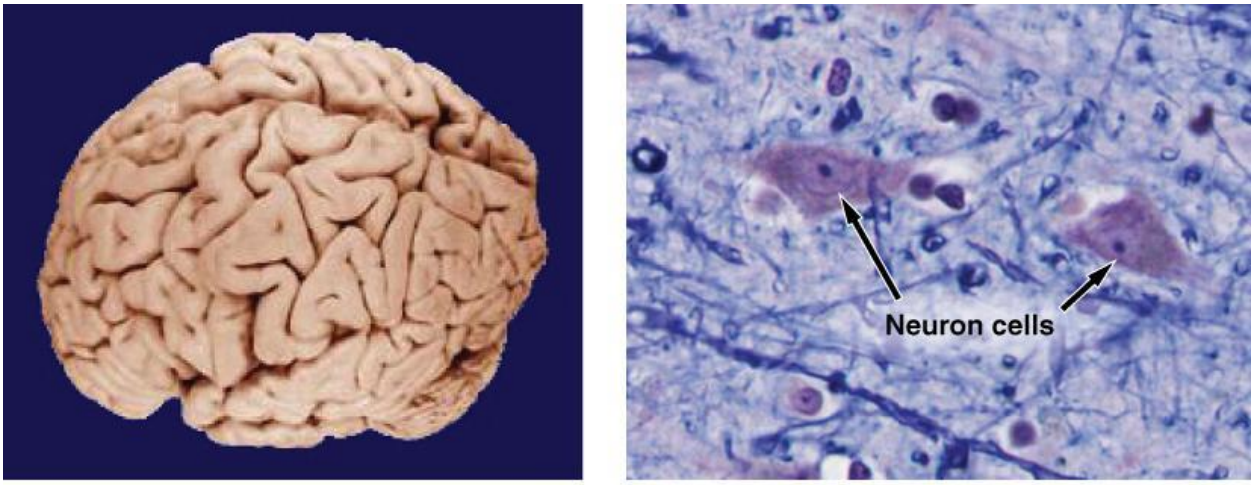


Figure 1.2 Gross and Microscopic Anatomy (a) Gross anatomy considers large structures such as the brain. (b) Microscopic anatomy can deal with the same structures, though at a different scale. This is a micrograph of nerve cells from the brain. LM × 1600. (credit a: "WriterHound"/Wikimedia Commons; credit b: Micrograph provided by the Regents of University of Michigan Medical School © 2012)

Anatomists take two general approaches to the study of the body's structures: regional and systemic. Regional anatomy is the study of the interrelationships of all of the structures in a specific body region, such as the abdomen. Studying regional anatomy helps us appreciate the interrelationships of body structures, such as how muscles, nerves, blood vessels, and other structures work together to serve a particular body region. In contrast, systemic anatomy is the study of the structures that make up a discrete body system—that is, a group of structures that work together to perform a unique body function. For example, a systemic anatomical study of the muscular system would consider all of the skeletal muscles of the body.

Whereas anatomy is about structure, physiology is about function. Human physiology is the scientific study of the chemistry and physics of the structures of the body and the ways in which they work together to support the functions of life. Much of the study of physiology centers on the body's tendency toward homeostasis. Homeostasis is the state of steady internal conditions maintained by living things. The study of physiology certainly includes observation, both with the naked eye and with microscopes, as well as manipulations and measurements. However, current advances in physiology usually depend on carefully designed laboratory experiments

that reveal the functions of the many structures and chemical compounds that make up the human body.

Like anatomists, physiologists typically specialize in a particular branch of physiology. For example, neurophysiology is the study of the brain, spinal cord, and nerves and how these work together to perform functions as complex and diverse as vision, movement, and thinking. Physiologists may work from the organ level (exploring, for example, what different parts of the brain do) to the molecular level (such as exploring how an electrochemical signal travels along nerves).

Form is closely related to function in all living things. For example, the thin flap of your eyelid can snap down to clear away dust particles and almost instantaneously slide back up to allow you to see again. At the microscopic level, the arrangement and function of the nerves and muscles that serve the eyelid allow for its quick action and retreat. At a smaller level of analysis, the function of these nerves and muscles likewise relies on the interactions of specific molecules and ions. Even the three-dimensional structure of certain molecules is essential to their function.

Your study of anatomy and physiology will make more sense if you continually relate the form of the structures you are studying to their function. In fact, it can be somewhat frustrating to attempt to study anatomy without an understanding of the physiology that a body structure supports. Imagine, for example, trying to appreciate the unique arrangement of the bones of the human hand if you had no conception of the function of the hand. Fortunately, your understanding of how the human hand manipulates tools—from pens to cell phones—helps you appreciate the unique alignment of the thumb in opposition to the four fingers, making your hand a structure that allows you to pinch and grasp objects and type text messages.

Chapter Attributions

Structural Organization of the Human Body

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

- Describe the structure of the human body in terms of six levels of organization
- List the eleven organ systems of the human body and identify at least one organ and one major function of each

Before you begin to study the different structures and functions of the human body, it is helpful to consider its basic architecture; that is, how its smallest parts are assembled into larger structures. It is convenient to consider the structures of the body in terms of fundamental levels of organization that increase in complexity: subatomic particles, atoms, molecules, organelles, cells, tissues, organs, organ systems, organisms and biosphere (Figure 1.3).

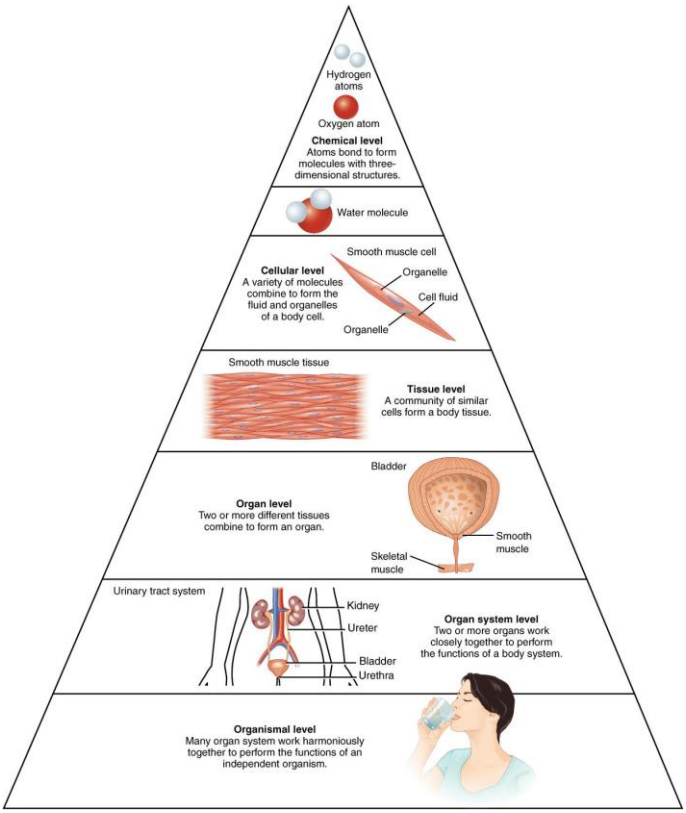


Figure 1.3 Levels of Structural Organization of the Human Body The organization of the body often is discussed in terms of six distinct levels of increasing complexity, from the smallest chemical building blocks to a unique human organism.

The Levels of Organization

To study the chemical level of organization, scientists consider the simplest building blocks of matter: subatomic particles, atoms and molecules. All matter in the universe is composed of one or more unique pure substances called elements, familiar examples of which are hydrogen, oxygen, carbon, nitrogen, calcium, and iron. The smallest unit of any of these pure substances (elements) is an atom. Atoms are made up of subatomic particles such as the proton, electron and neutron. Two or more atoms combine to form a molecule, such as the water molecules, proteins, and sugars found in living things. Molecules are the chemical building blocks of all body structures.

A **cell** is the smallest independently functioning unit of a living organism. Even bacteria, which are extremely small, independently-living organisms, have a cellular structure. Each bacterium is a single cell. All living structures of human anatomy contain cells, and almost all functions of human physiology are performed in cells or are initiated by cells.

A human cell typically consists of flexible membranes that enclose cytoplasm, a water-based cellular fluid together with a variety of tiny functioning units called **organelles**. In humans, as in all organisms, cells perform all functions of life. A **tissue** is a group of many similar cells (though sometimes composed of a few related types) that work together to perform a specific function. An **organ** is an anatomically distinct structure of the body composed of two or more tissue types. Each organ performs one or more specific physiological functions. An **organ system** is a group of organs that work together to perform major functions or meet physiological needs of the body.

This book covers eleven distinct organ systems in the human body (Figure 1.4 and Figure 1.5). Assigning organs to organ systems can be imprecise since organs that “belong” to one system can also have functions integral to another system. In fact, most organs contribute to more than one system.

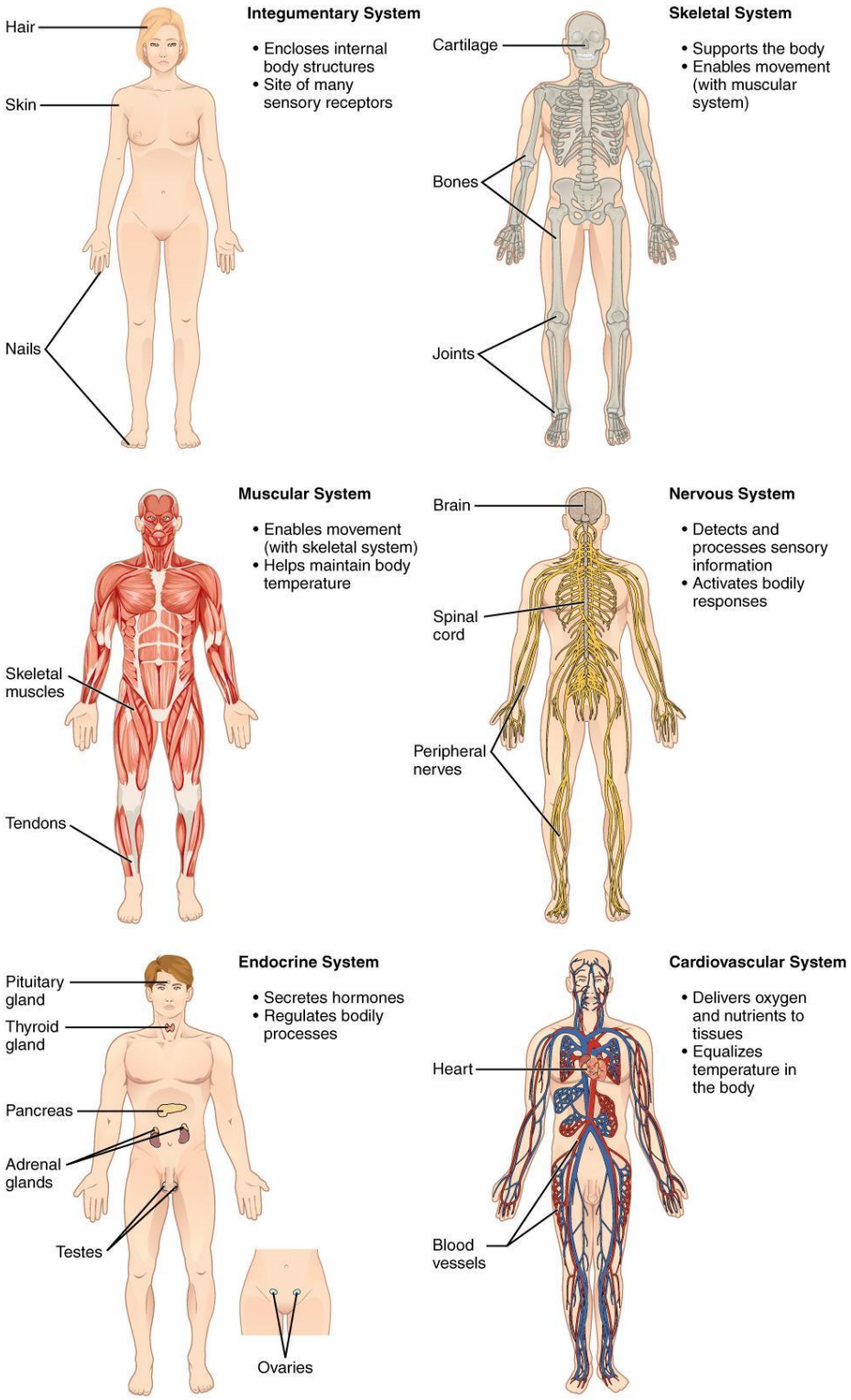


Figure 1.4 Organ Systems of the Human Body Organs that work together are grouped into organ systems.

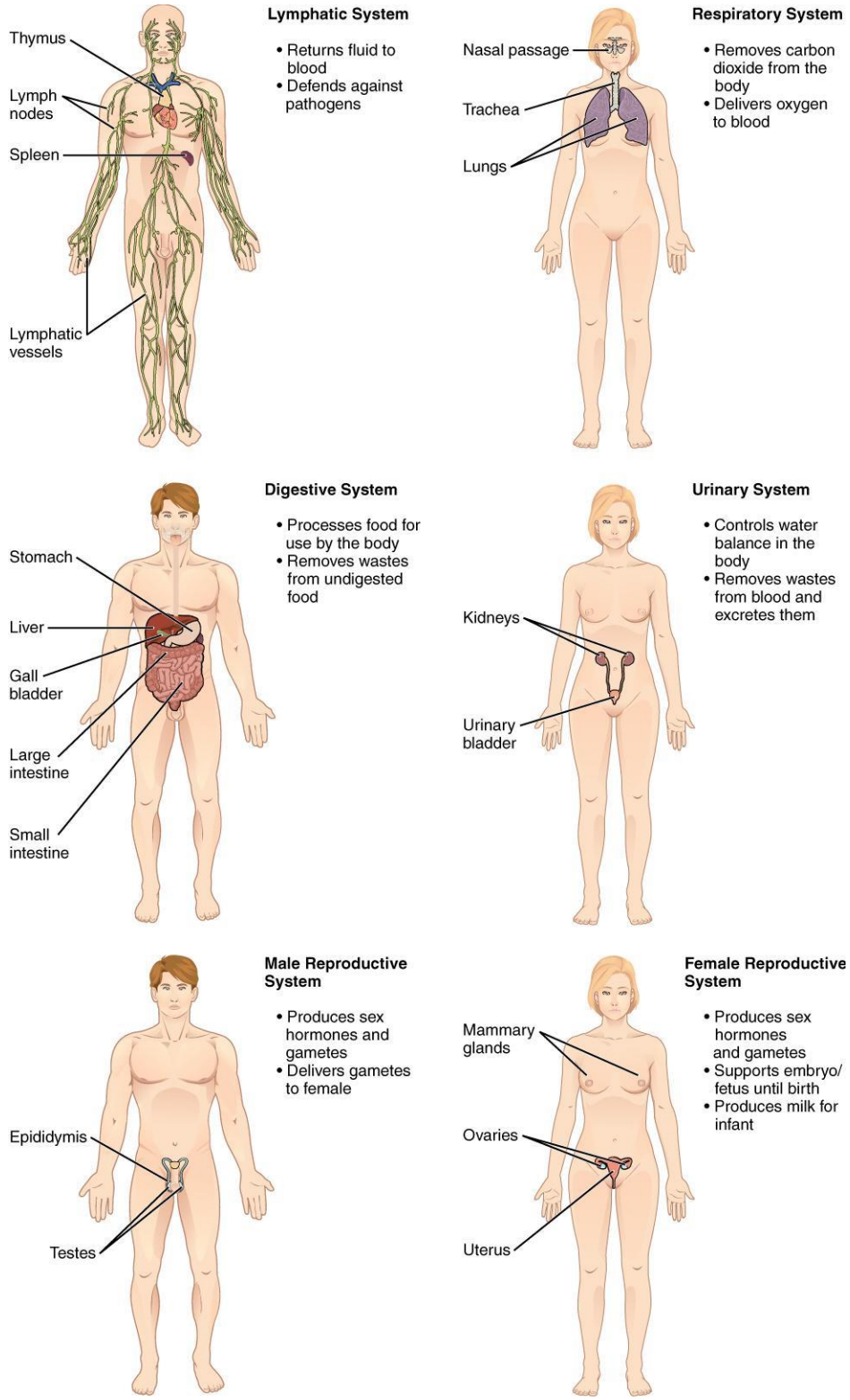


Figure 1.5 Organ Systems of the Human Body (continued) Organs that work together are grouped into organ systems.

The organism level is the highest level of organization. An **organism** is a living being that has a cellular structure and that can independently perform all physiologic functions necessary for life. In multicellular organisms, including humans, all cells, tissues, organs, and organ systems of the body work together to maintain the life and health of the organism.

Chapter Attributions

"Anatomy and Physiology" by J. Gordon Betts, Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, Peter DeSaix, OpenStax is licensed under CC BY 4.0. Access for free at https://openstax.org/books/anatomy-and-physiology/pages/1-2-structural-organization-of-the-human-body

2. Homeostasis

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

- Discuss the role of homeostasis in healthy functioning
- Contrast negative and positive feedback, giving one physiologic example of each mechanism

Maintaining homeostasis requires that the body continuously monitor its internal conditions. From body temperature to blood pressure to levels of certain nutrients, each physiological condition has a particular set point. A set point is the physiological value around which the normal range fluctuates. A normal range is the restricted set of values that is optimally healthful and stable. For example, the set point for normal human body temperature is approximately 37°C (98.6°F) Physiological parameters, such as body temperature and blood pressure, tend to fluctuate within a normal range a few degrees above and below that point. Control centers in the brain and other parts of the body monitor and react to deviations from homeostasis using negative feedback. Negative feedback is a mechanism that reverses a deviation from the set point. Therefore, negative feedback maintains body parameters within their normal range. The maintenance of homeostasis by negative feedback goes on throughout the body at all times, and an understanding of negative feedback is thus fundamental to an understanding of human physiology.

Negative Feedback

A negative feedback system has three basic components (Figure 1.10**a**). A **sensor**, also referred to a receptor, is a component of a feedback system that monitors a physiological value. This value is reported to the control center. The **control center** is the component in a feedback system that compares the value to the normal range. If the value deviates too much from the set point, then the control center activates an effector. An **effector** is the component in a feedback system that causes a change to reverse the situation and return the value to the normal range.

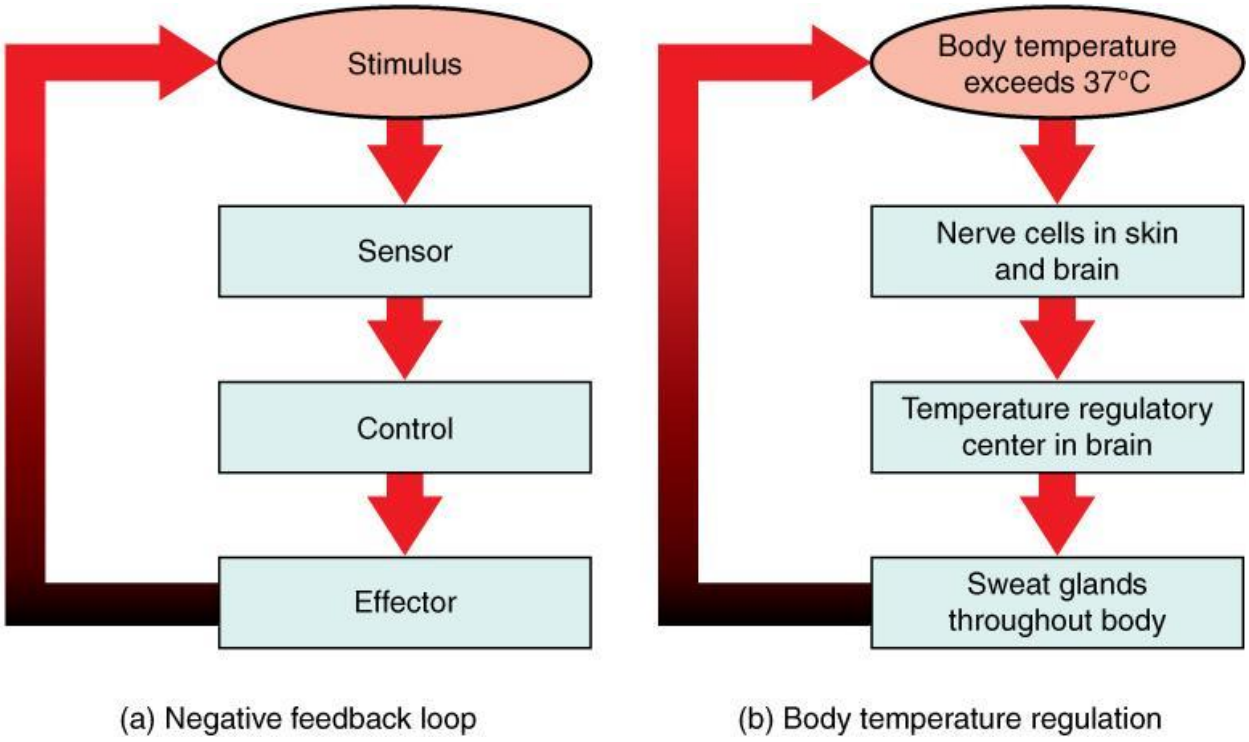


Figure 1.10 Negative Feedback Loop In a negative feedback loop, a stimulus—a deviation from a set point—is resisted through a physiological process that returns the body to homeostasis. (a) A negative feedback loop has four basic parts. (b) Body temperature is regulated by negative feedback.

In order to set the system in motion, a stimulus must drive a physiological parameter beyond its normal range (that is, beyond homeostasis). This stimulus is "heard" by a specific sensor. For example, in the control of blood glucose, specific endocrine cells in the pancreas detect excess glucose (the stimulus) in the bloodstream. These pancreatic beta cells respond to the increased level of blood glucose by releasing the hormone insulin into the bloodstream. The insulin signals skeletal muscle fibers, fat cells (adipocytes), and liver cells to take up the excess glucose, removing it from the bloodstream. As glucose concentration in the bloodstream drops, the decrease in concentration—the actual negative feedback—is detected by pancreatic

alpha cells, and insulin release stops. This prevents blood sugar levels from continuing to drop below the normal range.

Humans have a similar temperature regulation feedback system that works by promoting either heat loss or heat gain (Figure 1.10**b**). When the brain's temperature regulation center receives data from the sensors indicating that the body's temperature exceeds its normal range, it stimulates a cluster of brain cells referred to as the "heat-loss center." This stimulation has three major effects:

- Blood vessels in the skin begin to dilate allowing more blood from the body core to flow to the surface of the skin allowing the heat to radiate into the environment.
- As blood flow to the skin increases, sweat glands are activated to increase their output. As the sweat evaporates from the skin surface into the surrounding air, it takes heat with it.
- The depth of respiration increases, and a person may breathe through an open mouth instead of through the nasal passageways. This further increases heat loss from the lungs.

In contrast, activation of the brain's heat-gain center by exposure to cold reduces blood flow to the skin, and blood returning from the limbs is diverted into a network of deep veins. This arrangement traps heat closer to the body core and restricts heat loss. If heat loss is severe, the brain triggers an increase in random signals to skeletal muscles, causing them to contract and producing shivering. The muscle contractions of shivering release heat while using up ATP. The brain triggers the thyroid gland in the endocrine system to release thyroid hormone, which increases metabolic activity and heat production in cells throughout the body. The brain also signals the adrenal glands to release epinephrine (adrenaline), a hormone that causes the breakdown of glycogen into glucose, which can be used as an energy source. The breakdown of glycogen into glucose also results in increased metabolism and heat production.

INTERACTIVE LINK

Water concentration in the body is critical for proper functioning. A person's body retains very tight control on water levels without conscious control by the person. Watch this video (https://www.youtube.com/watch?v=vB7tSHqR1eY) to learn more about water concentration in the body. Which organ has primary control over the amount of water in the body?

Positive Feedback

Positive feedback intensifies a change in the body's physiological condition rather than reversing it. A deviation from the normal range results in more change, and the system moves farther away from the normal range. Positive feedback in the body is normal only when there is

a definite end point. Childbirth and the body's response to blood loss are two examples of positive feedback loops that are normal but are activated only when needed.

Childbirth at full term is an example of a situation in which the maintenance of the existing body state is not desired. Enormous changes in the mother's body are required to expel the baby at the end of pregnancy. And the events of childbirth, once begun, must progress rapidly to a conclusion or the life of the mother and the baby are at risk. The extreme muscular work of labor and delivery are the result of a positive feedback system (Figure 1.11).

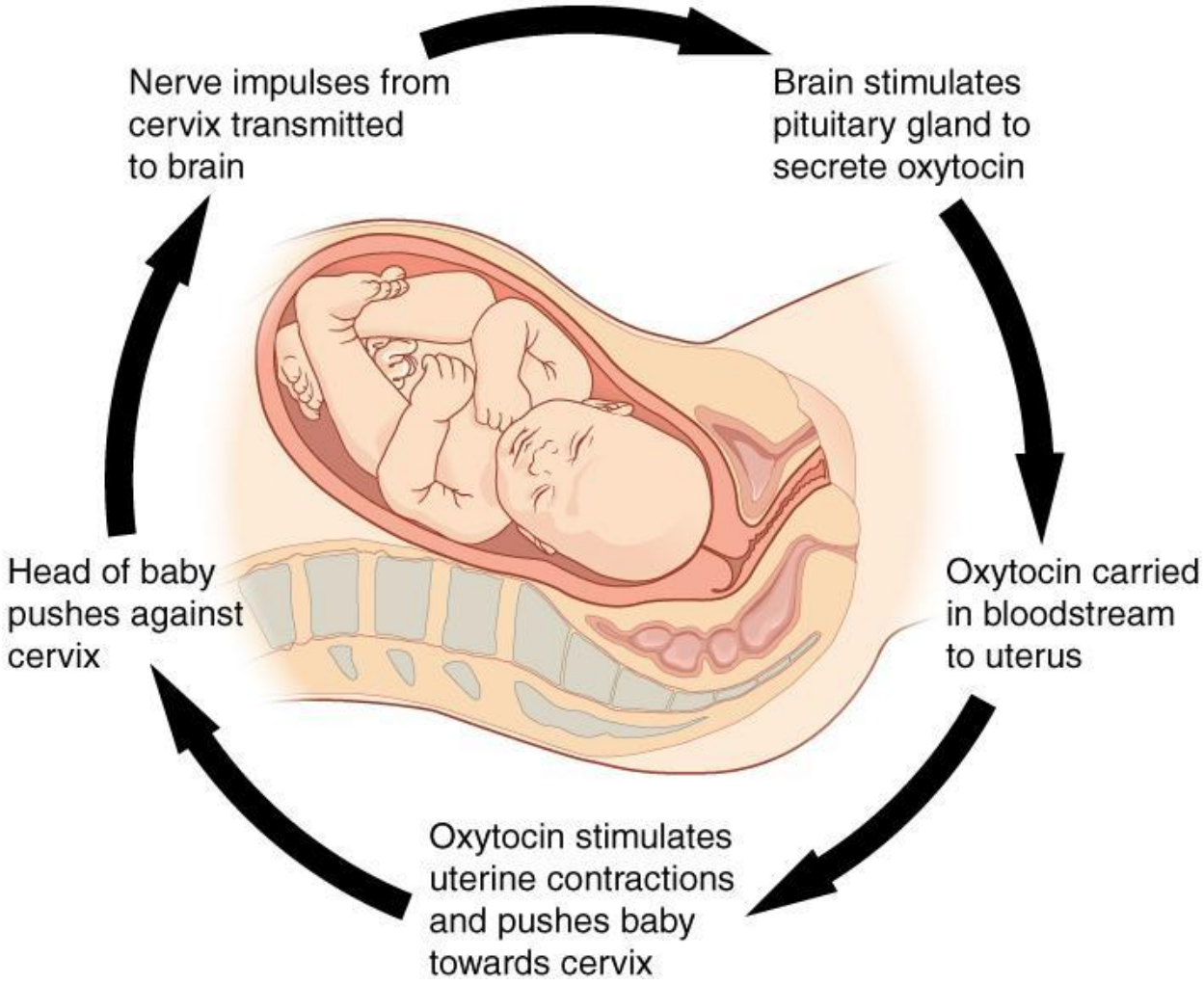


Figure 1.11 Positive Feedback Loop.

Normal childbirth is driven by a positive feedback loop. A positive feedback loop results in a change in the body's status, rather than a return to homeostasis.

The first contractions of labor (the stimulus) push the baby toward the cervix (the lowest part of the uterus). The cervix contains stretch-sensitive nerve cells that monitor the degree of stretching (the sensors). These nerve cells send messages to the brain, which in turn causes the pituitary gland at the base of the brain to release the hormone oxytocin into the bloodstream. Oxytocin causes stronger contractions of the smooth muscles in of the uterus (the effectors), pushing the baby further down the birth canal. This causes even greater stretching of the cervix. The cycle of stretching, oxytocin release, and increasingly more forceful contractions stops only when the baby is born. At this point, the stretching of the cervix halts, stopping the release of oxytocin.

A second example of positive feedback centers on reversing extreme damage to the body. Following a penetrating wound, the most immediate threat is excessive blood loss. Less blood circulating means reduced blood pressure and reduced perfusion (penetration of blood) to the brain and other vital organs. If perfusion is severely reduced, vital organs will shut down and the person will die. The body responds to this potential catastrophe by releasing substances in the injured blood vessel wall that begin the process of blood clotting. As each step of clotting occurs, it stimulates the release of more clotting substances. This accelerates the processes of clotting and sealing off the damaged area. Clotting is contained in a local area based on the tightly controlled availability of clotting proteins. This is an adaptive, life-saving cascade of events.

Chapter Attributions

3. Matter, Elements, Atoms

Elements and Atoms: The Building Blocks of Matter

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

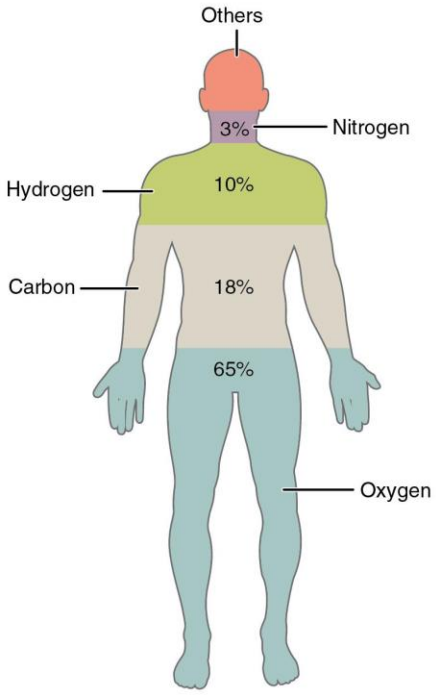
- Discuss the relationships between matter, mass, elements, compounds, atoms, and subatomic particles
- Distinguish between atomic number and mass number
- Identify the key distinction between isotopes of the same element
- Explain how electrons occupy electron shells and their contribution to an atom's relative stability

The substance of the universe—from a grain of sand to a star—is called matter. Scientists define matter as anything that occupies space and has mass. An object's mass and its weight are related concepts, but not quite the same. An object's mass is the amount of matter contained in the object, and the object's mass is the same whether that object is on Earth or in the zero-gravity environment of outer space. An object's weight, on the other hand, is its mass as affected by the pull of gravity. Where gravity strongly pulls on an object's mass its weight is greater than it is where gravity is less strong. An object of a certain mass weighs less on the moon, for example, than it does on Earth because the gravity of the moon is less than that of Earth. In other words, weight is variable, and is influenced by gravity. A piece of cheese that weighs a pound on Earth weighs only a few ounces on the moon.

Elements and Compounds

All matter in the natural world is composed of one or more of the 92 fundamental substances called elements. An element is a pure substance that is distinguished from all other matter by the fact that it cannot be created or broken down by ordinary chemical means. While your body can assemble many of the chemical compounds needed for life from their constituent elements, it cannot make elements. They must come from the environment. A familiar example of an element that you must take in is calcium (Ca++). Calcium is essential to the human body; it is absorbed and used for a number of processes, including strengthening bones. When you consume dairy products your digestive system breaks down the food into components small enough to cross into the bloodstream. Among these is calcium, which, because it is an element, cannot be broken down further. The elemental calcium in cheese, therefore, is the same as the calcium that forms your bones. Some other elements you might be familiar with are oxygen,

sodium, and iron. The elements in the human body are shown in Figure 2.2, beginning with the most abundant: oxygen (O), carbon (C), hydrogen (H), and nitrogen (N). Each element's name can be replaced by a one- or two-letter symbol; you will become familiar with some of these during this course. All the elements in your body are derived from the foods you eat and the air you breathe.



Element	Symbol	Percentage in Body
Oxygen	O	65.0
Carbon	C	18.5
Hydrogen	H	9.5
Nitrogen	N	3.2
Calcium	Ca	1.5
Phosphorus	P	1.0
Potassium	K	0.4
Sulfur	S	0.3
Sodium	Na	0.2
Chlorine	Cl	0.2
Magnesium	Mg	0.1
Trace elements include boron (B), chromium (Cr), cobalt (Co), copper (Cu), fluorine (F), iodine (I), iron (Fe), manganese (Mn), molybdenum (Mo), selenium (Se), silicon (Si), tin (Sn), vanadium (V), and zinc (Zn).		less than 1.0

Figure 2.2 Elements of the Human Body The main elements that compose the human body are shown from most abundant to least abundant.

In nature, elements rarely occur alone. Instead, they combine to form compounds. A compound is a substance composed of two or more elements joined by chemical bonds. For example, the compound glucose is an important body fuel. It is always composed of the same three elements: carbon, hydrogen, and oxygen. Moreover, the elements that make up any given compound always occur in the same relative amounts. In glucose, there are always six carbon and six oxygen units for every twelve hydrogen units. But what, exactly, are these "units" of elements?

Atoms and Subatomic Particles

An atom is the smallest quantity of an element that retains the unique properties of that element. In other words, an atom of hydrogen is a unit of hydrogen—the smallest amount of hydrogen that can exist. As you might guess, atoms are almost unfathomably small. The period at the end of this sentence is millions of atoms wide.

Atomic Structure and Energy

Atoms are made up of even smaller subatomic particles, three types of which are important: the proton, neutron, and electron. The number of positively-charged protons and non-charged ("neutral") neutrons, gives mass to the atom, and the number of each in the nucleus of the atom determine the element. The number of negatively-charged electrons that "spin" around the nucleus at close to the speed of light equals the number of protons. An electron has about 1/2000th the mass of a proton or neutron.

Figure 2.3 shows two models that can help you imagine the structure of an atom—in this case, helium (He). In the planetary model, helium's two electrons are shown circling the nucleus in a fixed orbit depicted as a ring. Although this model is helpful in visualizing atomic structure, in reality, electrons do not travel in fixed orbits, but whiz around the nucleus erratically in a so-called electron cloud.

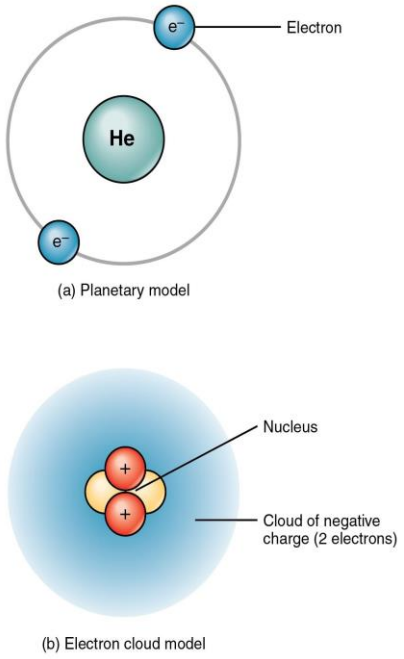


Figure 2.3 Two Models of Atomic Structure (a) In the planetary model, the electrons of helium are shown in fixed orbits, depicted as rings, at a precise distance from the nucleus, somewhat

like planets orbiting the sun. (b) In the electron cloud model, the electrons of carbon are shown in the variety of locations they would have at different distances from the nucleus over time.

An atom's protons and electrons carry electrical charges. Protons, with their positive charge, are designated p+. Electrons, which have a negative charge, are designated e–. An atom's neutrons have no charge: they are electrically neutral. Just as a magnet sticks to a steel refrigerator because their opposite charges attract, the positively charged protons attract the negatively charged electrons. This mutual attraction gives the atom some structural stability. The attraction by the positively charged nucleus helps keep electrons from straying far. The number of protons and electrons within a neutral atom are equal, thus, the atom's overall charge is balanced.

Atomic Number and Mass Number

An atom of carbon is unique to carbon, but a proton of carbon is not. One proton is the same as another, whether it is found in an atom of carbon, sodium (Na), or iron (Fe). The same is true for neutrons and electrons. So, what gives an element its distinctive properties—what makes carbon so different from sodium or iron? The answer is the unique quantity of protons each contains. Carbon by definition is an element whose atoms contain six protons. No other element has exactly six protons in its atoms. Moreover, all atoms of carbon, whether found in your liver or in a lump of coal, contain six protons. Thus, the atomic number, which is the number of protons in the nucleus of the atom, identifies the element. Because an atom usually has the same number of electrons as protons, the atomic number identifies the usual number of electrons as well.

In their most common form, many elements also contain the same number of neutrons as protons. The most common form of carbon, for example, has six neutrons as well as six protons, for a total of 12 subatomic particles in its nucleus. An element's mass number is the sum of the number of protons and neutrons in its nucleus. So the most common form of carbon's mass number is 12. (Electrons have so little mass that they do not appreciably contribute to the mass of an atom.) Carbon is a relatively light element. Uranium (U), in contrast, has a mass number of 238 and is referred to as a heavy metal. Its atomic number is 92

(it has 92 protons) but it contains 146 neutrons; it has the most mass of all the naturally occurring elements.

The periodic table of the elements, shown in Figure 2.4, is a chart identifying the 92 elements found in nature, as well as several larger, unstable elements discovered experimentally. The elements are arranged in order of their atomic number, with hydrogen and helium at the top of the table, and the more massive elements below. The periodic table is a useful device because for each element, it identifies the chemical symbol, the atomic number, and the mass number, while organizing elements according to their propensity to react with other elements. The number of protons and electrons in an element are equal. The number of protons and neutrons may be equal for some elements, but are not equal for all.

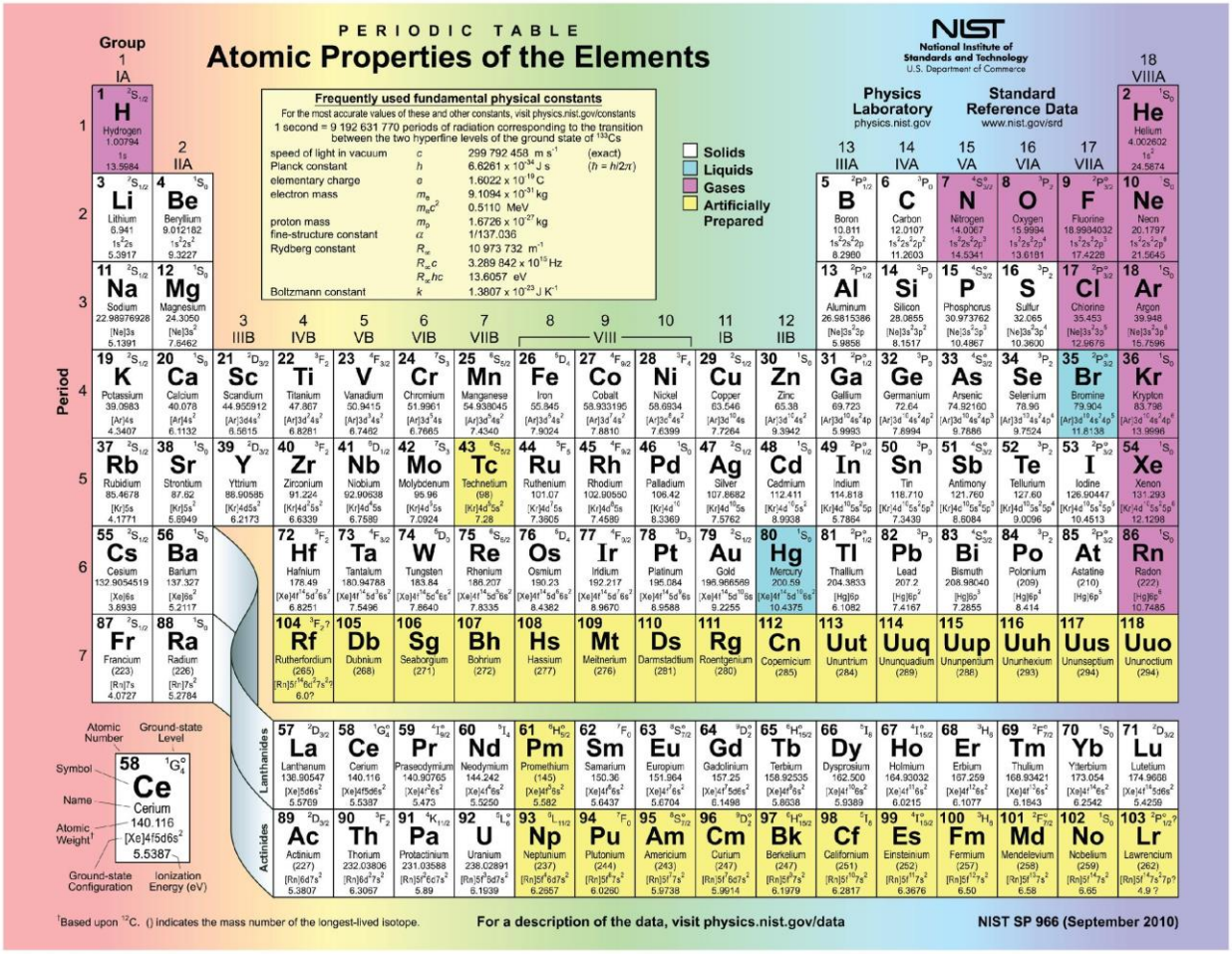


Figure 2.4 The Periodic Table of the Elements (credit: R.A. Dragoset, A. Musgrove, C.W. Clark, W.C. Martin)

INTERACTIVE LINK

Visit this website (https://ptable.com/) to view the periodic table. In the periodic table of the elements, elements in a single column have the same number of electrons that can participate in a chemical reaction. These electrons are known as "valence electrons." For example, the elements in the first column all have a single valence electron, an electron that can be "donated" in a chemical reaction with another atom. What is the meaning of a mass number shown in parentheses?

Isotopes

Although each element has a unique number of protons, it can exist as different isotopes. An **isotope** is one of the different forms of an element, distinguished from one another by different numbers of neutrons. The standard isotope of carbon is $_{12}C$, commonly called carbon twelve. $_{12}C$ has six protons and six neutrons, for a mass number of twelve. All of the isotopes of carbon have the same number of protons; therefore, $_{13}C$ has seven neutrons, and $_{14}C$ has eight neutrons. The different isotopes of an element can also be indicated with the mass number hyphenated (for example, C-12 instead of $_{12}C$). Hydrogen has three common isotopes, shown in Figure 2.5.

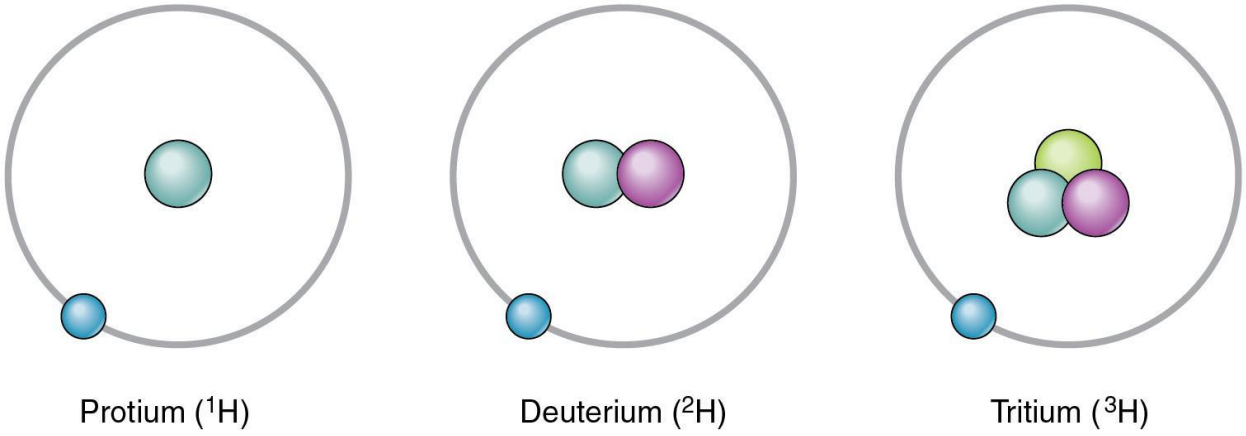


Figure 2.5 Isotopes of Hydrogen Protium, designated $_{1}H$, has one proton and no neutrons. It is by far the most abundant isotope of hydrogen in nature. Deuterium, designated $_{2}H$, has one proton and one neutron. Tritium, designated $_{3}H$, has two neutrons.

An isotope that contains more than the usual number of neutrons is referred to as a heavy isotope. An example is $_{14}C$. Heavy isotopes tend to be unstable, and unstable isotopes are radioactive. A **radioactive isotope** is an isotope whose nucleus readily decays, giving off subatomic particles and electromagnetic energy. Different radioactive isotopes (also called radioisotopes) differ in their half-life, the time it takes for half of any size sample of an isotope to decay. For example, the half-life of tritium—a radioisotope of hydrogen—is about 12 years, indicating it takes 12 years for half of the tritium nuclei in a sample to decay. Excessive exposure to radioactive isotopes can damage human cells and even cause cancer and birth defects,

but when exposure is controlled, some radioactive isotopes can be useful in medicine. For more information, see the Career Connections.

CAREER CONNECTION

Interventional Radiologist

The controlled use of radioisotopes has advanced medical diagnosis and treatment of disease. Interventional radiologists are physicians who treat disease by using minimally invasive techniques involving radiation. Many conditions that could once only be treated with a lengthy and traumatic operation can now be treated non-surgically, reducing the cost, pain, length of hospital stay, and recovery time for patients. For example, in the past, the only options for a patient with one or more tumors in the liver were surgery and chemotherapy (the administration of drugs to treat cancer). Some liver tumors, however, are difficult to access surgically, and others could require the surgeon to remove too much of the liver. Moreover, chemotherapy is highly toxic to the liver, and certain tumors do not respond well to it anyway. In some such cases, an interventional radiologist can treat the tumors by disrupting their blood supply, which they need if they are to continue to grow. In this procedure, called radioembolization, the radiologist accesses the liver with a fine needle, threaded through one of the patient's blood vessels. The radiologist then inserts tiny radioactive "seeds" into the blood vessels that supply the tumors. In the days and weeks following the procedure, the radiation emitted from the seeds destroys the vessels and directly kills the tumor cells in the vicinity of the treatment.

Radioisotopes emit subatomic particles that can be detected and tracked by imaging technologies. One of the most advanced uses of radioisotopes in medicine is the positron emission tomography (PET) scanner, which detects the activity in the body of a very small injection of radioactive glucose, the simple sugar that cells use for energy. The PET camera reveals to the medical team which of the patient's tissues are taking up the most glucose. Thus, the most metabolically active tissues show up as bright "hot spots" on the images (Figure 2.6). PET can reveal some cancerous masses because cancer cells consume glucose at a high rate to fuel their rapid reproduction.

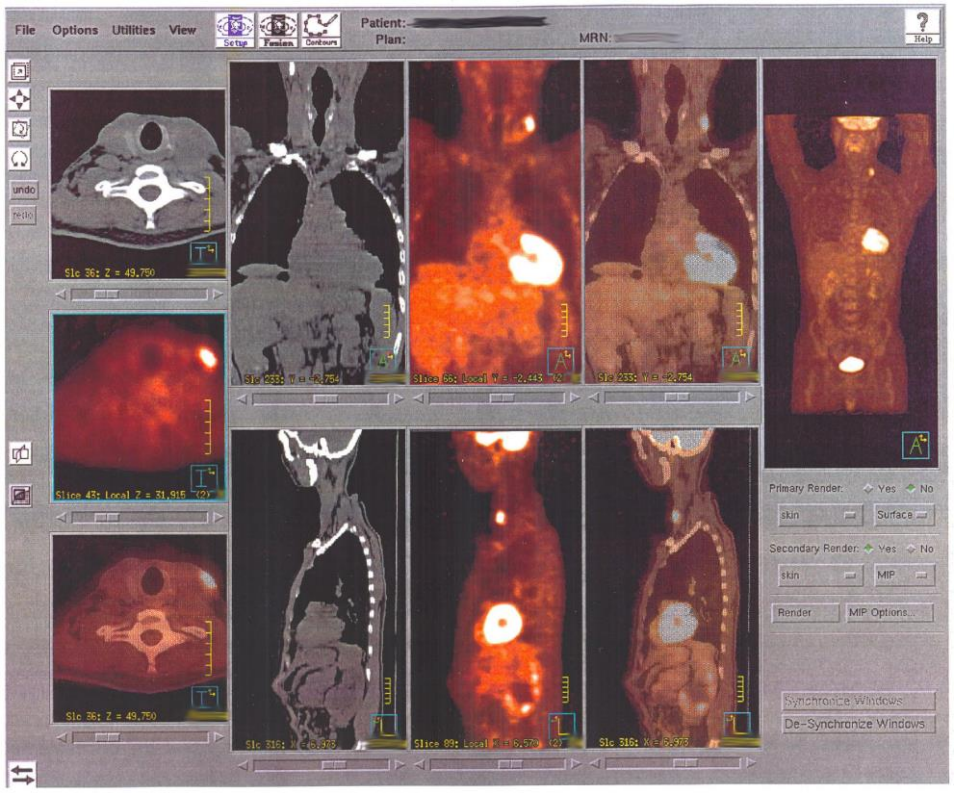


Figure 2.6 PET Scan PET highlights areas in the body where there is relatively high glucose use, which is characteristic of cancerous tissue. This PET scan shows sites of the spread of a large primary tumor to other sites.

The Behavior of Electrons

In the human body, atoms do not exist as independent entities. Rather, they are constantly reacting with other atoms to form and to break down more complex substances. To fully understand anatomy and physiology you must grasp how atoms participate in such reactions. The key is understanding the behavior of electrons.

Although electrons do not follow rigid orbits a set distance away from the atom's nucleus, they do tend to stay within certain regions of space called electron shells. An electron shell is a layer of electrons that encircle the nucleus at a distinct energy level.

The atoms of the elements found in the human body have from one to five electron shells, and all electron shells hold eight electrons except the first shell, which can only hold two. This configuration of electron shells is the same for all atoms. The precise number of shells depends on the number of electrons in the atom. Hydrogen and helium have just one and two electrons,

respectively. If you take a look at the periodic table of the elements, you will notice that hydrogen and helium are placed alone on either sides of the top row; they are the only elements that have just one electron shell (Figure 2.7). A second shell is necessary to hold the electrons in all elements larger than hydrogen and helium.

Lithium (Li), whose atomic number is 3, has three electrons. Two of these fill the first electron shell, and the third spills over into a second shell. The second electron shell can accommodate as many as eight electrons. Carbon, with its six electrons, entirely fills its first shell, and half-fills its second. With ten electrons, neon (Ne) entirely fills its two electron shells. Again, a look at the periodic table reveals that all of the elements in the second row, from lithium to neon, have just two electron shells. Atoms with more than ten electrons require more than two shells. These elements occupy the third and subsequent rows of the periodic table.

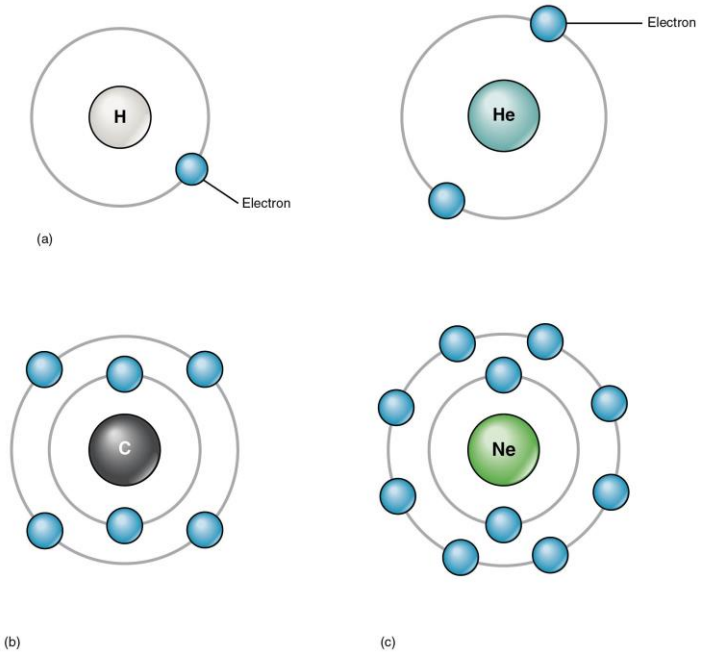


Figure 2.7 Electron Shells Electrons orbit the atomic nucleus at distinct levels of energy called electron shells. (a) With one electron, hydrogen only half-fills its electron shell. Helium also has a single shell, but its two electrons completely fill it. (b) The electrons of carbon completely fill its first electron shell, but only half-fills its second. (c) Neon, an element that does not occur in the body, has 10 electrons, filling both of its electron shells.

The factor that most strongly governs the tendency of an atom to participate in chemical reactions is the number of electrons in its valence shell. A valence shell is an atom's outermost

electron shell. If the valence shell is full, the atom is stable; meaning its electrons are unlikely to be pulled away from the nucleus by the electrical charge of other atoms. If the valence shell is not full, the atom is reactive; meaning it will tend to react with other atoms in ways that make the valence shell full. Consider hydrogen, with its one electron only half-filling its valence shell. This single electron is likely to be drawn into relationships with the atoms of other elements, so that hydrogen's single valence shell can be stabilized.

All atoms (except hydrogen and helium with their single electron shells) are most stable when there are exactly eight electrons in their valence shell. This principle is referred to as the octet rule, and it states that an atom will give up, gain, or share electrons with another atom so that it ends up with eight electrons in its own valence shell. For example, oxygen, with six electrons in its valence shell, is likely to react with other atoms in a way that results in the addition of two electrons to oxygen's valence shell, bringing the number to eight. When two hydrogen atoms each share their single electron with oxygen, covalent bonds are formed, resulting in a molecule of water, H_2O.

In nature, atoms of one element tend to join with atoms of other elements in characteristic ways. For example, carbon commonly fills its valence shell by linking up with four atoms of hydrogen. In so doing, the two elements form the simplest of organic molecules, methane, which also is one of the most abundant and stable carbon-containing compounds on Earth. As stated above, another example is water; oxygen needs two electrons to fill its valence shell. It commonly interacts with two atoms of hydrogen, forming H_2O. Incidentally, the name "hydrogen" reflects its contribution to water (hydro- = "water"; -gen = "maker"). Thus, hydrogen is the "water maker."

Chapter Attributions

Chemical Bonds

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

- Explain the relationship between molecules and compounds
- Distinguish between ions, cations, and anions
- Identify the key difference between ionic and covalent bonds
- Distinguish between nonpolar and polar covalent bonds
- Explain how water molecules link via hydrogen bonds

Atoms separated by a great distance cannot link; rather, they must come close enough for the electrons in their valence shells to interact. But do atoms ever actually touch one another? Most physicists would say no, because the negatively charged electrons in their valence shells repel one another. No force within the human body—or anywhere in the natural world—is strong enough to overcome this electrical repulsion. So when you read about atoms linking together or colliding, bear in mind that the atoms are not merging in a physical sense.

Instead, atoms link by forming a chemical bond. A bond is a weak or strong electrical attraction that holds atoms in the same vicinity. The new grouping is typically more stable—less likely to react again—than its component atoms were when they were separate. A more or less stable grouping of two or more atoms held together by chemical bonds is called a molecule. The bonded atoms may be of the same element, as in the case of H_2, which is called molecular hydrogen or hydrogen gas. When a molecule is made up of two or more atoms of different elements, it is called a chemical compound. Thus, a unit of water, or H_2O, is a compound, as is a single molecule of the gas methane, or CH_4.

Three types of chemical bonds are important in human physiology, because they hold together substances that are used by the body for critical aspects of homeostasis, signaling, and energy production, to name just a few important processes. These are ionic bonds, covalent bonds, and hydrogen bonds.

Ions and Ionic Bonds

Recall that an atom typically has the same number of positively charged protons and negatively charged electrons. As long as this situation remains, the atom is electrically neutral. But when an atom participates in a chemical reaction that results in the donation or acceptance of one or more electrons, the atom will then become positively or negatively charged. This happens frequently for most atoms in order to have a full valence shell, as described previously. This can happen either by gaining electrons to fill a shell that is more than half-full, or by giving away electrons to empty a shell that is less than half-full, thereby leaving the next smaller electron shell as the new, full, valence shell. An atom that has an electrical charge—whether positive or negative—is an ion.

INTERACTIVE LINK

Visit this website (https://education.jlab.org/frost/electroscope.html) to learn about electrical energy and the attraction/repulsion of charges. What happens to the charged electroscope when a conductor is moved between its plastic sheets, and why?

Potassium (K), for instance, is an important element in all body cells. Its atomic number is 19. It has just one electron in its valence shell. This characteristic makes potassium highly likely to participate in chemical reactions in which it donates one electron. (It is easier for potassium to donate one electron than to gain seven electrons.) The loss will cause the positive charge of potassium's protons to be more influential than the negative charge of potassium's electrons. In other words, the resulting potassium ion will be slightly positive. A potassium ion is written K^+, indicating that it has lost a single electron. A positively charged ion is known as a cation.

Now consider fluorine (F), a component of bones and teeth. Its atomic number is nine, and it has seven electrons in its valence shell. Thus, it is highly likely to bond with other atoms in such a way that fluorine accepts one electron (it is easier for fluorine to gain one electron than to donate seven electrons). When it does, its electrons will outnumber its protons by one, and it will have an overall negative charge. The ionized form of fluorine is called fluoride, and is written as F^-. A negatively charged ion is known as an anion.

Atoms that have more than one electron to donate or accept will end up with stronger positive or negative charges. A cation that has donated two electrons has a net charge of +2. Using

magnesium (Mg) as an example, this can be written Mg++ or Mg2+. An anion that has accepted two electrons has a net charge of –2. The ionic form of selenium (Se), for example, is typically written Se2–.

The opposite charges of cations and anions exert a moderately strong mutual attraction that keeps the atoms in close proximity forming an ionic bond. An ionic bond is an ongoing, close association between ions of opposite charge. The table salt you sprinkle on your food owes its existence to ionic bonding. As shown in Figure 2.8, sodium commonly donates an electron to chlorine, becoming the cation Na+. When chlorine accepts the electron, it becomes the chloride anion, Cl–. With their opposing charges, these two ions strongly attract each other.

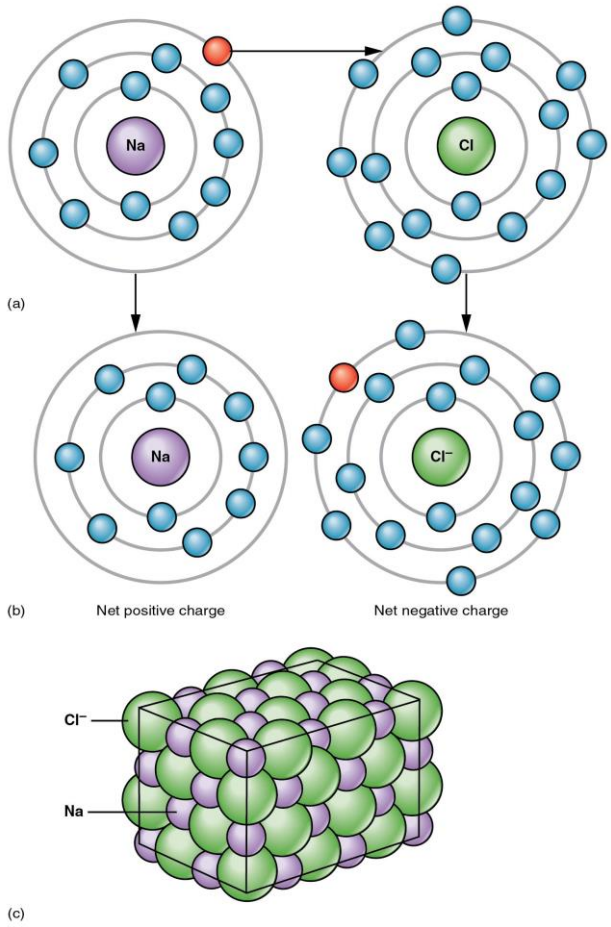


Figure 2.8 Ionic Bonding (a) Sodium readily donates the solitary electron in its valence shell to chlorine, which needs only one electron to have a full valence shell. (b) The opposite electrical charges of the resulting sodium cation and chloride anion result in the formation of a bond of

attraction called an ionic bond. (c) The attraction of many sodium and chloride ions results in the formation of large groupings called crystals.

Water is an essential component of life because it is able to break the ionic bonds in salts to free the ions. In fact, in biological fluids, most individual atoms exist as ions. These dissolved ions produce electrical charges within the body. The behavior of these ions produces the tracings of heart and brain function observed as waves on an electrocardiogram (EKG or ECG) or an electroencephalogram (EEG). The electrical activity that derives from the interactions of the charged ions is why they are also called electrolytes.

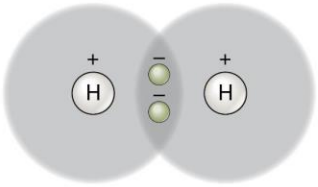
Covalent Bonds

Unlike ionic bonds formed by the attraction between a cation's positive charge and an anion's negative charge, molecules formed by a covalent bond share electrons in a mutually stabilizing relationship. Like next-door neighbors whose kids hang out first at one home and then at the other, the atoms do not lose or gain electrons permanently. Instead, the electrons move back and forth between the elements. Because of the close sharing of pairs of electrons (one electron from each of two atoms), covalent bonds are stronger than ionic bonds.

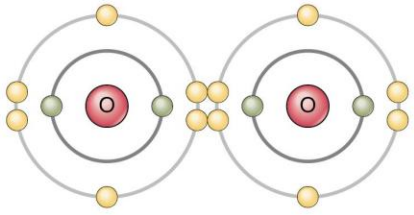
Nonpolar Covalent Bonds

Figure 2.9 shows several common types of covalent bonds. Notice that the two covalently bonded atoms typically share just one or two electron pairs, though larger sharings are possible. The important concept to take from this is that in covalent bonds, electrons in the outermost valence shell are shared to fill the valence shells of both atoms, ultimately stabilizing both of the atoms involved. In a single covalent bond, a single pair of electrons is shared between two atoms, while in a double covalent bond, two pairs of electrons are shared between two atoms. There even are triple covalent bonds, where three pairs of atoms are shared.

(a) A single covalent bond: hydrogen gas (H—H). Two atoms of hydrogen each share their solitary electron in a single covalent bond.



(b) A double covalent bond: oxygen gas (O═O). An atom of oxygen has six electrons in its valence shell; thus, two more would make it stable. Two atoms of oxygen achieve stability by sharing two pairs of electrons in a double covalent bond.



Molecule of oxygen gas (O_2)

(c) Two double covalent bonds: carbon dioxide (O═C═O). An atom of carbon has four electrons in its valence shell; thus, four more would make it stable. An atom of carbon and two atoms of oxygen achieve stability by sharing two electron pairs each, in two double covalent bonds.

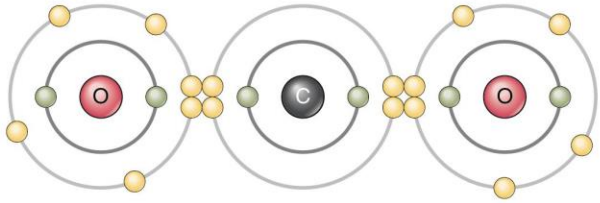


Figure 2.9 Covalent Bonding

You can see that the covalent bonds shown in Figure 2.9 are balanced. The sharing of the negative electrons is relatively equal, as is the electrical pull of the positive protons in the nucleus of the atoms involved. This is why covalently bonded molecules that are electrically balanced in this way are described as nonpolar; that is, no region of the molecule is either more positive or more negative than any other.

Polar Covalent Bonds

Groups of legislators with completely opposite views on a particular issue are often described as “polarized” by news writers. In chemistry, a polar molecule is a molecule that contains regions that have opposite electrical charges. Polar molecules occur when atoms share electrons unequally, in polar covalent bonds.

The most familiar example of a polar molecule is water (Figure 2.10). The molecule has three parts: one atom of oxygen, the nucleus of which contains eight protons, and two hydrogen

atoms, whose nuclei each contain only one proton. Because every proton exerts an identical positive charge, a nucleus that contains eight protons exerts a charge eight times greater than a nucleus that contains one proton. This means that the negatively charged electrons present in the water molecule are more strongly attracted to the oxygen nucleus than to the hydrogen nuclei. Each hydrogen atom's single negative electron therefore migrates toward the oxygen atom, making the oxygen end of their bond slightly more negative than the hydrogen end of their bond.

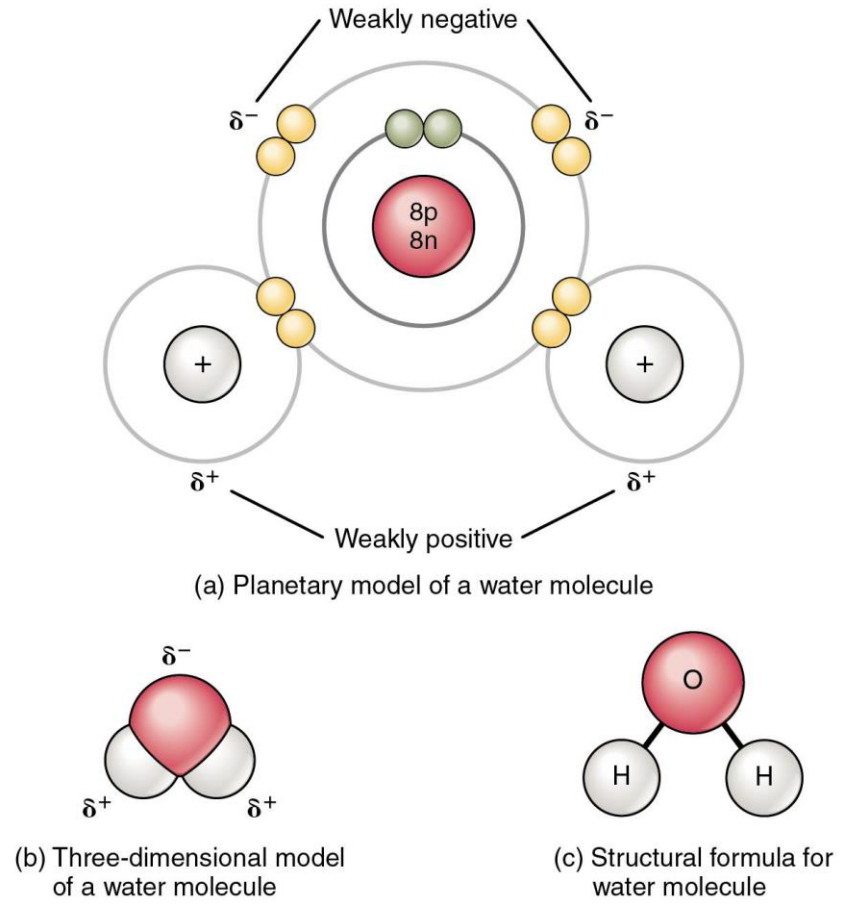


Figure 2.10 Polar Covalent Bonds in a Water Molecule

What is true for the bonds is true for the water molecule as a whole; that is, the oxygen region has a slightly negative charge and the regions of the hydrogen atoms have a slightly positive charge. These charges are often referred to as "partial charges" because the strength of the charge is less than one full electron, as would occur in an ionic bond. As shown in Figure 2.10, regions of weak polarity are indicated with the Greek letter delta (δ) and a plus (+) or minus (–) sign.

Even though a single water molecule is unimaginably tiny, it has mass, and the opposing electrical charges on the molecule pull that mass in such a way that it creates a shape somewhat like a triangular tent (see Figure 2.10**b**). This dipole, with the positive charges at one end formed by the hydrogen atoms at the "bottom" of the tent and the negative charge at the opposite end (the oxygen atom at the "top" of the tent) makes the charged regions highly likely to interact with charged regions of other polar molecules. For human physiology, the resulting bond is one of the most important formed by water—the hydrogen bond.

Hydrogen Bonds

A hydrogen bond is formed when a weakly positive hydrogen atom already bonded to one electronegative atom (for example, the oxygen in the water molecule) is attracted to another electronegative atom from another molecule. In other words, hydrogen bonds always include hydrogen that is already part of a polar molecule.

The most common example of hydrogen bonding in the natural world occurs between molecules of water. It happens before your eyes whenever two raindrops merge into a larger bead, or a creek spills into a river. Hydrogen bonding occurs because the weakly negative oxygen atom in one water molecule is attracted to the weakly positive hydrogen atoms of two other water molecules (Figure 2.11).

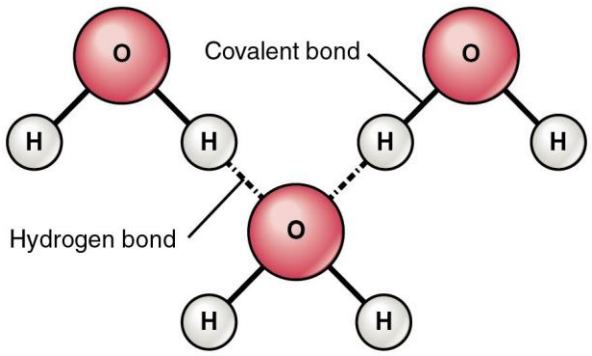


Figure 2.11 Hydrogen Bonds between Water Molecules Notice that the bonds occur between the weakly positive charge on the hydrogen atoms and the weakly negative charge on the oxygen atoms. Hydrogen bonds are relatively weak, and therefore are indicated with a dotted (rather than a solid) line.

Water molecules also strongly attract other types of charged molecules as well as ions. This explains why "table salt," for example, actually is a molecule called a "salt" in chemistry, which consists of equal numbers of positively-charged sodium (Na+) and negatively-charged chloride

(Cl–), dissolves so readily in water, in this case forming dipole-ion bonds between the water and the electrically-charged ions (electrolytes). Water molecules also repel molecules with nonpolar covalent bonds, like fats, lipids, and oils. You can demonstrate this with a simple kitchen experiment: pour a teaspoon of vegetable oil, a compound formed by nonpolar covalent bonds, into a glass of water. Instead of instantly dissolving in the water, the oil forms a distinct bead because the polar water molecules repel the nonpolar oil.

Chapter Attributions

"Anatomy and Physiology" by J. Gordon Betts, Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, Peter DeSaix, OpenStax is licensed under CC BY 4.0. Access for free at https://openstax.org/books/anatomy-and-physiology/pages/2-2-chemical-bonds

4. Bonding and Chemical Notation

(Review previous section on Chemical Bonds, pp. 26-33)

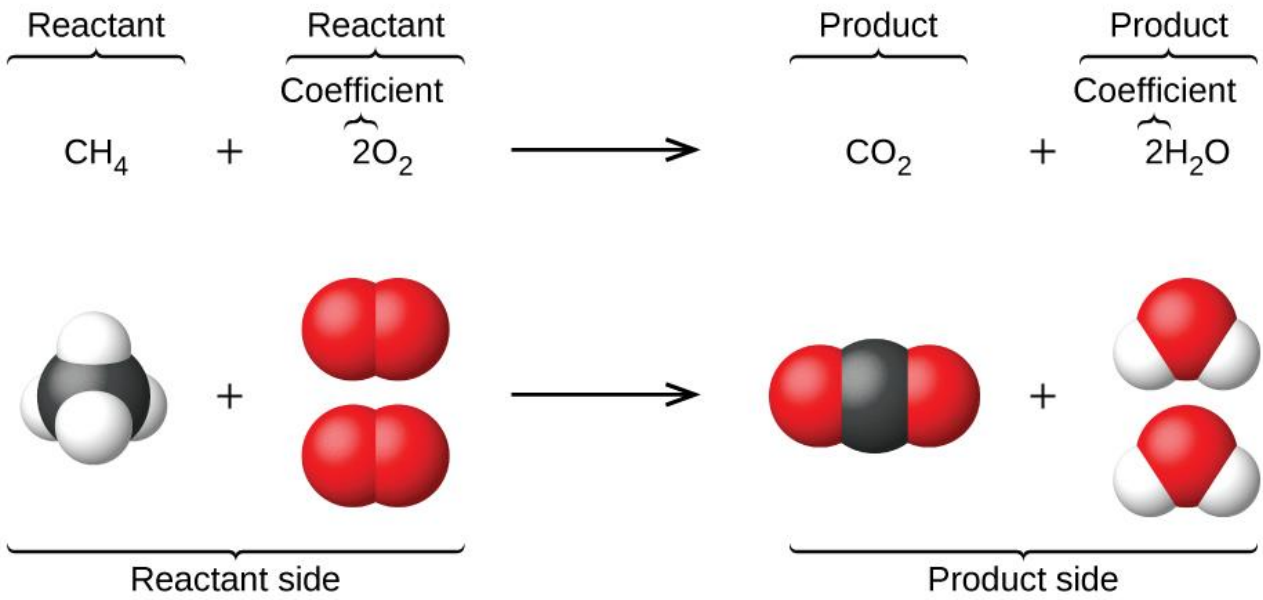
5. Balancing Equations

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

- Derive chemical equations from narrative descriptions of chemical reactions.
- Write and balance chemical equations in molecular, total ionic, and net ionic formats.

The preceding chapter introduced the use of element symbols to represent individual atoms. When atoms gain or lose electrons to yield ions, or combine with other atoms to form molecules, their symbols are modified or combined to generate chemical formulas that appropriately represent these species. Extending this symbolism to represent both the identities and the relative quantities of substances undergoing a chemical (or physical) change involves writing and balancing a **chemical equation**. Consider as an example the reaction between one methane molecule (CH_4) and two diatomic oxygen molecules (O_2) to produce one carbon dioxide molecule (CO_2) and two water molecules (H_2O). The chemical equation

representing this process is provided in the upper half of Figure, with space-filling molecular models shown in the lower half of the figure.



The reaction between methane and oxygen to yield carbon dioxide and water (shown at bottom) may be represented by a chemical equation using formulas (top).

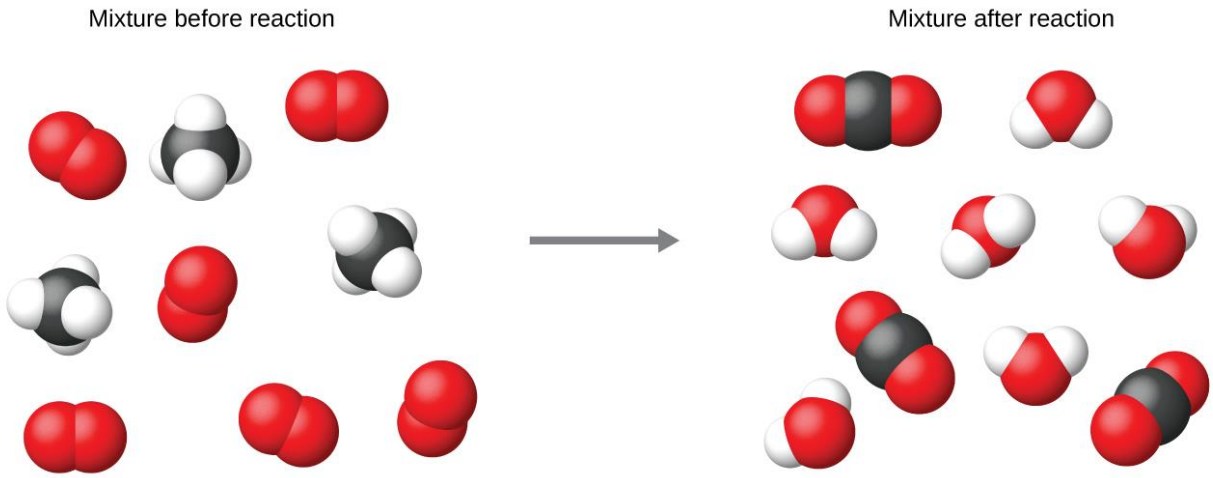
This example illustrates the fundamental aspects of any chemical equation:

1. The substances undergoing reaction are called **reactants**, and their formulas are placed on the left side of the equation.
2. The substances generated by the reaction are called **products**, and their formulas are placed on the right side of the equation.
3. Plus signs (+) separate individual reactant and product formulas, and an arrow (⟶)(⟶) separates the reactant and product (left and right) sides of the equation.
4. The relative numbers of reactant and product species are represented by **coefficients** (numbers placed immediately to the left of each formula). A coefficient of 1 is typically omitted.

It is common practice to use the smallest possible whole-number coefficients in a chemical equation, as is done in this example. Realize, however, that these coefficients represent the *relative* numbers of reactants and products, and, therefore, they may be correctly

interpreted as ratios. Methane and oxygen react to yield carbon dioxide and water in a 1:2:1:2 ratio. This ratio is satisfied if the numbers of these molecules are, respectively, 1-2-1-2, or 2-4-2-4, or 3-6-3-6, and so on (Figure). Likewise, these coefficients may be interpreted with regard to any amount (number) unit, and so this equation may be correctly read in many ways, including:

- *One* methane molecule and *two* oxygen molecules react to yield *one* carbon dioxide molecule and *two* water molecules.
- *One dozen* methane molecules and *two dozen* oxygen molecules react to yield *one dozen* carbon dioxide molecules and *two dozen* water molecules.
- *One mole* of methane molecules and *2 moles* of oxygen molecules react to yield *1 mole* of carbon dioxide molecules and *2 moles* of water molecules.



Regardless of the absolute numbers of molecules involved, the ratios between numbers of molecules of each species that react (the reactants) and molecules of each species that form (the products) are the same and are given by the chemical reaction equation.

Balancing Equations

The chemical equation described in section 4.1 is **balanced**, meaning that equal numbers of atoms for each element involved in the reaction are represented on the reactant and product sides. This is a requirement the equation must satisfy to be consistent with the law of conservation of matter. It may be confirmed by simply summing the numbers of atoms on

either side of the arrow and comparing these sums to ensure they are equal. Note that the number of atoms for a given element is calculated by multiplying the coefficient of any formula containing that element by the element's subscript in the formula. If an element appears in more than one formula on a given side of the equation, the number of atoms represented in each must be computed and then added together. For example, both product species in the example reaction, CO_2 and H_2O, contain the element oxygen, and so the number of oxygen atoms on the product side of the equation is

(1CO2molecule×2 O atomsCO2molecule)+(2H2O molecules×1 O atomH2O molecule)=4 O atoms(1CO2molecule×2 O atomsCO2molecule)+(2H2O molecules×1 O atomH2O molecule)=4 O atoms

The equation for the reaction between methane and oxygen to yield carbon dioxide and water is confirmed to be balanced per this approach, as shown here:

CH4+2O2⟶CO2+2H2OCH4+2O2⟶CO2+2H2O

Element	**Reactants**	**Products**	**Balanced?**
C	1 ×× 1 = 1	1 ×× 1 = 1	1 = 1, yes
H	4 ×× 1 = 4	2 ×× 2 = 4	4 = 4, yes
O	2 ×× 2 = 4	(1 ×× 2) + (2 ×× 1) = 4	4 = 4, yes

A balanced chemical equation often may be derived from a qualitative description of some chemical reaction by a fairly simple approach known as balancing by inspection. Consider as an example the decomposition of water to yield molecular hydrogen and oxygen. This process is represented qualitatively by an *unbalanced* chemical equation:

H2O⟶H2+O2(unbalanced)H2O⟶H2+O2(unbalanced)

Comparing the number of H and O atoms on either side of this equation confirms its imbalance:

Element	Reactants	Products	Balanced?
H	1 ×× 2 = 2	1 ×× 2 = 2	2 = 2, yes
O	1 ×× 1 = 1	1 ×× 2 = 2	1 ≠ 2, no

The numbers of H atoms on the reactant and product sides of the equation are equal, but the numbers of O atoms are not. To achieve balance, the *coefficients* of the equation may be changed as needed. Keep in mind, of course, that the *formula subscripts* define, in part, the identity of the substance, and so these cannot be changed without altering the qualitative meaning of the equation. For example, changing the reactant formula from H_2O to H_2O_2 would yield balance in the number of atoms, but doing so also changes the reactant's identity (it's now hydrogen peroxide and not water). The O atom balance may be achieved by changing the coefficient for H_2O to 2.

Element	Reactants	Products	Balanced?
H	**2** ×× 2 = 4	1 ×× 2 = 2	4 ≠ 2, no
O	2 ×× 1 = 2	1 ×× 2 = 2	2 = 2, yes

2H2O⟶H2+O2(unbalanced)2H2O⟶H2+O2(unbalanced)

The H atom balance was upset by this change, but it is easily reestablished by changing the coefficient for the H_2 product to 2.

2H2O⟶2H2+O2(balanced)2H2O⟶2H2+O2(balanced)

Element	Reactants	Products	Balanced?
H	2 ×× 2 = 4	**2** ×× 2 = 4	4 = 4, yes
O	2 ×× 1 = 2	1 ×× 2 = 2	2 = 2, yes

These coefficients yield equal numbers of both H and O atoms on the reactant and product sides, and the balanced equation is, therefore:

2H2O⟶2H2+O22H2O⟶2H2+O2

Balancing Chemical Equations. Write a balanced equation for the reaction of molecular nitrogen (N_2) and oxygen (O_2) to form dinitrogen pentoxide.

Solution. First, write the unbalanced equation.

N2+O2⟶N2O5(unbalanced)N2+O2⟶N2O5(unbalanced)

Next, count the number of each type of atom present in the unbalanced equation.

Element	**Reactants**	**Products**	**Balanced?**
N	1 ×× 2 = 2	1 ×× 2 = 2	2 = 2, yes
O	1 ×× 2 = 2	1 ×× 5 = 5	2 ≠ 5, no

Though nitrogen is balanced, changes in coefficients are needed to balance the number of oxygen atoms. To balance the number of oxygen atoms, a reasonable first attempt would be to change the coefficients for the O_2 and N_2O_5 to integers that will yield 10 O atoms (the least common multiple for the O atom subscripts in these two formulas).

N2+5O2⟶2N2O5(unbalanced)N2+5O2⟶2N2O5(unbalanced)

Element	**Reactants**	**Products**	**Balanced?**
N	1 ×× 2 = 2	**2** ×× 2 = 4	2 ≠ 4, no
O	**5** ×× 2 = 10	**2** ×× 5 = 10	10 = 10, yes

The N atom balance has been upset by this change; it is restored by changing the coefficient for the reactant N_2 to 2.

2N2+5O2⟶2N2O52N2+5O2⟶2N2O5

Element	Reactants	Products	Balanced?
N	**2** ×× 2 = 4	2 ×× 2 = 4	4 = 4, yes
O	5 ×× 2 = 10	2 ×× 5 = 10	10 = 10, yes

The numbers of N and O atoms on either side of the equation are now equal, and so the equation is balanced.

Check Your Learning. Write a balanced equation for the decomposition of ammonium nitrate to form molecular nitrogen, molecular oxygen, and water. (Hint: Balance oxygen last, since it is present in more than one molecule on the right side of the equation.)

ANSWER:

2NH4NO3⟶2N2+O2+4H2O2NH4NO3⟶2N2+O2+4H2O

It is sometimes convenient to use fractions instead of integers as intermediate coefficients in the process of balancing a chemical equation. When balance is achieved, all the equation's coefficients may then be multiplied by a whole number to convert the fractional coefficients to integers without upsetting the atom balance. For example, consider the reaction of ethane (C_2H_6) with oxygen to yield H_2O and CO_2, represented by the unbalanced equation:

C2H6+O2⟶H2O+CO2(unbalanced)C2H6+O2⟶H2O+CO2(unbalanced)

Following the usual inspection approach, one might first balance C and H atoms by changing the coefficients for the two product species, as shown:

C2H6+O2⟶3H2O+2CO2(unbalanced)C2H6+O2⟶3H2O+2CO2(unbalanced)

This results in seven O atoms on the product side of the equation, an odd number—no integer coefficient can be used with the O_2 reactant to yield an odd number, so a fractional coefficient, 72,72, is used instead to yield a provisional balanced equation:

C2H6+72O2⟶3H2O+2CO2C2H6+72O2⟶3H2O+2CO2

A conventional balanced equation with integer-only coefficients is derived by multiplying each coefficient by 2:

$$2C_2H_6+7O_2 \longrightarrow 6H_2O+4CO_2 2C_2H_6+7O_2 \longrightarrow 6H_2O+4CO_2$$

Finally with regard to balanced equations, recall that convention dictates use of the *smallest whole-number coefficients*. Although the equation for the reaction between molecular nitrogen and molecular hydrogen to produce ammonia is, indeed, balanced,

$$3N_2+9H_2 \longrightarrow 6NH_3 3N_2+9H_2 \longrightarrow 6NH_3$$

the coefficients are not the smallest possible integers representing the relative numbers of reactant and product molecules. Dividing each coefficient by the greatest common factor, 3, gives the preferred equation:

$$N_2+3H_2 \longrightarrow 2NH_3 N_2+3H_2 \longrightarrow 2NH_3$$

Use this interactive tutorial (https://phet.colorado.edu/en/simulation/balancing-chemical-equations) for additional practice balancing equations.

Additional Information in Chemical Equations

The physical states of reactants and products in chemical equations very often are indicated with a parenthetical abbreviation following the formulas. Common abbreviations include *s* for solids, *l* for liquids, *g* for gases, and *aq* for substances dissolved in water (*aqueous solutions*, as introduced in the preceding chapter). These notations are illustrated in the example equation here:

$$2Na(s)+2H_2O(l) \longrightarrow 2NaOH(aq)+H_2(g) 2Na(s)+2H_2O(l) \longrightarrow 2NaOH(aq)+H_2(g)$$

This equation represents the reaction that takes place when sodium metal is placed in water. The solid sodium reacts with liquid water to produce molecular hydrogen gas and the ionic compound sodium hydroxide (a solid in pure form, but readily dissolved in water).

Special conditions necessary for a reaction are sometimes designated by writing a word or symbol above or below the equation's arrow. For example, a reaction carried out by heating may be indicated by the uppercase Greek letter delta (Δ) over the arrow.

CaCO3(s)⟶ΔCaO(s)+CO2(g)CaCO3(s)⟶ΔCaO(s)+CO2(g)

Other examples of these special conditions will be encountered in more depth in later chapters.

Equations for Ionic Reactions

Given the abundance of water on earth, it stands to reason that a great many chemical reactions take place in aqueous media. When ions are involved in these reactions, the chemical equations may be written with various levels of detail appropriate to their intended use. To illustrate this, consider a reaction between ionic compounds taking place in an aqueous solution. When aqueous solutions of $CaCl_2$ and $AgNO_3$ are mixed, a reaction takes place producing aqueous $Ca(NO_3)_2$ and solid AgCl:

CaCl2(aq)+2AgNO3(aq)⟶Ca(NO3)2(aq)+2AgCl(s)CaCl2(aq)+2AgNO3(aq)⟶Ca(NO3)2(aq)+2AgCl(s)

This balanced equation, derived in the usual fashion, is called a **molecular equation** because it doesn't explicitly represent the ionic species that are present in solution. When ionic compounds dissolve in water, they may *dissociate* into their constituent ions, which are subsequently dispersed homogenously throughout the resulting solution (a thorough discussion of this important process is provided in the chapter on solutions). Ionic compounds dissolved in water are, therefore, more realistically represented as dissociated ions, in this case:

CaCl2(aq)⟶Ca2+(aq)+2Cl−(aq)2AgNO3(aq)⟶2Ag+(aq)+2NO3−(aq)Ca(NO3)2(aq)⟶Ca2+(aq)+2NO3−(aq)CaCl2(aq)⟶Ca2+(aq)+2Cl−(aq)2AgNO3(aq)⟶2Ag+(aq)+2NO3−(aq)Ca(NO3)2(aq)⟶Ca2+(aq)+2NO3−(aq)

Unlike these three ionic compounds, AgCl does not dissolve in water to a significant extent, as signified by its physical state notation, *s*.

Explicitly representing all dissolved ions results in a **complete ionic equation**. In this particular case, the formulas for the dissolved ionic compounds are replaced by formulas for their dissociated ions:

Ca2+(aq)+2Cl−(aq)+2Ag+(aq)+2NO3−(aq)⟶Ca2+(aq)+2NO3−(aq)+2Ag Cl(s)Ca2+(aq)+2Cl−(aq)+2Ag+(aq)+2NO3−(aq)⟶Ca2+(aq)+2NO3−(aq)+2AgCl(s)

Examining this equation shows that two chemical species are present in identical form on both sides of the arrow, $Ca^{2+}(aq)$ and NO3−(aq).NO3−(aq). These **spectator ions**—ions whose presence is required to maintain charge neutrality—are neither chemically nor physically changed by the process, and so they may be eliminated from the equation to yield a more succinct representation called a **net ionic equation**:

Ca2+(aq)+2Cl−(aq)+2Ag+(aq)+2NO3−(aq)⟶Ca2+(aq)+2NO3−(aq)+2AgCl(s)2Cl−(aq)+2Ag+(aq)⟶2AgCl(s)Ca2+(aq)+2Cl−(aq)+2Ag+(aq)+2NO3−(aq)⟶Ca2+(aq)+2NO3−(aq)+2AgCl(s)2Cl−(aq)+2Ag+(aq)⟶2AgCl(s)

Following the convention of using the smallest possible integers as coefficients, this equation is then written:

Cl−(aq)+Ag+(aq)⟶AgCl(s)Cl−(aq)+Ag+(aq)⟶AgCl(s)

This net ionic equation indicates that solid silver chloride may be produced from dissolved chloride and silver(I) ions, regardless of the source of these ions. These molecular and complete ionic equations provide additional information, namely, the ionic compounds used as sources of Cl^- and Ag^+.

Molecular and Ionic Equations. When carbon dioxide is dissolved in an aqueous solution of sodium hydroxide, the mixture reacts to yield aqueous sodium carbonate and liquid water. Write balanced molecular, complete ionic, and net ionic equations for this process.

Solution. Begin by identifying formulas for the reactants and products and arranging them properly in chemical equation form:

CO2(aq)+NaOH(aq)⟶Na2CO3(aq)+H2O(l)(unbalanced)CO2(aq)+NaOH(aq)⟶Na2CO3(aq)+H2O(l)(unbalanced)

Balance is achieved easily in this case by changing the coefficient for NaOH to 2, resulting in the molecular equation for this reaction:

CO2(aq)+2NaOH(aq)⟶Na2CO3(aq)+H2O(l)CO2(aq)+2NaOH(aq)⟶Na2CO3(aq)+H2O(l)

The two dissolved ionic compounds, NaOH and Na_2CO_3, can be represented as dissociated ions to yield the complete ionic equation:

CO2(aq)+2Na+(aq)+2OH−(aq)⟶2Na+(aq)+CO32−(aq)+H2O(l)CO2(aq)+2Na+(aq)+2OH−(aq)⟶2Na+(aq)+CO32−(aq)+H2O(l)

Finally, identify the spectator ion(s), in this case $Na^+(aq)$, and remove it from each side of the equation to generate the net ionic equation:

CO2(aq)+2Na+(aq)+2OH−(aq)⟶2Na+(aq)+CO32−(aq)+H2O(l)CO2(aq)+2OH−(aq)⟶CO32−(aq)+H2O(l)CO2(aq)+2Na+(aq)+2OH−(aq)⟶2Na+(aq)+CO32−(aq)+H2O(l)CO2(aq)+2OH−(aq)⟶CO32−(aq)+H2O(l)

Check Your Learning. Diatomic chlorine and sodium hydroxide (lye) are commodity chemicals produced in large quantities, along with diatomic hydrogen, via the electrolysis of brine, according to the following unbalanced equation:

NaCl(aq)+H2O(l)−→−−−−electricityNaOH(aq)+H2(g)+Cl2(g)NaCl(aq)+H2O(l)→electricityNaOH(aq)+H2(g)+Cl2(g)

Write balanced molecular, complete ionic, and net ionic equations for this process.

ANSWER:

2NaCl(aq)+2H2O(l)⟶2NaOH(aq)+H2(g)+Cl2(g)(molecular)2NaCl(aq)+2H2O(l)⟶2NaOH(aq)+H2(g)+Cl2(g)(molecular)
2Na+(aq)+2Cl−(aq)+2H2O(l)⟶2Na+(aq)+2OH−(aq)+H2(g)+Cl2(g)(complete ionic)2Na+(aq)+2Cl−(aq)+2H2O(l)⟶2Na+(aq)+2OH−(aq)+H2(g)+Cl2(g)(complete ionic)
2Cl−(aq)+2H2O(l)⟶2OH−(aq)+H2(g)+Cl2(g)(net ionic)2Cl−(aq)+2H2O(l)⟶2OH−(aq)+H2(g)+Cl2(g)(net ionic)

Key Concepts and Summary

Chemical equations are symbolic representations of chemical and physical changes. Formulas for the substances undergoing the change (reactants) and substances generated by the change (products) are separated by an arrow and preceded by integer coefficients indicating their relative numbers. Balanced equations are those whose coefficients result in equal numbers of atoms for each element in the reactants and products. Chemical reactions in aqueous solution that involve ionic reactants or products may be represented more realistically by complete ionic equations and, more succinctly, by net ionic equations.

Chemistry End of Chapter Exercises

What does it mean to say an equation is balanced? Why is it important for an equation to be balanced?

Consider molecular, complete ionic, and net ionic equations.

(a) What is the difference between these types of equations?

(b) In what circumstance would the complete and net ionic equations for a reaction be identical?

Balance the following equations:

(a) $PCl5(s)+H2O(l)⟶POCl3(l)+HCl(aq)PCl5(s)+H2O(l)⟶POCl3(l)+HCl(aq)$

(b) $Cu(s)+HNO3(aq)⟶Cu(NO3)2(aq)+H2O(l)+NO(g)Cu(s)+HNO3(aq)⟶Cu(NO3)2(aq)+H2O(l)+NO(g)$

(c) $H2(g)+I2(s)⟶HI(s)H2(g)+I2(s)⟶HI(s)$

(d) $Fe(s)+O2(g)⟶Fe2O3(s)Fe(s)+O2(g)⟶Fe2O3(s)$

(e) $Na(s)+H2O(l)⟶NaOH(aq)+H2(g)Na(s)+H2O(l)⟶NaOH(aq)+H2(g)$

(f) $(NH4)2Cr2O7(s)⟶Cr2O3(s)+N2(g)+H2O(g)(NH4)2Cr2O7(s)⟶Cr2O3(s)+N2(g)+H2O(g)$

(g) $P4(s)+Cl2(g)⟶PCl3(l)P4(s)+Cl2(g)⟶PCl3(l)$

(h) $PtCl4(s)⟶Pt(s)+Cl2(g)PtCl4(s)⟶Pt(s)+Cl2(g)$

Balance the following equations:

(a) Ag(s)+H2S(g)+O2(g)⟶Ag2S(s)+H2O(l)Ag(s)+H2S(g)+O2(g)⟶Ag2S(s)+H2O(l)

(b) P4(s)+O2(g)⟶P4O10(s)P4(s)+O2(g)⟶P4O10(s)

(c) Pb(s)+H2O(l)+O2(g)⟶Pb(OH)2(s)Pb(s)+H2O(l)+O2(g)⟶Pb(OH)2(s)

(d) Fe(s)+H2O(l)⟶Fe3O4(s)+H2(g)Fe(s)+H2O(l)⟶Fe3O4(s)+H2(g)

(e) Sc2O3(s)+SO3(l)⟶Sc2(SO4)3(s)Sc2O3(s)+SO3(l)⟶Sc2(SO4)3(s)

(f) Ca3(PO4)2(aq)+H3PO4(aq)⟶Ca(H2PO4)2(aq)Ca3(PO4)2(aq)+H3PO4(aq)⟶Ca(H2PO4)2(aq)

(g) Al(s)+H2SO4(aq)⟶Al2(SO4)3(s)+H2(g)Al(s)+H2SO4(aq)⟶Al2(SO4)3(s)+H2(g)

(h) TiCl4(s)+H2O(g)⟶TiO2(s)+HCl(g)TiCl4(s)+H2O(g)⟶TiO2(s)+HCl(g)

Write a balanced molecular equation describing each of the following chemical reactions.

(a) Solid calcium carbonate is heated and decomposes to solid calcium oxide and carbon dioxide gas.

(b) Gaseous butane, C_4H_{10}, reacts with diatomic oxygen gas to yield gaseous carbon dioxide and water vapor.

(c) Aqueous solutions of magnesium chloride and sodium hydroxide react to produce solid magnesium hydroxide and aqueous sodium chloride.

(d) Water vapor reacts with sodium metal to produce solid sodium hydroxide and hydrogen gas.

Write a balanced equation describing each of the following chemical reactions.

(a) Solid potassium chlorate, $KClO_3$, decomposes to form solid potassium chloride and diatomic oxygen gas.

(b) Solid aluminum metal reacts with solid diatomic iodine to form solid Al_2I_6.

(c) When solid sodium chloride is added to aqueous sulfuric acid, hydrogen chloride gas and aqueous sodium sulfate are produced.

(d) Aqueous solutions of phosphoric acid and potassium hydroxide react to produce aqueous potassium dihydrogen phosphate and liquid water.

Colorful fireworks often involve the decomposition of barium nitrate and potassium chlorate and the reaction of the metals magnesium, aluminum, and iron with oxygen.

(a) Write the formulas of barium nitrate and potassium chlorate.

(b) The decomposition of solid potassium chlorate leads to the formation of solid potassium chloride and diatomic oxygen gas. Write an equation for the reaction.

(c) The decomposition of solid barium nitrate leads to the formation of solid barium oxide, diatomic nitrogen gas, and diatomic oxygen gas. Write an equation for the reaction.

(d) Write separate equations for the reactions of the solid metals magnesium, aluminum, and iron with diatomic oxygen gas to yield the corresponding metal oxides. (Assume the iron oxide contains Fe^{3+} ions.)

Fill in the blank with a single chemical formula for a covalent compound that will balance the equation:

Aqueous hydrogen fluoride (hydrofluoric acid) is used to etch glass and to analyze minerals for their silicon content. Hydrogen fluoride will also react with sand (silicon dioxide).

(a) Write an equation for the reaction of solid silicon dioxide with hydrofluoric acid to yield gaseous silicon tetrafluoride and liquid water.

(b) The mineral fluorite (calcium fluoride) occurs extensively in Illinois. Solid calcium fluoride can also be prepared by the reaction of aqueous solutions of calcium chloride

and sodium fluoride, yielding aqueous sodium chloride as the other product. Write complete and net ionic equations for this reaction.

A novel process for obtaining magnesium from sea water involves several reactions. Write a balanced chemical equation for each step of the process.

(a) The first step is the decomposition of solid calcium carbonate from seashells to form solid calcium oxide and gaseous carbon dioxide.

(b) The second step is the formation of solid calcium hydroxide as the only product from the reaction of the solid calcium oxide with liquid water.

(c) Solid calcium hydroxide is then added to the seawater, reacting with dissolved magnesium chloride to yield solid magnesium hydroxide and aqueous calcium chloride.

(d) The solid magnesium hydroxide is added to a hydrochloric acid solution, producing dissolved magnesium chloride and liquid water.

(e) Finally, the magnesium chloride is melted and electrolyzed to yield liquid magnesium metal and diatomic chlorine gas.

From the balanced molecular equations, write the complete ionic and net ionic equations for the following:

(a) K2C2O4(aq)+Ba(OH)2(aq)⟶2KOH(aq)+BaC2O4(s)K2C2O4(aq)+Ba(OH)2(aq)⟶2KOH(aq)+BaC2O4(s)

(b) Pb(NO3)2(aq)+H2SO4(aq)⟶PbSO4(s)+2HNO3(aq)Pb(NO3)2(aq)+H2SO4(aq)⟶PbSO4(s)+2HNO3(aq)

(c) CaCO3(s)+H2SO4(aq)⟶CaSO4(s)+CO2(g)+H2O(l)CaCO3(s)+H2SO4(aq)⟶CaSO4(s)+CO2(g)+H2O(l)

Chapter Glossary

balanced equation. chemical equation with equal numbers of atoms for each element in the reactant and product

chemical equation. symbolic representation of a chemical reaction

coefficient. number placed in front of symbols or formulas in a chemical equation to indicate their relative amount

complete ionic equation. chemical equation in which all dissolved ionic reactants and products, including spectator ions, are explicitly represented by formulas for their dissociated ions

molecular equation. chemical equation in which all reactants and products are represented as neutral substances

net ionic equation. chemical equation in which only those dissolved ionic reactants and products that undergo a chemical or physical change are represented (excludes spectator ions)

product. substance formed by a chemical or physical change; shown on the right side of the arrow in a chemical equation

reactant. substance undergoing a chemical or physical change; shown on the left side of the arrow in a chemical equation

spectator ion. ion that does not undergo a chemical or physical change during a reaction, but its presence is required to maintain charge neutrality**.**

Chapter Attributions

6. Types of Energy (Chemical Reactions)

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

- Distinguish between kinetic and potential energy, and between exergonic and endergonic chemical reactions
- Identify four forms of energy important in human functioning
- Describe the three basic types of chemical reactions
- Identify several factors influencing the rate of chemical reactions

One characteristic of a living organism is metabolism, which is the sum total of all of the chemical reactions that go on to maintain that organism's health and life. The bonding processes you have learned thus far are anabolic chemical reactions; that is, they form larger molecules from smaller molecules or atoms. But recall that metabolism can proceed in another direction: in catabolic chemical reactions, bonds between components of larger molecules break, releasing smaller molecules or atoms. Both types of reaction involve exchanges not only of matter, but of energy.

The Role of Energy in Chemical Reactions

Chemical reactions require a sufficient amount of energy to cause the matter to collide with enough precision and force that old chemical bonds can be broken and new ones formed. In general, **kinetic energy** is the form of energy powering any type of matter in motion. Imagine you are building a brick wall. The energy it takes to lift and place one brick atop another is kinetic energy—the energy matter possesses because of its motion. Once the wall is in place, it stores potential energy. **Potential energy** is the energy of position, or the energy matter possesses because of the positioning or structure of its components. If the brick wall collapses, the stored potential energy is released as kinetic energy as the bricks fall.

In the human body, potential energy is stored in the bonds between atoms and molecules. **Chemical energy** is the form of potential energy in which energy is stored in chemical bonds. When those bonds are formed, chemical energy is invested, and when they break, chemical energy is released. Notice that chemical energy, like all energy, is neither created nor destroyed; rather, it is converted from one form to another. When you eat an energy bar before heading out the door for a hike, the honey, nuts, and other foods the bar contains are broken down and rearranged by your body into molecules that your muscle cells convert to kinetic energy.

Chemical reactions that release more energy than they absorb are characterized as exergonic. The catabolism of the foods in your energy bar is an example. Some of the chemical energy stored in the bar is absorbed into molecules your body uses for fuel, but some of it is released—for example, as heat. In contrast, chemical reactions that absorb more energy than they release are endergonic. These reactions require energy input, and the resulting molecule

stores not only the chemical energy in the original components, but also the energy that fueled the reaction. Because energy is neither created nor destroyed, where does the energy needed for endergonic reactions come from? In many cases, it comes from exergonic reactions.

Forms of Energy Important in Human Functioning

You have already learned that chemical energy is absorbed, stored, and released by chemical bonds. In addition to chemical energy, mechanical, radiant, and electrical energy are important in human functioning.

- Mechanical energy, which is stored in physical systems such as machines, engines, or the human body, directly powers the movement of matter. When you lift a brick into place on a wall, your muscles provide the mechanical energy that moves the brick.
- Radiant energy is energy emitted and transmitted as waves rather than matter. These waves vary in length from long radio waves and microwaves to short gamma waves emitted from decaying atomic nuclei. The full spectrum of radiant energy is referred to as the electromagnetic spectrum. The body uses the ultraviolet energy of sunlight to convert a compound in skin cells to vitamin D, which is essential to human functioning. The human eye evolved to see the wavelengths that comprise the colors of the rainbow, from red to violet, so that range in the spectrum is called “visible light.”
- Electrical energy, supplied by electrolytes in cells and body fluids, contributes to the voltage changes that help transmit impulses in nerve and muscle cells.

Characteristics of Chemical Reactions

All chemical reactions begin with a **reactant**, the general term for the one or more substances that enter into the reaction. Sodium and chloride ions, for example, are the reactants in the production of table salt. The one or more substances produced by a chemical reaction are called the **product**.

In chemical reactions, the components of the reactants—the elements involved and the number of atoms of each—are all present in the product(s). Similarly, there is nothing present in the products that are not present in the reactants. This is because chemical reactions are

governed by the law of conservation of mass, which states that matter cannot be created or destroyed in a chemical reaction.

Just as you can express mathematical calculations in equations such as 2 + 7 = 9, you can use chemical equations to show how reactants become products. As in math, chemical equations proceed from left to right, but instead of an equal sign, they employ an arrow or arrows indicating the direction in which the chemical reaction proceeds. For example, the chemical reaction in which one atom of nitrogen and three atoms of hydrogen produce ammonia would be written as $\mathrm{N} + 3\mathrm{H} \rightarrow \mathrm{NH_3}$. Correspondingly, the breakdown of ammonia into its components would be written as $\mathrm{NH_3} \rightarrow \mathrm{N} + 3\mathrm{H}$.

Notice that, in the first example, a nitrogen (N) atom and three hydrogen (H) atoms bond to form a compound. This anabolic reaction requires energy, which is then stored within the compound's bonds. Such reactions are referred to as synthesis reactions. A **synthesis reaction** is a chemical reaction that results in the synthesis (joining) of components that were formerly separate (Figure 2.12**a**). Again, nitrogen and hydrogen are reactants in a synthesis reaction that yields ammonia as the product. The general equation for a synthesis reaction is A + B→AB.

a) In a synthesis reaction, two components bond to make a larger molecule. Energy is required and is stored in the bond:

NOTE + BOOK ⟶ NOTEBOOK

b) In a decomposition reaction, bonds between components of a larger molecule are broken, resulting in smaller products:

BOOKWORM ⟶ BOOK + WORM

c) In an exchange reaction, bonds are both formed and broken such that the components of the reactants are rearranged:

NOTEBOOK + WORM ⟶ NOTE + BOOKWORM

Figure 2.12 The Three Fundamental Chemical Reactions The atoms and molecules involved in the three fundamental chemical reactions can be imagined as words.

In the second example, ammonia is catabolized into its smaller components, and the potential energy that had been stored in its bonds is released. Such reactions are referred to as decomposition reactions. A **decomposition reaction** is a chemical reaction that breaks down or

"de-composes" something larger into its constituent parts (see Figure 2.12**b**). The general equation for a decomposition reaction is: AB→A+B.

An **exchange reaction** is a chemical reaction in which both synthesis and decomposition occur, chemical bonds are both formed and broken, and chemical energy is absorbed, stored, and released (see Figure 2.12**c**). The simplest form of an exchange reaction might be: A+BC→AB+C. Notice that, to produce these products, B and C had to break apart in a decomposition reaction, whereas A and B had to bond in a synthesis reaction. A more complex exchange reaction might be: AB+CD→AC+BD. Another example might be: AB+CD→AD+BC.

In theory, any chemical reaction can proceed in either direction under the right conditions. Reactants may synthesize into a product that is later decomposed. Reversibility is also a quality of exchange reactions. For instance, A+BC→AB+C could then reverse to AB+C→A+BC. This reversibility of a chemical reaction is indicated with a double arrow: A+BC⇄AB+C. Still, in the human body, many chemical reactions do proceed in a predictable direction, either one way or the other. You can think of this more predictable path as the path of least resistance because, typically, the alternate direction requires more energy.

Factors Influencing the Rate of Chemical Reactions

If you pour vinegar into baking soda, the reaction is instantaneous; the concoction will bubble and fizz. But many chemical reactions take time. A variety of factors influence the rate of chemical reactions. This section, however, will consider only the most important in human functioning.

Properties of the Reactants

If chemical reactions are to occur quickly, the atoms in the reactants have to have easy access to one another. Thus, the greater the surface area of the reactants, the more readily they will interact. When you pop a cube of cheese into your mouth, you chew it before you swallow it. Among other things, chewing increases the surface area of the food so that digestive chemicals can more easily get at it. As a general rule, gases tend to react faster than liquids or solids, again because it takes energy to separate particles of a substance, and gases by definition already have space between their particles. Similarly, the larger the molecule, the greater the number

of total bonds, so reactions involving smaller molecules, with fewer total bonds, would be expected to proceed faster.

In addition, recall that some elements are more reactive than others. Reactions that involve highly reactive elements like hydrogen proceed more quickly than reactions that involve less reactive elements. Reactions involving stable elements like helium are not likely to happen at all.

Temperature

Nearly all chemical reactions occur at a faster rate at higher temperatures. Recall that kinetic energy is the energy of matter in motion. The kinetic energy of subatomic particles increases in response to increases in thermal energy. The higher the temperature, the faster the particles move, and the more likely they are to come in contact and react.

Concentration and Pressure

If just a few people are dancing at a club, they are unlikely to step on each other's toes. But as more and more people get up to dance—especially if the music is fast—collisions are likely to occur. It is the same with chemical reactions: the more particles present within a given space, the more likely those particles are to bump into one another. This means that chemists can speed up chemical reactions not only by increasing the **concentration** of particles—the number of particles in the space—but also by decreasing the volume of the space, which would correspondingly increase the pressure. If there were 100 dancers in that club, and the manager abruptly moved the party to a room half the size, the concentration of the dancers would double in the new space, and the likelihood of collisions would increase accordingly.

Enzymes and Other Catalysts

For two chemicals in nature to react with each other they first have to come into contact, and this occurs through random collisions. Because heat helps increase the kinetic energy of atoms, ions, and molecules, it promotes their collision. But in the body, extremely high heat—such as a very high fever—can damage body cells and be life-threatening. On the other hand, normal body temperature is not high enough to promote the chemical reactions that sustain life. That is where catalysts come in.

In chemistry, a **catalyst** is a substance that increases the rate of a chemical reaction without itself undergoing any change. You can think of a catalyst as a chemical change agent. They help increase the rate and force at which atoms, ions, and molecules collide, thereby increasing the probability that their valence shell electrons will interact.

The most important catalysts in the human body are enzymes. An **enzyme** is a catalyst composed of protein or ribonucleic acid (RNA), both of which will be discussed later in this chapter. Like all catalysts, enzymes work by lowering the level of energy that needs to be invested in a chemical reaction. A chemical reaction's **activation energy** is the "threshold" level of energy needed to break the bonds in the reactants. Once those bonds are broken, new arrangements can form. Without an enzyme to act as a catalyst, a much larger investment of energy is needed to ignite a chemical reaction (Figure 2.13).

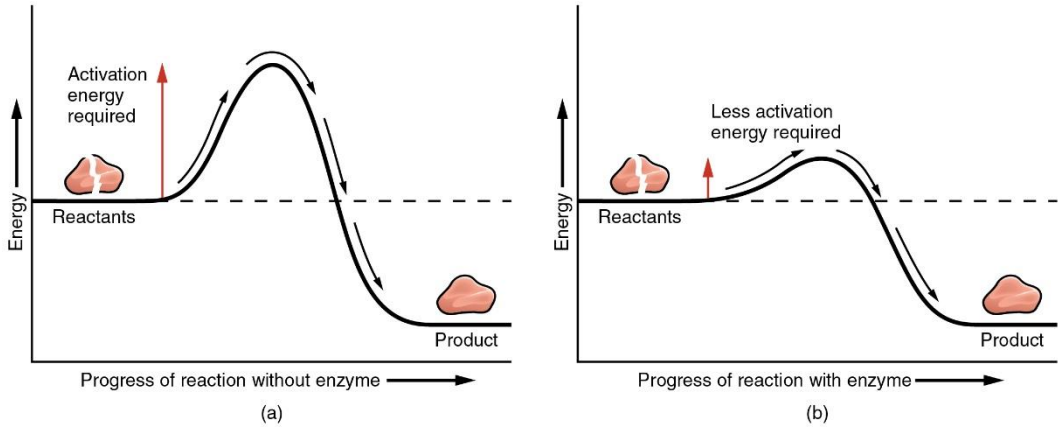


Figure 2.13 Enzymes Enzymes decrease the activation energy required for a given chemical reaction to occur. (a) Without an enzyme, the energy input needed for a reaction to begin is high. (b) With the help of an enzyme, less energy is needed for a reaction to begin.

Enzymes are critical to the body's healthy functioning. They assist, for example, with the breakdown of food and its conversion to energy. In fact, most of the chemical reactions in the body are facilitated by enzymes.

Chapter Attributions

"Anatomy and Physiology" by J. Gordon Betts, Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, Peter DeSaix,

7. Properties of Water

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

- 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? Why is water so important? 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. Water comprises approximately 60–70 percent of the human body. 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 are 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 generating pH. Understanding these characteristics of water helps to elucidate its importance in maintaining life.

Water's Polarity

One of water's important properties is that it is composed of polar molecules: the hydrogen and oxygen within water molecules (H_2O) form polar covalent bonds. While there is no net charge to a water molecule, water's polarity creates a slightly positive charge on hydrogen and a slightly negative charge on oxygen, contributing to water's properties of attraction. Water generates charges because oxygen is more electronegative than hydrogen, making it more

likely that a shared electron would be near the oxygen nucleus than the hydrogen nucleus, thus generating the partial negative charge near the **oxygen**.

As a result of water's polarity, each water molecule attracts other water molecules because of the opposite charges between water molecules, forming hydrogen bonds. Water also attracts or is attracted to other polar molecules and ions. We call a polar substance that interacts readily with or dissolves in water **hydrophilic** (hydro- = "water"; -philic = "loving"). In contrast, nonpolar molecules such as oils and fats do not interact well with water, as Figure 2.13 shows. A good example of this is vinegar and oil salad dressing (an acidic water solution). We call such nonpolar compounds **hydrophobic** (hydro- = "water"; -phobic = "fearing").

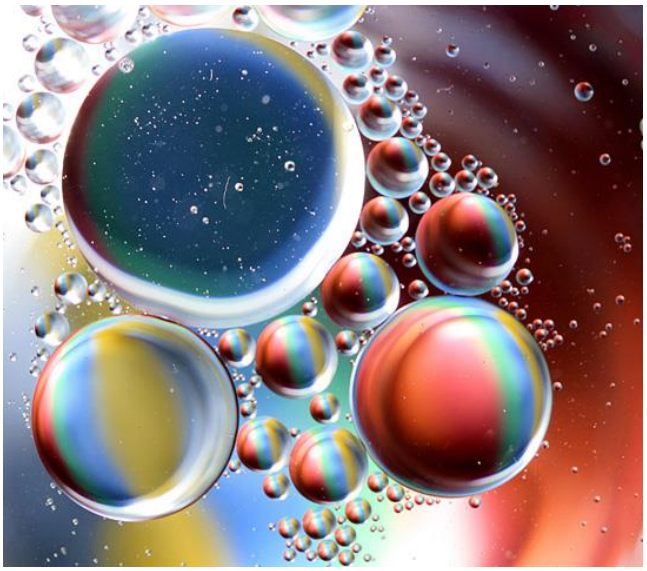


Figure 2.13 Oil and water do not mix. As this macro image of oil and water shows, oil does not dissolve in water but forms droplets instead. This is because it is a nonpolar compound. (credit: Gautam Dogra).

Water's States: Gas, Liquid, and Solid

The formation of hydrogen bonds is an important quality of the liquid water that is crucial to life as we know it. As water molecules make hydrogen bonds with each other, water takes on some unique chemical characteristics compared to other liquids and, 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 water molecules' motion (kinetic energy) causes the bonds to break due to the heat contained in the system. When the heat rises as water boils, the water molecules' higher kinetic energy causes the hydrogen bonds to break completely and allows water molecules to escape into the air as gas (steam or water vapor). Alternatively, when water temperature reduces and water freezes, the water molecules form a crystalline structure maintained by hydrogen bonding (there is not enough energy to break the hydrogen bonds) that makes ice less dense than liquid water, a phenomenon that we do not see when other liquids solidify.

Water's lower density in its solid form is due to the way hydrogen bonds orient as they freeze: the water molecules push farther apart compared to liquid water. With most other liquids, solidification when the temperature drops includes lowering kinetic energy between molecules, allowing them to pack even more tightly than in liquid form and giving the solid a greater density than the liquid.

The lower density of ice, as Figure 2.14 depicts, an anomaly causes it to float at the surface of liquid water, such as in an iceberg or ice cubes in a glass of water. In lakes and ponds, ice will form on the water's surface creating an insulating barrier that protects the animals and plant life in the pond from freezing. Without this insulating ice layer, plants and animals living in the pond would freeze in the solid block of ice and could not survive. The expansion of ice relative to liquid water causes the detrimental effect of freezing on living organisms. The ice crystals that form upon freezing rupture the delicate membranes essential for living cells to function, irreversibly damaging them. Cells can only survive freezing if another liquid like glycerol temporarily replaces the water in them.

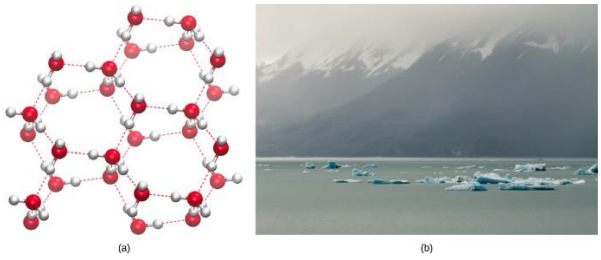


Figure 2.14 Hydrogen bonding makes ice less dense than liquid water. The (a) lattice structure of ice makes it less dense than the liquid water's freely flowing molecules, enabling it to (b) float on water. (credit a: modification of work by Jane Whitney, image created using Visual Molecular Dynamics (VMD) software[1]; credit b: modification of work by Carlos Ponte)

Water's High Heat Capacity

Water's high heat capacity is a property that hydrogen bonding among water molecules causes. Water has the highest **specific heat capacity** of any liquids. We define specific heat as the amount of heat one gram of a substance must absorb or lose to change its temperature by one degree Celsius. For water, this amount is one **calorie**. It therefore takes water a long time to heat and a long time to cool. In fact, water's specific heat capacity 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, warm blooded animals use water to more evenly disperse heat in their bodies: it acts in a similar manner to a car's cooling system, transporting heat from warm places to cool places, causing the body to maintain a more 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 (586 cal) is required to accomplish this change in water. This process occurs on the water's surface. As liquid water heats up, hydrogen bonding makes it difficult to separate the liquid water molecules from each other, which is required for it to enter its gaseous phase (steam). As a result, water acts as a heat sink or heat reservoir and requires much more heat to boil than does a liquid such as ethanol (grain alcohol), whose hydrogen bonding with other ethanol molecules is weaker than water's hydrogen bonding. Eventually, as water reaches its boiling point of 100° Celsius (212° Fahrenheit), 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, water's individual molecules acquire enough energy from other water molecules such that some surface water molecules can escape and vaporize: we call this process **evaporation**.

The fact that hydrogen bonds need to be broken for water to evaporate means that bonds use a substantial amount of energy in the process. As the water evaporates, energy is taken up by the process, cooling the environment where the evaporation is taking place. In many living organisms, including in humans, the evaporation of sweat, which is 90 percent water, allows the organism to cool so that it can maintain homeostasis of body temperature.

Water's Solvent Properties

Since water is a polar molecule with slightly positive and slightly negative charges, ions and polar molecules can readily dissolve in it. Therefore, we refer to water as a **solvent**, a substance capable of dissolving other polar molecules and ionic compounds. The charges associated with these molecules will form hydrogen bonds with water, surrounding the particle with water molecules. We refer to this as a **sphere of hydration**, or a hydration shell, as Figure 2.15 illustrates and serves to keep the particles separated or dispersed in the water.

When we add ionic compounds to water, the individual ions react with the water molecules' polar regions and their ionic bonds are disrupted in the process of **dissociation**. Dissociation occurs when atoms or groups of atoms break off from molecules and form ions. Consider table salt (NaCl, or sodium chloride): when we add NaCl crystals to water, the NaCl molecules dissociate into Na^+ and Cl^- ions, and spheres of hydration form around the ions, as Figure 2.15 illustrates. The partially negative charge of the water molecule's oxygen surrounds the positively charged sodium ion. The hydrogen's partially positive charge on the water molecule surrounds the negatively charged chloride ion.

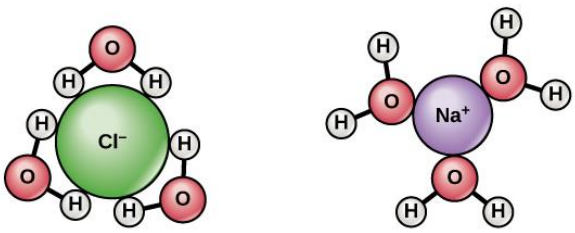


Figure 2.15 When we mix table salt (NaCl) in water, it forms spheres of hydration around the ions.

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 (because of hydrogen bonding), keeping the molecules together at the liquid-gas (water-air) interface, although there is no more room in the glass.

Cohesion allows for **surface tension**, the capacity of a substance to withstand rupturing when placed under tension or stress. This is also why water forms droplets when on a dry surface rather than flattening by gravity. When we place a small scrap of paper onto a water droplet, the paper floats on top even though paper is denser (heavier) than the water. Cohesion and surface tension keep the water molecules' hydrogen bonds intact and support the item floating on the top. It's even possible to "float" a needle on top of a glass of water if you place it gently without breaking the surface tension, as Figure 2.16 shows.

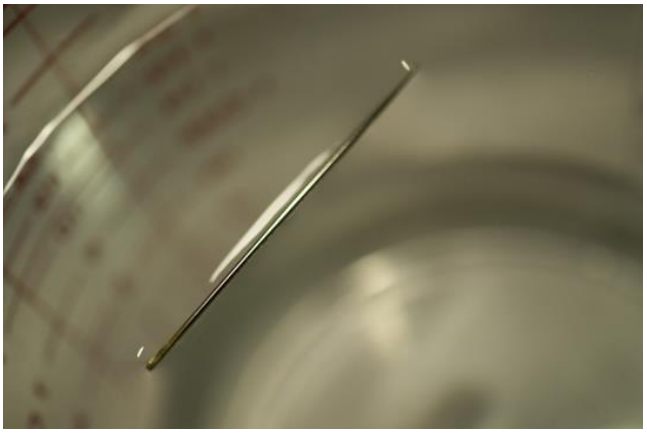


Figure 2.16 A needle's weight pulls the surface downward. At the same time, the surface tension pulls it up, suspending it on the water's surface preventing it from sinking. Notice the indentation in the water around the needle. (credit: Cory Zanker)

These cohesive forces are related to water's property of **adhesion**, or the attraction between water molecules and other molecules. This attraction is sometimes stronger than water's cohesive forces, especially when the water is exposed to charged surfaces such as those on the inside of thin glass tubes known as capillary tubes. We observe adhesion when water "climbs" up the tube placed in a glass of water: notice that the water appears to be higher on the tube's sides than in the middle. This is because the water molecules are attracted to the capillary's charged glass walls more than they are to each other and therefore adhere to it. We call this type of adhesion **capillary action**, as Figure 2.17 illustrates.

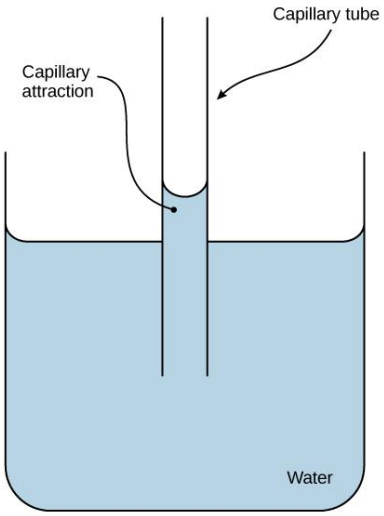


Figure 2.17 The adhesive forces exerted by the glass' internal surface exceeding the cohesive forces between the water molecules themselves causes capillary action in a glass tube. (credit: modification of work by Pearson-Scott Foresman, donated to the Wikimedia Foundation)

Why are cohesive and adhesive forces important for life? Cohesive and adhesive forces are important for transporting water from the roots to the leaves in plants. These forces create a “pull” on the water column. This pull results from the tendency of water molecules evaporating on the plant's surface to stay connected to water molecules below them, and so they are pulled along. Plants use this natural phenomenon to help transport water from their roots to their leaves. Without these properties of water, plants would be unable to receive the water and the dissolved minerals they require. In another example, insects such as the water strider, as Figure 2.18 shows, use the water's surface tension to stay afloat on the water's surface layer and even mate there.

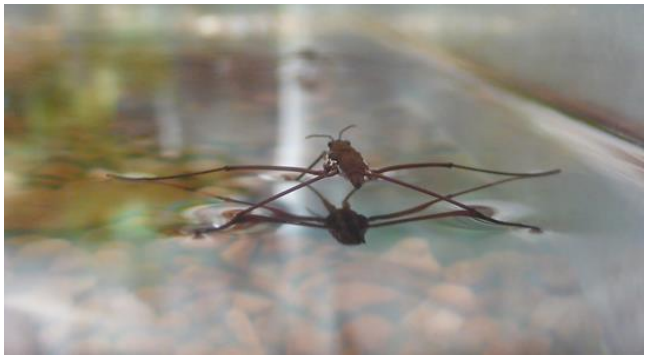


Figure 2.18 Water's cohesive and adhesive properties allow this water strider (*Gerris* sp.) to stay afloat. (credit: Tim Vickers)

pH, Buffers, Acids, and Bases

The pH of a solution indicates its acidity or alkalinity.

H2O(I)↔H+(aq)+OH−(aq)H2O(I)↔H+(aq)+OH-(aq)

You may have used **litmus** or pH paper, filter paper treated with a natural water-soluble dye for use as a pH indicator, tests how much acid (acidity) or base (alkalinity) exists in a solution. You might have even used some to test whether the water in a swimming pool is properly treated. In both cases, the pH test measures hydrogen ions' concentration in a given solution.

Hydrogen ions spontaneously generate in pure water by the dissociation (ionization) of a small percentage of water molecules into equal numbers of hydrogen (H^+) ions and hydroxide (OH^-) ions. While the hydroxide ions are kept in solution by their hydrogen bonding with other water molecules, the hydrogen ions, consisting of naked protons, immediately attract to un-ionized water molecules, forming hydronium ions (H_3O^+). Still, by convention, scientists refer to hydrogen ions and their concentration as if they were free in this state in liquid water.

The concentration of hydrogen ions dissociating from pure water is 1×10^{-7} moles H^+ ions per liter of water. Moles (mol) are a way to express the amount of a substance (which can be atoms, molecules, ions, etc.). One mole represents the atomic weight of a substance, expressed in grams, which equals the amount of the substance containing as many units as there are atoms in 12 grams of ^{12}C. Mathematically, one mole is equal to 6.02×10^{23} particles of the substance. Therefore, 1 mole of water is equal to 6.02×10^{23} water molecules. We calculate the pH as the negative of the base 10 logarithm of this concentration. The log10 of 1×10^{-7} is -7.0, and the negative of this number (indicated by the "p" of "pH") yields a pH of 7.0, which is also a neutral pH. The pH inside of human cells and blood are examples of two body areas where near-neutral pH is maintained.

Non-neutral pH readings result from dissolving acids or bases in water. Using the negative logarithm to generate positive integers, high concentrations of hydrogen ions yield a low pH number; whereas, low levels of hydrogen ions result in a high pH. An **acid** is a substance that

increases hydrogen ions' (H^+) concentration in a solution, usually by having one of its hydrogen atoms dissociate. A **base** provides either hydroxide ions (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 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 we place them in water, thereby raising the pH. An example of a weak basic solution is seawater, which has a pH near 8.0 This is close enough to a neutral pH that marine organisms have adapted in order to live and thrive in a saline environment.

The **pH scale** is, as we previously mentioned, an inverse logarithm and ranges from 0 to 14 (Figure 2.19). Anything below 7.0 (ranging from 0.0 to 6.9) is acidic, and anything above 7.0 (from 7.1 to 14.0) is alkaline. Extremes in pH in either direction from 7.0 are usually inhospitable to life. The pH inside cells (6.8) and the pH in the blood (7.4) are both very close to neutral. However, the environment in the stomach is highly acidic, with a pH of 1 to 2. As a result, how do stomach cells survive in such an acidic environment? How do they homeostatically maintain the near neutral pH inside them? The answer is that they cannot do it and are constantly dying. The stomach constantly produces new cells to replace dead ones, which stomach acids digest. Scientists estimate that the human body completely replaces the stomach lining every seven to ten days.

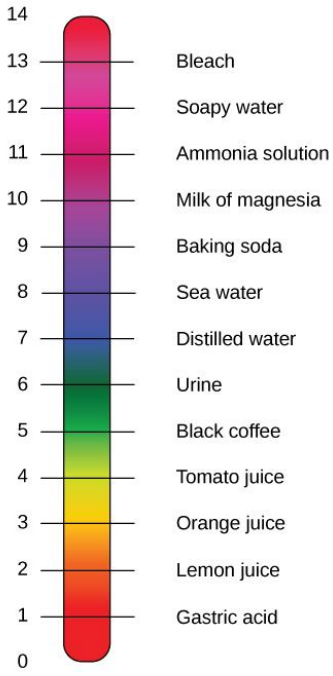


Figure 2.19 The pH scale measures hydrogen ions' (H^+) concentration in a solution. (credit: modification of work by Edward Stevens)

LINK TO LEARNING

Watch this video (https://www.youtube.com/embed/u837KYKyr9c) for a straightforward explanation of pH and its logarithmic scale.

How can organisms whose bodies 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 body's pH carefully maintained in the narrow range required for survival. Maintaining a constant blood pH is critical to a person's well-being. The buffer maintaining the pH of human blood involves carbonic acid (H_2CO_3), bicarbonate ion (HCO_3^-), and carbon dioxide (CO_2). When bicarbonate ions combine with free hydrogen ions and become carbonic acid, it removes hydrogen ions and moderates pH changes. Similarly, as Figure 2.20 shows, excess carbonic acid can convert to carbon dioxide gas which we exhale through the lungs. This prevents too many free hydrogen ions from building up in the blood and dangerously reducing the blood's pH. Likewise, if too much OH^- enters 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.

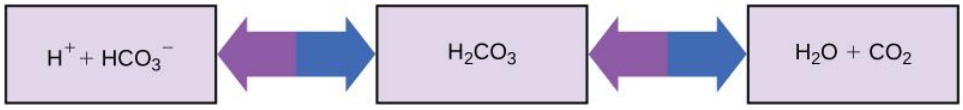


Figure 2.20 This diagram shows the body's buffering of blood pH levels. The blue arrows show the process of raising pH as more CO_2 is made. The purple arrows indicate the reverse process: the lowering of pH as more bicarbonate is created.

Other examples of buffers are antacids that some people use 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 who suffer "heartburn" after eating. Water's unique properties that contribute to this capacity to balance pH—as well as water's other characteristics—are essential to sustaining life on Earth.

LINK TO LEARNING

To learn more about water, visit the U.S. Geological Survey Water Science for Schools All About Water! Website: https://www.usgs.gov/special-topic/water-science-school

Chapter Attributions

8. Solutions, Solutes, pH and Buffer

Review previous chapter (Chapter 7. Property of Water, pages 54-65).

9. Biological Molecules

Carbon

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

- Explain why carbon is important for life

- Describe the role of functional groups in biological molecules

Many complex molecules called macromolecules, such as proteins, nucleic acids (RNA and DNA), carbohydrates, and lipids comprise cells. The macromolecules are a subset of **organic molecules** (any carbon-containing liquid, solid, or gas) that are especially important for life. The fundamental component for all of these macromolecules is carbon. The carbon atom has unique properties that allow it to form covalent bonds to as many as four different atoms, making this versatile element ideal to serve as the basic structural component, or "backbone," of the macromolecules.

Individual carbon atoms have an incomplete outermost electron shell. With an atomic number of 6 (six electrons and six protons), the first two electrons fill the inner shell, leaving four in the second shell. Therefore, carbon atoms can form up to four covalent bonds with other atoms to satisfy the octet rule. The methane molecule provides an example: it has the chemical formula CH_4. Each of its four hydrogen atoms forms a single covalent bond with the carbon atom by sharing a pair of electrons. This results in a filled outermost shell.

Hydrocarbons

Hydrocarbons are organic molecules consisting entirely of carbon and hydrogen, such as methane (CH_4) described above. We often use hydrocarbons in our daily lives as fuels—like the propane in a gas grill or the butane in a lighter. The many covalent bonds between the atoms in hydrocarbons store a great amount of energy, which releases when these molecules burn (oxidize). Methane, an excellent fuel, is the simplest hydrocarbon molecule, with a central carbon atom bonded to four different hydrogen atoms, as Figure 2.21 illustrates. The shape of its electron orbitals determines the shape of the methane molecule's geometry, where the atoms reside in three dimensions. The carbons and the four hydrogen atoms form a tetrahedron, with four triangular faces. For this reason, we describe methane as having tetrahedral geometry.

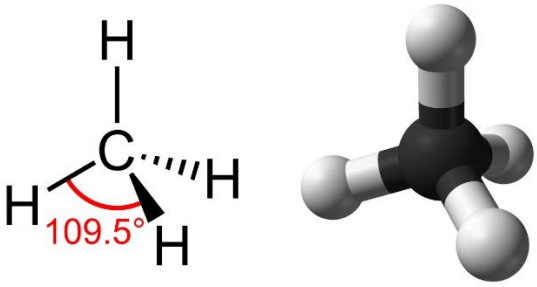


Figure 2.21 Methane has a tetrahedral geometry, with each of the four hydrogen atoms spaced 109.5° apart.

As the backbone of the large molecules of living things, 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, and each type of bond affects the molecule's geometry in a specific way. This three-dimensional shape or conformation of the large molecules of life (macromolecules) is critical to how they function.

Hydrocarbon Chains

Successive bonds between carbon atoms form hydrocarbon chains. These may be branched or unbranched. Furthermore, a molecule's different geometries of single, double, and triple covalent bonds alter the overall molecule's geometry as Figure 2.22 illustrates. The hydrocarbons ethane, ethene, and ethyne serve as examples of how different carbon-to-carbon bonds affect the molecule's geometry. The names of all three molecules start with the prefix "eth-," which is the prefix for two carbon hydrocarbons. The suffixes "-ane," "-ene," and "-yne" refer to the presence of single, double, or triple carbon-carbon bonds, respectively. Thus, propane, propene, and propyne follow the same pattern with three carbon molecules, butane, butene, and butyne for four carbon molecules, and so on. Double and triple bonds change the molecule's geometry: single bonds allow rotation along the bond's axis; whereas, double bonds lead to a planar configuration and triple bonds to a linear one. These geometries have a significant impact on the shape a particular molecule can assume.

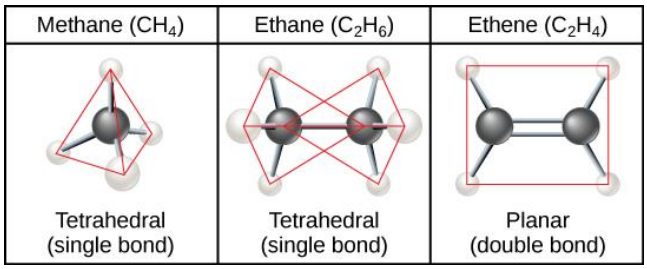


Figure 2.22 When carbon forms single bonds with other atoms, the shape is tetrahedral. When two carbon atoms form a double bond, the shape is planar, or flat. Single bonds, like those in ethane, are able to rotate. Double bonds, like those in ethene, cannot rotate, so the atoms on either side are locked in place.

Hydrocarbon Rings

So far, the hydrocarbons we have discussed have been **aliphatic hydrocarbons**, which consist of linear chains of carbon atoms. Another type of hydrocarbon, **aromatic hydrocarbons**, consists of closed rings of carbon atoms. We find ring structures in hydrocarbons, sometimes with the presence of double bonds, which we can see by comparing cyclohexane's structure to benzene in Figure 2.23. Examples of biological molecules that incorporate the benzene ring include some amino acids and cholesterol and its derivatives, including the hormones estrogen and testosterone. We also find the benzene ring in the herbicide 2,4-D. Benzene is a natural component of crude oil and has been classified as a carcinogen. Some hydrocarbons have both aliphatic and aromatic portions. Beta-carotene is an example of such a hydrocarbon.

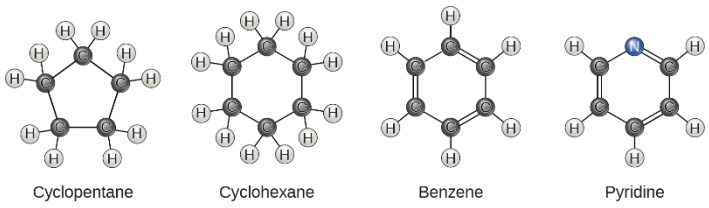


Figure 2.23 Carbon can form five- and six-membered rings. Single or double bonds may connect the carbons in the ring, and nitrogen may be substituted for carbon.

Isomers

The three-dimensional placement of atoms and chemical bonds within organic molecules is central to understanding their chemistry. We call molecules that share the same chemical

formula but differ in the placement (structure) of their atoms and/or chemical bonds **isomers**. **Structural isomers** (like butane and isobutene in Figure 2.24**a**) differ in the placement of their covalent bonds: both molecules have four carbons and ten hydrogens (C_4H_{10}), but the different atom arrangement within the molecules leads to differences in their chemical properties. For example, butane is suited for use as a fuel for cigarette lighters and torches; whereas, isobutene is suited for use as a refrigerant and a propellant in spray cans.

Geometric isomers, alternatively have similar placements of their covalent bonds but differ in how these bonds are made to the surrounding atoms, especially in carbon-to-carbon double bonds. In the simple molecule butene (C_4H_8), the two methyl groups (CH_3) can be on either side of the double covalent bond central to the molecule, as Figure 2.24**b** illustrates. When the carbons are bound on the same side of the double bond, this is the *cis* configuration. If they are on opposite sides of the double bond, it is a *trans* configuration. In the *trans* configuration, the carbons form a more or less linear structure; whereas, the carbons in the *cis* configuration make a bend (change in direction) of the carbon backbone.

VISUAL CONNECTION

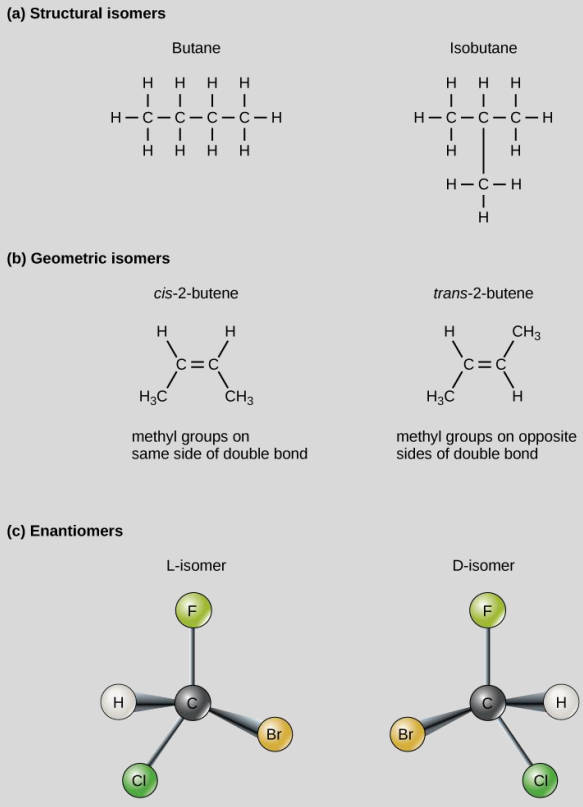


Figure 2.24 We call molecules that have the same number and type of atoms arranged differently isomers. (a) Structural isomers have a different covalent arrangement of atoms. (b) Geometric isomers have a different arrangement of atoms around a double bond. (c) Enantiomers are mirror images of each other.

Which of the following statements is false?

a. Molecules with the formulas CH_3CH_2COOH and $C_3H_6O_2$ could be structural isomers.

b. Molecules must have a double bond to be *cis-trans* isomers.

c. To be enantiomers, a molecule must have at least three different atoms or groups connected to a central carbon.

d. To be enantiomers, a molecule must have at least four different atoms or groups connected to a central carbon.

In triglycerides (fats and oils), long carbon chains known as fatty acids may contain double bonds, which can be in either the *cis* or *trans* configuration, as Figure 2.25 illustrates. Fats with at least one double bond between carbon atoms are unsaturated fats. When some of these bonds are in the *cis* configuration, the resulting bend in the chain's carbon backbone means that triglyceride molecules cannot pack tightly, so they remain liquid (oil) at room temperature. Alternatively, triglycerides with *trans* double bonds (popularly called trans fats), have relatively linear fatty acids that are able to pack tightly together at room temperature and form solid fats. In the human diet, trans fats are linked to an increased risk of cardiovascular disease, so many food manufacturers have reduced or eliminated their use in recent years. In contrast to unsaturated fats, we call triglycerides without double bonds between carbon atoms saturated fats, meaning that they contain all the hydrogen atoms available. Saturated fats are a solid at room temperature and usually of animal origin.

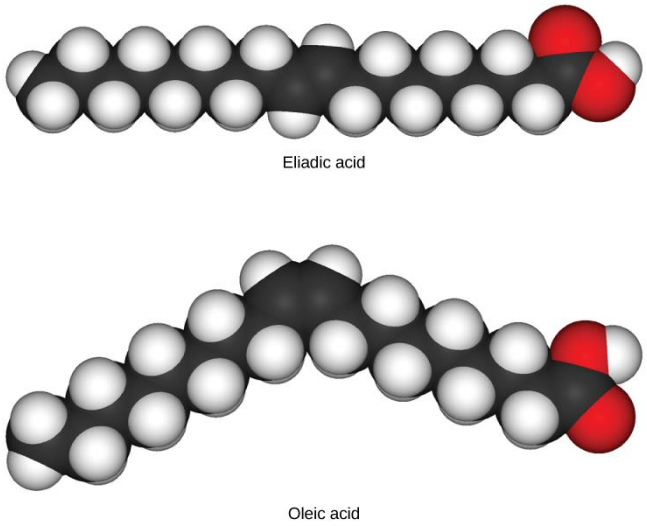


Figure 2.25 These space-filling models show a *cis* (oleic acid) and a *trans* (eliadic acid) fatty acid. Notice the bend in the molecule caused by the *cis* configuration.

Enantiomers

Enantiomers are molecules that share the same chemical structure and chemical bonds but differ in the three-dimensional placement of atoms so that they are non-superimposable mirror images. Figure 2.26 shows an amino acid alanine example, where the two structures are nonsuperimposable. In nature, the L-forms of amino acids are predominant in proteins. Some D forms of amino acids are seen in the cell walls of bacteria and polypeptides in other organisms. Similarly, the D-form of glucose is the main product of photosynthesis and we rarely see the molecule's L-form in nature.

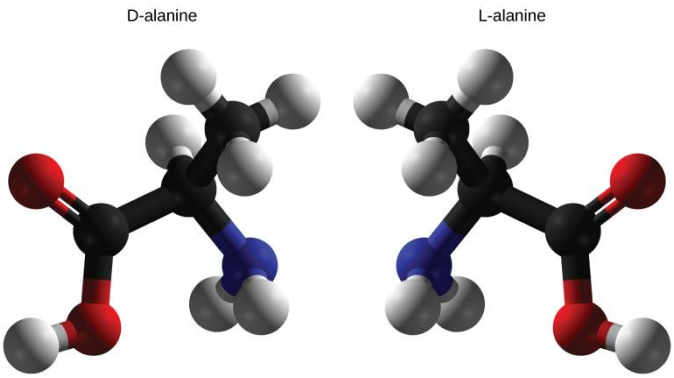


Figure 2.26 D-alanine and L-alanine are examples of enantiomers or mirror images. L-forms of amino acids are predominant in proteins.

Functional Groups

Functional groups are groups of atoms that occur within molecules and confer specific chemical properties to those molecules. We find them along the "carbon backbone" of macromolecules. Chains and/or rings of carbon atoms with the occasional substitution of an element such as nitrogen or oxygen form this carbon backbone. Molecules with other elements in their carbon backbone are **substituted hydrocarbons**.

The functional groups in a macromolecule are usually attached to the carbon backbone at one or several different places along its chain and/or ring structure. 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.

A functional group can participate in specific chemical reactions. Figure 2.27 shows some of the important functional groups in biological molecules. They include: hydroxyl, methyl, carbonyl, carboxyl, amino, phosphate, and sulfhydryl. These groups play an important role in forming molecules like DNA, proteins, carbohydrates, and lipids. We usually classify functional groups as hydrophobic or hydrophilic depending on their charge or polarity characteristics. An example of a hydrophobic group is the nonpolar methyl molecule. Among the hydrophilic functional groups is the carboxyl group in amino acids, some amino acid side chains, and the fatty acids that form triglycerides and phospholipids. This carboxyl group ionizes to release hydrogen ions (H^+) from the COOH group resulting in the negatively charged COO^- group. This contributes to the hydrophilic nature of whatever molecule on which it is found. Other functional groups, such as the carbonyl group, have a partially negatively charged oxygen atom that may form hydrogen bonds with water molecules, again making the molecule more hydrophilic.

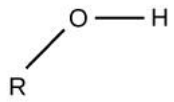
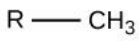
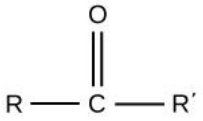
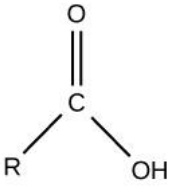
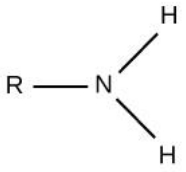
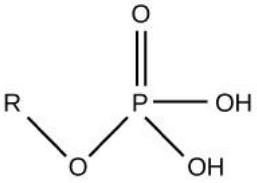
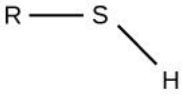
Functional Group	Structure	Properties
Hydroxyl		Polar
Methyl		Nonpolar
Carbonyl		Polar
Carboxyl		Charged, ionizes to release H^+. Since carboxyl groups can release H^+ ions into solution, they are considered acidic.
Amino		Charged, accepts H^+ to form NH_3^+. Since amino groups can remove H^+ from solution, 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 2.27 These functional groups are in many different biological molecules. R, also known as R-group, is an abbreviation for any group in which a carbon or hydrogen atom is attached to the rest of the molecule.

Hydrogen bonds between functional groups (within the same molecule or between different molecules) are important to the function of many macromolecules and help them to fold properly into and maintain the appropriate shape for functioning. Hydrogen bonds are also involved in various recognition processes, such as DNA complementary base pairing and the binding of an enzyme to its substrate, as Figure 2.28 illustrates.

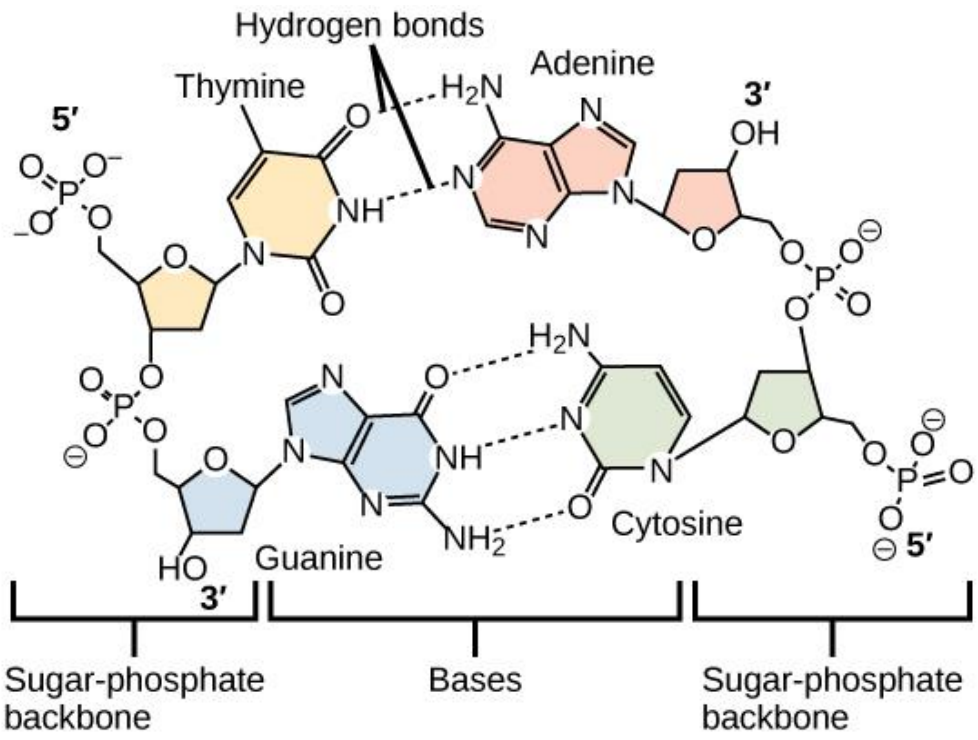


Figure 2.28 Hydrogen bonds connect two strands of DNA together to create the double-helix structure.

Chapter Attributions

"Biology 2e" Mary Ann Clark, Matthew Douglas, Jung Choi, OpenStax is licensed under CC BY 4.0. Access for free at https://openstax.org/books/biology-2e/pages/2-3-carbon

Synthesis of Biological Macromodules

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

- Understand macromolecule synthesis
- Explain dehydration (or condensation) and hydrolysis reactions

As you've learned, **biological macromolecules** are large molecules, necessary for life, that are built from smaller organic molecules. There are four major biological macromolecule classes (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.

Dehydration Synthesis

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 **dehydration synthesis**, which means "to put together while losing water."

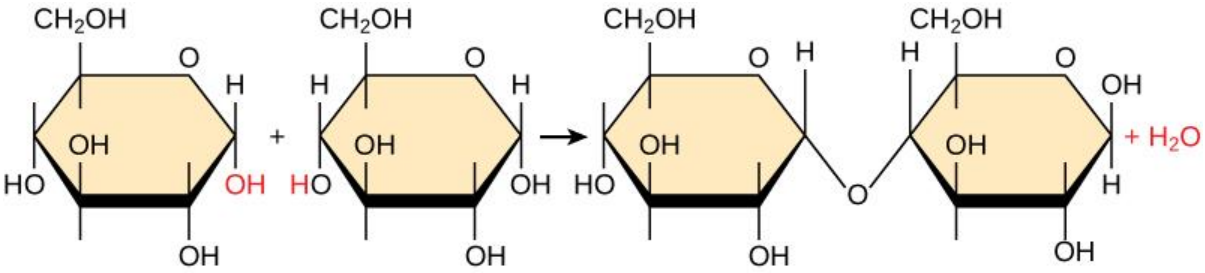


Figure 3.2 In the dehydration synthesis reaction above, two glucose molecules link to form the disaccharide maltose. In the process, it forms a water molecule.

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

Hydrolysis

Polymers break down into monomers during hydrolysis. A chemical reaction occurs when inserting a water molecule across the bond. Breaking a covalent bond with this water molecule in the compound achieves this (Figure 3.3). During these reactions, the polymer breaks into two components: one part gains a hydrogen atom (H+) and the other gains a hydroxyl molecule (OH–) from a split water molecule.

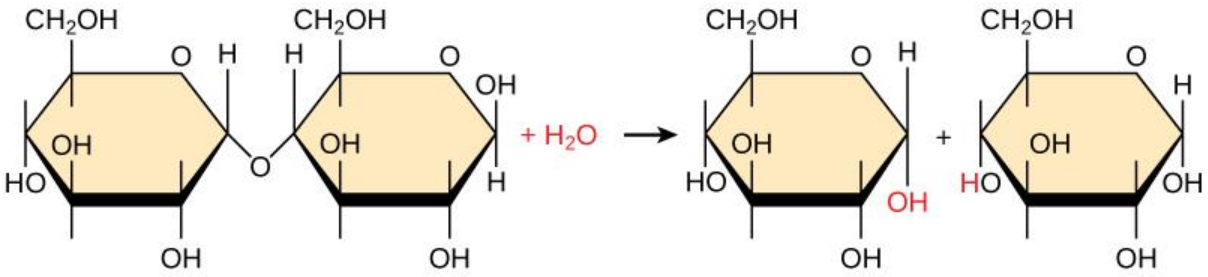


Figure 3.3 In the hydrolysis reaction here, the disaccharide maltose breaks down to form two glucose monomers by adding a water molecule. Note that this reaction is the reverse of the synthesis reaction in Figure 3.2.

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, catalytic enzymes in the digestive system hydrolyze or break down the food we ingest into smaller molecules. This allows cells in our body to easily absorb nutrients in the intestine. A specific enzyme breaks down each macromolecule. For instance, amylase, sucrase, lactase, or maltase break down carbohydrates. Enzymes called proteases, such as pepsin and peptidase, and hydrochloric acid break down proteins. Lipases break down lipids. These broken down macromolecules provide energy for cellular activities.

LINK TO LEARNING

Visit this site (https://www.youtube.com/watch?v=ZMTeqZLXBSo) to see visual representations of dehydration synthesis and hydrolysis.

Chapter Attributions

Carbohydrates

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

- Discuss the role of carbohydrates in cells and in the extracellular materials of animals and plants
- Explain carbohydrate classifications
- List common monosaccharides, disaccharides, and polysaccharides

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 carbohydrate sources that 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.

Molecular Structures

The stoichiometric formula $(CH_2O)_n$, where n is the number of carbons in the molecule represents **carbohydrates**. 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”). Scientists classify carbohydrates into three subtypes: monosaccharides, disaccharides, and polysaccharides.

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 an aldose, and if it has a ketone group (the functional group with the structure RC(=O)R'), it is a ketose. Depending on the number of carbons in the sugar, they can be trioses (three carbons), pentoses (five carbons), and/or hexoses (six carbons). Figure 3.4 illustrates monosaccharides.

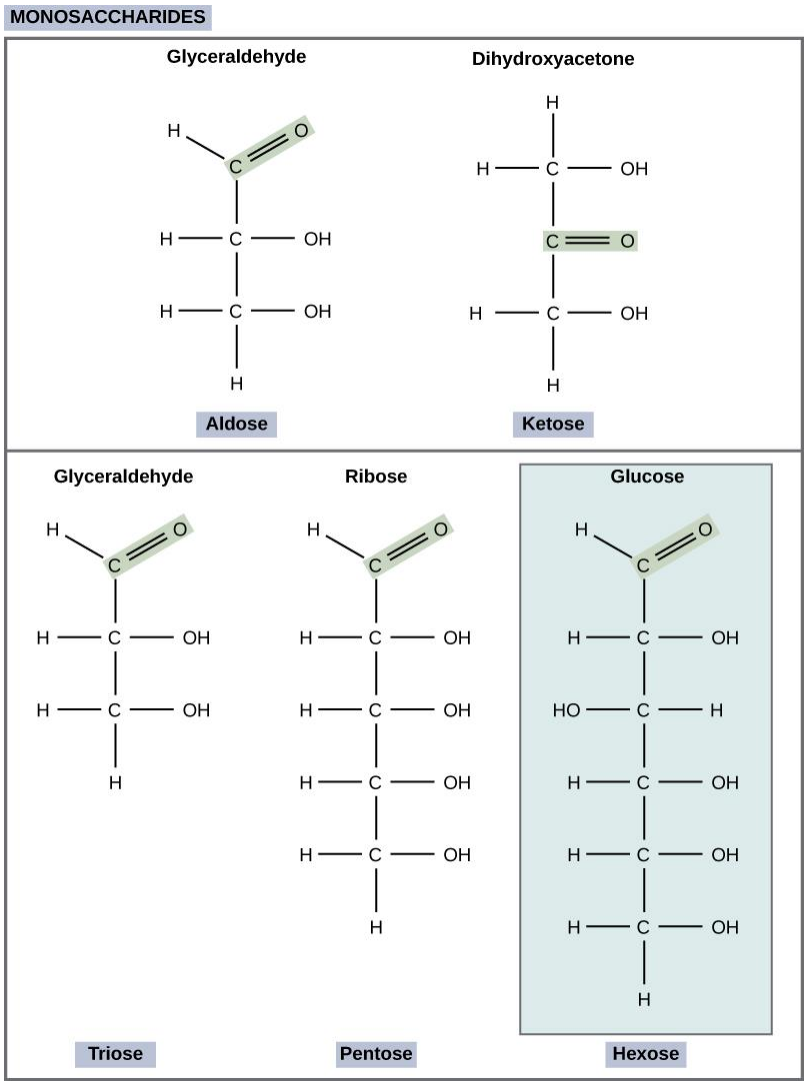


Figure 3.4 Scientists classify monosaccharides based on the position of their carbonyl group and the number of carbons in the backbone. Aldoses have a carbonyl group (indicated in green) at the end of the carbon chain, and ketoses have a carbonyl group in the middle of the carbon chain. Trioses, pentoses, 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 releases from glucose, and that energy helps make adenosine triphosphate (ATP). Plants synthesize glucose using carbon dioxide and water, and glucose in turn provides energy requirements for the plant. Humans and other animals that

feed on plants often store excess glucose as catabolized (cell breakdown of larger molecules) starch.

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 isomers) because of the different arrangement of functional groups around the asymmetric carbon. All these monosaccharides have more than one asymmetric carbon (Figure 3.5).

VISUAL CONNECTION

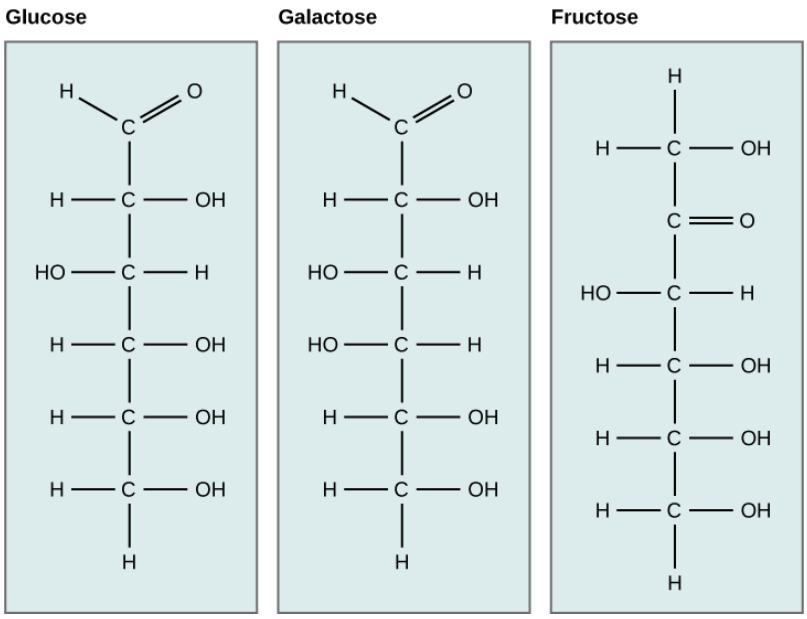


Figure 3.5 Glucose, galactose, and fructose are all hexoses. They are structural isomers, meaning they have the same chemical formula ($C_6H_{12}O_6$) but a different atom arrangement.

What kind of sugars are these, aldose or ketose?

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 in ring forms (Figure 3.6). Glucose in a ring form can have two different hydroxyl group arrangements (OH) around the anomeric carbon (carbon 1 that becomes

asymmetric in the ring formation process). If the hydroxyl group is below carbon number 1 in the sugar, it is in the alpha (*α*) position, and if it is above the plane, it is in the beta (*β*) position.

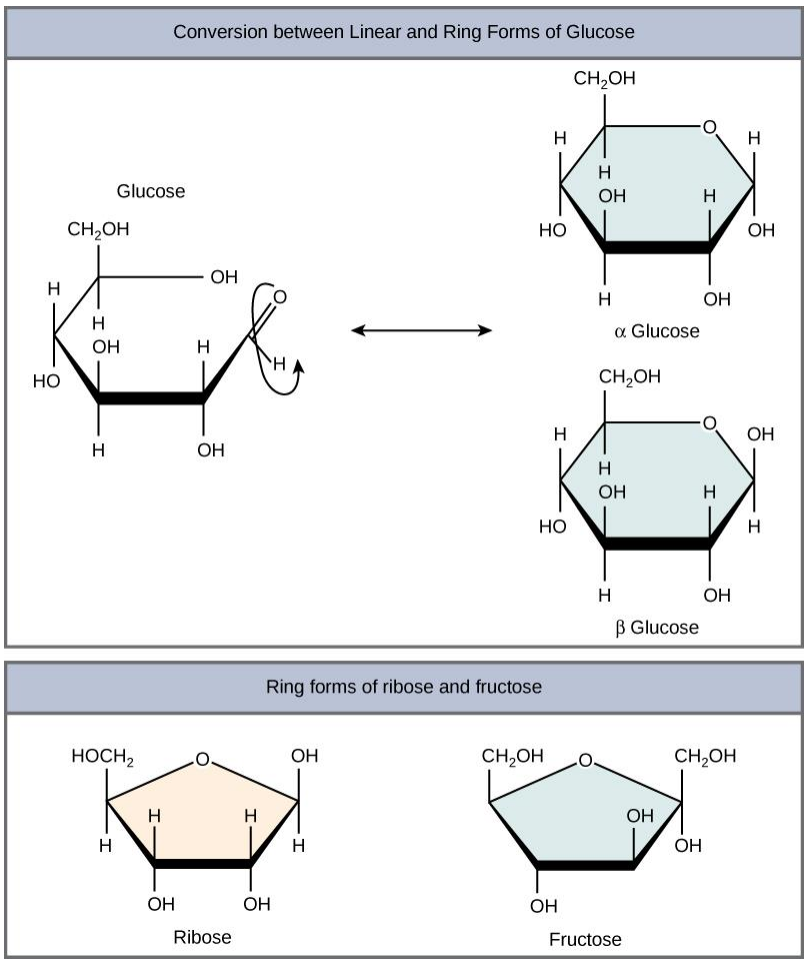


Figure 3.6 Five and six carbon monosaccharides exist in equilibrium between linear and ring forms. When the ring forms, the side chain it closes on locks into an *α* or *β* position. Fructose and ribose also form rings, although they form five-membered rings as opposed to the six-membered ring of glucose.

Disaccharides

Disaccharides (di- = “two”) form when two monosaccharides undergo a dehydration reaction (or a condensation reaction or dehydration synthesis). During this process, one monosaccharide's hydroxyl group combines with another monosaccharide's hydrogen, releasing a water molecule and forming a covalent bond. A covalent bond forms between a carbohydrate molecule and another molecule (in this case, between two monosaccharides).

Scientists call this a **glycosidic bond** (Figure 3.7). Glycosidic bonds (or glycosidic linkages) can be an alpha or beta type. An alpha bond is formed when the OH group on the carbon-1 of the first glucose is below the ring plane, and a beta bond is formed when the OH group on the carbon-1 is above the ring plane.

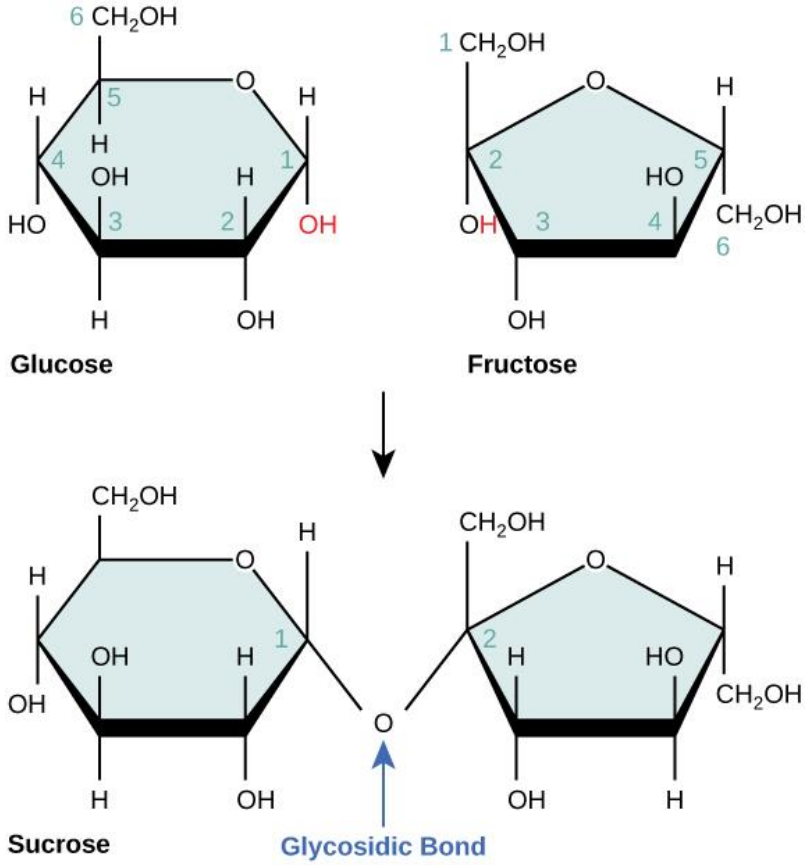


Figure 3.7 Sucrose forms when a glucose monomer and a fructose monomer join in a dehydration reaction to form a glycosidic bond. In the process, a water molecule is lost. By convention, the carbon atoms in a monosaccharide are numbered from the terminal carbon closest to the carbonyl group. In sucrose, a glycosidic linkage forms between carbon 1 in glucose and carbon 2 in fructose.

Common disaccharides include lactose, maltose, and sucrose (Figure 3.8). Lactose is a disaccharide consisting of the monomers glucose and galactose. It is 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 comprised of glucose and fructose monomers.

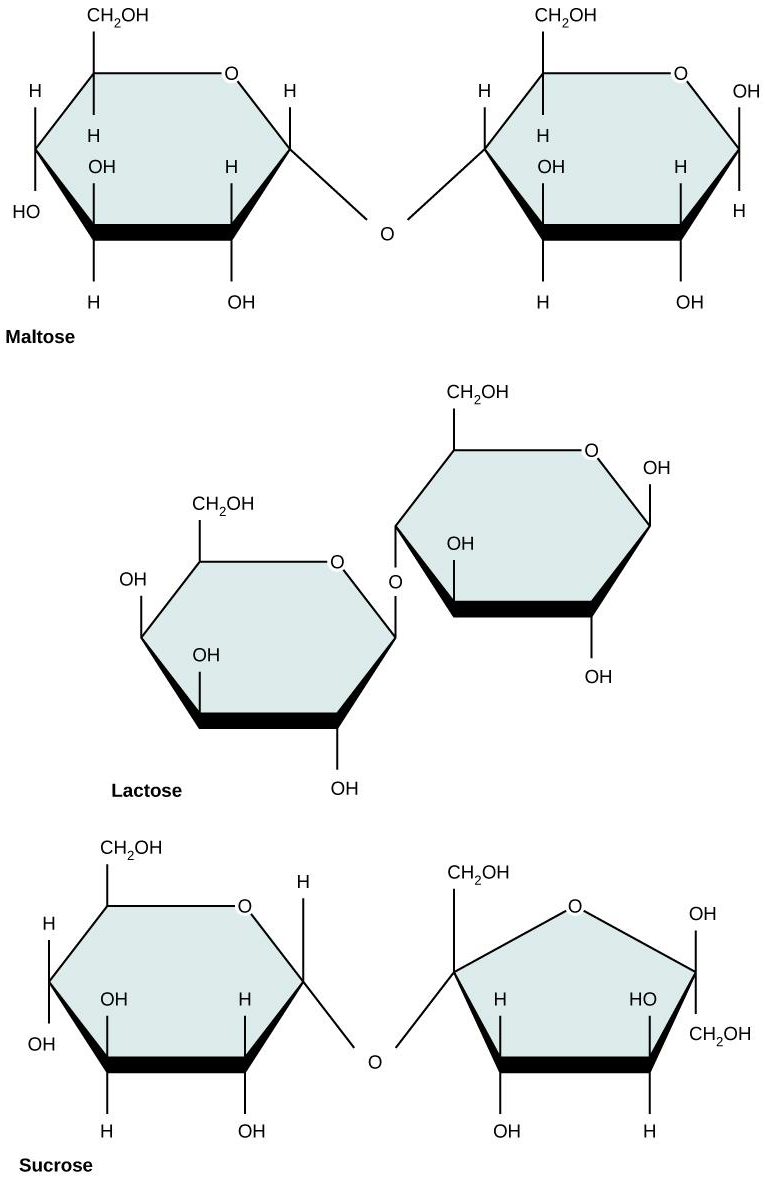


Figure 3.8 Common disaccharides include maltose (grain sugar), lactose (milk sugar), and sucrose (table sugar).

Polysaccharides

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

Plants store starch in the form of sugars. In plants, an amylose and amylopectic mixture (both glucose polymers) comprise these sugars. Plants are able to synthesize glucose, and they store the excess glucose, beyond their immediate energy needs, 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 food source for humans and animals. Enzymes break down the starch that humans consume. For example, an amylase present in saliva catalyzes, or breaks down this starch into smaller molecules, such as maltose and glucose. The cells can then absorb the glucose.

Glucose starch comprises 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 Figure 3.9 illustrates, unbranched glucose monomer chains (only α 1-4 linkages) form the starch; whereas, amylopectin is a branched polysaccharide (α 1-6 linkages at the branch points).

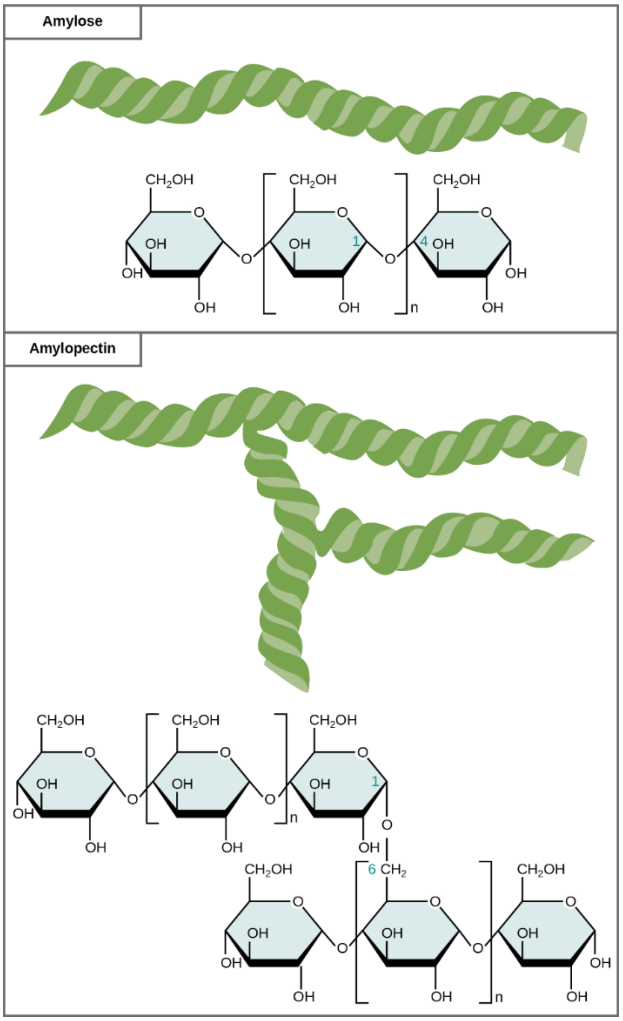


Figure 3.9 Amylose and amylopectin are two different starch forms. Unbranched glucose monomer chains comprise amylose by *α* 1-4 glycosidic linkages. Unbranched glucose monomer chains comprise amylopectin by *α* 1-4 and *α* 1-6 glycosidic linkages. Because of the way the subunits are joined, the glucose 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 comprised 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 breaks down to release glucose in a process scientists call glycogenolysis.

Cellulose is the most abundant natural biopolymer. Cellulose mostly comprises a plant's cell wall. This provides the cell structural support. Wood and paper are mostly cellulosic in nature. Glucose monomers comprise cellulose that *β* 1-4 glycosidic bonds link (Figure 3.10).

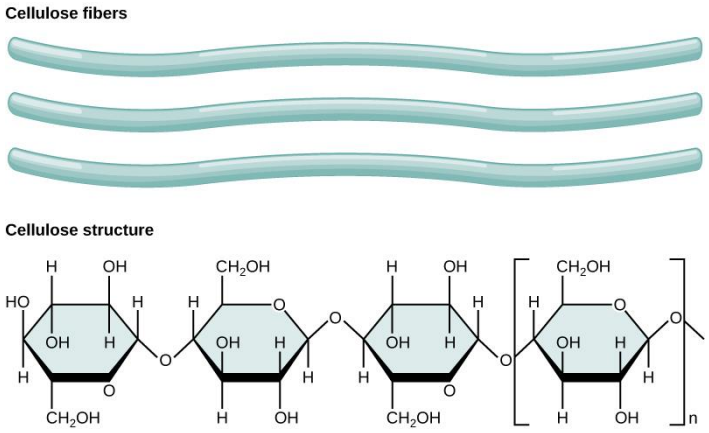


Figure 3.10 In cellulose, glucose monomers are linked in unbranched chains by *β* 1-4 glycosidic linkages. Because of the way the glucose subunits are joined, every glucose monomer is flipped relative to the next one resulting in a linear, fibrous structure.

As Figure 3.10 shows, 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 human digestive enzymes cannot break down the *β* 1-4 linkage, herbivores such as cows, koalas, and buffalos 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 some of these animals, certain species of bacteria and protists reside in the rumen (part of the herbivore's digestive system) and secrete the enzyme cellulase. The appendix of grazing animals also contains bacteria that digest cellulose, giving it an important role in ruminants' digestive systems. Cellulases can break down cellulose into glucose monomers that animals use as an energy source. Termites are also able to break down cellulose because of the presence of other organisms in their bodies that secrete cellulases.

Carbohydrates serve various functions in different animals. Arthropods (insects, crustaceans, and others) have an outer skeleton, the exoskeleton, which protects their internal body parts (as we see in the bee in Figure 3.11). This exoskeleton is made of the biological macromolecule **chitin**, which is a polysaccharide-containing nitrogen. It is made of repeating N-

acetyl-*β*-d-glucosamine units, which are a modified sugar. Chitin is also a major component of fungal cell walls. Fungi are neither animals nor plants and form a kingdom of their own in the domain Eukarya.

Figure 3.11 Insects have a hard outer exoskeleton made of chitin, a type of polysaccharide. (credit: Louise Docker)

CAREER CONNECTION

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 people increasingly seek out registered dietitians 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).

Benefits of Carbohydrates

Are carbohydrates good for you? Some often tell people who wish to lose weight that carbohydrates are bad and they should avoid them. 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.

As part of a well balanced diet, we should supplement carbohydrates with proteins, vitamins, and fats. Calorie-wise, a gram of carbohydrate provides 4.3 Kcal. For comparison, fats provide 9 Kcal/g, a less desirable ratio. Carbohydrates contain soluble and insoluble elements. The insoluble part, fiber, is mostly cellulose. Fiber has many uses. It promotes regular bowel movement by adding bulk, and it regulates the blood glucose consumption rate. Fiber also helps to remove excess cholesterol from the body. Fiber binds to the cholesterol in the small intestine, then attaches to the cholesterol and prevents the cholesterol particles from entering the bloodstream. Cholesterol then exits the body via the feces. Fiber-rich diets also have a protective role in reducing the occurrence of colon cancer. In addition, a meal containing whole grains and vegetables gives a feeling of fullness. As an immediate source of energy, glucose breaks down during the cellular respiration process, which produces ATP, the cell's energy currency. Without consuming carbohydrates, we reduce the availability of "instant energy". Eliminating carbohydrates from the diet is not the best way to lose weight. 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 more sensible way to lose weight.

LINK TO LEARNING

For an additional perspective on carbohydrates, explore "Biomolecules: the Carbohydrates": https://www.wisc-online.com/learn/general-education/anatomy-and-physiology2/ap21316/biomolecules-the-carbohydrates-video

Chapter Attributions

Lipids

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

- Describe the four major types of lipids
- Explain the role of fats in storing energy
- Differentiate between saturated and unsaturated fatty acids
- Describe phospholipids and their role in cells
- Define the basic structure of a steroid and some steroid functions
- Explain how cholesterol helps maintain the plasma membrane's fluid nature

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 (“water fearing”), 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 3.12). For example, they help keep aquatic birds and mammals dry when forming a protective layer over fur or feathers because of their water-repellant hydrophobic nature. 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 3.12 Hydrophobic lipids in aquatic mammals' fur, such as this river otter, protect them from the elements. (credit: Ken Bosma)

Fats and Oils

A fat molecule consists of two main components—glycerol and fatty acids. Glycerol is an organic compound (alcohol) with three carbons, five hydrogens, and three hydroxyl (OH) groups. Fatty acids have a long chain of hydrocarbons to which a carboxyl group is attached, hence the name “fatty acid.” The number of carbons in the fatty acid may range from 4 to 36. The most common are those containing 12–18 carbons. In a fat molecule, the fatty acids attach to each of the glycerol molecule's three carbons with an ester bond through an oxygen atom (Figure 3.13).

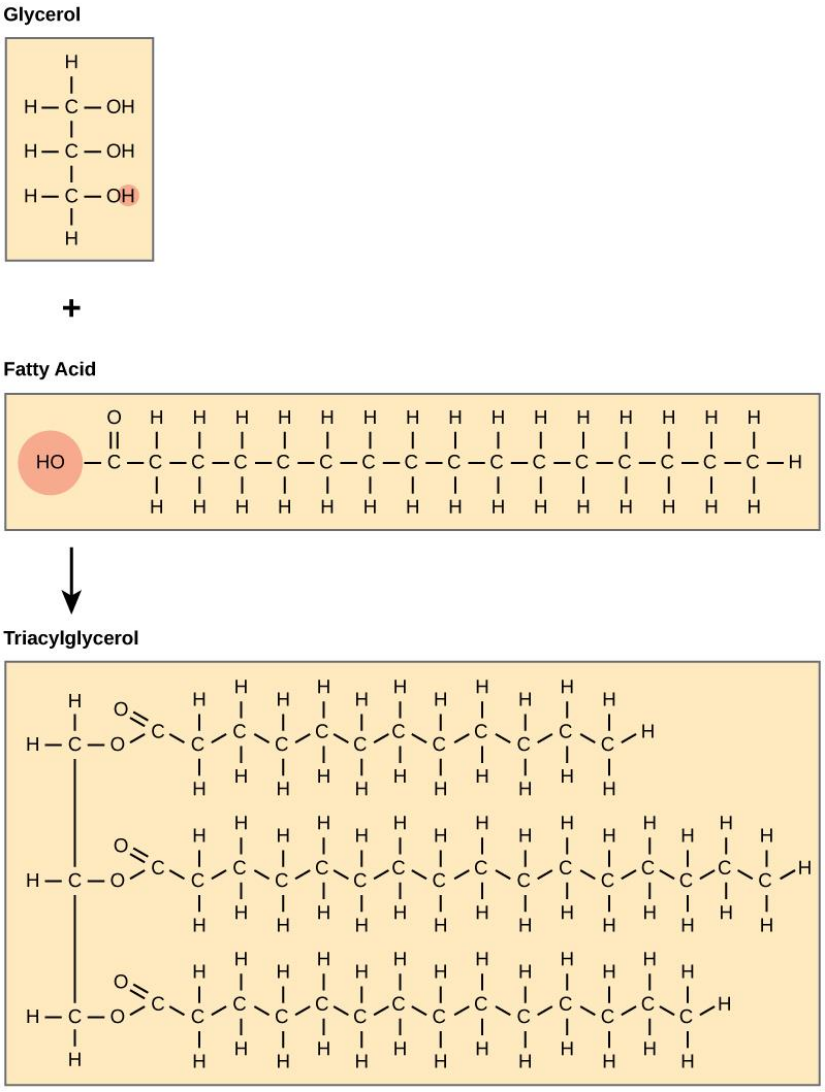


Figure 3.13 Joining three fatty acids to a glycerol backbone in a dehydration reaction forms triacylglycerol. Three water molecules release in the process.

During this ester bond formation, three water molecules are released. The three fatty acids in the triacylglycerol may be similar or dissimilar. We also call fats **triacylglycerols** or **triglycerides** because of their chemical structure. Some fatty acids have common names that specify their origin. For example, palmitic acid, a **saturated fatty acid**, is derived from the palm tree. Arachidic acid is derived from *Arachis hypogea,* the scientific name for groundnuts or peanuts.

Fatty acids may be saturated or unsaturated. In a fatty acid chain, if there are only single bonds between neighboring carbons in the hydrocarbon chain, the fatty acid is 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 3.14).

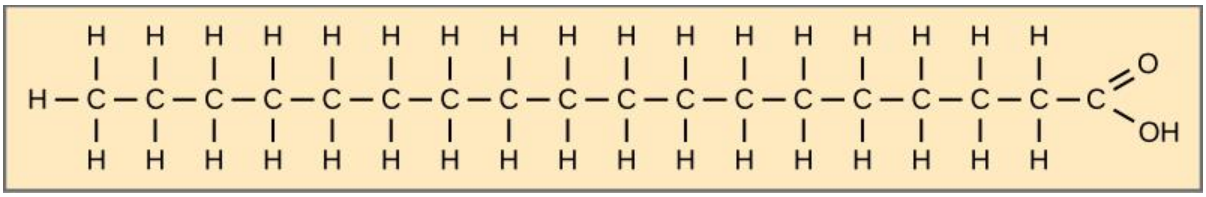


Figure 3.14 Stearic acid is a common saturated fatty acid.

When the hydrocarbon chain contains a double bond, the fatty acid is **unsaturated**. Oleic acid is an example of an unsaturated fatty acid (Figure 3.15).

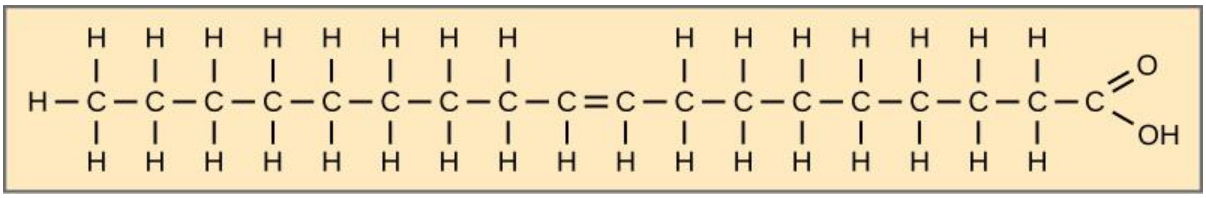


Figure 3.15 Oleic acid is a common unsaturated fatty acid.

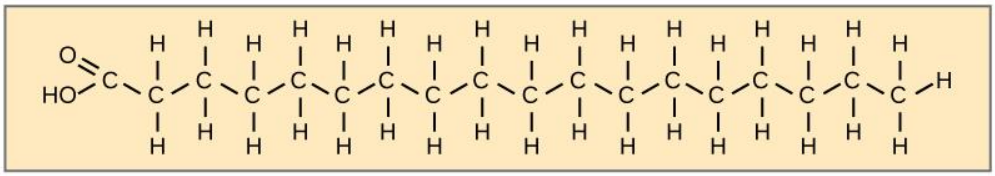
Most unsaturated fats are liquid at room temperature. We call these oils. If there is one double bond in the molecule, then it is a monounsaturated fat (e.g., olive oil), and if there is more than one double bond, then it is a polyunsaturated fat (e.g., canola oil).

When a fatty acid has no double bonds, it is a saturated fatty acid because it is not possible to add more hydrogen to the chain's carbon atoms. A fat may contain similar or different fatty

acids attached to glycerol. Long straight fatty acids with single bonds generally pack tightly and are solid at room temperature. Animal fats with stearic acid and palmitic acid (common in meat) and the fat with butyric acid (common in butter) are examples of saturated fats. Mammals store fats in specialized cells, or adipocytes, where fat globules occupy most of the cell's volume. Plants store fat or oil in many seeds and use them as a source of energy during seedling development. Unsaturated fats or oils are usually of plant origin and contain *cis* unsaturated fatty acids. *Cis* and *trans* indicate the configuration of the molecule around the double bond. If hydrogens are present in the same plane, it is a cis fat. If the hydrogen atoms are on two different planes, it is a **trans fat**. The *cis* double bond causes a bend or a "kink" that prevents the fatty acids from packing tightly, keeping them liquid at room temperature (Figure 3.16). Olive oil, corn oil, canola oil, and cod liver oil are examples of unsaturated fats. Unsaturated fats help to lower blood cholesterol levels; whereas, saturated fats contribute to plaque formation in the arteries.

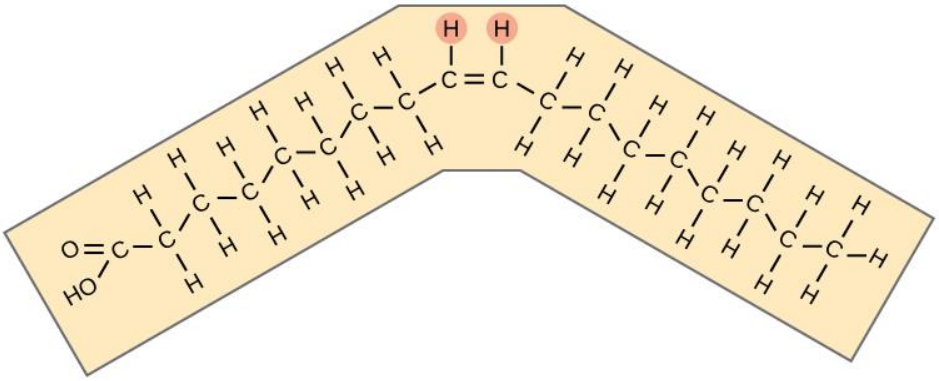
Saturated fatty acid

Stearic acid



Unsaturated fatty acids

Cis oleic acid



Trans oleic acid

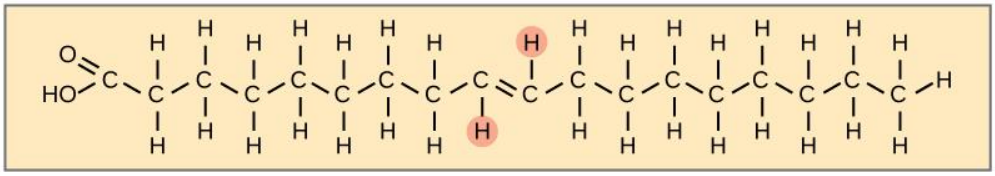


Figure 3.16 Saturated fatty acids have hydrocarbon chains connected by single bonds only. Unsaturated fatty acids have one or more double bonds. Each double bond may be in a *cis* or *trans* configuration. In the *cis* configuration, both hydrogens are on the same side of the hydrocarbon chain. In the *trans* configuration, the hydrogens are on opposite sides. A *cis* double bond causes a kink in the chain.

Trans Fats

The food industry artificially hydrogenates oils 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 convert 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 higher 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 using trans fats, and food labels are required to display the trans fat content.

Omega Fatty Acids

Essential fatty acids are those that the human body requires but does not synthesize. Consequently, they have to be supplemented through ingestion via the diet. **Omega**-3 fatty acids (like those in Figure 3.17) fall into this category and are one of only two known for humans (the other is omega-6 fatty acid). These are polyunsaturated fatty acids and are omega-3 because a double bond connects the third carbon from the hydrocarbon chain's end to its neighboring carbon.

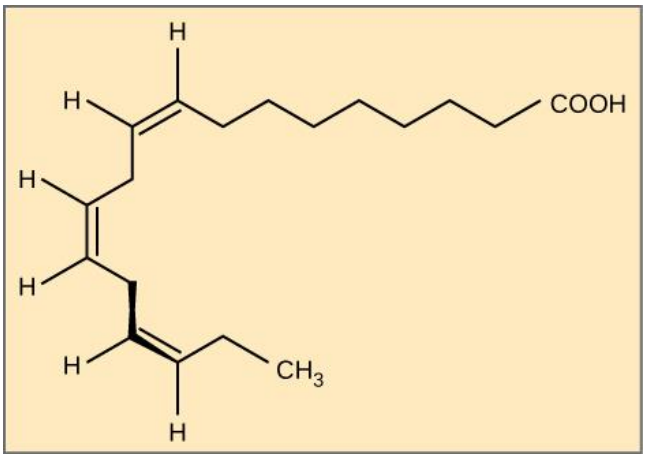


Figure 3.17 Alpha-linolenic acid is an example of an omega-3 fatty acid. It has three *cis* double bonds and, as a result, a curved shape. For clarity, the diagram does not show the carbons. Each singly bonded carbon has two hydrogens associated with it, which the diagram also does not show.

The farthest carbon away from the carboxyl group is numbered as the omega (ω) carbon, and if the double bond is between the third and fourth carbon from that end, it is an omega-3 fatty acid. Nutritionally important because the body does not make them, 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, lower triglycerides in the blood, decrease blood pressure, and prevent thrombosis by inhibiting blood clotting. They also reduce inflammation, and may help lower the risk of some cancers in animals.

Like carbohydrates, fats have received considerable 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, we should consume “healthy” fats in moderate amounts on a regular basis.

Waxes

Wax covers some aquatic birds' feathers and some plants' leaf surfaces. Because of waxes' hydrophobic nature, they prevent water from sticking on the surface (Figure 3.18). Long fatty acid chains esterified to long-chain alcohols comprise waxes.

Figure 3.18 Lipids comprise waxy coverings on some leaves. (credit: Roger Griffith)

Phospholipids

Phospholipids are major plasma membrane constituents that comprise cells' outermost layer. Like fats, they are comprised of fatty acid chains attached to a glycerol or sphingosine

backbone. However, instead of three fatty acids attached as in triglycerides, there are two fatty acids forming diacylglycerol, and a modified phosphate group occupies the glycerol backbone's third carbon (Figure 3.19). A phosphate group alone attached to a diaglycerol does not qualify as a phospholipid. It is phosphatidate (diacylglycerol 3-phosphate), the precursor of phospholipids. An alcohol modifies the phosphate group. Phosphatidylcholine and phosphatidylserine are two important phospholipids that are in plasma membranes.

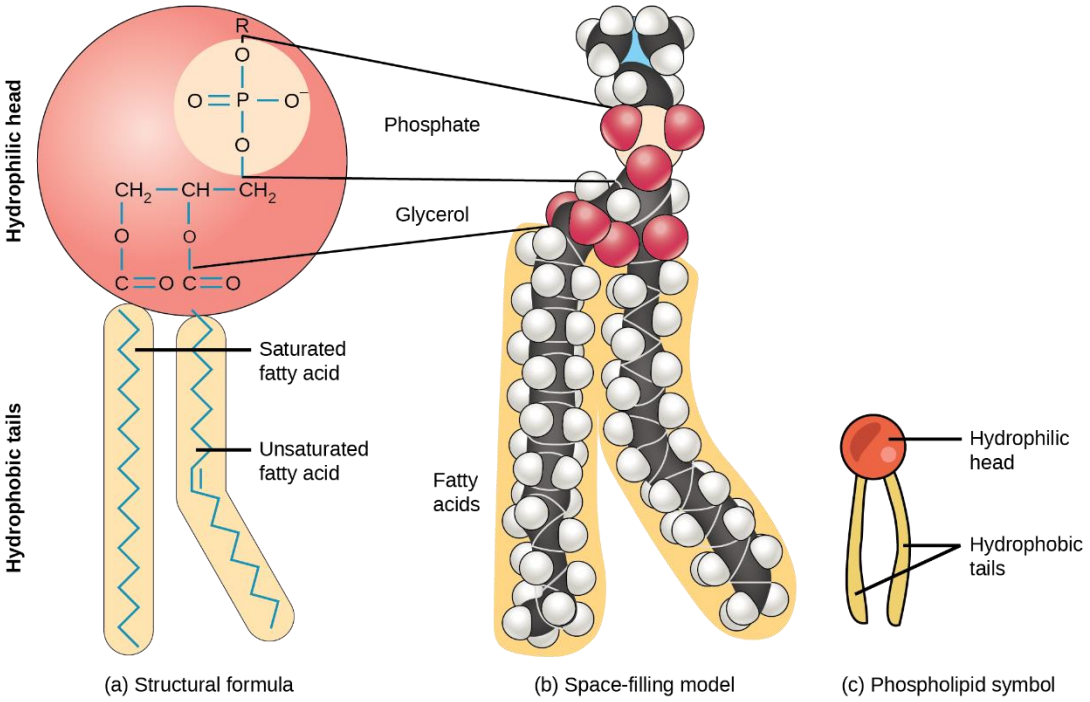


Figure 3.19 A phospholipid is a molecule with two fatty acids and a modified phosphate group attached to a glycerol backbone. Adding a charged or polar chemical group may modify the phosphate.

A phospholipid is an amphipathic molecule, meaning it has a hydrophobic and a hydrophilic part. The fatty acid chains are hydrophobic and cannot interact with water; whereas, the phosphate-containing group is hydrophilic and interacts with water (Figure 3.20).

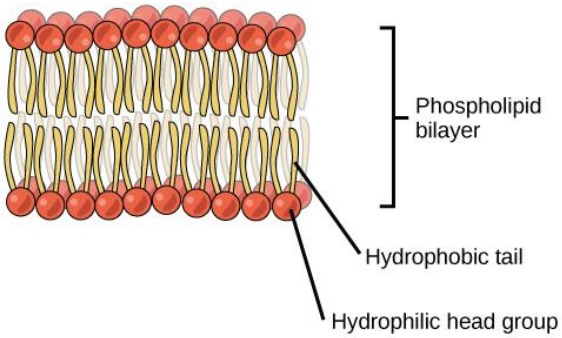


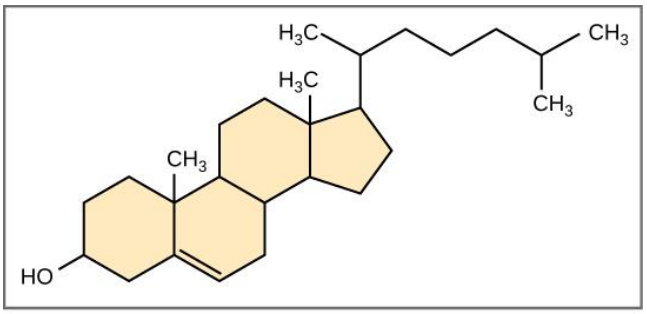
Figure 3.20 The phospholipid bilayer is the major component of all cellular membranes. The hydrophilic head groups of the phospholipids face the aqueous solution. The hydrophobic tails are sequestered in the middle of the bilayer.

The head is the hydrophilic part, and the tail contains the hydrophobic fatty acids. In a membrane, a bilayer of phospholipids forms the structure's matrix, phospholipids' fatty acid tails face inside, away from water; whereas, the phosphate group faces the outside, aqueous side (Figure 3.20).

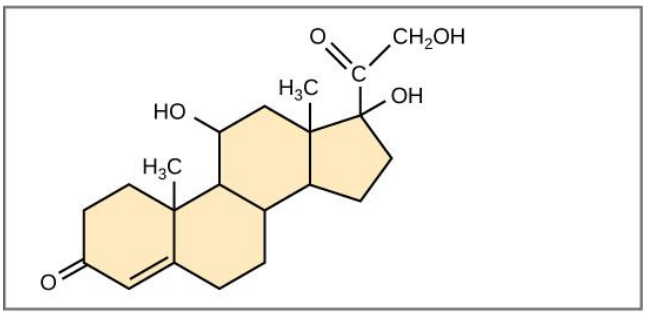
Phospholipids are responsible for the plasma membrane's dynamic nature. If a drop of phospholipids is placed in water, it spontaneously forms a structure that scientists call a micelle, where the hydrophilic phosphate heads face the outside and the fatty acids face the structure's interior.

Steroids

Unlike the phospholipids and fats that we discussed earlier, **steroids** have a fused ring structure. Although they do not resemble the other lipids, scientists group them 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 3.21). Many steroids also have the –OH functional group, which puts them in the alcohol classification (sterols).



Cholesterol



Cortisol

Figure 3.21 Four fused hydrocarbon rings comprise steroids such as cholesterol and cortisol.

Cholesterol is the most common steroid. The liver synthesizes cholesterol and is the precursor to many steroid hormones such as testosterone and estradiol, which gonads and endocrine glands secrete. It is also the precursor to Vitamin D. Cholesterol is also the precursor of bile salts, which help emulsifying fats and their subsequent absorption by cells. Although lay people often speak negatively about cholesterol, it is necessary for the body's proper functioning. Sterols (cholesterol in animal cells, phytosterol in plants) are components of the plasma membrane of cells and are found within the phospholipid bilayer.

LINK TO LEARNING

For an additional perspective on lipids, explore the interactive animation “Biomolecules: The Lipids”: https://www.wisc-online.com/learn/natural-science/life-science/ap13204/biomolecules---the-lipids

Chapter Attributions

Proteins

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

- Describe the functions proteins perform in the cell and in tissues
- Discuss the relationship between amino acids and proteins
- Explain the four levels of protein organization
- Describe the ways in which protein shape and function are linked

Proteins are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules. 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. They are all, however, amino acid polymers arranged in a linear sequence.

Types and Functions of Proteins

Enzymes, which living cells produce, are catalysts in biochemical reactions (like digestion) and are usually complex or conjugated proteins. Each enzyme is specific for the substrate (a reactant that binds to an enzyme) upon which it acts. The enzyme may help in breakdown, rearrangement, or synthesis reactions. We call enzymes that break down their substrates catabolic enzymes. Those that build more complex molecules from their substrates are anabolic enzymes, and enzymes that affect the rate of reaction are catalytic enzymes. Note that all enzymes increase the reaction rate and, therefore, are organic catalysts. An example of an enzyme is salivary amylase, which hydrolyzes its substrate amylose, a component of starch.

Hormones are chemical-signaling molecules, usually small proteins or steroids, secreted by endocrine cells that act to control or regulate specific physiological processes, including growth, development, metabolism, and reproduction. For example, insulin is a protein hormone that helps regulate the blood glucose level. Table 3.1 lists the primary types and functions of proteins.

Protein Types and Functions

Type	Examples	Functions
Digestive Enzymes	Amylase, lipase, pepsin, trypsin	Help in food by catabolizing nutrients into units
Transport	Hemoglobin, albumin	Carry substances in the blood or lymph thr body
Structural	Actin, tubulin, keratin	Construct different structures, like the cyt
Hormones	Insulin, thyroxine	Coordinate different body systems' activit
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 embryo deve the seedling

Table3.1

Proteins have different shapes and molecular weights. Some proteins are globular in shape; whereas, others are fibrous in nature. For example, hemoglobin is a globular protein, but collagen, located in our skin, is a fibrous protein. Protein shape is critical to its function, and many different types of chemical bonds maintain this shape. Changes in temperature, pH, and exposure to chemicals may lead to permanent changes in the protein's shape, leading to loss of function, or **denaturation**. Different arrangements of the same 20 types of amino acids comprise all proteins. Two rare new amino acids were discovered recently (selenocystein and pirrolysine), and additional new discoveries may be added to the list.

Amino Acids

Amino acids are the monomers that comprise proteins. Each amino acid has the same fundamental structure, which consists of a central carbon atom, or the alpha (α) carbon, bonded to an amino group (NH_2), a carboxyl group (COOH), and to a hydrogen atom. Every amino acid also has another atom or group of atoms bonded to the central atom known as the R group (Figure 3.22).

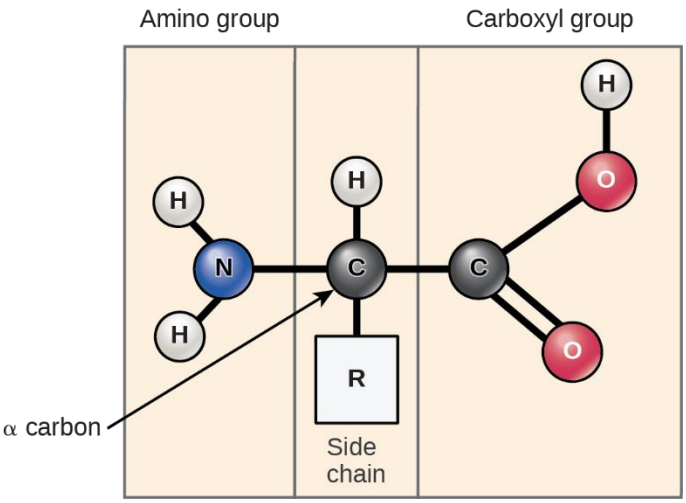


Figure 3.22 Amino acids have a central asymmetric carbon to which an amino group, a carboxyl group, a hydrogen atom, and a side chain (R group) are attached.

Scientists use the name "amino acid" because these acids contain both amino group and carboxyl-acid-group in their basic structure. As we mentioned, there are 20 common amino acids present in proteins. Nine of these are essential amino acids in humans because the human body cannot produce them and we obtain them from our diet. For each amino acid, the R group (or side chain) is different (Figure 3.23).

VISUAL CONNECTION

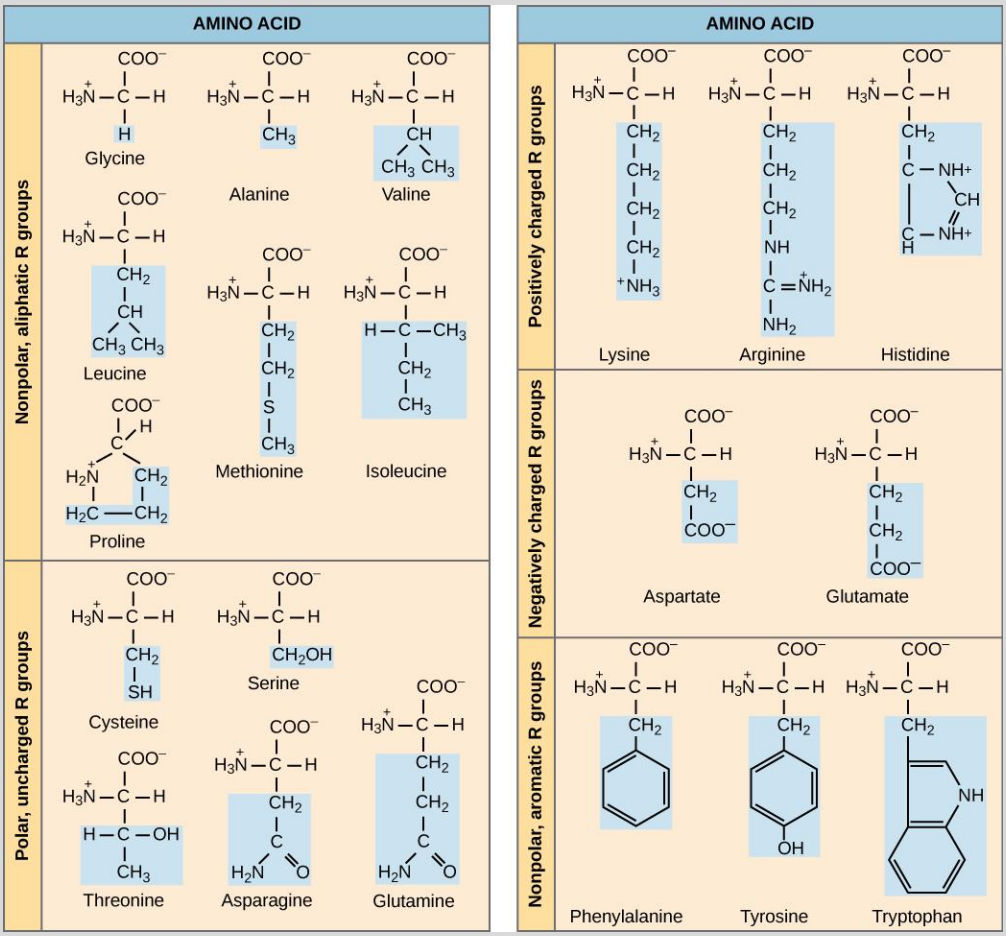


Figure 3.23 There are 20 common amino acids commonly found in proteins, each with a different R group (variant group) that determines its chemical nature.

Which categories of amino acid would you expect to find on a soluble protein's surface and which would you expect to find in the interior? What distribution of amino acids would you expect to find in a protein embedded in a lipid bilayer?

The chemical nature of the side chain determines the amino acid's nature (that is, whether it is acidic, basic, polar, or nonpolar). For example, the amino acid glycine has a hydrogen atom as the R group. 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, and

therefore these amino acids are also basic amino acids. Proline has an R group that is linked to the amino group, forming a ring-like structure. Proline is an exception to the amino acid's standard structure since its amino group is not separate from the side chain (Figure 3.23).

A single upper case letter or a three-letter abbreviation represents amino acids. For example, the letter V or the three-letter symbol val represent valine. Just as some fatty acids are essential to a diet, some amino acids also are necessary. These essential amino acids in humans include isoleucine, leucine, and cysteine. Essential amino acids refer to those necessary to build proteins in the body, but not those that the body produces. Which amino acids are essential varies from organism to organism.

The sequence and the number of amino acids ultimately determine the protein's shape, size, and function. A covalent bond, or **peptide bond**, attaches to each amino acid, which a dehydration reaction forms. One amino acid's carboxyl group and the incoming amino acid's amino group combine, releasing a water molecule. The resulting bond is the peptide bond (Figure 3.24).

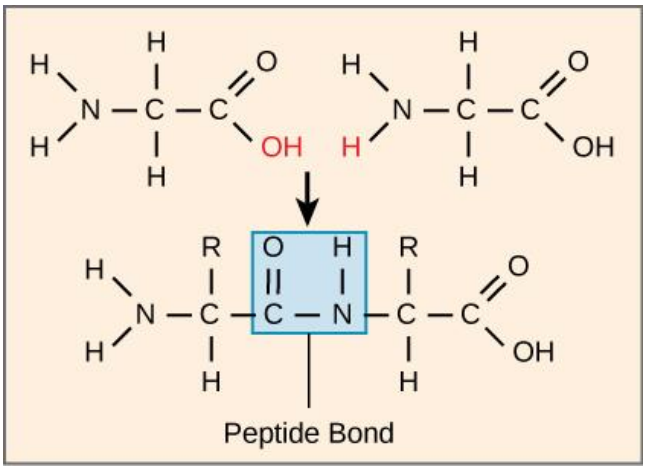


Figure 3.24 Peptide bond formation is a dehydration synthesis reaction. The carboxyl group of one amino acid is linked to the incoming amino acid's amino group. In the process, it releases a water molecule.

The products that such linkages form are peptides. As more amino acids join to this growing chain, the resulting chain is a polypeptide. Each polypeptide has a free amino group at one end. This end the N terminal, or the amino terminal, and the other end has a free carboxyl group,

also the C or carboxyl terminal. While the terms polypeptide and protein are sometimes used interchangeably, a polypeptide is technically a polymer of amino acids, whereas the term protein is used for a polypeptide or polypeptides that have combined together, often have bound non-peptide prosthetic groups, have a distinct shape, and have a unique function. After protein synthesis (translation), most proteins are modified. These are known as post-translational modifications. They may undergo cleavage, phosphorylation, or may require adding other chemical groups. Only after these modifications is the protein completely functional.

EVOLUTION CONNECTION

The Evolutionary Significance of Cytochrome c

Cytochrome c is an important component of the electron transport chain, a part of cellular respiration, and it is normally located in the cellular organelle, the mitochondrion. This protein has a heme prosthetic group, and the heme's central ion alternately reduces and oxidizes during electron transfer. Because this essential protein's role in producing cellular energy is crucial, it has changed very little over millions of years. Protein sequencing has shown that there is a considerable amount of cytochrome c amino acid sequence homology among different species. In other words, we can assess evolutionary kinship by measuring the similarities or differences among various species' DNA or protein sequences.

Scientists have determined that human cytochrome c contains 104 amino acids. For each cytochrome c molecule from different organisms that scientists have sequenced to date, 37 of these amino acids appear in the same position in all cytochrome c samples. This indicates that there may have been a common ancestor. On comparing the human and chimpanzee protein sequences, scientists did not find a sequence difference. When researchers compared human and rhesus monkey sequences, the single difference was in one amino acid. In another comparison, human to yeast sequencing shows a difference in the 44th position.

Protein Structure

As we discussed earlier, a protein's shape is critical to its function. For example, an enzyme can bind to a specific substrate at an 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.

Primary Structure

Amino acids' unique sequence in a polypeptide chain is its **primary structure**. For example, the pancreatic hormone insulin has two polypeptide chains, A and B, and they are linked together by disulfide bonds. The N terminal amino acid of the A chain is glycine; whereas, the C terminal amino acid is asparagine (Figure 3.25). The amino acid sequences in the A and B chains are unique to insulin.

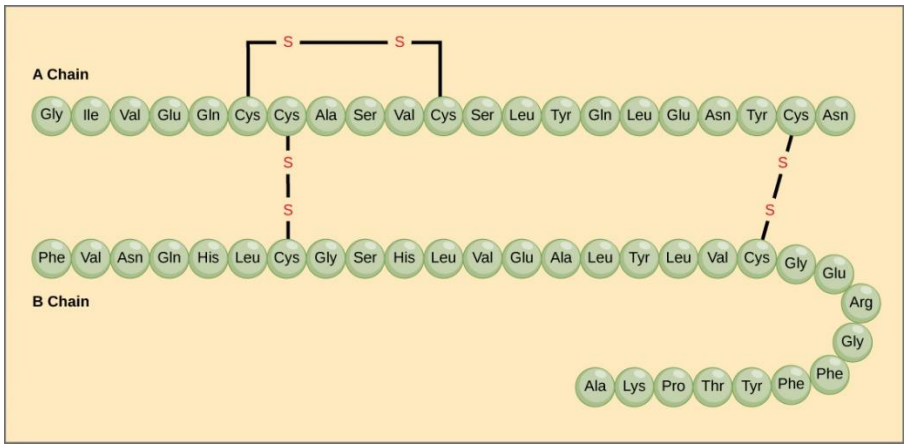


Figure 3.25 Bovine serum insulin is a protein hormone comprised of two peptide chains, A (21 amino acids long) and B (30 amino acids long). In each chain, three-letter abbreviations that represent the amino acids' names in the order they are present indicate primary structure. The amino acid cysteine (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. Note that all disulfide bonds are the same length, but we have drawn them different sizes for clarity.

The gene encoding the protein ultimately determines the unique sequence for every protein. A change in nucleotide sequence of the gene's coding region may lead to adding a different amino acid to the growing polypeptide chain, causing a change in protein structure and function. In sickle cell anemia, the hemoglobin *β* chain (a small portion of which we show in Figure 3.26) has a single amino acid substitution, causing a change in protein structure and function. Specifically, valine in the *β* chain substitutes the amino acid glutamic. What is most

remarkable to consider is that a hemoglobin molecule is comprised of two alpha and two beta chains that each consist of about 150 amino acids. The molecule, therefore, has about 600 amino acids. The structural difference between a normal hemoglobin molecule and a sickle cell molecule—which dramatically decreases life expectancy—is a single amino acid of the 600. What is even more remarkable is that three nucleotides each encode those 600 amino acids, and a single base change (point mutation), 1 in 1800 bases causes the mutation.

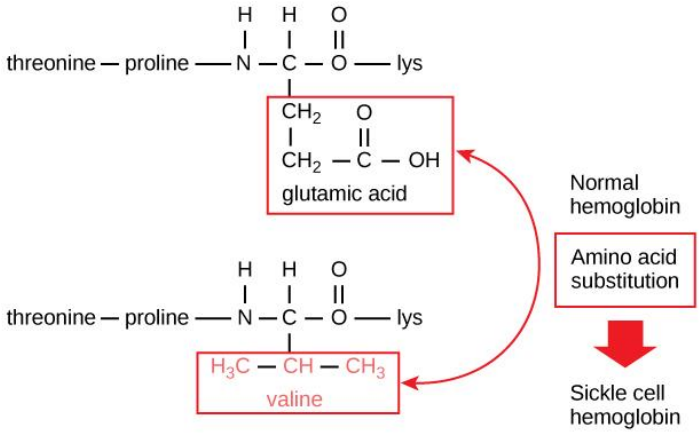


Figure 3.26 The beta chain of hemoglobin is 147 residues in length, yet a single amino acid substitution leads to sickle cell anemia. In normal hemoglobin, the amino acid at position seven is glutamate. In sickle cell hemoglobin, a valine replaces glutamate.

Because of this change of one amino acid in the chain, hemoglobin molecules form long fibers that distort the biconcave, or disc-shaped, red blood cells and causes them to assume a crescent or “sickle” shape, which clogs blood vessels (Figure 3.27). This can lead to myriad serious health problems such as breathlessness, dizziness, headaches, and abdominal pain for those affected by this disease.

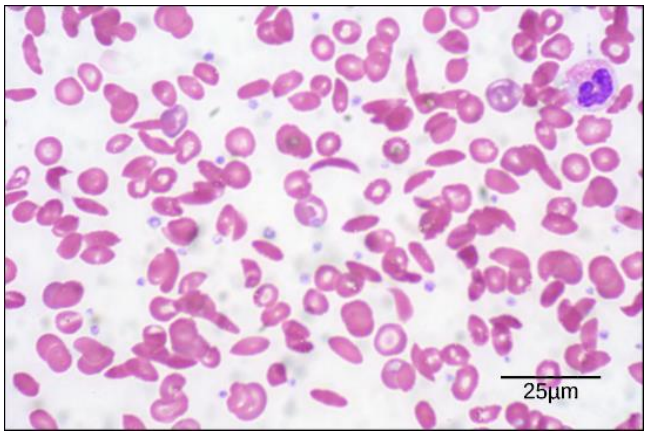


Figure 3.27 In this blood smear, visualized at 535x magnification using bright field microscopy, sickle cells are crescent shaped, while normal cells are disc-shaped. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

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 3.28). Both structures are held in shape by hydrogen bonds. The hydrogen bonds form between the oxygen atom in the carbonyl group in one amino acid and another amino acid that is four amino acids farther along the chain.

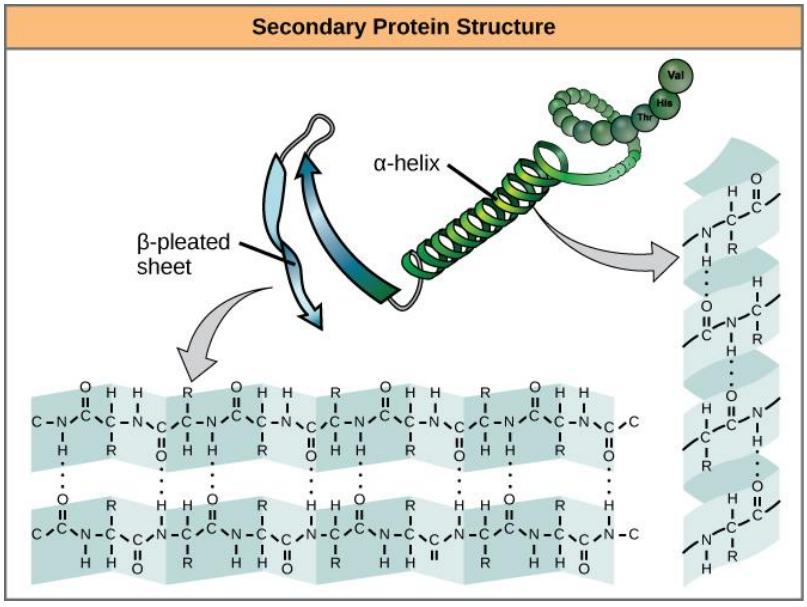


Figure 3.28 The α-helix and β-pleated sheet are secondary structures of proteins that form because of hydrogen bonding between carbonyl and amino groups in the peptide backbone. Certain amino acids have a propensity to form an α-helix, while others have a propensity to form a β-pleated sheet.

Every helical turn in an alpha helix has 3.6 amino acid residues. The polypeptide's R groups (the variant groups) protrude out from the α-helix chain. In the β-pleated sheet, hydrogen bonding between atoms on the polypeptide chain's backbone form the "pleats". The R groups are attached to the carbons and extend above and below the pleat's folds. The pleated segments align parallel or antiparallel to each other, and hydrogen bonds form between the partially positive nitrogen atom in the amino group and the partially negative oxygen atom in the peptide backbone's carbonyl group. The α-helix and β-pleated sheet structures are in most globular and fibrous proteins and they play an important structural role.

Tertiary Structure

The polypeptide's unique three-dimensional structure is its **tertiary structure** (Figure 3.29). This structure is in part due to chemical interactions at work on the polypeptide chain. Primarily, the interactions among R groups create the protein's complex three-dimensional tertiary structure. The nature of the R groups in the amino acids involved can counteract forming the hydrogen bonds we described for standard secondary structures. For example, R groups with like charges repel each other and those with unlike charges are attracted to each other (ionic bonds). When protein folding takes place, the nonpolar amino acids' hydrophobic R groups lie in the protein's interior; whereas, the hydrophilic R groups lie on the outside. Scientists also call the former interaction types hydrophobic interactions. Interaction between cysteine side chains forms disulfide linkages in the presence of oxygen, the only covalent bond that forms during protein folding.

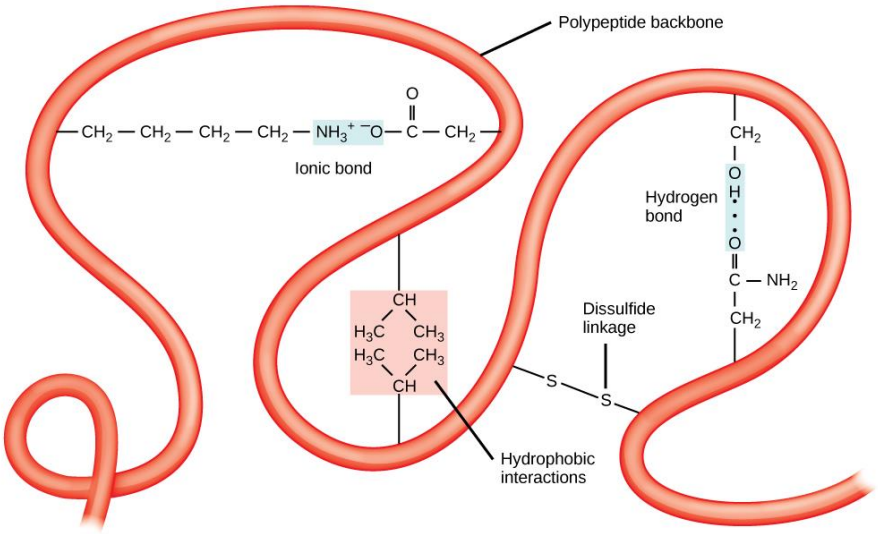


Figure 3.29 A variety of chemical interactions determine the proteins' tertiary structure. These include hydrophobic interactions, ionic bonding, hydrogen bonding, and disulfide linkages.

All of these interactions, weak and strong, determine the protein's final three-dimensional shape. When a protein loses its three-dimensional shape, it may no longer be functional.

Quaternary Structure

In nature, some proteins form from several polypeptides, or subunits, and the interaction of these subunits forms the **quaternary structure**. Weak interactions between the subunits help to stabilize the overall structure. For example, insulin (a globular protein) has a combination of hydrogen and disulfide bonds that cause it to mostly clump into a ball shape. Insulin starts out as a single polypeptide and loses some internal sequences in the presence of post-translational modification after forming the disulfide linkages that hold the remaining chains together. Silk (a fibrous protein), however, has a *β*-pleated sheet structure that is the result of hydrogen bonding between different chains.

Figure 3.30 illustrates the four levels of protein structure (primary, secondary, tertiary, and quaternary).

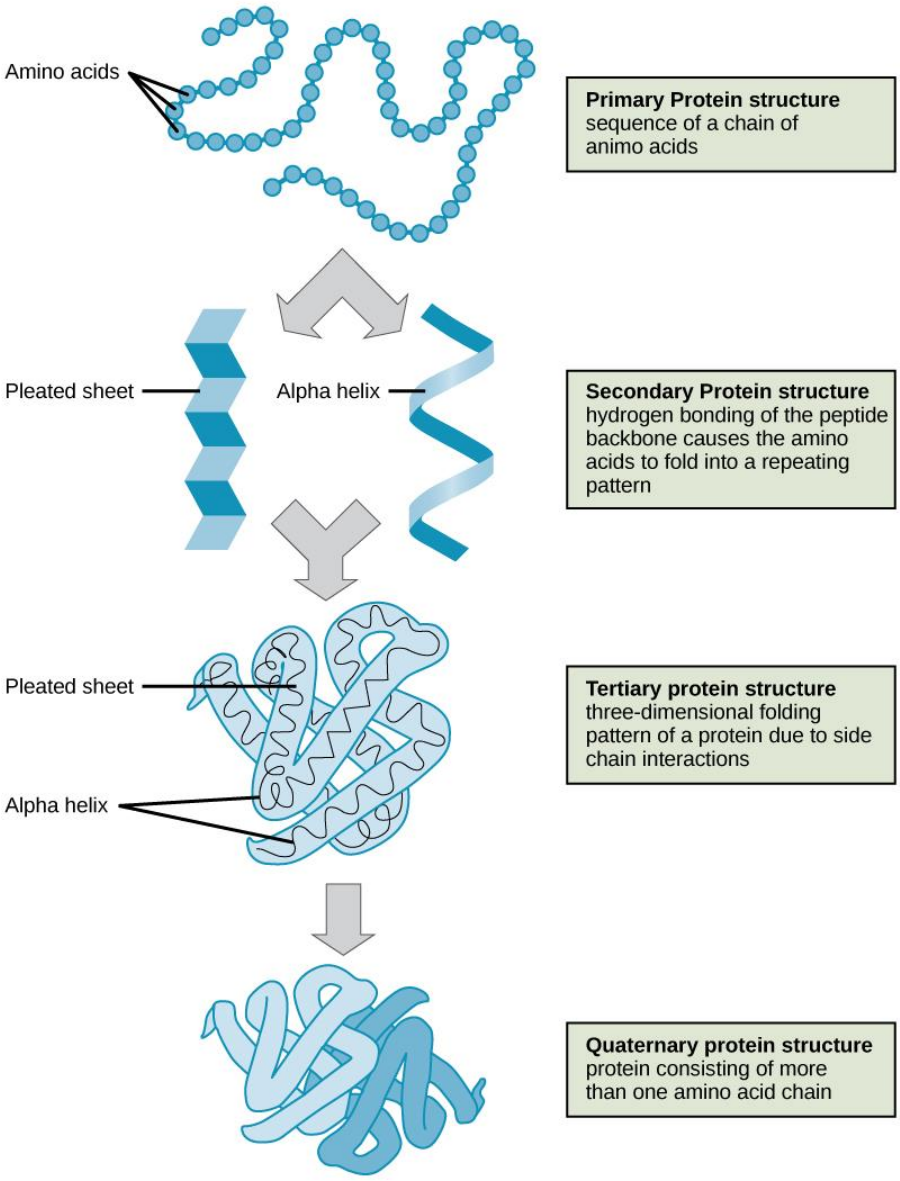


Figure 3.30 Observe the four levels of protein structure in these illustrations. (credit: modification of work by National Human Genome Research Institute)

Denaturation and Protein Folding

Each protein has its own unique sequence and shape that chemical interactions hold together. If the protein is subject to changes in temperature, pH, or exposure to chemicals, the protein structure may change, losing its shape without losing its primary sequence in what scientists call denaturation. Denaturation is often reversible because the polypeptide's primary structure is conserved in the process if the denaturing agent is removed, allowing the protein to resume

its function. Sometimes denaturation is irreversible, leading to loss of function. One example of irreversible protein denaturation is frying an egg. The albumin protein in the liquid egg white denatures when placed in a hot pan. Not all proteins denature at high temperatures. For instance, bacteria that survive in hot springs have proteins that function at temperatures close to boiling. The stomach is also very acidic, has a low pH, and denatures proteins as part of the digestion process; however, the stomach's digestive enzymes retain their activity under these conditions.

Protein folding is critical to its function. Scientists originally thought that the proteins themselves were responsible for the folding process. Only recently researchers discovered that often they receive assistance in the folding process from protein helpers, or **chaperones** (or chaperonins) that associate with the target protein during the folding process. They act by preventing polypeptide aggregation that comprise the complete protein structure, and they disassociate from the protein once the target protein is folded.

LINK TO LEARNING

For an additional perspective on proteins, view this animation called “Biomolecules: The Proteins” https://www.wisc-online.com/learn/natural-science/life-science/ap13304/biomolecules---the-proteins

Chapter Attributions

Nucleic Acids

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

- Describe nucleic acids' structure and define the two types of nucleic acids
- Explain DNA's structure and role
- Explain RNA's structure and roles

Nucleic acids are the most important macromolecules for the continuity of life. They carry the cell's genetic blueprint and carry instructions for its functioning.

DNA and RNA

The two main types of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material in all living organisms, ranging from single-celled bacteria to multicellular mammals. It is in the nucleus of eukaryotes and in the organelles, chloroplasts, and mitochondria. In prokaryotes, the DNA is not enclosed in a membranous envelope.

The cell's entire genetic content is its genome, and the study of genomes is genomics. In eukaryotic cells but not in prokaryotes, 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. The DNA molecules never leave the nucleus but instead use an intermediary to communicate with the rest of the cell. This intermediary is the **messenger RNA (mRNA)**. Other types of RNA—like rRNA, tRNA, and microRNA—are involved in protein synthesis and its regulation.

DNA and RNA are comprised of monomers that scientists call **nucleotides**. The nucleotides combine with each other to form a **polynucleotide**, DNA or RNA. Three components comprise each nucleotide: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate group (Figure 3.31). Each nitrogenous base in a nucleotide is attached to a sugar molecule, which is attached to one or more phosphate groups.

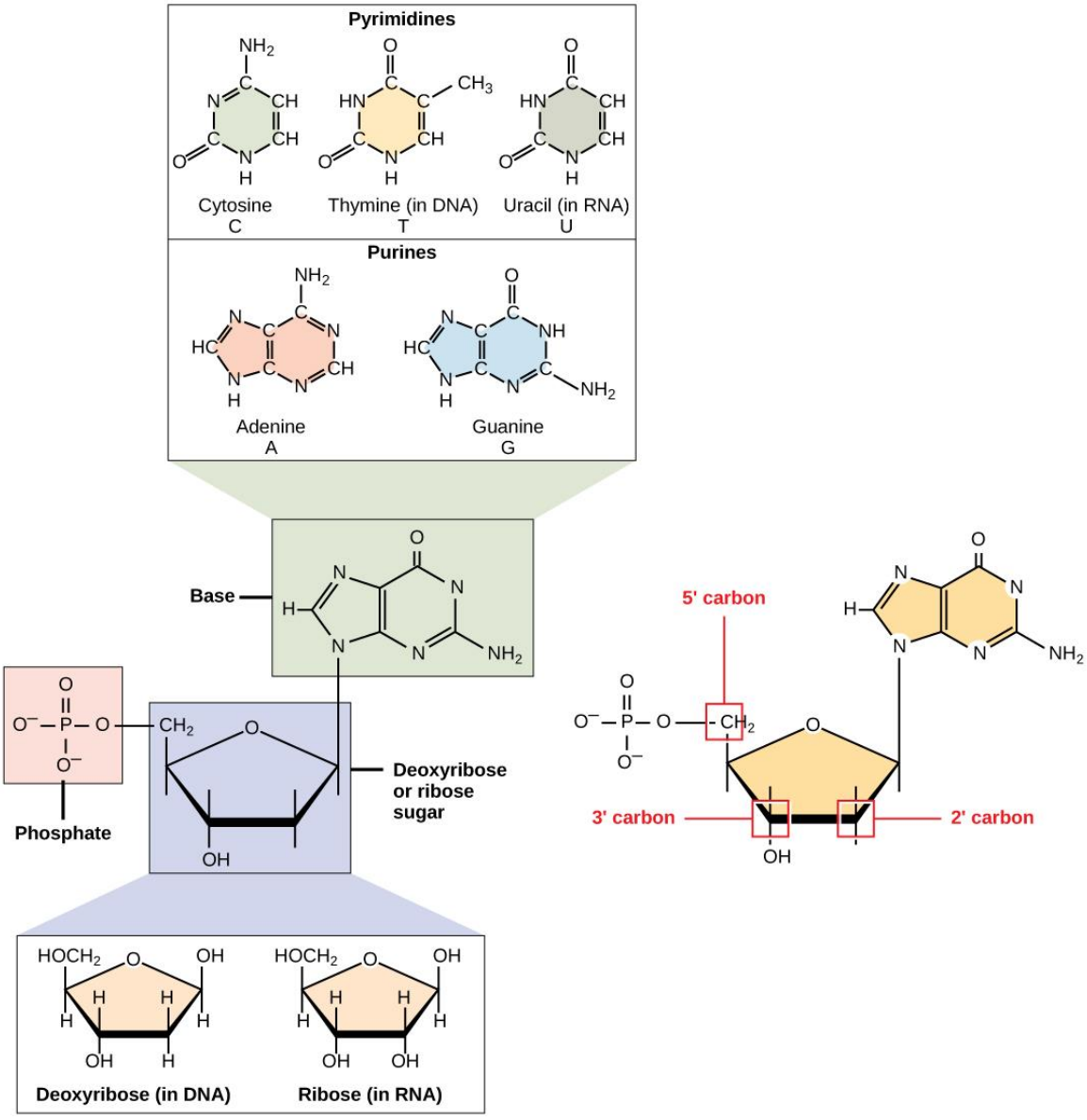


Figure 3.31 Three components comprise a nucleotide: a nitrogenous base, a pentose sugar, and one or more phosphate groups. Carbon residues in the pentose are numbered 1′ through 5′ (the prime distinguishes these residues from those in the base, which are numbered without using a prime notation). The base is attached to the ribose's 1′ position, and the phosphate is attached to the 5′ position. When a polynucleotide forms, the incoming nucleotide's 5′ phosphate attaches to the 3′ hydroxyl group at the end of the growing chain. Two types of pentose are in nucleotides, deoxyribose (found in DNA) and ribose (found in RNA). Deoxyribose is similar in structure to ribose, but it has an H instead of an OH at the 2′ position. We can divide

bases into two categories: purines and pyrimidines. Purines have a double ring structure, and pyrimidines have a single ring.

The nitrogenous bases, important components of nucleotides, are organic molecules and are so named because they contain carbon and nitrogen. They are bases because they contain an amino group that has the potential of binding an extra hydrogen, and thus decreasing the hydrogen ion concentration in its environment, making it more basic. Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G) cytosine (C), and thymine (T).

Scientists classify adenine and guanine as **purines**. The purine's primary structure is two carbon-nitrogen rings. Scientists classify cytosine, thymine, and uracil as **pyrimidines** which have a single carbon-nitrogen ring as their primary structure (Figure 3.31). Each of these basic carbon-nitrogen rings has different functional groups attached to it. In molecular biology shorthand, we know the nitrogenous bases by their symbols A, T, G, C, and U. DNA contains A, T, G, and C; whereas, RNA contains A, U, G, and C.

The pentose sugar in DNA is deoxyribose, and in RNA, the sugar is ribose (Figure 3.31). The difference between the sugars is the presence of the hydroxyl group on the ribose's second carbon and hydrogen on the deoxyribose's second carbon. The carbon atoms of the sugar molecule are numbered as 1′, 2′, 3′, 4′, and 5′ (1′ is read as “one prime”). The phosphate residue attaches to the hydroxyl group of the 5′ carbon of one sugar and the hydroxyl group of the 3′ carbon of the sugar of the next nucleotide, which forms a 5′–3′ **phosphodiester** linkage. A simple dehydration reaction like the other linkages connecting monomers in macromolecules does not form the phosphodiester linkage. Its formation involves removing two phosphate groups. A polynucleotide may have thousands of such phosphodiester linkages.

DNA Double-Helix Structure

DNA has a double-helix structure (Figure 3.32). The sugar and phosphate lie on the outside of the helix, forming the DNA's backbone. The nitrogenous bases are stacked in the interior, like a pair of staircase steps. Hydrogen bonds bind the pairs to each other. Every base pair in the double helix is separated from the next base pair by 0.34 nm. The helix's two strands run in opposite directions, meaning that the 5′ carbon end of one strand will face the 3′ carbon end of

its matching strand. (Scientists call this an antiparallel orientation and is important to DNA replication and in many nucleic acid interactions.)

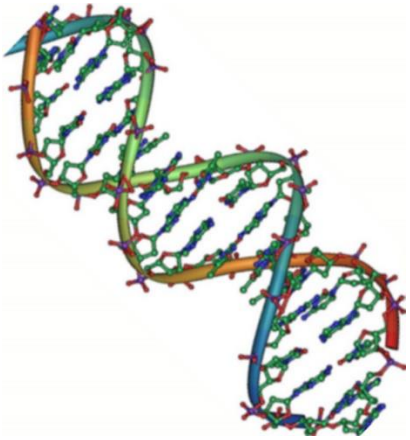


Figure 3.32 Native DNA is an antiparallel double helix. The phosphate backbone (indicated by the curvy lines) is on the outside, and the bases are on the inside. Each base from one strand interacts via hydrogen bonding with a base from the opposing strand. (credit: Jerome Walker/Dennis Myts)

Only certain types of base pairing are allowed. For example, a certain purine can only pair with a certain pyrimidine. This means A can pair with T, and G can pair with C, as Figure 3.33 shows. This is the base complementary rule. In other words, the DNA strands are complementary to each other. If the sequence of one strand is AATTGGCC, the complementary strand would have the sequence TTAACCGG. During DNA replication, each strand copies itself, resulting in a daughter DNA double helix containing one parental DNA strand and a newly synthesized strand.

VISUAL CONNECTION

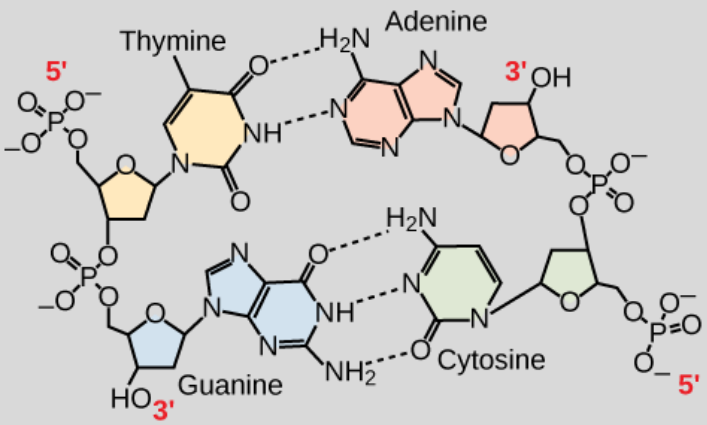


Figure 3.33 In a double stranded DNA molecule, the two strands run antiparallel to one another so that one strand runs 5′ to 3′ and the other 3′ to 5′. The phosphate backbone is located on the outside, and the bases are in the middle. Adenine forms hydrogen bonds (or base pairs) with thymine, and guanine base pairs with cytosine.

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

RNA

Ribonucleic acid, or RNA, is mainly involved in the process of protein synthesis under the direction of DNA. RNA is usually single-stranded and is comprised of ribonucleotides that are linked by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, U, G, and C), and the phosphate group.

There are four major types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and microRNA (miRNA). The first, mRNA, carries the message from DNA, which controls all of the cellular activities in a cell. If a cell requires synthesizing a certain protein, the gene for this product turns “on” and the messenger RNA synthesizes in the nucleus. The RNA base sequence is complementary to the DNA's coding sequence from which it has been copied. However, in RNA, the base T is absent and U is present instead. If the DNA strand has a sequence AATTGCGC, the sequence of the complementary RNA is UUAACGCG. In the cytoplasm, the mRNA interacts with ribosomes and other cellular machinery (Figure 3.34).

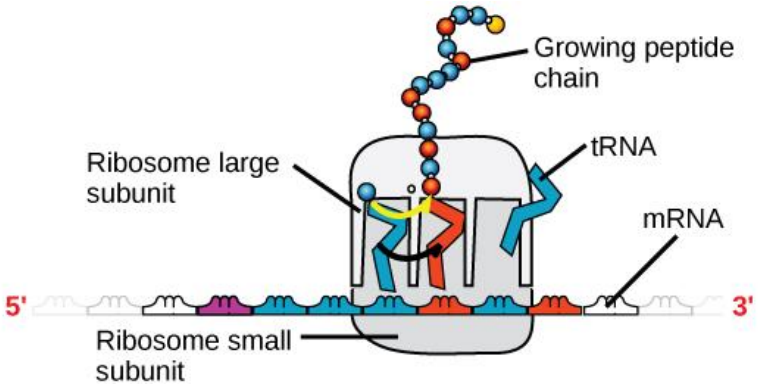


Figure 3.34 A ribosome has two parts: a large subunit and a small subunit. The mRNA sits in between the two subunits. A tRNA molecule recognizes a codon on the mRNA, binds to it by complementary base pairing, and adds the correct amino acid to the growing peptide chain.

The mRNA is read in sets of three bases known as codons. Each codon codes for a single amino acid. In this way, the mRNA is read and the protein product is made. **Ribosomal RNA (rRNA)** is a major constituent of ribosomes on which the mRNA binds. The rRNA ensures the proper alignment of the mRNA and the Ribosomes. The ribosome's rRNA also has an enzymatic activity (peptidyl transferase) and catalyzes peptide bond formation between two aligned amino acids. **Transfer RNA (tRNA)** is one of the smallest of the four types of RNA, usually 70–90 nucleotides long. It carries the correct amino acid to the protein synthesis site. It is the base pairing between the tRNA and mRNA that allows for the correct amino acid to insert itself in the polypeptide chain. MicroRNAs are the smallest RNA molecules and their role involves regulating gene expression by interfering with the expression of certain mRNA messages. Table 3.2 summarizes DNA and RNA features.

DNA and RNA Features

	DNA	RNA
Function	Carries genetic information	Involved in protein synthesis
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

Table3.2

Even though the RNA is single stranded, most RNA types show extensive intramolecular base pairing between complementary sequences, creating a predictable three-dimensional structure essential for their function.

As you have learned, information flow in an organism takes place from DNA to RNA to protein. DNA dictates the structure of mRNA in a process scientists call **transcription**, and RNA dictates the protein's structure in a process scientists call **translation**. This is the Central Dogma of Life, which holds true for all organisms; however, exceptions to the rule occur in connection with viral infections.

Chapter Attributions

ATP: Adenosine Triphosphate

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

- Explain ATP's role as the cellular energy currency
- Describe how energy releases through ATP hydrolysis

Even exergonic, energy-releasing 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, from where does energy to power such reactions come? The answer lies with an energy-supplying molecule scientists call **adenosine triphosphate**, or **ATP**. This is a small, relatively simple molecule (Figure 6.13), but within some of its bonds, it contains the potential for a quick burst of energy that can be harnessed to perform cellular work. Think of this molecule as the cells' primary energy currency in much the same way that money is the currency that people exchange for things they need. ATP powers the majority of energy-requiring cellular reactions.

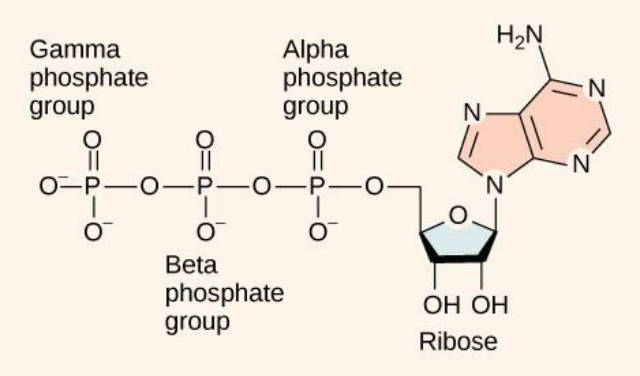


Figure 6.13 ATP is the cell's primary energy currency. It has an adenosine backbone with three phosphate groups attached.

As its name suggests, adenosine triphosphate is comprised of adenosine bound to three phosphate groups (Figure 6.13). Adenosine is a nucleoside consisting of the nitrogenous base adenine and a five-carbon sugar, ribose. The three phosphate groups, in order of closest to furthest from the ribose sugar, are alpha, beta, and gamma. Together, these chemical groups constitute an energy powerhouse. However, not all bonds within this molecule exist in a particularly high-energy state. Both bonds that link the phosphates are equally high-energy bonds (**phosphoanhydride bonds**) that, when broken, release sufficient energy to power a variety of cellular reactions and processes. These high-energy bonds are the bonds between the second and third (or beta and gamma) phosphate groups and between the first and second phosphate groups. These bonds are “high-energy” 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 using a water molecule, it is a hydrolysis reaction. In other words, ATP hydrolyzes into ADP in the following reaction:

ATP+H2O→ADP+Pi+free energysize 12{{ATP} + H rSub { size 8{2} } O ADP + P rSub { size 8{i}} + {free energy} } {}

Like most chemical reactions, ATP to ADP hydrolysis is reversible. The reverse reaction regenerates ATP from ADP + P_i. Cells rely on ATP regeneration just as people rely on regenerating spent money through some sort of income. Since ATP hydrolysis releases energy, ATP regeneration must require an input of free energy. This equation expresses ATP formation:

ADP+Pi+free energy→ATP+H2Osize 12{{ATP} + H rSub { size 8{2} } O ADP + P rSub { size 8{i}} + {free energy} } {}

Two prominent questions remain with regard to using ATP as an energy source. Exactly how much free energy releases with ATP hydrolysis, and how does that free energy do cellular work? The calculated ΔG for the hydrolysis of one ATP mole into ADP and P_i is −7.3 kcal/mole (−30.5 kJ/mol). Since this calculation is true under standard conditions, one would expect a different value exists under cellular conditions. In fact, the ΔG for one ATP mole's hydrolysis in a living cell is almost double the value at standard conditions: –14 kcal/mol (−57 kJ/mol).

ATP is a highly unstable molecule. Unless quickly used to perform work, ATP spontaneously dissociates into ADP + P_i, and the free energy released during this process is lost as heat. The second question we posed above discusses how ATP hydrolysis energy release performs work inside the cell. This depends on a strategy scientists call energy coupling. Cells couple the ATP hydrolysis' exergonic reaction allowing them to proceed. One example of energy coupling using ATP involves a transmembrane ion pump that is extremely important for cellular function. This sodium-potassium pump (Na^+/K^+ pump) drives sodium out of the cell and potassium into the cell (Figure 6.14). A large percentage of a cell's ATP powers this pump, because cellular processes bring considerable sodium into the cell and potassium out of it. The pump works constantly to stabilize cellular concentrations of sodium and potassium. In order for the pump to turn one cycle (exporting three Na+ ions and importing two K^+ ions), one ATP molecule must hydrolyze. When ATP hydrolyzes, its gamma phosphate does not simply float away, but it actually transfers onto the pump protein. Scientists call this process of a phosphate group binding to a molecule phosphorylation. As with most ATP hydrolysis cases, a phosphate from ATP transfers onto another molecule. In a phosphorylated state, the Na^+/K^+ pump has more free energy and is triggered to undergo a conformational change. This change allows it to release Na^+ to the cell's outside. It then binds extracellular K^+, which, through another conformational change, causes the phosphate to detach from the pump. This phosphate release triggers the K^+ to release to the cell's inside. Essentially, the energy released from the ATP hydrolysis couples with the energy required to power the pump and transport Na^+ and K^+ ions. ATP performs cellular work using this basic form of energy coupling through phosphorylation.

VISUAL CONNECTION

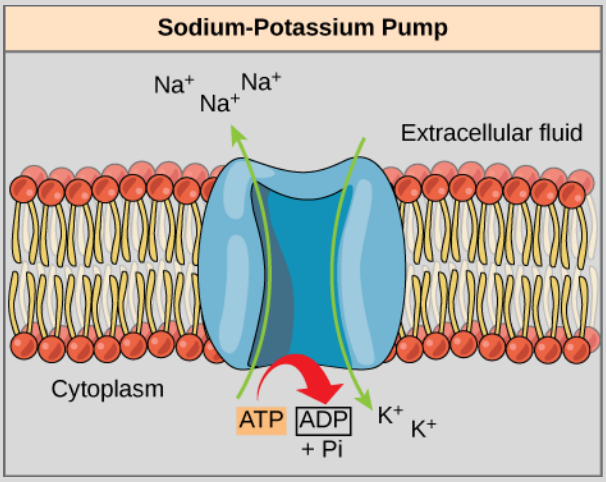


Figure 6.14 The sodium-potassium pump is an example of energy coupling. The energy derived from exergonic ATP hydrolysis pumps sodium and potassium ions across the cell membrane.

One ATP molecule's hydrolysis releases 7.3 kcal/mol of energy ($\Delta G = -7.3$ kcal/mol of energy). If it takes 2.1 kcal/mol of energy to move one Na^+ across the membrane ($\Delta G = +2.1$ kcal/mol of energy), how many sodium ions could one ATP molecule's hydrolysis move?

Often during cellular metabolic reactions, such as nutrient synthesis and breakdown, certain molecules must alter slightly in their conformation to become substrates for the next step in the reaction series. One example is during the very first steps of cellular respiration, when a sugar glucose molecule breaks down in the process of glycolysis. In the first step, ATP is required to phosphorylze glucose, creating a high-energy but unstable intermediate. This phosphorylation reaction powers a conformational change that allows the phosphorylated glucose molecule to convert to the phosphorylated sugar fructose. Fructose is a necessary intermediate for glycolysis to move forward. Here, ATP hydrolysis' exergonic reaction couples with the endergonic reaction of converting glucose into a phosphorylated intermediate in the pathway. Once again, the energy released by breaking a phosphate bond within ATP was used for phosphorylyzing another molecule, creating an unstable intermediate and powering an important conformational change.

LINK TO LEARNING

See an interactive animation of the ATP-producing glycolysis process at this site: http://www.science.smith.edu/departments/Biology/Bio231/glycolysis.html

Chapter Attributions

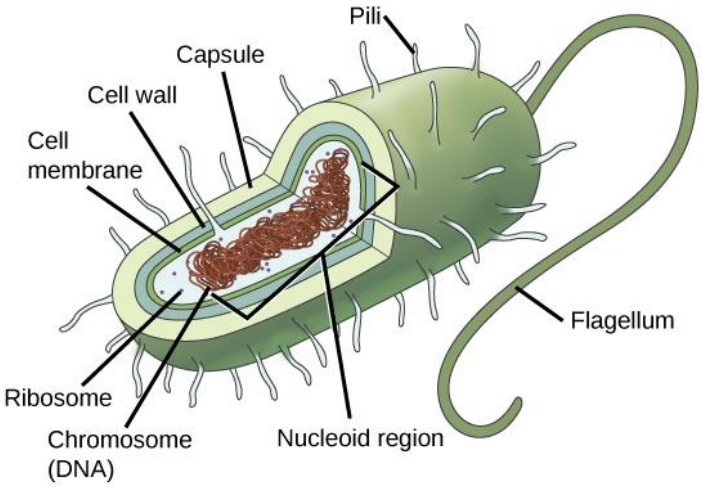
10. Cell Theory

Cells fall into one of two broad categories: prokaryotic and eukaryotic. The predominantly single-celled organisms of the domains Bacteria and Archaea are classified as prokaryotes (*pro-* = before; *-karyon-* = nucleus). Animal cells, plant cells, fungi, and protists are eukaryotes (*eu-* = true).

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 region within the cell in which other cellular components are found; 3) DNA, the genetic material of the cell; and 4) ribosomes, particles that synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

A **prokaryotic cell** is a simple, single-celled (unicellular) organism that lacks a nucleus, or any other membrane-bound organelle. We will shortly come to see that this is significantly different in eukaryotes. Prokaryotic DNA is found in the central part of the cell: a darkened region called the nucleoid (Figure).



This figure shows the generalized structure of a prokaryotic cell.

Eukaryotic Cells

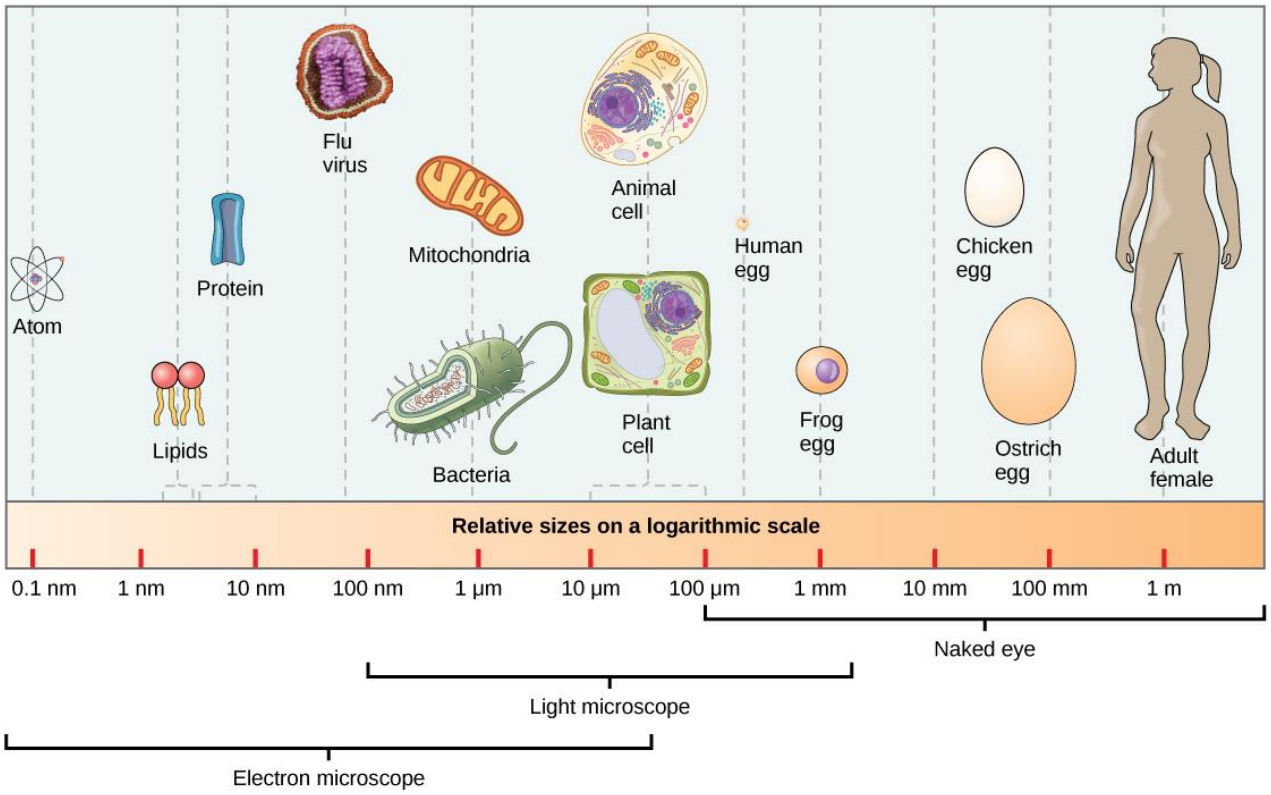
In nature, the relationship between form and function is apparent at all levels, including the level of the cell, and this will become clear as we explore eukaryotic cells. The principle “form follows function” is found in many contexts. For example, birds and fish have streamlined bodies that allow them to move quickly through the medium in which they live, be it air or water. It means that, in general, one can deduce the function of a structure by looking at its form, because the two are matched.

A **eukaryotic cell** is a cell that has a membrane-bound nucleus and other membrane-bound compartments or sacs, called **organelles**, which have specialized functions. The word eukaryotic means “true kernel” or “true nucleus,” alluding to the presence of the membrane-bound nucleus in these cells. The word “organelle” means “little organ,” and, as already mentioned, organelles have specialized cellular functions, just as the organs of your body have specialized functions.

Cell Size

At 0.1–5.0 micrometers (µm; 1/1,000,000 of a meter) in diameter, prokaryotic cells are significantly smaller than eukaryotic cells, which have diameters ranging from 10–100 µm (Figure). The small size of prokaryotes allows ions and organic molecules that enter them to

quickly spread to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly move out. However, larger eukaryotic cells have evolved different structural adaptations to enhance cellular transport. Indeed, the large size of these cells would not be possible without these adaptations. In general, cell size is limited because volume increases much more quickly than does cell surface area. This is because volume is a cubic dimension and surface area is a squared dimension. For example, if X=2, then the surface area (x squared) is 4 and the volume (x cubed) is 8. If x =3, then the surface area is 9 and the volume is 27. As a cell becomes larger, it becomes more and more difficult for the cell to acquire sufficient materials to support the processes inside the cell, because the relative size of the surface area across which materials must be transported declines.



This figure shows the relative sizes of different kinds of cells and cellular components. An adult human is shown for comparison. Note that a light microscope is required to view both prokaryotic and eukaryotic cells. Note: 1000 nanometers equals 1 micrometer, 1000 micrometers equals 1 millimeter, and 1000 millimeters equals one meter.

Section Summary

Prokaryotes are predominantly single-celled organisms of the domains Bacteria and Archaea. All prokaryotes have plasma membranes, cytoplasm, ribosomes, a cell wall, DNA, and lack membrane-bound organelles. Many also have polysaccharide capsules. Prokaryotic cells range in diameter from 0.1–5.0 µm.

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. Eukaryotic cells tend to be 10 to 100 times the size of prokaryotic cells.

Multiple Choice

Which of these do all prokaryotes and eukaryotes share?

a. nuclear envelope

b. cell walls

c. organelles

d. plasma membrane

A typical prokaryotic cell _____ compared to a eukaryotic cell.

a. is smaller in size by a factor of 100

b. is similar in size

c. is smaller in size by a factor of one million

d. is larger in size by a factor of 10

Free Response

What are organelles and which type of cells (prokaryotic or eukaryotic) contains them?

Glossary

eukaryotic cell. a cell that has a membrane-bound nucleus and several other membrane-bound compartments or sacs

organelle. a membrane-bound compartment or sac within a cell

prokaryotic cell. a unicellular organism that lacks a nucleus or any other membrane-bound organelle

Chapter Attributions

11. Cell Membranes and Organelles

Components and Structure

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

- Understand the cell membrane fluid mosaic model
- Describe phospholipid, protein, and carbohydrate functions 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 5.1 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 and white blood cells, to change shape as they pass through narrow capillaries. These are the more obvious plasma membrane functions. In addition, the plasma membrane's surface 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 immune response's "self" versus "non-self" distinction.

Among the most sophisticated plasma membrane functions is the ability for complex, integral proteins, receptors to transmit signals. These proteins act both as extracellular input receivers

and as intracellular processing activators. 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, viruses hijack receptors (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 signal transduction process to malfunction with disastrous consequences.

Fluid Mosaic Model

Scientists identified the plasma membrane in the 1890s, and its chemical components in 1915. The principal components they identified were lipids and proteins. In 1935, Hugh Davson and James Danielli proposed the plasma membrane's structure. This was the first model that others in the scientific community widely accepted. It was based on the plasma membrane's "railroad track" appearance in early electron micrographs. Davson and Danielli theorized that the plasma membrane's structure resembles a sandwich. They made the analogy of proteins to bread, and lipids to the filling. In the 1950s, advances in microscopy, notably transmission electron microscopy (TEM), allowed researchers to see that the plasma membrane's core consisted of a double, rather than a single, layer. In 1972, S.J. Singer and Garth L. Nicolson proposed a new model that provides microscopic observations and better explains plasma membrane function.

The explanation, the **fluid mosaic model**, has evolved somewhat over time, but it still best accounts for plasma membrane structure and function as we now understand them. The fluid mosaic model describes the plasma membrane structure 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 5.2).

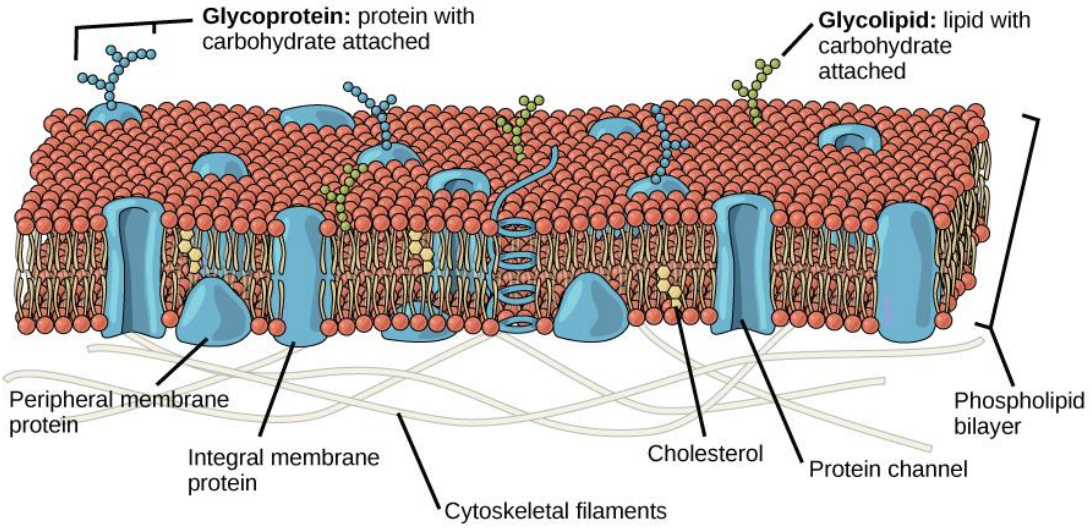


Figure 5.2 The plasma membrane fluid mosaic model describes the plasma membrane as a fluid combination of phospholipids, cholesterol, and proteins. Carbohydrates attached to lipids (glycolipids) and to proteins (glycoproteins) extend from the membrane's outward-facing surface.

A plasma membrane's principal components are lipids (phospholipids and cholesterol), proteins, and carbohydrates attached to some of the lipids and proteins. A phospholipid is a molecule consisting of glycerol, two fatty acids, and a phosphate-linked head group. Cholesterol, another lipid comprised of four fused carbon rings, is situated alongside the phospholipids in the membrane's core. The protein, lipid, and carbohydrate proportions 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 (of all types) account for about 40 percent, and carbohydrates comprise the remaining 10 percent. However, protein and lipid concentration varies with different cell membranes. For example, myelin, an outgrowth of specialized cells' membrane that insulates the peripheral nerves' axons, contains only 18 percent protein and 76 percent lipid. The mitochondrial inner membrane contains 76 percent protein and only 24 percent lipid. The plasma membrane of human red blood cells is 30 percent lipid. Carbohydrates are present only on the plasma membrane's exterior surface and are attached to proteins, forming **glycoproteins**, or attached to lipids, forming **glycolipids**.

Phospholipids

The membrane's main fabric comprises amphiphilic, phospholipid molecules.
The **hydrophilic** or “water-loving” areas of these molecules (which look like a collection of balls in an artist’s rendition of the model) (Figure 5.2) are in contact with the aqueous fluid both inside and outside the cell. **Hydrophobic**, or water-hating molecules, tend to be non-polar. They interact with other non-polar molecules in chemical reactions, but generally do not interact with polar molecules. When placed in water, hydrophobic molecules tend to form a ball or cluster. The phospholipids' hydrophilic regions form hydrogen bonds with water and other polar molecules on both the cell's exterior and interior. Thus, the membrane surfaces that face the cell's interior and exterior are hydrophilic. In contrast, the cell membrane's interior is hydrophobic and will not interact with water. Therefore, phospholipids form an excellent two-layer cell membrane that separates fluid within the cell from the fluid outside the cell.

A phospholipid molecule (Figure 5.3) consists of a three-carbon glycerol backbone with two fatty acid molecules attached to carbons 1 and 2, and a phosphate-containing group attached to the third carbon. This arrangement gives the overall molecule a head area (the phosphate-containing group), which has a polar character or negative charge, and a tail area (the fatty acids), which has no charge. The head can form hydrogen bonds, but the tail cannot. Scientists call a molecule with a positively or negatively charged area and an uncharged, or non-polar, area **amphiphilic** or “dual-loving.”

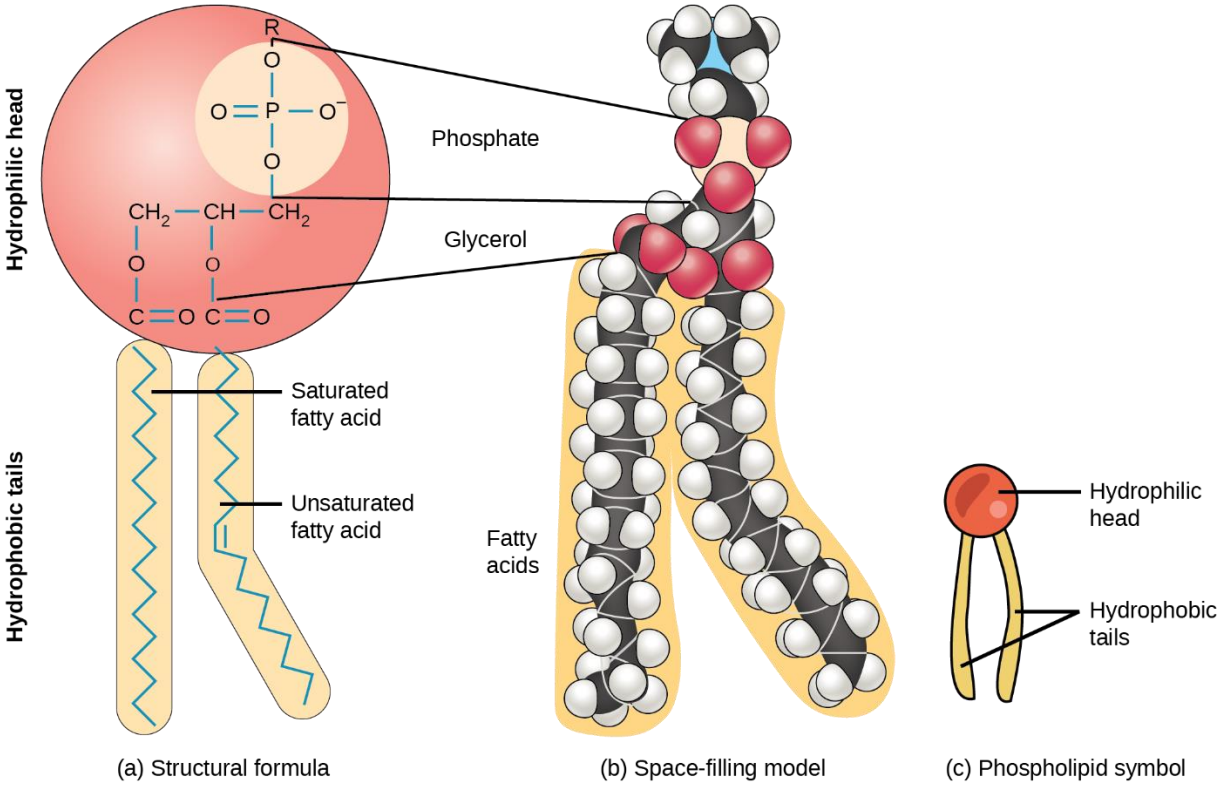


Figure 5.3 A hydrophilic head and two hydrophobic tails comprise this phospholipid molecule. The hydrophilic head group consists of a phosphate-containing group attached to a glycerol molecule. The hydrophobic tails, each containing either a saturated or an unsaturated fatty acid, are long hydrocarbon chains.

This characteristic is vital to the plasma membrane's structure because, in water, phospholipids arrange themselves with their hydrophobic tails facing each other and their hydrophilic heads facing out. In this way, they form a lipid bilayer—a double layered phospholipid barrier that separates the water and other materials on one side from the water and other materials on the other side. Phosopholipids heated in an aqueous solution usually spontaneously form small spheres or droplets (micelles or liposomes), with their hydrophilic heads forming the exterior and their hydrophobic tails on the inside (Figure 5.4).

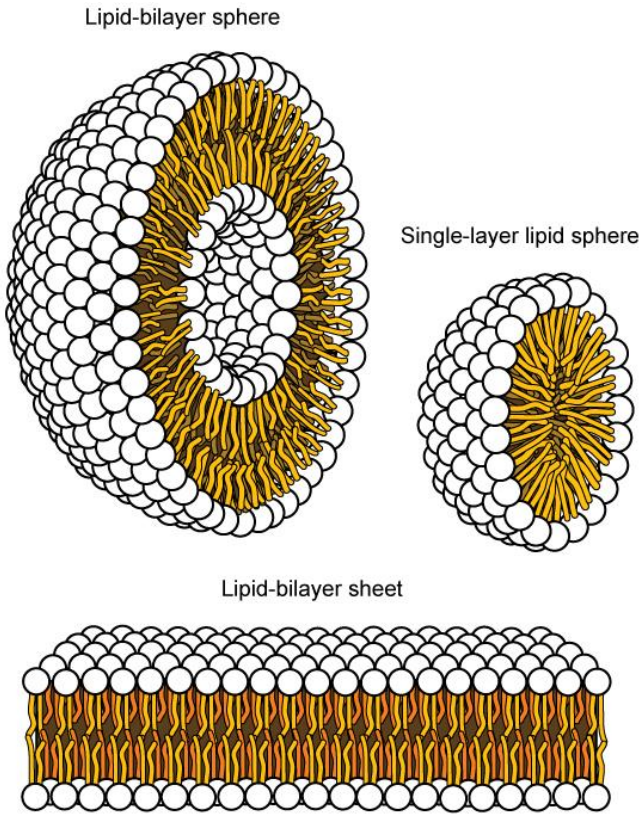


Figure 5.4 In an aqueous solution, phospholipids usually arrange themselves with their polar heads facing outward and their hydrophobic tails facing inward. (credit: modification of work by Mariana Ruiz Villareal)

Proteins

Proteins comprise the plasma membranes' second major component. **Integral proteins**, or integrins, as their name suggests, integrate completely into the membrane structure, and their hydrophobic membrane-spanning regions interact with the phospholipid bilayer's hydrophobic region (Figure 5.2). 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 to the other, and are exposed on either side. Up to 12 single protein segments comprise some complex proteins, which are extensively folded and embedded in the membrane (Figure 5.5). This protein type has a hydrophilic region or regions, and one or several mildly hydrophobic regions. This arrangement of protein regions orients the protein alongside the phospholipids, with the protein's hydrophobic region adjacent to the phosopholipids' tails and the protein's hydrophilic

region or regions protruding from the membrane and in contact with the cytosol or extracellular fluid.

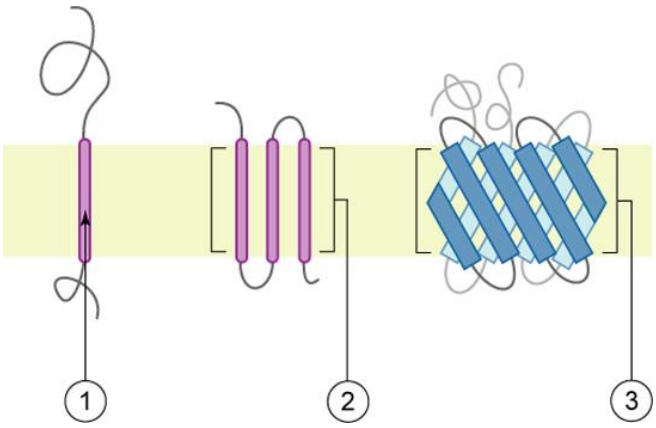


Figure 5.5 Integral membrane proteins may have one or more alpha-helices that span the membrane (examples 1 and 2), or they may have beta-sheets that span the membrane (example 3). (credit: “Foobar”/Wikimedia Commons)

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

Carbohydrates

Carbohydrates are the third major plasma membrane component. They are always on the cells' exterior surface and are bound either to proteins (forming glycoproteins) or to lipids (forming glycolipids) (Figure 5.2). 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 for cell recognition, much the way that the facial features unique to each person allow individuals to recognize him or her. This recognition function is very important to cells, as it allows the immune system to differentiate between body cells (“self”) and foreign cells or tissues (“non-self”). Similar glycoprotein and glycolipid types are on the

surfaces of viruses and may change frequently, preventing immune cells from recognizing and attacking them.

We collectively refer to these carbohydrates on the cell's exterior surface—the carbohydrate components of both glycoproteins and glycolipids—as the glycocalyx (meaning “sugar coating”). The glycocalyx is highly hydrophilic and attracts large amounts of water to the cell's surface. This aids in the cell's interaction with its watery environment and in the cell’s ability to obtain substances dissolved in the water. As we discussed above, the glycocalyx is also important for cell identification, self/non-self determination, and embryonic development, and is used in cell to cell attachments to form tissues.

EVOLUTION CONNECTION

How Viruses Infect Specific Organs

Glycoprotein and glycolipid patterns on the cells' surfaces 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 5.6). 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 virus' activity. Unfortunately, these recognition sites on HIV change at a rapid rate because of mutations, making 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 differences in these recognition sites distinguish. 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 surface patterns' new variations. In the case of HIV, the problem is compounded because the virus specifically infects and destroys cells involved in the immune response, further incapacitating the host.

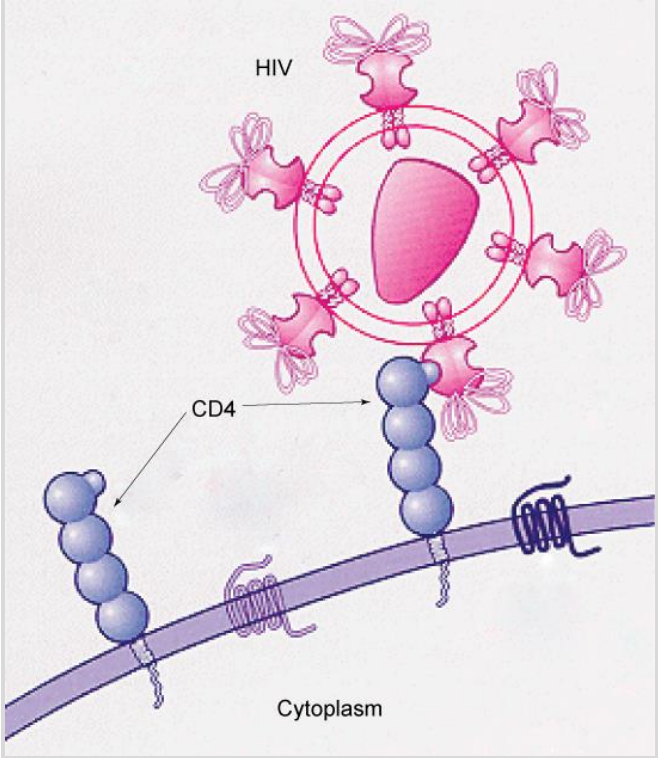


Figure 5.6 HIV binds to the CD4 receptor, a glycoprotein on T cell surfaces. (credit: modification of work by NIH, NIAID)

Membrane Fluidity

The membrane's mosaic characteristic 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 one extracts the needle.

The membrane's mosaic characteristics 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 carbon string of approximately 30 degrees (Figure 5.3).

Thus, if decreasing temperatures compress saturated fatty acids with their straight tails, 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 membrane's relative fluidity is particularly important in a cold environment. A cold environment usually compresses membranes comprised 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 lower temperature.

LINK TO LEARNING

Visit this site (https://www.youtube.com/watch?v=LKN5sq5dtW4) to see animations of the membranes' fluidity and mosaic quality.

Animals have an additional membrane constituent that assists in maintaining fluidity. Cholesterol, which lies alongside the phospholipids in the membrane, tends to dampen temperature effects 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 temperature range 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.

Plasma Membrane Components and Functions

Component	Location
Phospholipid	Main membrane fabric
Cholesterol	Attached between phospholipids and between the two p layers
Integral proteins (for example, integrins)	Embedded within the phospholipid layer(s); may or may penetrate through both layers
Peripheral proteins	On the phospholipid bilayer's inner or outer surface; not within the phospholipids
Carbohydrates (components of glycoproteins and glycolipids)	Generally attached to proteins on the outside membran

Table5.1

CAREER CONNECTION

Immunologist

The variations in peripheral proteins and carbohydrates that affect a cell's recognition sites are of prime interest in immunology. In developing vaccines, researchers have been able to conquer many infectious diseases, such as smallpox, polio, diphtheria, and tetanus.

Immunologists 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 also 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, researchers did not understand the importance of having a healthy immune system in preventing cancer.

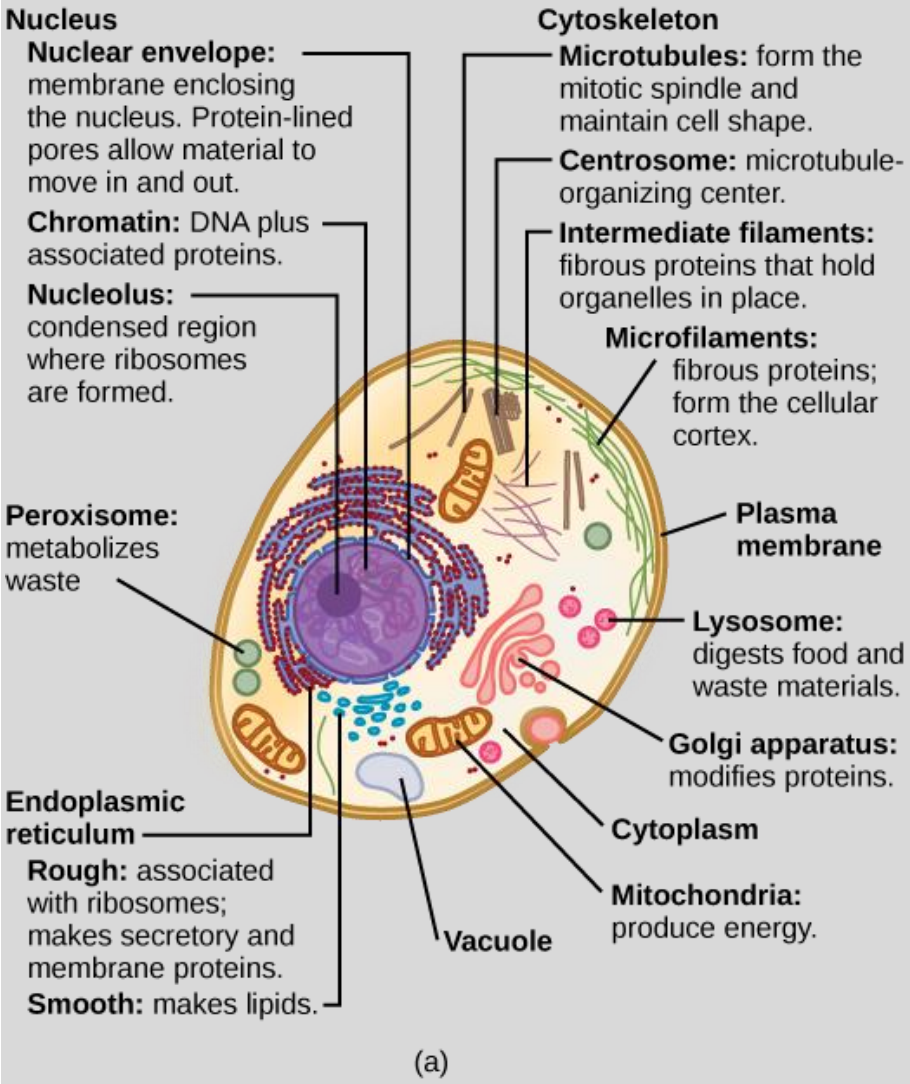
To work as an immunologist, one must have a PhD or MD. In addition, immunologists undertake at least two to three years of training in an accredited program and must pass the American Board of Allergy and Immunology exam. Immunologists must possess knowledge of the human body's function as they relate to issues beyond immunization, and knowledge of pharmacology and medical technology, such as medications, therapies, test materials, and surgical procedures.

Chapter Attributions

A More Detailed Look at Eukaryotic Cells

At this point, it should be clear that eukaryotic cells have a more complex structure than do prokaryotic cells. Organelles and other cellular components allow for various functions to occur in the cell at the same time. Before discussing the functions of organelles within a eukaryotic cell, let us first examine two important components of the cell: the plasma membrane and the cytoplasm.

ART CONNECTION

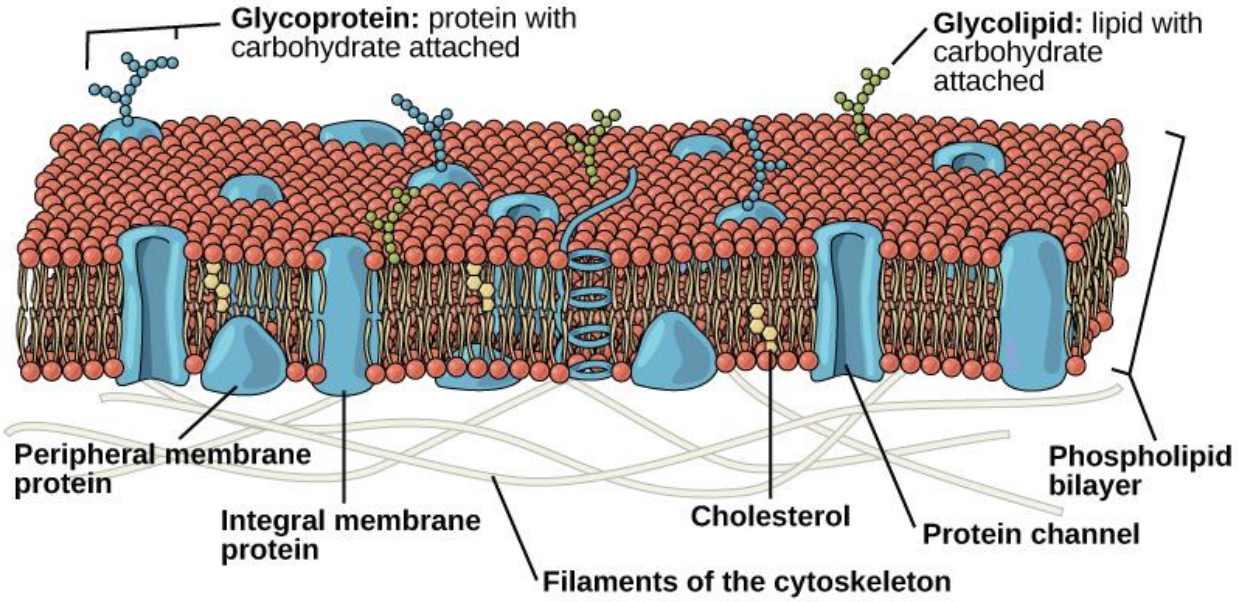


This figure shows a typical animal cell.

The Plasma Membrane

Like prokaryotes, eukaryotic cells have a **plasma membrane** (Figure) made up of a phospholipid bilayer with embedded proteins that separates the internal contents of the cell from its surrounding environment. A phospholipid is a lipid molecule composed of two fatty acid chains, a glycerol backbone, and a phosphate group. The plasma membrane regulates the passage of some substances, such as organic molecules, ions, and water, preventing the passage of some

to maintain internal conditions, while actively bringing in or removing others. Other compounds move passively across the membrane.



The plasma membrane is a phospholipid bilayer with embedded proteins. There are other components, such as cholesterol and carbohydrates, which can be found in the membrane in addition to phospholipids and protein. The phrase "fluid mosaic" is used to describe the structure of the plasma membrane because it is dynamic and contains numerous components.

The plasma membranes of cells that specialize in absorption are folded into fingerlike projections called microvilli (singular = microvillus). This folding increases the surface area of the plasma membrane. Such cells are typically found lining the small intestine, the organ that absorbs nutrients from digested food. This is an excellent example of form matching the function of a structure.

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, and thus, afflicted individuals cannot absorb nutrients. This leads to malnutrition, cramping, and diarrhea. Patients suffering from celiac disease must follow a gluten-free diet.

The Cytoplasm

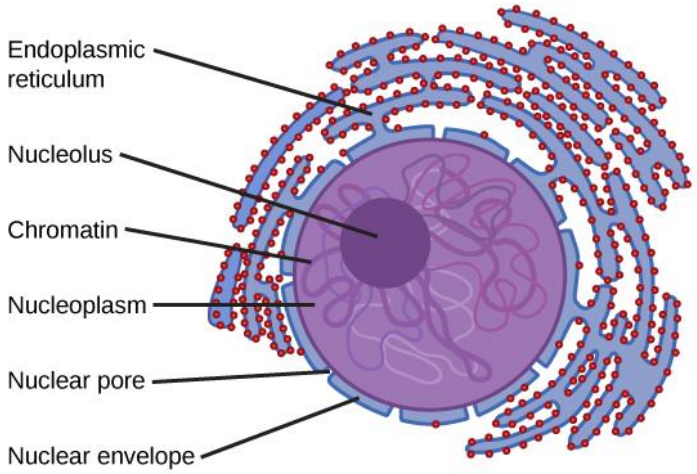
The **cytoplasm** comprises the contents of a cell between the plasma membrane and the nuclear envelope (a structure to be discussed shortly). It is made up of organelles suspended in the gel-like **cytosol**, the cytoskeleton, and various chemicals. Even though the cytoplasm consists of 70 to 80 percent water, it has a semi-solid consistency, which comes from the proteins within it. However, proteins are not the only organic molecules found in the cytoplasm. Glucose and other simple sugars, polysaccharides, amino acids, nucleic acids, fatty acids, and derivatives of glycerol are found there too. Ions of sodium, potassium, calcium, and many other elements are also dissolved in the cytoplasm. Many metabolic reactions, including protein synthesis, take place in the cytoplasm.

The Endomembrane System

The **endomembrane system** (*endo* = within) is a group of membranes and organelles (Figure) in eukaryotic cells that work together to modify, package, and transport lipids and proteins. It includes the nuclear envelope, lysosomes, and vesicles, the endoplasmic reticulum and Golgi apparatus, which we will cover shortly. Although not technically *within* the cell, the plasma membrane is included in the endomembrane system because, as you will see, it interacts with the other endomembranous organelles.

The Nucleus

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



The outermost boundary of the nucleus is the nuclear envelope. Notice that the nuclear envelope consists of two phospholipid bilayers (membranes)—an outer membrane and an inner membrane—in contrast to the plasma membrane (Figure), which consists of only one phospholipid bilayer. (credit: modification of work by NIGMS, NIH)

The **nuclear envelope** is a double-membrane structure that constitutes the outermost portion of the nucleus (Figure). 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 the cytoplasm. The DNA in the nucleus is too large to fit through the pores.

To understand chromatin, it is helpful to first consider chromosomes. Chromosomes are structures within the nucleus that are made up of DNA (the hereditary material) and proteins. This combination of DNA and proteins is called chromatin. In eukaryotes, chromosomes are linear structures. Every species has a specific number of chromosomes in the nucleus of its body cells. For example, in humans, the chromosome number is 46, whereas in fruit flies, the chromosome number is eight.

Chromosomes are only visible and distinguishable from one another when the cell is getting ready to divide. This is because the DNA condenses or compacts in preparation for cell division. When the cell is in the growth and maintenance phases of its life cycle, the chromosomes resemble an unwound, jumbled bunch of threads and the DNA is more accessible to be used to make proteins.

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. A darkly staining area within the nucleus, called the **nucleolus** (plural = nucleoli), aggregates the ribosomal RNA with associated proteins to assemble the ribosomal subunits that are then transported through the nuclear pores into the cytoplasm.

The Endoplasmic Reticulum

The **endoplasmic reticulum (ER)** (Figure) is a series of interconnected membranous tubules that collectively modify proteins and synthesize lipids. However, these two functions are performed in separate areas of the endoplasmic reticulum: the rough endoplasmic reticulum and the smooth endoplasmic reticulum, respectively.

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

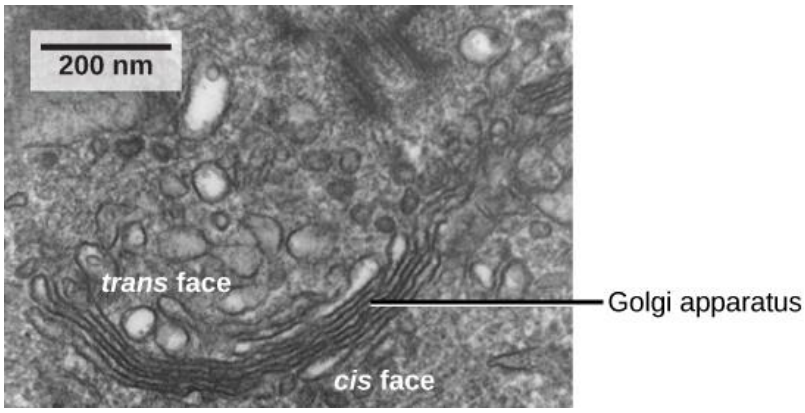
The ribosomes synthesize proteins while attached to the ER, resulting in transfer of their newly synthesized proteins into the lumen of the RER where they undergo modifications such as folding or addition of sugars. The RER also makes phospholipids for cell membranes.

If the phospholipids or modified proteins are not destined to stay in the RER, they will be packaged within vesicles and transported from the RER by budding from the membrane (Figure). Since the RER is engaged in modifying proteins that will be secreted from the cell, it is abundant in cells that secrete proteins, such as the liver.

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

The Golgi Apparatus

We have already mentioned that vesicles can bud from the ER, but where do the vesicles go? Before reaching their final destination, the lipids or proteins within the transport vesicles need to be sorted, packaged, and tagged so that they wind up in the right place. The sorting, tagging, packaging, and distribution of lipids and proteins take place in the **Golgi apparatus** (also called the Golgi body), a series of flattened membranous sacs (Figure).



The Golgi apparatus in this transmission electron micrograph of a white blood cell is visible as a stack of semicircular flattened rings in the lower portion of this image. Several vesicles can be seen near the Golgi apparatus. (credit: modification of work by Louisa Howard; scale-bar data from Matt Russell)

The Golgi apparatus has a receiving face near the endoplasmic reticulum and a releasing face on the side away from the ER, toward the cell membrane. The transport vesicles that form from the ER travel to the receiving face, fuse with it, and empty their contents into the lumen of the Golgi apparatus. As the proteins and lipids travel through the Golgi, they undergo further

modifications. The most frequent modification is the addition of short chains of sugar molecules. The newly modified proteins and lipids are then tagged with small molecular groups to enable them to be routed to their proper destinations.

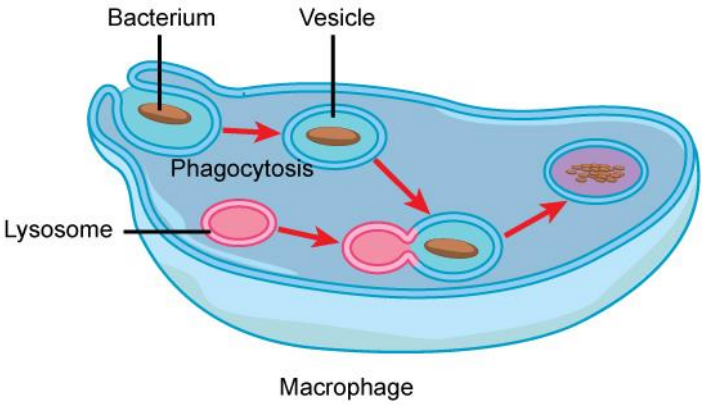
Finally, the modified and tagged proteins are packaged into vesicles that bud from the opposite face of the Golgi. While some of these vesicles, transport vesicles, deposit their contents into other parts of the cell where they will be used, others, secretory vesicles, fuse with the plasma membrane and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

The amount of Golgi in different cell types again illustrates that form follows function within cells. Cells that engage in a great deal of secretory activity (such as cells of the salivary glands that secrete digestive enzymes or cells of the immune system that secrete antibodies) have an abundant number of Golgi.

Lysosomes

In animal cells, the **lysosomes** are the cell's "garbage disposal." Digestive enzymes within the lysosomes aid the breakdown of proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles. In single-celled eukaryotes, lysosomes are important for digestion of the food they ingest and the recycling of organelles. These enzymes are active at a much lower pH (more acidic) than those located in the cytoplasm. Many reactions that take place in the cytoplasm could not occur at a low pH, thus the advantage of compartmentalizing the eukaryotic cell into organelles is apparent.

Lysosomes also use their hydrolytic enzymes to destroy disease-causing organisms that might enter the cell. A good example of this occurs in a group of white blood cells called macrophages, which are part of your body's immune system. In a process known as phagocytosis, a section of the plasma membrane of the macrophage invaginates (folds in) and engulfs a pathogen. The invaginated section, with the pathogen inside, then pinches itself off from the plasma membrane and becomes a vesicle. The vesicle fuses with a lysosome. The lysosome's hydrolytic enzymes then destroy the pathogen (Figure).

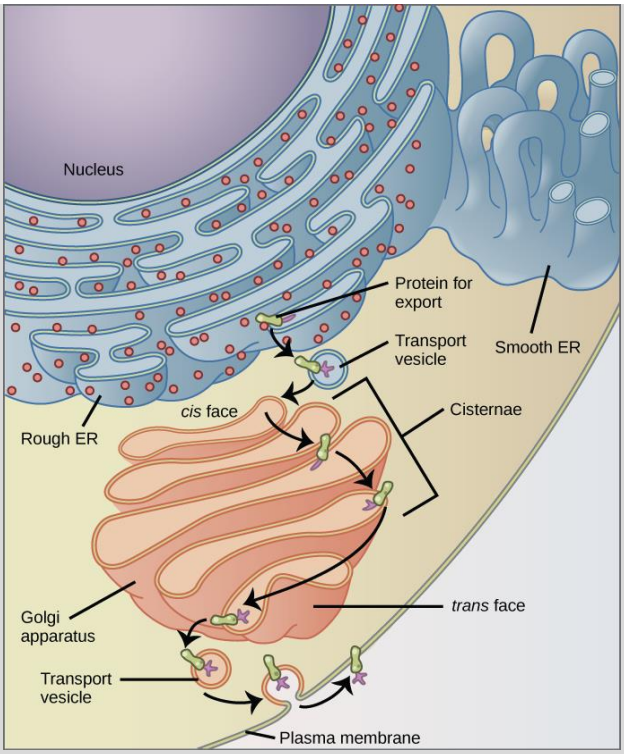


A macrophage has phagocytized a potentially pathogenic bacterium into a vesicle, which then fuses with a lysosome within the cell so that the pathogen can be destroyed. Other organelles are present in the cell, but for simplicity, are not shown.

Vesicles

Vesicles are membrane-bound sacs that function in storage and transport. Vesicles can fuse with other membranes within the cell system.

ART CONNECTION



The endomembrane system works to modify, package, and transport lipids and proteins. (credit: modification of work by Magnus Manske)

Ribosomes

Ribosomes are the cellular structures responsible for protein synthesis. When viewed through an electron microscope, free ribosomes appear as either clusters or single tiny dots floating freely in the cytoplasm. Ribosomes may be attached to either the cytoplasmic side of the plasma membrane or the cytoplasmic side of the endoplasmic reticulum. Electron microscopy has shown that ribosomes consist of large and small subunits. Ribosomes are enzyme complexes that are responsible for protein synthesis.

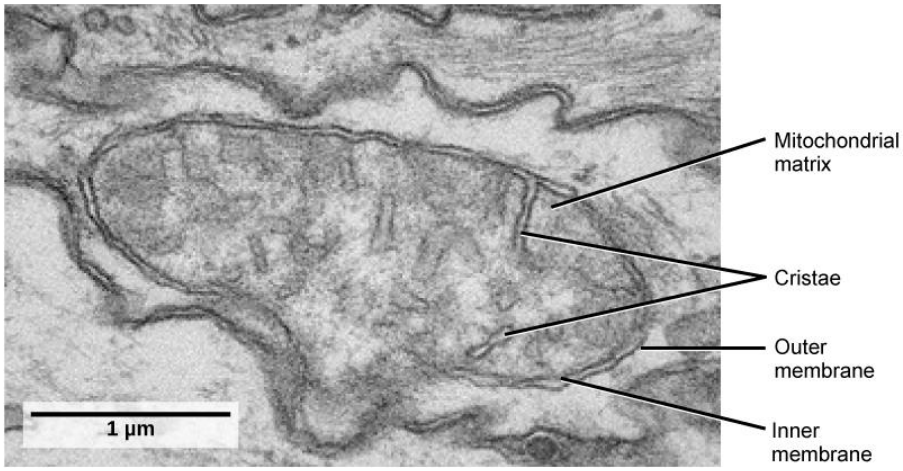
Because protein synthesis is essential for all cells, ribosomes are found in practically every cell, although they are smaller in prokaryotic cells. They are particularly abundant in immature red

blood cells for the synthesis of hemoglobin, which functions in the transport of oxygen throughout the body.

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. The formation of ATP from the breakdown of glucose is known as cellular respiration. Mitochondria are oval-shaped, double-membrane organelles (Figure) that have their own ribosomes and DNA. Each membrane is a phospholipid bilayer embedded with proteins. The inner layer has folds called cristae, which increase the surface area of the inner membrane. The area surrounded by the folds is called the mitochondrial matrix. The cristae and the matrix have different roles in cellular respiration.

In keeping with our theme of form following function, it is important to point out that muscle cells have a very high concentration of mitochondria because muscle cells need a lot of energy to contract.



This transmission electron micrograph shows a mitochondrion as viewed with an electron microscope. Notice the inner and outer membranes, the cristae, and the mitochondrial matrix. (credit: modification of work by Matthew Britton; scale-bar data from Matt Russell)

Section Summary

Like a prokaryotic cell, a eukaryotic cell has a plasma membrane, cytoplasm, and ribosomes, but a eukaryotic cell is typically larger than a prokaryotic cell, has a true nucleus (meaning its DNA is surrounded by a membrane), and has other membrane-bound organelles that allow for compartmentalization of functions. The plasma membrane is a phospholipid bilayer embedded with proteins. The nucleolus within the nucleus is the site for ribosome assembly. Ribosomes are found in the cytoplasm or are attached to the cytoplasmic side of the plasma membrane or endoplasmic reticulum. They perform protein synthesis. Mitochondria perform cellular respiration and produce ATP. Vesicles are storage and transport compartments.

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

Art Connections

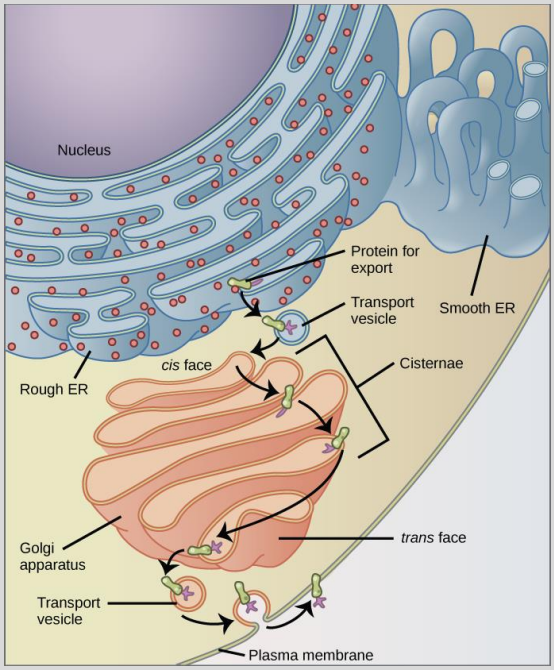


Figure Why does the *cis* face of the Golgi not face the plasma membrane?

Multiple Choice

Which of the following is found both in eukaryotic and prokaryotic cells?

a. nucleus

b. mitochondrion

c. vessicle

d. ribosome

Which of the following is not a component of the endomembrane system?

a. mitochondrion

b. Golgi apparatus

c. endoplasmic reticulum

d. lysosome

Free Response

In the context of cell biology, what do we mean by form follows function? What are at least two examples of this concept?

Glossary

cell wall. a rigid cell covering made of cellulose in plants, peptidoglycan in bacteria, non-peptidoglycan compounds in Archaea, and chitin in fungi that protects the cell, provides structural support, and gives shape to the cell

central vacuole. a large plant cell organelle that acts as a storage compartment, water reservoir, and site of macromolecule degradation

chloroplast. a plant cell organelle that carries out photosynthesis

cilium. (plural: cilia) a short, hair-like structure that extends from the plasma membrane in large numbers and is used to move an entire cell or move substances along the outer surface of the cell

cytoplasm. the entire region between the plasma membrane and the nuclear envelope, consisting of organelles suspended in the gel-like cytosol, the cytoskeleton, and various chemicals

cytoskeleton. the network of protein fibers that collectively maintains the shape of the cell, secures some organelles in specific positions, allows cytoplasm and vesicles to move within the cell, and enables unicellular organisms to move

cytosol. the gel-like material of the cytoplasm in which cell structures are suspended

desmosome. a linkage between adjacent epithelial cells that forms when cadherins in the plasma membrane attach to intermediate filaments

endomembrane system. the group of organelles and membranes in eukaryotic cells that work together to modify, package, and transport lipids and proteins

endoplasmic reticulum (ER). a series of interconnected membranous structures within eukaryotic cells that collectively modify proteins and synthesize lipids

extracellular matrix. the material, primarily collagen, glycoproteins, and proteoglycans, secreted from animal cells that holds cells together as a tissue, allows cells to communicate with each other, and provides mechanical protection and anchoring for cells in the tissue

flagellum. (plural: flagella) the long, hair-like structure that extends from the plasma membrane and is used to move the cell

gap junction. a channel between two adjacent animal cells that allows ions, nutrients, and other low-molecular weight substances to pass between the cells, enabling the cells to communicate

Golgi apparatus. a eukaryotic organelle made up of a series of stacked membranes that sorts, tags, and packages lipids and proteins for distribution

lysosome. an organelle in an animal cell that functions as the cell's digestive component; it breaks down proteins, polysaccharides, lipids, nucleic acids, and even worn-out organelles

mitochondria. (singular: mitochondrion) the cellular organelles responsible for carrying out cellular respiration, resulting in the production of ATP, the cell's main energy-carrying molecule

nuclear envelope. the double-membrane structure that constitutes the outermost portion of the nucleus

nucleolus. the darkly staining body within the nucleus that is responsible for assembling ribosomal subunits

nucleus. the cell organelle that houses the cell's DNA and directs the synthesis of ribosomes and proteins

peroxisome. a small, round organelle that contains hydrogen peroxide, oxidizes fatty acids and amino acids, and detoxifies many poisons

plasma membrane. a phospholipid bilayer with embedded (integral) or attached (peripheral) proteins that separates the internal contents of the cell from its surrounding environment

plasmodesma. (plural: plasmodesmata) a channel that passes between the cell walls of adjacent plant cells, connects their cytoplasm, and allows materials to be transported from cell to cell

ribosome. a cellular structure that carries out protein synthesis

rough endoplasmic reticulum (RER). the region of the endoplasmic reticulum that is studded with ribosomes and engages in protein modification

smooth endoplasmic reticulum (SER). the region of the endoplasmic reticulum that has few or no ribosomes on its cytoplasmic surface and synthesizes carbohydrates, lipids, and steroid hormones; detoxifies chemicals like pesticides, preservatives, medications, and environmental pollutants, and stores calcium ions

tight junction. a firm seal between two adjacent animal cells created by protein adherence

vacuole. a membrane-bound sac, somewhat larger than a vesicle, that functions in cellular storage and transport

vesicle. a small, membrane-bound sac that functions in cellular storage and transport; its membrane is capable of fusing with the plasma membrane and the membranes of the endoplasmic reticulum and Golgi apparatus

Chapter Attributions

12. Membrane Permeability, Transport and Osmosis

Passive Transport

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

- Explain why and how passive transport occurs
- Understand the osmosis and diffusion processes
- Define tonicity and 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. 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. Red blood cells use some of their energy doing just that. Most cells spend the majority of their energy to maintain an imbalance of sodium and potassium ions between the cell's interior and exterior, as well as on protein synthesis.

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 single substance concentration range has a **concentration gradient**.

Selective Permeability

Plasma membranes are asymmetric: the membrane's interior is not identical to its exterior. There is a considerable difference between the array of phospholipids and proteins between the two leaflets that form a membrane. On the membrane's interior, some proteins serve to anchor the membrane to cytoskeleton's fibers. There are peripheral proteins on the membrane's exterior that bind extracellular matrix elements. Carbohydrates, attached to lipids or proteins, are also on the plasma membrane's exterior surface. These carbohydrate complexes help the cell bind required substances in the extracellular fluid. This adds considerably to plasma membrane's selective nature (Figure 5.7).

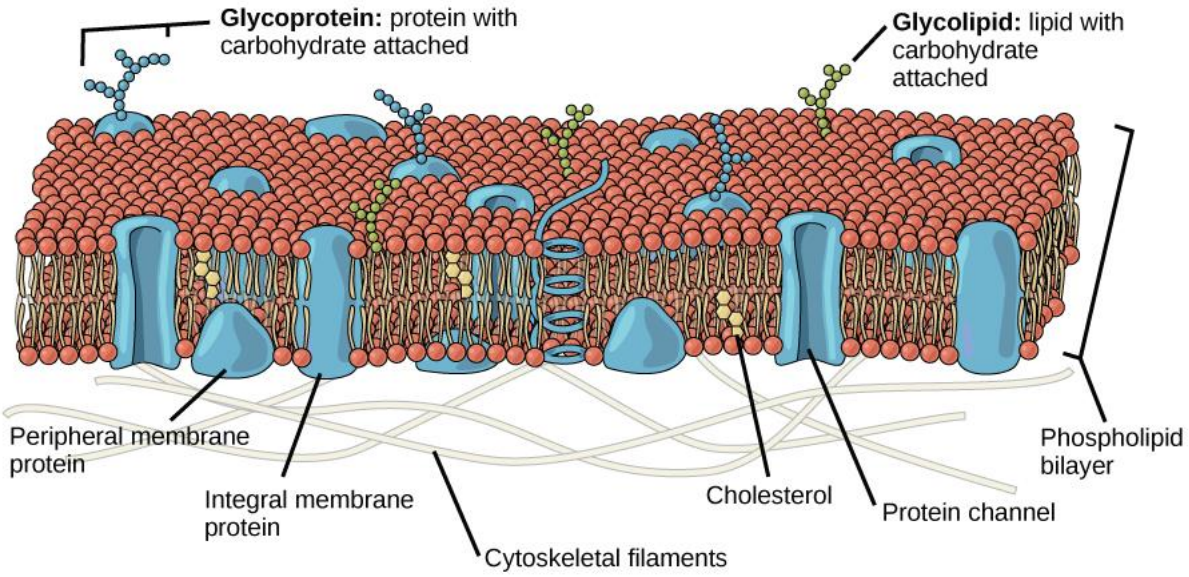


Figure 5.7 The plasma membrane's exterior surface is not identical to its interior surface.

Recall that plasma membranes are amphiphilic: They have hydrophilic and hydrophobic regions. This characteristic helps move some materials through the membrane and hinders the movement of others. Non-polar and lipid-soluble material with a low molecular weight can easily slip through the membrane's hydrophobic lipid core. 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 readily transport themselves into the body's tissues and organs. Oxygen and carbon dioxide molecules have no charge and pass through membranes by simple diffusion.

Polar substances present problems for the membrane. While some polar molecules connect easily with the cell's outside, they cannot readily pass through the plasma membrane's lipid core. Additionally, while small ions could easily slip through the spaces in the membrane's mosaic, their charge prevents them from doing so. Ions such as sodium, potassium, calcium, and chloride must have special means of penetrating plasma membranes. Simple sugars and amino acids also need the help of various transmembrane proteins (channels) to transport themselves across plasma membranes.

Diffusion

Diffusion is a passive process of transport. A single substance moves from a high concentration to a low concentration area 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 room's edges. The ammonia vapor will diffuse, or spread away, from the bottle, and gradually, increasingly 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 5.8). Diffusion expends no energy. On the contrary, concentration gradients are a form of potential energy, which dissipates as the gradient is eliminated.

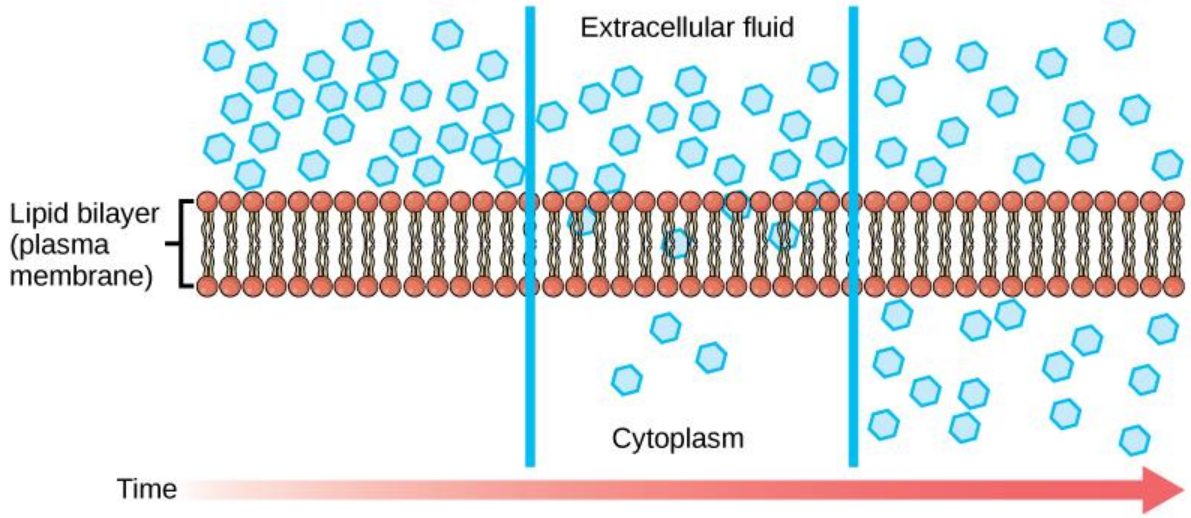


Figure 5.8 Diffusion through a permeable membrane moves a substance from a high concentration area (extracellular fluid, in this case) down its concentration gradient (into the cytoplasm). (credit: modification of work by Mariana Ruiz Villareal)

Each separate substance in a medium, such as the extracellular fluid, has its own concentration gradient, independent of other materials' concentration gradients. In addition, each substance will diffuse according to that gradient. Within a system, there will be different diffusion rates of various 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 molecule diffusion through whatever medium in which they are localized. A substance moves into any space available to it until it evenly distributes itself throughout. 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. We call this lack of a concentration gradient in which the substance has no net movement dynamic equilibrium. While diffusion will go forward in the presence of a substance's concentration gradient, several factors affect the diffusion rate.

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

A variation of diffusion is the process of filtration. In filtration, material moves according to its concentration gradient through a membrane. Sometimes pressure enhances the diffusion rate,

causing the substances to filter more rapidly. This occurs in the kidney, where blood pressure forces large amounts of water and accompanying dissolved substances, or **solutes**, out of the blood and into the renal tubules. The diffusion rate in this instance is almost totally dependent on pressure. One of the effects of high blood pressure is the appearance of protein in the urine, which abnormally high pressure "squeezes through".

Facilitated transport

In **facilitated transport**, or 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 polar molecule ions that the cell membrane's hydrophobic parts repel. Facilitated transport proteins shield these materials from the membrane's repulsive force, allowing them to diffuse into the cell.

The transported material first attaches to protein or glycoprotein receptors on the plasma membrane's exterior surface. This allows removal of material from the extracellular fluid that the cell needs. The substances then pass to specific integral proteins that facilitate their passage. Some of these integral proteins are collections of beta-pleated sheets that form a pore or channel through the phospholipid bilayer. Others are carrier proteins which bind with the substance and aid its diffusion through the membrane.

Channels

The integral proteins involved in facilitated transport are **transport proteins**, and they function as either channels for the material or carriers. In both cases, they are transmembrane proteins. Channels are specific for the transported 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 (Figure 5.9). Passage through the channel allows polar compounds to avoid the plasma membrane's nonpolar central layer 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.

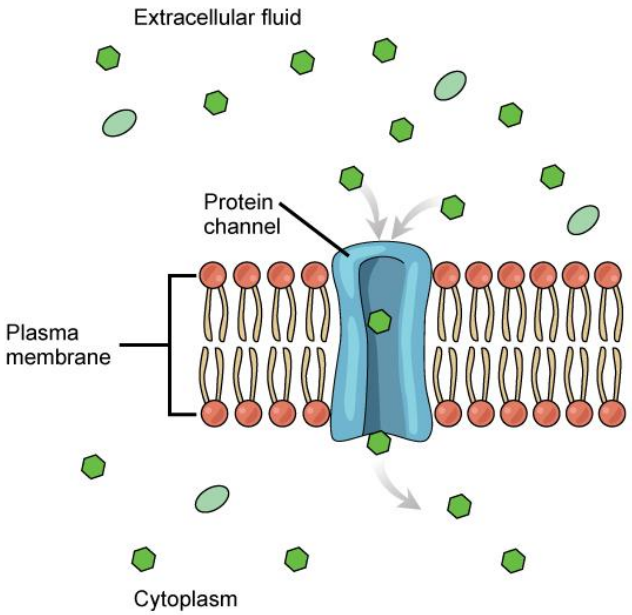


Figure 5.9 Facilitated transport moves substances down their concentration gradients. They may cross the plasma membrane with the aid of channel proteins. (credit: modification of work by Mariana Ruiz Villareal)

Channel proteins are either open at all times or they are “gated,” which controls the channel's opening. When a particular ion attaches to the channel protein it may control the opening, or other mechanisms or substances may be involved. In some tissues, sodium and chloride ions pass freely through open channels; whereas, in other tissues a gate must open to allow passage. An example of this occurs in the kidney, where there are both channel forms in different parts of the renal tubules. Cells involved in transmitting electrical impulses, such as nerve and muscle cells, have gated channels for sodium, potassium, and calcium in their membranes. Opening and closing these channels changes the relative concentrations on opposing sides of the membrane of these ions, resulting in facilitating electrical transmission along membranes (in the case of nerve cells) or in muscle contraction (in the case of muscle cells).

Carrier Proteins

Another type of protein embedded in the plasma membrane is a **carrier protein**. This aptly named protein binds a substance and, thus triggers a change of its own shape, moving the bound molecule from the cell's outside to its interior (Figure 5.10). Depending on the gradient,

the material may move in the opposite direction. Carrier proteins are typically specific for a single substance. This selectivity adds to the plasma membrane's overall selectivity. Scientists poorly understand the exact mechanism for the change of shape. Proteins can change shape when their hydrogen bonds are affected, but this may not fully explain this mechanism. Each carrier protein is specific to one substance, and there are a finite number of these proteins in any membrane. This can cause problems in transporting enough 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 transport rate.

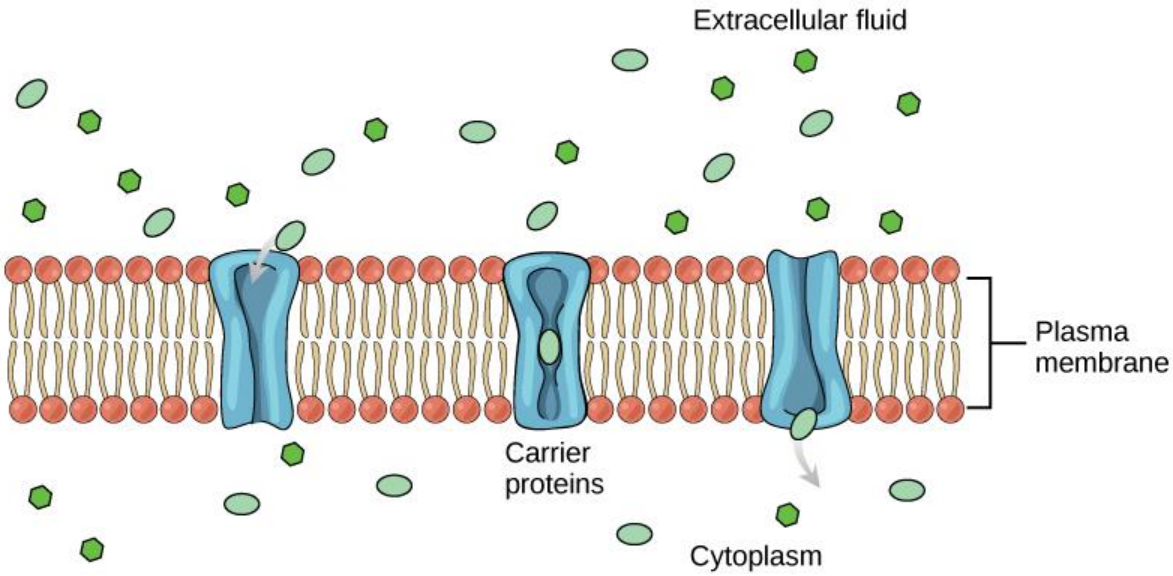


Figure 5.10 Some substances are able to move down their concentration gradient across the plasma membrane with the aid of carrier proteins. Carrier proteins change shape as they move molecules across the membrane. (credit: modification of work by Mariana Ruiz Villareal)

An example of this process occurs in the kidney. In one part, the kidney filters glucose, water, salts, ions, and amino acids that the body requires. This filtrate, which includes glucose, then reabsorbs 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 the body excretes this through urine. In a diabetic individual, the term is “spilling glucose into the urine.” A different group of carrier proteins, glucose transport

proteins, or GLUTs, are involved in transporting glucose and other hexose sugars through plasma membranes within the body.

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

Osmosis

Osmosis is the movement of water through a semipermeable membrane according to the water's concentration gradient across the membrane, which is inversely proportional to the solutes' concentration. While diffusion transports material across membranes and within cells, osmosis transports *only water* across a membrane and the membrane limits the solutes' diffusion in the water. Not surprisingly, the aquaporins that facilitate water movement play a large role in osmosis, most prominently in red blood cells and the membranes of kidney tubules.

Mechanism

Osmosis is a special case of diffusion. Water, like other substances, moves from an area of high concentration to one of low concentration. An obvious question is what makes water move at all? Imagine a beaker with a semipermeable membrane separating the two sides or halves (Figure 5.11). On both sides of the membrane the water level is the same, but there are different dissolved substance concentrations, or **solute**, that cannot cross the membrane (otherwise the solute crossing the membrane would balance concentrations on each side). If the solution's volume on both sides of the membrane is the same, but the solute's concentrations are different, then there are different amounts of water, the solvent, on either side of the membrane.

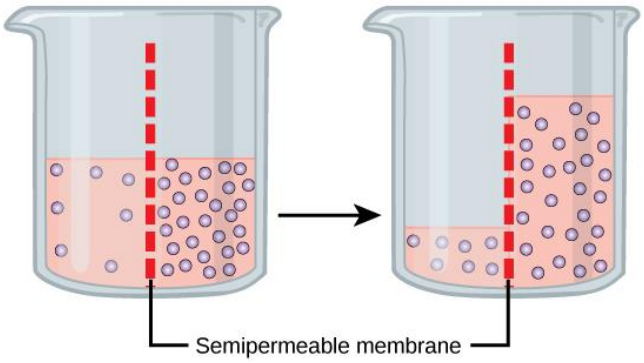


Figure 5.11 In osmosis, water always moves from an area of higher water concentration to one of lower concentration. In the diagram, the solute cannot pass through the selectively permeable membrane, but the water can.

To illustrate this, imagine two full water glasses. 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 sugar amount 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 solute mixture 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 water's concentration gradient goes to zero or until the water's hydrostatic pressure balances the osmotic pressure. Osmosis proceeds constantly in living systems.

Tonicity

Tonicity describes how an extracellular solution can change a cell's volume by affecting osmosis. A solution's tonicity often directly correlates with the solution's osmolarity. **Osmolarity** describes the solution's total solute concentration. 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 a membrane permeable to water, though not to the solute separates two different osmolarities, water will move from the membrane's side with lower osmolarity (and more water) to the side with higher osmolarity (and less water). This effect makes sense if you remember that the solute cannot move across the membrane, and thus the only component in the system that can move—the water—moves along its own concentration gradient. An important distinction that concerns living systems is that osmolarity measures the number of particles (which may be molecules) in a solution. Therefore, a solution that is cloudy with cells may have a lower osmolarity than a solution that is clear, if the second solution contains more dissolved molecules than there are cells.

Hypotonic Solutions

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

Hypertonic Solutions

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

Isotonic Solutions

In an **isotonic** solution, the extracellular fluid has the same osmolarity as the cell. If the cell's osmolarity matches that of the extracellular fluid, there will be no net movement of water into or out of the cell, although water will still move in and out. Blood cells and plant cells in hypertonic, isotonic, and hypotonic solutions take on characteristic appearances (Figure 5.12).

VISUAL CONNECTION

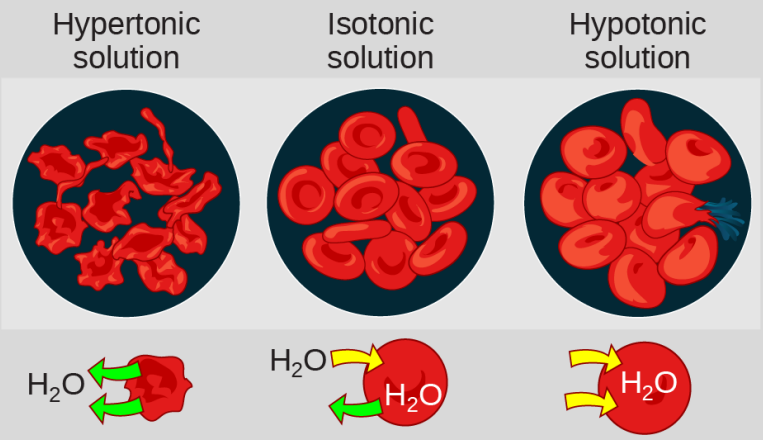


Figure 5.12 Osmotic pressure changes red blood cells' shape in hypertonic, isotonic, and hypotonic solutions. (credit: Mariana Ruiz Villareal)

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?

LINK TO LEARNING

For a video illustrating the diffusion process in solutions, visit this site: https://commons.wikimedia.org/wiki/File:Dispersion.gif

Tonicity in Living Systems

In a hypotonic environment, water enters a cell, and the cell swells. In an isotonic condition, the relative solute and solvent concentrations are equal on both membrane sides. There is no net water movement; therefore, there is no change in the cell's size. In a hypertonic solution, water leaves a cell and the cell shrinks. If either the hypo- or hyper- condition goes to excess, the cell’s functions become compromised, and the cell may be destroyed.

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

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

Various living things have ways of controlling the effects of osmosis—a mechanism we call osmoregulation. Some organisms, such as plants, fungi, bacteria, and some protists, have cell walls that surround the plasma membrane and prevent cell lysis in a hypotonic solution. The plasma membrane can only expand to the cell wall's limit, so the cell will not lyse. The cytoplasm in plants is always slightly hypertonic to the cellular environment, and water will always enter a cell if water is available. This water inflow produces turgor pressure, which stiffens the plant's cell walls (Figure 5.13). In nonwoody plants, turgor pressure supports the plant. Conversly, if you do not water the plant, the extracellular fluid will become hypertonic, causing water to leave the cell. In this condition, the cell does not shrink because the cell wall is not flexible. However, the cell membrane detaches from the wall and constricts the cytoplasm. We call this **plasmolysis**. Plants lose turgor pressure in this condition and wilt (Figure 5.14).

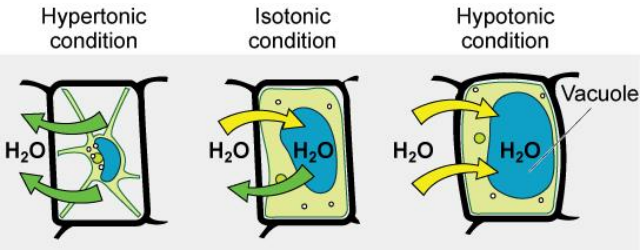


Figure 5.13 The turgor pressure within a plant cell depends on the solution's tonicity in which it is bathed. (credit: modification of work by Mariana Ruiz Villareal)

Figure 5.14 Without adequate water, the plant on the left has lost turgor pressure, visible in its wilting. Watering the plant (right) will restore the turgor pressure. (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 5.15).

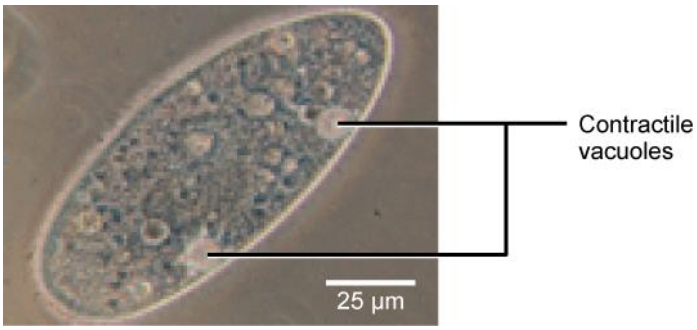


Figure 5.15 A paramecium's contractile vacuole, here visualized using bright field light microscopy at 480x magnification, continuously pumps water out of the organism's body to keep it from bursting in a hypotonic medium. (credit: modification of work by NIH; scale-bar data from Matt Russell)

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 water amount in the body. Osmoreceptors are specialized cells in the brain that monitor solute concentration in the blood. If the solute levels increase beyond a certain range, a hormone releases that slows water loss through the kidney and dilutes the blood to safer levels. Animals also have high albumin concentrations, which the liver produces, 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.

Chapter Attributions

Active Transport

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

- Understand how electrochemical gradients affect ions
- Distinguish between primary active transport and secondary active transport

Active transport mechanisms require 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 substance's concentration 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.

Electrochemical Gradient

We have discussed simple concentration gradients—a substance's differential concentrations 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 in which they are bathed, and at the same time, cells have higher concentrations of potassium (K^+) and lower concentrations of sodium (Na^+) than the extracellular fluid. Thus in a living cell, the concentration gradient of Na^+ tends to drive it into the cell, and its electrical gradient (a positive ion) also drives it inward to the negatively charged interior. However, the situation is more complex for other elements such as potassium. The electrical gradient of K^+, a positive ion, also drives it into the cell, but the concentration gradient of K^+ drives K^+ *out* of the cell (Figure 5.16). We call the combined concentration gradient and electrical charge that affects an ion its **electrochemical gradient**.

VISUAL CONNECTION

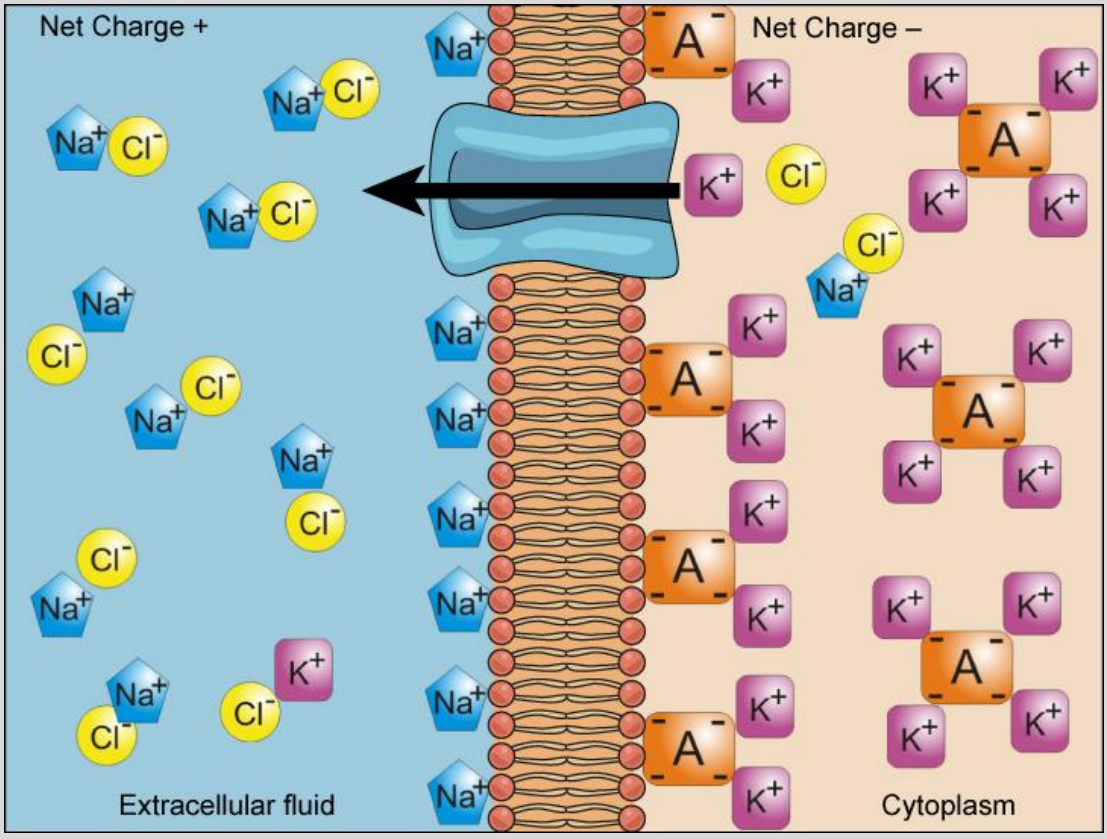


Figure 5.16 Electrochemical gradients arise from the combined effects of concentration gradients and electrical gradients. Structures labeled A represent proteins. (credit: "Synaptitude"/Wikimedia Commons)

Injecting a potassium solution into a person's blood is lethal. This is how capital punishment and euthanasia subjects die. 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. This energy comes from ATP generated through the cell's metabolism. Active transport mechanisms, or **pumps**, work against electrochemical gradients. Small substances constantly pass through plasma membranes. Active transport maintains concentrations of ions and other substances that living cells require in the face of these passive movements. A cell may spend much of its metabolic energy supply maintaining these processes. (A red blood cell uses most of its metabolic energy to maintain the imbalance between exterior and interior sodium and potassium levels that the cell requires.) Because active transport mechanisms depend on a cell's metabolism for energy, they are sensitive to many metabolic poisons that interfere with the ATP supply.

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

Carrier Proteins for Active Transport

An important membrane adaption for active transport is the presence of specific carrier proteins or pumps to facilitate movement: there are three protein types or **transporters** (Figure 5.17). A **uniporter** carries one specific ion or molecule. A **symporter** carries two different ions or molecules, both in the same direction. An **antiporter** also carries two different ions or molecules, but in different directions. All of these transporters can also transport small, uncharged organic molecules like glucose. These three types of carrier proteins are also in

facilitated diffusion, but they do not require ATP to work in that process. Some examples of pumps for active transport are Na^+-K^+ ATPase, which carries sodium and potassium ions, and H^+-K^+ ATPase, which carries hydrogen and potassium ions. Both of these are antiporter carrier proteins. Two other carrier proteins are Ca^{2+} ATPase and H^+ ATPase, which carry only calcium and only hydrogen ions, respectively. Both are pumps.

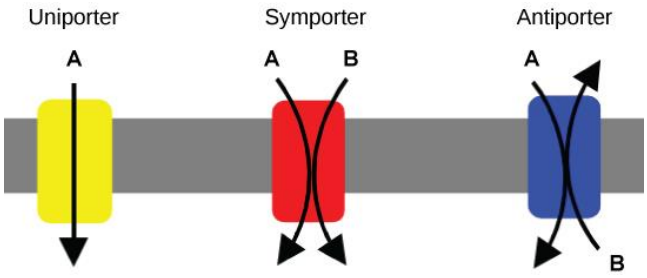


Figure 5.17 A uniporter carries one molecule or ion. A symporter carries two different molecules or ions, both in the same direction. An antiporter also carries two different molecules or ions, but in different directions. (credit: modification of work by "Lupask"/Wikimedia Commons)

Primary Active Transport

The primary active transport that functions with the active transport of sodium and potassium allows secondary active transport to occur. The second transport method is still active because it depends on using energy as does primary transport (Figure 5.18).

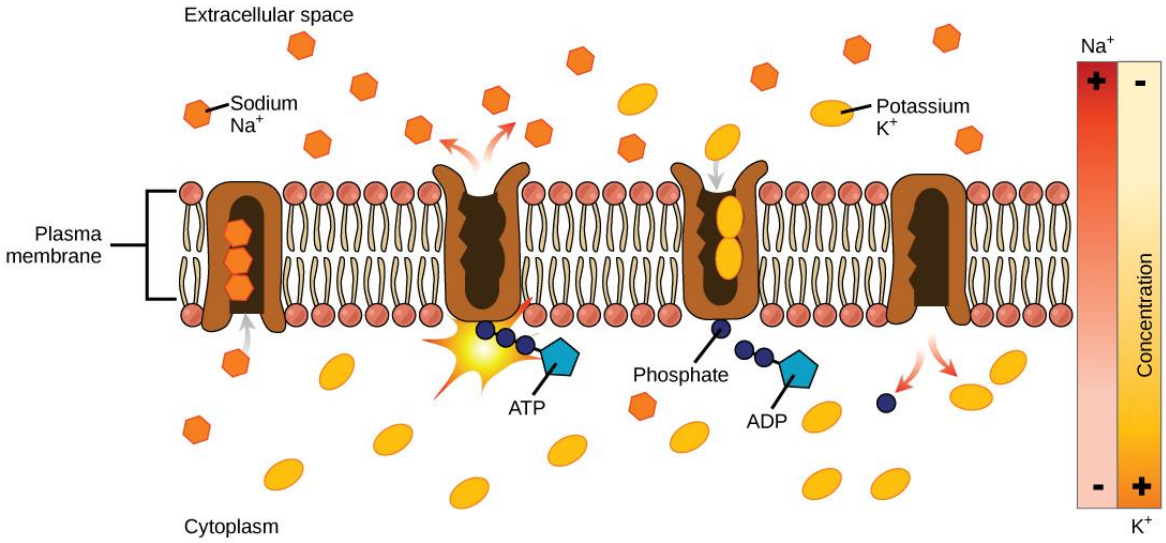


Figure 5.18 Primary active transport moves ions across a membrane, creating an electrochemical gradient (electrogenic transport). (credit: modification of work by Mariana Ruiz Villareal)

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

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

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

an **electrogenic pump** (a pump that creates a charge imbalance), creating an electrical imbalance across the membrane and contributing to the membrane potential.

LINK TO LEARNING

Watch this video (bit.ly/33SjlJ1) to see an active transport simulation in a sodium-potassium ATPase.

Secondary Active Transport (Co-transport)

Secondary active transport brings sodium ions, and possibly other compounds, into the cell. As sodium ion concentrations build outside of the plasma membrane because of the primary active transport process, this creates an electrochemical gradient. If a channel protein exists and is open, the sodium ions will pull through the membrane. This movement transports other substances that can attach themselves to the transport protein through the membrane (Figure 5.19). Many amino acids, as well as glucose, enter a cell this way. This secondary process also stores high-energy hydrogen ions in the mitochondria of plant and animal cells in order to produce ATP. The potential energy that accumulates in the stored hydrogen ions translates into kinetic energy as the ions surge through the channel protein ATP synthase, and that energy then converts ADP into ATP.

VISUAL CONNECTION

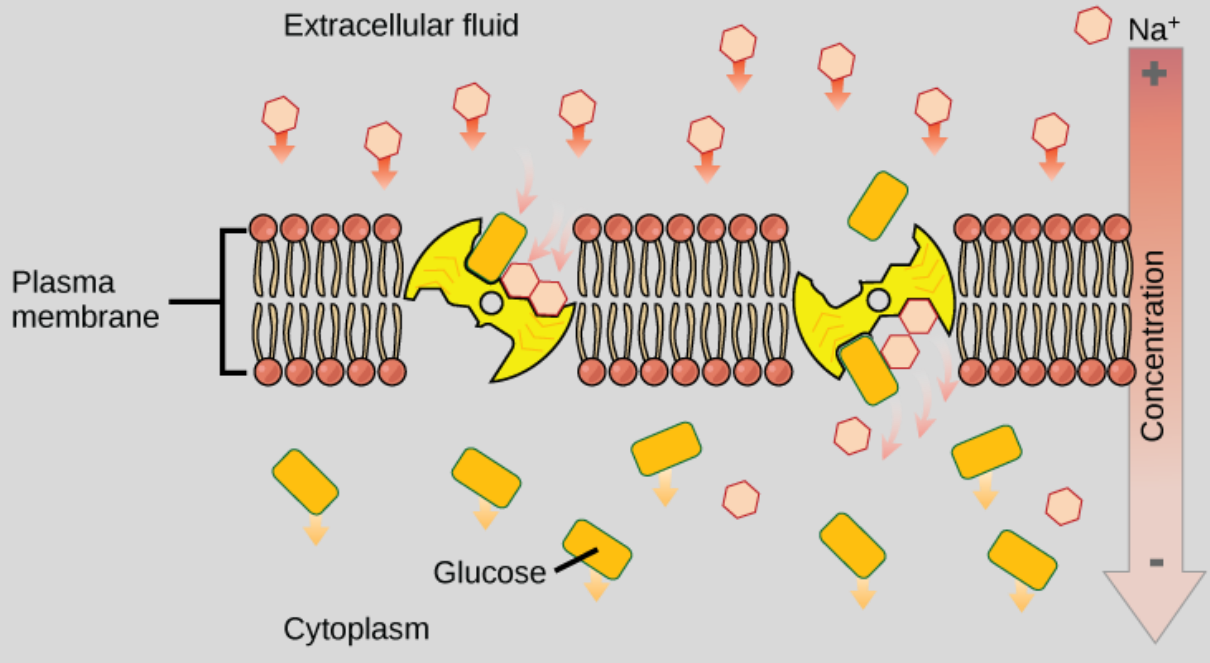


Figure 5.19 An electrochemical gradient, which primary active transport creates, can move other substances against their concentration gradients, a process scientists call co-transport or secondary active transport. (credit: modification of work by Mariana Ruiz Villareal)

If the pH outside the cell decreases, would you expect the amount of amino acids transported into the cell to increase or decrease?

Chapter Attributions

Bulk Transport

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

- 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 5.2 for examples). Some cells are even capable of engulfing entire unicellular microorganisms. You might have correctly hypothesized that when a cell uptakes and releases large particles, it requires energy. A large particle, however, cannot pass through the membrane, even with energy that the cell supplies.

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 endocytosis variations, but all share a common characteristic: the cell's plasma membrane invaginates, forming a pocket around the target particle. The pocket pinches off, resulting in the particle containing itself in a newly created intracellular vesicle formed from the plasma membrane.

Phagocytosis

Phagocytosis (the condition of “cell eating”) is the process by which a cell takes in large particles, such as other cells or relatively large particles. For example, when microorganisms invade the human body, a type of white blood cell, a neutrophil, will remove the invaders through this process, surrounding and engulfing the microorganism, which the neutrophil then destroys (Figure 5.20).

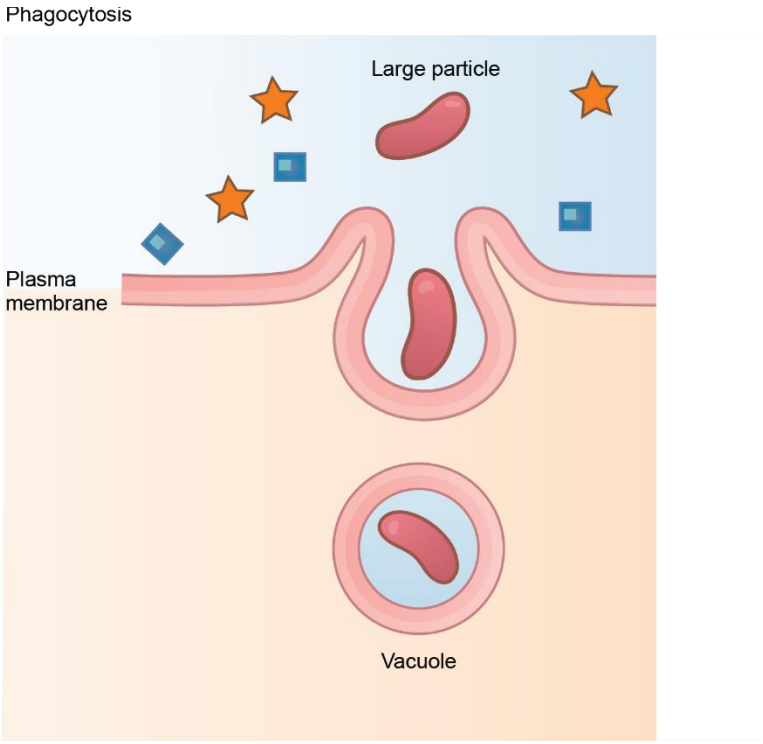


Figure 5.20 In phagocytosis, the cell membrane surrounds the particle and engulfs it. (credit: modification of work by Mariana Ruiz Villareal)

In preparation for phagocytosis, a portion of the plasma membrane's inward-facing surface becomes coated with the protein **clathrin**, which stabilizes this membrane's section. The membrane's coated portion then extends from the cell's body 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 breaking down the material in the newly formed compartment (endosome). When accessible nutrients from the vesicular contents' degradation 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

A variation of endocytosis is **pinocytosis**. This literally means “cell drinking”. Discovered by Warren Lewis in 1929, this American embryologist and cell biologist described a process whereby he assumed that the cell was purposefully taking in extracellular fluid. In reality, this is a process that takes 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 5.21).

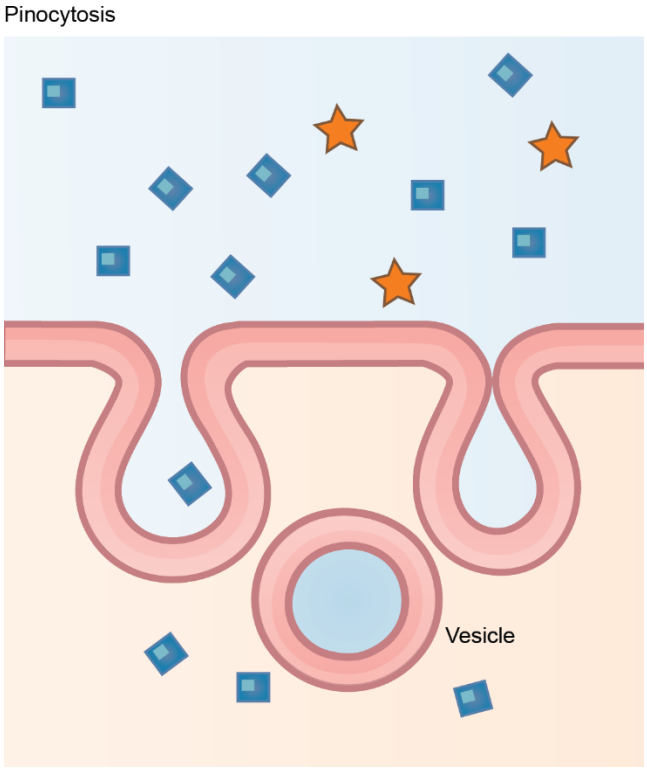


Figure 5.21 In pinocytosis, the cell membrane invaginates, surrounds a small volume of fluid, and pinches off. (credit: modification of work by Mariana Ruiz Villareal)

A variation of pinocytosis is **potocytosis**. This process uses a coating protein, **caveolin**, on the plasma membrane's cytoplasmic side, which performs a similar function to clathrin. The cavities in the plasma membrane that form the vacuoles have membrane receptors and lipid rafts in addition to caveolin. The vacuoles or vesicles formed in caveolae (singular caveola) are smaller

than those in pinocytosis. Potocytosis brings small molecules into the cell and transports them through the cell for their release on the other side, a process we call transcytosis.

Receptor-mediated Endocytosis

A targeted variation of endocytosis employs receptor proteins in the plasma membrane that have a specific binding affinity for certain substances (Figure 5.22).

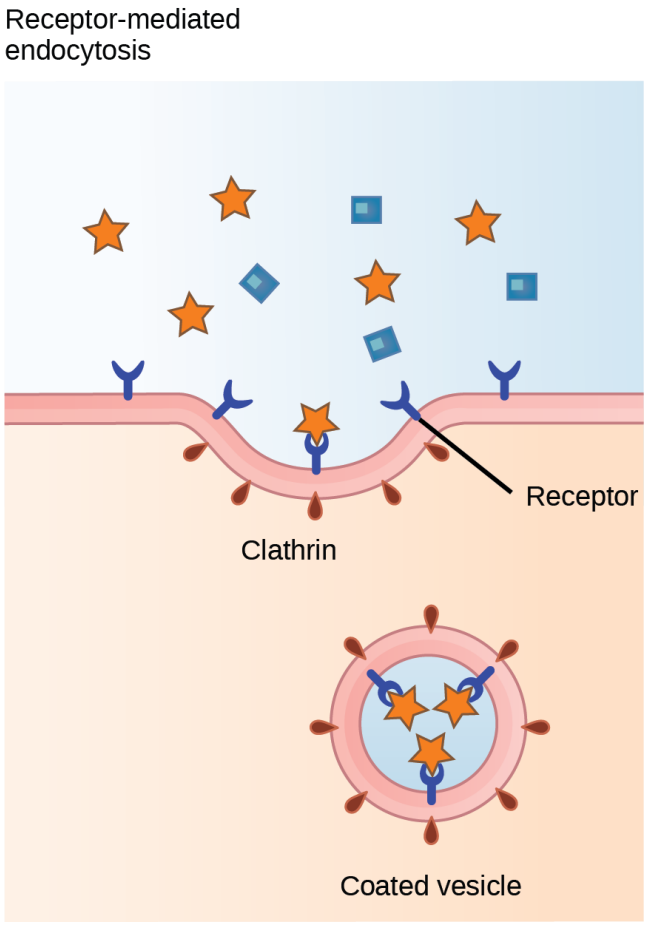


Figure 5.22 In receptor-mediated endocytosis, the cell's uptake of substances targets a single type of substance that binds to the receptor on the cell membrane's external surface. (credit: modification of work by Mariana Ruiz Villareal)

In **receptor-mediated endocytosis**, as in phagocytosis, clathrin attaches to the plasma membrane's cytoplasmic side. If a compound's uptake is dependent on receptor-mediated endocytosis and the process is ineffective, the material will not be removed from the tissue fluids or blood. Instead, it will stay in those fluids and increase in concentration. The failure of

receptor-mediated endocytosis causes some human diseases. For example, receptor mediated endocytosis removes low density lipoprotein or LDL (or "bad" cholesterol) from the blood. 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.

Although receptor-mediated endocytosis is designed to bring specific substances that are normally 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.

LINK TO LEARNING

See receptor-mediated endocytosis in action, and click on different parts for a focused animation: https://www.youtube.com/watch?v=hLbjLWNA5c0

Exocytosis

The reverse process of moving material into a cell is the process of exocytosis. **Exocytosis** is the opposite of the processes we discussed above in that its purpose is to expel material from the cell into the extracellular fluid. Waste material is enveloped in a membrane and fuses with the plasma membrane's interior. This fusion opens the membranous envelope on the cell's exterior, and the waste material expels into the extracellular space (Figure 5.23). Other examples of cells releasing molecules via exocytosis include extracellular matrix protein secretion and neurotransmitter secretion into the synaptic cleft by synaptic vesicles.

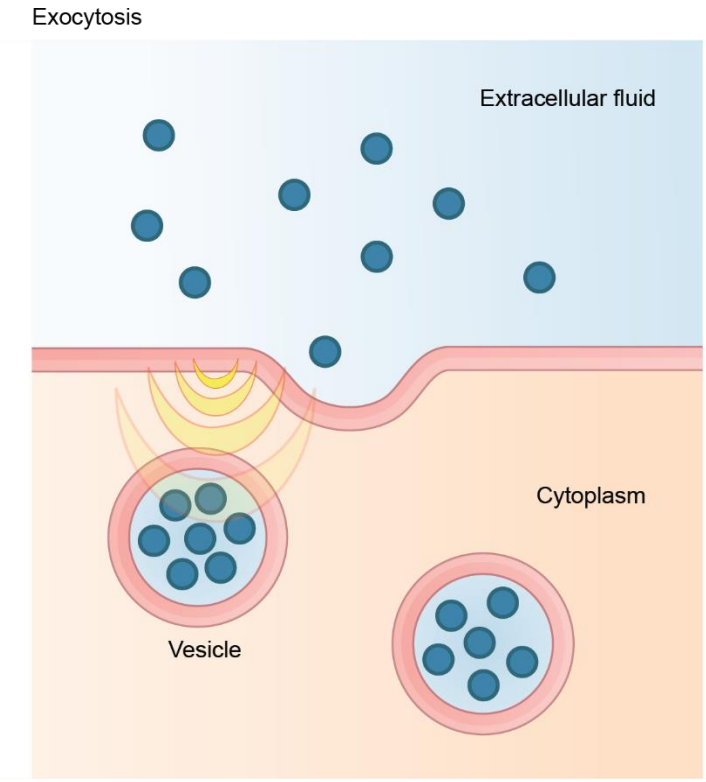


Figure 5.23 In exocytosis, vesicles containing substances fuse with the plasma membrane. The contents then release to the cell's exterior. (credit: modification of work by Mariana Ruiz Villareal)

Methods of Transport, Energy Requirements, and Types of Transported Material

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

Methods of Transport, Energy Requirements, and Types of Transported Material

Transport Method	Active/Passive	Material Transported
Secondary active transport	Active	Amino acids, lactose
Phagocytosis	Active	Large macromolecules, whole cells, or cellular s
Pinocytosis and potocytosis	Active	Small molecules (liquids/water)
Receptor-mediated endocytosis	Active	Large quantities of macromolecules

Table5.2

Chapter Attributions

13. DNA and Replication

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

- Describe the structure and features of the nuclear membrane
- List the contents of the nucleus
- Explain the organization of the DNA molecule within the nucleus
- Describe the process of DNA replication

The nucleus is the largest and most prominent of a cell's organelles (Figure 3.19). The nucleus is generally considered the control center of the cell because it stores all of the genetic instructions for manufacturing proteins. Interestingly, some cells in the body, such as muscle cells, contain more than one nucleus (Figure 3.20), which is known as multinucleated. Other cells, such as mammalian red blood cells (RBCs), do not contain nuclei at all. RBCs eject their

nuclei as they mature, making space for the large numbers of hemoglobin molecules that carry oxygen throughout the body (Figure 3.21). Without nuclei, the life span of RBCs is short, and so the body must produce new ones constantly.

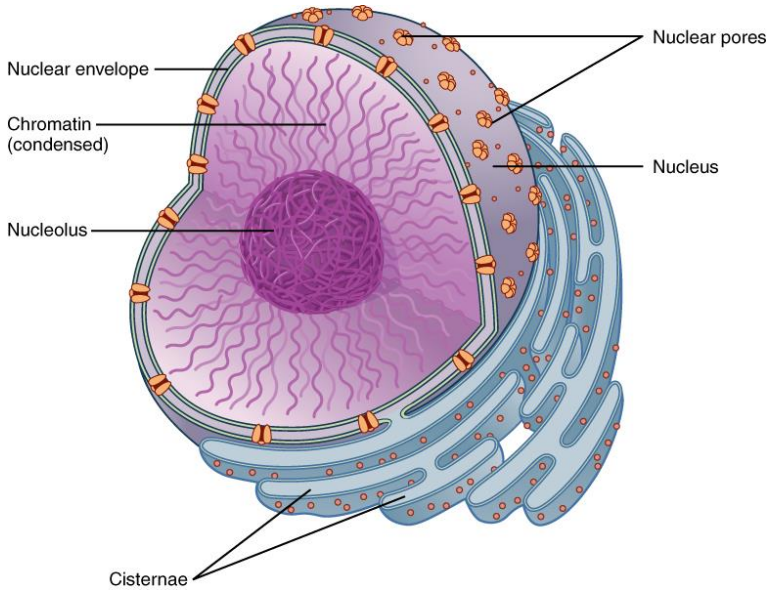


Figure 3.19 The Nucleus The nucleus is the control center of the cell. The nucleus of living cells contains the genetic material that determines the entire structure and function of that cell.

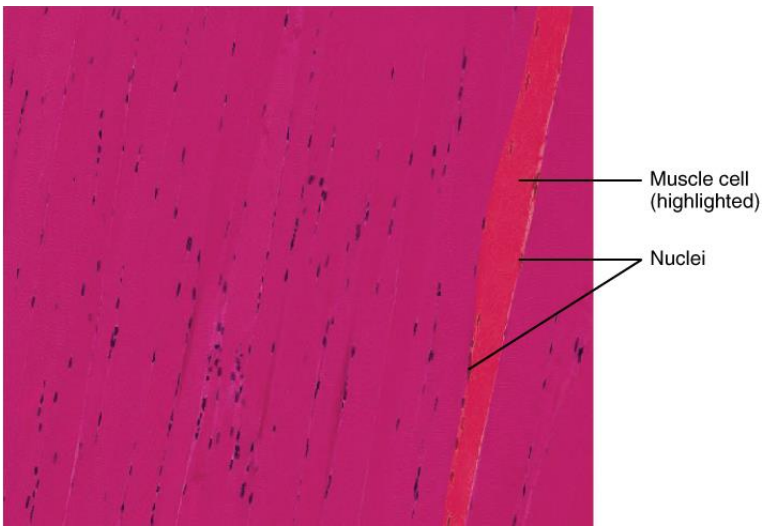


Figure 3.20 Multinucleate Muscle Cell Unlike cardiac muscle cells and smooth muscle cells, which have a single nucleus, a skeletal muscle cell contains many nuclei, and is referred to as "multinucleated." These muscle cells are long and fibrous (often referred to as muscle fibers). During development, many smaller cells fuse to form a mature muscle fiber. The nuclei of the

fused cells are conserved in the mature cell, thus imparting a multinucleate characteristic to mature muscle cells. LM × 104.3. (Micrograph provided by the Regents of University of Michigan Medical School © 2012)

INTERACTIVE LINK

View the University of Michigan WebScope (http://bit.ly/2Muv62q) to explore the tissue sample in greater detail.

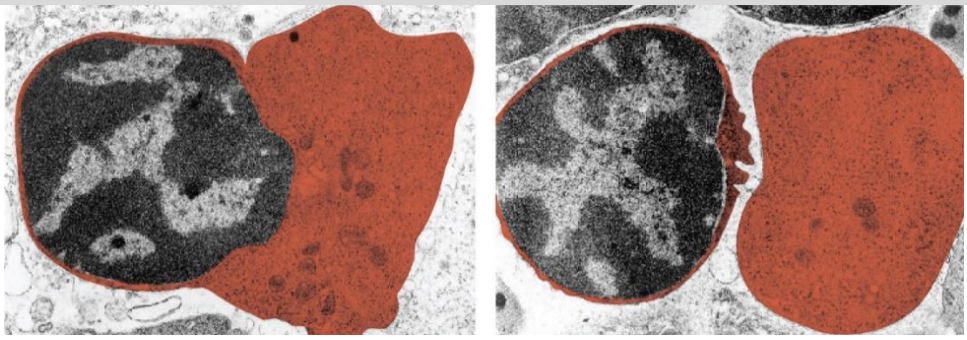


Figure 3.21 Red Blood Cell Extruding Its Nucleus Mature red blood cells lack a nucleus. As they mature, erythroblasts extrude their nucleus, making room for more hemoglobin. The two panels here show an erythroblast before and after ejecting its nucleus, respectively. (credit: modification of micrograph provided by the Regents of University of Michigan Medical School © 2012)

Inside the nucleus lies the blueprint that dictates everything a cell will do and all of the products it will make. This information is stored within DNA. The nucleus sends “commands” to the cell via molecular messengers that translate the information from DNA. Each cell in your body (with the exception of germ cells) contains the complete set of your DNA. When a cell divides, the DNA must be duplicated so that the each new cell receives a full complement of DNA. The following section will explore the structure of the nucleus and its contents, as well as the process of DNA replication.

Organization of the Nucleus and Its DNA

Like most other cellular organelles, the nucleus is surrounded by a membrane called the **nuclear envelope**. This membranous covering consists of two adjacent lipid bilayers with a thin fluid space in between them. Spanning these two bilayers are nuclear pores. A **nuclear**

pore is a tiny passageway for the passage of proteins, RNA, and solutes between the nucleus and the cytoplasm. Proteins called pore complexes lining the nuclear pores regulate the passage of materials into and out of the nucleus.

Inside the nuclear envelope is a gel-like nucleoplasm with solutes that include the building blocks of nucleic acids. There also can be a dark-staining mass often visible under a simple light microscope, called a **nucleolus** (plural = nucleoli). The nucleolus is a region of the nucleus that is responsible for manufacturing the RNA necessary for construction of ribosomes. Once synthesized, newly made ribosomal subunits exit the cell's nucleus through the nuclear pores.

The genetic instructions that are used to build and maintain an organism are arranged in an orderly manner in strands of DNA. Within the nucleus are threads of **chromatin** composed of DNA and associated proteins (Figure 3.22). Along the chromatin threads, the DNA is wrapped around a set of **histone** proteins. A **nucleosome** is a single, wrapped DNA-histone complex. Multiple nucleosomes along the entire molecule of DNA appear like a beaded necklace, in which the string is the DNA and the beads are the associated histones. When a cell is in the process of division, the chromatin condenses into chromosomes, so that the DNA can be safely transported to the "daughter cells." The **chromosome** is composed of DNA and proteins; it is the condensed form of chromatin. It is estimated that humans have almost 22,000 genes distributed on 46 chromosomes.

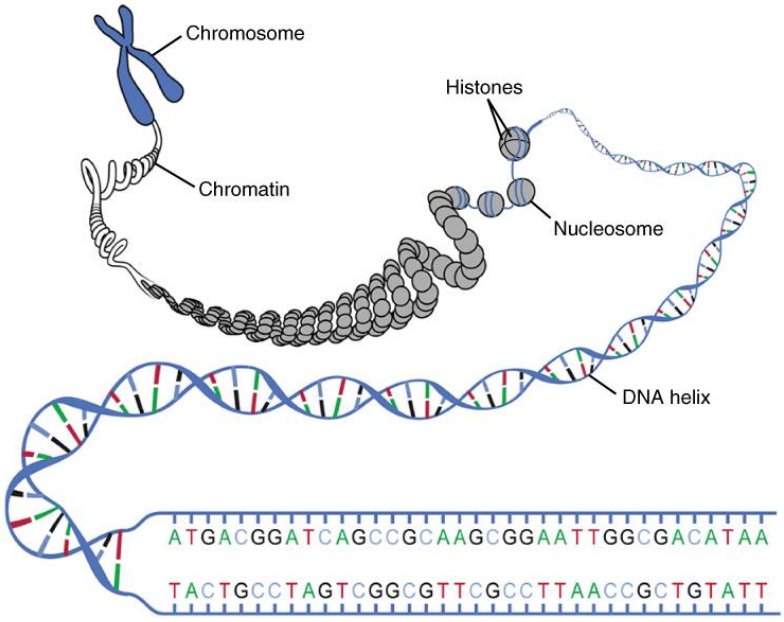


Figure 3.22 DNA Macrostructure Strands of DNA are wrapped around supporting histones. These proteins are increasingly bundled and condensed into chromatin, which is packed tightly into chromosomes when the cell is ready to divide.

DNA Replication

In order for an organism to grow, develop, and maintain its health, cells must reproduce themselves by dividing to produce two new daughter cells, each with the full complement of DNA as found in the original cell. Billions of new cells are produced in an adult human every day. Only very few cell types in the body do not divide, including nerve cells, skeletal muscle fibers, and cardiac muscle cells. The division time of different cell types varies. Epithelial cells of the skin and gastrointestinal lining, for instance, divide very frequently to replace those that are constantly being rubbed off of the surface by friction.

A DNA molecule is made of two strands that "complement" each other in the sense that the molecules that compose the strands fit together and bind to each other, creating a double-stranded molecule that looks much like a long, twisted ladder. Each side rail of the DNA ladder is composed of alternating sugar and phosphate groups (Figure 3.23). The two sides of the ladder are not identical, but are complementary. These two backbones are bonded to each other across pairs of protruding bases, each bonded pair forming one "rung," or cross member. The four DNA bases are adenine (A), thymine (T), cytosine (C), and guanine (G). Because of their shape and charge, the two bases that compose a pair always bond together. Adenine always binds with thymine, and cytosine always binds with guanine. The particular sequence of bases along the DNA molecule determines the genetic code. Therefore, if the two complementary strands of DNA were pulled apart, you could infer the order of the bases in one strand from the bases in the other, complementary strand. For example, if one strand has a region with the sequence AGTGCCT, then the sequence of the complementary strand would be TCACGGA.

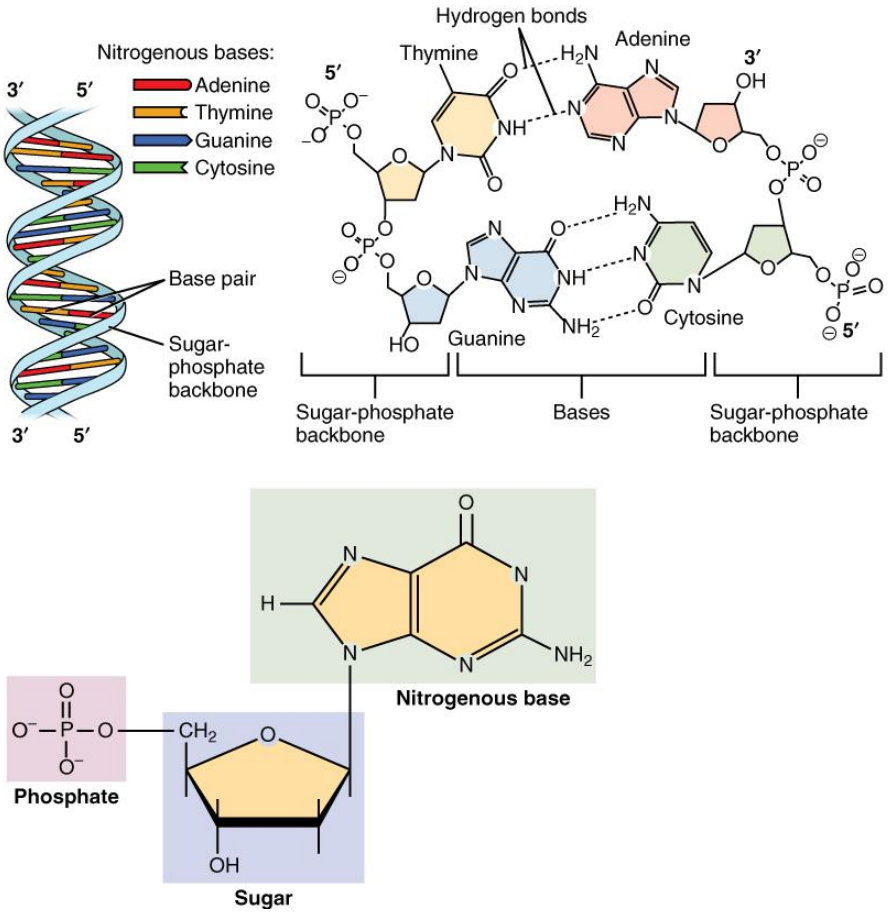


Figure 3.23 Molecular Structure of DNA The DNA double helix is composed of two complementary strands. The strands are bonded together via their nitrogenous base pairs using hydrogen bonds.

DNA replication is the copying of DNA that occurs before cell division can take place. After a great deal of debate and experimentation, the general method of DNA replication was deduced in 1958 by two scientists in California, Matthew Meselson and Franklin Stahl. This method is illustrated in Figure 3.24 and described below.

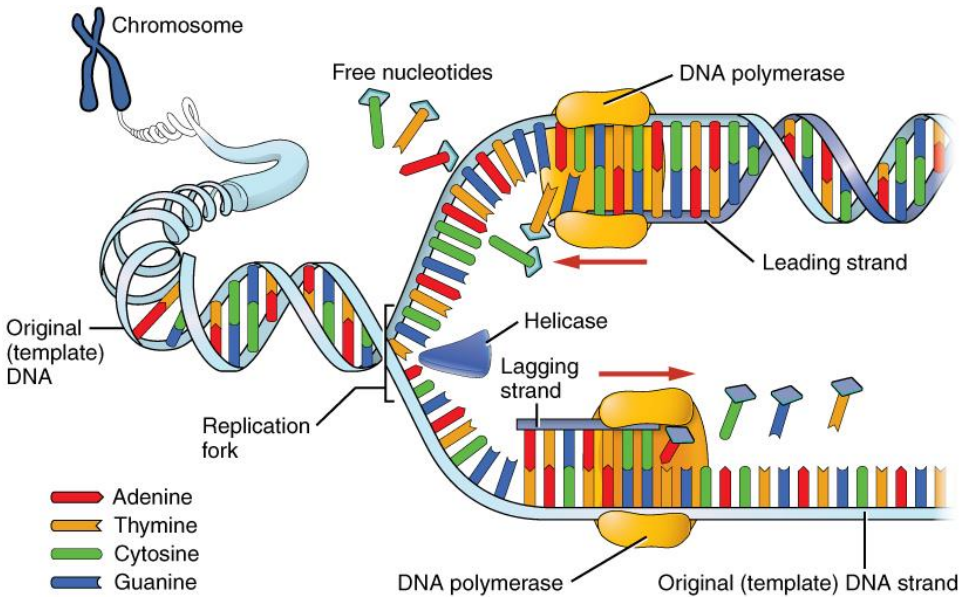


Figure 3.24 DNA Replication DNA replication faithfully duplicates the entire genome of the cell. During DNA replication, a number of different enzymes work together to pull apart the two strands so each strand can be used as a template to synthesize new complementary strands. The two new daughter DNA molecules each contain one pre-existing strand and one newly synthesized strand. Thus, DNA replication is said to be "semiconservative."

Stage 1: Initiation. The two complementary strands are separated, much like unzipping a zipper. Special enzymes, including **helicase**, untwist and separate the two strands of DNA.

Stage 2: Elongation. Each strand becomes a template along which a new complementary strand is built. **DNA polymerase** brings in the correct bases to complement the template strand, synthesizing a new strand base by base. A DNA polymerase is an enzyme that adds free nucleotides to the end of a chain of DNA, making a new double strand. This growing strand continues to be built until it has fully complemented the template strand.

Stage 3: Termination. Once the two original strands are bound to their own, finished, complementary strands, DNA replication is stopped and the two new identical DNA molecules are complete.

Each new DNA molecule contains one strand from the original molecule and one newly synthesized strand. The term for this mode of replication is "semiconservative," because half of the original DNA molecule is conserved in each new DNA molecule. This process continues until

the cell's entire **genome**, the entire complement of an organism's DNA, is replicated. As you might imagine, it is very important that DNA replication take place precisely so that new cells in the body contain the exact same genetic material as their parent cells. Mistakes made during DNA replication, such as the accidental addition of an inappropriate nucleotide, have the potential to render a gene dysfunctional or useless. Fortunately, there are mechanisms in place to minimize such mistakes. A DNA proofreading process enlists the help of special enzymes that scan the newly synthesized molecule for mistakes and corrects them. Once the process of DNA replication is complete, the cell is ready to divide. You will explore the process of cell division later in the chapter.

INTERACTIVE LINK

Watch this video (https://www.youtube.com/watch?v=FBmO_rmXxIw) to learn about DNA replication. DNA replication proceeds simultaneously at several sites on the same molecule. What separates the base pair at the start of DNA replication?

14. Protein Synthesis

Transcription

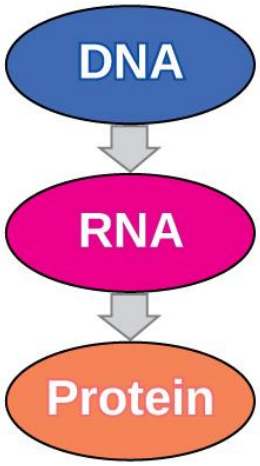
By the end of this section, you will be able to:

- Explain the central dogma
- Explain the main steps of transcription
- Describe how eukaryotic mRNA is processed

The second function of DNA (the first was replication) is to provide the information needed to construct the proteins necessary so that the cell can perform all of its functions. To do this, the DNA is “read” or transcribed into an **mRNA** molecule. The mRNA then provides the code to form a protein by a process called translation. Through the processes of transcription and translation, a protein is built with a specific sequence of amino acids that was originally encoded in the DNA. This module discusses the details of transcription.

The Central Dogma of Molecular Biology: DNA Encodes RNA; RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the central dogma (Figure), which states that genes specify the sequences of mRNAs, which in turn specify the sequences of proteins.



The central dogma of molecular biology states that DNA encodes RNA, which in turn encodes protein.

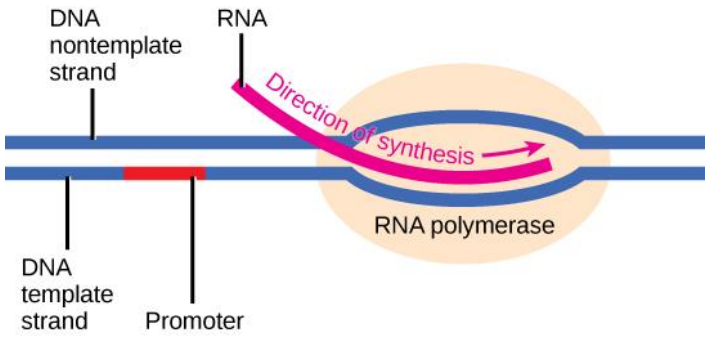
The copying of DNA to mRNA (i.e. transcription) is relatively straightforward, with one nucleotide being added to the mRNA strand for every complementary nucleotide read in the DNA strand. The translation to protein is more complex because groups of three mRNA nucleotides correspond to one amino acid of the protein sequence. However, as we shall see in the next module, the translation to protein is still systematic, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on. The groups of three nucleotides that specify an amino acid are called codons.

Transcription: from DNA to mRNA

With the genes bound in the nucleus, transcription occurs in the nucleus of the cell and the mRNA transcript must be transported to the cytoplasm. Transcription occurs in three main stages: initiation, elongation, and termination.

Initiation

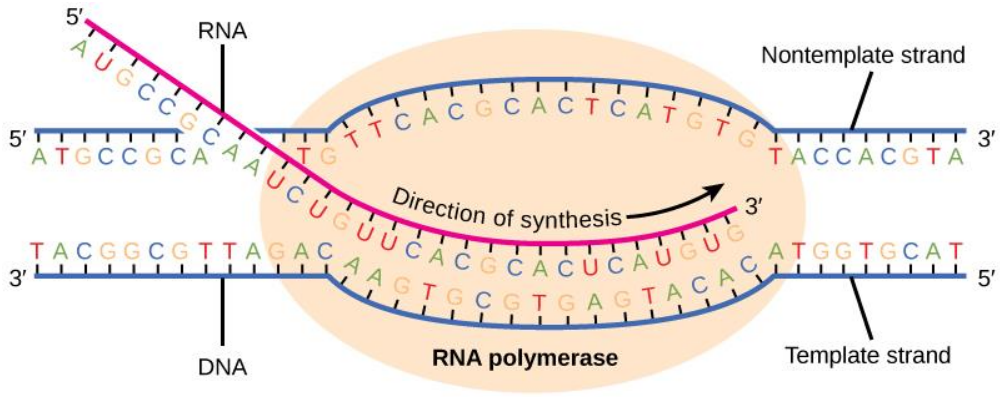
Transcription requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a **transcription bubble**. The DNA sequence onto which the proteins and enzymes involved in transcription bind to initiate the process is called a **promoter**. In most 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 of the time, some of the time, or hardly at all (Figure).



The initiation of transcription begins when DNA is unwound, forming a transcription bubble. Enzymes and other proteins involved in transcription bind at the promoter.

Elongation

Transcription always proceeds from one of the two DNA strands, which is called the **template strand**. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the **nontemplate strand**, with the exception that RNA contains a uracil (U) in place of the thymine (T) found in DNA. During elongation, an enzyme called **RNA polymerase** proceeds along the DNA template adding nucleotides by base pairing with the DNA template in a manner similar to DNA replication, with the difference that an RNA strand is being synthesized that does not remain bound to the DNA template. As elongation proceeds, the DNA is continuously unwound ahead of the enzyme and rewound behind it (Figure).



During elongation, RNA polymerase tracks along the DNA template, synthesizes mRNA in the 5' to 3' direction, and unwinds then rewinds the DNA as it is read.

Termination

Once a gene is transcribed, the RNA polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals, but both involve repeated nucleotide sequences in the DNA template that result in RNA polymerase stalling, leaving the DNA template, and freeing the mRNA transcript.

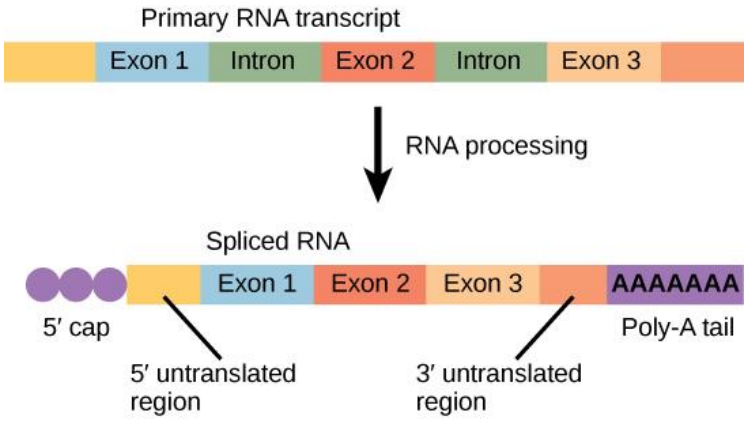
Eukaryotic RNA Processing

The newly transcribed eukaryotic mRNAs must undergo several processing steps before they can be transferred from the nucleus to the cytoplasm and translated into a protein.

The mRNA transcript is first coated in RNA-stabilizing proteins to prevent it from degrading while it is processed and exported out of the nucleus. This occurs while the pre-mRNA still is being synthesized by adding a special nucleotide “cap” to the 5' end of the growing transcript. In addition to preventing degradation, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

Once elongation is complete, an enzyme then adds a string of approximately 200 adenine residues to the 3' end, called the poly-A tail. This modification further protects the pre-mRNA from degradation and signals to cellular factors that the transcript needs to be exported to the cytoplasm.

Eukaryotic genes are composed of protein-coding sequences called **exons** (*ex*-on signifies that they are *ex*pressed) and *int*ervening sequences called **introns** (*int*-ron denotes their *int*ervening role). Introns are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins. It is essential that all of a pre-mRNA's introns be completely and precisely removed before protein synthesis so that the exons join together to code for the correct amino acids. If the process errs by even a single nucleotide, the sequence of the rejoined exons would be shifted, and the resulting protein would be nonfunctional. The process of removing introns and reconnecting exons is called **splicing** (Figure). Introns are removed and degraded while the pre-mRNA is still in the nucleus.



Eukaryotic mRNA contains introns that must be spliced out. A 5' cap and 3' tail are also added.

Section Summary

mRNA synthesis is initiated at a promoter sequence on the DNA template. Elongation synthesizes new mRNA (called a pre-mRNA). Termination liberates the mRNA and occurs by mechanisms that stall the RNA polymerase and cause it to fall off the DNA template. Newly transcribed mRNAs are modified with a cap and a poly-A tail. These structures protect the mature mRNA from degradation and help export it from the nucleus. mRNAs also undergo splicing, in which introns are removed and exons are reconnected with single-nucleotide accuracy. Only finished mRNAs are exported from the nucleus to the cytoplasm.

Multiple Choice

A promoter is _____.

a. a specific sequence of DNA nucleotides

b. a specific sequence of RNA nucleotides

c. a protein that binds to DNA

d. an enzyme that synthesizes RNA

Portions of eukaryotic mRNA sequence that are removed during RNA processing are _____.

a. exons

b. caps

c. poly-A tails

d. introns

Glossary

exon. a sequence present in protein-coding mRNA after completion of pre-mRNA splicing

intron. non–protein-coding intervening sequences that are spliced from mRNA during processing

mRNA. messenger RNA; a form of RNA that carries the nucleotide sequence code for a protein sequence that is translated into a polypeptide sequence

nontemplate strand. the strand of DNA that is not used to transcribe mRNA; this strand is identical to the mRNA except that T nucleotides in the DNA are replaced by U nucleotides in the mRNA

promoter. a sequence on DNA to which RNA polymerase and associated factors bind and initiate transcription

RNA polymerase. an enzyme that synthesizes an RNA strand from a DNA template strand

splicing. the process of removing introns and reconnecting exons in a pre-mRNA

template strand. the strand of DNA that specifies the complementary mRNA molecule

transcription bubble. the region of locally unwound DNA that allows for transcription of mRNA

Chapter Attributions

"Human Biology" by Willy Cushwa, OpenStax is licensed under CC BY 4.0. Access for free at https://cnx.org/contents/5ZI71dr1@3.2:utN40dRD@1/Transcription

Translation

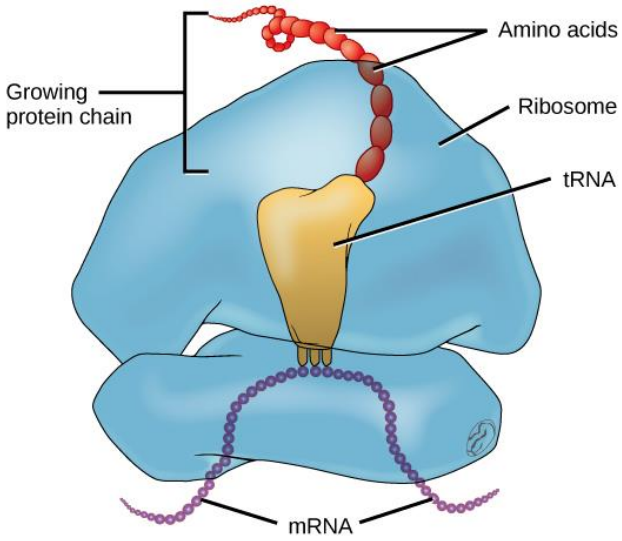
By the end of this section, you will be able to:

- Describe the different steps in protein synthesis
- Discuss the role of ribosomes in protein synthesis
- Describe the genetic code and how the nucleotide sequence determines the amino acid and the protein sequence

The synthesis of proteins is one of a cell's most energy-consuming metabolic processes. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform a wide variety of the functions of a cell. The process of translation, or protein synthesis, involves decoding an mRNA message into a polypeptide product. Amino acids are covalently strung together in lengths ranging from approximately 50 amino acids to more than 1,000.

The Protein Synthesis Machinery

In addition to the mRNA template, many other molecules contribute to the process of translation. The composition of each component may vary across species; for instance, ribosomes may consist of different numbers of ribosomal RNAs (**rRNA**) and polypeptides depending on the organism. 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 enzymatic factors (Figure).



The protein synthesis machinery includes the large and small subunits of the ribosome, mRNA, and tRNA. (credit: modification of work by NIGMS, NIH)

Ribosomes are located in the cytoplasm and endoplasmic reticulum of eukaryotes. Ribosomes are made up of a large and a small subunit that come together for translation. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds **tRNAs**, a type of RNA molecule that brings amino acids to the growing chain of the polypeptide. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction.

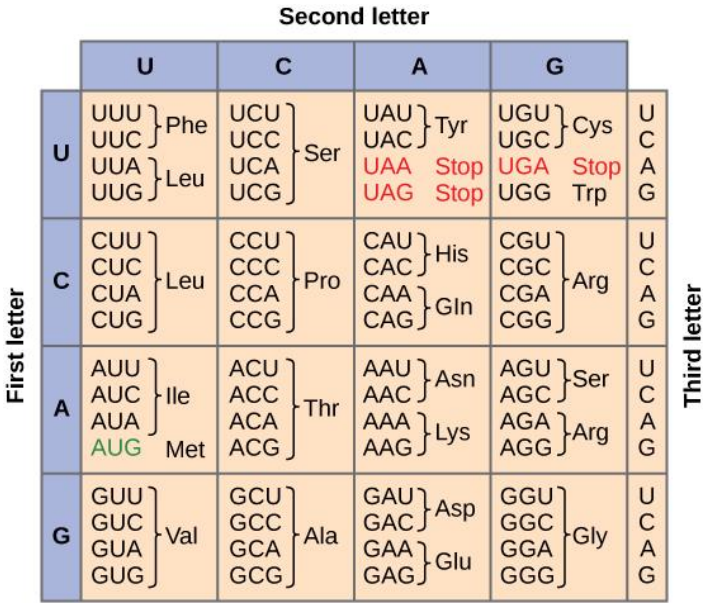
Depending on the species, 40 to 60 types of tRNA exist in the cytoplasm. Serving as adaptors, specific tRNAs bind to sequences on the mRNA template and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually "translate" the language of RNA into the language of proteins. For each tRNA to function, it must have its specific amino acid bonded to it. In the process of tRNA "charging," each tRNA molecule is bonded to its correct amino acid.

The Genetic Code

To summarize what we know to this point, the cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template converts nucleotide-based genetic information into a protein product. Protein sequences consist of 20 commonly occurring amino

acids; therefore, it can be said that the protein alphabet consists of 20 letters. Each amino acid is defined by a three-nucleotide sequence called the triplet **codon**. The relationship between a nucleotide codon and its corresponding amino acid is called the **genetic code**.

Given the different numbers of “letters” in the mRNA and protein “alphabets,” combinations of nucleotides corresponded to single amino acids. Using a three-nucleotide code means that there are a total of 64 (4 × 4 × 4) possible combinations; therefore, a given amino acid is encoded by more than one nucleotide triplet (Figure).



This figure shows the genetic code for translating each nucleotide triplet, or codon, in mRNA into an amino acid or a termination signal in a nascent protein. (credit: modification of work by NIH)

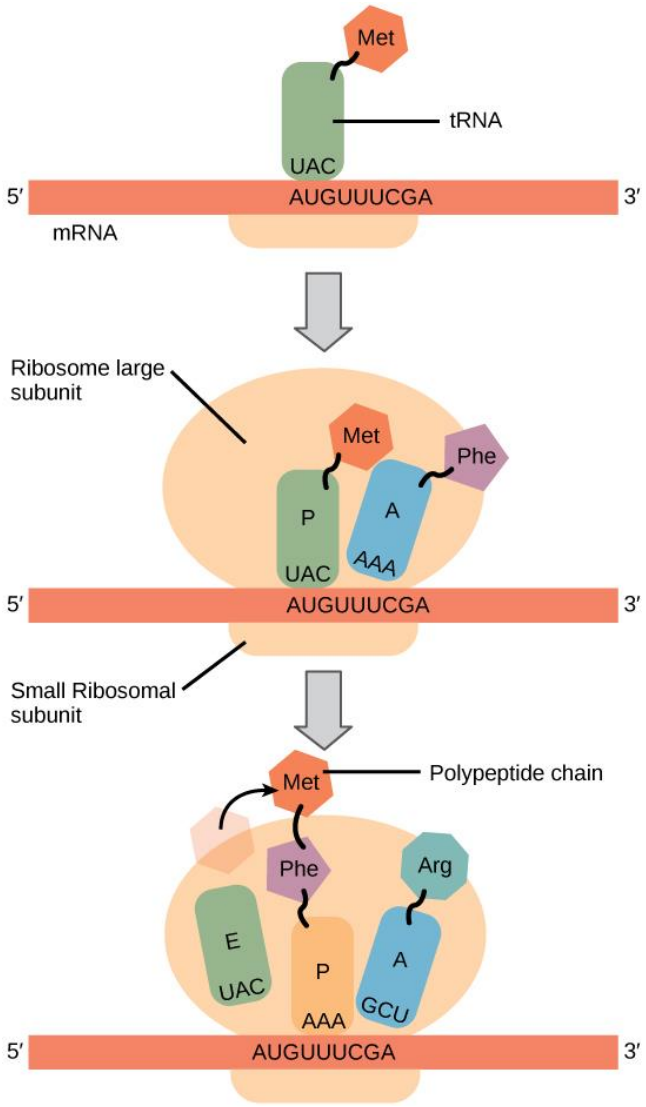
Three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **stop codons**. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the **start codon** to initiate translation. The reading frame for translation is set by the AUG start codon near the 5' end of the mRNA. The genetic code is universal. With a few exceptions, virtually all species use the same genetic code for protein synthesis, which is powerful evidence that all life on Earth shares a common origin.

The Mechanism of Protein Synthesis

Just as with mRNA synthesis, protein synthesis can be divided into three phases: initiation, elongation, and termination.

Protein synthesis begins with the formation of an initiation complex. This complex involves the small ribosome subunit, the mRNA template, and a tRNA that interacts with the AUG start codon, and is linked to the amino acid methionine.

In polypeptide elongation the large ribosomal subunit consists of three compartments: P, A, and E sites. The A site binds incoming charged tRNAs (tRNAs with their attached specific amino acids). The P site binds charged tRNAs carrying amino acids that have formed bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The E site releases dissociated tRNAs so they can be recharged with free amino acids. The ribosome shifts one codon at a time, catalyzing each process that occurs in the three sites. With each step, a charged tRNA enters the complex, the polypeptide becomes one amino acid longer, and an uncharged tRNA departs.



Translation begins when a tRNA anticodon recognizes a codon on the mRNA. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is encountered. When the ribosome encounters the stop codon, the growing polypeptide is released and the ribosome subunits dissociate and leave the mRNA. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

CONCEPT IN ACTION

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at this site (https://learn.genetics.utah.edu/content/basics/transcribe/).

Section Summary

The central dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is the correspondence between the three-nucleotide mRNA codon and an amino acid. The genetic code is "translated" by the tRNA molecules, which associate a specific codon with a specific amino acid. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three stop codons. This means that more than one codon corresponds to an amino acid. Almost every species on the planet uses the same genetic code.

The players in translation include the mRNA template, ribosomes, tRNAs, and various enzymatic factors. The small ribosomal subunit binds to the mRNA template. Translation begins at the initiating AUG on the mRNA. The formation of bonds occurs between sequential amino acids specified by the mRNA template according to the genetic code. The ribosome accepts charged tRNAs, and as it steps along the mRNA, it catalyzes bonding between the new amino acid and the end of the growing polypeptide. The entire mRNA is translated in three-nucleotide "steps" of the ribosome. When a stop codon is encountered, a release factor binds and dissociates the components and frees the new protein.

Multiple Choice

The RNA components of ribosomes are synthesized in the _____.

a. cytoplasm

b. nucleus

c. nucleolus

d. endoplasmic reticulum

How long would the peptide be that is translated from this MRNA sequence: 5'-AUGGGCUACCGA-3'?

a. 0

b. 2

c. 3

d. 4

Free Response

Transcribe and translate the following DNA sequence (nontemplate strand): 5'-ATGGCCGGTTATTAAGCA-3'

Glossary

codon. three consecutive nucleotides in mRNA that specify the addition of a specific amino acid or the release of a polypeptide chain during translation

genetic code. the amino acids that correspond to three-nucleotide codons of mRNA

rRNA. ribosomal RNA; molecules of RNA that combine to form part of the ribosome

stop codon. one of the three mRNA codons that specifies termination of translation

start codon. the AUG (or, rarely GUG) on an mRNA from which translation begins; always specifies methionine

tRNA. transfer RNA; an RNA molecule that contains a specific three-nucleotide anticodon sequence to pair with the mRNA codon and also binds to a specific amino acid

Chapter Attributions

15. Cell Cycle and Mitosis

The Cell Cycle

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

- Describe the three stages of interphase
- Discuss the behavior of chromosomes during karyokinesis/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. Cells on the path to cell division proceed through a series of precisely timed and carefully regulated stages of growth, DNA replication, and nuclear and cytoplasmic division that ultimately produces two identical (clone) cells. The cell cycle has two major phases: interphase and the mitotic phase (Figure 10.5). During **interphase**, the cell grows and DNA is replicated. During the **mitotic phase**, the replicated DNA and cytoplasmic contents are separated, and the cell cytoplasm is typically partitioned by a third process of the cell cycle called **cytokinesis**. We should note, however, that interphase and mitosis (kayrokinesis) may take place without cytokinesis, in which case cells with multiple nuclei (multinucleate cells) are produced.

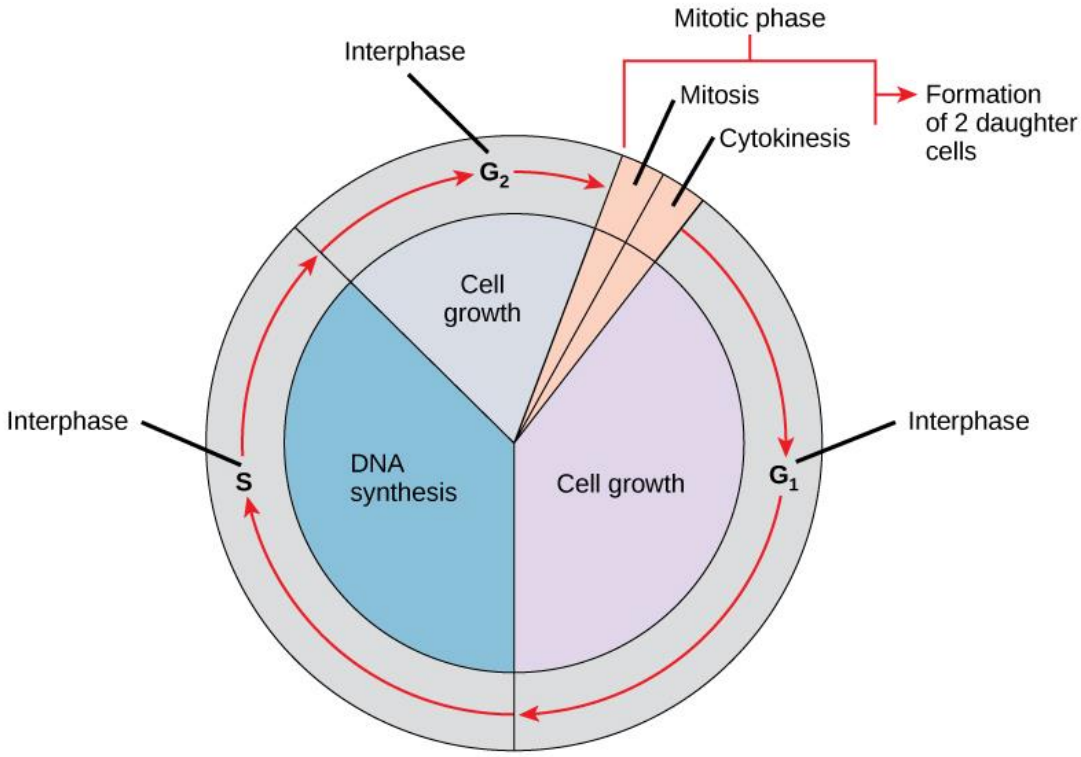


Figure 10.5 The cell cycle in multicellular organisms consists of interphase and the mitotic phase. During interphase, the cell grows and the nuclear DNA is duplicated. Interphase is followed by the mitotic phase. During the mitotic phase, the duplicated chromosomes are segregated and distributed into daughter nuclei. Following mitosis, the cytoplasm is usually divided as well by cytokinesis, resulting in two genetically identical daughter cells.

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 stages of interphase are called *G_1, S, and G_2.*

G_1 Phase (First Gap)

The first stage of interphase is called the **G_1 phase** (first gap) because, from a microscopic point of view, little change is visible. However, during the G_1 stage, the cell is quite active at the biochemical level. The cell is accumulating the building blocks of chromosomal DNA and the associated proteins 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 also duplicated during the S phase. The two centrosomes of homologous chromosomes will give rise to the **mitotic spindle**, the apparatus that orchestrates the movement of chromosomes during mitosis. For example, roughly at the center of each animal cell, the centrosomes are associated with a pair of rod-like objects, the **centrioles**, which are positioned at right angles to each other. Centrioles help organize cell division. We should note, however, that centrioles are not present in the centrosomes of other eukaryotic organisms, such as plants and most fungi.

G_2 Phase (Second Gap)

In the **G_2 phase**, the cell replenishes its energy stores and synthesizes proteins necessary for chromosome manipulation and movement. Some cell organelles are duplicated, and the cytoskeleton is dismantled to provide resources for the mitotic phase. There may be additional cell growth during G_2. The final preparations for the mitotic phase must be completed before the cell is able to enter the first stage of mitosis.

The Mitotic Phase

The mitotic phase is a multistep process during which the duplicated chromosomes are aligned, separated, and move into two new, identical daughter cells. The first portion of the mitotic phase is called **karyokinesis**, or nuclear division. As we have just seen, the second portion of the mitotic phase (and often viewed as a process separate from and following mitosis) is called cytokinesis—the physical separation of the cytoplasmic components into the two daughter cells.

LINK TO LEARNING

Revisit the stages of mitosis at this site:
http://www.biology.arizona.edu/cell_bio/tutorials/cell_cycle/cells3.html

Karyokinesis (Mitosis)

Karyokinesis, also known as **mitosis**, is divided into a series of phases—prophase, prometaphase, metaphase, anaphase, and telophase—that result in the division of the cell nucleus.

VISUAL CONNECTION

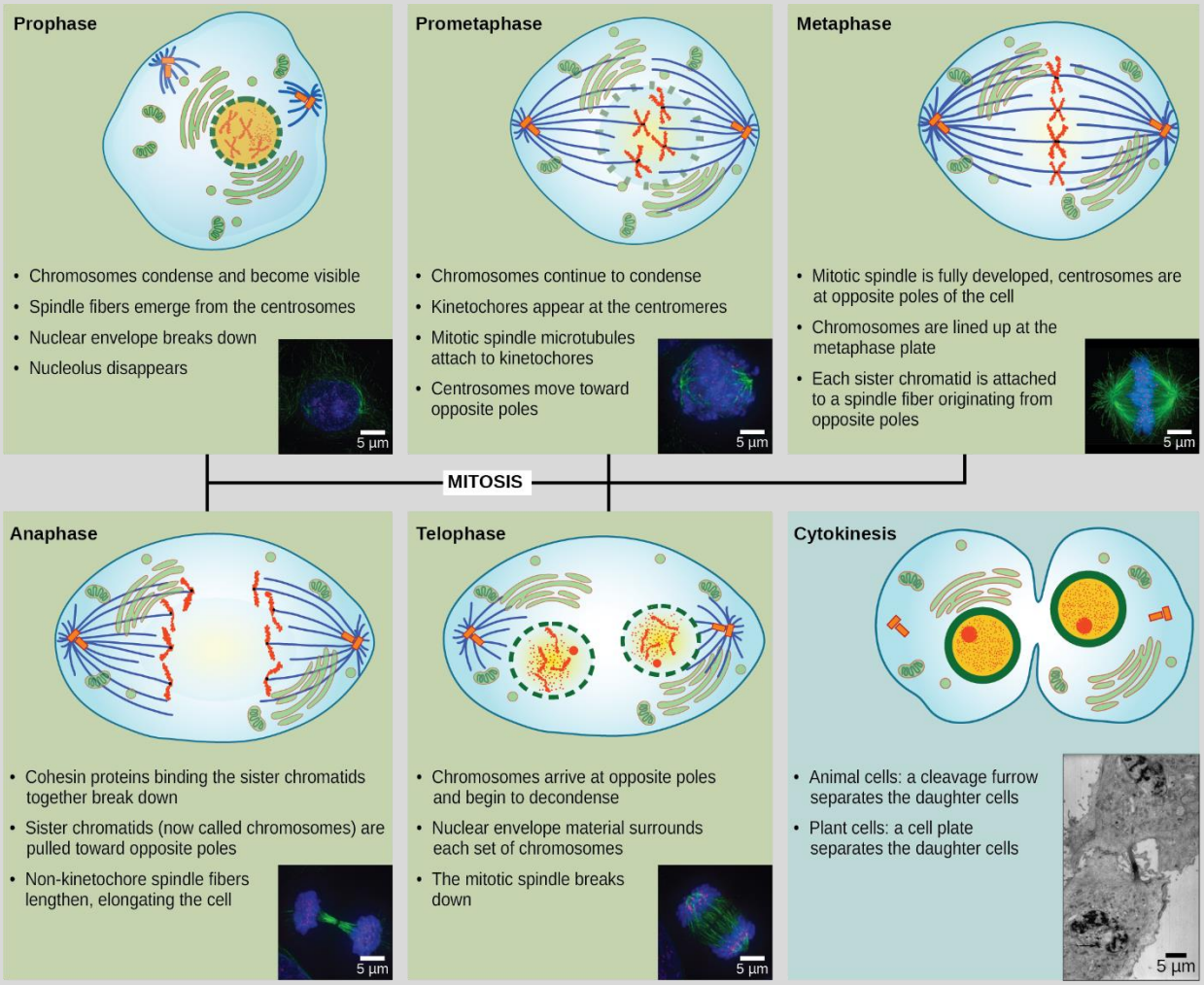


Figure 10.6 Karyokinesis (or mitosis) is divided into five stages—prophase, prometaphase, metaphase, anaphase, and telophase. The pictures at the bottom were taken by fluorescence microscopy (hence, the black background) of cells artificially stained by fluorescent dyes: blue fluorescence indicates DNA (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)

Which of the following is the correct order of events in mitosis?

a. Sister chromatids line up at the metaphase plate. The kinetochore becomes attached to the mitotic spindle. The nucleus reforms and the cell divides. Cohesin proteins break down and the sister chromatids separate.
b. The kinetochore becomes attached to the mitotic spindle. Cohesin proteins break down and the sister chromatids separate. Sister chromatids line up at the metaphase plate. The nucleus reforms and the cell divides.
c. The kinetochore becomes attached to the cohesin proteins. Sister chromatids line up at the metaphase plate. The kinetochore breaks down and the sister chromatids separate. The nucleus reforms and the cell divides.
d. The kinetochore becomes attached to the mitotic spindle. Sister chromatids line up at the metaphase plate. Cohesin proteins break down and the sister chromatids separate. The nucleus reforms and the cell divides.

Prophase (the "first phase"): the nuclear envelope starts to dissociate into small vesicles, and the membranous organelles (such as the Golgi complex [Golgi apparatus] and the endoplasmic reticulum), fragment and disperse toward the periphery of the cell. The nucleolus disappears (disperses) as well, and the centrosomes begin to move 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 now become visible under a light microscope.

Prometaphase (the "first change phase"): Many processes that began in prophase continue to advance. The remnants of the nuclear envelope fragment further, and the mitotic spindle continues to develop as more microtubules assemble and stretch across the length of the former nuclear area. Chromosomes become even more condensed and discrete. Each sister chromatid develops a protein structure called a **kinetochore** in its centromeric region (Figure 10.7). The proteins of the kinetochore attract and bind to the 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.

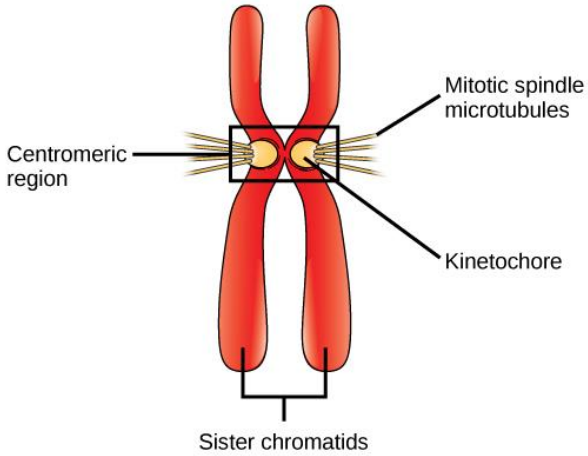


Figure 10.7 During prometaphase, mitotic spindle microtubules from opposite poles attach to each sister chromatid at the kinetochore. In anaphase, the connection between the sister chromatids breaks down, and the microtubules pull the chromosomes toward opposite poles.

Metaphase (the "change phase"): All the chromosomes are aligned in a plane called the **metaphase plate**, or the equatorial plane, roughly 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.

Anaphase ("upward phase"): The cohesin proteins degrade, and the sister chromatids separate at the centromere. Each chromatid, now called a single 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.

Telophase (the “distance phase”): the chromosomes reach the opposite poles and begin to *decondense* (unravel), relaxing once again into a stretched-out chromatin configuration. The mitotic spindles are depolymerized into tubulin monomers that will be used to assemble cytoskeletal components for each daughter cell. Nuclear envelopes form around the chromosomes, and nucleosomes appear within the nuclear area.

Cytokinesis

Cytokinesis, or “cell motion,” is sometimes viewed as 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 However, as we have seen earlier, cytokinesis can also be viewed as a separate phase, which may or may not take place following mitosis. If cytokinesis does take place, cell 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 animal cells, cytokinesis typically starts during late anaphase. A contractile ring composed of actin filaments forms just inside the plasma membrane at the former metaphase plate. The actin filaments pull the equator of the cell inward, forming a fissure. This fissure is called the **cleavage furrow**. The furrow deepens as the actin ring contracts, and eventually the membrane is cleaved in two (Figure 10.8).

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. The Golgi membranes become parts of the plasma membrane on either side of the new cell wall (Figure 10.8).

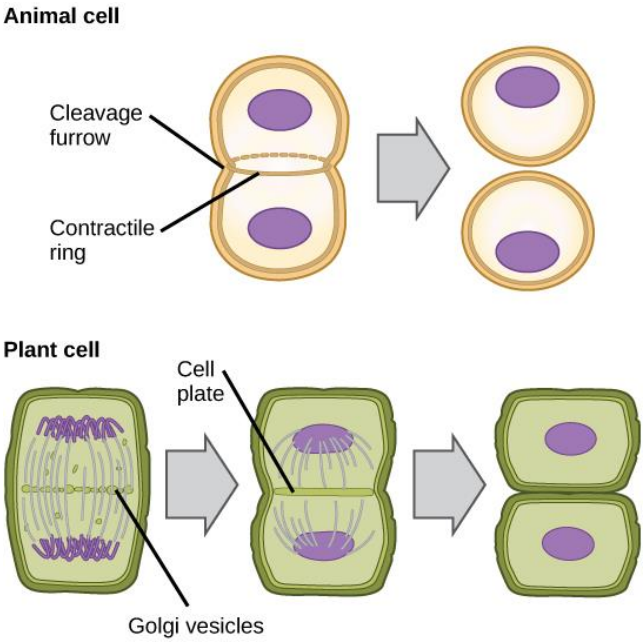


Figure 10.8 During cytokinesis in animal cells, a ring of actin filaments forms at the metaphase plate. The ring contracts, forming a cleavage furrow, which divides the cell in two. In plant cells, Golgi vesicles coalesce at the former metaphase plate, forming a phragmoplast. A cell plate formed by the 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.

G_0 Phase

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, and cytokinesis. Cells in **G_0 phase** 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 due to environmental conditions such as availability of nutrients, or stimulation by growth factors. The cell will remain in this phase until conditions improve or until an external signal triggers the onset of G_1. Other cells that never or rarely divide, such as mature cardiac muscle and nerve cells, remain in G_0 permanently.

SCIENTIFIC METHOD CONNECTION

Determine the Time Spent in Cell-Cycle Stages

Problem: How long does a cell spend in interphase compared to each stage of mitosis?

Background: A prepared microscope slide of whitefish blastula cross-sections will show cells arrested in various stages of the cell cycle. (Note: It is not visually possible to separate the stages of interphase from each other, but the mitotic stages are readily identifiable.) If 100 cells are examined, the number of cells in each identifiable cell-cycle stage will give an estimate of the time it takes for the cell to complete that stage.

Problem Statement: Given the events included in all of interphase and those that take place in each stage of mitosis, estimate the length of each stage based on a 24-hour cell cycle. Before proceeding, state your hypothesis.

Test your hypothesis: Test your hypothesis by doing the following:

1. Place a fixed and stained microscope slide of whitefish blastula cross-sections under the scanning objective of a light microscope.
2. Locate and focus on one of the sections using the low-power objective of your microscope. Notice that the section is a circle composed of dozens of closely packed individual cells.
3. Switch to the medium-power objective and refocus. With this objective, individual cells are clearly visible, but the chromosomes will still be very small.
4. Switch to the high-power objective and slowly move the slide left to right, and up and down to view all the cells in the section (Figure 10.9). As you scan, you will notice that most of the cells are not undergoing mitosis but are in the interphase period of the cell cycle.

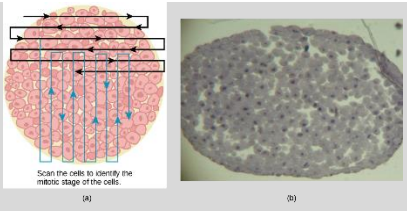


Figure 10.9 Slowly scan whitefish blastula cells with the high-power objective as illustrated in image (a) to identify their mitotic stage. (b) A microscopic image of the scanned cells is shown. (credit “micrograph”: modification of work by Linda Flora; scale-bar data from Matt Russell)

5. Practice identifying the various stages of the cell cycle, using the drawings of the stages as a guide.
6. Once you are confident about your identification, begin to record the stage of each cell you encounter as you scan left to right, and top to bottom across the blastula section.
7. Keep a tally of your observations and stop when you reach 100 cells identified.
8. The larger the sample size (total number of cells counted), the more accurate the results. If possible, gather and record group data prior to calculating percentages and making estimates.

Record your observations: Make a table similar to Table 10.1 within which to record your observations.

Results of Cell Stage Identification

Phase or Stage	Individual Totals	Group Totals	Percent
Interphase			
Prophase			
Metaphase			
Anaphase			
Telophase			
Cytokinesis			
Totals	100	100	100 percent

Table10.1

Analyze your data/report your results: To find the length of time whitefish blastula cells spend in each stage, multiply the percent (recorded as a decimal) by 24 hours. Make a table similar to Table 10.2 to illustrate your data.

Estimate of Cell Stage Length

Phase or Stage	Percent	Time in Hours
Interphase		
Prophase		
Metaphase		
Anaphase		
Telophase		
Cytokinesis		

Table10.2

Draw a conclusion: Did your results support your estimated times? Were any of the outcomes unexpected? If so, discuss those events in that stage that may have contributed to the calculated time.

Chapter Attributions

Cancer and the Cell Cycle

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

- 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 redundancy and 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. If changes to the 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, however, 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 (“-oma”) can result.

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, however, 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. However, if the atypical daughter cells are able to undergo further cell divisions, subsequent generations of cells may accumulate even more mutations, some possibly 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.

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 50 percent of all human tumor cells. This discovery is not surprising in light of the multiple roles that the p53 protein plays at the G_1 checkpoint. A cell with a faulty p53 may fail to detect errors present in the genomic DNA (Figure 10.14). 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, however, cannot trigger apoptosis.

VISUAL CONNECTION

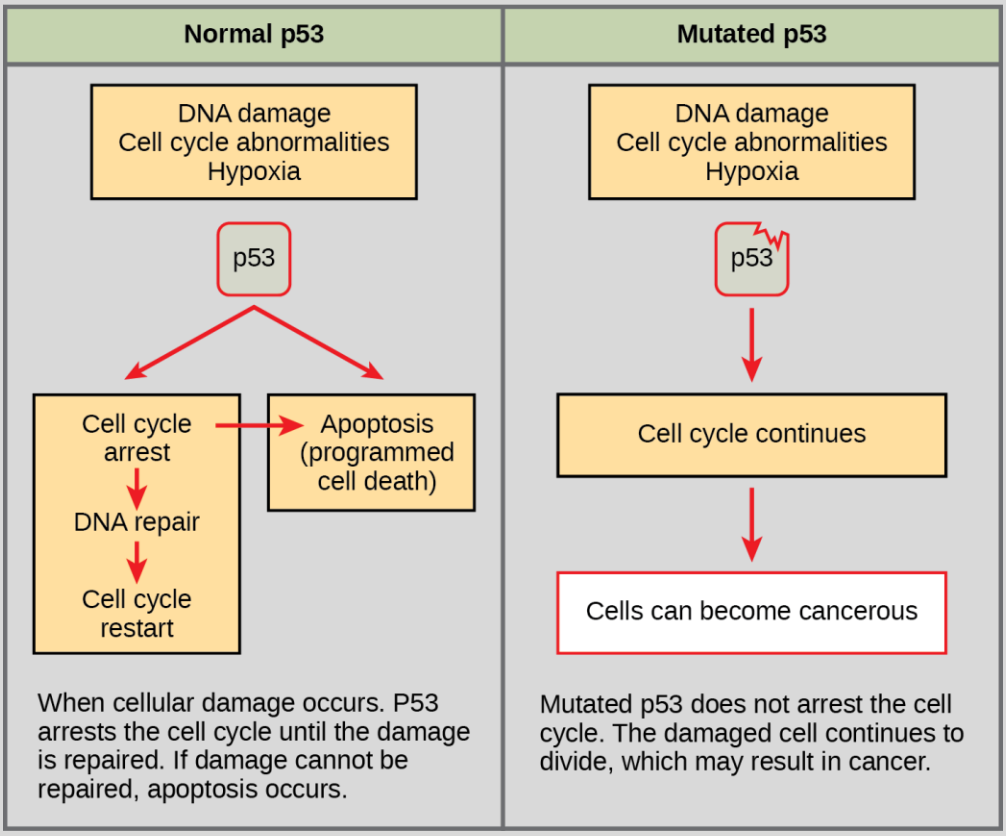


Figure 10.14 The role of normal p53 is to monitor DNA and the supply of oxygen (hypoxia is a condition of reduced oxygen supply). If damage is detected, p53 triggers repair mechanisms. If repairs are unsuccessful, p53 signals apoptosis. A cell with an abnormal p53 protein cannot repair damaged DNA and thus cannot signal apoptosis. Cells with abnormal p53 can become cancerous. (credit: modification of work by Thierry Soussi)

Human papillomavirus can cause cervical cancer. The virus encodes E6, a protein that binds p53. Based on this fact and what you know about p53, what effect do you think E6 binding has on p53 activity?

a. E6 activates p53

b. E6 inactivates p53

c. E6 mutates p53

d. E6 binding marks p53 for degradation

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 G_1 checkpoint is severely compromised and the cell proceeds directly from G_1 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.

LINK TO LEARNING

Watch an animation of how cancer results from errors in the cell cycle: https://www.youtube.com/embed/RZhL7LDPk8w

Chapter Attributions

16. Meiosis and Gametogenesis

The Process of Meosis

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

- Describe the behavior of chromosomes during meiosis, and the differences between the first and second meiotic divisions

- Describe the cellular events that take place during meiosis
- Explain the differences between meiosis and mitosis
- Explain the mechanisms within the meiotic process that produce genetic variation among the haploid gametes

Sexual reproduction requires the union of two specialized cells, called **gametes**, each of which contains one set of chromosomes. When gametes unite, they form a **zygote**, or fertilized egg that contains two sets of chromosomes. (Note: Cells that contain one set of chromosomes are called **haploid**; cells containing two sets of chromosomes are called **diploid**.) If the reproductive cycle is to continue for any sexually reproducing species, then the diploid cell must somehow reduce its number of chromosome sets to produce haploid gametes; otherwise, the number of chromosome sets will double with every future round of fertilization. Therefore, sexual reproduction requires a nuclear division that reduces the number of chromosome sets by half.

Most animals and plants and many unicellular organisms are diploid and therefore have two sets of chromosomes. In each **somatic cell** of the organism (all cells of a multicellular organism except the gametes or reproductive cells), the nucleus contains two copies of each chromosome, called **homologous chromosomes**. Homologous chromosomes are matched pairs containing the same genes in identical locations along their lengths. Diploid organisms inherit one copy of each homologous chromosome from each parent.

Meiosis is the *nuclear division* that forms haploid cells from diploid cells, and it employs many of the same cellular mechanisms as mitosis. However, as you have learned, **mitosis** produces daughter cells whose nuclei are genetically identical to the original parent nucleus. In mitosis, both the parent and the daughter nuclei are at the same "ploidy level"—diploid in the case of most multicellular most animals. Plants use mitosis to grow as sporophytes, and to grow and produce eggs and sperm as gametophytes; so they use mitosis for both haploid and diploid cells (as well as for all other ploidies). In meiosis, the starting nucleus is always diploid and the daughter nuclei that result are haploid. To achieve this reduction in chromosome number, meiosis consists of one round of chromosome replication followed by two rounds of nuclear division. Because the events that occur during each of the division stages are analogous to the events of mitosis, the same stage names are assigned. However, because there are two rounds

of division, the major process and the stages are designated with a "I" or a "II." Thus, **meiosis I** is the first round of meiotic division and consists of prophase I, prometaphase I, and so on. Likewise, **Meiosis II** (during which the second round of meiotic division takes place) includes prophase II, prometaphase II, and so on.

Meiosis I

Meiosis is preceded by an interphase consisting of G_1, S, and G_2 phases, which are nearly identical to the phases preceding mitosis. The G_1 phase (the "first gap phase") is focused on cell growth. During the S phase—the second phase of interphase—the cell copies or *replicates* the DNA of the chromosomes. Finally, in the G_2 phase (the "second gap 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—*sister chromatids* that are held together at the centromere by **cohesin** proteins, which hold the chromatids together until anaphase II.

Prophase I

Early in prophase I, before the chromosomes can be seen clearly with a microscope, 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 closer together. Recall that in mitosis, homologous chromosomes do not pair together. The **synaptonemal complex**, a lattice of proteins between the homologous chromosomes, first forms at specific locations and then spreads outward to cover the entire length of the chromosomes. The tight pairing of the homologous chromosomes is called *synapsis*. In **synapsis**, the genes on the chromatids of the homologous chromosomes are aligned precisely with each other. The synaptonemal complex supports the exchange of chromosomal segments between homologous nonsister chromatids—a process called **crossing over**. Crossing over can be observed visually after the exchange as *chiasmata* (singular = chiasma) (Figure 11.2).

In humans, even though the X and Y sex chromosomes are not completely homologous (that is, most of their genes differ), there is 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.

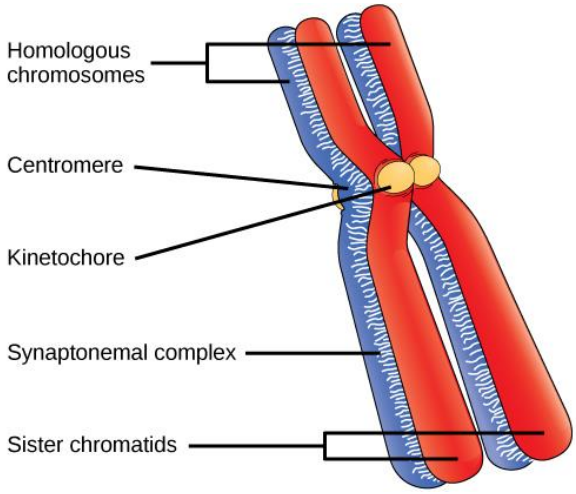


Figure 11.2 Early in prophase I, homologous chromosomes come together to form a synapse. The chromosomes are bound tightly together and in perfect alignment by a protein lattice called a synaptonemal complex and by cohesin proteins at the centromere.

Located at intervals along the synaptonemal complex are large protein assemblies called **recombination nodules**. These assemblies mark the points of later chiasmata and mediate the multistep process of **crossover**—or genetic recombination—between the nonsister chromatids. Near the recombination nodule, the double-stranded DNA of each chromatid is cleaved, the cut ends are modified, and a new connection is made between the nonsister chromatids. As prophase I progresses, the synaptonemal complex begins to break down and the chromosomes begin to condense. When the synaptonemal complex is gone, the homologous chromosomes remain attached to each other at the centromere and at chiasmata. The chiasmata remain until anaphase I. The number of chiasmata varies according to the species and the length of the chromosome. There must be at least one chiasma per chromosome for proper separation of homologous chromosomes during meiosis I, but there may be as many as 25. Following crossover, the synaptonemal complex breaks down and the cohesin connection between homologous pairs is removed. At the end of prophase I, the pairs are held together only at the chiasmata (Figure 11.3). These pairs are called **tetrads** because the four sister chromatids of each pair of homologous chromosomes are now visible.

The crossover events are the first source of genetic variation in the nuclei produced by meiosis. A single crossover event between homologous nonsister chromatids leads to a reciprocal exchange of equivalent DNA between a maternal chromosome and a paternal chromosome. When a recombinant sister chromatid is moved into a gamete cell it will carry some DNA from one parent and some DNA from the other parent. The recombinant chromatid has a combination of maternal and paternal genes that did not exist before the crossover. Crossover events can occur almost anywhere along the length of the synapsed chromosomes. Different cells undergoing meiosis will therefore produce different recombinant chromatids, with varying combinations of maternal and parental genes. Multiple crossovers in an arm of the chromosome have the same effect, exchanging segments of DNA to produce genetically recombined chromosomes.

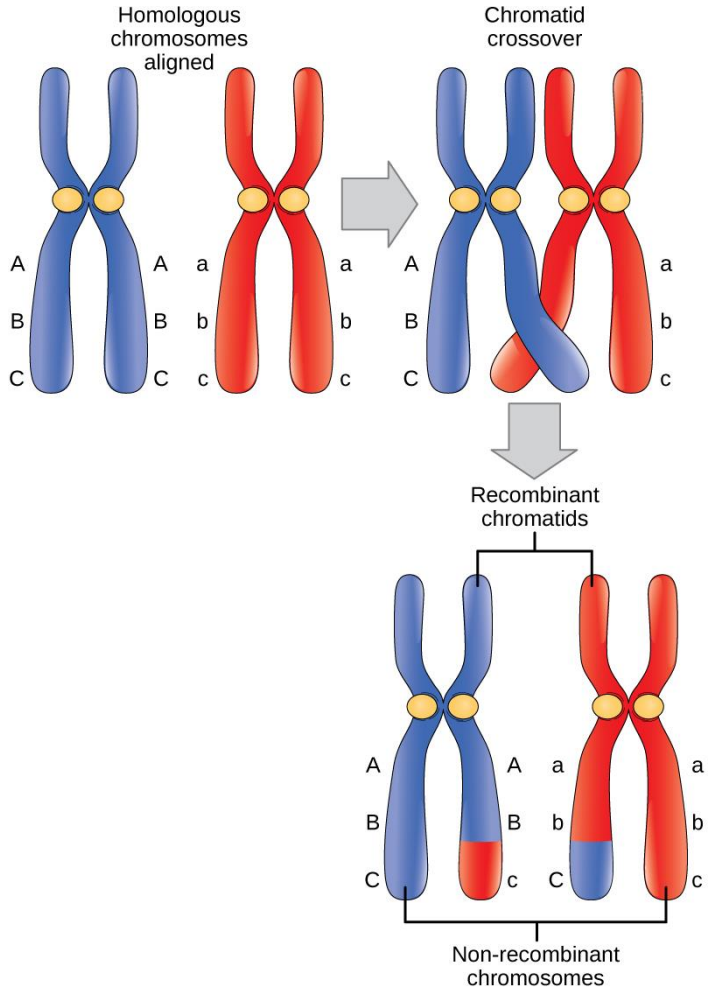


Figure 11.3 Crossover occurs between *nonsister chromatids of homologous chromosomes.* The result is an exchange of genetic material between homologous chromosomes.

Prometaphase I

The key event in prometaphase I is the attachment of the spindle fiber microtubules to the kinetochore proteins at the centromeres. Kinetochore proteins are multiprotein complexes that bind the centromeres of a chromosome to the microtubules of the mitotic spindle. Microtubules grow from microtubule-organizing centers (MTOCs). In animal cells, MTOCs are centrosomes located at opposite poles of the cell. The microtubules from each pole move toward the middle of the cell and attach to one of the kinetochores of the two fused homologous chromosomes. Each member of the homologous pair attaches to a microtubule extending from opposite poles of the cell so that in the next phase, the microtubules can pull the homologous pair apart. A spindle fiber that has attached to a kinetochore is called a *kinetochore microtubule*. At the end of prometaphase I, each tetrad is attached to microtubules from both poles, with one homologous chromosome facing each pole. The homologous chromosomes are still held together at the chiasmata. In addition, the nuclear membrane has broken down entirely.

Metaphase I

During metaphase I, the homologous chromosomes are arranged at the **metaphase plate**—roughly in the midline of the cell, with the kinetochores facing opposite poles. The homologous pairs orient themselves randomly at the equator. For example, if the two homologous members of chromosome 1 are labeled *a* and *b*, then the chromosomes could line up a-b or b-a. This is important in determining the genes carried by a gamete, as each will only receive one of the two homologous chromosomes. (Recall that after crossing over takes place, homologous chromosomes are not identical. They contain slight differences in their genetic information, causing each gamete to have a unique genetic makeup.)

The randomness in the alignment of recombined chromosomes at the metaphase plate, coupled with the crossing over events between nonsister chromatids, are responsible for much of the genetic variation in the offspring. To clarify this further, remember that the homologous chromosomes of a sexually reproducing organism are originally inherited as two separate sets,

one from each parent. Using humans as an example, one set of 23 chromosomes is present in the egg donated by the mother. The father provides the other set of 23 chromosomes in the sperm that fertilizes the egg. Every cell of the multicellular offspring has copies of the original two sets of homologous chromosomes. In prophase I of meiosis, the homologous chromosomes form the tetrads. In metaphase I, these pairs line up at the midway point between the two poles of the cell to form the metaphase plate. Because there is an equal chance that a microtubule fiber will encounter a maternally or paternally inherited chromosome, the arrangement of the tetrads at the metaphase plate is random. Thus, any maternally inherited chromosome may face either pole. Likewise, any paternally inherited chromosome may also face either pole. *The orientation of each tetrad is independent of the orientation of the other 22 tetrads.*

This event—the *random* (or *independent*) assortment of homologous chromosomes at the metaphase plate—is the second mechanism that introduces variation into the gametes or spores. In each cell that undergoes meiosis, the arrangement of the tetrads is different. The number of variations is dependent on the number of chromosomes making up a set. There are two possibilities for orientation at the metaphase plate; the possible number of alignments therefore equals 2^n in a diploid cell, where n is the number of chromosomes per haploid set. Humans have 23 chromosome pairs, which results in over eight million (2^{23}) possible genetically-distinct gametes just from the random alignment of chromosomes at the metaphase plate. This number does not include the variability that was previously produced by crossing over between the nonsister chromatids. Given these two mechanisms, it is highly unlikely that any two haploid cells resulting from meiosis will have the same genetic composition (Figure 11.4).

To summarize, meiosis I creates genetically diverse gametes in two ways. First, during prophase I, crossover events between the nonsister chromatids of each homologous pair of chromosomes generate recombinant chromatids with new combinations of maternal and paternal genes. Second, the random assortment of tetrads on the metaphase plate produces unique combinations of maternal and paternal chromosomes that will make their way into the gametes.

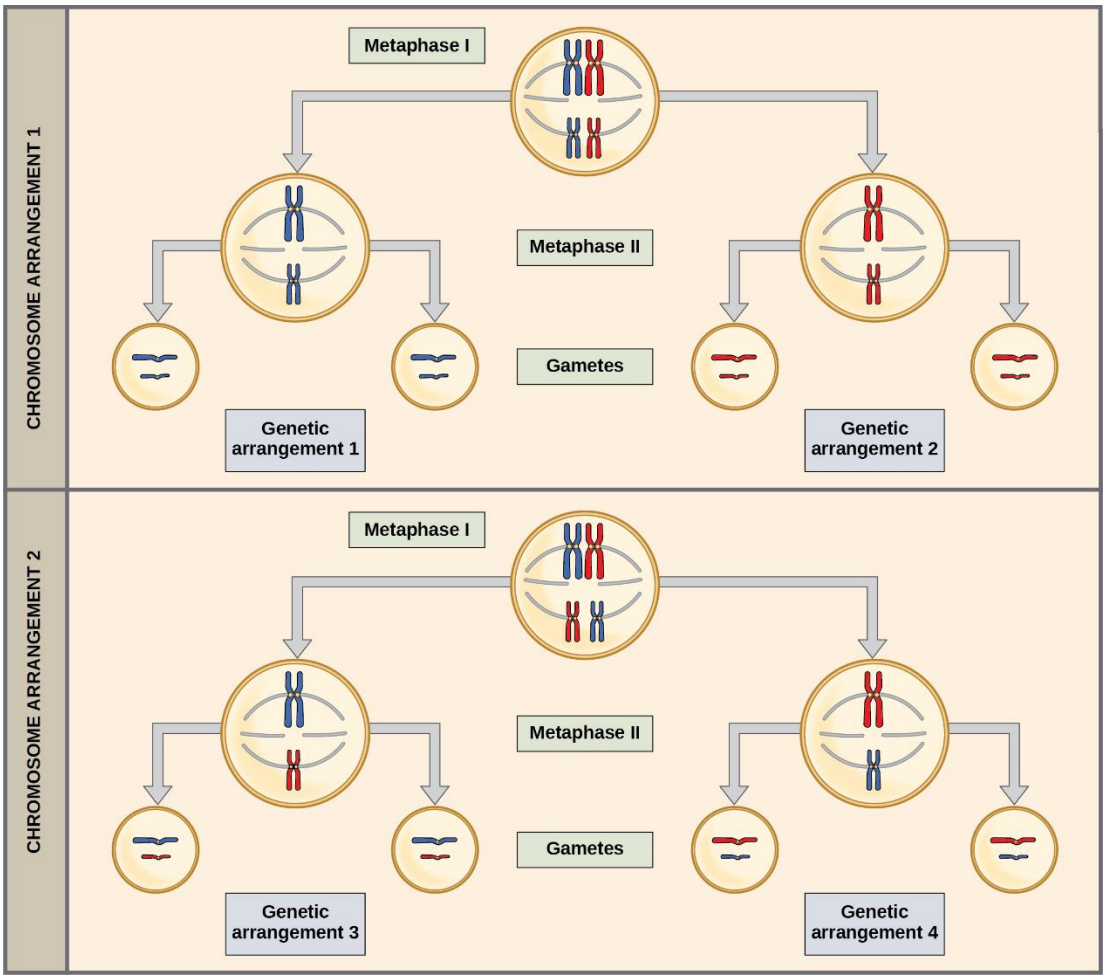


Figure 11.4 Random, independent assortment during metaphase I can be demonstrated by considering a cell with a set of two chromosomes (n = 2). In this case, there are two possible arrangements at the equatorial plane in metaphase I. The total possible number of different gametes is 2^n, where n equals the number of chromosomes in a set. In this example, there are four possible genetic combinations for the gametes. With n = 23 in human cells, there are over eight million possible combinations of paternal and maternal chromosomes.

Anaphase I

In anaphase I, the microtubules pull the linked chromosomes apart. The sister chromatids remain tightly bound together at the centromere. The chiasmata are broken in anaphase I as the microtubules attached to the fused kinetochores pull the homologous chromosomes apart (Figure 11.5).

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 separated sets of chromatids produced during 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 result of the first meiotic division of a diploid cell. The cells are haploid because at each pole, there is just one of each pair of the homologous chromosomes. Therefore, only one full set of the chromosomes is present. This is why the cells are considered haploid—there is only one chromosome set, *even though each chromosome 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.

LINK TO LEARNING

Review the process of meiosis, observing how chromosomes align and migrate, at Meiosis: An Interactive Animation (https://www.cellsalive.com/meiosis.htm)

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 are similar to mitosis, except that each dividing cell has only one set of homologous chromosomes, each with two chromatids. Therefore, each cell has half the number of sister chromatids to separate out as a diploid cell undergoing mitosis. In terms of

chromosomal content, cells at the start of meiosis II are similar to haploid cells in G_2, preparing to undergo mitosis.

Prophase II

If the chromosomes decondensed in telophase I, they condense again. If nuclear envelopes were formed, they fragment into vesicles. The MTOCs that were duplicated during interkinesis move away from each other toward opposite poles, and new spindles are formed.

Prometaphase II

The nuclear envelopes are completely broken down, and the spindle is fully formed. Each sister chromatid forms an individual kinetochore that attaches to microtubules from opposite poles.

Metaphase II

The sister chromatids are maximally condensed and aligned at the equator of the cell.

Anaphase II

The sister chromatids are pulled apart by the kinetochore microtubules and move toward opposite poles. Nonkinetochore microtubules elongate the cell.

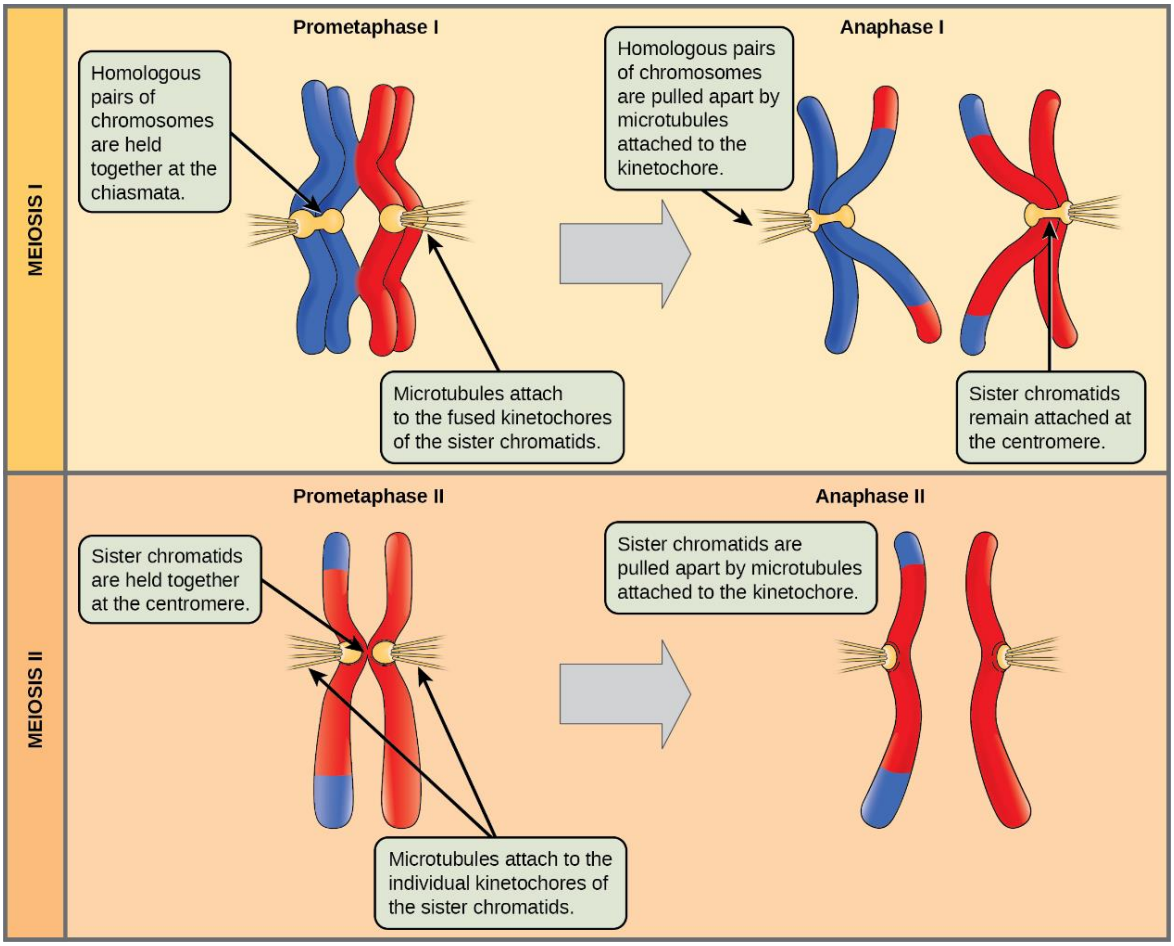


Figure 11.5 The process of chromosome alignment differs between meiosis I and meiosis II. In prometaphase I, microtubules attach to the fused kinetochores of homologous chromosomes, and the homologous chromosomes are arranged at the midline of the cell (the metaphase plate) in metaphase I. In anaphase I, the homologous chromosomes separate. In prometaphase II, microtubules attach to the kinetochores of sister chromatids, and the sister chromatids are arranged at the midpoint of the cells in metaphase II. In anaphase II, the sister chromatids separate.

Telophase II and Cytokinesis

The chromosomes arrive at opposite poles and begin to decondense. Nuclear envelopes form around the chromosomes. If the parent cell was diploid, as is most commonly the case, then cytokinesis now separates the two cells into four unique haploid cells. The cells produced are genetically unique because of the random assortment of paternal and maternal homologs and

because of the recombination 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 11.6.

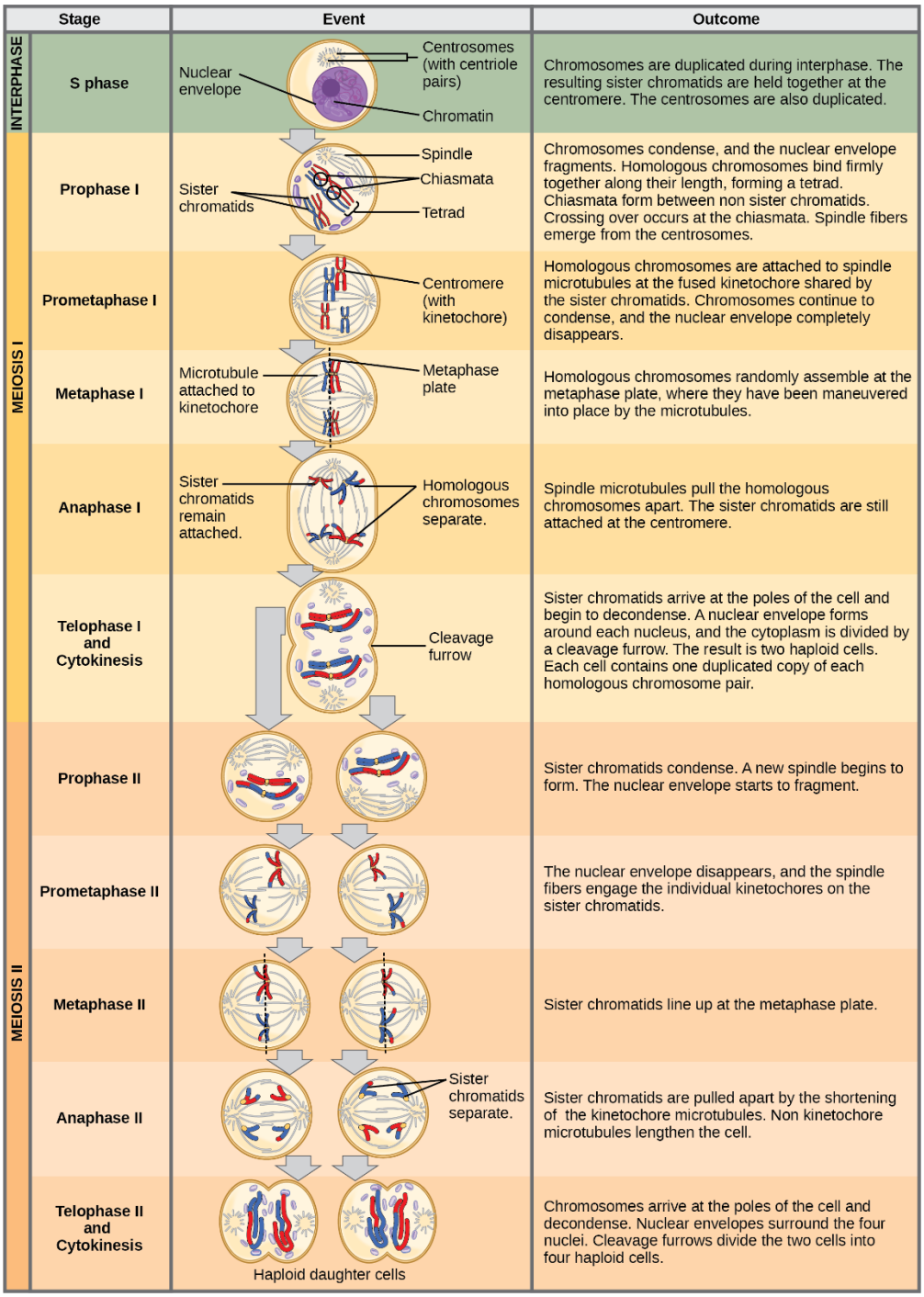


Figure 11.6 An animal cell with a diploid number of four (2*n* = 4) proceeds through the stages of meiosis to form four haploid daughter cells.

Comparing Meiosis and Mitosis

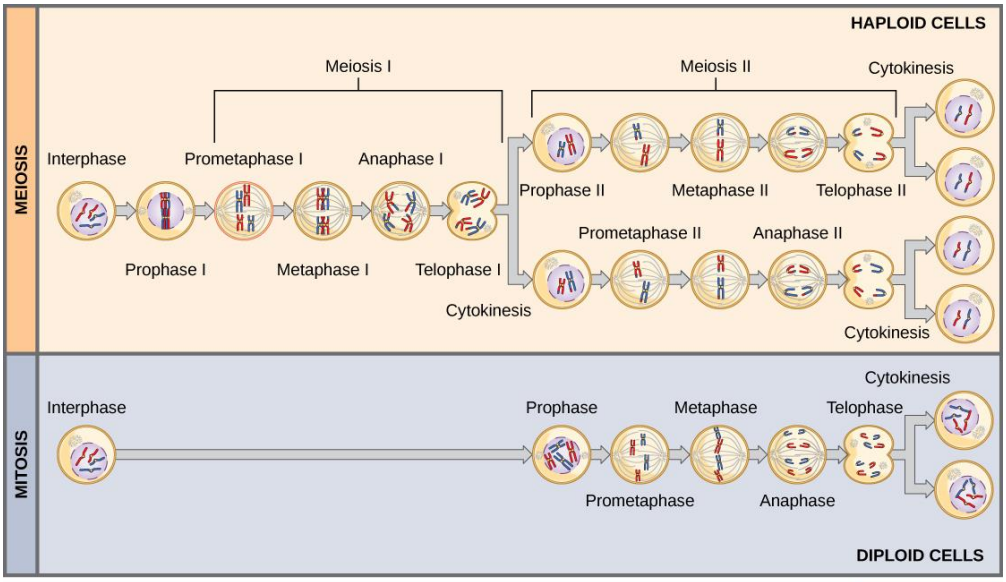
Mitosis and meiosis are both forms of division of the nucleus in eukaryotic cells. They share some similarities, but also exhibit a number of important and distinct differences that lead to very different outcomes (Figure 11.7). Mitosis is a single nuclear division that results in two nuclei that are usually partitioned into two new cells. The nuclei resulting from a mitotic division are genetically identical to the original nucleus. They have the same number of sets of chromosomes: one set in the case of haploid cells and two sets in the case of diploid cells. In contrast, meiosis consists of two nuclear divisions resulting in four nuclei that are usually partitioned into four new, genetically distinct cells. The four nuclei produced during meiosis are not genetically identical, and they contain one chromosome set only. This is half the number of chromosome sets in the original cell, which is diploid.

The main differences between mitosis and meiosis occur in meiosis I, which is a very different nuclear division than mitosis. In meiosis I, the homologous chromosome pairs physically meet and are bound together with the synaptonemal complex. Following this, the chromosomes develop chiasmata and undergo crossover between nonsister chromatids. In the end, the chromosomes line up along the metaphase plate as tetrads—with kinetochore fibers from opposite spindle poles attached to each kinetochore of a homolog to form a tetrad. *All of these events occur only in meiosis I.*

When the chiasmata resolve and the tetrad is broken up with the homologs moving to one pole or another, the ploidy level—the number of sets of chromosomes in each future nucleus—has been reduced from two to one. For this reason, meiosis I is referred to as a **reductional division**. There is no such reduction in ploidy level during mitosis.

Meiosis II is analogous to a mitotic division. In this case, the duplicated chromosomes (only one set of them) line up on the metaphase plate with divided kinetochores attached to kinetochore fibers from opposite poles. During anaphase II, as in mitotic anaphase, the kinetochores divide and one sister chromatid—now referred to as a chromosome—is pulled to one pole while the other sister chromatid is pulled to the other pole. If it were not for the fact that there had been crossover, the two products of each individual meiosis II division would be identical (as in mitosis). Instead, they are different because there has always been at least one crossover per chromosome. Meiosis II is not a reduction division because although there are fewer copies of

the genome in the resulting cells, there is still one set of chromosomes, as there was at the end of meiosis I.



						OUTCOME
PROCESS	DNA synthesis	Synapsis of homologous chromosomes	Crossover	Homologous chromosomes line up at metaphase plate	Sister chromatids line up at metaphase plate	Number and genetic composition of daughter cells
MEIOSIS	Occurs in S phase of interphase	During prophase I	During prophase I	During metaphase I	During metaphase II	Four haploid cells at the end of meiosis II
MITOSIS	Occurs in S phase of interphase	Does not occur in mitosis	Does not occur in mitosis	Does not occur in mitosis	During metaphase	Two diploid cells at the end of mitosis

Figure 11.7 Meiosis and mitosis are both preceded by one cycle 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.

EVOLUTION CONNECTION

The Mystery of the Evolution of Meiosis

Some characteristics of organisms are so widespread and fundamental that it is sometimes difficult to remember that they evolved like other simple traits. Meiosis is such an extraordinarily complex series of cellular events that biologists have had trouble testing hypotheses concerning how it may have evolved. Although meiosis is inextricably entwined

with sexual reproduction and its advantages and disadvantages, it is important to separate the questions of the evolution of meiosis and the evolution of sex, because early meiosis may have been advantageous for different reasons than it is now. Thinking outside the box and imagining what the early benefits from meiosis might have been is one approach to uncovering how it may have evolved.

Meiosis and mitosis share obvious cellular processes, and it makes sense that meiosis evolved from mitosis. The difficulty lies in the clear differences between meiosis I and mitosis. Adam Wilkins and Robin Holliday[1] summarized the unique events that needed to occur for the evolution of meiosis from mitosis. These steps are homologous chromosome pairing and synapsis, crossover exchanges, sister chromatids remaining attached during anaphase, and suppression of DNA replication in interphase. They argue that the first step is the hardest and most important and that understanding how it evolved would make the evolutionary process clearer. They suggest genetic experiments that might shed light on the evolution of synapsis.

There are other approaches to understanding the evolution of meiosis in progress. Different forms of meiosis exist in single-celled protists. Some appear to be simpler or more "primitive" forms of meiosis. Comparing the meiotic divisions of different protists may shed light on the evolution of meiosis. Marilee Ramesh and colleagues[2] compared the genes involved in meiosis in protists to understand when and where meiosis might have evolved. Although research is still ongoing, recent scholarship into meiosis in protists suggests that some aspects of meiosis may have evolved later than others. This kind of genetic comparison can tell us what aspects of meiosis are the oldest and what cellular processes they may have borrowed from in earlier cells.

LINK TO LEARNING

Click through the steps of this interactive animation to compare the meiotic process of cell division to that of mitosis at How Cells Divide (https://www.cellsalive.com/meiosis.htm).

Footnotes

- 1 Adam S. Wilkins and Robin Holliday, "The Evolution of Meiosis from Mitosis," *Genetics* 181 (2009): 3–12.

- 2 Marilee A. Ramesh, Shehre-Banoo Malik and John M. Logsdon, Jr, “A Phylogenetic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis,” *Current Biology* 15 (2005):185–91.

Chapter Attributions

"Biology 2e" Mary Ann Clark, Matthew Douglas, Jung Choi, OpenStax is licensed under CC BY 4.0. Access for free at https://openstax.org/books/biology-2e/pages/11-1-the-process-of-meiosis

Sexual Reproduction

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

- Explain that meiosis and sexual reproduction are highly evolved traits
- Identify variation among offspring as a potential evolutionary advantage of sexual reproduction
- Describe the three different life-cycle types among sexually reproducing multicellular organisms.

Sexual reproduction was likely 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 also 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 should be similarly successful. There is also the obvious benefit to an organism that can produce offspring whenever circumstances are favorable by asexual budding, fragmentation, or by producing eggs asexually. These methods of reproduction do not require 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. In sexual populations, the males are not producing the offspring themselves, so hypothetically an asexual population could grow twice as fast.

However, multicellular organisms that exclusively depend on asexual reproduction are exceedingly rare. Why are meiosis and sexual reproductive strategies so common? These are important (and as yet unanswered) questions in biology, even though they have been the focus of much research beginning in the latter half of the 20th 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. Thus, on average, a sexually reproducing population will leave more descendants than an otherwise similar asexually reproducing population. The only source of variation in asexual organisms is mutation. Mutations that take place during the formation of germ cell lines are also the ultimate source of variation in sexually reproducing organisms. However, in contrast to mutation during asexual reproduction, the mutations during sexual reproduction can be 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.

EVOLUTION CONNECTION

The Red Queen Hypothesis

Genetic variation is the outcome of sexual reproduction, but why are ongoing variations necessary, even under seemingly stable environmental conditions? Enter the Red Queen hypothesis, first proposed by Leigh Van Valen in 1973.[3] The concept was named in reference to the Red Queen's race in Lewis Carroll's book, *Through the Looking-Glass*.

All species **coevolve** (evolve together) 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 a reproductive edge over close competitors, predators, parasites, or even prey. However, survival of any given genotype or phenotype in a population is dependent on the reproductive fitness of other genotypes or phenotypes within a given species. The only method that will allow a coevolving species to maintain its own share of the resources is to also *continually improve its* **fitness** (the capacity of the members to produce more reproductively viable offspring relative to others within a species). 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 coevolution between competing species.

Life Cycles of Sexually Reproducing Organisms

Fertilization and meiosis alternate in sexual **life cycles**. What happens between these two events depends on the organism's "reproductive strategy." The process of meiosis reduces the chromosome number by half. Fertilization, the joining of two haploid gametes, restores the diploid condition. Some organisms have a multicellular diploid stage that is most obvious and only produce haploid reproductive cells. Animals, including humans, have this type of life cycle. Other organisms, such as fungi, have a multicellular haploid stage that is most obvious. Plants and some algae have alternation of generations, in which they have multicellular diploid and haploid life stages that are apparent to different degrees depending on the group.

Nearly all animals employ a diploid-dominant life-cycle strategy in which the only haploid cells produced by the organism 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 germ cell line and meiosis to produce haploid 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 11.8).

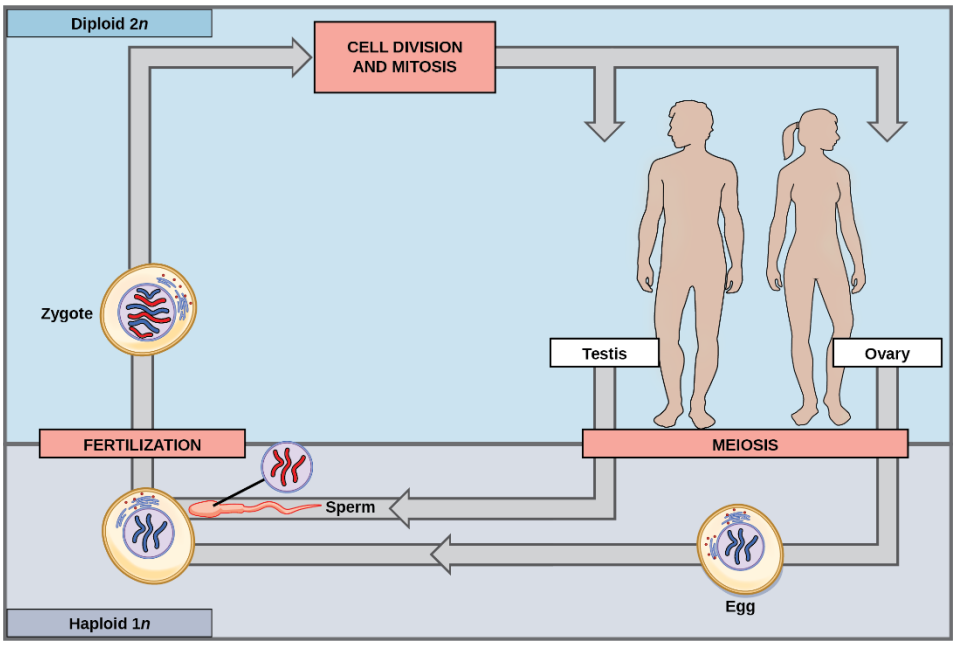


Figure 11.8 In animals, sexually reproducing adults form haploid gametes from diploid germ cells. Fusion of the gametes gives rise to a fertilized egg cell, or zygote. The zygote will undergo multiple rounds of mitosis to produce a multicellular offspring. The germ cells are generated early in the development of the zygote.

Most fungi and algae employ a life-cycle type in which the "body" of the organism—the ecologically important part of the life cycle—is haploid. 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 these spores are haploid like the "parents," they contain a new genetic combination from two parents. The spores can remain dormant for various time periods. Eventually, when conditions are favorable, the spores form multicellular haploid structures through many rounds of mitosis (Figure 11.9).

VISUAL CONNECTION

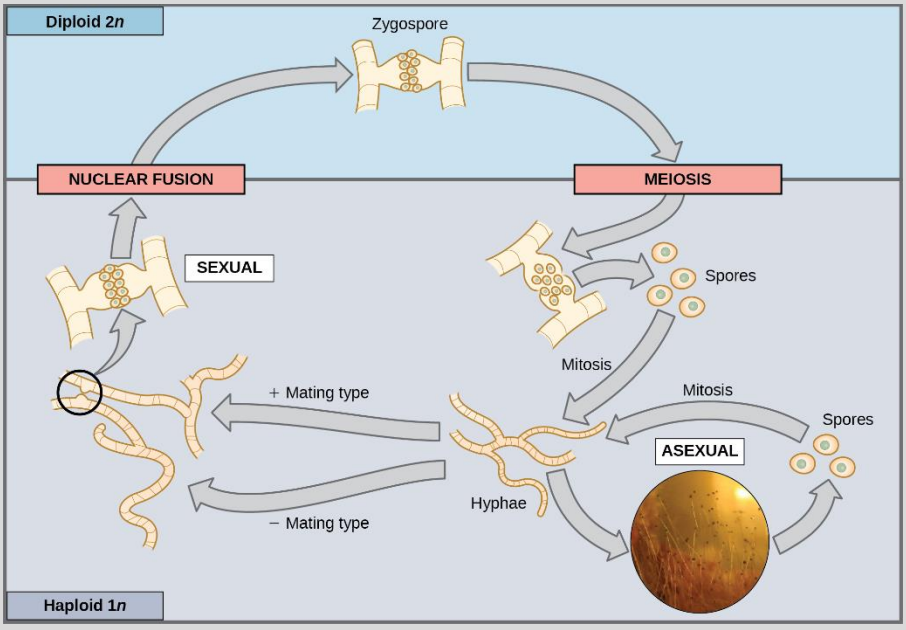


Figure 11.9 Fungi, such as black bread mold (*Rhizopus nigricans*), have a haploid multicellular stage that produces specialized haploid cells by mitosis that fuse to form a diploid zygote. The haploid multicellular stage produces specialized haploid cells by mitosis that fuse to form a diploid zygote. The zygote undergoes meiosis to produce haploid spores. Each spore gives rise to a multicellular haploid organism by mitosis. Above, different mating hyphae types (denoted as + and –) join to form a zygospore through nuclear fusion. (credit "zygomycota" micrograph: modification of work by "Fanaberka"/Wikimedia Commons)

If a mutation occurs so that a fungus is no longer able to produce a minus mating type, will it still be able to reproduce?

The third life-cycle type, employed by some algae and all plants, is a blend of the haploid-dominant and diploid-dominant extremes. Species with alternation of generations 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 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 11.10).

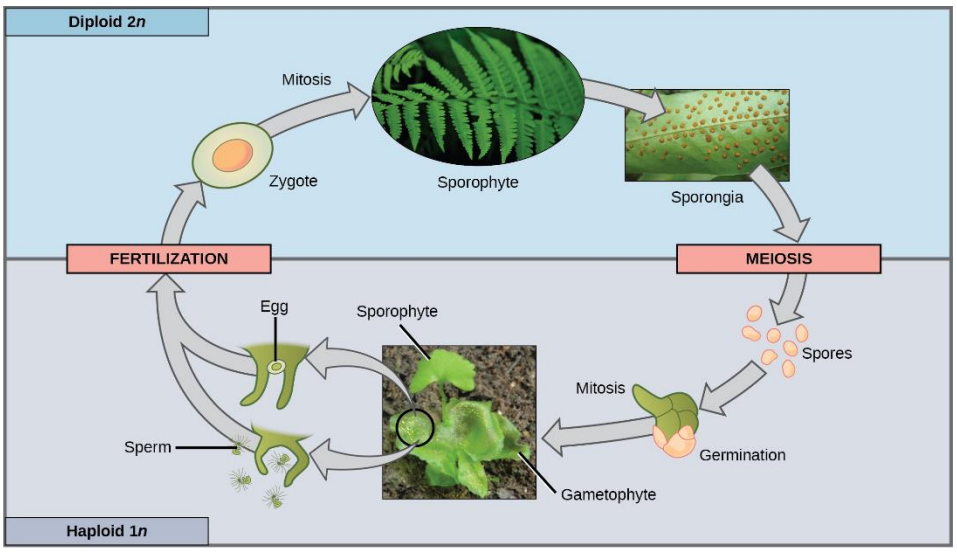


Figure 11.10 Plants have a life cycle that alternates between a multicellular haploid organism and a multicellular diploid organism. In some plants, such as ferns, both the haploid and diploid plant stages are free-living. The diploid plant is called a sporophyte because it produces haploid spores by meiosis. The spores develop into multicellular, haploid plants that are called *gametophytes* because they produce gametes. The gametes of two individuals will fuse to form a diploid zygote that becomes the sporophyte. (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)

Although all plants utilize some version of the alternation of generations, the relative size of the sporophyte and the gametophyte and the relationship between them vary greatly. In plants such as moss, the gametophyte organism is the free-living plant and the sporophyte is physically dependent on the gametophyte. In other plants, such as ferns, both the gametophyte and sporophyte plants are free-living; however, the sporophyte is much larger. In seed plants, such as magnolia trees and daisies, the gametophyte is composed of only a few cells and, in the case of the female gametophyte, is completely retained within the sporophyte.

Sexual reproduction takes many forms in multicellular organisms. 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.

Chapter Attributions

17. Genetics and Patterns of Inheritance

Characteristics and Traits

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

- Explain the relationship between genotypes and phenotypes in dominant and recessive gene systems
- 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
- Identify non-Mendelian inheritance patterns such as incomplete dominance, codominance, recessive lethals, multiple alleles, and sex linkage

Physical characteristics are expressed through genes carried on chromosomes. The genetic makeup of peas consists of two similar, or homologous, copies of each chromosome, one from each parent. Each pair of homologous chromosomes has the same linear order of genes. In other words, peas are diploid organisms in that they have two copies of each chromosome. The same is true for many other plants and for virtually all animals. Diploid organisms produce haploid gametes, which contain one copy of each homologous chromosome that unite at fertilization to create a diploid zygote.

For cases in which a single gene controls a single characteristic, a diploid organism has two genetic copies that may or may not encode the same version of that characteristic. Gene variants that arise by mutation and exist at the same relative locations on homologous chromosomes are called **alleles**. Mendel examined the inheritance of genes with just two allele forms, but it is common to encounter more than two alleles for any given gene in a natural population.

Phenotypes and Genotypes

Two alleles for a given gene in a diploid organism are 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, consisting of both physically visible and non-expressed alleles, 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 F_1 hybrid offspring had yellow pods. That is, the hybrid offspring were phenotypically identical to the true-breeding parent with yellow pods. However, we know that the allele donated by the parent with green pods was not simply lost because it reappeared in some of the F_2 offspring. Therefore, the F_1 plants must have been genotypically different from the parent with yellow pods.

The P_1 plants that Mendel used in his experiments were each homozygous for the trait he was studying. Diploid organisms that are **homozygous** at a given gene, or locus, have two identical alleles for that gene on their homologous chromosomes. Mendel's parental pea plants always bred true because both of the gametes produced carried the same trait. When P_1 plants with contrasting traits were cross-fertilized, all of the offspring were **heterozygous** for the contrasting trait, meaning that their genotype reflected that they had different alleles for the gene being examined.

Dominant and Recessive Alleles

Our discussion of homozygous and heterozygous organisms brings us to why the F_1 heterozygous offspring were identical to one of the parents, rather than expressing both alleles. In all seven pea-plant characteristics, one of the two contrasting alleles was dominant,

and the other was recessive. Mendel called the dominant allele the expressed unit factor; the recessive allele was referred to as the latent unit factor. We now know that these so-called unit factors are actually genes on homologous chromosome pairs. For a gene that is expressed in a dominant and recessive pattern, homozygous dominant and heterozygous organisms will look identical (that is, they will have different genotypes but the same phenotype). The recessive allele will only be observed in homozygous recessive individuals (Table 12.4).

Human Inheritance in Dominant and Recessive Patterns

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

Table12.4

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, violet is the dominant trait for a pea plant's flower color, so the flower-color gene would be abbreviated as *V* (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 violet flowers as *VV*, a homozygous recessive pea plant with white flowers as *vv*, and a heterozygous pea plant with violet flowers as *Vv*.

The Punnett Square Approach for a Monohybrid Cross

When fertilization occurs between two true-breeding parents that differ in only one characteristic, the process is called a **monohybrid** cross, and the resulting offspring are monohybrids. Mendel performed seven monohybrid crosses involving contrasting traits for each characteristic. On the basis of his results in F_1 and F_2 generations, Mendel postulated that each parent in the monohybrid cross contributed one of two paired unit factors to each offspring, and every possible combination of unit factors was equally likely.

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. A **Punnett square**, devised by the British geneticist Reginald Punnett, can be drawn that applies the rules of probability to predict the possible outcomes of a genetic cross or mating and their expected frequencies. 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 12.4).

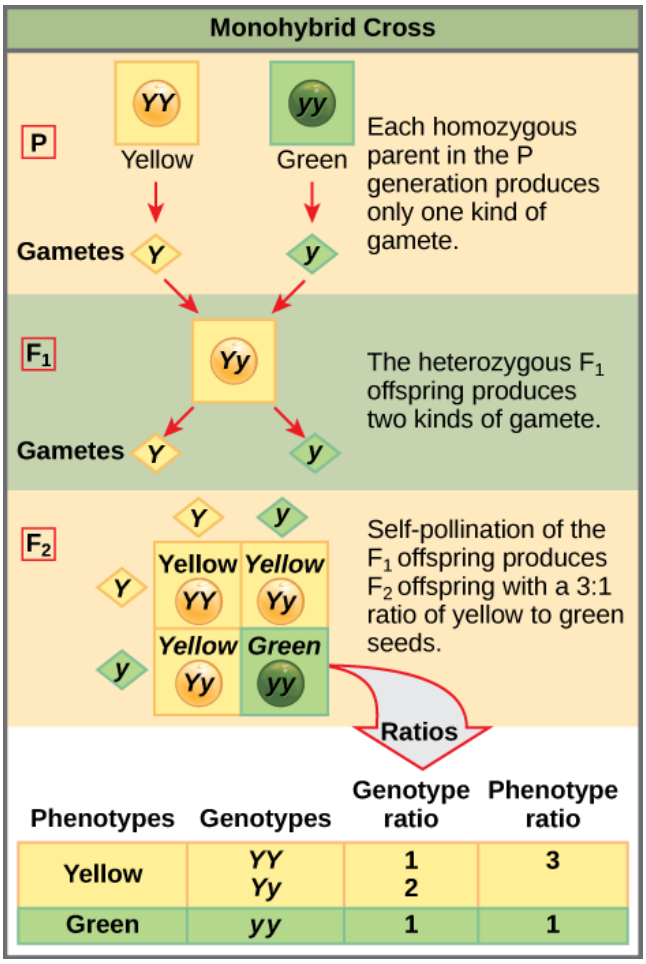


Figure 12.4 In the P generation, pea plants that are true-breeding for the dominant yellow phenotype are crossed with plants with the recessive green phenotype. This cross produces F_1 heterozygotes with a yellow phenotype. Punnett square analysis can be used to predict the genotypes of the F_2 generation.

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 12.4). 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. They are grouped together. 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 12.4). Furthermore, because the *YY* and *Yy* offspring have yellow seeds and are phenotypically identical, applying the sum rule of probability, 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 F_2 generation resulting from crosses for individual traits.

Mendel validated these results by performing an F_3 cross in which he self-crossed the dominant- and recessive-expressing F_2 plants. When he self-crossed the plants expressing green seeds, all of the offspring had green seeds, confirming that all green seeds had homozygous genotypes of *yy*. When he self-crossed the F_2 plants expressing yellow seeds, he found that one-third of the plants bred true, and two-thirds of the plants segregated at a 3:1 ratio of yellow:green seeds. In this case, the true-breeding plants had homozygous (*YY*) genotypes, whereas the segregating plants corresponded to the heterozygous (*Yy*) genotype. When these plants self-fertilized, the outcome was just like the F_1 self-fertilizing cross.

The Test Cross Distinguishes the Dominant Phenotype

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, the dominant-expressing organism is crossed with an organism that is homozygous recessive for the same characteristic. If the dominant-expressing organism is a homozygote, then all F_1 offspring will be heterozygotes expressing the dominant trait (Figure 12.5). Alternatively, if the dominant expressing organism is a heterozygote, the F_1 offspring will exhibit a 1:1 ratio of heterozygotes and recessive homozygotes (Figure 12.5). The test cross further validates Mendel's postulate that pairs of unit factors segregate equally.

VISUAL CONNECTION

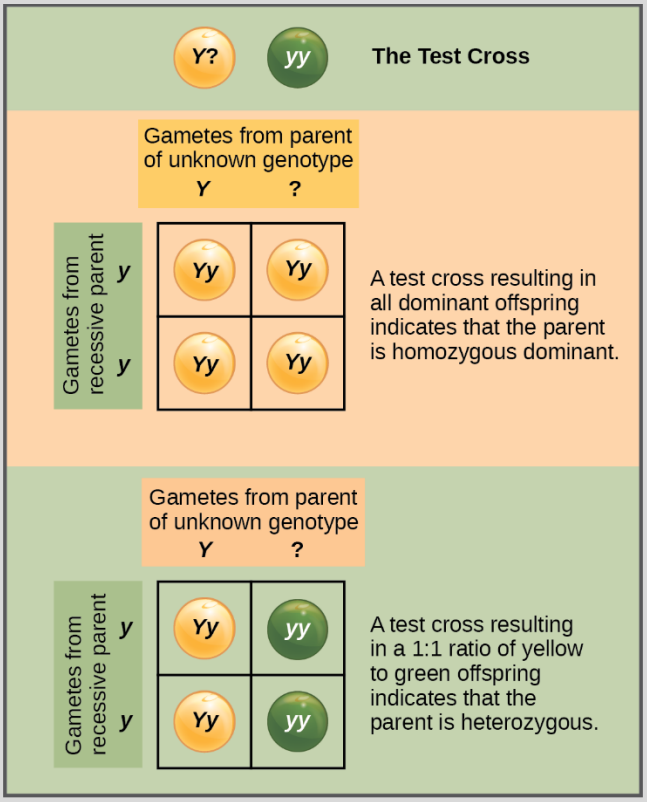


Figure 12.5 A test cross can be performed to determine whether an organism expressing a dominant trait is a homozygote or a heterozygote.

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?

Many human diseases are genetically inherited. 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 (Figure 12.6).

VISUAL CONNECTION

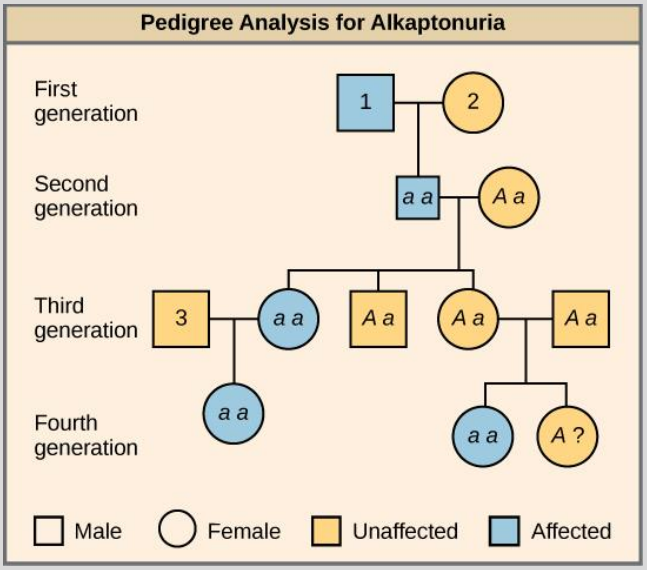


Figure 12.6 Alkaptonuria is a recessive genetic disorder in which two amino acids, phenylalanine and tyrosine, are not properly metabolized. Affected individuals may have darkened skin and brown urine, and may suffer joint damage and other complications. In this pedigree, individuals with the disorder 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, they 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.

What are the genotypes of the individuals labeled 1, 2, and 3?

Alternatives to Dominance and Recessiveness

Mendel's experiments with pea plants suggested that: (1) two "units" or alleles exist for every gene; (2) alleles maintain their integrity in each generation (no blending); and (3) in the presence of the dominant allele, the recessive allele is hidden and makes no contribution to the phenotype. Therefore, recessive alleles can be "carried" and not expressed by individuals. Such heterozygous individuals are sometimes referred to as "carriers." Further genetic studies in other plants and animals have shown that much more complexity exists, but that the

fundamental principles of Mendelian genetics still hold true. In the sections to follow, we consider some of the extensions of Mendelism. If Mendel had chosen an experimental system that exhibited these genetic complexities, it's possible that he would not have understood what his results meant.

Incomplete Dominance

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 12.7), a cross between a homozygous parent with white flowers (C^WC^W) and a homozygous parent with red flowers (C^RC^R) will produce offspring with pink flowers (C^RC^W). (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 C^RC^R:2 C^RC^W:1 C^WC^W, and the phenotypic ratio would be 1:2:1 for red:pink:white.

Figure 12.7 These pink flowers of a heterozygote snapdragon result from incomplete dominance. (credit: “storebukkebruse”/Flickr)

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 (L^ML^M and L^NL^N) express either the M or the N allele, and heterozygotes (L^ML^N) 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 12.8). 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.

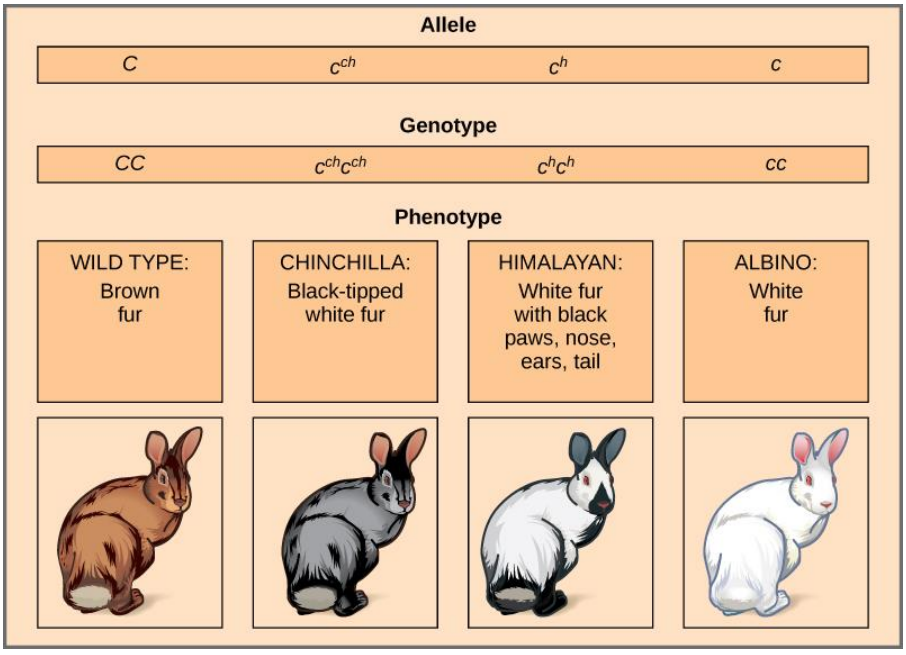


Figure 12.8 Four different alleles exist for the rabbit coat color (*C*) gene.

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 the allelic series in rabbits, the wild-type allele may supply a given dosage of fur pigment, whereas the mutants supply a lesser dosage or none at all. Interestingly, the Himalayan phenotype 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.

Alternatively, one mutant allele can be dominant over all other phenotypes, including the wild type. This may occur when the mutant allele somehow interferes with the genetic message so that even a heterozygote with one wild-type allele copy expresses the mutant phenotype. One way in which the mutant allele can interfere is by enhancing the function of the wild-type gene product or changing its distribution in the body. One example of this is the *Antennapedia* mutation in *Drosophila* (Figure 12.9). In this case, the mutant allele expands the distribution of the gene product, and as a result, the *Antennapedia* heterozygote develops legs on its head where its antennae should be.

Figure 12.9 As seen in comparing the wild-type *Drosophila* (left) and the *Antennapedia* mutant (right), the *Antennapedia* mutant has legs on its head in place of antennae.

EVOLUTION CONNECTION

Multiple Alleles Confer Drug Resistance in the Malaria Parasite

Malaria is a parasitic disease in humans that is transmitted by infected female mosquitoes, including *Anopheles gambiae* (Figure 12.10**a**), and is characterized by cyclic high fevers, chills, flu-like symptoms, and severe anemia. *Plasmodium falciparum* and *P. vivax* are the most common causative agents of malaria, and *P. falciparum* is the most deadly (Figure 12.10**b**). When promptly and correctly treated, *P. falciparum* malaria has a mortality rate of 0.1 percent. 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.

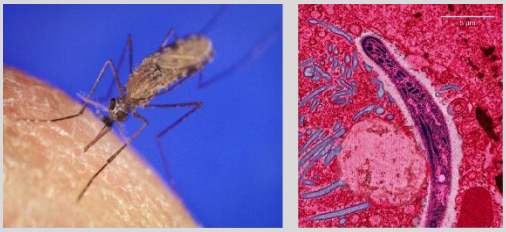


Figure 12.10 The (a) *Anopheles gambiae*, or African malaria mosquito, acts as a vector in the transmission to humans of the malaria-causing parasite (b) *Plasmodium falciparum*, here visualized using false-color transmission electron microscopy. (credit a: James D. Gathany; credit b: Ute Frevert; false color by Margaret Shear; scale-bar data from Matt Russell)

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 is infectious to 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 pathogens 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.[2]

X-Linked Traits

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 being examined is present on the X chromosome, but not on the Y chromosome, it is said to be **X-linked**.

Eye color in *Drosophila* was one of the first X-linked traits to be identified. Thomas Hunt Morgan mapped this trait to the X chromosome in 1910. Like humans, *Drosophila* males have an XY chromosome pair, and females are XX. In flies, the wild-type eye color is red (X^W) and it is dominant to white eye color (X^w) (Figure 12.11). Because of the location of the eye-color gene, reciprocal crosses do not produce the same offspring ratios. 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. *Drosophila* males lack a second allele copy on the Y chromosome; that is, their genotype can only be X^WY or X^wY. In contrast, females have two allele copies of this gene and can be X^WX^W, X^WX^w, or X^wX^w.

Figure 12.11 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 left are brown, cinnabar, sepia, vermilion, white, and red. Red eye color is wild-type and is dominant to white eye color.

In an X-linked cross, the genotypes of F_1 and F_2 offspring depend on whether the recessive trait was expressed by the male or the female in the P_1 generation. With regard to *Drosophila* eye color, when the P_1 male expresses the white-eye phenotype and the female is homozygous red-eyed, all members of the F_1 generation exhibit red eyes (Figure 12.12). The F_1 females are heterozygous (X^WX^w), and the males are all X^WY, having received their X chromosome from the homozygous dominant P_1 female and their Y chromosome from the P_1 male. A subsequent cross between the X^WX^w female and the X^WY male would produce only red-eyed females (with X^WX^W or X^WX^w genotypes) and both red- and white-eyed males (with X^WY or X^wY genotypes). Now, consider a cross between a homozygous white-eyed female and a male with red eyes. The F_1 generation would exhibit only heterozygous red-eyed females (X^WX^w) and only white-eyed males (X^wY). Half of the F_2 females would be red-eyed (X^WX^w) and half would be white-eyed (X^wX^w). Similarly, half of the F_2 males would be red-eyed (X^WY) and half would be white-eyed (X^wY).

VISUAL CONNECTION

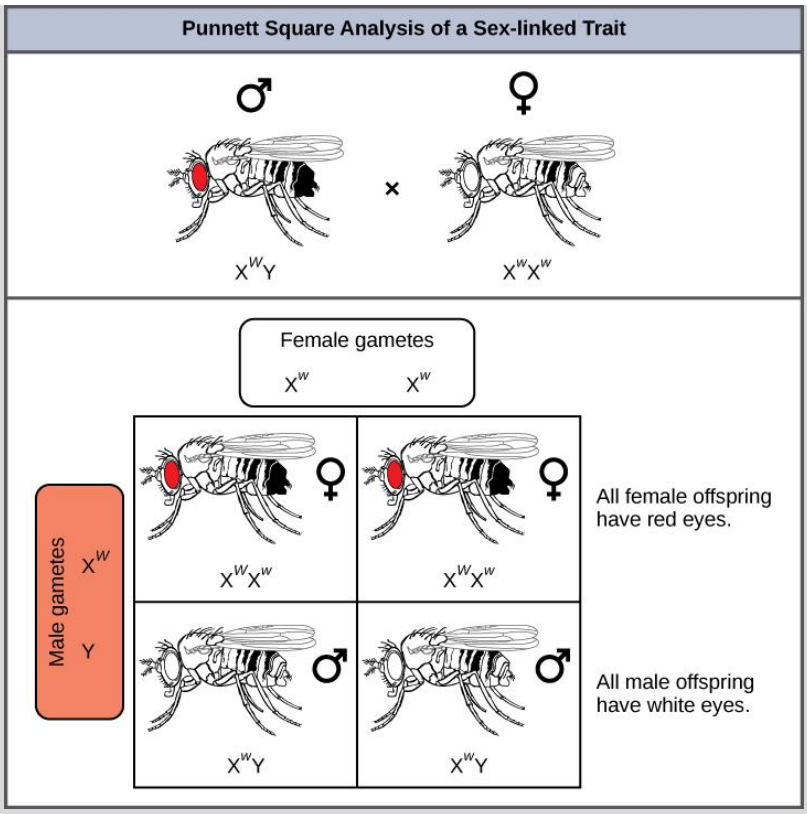


Figure 12.12 Punnett square analysis is used to determine the ratio of offspring from a cross between a red-eyed male fruit fly and a white-eyed female fruit fly.

What ratio of offspring would result from a cross between a white-eyed male and a female that is heterozygous for red eye color?

Discoveries in fruit fly genetics can be applied to human genetics. When a female parent is homozygous for a recessive X-linked trait, she will pass the trait on to 100 percent of her offspring. Her male offspring are, therefore, destined to express the trait, as they will inherit their father's Y chromosome. In humans, the alleles for certain conditions (some forms of color blindness, hemophilia, and muscular dystrophy) are X-linked. Females who are heterozygous for these diseases are said to be carriers and may not exhibit any phenotypic effects. These females will pass the disease to half of their sons and will pass carrier status to half of their daughters; therefore, recessive X-linked traits appear more frequently in males than females.

In some groups of organisms with sex chromosomes, the sex 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 include red-green color blindness, and Types A and B hemophilia. 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 12.13). Although some Y-linked recessive disorders exist, typically they are associated with infertility in males and are therefore not transmitted to subsequent generations.

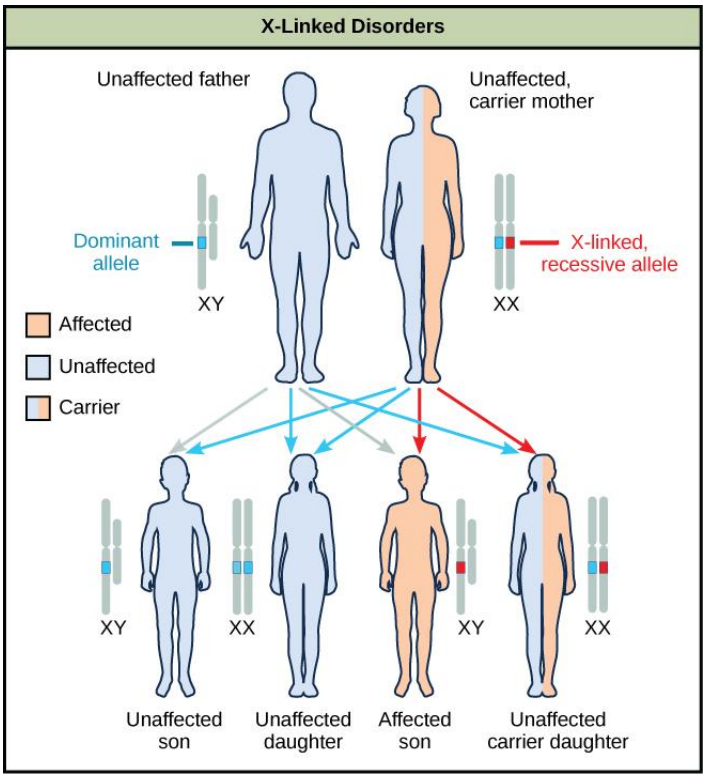


Figure 12.13 The son of a woman who is a carrier of a recessive X-linked disorder will have a 50 percent chance of being affected. A daughter will not be affected, but she will have a 50 percent chance of being a carrier like her mother.

LINK TO LEARNING

Watch this video to learn more about sex-linked traits: https://www.youtube.com/embed/-ROhfKyxgCo

Lethality

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 as long as individuals with this allele also have a wild-type, functional copy. The wild-type allele functions at a capacity sufficient to sustain life and is therefore considered to be dominant over the nonfunctional allele. However, consider two heterozygous parents that have a genotype of wild-type/nonfunctional mutant for a hypothetical essential gene. In one quarter of their offspring, we would expect to observe individuals that are homozygous recessive for the nonfunctional allele. Because the gene is essential, these individuals might fail to develop past fertilization, die *in utero*, or die later in life, depending on what life stage requires this gene. An inheritance pattern in which an allele is only lethal in the homozygous form and in which the heterozygote may be normal or have some altered nonlethal phenotype is referred to as **recessive lethal**.

For crosses between heterozygous individuals with a recessive lethal allele that causes death before birth when homozygous, only wild-type homozygotes and heterozygotes would be observed. The genotypic ratio would therefore be 2:1. In other instances, the recessive lethal allele might also exhibit a dominant (but not lethal) phenotype in the heterozygote. For instance, the recessive lethal *Curly* allele in *Drosophila* affects wing shape in the heterozygote form but is lethal in the homozygote.

A single copy of the wild-type allele is not always sufficient for normal functioning or even survival. The **dominant lethal** inheritance pattern is one in which an allele is lethal both in the homozygote and the heterozygote; this allele can only be transmitted if the lethality phenotype occurs after reproductive age. Individuals with mutations that result in dominant lethal alleles

fail to survive even in the heterozygote form. Dominant lethal alleles are very rare because, as you might expect, the allele only lasts one generation and is not transmitted. However, just as the recessive lethal allele might not immediately manifest the phenotype of death, dominant lethal alleles also might not be expressed until adulthood. Once the individual reaches reproductive age, the allele may be unknowingly passed on, resulting in a delayed death in both generations. An example of this in humans is Huntington's disease, in which the nervous system gradually wastes away (Figure 12.14). People who are heterozygous for the dominant Huntington allele (*Hh*) will inevitably develop the fatal disease. However, the onset of Huntington's 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.

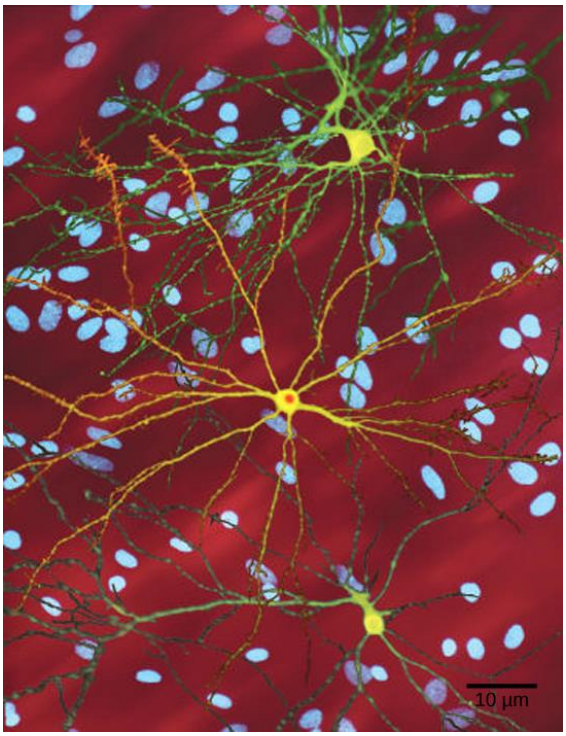


Figure 12.14 The neuron in the center of this micrograph (yellow) has nuclear inclusions characteristic of Huntington's disease (orange area in the center of the neuron). Huntington's disease occurs when an abnormal dominant allele for the Huntington gene is present. (credit: Dr. Steven Finkbeiner, Gladstone Institute of Neurological Disease, The Taube-Koret Center for Huntington's Disease Research, and the University of California San Francisco/Wikimedia)

Footnotes

- 2 Sumiti Vinayak, et al., “Origin and Evolution of Sulfadoxine Resistant *Plasmodium falciparum*,” *Public Library of Science Pathogens* 6, no. 3 (2010): e1000830, doi:10.1371/journal.ppat.1000830.

Chapter Attributions

Laws of Inheritance

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

- Explain Mendel’s law of segregation and independent assortment in terms of genetics and the events of meiosis
- Use the forked-line method and the probability rules to calculate the probability of genotypes and phenotypes from multiple gene crosses
- Explain the effect of linkage and recombination on gamete genotypes
- Explain the phenotypic outcomes of epistatic effects between genes

Mendel generalized the results of his pea-plant experiments into four postulates, some of which are sometimes called “laws,” that describe the basis of dominant and recessive inheritance in diploid organisms. As you have learned, more complex extensions of Mendelism exist that do not exhibit the same F_2 phenotypic ratios (3:1). Nevertheless, these laws summarize the basics of classical genetics.

Pairs of Unit Factors, or Genes

Mendel proposed first that paired unit factors of heredity were transmitted faithfully from generation to generation by the dissociation and reassociation of paired factors during gametogenesis and fertilization, respectively. After he crossed peas with contrasting traits and found that the recessive trait resurfaced in the F_2 generation, Mendel deduced that hereditary

factors must be inherited as discrete units. This finding contradicted the belief at that time that parental traits were blended in the offspring.

Alleles Can Be Dominant or Recessive

Mendel's **law of dominance** states that in a heterozygote, one trait will conceal the presence of another trait for the same characteristic. Rather than both alleles contributing to a phenotype, the dominant allele will be expressed exclusively. 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 12.15), and these offspring will breed true when self-crossed.

Since Mendel's experiments with pea plants, researchers have found that the law of dominance does not always hold true. Instead, several different patterns of inheritance have been found to exist.

Figure 12.15 The child in the photo expresses albinism, a recessive trait.

Equal Segregation of Alleles

Observing that true-breeding pea plants with contrasting traits gave rise to F_1 generations that all expressed the dominant trait and F_2 generations that expressed the dominant and recessive

traits in a 3:1 ratio, Mendel proposed the **law of segregation**. This law states that paired unit factors (genes) must segregate equally into gametes such that offspring have an equal likelihood of inheriting either factor. For the F_2 generation of a monohybrid cross, the following three possible combinations of genotypes could result: homozygous dominant, heterozygous, or homozygous recessive. Because heterozygotes could arise from two different pathways (receiving one dominant and one recessive allele from either parent), and because heterozygotes and homozygous dominant individuals are phenotypically identical, the law supports Mendel's observed 3:1 phenotypic ratio. The equal segregation of alleles is the reason we can apply the Punnett square to accurately predict the offspring of parents with known genotypes. The physical basis of Mendel's law 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. The role of the meiotic segregation of chromosomes in sexual reproduction was not understood by the scientific community during Mendel's lifetime.

Independent Assortment

Mendel's **law 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 the **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 law 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 F_1 generation of offspring all are *YyRr* (Figure 12.16).

VISUAL CONNECTION

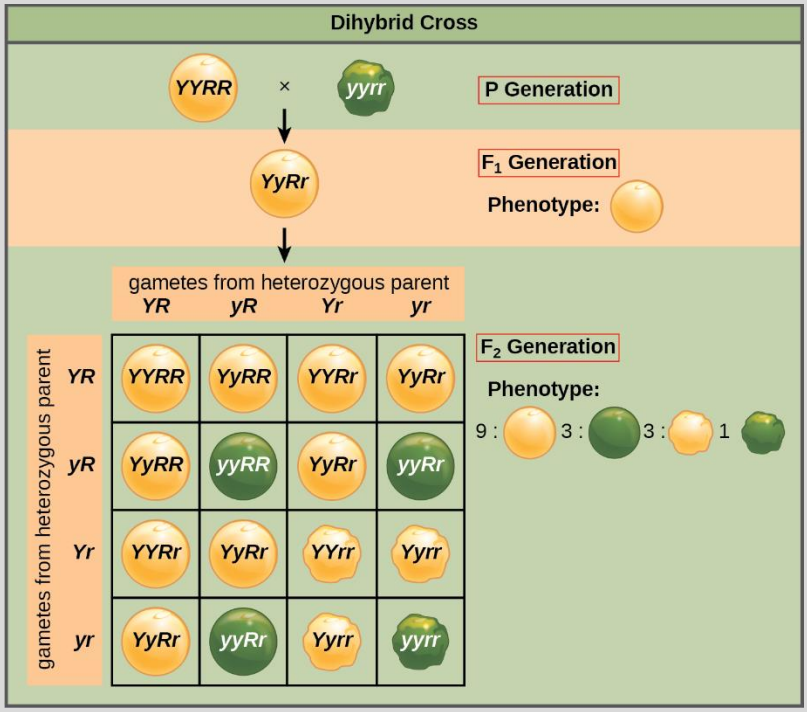


Figure 12.16 This dihybrid cross of pea plants involves the genes for seed color and texture.

In pea plants, purple flowers (P) are dominant to white flowers (p) and yellow peas (Y) are dominant to green peas (y). What are the possible genotypes and phenotypes for a cross between PpYY and ppYy pea plants? How many squares do you need to do a Punnett square analysis of this cross?

For the F2 generation, the law 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 law 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 (Figure 12.16) 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 12.16). These are the offspring ratios we would expect, assuming we performed the crosses with a large enough sample size.

Because of independent assortment and dominance, the 9:3:3:1 dihybrid phenotypic ratio can be collapsed into two 3:1 ratios, characteristic of any monohybrid cross that follows a dominant and recessive pattern. Ignoring seed color and considering only seed texture in the above dihybrid cross, we would expect that three quarters of the F_2 generation offspring would be round, and one quarter would be wrinkled. Similarly, isolating only seed color, we would assume that three quarters of the F_2 offspring would be yellow and one quarter would be green. The sorting of alleles for texture and color are independent events, so we can apply the product rule. Therefore, the proportion of round and yellow F_2 offspring is expected to be (3/4) × (3/4) = 9/16, and the proportion of wrinkled and green offspring is expected to be (1/4) × (1/4) = 1/16. These proportions are identical to those obtained using a Punnett square. Round, green and wrinkled, yellow offspring can also be calculated using the product rule, as each of these genotypes includes one dominant and one recessive phenotype. Therefore, the proportion of each is calculated as (3/4) × (1/4) = 3/16.

The law of independent assortment also indicates that a cross between yellow, wrinkled (*YYrr*) and green, round (*yyRR*) parents would yield the same F_1 and F_2 offspring as in the *YYRR* x *yyrr* cross.

The physical basis for the law 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.

Forked-Line Method

When more than two genes are being considered, the Punnett-square method becomes unwieldy. For instance, examining a cross involving four genes would require a 16 × 16 grid containing 256 boxes. It would be extremely cumbersome to manually enter each genotype. For more complex crosses, the forked-line and probability methods are preferred.

To prepare a forked-line diagram for a cross between F_1 heterozygotes resulting from a cross between *AABBCC* and *aabbcc* parents, we first create rows equal to the number of genes being considered, and then segregate the alleles in each row on forked lines according to the probabilities for individual monohybrid crosses (Figure 12.17). We then multiply the values

along each forked path to obtain the F_2 offspring probabilities. Note that this process is a diagrammatic version of the product rule. The values along each forked pathway can be multiplied because each gene assorts independently. For a trihybrid cross, the F_2 phenotypic ratio is 27:9:9:9:3:3:3:1.

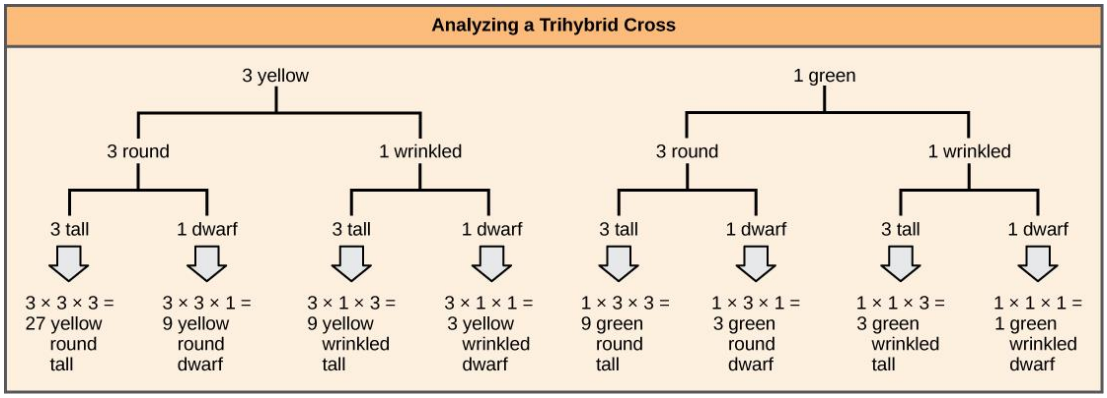


Figure 12.17 The forked-line method can be used to analyze a trihybrid cross. Here, the probability for color in the F_2 generation occupies the top row (3 yellow:1 green). The probability for shape occupies the second row (3 round: 1 wrinkled), and the probability for height occupies the third row (3 tall:1 dwarf). The probability for each possible combination of traits is calculated by multiplying the probability for each individual trait. Thus, the probability of F_2 offspring having yellow, round, and tall traits is 3 × 3 × 3, or 27.

Probability Method

While the forked-line method is a diagrammatic approach to keeping track of probabilities in a cross, the probability method gives the proportions of offspring expected to exhibit each phenotype (or genotype) without the added visual assistance. Both methods make use of the product rule and consider the alleles for each gene separately. Earlier, we examined the phenotypic proportions for a trihybrid cross using the forked-line method; now we will use the probability method to examine the genotypic proportions for a cross with even more genes.

For a trihybrid cross, writing out the forked-line method is tedious, albeit not as tedious as using the Punnett-square method. To fully demonstrate the power of the probability method, however, we can consider specific genetic calculations. For instance, for a tetrahybrid cross between individuals that are heterozygotes for all four genes, and in which all four genes are

sorting independently and in a dominant and recessive pattern, what proportion of the offspring will be expected to be homozygous recessive for all four alleles? Rather than writing out every possible genotype, we can use the probability method. We know that for each gene, the fraction of homozygous recessive offspring will be 1/4. Therefore, multiplying this fraction for each of the four genes, (1/4) × (1/4) × (1/4) × (1/4), we determine that 1/256 of the offspring will be quadruply homozygous recessive.

For the same tetrahybrid cross, what is the expected proportion of offspring that have the dominant phenotype at all four loci? We can answer this question using phenotypic proportions, but let's do it the hard way—using genotypic proportions. The question asks for the proportion of offspring that are 1) homozygous dominant at *A* or heterozygous at *A,* and 2) homozygous at *B* or heterozygous at *B*, and so on. Noting the "or" and "and" in each circumstance makes clear where to apply the sum and product rules. The probability of a homozygous dominant at *A* is 1/4 and the probability of a heterozygote at *A* is 1/2. The probability of the homozygote or the heterozygote is 1/4 + 1/2 = 3/4 using the sum rule. The same probability can be obtained in the same way for each of the other genes, so that the probability of a dominant phenotype at *A* and *B* and *C* and *D* is, using the product rule, equal to 3/4 × 3/4 × 3/4 × 3/4, or 27/64. If you are ever unsure about how to combine probabilities, returning to the forked-line method should make it clear.

Rules for Multihybrid Fertilization

Predicting the genotypes and phenotypes of offspring from given crosses is the best way to test your knowledge of Mendelian genetics. Given a multihybrid cross that obeys independent assortment and follows a dominant and recessive pattern, several generalized rules exist; you can use these rules to check your results as you work through genetics calculations (Table 12.5). To apply these rules, first you must determine *n*, the number of heterozygous gene pairs (the number of genes segregating two alleles each). For example, a cross between *AaBb* and *AaBb* heterozygotes has an *n* of 2. In contrast, a cross between *AABb* and *AABb* has an *n* of 1 because *A* is not heterozygous.

General Rules for Multihybrid Crosses

General Rule	Number of Heterozygous Gene Pairs
Number of different F_1 gametes	2^n
Number of different F_2 genotypes	3^n
Given dominant and recessive inheritance, the number of different F_2 phenotypes	2^n

Table12.5

Linked Genes Violate the Law of Independent Assortment

Although all of Mendel's pea characteristics behaved according to the law of independent assortment, we now know that some allele combinations are not inherited independently of each other. Genes that are located on separate non-homologous chromosomes will always sort independently. However, each chromosome contains hundreds or thousands of genes, organized linearly on chromosomes like beads on a string. The segregation of alleles into gametes can be influenced by **linkage**, in which genes that are located physically close to each other on the same chromosome are more likely to be inherited as a pair. However, because of the process of recombination, or "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 linear order. The alleles may differ on homologous chromosome pairs, but the genes to which they correspond do not. In preparation for the first division of meiosis, homologous chromosomes replicate and synapse. Like genes on the homologs align with each other. At this stage, segments of homologous chromosomes exchange linear segments of genetic material (Figure 12.18). This process is called *recombination*, or crossover, and it is a common genetic process. Because the genes are

aligned during recombination, 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.

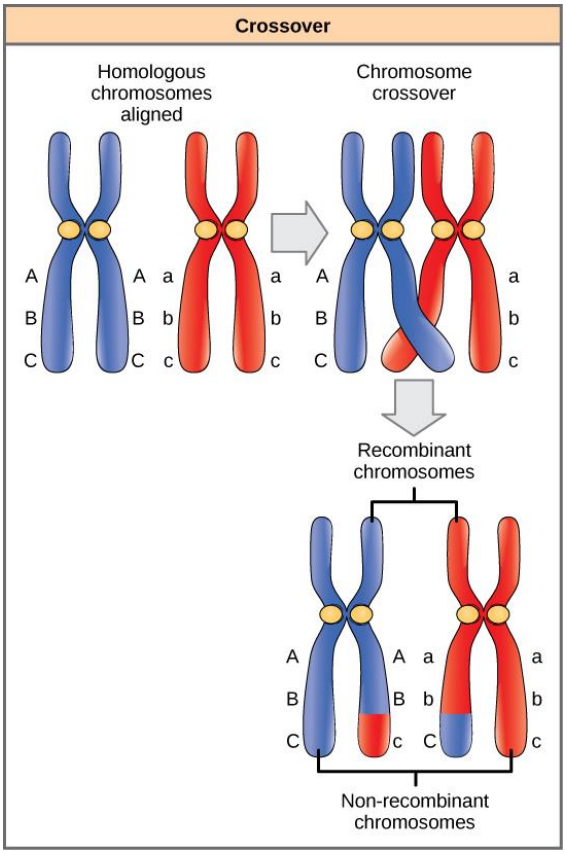


Figure 12.18 The process of crossover, or recombination, occurs when two homologous chromosomes align during meiosis and exchange a segment of genetic material. Here, the alleles for gene C were exchanged. The result is two recombinant and two non-recombinant chromosomes.

When two genes are located in close proximity on the same chromosome, they are considered linked, and their alleles tend 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 one homologous chromosome has alleles for tall plants and red flowers, and the other chromosome has genes 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. But unlike if the genes were on different chromosomes, 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 one or more crossovers between them increases, and the genes behave more like they are on separate chromosomes. Geneticists have used the proportion of recombinant gametes (the ones not like the parents) as a measure of how far apart genes are on a chromosome. Using this information, they have constructed elaborate maps of genes on chromosomes for well-studied organisms, including humans.

Mendel's seminal publication makes no mention of linkage, and many researchers have questioned whether he encountered linkage but chose not to publish those crosses out of concern that they would invalidate his independent assortment postulate. The garden pea has seven chromosomes, and some have suggested that his choice of seven characteristics was not a coincidence. However, even if the genes he examined were not located on separate chromosomes, it is possible that he simply did not observe linkage because of the extensive shuffling effects of recombination.

SCIENTIFIC METHOD CONNECTION

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 pea plants mature in one growing season, and 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. Upon plant

maturation, the plants 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 laws. When the true-breeding parents are crossed, all of the F_1 offspring are tall and have inflated pods, which indicates that the tall and inflated traits are dominant over the dwarf and constricted traits, respectively. A self-cross of the F_1 heterozygotes results in 2,000 F_2 progeny.

Test the hypothesis: Because each trait pair sorts independently, the ratios of tall:dwarf and inflated:constricted are each expected to be 3:1. The tall/dwarf trait pair is called *T/t*, and the inflated/constricted trait pair is designated *I/i*. Each member of the F_1 generation therefore has a genotype of *TtIi*. Construct a grid analogous to Figure 12.16, in which you cross two *TtIi* individuals. Each individual can donate four combinations of two traits: *TI*, *Ti*, *tI*, or *ti*, meaning that there are 16 possibilities of offspring genotypes. Because the *T* and *I* alleles are dominant, any individual having one or two of those alleles will express the tall or inflated phenotypes, respectively, regardless if they also have a *t* or *i* allele. Only individuals that are *tt* or *ii* will express the dwarf and constricted alleles, respectively. As shown in Figure 12.19, you predict that you will observe the following offspring proportions: tall/inflated:tall/constricted:dwarf/inflated:dwarf/constricted in 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.

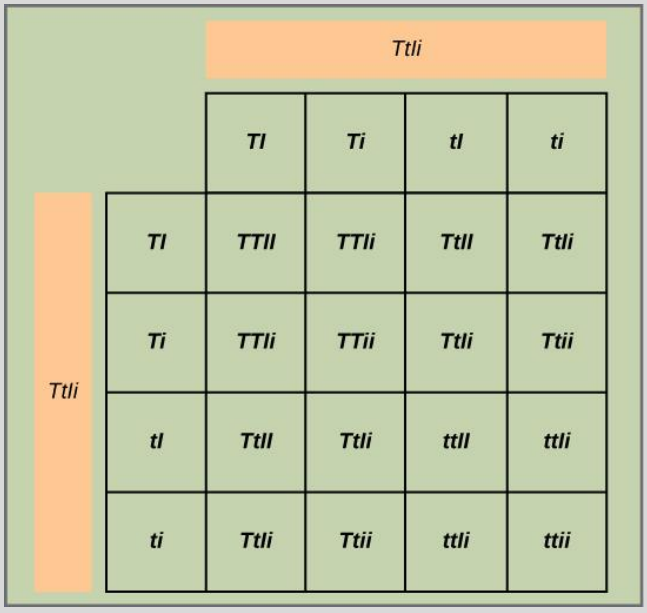


Figure 12.19 This figure shows all possible combinations of offspring resulting from a dihybrid cross of pea plants that are heterozygous for the tall/dwarf and inflated/constricted alleles.

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?

Analyze your data: You observe the following plant phenotypes in the F_2 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 laws.

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?

Epistasis

Mendel's studies in pea plants implied that the sum of an individual's phenotype was controlled by genes (or as he called them, unit factors), such 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.

LINK TO LEARNING

Eye color in humans is determined by multiple genes. Use the Eye Color Calculator (http://www.athro.com/evo/inherit.html) to predict the eye color of children from parental eye color.

In some cases, several genes can contribute to aspects of a common phenotype without their gene products ever directly interacting. In the case of organ development, for instance, genes may be expressed sequentially, with each gene adding to the complexity and specificity of the organ. 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, with one gene modifying the expression of another.

In **epistasis**, the interaction between genes is antagonistic, such that one gene masks or interferes with the expression of another. “Epistasis” is a word composed of Greek roots that mean “standing upon.” The alleles that are being masked or silenced are said to be hypostatic to the epistatic alleles that are doing the masking. 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* (Figure 12.20). Therefore, the genotypes *AAcc*, *Aacc*, and *aacc* all produce the same 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 12.20). In this case, the *C* gene is epistatic to the *A* gene.

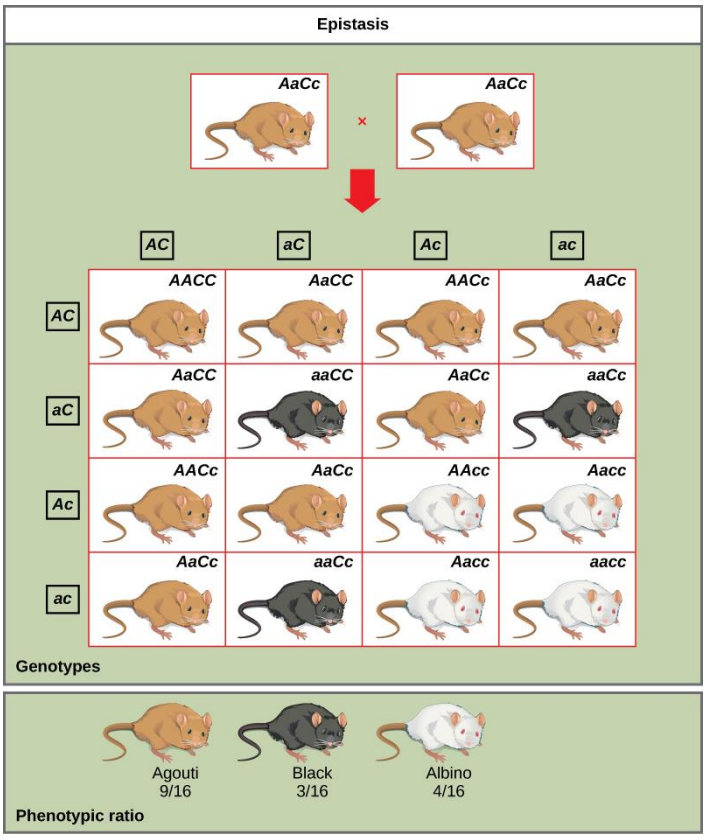


Figure 12.20 In mice, the mottled agouti coat color (*A*) is dominant to a solid coloration, such as black or gray. A gene at a separate locus (*C*) is responsible for pigment production. The recessive *c* allele does not produce pigment, and a mouse with the homozygous recessive *cc* genotype is albino 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 noninteracting 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.

LINK TO LEARNING

For an excellent review of Mendel's experiments and to perform your own crosses and identify patterns of inheritance, visit the Mendel's Peas (http://www2.edc.org/weblabs/mendel/MendelMenu.html) web lab.

Chapter Attributions

Chromosomal Theory and Genetic Linkage

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

- Discuss Sutton's Chromosomal Theory of Inheritance
- Describe genetic linkage
- Explain the process of homologous recombination, or crossing over
- Describe chromosome creation
- Calculate the distances between three genes on a chromosome using a three-point test cross

Long before scientists visualized chromosomes under a microscope, the father of modern genetics, Gregor Mendel, began studying heredity in 1843. With improved microscopic techniques during the late 1800s, cell biologists could stain and visualize subcellular structures with dyes and observe their actions during cell division and meiosis. With each mitotic division, chromosomes replicated, condensed from an amorphous (no constant shape) nuclear mass into distinct X-shaped bodies (pairs of identical sister chromatids), and migrated to separate cellular poles.

Chromosomal Theory of Inheritance

The speculation that chromosomes might be the key to understanding heredity led several scientists to examine Mendel's publications and reevaluate his model in terms of chromosome behavior during mitosis and meiosis. In 1902, Theodor Boveri observed that proper sea urchin embryonic development does not occur unless chromosomes are present. That same year, Walter Sutton observed chromosome separation into daughter cells during meiosis (Figure 13.2). Together, these observations led to the **Chromosomal Theory of Inheritance**, which identified chromosomes as the genetic material responsible for Mendelian inheritance.

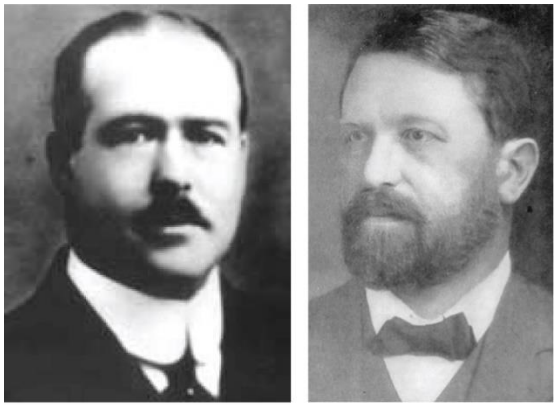


Figure 13.2 (a) Walter Sutton and (b) Theodor Boveri developed the Chromosomal Theory of Inheritance, which states that chromosomes carry the unit of heredity (genes).

The Chromosomal Theory of Inheritance was consistent with Mendel's laws, which the following observations supported:

- During meiosis, homologous chromosome pairs migrate as discrete structures that are independent of other chromosome pairs.
- Chromosome sorting from each homologous pair into pre-gametes appears to be random.
- Each parent synthesizes gametes that contain only half their chromosomal complement.
- Even though male and female gametes (sperm and egg) differ in size and morphology, they have the same number of chromosomes, suggesting equal genetic contributions from each parent.

- The gametic chromosomes combine during fertilization to produce offspring with the same chromosome number as their parents.

Despite compelling correlations between chromosome behavior during meiosis and Mendel's abstract laws, scientists proposed the Chromosomal Theory of Inheritance long before there was any direct evidence that chromosomes carried traits. Critics pointed out that individuals had far more independently segregating traits than they had chromosomes. It was only after several years of carrying out crosses with the fruit fly, *Drosophila melanogaster*, that Thomas Hunt Morgan provided experimental evidence to support the Chromosomal Theory of Inheritance.

Genetic Linkage and Distances

Mendel's work suggested that traits are inherited independently of each other. Morgan identified a 1:1 correspondence between a segregating trait and the X chromosome, suggesting that random chromosome segregation was the physical basis of Mendel's model. This also demonstrated that linked genes disrupt Mendel's predicted outcomes. That each chromosome can carry many linked genes explains how individuals can have many more traits than they have chromosomes. However, researchers in Morgan's laboratory suggested that alleles positioned on the same chromosome were not always inherited together. During meiosis, linked genes somehow became unlinked.

Homologous Recombination

In 1909, Frans Janssen observed chiasmata—the point at which chromatids are in contact with each other and may exchange segments—prior to the first meiosis division. He suggested that alleles become unlinked and chromosomes physically exchange segments. As chromosomes condensed and paired with their homologs, they appeared to interact at distinct points. Janssen suggested that these points corresponded to regions in which chromosome segments exchanged. We now know that the pairing and interaction between homologous chromosomes, or synapsis, does more than simply organize the homologs for migration to separate daughter cells. When synapsed, homologous chromosomes undergo reciprocal physical exchanges at their arms in **homologous recombination**, or more simply, "crossing over."

To better understand the type of experimental results that researchers were obtaining at this time, consider a heterozygous individual that inherited dominant maternal alleles for two genes on the same chromosome (such as *AB*) and two recessive paternal alleles for those same genes (such as *ab*). If the genes are linked, one would expect this individual to produce gametes that are either *AB* or *ab* with a 1:1 ratio. If the genes are unlinked, the individual should produce *AB*, *Ab*, *aB*, and *ab* gametes with equal frequencies, according to the Mendelian concept of independent assortment. Because they correspond to new allele combinations, the genotypes Ab and aB are **nonparental types** that result from homologous recombination during meiosis. **Parental types** are progeny that exhibit the same allelic combination as their parents. Morgan and his colleagues, however, found that when they test crossed such heterozygous individuals to a homozygous recessive parent (*AaBb* × *aabb*), both parental and nonparental cases occurred. For example, 950 offspring might be recovered that were either *AaBb* or *aabb*, but 50 offspring would also result that were either *Aabb* or *aaBb*. These results suggested that linkage occurred most often, but a significant minority of offspring were the products of recombination.

VISUAL CONNECTION

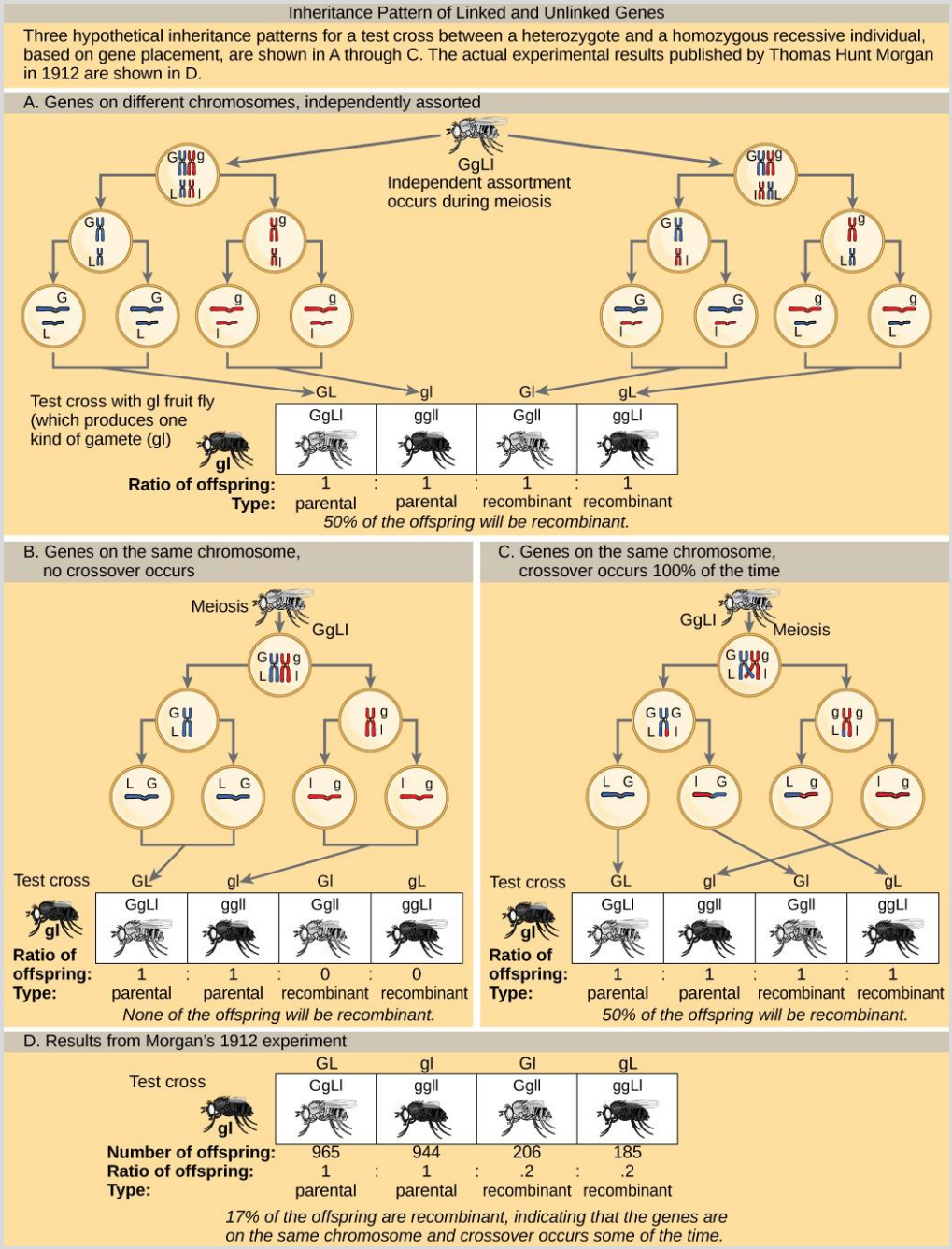


Figure 13.3 This figure shows unlinked and linked gene inheritance patterns. In (a), two genes are located on different chromosomes so independent assortment occurs during meiosis. The offspring have an equal chance of being the parental type (inheriting the same combination of traits as the parents) or a nonparental type (inheriting a different combination of traits than the parents). In (b), two genes are very close together on the same chromosome so that no crossing

over occurs between them. Therefore, the genes are always inherited together and all the offspring are the parental type. In (c), two genes are far apart on the chromosome such that crossing over occurs during every meiotic event. The recombination frequency will be the same as if the genes were on separate chromosomes. (d) The actual recombination frequency of fruit fly wing length and body color that Thomas Morgan observed in 1912 was 17 percent. A crossover frequency between 0 percent and 50 percent indicates that the genes are on the same chromosome and crossover sometimes occurs.

In a test cross for two characteristics such as the one here, can the recombinant offspring's predicted frequency be 60 percent? Why or why not?

Genetic Maps

Janssen did not have the technology to demonstrate crossing over so it remained an abstract idea that scientists did not widely believe. Scientists thought chiasmata were a variation on synapsis and could not understand how chromosomes could break and rejoin. Yet, the data were clear that linkage did not always occur. Ultimately, it took a young undergraduate student and an “all-nighter” to mathematically elucidate the linkage and recombination problem.

In 1913, Alfred Sturtevant, a student in Morgan’s laboratory, gathered results from researchers in the laboratory, and took them home one night to mull them over. By the next morning, he had created the first “chromosome map,” a linear representation of gene order and relative distance on a chromosome (Figure 13.4).

VISUAL CONNECTION

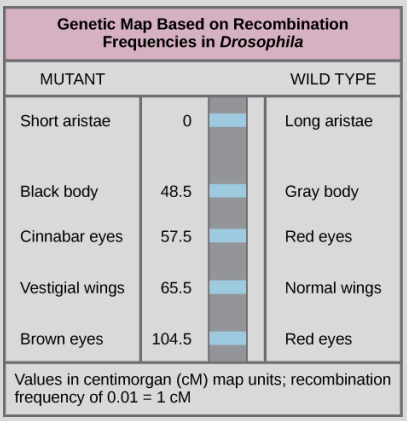


Figure 13.4 This genetic map orders *Drosophila* genes on the basis of recombination frequency.

Which of the following statements is true?

a. Recombination of the body color and red/cinnabar eye alleles will occur more frequently than recombination of the alleles for wing length and aristae length.

b. Recombination of the body color and aristae length alleles will occur more frequently than recombination of red/brown eye alleles and the aristae length alleles.

c. Recombination of the gray/black body color and long/short aristae alleles will not occur.

d. Recombination of the red/brown eye and long/short aristae alleles will occur more frequently than recombination of the alleles for wing length and body color.

As Figure 13.4 shows, by using recombination frequency to predict genetic distance, we can infer the relative gene order on chromosome 2. The values represent map distances in centimorgans (cM), which correspond to recombination frequencies (in percent). Therefore, the genes for body color and wing size were 65.5 − 48.5 = 17 cM apart, indicating that the maternal and paternal alleles for these genes recombine in 17 percent of offspring, on average.

To construct a chromosome map, Sturtevant assumed that genes were ordered serially on threadlike chromosomes. He also assumed that the incidence of recombination between two homologous chromosomes could occur with equal likelihood anywhere along the chromosome's length. Operating under these assumptions, Sturtevant postulated that alleles that were far apart on a chromosome were more likely to dissociate during meiosis simply because there was a larger region over which recombination could occur. Conversely, alleles that were close to each other on the chromosome were likely to be inherited together. The average number of crossovers between two alleles—that is, their **recombination frequency**—correlated with their genetic distance from each other, relative to the locations of other genes on that chromosome. Considering the example cross between *AaBb* and *aabb* above, we could calculate the recombination's frequency as 50/1000 = 0.05. That is, the likelihood of a crossover between genes *A/a* and *B/b* was 0.05, or 5 percent. Such a result would indicate that the genes were definitively linked, but that they were far enough apart for crossovers to occasionally

occur. Sturtevant divided his genetic map into map units, or **centimorgans (cM)**, in which a 0,01 recombination frequency corresponds to 1 cM.

By representing alleles in a linear map, Sturtevant suggested that genes can range from linking perfectly (recombination frequency = 0) to unlinking perfectly (recombination frequency = 0.5) when genes are on different chromosomes or genes separate very far apart on the same chromosome. Perfectly unlinked genes correspond to the frequencies Mendel predicted to assort independently in a dihybrid cross. A 0.5 recombination frequency indicates that 50 percent of offspring are recombinants and the other 50 percent are parental types. That is, every type of allele combination is represented with equal frequency. This representation allowed Sturtevant to additively calculate distances between several genes on the same chromosome. However, as the genetic distances approached 0.50, his predictions became less accurate because it was not clear whether the genes were very far apart on the same or on different chromosomes.

In 1931, Barbara McClintock and Harriet Creighton demonstrated the crossover of homologous chromosomes in corn plants. Weeks later, Curt Stern demonstrated microscopically homologous recombination in *Drosophila*. Stern observed several X-linked phenotypes that were associated with a structurally unusual and dissimilar X chromosome pair in which one X was missing a small terminal segment, and the other X was fused to a piece of the Y chromosome. By crossing flies, observing their offspring, and then visualizing the offspring's chromosomes, Stern demonstrated that every time the offspring allele combination deviated from either of the parental combinations, there was a corresponding exchange of an X chromosome segment. Using mutant flies with structurally distinct X chromosomes was the key to observing the products of recombination because DNA sequencing and other molecular tools were not yet available. We now know that homologous chromosomes regularly exchange segments in meiosis by reciprocally breaking and rejoining their DNA at precise locations.

LINK TO LEARNING

Review Sturtevant's process to create a genetic map on the basis of recombination frequencies here: https://dnalc.cshl.edu/view/16280-Animation-11-Genes-get-shuffled-when-chromosomes-exchange-pieces-.html

Mendel's Mapped Traits

Homologous recombination is a common genetic process, yet Mendel never observed it. Had he investigated both linked and unlinked genes, it would have been much more difficult for him to create a unified model of his data on the basis of probabilistic calculations. Researchers who have since mapped the seven traits that Mendel investigated onto a pea plant genome's seven chromosomes have confirmed that all the genes he examined are either on separate chromosomes or are sufficiently far apart as to be statistically unlinked. Some have suggested that Mendel was enormously lucky to select only unlinked genes; whereas, others question whether Mendel discarded any data suggesting linkage. In any case, Mendel consistently observed independent assortment because he examined genes that were effectively unlinked.

Chapter Attributions

Chromosomal Basis of Inherited Disorders

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

- Describe how a karyogram is created
- Explain how nondisjunction leads to disorders in chromosome number
- Compare disorders that aneuploidy causes
- Describe how errors in chromosome structure occur through inversions and translocations

Inherited disorders can arise when chromosomes behave abnormally during meiosis. We can divide chromosome disorders into two categories: abnormalities in chromosome number and chromosomal structural rearrangements. Because even small chromosome segments can span many genes, chromosomal disorders are characteristically dramatic and often fatal.

Chromosome Identification

Chromosome isolation and microscopic observation forms the basis of cytogenetics and is the primary method by which clinicians detect chromosomal abnormalities in humans.
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**. Another name is an ideogram (Figure 13.5).

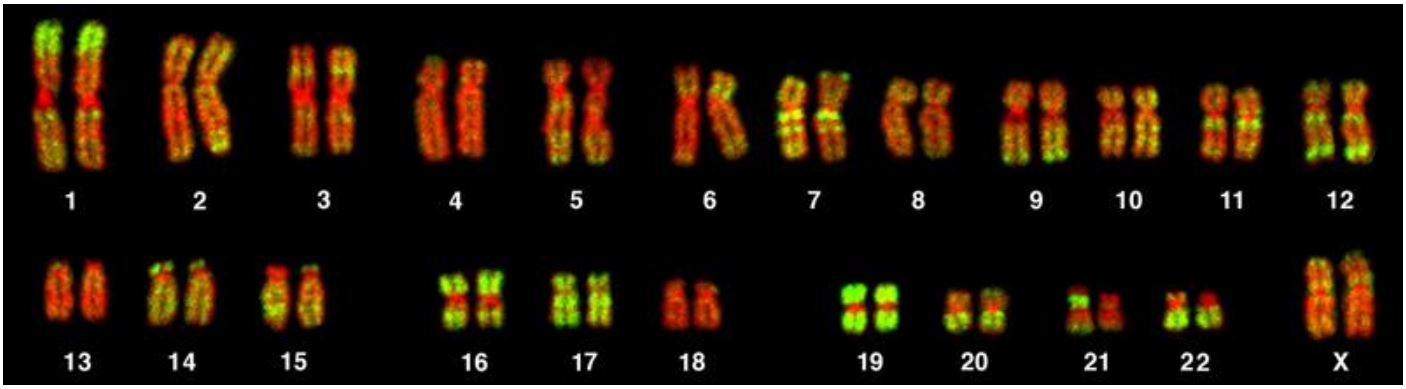


Figure 13.5 This karyotype is of a female human. Notice that homologous chromosomes are the same size, and have the same centromere positions and banding patterns. A human male would have an XY chromosome pair instead of the XX pair. (credit: Andreas Blozer et al)

In a given species, we can identify chromosomes by their number, size, centromere position, and banding pattern. In a human karyotype, **autosomes** or "body chromosomes" (all of the non–sex chromosomes) are generally organized in approximate order of size from largest (chromosome 1) to smallest (chromosome 22). The X and Y chromosomes are not autosomes. However, chromosome 21 is actually shorter than chromosome 22. Researchers discovered this after naming Down syndrome as trisomy 21, reflecting how this disease results from possessing one extra chromosome 21 (three total). Not wanting to change the name of this important disease, scientists retained the numbering of chromosome 21 despite describing it having the shortest set of chromosomes. We may designate the chromosome "arms" projecting from either end of the centromere as short or long, depending on their relative lengths. We abbreviate the short arm *p* (for "petite"); whereas, we abbreviate the long arm *q* (because it

follows “p” alphabetically). Numbers further subdivide and denote each arm. Using this naming system, we can describe chromosome locations consistently in the scientific literature.

CAREER CONNECTION

Geneticists Use Karyograms to Identify Chromosomal Aberrations

Although we refer to Mendel 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 geneticists can identify traits characterized by chromosomal abnormalities from a single cell. To observe an individual’s karyotype, a geneticist first collects a person’s cells (like white blood cells) from a blood sample or other tissue. In the laboratory, he or she stimulates the isolated cells to begin actively dividing. The geneticist then applies the chemical colchicine to cells to arrest condensed chromosomes in metaphase. The geneticist then induces swelling in the cells using a hypotonic solution so the chromosomes spread apart. Finally, the geneticist preserves the sample in a fixative and applies it to a slide.

The geneticist then stains chromosomes with one of several dyes to better visualize each chromosome pair's distinct and reproducible banding patterns. Following staining, the geneticist views the chromosomes 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 23 chromosome pairs. An experienced geneticist can identify each band. In addition to the banding patterns, geneticists further identify chromosomes on the basis of size and centromere location. To obtain the classic depiction of the karyotype in which homologous chromosome pairs align in numerical order from longest to shortest, the geneticist obtains a digital image, identifies each chromosome, and manually arranges the chromosomes into this pattern (Figure 13.5).

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 one identifies by a third copy of chromosome 21, and Turner Syndrome, which has the presence of only one X chromosome in women instead of the normal two characterizes. Geneticists can also identify large DNA deletions or insertions. For instance, geneticists can identify Jacobsen

Syndrome—which involves distinctive facial features as well as heart and bleeding defects—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 one could only infer by performing crosses and observing the traits that offspring expressed. By observing a karyogram, today's geneticists can actually visualize an individual's chromosomal composition to confirm or predict genetic abnormalities in offspring, even before birth.

Chromosome Number Disorders

Of all of the chromosomal disorders, chromosome number abnormalities are the most obviously identifiable from a karyogram. Chromosome number disorders include duplicating or losing entire chromosomes, as well as changes in the number of complete sets of chromosomes. They are caused by **nondisjunction**, which occurs when homologous chromosome pairs or sister chromatids fail to separate during meiosis. Misaligned or incomplete synapsis, or a spindle apparatus dysfunction that facilitates chromosome migration, can cause nondisjunction. The risk of nondisjunction occurring increases with the parents' age.

Nondisjunction can occur during either meiosis I or II, with differing results (Figure 13.6). If homologous chromosomes fail to separate during meiosis I, the result is two gametes that lack that particular chromosome and two gametes with two chromosome copies. If sister chromatids fail to separate during meiosis II, the result is one gamete that lacks that chromosome, two normal gametes with one chromosome copy, and one gamete with two chromosome copies.

VISUAL CONNECTION

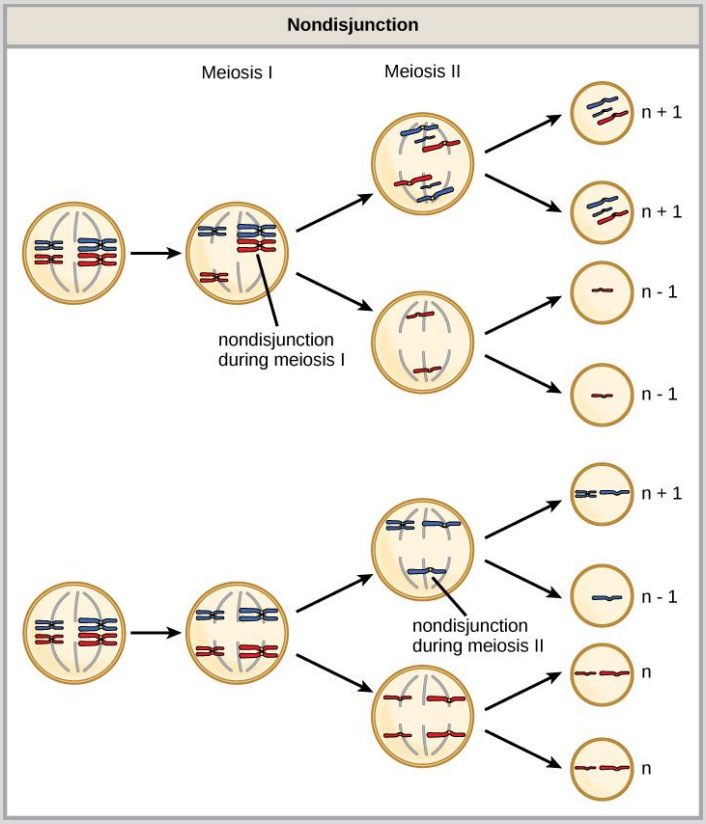


Figure 13.6 Nondisjunction occurs when homologous chromosomes or sister chromatids fail to separate during meiosis, resulting in an abnormal chromosome number. Nondisjunction may occur during meiosis I or meiosis II.

Which of the following statements about nondisjunction is true?

a. Nondisjunction only results in gametes with n+1 or n–1 chromosomes.

b. Nondisjunction occurring during meiosis II results in 50 percent normal gametes.

c. Nondisjunction during meiosis I results in 50 percent normal gametes.

d. Nondisjunction always results in four different kinds of gametes.

Aneuploidy

Scientists call an individual with the appropriate number of chromosomes for their species **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**, a

term that includes **monosomy** (losing one chromosome) or **trisomy** (gaining an extraneous chromosome). Monosomic human zygotes missing any one copy of an autosome invariably fail to develop to birth because they lack essential genes. This underscores the importance of "gene dosage" in humans. Most autosomal trisomies also fail to develop to birth; however, duplications of some smaller chromosomes (13, 15, 18, 21, or 22) can result in offspring that survive for several weeks to many years. Trisomic individuals suffer from a different type of genetic imbalance: an excess in gene dose. Individuals with an extra chromosome may synthesize an abundance of the gene products, which that chromosome encodes. This extra dose (150 percent) 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. Short stature and stunted digits, facial distinctions that include a broad skull and large tongue, and significant developmental delays characterize individuals with this inherited disorder. We can correlate the incidence of Down syndrome with maternal age. Older women are more likely to become pregnant with fetuses carrying the trisomy 21 genotype (Figure 13.7).

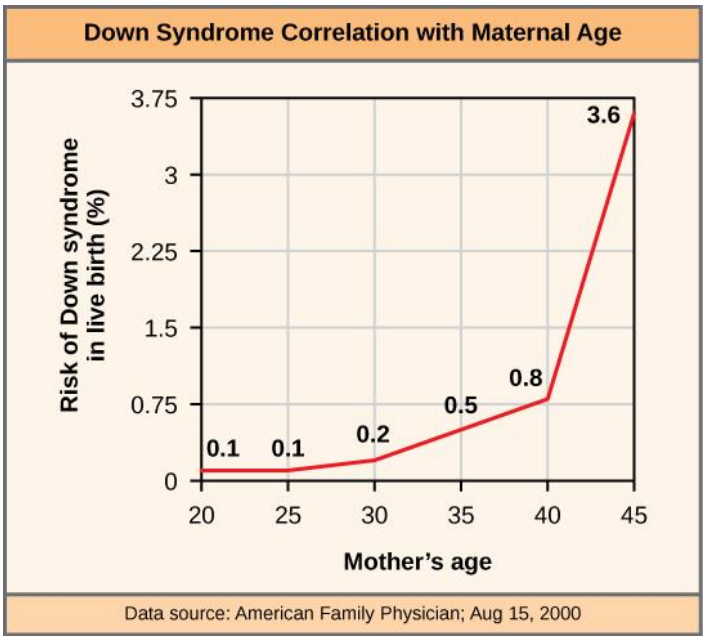


Figure 13.7 The incidence of having a fetus with trisomy 21 increases dramatically with maternal age.

LINK TO LEARNING

Visualize adding a chromosome that leads to Down syndrome in this video simulation: https://www.youtube.com/watch?v=EA0qxhR2oOk

Polyploidy

We call an individual with more than the correct number of chromosome sets (two for diploid species) **polyploid**. For instance, fertilizing an abnormal diploid egg with a normal haploid sperm would yield a triploid zygote. Polyploid animals are extremely rare, with only a few examples among the flatworms, crustaceans, amphibians, fish, and lizards. Polyploid animals are sterile because meiosis cannot proceed normally and instead produces mostly aneuploid daughter cells that cannot yield viable zygotes. Rarely, polyploid animals can reproduce asexually by haplodiploidy, in which an unfertilized egg divides mitotically to produce offspring. 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 13.8).

Figure 13.8 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)

Sex Chromosome Nondisjunction in Humans

Humans display dramatic deleterious effects with autosomal trisomies and monosomies. Therefore, it may seem counterintuitive that human females and males can function normally,

despite carrying different numbers of the X chromosome. Rather than a gain or loss of autosomes, variations in the number of sex chromosomes occur with relatively mild effects. In part, this happens because of the molecular process **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 quiescent (dormant) structure, or a Barr body. The chance that an X chromosome (maternally or paternally derived) inactivates in each cell is random, but once this occurs, all cells derived from that one will have the same inactive X chromosome or Barr body. By this process, females compensate for their double genetic dose of X chromosome. In so-called “tortoiseshell” cats, we observe embryonic X inactivation as color variegation (Figure 13.9). 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 inactivates in that region's embryonic cell progenitor.

Figure 13.9 In cats, the gene for coat color is located on the X chromosome. In female cats' embryonic development, one of the two X chromosomes randomly inactivates in each cell, resulting in a tortoiseshell pattern if the cat has two different alleles for coat color. Male cats, having only one X chromosome, never exhibit a tortoiseshell coat color. (credit: Michael Bodega)

An individual carrying an abnormal number of X chromosomes will inactivate all but one X chromosome in each of her cells. 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 typically occur with mild mental and physical defects, as well as sterility. If the X chromosome is absent altogether, the individual will not develop in utero.

Scientists have identified and characterized several errors in sex chromosome number. Individuals with three X chromosomes, triplo-X, 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. We see this 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.

Duplications and Deletions

In addition to losing or gaining an entire chromosome, a chromosomal segment may duplicate or lose itself. 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 that occurs with nervous system abnormalities and identifiable physical features that result from a deletion of most 5p (the small arm of chromosome 5) (Figure 13.10). Infants with this genotype emit a characteristic high-pitched cry on which the disorder’s name is based.

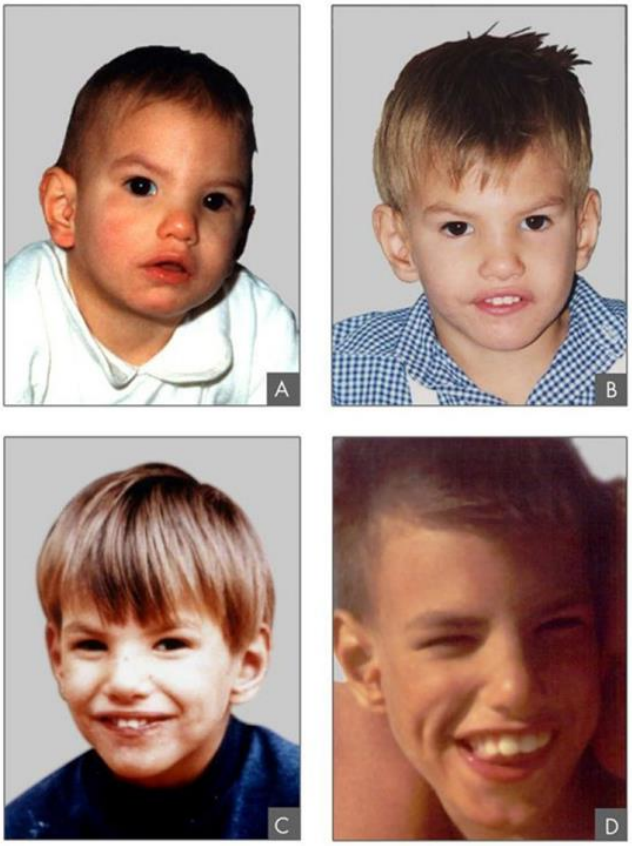


Figure 13.10 This figure shows an individual with cri-du-chat syndrome at two, four, nine, and 12 years of age. (credit: Paola Cerruti Mainardi)

Chromosomal Structural Rearrangements

Cytologists have characterized numerous structural rearrangements in chromosomes, but chromosome inversions and translocations are the most common. We can identify both during meiosis by the adaptive pairing of rearranged chromosomes with their former homologs to maintain appropriate gene alignment. If the genes on two homologs are not oriented correctly, a recombination event could result in losing genes from one chromosome and gaining genes on the other. This would produce aneuploid gametes.

Chromosome Inversions

A **chromosome inversion** is the detachment, 180° rotation, and reinsertion of part of a chromosome. Inversions may occur in nature as a result of mechanical shear, or from transposable elements' action (special DNA sequences capable of facilitating rearranging

chromosome segments with the help of enzymes that cut and paste DNA sequences). Unless they disrupt a gene sequence, inversions only change gene orientation and are likely to have more mild effects than aneuploid errors. However, altered gene orientation can result in functional changes because regulators of gene expression could move 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 the centromere (Figure 13.11). A pericentric inversion that is asymmetric about the centromere can change the chromosome arms' relative lengths, making these inversions easily identifiable.

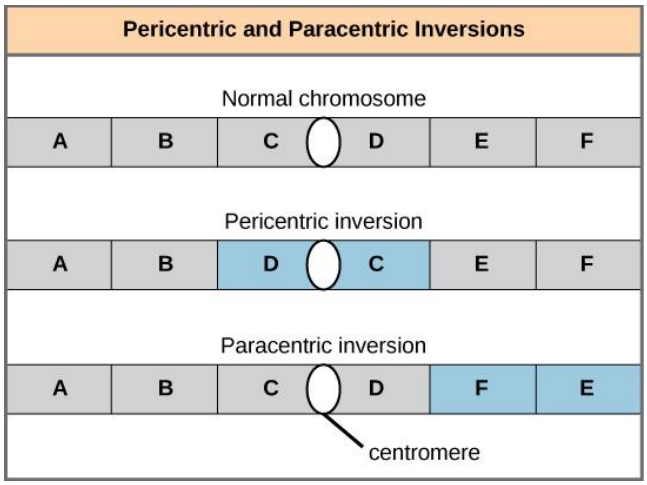


Figure 13.11 Pericentric inversions include the centromere, and paracentric inversions do not. A pericentric inversion can change the chromosome arms' relative lengths. A paracentric inversion cannot.

When one homologous chromosome undergoes an inversion but the other does not, the individual is an inversion heterozygote. To maintain point-for-point synapsis during meiosis, one homolog must form a loop, and the other homolog must mold around it. Although this topology can ensure that the genes correctly align, it also forces the homologs to stretch and can occur with imprecise synapsis regions (Figure 13.12).

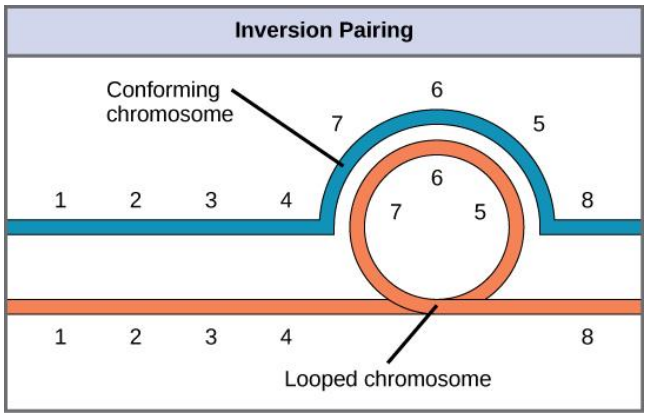


Figure 13.12 When one chromosome undergoes an inversion but the other does not, one chromosome must form an inverted loop to retain point-for-point interaction during synapsis. This inversion pairing is essential to maintaining gene alignment during meiosis and to allow for recombination.

EVOLUTION CONNECTION

The Chromosome 18 Inversion

Not all chromosomes' structural rearrangements produce nonviable, impaired, or infertile individuals. In rare instances, such a change can result in new species evolving. In fact, a pericentric inversion in chromosome 18 appears to have contributed to human evolution. This inversion is not present in our closest genetic relatives, the chimpanzees. Humans and chimpanzees differ cytogenetically by pericentric inversions on several chromosomes and by the fusion of two separate chromosomes in chimpanzees that correspond to chromosome two in humans.

Scientists believe the pericentric chromosome 18 inversion 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 an ancestral human 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. We do not know how this inversion contributed to hominid evolution, but it appears to be a significant factor in the divergence of humans from other primates.[1]

Translocations

A **translocation** occurs when a chromosome segment dissociates and reattaches to a different, nonhomologous 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 occurred with several cancers and with schizophrenia. Reciprocal translocations result from exchanging chromosome segments between two nonhomologous chromosomes such that there is no genetic information gain or loss (Figure 13.13).

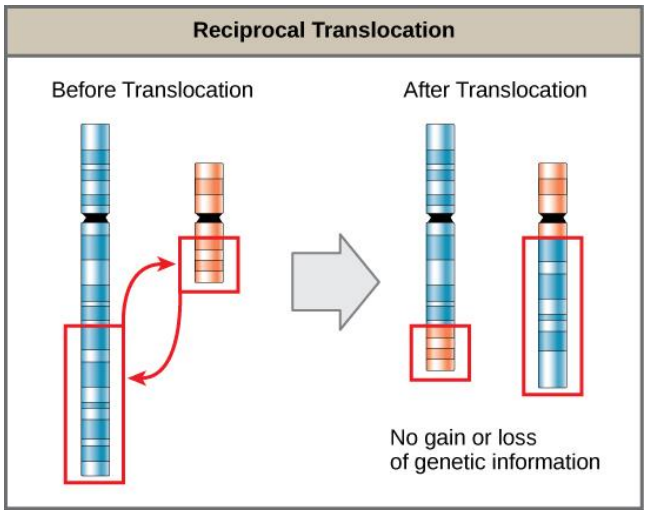


Figure 13.13 A reciprocal translocation occurs when a DNA segment transfers from one chromosome to another, nonhomologous chromosome. (credit: modification of work by National Human Genome Research/USA)

Footnotes

- 1 Violaine Goidts et al., “Segmental 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

Chapter Attributions