5

The Structure and Function of Large Biological Molecules

▲ **Figure 5.1 Why do scientists study the structures of macromolecules?**

KEY CONCEPTS

OVERVIEW

The Molecules of Life

Given the rich complexity of life on Earth, we might expect organisms to have an enormous diversity of molecules. Remarkably, however, the critically important large molecules of all living things—from bacteria to elephants—fall into just four main classes: carbohydrates, lipids, proteins, and nucleic acids. On the molecular scale, members of three of these classes—carbohydrates, proteins, and nucleic acids—are huge and are therefore called **macromolecules**. For example, a protein may consist of thousands of atoms that form a molecular colossus with a mass well over 100,000 daltons. Considering the size and complexity of macromolecules, it is noteworthy that biochemists have determined the detailed structure of so many of them. The scientist in the foreground of **Figure 5.1** is using 3-D glasses to help her visualize the structure of the protein displayed on her screen.

The architecture of a large biological molecule helps explain how that molecule works. Like water and simple organic molecules, large biological molecules exhibit unique emergent properties arising from the orderly arrangement of their atoms. In this chapter, we'll first consider how macromolecules are built. Then we'll examine the structure and function of all four classes of large biological molecules: carbohydrates, lipids, proteins, and nucleic acids.

CONCEPT **5.1**

Macromolecules are polymers, built from monomers

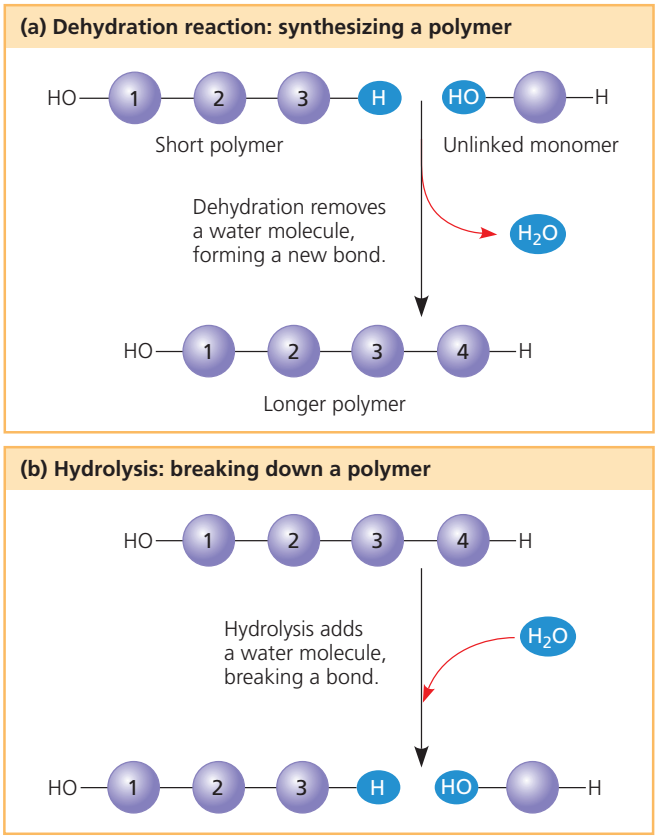
The macromolecules in three of the four classes of life's organic compounds—carbohydrates, proteins, and nucleic acids—are chain-like molecules called polymers (from the Greek *polys*, many, and *meros*, part). A **polymer** is a long molecule consisting of many similar or identical building blocks linked by covalent bonds, much as a train consists of a chain of cars. The repeating units that serve as the building blocks of a polymer are smaller molecules called **monomers** (from the Greek *monos*, single). Some of the molecules that serve as monomers also have other functions of their own.

The Synthesis and Breakdown of Polymers

Although each class of polymer is made up of a different type of monomer, the chemical mechanisms by which cells make and break down polymers are basically the same in all cases. In cells, these processes are facilitated by **enzymes**, specialized macromolecules that speed up chemical reactions. Monomers are connected by a reaction in which two molecules are covalently bonded to each other, with the loss of a water molecule; this is known as a **dehydration reaction** **(Figure 5.2a)**. When a bond forms between two monomers, each monomer contributes part of the water molecule that is released during the reaction: One monomer provides a hydroxyl group (—OH), while the other provides a hydrogen (—H). This reaction is repeated as monomers are added to the chain one by one, making a polymer.

Polymers are disassembled to monomers by **hydrolysis**, a process that is essentially the reverse of the dehydration reac-

▼ **Figure 5.2 The synthesis and breakdown of polymers.**



tion **(Figure 5.2b)**. Hydrolysis means to break using water (from the Greek *hydro*, water, and *lysis*, break). The bond between the monomers is broken by the addition of a water molecule, with the hydrogen from the water attaching to one monomer and the hydroxyl group attaching to the adjacent monomer. An example of hydrolysis working within our bodies is the process of digestion. The bulk of the organic material in our food is in the form of polymers that are much too large to enter our cells. Within the digestive tract, various enzymes attack the polymers, speeding up hydrolysis. The released monomers are then absorbed into the bloodstream for distribution to all body cells. Those cells can then use dehydration reactions to assemble the monomers into new, different polymers that can perform specific functions required by the cell.

The Diversity of Polymers

Each cell has thousands of different macromolecules; the collection varies from one type of cell to another even in the same organism. The inherent differences between human siblings reflect small variations in polymers, particularly DNA and proteins. Molecular differences between unrelated individuals are more extensive and those between species greater still. The diversity of macromolecules in the living world is vast, and the possible variety is effectively limitless.

What is the basis for such diversity in life's polymers? These molecules are constructed from only 40 to 50 common monomers and some others that occur rarely. Building a huge variety of polymers from such a limited number of monomers is analogous to constructing hundreds of thousands of words from only 26 letters of the alphabet. The key is arrangement—the particular linear sequence that the units follow. However, this analogy falls far short of describing the great diversity of macromolecules because most biological polymers have many more monomers than the number of letters in the longest word. Proteins, for example, are built from 20 kinds of amino acids arranged in chains that are typically hundreds of amino acids long. The molecular logic of life is simple but elegant: Small molecules common to all organisms are ordered into unique macromolecules.

Despite this immense diversity, molecular structure and function can still be grouped roughly by class. Let's examine each of the four major classes of large biological molecules. For each class, the large molecules have emergent properties not found in their individual building blocks.

CONCEPT CHECK 5.1

1. What are the four main classes of large biological molecules? Which class does not consist of polymers?
2. How many molecules of water are needed to completely hydrolyze a polymer that is ten monomers long?
3. **WHAT IF?** Suppose you eat a serving of fish. What reactions must occur for the amino acid monomers in the protein of the fish to be converted to new proteins in your body?

For suggested answers, see Appendix A.

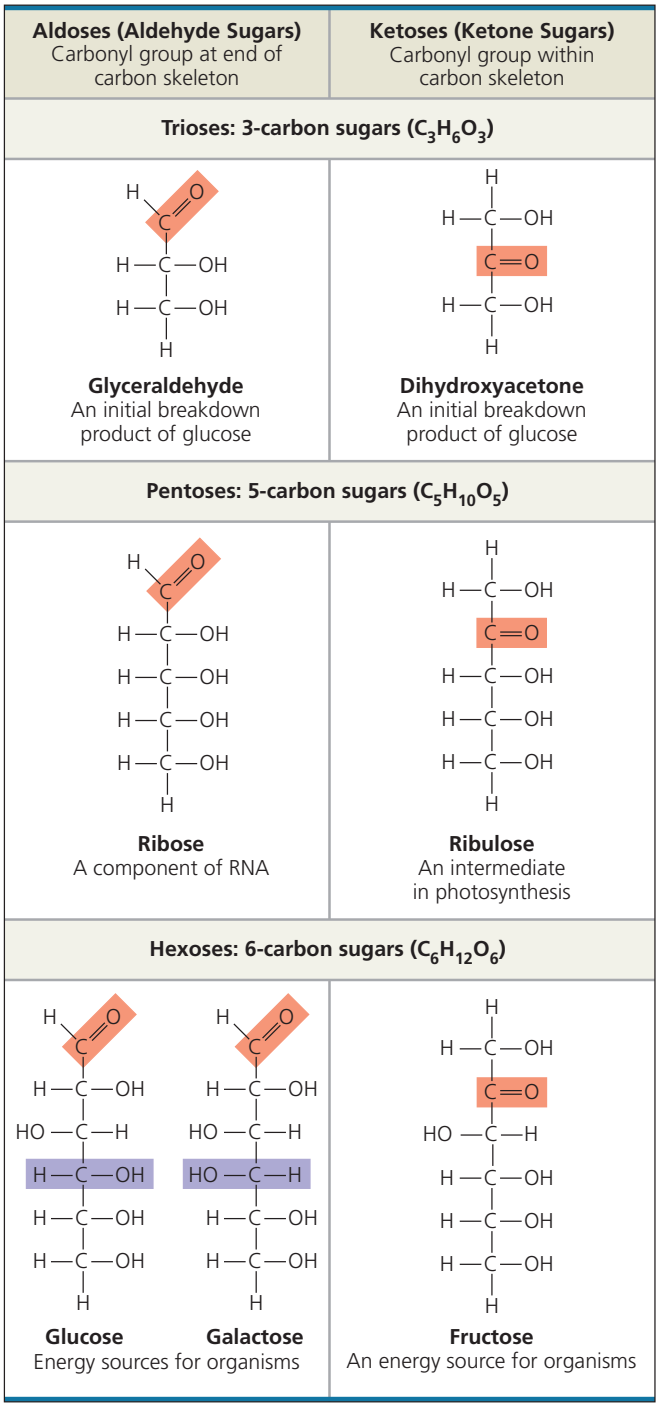
CONCEPT 5.2

Carbohydrates serve as fuel and building material

Carbohydrates include both sugars and polymers of sugars. The simplest carbohydrates are the monosaccharides, or simple sugars; these are the monomers from which more complex carbohydrates are constructed. Disaccharides are double sugars, consisting of two monosaccharides joined by a covalent bond. Carbohydrates also include macromolecules called polysaccharides, polymers composed of many sugar building blocks.

Sugars

Monosaccharides (from the Greek *monos*, single, and *sacchar*, sugar) generally have molecular formulas that are some multiple of the unit CH_2O. Glucose ($C_6H_{12}O_6$), the most common monosaccharide, is of central importance in the chemistry



▲ **Figure 5.3 The structure and classification of some monosaccharides.** Sugars vary in the location of their carbonyl groups (orange), the length of their carbon skeletons, and the spatial arrangement around asymmetric carbons (compare, for example, the purple portions of glucose and galactose).

MAKE CONNECTIONS *In the 1970s, a process was developed that converts the glucose in corn syrup to its sweeter isomer, fructose. High-fructose corn syrup, a common ingredient in soft drinks and processed food, is a mixture of glucose and fructose. What type of isomers are glucose and fructose? See Figure 4.7, p. 62.*

of life. In the structure of glucose, we can see the trademarks of a sugar: The molecule has a carbonyl group (C=O) and multiple hydroxyl groups (—OH) **(Figure 5.3)**. Depending on the location of the carbonyl group, a sugar is either an aldose (aldehyde sugar) or a ketose (ketone sugar). Glucose, for example, is an aldose; fructose, an isomer of glucose, is a ketose. (Most names for sugars end in *-ose*.) Another criterion for classifying sugars is the size of the carbon skeleton, which ranges from three to seven carbons long. Glucose, fructose, and other sugars that have six carbons are called hexoses. Trioses (three-carbon sugars) and pentoses (five-carbon sugars) are also common.

Still another source of diversity for simple sugars is in the spatial arrangement of their parts around asymmetric carbons. (Recall that an asymmetric carbon is a carbon attached to four different atoms or groups of atoms.) Glucose and galactose, for example, differ only in the placement of parts around one asymmetric carbon (see the purple boxes in Figure 5.3). What seems like a small difference is significant enough to give the two sugars distinctive shapes and behaviors.

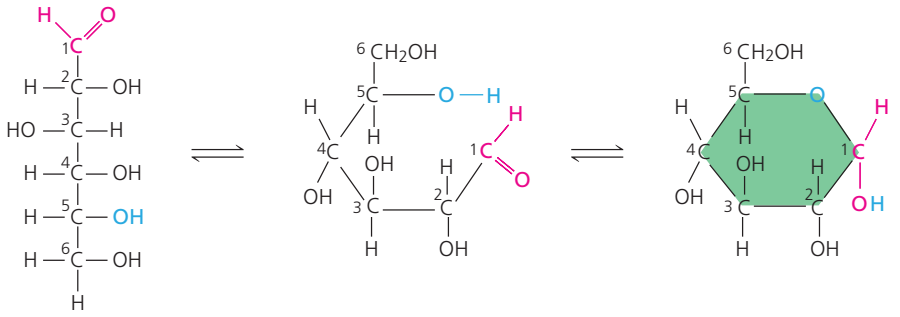
Although it is convenient to draw glucose with a linear carbon skeleton, this representation is not completely accurate. In aqueous solutions, glucose molecules, as well as most other five- and six-carbon sugars, form rings **(Figure 5.4)**.

Monosaccharides, particularly glucose, are major nutrients for cells. In the process known as cellular respiration, cells extract energy in a series of reactions starting with glucose molecules. Simple-sugar molecules are not only a major fuel for cellular work, but their carbon skeletons also serve as raw material for the synthesis of other types of small organic molecules, such as amino acids and fatty acids. Sugar molecules that are not immediately used in these ways are generally incorporated as monomers into disaccharides or polysaccharides.

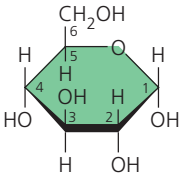
A **disaccharide** consists of two monosaccharides joined by a **glycosidic linkage**, a covalent bond formed between two monosaccharides by a dehydration reaction. For example, maltose is a disaccharide formed by the linking of two molecules of glucose **(Figure 5.5a)**. Also known as malt sugar, maltose is an ingredient used in brewing beer. The most prevalent disaccharide is sucrose, which is table sugar. Its two monomers are glucose and fructose **(Figure 5.5b)**. Plants generally transport carbohydrates from leaves to roots and other nonphotosynthetic organs in the form of sucrose. Lactose, the sugar present in milk, is another disaccharide, in this case a glucose molecule joined to a galactose molecule.

Polysaccharides

Polysaccharides are macromolecules, polymers with a few hundred to a few thousand monosaccharides joined by glycosidic linkages. Some polysaccharides serve as storage material, hydrolyzed as needed to provide sugar for cells. Other polysaccharides serve as building material for structures that



(a) Linear and ring forms. Chemical equilibrium between the linear and ring structures greatly favors the formation of rings. The carbons of the sugar are numbered 1 to 6, as shown. To form the glucose ring, carbon 1 bonds to the oxygen attached to carbon 5.

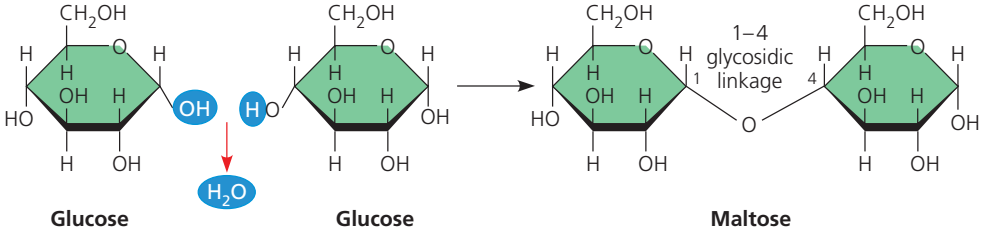


(b) Abbreviated ring structure. Each corner represents a carbon. The ring's thicker edge indicates that you are looking at the ring edge-on; the components attached to the ring lie above or below the plane of the ring.

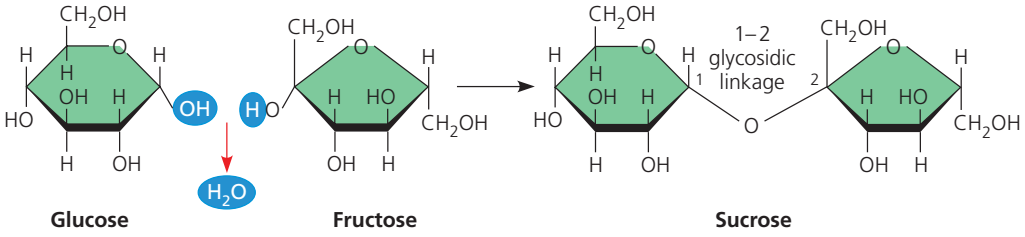
▲ Figure 5.4 Linear and ring forms of glucose.

DRAW IT *Start with the linear form of fructose (see Figure 5.3) and draw the formation of the fructose ring in two steps. First, number the carbons starting at the top of the linear structure. Then attach carbon 5 via its oxygen to carbon 2. Compare the number of carbons in the fructose and glucose rings.*

(a) Dehydration reaction in the synthesis of maltose. The bonding of two glucose units forms maltose. The glycosidic linkage joins the number 1 carbon of one glucose to the number 4 carbon of the second glucose. Joining the glucose monomers in a different way would result in a different disaccharide.



(b) Dehydration reaction in the synthesis of sucrose. Sucrose is a disaccharide formed from glucose and fructose. Notice that fructose, though a hexose like glucose, forms a five-sided ring.



▲ Figure 5.5 Examples of disaccharide synthesis.

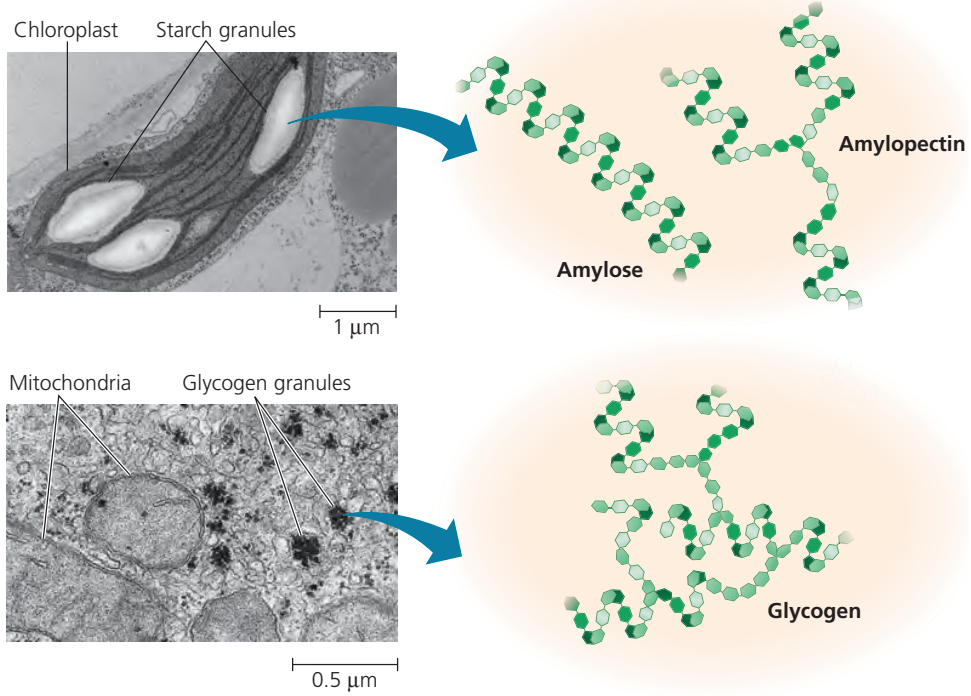
DRAW IT *Referring to Figure 5.4, number the carbons in each sugar in this figure. Show how the numbering is consistent with the name of the glycosidic linkage in each disaccharide.*

protect the cell or the whole organism. The architecture and function of a polysaccharide are determined by its sugar monomers and by the positions of its glycosidic linkages.

Storage Polysaccharides

Both plants and animals store sugars for later use in the form of storage polysaccharides. Plants store **starch**, a polymer of glucose monomers, as granules within cellular structures known as plastids, which include chloroplasts. Synthesizing starch enables the plant to stockpile surplus glucose. Because glucose is a major cellular fuel, starch represents stored energy. The sugar can later be withdrawn from this carbohydrate "bank" by hydrolysis, which breaks the bonds between the glucose monomers. Most animals, including humans, also have enzymes that can hydrolyze plant starch, making glucose available as a nutrient for cells. Potato tubers and grains—the fruits of wheat, maize (corn), rice, and other grasses—are the major sources of starch in the human diet.

Most of the glucose monomers in starch are joined by 1–4 linkages (number 1 carbon to number 4 carbon), like the glucose units in maltose (see Figure 5.5a). The simplest form of starch, amylose, is unbranched. Amylopectin, a more complex



(a) Starch: a plant polysaccharide. This micrograph shows part of a plant cell with a chloroplast, the cellular organelle where glucose is made and then stored as starch granules. Amylose (unbranched) and amylopectin (branched) are two forms of starch.

(b) Glycogen: an animal polysaccharide. Animal cells stockpile glycogen as dense clusters of granules within liver and muscle cells, as shown in this micrograph of part of a liver cell. Mitochondria are cellular organelles that help break down glucose released from glycogen. Note that glycogen is more branched than amylopectin starch.

▲ **Figure 5.6 Storage polysaccharides of plants and animals.** These examples, starch and glycogen, are composed entirely of glucose monomers, represented here by hexagons. Because of the angle of the 1–4 linkages, the polymer chains tend to form helices in unbranched regions.

starch, is a branched polymer with 1–6 linkages at the branch points. Both of these starches are shown in **Figure 5.6a**.

Animals store a polysaccharide called **glycogen**, a polymer of glucose that is like amylopectin but more extensively branched **(Figure 5.6b)**. Humans and other vertebrates store glycogen mainly in liver and muscle cells. Hydrolysis of glycogen in these cells releases glucose when the demand for sugar increases. This stored fuel cannot sustain an animal for long, however. In humans, for example, glycogen stores are depleted in about a day unless they are replenished by consumption of food. This is an issue of concern in low-carbohydrate diets.

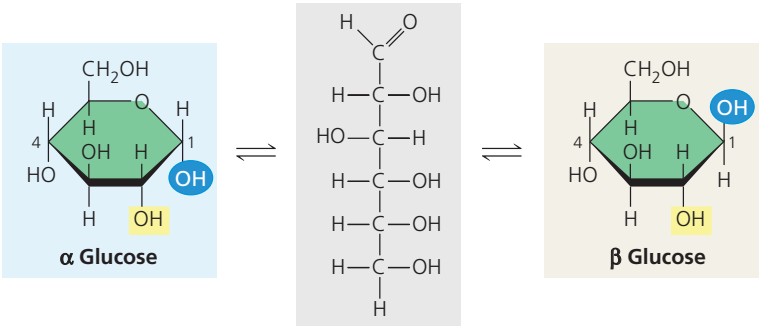
Structural Polysaccharides

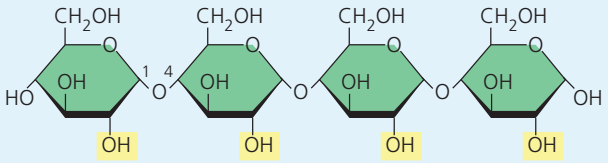
Organisms build strong materials from structural polysaccharides. For example, the polysaccharide called **cellulose** is a major component of the tough walls that enclose plant cells. On a global scale, plants produce almost 10^{14} kg (100 billion tons) of cellulose per year; it is the most abundant organic compound on Earth. Like starch, cellulose is a polymer of glucose, but the glycosidic linkages in these two polymers differ. The difference is based on the fact that there are actually two slightly different ring structures for glucose **(Figure 5.7a)**. When glucose forms a ring, the hydroxyl group attached to the number 1 carbon is positioned either below or above the plane of the ring. These two ring forms for glucose are called alpha (α) and beta (β), respectively. In starch, all the glucose monomers are in the α configuration **(Figure 5.7b)**, the arrangement we saw in Figures 5.4 and 5.5. In contrast, the glucose monomers of cellulose are all in the β configuration, making every glucose monomer "upside down" with respect to its neighbors **(Figure 5.7c)**.

The differing glycosidic linkages in starch and cellulose give the two molecules distinct three-dimensional shapes. Whereas certain starch molecules are largely helical, a cellulose molecule is straight. Cellulose is never branched, and some hydroxyl groups on its glucose monomers are free to hydrogen-bond with the hydroxyls of other cellulose molecules lying parallel to it. In plant cell walls, parallel cellulose molecules held together in this way are grouped into units called microfibrils **(Figure 5.8)**. These cable-like microfibrils are a strong building material for plants and an important substance for humans because cellulose is the major constituent of paper and the only component of cotton.

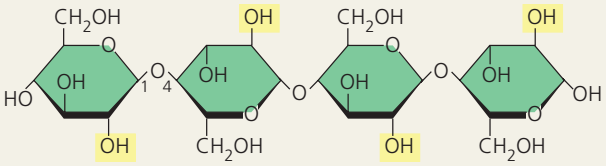
Enzymes that digest starch by hydrolyzing its α linkages are unable to hydrolyze the β linkages of cellulose because of the distinctly different shapes of these two molecules. In fact, few organisms possess enzymes that can digest cellulose. Animals, including humans, do not; the cellulose in our food passes through the digestive tract and is eliminated with the feces. Along the way, the cellulose abrades the wall of the digestive tract and stimulates the lining to secrete mucus, which aids in the smooth passage of food through the tract. Thus, although cellulose is not a nutrient for humans, it is an important part of a healthful diet. Most fresh fruits, vegetables, and whole grains are rich in cellulose. On food packages, "insoluble fiber" refers mainly to cellulose.

(a) α and β glucose ring structures. These two interconvertible forms of glucose differ in the placement of the hydroxyl group (highlighted in blue) attached to the number 1 carbon.



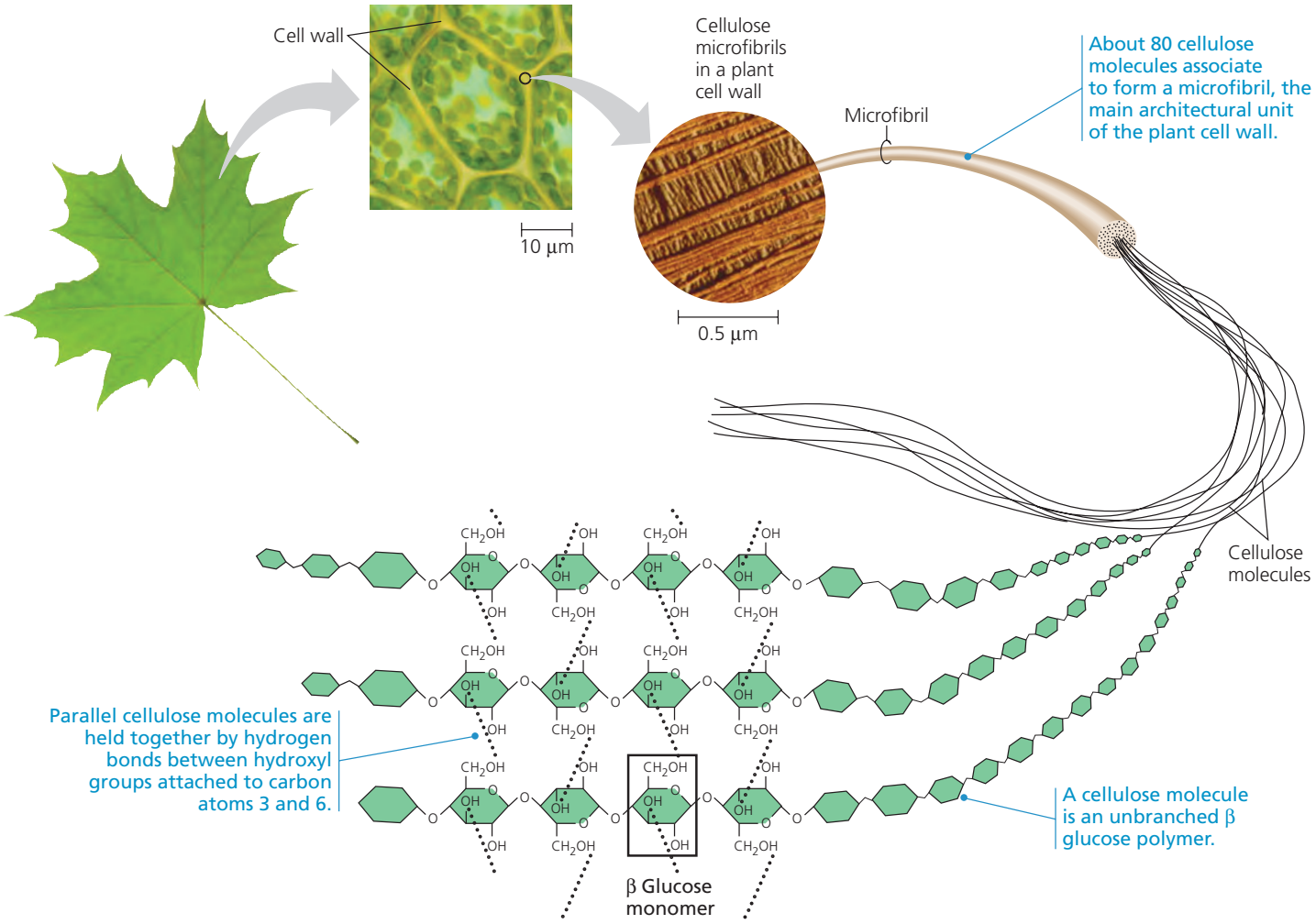


(b) Starch: 1–4 linkage of α glucose monomers. All monomers are in the same orientation. Compare the positions of the —OH groups highlighted in yellow with those in cellulose (c).



(c) Cellulose: 1–4 linkage of β glucose monomers. In cellulose, every β glucose monomer is upside down with respect to its neighbors.

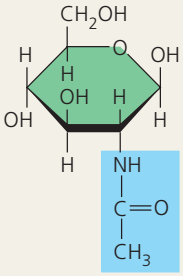
▲ **Figure 5.7 Starch and cellulose structures.**



▲ **Figure 5.8 The arrangement of cellulose in plant cell walls.**

Some microorganisms can digest cellulose, breaking it down into glucose monomers. A cow harbors cellulose-digesting prokaryotes and protists in its stomach. These microbes hydrolyze the cellulose of hay and grass and convert the glucose to other compounds that nourish the cow. Similarly, a termite, which is unable to digest cellulose by itself, has prokaryotes or protists living in its gut that can make a meal of wood. Some fungi can also digest cellulose, thereby helping recycle chemical elements within Earth's ecosystems.

Another important structural polysaccharide is **chitin**, the carbohydrate used by arthropods (insects, spiders, crustaceans, and related animals) to build their exoskeletons **(Figure 5.9)**. An exoskeleton is a hard case that surrounds the soft parts of an animal. Pure chitin is leathery and flexible, but it becomes hardened when encrusted with calcium carbonate, a salt. Chitin is also found in many fungi, which use this polysaccharide rather than cellulose as the building material for their cell walls. Chitin is similar to cellulose, with β linkages, except that the glucose monomer of chitin has a nitrogen-containing appendage (see Figure 5.9, top right).



◀ The structure of the chitin monomer

◀ Chitin forms the exoskeleton of arthropods. This cicada is molting, shedding its old exoskeleton and emerging in adult form.

▲ Chitin is used to make a strong and flexible surgical thread that decomposes after the wound or incision heals.

▲ **Figure 5.9 Chitin, a structural polysaccharide.**

CONCEPT CHECK 5.2

1. Write the formula for a monosaccharide that has three carbons.
2. A dehydration reaction joins two glucose molecules to form maltose. The formula for glucose is $C_6H_{12}O_6$. What is the formula for maltose?
3. **WHAT IF?** After a cow is given antibiotics to treat an infection, a vet gives the animal a drink of "gut culture" containing various prokaryotes. Why is this necessary?

For suggested answers, see Appendix A.

CONCEPT 5.3

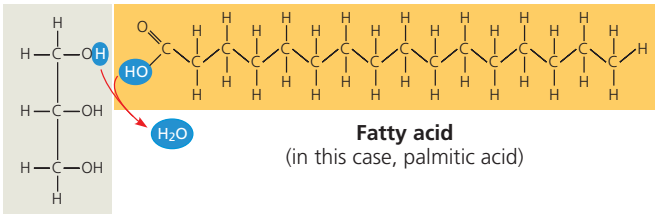
Lipids are a diverse group of hydrophobic molecules

Lipids are the one class of large biological molecules that does not include true polymers, and they are generally not big enough to be considered macromolecules. The compounds called **lipids** are grouped together because they share one important trait: They mix poorly, if at all, with water. The hydrophobic behavior of lipids is based on their molecular structure. Although they may have some polar bonds associated with oxygen, lipids consist mostly of hydrocarbon regions. Lipids are varied in form and function. They include waxes and certain pigments, but we will focus on the most biologically important types of lipids: fats, phospholipids, and steroids.

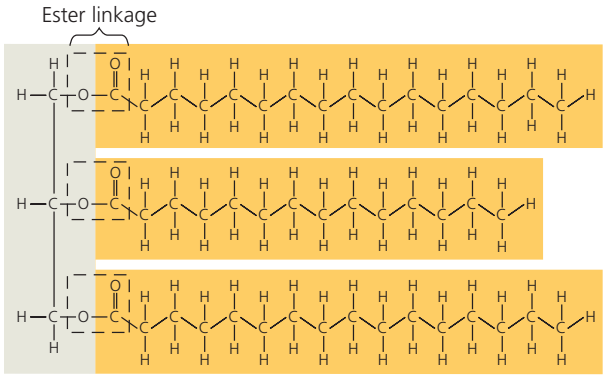
Fats

Although fats are not polymers, they are large molecules assembled from smaller molecules by dehydration reactions. A **fat** is constructed from two kinds of smaller molecules: glycerol and fatty acids **(Figure 5.10a)**. Glycerol is an alcohol; each of its three carbons bears a hydroxyl group. A **fatty acid** has a long carbon skeleton, usually 16 or 18 carbon atoms in length. The carbon at one end of the skeleton is part of a carboxyl group, the functional group that gives these molecules the name fatty *acid*. The rest of the skeleton consists of a hydrocarbon chain. The relatively nonpolar C—H bonds in the hydrocarbon chains of fatty acids are the reason fats are hydrophobic. Fats separate from water because the water molecules hydrogen-bond to one another and exclude the fats. This is the reason that vegetable oil (a liquid fat) separates from the aqueous vinegar solution in a bottle of salad dressing.

In making a fat, three fatty acid molecules are each joined to glycerol by an ester linkage, a bond between a hydroxyl group and a carboxyl group. The resulting fat, also called a **triacylglycerol**, thus consists of three fatty acids linked to one glycerol molecule. (Still another name for a fat is



(a) One of three dehydration reactions in the synthesis of a fat



(b) Fat molecule (triacylglycerol)

▲ Figure 5.10 The synthesis and structure of a fat, or triacylglycerol. The molecular building blocks of a fat are one molecule of glycerol and three molecules of fatty acids. **(a)** One water molecule is removed for each fatty acid joined to the glycerol. **(b)** A fat molecule with three fatty acid units, two of them identical. The carbons of the fatty acids are arranged zigzag to suggest the actual orientations of the four single bonds extending from each carbon (see Figure 4.3a).

triglyceride, a word often found in the list of ingredients on packaged foods.) The fatty acids in a fat can be the same, or they can be of two or three different kinds, as in **Figure 5.10b**.

The terms *saturated fats* and *unsaturated fats* are commonly used in the context of nutrition **(Figure 5.11)**. These terms refer to the structure of the hydrocarbon chains of the fatty acids. If there are no double bonds between carbon atoms composing a chain, then as many hydrogen atoms as possible are bonded to the carbon skeleton. Such a structure is said to be *saturated* with hydrogen, and the resulting fatty acid therefore called a **saturated fatty acid** **(Figure 5.11a)**. An **unsaturated fatty acid** has one or more double bonds, with one fewer hydrogen atom on each double-bonded carbon. Nearly all double bonds in naturally occurring fatty acids are *cis* double bonds, which cause a kink in the hydrocarbon chain wherever they occur **(Figure 5.11b)**. (See Figure 4.7 to remind yourself about *cis* and *trans* double bonds.)

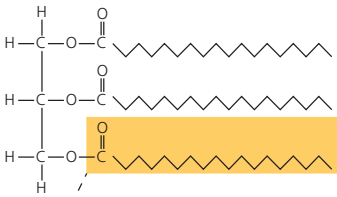
A fat made from saturated fatty acids is called a saturated fat. Most animal fats are saturated: The hydrocarbon chains of their fatty acids—the "tails" of the fat molecules—lack double bonds, and their flexibility allows the fat molecules to pack together tightly. Saturated animal fats—such as lard and butter—are solid at room temperature. In contrast, the fats of plants

▼ Figure 5.11 Saturated and unsaturated fats and fatty acids.

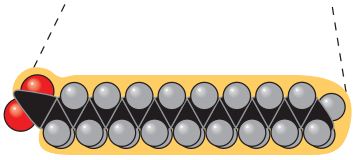
(a) Saturated fat

At room temperature, the molecules of a saturated fat, such as the fat in butter, are packed closely together, forming a solid.

Structural formula of a saturated fat molecule (Each hydrocarbon chain is represented as a zigzag line, where each bend represents a carbon atom and hydrogens are not shown.)



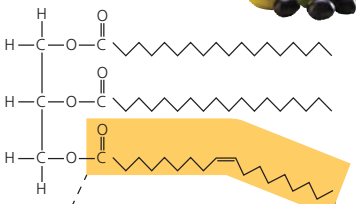
Space-filling model of stearic acid, a saturated fatty acid (red = oxygen, black = carbon, gray = hydrogen)



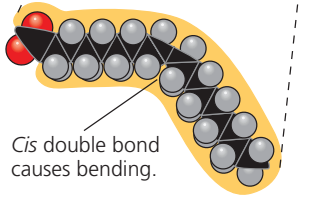
(b) Unsaturated fat

At room temperature, the molecules of an unsaturated fat such as olive oil cannot pack together closely enough to solidify because of the kinks in some of their fatty acid hydrocarbon chains.

Structural formula of an unsaturated fat molecule



Space-filling model of oleic acid, an unsaturated fatty acid



and fishes are generally unsaturated, meaning that they are built of one or more types of unsaturated fatty acids. Usually liquid at room temperature, plant and fish fats are referred to as oils—olive oil and cod liver oil are examples. The kinks where the *cis* double bonds are located prevent the molecules from packing together closely enough to solidify at room temperature. The phrase "hydrogenated vegetable oils" on food labels means that unsaturated fats have been synthetically

converted to saturated fats by adding hydrogen. Peanut butter, margarine, and many other products are hydrogenated to prevent lipids from separating out in liquid (oil) form.

A diet rich in saturated fats is one of several factors that may contribute to the cardiovascular disease known as atherosclerosis. In this condition, deposits called plaques develop within the walls of blood vessels, causing inward bulges that impede blood flow and reduce the resilience of the vessels. Recent studies have shown that the process of hydrogenating vegetable oils produces not only saturated fats but also unsaturated fats with *trans* double bonds. These **trans fats** may contribute more than saturated fats to atherosclerosis (see Chapter 42) and other problems. Because trans fats are especially common in baked goods and processed foods, the U.S. Department of Agriculture requires nutritional labels to include information on trans fat content. Some U.S. cities and at least one country—Denmark—have even banned the use of trans fats in restaurants.

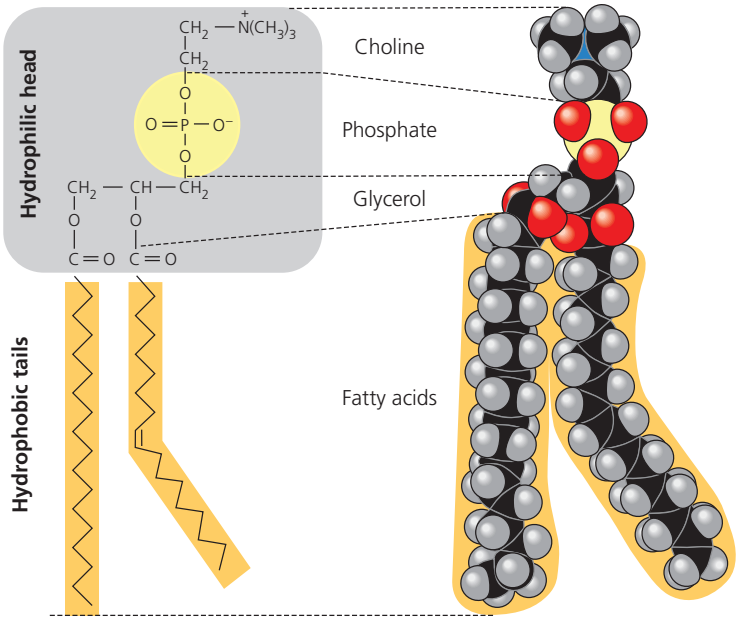
Certain unsaturated fatty acids must be supplied in the human diet because they cannot be synthesized in the body. These essential fatty acids include the omega-3 fatty acids, which are required for normal growth in children and appear to protect against cardiovascular disease in adults. Fatty fish and certain nuts and vegetable oils are rich in omega-3 fatty acids (so named because they have a double bond at the third carbon-carbon bond from the end of the hydrocarbon chain).

The major function of fats is energy storage. The hydrocarbon chains of fats are similar to gasoline molecules and just as rich in energy. A gram of fat stores more than twice as much energy as a gram of a polysaccharide, such as starch. Because plants are relatively immobile, they can function with bulky energy storage in the form of starch. (Vegetable oils are generally obtained from seeds, where more compact storage is an asset to the plant.) Animals, however, must carry their energy stores with them, so there is an advantage to having a more compact reservoir of fuel—fat. Humans and other mammals stock their long-term food reserves in adipose cells (see Figure 4.6a), which swell and shrink as fat is deposited and withdrawn from storage. In addition to storing energy, adipose tissue also cushions such vital organs as the kidneys, and a layer of fat beneath the skin insulates the body. This subcutaneous layer is especially thick in whales, seals, and most other marine mammals, protecting them from cold ocean water.

Phospholipids

Cells could not exist without another type of lipid—**phospholipids** (Figure 5.12). Phospholipids are essential for cells because they make up cell membranes. Their structure provides a classic example of how form fits function at the molecular level. As shown in Figure 5.12, a phospholipid is similar to a fat molecule but has only two fatty acids attached to glycerol rather than three. The third hydroxyl group of glycerol is joined to a phosphate group, which has a negative electrical charge in the cell. Additional small molecules, which are usually charged or polar, can be linked to the phosphate group to form a variety of phospholipids.

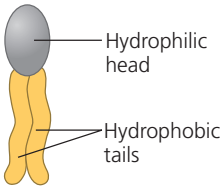
The two ends of phospholipids show different behavior toward water. The hydrocarbon tails are hydrophobic and are excluded from water. However, the phosphate group and its attachments form a hydrophilic head that has an affinity for water. When phospholipids are added to water, they self-assemble into double-layered structures called "bilayers," shielding their hydrophobic portions from water (Figure 5.13).



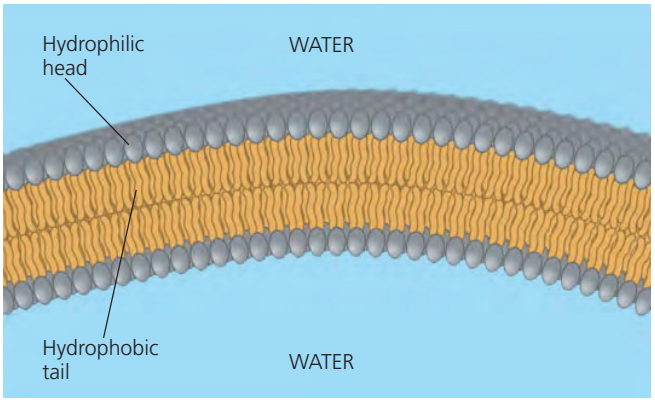
(a) Structural formula (b) Space-filling model

◀ **Figure 5.12 The structure of a phospholipid.** A phospholipid has a hydrophilic (polar) head and two hydrophobic (nonpolar) tails. Phospholipid diversity is based on differences in the two fatty acids and in the groups attached to the phosphate group of the head. This particular phospholipid, called a phosphatidylcholine, has an attached choline group. The kink in one of its tails is due to a *cis* double bond. Shown here are **(a)** the structural formula, **(b)** the space-filling model (yellow = phosphorus, blue = nitrogen), and **(c)** the symbol for a phospholipid that will appear throughout this book. (In most figures, this symbol will be used to represent a phospholipid with either saturated or unsaturated tails.)

DRAW IT *Draw an oval around the hydrophilic head of the space-filling model.*



(c) Phospholipid symbol

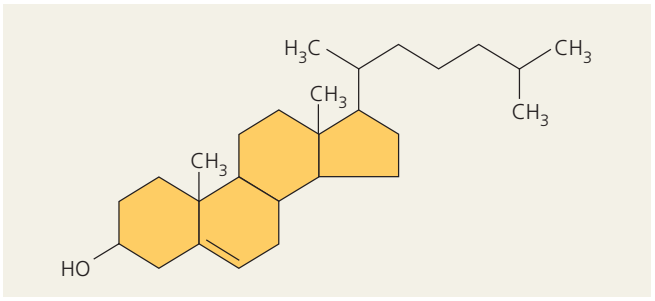


▲ **Figure 5.13 Bilayer structure formed by self-assembly of phospholipids in an aqueous environment.** The phospholipid bilayer shown here is the main fabric of biological membranes. Note that the hydrophilic heads of the phospholipids are in contact with water in this structure, whereas the hydrophobic tails are in contact with each other and remote from water.

At the surface of a cell, phospholipids are arranged in a similar bilayer. The hydrophilic heads of the molecules are on the outside of the bilayer, in contact with the aqueous solutions inside and outside of the cell. The hydrophobic tails point toward the interior of the bilayer, away from the water. The phospholipid bilayer forms a boundary between the cell and its external environment; in fact, cells could not exist without phospholipids.

Steroids

Steroids are lipids characterized by a carbon skeleton consisting of four fused rings. Different steroids, such as cholesterol and the vertebrate sex hormones, are distinguished by the particular chemical groups attached to this ensemble of rings **(Figure 5.14)**. **Cholesterol** is a crucial molecule in animals. It is a common component of animal cell membranes and is also the precursor from which other steroids are synthesized. In vertebrates, cholesterol is synthesized in the liver and obtained from the diet. A high level of cholesterol in the blood may contribute to atherosclerosis. In fact, both saturated fats and trans fats exert their negative impact on health by affecting cholesterol levels.



▲ **Figure 5.14 Cholesterol, a steroid.** Cholesterol is the molecule from which other steroids, including the sex hormones, are synthesized. Steroids vary in the chemical groups attached to their four interconnected rings (shown in gold).

MAKE CONNECTIONS *Compare cholesterol with the sex hormones shown in Concept 4.3 on p. 63. Circle the chemical groups that cholesterol has in common with estradiol; put a square around the chemical groups that cholesterol has in common with testosterone.*

CONCEPT CHECK 5.3

1. Compare the structure of a fat (triglyceride) with that of a phospholipid.
2. Why are human sex hormones considered lipids?
3. **WHAT IF?** Suppose a membrane surrounded an oil droplet, as it does in the cells of plant seeds. Describe and explain the form it might take.

For suggested answers, see Appendix A.

CONCEPT 5.4

Proteins include a diversity of structures, resulting in a wide range of functions

Nearly every dynamic function of a living being depends on proteins. In fact, the importance of proteins is underscored by their name, which comes from the Greek word *proteios*, meaning "first," or "primary." Proteins account for more than 50% of the dry mass of most cells, and they are instrumental in almost everything organisms do. Some proteins speed up chemical reactions, while others play a role in defense, storage, transport, cellular communication, movement, or structural support. **Figure 5.15**, on the next page, shows examples of proteins with these functions, which you'll learn more about in later chapters.

Life would not be possible without enzymes, most of which are proteins. Enzymatic proteins regulate metabolism by acting as **catalysts**, chemical agents that selectively speed up chemical reactions without being consumed by the reaction. Because an enzyme can perform its function over and over again, these molecules can be thought of as workhorses that keep cells running by carrying out the processes of life.

A human has tens of thousands of different proteins, each with a specific structure and function; proteins, in fact, are the most structurally sophisticated molecules known. Consistent with their diverse functions, they vary extensively in structure, each type of protein having a unique three-dimensional shape.

Polypeptides

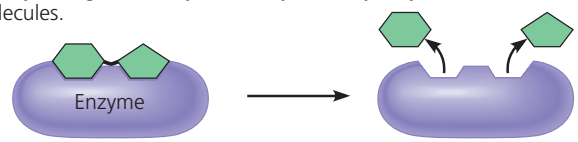
Diverse as proteins are, they are all unbranched polymers constructed from the same set of 20 amino acids. Polymers of amino acids are called **polypeptides**. A **protein** is a biologically functional molecule that consists of one or more polypeptides, each folded and coiled into a specific three-dimensional structure.

▼ Figure 5.15 An overview of protein functions.

Enzymatic proteins

Function: Selective acceleration of chemical reactions

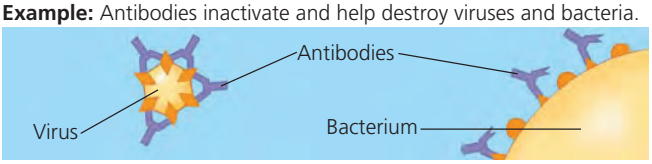
Example: Digestive enzymes catalyze the hydrolysis of bonds in food molecules.



Defensive proteins

Function: Protection against disease

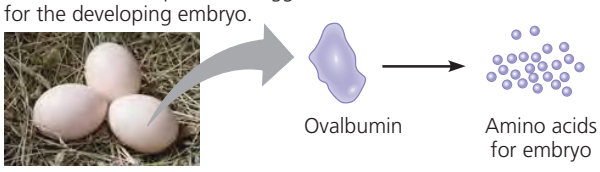
Example: Antibodies inactivate and help destroy viruses and bacteria.



Storage proteins

Function: Storage of amino acids

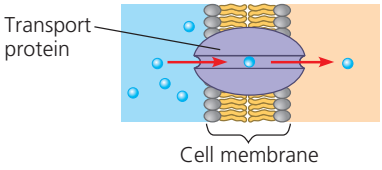
Examples: Casein, the protein of milk, is the major source of amino acids for baby mammals. Plants have storage proteins in their seeds. Ovalbumin is the protein of egg white, used as an amino acid source for the developing embryo.



Transport proteins

Function: Transport of substances

Examples: Hemoglobin, the iron-containing protein of vertebrate blood, transports oxygen from the lungs to other parts of the body. Other proteins transport molecules across cell membranes.



Hormonal proteins

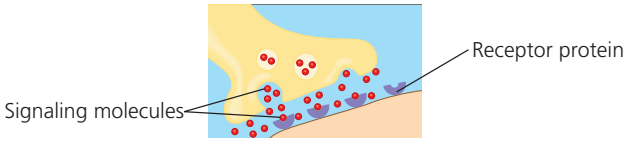
Function: Coordination of an organism's activities

Example: Insulin, a hormone secreted by the pancreas, causes other tissues to take up glucose, thus regulating blood sugar concentration.

Receptor proteins

Function: Response of cell to chemical stimuli

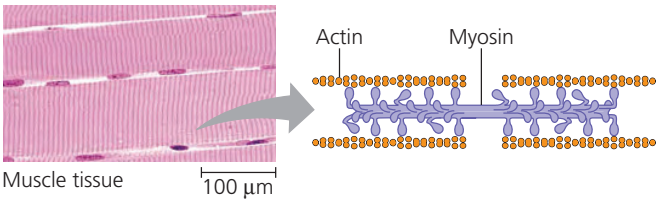
Example: Receptors built into the membrane of a nerve cell detect signaling molecules released by other nerve cells.



Contractile and motor proteins

Function: Movement

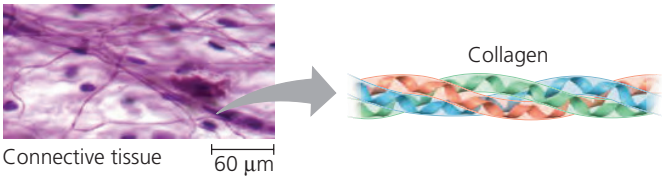
Examples: Motor proteins are responsible for the undulations of cilia and flagella. Actin and myosin proteins are responsible for the contraction of muscles.



Structural proteins

Function: Support

Examples: Keratin is the protein of hair, horns, feathers, and other skin appendages. Insects and spiders use silk fibers to make their cocoons and webs, respectively. Collagen and elastin proteins provide a fibrous framework in animal connective tissues.



Amino Acid Monomers

All amino acids share a common structure. An **amino acid** is an organic molecule possessing both an amino group and a carboxyl group (see Figure 4.9). The illustration at the right shows the general formula for an amino acid. At the center of the amino acid is an asymmetric carbon atom called the *alpha (α) carbon*. Its four different partners are an amino group, a carboxyl group, a hydrogen atom, and a variable group symbolized by R. The R group, also called the side chain, differs with each amino acid.

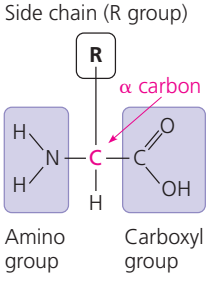
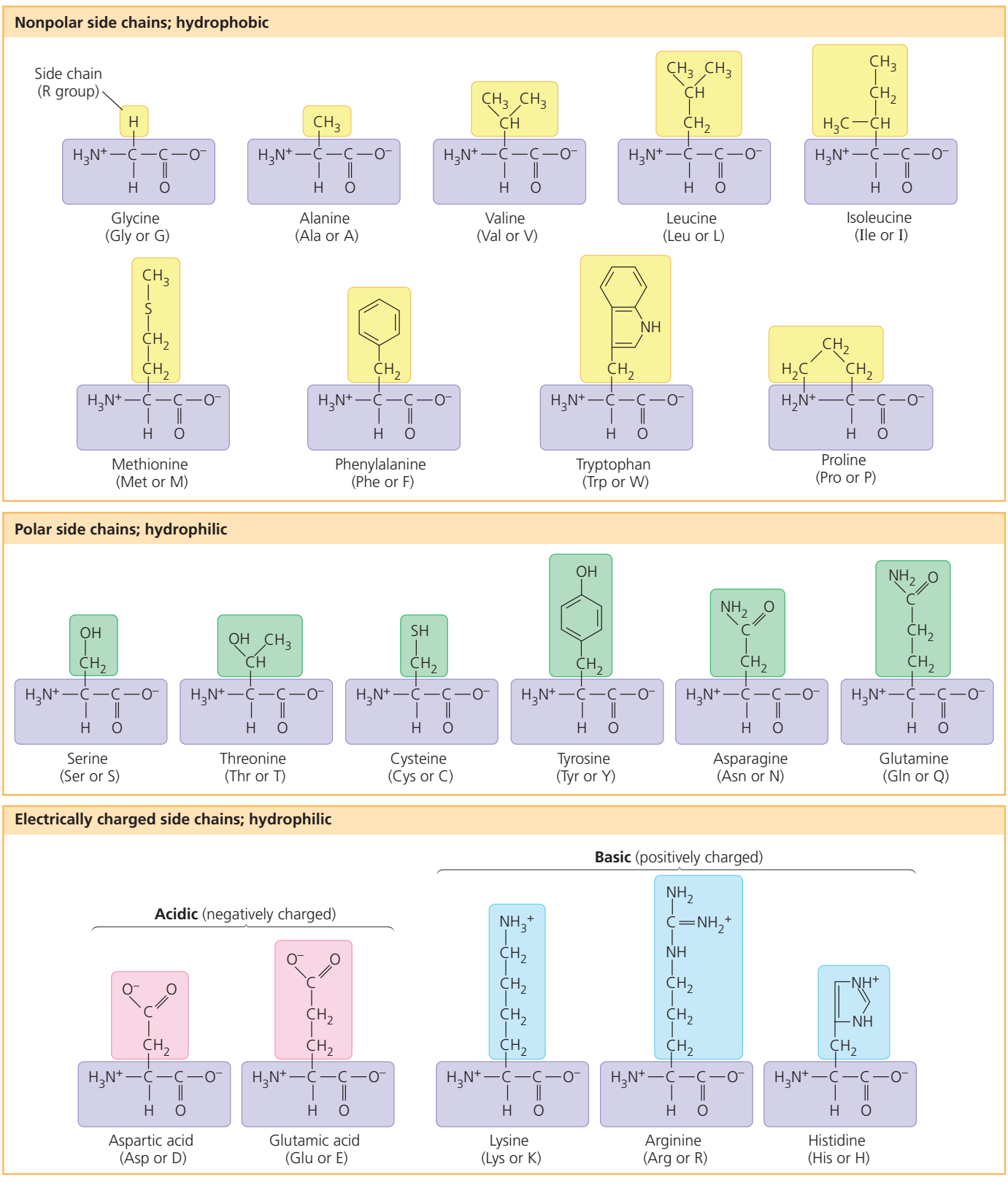


Figure 5.16 shows the 20 amino acids that cells use to build their thousands of proteins. Here the amino groups and carboxyl groups are all depicted in ionized form, the way they usually exist at the pH found in a cell. The side chain (R group) may be as simple as a hydrogen atom, as in the amino acid glycine, or it may be a carbon skeleton with various functional groups attached, as in glutamine.

▼ **Figure 5.16 The 20 amino acids of proteins.** The amino acids are grouped here according to the properties of their side chains (R groups) and shown in their prevailing ionic forms at pH 7.2, the pH within a cell. The three-letter and one-letter abbreviations for the amino acids are in parentheses. All amino acids used in proteins are L enantiomers, the form shown here (see Figure 4.7).

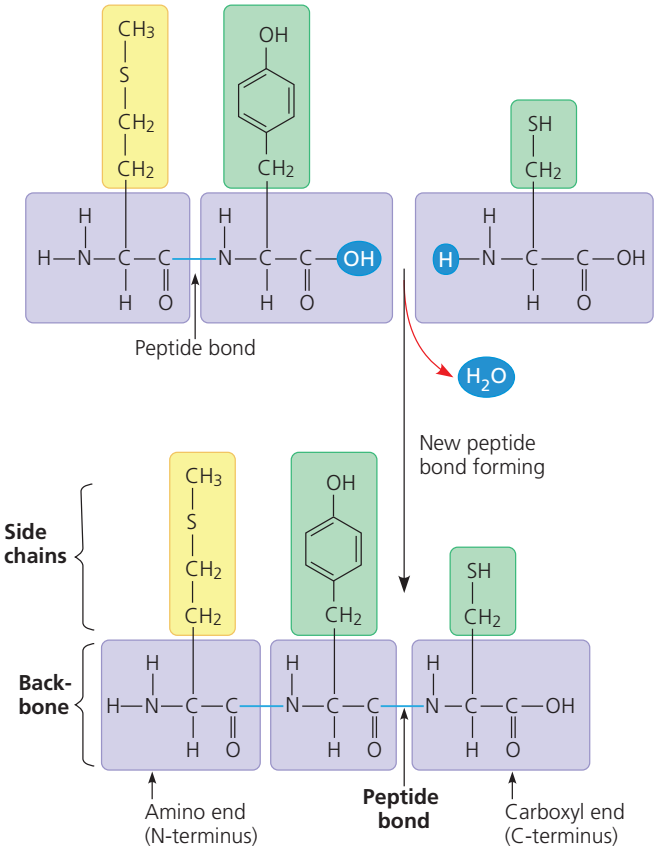


The physical and chemical properties of the side chain determine the unique characteristics of a particular amino acid, thus affecting its functional role in a polypeptide. In Figure 5.16, the amino acids are grouped according to the properties of their side chains. One group consists of amino acids with nonpolar side chains, which are hydrophobic. Another group consists of amino acids with polar side chains, which are hydrophilic. Acidic amino acids are those with side chains that are generally negative in charge owing to the presence of a carboxyl group, which is usually dissociated (ionized) at cellular pH. Basic amino acids have amino groups in their side chains that are generally positive in charge. (Notice that *all* amino acids have carboxyl groups and amino groups; the terms *acidic* and *basic* in this context refer only to groups on the side chains.) Because they are charged, acidic and basic side chains are also hydrophilic.

Amino Acid Polymers

Now that we have examined amino acids, let's see how they are linked to form polymers **(Figure 5.17)**. When two amino acids are positioned so that the carboxyl group of one is adjacent to the amino group of the other, they can become joined by a dehydration reaction, with the removal of a water molecule. The resulting covalent bond is called a **peptide bond**. Repeated over and over, this process yields a polypeptide, a polymer of many amino acids linked by peptide bonds.

The repeating sequence of atoms highlighted in purple in Figure 5.17 is called the polypeptide backbone. Extending from this backbone are the different side chains (R groups) of the amino acids. Polypeptides range in length from a few amino acids to a thousand or more. Each specific polypeptide has a unique linear sequence of amino acids. Note that one end of the polypeptide chain has a free amino group, while the opposite end has a free carboxyl group. Thus, a polypeptide of any length has a single amino end (N-terminus) and a single carboxyl end (C-terminus). In a polypeptide of any significant size, the side chains far outnumber the terminal groups, so the chemical nature of the molecule as a whole is determined by the kind and sequence of the side chains. The immense variety of polypeptides in nature illustrates an important concept introduced earlier—that cells can make many different polymers by linking a limited set of monomers into diverse sequences.



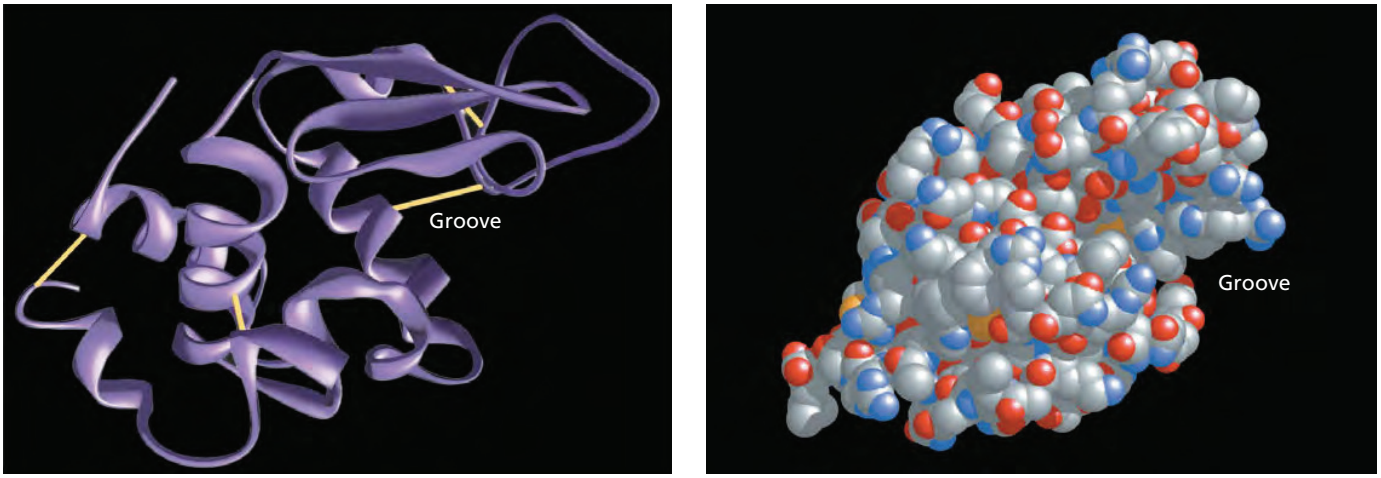
▲ **Figure 5.17 Making a polypeptide chain.** Peptide bonds are formed by dehydration reactions, which link the carboxyl group of one amino acid to the amino group of the next. The peptide bonds are formed one at a time, starting with the amino acid at the amino end (N-terminus). The polypeptide has a repetitive backbone (purple) to which the amino acid side chains (yellow and green) are attached.

DRAW IT *Circle and label the carboxyl and amino groups that will form the new peptide bond.*

Protein Structure and Function

The specific activities of proteins result from their intricate three-dimensional architecture, the simplest level of which is the sequence of their amino acids. The pioneer in determining the amino acid sequence of proteins was Frederick Sanger, who, with his colleagues at Cambridge University in England, worked on the hormone insulin in the late 1940s and early 1950s. He used agents that break polypeptides at specific places, followed by chemical methods to determine the amino acid sequence in these small fragments. Sanger and his co-workers were able, after years of effort, to reconstruct the complete amino acid sequence of insulin. Since then, most of the steps involved in sequencing a polypeptide have been automated.

Once we have learned the amino acid sequence of a polypeptide, what can it tell us about the three-dimensional structure (commonly referred to simply as the "structure") of the protein and its function? The term *polypeptide* is not synonymous with the term *protein*. Even for a protein consisting of a single polypeptide, the relationship is somewhat analogous to that between a long strand of yarn and a sweater of particular size and shape that can be knit from the yarn. A functional protein is not *just* a polypeptide chain, but one or more polypeptides precisely twisted, folded, and coiled into a molecule of unique shape **(Figure 5.18)**. And it is the amino acid sequence of each polypeptide that determines what



(a) A **ribbon model** shows how the single polypeptide chain folds and coils to form the functional protein. (The yellow lines represent disulfide bridges that stabilize the protein's shape; see Figure 5.20.)

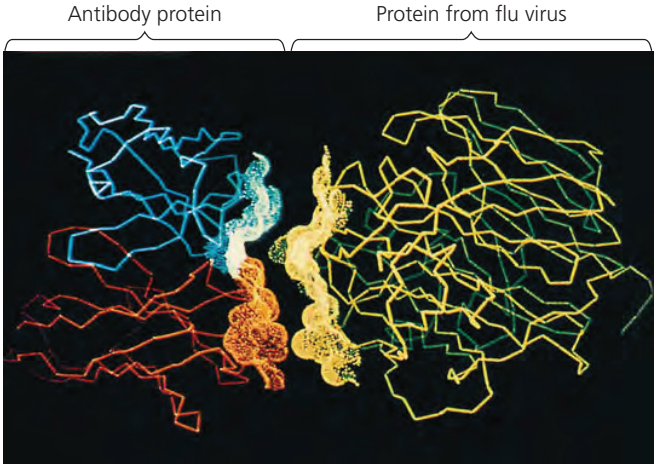
(b) A **space-filling model** shows more clearly the globular shape seen in many proteins, as well as the specific three-dimensional structure unique to lysozyme.

▲ **Figure 5.18 Structure of a protein, the enzyme lysozyme.** Present in our sweat, tears, and saliva, lysozyme is an enzyme that helps prevent infection by binding to and destroying specific molecules on the surface of many kinds of bacteria. The groove is the part of the protein that recognizes and binds to the target molecules on bacterial walls.

three-dimensional structure the protein will have under normal cellular conditions.

When a cell synthesizes a polypeptide, the chain generally folds spontaneously, assuming the functional structure for that protein. This folding is driven and reinforced by the formation of a variety of bonds between parts of the chain, which in turn depends on the sequence of amino acids. Many proteins are roughly spherical (*globular proteins*), while others are shaped like long fibers (*fibrous proteins*). Even within these broad categories, countless variations exist.

A protein's specific structure determines how it works. In almost every case, the function of a protein depends on its ability to recognize and bind to some other molecule. In an especially striking example of the marriage of form and function, **Figure 5.19** shows the exact match of shape between an antibody (a protein in the body) and the particular foreign substance on a flu virus that the antibody binds to and marks for destruction. In Chapter 43, you'll learn more about how the immune system generates antibodies that match the shapes of specific foreign molecules so well. Also, you may recall from Chapter 2 that natural signaling molecules called endorphins bind to specific receptor proteins on the surface of brain cells in humans, producing euphoria and relieving pain. Morphine, heroin, and other opiate drugs are able to mimic endorphins because they all share a similar shape with endorphins and can thus fit into and bind to endorphin receptors in the brain. This fit is very specific, something like a lock and key (see Figure 2.18). Thus, the function of a protein—for instance, the ability of a receptor protein to bind to a particular pain-relieving signaling molecule—is an emergent property resulting from exquisite molecular order.



▲ **Figure 5.19 An antibody binding to a protein from a flu virus.** A technique called X-ray crystallography was used to generate a computer model of an antibody protein (blue and orange, left) bound to a flu virus protein (green and yellow, right). Computer software was then used to back the images away from each other, revealing the exact complementarity of shape between the two protein surfaces.

Four Levels of Protein Structure

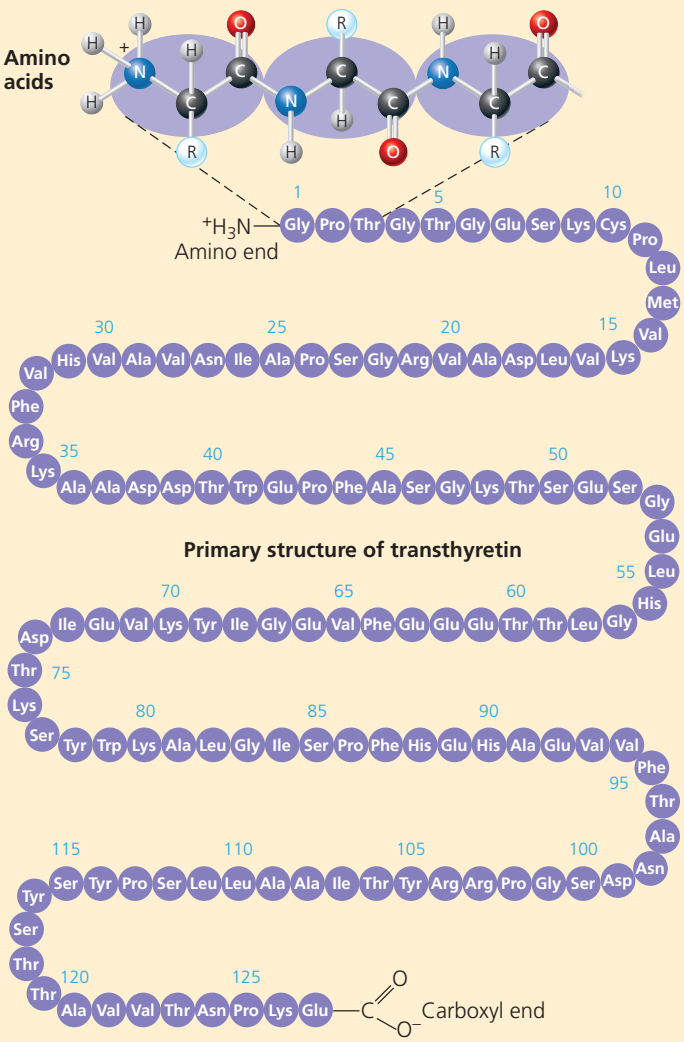
With the goal of understanding the function of a protein, learning about its structure is often productive. In spite of their great diversity, all proteins share three superimposed levels of structure, known as primary, secondary, and tertiary structure. A fourth level, quaternary structure, arises when a protein consists of two or more polypeptide chains. **Figure 5.20**, on the following two pages, describes these four levels of protein structure. Be sure to study this figure thoroughly before going on to the next section.

▼ Figure 5.20

Exploring Levels of Protein Structure

Primary Structure

Linear chain of amino acids



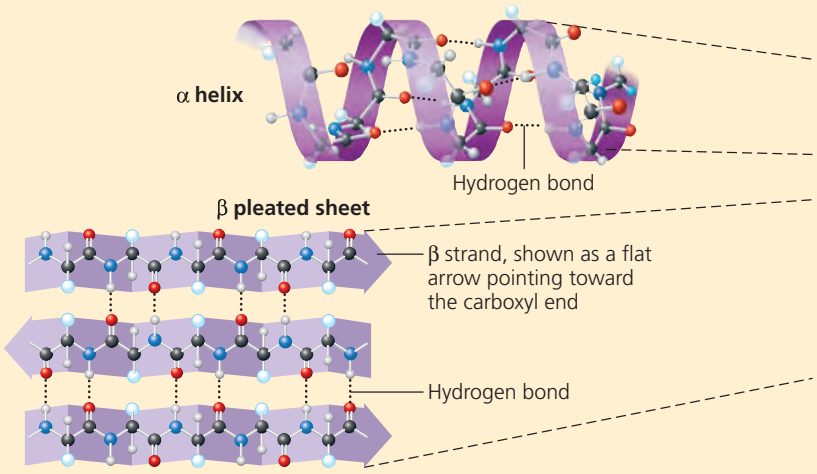
Primary structure of transthyretin

The **primary structure** of a protein is a linked series of amino acids with a unique sequence. As an example, let's consider transthyretin, a globular blood protein that transports vitamin A and one of the thyroid hormones throughout the body. Transthyretin is made up of four identical polypeptide chains, each composed of 127 amino acids. Shown here is one of these chains unraveled for a closer look at its primary structure. Each of the 127 positions along the chain is occupied by one of the 20 amino acids, indicated here by its three-letter abbreviation.

The primary structure is like the order of letters in a very long word. If left to chance, there would be 20^{127} different ways of making a polypeptide chain 127 amino acids long. However, the precise primary structure of a protein is determined not by the random linking of amino acids, but by inherited genetic information. The primary structure in turn dictates secondary and tertiary structure, due to the chemical nature of the backbone and the side chains (R groups) of the amino acids positioned along the chain.

Secondary Structure

Regions stabilized by hydrogen bonds between atoms of the polypeptide backbone



Most proteins have segments of their polypeptide chains repeatedly coiled or folded in patterns that contribute to the protein's overall shape. These coils and folds, collectively referred to as **secondary structure**, are the result of hydrogen bonds between the repeating constituents of the polypeptide backbone (not the amino acid side chains). Within the backbone, the oxygen atoms have a partial negative charge, and the hydrogen atoms attached to the nitrogens have a partial positive charge (see Figure 2.16); therefore, hydrogen bonds can form between these atoms. Individually, these hydrogen bonds are weak, but because they are repeated many times over a relatively long region of the polypeptide chain, they can support a particular shape for that part of the protein.

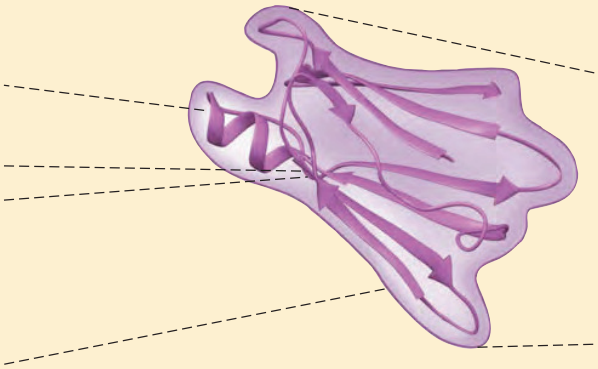
One such secondary structure is the **α helix**, a delicate coil held together by hydrogen bonding between every fourth amino acid, shown above. Although each transthyretin polypeptide has only one α helix region (see tertiary structure on the next page), other globular proteins have multiple stretches of α helix separated by nonhelical regions (see hemoglobin on the next page). Some fibrous proteins, such as α-keratin, the structural protein of hair, have the α helix formation over most of their length.

The other main type of secondary structure is the **β pleated sheet**. As shown above, in this structure two or more strands of the polypeptide chain lying side by side (called β strands) are connected by hydrogen bonds between parts of the two parallel polypeptide backbones. β pleated sheets make up the core of many globular proteins, as is the case for transthyretin (see tertiary structure on the next page), and dominate some fibrous proteins, including the silk protein of a spider's web. The teamwork of so many hydrogen bonds makes each spider silk fiber stronger than a steel strand of the same weight.

▼ Spiders secrete silk fibers made of a structural protein containing β pleated sheets, which allow the spider web to stretch and recoil.

Tertiary Structure

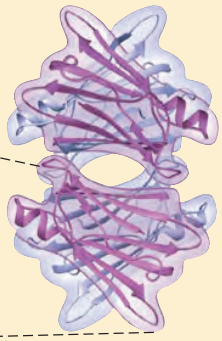
Three-dimensional shape stabilized by interactions between side chains



Transthyretin polypeptide

Quaternary Structure

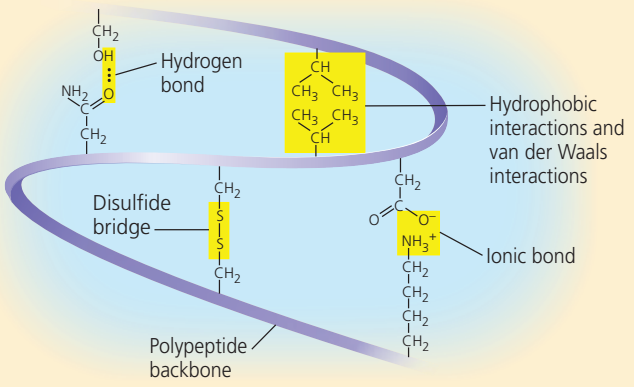
Association of multiple polypeptides, forming a functional protein



Transthyretin protein (four identical polypeptides)

Superimposed on the patterns of secondary structure is a protein's tertiary structure, shown above in a ribbon model of the transthyretin polypeptide. While secondary structure involves interactions between backbone constituents, **tertiary structure** is the overall shape of a polypeptide resulting from interactions between the side chains (R groups) of the various amino acids. One type of interaction that contributes to tertiary structure is—somewhat misleadingly—called a **hydrophobic interaction**. As a polypeptide folds into its functional shape, amino acids with hydrophobic (nonpolar) side chains usually end up in clusters at the core of the protein, out of contact with water. Thus, a "hydrophobic interaction" is actually caused by the exclusion of nonpolar substances by water molecules. Once nonpolar amino acid side chains are close together, van der Waals interactions help hold them together. Meanwhile, hydrogen bonds between polar side chains and ionic bonds between positively and negatively charged side chains also help stabilize tertiary structure. These are all weak interactions in the aqueous cellular environment, but their cumulative effect helps give the protein a unique shape.

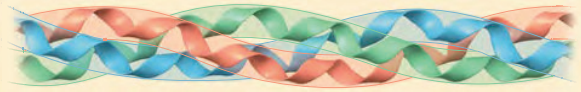
Covalent bonds called **disulfide bridges** may further reinforce the shape of a protein. Disulfide bridges form where two cysteine monomers, which have sulfhydryl groups (—SH) on their side chains (see Figure 4.9), are brought close together by the folding of the protein. The sulfur of one cysteine bonds to the sulfur of the second, and the disulfide bridge (—S—S—) rivets parts of the protein together (see yellow lines in Figure 5.18a). All of these different kinds of interactions can contribute to the tertiary structure of a protein, as shown here in a small part of a hypothetical protein:



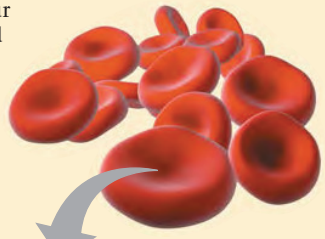
Some proteins consist of two or more polypeptide chains aggregated into one functional macromolecule. **Quaternary structure** is the overall protein structure that results from the aggregation of these polypeptide subunits. For example, shown above is the complete globular transthyretin protein, made up of its four polypeptides.

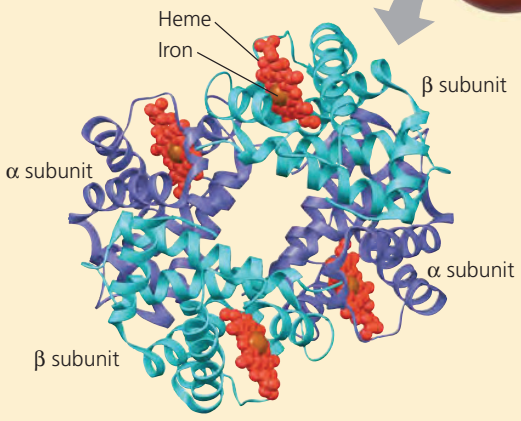
Another example is collagen, shown below, which is a fibrous protein that has three identical helical polypeptides intertwined into a larger triple helix, giving the long fibers great strength. This suits collagen fibers to their function as the girders of connective tissue in skin, bone, tendons, ligaments, and other body parts. (Collagen accounts for 40% of the protein in a human body.)

Collagen

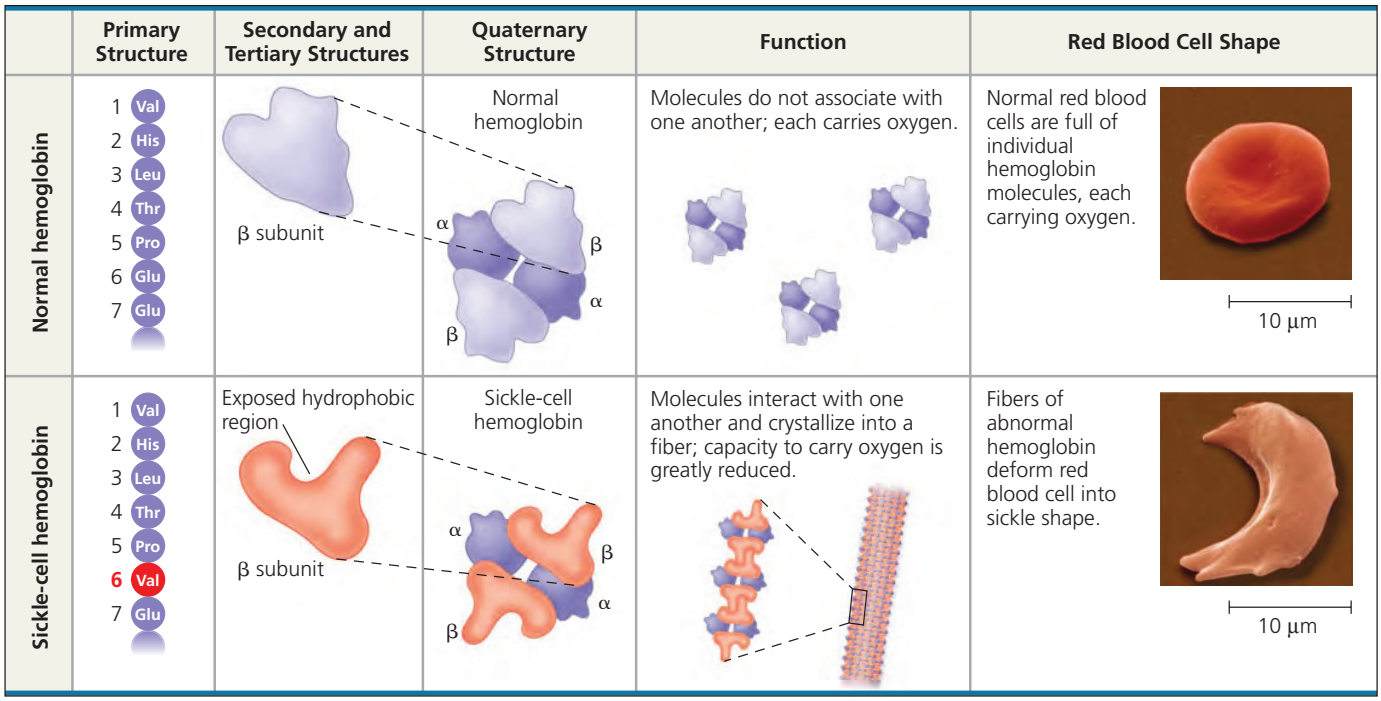


Hemoglobin, the oxygen-binding protein of red blood cells shown below, is another example of a globular protein with quaternary structure. It consists of four polypeptide subunits, two of one kind (α) and two of another kind (β). Both α and β subunits consist primarily of α-helical secondary structure. Each subunit has a nonpolypeptide component, called heme, with an iron atom that binds oxygen.





Hemoglobin



▲ Figure 5.21 A single amino acid substitution in a protein causes sickle-cell disease.

MAKE CONNECTIONS *Considering the chemical characteristics of the amino acids valine and glutamic acid (see Figure 5.16), propose a possible explanation for the dramatic effect on protein function that occurs when valine is substituted for glutamic acid.*

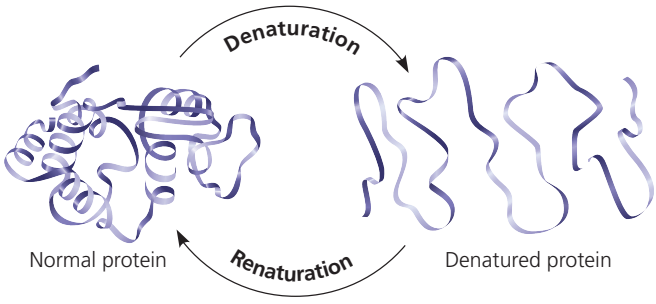
Sickle-Cell Disease: A Change in Primary Structure

Even a slight change in primary structure can affect a protein's shape and ability to function. For instance, **sickle-cell disease**, an inherited blood disorder, is caused by the substitution of one amino acid (valine) for the normal one (glutamic acid) at a particular position in the primary structure of hemoglobin, the protein that carries oxygen in red blood cells. Normal red blood cells are disk-shaped, but in sickle-cell disease, the abnormal hemoglobin molecules tend to crystallize, deforming some of the cells into a sickle shape **(Figure 5.21)**. A person with the disease has periodic "sickle-cell crises" when the angular cells clog tiny blood vessels, impeding blood flow. The toll taken on such patients is a dramatic example of how a simple change in protein structure can have devastating effects on protein function.

What Determines Protein Structure?

You've learned that a unique shape endows each protein with a specific function. But what are the key factors determining protein structure? You already know most of the answer: A polypeptide chain of a given amino acid sequence can spontaneously arrange itself into a three-dimensional shape determined and maintained by the interactions responsible for secondary and tertiary structure. This folding normally occurs as the protein is being synthesized in the crowded environment within a cell, aided by other proteins. However, protein structure also depends on the physical and chemical conditions of the protein's environment. If the pH, salt concentration, temperature, or other aspects of its environment are altered, the weak chemical bonds and interactions within a protein may be destroyed, causing the protein to unravel and lose its native shape, a change called **denaturation** **(Figure 5.22)**. Because it is misshapen, the denatured protein is biologically inactive.

Most proteins become denatured if they are transferred from an aqueous environment to a nonpolar solvent, such as



▲ Figure 5.22 Denaturation and renaturation of a protein. High temperatures or various chemical treatments will denature a protein, causing it to lose its shape and hence its ability to function. If the denatured protein remains dissolved, it can often renature when the chemical and physical aspects of its environment are restored to normal.

ether or chloroform; the polypeptide chain refolds so that its hydrophobic regions face outward toward the solvent. Other denaturation agents include chemicals that disrupt the hydrogen bonds, ionic bonds, and disulfide bridges that maintain a protein's shape. Denaturation can also result from excessive heat, which agitates the polypeptide chain enough to overpower the weak interactions that stabilize the structure. The white of an egg becomes opaque during cooking because the denatured proteins are insoluble and solidify. This also explains why excessively high fevers can be fatal: Proteins in the blood can denature at very high body temperatures.

When a protein in a test-tube solution has been denatured by heat or chemicals, it can sometimes return to its functional shape when the denaturing agent is removed. We can conclude that the information for building specific shape is intrinsic to the protein's primary structure. The sequence of amino acids determines the protein's shape—where an α helix can form, where β pleated sheets can exist, where disulfide bridges are located, where ionic bonds can form, and so on. But how does protein folding occur in the cell?

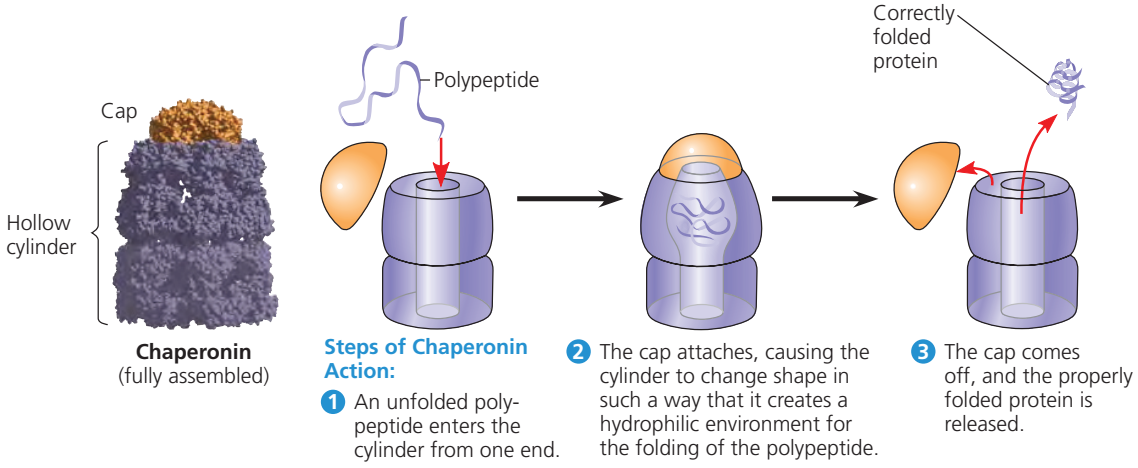
Protein Folding in the Cell

Biochemists now know the amino acid sequence for more than 10 million proteins and the three-dimensional shape for more than 20,000. Researchers have tried to correlate the primary structure of many proteins with their three-dimensional structure to discover the rules of protein folding. Unfortunately, however, the protein-folding process is not that simple. Most proteins probably go through several intermediate structures on their way to a stable shape, and looking at the mature structure does not reveal the stages of folding required to achieve that form. However, biochemists have developed methods for tracking a protein through such stages.

Crucial to the folding process are **chaperonins** (also called chaperone proteins), protein molecules that assist in the proper folding of other proteins **(Figure 5.23)**. Chaperonins do not specify the final structure of a polypeptide. Instead, they keep the new polypeptide segregated from "bad influences" in the cytoplasmic environment while it folds spontaneously. The chaperonin shown in Figure 5.23, from the bacterium *E. coli*, is a giant multiprotein complex shaped like a hollow cylinder. The cavity provides a shelter for folding polypeptides. In the past decade, researchers have discovered molecular systems that interact with chaperonins and check whether proper folding has occurred. Such systems either refold the misfolded proteins correctly or mark them for destruction.

Misfolding of polypeptides is a serious problem in cells. Many diseases, such as Alzheimer's, Parkinson's, and mad cow disease, are associated with an accumulation of misfolded proteins. In fact, misfolded versions of the transthyretin protein featured in Figure 5.20 have been implicated in several diseases, including one form of senile dementia.

Even when scientists have a correctly folded protein in hand, determining its exact three-dimensional structure is not simple, for a single protein molecule has thousands of atoms. The first 3-D structures were worked out in 1959 for hemoglobin and a related protein. The method that made these feats possible was **X-ray crystallography**, which has since been used to determine the 3-D structure of many other proteins. In a recent example, Roger Kornberg and his colleagues at Stanford University used this method to elucidate the structure of RNA polymerase, an enzyme that plays a crucial role in the expression of genes (**Figure 5.24**, on the next page). Another method for analyzing protein structure is nuclear magnetic resonance (NMR) spectroscopy, which does not require protein crystallization. A still newer approach employs bioinformatics (see Chapter 1) to predict the 3-D structure of polypeptides from their amino acid sequence. X-ray crystallography, NMR spectroscopy, and bioinformatics are complementary approaches to understanding protein structure and function.

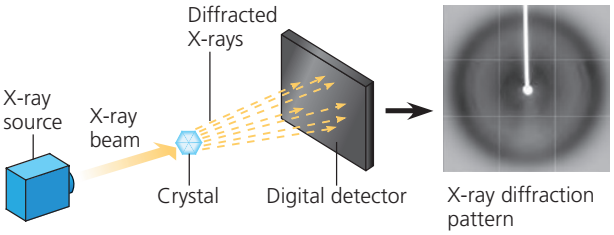


▶ Figure 5.23 A chaperonin in action. The computer graphic (left) shows a large chaperonin protein complex from the bacterium *E. coli*. It has an interior space that provides a shelter for the proper folding of newly made polypeptides. The complex consists of two proteins: One protein is a hollow cylinder; the other is a cap that can fit on either end.

▼ **Figure 5.24** **INQUIRY**

What can the 3-D shape of the enzyme RNA polymerase II tell us about its function?

EXPERIMENT In 2006, Roger Kornberg was awarded the Nobel Prize in Chemistry for using X-ray crystallography to determine the 3-D shape of RNA polymerase II, which binds to the DNA double helix and synthesizes RNA. After crystallizing a complex of all three components, Kornberg and his colleagues aimed an X-ray beam through the crystal. The atoms of the crystal diffracted (bent) the X-rays into an orderly array that a digital detector recorded as a pattern of spots called an X-ray diffraction pattern.



RESULTS Using data from X-ray diffraction patterns, as well as the amino acid sequence determined by chemical methods, Kornberg and colleagues built a 3-D model of the complex with the help of computer software.

CONCLUSION By analyzing their model, the researchers developed a hypothesis about the functions of different regions of RNA polymerase II. For example, the region above the DNA may act as a clamp that holds the nucleic acids in place. (You'll learn more about this enzyme in Chapter 17.)

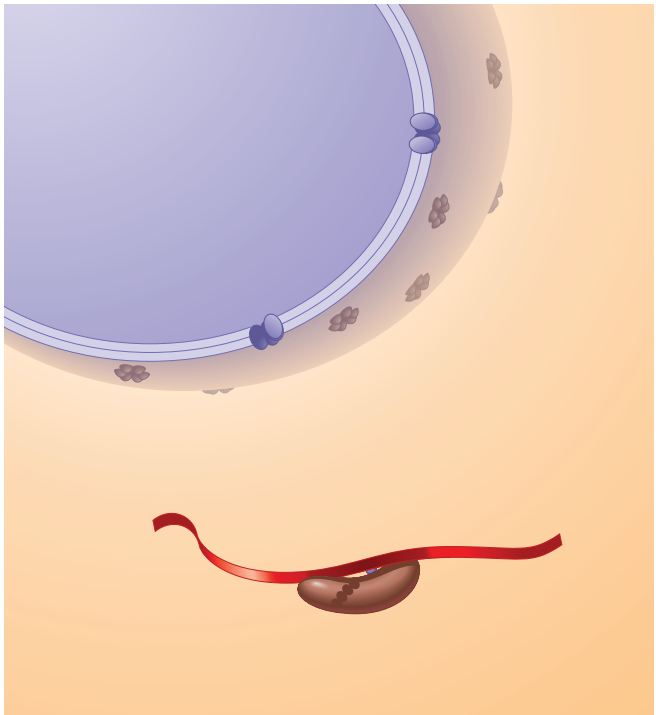
SOURCE A. L. Gnatt et al., Structural basis of transcription: an RNA polymerase II elongation complex at 3.3Å, *Science* 292:1876–1882 (2001).

WHAT IF? If you were an author of the paper and were describing the model, what type of protein structure would you call the small polypeptide spirals in RNA polymerase II?

CONCEPT CHECK 5.4

1. Why does a denatured protein no longer function normally?
2. What parts of a polypeptide participate in the bonds that hold together secondary structure? Tertiary structure?
3. **WHAT IF?** Where would you expect a polypeptide region that is rich in the amino acids valine, leucine, and isoleucine to be located in the folded polypeptide? Explain.

For suggested answers, see Appendix A.



▲ **Figure 5.25** **DNA → RNA → protein.** In a eukaryotic cell, DNA in the nucleus programs protein production in the cytoplasm by dictating synthesis of messenger RNA (mRNA). (The cell nucleus is actually much larger relative to the other elements of this figure.)

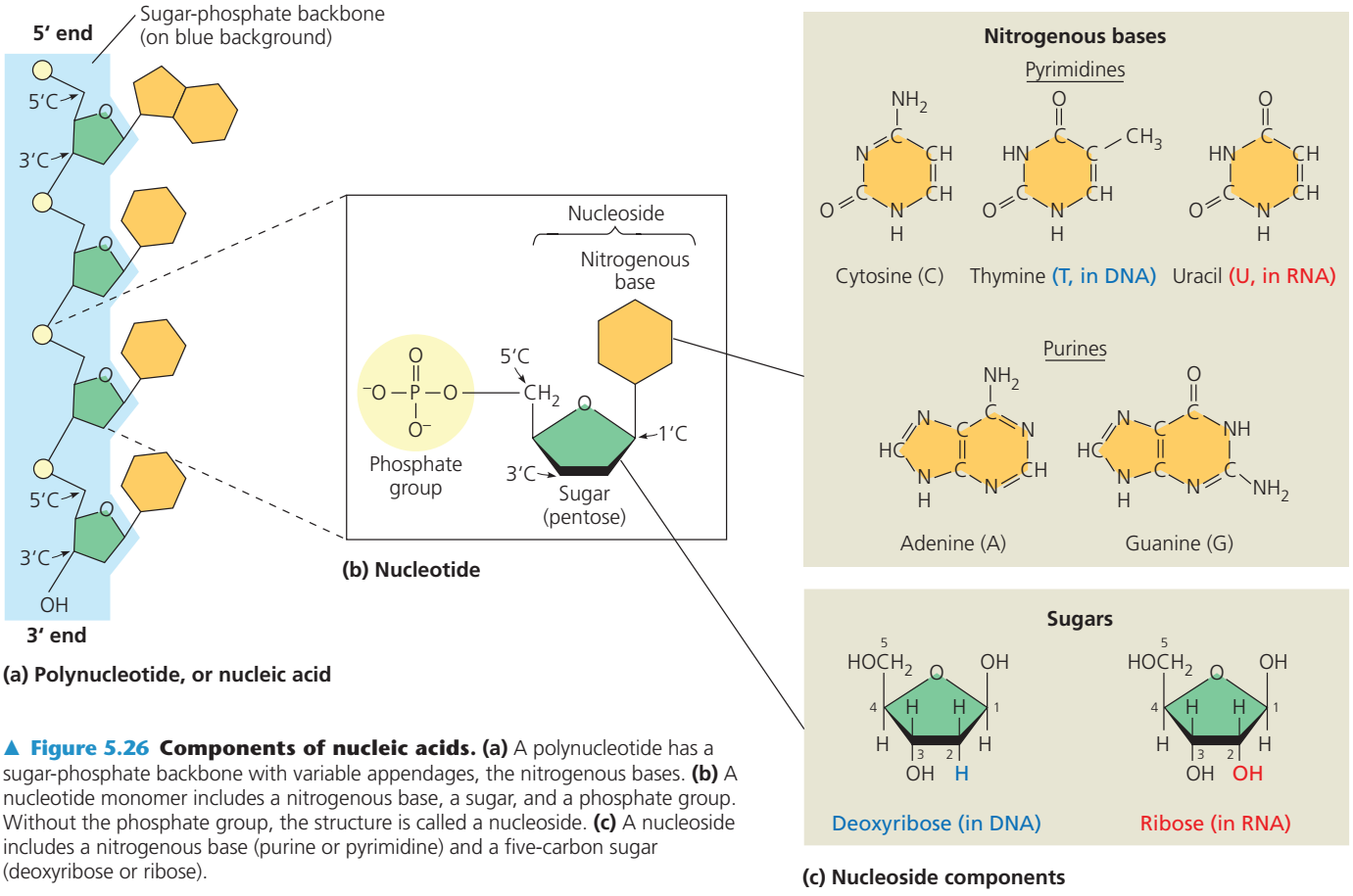
programs all the cell's activities. The DNA, however, is not directly involved in running the operations of the cell, any more than computer software by itself can print a bank statement or read the bar code on a box of cereal. Just as a printer is needed to print out a statement and a scanner is needed to read a bar code, proteins are required to implement genetic programs. The molecular hardware of the cell—the tools for biological functions—consists mostly of proteins. For example, the oxygen carrier in red blood cells is the protein hemoglobin, not the DNA that specifies its structure.

How does RNA, the other type of nucleic acid, fit into gene expression, the flow of genetic information from DNA to proteins? Each gene along a DNA molecule directs synthesis of a type of RNA called *messenger RNA* (*mRNA*). The mRNA molecule interacts with the cell's protein-synthesizing machinery to direct production of a polypeptide, which folds into all or part of a protein. We can summarize the flow of genetic information as DNA → RNA → protein (see Figure 5.25). The sites of protein synthesis are tiny structures called ribosomes. In a eukaryotic cell, ribosomes are in the cytoplasm, but DNA resides in the nucleus. Messenger RNA conveys genetic instructions for building proteins from the nucleus to the cytoplasm. Prokaryotic cells lack nuclei but still use mRNA to convey a message from the DNA to ribosomes and other cellular equipment that translate the coded information into amino acid sequences. In recent years, the spotlight has been turned on other, previously unknown types of RNA that play many other roles in the cell. As is so often true in biology, the story is still being written! You'll hear more about the newly discovered functions of RNA molecules in Chapter 18.

The Components of Nucleic Acids

Nucleic acids are macromolecules that exist as polymers called **polynucleotides** (Figure 5.26a). As indicated by the name, each polynucleotide consists of monomers called **nucleotides**. A nucleotide, in general, is composed of three parts: a nitrogen-containing (nitrogenous) base, a five-carbon sugar (a pentose), and one or more phosphate groups (Figure 5.26b). In a polynucleotide, each monomer has only one phosphate group. The portion of a nucleotide without any phosphate groups is called a *nucleoside*.

To build a nucleotide, let's first consider the nitrogenous bases (Figure 5.26c). Each nitrogenous base has one or two rings that include nitrogen atoms. (They are called nitrogenous *bases* because the nitrogen atoms tend to take up H^+



▲ **Figure 5.26 Components of nucleic acids. (a)** A polynucleotide has a sugar-phosphate backbone with variable appendages, the nitrogenous bases. **(b)** A nucleotide monomer includes a nitrogenous base, a sugar, and a phosphate group. Without the phosphate group, the structure is called a nucleoside. **(c)** A nucleoside includes a nitrogenous base (purine or pyrimidine) and a five-carbon sugar (deoxyribose or ribose).

from solution, thus acting as bases.) There are two families of nitrogenous bases: pyrimidines and purines. A **pyrimidine** has one six-membered ring of carbon and nitrogen atoms. The members of the pyrimidine family are cytosine (C), thymine (T), and uracil (U). **Purines** are larger, with a six-membered ring fused to a five-membered ring. The purines are adenine (A) and guanine (G). The specific pyrimidines and purines differ in the chemical groups attached to the rings. Adenine, guanine, and cytosine are found in both DNA and RNA; thymine is found only in DNA and uracil only in RNA.

Now let's add a sugar to the nitrogenous base. In DNA the sugar is **deoxyribose**; in RNA it is **ribose** (see Figure 5.26c). The only difference between these two sugars is that deoxyribose lacks an oxygen atom on the second carbon in the ring; hence the name *deoxy*ribose. To distinguish the numbers of the sugar carbons from those used for the ring atoms of the attached nitrogenous base, the sugar carbon numbers of a nucleoside or nucleotide have a prime (′) after them. Thus, the second carbon in the sugar ring is the 2′ ("2 prime") carbon, and the carbon that sticks up from the ring is called the 5′ carbon.

So far, we have built a nucleoside (nitrogenous base plus sugar). To complete the construction of a nucleotide, we attach a phosphate group to the 5′ carbon of the sugar (see Figure 5.26b). The molecule is now a nucleoside monophosphate, better known as a nucleotide.

Nucleotide Polymers

Now we can see how these nucleotides are linked together to build a polynucleotide. Adjacent nucleotides are joined by a phosphodiester linkage, which consists of a phosphate group that links the sugars of two nucleotides. This bonding results in a backbone with a repeating pattern of sugar-phosphate units (see Figure 5.26a). (Note that the nitrogenous bases are not part of the backbone.) The two free ends of the polymer are distinctly different from each other. One end has a phosphate attached to a 5′ carbon, and the other end has a hydroxyl group on a 3′ carbon; we refer to these as the 5′ end and the 3′ end, respectively. We can say that a polynucleotide has a built-in directionality along its sugar-phosphate backbone, from 5′ to 3′, somewhat like a one-way street. All along this sugar-phosphate backbone are appendages consisting of the nitrogenous bases.

The sequence of bases along a DNA (or mRNA) polymer is unique for each gene and provides very specific information to the cell. Because genes are hundreds to thousands of nucleotides long, the number of possible base sequences is effectively limitless. A gene's meaning to the cell is encoded in its specific sequence of the four DNA bases. For example, the sequence 5′-AGGTAACTT-3′ means one thing, whereas the sequence 5′-CGCTTTAAC-3′ has a different meaning. (Entire genes, of course, are much longer.) The linear order of bases in a gene specifies the amino acid sequence—the primary structure—of a protein, which in turn specifies that protein's three-dimensional structure and its function in the cell.

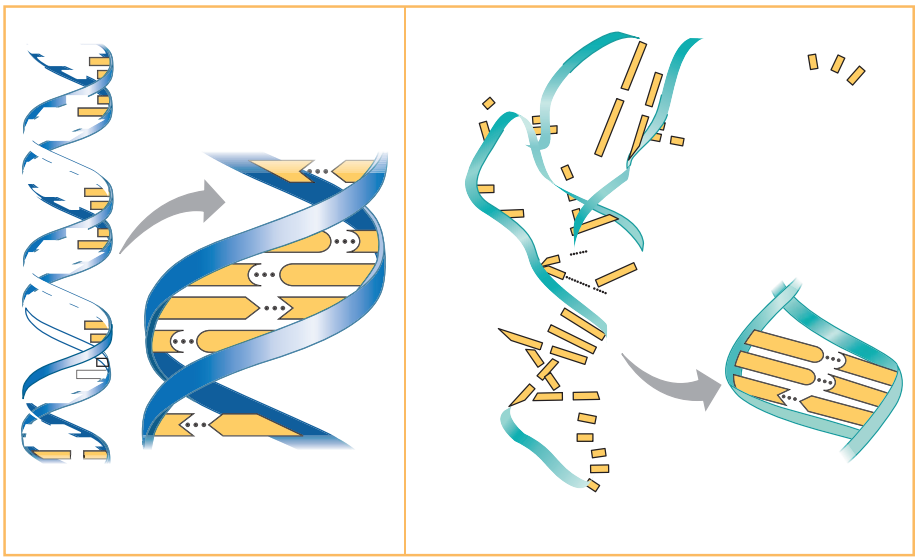
The Structures of DNA and RNA Molecules

RNA molecules usually exist as single polynucleotide chains like the one shown in Figure 5.26a. In contrast, DNA molecules have two polynucleotides, or "strands," that spiral around an imaginary axis, forming a **double helix** **(Figure 5.27a)**. The two sugar-phosphate backbones run in opposite 5′ → 3′ directions from each other; this arrangement is referred to as **antiparallel**, somewhat like a divided highway. The sugar-phosphate backbones are on the outside of the helix, and the nitrogenous bases are paired in the interior of the helix. The two strands are held together by hydrogen bonds between the paired bases (see Figure 5.27a). Most DNA molecules are very long, with thousands or even millions of base pairs. One long DNA double helix includes many genes, each one a particular segment of the molecule.

Only certain bases in the double helix are compatible with each other. Adenine (A) always pairs with thymine (T), and guanine (G) always pairs with cytosine (C). If we were to read the sequence of bases along one strand of the double helix, we would know the sequence of bases along the other strand. If a stretch of one strand has the base sequence 5′-AGGTCCG-3′, then the base-pairing rules tell us that the same stretch of the other strand must have the sequence 3′-TCCAGGC-5′. The two strands of the double helix are *complementary*, each the predictable counterpart of the other. It is this feature of DNA that makes it possible to generate two identical copies of each DNA molecule in a cell that is preparing to divide. When the cell divides, the copies are distributed to the daughter cells, making them genetically identical to the parent cell. Thus, the structure of DNA accounts for its function of transmitting genetic information whenever a cell reproduces.

Complementary base pairing can also occur between parts of two RNA molecules or even between two stretches of nucleotides in the *same* RNA molecule. In fact, base pairing within an RNA molecule allows it to take on the particular three-dimensional shape necessary for its function. Consider, for example, the type of RNA called *transfer RNA* (*tRNA*), which brings amino acids to the ribosome during the synthesis of a polypeptide. A tRNA molecule is about 80 nucleotides in length. Its functional shape results from base pairing between nucleotides where complementary stretches of the molecule run antiparallel to each other **(Figure 5.27b)**.

Note that in RNA, adenine (A) pairs with uracil (U); thymine (T) is not present in RNA. Another difference between RNA and DNA is that DNA almost always exists as a double helix, whereas RNA molecules are more variable in shape. This variability arises because the extent and location of complementary base pairing within an RNA molecule differs in different types of RNA, as you will see in Chapter 17.



▶ **Figure 5.27 The structures of DNA and tRNA molecules. (a)** The DNA molecule is usually a double helix, with the sugar-phosphate backbones of the antiparallel polynucleotide strands (symbolized here by blue ribbons) on the outside of the helix. Holding the two strands together are pairs of nitrogenous bases attached to each other by hydrogen bonds. As illustrated here with symbolic shapes for the bases, adenine (A) can pair only with thymine (T), and guanine (G) can pair only with cytosine (C). Each DNA strand in this figure is the structural equivalent of the polynucleotide diagrammed in Figure 5.26a. **(b)** A tRNA molecule has a roughly L-shaped structure, with complementary base pairing of antiparallel stretches of RNA. In RNA, A pairs with U.

DNA and Proteins as Tape Measures of Evolution

EVOLUTION We are accustomed to thinking of shared traits, such as hair and milk production in mammals, as evidence of shared ancestors. Because we now understand that DNA carries heritable information in the form of genes, we can see that genes and their products (proteins) document the hereditary background of an organism. The linear sequences of nucleotides in DNA molecules are passed from parents to offspring; these sequences determine the amino acid sequences of proteins. Siblings have greater similarity in their DNA and proteins than do unrelated individuals of the same species. If the evolutionary view of life is valid, we should be able to extend this concept of "molecular genealogy" to relationships between species: We should expect two species that appear to be closely related based on fossil and anatomical evidence to also share a greater proportion of their DNA and protein sequences than do more distantly related species. In fact, that is the case. An example is the comparison of the β polypeptide chain of human hemoglobin with the corresponding hemoglobin polypeptide in other vertebrates. In this chain of 146 amino acids, humans and gorillas differ in just 1 amino acid, while humans and frogs differ in 67 amino acids. Molecular biology has added a new tape measure to the toolkit biologists use to assess evolutionary kinship.

The Theme of Emergent Properties in the Chemistry of Life: *A Review*

Recall that life is organized along a hierarchy of structural levels (see Figure 1.4). With each increasing level of order, new properties emerge. In Chapters 2–5, we have dissected the chemistry of life. But we have also begun to develop a more integrated view of life, exploring how properties emerge with increasing order.

We have seen that water's behavior results from the interactions of its molecules, each an ordered arrangement of hydrogen and oxygen atoms. We reduced the complexity and diversity of organic compounds to carbon skeletons and appended chemical groups. We saw that macromolecules are assembled from small organic molecules, taking on new properties. By completing our overview with an introduction to macromolecules and lipids, we have built a bridge to Unit Two, where we will study cell structure and function. We will keep a balance between the need to reduce life to simpler processes and the ultimate satisfaction of viewing those processes in their integrated context.

CONCEPT CHECK 5.5

1. **DRAW IT** Go to Figure 5.26a and, for the top three nucleotides, number all the carbons in the sugars, circle the nitrogenous bases, and star the phosphates.
2. **DRAW IT** In a DNA double helix, a region along one DNA strand has this sequence of nitrogenous bases: 5′-TAGGCCT-3′. Copy this sequence, and write down its complementary strand, clearly indicating the 5′ and 3′ ends of the complementary strand.
3. **WHAT IF?** (a) Suppose a substitution occurred in one DNA strand of the double helix in question 2, resulting in

   5′-TAAGCCT-3′
   3′-ATCCGGA-5′

   Copy these two strands, and circle and label the mismatched bases. (b) If the modified top strand is used by the cell to construct a complementary strand, what would that matching strand be?

For suggested answers, see Appendix A.

5 CHAPTER REVIEW

SUMMARY OF KEY CONCEPTS

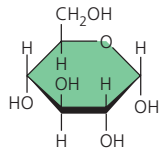
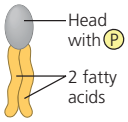
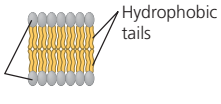
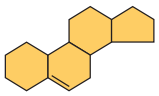
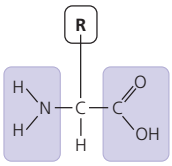
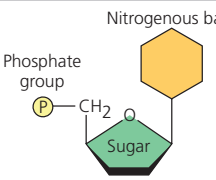
CONCEPT 5.1

Macromolecules are polymers, built from monomers (pp. 68–69)

- Carbohydrates, proteins, and nucleic acids are **polymers**, chains of **monomers**. The components of lipids vary.

Monomers form larger molecules by **dehydration reactions**, in which water molecules are released. Polymers can disassemble by the reverse process, **hydrolysis**. An immense variety of polymers can be built from a small set of monomers.

? *What is the fundamental basis for the differences between carbohydrates, proteins, and nucleic acids?*

Large Biological Molecules	Components	Examples	Functions
CONCEPT 5.2 **Carbohydrates serve as fuel and building material (pp. 69–74)** ? *Compare the composition, structure, and function of starch and cellulose. What role do starch and cellulose play in the human body?*	 Monosaccharide monomer	**Monosaccharides:** glucose, fructose	Fuel; carbon sources that can be converted to other molecules or combined into polymers
		Disaccharides: lactose, sucrose	
		Polysaccharides: • Cellulose (plants) • Starch (plants) • Glycogen (animals) • Chitin (animals and fungi)	• Strengthens plant cell walls • Stores glucose for energy • Stores glucose for energy • Strengthens exoskeletons and fungal cell walls
CONCEPT 5.3 **Lipids are a diverse group of hydrophobic molecules (pp. 74–77)** ? *Why are lipids not considered to be macromolecules or polymers?*		**Triacylglycerols** (fats or oils): glycerol + 3 fatty acids	Important energy source 
		**Phospholipids:** phosphate group + 2 fatty acids	Lipid bilayers of membranes 
	 Steroid backbone	**Steroids:** four fused rings with attached chemical groups	• Component of cell membranes (cholesterol) • Signaling molecules that travel through the body (hormones)
CONCEPT 5.4 **Proteins include a diversity of structures, resulting in a wide range of functions (pp. 77–86)** ? *Proteins are the most structurally and functionally diverse class of biological molecules. Explain the basis for this diversity.*	 Amino acid monomer (20 types)	• Enzymes • Structural proteins • Storage proteins • Transport proteins • Hormones • Receptor proteins • Motor proteins • Defensive proteins	• Catalyze chemical reactions • Provide structural support • Store amino acids • Transport substances • Coordinate organismal responses • Receive signals from outside cell • Function in cell movement • Protect against disease
CONCEPT 5.5 **Nucleic acids store, transmit, and help express hereditary information (pp. 86–89)** ? *What role does complementary base pairing play in the functions of nucleic acids?*		**DNA:**  • Sugar = deoxyribose • Nitrogenous bases = C, G, A, T • Usually double-stranded	Stores hereditary information
		RNA:  • Sugar = ribose • Nitrogenous bases = C, G, A, U • Usually single-stranded	Various functions during gene expression, including carrying instructions from DNA to ribosomes

TEST YOUR UNDERSTANDING

LEVEL 1: KNOWLEDGE/COMPREHENSION

1. Which of the following categories includes all others in the list?
   a. monosaccharide
   b. disaccharide
   c. starch
   d. carbohydrate
   e. polysaccharide

2. The enzyme amylase can break glycosidic linkages between glucose monomers only if the monomers are in the α form. Which of the following could amylase break down?
   a. glycogen, starch, and amylopectin
   b. glycogen and cellulose
   c. cellulose and chitin
   d. starch and chitin
   e. starch, amylopectin, and cellulose

3. Which of the following statements concerning *unsaturated* fats is true?
   a. They are more common in animals than in plants.
   b. They have double bonds in the carbon chains of their fatty acids.
   c. They generally solidify at room temperature.
   d. They contain more hydrogen than do saturated fats having the same number of carbon atoms.
   e. They have fewer fatty acid molecules per fat molecule.

4. The structural level of a protein *least* affected by a disruption in hydrogen bonding is the
   a. primary level.
   b. secondary level.
   c. tertiary level
   d. quaternary level.
   e. All structural levels are equally affected.

5. Enzymes that break down DNA catalyze the hydrolysis of the covalent bonds that join nucleotides together. What would happen to DNA molecules treated with these enzymes?
   a. The two strands of the double helix would separate.
   b. The phosphodiester linkages of the polynucleotide backbone would be broken.
   c. The purines would be separated from the deoxyribose sugars.
   d. The pyrimidines would be separated from the deoxyribose sugars.
   e. All bases would be separated from the deoxyribose sugars.

LEVEL 2: APPLICATION/ANALYSIS

6. The molecular formula for glucose is $C_6H_{12}O_6$. What would be the molecular formula for a polymer made by linking ten glucose molecules together by dehydration reactions?
   a. $C_{60}H_{120}O_{60}$
   b. $C_6H_{12}O_6$
   c. $C_{60}H_{102}O_{51}$
   d. $C_{60}H_{100}O_{50}$
   e. $C_{60}H_{111}O_{51}$

7. Which of the following pairs of base sequences could form a short stretch of a normal double helix of DNA?
   a. 5′-purine-pyrimidine-purine-pyrimidine-3′ with 3′-purine-pyrimidine-purine-pyrimidine-5′
   b. 5′-AGCT-3′ with 5′-TCGA-3′
   c. 5′-GCGC-3′ with 5′-TATA-3′
   d. 5′-ATGC-3′ with 5′-GCAT-3′
   e. All of these pairs are correct.

8. Construct a table that organizes the following terms, and label the columns and rows.

phosphodiester linkages	polypeptides	monosaccharides
peptide bonds	triacylglycerols	nucleotides
glycosidic linkages	polynucleotides	amino acids
ester linkages	polysaccharides	fatty acids

9. **DRAW IT** Copy the polynucleotide strand in Figure 5.26a and label the bases G, T, C, and T, starting from the 5′ end. Assuming this is a DNA polynucleotide, now draw the complementary strand, using the same symbols for phosphates (circles), sugars (pentagons), and bases. Label the bases. Draw arrows showing the 5′ → 3′ direction of each strand. Use the arrows to make sure the second strand is antiparallel to the first. *Hint*: After you draw the first strand vertically, turn the paper upside down; it is easier to draw the second strand from the 5′ toward the 3′ direction as you go from top to bottom.

LEVEL 3: SYNTHESIS/EVALUATION

10. **EVOLUTION CONNECTION**
    Comparisons of amino acid sequences can shed light on the evolutionary divergence of related species. If you were comparing two living species, would you expect all proteins to show the same degree of divergence? Why or why not?

11. **SCIENTIFIC INQUIRY**
    Suppose you are a research assistant in a lab studying DNA-binding proteins. You have been given the amino acid sequences of all the proteins encoded by the genome of a certain species and have been asked to find candidate proteins that could bind DNA. What type of amino acids would you expect to see in such proteins? Why?

12. **SCIENCE, TECHNOLOGY, AND SOCIETY**
    Some amateur and professional athletes take anabolic steroids to help them "bulk up" or build strength. The health risks of this practice are extensively documented. Apart from health considerations, how do you feel about the use of chemicals to enhance athletic performance? Is an athlete who takes anabolic steroids cheating, or is such use part of the preparation required to succeed in competition? Explain.

13. **WRITE ABOUT A THEME**
    Structure and Function Proteins, which have diverse functions in a cell, are all polymers of the same subunits—amino acids. Write a short essay (100–150 words) that discusses how the structure of amino acids allows this one type of polymer to perform so many functions.

For selected answers, see Appendix A.

MasteringBIOLOGY® www.masteringbiology.com

1. MasteringBiology® Assignments

Tutorials Types of Carbohydrates • Amino Acid Functional Groups • Levels of Structure in Proteins • Nucleic Acid Building Blocks • The Double Helix

Activities Condensation and Hydrolysis Reactions • Making and Breaking Polymers • Carbohydrate Structure and Function • Carbohydrates • Lipids • Protein Functions • Protein Structure • Nucleic Acid Structure • Structure of RNA and DNA

Questions Student Misconceptions • Reading Quiz • Multiple Choice • End-of-Chapter

2. eText

Read your book online, search, take notes, highlight text, and more.

3. The Study Area

Practice Tests • Cumulative Test • BioFlix 3-D Animations • MP3 Tutor Sessions • Videos • Activities • Investigations • Lab Media • Audio Glossary • Word Study Tools • Art

UNIT 2

The Cell

An Interview with

Bonnie L. Bassler

Bonnie Bassler loves her life as a biologist scrutinizing the secret lives of bacteria. For the past 20 years or so, Bonnie and her lab (her "gang," as she calls them) have made momentous discoveries about how bacterial cells use chemicals to communicate with each other in a process called quorum sensing. Dr. Bassler has a B.S. in Biochemistry from the University of California at Davis and a Ph.D. in Biochemistry from The Johns Hopkins University. Among her many awards and honors, she has received a MacArthur Foundation Fellowship and is a member of the National Academy of Sciences. She is the 2010–2011 President of the American Society for Microbiology, the largest specialized life science organization in the world. At Princeton University since 1994, she is currently the Squibb Professor in Molecular Biology and an Investigator of the Howard Hughes Medical Institute.

How did you get started in science?

I've always been interested in nature and animals, and in puzzles and mystery books—I really like figuring things out. As an undergraduate at UC Davis, I worked in a lab on a bacterial project while taking courses in both biochemistry and genetics. Then, as a graduate student at Johns Hopkins, I learned a lot of biochemistry while studying marine bacteria. The bacteria belong to the genus *Vibrio*, and I was working on chemotaxis, movement by cells toward food or away from noxious chemicals in the environment.

What are the advantages of using bacteria for research in cell biology?

Bacteria have been the foundation of molecular biology for the last 100 years because they're accessible. They grow fast, they form clones of identical cells, and they're amenable to biochemical and genetic analyses. Most of what we initially learned about molecular biology—about genes and proteins and other biomolecules—came from work done on bacteria. Because of evolutionary history, the most basic and ancient life processes that happen in bacteria also happen in humans and other higher organisms. Humans have more bells and whistles, more proteins, more sophistication, more complexity. But if you want to understand the basic components of a process, very often you can use bacteria to do that. Also, working with bacteria fits my personality. I prefer having 10 billion bacterial offspring the day after an experiment to having to wait weeks or months for a small number of baby mice. Every morning there's a surprise waiting for me in the incubator!

What is quorum sensing, and how did you first hear about it?

When I was finishing my graduate work, I heard a talk by Mike Silverman, a scientist with the Agouron Institute in San Diego, about how bacteria "talk" to each other, "count" their own numbers, and coordinate their behavior. Mike had been working on a light-producing (bioluminescent) marine bacterium called *Vibrio fischeri* that lives symbiotically inside a variety of marine animals. The animal provides nutrients for the bacteria, which live in an enclosed space within the animal's body. In return, the bacteria provide light that benefits the animal—by scaring away predators or attracting prey or a mate. But if only a small number of bacteria are present, they do not make light—producing light would waste energy because the light wouldn't be visible. The word "quorum" means "the number needed to do something," and bacteria can sense whether there is a quorum or not and act accordingly.

The way quorum sensing works is that bacteria release certain signaling chemicals into the environment. As the bacterial cells increase in number, the molecules reach a concentration at which many of them bind to receptor proteins on the surface of or inside the bacteria. The signaling molecule fits together with the receptor like a key in a lock. In the case of the surface receptors, each receptor molecule has a part on the outside of the cell and a part on the inside. The signaling chemical binds to the outer part of the receptor, "tickling" the protein so that it makes something happen inside the cell. For instance, in *Vibrio fischeri*, binding of signaling molecules ultimately turns on genes that code for enzymes that make light. Mike had worked out this mechanism of how cells of *Vibrio fischeri* turn on light in synchrony.

It's important to understand that back then, we just didn't think about bacteria like that—we thought bacteria ignored each other and did their own thing as solitary cells. I was totally fascinated. I thought, "He's either crazy or he's brilliant—but I just have to work on that." I went up to the podium after his talk and begged him to let me be his postdoc. Finally he said yes, even though he was a geneticist and I was a biochemist! He took a chance on me.

How does a genetic approach differ from a biochemical one?

Geneticists make lots of mutant organisms, then think up clever strategies to find the ones with mutations in the genes they're interested in. In the case of quorum sensing in bioluminescent bacteria, you look for cells that remain dark. If you have mutated genes involved in quorum sensing, you would expect the bacteria not to make light because light emission depends on the cells communicating with each other. Eventually, you would hope to identify the components that function in normal light-emitting bacteria but not in the mutants. Biochemists, on the other hand, start by isolating molecules and studying their properties directly. Genetics and biochemistry are complementary approaches. I'm glad I know both because the combination is more effective than either approach by itself.

What did you learn about quorum sensing as a postdoc?

In Mike's lab, I worked on another species of bioluminescent *Vibrio* called *Vibrio harveyi*. Because these bacteria are free-living in the ocean, we thought their quorum-sensing molecular circuitry might be more complicated than that of *Vibrio fischeri*. What I found was that *Vibrio harveyi* has two parallel systems for quorum sensing, one that senses cells of the same species and one that counts bacteria of other species. Fast-forwarding a decade or so, this second system

seems to be present in *many* bacteria, and the second signaling molecule appears to be universal. So, apparently bacteria can measure the ratios of these two signals, and they're saying, "How many of us and how many of you are there?" Then they do different things, depending on who is in the majority. And this isn't just restricted to bioluminescence. Other bacterial behaviors are also controlled by quorum sensing, such as forming an organized thin layer (called a biofilm) on your teeth or coordinating a virulent infection.

Tell us more about biofilms.

We used to think that most bacteria lived as individual cells suspended in liquid environments. But we now understand that in the wild, they live attached to surfaces in biofilms, and they secrete carbohydrates and other molecules that form a protective slime on the biofilm surface. Most of us have noticed the biofilm coating our teeth every morning. Believe it or not, there are about 600 bacterial species in that biofilm just trying to make a living, getting nutrients from us, but the side effect is that we get cavities. And when someone has a lung infection or an implant or heart valve that harbors an infection, the bacteria are growing as a biofilm in the lungs or on the introduced device. So we now understand why these infections are so hard to treat: It's because the slime on the biofilm is providing a protective shield that antibiotics can't penetrate.

What questions are you and your lab asking now?

My group is interested in how information outside an organism gets inside so that the organism does the right thing at the right time. We work on bacteria because they're simple, but we hope that we will have insights for people working on higher organisms. And we're curious about how collective behaviors first evolved on Earth. How did multicellularity come about? We know that the first organisms were bacteria, but how did they begin to do things together? How did groups of cells in your body come to act like a liver or a heart? We're very interested in how the flow of information through networks facilitates multicellularity.

Are there applications for the basic research you do?

When you're asking fundamental questions, you hope that surprises, things you never thought of, will come out of it. Now that we know that bacteria talk to each other and perform group activities, the question is whether we could interfere with these conversations for therapeutic purposes. Could we make molecules that keep bacteria from "talking" or "hearing"? Maybe these would be new antibiotics. Biofilms are a terrible problem in medicine and dental health, and now that we are starting to know the molecular basis for their formation, maybe we can learn how to prevent them from forming.

Bacteria get a lot of bad press for the negative things they do. On the other hand, bacteria also do many miraculous things that keep us alive; they are working for us every instant of our lives. You are covered with a bacterial biofilm that acts as invisible body armor—these good bacteria occupy all the spaces on your skin, preventing invading bacteria from attaching. Throughout your gut you have a huge mass of bacteria, and they're making vitamins for you and helping you digest your food. So all biofilms aren't bad—and for the good biofilms, what if we could find a molecule to make quorum sensing better? Rather than an antibiotic, this would be a probiotic.

Where do you think this field is going?

I think we'll be turning our attention to the possibility of communication between organisms from different kingdoms and different domains. Bacteria have been around for over 4 billion years and have probably been living with multicellular eukaryotic hosts for hundreds of millions of years. So why wouldn't these hosts have evolved strategies to listen in, say, to the conversation being carried out by a group of pathogenic bacteria? Does our immune system "hear" bacterial signaling molecules? Do hosts actively prevent quorum sensing among pathogenic bacteria? Do they tune in and help the good bacteria? I think this is going to be a dialogue, not a monologue.

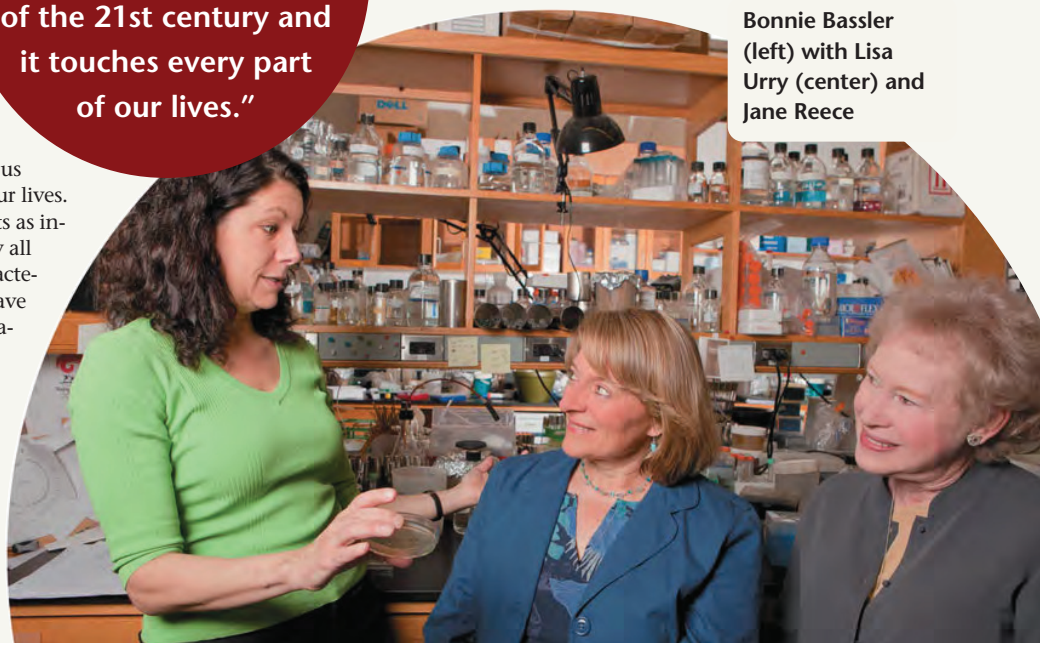
What do you enjoy most about your life as a scientist?

I love what I work on. I figured out as a postdoc how much fun this life in science is—that it is not about me against other scientists or who is going to discover something first. Instead, it's me against this bacterium, and we are in it head-to-head for the rest of our lives, in a contest of wills between bacteria trying to keep their secrets and me trying to discover them. Also, the basic question of how groups work together fascinates me. I work with a fantastic group of students and we share everything—everybody gives everybody everything, and then we all get more. That's quorum sensing! Both my molecular and nonmolecular lives involve getting the group to do more than the individual. I love that parallelism! My gang of students show me their data, and it's my job to help them figure out the science and get on with their careers. I'm so lucky—having 24 hands and 12 brains is so much better than two hands and one brain. The science is always changing, and trying to keep up with these young and tireless people is hugely challenging and rewarding.

What is your advice to an undergraduate who is considering a career in biology?

For undergraduates who are considering a life in science, my advice is to work on something that you are passionate about. Don't be limited by thinking that bench science is the only thing a scientist can do. There are so many potential careers for a biologist. You could work on Capitol Hill as a scientific advisor or policymaker. You could teach. You could be a lawyer. You could be a writer who helps the public understand science. You could work on science education at the kindergarten level. Figure out your particular combination of personality traits and what you really love doing as a scientist; then make that niche for yourself and bring science to that career. The sky's the limit for biologists because biology is the science of the 21st century and it touches every part of our lives.

"The sky's the limit for biologists because biology is the science of the 21st century and it touches every part of our lives."



Bonnie Bassler (left) with Lisa Urry (center) and Jane Reece