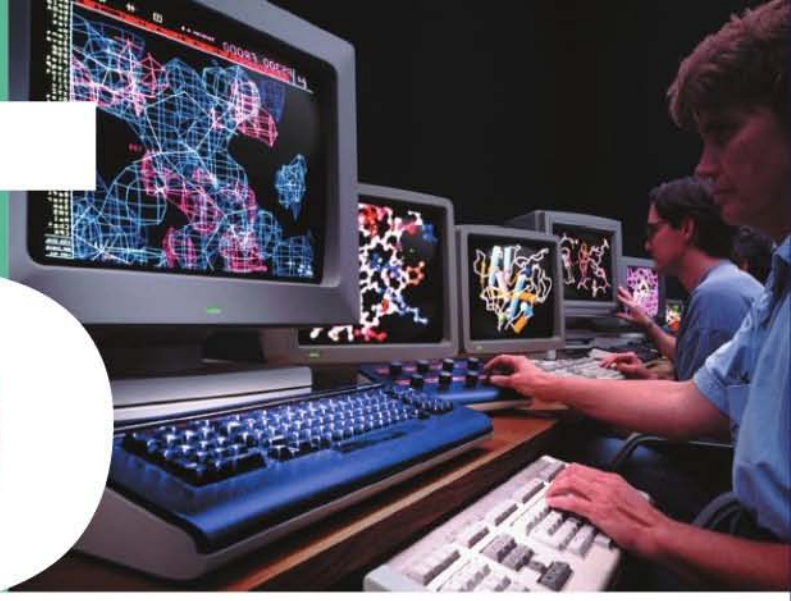
5 The Structure and Function of Large Biological Molecules



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

KEY CONCEPTS

5.1 Macromolecules are polymers, built from monomers
5.2 Carbohydrates serve as fuel and building material
5.3 Lipids are a diverse group of hydrophobic molecules
5.4 Proteins have many structures, resulting in a wide range of functions
5.5 Nucleic acids store and transmit hereditary information

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 thus 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 structures of so many of them **(Figure 5.1)**.

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

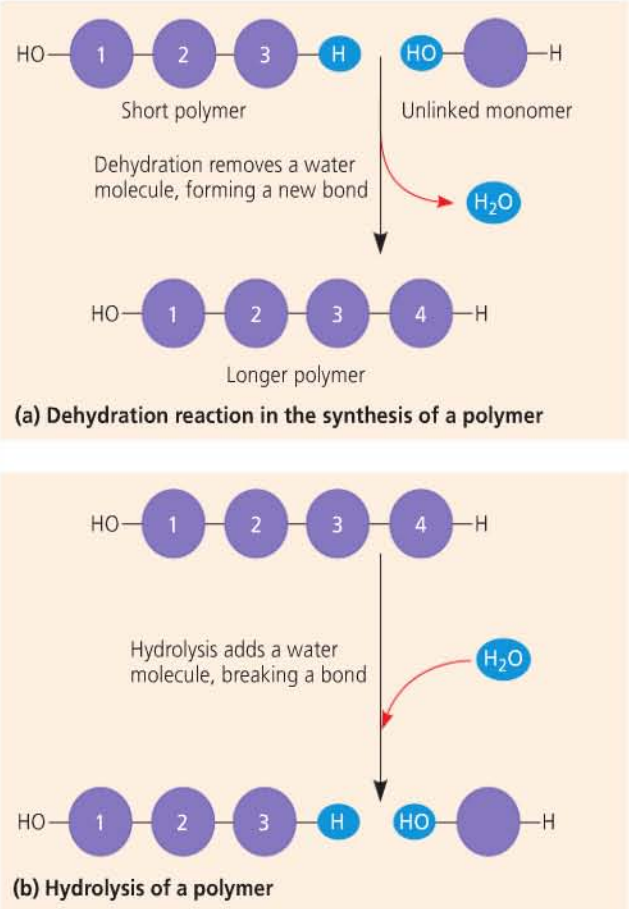
Macromolecules are polymers, built from monomers

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 *meris*, 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**. Some of the molecules that serve as monomers also have other functions of their own.

The Synthesis and Breakdown of Polymers

The classes of polymers differ in the nature of their monomers, but the chemical mechanisms by which cells make and break down polymers are basically the same in all cases **(Figure 5.2)**. Monomers are connected by a reaction in which two molecules are covalently bonded to each other through loss of a water molecule; this is known as a **condensation reaction**, specifically a **dehydration reaction**, because water is the molecule that is lost **(Figure 5.2a)**. When a bond forms between two monomers, each monomer contributes part of the water molecule that is lost: One molecule provides a hydroxyl group (—OH), while the other provides a hydrogen (—H). This reaction can be repeated as monomers are added to the chain one by one, making a polymer. The dehydration process is facilitated by **enzymes**, specialized macromolecules that speed up chemical reactions in cells.

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



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

reaction **(Figure 5.2b)**. Hydrolysis means to break using water (from the Greek *hydro*, water, and *lysis*, break). Bonds between the monomers are broken by the addition of water molecules, with a hydrogen from the water attaching to one monomer and a 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 kinds of macromolecules; the collection varies from one type of cell to another even in the same organism. The inherent differences between human siblings reflect 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 look at 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?
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 green beans. What reactions must occur for the amino acid monomers in the protein of the beans to be converted to 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, also known as simple sugars. 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

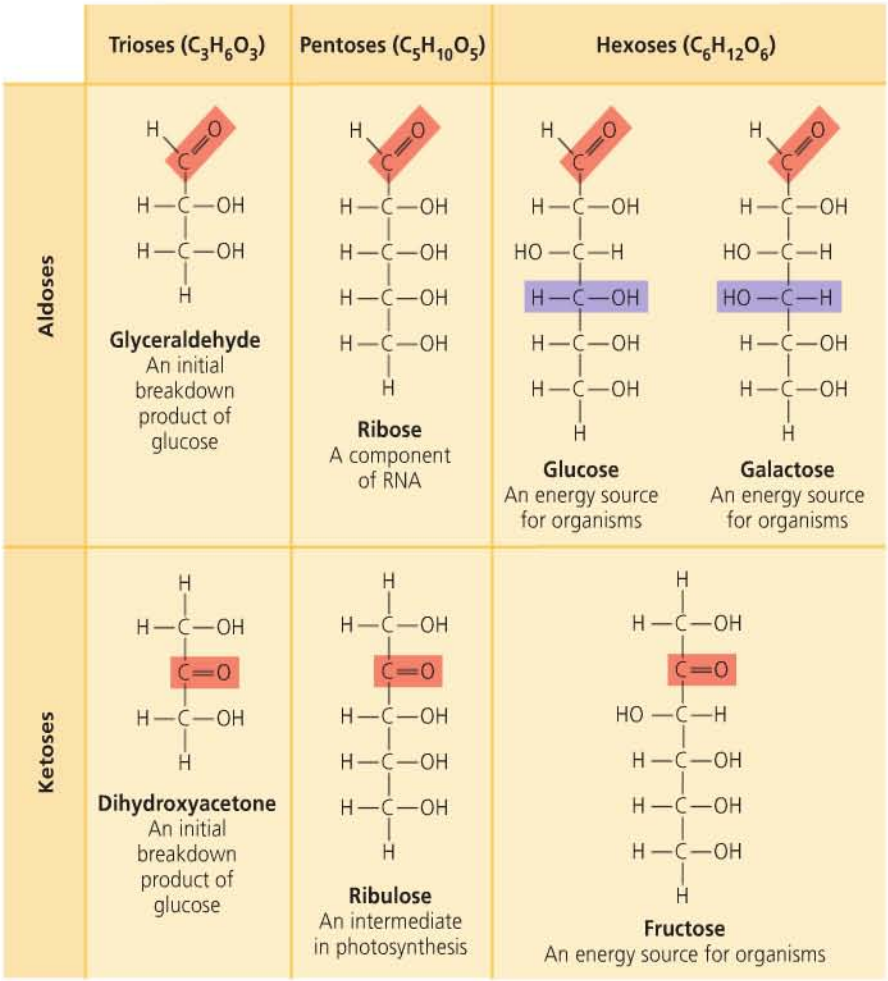
some multiple of the unit CH_2O **(Figure 5.3)**. Glucose ($C_6H_{12}O_6$), the most common monosaccharide, is of central importance in the chemistry of life. In the structure of glucose, we can see the trademarks of a sugar: The molecule has a carbonyl group (>C=O) and multiple hydroxyl groups (—OH). 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, a structural 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.

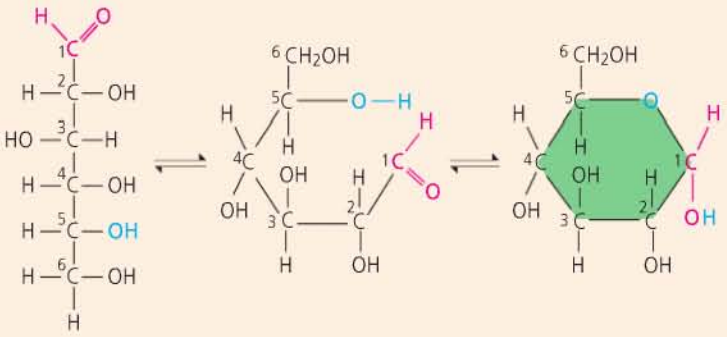
Although it is convenient to draw glucose with a linear carbon skeleton, this representation is not completely accurate. In aqueous solutions, glucose molecules, as well as most other sugars, form rings **(Figure 5.4)**.

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. Not only are simple-sugar molecules a major fuel for cellular work; 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.

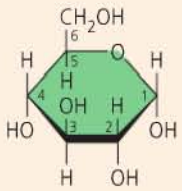
A **disaccharide** consists of two monosaccharides joined by a **glycosidic linkage**, a covalent bond formed between two monosaccharides by a dehydration reaction. For example, maltose is a disaccharide formed by the linking of two molecules of glucose **(Figure 5.5a)**. Also known as malt sugar, maltose 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.



▶ **Figure 5.3 The structure and classification of some monosaccharides.** Sugars may be aldoses (aldehyde sugars, top row) or ketoses (ketone sugars, bottom row), depending on the location of the carbonyl group (dark orange). Sugars are also classified according to the length of their carbon skeletons. A third point of variation is the spatial arrangement around asymmetric carbons (compare, for example, the purple portions of glucose and galactose).



(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.

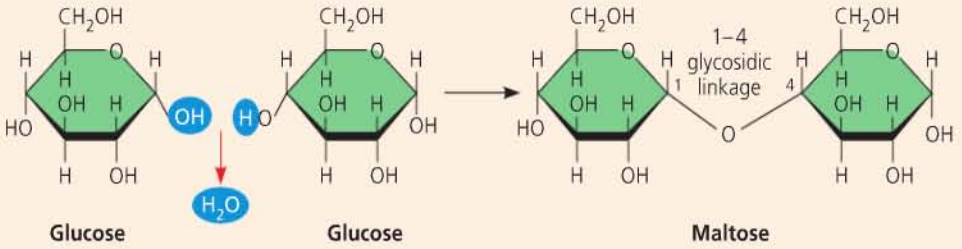


(b) Abbreviated ring structure. Each corner represents a carbon. The ring's thicker edge indicates that you are looking at the ring edge-on; the components attached to the ring lie above or below the plane of the ring.

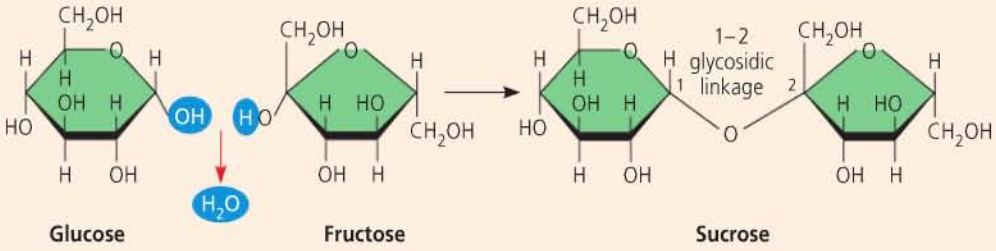
▲ **Figure 5.4 Linear and ring forms of glucose.**

DRAW IT *Start with the linear form of fructose (see Figure 5.3) and draw the formation of the fructose ring in two steps. Number the carbons. Attach carbon 5 via 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.**

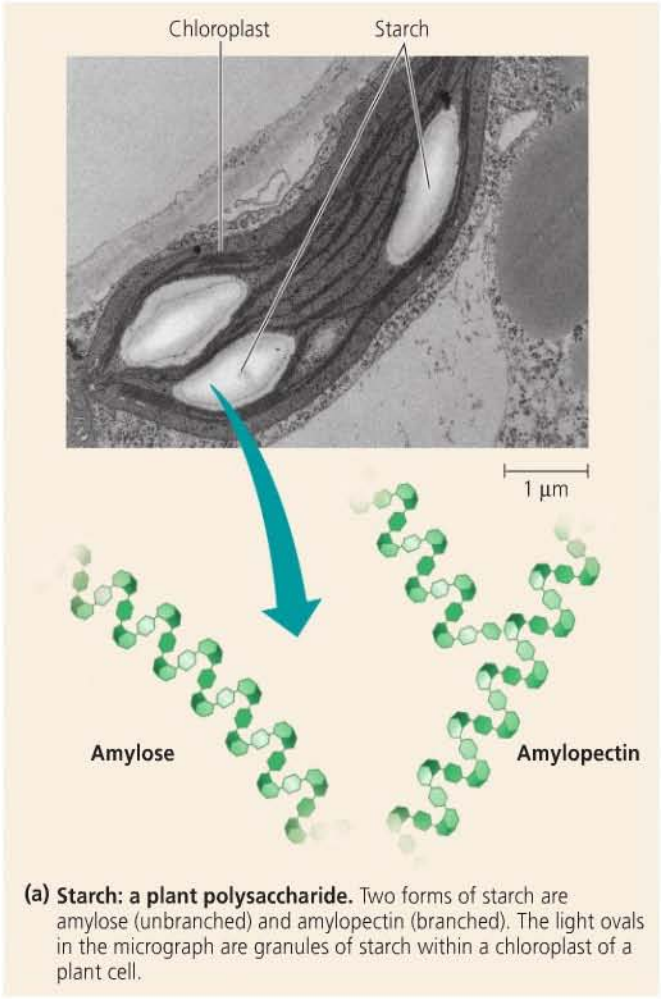
Polysaccharides

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 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.

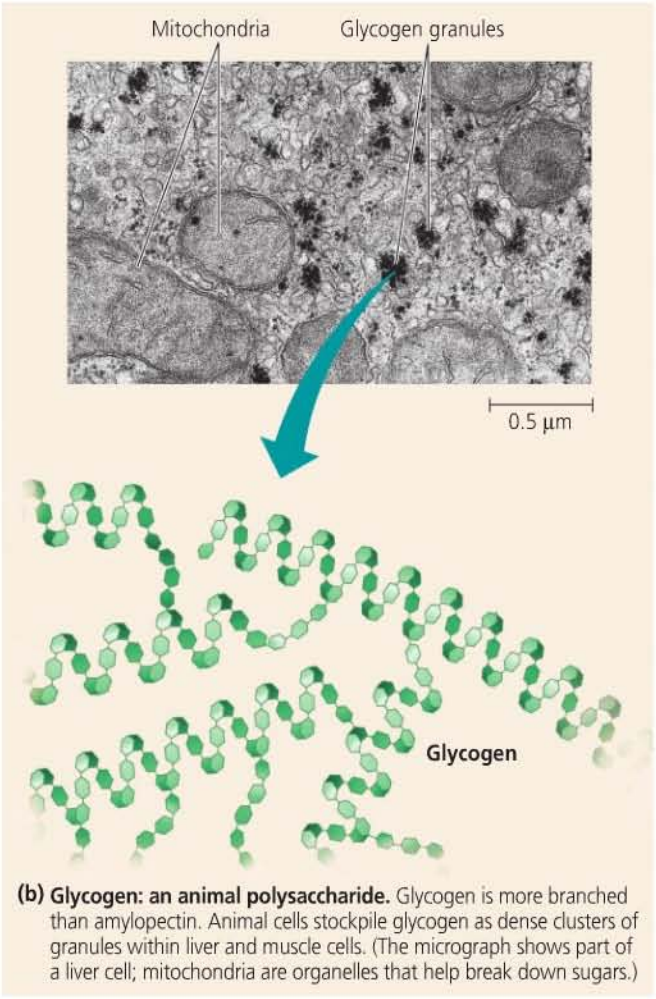
Storage Polysaccharides

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 angle of these



(a) Starch: a plant polysaccharide. Two forms of starch are amylose (unbranched) and amylopectin (branched). The light ovals in the micrograph are granules of starch within a chloroplast of a plant cell.



(b) Glycogen: an animal polysaccharide. Glycogen is more branched than amylopectin. Animal cells stockpile glycogen as dense clusters of granules within liver and muscle cells. (The micrograph shows part of a liver cell; mitochondria are organelles that help break down sugars.)

▲ **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 their molecular structure, the polymer chains tend to form helices.

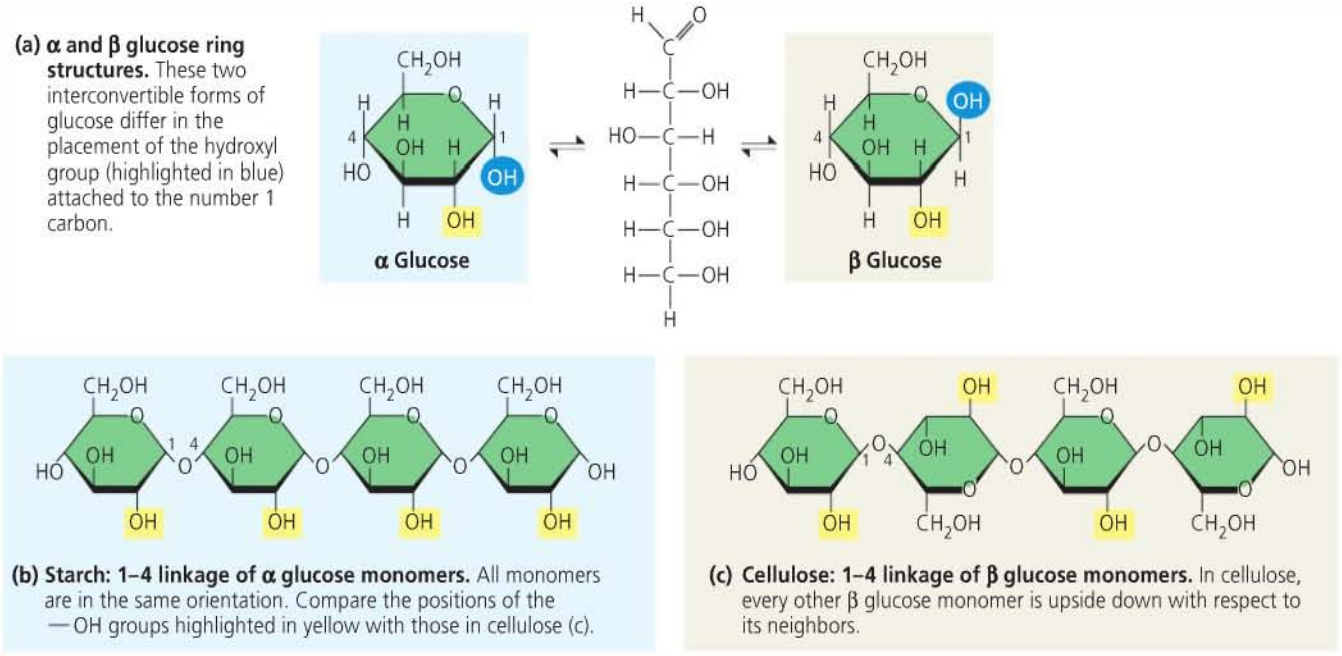
bonds makes the polymer helical. The simplest form of starch, amylose, is unbranched **(Figure 5.6a)**. Amylopectin, a more complex starch, is a branched polymer with 1–6 linkages at the branch points.

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.

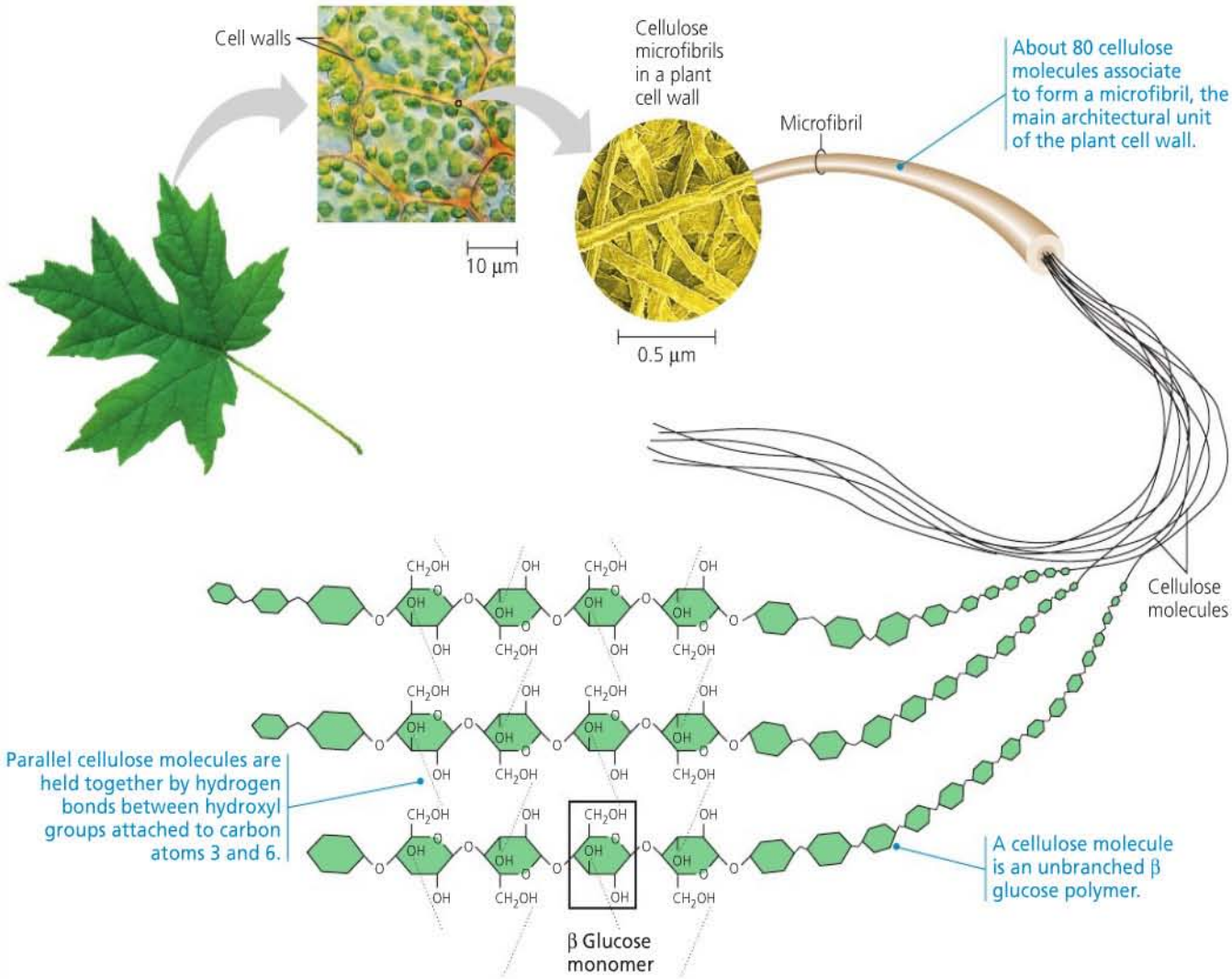
Structural Polysaccharides

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 other glucose monomer upside down with respect to its neighbors **(Figure 5.7c)**.

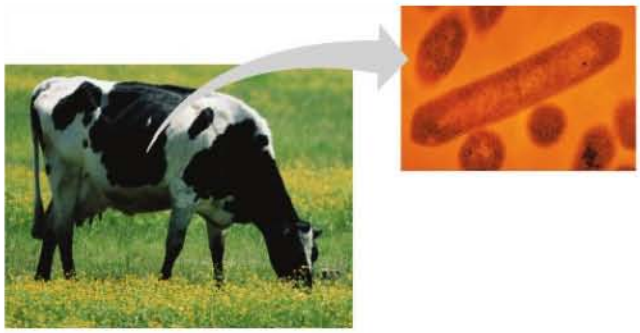
The differing glycosidic linkages in starch and cellulose give the two molecules distinct three-dimensional shapes. Whereas a starch molecule is mostly helical, a cellulose molecule is straight. Cellulose is never branched, and some hydroxyl groups on its glucose monomers are free to hydrogen-bond with 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



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



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



▲ **Figure 5.9 Cellulose-digesting prokaryotes are found in grazing animals such as this cow.**

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. 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.

Some prokaryotes can digest cellulose, breaking it down into glucose monomers. A cow harbors cellulose-digesting prokaryotes in its rumen, the first compartment in its stomach **(Figure 5.9)**. The prokaryotes hydrolyze the cellulose of hay and grass and convert the glucose to other nutrients that nourish the cow. Similarly, a termite, which is unable to digest cellulose by itself, has prokaryotes 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 5.10)**. 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, except that the glucose monomer of chitin has a nitrogen-containing appendage (see Figure 5.10a).

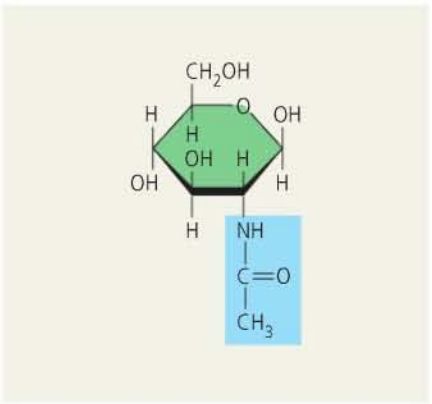
CONCEPT CHECK 5.2

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?** What would happen if a cow were given antibiotics that killed all the prokaryotes in its stomach?

For suggested answers, see Appendix A.

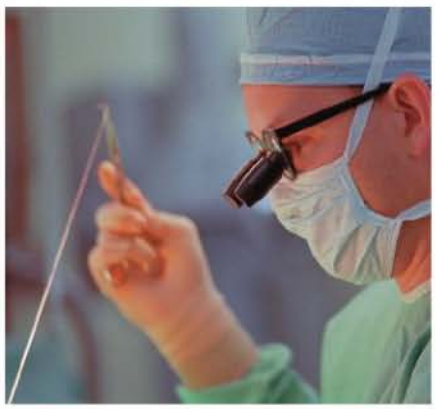
CONCEPT 5.3
Lipids are a diverse group of hydrophobic molecules

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.



(a) The structure of the chitin monomer.

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



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

▲ **Figure 5.10 Chitin, a structural polysaccharide.**

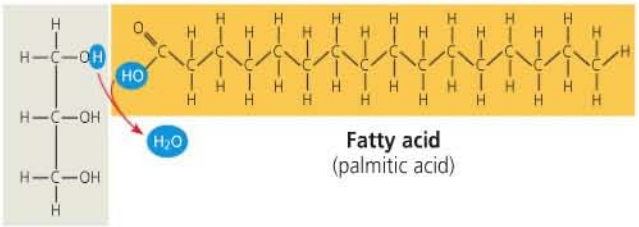
Lipids are varied in form and function. They include waxes and certain pigments, but we will focus on the most biologically important types of lipids: fats, phospholipids, and steroids.

Fats

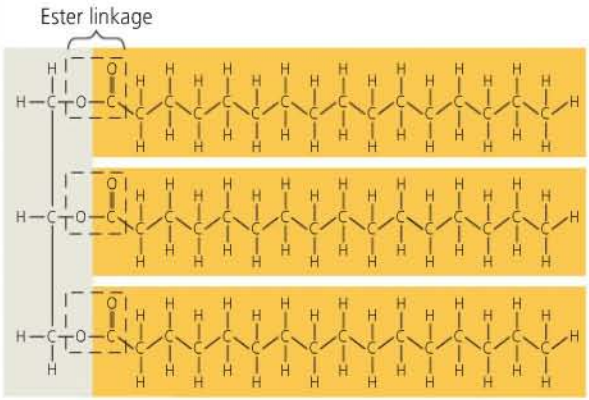
Although fats are not polymers, they are large molecules assembled from a few smaller molecules by dehydration reactions. A **fat** is constructed from two kinds of smaller molecules: glycerol and fatty acids **(Figure 5.11a)**. Glycerol is an alcohol with three carbons, each bearing a hydroxyl group. A **fatty acid** has a long carbon skeleton, usually 16 or 18 carbon atoms in length. The carbon at one end of the fatty acid is part of a carboxyl group, the functional group that gives these molecules the name fatty *acid.* Attached to the carboxyl group is a long 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 each join 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 *triglyceride*, a word often found in the list of ingredients on packaged foods.) The fatty acids in a fat can be the same, as in **Figure 5.11b**, or they can be of two or three different kinds.

Fatty acids vary in length and in the number and locations of double bonds. The terms *saturated fats* and *unsaturated fats* are commonly used in the context of nutrition **(Figure 5.12)**.

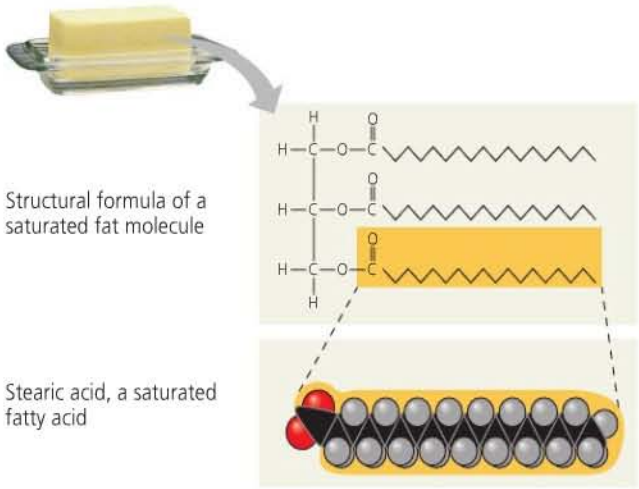


(a) Dehydration reaction in the synthesis of a fat

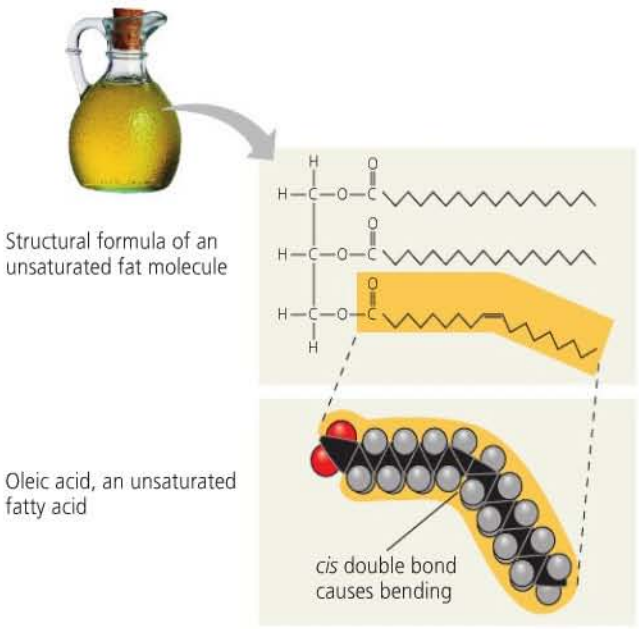


(b) Fat molecule (triacylglycerol)

▲ **Figure 5.11 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 identical fatty acid units. The carbons of the fatty acids are arranged zig-zag to suggest the actual orientations of the four single bonds extending from each carbon (see Figure 4.3a).



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



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

▲ **Figure 5.12 Examples of saturated and unsaturated fats and fatty acids.** The structural formula for each fat follows a common chemical convention of omitting the carbons and attached hydrogens of the hydrocarbon regions. In the space-filling models of the fatty acids, black = carbon, gray = hydrogen, and red = oxygen.

These terms refer to the structure of the hydrocarbon chains of the fatty acids. If there are no double bonds between carbon atoms composing the chain, then as many hydrogen atoms as possible are bonded to the carbon skeleton. Such a structure is described as being *saturated* with hydrogen, so the resulting fatty acid is called a **saturated fatty acid** (Figure 5.12a). An **unsaturated fatty acid** has one or more double bonds, formed by the removal of hydrogen atoms from the carbon skeleton. The fatty acid will have a kink in its hydrocarbon chain wherever a *cis* double bond occurs (Figure 5.12b).

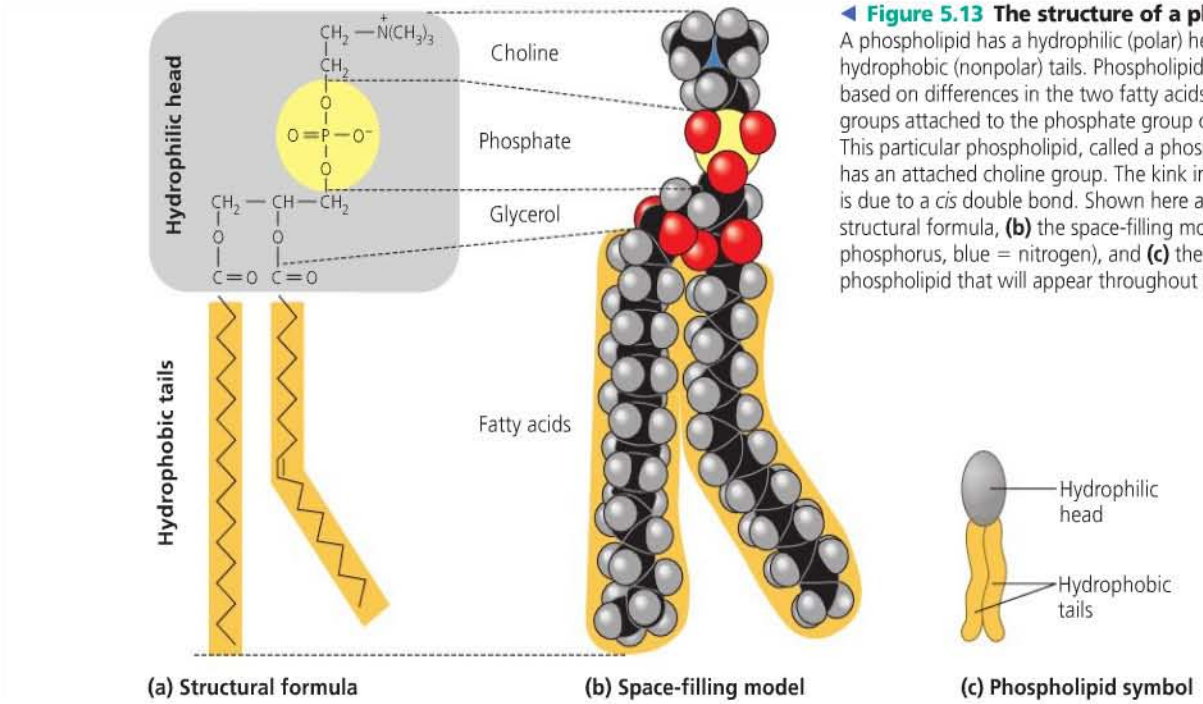
A fat made from saturated fatty acids is called a saturated fat. Most animal fats are saturated: The hydrocarbon chains of their fatty acids—the "tails" of the fat molecules—lack double bonds, and their flexibility allows the fat molecules to pack together tightly. Saturated animal fats—such as lard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

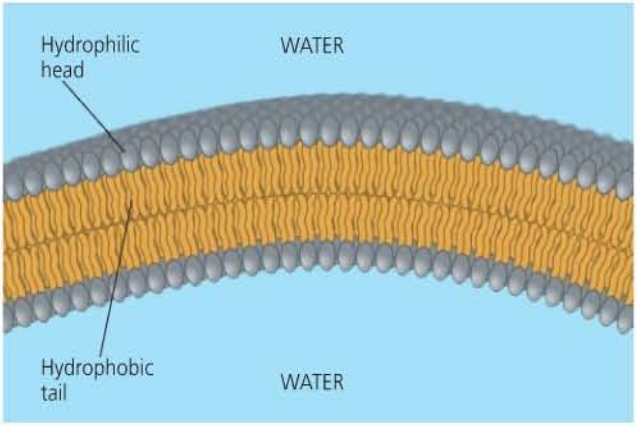
Fat has come to have such a negative connotation in our culture that you might wonder what useful purpose fats serve. The major function of fats is energy storage. The hydrocarbon chains of fats are similar to gasoline molecules and just as rich in energy. A gram of fat stores more than twice as much 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

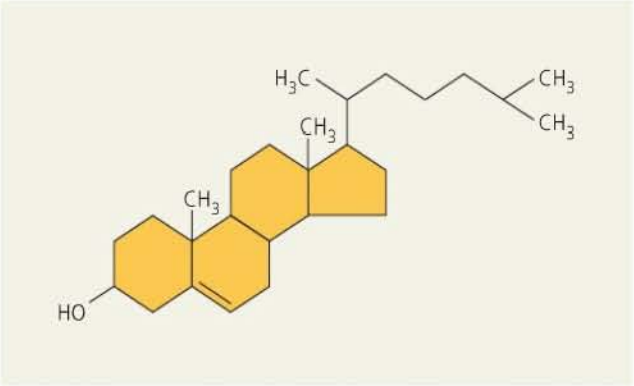
Cells could not exist without another type of lipid—**phospholipids** (Figure 5.13). Phospholipids are essential



◀ **Figure 5.13 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.



▲ **Figure 5.14 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.



▲ **Figure 5.15 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).

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 5.13, 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. 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 aggregates—bilayers—that shield their hydrophobic portions from water **(Figure 5.14)**.

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

Many hormones, as well as cholesterol, are **steroids**, which are lipids characterized by a carbon skeleton consisting of four fused rings **(Figure 5.15)**. Different steroids vary in the chemical groups attached to this ensemble of rings. **Cholesterol** 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. Many hormones, including vertebrate sex hormones, are steroids produced from cholesterol (see Figure 4.9). Thus, cholesterol is a crucial molecule in animals, although a high level of it in the blood may contribute to atherosclerosis. Both saturated fats and trans fats exert their negative impact on health by affecting cholesterol levels.

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 have many structures, resulting in a wide range of functions

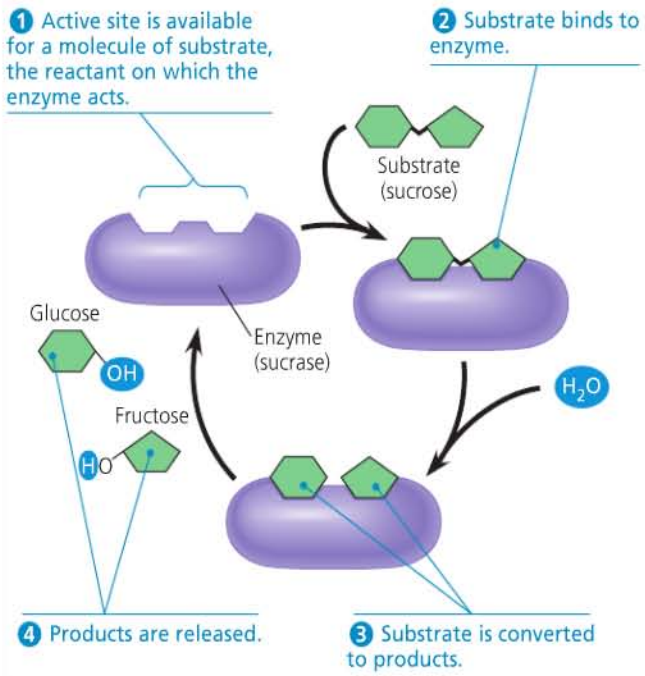
Nearly every dynamic function of a living being depends on proteins. In fact, the importance of proteins is underscored by their name, which comes from the Greek word *proteios*, meaning "first place." Proteins account for more than 50% of the dry mass of most cells, and they are instrumental in almost everything organisms do. Some proteins speed up chemical reactions, while others play a role in structural support,

Table 5.1 An Overview of Protein Functions

Type of Protein	Function	Examples
Enzymatic proteins	Selective acceleration of chemical reactions	Digestive enzymes catalyze the hydrolysis of the polymers in food.
Structural proteins	Support	Insects and spiders use silk fibers to make their cocoons and webs, respectively. Collagen and elastin provide a fibrous framework in animal connective tissues. Keratin is the protein of hair, horns, feathers, and other skin appendages.
Storage proteins	Storage of amino acids	Ovalbumin is the protein of egg white, used as an amino acid source for the developing embryo. Casein, the protein of milk, is the major source of amino acids for baby mammals. Plants have storage proteins in their seeds.
Transport proteins	Transport of other substances	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	Coordination of an organism's activities	Insulin, a hormone secreted by the pancreas, helps regulate the concentration of sugar in the blood of vertebrates.
Receptor proteins	Response of cell to chemical stimuli	Receptors built into the membrane of a nerve cell detect chemical signals released by other nerve cells.
Contractile and motor proteins	Movement	Actin and myosin are responsible for the contraction of muscles. Other proteins are responsible for the undulations of the organelles called cilia and flagella.
Defensive proteins	Protection against disease	Antibodies combat bacteria and viruses.

storage, transport, cellular communication, movement, and defense against foreign substances **(Table 5.1)**.

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 **(Figure 5.16)**.



▲ **Figure 5.16 The catalytic cycle of an enzyme.** The enzyme sucrase accelerates hydrolysis of sucrose into glucose and fructose. Acting as a catalyst, the sucrase protein is not consumed during the cycle, but remains available for further catalysis.

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 polymers constructed from the same set of 20 amino acids. Polymers of amino acids are called **polypeptides**. A **protein** consists of one or more polypeptides, each folded and coiled into a specific three-dimensional structure.

Amino Acid Monomers

All amino acids share a common structure. **Amino acids** are organic molecules possessing both carboxyl and amino groups (see Chapter 4). 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.

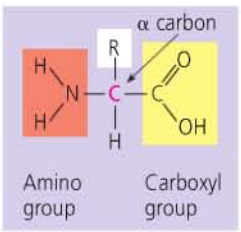
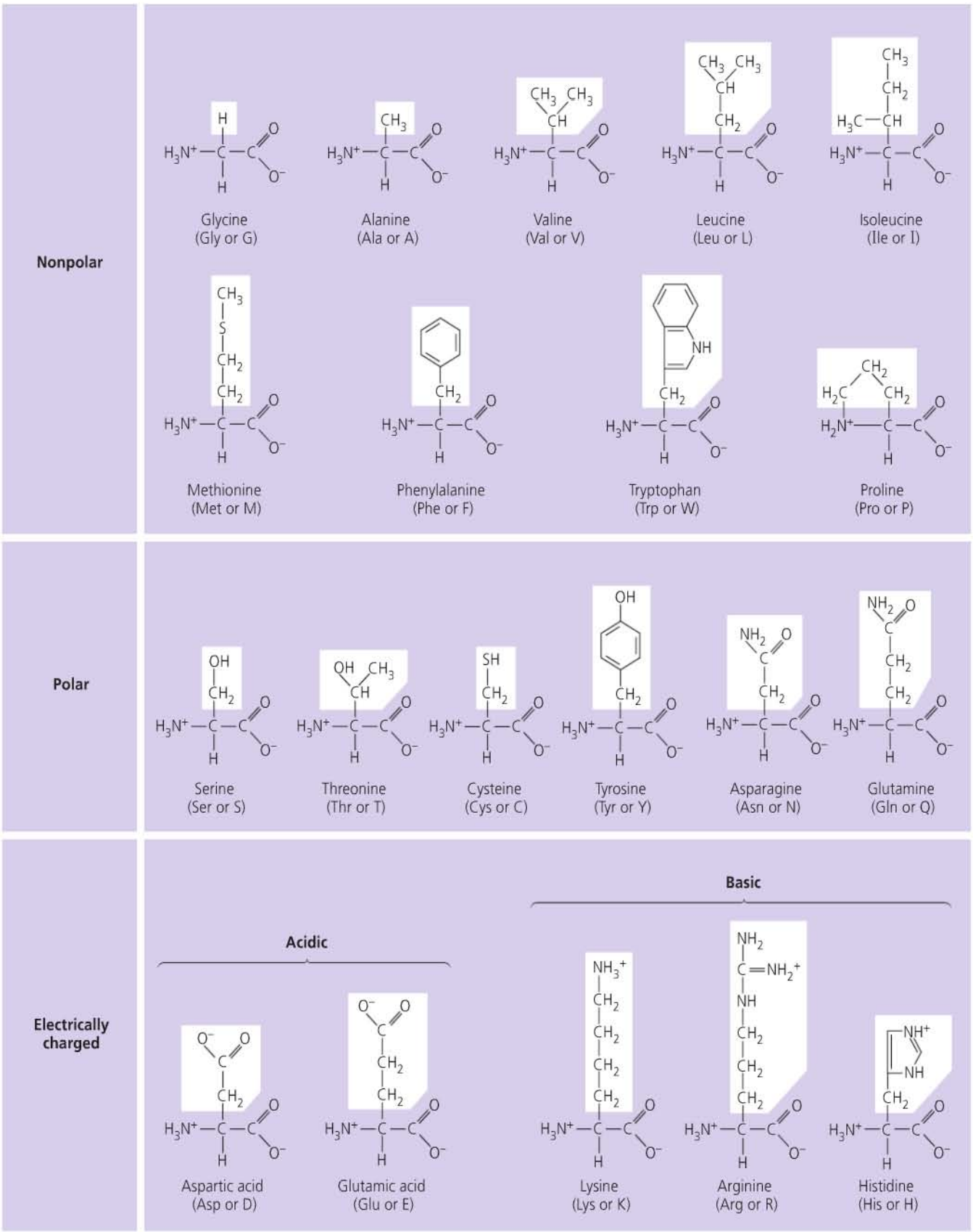


Figure 5.17 shows the 20 amino acids that cells use to build their thousands of proteins. Here the amino and carboxyl



▲ **Figure 5.17 The 20 amino acids of proteins.** The amino acids are grouped here according to the properties of their side chains (R groups), highlighted in white. The amino acids are shown in their prevailing ionic forms at pH 7.2, the pH within a cell. The three-letter and more commonly used one-letter abbreviations for the amino acids are in parentheses. All the amino acids used in proteins are the same enantiomer, called the L form, as shown here (see Figure 4.7).

groups are all depicted in ionized form, the way they usually exist at the pH in a cell. The R group may be as simple as a hydrogen atom, as in the amino acid glycine (the one amino acid lacking an asymmetric carbon, since two of its α carbon's partners are hydrogen atoms), or it may be a carbon skeleton with various functional groups attached, as in glutamine. (Organisms do have other amino acids, some of which are occasionally found in proteins. Because these are relatively rare, they are not shown in Figure 5.17.)

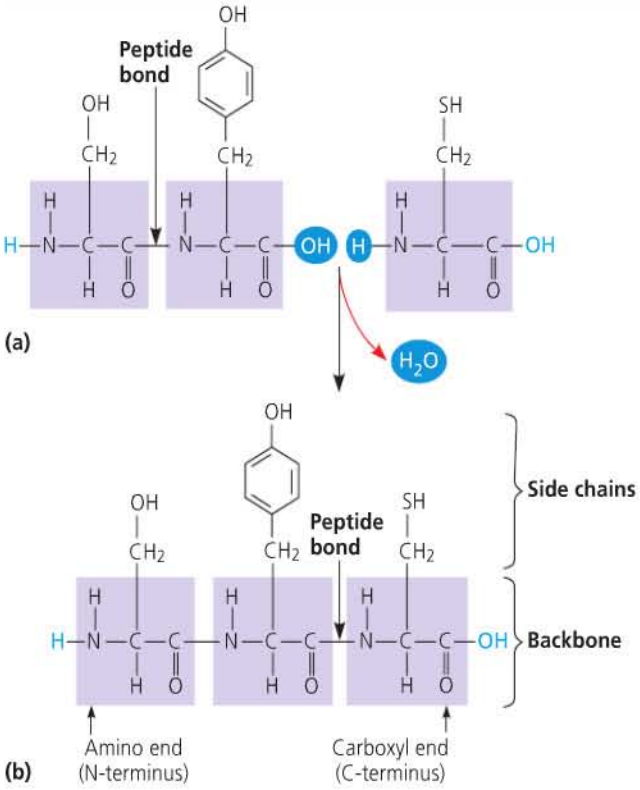
The physical and chemical properties of the side chain determine the unique characteristics of a particular amino acid, thus affecting its functional role in a polypeptide. In Figure 5.17, the amino acids are grouped according to the properties of their side chains. One group consists of amino acids 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 5.18)**. 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. At one end of the polypeptide chain is a free amino group; at the opposite end is a free carboxyl group. Thus, the chain has an amino end (N-terminus) and a carboxyl end (C-terminus). The repeating sequence of atoms highlighted in purple in Figure 5.18b is called the polypeptide backbone. Extending from this backbone are different kinds of appendages, the side chains of the amino acids. Polypeptides range in length from a few monomers to a thousand or more. Each specific polypeptide has a unique linear sequence of amino acids. 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.

Protein Structure and Function

The specific activities of proteins result from their intricate three-dimensional architecture, the simplest level of which is

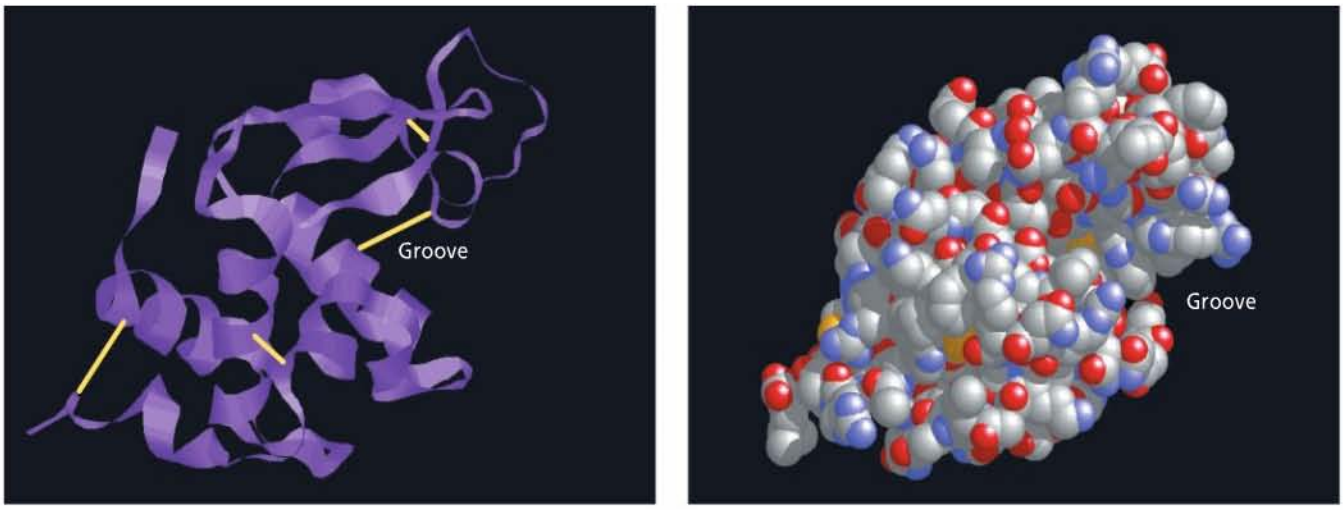


▲ **Figure 5.18 Making a polypeptide chain. (a)** Peptide bonds formed by dehydration reactions link the carboxyl group of one amino acid to the amino group of the next. **(b)** The peptide bonds are formed one at a time, starting with the amino acid at the amino end (N-terminus). The polypeptide has a repetitive backbone (purple) to which the amino acid side chains are attached.

DRAW IT *In (a), circle and label the carboxyl and amino groups that will form the peptide bond shown in (b).*

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 5.19)**. And it



(a) A **ribbon model** shows how the single polypeptide chain folds and coils to form the functional protein. (The yellow lines represent cross-linking bonds between cysteines that stabilize the protein's shape.)

(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 5.19 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.

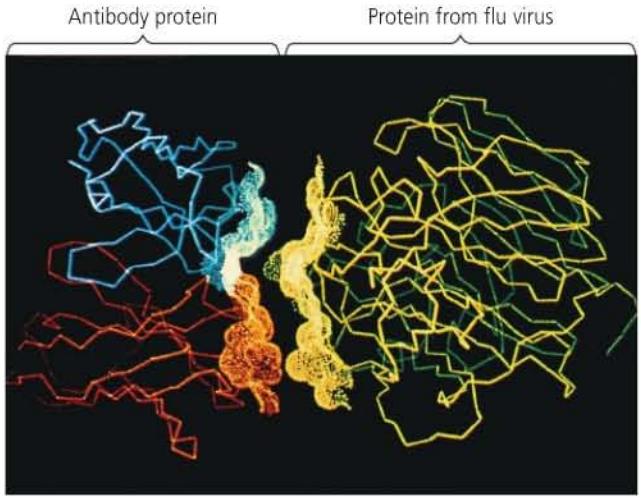
is the amino acid sequence of each polypeptide that determines what three-dimensional structure the protein will have.

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 5.20** 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. A second example is an enzyme, which must recognize and bind closely to its substrate, the substance the enzyme works on (see Figure 5.16). Also, you learned in 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.

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 5.21**, 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.

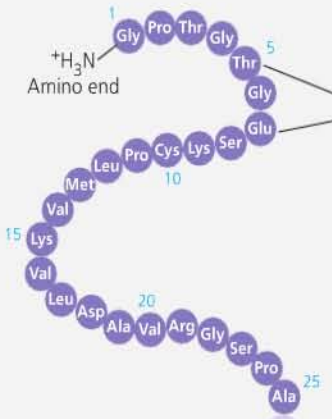


▲ **Figure 5.20 An antibody binding to a protein from a flu virus.** A technique called X-ray crystallography was used to generate a computer model of an antibody protein (blue and orange, left) bound to a flu virus protein (green and yellow, right). Computer software was then used to back the images away from each other, revealing the exact complementarity of shape between the two protein surfaces.

▼ Figure 5.21

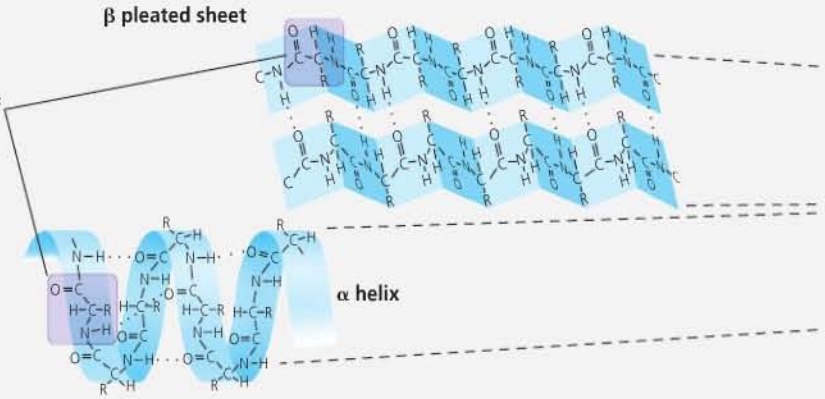
Exploring Levels of Protein Structure

Primary Structure



The **primary structure** of a protein is its unique sequence of amino acids. As an example, let's consider transthyretin, a globular protein found in the blood that transports vitamin A and one of the thyroid hormones throughout the body. Each of the four identical polypeptide chains that together make up transthyretin is 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.

Secondary Structure



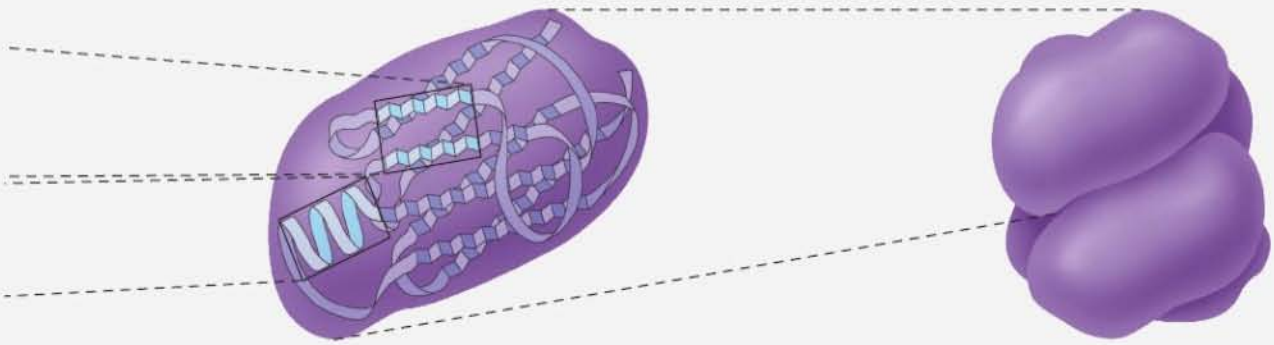
Most proteins have segments of their polypeptide chains repeatedly coiled or folded in patterns that contribute to the protein's overall shape. These coils and folds, collectively referred to as **secondary structure**, are the result of hydrogen bonds between the repeating constituents of the polypeptide backbone (not the amino acid side chains). Both the oxygen and the nitrogen atoms of the backbone are electronegative, with partial negative charges (see Figure 2.16). The weakly positive hydrogen atom attached to the nitrogen atom has an affinity for the oxygen atom of a nearby peptide bond. 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 transthyretin has only one α helix region (see tertiary structure), other globular proteins have multiple stretches of α helix separated by nonhelical regions. 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 regions of the polypeptide chain lying side by side are connected by hydrogen bonds between parts of the two parallel polypeptide backbones. Pleated sheets make up the core of many globular proteins, as is the case for transthyretin, 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.

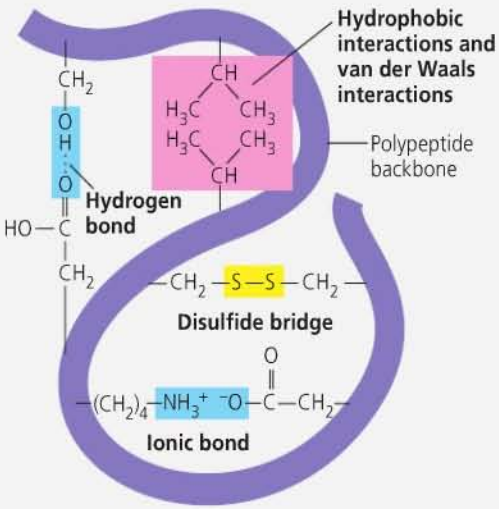
Tertiary Structure

Quaternary Structure

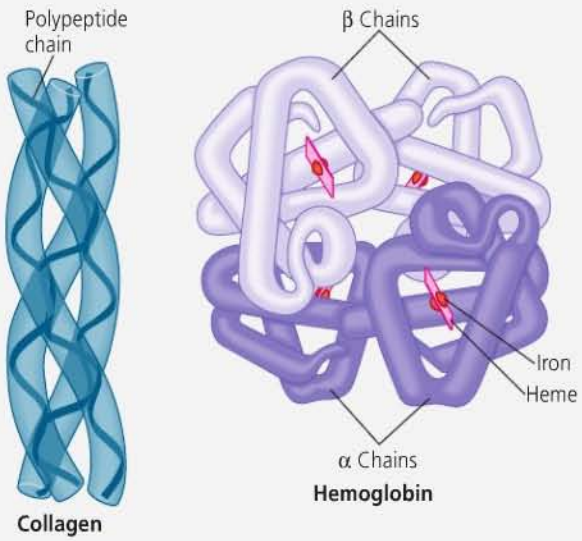


Superimposed on the patterns of secondary structure is a protein's **tertiary structure**, shown above for the transthyretin polypeptide. While secondary structure involves interactions between backbone constituents, tertiary structure is the overall shape of a polypeptide resulting from interactions 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, what we call a hydrophobic interaction is actually caused by the action of water molecules, which exclude nonpolar substances as they form hydrogen bonds with each other and with hydrophilic parts of the protein. 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, but their cumulative effect helps give the protein a unique shape.

The shape of a protein may be reinforced further by covalent bonds called **disulfide bridges**. Disulfide bridges form where two cysteine monomers, amino acids with sulfhydryl groups (—SH) on their side chains (see Figure 4.10), 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 5.19a). All of these different kinds of bonds can occur in one 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 left, which is a fibrous protein that has helical subunits 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). Hemoglobin, the oxygen-binding protein of red blood cells shown below right, is another example of a globular protein with quaternary structure. It consists of four polypeptide subunits, two of one kind ("α chains") and two of another kind ("β chains"). 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.

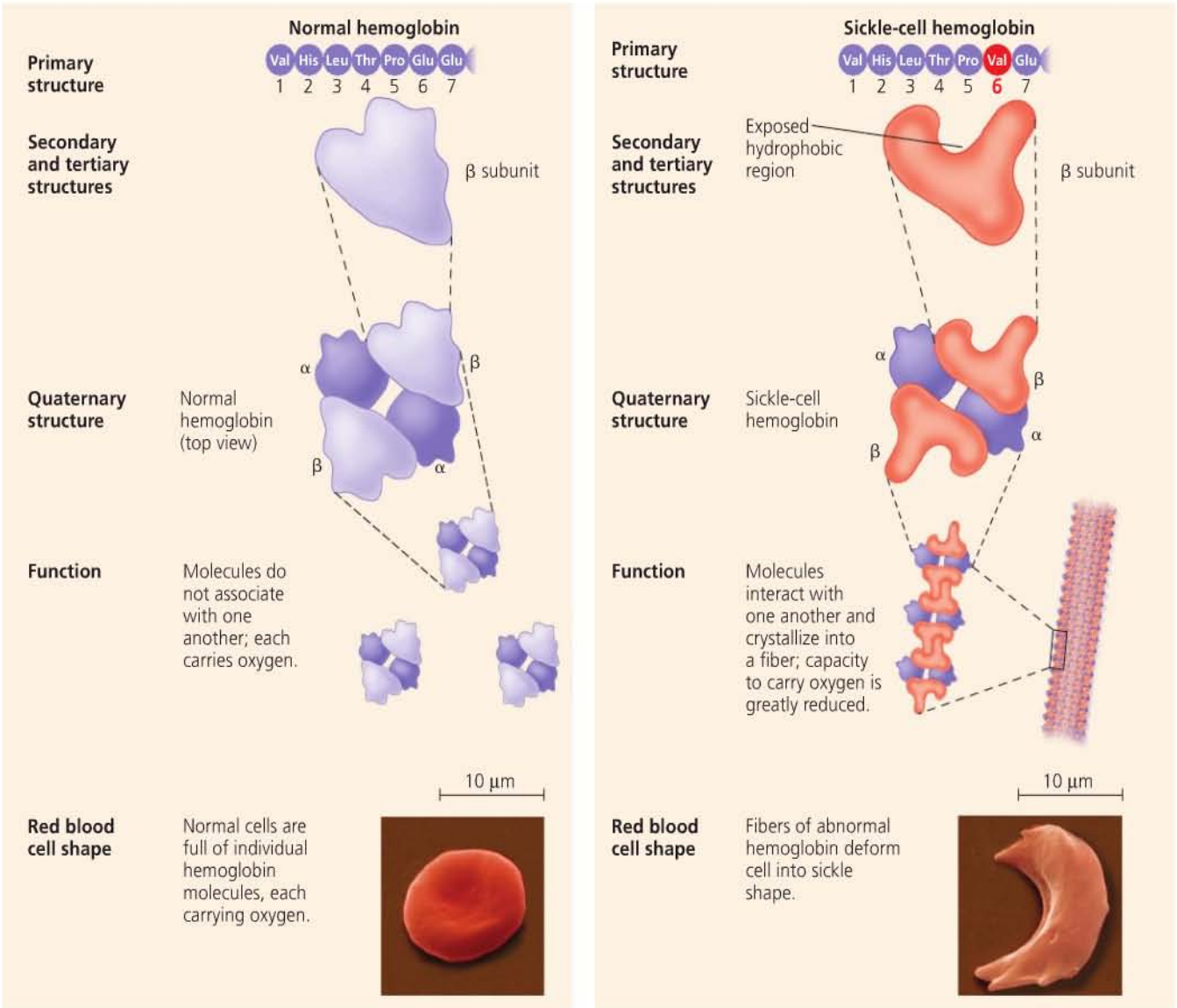


Sickle-Cell Disease: A Change in Primary Structure

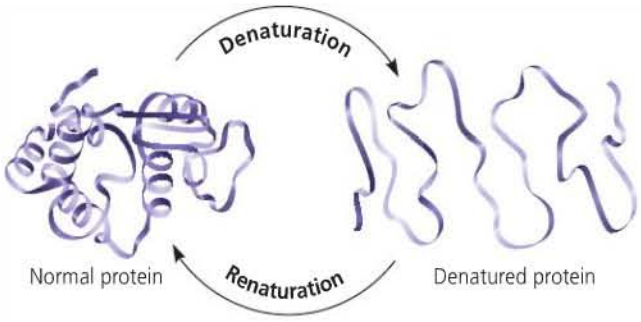
Even a slight change in primary structure can affect a protein's shape and ability to function. For instance, *sickle-cell disease*, an inherited blood disorder, is caused by the substitution of one amino acid (valine) for the normal one (glutamic acid) at a particular position in the primary structure 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 5.22)**. The life of someone with the disease is punctuated by "sickle-cell crises," which occur 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 within the cell. 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 protein may unravel and lose its native shape, a change called **denaturation (Figure 5.23)**. Because it is misshapen, the denatured protein is biologically inactive.



▲ **Figure 5.22 A single amino acid substitution in a protein causes sickle-cell disease.** To show fiber formation clearly, the orientation of the hemoglobin molecule here is different from that in Figure 5.21.



▲ **Figure 5.23 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.

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

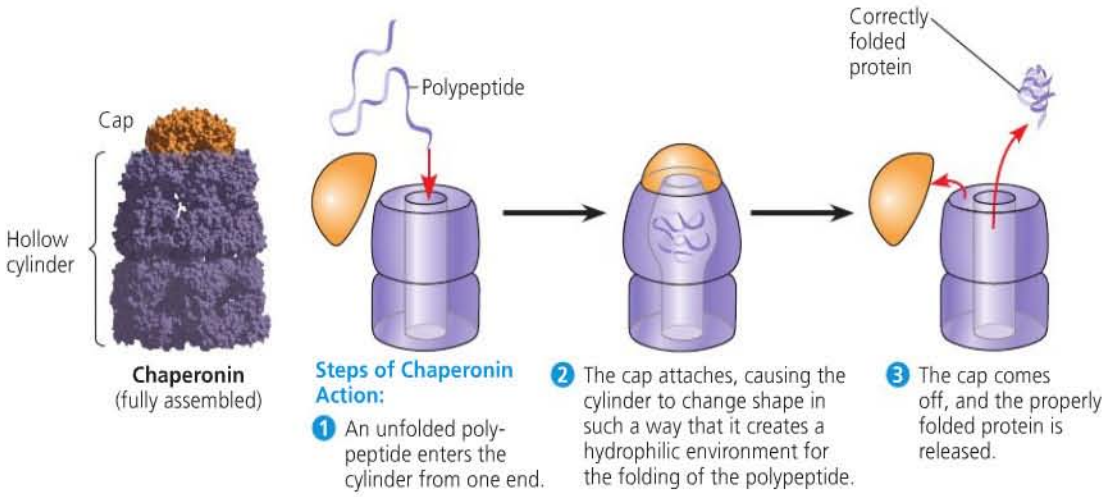
When a protein in a test-tube solution has been denatured by heat or chemicals, it can sometimes return to its functional shape when the denaturing agent is removed. We can conclude that the information for building specific shape is intrinsic to the protein's primary structure. The sequence of amino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

Protein Folding in the Cell

Biochemists now know the amino acid sequences of more than 1.2 million proteins and the three-dimensional shapes of about 8,500. One would think that by correlating the primary structures of many proteins with their three-dimensional structures, it would be relatively easy to discover the rules of protein folding. Unfortunately, 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. Researchers have also discovered **chaperonins** (also called chaperone proteins), protein molecules that assist in the proper folding of other proteins **(Figure 5.24)**. 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 5.24, from the bacterium *E. coli*, is a giant multiprotein complex shaped like a hollow cylinder. The cavity provides a shelter for folding polypeptides.

Misfolding of polypeptides is a serious problem in cells. Many diseases, such as Alzheimer's and Parkinson's, are associated with an accumulation of misfolded proteins. Recently, researchers have begun to shed light on molecular systems in the cell that interact with chaperonins and check whether proper folding has occurred. Such systems either refold the misfolded proteins correctly or mark them for destruction.

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 structures of many other proteins. In a recent example, Roger Kornberg and his colleagues at Stanford University used this method in order to elucidate the structure of RNA polymerase, an enzyme that plays a crucial role in the expression



▶ **Figure 5.24 A chaperonin in action.** The computer graphic (left) shows a large chaperonin protein complex. 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.

of genes (Figure 5.25). Another method now in use is nuclear magnetic resonance (NMR) spectroscopy, which does not require protein crystallization.

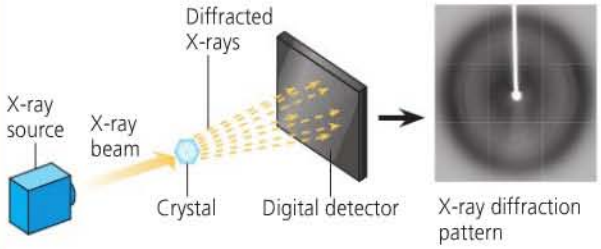
A still newer approach uses bioinformatics (see Chapter 1) to predict the 3-D structures of polypeptides from their amino acid sequences. In 2005, researchers in Austria used computers to analyze the sequences and structures of 129 common plant protein allergens. They were able to classify all of these allergens into 20 families of proteins out of 3,849 families and 65% into just 4 families. This result suggests that the shared structures in these protein families may play a role in generating allergic reactions. These structures may provide targets for new allergy medications.

X-ray crystallography, NMR spectroscopy, and bioinformatics are complementary approaches to understanding protein structure. Together they have also given us valuable hints about protein function.

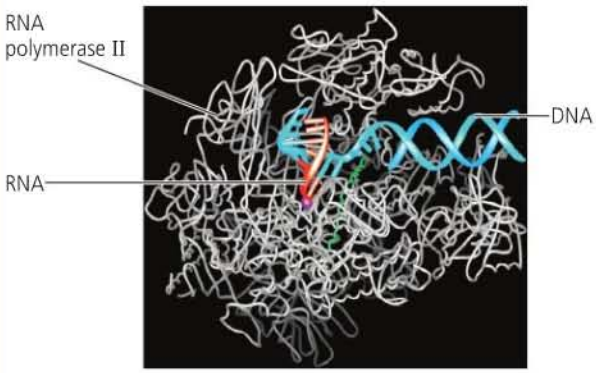
▼ Figure 5.25 Inquiry

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. (Youll 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 green polypeptide spiral in the center?

CONCEPT CHECK 5.4

1. Why does a denatured protein no longer function normally?
2. What parts of a polypeptide chain participate in the bonds that hold together secondary structure? What parts participate in tertiary structure?
3. **WHAT IF?** If a genetic mutation changes primary structure, how might it destroy the protein's function?

For suggested answers, see Appendix A.

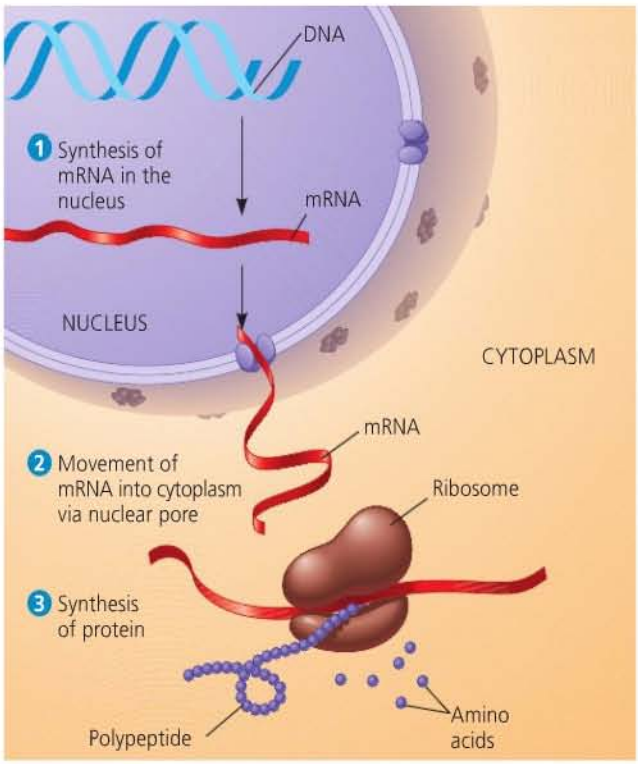
CONCEPT 5.5
Nucleic acids store and transmit hereditary information

If the primary structure of polypeptides determines a protein's shape, what determines primary structure? The amino acid sequence of a polypeptide is programmed by a unit of inheritance known as a **gene**. Genes consist of DNA, a polymer belonging to the class of compounds known as **nucleic acids**.

The Roles of Nucleic Acids

The two types of nucleic acids, **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**, enable living organisms to reproduce their complex components from one generation to the next. Unique among molecules, DNA provides directions for its own replication. DNA also directs RNA synthesis and, through RNA, controls protein synthesis (Figure 5.26).

DNA is the genetic material that organisms inherit from their parents. Each chromosome contains one long DNA molecule, usually carrying several hundred or more genes. When a cell reproduces itself by dividing, its DNA molecules are copied and passed along from one generation of cells to the next. Encoded in the structure of DNA is the information that 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



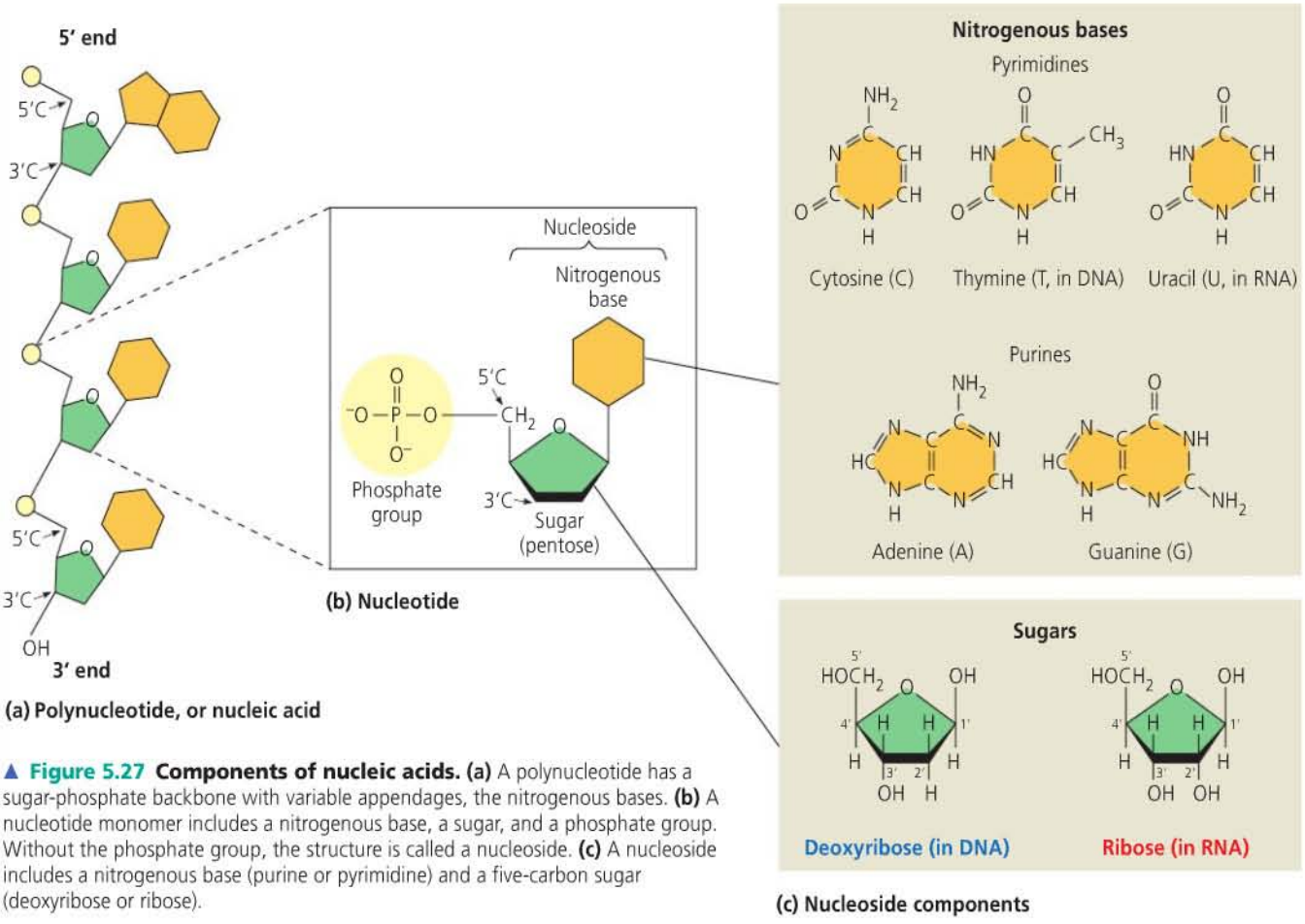
▲ **Figure 5.26 DNA → RNA → protein.** In a eukaryotic cell, DNA in the nucleus programs protein production in the cytoplasm by dictating synthesis of messenger RNA (mRNA). (The cell nucleus is actually much larger relative to the other elements of this figure.)

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 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 5.26). 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 RNA to convey a message from the DNA to ribosomes and other cellular equipment that translate the coded information into amino acid sequences. RNA also plays many other roles in the cell.

The Structure of Nucleic Acids

Nucleic acids are macromolecules that exist as polymers called **polynucleotides** (Figure 5.27a). As indicated by the



▲ **Figure 5.27 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).

name, each polynucleotide consists of monomers called **nucleotides**. A nucleotide is itself composed of three parts: a nitrogenous base, a five-carbon sugar (a pentose), and a phosphate group **(Figure 5.27b)**. The portion of this unit without the phosphate group is called a *nucleoside*.

Nucleotide Monomers

To build a nucleotide, let's first consider the two components of the nucleoside: the nitrogenous base and the sugar **(Figure 5.27c)**. There are two families of nitrogenous bases: pyrimidines and purines. A **pyrimidine** has a six-membered ring of carbon and nitrogen atoms. (The nitrogen atoms tend to take up H^+ from solution, which explains why it's called a nitrogenous *base*.) 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 types of nucleic acid; thymine is found only in DNA and uracil only in RNA.

The sugar connected to the nitrogenous base is **ribose** in the nucleotides of RNA and **deoxyribose** in DNA (see Figure 5.27c). 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. Because the atoms in both the nitrogenous base and the sugar are numbered, the sugar atoms have a prime (′) after the number to distinguish 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. To complete the construction of a nucleotide, we attach a phosphate group to the 5′ carbon of the sugar (see Figure 5.27b). 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 5.27a). 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 the DNA strand 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 AGGTAACTT means one thing, whereas the sequence CGCTTTAAC 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 function in the cell.

The DNA Double Helix

The RNA molecules of cells consist of a single polynucleotide chain like the one shown in Figure 5.27. In contrast, cellular DNA molecules have two polynucleotides that spiral around an imaginary axis, forming a **double helix (Figure 5.28)**. James Watson and Francis Crick, working at Cambridge University, first proposed the double helix as the three-dimensional structure of DNA in 1953. The two sugar-phosphate backbones run in opposite 5′ → 3′ directions from each other, an arrangement 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 polynucleotides, or strands, as they are called, are held together by hydrogen bonds between the paired bases and by van der Waals interactions between the stacked bases. Most DNA molecules are very long, with thousands or even millions of base pairs connecting the two chains. 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 as we traveled the length 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 possible the precise copying of genes that is responsible for inheritance (see Figure 5.28). In preparation for cell division, each of the two strands of a DNA molecule serves as a template to order nucleotides into a new complementary strand. The result is two identical copies of the original double-stranded DNA molecule, which are then distributed to the two daughter cells. Thus, the structure of DNA accounts for its function in transmitting genetic information whenever a cell reproduces.

DNA and Proteins as Tape Measures of 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 a polypeptide chain of human hemoglobin with the corresponding hemoglobin polypeptide in five 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. (Clearly, these changes do not make the protein nonfunctional.) Molecular biology has added a new tape measure to the toolkit biologists use to assess evolutionary kinship.

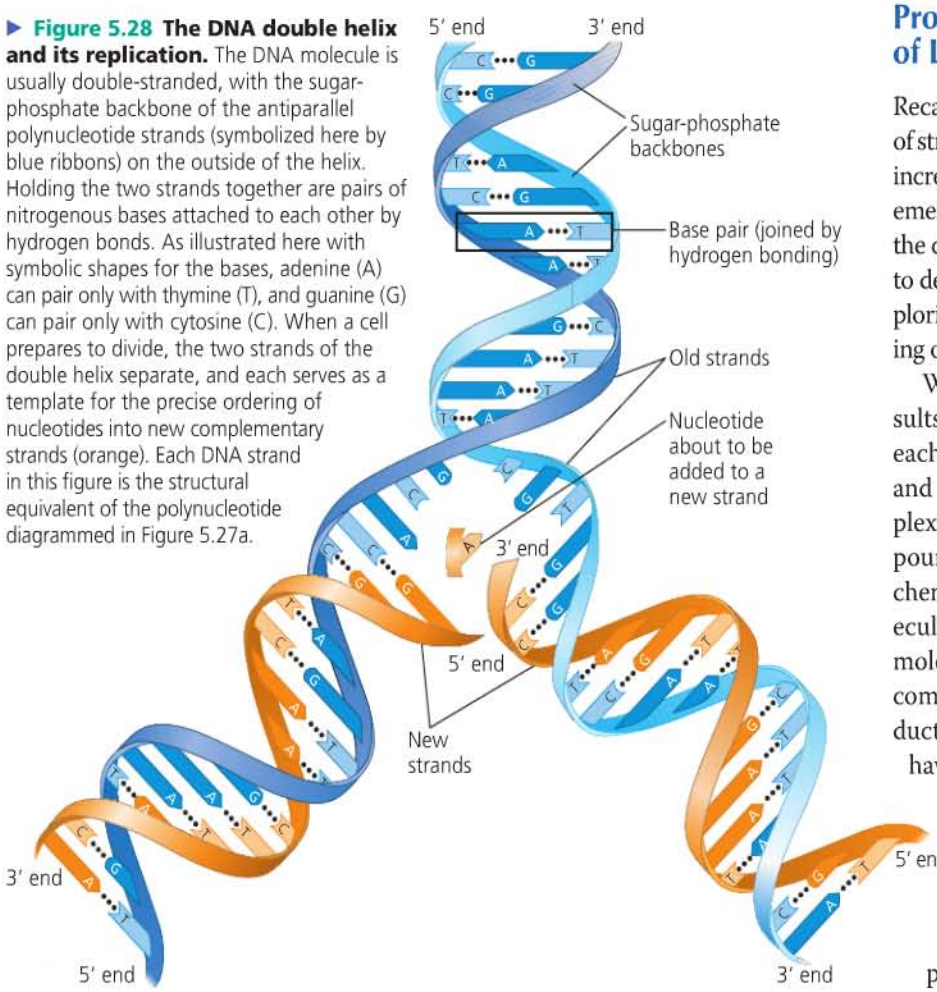
CONCEPT CHECK 5.5

1. Go to Figure 5.27a and number all the carbons in the sugars for the top three nucleotides; circle the nitrogenous bases and star the phosphates.
2. In a DNA double helix, a region along one DNA strand has this sequence of nitrogenous bases: 5′-TAGGCCT-3′. Write down this strand and 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′

   Draw the two strands; circle and label the mismatched bases. (b) If the modified top strand is replicated, what would its matching strand be?

For suggested answers, see Appendix A.



▶ **Figure 5.28 The DNA double helix and its replication.** The DNA molecule is usually double-stranded, with the sugar-phosphate backbone 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). When a cell prepares to divide, the two strands of the double helix separate, and each serves as a template for the precise ordering of nucleotides into new complementary strands (orange). Each DNA strand in this figure is the structural equivalent of the polynucleotide diagrammed in Figure 5.27a.

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 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.

Chapter 5 Review

MEDIA Go to the Study Area at **www.masteringbio.com** for BioFlix 3-D Animations, MP3 Tutors, Videos, Practice Tests, an eBook, and more.

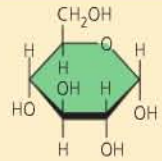
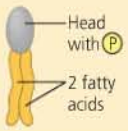
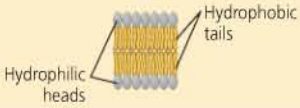
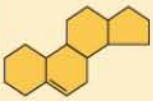
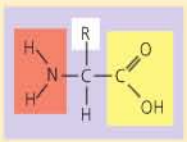
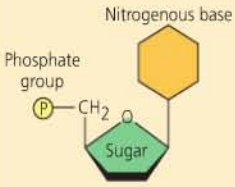
SUMMARY OF KEY CONCEPTS

CONCEPT 5.1
Macromolecules are polymers, built from monomers (pp. 68–69)

▶ **The Synthesis and Breakdown of Polymers** 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.

▶ **The Diversity of Polymers** An immense variety of polymers can be built from a small set of monomers.

MEDIA
Activity Making and Breaking Polymers

Large Biological Molecules	Components	Examples	Functions
Concept 5.2 **Carbohydrates serve as fuel and building material (pp. 69–74)** MEDIA Activity Models of Glucose; Activity Carbohydrates	 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)** MEDIA Activity Lipids		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 have many structures, resulting in a wide range of functions (pp. 77–86)** MEDIA MP3 Tutor Protein Structure and Function; Activity Protein Functions; Activity Protein Structure; Biology Labs On-Line HemoglobinLab	 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 and transmit hereditary information (pp. 86–89)** MEDIA Activity Nucleic Acid Functions; Activity Nucleic Acid Structure	 Nucleotide monomer	DNA:  • Sugar = deoxyribose • Nitrogenous bases = C, G, A, T • Usually double-stranded	Stores all hereditary information
		RNA:  • Sugar = ribose • Nitrogenous bases = C, G, A, U • Usually single-stranded	Carries protein-coding instructions from DNA to protein-synthesizing machinery

TESTING YOUR KNOWLEDGE

SELF-QUIZ

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

2. 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. $C_{60}H_{111}O_{51}$

3. The enzyme amylase can break glycosidic linkages between glucose monomers only if the monomers are 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

4. 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 saturated fats having the same number of carbon atoms.
   e. They have fewer fatty acid molecules per fat molecule.

5. 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.

6. 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.

7. 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 between deoxyribose sugars 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.

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** Draw the polynucleotide strand in Figure 5.27a and label the bases G, T, C, and T, starting from the 5′ end. Now draw the complementary strand of the double helix, 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.

For Self-Quiz answers, see Appendix A.

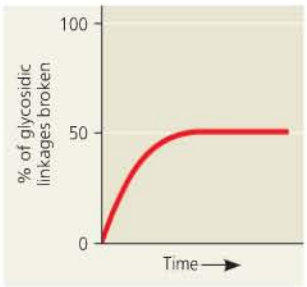
MEDIA Visit the Study Area at **www.masteringbio.com** for a Practice Test.

EVOLUTION CONNECTION

10. Comparisons of amino acid sequences can shed light on the evolutionary divergence of related species. Would you expect all the proteins of a given set of living species to show the same degree of divergence? Why or why not?

SCIENTIFIC INQUIRY

11. During the Napoleonic Wars in the early 1800s, there was a sugar shortage in Europe because supply ships could not enter blockaded harbors. To create artificial sweeteners, German scientists hydrolyzed wheat starch. They did this by adding hydrochloric acid to heated starch solutions, breaking some of the glycosidic linkages between the glucose monomers. The graph here shows the percentage of glycosidic linkages broken over time. Why do you think consumers found the sweetener to be less sweet than sugar? Sketch a glycosidic linkage in starch using Figures 5.5a and 5.7b for reference. Show how the acid was able to break this bond. Why do you think the acid broke only 50% of the linkages in the wheat starch?



Biological Inquiry: A Workbook of Investigative Cases Explore large biological molecules further with the case "Picture Perfect."

SCIENCE, TECHNOLOGY, AND SOCIETY

12. 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 that is required to succeed in competition? Explain.

UNIT 2

The Cell

AN INTERVIEW WITH

Paul Nurse

British biologist Sir Paul Nurse shared a Nobel Prize in 2001 with Leland H. Hartwell and R. Timothy Hunt for groundbreaking discoveries about how the eukaryotic cell controls its reproduction. Educated at the Universities of Birmingham and East Anglia, Dr. Nurse was a professor at the University of Oxford and a researcher at the Imperial Cancer Research Fund (now called Cancer Research UK), eventually heading up the latter organization. In 2003, he left the United Kingdom for Rockefeller University in New York City, where he serves as president while also running an active research laboratory.

Tell us about Rockefeller University.

Rockefeller is basically a research institute that gives Ph.D.'s. We have no departments; instead, we have between 70 and 80 independent laboratories. Our sole purpose is to do the highest-quality research relevant to biomedicine. We have physicists, chemists, and mathematicians, as well as biologists and biomedical scientists, but they all tend to be interested in biological problems. It's a wonderfully stimulating environment— and rather anarchic, but I'm happy with that. In my role as president, I focus very much on having the best people around, and then just letting them loose to investigate what they want. I'm very privileged to be here.

How did you first become interested in science, and biology in particular?

When I was a child, I had a long walk to my school through a couple of parks, and I liked to look at the plants, the birds, and the insects. I remember wondering why certain plants seemed to have big leaves when they grew in the shade and small leaves when they grew in sunlight. I also had an interest in astronomy, which I still have. So the origin of my interest in science was really through natural history.

I wasn't the greatest of students at school. My parents were working class, and we didn't have many books at home. It took me quite a while to get up to speed. But I had a tremendous curiosity to learn about the world around me. Then, as a teenager, I had a science teacher who really encouraged me, named Keith Neil. I did a number of biology projects with him—a fruit fly project, censuses of butterflies and beetles, and a trout egg-laying project in the lab. After I got the Nobel Prize, I nominated him for a mentoring award and did a TV program with him. I think my interest in biology was also influenced by a sense that there were so many unanswered questions in biology that even an ordinary mortal could contribute to the field in some way!

At university, I was at first going to concentrate on ecology and evolution, but I ended up going for subjects that seemed easier to study in a laboratory—cell biology and genetics.

What led you to study the cell cycle?

As a graduate student, I asked myself, "What's really important in biology?" I thought it would be important to focus on the features that distinguish life from nonlife, and a key one of those is the ability of an organism to reproduce itself. You see reproduction in its simplest form in the division of a cell because the cell is the simplest unit of life.

What did you want to find out about cell division?

The cell cycle is a series of events from the "birth" of a cell to its later division into two cells. I wanted to understand how the different events in the cell cycle were coordinated—what controlled the progression of cell-cycle events. In fact, I have ended up spending most of my life trying to answer that question!

There were two alternative ideas about how the cell cycle works: One was that the cell simply proceeds automatically through the events in the pathway, from A to B to C to D, and so forth. Another was that a smaller number of rate-limiting steps within the pathway serve as control points. These control points would determine how rapidly the cell cycle progressed.

How did you proceed to figure out which hypothesis was correct?

My inspiration came from Lee Hartwell, who had used genetics in budding yeast (the yeast used by bakers and brewers) to find mutants with defects in the cell cycle. This is the classical genetic approach. You look for what stops a system from working to tell you how it should be working. I used Lee's work as a template, but I used a different sort of yeast, called fission yeast, which is actually not closely related to budding yeast. Like Lee, I isolated mutants with defects in genes that defined particular steps in the cell-cycle process. We called them *cdc* genes—*cdc* for *c*ell *d*ivision *c*ycle.

In reproducing, a cell of fission yeast first grows to twice its starting size and then divides in two. I looked for mutants that couldn't divide, where the cells just got bigger and bigger. Finding such mutants told us that the genes defective in these cells were necessary for the cells to divide. And when you find a number of different kinds of mutants that are defective in a process, the obvious interpretation is that the process is a pathway of events, each of which is controlled by one or more genes. So the cell cycle could have worked like a straightforward pathway, with each step simply leading to the next.

But one day I happened to notice a different sort of mutant under the microscope: yeast cells that were dividing but at an unusually small size. I realized that if a cell went through the cell cycle faster than normal, it would reach the end of the cell cycle before it had doubled in size. The mutant I'd spotted had only a single mutation, in a single gene, but it was sufficient to make the whole cell cycle go faster. That meant there had to be at least some major rate-limiting steps in the cell cycle, because if they didn't exist you couldn't make the cell cycle go faster. So the second hypothesis seemed to be correct.

All of that came out of just looking at those small mutant cells for five minutes. I could say that most of the next twenty years of my career was based on those five minutes!

It's essential to add that many other scientists have contributed to the field of cell cycle regulation, working in many different places. In

addition to Lee Hartwell, who is in Seattle, these people included Yoshio Masui in Toronto and Jim Maller in Colorado, who both worked with frog eggs, and my longtime friend Tim Hunt, in England, who studied the eggs of sea urchins.

You focused in on a mutant with a defect in a gene you called *cdc2*. What does this gene do?

It turned out that the *cdc2* gene codes for an enzyme called a protein kinase. Protein kinases are heavily used by cells as a means of regulating what other proteins do. There are hundreds of different kinds of these enzymes—over 500 in human cells, for example. What they do is phosphorylate other proteins: They take phosphates from ATP molecules and transfer them to proteins. Phosphate groups are big lumps of negative charge, and they can change the shape of a protein and therefore its properties. So showing that *cdc2* coded for a protein kinase was important in identifying protein phosphorylation as a key regulatory mechanism in the cell cycle. Eventually we showed that, in fission yeast, the *cdc2* protein kinase is used fairly early in the cell cycle, where it controls the replication of DNA, and then again later in the cycle, when the replicated chromosomes are ready to separate from each other in mitosis. This is soon before the cell divides in two.

Do similar enzymes control the cell cycle in other kinds of organisms?

A postdoc in my lab, Melanie Lee, was able to track down the human equivalent of the yeast *cdc2* gene, by showing that it was able to substitute for a defective *cdc2* gene in a yeast cell. After we put the human gene in a yeast cell that had a defective *cdc2* gene, the yeast cell was able to divide normally!

It turns out that the Cdc2 protein and many others involved in cell-cycle regulation are very similar in all eukaryotic organisms. What this has to mean is that this system of controlling cell division must have evolved very soon after eukaryotic cells first appeared on the planet, and that it was so crucial to cell survival that it remained unchanged. We have found the *cdc2* gene, for example, in every eukaryotic organism we've looked at.

What is the medical relevance of research on the cell cycle?

The main medical relevance is to cancer, because cancer occurs when cells grow and divide out of control. Now, growth and division is a good thing in the right place at the right time; it's how a fertilized human egg develops into a baby and how wounds in the body are repaired. But if you start getting growth and division in the wrong place at the wrong time, then you can get tumors that destroy the function of the organs or tissue in which they are located.

But maybe, in the end, the more important connection to cancer has to do with what's called genome stability. You have to precisely replicate all your genes in every cell cycle and then separate them precisely into the two progeny cells—that's what the cell cycle is all about. If there are mistakes, if the DNA does not replicate properly or the chromosomes don't separate properly, you end up generating genome instability, in which the number of chromosomes may be altered and parts of chromosomes rearranged. Such changes can lead to cancer. So understanding how genome stability is maintained is crucial for understanding how cancer arises.

What is your approach to mentoring young scientists in your lab? And what about collaboration with other labs?

I've always run a pretty disorganized lab. With students and postdocs, I look upon my job as not so much directing them as helping them follow their own interests, although I do try to keep them from falling into too many elephant traps. The lab is a bit inefficient, to be honest, because we're constantly starting new projects. But since people take these projects away with them when they leave, this practice helps the field expand very fast.

My collaboration with people outside my own lab is mostly in the form of talking. I find I think much better when I can bounce around ideas with other people. This sort of conversation is especially useful when there is a very honest and open relationship—it's good to be sufficiently comfortable with someone to be able to say, "What you just said was stupid." I've benefited from such frank discussions for decades with Tim Hunt, for example, but we've never published a paper together. Tim discovered another kind of cell-cycle control protein, called cyclin, which works in partnership with protein kinases like Cdc2. The protein kinases we've been talking about in this interview are therefore called cyclin-dependent kinases, or CDKs.

What responsibilities do scientists have toward society?

I always say that scientists need to have a "license to operate." We have to earn that license; we cannot assume society will be pro-science. When I was younger, I used to think: Well, I'm doing this because I'm curious, and science should be supported because it's an important cultural endeavor, something like art or music. But we have to realize that if we can justify science only in cultural terms, science budgets will plummet. The public and their governmental representatives want to use scientific discoveries to benefit humankind, and that's completely reasonable. I think it's critical that scientists communicate effectively with the public so that we can influence policymakers in government. Most importantly, we scientists have to listen to the public and understand how they see the issues; we need to have a dialogue with the public. Without that dialogue, we just don't know what misunderstandings are out there. I think we need more grassroots involvement by scientists.

. . . this system of controlling cell division must have evolved very soon after eukaryotic cells first appeared on the planet.

InquiryinAction

Learn about an experiment by Paul Nurse and colleagues in Inquiry Figure 12.16 on page 240.

Paul Nurse and Jane Reece