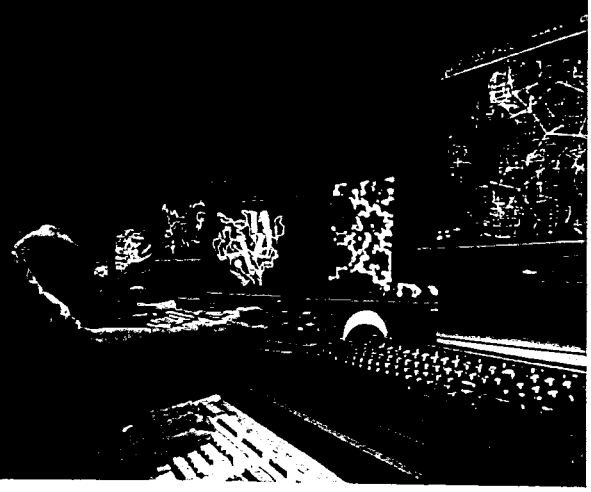
The Structure and Function of Macromolecules



▲ **Figure 5.1 Scientists working with computer models of proteins.**

Key Concepts

- **5.1** Most 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

We have seen how the concept of emergent properties applies to water and relatively simple organic molecules. Each type of small molecule has unique properties arising from the orderly arrangement of its atoms. Another level in the hierarchy of biological organization is reached when small organic molecules are joined inside cells, forming larger molecules. The four main classes of large biological molecules are carbohydrates, lipids, proteins, and nucleic acids. Many of these cellular molecules are, on the molecular scale, huge. For example, a protein may consist of thousands of covalently connected atoms that form a molecular colossus with a mass of over 100,000 daltons. Biologists use the term **macromolecule** for such giant molecules.

Considering the size and complexity of macromolecules, it is remarkable that biochemists have determined the detailed structures of so many of them **(Figure 5.1)**. The architecture of a macromolecule helps explain how that molecule works. Life's large molecules are the main subject of this chapter. For these molecules, as at all levels in the biological hierarchy, form and function are inseparable.

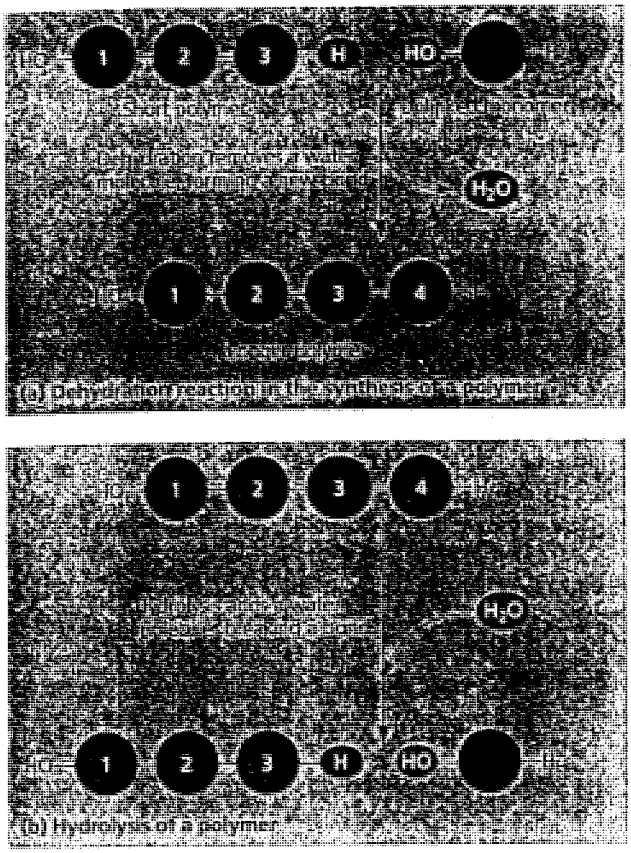
Concept 5.1

Most macromolecules are polymers, built from monomers

The large molecules 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 small 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 polymeric macromolecules differ in the nature of their monomers, but the chemical mechanisms by which cells make and break 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 called a **condensation reaction**, specifically a **dehydration reaction**, because the molecule lost is water **(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). In making a polymer, this reaction is repeated as monomers are added to the chain one by one. The cell must expend energy to carry out these dehydration reactions, and the process oc-



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

curs only with the help of enzymes, specialized proteins that speed up chemical reactions in cells.

Polymers are disassembled to monomers by **hydrolysis**, a process that is essentially the reverse of the dehydration reaction **(Figure 5.2b)**. Hydrolysis means to break with water (from the Greek *hydro*, water, and *lysis*, break). Bonds between monomers are broken by the addition of water molecules, a hydrogen from the water attaching to one monomer and a hydroxyl group attaching to the adjacent monomer. An example of hydrolysis working in 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 polymers that differ from the ones that were digested. The new polymers 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 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 an enormous variety of polymers from such a limited list of monomers is analogous to constructing hundreds of thousands of words from only 26 letters of the alphabet. The key is arrangement—variation in the linear sequence that the units follow. However, this analogy falls far short of describing the great diversity of macromolecules, because most biological polymers are much longer than 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.

We are now ready to investigate the specific structures and functions of the four major classes of organic compounds found in cells. For each class, we will see that the large molecules have emergent properties not found in their individual building blocks.

Concept Check

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 10 monomers long?
3. After you eat a slice of apple, which reactions must occur for the amino acid monomers in the protein of the apple to be converted into proteins in your body?

For suggested answers, see Appendix A.

Concept

Carbohydrates serve as fuel and building material

Carbohydrates include both sugars and the polymers of sugars. The simplest carbohydrates are the monosaccharides, or single sugars, also known as simple sugars. Disaccharides are double sugars, consisting of two monosaccharides joined by a condensation reaction. The carbohydrates that are macromolecules are polysaccharides, polymers composed of many sugar building blocks.

Sugars

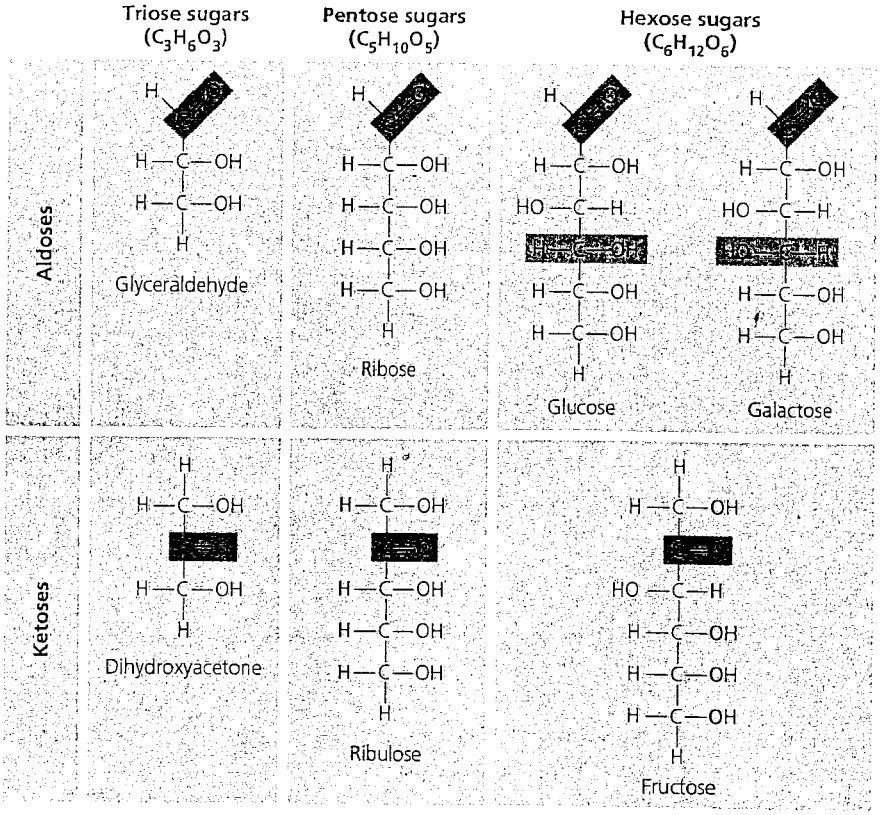
Monosaccharides (from the Greek *monos*, single, and *sacchar*, sugar) generally have molecular formulas that are some multiple of the unit CH_2O **(Figure 5.3)**. Glucose ($C_6H_{12}O_6$), the most common monosaccharide, is of central importance in the chemistry of life. In the structure of glucose, 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 from Chapter 4 that an asymmetric carbon is a carbon attached to four different kinds of partners.) 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.

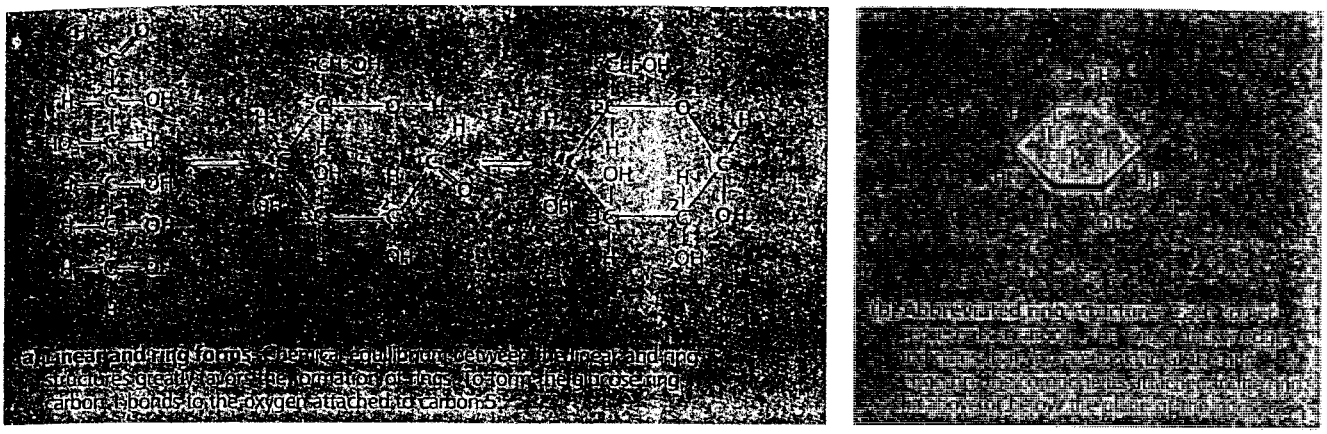
Although it is convenient to draw glucose with a linear carbon skeleton, this representation is not completely accurate. In aqueous solutions, glucose molecules, as well as most other sugars, form rings **(Figure 5.4)**.

Monosaccharides, particularly glucose, are major nutrients for cells. In the process known as cellular respiration, cells extract the energy stored in glucose molecules. Not only are simple sugar molecules a major fuel for cellular work, but their carbon skeletons 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.

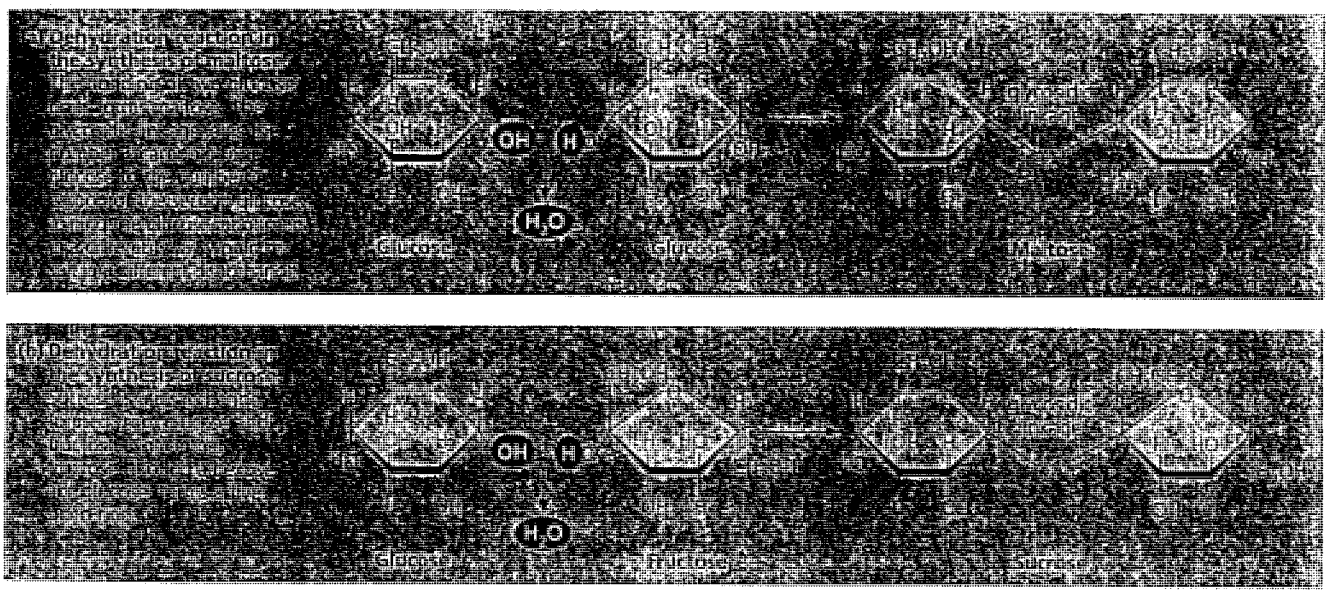
A **disaccharide** consists of two monosaccharides joined by a **glycosidic linkage**, a covalent bond formed between two monosaccharides by a dehydration reaction. For example, maltose is a disaccharide formed by the linking of two molecules of glucose **(Figure 5.5a)**. Also known as malt sugar, maltose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



▲ Figure 5.4 Linear and ring forms of glucose.



▲ Figure 5.5 Examples of disaccharide synthesis.

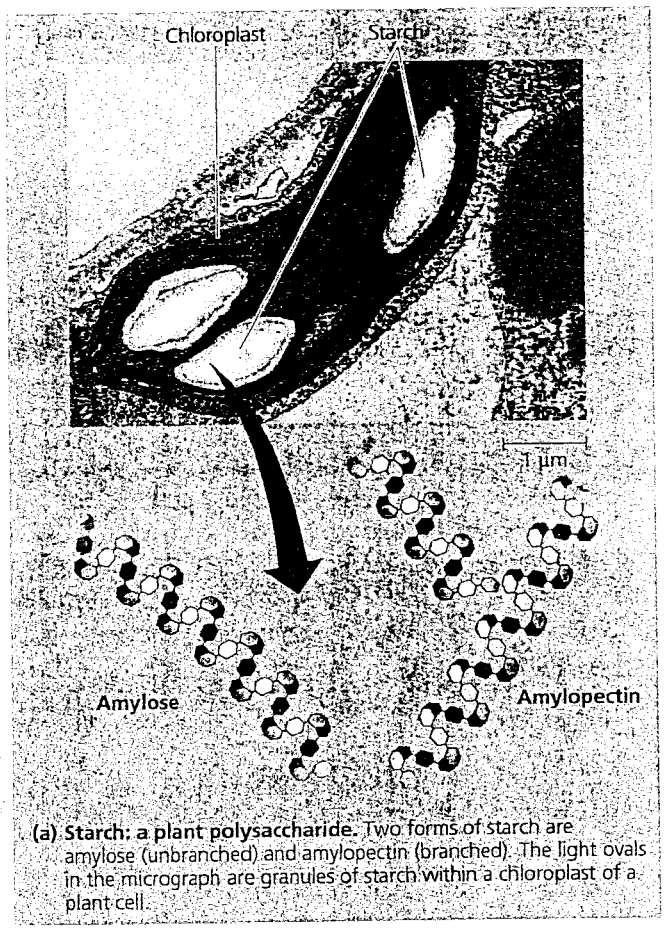
Polysaccharides

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.

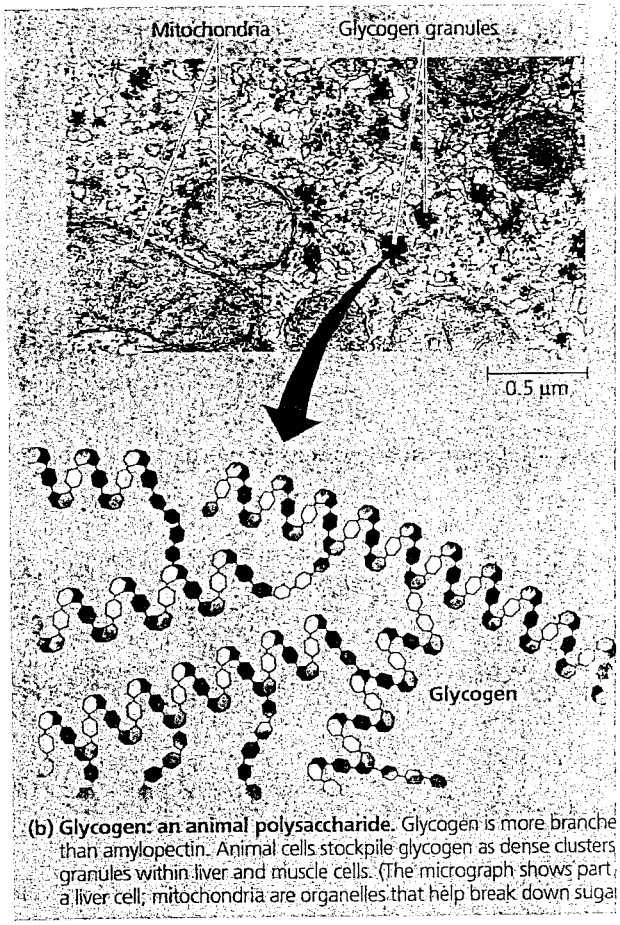
Storage Polysaccharides

Starch, a storage polysaccharide of plants, is a polymer consisting entirely of glucose monomers. Most of these monomers 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 bonds makes the polymer helical. The simplest form of starch, amylose, is unbranched. Amylopectin, a more complex form of starch, is a branched polymer with 1–6 linkages at the branch points.

Plants store starch as granules within cellular structures called plastids, which include chloroplasts **(Figure 5.6a)**. 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



(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 branche than amylopectin. Animal cells stockpile glycogen as dense clusters granules within liver and muscle cells. (The micrograph shows part a liver cell; mitochondria are organelles that help break down suga

▲ **Figure 5.6 Storage polysaccharides of plants and animals.** These examples, starch and glycogen, are composed entirely of glucose monomers, represented here by hexagons. Due to their molecular structure, the polymer chains tend to form helices.

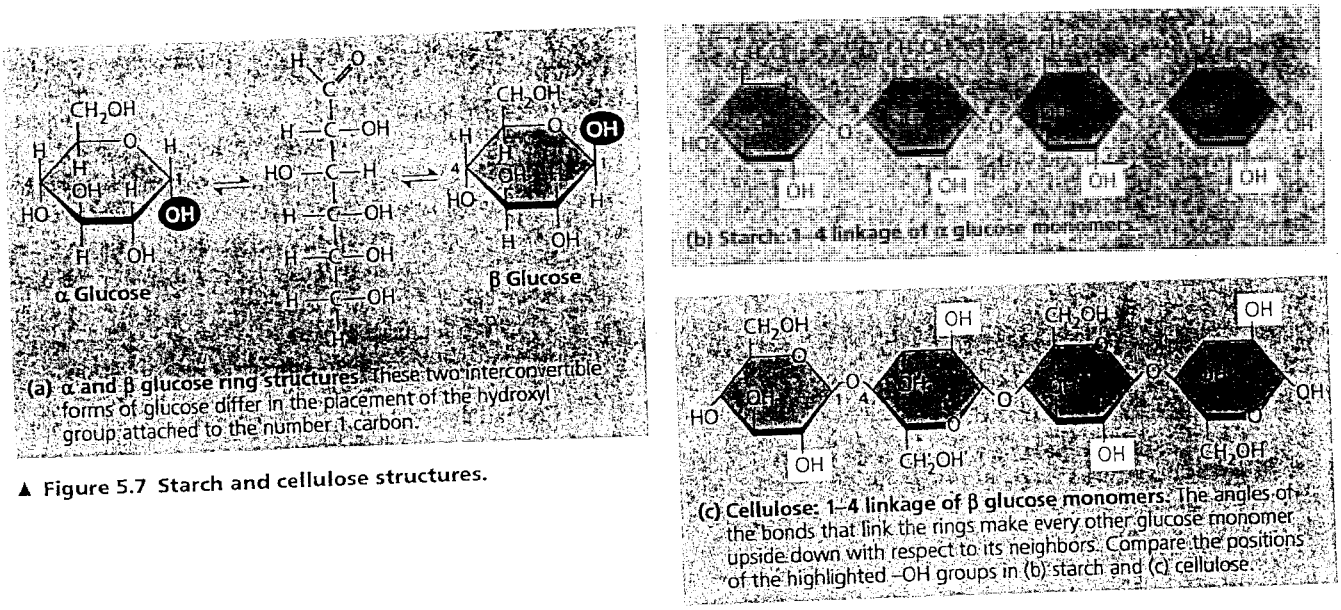
available as a nutrient for cells. Potato tubers and grains—the fruits of wheat, corn, rice, and other grasses—are the major sources of starch in the human diet.

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.

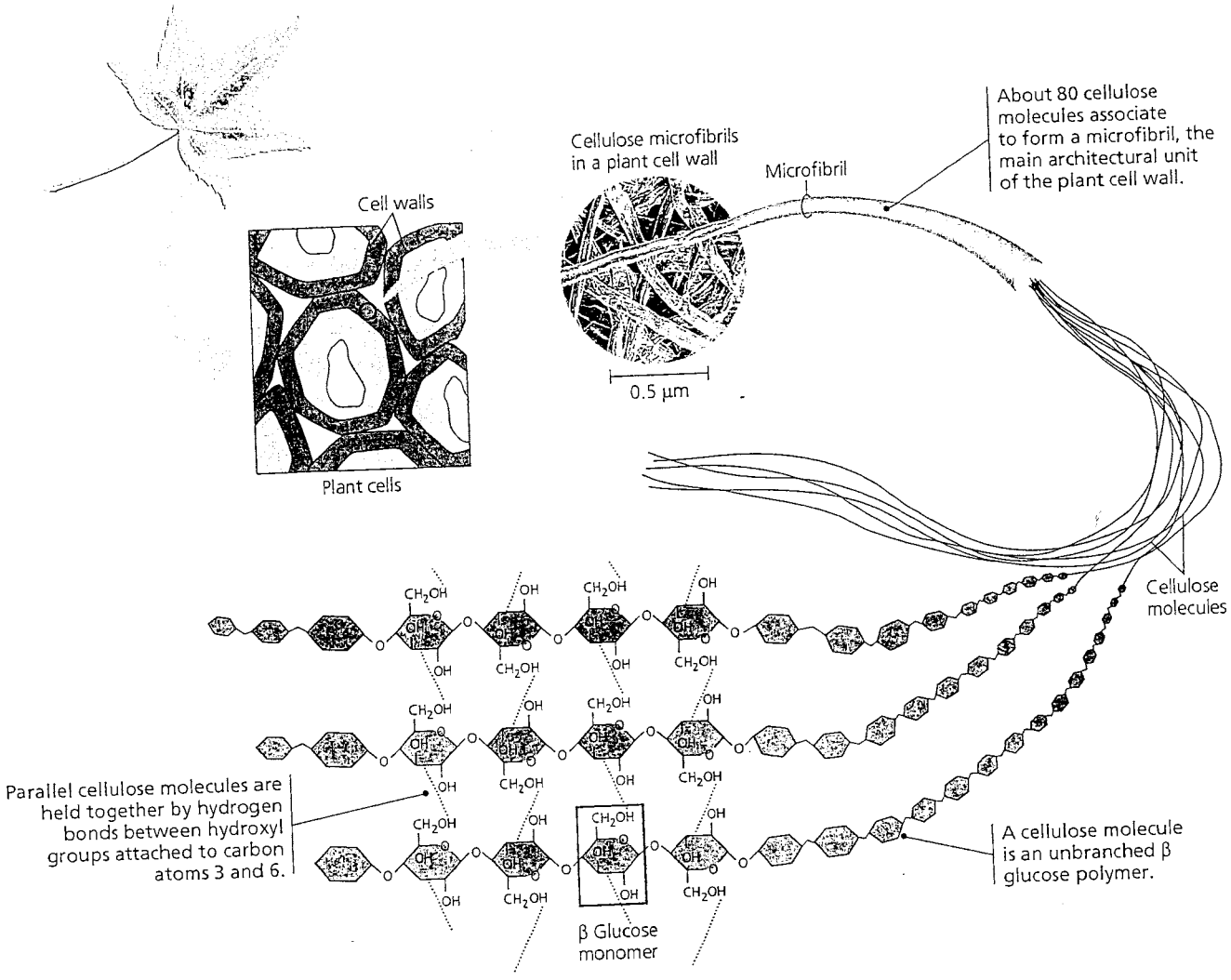
Structural Polysaccharides

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^{11} (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 fer. The difference is based on the fact that there are actu two slightly different ring structures for glucose **(Figure 5.** When glucose forms a ring, the hydroxyl group attache the number 1 carbon is positioned either below or above plane of the ring. These two ring forms for glucose are ca alpha (α) and beta (β), respectively. In starch, all the glu monomers are in the α configuration **(Figure 5.7b)**, arrangement we saw in Figures 5.4 and 5.5. In contrast glucose monomers of cellulose are all in the β configura making every other glucose monomer upside down witl spect to its neighbors **(Figure 5.7c)**.

The differing glycosidic linkages in starch and cellulose the two molecules distinct three-dimensional shapes. Wh a starch molecule is mostly helical, a cellulose molecu straight (and never branched), and its hydroxyl groups free to hydrogen-bond with the hydroxyls of other cellu molecules lying parallel to it. In plant cell walls, parallel c lose molecules held together in this way are grouped into called microfibrils **(Figure 5.8)**. These cable-like microf



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



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

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

are a strong building material for plants as well as for humans, who use wood, which is rich in cellulose, for lumber.

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 microbes can digest cellulose, breaking it down to glucose monomers. A cow harbors cellulose-digesting bacteria in the rumen, the first compartment in its stomach **(Figure 5.9)**. The bacteria 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 microbes 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, 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

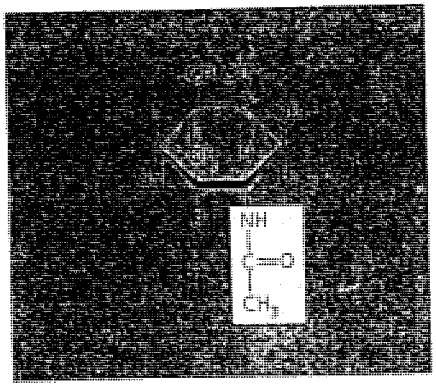
1. Write the formula for a monosaccharide that has three carbons.
2. A dehydration reaction joins two glucose molecules to form maltose. The formula for glucose is $C_6H_{12}O_6$. What is the formula for maltose?
3. Compare and contrast starch and cellulose.

For suggested answers, see Appendix A.

Concept

Lipids are a diverse group of hydrophobic molecules

Lipids are the one class of large biological molecules that doe not consist of polymers. The compounds called **lipids** ar grouped together because they share one important trai



(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.

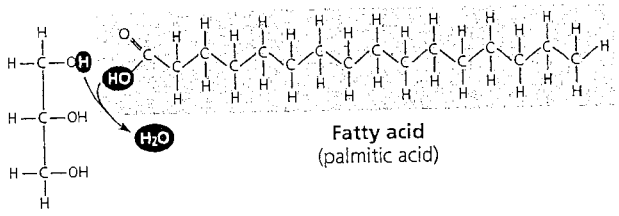
(c) Chitin is used to make a strong and flexib surgical thread that decomposes after the wound or incision heals.

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

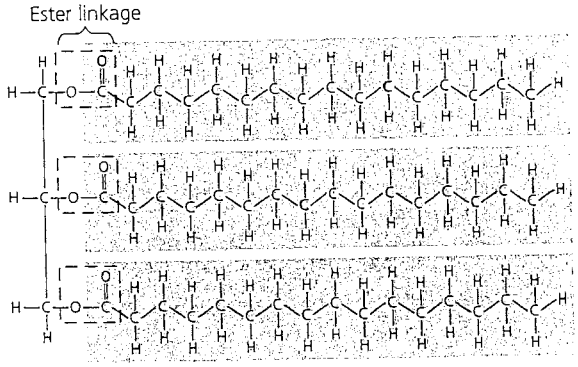
behavior of lipids is based on their molecular structure. Although they may have some polar bonds associated with oxygen, lipids consist mostly of hydrocarbons. Smaller than true (polymeric) macromolecules, lipids are a highly varied group in both form and function. Lipids include waxes and certain pigments, 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, and they are assembled from 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. At one end of the fatty acid is 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 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



(a) Dehydration reaction in the synthesis of a fat



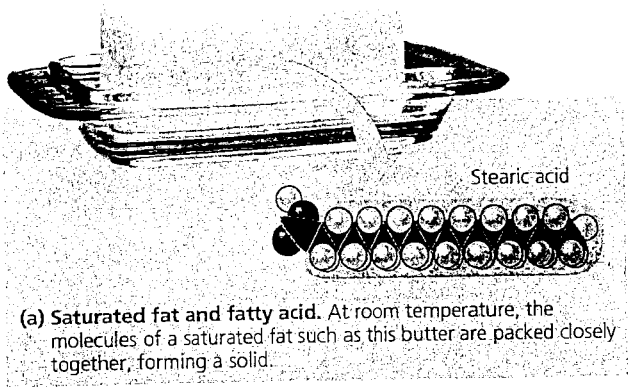
(b) Fat molecule (triacylglycerol)

▲ **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).

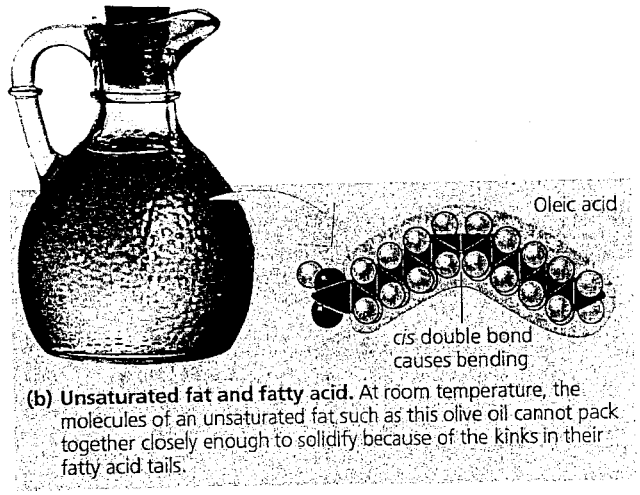
nomenon is the separation of vegetable oil (a liquid fat) from the aqueous vinegar solution in a bottle of salad dressing.

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)**. 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**



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



(b) Unsaturated fat and fatty acid. 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 their fatty acid tails.

▲ **Figure 5.12 Examples of saturated and unsaturated fats and fatty acids.**

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)**.

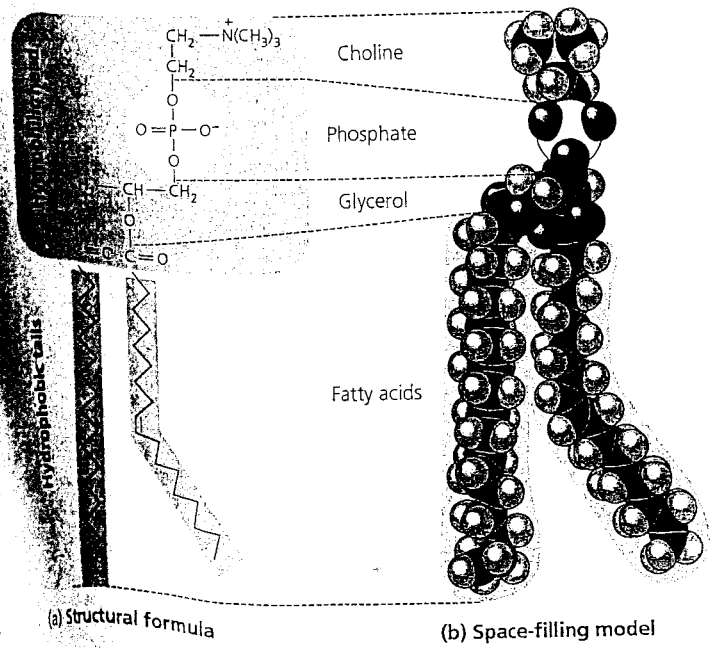
A fat made from saturated fatty acids is called a saturated fat. Most animal fats are saturated: The hydrocarbon chains of their fatty acids—the "tails" of the fat molecules—lack double bonds, and the molecules can pack tightly, side by side. Saturated animal fats—such as lard and butter—are solid at 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* fat molecules may contribute more than saturated fats to atherosclerosis (see Chapter 42) and other problems.

Fat has come to have such a negative connotation in our culture that you might wonder whether fats serve any useful purpose. 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 4.6b), 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

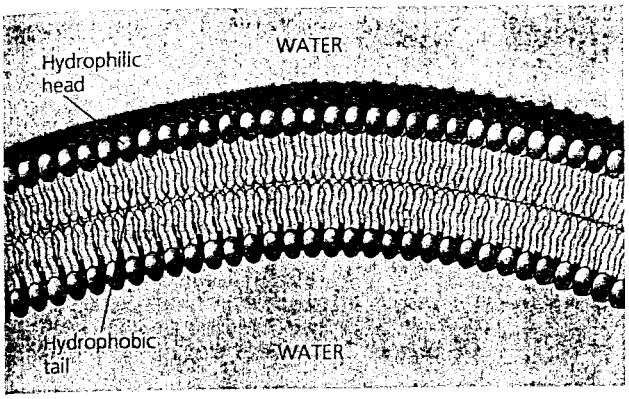
A **phospholipid**, as shown in **Figure 5.13**, is similar to a fat, but has only two fatty acids attached to glycerol rather than three. The third hydroxyl group of glycerol is joined to a phosphate



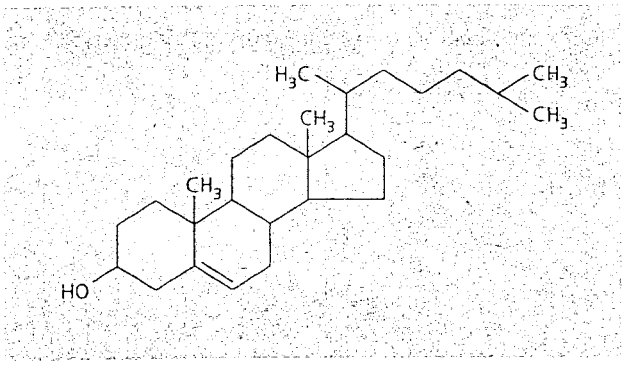
(a) Structural formula (b) Space-filling model

◀ 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. **(a)** The structural formula follows a common chemical convention of omitting the carbons and attached hydrogens of the hydrocarbon tails. **(b)** In the space-filling model, black = carbon, gray = hydrogen, red = oxygen, yellow = phosphorus, and blue = nitrogen. **(c)** This symbol for a phospholipid will appear throughout the book.

(c) Phospholipid symbol



▲ **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 functional groups attached to their four interconnected rings (shown in gold).

group, which has a negative electrical charge. Additional small molecules, usually charged or polar, can be linked to the phosphate group to form a variety of phospholipids.

Phospholipids show ambivalent behavior toward water. Their 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 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, phospholipids are major components of all cell membranes. This behavior provides another example of how form fits function at the molecular level.

Steroids

Steroids are lipids characterized by a carbon skeleton consisting of four fused rings **(Figure 5.15)**. Different steroids vary in the functional groups attached to this ensemble of rings. One steroid, **cholesterol**, is a common component of animal cell membranes and is also the precursor from which other steroids are synthesized. 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

1. Compare the structure of a fat (triglyceride) with that of a phospholipid.
2. How do saturated fats differ from unsaturated fats, both in structure and in behavior?
3. Why are human sex hormones considered to be lipids?

For suggested answers, see Appendix A.

Concept

Proteins have many structures, resulting in a wide range of functions

The importance of proteins is implied 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, storage, transport, cellular communications, movement, and defense against foreign substances (**Table 5.1**, on the next page).

The most important type of protein may be **enzymes**. Enzymatic proteins regulate metabolism by acting as **catalysts**, chemical agents that selectively speed up chemical reactions in

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 homone 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 movement 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.

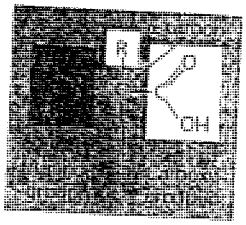
the cell without being consumed by the reaction **(Figure 5.16)**. Because an enzyme can perform its function over and over again, these molecules can be thought of as workhorses that keep cells running by carrying out the processes of life.

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, or conformation.

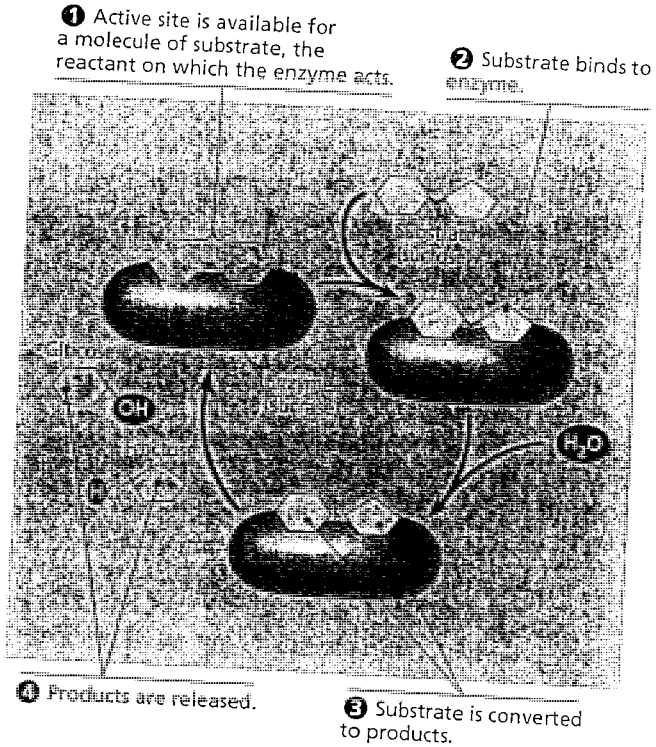
Polypeptides

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 folded and coiled into specific conformations.

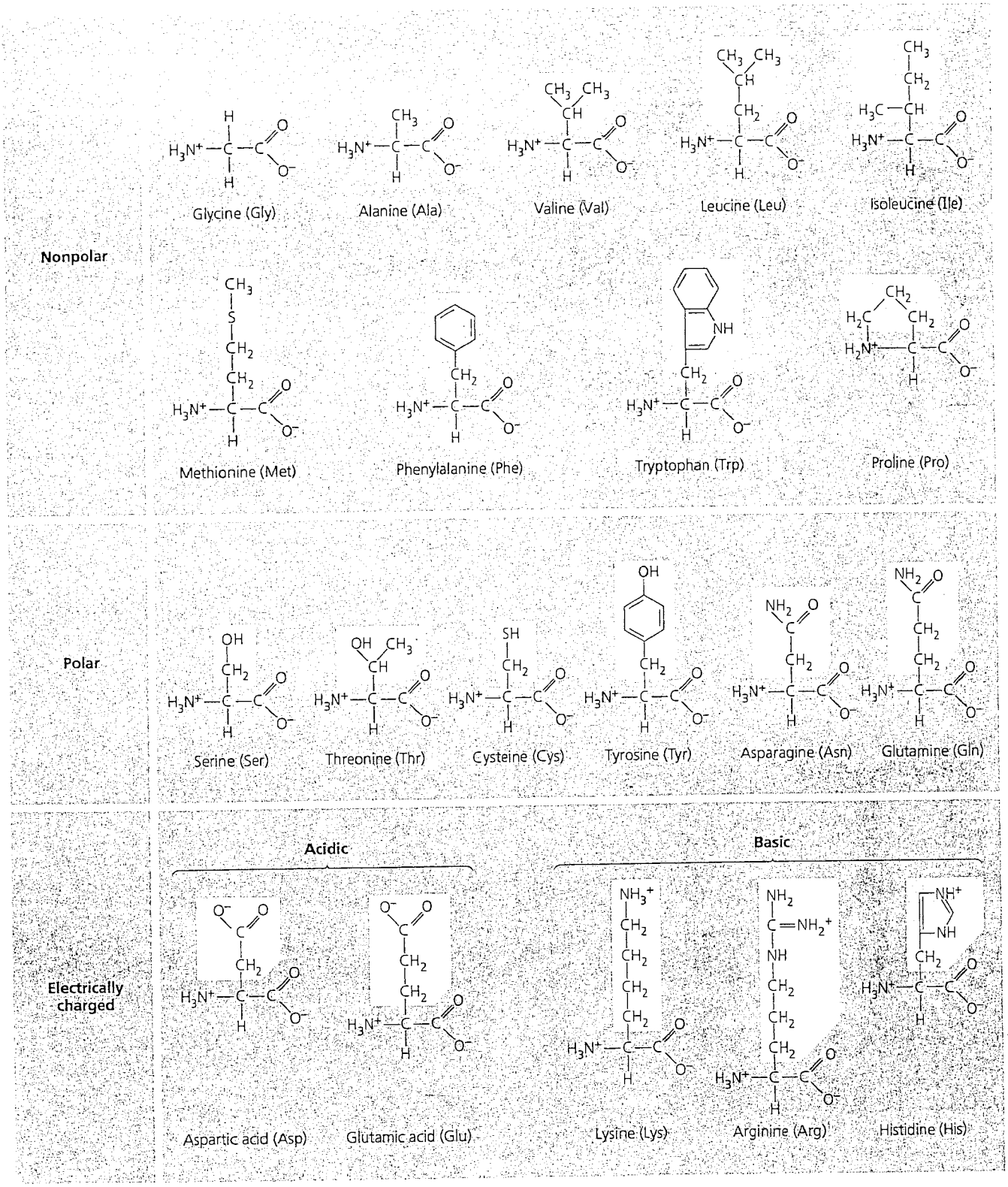
Amino Acid Monomers



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



▲ **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 is available for further catalysis.



▲ 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 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).

build their thousands of proteins. Here the amino and carboxyl 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.)

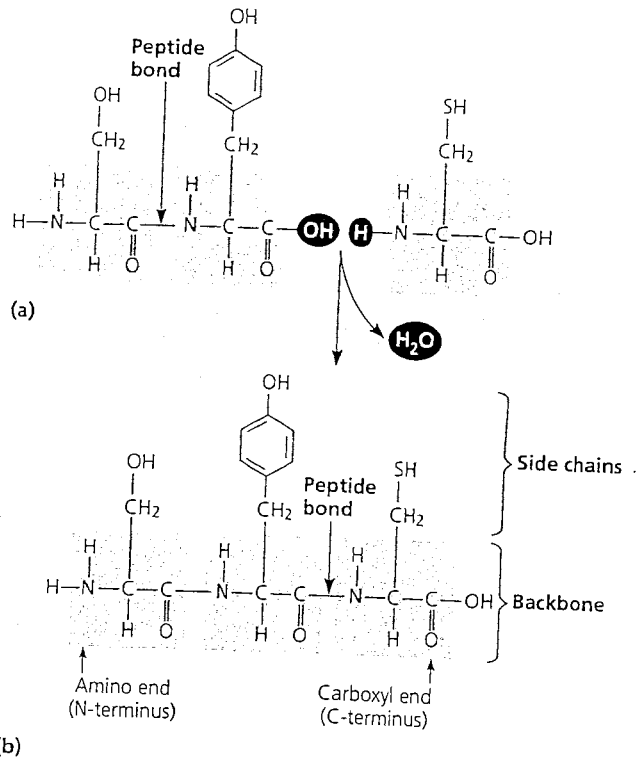
The physical and chemical properties of the side chain determine the unique characteristics of a particular amino acid. In Figure 5.17, the amino acids are grouped according to the properties of their side chains. One group consists of amino acids with nonpolar side chains, which are hydrophobic. Another 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, an enzyme can cause them to join by catalyzing 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. Attached to 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.

Determining the Amino Acid Sequence of a Polypeptide

The pioneer in determining the amino acid sequence of proteins was Frederick Sanger, who, with his colleagues at



▲ **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.

Cambridge University in England, worked on the hormone insulin in the late 1940s and early 1950s. His approach was to use protein-digesting enzymes and other catalysts that break polypeptides at specific places rather than completely hydrolyzing the chains to amino acids. Treatment with one of these agents cleaves a polypeptide into fragments (each consisting of multiple amino acid subunits) that can be separated by a technique called chromatography. Hydrolysis with a different agent breaks the polypeptide at different sites, yielding a second group of fragments. Sanger used chemical methods to determine the sequence of amino acids in these small fragments. Then he searched for overlapping regions among the pieces obtained by hydrolyzing with the different agents. Consider, for instance, two fragments with the following sequences:

Cys-Ser-Leu-Tyr-Gln-Leu
Tyr-Gln-Leu-Glu-Asn

We can deduce from the overlapping regions that the intact polypeptide contains in its primary structure the following segment:

Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn

Just as we could reconstruct this sentence from a collection of fragments with overlapping sequences of letters, Sanger and his co-workers were able, after years of effort, to reconstruct the complete primary structure of insulin. Since then, most of the steps involved in sequencing a polypeptide have been automated.

Protein Conformation and Function

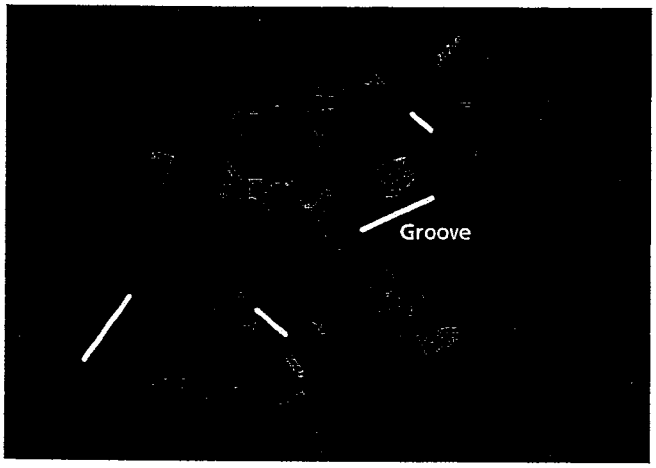
Once we have learned the amino acid sequence of a polypeptide, what can it tell us about protein conformation and function? The term *polypeptide* is not quite synonymous with the term *protein*. Even for a protein consisting of a single polypeptide, the relationship is somewhat analogous to that between a long strand of yarn and a sweater of particular size and shape that one can 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)**. It is the amino acid sequence of a polypeptide that determines what three-dimensional conformation the protein will take.

When a cell synthesizes a polypeptide, the chain generally folds spontaneously, assuming the functional conformation 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 globular (roughly spherical), while others are fibrous in shape. Even within these broad categories, countless variations are possible.

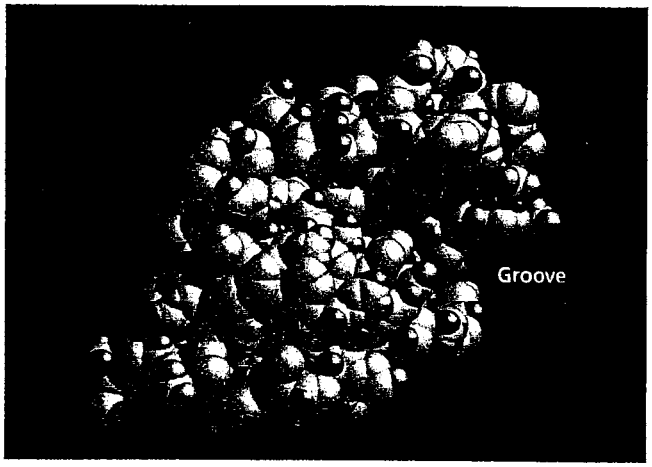
A protein's specific conformation determines how it works. In almost every case, the function of a protein depends on its ability to recognize and bind to some other molecule. For instance, an antibody (a protein) binds to a particular foreign substance that has invaded the body, and an enzyme (another type of protein) recognizes and binds to its substrate, the substance the enzyme works on. In Chapter 2, you learned that natural signal 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 2.17). Thus, the function of a protein—for instance, the ability of a receptor protein to identify and associate with a particular pain-relieving signal molecule—is an emergent property resulting from exquisite molecular order.

Four Levels of Protein Structure

In the complex architecture of a protein, we can recognize three superimposed levels of structure, known as primary, secondary, and tertiary structure. A fourth level, quaternary structure, arises when a protein consists of two or more polypeptide chains. **Figure 5.20**, on the following two pages, describes these four levels of protein structure. Be sure to study this figure thoroughly before going on to the next section.



(a) A **ribbon model** shows how the single polypeptide chain folds and coils to form the functional protein. (The yellow lines represent one type of chemical bond that stabilizes the protein's shape.)

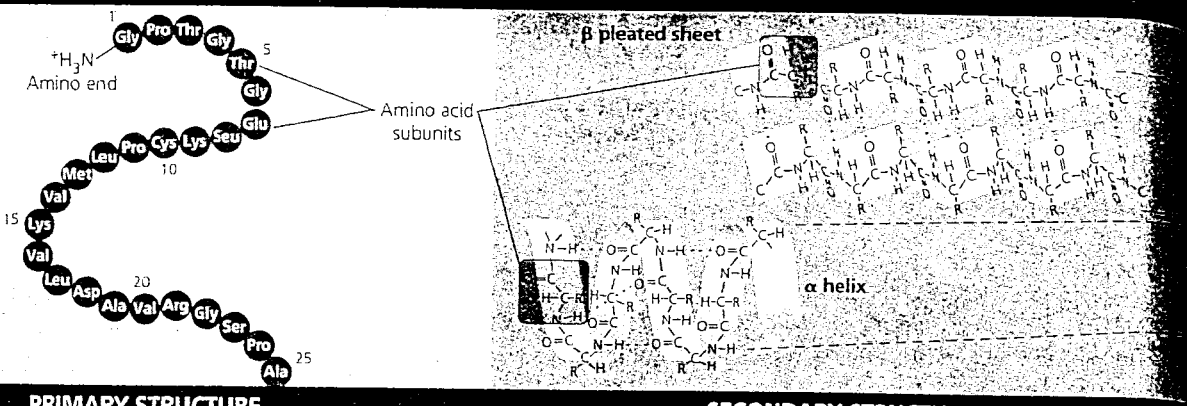


(b) A **space-filling model** shows more clearly the globular shape seen in many proteins, as well as the specific conformation unique to lysozyme.

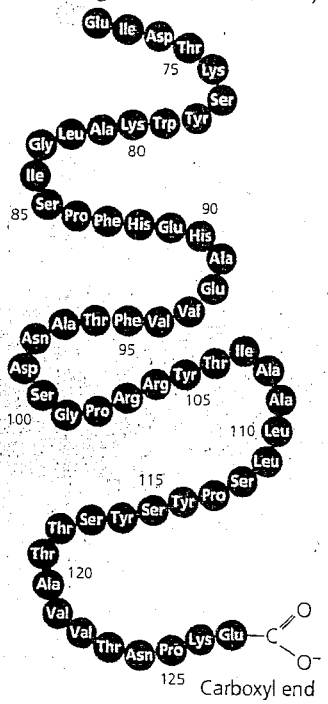
▲ **Figure 5.19 Conformation of a protein, the enzyme lysozyme.** Present in our sweat, tears, and saliva, lysozyme is an enzyme that helps prevent infection by binding to and destroying specific molecules on the surface of many kinds of bacteria. The groove is the part of the protein that recognizes and binds to the target molecules on bacterial walls.

Figure 5.20

Levels of Protein 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 a particular thyroid hormone 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. A specific one of the 20 amino acids, indicated here by its three-letter abbreviation, occupies each of the 127 positions along the chain. 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.



Most proteins have segments of their polypeptide chains repeatedly coiled or folded in patterns that contribute to the protein's overall conformation. 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 2.15). 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.

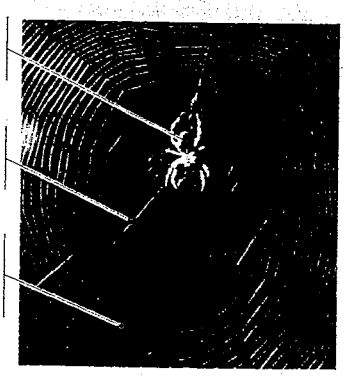
One such secondary structure is the **α helix**, a delicate coil held together by hydrogen bonding between every fourth amino acid, shown above for transthyretin. Although transthyretin has only one α helix region (see tertiary structure), other globular proteins have multiple stretches of α helix separated 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.

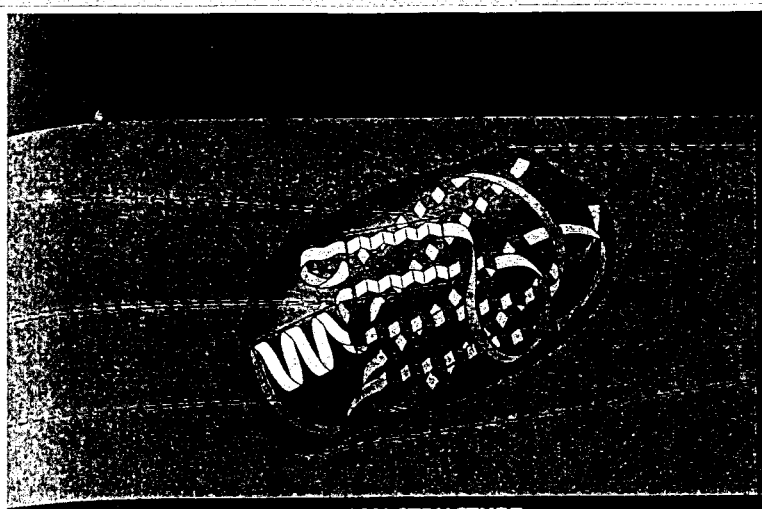
Abdominal glands of the spider secrete silk fibers that form the web.

The radiating strands, made of dry silk fibers, maintain the shape of the web.

The spiral strands (capture strands) are elastic, stretching in response to wind, rain, and the touch of insects.



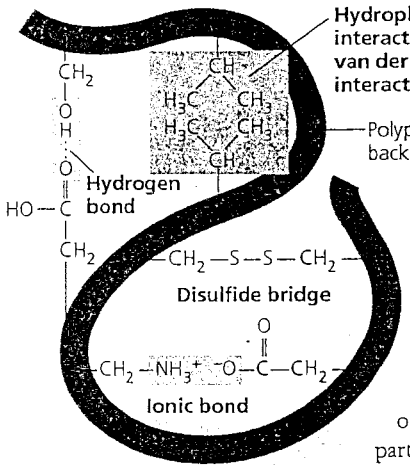
Spider silk: a structural protein containing β pleated sheets



TERTIARY STRUCTURE

QUATERNARY STRUCTURE

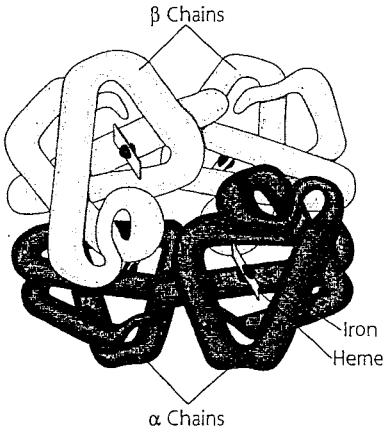
Superimposed on the patterns of secondary structure is a protein's **tertiary structure**, shown above for the transthyretin polypeptide. Rather than involving interactions between backbone constituents, tertiary structure is the overall shape of a polypeptide resulting from interactions between the 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 conformation, 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 conformation 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, 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 above 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 on the right, 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, 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.

Collagen



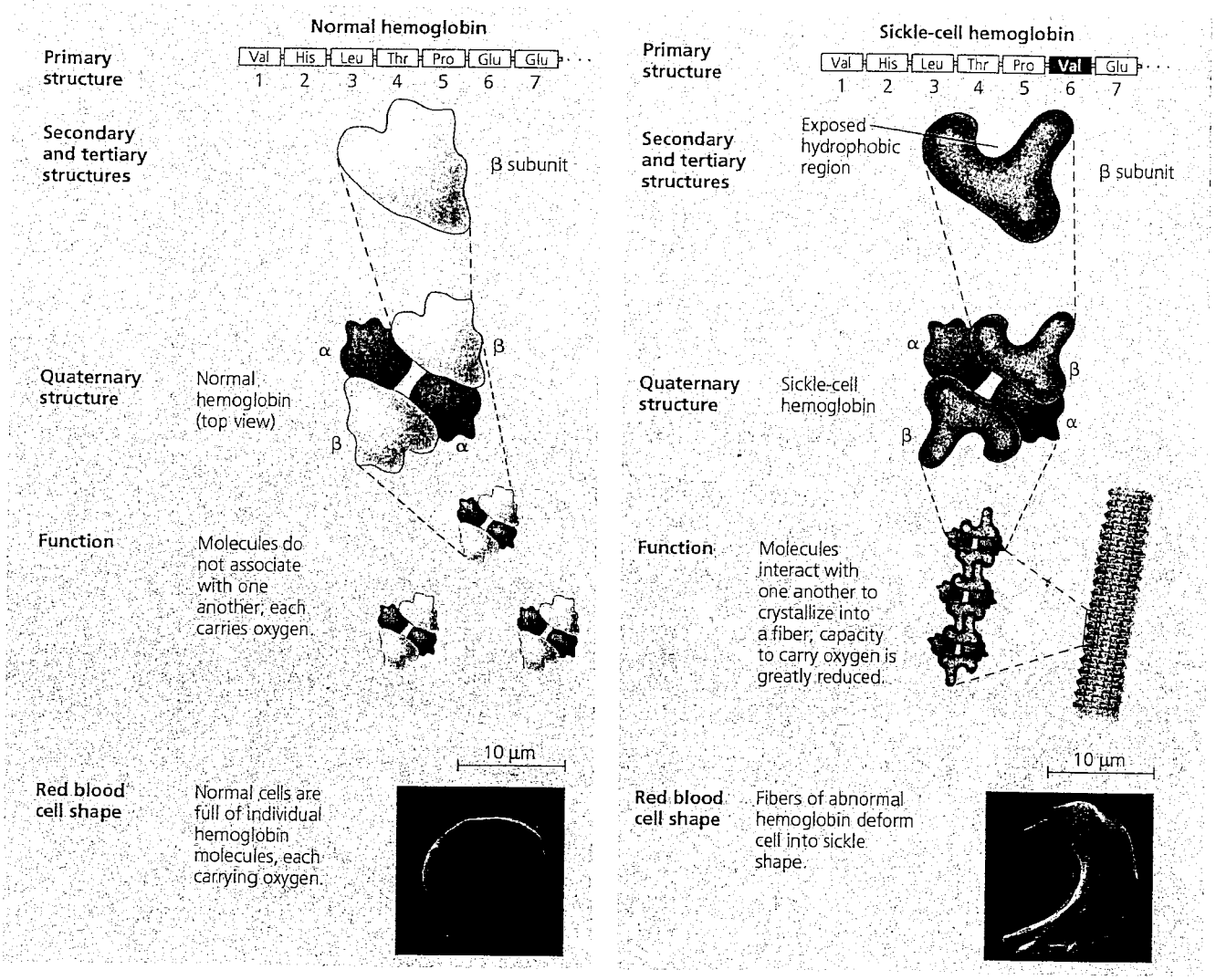
Hemoglobin

Sickle-Cell Disease: A Simple Change in Primary Structure

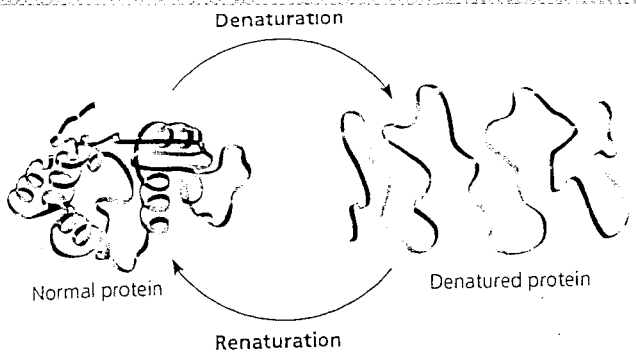
Even a slight change in primary structure can affect a protein's conformation and ability to function. For instance, the substitution of one amino acid (valine) for the normal one (glutamic acid) at a particular position in the primary structure of hemoglobin, the protein that carries oxygen in red blood cells, can cause *sickle-cell disease*, an inherited blood disorder. Normal red blood cells are disk-shaped, but in sickle-cell disease, the abnormal hemoglobin molecules tend to crystallize, deforming some of the cells into a sickle shape **(Figure 5.21)**. 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 Conformation?

You've learned that a unique shape endows each protein with a specific function. But what are the key factors determining protein conformation? 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 conformation 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 conformation, a change called **denaturation (Figure 5.22)**. Because it is misshapen, the denatured protein is biologically inactive.



▲ Figure 5.21 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 5.20.



▲ **Figure 5.22 Denaturation and renaturation of a protein.** High temperatures or various chemical treatments will denature a protein, causing it to lose its conformation 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.

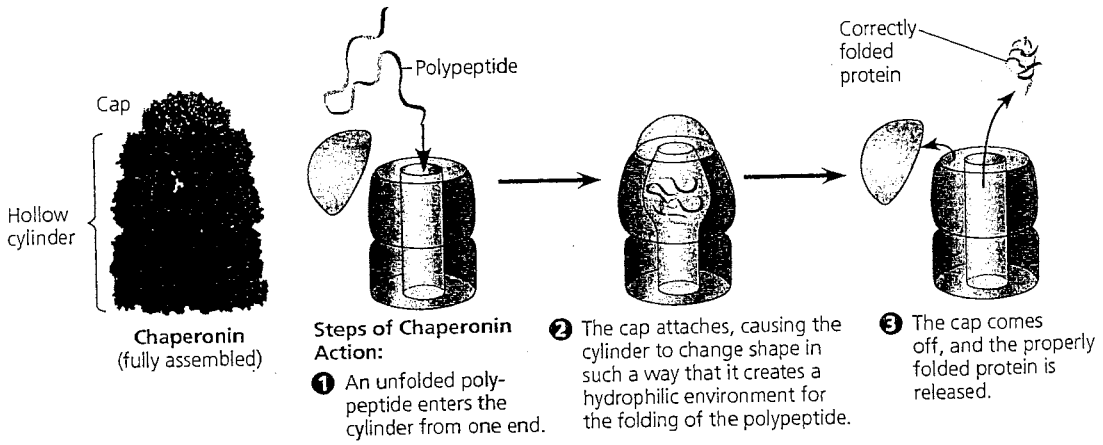
Most proteins become denatured if they are transferred from an aqueous environment to an organic solvent, such as ether or chloroform; the polypeptide chain refolds so that its hydrophobic regions face outward toward the solvent. Other denaturation agents include chemicals that disrupt the hydrogen bonds, 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 conformation. The white of an egg becomes opaque during cooking because the denatured proteins are insoluble and solidify. This also explains why extremely high fevers can be fatal: Proteins in the blood become denatured by such high body temperatures.

When a protein in a test-tube solution has been denatured by heat or chemicals, it will often 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 conformation—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. However, in the crowded environment inside a cell, correct folding may be more of a problem than it is in a test tube.

The Protein-Folding Problem

Biochemists now know the amino acid sequences of more than 875,000 proteins and the three-dimensional shapes of about 7,000. One would think that by correlating the primary structures of many proteins with their conformations, it would be relatively easy to discover the rules of protein folding. Unfortunately, the protein-folding problem is not that simple. Most proteins probably go through several intermediate states on their way to a stable conformation, and looking at the mature conformation does not reveal the stages of folding required to achieve that form. However, biochemists have developed methods for tracking a protein through its intermediate stages of folding. Researchers have also discovered **chaperonins** (also called chaperone proteins), protein molecules that assist the proper folding of other proteins **(Figure 5.23)**. Chaperonins do not actually specify the correct final structure of a polypeptide. Instead, they work by keeping the new polypeptide segregated from "bad influences" in the cytoplasmic environment while it folds spontaneously. The well-studied chaperonin shown in Figure 5.23, from the bacterium *E. coli*, is a giant multiprotein complex shaped like a hollow cylinder. The cavity provides a shelter for folding polypeptides of various types.

Even when scientists have an actual protein in hand, determining its exact three-dimensional structure is not simple, for a single protein molecule is built of thousands of atoms. X-ray **crystallography** is an important method used to determine a protein's three-dimensional structure **(Figure 5.24)**. Another method that has recently been applied to this problem is nuclear magnetic resonance (NMR) spectroscopy, which does not require protein crystallization. These approaches have contributed greatly to our understanding of protein structure and have also given us valuable hints about protein function.



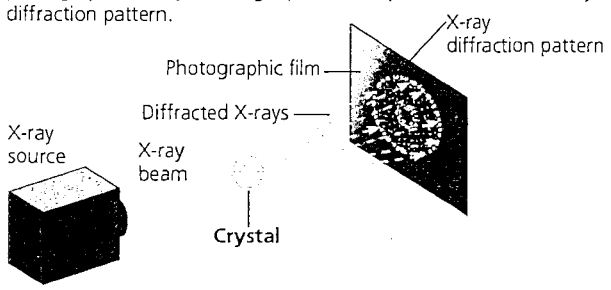
▶ **Figure 5.23 A chaperonin in action.** The computer graphic (left) shows a large chaperonin protein complex with an interior space that provides a shelter for the proper folding of newly made polypeptides. The complex consists of two proteins: One protein is a hollow cylinder; the other is a cap that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

Figure 5.24

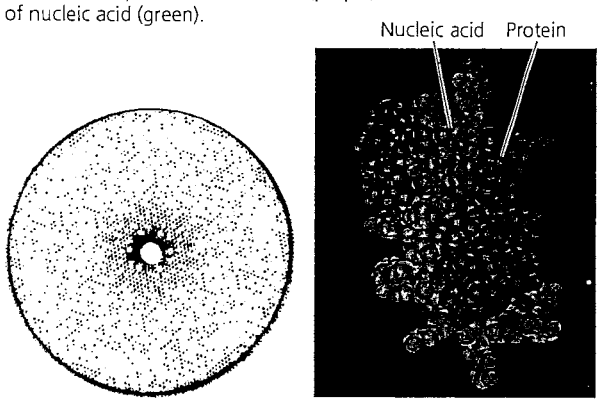
X-Ray Crystallography

APPLICATION Scientists use X-ray crystallography to determine the three-dimensional structure of macromolecules such as nucleic acids and proteins. In this figure we will examine how researchers at the University of California, Riverside, determined the structure of the protein ribonuclease, an enzyme whose function involves binding to a nucleic acid molecule.

TECHNIQUE Researchers aim an X-ray beam through the crystallized protein. The atoms of the crystal diffract (deflect) the X-rays into an orderly array. The diffracted X-rays expose photographic film, producing a pattern of spots known as an X-ray diffraction pattern.



RESULTS Using data from X-ray diffraction patterns, as well as the amino acid sequence determined by chemical methods, scientists build a 3D computer model of the protein, such as this model of the protein ribonuclease (purple) bound to a short strand of nucleic acid (green).



(a) X-ray diffraction pattern (b) 3D computer model

Concept Check

1. Why does a denatured protein no longer function normally?
2. Differentiate between secondary and tertiary structure by describing the parts of the polypeptide chain that participate in the bonds that hold together each level of structure.
3. A genetic mutation can change a protein's primary structure. How can this destroy the protein's function?

For suggested answers, see Appendix A.

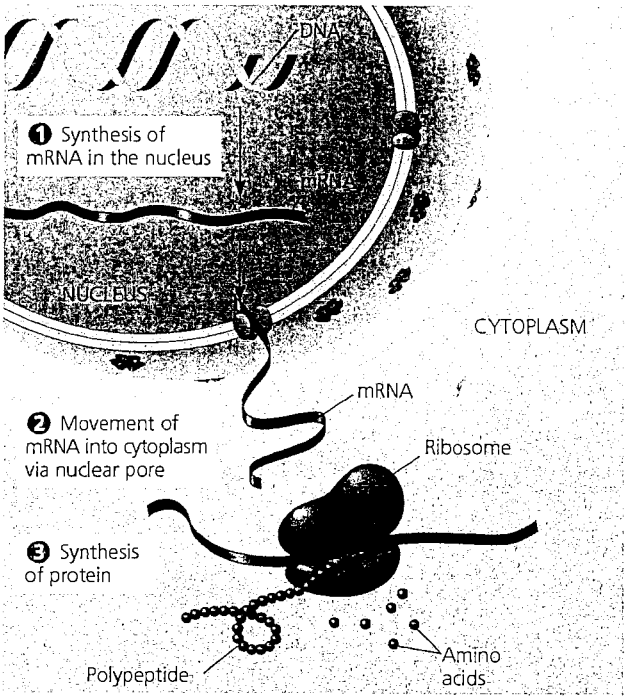
Concept

Nucleic acids store and transmit hereditary information

If the primary structure of polypeptides determines the conformation of a protein, 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, which is a polymer belonging to the class of compounds known as nucleic acids.

The Roles of Nucleic Acids

There are two types of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). These are the molecules that 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 5.25)**.



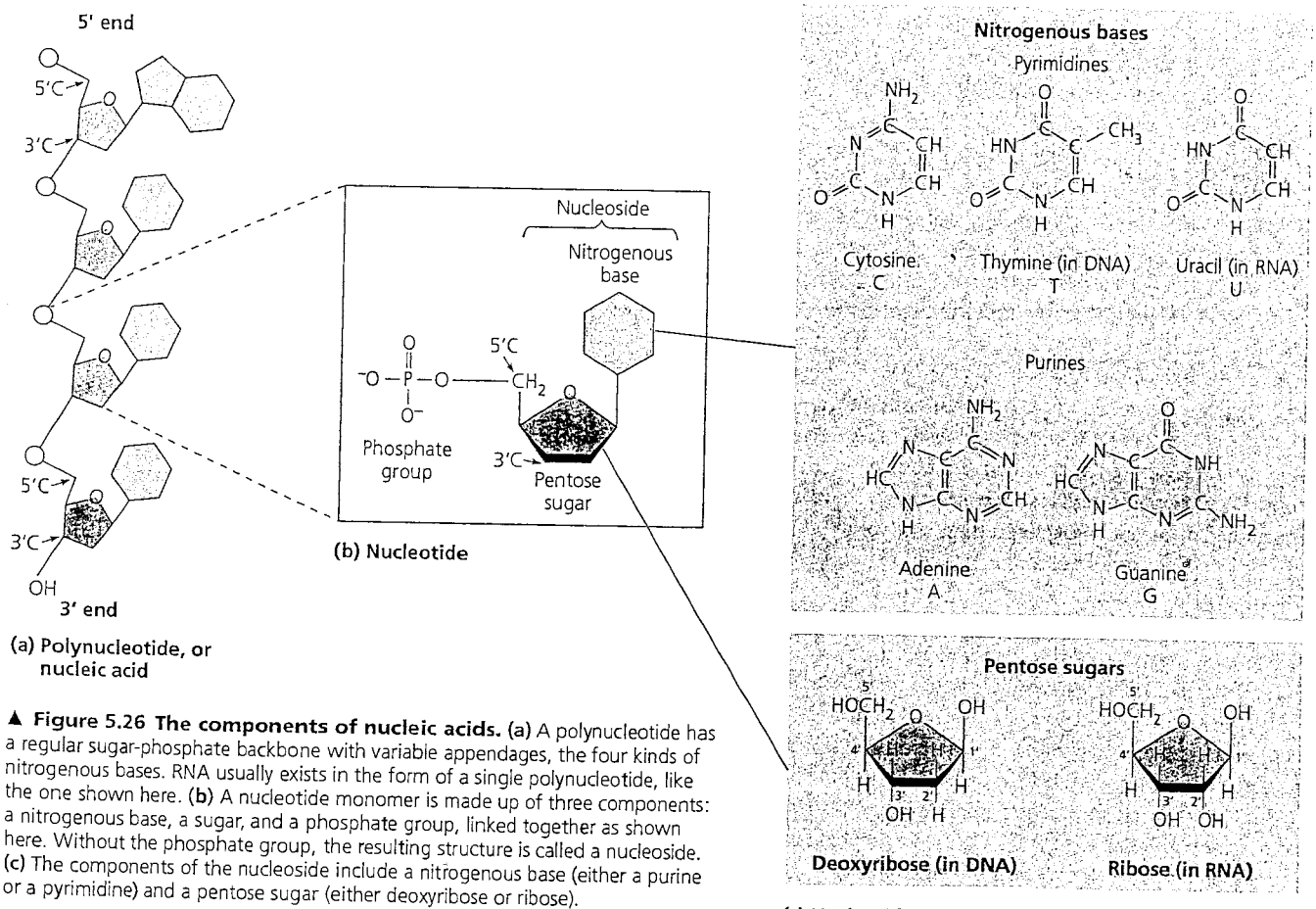
▲ **Figure 5.25 DNA → RNA → protein: a diagrammatic overview of information flow in a cell.** In a eukaryotic cell, DNA in the nucleus programs protein production in the cytoplasm by dictating the synthesis of messenger RNA (mRNA), which travels to the cytoplasm and binds to ribosomes. As a ribosome (greatly enlarged in this drawing) moves along the mRNA, the genetic message is translated into a polypeptide of specific amino acid sequence.

DNA is the genetic material that organisms inherit from their parents. Each chromosome contains one long DNA molecule, usually consisting of from several hundred to more than a thousand 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 to implement genetic programs. The molecular hardware of the cell—the tools for most biological functions—consists of proteins. For example, the oxygen carrier in the blood 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 the length of a DNA molecule directs the synthesis of a type of RNA called *messenger RNA* (mRNA). The mRNA molecule then interacts with the cell's protein-synthesizing machinery to direct the production of a polypeptide. We can summarize the flow of genetic information as DNA → RNA → protein (see Figure 5.25). The actual sites of protein synthesis are cellular structures called ribosomes. In a eukaryotic cell, ribosomes are located in the cytoplasm, but DNA resides in the nucleus. Messenger RNA conveys the genetic instructions for building proteins from the nucleus to the cytoplasm. Prokaryotic cells lack nuclei, but they still use RNA to send a message from the DNA to the ribosomes and other equipment of the cell that translate the coded information into amino acid sequences.

The Structure of Nucleic Acids

Nucleic acids are macromolecules that exist as polymers called **polynucleotides** **(Figure 5.26a)**. As indicated by the name, each polynucleotide consists of monomers called **nucleotides**. A nucleotide is itself composed of three parts: a nitrogenous base, a pentose (five-carbon sugar), and a phosphate group **(Figure 5.26b)**. The portion of this unit without the phosphate group is called a *nucleoside*.



▲ **Figure 5.26 The components of nucleic acids.** **(a)** A polynucleotide has a regular sugar-phosphate backbone with variable appendages, the four kinds of nitrogenous bases. RNA usually exists in the form of a single polynucleotide, like the one shown here. **(b)** A nucleotide monomer is made up of three components: a nitrogenous base, a sugar, and a phosphate group, linked together as shown here. Without the phosphate group, the resulting structure is called a nucleoside. **(c)** The components of the nucleoside include a nitrogenous base (either a purine or a pyrimidine) and a pentose sugar (either deoxyribose or ribose).

Nucleotide Monomers

To build a nucleotide, let's first consider the two components of the nucleoside: the nitrogenous base and the sugar (**Figure 5.26c**). 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 the term *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 functional 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 pentose connected to the nitrogenous base is **ribose** in the nucleotides of RNA and **deoxyribose** in DNA (see Figure 5.26c). The only difference between these two sugars is that deoxyribose lacks an oxygen atom on the second carbon in the ring; hence its name. 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 5.26b). The molecule is now a nucleoside monophosphate, better known as a nucleotide.

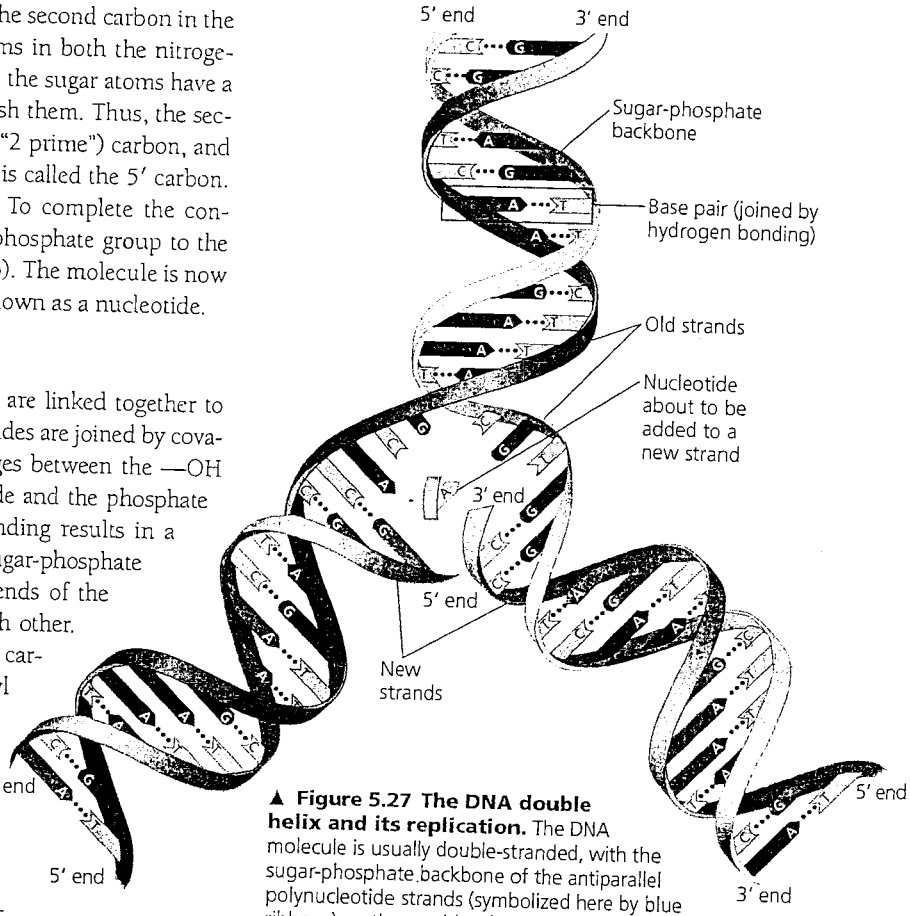
Nucleotide Polymers

Now we can see how these nucleotides are linked together to build a polynucleotide. Adjacent nucleotides are joined by covalent bonds called phosphodiester linkages between the —OH group on the 3′ carbon of one nucleotide and the phosphate on the 5′ carbon of the next. This bonding results in a backbone with a repeating pattern of sugar-phosphate units (see Figure 5.26a). 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. So 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. 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. (Real 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 conformation 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 5.26. In contrast, cellular DNA molecules have two polynucleotides that spiral around an imaginary axis, forming a **double helix** (**Figure 5.27**). James Watson and Francis Crick, working at Cambridge University,



▲ Figure 5.27 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 5.26a.

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 5.27). 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. For example, if we compare a polypeptide chain of human hemoglobin with the corresponding hemoglobin polypeptide in five other vertebrates, we find the following. In this chain of 146 amino acids, humans and gorillas differ in just 1 amino acid, humans and gibbons differ in 2 amino acids, and humans and rhesus monkeys differ in 8 amino acids. More distantly related species have chains that are less similar. Humans and mice differ in 27 amino acids, and humans and frogs differ in 67 amino acids. Molecular biology has added a new tape measure to the toolkit biologists use to assess evolutionary kinship.

Concept Check

1. Go to Figure 5.26a 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′. List the base sequence along the other strand of the molecule, clearly indicating the 5′ and 3′ ends of this strand.

For suggested answers, see Appendix A.

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 1.3). With each increasing level of order, new properties emerge in addition to those of the component parts. In Chapters 2–5, we have dissected the chemistry of life using the strategy of the reductionist. But we have also begun to develop a more integrated view of life as we have seen how properties emerge with increasing order.

We have seen that the unusual behavior of water, so essential to life on Earth, results from interactions of the water molecules, themselves an ordered arrangement of hydrogen and oxygen atoms. We reduced the great complexity and diversity of organic compounds to the chemical characteristics of carbon, but we also saw that the unique properties of organic compounds are related to the specific structural arrangements of carbon skeletons and their appended functional groups. We learned that small organic molecules are often assembled into giant molecules, but we also discovered that a macromolecule does not behave like a composite of its monomers but rather takes on additional properties owing to the interactions between those monomers.

By completing our overview of the molecular basis of life with an introduction to the important classes of macromolecules that build living cells, we have built a bridge to Unit Two, where we will study the cell's structure and function. We will maintain our balance between the need to reduce life to a conglomerate of simpler processes and the ultimate satisfaction of viewing those processes in their integrated context.