Biochemistry for Biotechnology students

(Second year)

(First course)

Preparation by Dr. Ali W. Al-Ani

Reference:

INTRODUCTION TO GENERAL, ORGANIC, AND BIOCHEMISTRY

Tenth Edition

Morris Hein
Mount San Antonio College

Scott Pattison
Ball State University

Susan Arena
University of Illinois, Urbana-Champaign

- Outline

- Carbohydrates
- Lipids
- Amino Acids, Polypeptides, and Proteins
- Nucleic Acids and Heredity

First lecture

Carbohydrates

Green plants turn H_2O, CO_2, and sunlight into carbohydrates.

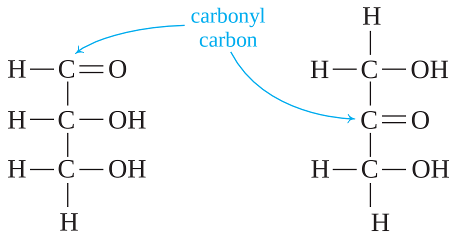
Chapter Outline

27.1 Carbohydrates: A First Class of Biochemicals

Carbohydrates are among the most widespread and important biochemicals. Most of the matter in plants, except water, consists of these substances. Carbohydrates are one of the three principal classes of energy-yielding nutrients; the other two are fats and proteins. Because of their widespread distribution and their role in many vital metabolic processes such as photosynthesis, carbohydrates have been scrutinized by scientists for over 150 years.

The name *carbohydrates* was given to this class of compounds many years ago by French scientists, who called them *hydrates de carbone* because their empirical formulas approximated $(C\#H_2O)_n$. It was found later that not all substances classified as carbohydrates conform to this formula (e.g., rhamnose, $C_6H_{12}O_5$, and deoxyribose, $C_5H_{10}O_4$). It seems clear that carbohydrates are not simply hydrated carbon; they are complex substances that contain from three to many thousands of carbon atoms. **Carbohydrates** are generally defined as polyhydroxy aldehydes or polyhydroxy ketones or substances that yield these compounds when hydrolyzed.

The simplest carbohydrates are glyceraldehyde and dihydroxyacetone:



glyceraldehyde, $C_3H_6O_3$

dihydroxyacetone, $C_3H_6O_3$

These substances are "polyhydroxy" because each molecule has more than one hydroxyl group. Glyceraldehyde contains a carbonyl carbon in a terminal position and is therefore an aldehyde. The internal carbonyl of dihydroxyacetone identifies it as a ketone. Much of the chemistry and biochemistry of carbohydrates can be understood from a basic knowledge of the chemistry of the hydroxyl and carbonyl functional groups. (See Chapters 22 and 23 for review(.

When sugar (sucrose) and sulfuric acid (left) are combined, water is removed from the sugar, leaving a black column of carbon (right.(

27.2 Classification of Carbohydrates

monosaccharide

disaccharide

A carbohydrate is classified as a monosaccharide, a disaccharide, an oligosaccharide, or a polysaccharide, depending on the number of monosaccharide units linked to form the molecule. A **monosaccharide** is a carbohydrate that cannot be hydrolyzed to simpler carbohydrate units. The monosaccharide is the basic carbohydrate unit of cellular metabolism. A **disaccharide** yields two monosaccharides—either alike or different—when hydrolyzed:

$$\text{disaccharide} + \text{water} \xrightarrow{\text{H}^+ \text{ or enzymes}} 2 \text{ monosaccharides}$$

Disaccharides are often used by plants or animals to transport monosaccharides from one cell to another. The monosaccharides and disaccharides generally have names ending in *-ose*—for example, glucose, sucrose, and lactose. These water-soluble carbohydrates, which have a characteristically sweet taste, are also called *sugars*.

oligosaccharide

An **oligosaccharide** has two to six monosaccharide units linked together. *Oligo* comes from the Greek word *oligos*, which means "small or few." Free oligosaccharides that contain more than two monosaccharide units are rarely found in nature.

polysaccharide

A **polysaccharide** is a macromolecular substance that can be hydrolyzed to yield many monosaccharide units:

$$\text{polysaccharide} + \text{water} \xrightarrow{\text{H}^+ \text{ or enzymes}} \text{many monosaccharide units}$$

Polysaccharides are important as structural supports, particularly in plants, and also serve as a storage depot for monosaccharides, which cells use for energy.

Carbohydrates are vital to life, so it is not surprising that they have also been classified with respect to their impact on metabolism. Originally, mono- and disaccharides were grouped together as the *simple sugars*—carbohydrates that were easily digested and that produced a rapid increase in blood sugar after a meal. The oligo- and polysaccharides were placed in the *complex carbohydrate* category—carbohydrates that would release glucose more slowly during digestion. More research has shown that dietary carbohydrates' effect on blood glucose is more complex than these categories suggest. Today's nutritionists rank carbohydrates via the glycemic index. This is an experimental measure of a carbohydrate's impact on blood sugar relative to a standard, either glucose or white bread. It has been shown that a diet of carbohydrates with a higher glycemic index correlates with greater risk of diabetes (Type 2) and coronary heart disease. That is, if a carbohydrate has a low glycemic index, it is probably a healthier food.

Second lecture

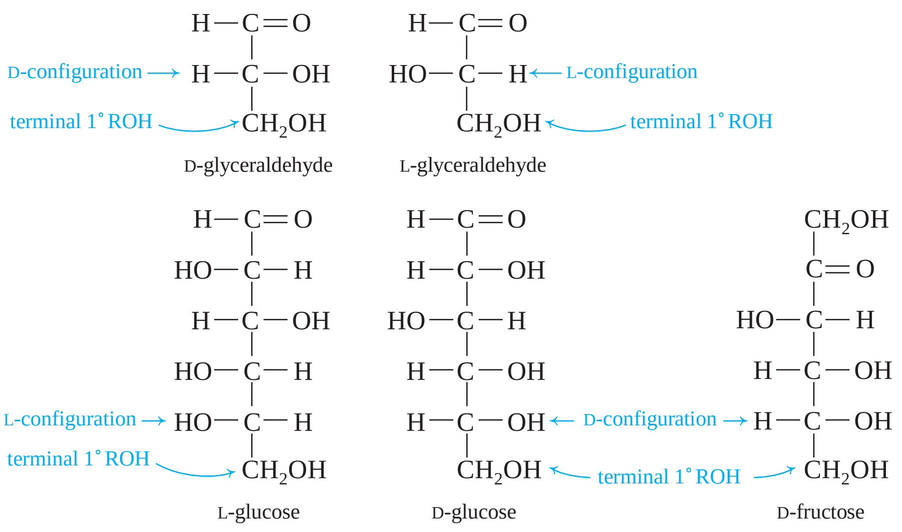
Carbohydrates can also be classified in other ways. A monosaccharide might be described with respect to several of these categories:

.1As a triose, tetrose, pentose, hexose, or heptose:

Trioses	$C_3H_6O_3$	Hexoses	$C_6H_{12}O_6$
Tetroses	$C_4H_8O_4$	Heptoses	$C_7H_{14}O_7$
Pentoses	$C_5H_{10}O_5$		

Theoretically, a monosaccharide can have any number of carbons greater than three, but only monosaccharides of three to seven carbons are commonly found in the biosphere.

2. As an aldose or ketose, depending on whether an aldehyde group (¬ CHO) or keto group (CO)is present. For ketoses, the C O is normally located on carbon .2
3. As a D or L isomer, depending on the spatial orientation of the ¬ H and ¬ OH groups attached to the carbon atom adjacent to the terminal primary alcohol group. When the ¬ OH is written to the right of this carbon in the projection formula (see Section 26.4), the D isomer is represented. When this ¬ OH is written to the left, the L isomer is represented. The reference compounds for this classification are the trioses D-glyceraldehyde and L-glyceraldehyde, whose formulas follow. Also shown are two aldohexoses (D- and L-glucose) and a ketohexose (D-fructose.(

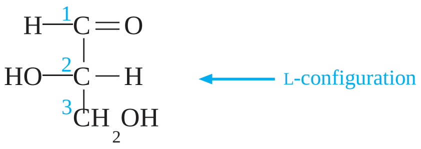


Remember, enantiomers are mirror-image isomers.

The letters D and L do not in any way refer to the direction of optical rotation of a carbohydrate. The D and L forms of any specific compound are enantiomers (e.g,. D- and L-glucose.(

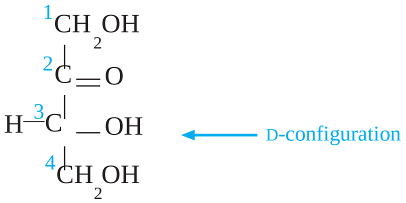
.4As a (+) or (-) isomer, depending on whether the monosaccharide rotates the plane of polarized light to the right (+) or to the left (-). (See Section (.26.5

)adjacent to the terminal CH_2OH) is on the left. The aldehyde group is carbon :1



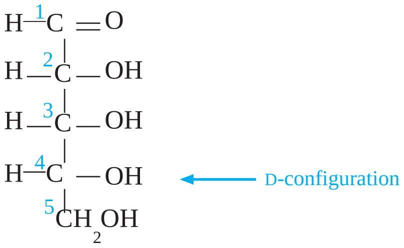
an L-aldotriose

)b) Tetrose indicates a four-carbon carbohydrate; keto indicates a ketone group (on carbon 2); D- indicates that the ¬ OH on carbon 3 (adjacent to the terminal CH_2OH) is on the right. Carbons 1 and 4 have the configuration of primary alcohols:



a D-ketotetrose

)c) Pentose indicates a five-carbon carbohydrate; aldo indicates an aldehyde group (on carbon 1); D- indicates that the ¬ OH on carbon 4 (adjacent to the terminal CH_2OH) is on the right. The orientation of the ¬ OH groups on carbons 2 and 3 is not specified here and therefore can be written in either direction for this problem:



a D-aldopentose

More Practice?
Try Paired Exercises 1–.6

Practice 27.1

Write the projection formula for an L-ketopentose.

Practice 27.2

In a projection formula for a D-aldotriose, is the ¬ OH of the secondary alcohol carbon written on the right or the left side?

27.3 Importance of Carbohydrates for Life

As we stated earlier, carbohydrates are the most abundant organic chemical in nature—they must be molecules of exceptional utility. But what makes them so special? Why are they so important in biochemistry?

If a carbohydrate carbon has less than 4 bonds to oxygen, it can be oxidized further.

Carbohydrates are used by essentially all cells as an energy source. Almost all cells, from the most primitive bacteria to complex human cells, have the

Third lecture

Polar bears have a large reserve of lipids.

Lipids

Chapter Outline

The behavior of lipids in water—the fact that they are insoluble—is a key to their importance in nature. For example, an oil slick can spread for many square miles on the surface of the ocean partly because oil and water don't mix. This same principle enables cells to surround themselves with a thin film of lipid, the cell membrane. We protect a fine wood floor with wax, another lipid, because we can depend on this material to adhere to the floor and not dissolve in water. A lipid's stickiness and water insolubility also create diseases such as atherosclerosis, where arteries become partially clogged by cholesterol-containing lipids. Lipids bring both benefits and problems—these molecules are truly a mixed blessing

28.1 Lipids: Hydrophobic Molecules

lipids

Lipids are water-insoluble, oily, or greasy biochemical compounds that can be extracted from cells by nonpolar solvents such as ether, chloroform, or benzene. Unlike carbohydrates, lipids share no common chemical structure. Instead, these biochemicals share the physical property of not dissolving in water.

What makes a molecule such as a lipid insoluble in water? It is perhaps easiest to answer this question by establishing what properties make a molecule water soluble. A good generalization to remember is "like dissolves like." That is, to dissolve in water a solute must be "like" water. The solute should be relatively small so that it can move quickly and freely through the solution like a water molecule. Also, the solute should bond to water molecules just as water molecules bond to each other. Sugars and many salts are water soluble because they form relatively strong bonds to water and are small enough that they can move easily through the solution. *Hydrophilic* ("water loving") is an adjective used to describe water-soluble solutes.

In contrast, lipids are "unlike" water molecules and are water insoluble. Lipids interact weakly with water molecules because they are composed primarily of nonpolar alkyl groups. Also, lipids tend to be very large compared with water molecules (molar mass = 18 g/mole). Typically, lipids have molar masses larger than about 200 g/mole. Thus, lipids are too big and don't attract water strongly enough to be water soluble. They are classified as *hydrophobic* ("water fearing") to designate their strong tendency to move away from water as if moving away from something fearful.

Consider fatty acids, which are common components of lipids. As shown in Table 28.1, as the fatty acids get larger, the water solubility of the fatty acid decreases dramatically. The four-carbon compound butyric acid is infinitely soluble in water, while fatty acids with more than about ten carbons long are insoluble in water. The fatty acids become increasingly "unlike" water as they become larger and more nonpolar.

TABLE 28.1 | Some Naturally Occurring Fatty Acids

Fatty acid	Number of C atoms	Formula	Solubility)g/100 g water)	Melting point (°C(
Saturated acids				
Butyric acid	4	$CH_3CH_2CH_2COOH$	q	7.9-
Caproic acid	6	$CH_3(CH_2)_4COOH$	1.08	3.4-
Caprylic acid	8	$CH_3(CH_2)_6COOH$	0.07	17
Capric acid	10	$CH_3(CH_2)_8COOH$	0.015	31
Lauric acid	12	$CH_3(CH_2)_{10}COOH$	insoluble	44
Myristic acid	14	$CH_3(CH_2)_{12}COOH$	insoluble	59
Palmitic acid	16	$CH_3(CH_2)_{14}COOH$	insoluble	63
Stearic acid	18	$CH_3(CH_2)_{16}COOH$	insoluble	70
Arachidic acid	20	$CH_3(CH_2)_{18}COOH$	insoluble	76
Unsaturated acids*				
Palmitoleic acid	16	$CH_3(CH_2)_5CH$“ $CH(CH_2)_7COOH$	—	0.5
Oleic acid	18	$CH_3(CH_2)_7CH$“ $CH(CH_2)_7COOH$	—	14
Linoleic acid	18	$CH_3(CH_2)_4CH$“ $CHCH_2CH$“ $CH(CH_2)_7COOH$	—	12-
Linolenic acid	18	CH_3CH_2CH“ $CHCH_2CH$“ $CHCH_2CH$“ $CH(CH_2)_7COOH$	—	11-
Arachidonic acid	20	$CH_3(CH_2)_4(CH$“ $CHCH_2)_4CH_2CH_2COOH$	—	50-
Eicosapentaenoic acid	20	$CH_3CH_2(CH$“ $CHCH_2)_5CH_2CH_2COOH$	—	60-
Docosahexaenoic acid	22	$CH_3CH_2(CH$“ $CHCH_2)_6CH_2COOH$	—	45-

*Omega (m) is the last letter in the Greek alphabet. Correspondingly, the last carbon atom in a carbon chain of a compound is often referred to as the omega carbon. In reference to unsaturated carboxylic acids, omega plus a number (e.g., m-3) indicates the location of the first carbon–carbon double bond, counting from the omega carbon.

The omega designations of these acids are palmitoleic (m-7); oleic (m-9); linoleic (m-6); linolenic (m-3); arachidonic (m-6); eicosapentaenoic (m-3); docosahexaenoic (m-.(3

28.2 Classification of Lipids

Lipids are hydrophobic molecules; their structures are relatively large and nonpolar. Yet, within this broad description, lipid structures vary markedly. The following classification scheme recognizes important structural differences:

1. **Simple lipids**
   (a) *Fats and oils:* esters of fatty acids and glycerol.
   (b) *Waxes:* esters of high-molar-mass fatty acids and high-molar-mass alcohols.
2. **Compound lipids**
   (a) *Phospholipids:* substances that yield glycerol, phosphoric acid, fatty acids, and a nitrogen-containing base upon hydrolysis.
   (b) *Sphingolipids:* substances that yield an unsaturated amino alcohol (sphingosine), a long-chain fatty acid, and either a carbohydrate or phosphate and a nitrogen base upon hydrolysis.
   (c) *Glycolipids:* substances that yield sphingosine, a fatty acid, and a carbohydrate upon hydrolysis.
3. **Steroids**
   Substances that possess the steroid nucleus, which is a 17-carbon structure consisting of four fused carbocyclic rings. Cholesterol and several hormones are in this class.
4. **Miscellaneous lipids**
   Substances that do not fit into the preceding classifications; these include the fat-soluble vitamins A, D, E, and K, and lipoproteins.

The most abundant lipids are the fats and oils. These substances constitute one of the three important classes of foods. In the discussion that follows, we examine fats and oils as they pertain to biochemistry. A more complete consideration of the properties and composition of various fats and oils is given in Section .24.10

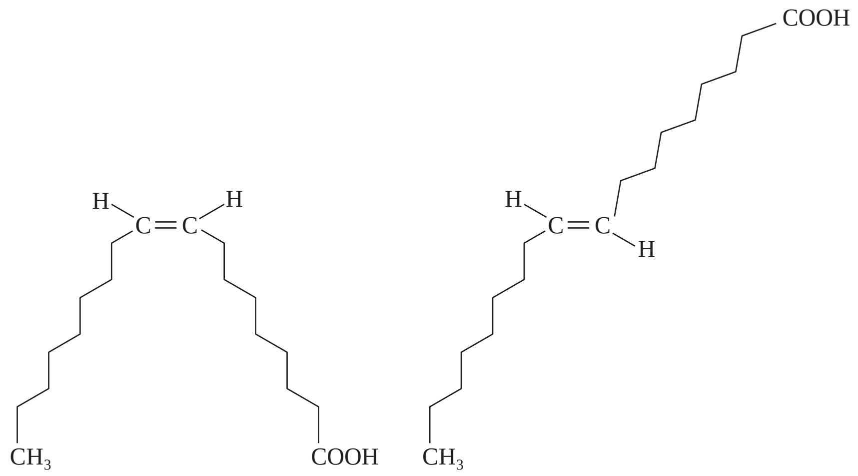
Fourth lecture

28.3 Simple Lipids

Fatty Acids

Fatty acids, which form part of most lipids, are carboxylic acids with long, hydrophobic carbon chains. The formulas for some of the most common fatty acids are shown in Table 28.1. All these fatty acids are straight-chain compounds with an even number of carbon atoms. Seven of the fatty acids in this table—palmitoleic, oleic, linoleic, linolenic, arachidonic eicosapentaenoic acid and docosahexaenoic acid—are unsaturated, having carbon–carbon double bonds in their structures. Animal and higher plant cells produce lipids in which palmitic, oleic, linoleic, and stearic acids predominate. Over one-half of plant and animal fatty acids are unsaturated plant lipids, which tend to be more unsaturated than their animal counterparts.

Double bonds impart some special characteristics to the unsaturated fatty acids. Remember that the presence of double bonds raises the possibility of geometric isomerism (Section 20.3). Unsaturated fatty acids may be either *cis* or *trans* isomers. To illustrate the effect of these double bonds on fatty acid structure, the two fatty acids in Figure 28.1 are portrayed in a simplified manner, with each of the many ¬ CH_2 ¬ groups as an apex at the intersection between two single bonds. Note that the *trans* isomer is almost a linear molecule, while the double bond in the *cis* isomer introduces a kink into the fatty acid structure. Unsaturated fatty acids found in nature are almost always *cis* isomers. These kinked fatty acids cannot stack closely together and hence do not solidify easily. As shown in Table 28.1, unsaturated fatty acids have lower melting points than saturated fatty acids of a similar size. Cooking oils are liquids at room temperature because a high percentage of their fatty acids are unsaturated. In like manner, biological membranes are very fluid because of the presence of fatty acid *cis* isomers. (See Section (.28.6



cis-oleic acid

trans-oleic acid

Figure 28.1
Cis- and *trans*-oleic acids. Notice the bent structure of the *cis* isomer.

Scientists continue to learn more about the nutritional importance of unsaturated fatty acids. Recently it has become clear that *trans* unsaturated fatty acids are harmful. Consumption of these isomers can cause cardiovascular problems, and the food industry has moved to eliminate *trans* fats from all prepared foods. The Nutrition Facts labels on food packages must report the *trans* fat content.

Certain specific polyunsaturated fatty acids are required in the human diet for good health; these are essential fatty acids. They fall into two categories, the omega-3 and omega-6 fatty acids, based on how close the last double bond is to the terminal methyl group [the omega (m) position]. Linoleic acid is a common omega-6 fatty acid

This unusual numbering system emphasizes double-bond locations that make these fatty acids essential. Other omega-3 and omega-6 fatty acids are listed in Table 28.1. Good health requires a diet that supplies both classes of these polyunsaturated fatty acids. It has only recently become clear why these fatty acids are essential.

Eicosanoids

Certain fatty acids, as well as other lipids, are biochemical precursors of several classes of hormones. The well-known steroid hormones are synthesized from cholesterol; they will be discussed later in this chapter. The omega-6 and omega-3 fatty acids are also used to make hormones, the most common of which are the eicosanoids. These hormones are derived from fatty acids with 20 carbon atoms, either the omega-6 (arachidonic acid) or the omega-3 (eicosapentaenoic acid). The term *eicosanoid* is derived from the Greek word for “twenty” (*eikosi*). Prostaglandins are perhaps the best known of the eicosanoid class, which also includes leukotrienes, prostacyclins, and thromboxanes.

Synthesis of these hormones starts with the release of either omega-6 or omega-3 fatty acids from the cell membranes. Many different environmental changes can cause release of these fatty acids. Examples include a local infection, an allergic reaction, or tissue damage. Once the fatty acids are released, enzymes add oxygen to the double bonds in a complex series of reactions. (See Figure 28.2.) The enzyme cyclooxygenase starts the process that leads to prostaglandins, thromboxanes, and prostatcyclins. As the enzyme’s name suggests, all the product structures are partly cyclic. The enzyme lipoxygenase starts a parallel set of reactions that lead to the fourth class of

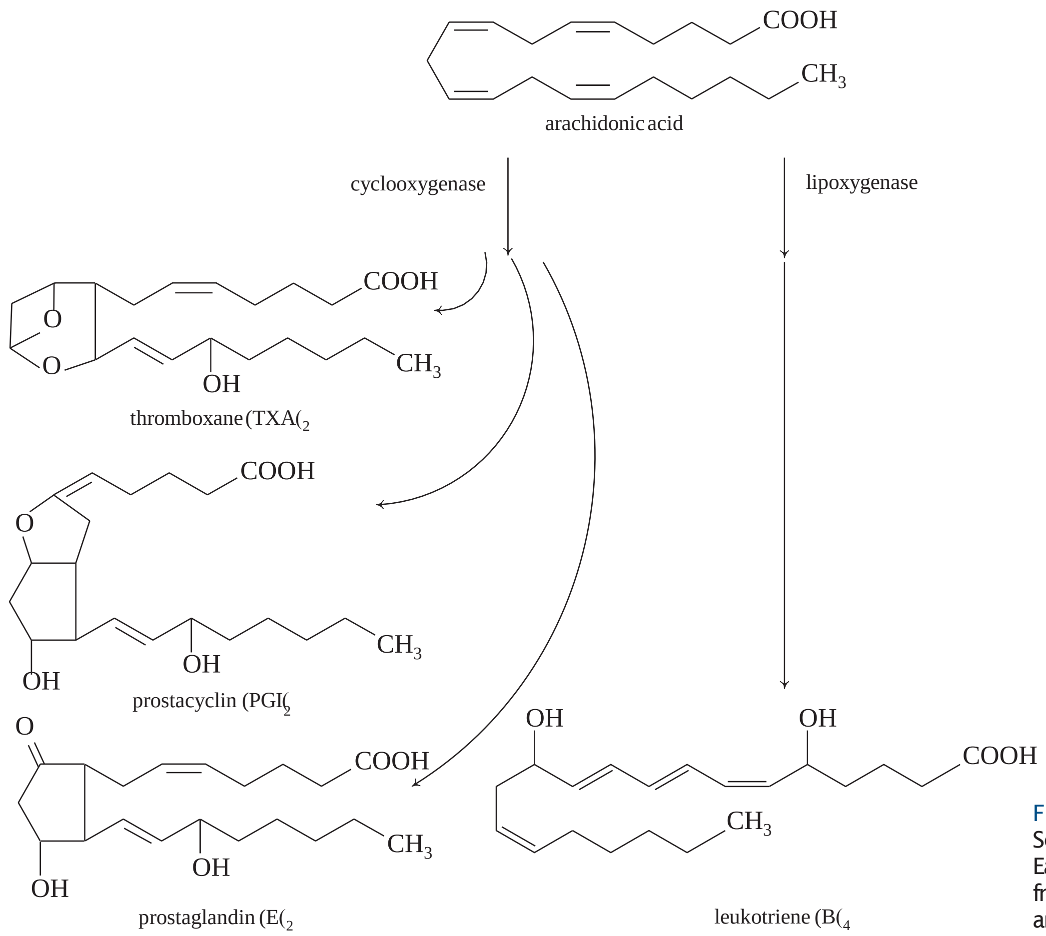


Figure 28.2
Several examples of eicosanoids. Each of these molecules is derived from the omega-6 fatty acid, arachidonic acid.

Fifth lecture

Amino Acids, Polypeptides, and Proteins

The spider has three sets of spinnerets that produce protein-containing fluids that harden as they are drawn out to form silk threads.

Chapter Outline

1. The Structure–Function Connection
2. The Nature of Amino Acids
3. Essential Amino Acids
4. D-Amino Acids and L-Amino Acids
5. Amphoterism
6. Formation of Polypeptides
7. Protein Structure
8. Protein Functions
9. Some Examples of Proteins and Their Structures

29.10 Loss of Protein Structure

11. Tests for Proteins and Amino Acids
12. Determination of the Primary Structure of Polypeptides

Proteins are present in every living cell. Their very name, derived from the Greek word *proteios*, which means “holding first place,” signifies the importance of these substances. Think of the startling properties of these molecules. Spider-web protein is many times stronger than the toughest steel; hair, feathers, and hooves are all made from one related group of proteins; another protein provides the glass-clear lens material found in the eyes and needed for vision. Very small quantities (milligram amounts) of certain proteins missing from the blood can signify that a person’s metabolic processes are out of control. Juvenile-onset diabetes mellitus results from a lack of the insulin protein. Dwarfism can arise when the growth hormone protein is lacking. A special “antifreeze” blood protein allows Antarctic fish to survive at body temperatures below freezing.

This list could go on and on, but what is perhaps most amazing is that the great variety of proteins are made from the same, relatively small, group of amino acids. By using various amounts of these amino acids in different sequences, nature creates biochemical compounds that are essential to the many functions needed to sustain life. Another aspect of proteins is their importance in our nourishment. Proteins are one of the three major classes of foods. The other two, carbohydrates and fats, are needed for energy; proteins are needed for growth and maintenance of body tissue. Some common foods with high (over 10%) protein content are fish, beans, nuts, cheese, eggs, poultry, and meat. These foods tend to be scarce and relatively expensive. Therefore, proteins are the class of foods that is least available to the undernourished people of the world. Hence, the question of how to secure an adequate supply of high-quality protein for an ever-increasing population is one of the world’s

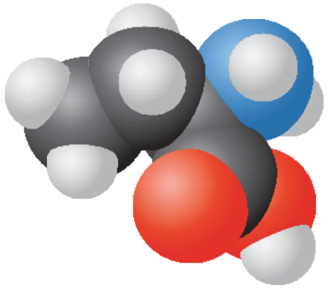
29.1 The Structure–Function Connection

Proteins function as structural materials and as enzymes (catalysts) that regulate the countless chemical reactions taking place in every living organism, including the reactions involved in the decomposition and synthesis of proteins.

All proteins are polymeric substances that yield amino acids on hydrolysis. Those that yield only amino acids when hydrolyzed are classified as **simple proteins**; those that yield amino acids and one or more additional products are classified as **conjugated proteins**. There are approximately 200 different known amino acids in nature. Some are found in only one particular species of plant or animal, others in only a few life-forms. But 20 of these amino acids are found in almost all proteins. Furthermore, these same 20 amino acids are used by all forms of life in the synthesis of proteins. All proteins contain carbon, hydrogen, oxygen, and nitrogen. Some proteins contain additional elements, usually sulfur, phosphorus, iron, copper, or zinc. The significant presence of nitrogen in all proteins sets them apart from carbohydrates and lipids. The average nitrogen content of proteins is about .16%

Proteins are highly specific in their functions. The amino acid units in a given protein molecule are arranged in a definite sequence. An amazing fact about proteins is that, in some cases, if just one of the hundreds or thousands of amino acid units is missing or out of place, the biological function of that protein is seriously damaged or destroyed.

produces a protein strong enough to form a horse's hoof; a different sequence produces a protein capable of absorbing oxygen in the lungs and releasing it to needy cells; and yet another sequence produces a hormone capable of directing carbohydrate metabolism for an entire organism. Full understanding of the function of a protein requires an understanding of its structure.

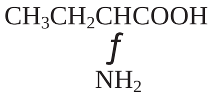


Spacefilling model of a-amino butyric acid.

29.2The Nature of Amino Acids

Each amino acid has at least two functional groups: an amino group (¬ NH_2) and a carboxyl group (¬ COOH). The amino acids found in proteins are called **alpha (a) amino acids** because the amino group is attached to the first or (a) carbon atom adjacent to the carboxyl group. The beta (þ) position is the next adjacent carbon, the gamma (μ) position the next, and so on. The following formula represents an a-amino acid:

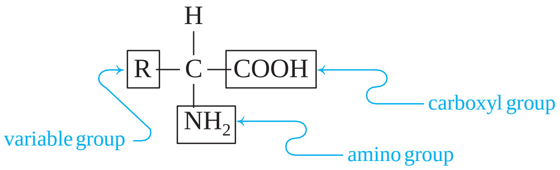
alpha (a) amino acid



$$\underset{\displaystyle NH_2}{CH_3CH_2\overset{}{C}HCOOH}$$

a-amino butyric acid

Amino acids as a whole are represented by this general formula:



The portion of the molecule designated R is commonly referred to as the *amino acid side chain*. It is not restricted to alkyl groups and may contain (a) open-chain, cyclic, or aromatic hydrocarbon groups; (b) additional amino or carboxyl groups; (c) hydroxyl groups; or (d) sulfur-containing groups.

Amino acids are divided into four groups based on the characteristics of the amino acid side chains. This classification has been chosen to emphasize the importance of the side chains in protein structure.

1. *Nonpolar amino acids*: The side chain for each of these amino acids is either aliphatic or aromatic in nature and is hydrophobic. These amino acids don't interact with water well and are commonly found buried in the middle of a protein structure.
2. *Polar, uncharged amino acids*: The side chains of these amino acids contain polar bonds from functional groups such as alcohols or amides. These groups are attracted to water and are, thus, hydrophilic. The exterior surface of water-soluble proteins is coated with these amino acids.
3. *Acidic amino acids*: Each of these amino acids has a side chain that contains a carboxylic acid. So the side chain can donate a hydrogen ion and is classified as an acid. Under common physiological conditions ($pH = 7$), these side chains lose their hydrogen ions and form carboxylate anions. Acidic amino acids are also known as "negatively charged" amino acids because of this trait.
4. *Basic amino acids*: The side chain of each of these amino acids contains a nitrogen that can act as a base. These side chains can accept a hydrogen ion under common physiological conditions ($pH = 7$) and become cations. Thus, basic amino acids are also known as "positively charged" amino acids.

Cerebrum nerve synapse. Neurotransmitters, many of which are amino acids, cross these synapses to send messages between cells.

The names, formulas, and abbreviations of the common amino acids are given in Table .29.1Two of the these (aspartic acid and glutamic acid) are classified as acidic, three (lysine, arginine, and histidine) are basic, and the remainder as neutral amino acids.

Sixth lecture

TABLE 29.1 | Common Amino Acids Derived from Proteins

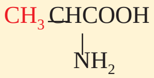
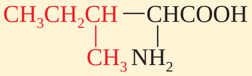
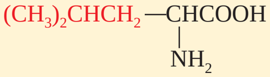
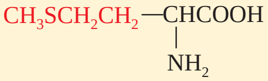
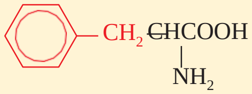
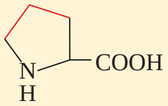
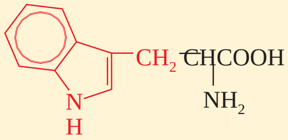
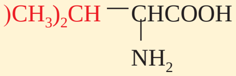
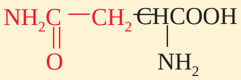
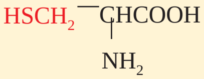
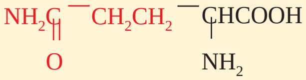
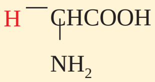
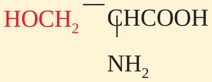
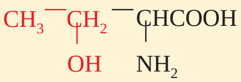
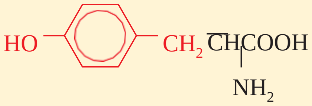
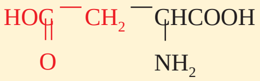
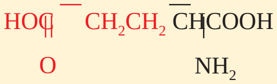
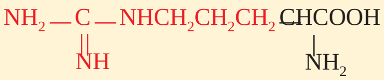
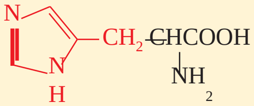
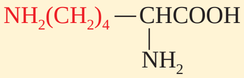
Name	Three-letter abbreviation	One-letter abbreviation	Formula
Nonpolar			
Alanine	Ala	A	 $CH_3\underset{NH_2}{CH}COOH$
Isoleucine	Ile	I	 $CH_3CH_2\underset{CH_3}{CH}-\underset{NH_2}{CH}COOH$
Leucine	Leu	L	 $(CH_3)_2CHCH_2-\underset{NH_2}{CH}COOH$
Methionine	Met	M	 $CH_3SCH_2CH_2-\underset{NH_2}{CH}COOH$
Phenylalanine	Phe	F	 $-CH_2\underset{NH_2}{CH}COOH$
Proline	Pro	P	
Tryptophan	Try	W	 $CH_2\underset{NH_2}{CH}COOH$
Valine	Val	V	 $)CH_3)_2CH-\underset{NH_2}{CH}COOH$
Polar, Uncharged			
Asparagine	Asp	N	 $NH_2\underset{O}{\underset{\|}{C}}-CH_2\underset{NH_2}{CH}COOH$
Cysteine	Cys	C	 $HSCH_2-\underset{NH_2}{CH}COOH$
Glutamine	Gln	Q	 $NH_2\underset{O}{\underset{\|}{C}}-CH_2CH_2-\underset{NH_2}{CH}COOH$
Glycine	Gly	G	 $H-\underset{NH_2}{CH}COOH$
Serine	Ser	S	 $HOCH_2-\underset{NH_2}{CH}COOH$
Threonine	Thr	T	 $CH_3-\underset{OH}{CH_2}-\underset{NH_2}{CH}COOH$
Tyrosine	Tyr	Y	 $HO-\ \ -CH_2\underset{NH_2}{CH}COOH$
Acidic			
Aspartic acid	Asp	D	 $HO\underset{O}{\underset{\|}{C}}-CH_2-\underset{NH_2}{CH}COOH$
Glutamic acid	Glu	E	 $HO\underset{O}{\underset{\|}{C}}-CH_2CH_2\underset{NH_2}{CH}COOH$

TABLE 29.1 | Continued

Name	Three-letter abbreviation	One-letter abbreviation	Formula
Basic			
Arginine	Ar	R	 $\mathrm{NH_2-\underset{\underset{NH}{\|}}{C}-NHCH_2CH_2CH_2\underset{\underset{NH_2}{\vert}}{C}HCOOH}$
Histidine	g	H	
	His		
Lysine		K	 $\mathrm{NH_2(CH_2)_4-\underset{\underset{NH_2}{\vert}}{C}HCOOH}$
	Lys		

Perhaps the most important role played by amino acids is as the building blocks for proteins. However, selected amino acids also have physiological importance on their own. Many neurotransmitters are amino acids or their derivatives. Glycine and glutamic acids function as chemical messengers between nerve cells in some organisms. Tyrosine is converted to the very important neurotransmitter dopamine. A deficiency of this amino acid derivative causes Parkinson's disease, which can be relieved by another compound formed from tyrosine, L-dopa. Tyrosine is also the parent compound for the "flight-or-fight" hormone epinephrine (adrenaline) and the metabolic hormone thyroxine. Histidine, which is converted in the body to histamine, is yet another amino acid with an important physiological role. Histamine causes the stomach lining to secrete HCl but is probably best known for causing many of the symptoms associated with tissue inflammation and colds; this is the reason antihistamines are such popular and important over-the-counter medications.

Antihistamines lessen the symptoms of colds and allergies by reducing tissue inflammation.

Practice 29.1

Write the structures for the simplest a-amino acid and þ-amino acid.

29.3 Essential Amino Acids

During digestion, protein is broken down into its constituent amino acids, which supply much of the body's need for amino acids. (See Chapter 32.) Ten of the amino acids are **essential amino acids** (see Table 29.2) because they are essential to the normal functioning of the human body. Since the body is not capable of synthesizing them, they must be supplied in our diets if we are to enjoy normal health. Some animals require other amino acids in addition to those listed for humans.

TABLE 29.2 | Essential Amino Acids for Humans

Arginine
Histidine
Isoleucine
Leucine
Lysine
Methionine
Phenylalanine
Threonine
Tryptophan
Valine

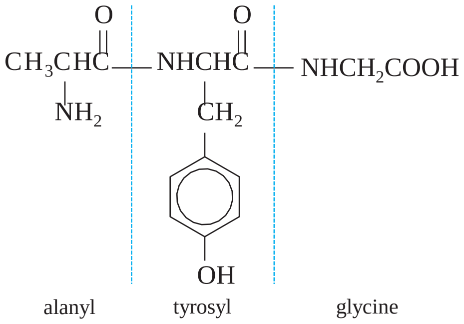
essential amino acids

On a nutritional basis, proteins are classified as *complete* or *incomplete*. A complete protein supplies all the essential amino acids; an incomplete protein is deficient in one or more essential amino acids. Many proteins, especially those from vegetable sources, are incomplete. For example, protein from corn (maize) is deficient in lysine. The nutritional quality of such vegetable proteins can be greatly improved by supplementing them with the essential amino acids that are lacking, if these can be synthesized at reasonable costs. Lysine, methionine, and tryptophan are now used to enrich human food and livestock feed as a way to extend the world's limited supply of high-quality food protein. In still another approach to the problem of obtaining more high-quality protein, plant breeders have developed maize varieties with greatly improved lysine content. Genetic engineering may hold the key to further significant improvements in plant protein quality.

Vegetarians must carefully choose combinations of foods that include all essential amino acids.

Seventh lecture

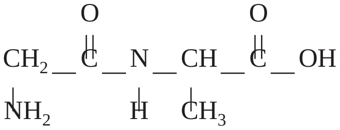
Peptides are named as acyl derivatives of the C-terminal amino acid, with the C-terminal unit keeping its complete name. The *-ine* ending of all but the C-terminal amino acid is changed to *-yl*, and these are listed in the order in which they appear, starting with the N-terminal amino acid:



Ala-Tyr-Gly

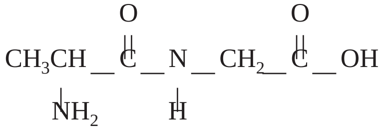
Hence, Ala-Tyr-Gly is called alanyltyrosylglycine. The name of Arg-Glu-His-Ala is arginylglutamylhistidylalanine.

Alanine and glycine can form two different dipeptides, Gly-Ala and Ala-Gly, using each amino acid only once:



$$H_2N{-}CH_2{-}C(=O){-}NH{-}CH(CH_3){-}C(=O){-}OH$$

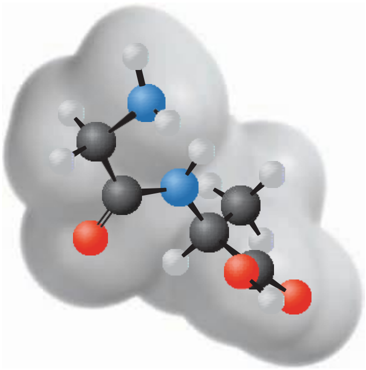
and



$$CH_3CH(NH_2){-}C(=O){-}NH{-}CH_2{-}C(=O){-}OH$$

glycylalanine (Gly-Ala)

alanylglycine (Ala-Gly(



Spacefilling model of glycylalanine: Can you see the amide bond?

If three different amino acids react—for example, glycine, alanine, and threonine—six tripeptides, in which each amino acid appears only once, are possible:

Gly-Ala-Thr
Gly-Thr-Ala
Ala-Thr-Gly
Ala-Gly-Thr
Thr-Ala-Gly
Thr-Gly-Ala

TABLE 29.4 | Primary Structures and Functions of Some Biological Polypeptides*

Name	Primary structure	General biological function
Substance P	Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH_2	Is a pain-producing agent
Bradykinin	Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg	Affects tissue inflammation and blood pressure
Angiotensin II	Asp-Arg-Val-Tyr-Val-His-Pro-Phe	Maintains water balance and blood pressure
Leu-enkephalin Met-enkephalin	Tyr-Gly-Gly-Phe-Leu Tyr-Gly-Gly-Phe-Met	Relieves pain, produces sense of well-being
Vasopressin	S-S Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly-NH_2	Increases blood pressure, decreases kidney water excretion
Oxytocin	S-S Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH_2	Initiates childbirth labor, causes mammary gland milk release, affects kidney excretion of water and sodium
Thyrotropin-releasing hormone	Glu-His-Pro	Stimulates release of hormones from the pituitary
Orexin	33amino acids long	Causes wakefulness
Neuropeptide FF	Phe-Leu-Phe-Gln-Pro-Gln-Arg-Phe-NH_2	Modulates pain sensations
Neurotensin	Glu-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu	Involved in brain memory functions
Neuropeptide Y	36amino acids long	Stimulates eating
Endomorphin	Tyr-Pro-Trp-Phe-NH_2	Acts as a morphine-like analgesic

The number of possible peptides rises very rapidly as the number of amino acid units increases. For example, there are 120 (1 * 2 * 3 * 4 * 5 = 120) different ways to combine five different amino acids to form a pentapeptide, using each amino acid only once in each molecule. If the same constraints are applied to 15 different amino acids, the number of possible combinations is greater than 1 trillion (10^{12})! Since a protein molecule can contain several hundred amino acid units, with individual amino acids occurring several times, the number of possible combinations from 20 amino acids is simply beyond imagination.

There are a number of small, naturally occurring polypeptides with significant biochemical functions. (Over 30 different peptides are known at present.) In general, these substances serve as hormones or nerve transmitters. Their functions range from controlling pain and pleasure responses in the brain to controlling smooth muscle contraction or kidney fluid excretion rates. (See Table (.29.4

The amino acid sequence and chain length give a polypeptide its biological effectiveness and specificity. For example, recent research has shown that the effects of opiates (opium derivatives) on the brain are also exhibited by two naturally occurring pentapeptides: Leu-enkephalin (Tyr-Gly-Gly-Phe-Leu) and Met-enkephalin (Tyr-Gly-Gly-Phe-Met). These two pentapeptides are natural painkillers. Alterations of the amino acid sequence—which alter the side-chain characteristics—cause drastic changes in the analgesic effects of Leu-enkephalin and Met-enkephalin. The substitution of L-alanine for either of the glycine residues in these compounds (simply changing one side-chain group from ¬ H to ¬ CH_3) causes approximately a thousand-fold decrease in their effectiveness as painkillers! The substitution of L-tyrosine for L-phenylalanine causes a comparable loss of activity. Even the substitution of D-tyrosine for the L-tyrosine residue causes a considerable loss in the analgesic effectiveness of the pentapeptides.

It is clearly evident that a particular sequence of amino acid residues is essential for proper polypeptide function. This sequence aligns the side-chain characteristics (large or small; polar or nonpolar; acidic, basic, or neutral) in the proper positions for a specific polypeptide function.

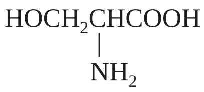
Oxytocin and vasopressin are similar nonapeptides, differing at only two positions in their primary structure. (See Table 29.4.) Yet their biological functions differ dramatically. Oxytocin controls uterine contractions during labor in childbirth and also causes contraction of the smooth muscles of the mammary glands, resulting in milk excretion. Vasopressin in high concentration raises the blood pressure and has been used to treat surgical shock. Vasopressin is also an antidiuretic, regulating the excretion of fluid by the kidneys. The absence of vasopressin leads to diabetes insipidus. This condition is characterized by the excretion of up to 30 L of urine per day, but it can be controlled by the administration of vasopressin or any of its derivatives.

The isolation and synthesis of oxytocin and vasopressin was accomplished by Vincent du Vigneaud (1901–1978) and coworkers at Cornell University. Du Vigneaud was awarded the Nobel Prize in chemistry in 1955 for this work. Synthetic oxytocin is indistinguishable from the natural material. It is available commercially and is used to induce labor in the late stages of pregnancy.

Example 29.2

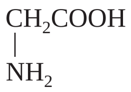
Write the structure of the tripeptide Ser-Gly-Ala.

SOLUTION First, write the structures of the three amino acids in this tripeptide:



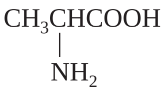
$$\mathrm{HOCH_2\underset{\underset{\displaystyle NH_2}{|}}{C}HCOOH}$$

serine (Ser(



$$\mathrm{\underset{\underset{\displaystyle NH_2}{|}}{C}H_2COOH}$$

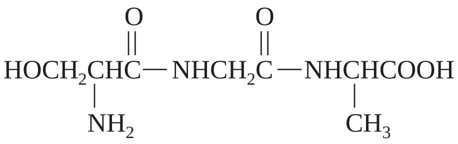
glycine (Gly(



$$\mathrm{CH_3\underset{\underset{\displaystyle NH_2}{|}}{C}HCOOH}$$

alanine (Ala(

By convention, the amino acid residue written at the left end of the tripeptide has a free amino group, while the residue at the right end has a free carboxylic acid group. Now split out a water molecule in two places: between the carboxyl group of serine and the amino group of glycine and between the carboxyl group of glycine and the amino group of alanine. When the amino acids are connected by peptide linkages, the following structure results:



$$\mathrm{HOCH_2\underset{\underset{\displaystyle NH_2}{|}}{C}H\overset{\overset{\displaystyle O}{\|}}{C}{-}NHCH_2\overset{\overset{\displaystyle O}{\|}}{C}{-}NH\underset{\underset{\displaystyle CH_3}{|}}{C}HCOOH}$$

Practice 29.4 Organic Review (see Section (25.3

Amide links form when carboxylic acids react with amines. Show the reaction between Gly, Asp, and Ser to form the tripeptide Gly-Asp-Ser. Remember to show all reactants and products.

More Practice?
Try Paired Exercises 9–.20

Eighth lecture

Nucleic Acids and Heredity

The science of genetics began when Gregor Mendel studied pea plants in the mid-nineteenth century.

Chapter Outline

The advent of genetic engineering has raised a similar specter. Genetic engineers work with the molecules that code life—the nucleic acids. By changing the code, they can produce new life-forms. Already, bacteria have been altered to make needed human proteins. Recently, both cows and goats have been genetically engineered to produce a human protein in their milk.

The scientists involved in these initial programs have followed careful protocols and have produced valuable medicines. However, the day may come when we can decide whether humans should be made smarter or stronger via genetic engineering. How this decision will be made and what it will be are the topics of heated discussions and much controversy; the potential ramifications are enormous—and not only in the scientific community. The power to even consider such decisions and to possibly open "Pandora's box" in ways that were once only the stuff of science fiction is the result of our understanding of the biochemistry of nucleic acids.

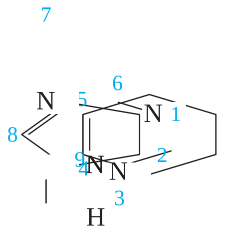
31.1 Molecules of Heredity—A Link

The question of how hereditary material duplicates itself was one of the most baffling problems of biology for many years. Generations of biologists attempted in vain to solve this problem and to answer the question "Why are the offspring of a species undeniably of that species?" Many thought the chemical basis for heredity lay in the structure of the proteins. But no one was able to provide evidence showing how protein could reproduce itself. The answer to the heredity question was finally found in the structure of the nucleic acids.

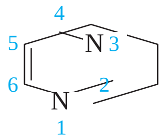
The unit structure of all living things is the cell. Suspended in the nuclei of cells are chromosomes, which consist largely of proteins and nucleic acids. A simple protein bonded to a nucleic acid is called a **nucleoprotein**. **Nucleic acids** are polymers of nucleotides and contain either the sugar *deoxyribose* or the sugar *ribose*. Accordingly, they are called deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), respectively. Although many of us think of DNA as a recent discovery, it was actually discovered in 1869 by the Swiss physiologist Friedrich Miescher (1844–1895), who extracted it from the nuclei of cells.

31.2 Bases and Nucleosides

Nucleic acids are complex chemicals that combine several different classes of smaller molecules. As with many complex structures, it is easier to understand the whole by first studying its component parts. We begin our examination of nucleic acids by learning about a critical part of these molecules, two classes of heterocyclic bases called the *purines* and the *pyrimidines*:

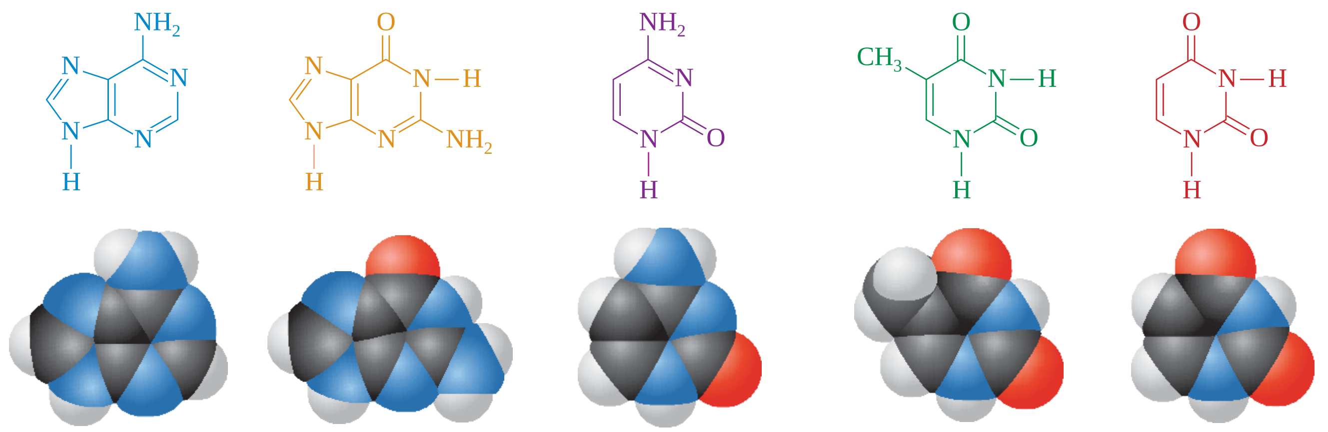


purine, $C_5H_4N_4$



pyrimidine, $C_4H_4N_2$

The parent compounds are related in structure: The pyrimidine is a six-membered heterocyclic ring, while the purine contains both a five- and six-membered ring. The nitrogen atoms cause these compounds to be known as *heterocyclics* (the rings are



adenine
-6)aminopurine(

guanine
-2)amino-6-oxypurine(

cytosine
-2)oxy-4-amino-pyrimidine(

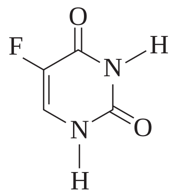
thymine
-2,4)dioxy-5-methyl-pyrimidine(

uracil
-2,4)dioxy-pyrimidine(

Figure 31.1
Purine and pyrimidine bases found in living matter.

made up of more than just carbon atoms) and also as bases. Like the ammonia nitrogen, heterocyclics react with hydrogen ions to make a solution more basic.

Five major bases are commonly found in nucleic acids—two purine bases (adenine and guanine) and three pyrimidine bases (cytosine, thymine, and uracil). Figure 31.1 gives one stable form for each compound. These bases must be available to the cell in order to reproduce genetic information (DNA). The importance of the bases is underscored by their use in cancer chemotherapy. Specific "modified" bases are used to kill fast-growing cancer cells. Perhaps the most common anticancer drug in this category is 5-fluorouracil:



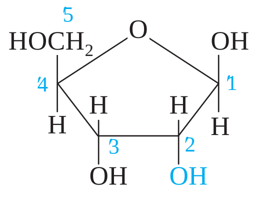
-5fluorouracil

Note that 5-fluorouracil is similar to thymine: It acts to inhibit the enzyme that catalyzes the formation of thymine. Cancer cells require a rapid synthesis of DNA. Without sufficient thymine to form DNA, cancer cells die.

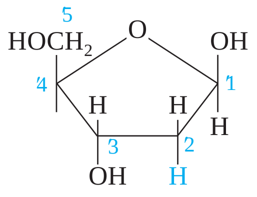
The natural bases differ one from another in their ring substituents. Each base has a lowermost nitrogen, which is bonded to a hydrogen as well as two carbons. This specific ¬ NH shares chemical similarities with an alcohol (¬ OH) group. Just as two sugars molecules can be linked when an alcohol of one monosaccharide reacts with a second monosaccharide (see Chapter 27), so a purine or pyrimidine molecule can be bonded to a sugar molecule by a reaction with the ¬ NH group.

A **nucleoside** is formed when either a purine or pyrimidine base is linked to a sugar molecule, usually D-ribose or D-2¿-deoxyribose:

nucleoside

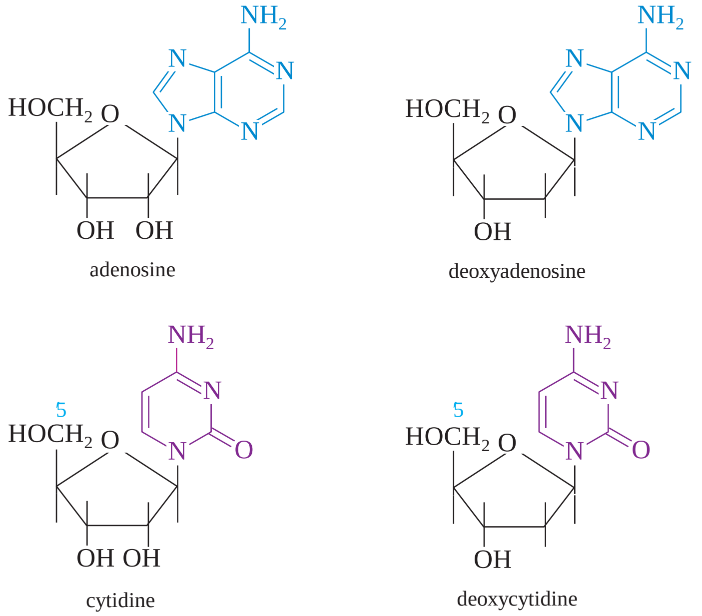


D-ribose



D-2 ′deoxyribose

Ninth lecture



adenosine deoxyadenosine

cytidine deoxycytidine

Figure 31.2
Structures of typical ribonucleosides and deoxyribonucleosides.

The base and sugar are bonded together between carbon 1¿ of the sugar and either the purine nitrogen at position 9 or the pyrimidine nitrogen at position 1 by splitting out a molecule of water. Typical structures of nucleosides are shown in Figure 31.2. A prime is added to the position number to differentiate the sugar numbering system from the purine or pyrimidine numbering system.

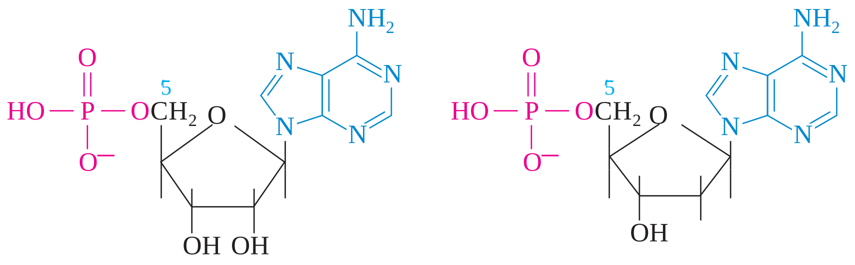
The name of each nucleoside emphasizes the importance of the base to the chemistry of the molecule. Thus, adenine and D-ribose react to yield adenosine, whereas cytosine and D-2¿-deoxyribose yield deoxycytidine. The root of the nucleoside name derives from the purine or pyrimidine name. The compositions of the common ribonucleosides and the deoxyribonucleosides are given in Table .31.1

Practice 31.1

Draw the structures of (a) guanosine and (b) deoxycytidine.

TABLE 31.1 | Composition of Ribonucleosides and Deoxyribonucleosides

Name	Composition	Abbreviation
Adenosine	Adenine–ribose	A
Deoxyadenosine	Adenine–deoxyribose	dA
Guanosine	Guanine–ribose	G
Deoxyguanosine	Guanine–deoxyribose	dG
Cytidine	Cytosine–ribose	C
Deoxycytidine	Cytosine–deoxyribose	dC
Thymidine	Thymine–ribose	T
Deoxythymidine	Thymine–deoxyribose	dT
Uridine	Uracil–ribose	U
Deoxyuridine	Uracil–deoxyribose	dU



adenosine-5′-monophosphate
(AMP(

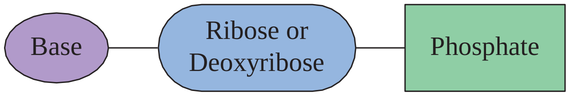
deoxyadenosine-5′-monophosphate
(dAMP(

Figure 31.3
Examples of nucleotides.

31.3 Nucleotides: Phosphate Esters

A more complex set of biological molecules is formed by linking phosphate groups to nucleosides. Phosphate esters of nucleosides are termed **nucleotides**. These molecules consist of a purine or a pyrimidine base linked to a sugar, which in turn is bonded to at least one phosphate group:

nucleotide

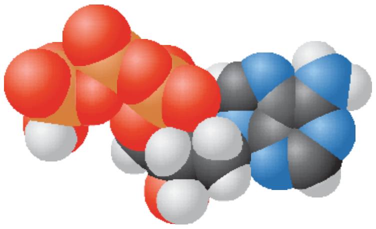


The ester may be a monophosphate, a diphosphate, or a triphosphate. When two or more phosphates are linked together, a high-energy phosphate anhydride bond is formed. (See Section 31.4.) The ester linkage may be to the hydroxyl group of position 2¿, 3¿, or 5¿ of ribose or to position 3¿ or 5¿ of deoxyribose. Examples of nucleotide structures are shown in Figure .31.3

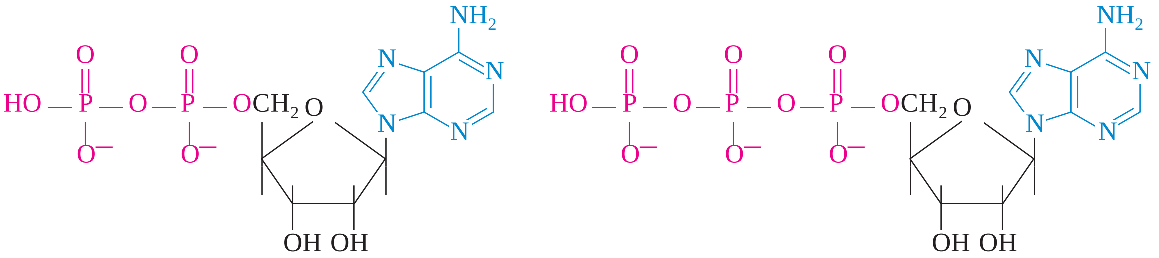
Nucleotide abbreviations start with the corresponding nucleoside abbreviation. (See Table 31.1.) The letters MP (monophosphate) can be added to any of these to designate the corresponding nucleotide. Thus, GMP is guanosine monophosphate. A lowercase d is placed in front of GMP if the nucleotide contains the deoxyribose sugar (dGMP). When the letters such as AMP or GMP are given, it is generally understood that the phosphate group is attached to position 5¿ of the ribose unit (5¿-AMP). If attachment is elsewhere, it will be designated, for example, as 3¿-AMP.

Two other important adenosine phosphate esters are **adenosine diphosphate (ADP)** and **adenosine triphosphate (ATP)**. Note that the letters DP are used for diphosphate and TP for triphosphate. In these molecules, the phosphate groups are linked together. The structures are similar to AMP, except that they contain two and three phosphate residues, respectively. (See Figure 31.4.) All the nucleosides form mono-, di-, and triphosphate nucleotides.

ADP (adenosine diphosphate)
ATP (adenosine triphosphate(



Spacefilling model of ATP.



adenosine-5′-diphosphate
(ADP(

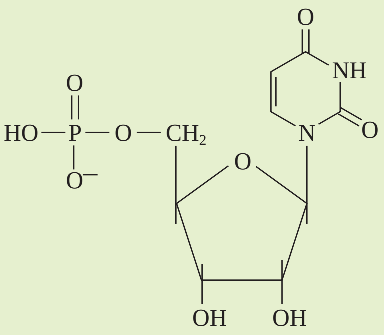
adenosine-5′-triphosphate
(ATP(

Figure 31.4
Structures of ADP and ATP.

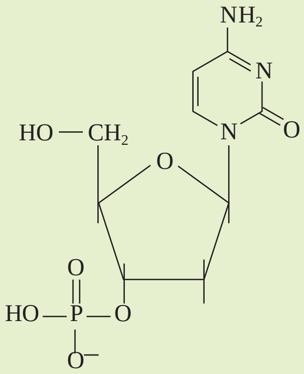
Tenth lecture

Practice 31.2

Name and give abbrevations for the following structures:



)a(

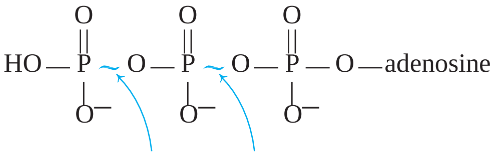


)b(

31.4 High-Energy Nucleotides

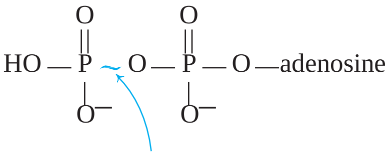
Muscle movements, including heartbeats, depend on energy from ATP.

Nucleotides have a central role in the energy transfers in many metabolic processes. ATP and ADP are especially important in these processes, since these two nucleotides store and release energy to the cells and tissues. The source of energy is the foods we eat, particularly carbohydrates and fats. Energy is released as the carbons from these foods are oxidized. (See Chapter 33.) Part of this energy is used to maintain body temperature, and part is stored in the phosphate anhydride bonds of such molecules as ADP and ATP. Because a relatively large amount of energy is stored in these bonds, they are known as high-energy phosphate anhydride bonds:



$$\mathrm{HO{-}\overset{\overset{O}{\parallel}}{\underset{\underset{O^-}{|}}{P}}\sim O{-}\overset{\overset{O}{\parallel}}{\underset{\underset{O^-}{|}}{P}}\sim O{-}\overset{\overset{O}{\parallel}}{\underset{\underset{O^-}{|}}{P}}{-}O{-}adenosine}$$

High-energy phosphate anhydride bonds
(ATP(



$$\mathrm{HO{-}\overset{\overset{O}{\parallel}}{\underset{\underset{O^-}{|}}{P}}\sim O{-}\overset{\overset{O}{\parallel}}{\underset{\underset{O^-}{|}}{P}}{-}O{-}adenosine}$$

High-energy anhydride bond
(ADP(

Energy is released during the hydrolysis of high-energy phosphate anhydride bonds in ADP and ATP. In the hydrolysis, ATP forms ADP and inorganic phosphate (P_i), yielding about 35 kJ of energy per mole of ATP:

$$\mathrm{ATP + H_2O \xrightleftharpoons[\text{energy storage}]{\text{energy utilization}} ADP + P_i + {\sim}35\,kJ}$$

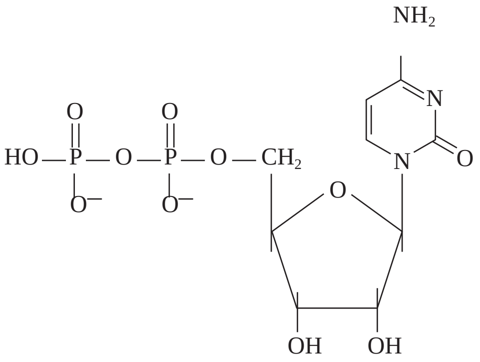
The hydrolysis reaction is reversible, with ADP being converted to ATP by still higher energy molecules. In this manner, energy is supplied to the cells from ATP, and energy is stored by the synthesis of ATP. Processes such as muscle movement, nerve sensations, vision, and even the maintenance of our heartbeats are all dependent on energy from ATP.

Example 31.1

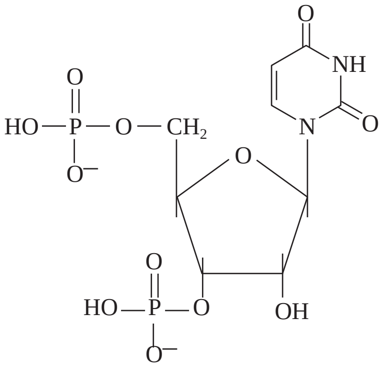
Draw the structures of (a) CDP and (b) uridine-3¿,5¿-diphosphate.

SOLUTION)a) When no numbers are given in a nucleotide abbreviation, it is understood that the phosphates are connected at the 5¿-position (i.e,.

CDP = cytidine-5¿-diphosphate). In the absence of the prefix *deoxy-* (abbreviation d-), ribose is the sugar. Cytidine (C) signals the base, cytosine, connected to the sugar. Finally, diphosphate (DP) gives the number of phosphates in the nucleotide. The structure is as follows:



)b) The full name provides all the information needed to draw the nucleotide structure. In the absence of the prefix *deoxy*, the sugar is ribose. *Uridine* signals the base uracil, connected to the sugar. -¿5,¿3-*diphosphate* gives the number and locations of the phosphates. The structure is as follows:



Practice 31.3 **Organic Review (see Section (24.13**

Draw the structures for the following: (a) guanosine-5¿-diphosphate; (b) guanosine-3¿,5¿-diphosphate. Phosphate esters and anhydrides are important nucleotide components. Mark the esters with * and the anhydrides with ** in the two nucleotides you have drawn.

More Practice?
Try Paired Exercises .1-12

31.5Polynucleotides; Nucleic Acids

Starting with two nucleotides, a dinucleotide is formed by splitting out a molecule of water between the ¬ OH of the phosphate group of one nucleotide and the ¬ OH on carbon 3¿ of the ribose or deoxyribose of the other nucleotide. Then, another and another nucleotide can be added in the same manner, until a polynucleotide chain is formed. Each nucleotide is linked to its neighbors by phosphate ester bonds. (See Section (.24.13

Two series of polynucleotide chains are known, one containing D-ribose and the other D-2¿-deoxyribose. One polymeric chain consists of the monomers AMP, GMP, CMP, and UMP and is known as a *polyribonucleotide*. The other chain contains the