

Biotechnology Semester I

Course title: Biomolecules and Cell Biology

Unit I:

Water and its physicochemical properties

Water is the most abundant substance in living systems, making up 70% or more of the weight of most organisms. The first living organisms doubtless arose in an aqueous environment, and the course of evolution has been shaped by the properties of the aqueous medium in which life began. The attractive forces between water molecules and the slight tendency of water to ionize are of crucial importance to the structure and function of biomolecules. The water molecule and its ionization products, H^+ and OH^- , profoundly influence the structure, self-assembly, and properties of all cellular components, including proteins, nucleic acids, and lipids. The noncovalent interactions responsible for the strength and specificity of "recognition" among biomolecules are decisively influenced by the solvent properties of water, including its ability to form hydrogen bonds with itself and with solutes.

Physical properties

Hydrogen Bonding

Hydrogen bonds between water molecules provide the cohesive forces that make water a liquid at room temperature and that favor the extreme ordering of molecules that is typical of crystalline water (ice). Polar biomolecules dissolve readily in water because they can replace water-water interactions with more energetically favorable water-solute interactions. In contrast, nonpolar biomolecules interfere with water-water interactions but are unable to form water-solute interactions, consequently, nonpolar molecules are poorly soluble in water. In aqueous solutions, nonpolar molecules tend to cluster together. Hydrogen bonds and ionic, hydrophobic (Greek, "water-fearing"), and van der Waals interactions are individually weak, but collectively they have a very significant influence on the three-dimensional structures of proteins, nucleic acids, polysaccharides, and membrane lipids.

Water has a higher melting point, boiling point, and heat of vaporization than most other common solvents (Table 1).

TABLE 1 Melting Point, Boiling Point, and Heat of Vaporization of Some Common Solvents

	<i>Melting point (°C)</i>	<i>Boiling point (°C)</i>	<i>Heat of vaporization</i>
	<i>(J/g)*</i>		
Water	0	100	2,260
Methanol (CH_3OH)	-98	65	1,100
Ethanol ($\text{CH}_3\text{CH}_2\text{OH}$)	-117	78	854
Propanol ($\text{CH}_3\text{CH}_2\text{CH}_2\text{OH}$)	-127	97	687
Butanol ($\text{CH}_3(\text{CH}_2)_2\text{CH}_2\text{OH}$)	-90	117	590
Acetone (CH_3COCH_3)	-95	56	523
Hexane ($\text{CH}_3(\text{CH}_2)_4\text{CH}_3$)	-98	69	423
Benzene (C_6H_6)	6	80	394
Butane ($\text{CH}_3(\text{CH}_2)_2\text{CH}_3$)	-135	-0.5	381

These unusual properties are a consequence of attractions between adjacent water molecules that give liquid water great internal cohesion. A look at the electron structure of the H₂O molecule reveals the cause of these intermolecular attractions. Each hydrogen atom of a water molecule shares an electron pair with the central oxygen atom. The geometry of the molecule is dictated by the shapes of the outer electron orbitals of the oxygen atom, which are similar to the *sp*³ bonding orbitals of carbon (see Fig. 1–14). These orbitals describe a rough tetrahedron, with a hydrogen atom at each of two corners and unshared electron pairs at the other two corners (Fig. 1a). The H-O-H bond angle is 104.5°, slightly less than the 109.5° of a perfect tetrahedron because of crowding by the nonbonding orbitals of the oxygen atom.

The oxygen nucleus attracts electrons more strongly than does the hydrogen nucleus (a proton); that is, oxygen is more electronegative. The sharing of electrons between H and O is therefore unequal; the electrons are more often in the vicinity of the oxygen atom than of the hydrogen. The result of this unequal electron sharing is two electric dipoles in the water molecule, one along each of the H-O bonds; each hydrogen bears a partial positive charge and the oxygen atom bears a partial negative charge equal to the sum of the two partial positives. As a result, there is an electrostatic attraction between the oxygen atom of one water molecule and the hydrogen of another (Fig. 1c), called a **hydrogen bond**. Hydrogen bonds are relatively weak. Those in liquid water have a **bond dissociation energy** (the energy required to break a bond) of about 23 kJ/mol, compared with 470 kJ/mol for the covalent O-H bond in water or 348 kJ/mol for a covalent C-C bond.

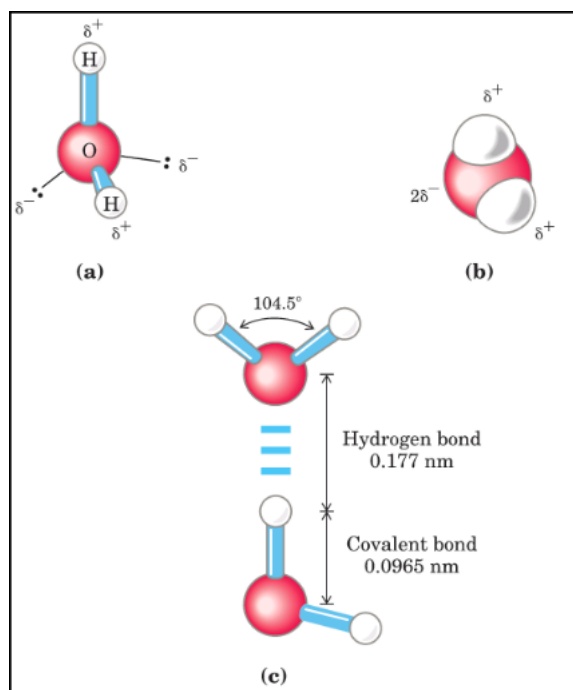


FIGURE 1 Structure of the water molecule. The dipolar nature of the H₂O molecule is shown by (a) ball-and-stick and (b) space-filling models. The dashed lines in (a) represent the nonbonding orbitals. There is

a nearly tetrahedral arrangement of the outer-shell electron pairs around the oxygen atom; the two hydrogen atoms have localized partial positive charges and the oxygen atom has a partial negative charge. (c) Two H_2O molecules joined by a hydrogen bond between the oxygen atom of the upper molecule and a hydrogen atom of the lower one. Hydrogen bonds are longer and weaker than covalent O-H bonds.

Chemical properties

Interaction of water with Polar Solutes

Hydrogen bonds are not unique to water. They readily form between an electronegative atom (the hydrogen acceptor, usually oxygen or nitrogen with a lone pair of electrons) and a hydrogen atom covalently bonded to another electronegative atom (the hydrogen donor) in the same or another molecule (Fig.2). Hydrogen atoms covalently bonded to carbon atoms do not participate in hydrogen bonding, because carbon is only slightly more electronegative than hydrogen and thus the C-H bond is only very weakly polar. The distinction explains why butanol ($\text{CH}_3(\text{CH}_2)_2\text{CH}_2\text{OH}$) has a relatively high boiling point of 117°C , whereas butane ($\text{CH}_3(\text{CH}_2)_2\text{CH}_3$) has a boiling point of only -0.5°C . Butanol has a polar hydroxyl group and thus can form intermolecular hydrogen bonds. Uncharged but polar biomolecules such as sugars dissolve readily in water because of the stabilizing effect of hydrogen bonds between the hydroxyl groups or carbonyl oxygen of the sugar and the polar water molecules. Alcohols, aldehydes, ketones, and compounds containing NOH bonds all form hydrogen bonds with water molecules (Fig.3) and tend to be soluble in water. Hydrogen bonds are strongest when the bonded molecules are oriented to maximize electrostatic interaction, which occurs when the hydrogen atom and the two atoms that share it are in a straight line—that is, when the acceptor atom is in line with the covalent bond between the donor atom and H (Fig.5). Hydrogen bonds are thus highly directional and capable of holding two hydrogen-bonded molecules or groups in a specific geometric arrangement. As we shall see later, this property of hydrogen bonds confers very precise three dimensional structures on protein and nucleic acid molecules, which have many intramolecular hydrogen bonds.

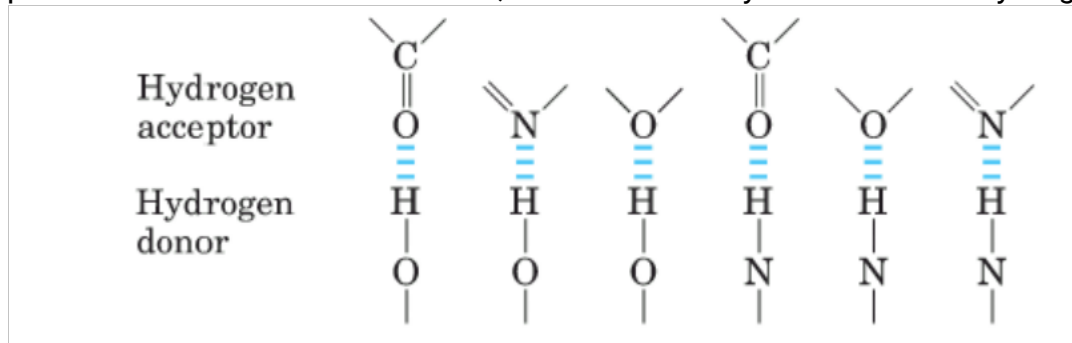


FIGURE 2 Common hydrogen bonds in biological systems. The hydrogen acceptor is usually oxygen or nitrogen; the hydrogen donor is another electronegative atom.

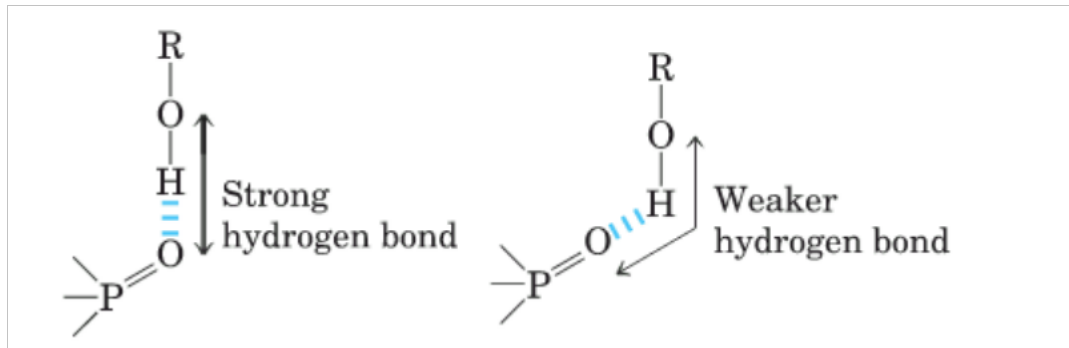


FIGURE 3 Directionality of the hydrogen bond. The attraction between the partial electric charges (see Fig. 2-1) is greatest when the three atoms involved (in this case O, H, and O) lie in a straight line. When the hydrogen-bonded moieties are structurally constrained (as when they are parts of a single protein molecule, for example), this ideal geometry may not be possible and the resulting hydrogen bond is weaker.

Interaction of water with Charged Solutes

Water is a polar solvent. It readily dissolves most biomolecules, which are generally charged or polar compounds that dissolve easily in water are **hydrophilic** (Greek, “water-loving”). In contrast, nonpolar solvents such as chloroform and benzene are poor solvents for polar biomolecules but easily dissolve those that are **hydrophobic** nonpolar molecules such as lipids and waxes. Water dissolves salts such as NaCl by hydrating and stabilizing the Na⁺ and Cl⁻ ions, weakening the electrostatic interactions between them and thus counteracting their tendency to associate in a crystalline lattice. The same factors apply to charged biomolecules, compounds with functional groups such as ionized carboxylic acids, protonated amines and phosphate esters or anhydrides. Water readily dissolves such compounds by replacing solute-solute hydrogen bonds with solute-water hydrogen bonds, thus screening the electrostatic interactions between solute molecules.

Interaction of water with nonpolar Solutes

When water is mixed with benzene or hexane, two phases form; neither liquid is soluble in the other. Nonpolar compounds such as benzene and hexane are hydrophobic—they are unable to undergo energetically favorable interactions with water molecules, and they interfere with the hydrogen bonding among water molecules. All molecules or ions in aqueous solution interfere with the hydrogen bonding of some water molecules in their immediate vicinity, but polar or charged solutes (such as NaCl) compensate for lost water-water hydrogen bonds by forming new solute-water interactions. **Amphipathic** compounds contain regions that are polar (or charged) and regions that are nonpolar. When an amphipathic compound is mixed with water, the polar, hydrophilic region interacts favorably with the solvent and tends to dissolve, but the nonpolar, hydrophobic region tends to avoid contact with the water (Fig. 4a). The nonpolar regions of the molecules cluster together to present the smallest hydrophobic area to the aqueous solvent, and the polar regions are arranged to maximize their interaction with the solvent (Fig. 4b). These stable structures of amphipathic compounds in water, called **micelles**, may contain hundreds or thousands of molecules. The forces that hold the nonpolar regions of the molecules together are called **hydrophobic interactions**. The strength of hydrophobic interactions is not due to any

intrinsic attraction between nonpolar moieties. Rather, it results from the system's achieving greatest thermodynamic stability by minimizing the number of ordered water molecules required to surround hydrophobic portions of the solute molecules. Many biomolecules are amphipathic; proteins, pigments, certain vitamins, and the sterols and phospholipids of membranes all have polar and nonpolar surface regions. Structures composed of these molecules are stabilized by hydrophobic interactions among the nonpolar regions. Hydrophobic interactions among lipids, and between lipids and proteins, are the most important determinants of structure in biological membranes. Hydrophobic interactions between nonpolar amino acids also stabilize the three-dimensional structures of proteins. Hydrogen bonding between water and polar solutes also causes some ordering of water molecules, but the effect is less significant than with nonpolar solutes.

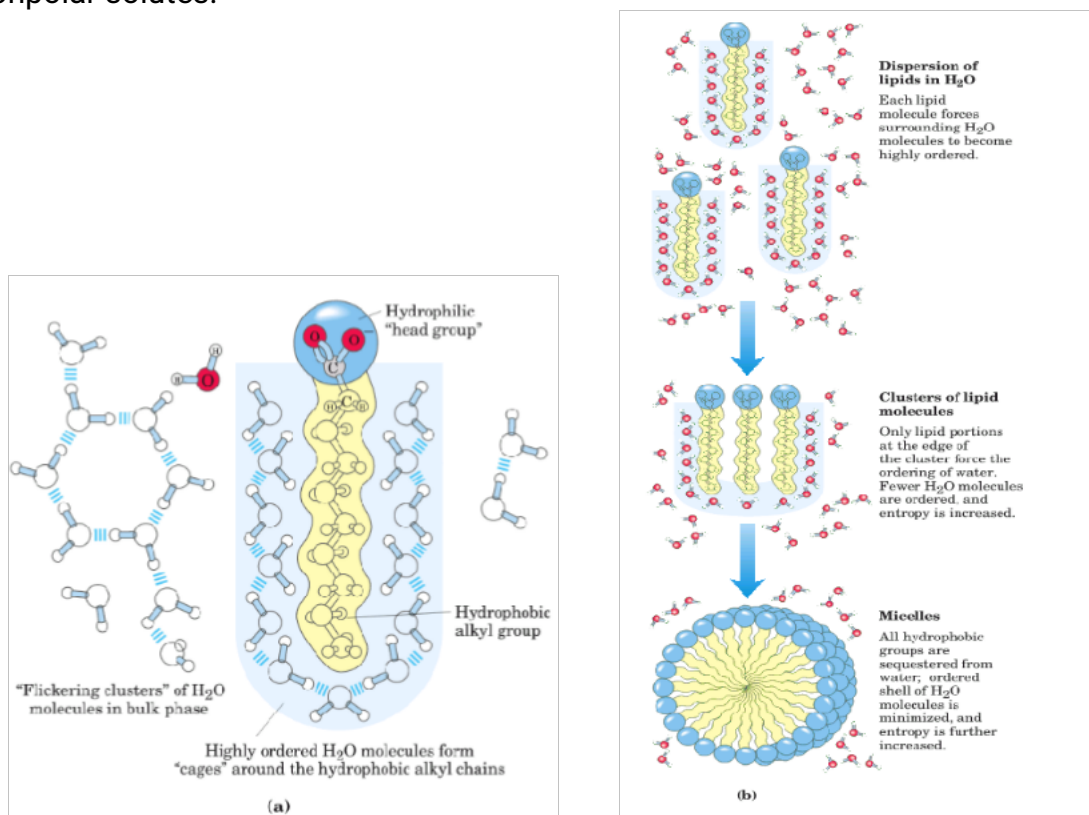


FIGURE 4 Amphipathic compounds in aqueous solution. (a) Longchain fatty acids have very hydrophobic alkyl chains, each of which is surrounded by a layer of highly ordered water molecules. **(b)** By clustering together in micelles, the fatty acid molecules expose the smallest possible hydrophobic surface area to the water, and fewer water molecules are required in the shell of ordered water. The energy gained by freeing immobilized water molecules stabilizes the micelle.

Ionization of Water, Weak Acids, and Weak Bases

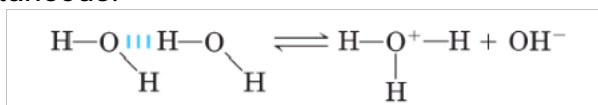
Like all reversible reactions, the ionization of water can be described by an equilibrium constant. When weak acids are dissolved in water, they contribute H^+ by ionizing; weak bases consume H^+ by becoming protonated. These processes are also governed by equilibrium constants. The total hydrogen ion concentration from all sources is experimentally measurable and is expressed as the pH of the solution. To predict the state of ionization of solutes in water, we must take into account the

relevant equilibrium constants for each ionization reaction.

Water molecules have a slight tendency to undergo reversible ionization to yield a hydrogen ion (a proton) and a hydroxide ion, giving the equilibrium.



The dissociation product of water as H^+ , free protons do not exist in solution; hydrogen ions formed in water are immediately hydrated to **hydronium ions** (H_3O^+). Hydrogen bonding between water molecules makes the hydration of dissociating protons virtually instantaneous.



The ionization of water can be measured by its electrical conductivity; pure water carries electrical current as H^+ migrates toward the cathode and OH^- toward the anode. The movement of hydronium and hydroxide ions in the electric field is anomalously fast compared with that of other ions such as Na^+ , K^+ , and Cl^- . Because reversible ionization is crucial to the role of water in cellular function, the extent of ionization of water is expressed in quantitative terms.

The position of equilibrium of any chemical reaction is given by its equilibrium constant, K_{eq} (sometimes expressed simply as K). For the generalized reaction



an equilibrium constant can be defined in terms of the concentrations of reactants (A and B) and products (C and D) at equilibrium:

$$K_{\text{eq}} = \frac{[\text{C}][\text{D}]}{[\text{A}][\text{B}]}$$

Strictly speaking, the concentration terms should be the activities, or effective concentrations in nonideal solutions, of each species. Except in very accurate work, however, the equilibrium constant may be approximated by measuring the concentrations at equilibrium. For reasons beyond the scope of this discussion, equilibrium constants are dimensionless.

The equilibrium constant is fixed and characteristic for any given chemical reaction at a specified temperature. It defines the composition of the final equilibrium mixture, regardless of the starting amounts of reactants and products. Conversely, we can calculate the equilibrium constant for a given reaction at a given temperature if the equilibrium concentrations of all its reactants and products are known.

Equilibrium Constant

The degree of ionization of water at equilibrium is small; at 25 °C only about two of every 109 molecules in pure water are ionized at any instant. The equilibrium constant for the reversible ionization of water is:

$$K_{\text{eq}} = \frac{[\text{H}^+][\text{OH}^-]}{[\text{H}_2\text{O}]}$$

In pure water at 25 °C, the concentration of water is 55.5 M (grams of H_2O in 1 L divided by its gram molecular weight: $(1,000 \text{ g/L})/(18.015 \text{ g/mol})$) and is essentially constant in relation to the very low concentrations of H^+ and OH^- , namely, $1 \times 10^{-7} \text{ M}$. Accordingly, we

can substitute 55.5 M in the above equilibrium constant expression to yield:

$$K_{\text{eq}} = \frac{[\text{H}^+][\text{OH}^-]}{55.5 \text{ M}},$$

Which, on rearranging, becomes:

$$(55.5 \text{ M})(K_{\text{eq}}) = [\text{H}^+][\text{OH}^-] = K_w$$

Where K_w designates the product $(55.5 \text{ M})(K_{\text{eq}})$, the **ion product of water** at 25 °C.

The value for K_{eq} , determined by electrical-conductivity measurements of pure water, is $1.8 \times 10^{-16} \text{ M}$ at 25° C. Substituting this value for K_{eq} in the above equation gives the value of the ion product of water:

$$\begin{aligned} K_w &= [\text{H}^+][\text{OH}^-] = (55.5 \text{ M})(1.8 \times 10^{-16} \text{ M}) \\ &= 1.0 \times 10^{-14} \text{ M}^2 \end{aligned}$$

Thus the product $[\text{H}^+][\text{OH}^-]$ in aqueous solutions at 25° C always equals $1 \times 10^{-14} \text{ M}^2$. When there are exactly equal concentrations of H^+ and OH^- , as in pure water, the solution is said to be at **neutral pH**. At this pH, the concentration of H^+ and OH^- can be calculated from the ion product of water as follows:

$$K_w = [\text{H}^+][\text{OH}^-] = [\text{H}^+]^2$$

Solving for $[\text{H}^+]$ gives

$$\begin{aligned} [\text{H}^+] &= \sqrt{K_w} = \sqrt{1 \times 10^{-14} \text{ M}^2} \\ [\text{H}^+] &= [\text{OH}^-] = 10^{-7} \text{ M} \end{aligned}$$

As the ion product of water is constant, whenever $[\text{H}^+]$ is greater than $1 \times 10^{-7} \text{ M}$, $[\text{OH}^-]$ must become less than $1 \times 10^{-7} \text{ M}$, and vice versa. When $[\text{H}^+]$ is very high, as in a solution of hydrochloric acid, $[\text{OH}^-]$ must be very low. From the ion product of water we can calculate $[\text{H}^+]$ if we know $[\text{OH}^-]$, and vice versa.

pH

The ion product of water, K_w , is the basis for the pH scale (Table 2–6). It is a convenient means of designating the concentration of H^+ (and thus of OH^-) in any aqueous solution in the range between 1.0 M H^+ and 1.0 M OH^- . The term pH is defined by the expression

$$\text{pH} = \log \frac{1}{[\text{H}^+]} = -\log [\text{H}^+]$$

The symbol p denotes “negative logarithm of.” For a precisely neutral solution at 25° C, in which the concentration of hydrogen ions is $1.0 \times 10^{-7} \text{ M}$, the pH can be calculated as follows:

$$\begin{aligned} \text{pH} &= \log \frac{1}{1.0 \times 10^{-7}} = \log (1.0 \times 10^7) \\ &= \log 1.0 + \log 10^7 = 0 + 7 = 7 \end{aligned}$$

TABLE 2 The pH Scale

[H ⁺] (M)	pH	[OH ⁻] (M)	pOH*
10 ⁰ (1)		0	10 ⁻¹⁴
10 ⁻¹	1	10 ⁻¹³	13
10 ⁻²	2	10 ⁻¹²	12
10 ⁻³	3	10 ⁻¹¹	11
10 ⁻⁴	4	10 ⁻¹⁰	10
10 ⁻⁵	5	10 ⁻⁹	9
10 ⁻⁶	6	10 ⁻⁸	8
10 ⁻⁷	7	10 ⁻⁷	7
10 ⁻⁸	8	10 ⁻⁶	6
10 ⁻⁹	9	10 ⁻⁵	5
10 ⁻¹⁰	10	10 ⁻⁴	4
10 ⁻¹¹	11	10 ⁻³	3
10 ⁻¹²	12	10 ⁻²	2
10 ⁻¹³	13	10 ⁻¹	1
10 ⁻¹⁴	14	10 ⁰ (1)	0

*The expression pOH is sometimes used to describe the basicity, or OH⁻ concentration, of a solution; pOH is defined by the expression $\text{pOH} = -\log [\text{OH}^-]$, which is analogous to the expression for pH. Note that in all cases, $\text{pH} + \text{pOH} = 14$.

The value of 7 for the pH of a precisely neutral solution is not an arbitrarily chosen figure; it is derived from the absolute value of the ion product of water at 25° C, which by convenient coincidence is a round number. Solutions having a pH greater than 7 are alkaline or basic; the concentration of OH⁻ is greater than that of H⁺. Conversely, solutions having a pH less than 7 are acidic. Note that the pH scale is logarithmic, not arithmetic. To say that two solutions differ in pH by 1 pH unit means that one solution has ten times the H⁺ concentration of the other, but it does not tell us the absolute magnitude of the difference. Figure 5 gives the pH of some common aqueous fluids. A cola drink (pH 3.0) or red wine (pH 3.7) has an H⁺ concentration approximately 10,000 times that of blood (pH 7.4). The pH of an aqueous solution can be approximately measured using various indicator dyes, including litmus, phenolphthalein, and phenol red, which undergo color changes as a proton dissociates from the dye molecule.

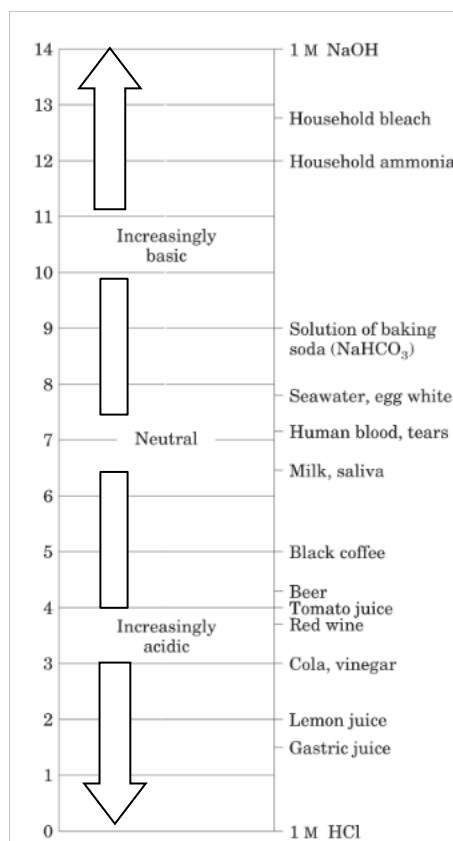


Figure 5 The pH of some aqueous fluids

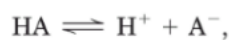
Dissociation constants of weak acids and bases

Hydrochloric, sulfuric, and nitric acids, commonly called strong acids, are completely ionized in dilute aqueous solutions; the strong bases NaOH and KOH are also completely ionized. Of more interest to biochemists is the behavior of weak acids and bases—those not completely ionized when dissolved in water. These are common in biological systems and play important roles in metabolism and its regulation.

Acids may be defined as proton donors and bases as proton acceptors. A proton donor and its corresponding proton acceptor make up a **conjugate acid-base pair** (Fig.16). Acetic acid (CH_3COOH), a proton donor, and the acetate anion (CH_3COO^-), the corresponding proton acceptor, constitute a conjugate acid base pair, related by the reversible reaction



Each acid has a characteristic tendency to lose its proton in an aqueous solution. The stronger the acid, the greater is its tendency to lose its proton. The tendency of any acid (HA) to lose a proton and form its conjugate base (A^-) is defined by the equilibrium constant (K_{eq}) for the reversible reaction



which is

$$K_{\text{eq}} = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]} = K_{\text{a}}$$

Equilibrium constants for ionization reactions are usually called ionization or **dissociation constants**, often designated K_a . The dissociation constants of some acids are given in Figure 6. Stronger acids, such as phosphoric and carbonic acids, have larger dissociation constants; weaker acids, such as monohydrogen phosphate (HPO_4^{2-}), have smaller dissociation constants.

Also included in Figure 16 are values of **pKa**, which is analogous to pH and is defined by the equation

$$\text{p}K_a = \log \frac{1}{K_a} = -\log K_a$$

The stronger the tendency to dissociate a proton, the stronger is the acid and the lower its pKa. As we shall now see, the pKa of any weak acid can be determined quite easily.

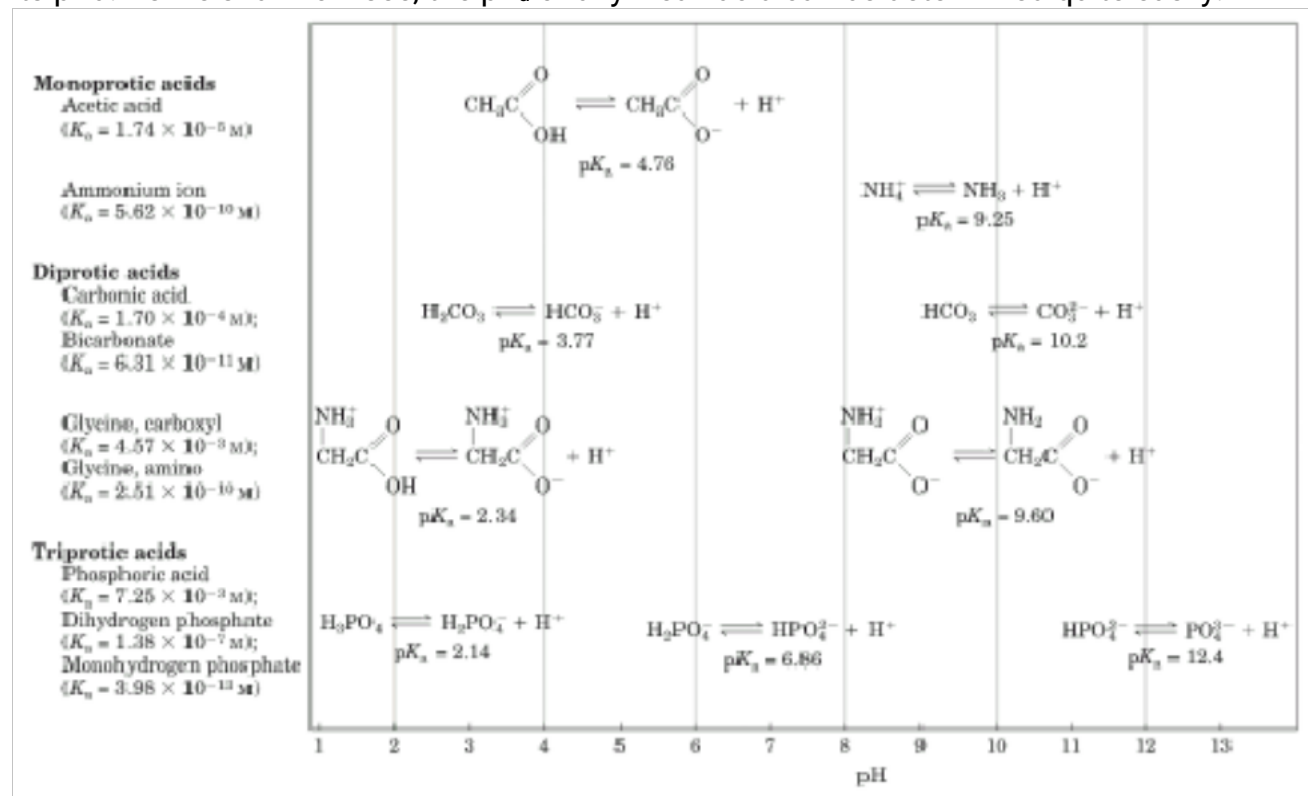


FIGURE 6 Conjugate acid-base pairs consist of a proton donor and a proton acceptor. Some compounds, such as acetic acid and ammonium ion, are monoprotic; they can give up only one proton. Others are diprotic (H_2CO_3 (carbonic acid) and glycine) or triprotic (H_3PO_4 (phosphoric acid)). The dissociation reactions for each pair are shown where they occur along a pH gradient. The equilibrium or dissociation constant (K_a) and its negative logarithm, the pKa, are shown for each reaction.

Buffers

Buffers are aqueous systems that tend to resist changes in pH when small amounts of acid (H^+) or base (OH^-) are added. A buffer system consists of a weak acid (the proton donor) and its conjugate base (the proton acceptor).

Buffering results from two reversible reaction equilibria occurring in a solution of nearly equal concentrations of a proton donor and its conjugate proton acceptor. Figure 7 explains how a buffer system works. Whenever H^+ or OH^- is added to a buffer, the result is a small change in the ratio of the relative concentrations of the weak acid and

its anion and thus a small change in pH. The decrease in concentration of one component of the system is balanced exactly by an increase in the other. The sum of the buffer components does not change, only their ratio.

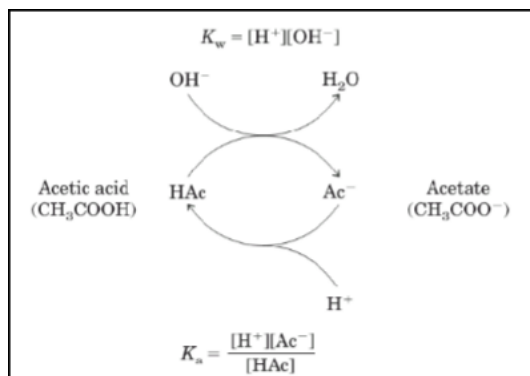


FIGURE 7 The acetic acid–acetate pair as a buffer system. The system is capable of absorbing either H^+ or OH^- through the reversibility of the dissociation of acetic acid. The proton donor, acetic acid (HAc), contains a reserve of bound H^+ , which can be released to neutralize an addition of OH^- to the system, forming H_2O . This happens because the product $[\text{H}^+][\text{OH}^-]$ transiently exceeds K_w ($1 \times 10^{-14} \text{ M}^2$). The equilibrium quickly adjusts so that this product equals $1 \times 10^{-14} \text{ M}^2$ (at 25°C), thus transiently reducing the concentration of H^+ . But now the quotient $[\text{H}^+][\text{Ac}^-]/[\text{HAc}]$ is less than K_a , so HAc dissociates further to restore equilibrium. Similarly, the conjugate base, Ac^- , can react with H^+ ions added to the system; again, the two ionization reactions simultaneously come to equilibrium. Thus a conjugate acid-base pair, such as acetic acid and acetate ion, tends to resist a change in pH when small amounts of acid or base are added. Buffering action is simply the consequence of two reversible reactions taking place simultaneously and reaching their points of equilibrium as governed by their equilibrium constants, K_w and K_a .

Relation between pH, pKa, and buffer concentration

The relation between pH, pKa, and buffer concentration is given by **Henderson-Hasselbalch equation**, which is important for understanding buffer action and acid-base balance in the blood and tissues of vertebrates. This equation is simply a useful way of restating the expression for the dissociation constant of an acid. For the dissociation of a weak acid HA into H^+ and A^- , the Henderson-Hasselbalch equation can be derived as follows:

$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]}$$

First solve for $[\text{H}^+]$:

$$[\text{H}^+] = K_a \frac{[\text{HA}]}{[\text{A}^-]}$$

Then take the negative logarithm of both sides:

$$-\log [\text{H}^+] = -\log K_a - \log \frac{[\text{HA}]}{[\text{A}^-]}$$

Substitute pH for $-\log [\text{H}^+]$ and $\text{p}K_a$ for $-\log K_a$:

$$\text{pH} = \text{p}K_a - \log \frac{[\text{HA}]}{[\text{A}^-]}$$

Now invert $-\log [\text{HA}]/[\text{A}^-]$, which involves changing its sign, to obtain the Henderson-Hasselbalch equation:

$$\text{pH} = \text{p}K_a + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

Stated more generally,

$$\text{pH} = \text{p}K_a + \log \frac{[\text{proton acceptor}]}{[\text{proton donor}]}$$

This equation fits the titration curve of all weak acids and enables us to deduce a number of important quantitative relationships. For example, it shows why the pKa of a weak acid is equal to the pH of the solution at the midpoint of its titration. At that point, $[\text{HA}]$ equals $[\text{A}^-]$, and

$$\text{pH} = \text{pK}_a + \log 1 = \text{pK}_a + 0 = \text{pK}_a$$

the Henderson-Hasselbalch equation also allows us to (1) calculate pKa, given pH and the molar ratio of proton donor and acceptor; (2) calculate pH, given pKa and the molar ratio of proton donor and acceptor; and (3) calculate the molar ratio of proton donor and acceptor, given pH and pKa.

Amino acids

Proteins are polymers of amino acids, with each **amino acid residue** joined to its neighbor by a specific type of covalent bond. (The term “residue” reflects the loss of the elements of water when one amino acid is joined to another.) Proteins can be broken down (hydrolyzed) to their constituent amino acids by a variety of methods, and the earliest studies of proteins naturally focused on the free amino acids derived from them. Twenty different amino acids are commonly found in proteins. The first to be discovered was asparagine, in 1806.

All 20 of the common amino acids are α -amino acids. They have a carboxyl group and an amino group bonded to the same carbon atom (the α -carbon) (Fig. 1). They differ from each other in their side chains, or **R groups**, which vary in structure, size, and electric charge, and which influence the solubility of the amino acids in water.

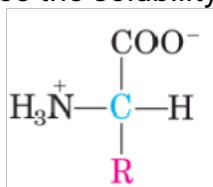


FIGURE 1 General structure of an amino acid. This structure is common to all but one of the α -amino acids. (Proline, a cyclic amino acid, is the exception.) The R group or side chain (red) attached to the α -carbon (blue) is different in each amino acid.

Classification of amino acids

Amino acids are classified into five main classes based on the properties of their R groups, in particular, their **polarity**, or tendency to interact with water at biological pH (near pH 7.0). The polarity of the R groups varies widely, from nonpolar and hydrophobic (water-insoluble) to highly polar and hydrophilic (water-soluble).

The structures of the 20 common amino acids are shown in Figure 2. Within each class there are gradations of polarity, size, and shape of the R groups.

1. Nonpolar, Aliphatic R Groups: The R groups in this class of amino acids are nonpolar and hydrophobic. The side chains of **alanine**, **valine**, **leucine**, and **isoleucine** tend to cluster together within proteins, stabilizing protein structure by means of hydrophobic interactions. **Glycine** has the simplest structure. Although it is formally nonpolar, its very small side chain makes no real contribution to hydrophobic interactions. **Methionine**, one of the two sulfur-containing amino acids, has a nonpolar thioether group in its side chain. **Proline** has an aliphatic side chain with a distinctive cyclic structure. The secondary amino (imino) group of proline residues is held in a rigid conformation that reduces the structural flexibility of polypeptide regions containing

proline.

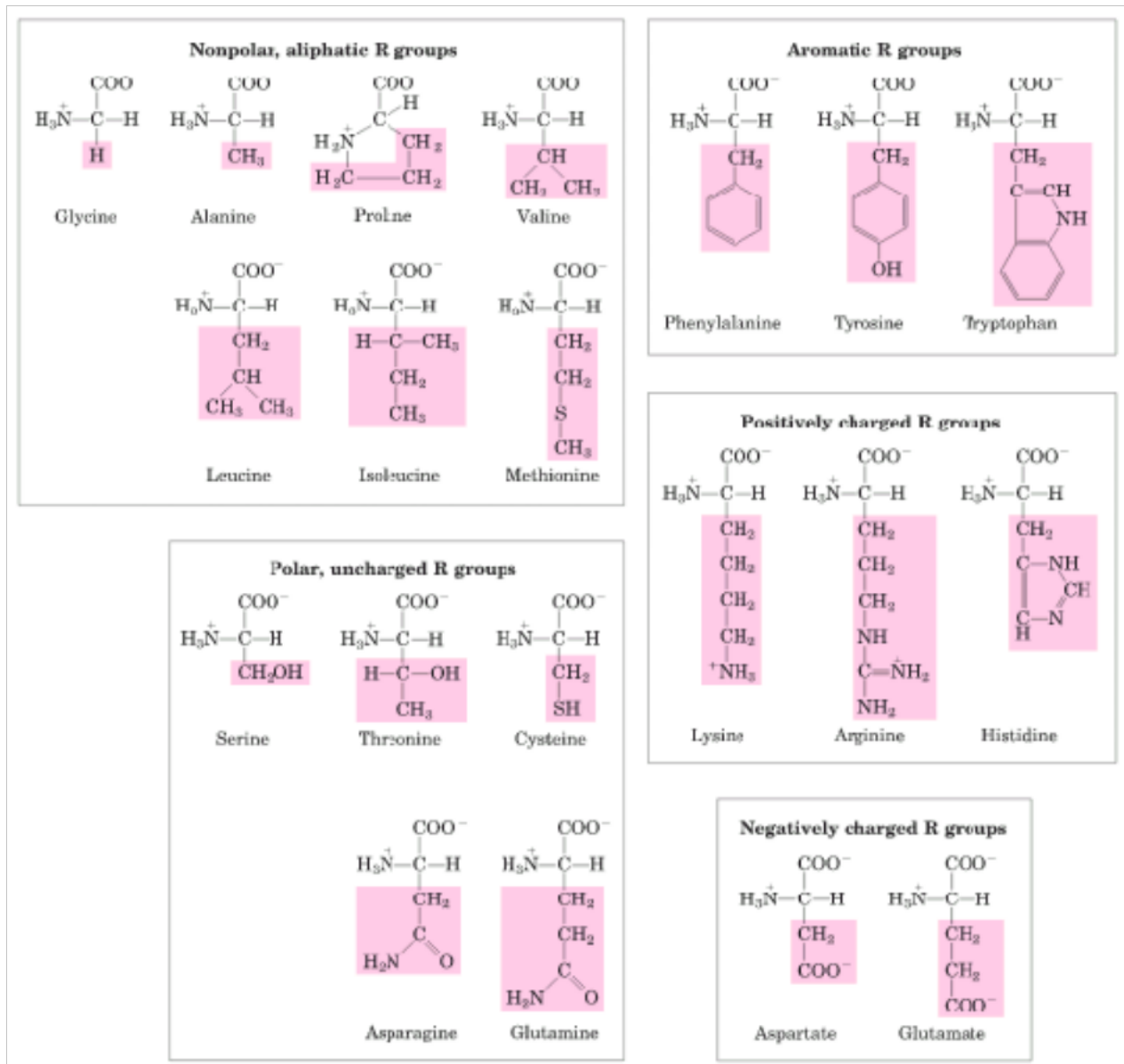


FIGURE 2 The 20 common amino acids of proteins. The structural formulas show the state of ionization that would predominate at pH 7.0. The unshaded portions are those common to all the amino acids; the portions shaded in red are the R groups. Although the R group of histidine is shown uncharged, its pKa is such that a small but significant fraction of these groups are positively charged at pH 7.0.

2. Aromatic R Groups: Phenylalanine, tyrosine, and tryptophan, with their aromatic side chains, are relatively nonpolar (hydrophobic). All can participate in hydrophobic interactions. The hydroxyl group of tyrosine can form hydrogen bonds, and it is an important functional group in some enzymes. Tyrosine and tryptophan are significantly more polar than phenylalanine, because of the tyrosine hydroxyl group and the nitrogen

of the tryptophan indole ring. Tryptophan and tyrosine, and to a much lesser extent phenylalanine, absorb ultraviolet light.

3. Polar, Uncharged R Groups: The R groups of these amino acids are more soluble in water, or more hydrophilic, than those of the nonpolar amino acids, because they contain functional groups that form hydrogen bonds with water. This class of amino acids includes **serine, threonine, cysteine, asparagine, and glutamine**. The polarity of serine and threonine is contributed by their hydroxyl groups; that of cysteine by its sulfhydryl group; and that of asparagine and glutamine by their amide groups.

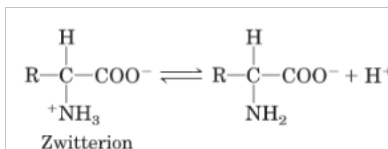
Asparagine and glutamine are the amides of two other amino acids also found in proteins, aspartate and glutamate, respectively, to which asparagine and glutamine are easily hydrolyzed by acid or base. **Cysteine** is readily oxidized to form a covalently linked dimeric amino acid called cystine, in which two cysteine molecules or residues are joined by a disulfide bond (Fig.2). The disulfide-linked residues are strongly hydrophobic (nonpolar). Disulfide bonds play a special role in the structures of many proteins by forming covalent links between parts of a protein molecule or between two different polypeptide chains.

4. Positively Charged (Basic) R Groups: The most hydrophilic R groups are those that are either positively or negatively charged. The amino acids in which the R groups have significant positive charge at pH 7.0 are **lysine**, which has a second primary amino group at the ϵ -position on its aliphatic chain; **arginine**, which has a positively charged guanidino group; and **histidine**, which has an imidazole group. Histidine is the only common amino acid having an ionizable side chain.

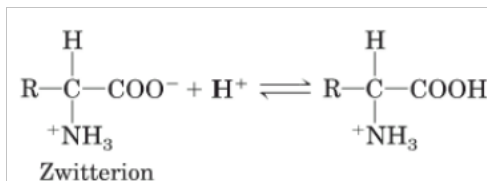
5. Negatively Charged (Acidic) R Groups: The two amino acids having R groups with a net negative charge at pH 7.0 are **aspartate** and **glutamate**, each of which has a second carboxyl group.

Zwitterion

When an amino acid is dissolved in water, it exists in solution as the dipolar ion, or **zwitterions** (German for “hybrid ion”). A zwitterion can act as either an acid (proton donor):

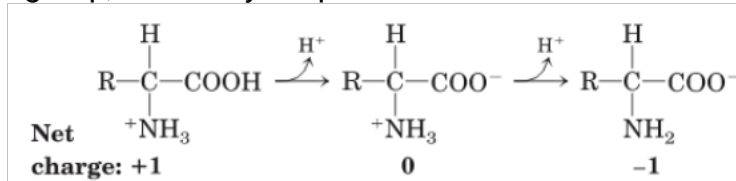


or a base (proton acceptor):



Substances having this dual nature are **amphoteric** and are often called **ampholytes** (from “amphoteric electrolytes”). A simple monoamino monocarboxylic α -amino acid,

such as alanine, is a diprotic acid when fully protonated, it has two groups, the COOH group and the -NH_3^+ group, that can yield protons:



Isoelectric point (pI):

The characteristic pH at which the net electric charge is zero is called the **isoelectric point** or **isoelectric pH**, designated **pI**. For glycine, which has no ionizable group in its side chain, the isoelectric point is simply the arithmetic mean of the two pK_a values:

Glycine has a net negative charge at any pH above its pI and will thus move toward the positive electrode (the anode) when placed in an electric field. At any pH below its pI, glycine has a net positive charge and will move toward the negative electrode (the cathode). The farther the pH of a glycine solution is from its isoelectric point, the greater the net electric charge of the population of glycine molecules. At pH 1.0, for example, glycine exists almost entirely as the form

$^+\text{H}_3\text{N-CH}_2\text{-COOH}$, with a net positive charge of 1.0. At pH 2.34, where there is an equal mixture of $^+\text{H}_3\text{N-CH}_2\text{-COOH}$ and $^+\text{H}_3\text{N-CH}_2\text{-COO}^-$, the average or net positive charge is 0.5. The sign and the magnitude of the net charge of any amino acid at any pH can be predicted in the same way.

Peptide bond:

Two amino acid molecules can be covalently joined through a substituted amide linkage, termed a **peptide bond**, to yield a dipeptide. Such a linkage is formed by removal of the elements of water (dehydration) from the α -carboxyl group of one amino acid and the α -amino group of another (Fig. 3). Peptide bond formation is an example of a condensation reaction, a common class of reactions in living cells.

Three amino acids can be joined by two peptide bonds to form a tripeptide; similarly, amino acids can be linked to form tetrapeptides, pentapeptides, and so forth. When a few amino acids are joined in this fashion, the structure is called an **oligopeptide**. When many amino acids are joined, the product is called a **polypeptide**. Proteins may have thousands of amino acid residues. Although the terms “protein” and “polypeptide” are sometimes used interchangeably, molecules referred to as polypeptides generally have molecular weights below 10,000, and those called proteins have higher molecular weights.

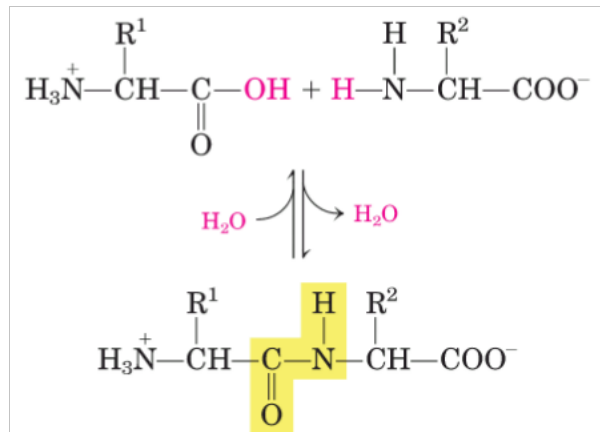


FIGURE 3 Formation of a peptide bond by condensation. The α -amino group of one amino acid (with R2 group) acts as a nucleophile to displace the hydroxyl group of another amino acid (with R1 group), forming a peptide bond (shaded in yellow). Amino groups are good nucleophiles, but the hydroxyl group is a poor leaving group and is not readily displaced. At physiological pH, the reaction shown does not occur to any appreciable extent.

Structure of proteins:

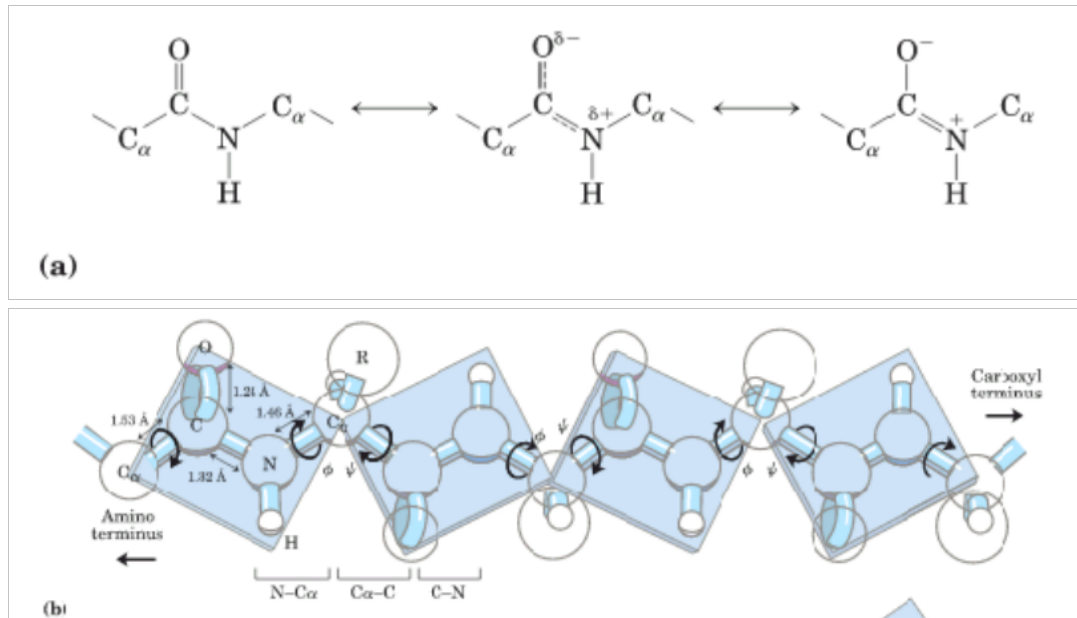
Four levels of protein structure are commonly defined (Fig. 3–16). A description of all covalent bonds (mainly peptide bonds and disulfide bonds) linking amino acid residues in a polypeptide chain is its **primary structure**. The most important element of primary structure is the *sequence* of amino acid residues. **Secondary structure** refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns. **Tertiary structure** describes all aspects of the three-dimensional folding of a polypeptide. When a protein has two or more polypeptide subunits, their arrangement in space is referred to as **quaternary structure**.

Primary structure

The primary structure of a polypeptide is its amino acid sequence. The amino acids are connected by peptide bonds. The structure of polypeptide determines the higher levels of structural organization.

The α -carbons of adjacent amino acid residues are separated by three covalent bonds, arranged as $\text{C}_\alpha\text{-C-N-C}_\alpha$. X-ray diffraction studies of crystals of amino acids and of simple dipeptides and tripeptides demonstrated that the peptide C-N bond is somewhat shorter than the C-N bond in a simple amine and that the atoms associated with the peptide bond are coplanar. This indicated a resonance or partial sharing of two pairs of electrons between the carbonyl oxygen and the amide nitrogen (Fig. 1a). The oxygen has a partial negative charge and the nitrogen a partial positive charge, setting up a small electric dipole. The six atoms of the **peptide group** lie in a single plane, with the oxygen atom of the carbonyl group and the hydrogen atom of the amide nitrogen trans to each other. From these findings Pauling and Corey concluded that the peptide C-N bonds are unable to rotate freely because of their partial double-bond character. Rotation is permitted about the N-C_α and the $\text{C}_\alpha\text{-C}$ bonds. The backbone of a polypeptide chain can thus be pictured as a series of rigid planes with consecutive planes sharing a common point of rotation at C_α (Fig. 1b). The rigid peptide bonds limit the range of conformations that can be assumed by a polypeptide chain. By convention, the bond

angles resulting from rotations at C_α are labeled ϕ (phi) for the $N-C_\alpha$ bond and ψ (psi) for the $C_\alpha-C$ bond. Again by convention, both ϕ and ψ are defined as 180° when the polypeptide is in its fully extended conformation and all peptide groups are in the same plane (Fig. 4–2b). In principle, ϕ and ψ can have any value between -180° and $+180^\circ$, but many values are prohibited by steric interference between atoms in the polypeptide backbone and amino acid side chains. The conformation in which both ϕ and ψ are 0° (Fig. 2c) is prohibited for this reason; this conformation is used merely as a reference point for describing the angles of rotation. Allowed values for ϕ and ψ are graphically revealed when ψ is plotted versus ϕ in a **Ramachandran plot** (Fig. 3), introduced by **G. N. Ramachandran**.



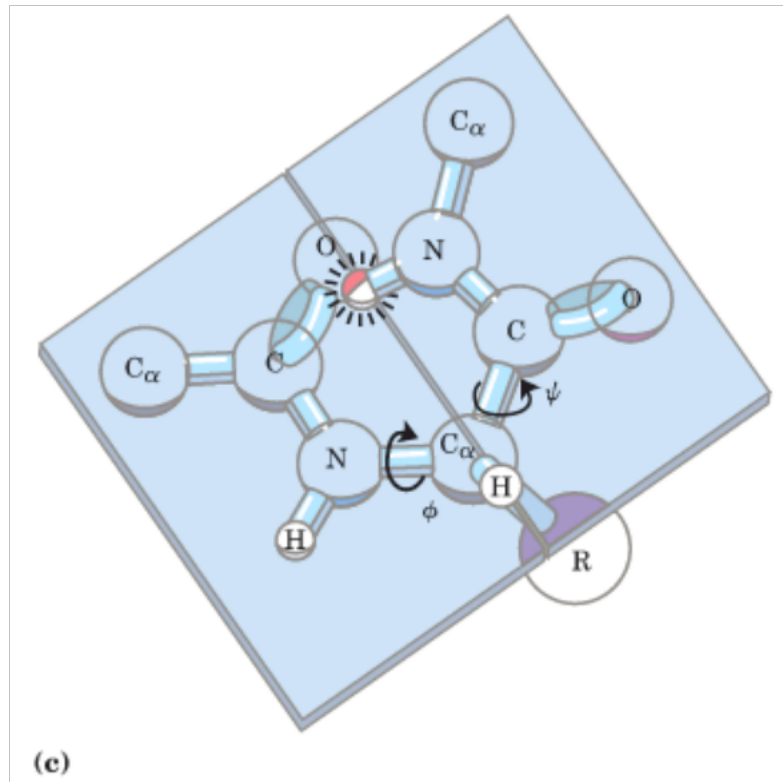


FIGURE 2 The planar peptide group. (a) Each peptide bond has some double-bond character due to resonance and cannot rotate. (b) Three bonds separate sequential α -carbons in a polypeptide chain. The N - C_α and C_α - C bonds can rotate, with bond angles designated ϕ and ψ , respectively. The peptide CON bond is not free to rotate. Other single bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups. In the conformation shown, ψ and ϕ are 180° (or -180°). As one looks out from the α -carbon, the ϕ and ψ angles increase as the carbonyl or amide nitrogens (respectively) rotate clockwise. (c) By convention, both ϕ and ψ are defined as 0° when the two peptide bonds flanking that α -carbon are in the same plane and positioned as shown. In a protein, this conformation is prohibited by steric overlap between an α -carbonyl oxygen and an α -amino hydrogen atom. To illustrate the bonds between atoms, the balls representing each atom are smaller than the van der Waals radii for this scale. $1 \text{ \AA} = 0.1 \text{ nm}$.

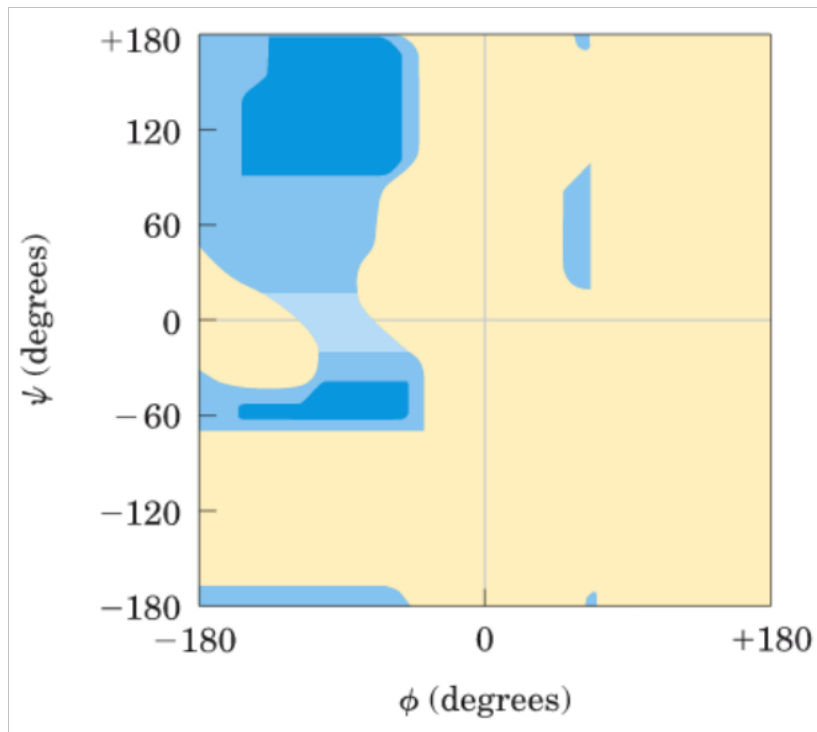


FIGURE 3 Ramachandran plot for L-Ala residues. The conformations of peptides are defined by the values of ϕ and ψ . Conformations deemed possible are those that involve little or no steric interference, based on calculations using known van der Waals radii and bond angles. The areas shaded dark blue reflect conformations that involve no steric overlap and thus are fully allowed; medium blue indicates conformations allowed at the extreme limits for unfavorable atomic contacts; the lightest blue area reflects conformations that are permissible if a little flexibility is allowed in the bond angles. The asymmetry of the plot results from the L stereochemistry of the amino acid residues. The plots for other L-amino acid residues with unbranched side chains are nearly identical. The allowed ranges for branched amino acid residues such as Val, Ile, and Thr are somewhat smaller than for Ala. The Gly residue, which is less sterically hindered, exhibits a much broader range of allowed conformations. The range for Pro residues is greatly restricted because ϕ is limited by the cyclic side chain to the range of -35° to -85° .

Secondary structure:

The most common type of secondary structure (2nd structure) are the α -helix and β -pleated sheet. Both α -helix and β -pleated sheet patterns are stabilized by hydrogen bonds between the carbonyl and N-H group in the polypeptide's backbone.

α -helix: The α -helix is a rigid, rod like structure that forms when a polypeptide chain twists in a helical conformation. The screw sense of α -helix can be right handed or left handed. However right-handed helices are energetically more favorable. In almost all proteins, the helical twist of the α -helix is right-handed. There are 3.6 amino acid residues per turn of the helix and the pitch is 0.4 nm. Each residue is related to the next one by a rise of 1.5 Å along the helix axis. A single turn of α -helix involves 13 atoms from O to the H of the H bond. For this reason, the α -helix is referred as 3.6₁₃ helix. Length of α -helix is usually 10-15 amino acid residues. Intrachain hydrogen bonds between the N-H group of each amino acid and carbonyl group of the amino acid four residues away.

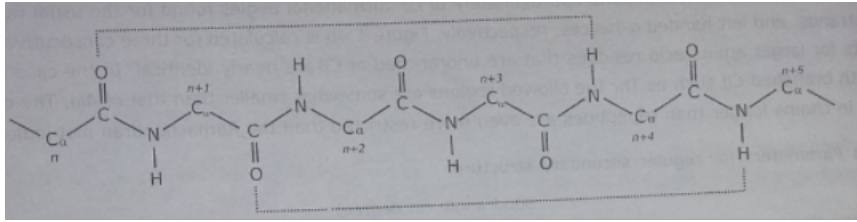


Figure: The hydrogen bonding arrangement in the α -helix

Except for the amino acids near the ends of an α -helix, all the main chain CO and NH groups are hydrogen bonded. The side chains of amino acids extended outward from the helix. All H bonds lie parallel to the helix axis and point in the same direction.

Helices can be formed from either D- or L-amino acids, but a given helix must be composed entirely of amino acids can form either right or left handed α -helices.

Amino acids have different properties for forming α -helices. Amino acid residues such as alanine, glutamine, glutamate, leucine, methionine, arginine show higher tendency to form α -helices. Proline tends to disrupt α -helices because it lacks an $-NH$ group and because its ring structure restricts its ϕ value to near -60° .

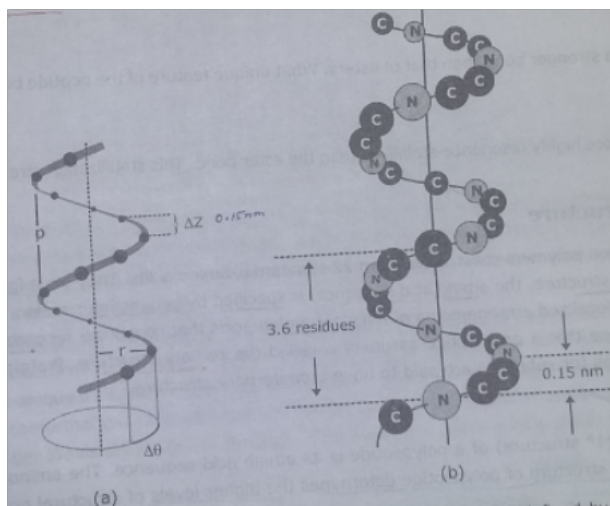


Figure a) describing the geometry of α -helix, b) the right handed α -helix

β -pleated sheets:

β -pleated sheets form when two or more polypeptide chain segment line up side by side. Each individual segment is referred as a β -strand. Rather than being coiled, each β -strand is fully extended. The distance between adjacent amino acids along a β -strand is approximately 3.5 \AA along with a distance of 1.5 \AA along an α -helix. β -pleated sheets are stabilized by interchain hydrogen bonds that form between the polypeptide backbone N-H and carbonyl groups of adjacent strands. Adjacent strand can be either parallel or antiparallel. In parallel β -pleated sheet structures, the polypeptide chains are arranged in the same direction. However in antiparallel β -pleated sheet chains run in opposite directions. Antiparallel β -sheets are more stable than parallel β -sheets because fully collinear hydrogen bonds form.

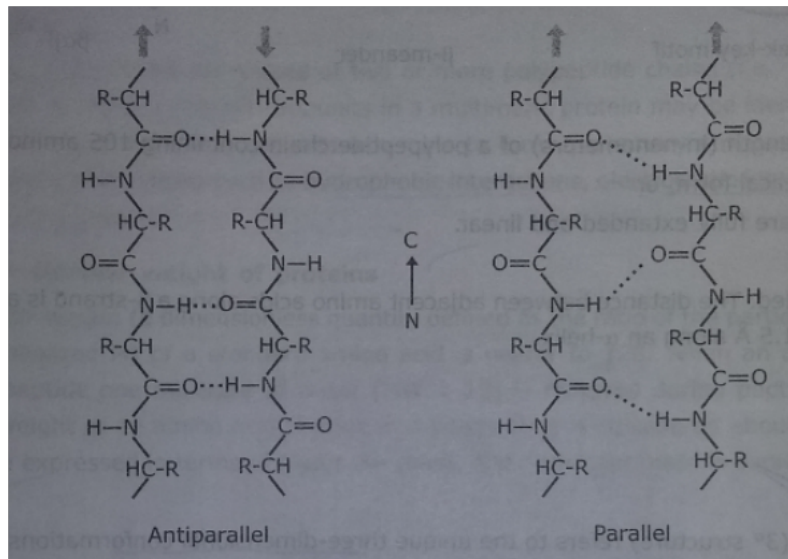


Figure: In an antiparallel sheet, adjacent sheets run in opposite direction. Hydrogen bonds between NH and CO groups connect each amino acid to a single amino acid on an adjacent strand. In parallel sheet, adjacent strands run in the same direction. Hydrogen bonds connect each amino acid on one strand with two different amino acids on the adjacent strand. For each amino acid, the NH group is hydrogen bonded to the CO group of one amino acid on adjacent strand, where the CO group is hydrogen bonded to NH group on the amino acid two residues farther along the chain.

Tertiary structure

The term tertiary structure (3rd) refers to the unique three dimensional conformations that globular proteins assume as a consequence of the interactions between the side chains in their primary structure. All information needed to fold the protein into its native structure is contained within the primary structure of the peptide chain itself.

The following types of covalent and non covalent interactions stabilize tertiary structure:

1. Hydrophobic interactions
2. Electrostatic interactions
3. Hydrogen bonds
4. Van der Waals forces of interaction
5. Covalent bond: The most prominent covalent bonds in tertiary structure are the disulphide bridges found in many extracellular proteins. Cysteine has a thiol group that is unique among 20 amino acids in that it often forms a disulfide bond to another cysteine residue through the oxidation of their thiol groups. This dimeric compound is referred to as the amino acid cystine.

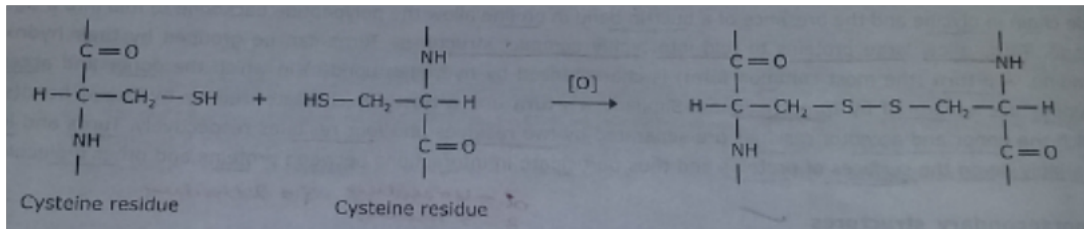


Figure: Structure of cysteine

The fundamental unit of tertiary structure is the domain. It is a polypeptide chain that can fold independently into a stable three dimensional tertiary structure. It is also functional unit of protein. Proteins may comprise a single domain or more than one domain. A small protein often consists of only one domain.

Bioenergetics and thermodynamics

Bioenergetics is the quantitative study of the energy transductions that occur in living cells and of the nature and function of the chemical processes underlying these transductions.

Thermodynamic principles:

First law of thermodynamics: The first law of thermodynamics describes the principle of the conservation of energy: in any physical or chemical change, the total amount of energy in the universe remains constant, i.e the total of energy of a system, including surroundings, remain constant, although the form of the energy may change.

Mathematically it can be expressed as:

$$\Delta U = \Delta q - \Delta w$$

ΔU is the change in internal energy,

Δq the heat exchanged from the surroundings,

Δw the work done by the system

If Δq is positive, heat has been transferred to the system, giving an increase in internal energy. When Δq is negative, heat has been transferred to the surroundings, giving a decrease in internal energy. When Δw is positive, work has been done by the system, giving a decrease in internal energy. When Δw is negative, work has been done by surroundings, giving an increase in internal energy.

Second law of thermodynamics: The second law of thermodynamics states that the total entropy of a system must increase if a process is to occur spontaneously.

Mathematically it can be expressed as:

$$\Delta S \leq \Delta q/T$$

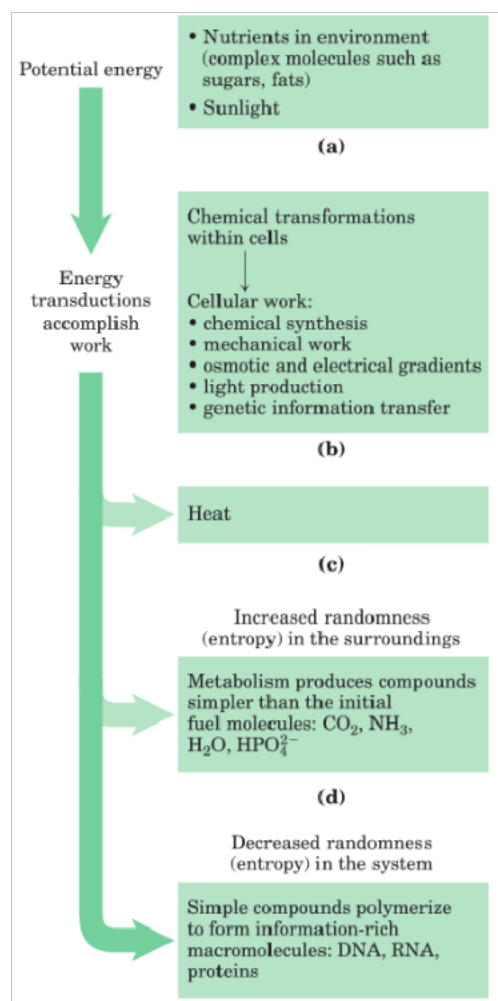
Where ΔS is the change in the entropy of the system

Open, closed and isolated system

Living organisms consist of collections of molecules much more highly

organized than the surrounding materials from which they are constructed, and organisms maintain and produce order, seemingly oblivious to the second law of thermodynamics. But living organisms do not violate the second law; they operate strictly within it. The reacting system is the collection of matter that is undergoing a particular chemical or physical process; it may be an organism, a cell, or two reacting compounds. The reacting system and its surroundings together constitute the universe. Some chemical or physical processes can be carried out in isolated or closed systems, in which no material or energy is exchanged with the surroundings. Living cells and organisms, however, are open systems, exchanging both material and energy with their surroundings; living systems are never at equilibrium with their surroundings, and the constant transactions between system and surroundings explain how organisms can create order within themselves while operating within the second law of thermodynamics.

For chemical reactions occurring in solution, a **system** is defined as all the reactants and products present, the solvent that contains them, and the immediate atmosphere. In short, a system is everything within a defined region of space. The system and its surroundings together constitute the **universe**. There are three types of system as:



Isolated system: If the system exchanges neither matter nor energy with its surroundings, it is said to be isolated system.

Closed system: If the system exchanges energy but not matter with its surroundings, it is said to be closed system.

Open system: if the system exchanges both energy and matter with its surroundings, it is an open system. A living organism is an open system; it exchanges both matter and energy with its surroundings. Living organisms derive energy from their surroundings in two ways: (1) they take up chemical fuels (such as glucose) from the environment and extract energy by oxidizing them or (2) they absorb energy from sunlight.

The first law of thermodynamics, developed from physics and chemistry but fully valid for biological systems as well, describes the principle of the conservation of energy: *in any physical or chemical change, the total amount of energy in the universe remains constant, although the form of the energy may change*. Cells are consummate transducers of energy, capable of interconverting chemical, electromagnetic, mechanical, and osmotic energy with great efficiency (Fig.1).

FIGURE 1; Some energy interconversion in living organisms. During metabolic energy transductions, the randomness of

the system plus surroundings (expressed quantitatively as entropy) increases as the potential energy of complex nutrient molecules decreases. **(a)** Living organisms extract energy from their surroundings; **(b)** convert some of it into useful forms of energy to produce work; **(c)** return some energy to the surroundings as heat; and **(d)** release end-product molecules that are less well organized than the starting fuel, increasing the entropy of the universe. One effect of all these transformations is **(e)** increased order (decreased randomness) in the system in the form of complex macromolecules.

Gibb's free energy:

Free energy or Gibb's free energy indicates the portion of the total energy of a system that is available for useful work (also known as chemical potential). The change in free energy is denoted as ΔG . Under constant temperature and pressure, the relationship between free energy change (ΔG) of a reacting system and the change in entropy (ΔS) is expressed by following equation:

$$\Delta G = \Delta H - T\Delta S$$

Where ΔH is the change in enthalpy and T is absolute temperature. ΔH is the measure of change in heat content of reactants and products. The change in the free energy, ΔG , can be used to predict the direction of a reaction at constant temperature and pressure.

If ΔG is negative, the reaction proceeds spontaneously with the loss of free energy (**exergonic**).

ΔG is positive, the reaction proceeds only when free energy can be gained (**endergonic**), ΔG is 0, the system is at equilibrium, both forward and reverse reactions occur at equal rates,

ΔG of the reaction $A \rightarrow B$ depends on the concentration of reactant and product. At constant temperature and pressure, the following equation can be derived:

$$\Delta G = \Delta G^0 + RT \ln \frac{[B]}{[A]}$$

Where ΔG^0 is the standard free energy change,

R is the gas constant

T is the absolute temperature,

$[A]$ and $[B]$ are the actual concentrations of reactant and product

Enthalpy, H , is the heat content of the reacting system. It reflects the number and kinds of chemical bonds in the reactants and products. When a chemical reaction releases heat, it is said to be exothermic; the heat content of the products is less than that of the reactants and ΔH has, by convention, a negative value. Reacting systems that take up heat from their surroundings are endothermic and have positive values of ΔH .

Entropy, S , is a quantitative expression for the randomness or disorder in a system. When the products of a reaction are less complex and more disordered than the reactants, the reaction is said to proceed with a gain in entropy.

The units of ΔG and ΔH are joules/mole or calories/mole (recall that 1 cal = 4.184 J); units of entropy are joules/mole.Kelvin (J/mol.K).

Under the conditions existing in biological systems (including constant temperature and pressure), changes in free energy, enthalpy, and entropy are related to each other quantitatively by the equation

$$\Delta G = \Delta H - T\Delta S$$

in which ΔG is the change in Gibbs free energy of the reacting system, ΔH is the change in enthalpy of the system, T is the absolute temperature, and ΔS is the change in entropy of the system. By convention, ΔS has a positive sign when entropy increases and ΔH , as

noted above, has a negative sign when heat is released by the system to its surroundings. Either of these conditions, which are typical of favorable processes, tends to make ΔG negative. In fact, ΔG of a spontaneously reacting system is always negative. The second law of thermodynamics states that the entropy *of the universe* increases during all chemical and physical processes, but it does not require that the entropy increase take place *in the reacting system* itself. The order produced within cells as they grow and divide is more than compensated for by the disorder they create in their surroundings in the course of growth and division. In short, living organisms preserve their internal order by taking from the surroundings free energy in the form of nutrients or sunlight, and returning to their surroundings an equal amount of energy as heat and entropy.

Carbohydrates

The carbohydrates, often termed as sugars, are the '*staff of life*' for most organisms. On the basis of mass, they are the most abundant class of biomolecules in nature. Carbohydrates are also known as *saccharides* (*sakcharon* G = sugar or sweetness) since many of those of relatively small molecular weight have a sweet taste, although this is not true of those with large molecules. They are widely distributed molecules (*moles* L = mass) in both plant and animal tissues. They are indispensable for living organisms, serving as skeletal structures in plants and also in insects and crustaceans. Plants are considerably richer in carbohydrates in comparison to the animals.

The term 'carbohydrate' was originally coined for this class of compounds as most of them were 'hydrates of carbon' or could be represented by the general formula, $C_x(H_2O)_y$. Later, it was found that some of them, such as deoxyribose ($C_5H_{10}O_4$) and rhamnose ($C_6H_{12}O_5$) do not have the required ratio of hydrogen to oxygen. In addition, certain other carbohydrates are now known to possess nitrogen (e.g., glucosamine, $C_6H_{13}O_5N$), phosphorus or sulfur also and obviously do not coincide with the above general formula. Furthermore, formaldehyde ($H.CHO$ or CH_2O), acetic acid ($CH_3.COOH$ or $C_2H_4O_2$) and lactic acid ($CH_3.CHOH.COOH$ or $C_3H_6O_3$) which have C, H and O and the ratio of H : O is also the same as in water, but are not a carbohydrates. Hence, the continued usage of the term 'carbohydrate' is for convenience rather than exactness. To accommodate a wide variety of compounds, the carbohydrates are nowadays broadly defined as polyhydroxy aldehydes or ketones and their derivatives or as substances that yield one of these compounds on hydrolysis.

NOMENCLATURE AND DEFINITION

The term 'carbohydrate' was originally coined for this class of compounds as most of them were 'hydrates of carbon' or could be represented by the general formula, $C_x(H_2O)_y$. Later, it was found that some of them, such as deoxyribose ($C_5H_{10}O_4$) and rhamnose ($C_6H_{12}O_5$) do not have the required ratio of hydrogen to oxygen. In addition, certain other carbohydrates are now known to possess nitrogen (e.g., glucosamine, $C_6H_{13}O_5N$), phosphorus or sulfur also and obviously do not coincide with the above general formula. Furthermore, formaldehyde ($H.CHO$ or CH_2O), acetic acid ($CH_3.COOH$ or $C_2H_4O_2$) and lactic acid ($CH_3.CHOH.COOH$ or $C_3H_6O_3$) which have C, H and O and the ratio of H : O is also the same as in water, but are not a carbohydrates. Hence, the continued usage of the term 'carbohydrate' is for convenience rather than exactness.

To accommodate a wide variety of compounds, the carbohydrates are nowadays broadly defined as polyhydroxy aldehydes or ketones and their derivatives or as substances that yield one of these compounds on hydrolysis.

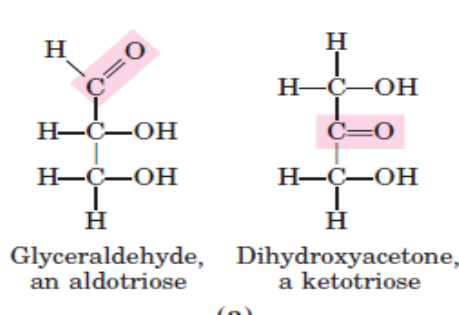
CLASSIFICATION

Carbohydrates are usually classified in 3 groups :

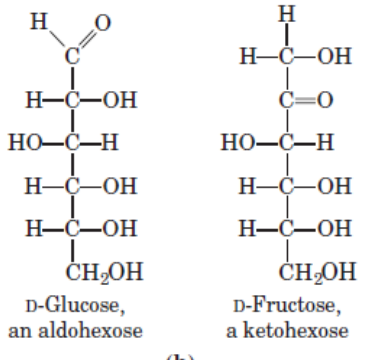
- A. **Monosaccharides or Monosaccharoses** (monoG = one; sakcharonG = sugar).
The monosaccharides, often called simple sugars, are compounds which possess a free aldehyde ($-CHO$) or ketone ($=CO$) group and 2 or more hydroxyl ($-OH$) groups. They are, in fact, the simplest sugars and cannot be hydrolyzed into smaller units. Their general formula is $C_n(H_2O)_n$ or $C_nH_{2n}O_n$. The monosaccharides may be subdivided into trioses, tetroses, pentoses, hexoses,

heptoses etc., depending upon the number of carbon atoms they possess; and as aldoses or ketoses, depending upon whether they contain aldehyde or ketone group. Some important examples are :

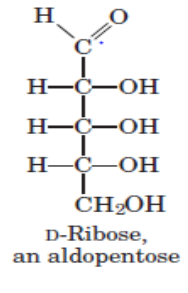
Name	Formula	Aldoses	Ketoses
		(Aldo sugars)	(Keto sugars)
Trioses	$C_3H_6O_3$	Glycerose	Dihydroxyacetone
Tetroses	$C_4H_8O_4$	Erythrose	Erythrulose
Pentoses	$C_5H_{10}O_5$	Ribose	Ribulose
Hexoses	$C_6H_{12}O_6$	Glucose	Fructose
Heptoses	$C_7H_{14}O_7$	Glucoheptose	Sodoheptulose



(a)



(b)



(c)

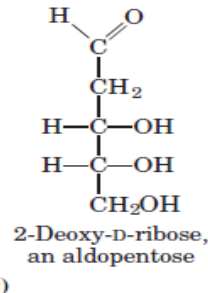


Figure 1 **Representative monosaccharides.** (a) Two trioses, an aldose and a ketose. The carbonyl group in each is shaded. (b) Two common hexoses. (c) The pentose components of nucleic acids. D-Ribose is a component of ribonucleic acid (RNA). 2-Deoxy-D-ribose is a component of deoxyribonucleic acid (DNA).

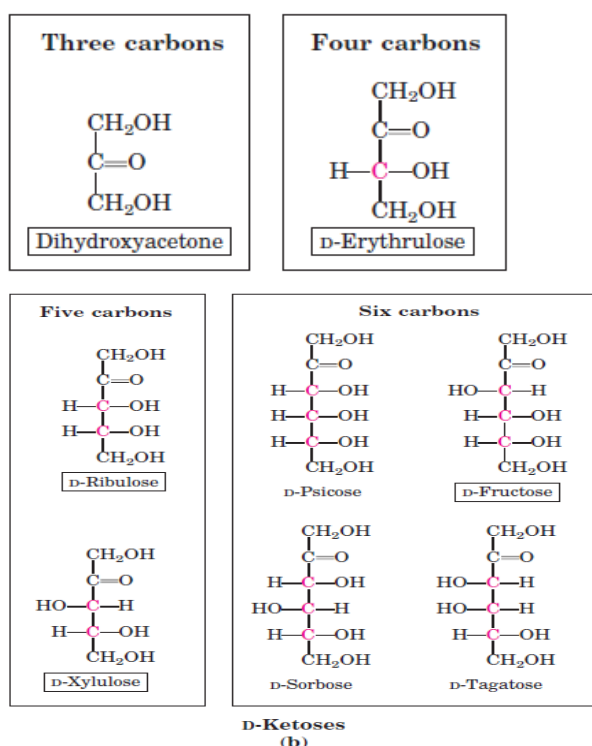
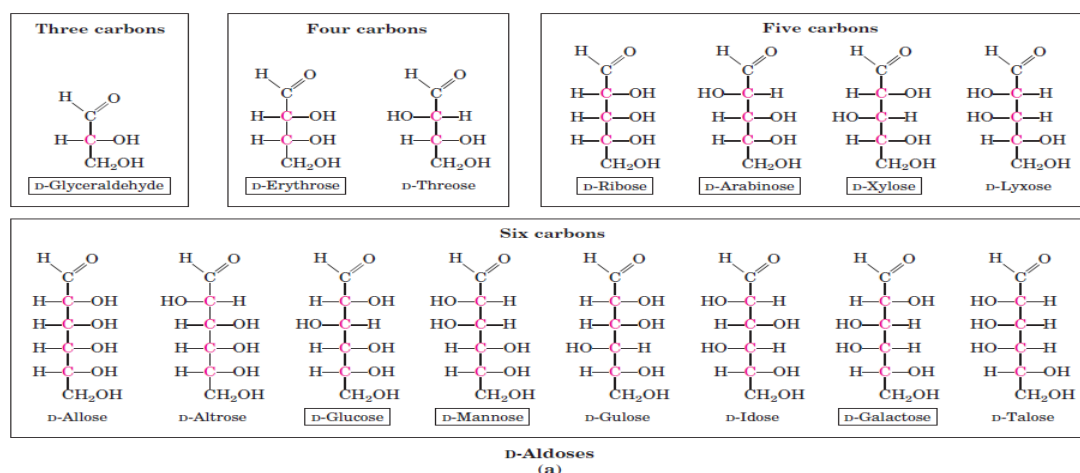


FIGURE 2 Aldoses and ketoses.

The series of (a) D-aldoses and (b) D-ketoses having from three to six carbon atoms, shown as projection formulas. The carbon atoms in red are chiral centers. In all these D isomers, the chiral carbon most distant from the carbonyl carbon has the same configuration as the chiral carbon in D-glyceraldehyde. The sugars named in boxes are the most common in nature.

Isomerism in carbohydrates

Monosaccharides Have Asymmetric Centers:

All the monosaccharides except dihydroxyacetone contain one or more asymmetric (chiral) carbon atoms and thus occur in optically active isomeric forms (pp. 17–19). The simplest aldose, glyceraldehyde, contains one chiral center (the middle carbon atom) and therefore has two different optical isomers, or enantiomers (Fig. 3).

By convention, one of these two forms is designated the D isomer, the other the L isomer. As for other biomolecules with chiral centers, the absolute configurations of sugars are known from x-ray crystallography. To represent three-dimensional sugar structures on paper, we often use Fischer projection formulas (Fig. 3). In general, a molecule with n chiral centers can have 2^n stereoisomers. Glyceraldehyde has $2^1 = 2$; the aldohexoses, with four chiral centers, have $2^4 = 16$ stereoisomers. The stereoisomers of monosaccharides of each carbon-chain length can be divided into

two groups that differ in the configuration about the chiral center most distant from the carbonyl carbon. Those in which the configuration at this reference carbon is the same as that of D-glyceraldehyde are designated D isomers, and those with the same configuration as L-glyceraldehyde are L isomers. When the hydroxyl group on the reference carbon is on the right in the projection formula, the sugar is the D isomer; when on the left, it is the L isomer. Of the 16 possible aldohexoses, eight are D forms and eight are L. Most of the hexoses of living organisms are D isomers.

Figure 2 shows the structures of the D stereoisomers of all the aldoses and ketoses having three to six carbon atoms. Two sugars that differ only in the configuration around one carbon atom are called epimers; D-glucose and D-mannose, which differ only in the stereochemistry at C-2, are epimers, as are D-glucose and D-galactose (which differ at C-4) (Fig. 4).

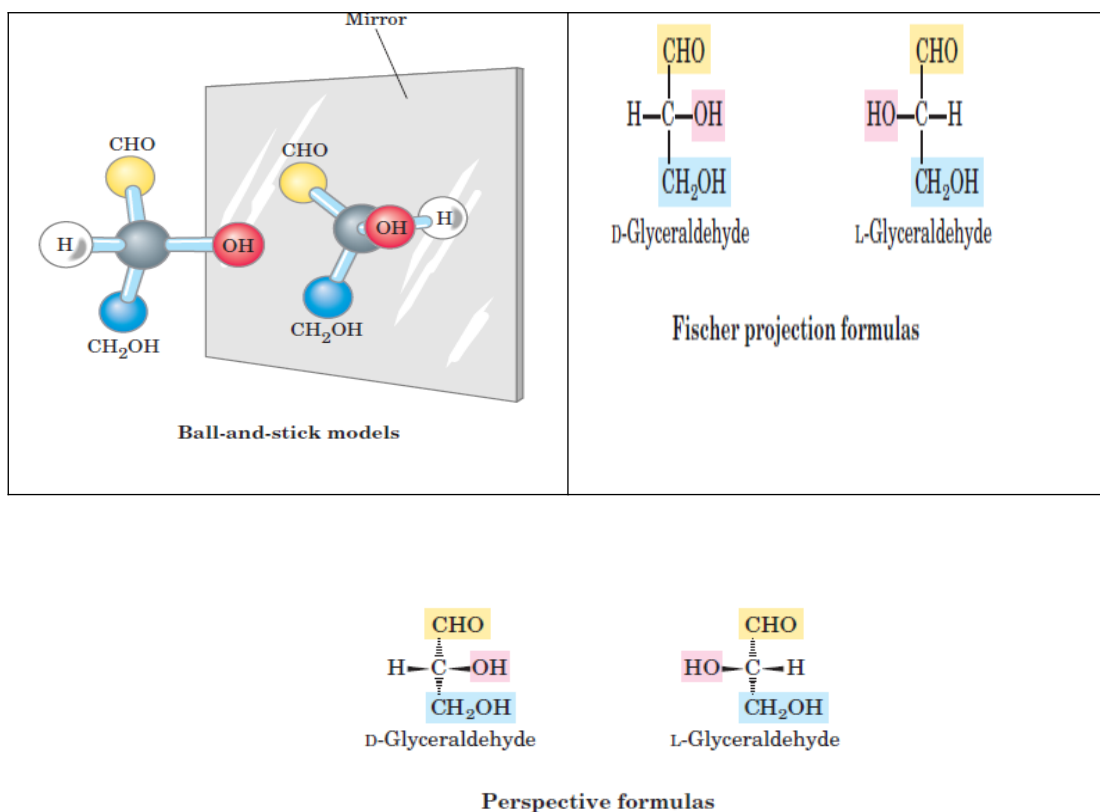


FIGURE 3. Three ways to represent the two stereoisomers of glyceraldehyde. The stereoisomers are mirror images of each other. Ball- and- stick models show the actual configuration of molecules. By convention, in Fischer projection formulas, horizontal bonds project out of the plane of the paper, toward the reader; vertical bonds project behind the plane of the paper, away from the reader.

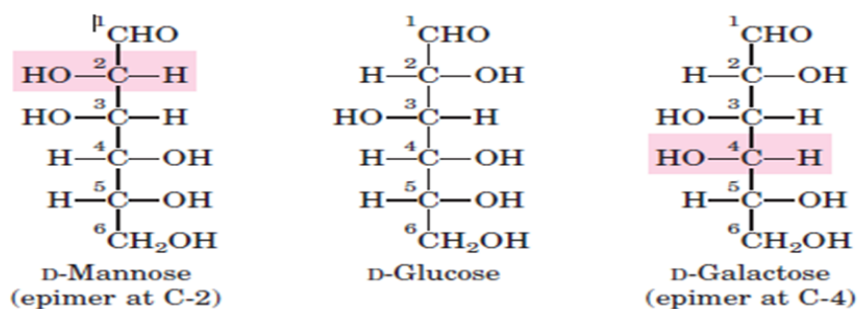


Fig. 4 Epimers. D-Glucose and two of its epimers are shown as projection formulas. Each epimer differs from D-glucose in the configuration at one chiral center (shaded).

The Common Monosaccharides Have Cyclic Structures

In aqueous solution, aldotetroses and all monosaccharides with five or more carbon atoms in the backbone occur predominantly as cyclic (ring) structures in which the carbonyl group has formed a covalent bond with the oxygen of a hydroxyl group along the chain. The formation of these ring structures is the result of a general reaction between alcohols and aldehydes or ketones to form derivatives called hemiacetals or hemiketals (Fig. 5), which contain an additional asymmetric carbon atom and thus can exist in two stereoisomeric forms. For example, D-glucose exists in solution as an intramolecular hemiacetal in which the free hydroxyl group at C-5 has reacted with the aldehydic C-1, rendering the latter carbon asymmetric and producing two stereoisomers, designated α and β (Fig. 6). These six-membered ring compounds are called pyranoses because they resemble the six-membered ring compound pyran (Fig. 7-7). The systematic names for the two ring forms of D-glucose are α -D-glucopyranose and β -D-glucopyranose.

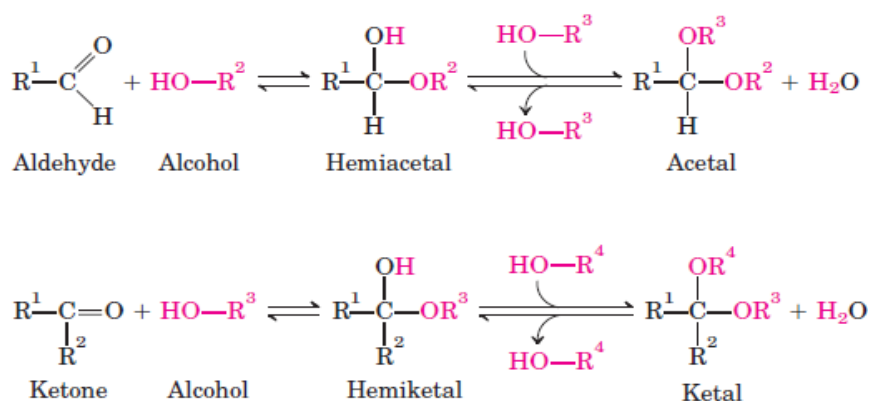


Fig. 5 Formation of hemiacetals and hemiketals. An aldehyde or ketone can react with an alcohol in a 1:1 ratio to yield a hemiacetal or hemiketal, respectively, creating a new chiral center at the carbonyl carbon. Substitution of a second alcohol molecule produces an acetal or ketal. When the second alcohol is part of another

sugar molecule, the bond produced is a glycosidic bond.

Aldohexoses also exist in cyclic forms having five membered rings, which, because they resemble the five membered ring compound furan, are called furanoses. Isomeric forms of monosaccharides that differ only in their configuration about the hemiacetal or hemiketal carbon atom are called anomers. The hemiacetal (or carbonyl) carbon atom is called the anomeric carbon. The α and β anomers of D-glucose interconvert in aqueous solution by a process called mutarotation. Thus, a solution of α -D-glucose and a solution of β -D-glucose eventually form identical equilibrium mixtures having identical optical properties. This mixture consists of about one-third α -D-glucose, two-thirds β -D-glucose, and very small amounts of the linear and five-membered ring (glucofuranose) forms. Ketohexoses also occur in α and β anomeric forms. In these compounds the hydroxyl group at C-5 (or C-6) reacts with the keto group at C-2, forming a furanose (or pyranose) ring containing a hemiketal linkage (Fig. 5). D-Fructose readily forms the furanose ring (Fig. 7); the more common anomer of this sugar in combined forms or in derivatives is β -D-fructofuranose.

Haworth perspective formulas like those in Fig. 7 are commonly used to show the stereochemistry of ring forms of monosaccharides. However, the six-membered pyranose ring is not planar, as Haworth perspectives suggest, but tends to assume either of two "chair" conformations (Fig. 8).

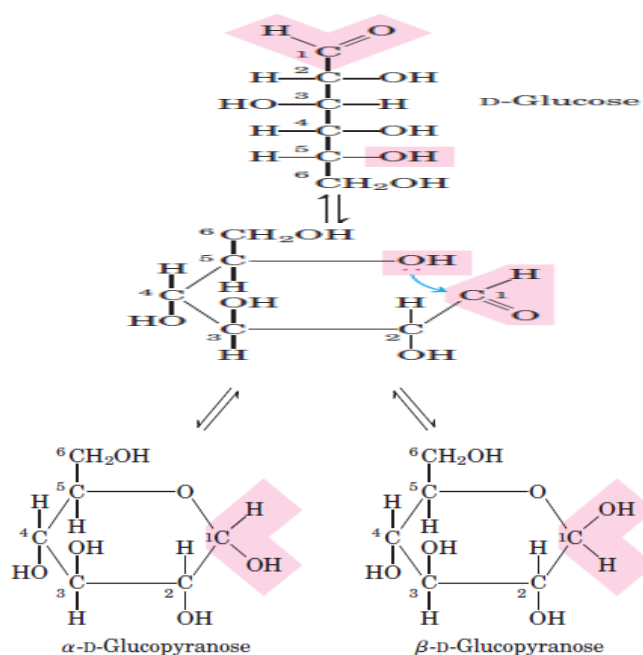


Fig.6. Formation of the two cyclic forms of D-glucose. Reaction between the aldehyde group at C-1 and the hydroxyl group at C-5 forms a hemiacetal linkage, producing either of two stereoisomers, the α and β anomers, which differ only in the stereochemistry around the hemiacetal carbon. The interconversion of α and β

anomers is called mutarotation.

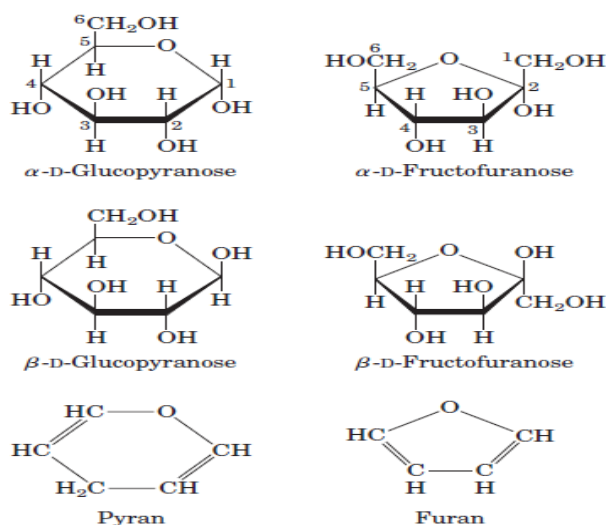


Fig. 7. Pyranoses and furanoses. The pyranose forms of D-glucose and the furanose forms of D-fructose are shown here as Haworth perspective formulas. The edges of the ring nearest the reader are represented by bold lines. Hydroxyl groups below the plane of the ring in these Haworth perspectives would appear at the right side of a Fischer projection (compare with Fig. 6). Pyran and furan are shown for comparison.

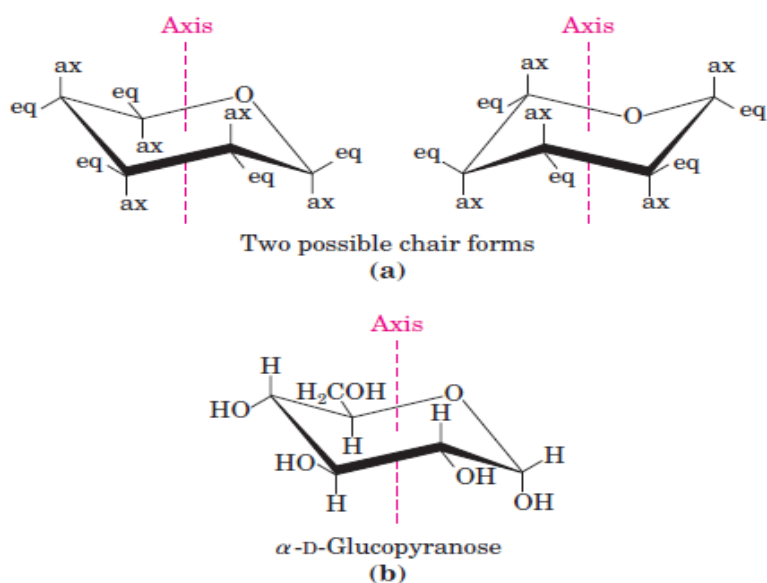


Fig. 8. Conformational formulas of pyranoses. (a) Two chair forms of the pyranose ring. Substituents on the ring carbons may be either axial (ax), projecting parallel to the vertical axis through the ring, or equatorial (eq), projecting roughly perpendicular to this axis. Two conformers such as these are not readily interconvertible without breaking the ring. However, when the molecule is "stretched" (by atomic force microscopy), an input of about 46 kJ of energy per mole of sugar can force the interconversion of chair forms. Generally, substituents in the equatorial positions are less sterically hindered by neighboring substituents, and conformers with bulky substituents in equatorial positions are favored. Another conformation,

the “boat” (not shown), is seen only in derivatives with very bulky substituents. (b) A chair conformation of -D-glucopyranose.

B. Oligosaccharides or Oligosaccharoses (oligo = few).

These are compound sugars that yield 2 to 10 molecules of the same or different monosaccharides on hydrolysis. Accordingly, an oligosaccharide yielding 2 molecules of monosaccharide on hydrolysis is designated as a disaccharide, and the one yielding 3 molecules of monosaccharide as a trisaccharide and so on. The general formula of disaccharides is $C_n(H_2O)_n - 1$ and that of trisaccharides is $C_n(H_2O)_n - 2$ and so on. A few examples are :

Disaccharides – Sucrose, Lactose, Maltose, Cellobiose, Trehalose, Gentiobiose, Melibiose

Trisaccharides – Rhamnose, Gentianose, Raffinose (= Melitose), Rabinose, Melezitose

Tetrasaccharides – Stachyose, Scorodose

Pentasaccharide – Verbascose

Disaccharides Contain a Glycosidic Bond

Disaccharides (such as maltose, lactose, and sucrose) consist of two monosaccharides joined covalently by an O-glycosidic bond, which is formed when a hydroxyl group of one sugar reacts with the anomeric carbon of the other (Fig. 7–11). This reaction represents the formation of an acetal from a hemiacetal (such as glucopyranose) and an alcohol (a hydroxyl group of the second sugar molecule) (Fig. 7–5). Glycosidic bonds are readily hydrolyzed by acid but resist cleavage by base. Thus disaccharides can be hydrolyzed to yield their free monosaccharide components by boiling with dilute acid. N-glycosyl bonds join the anomeric carbon of a sugar to a nitrogen atom in glycoproteins (see Fig. 7–31) and nucleotides (see Fig. 8–1).

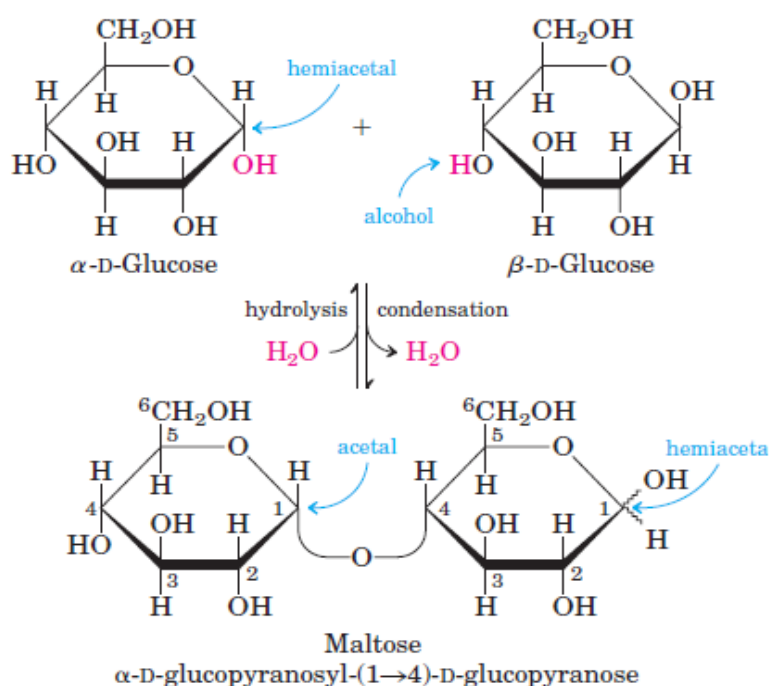


Fig. 11 Formation of maltose. A disaccharide is formed from two monosaccharides (here,

two molecules of D-glucose) when an -OH (alcohol) of one glucose molecule (right) condenses with the intramolecular hemiacetal of the other glucose molecule (left), with elimination of H₂O and formation of an O-glycosidic bond. The reversal of this reaction is hydrolysis—attack by H₂O on the glycosidic bond.

C. Polysaccharides or Polysaccharoses (poly = many).

These are also compound sugars and yield more than 10 molecules of monosaccharides on hydrolysis. These may be further classified depending on whether the monosaccharide molecules produced as a result of the hydrolysis of polysaccharides are of the same type (homopolysaccharides) or of different types (heteropolysaccharides). Their general formula is (C₆H₁₀O₅)_x. Some common examples are :

Homopolysaccharides: – Starch, Glycogen, Inulin, Cellulose, Pectin, and Chitin.

Heteropolysaccharides: – “Specific soluble sugar” of pneumococcus type III, Hyaluronic acid, Chondrotin.

Starch contains two types of glucose polymer, amylose and amylopectin (Fig. 15). The former consists of long, unbranched chains of D-glucose residues connected by (α1 → 4) linkages. Such chains vary in molecular weight from a few thousand to more than a million. Amylopectin also has a high molecular weight (up to 100 million) but unlike amylose is highly branched. The glycosidic linkages joining successive glucose residues in amylopectin chains are (α1 → 4); the branch points (occurring every 24 to 30 residues) are (α1 → 6) linkages.

Glycogen is the main storage polysaccharide of animal cells. Like amylopectin, glycogen is a polymer of (α1 → 4)-linked subunits of glucose, with (α1 → 6)-linked branches, but glycogen is more extensively branched (on average, every 8 to 12 residues) and more compact than starch. Glycogen is especially abundant in the liver, where it may constitute as much as 7% of the wet weight; it is also present in skeletal muscle. In hepatocytes glycogen is found in large granules.

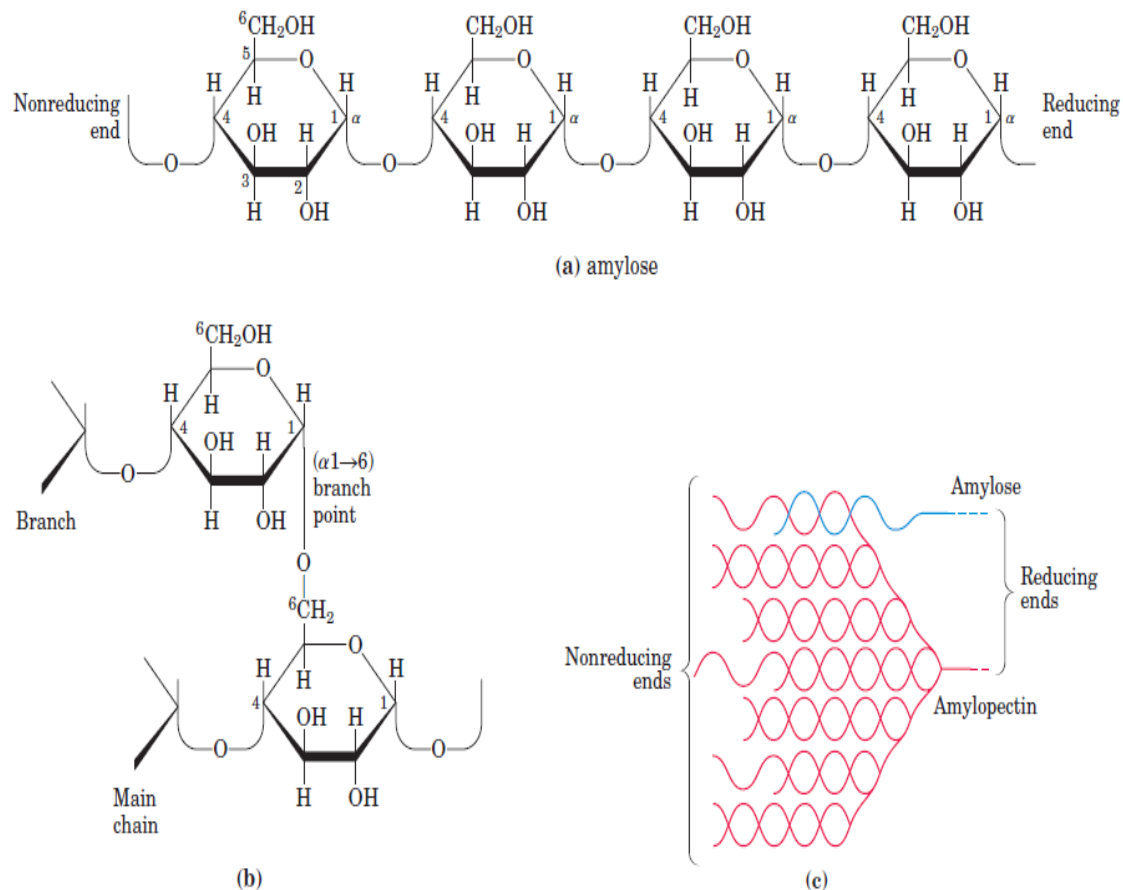


Fig.15 Amylose and amylopectin, the polysaccharides of starch. (a) A short segment of amylose, a linear polymer of D-glucose residues in ($\alpha 1 \rightarrow 4$) linkage. A single chain can contain several thousand glucose residues. Amylopectin has stretches of similarly linked residues between branch points. (b) An ($\alpha 1 \rightarrow 6$) branch point of amylopectin. (c) A cluster of amylose and amylopectin like that believed to occur in starch granules. Strands of amylopectin form double helical structures with each other or with amylose strands. Glucose residues at the non-reducing ends of the outer branches are removed enzymatically during the mobilization of starch for energy production. Glycogen has a similar structure but is more highly branched and more compact.

Cellulose

Cellulose Fig. 16. is not only the most abundant extracellular structural polysaccharide of the plant world but is also undoubtedly the most abundant of all biomolecules in the biosphere. It is present in all land plants, but is completely lacking in meat, egg, fish and milk. It is, however, not metabolized by the human system. Cellulose occurs in the cell walls of plants where it contributes in a major way to the structure of the organism. Cellulose is often found associated with other structural substances such as lignin. Plant residues in soil consist of 40- 70% cellulose. It occurs in nearly pure form in cotton (94%), flax (90%) and in the wood of many plants (60%). It is not usually found in bacterial cell walls, but it is the substance which keeps the cells of *Sarcina ventriculi*

together in large packets or bunches.

The molecular weight of cellulose ranges between 200,000 and 2,000,000, thus corresponding to 1,250–12,500 glucose residues per molecule. It may be formed by taking out a molecule of water from the glycosidic OH group on carbon atom 1 of one β -D-glucose molecule and the alcoholic OH group on carbon atom 4 of the adjacent β -D-glucose molecule. It, thus, resembles in structure with amylose except that the glucose units are linked together by β -1, 4-glucoside linkages. Cellulose may, henceforth, be regarded as an anhydride of β -D-glucose units.

It is a fibrous, tough, white solid, insoluble in water but soluble in ammoniacal cupric hydroxide solution (Schweitzer's reagent). It gives no colour with iodine and lacks sweetness. Although insoluble in water, cellulose absorbs water and adds to the bulk of the fecal matter and facilitates its removal.

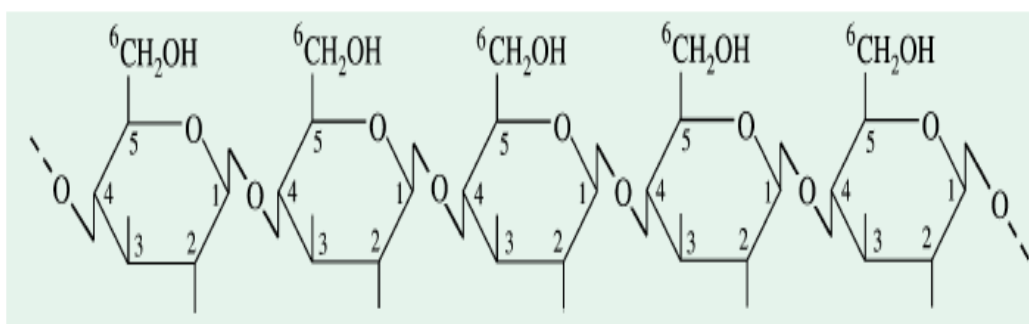


Fig. 16. Structure of Cellulose

Glycoproteins

Glycoproteins have one or several oligosaccharides of varying complexity joined covalently to a protein. They are found on the outer face of the plasma membrane, in the extracellular matrix, and in the blood. Inside cells they are found in specific organelles such as Golgi complexes, secretory granules, and lysosomes. The oligosaccharide portions of glycoproteins are less monotonous than the glycosaminoglycan chains of proteoglycans; they are rich in information, forming highly specific sites for recognition and high-affinity binding by other proteins.

Glycoproteins are carbohydrate-protein conjugates in which the carbohydrate moieties are smaller and more structurally diverse than the glycosaminoglycans of proteoglycans. The carbohydrate is attached at its anomeric carbon through a glycosidic link to the -OH of a Ser or Thr residue (O-linked), or through an N-glycosyl link to the amide nitrogen of an Asn residue (N-linked) (Fig. 17). Some glycoproteins have a single oligosaccharide chain, but many have more than one; the carbohydrate may constitute from 1% to 70% or more of the glyco protein by mass. The structures of a large number of O- and N-linked oligosaccharides from a variety of glycoproteins are known.

Many of the proteins secreted by eukaryotic cells are glycoproteins, including most of the proteins of blood. For example, immunoglobulins (antibodies) and certain hormones, such as follicle-stimulating hormone, luteinizing hormone, and

thyroid-stimulating hormone, are glycoproteins. Many milk proteins, including lactalbumin, and some of the proteins secreted by the pancreas (such as ribonuclease) are glycosylated, as are most of the proteins contained in lysosomes.

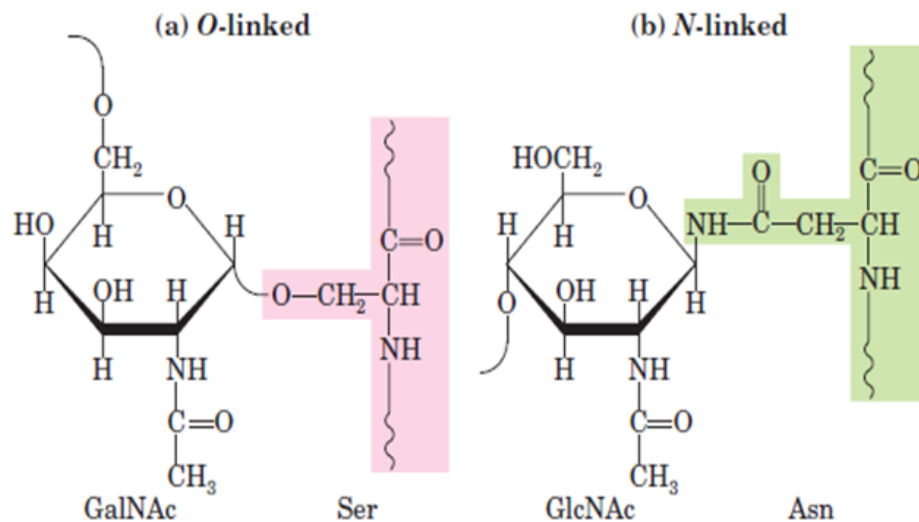


Fig. 17. Oligosaccharide linkages in glycoproteins. (a) O-linked oligosaccharides have a glycosidic bond to the hydroxyl group of Ser or Thr residues (shaded pink), illustrated here with GalNAc as the sugar at the reducing end of the oligosaccharide. One simple chain and one complex chain are shown. (b) N-linked oligosaccharides have an N-glycosyl bond to the amide nitrogen of an Asn residue (shaded green), illustrated here with GlcNAc as the terminal sugar. Three common types of oligosaccharide chains that are N-linked in glycoproteins are shown.

Peptidoglycan

Peptidoglycan, Fig.18 or murein, is an enormous mesh like polymer composed of many identical subunits. The rigid component of bacterial cell walls is a heteropolymer of alternating ($\beta 1 \rightarrow 4$)-linked N-acetylglucosamine and N-acetylmuramic acid residues as peptide chain of four alternating D- and L-amino acids is connected to the carboxyl group of Nacetylmuramic acid. Many bacteria replace meso-diaminopimelic acid with another diaminoacid, usually L-lysine. The linear polymers lie side by side in the cell wall, crosslinked by short peptides, the exact structure of which depends on the bacterial species. The peptide cross-links weld the polysaccharide chains into a strong sheath that envelops the entire cell and prevents cellular swelling and lysis due to the osmotic entry of water. The enzyme lysozyme kills bacteria by hydrolyzing the ($\beta 1 \rightarrow 4$) glycosidic bond between N-acetylglucosamine and Nacetylmuramic acid (see Fig. 6-24). Lysozyme is notably present in tears, presumably as a defense against bacterial

infections of the eye. It is also produced by certain bacterial viruses to ensure their release from the host bacterial cell, an essential step of the viral infection cycle. Penicillin and related antibiotics kill bacteria by preventing synthesis of the cross-links, leaving the cell wall too weak to resist osmotic lysis.

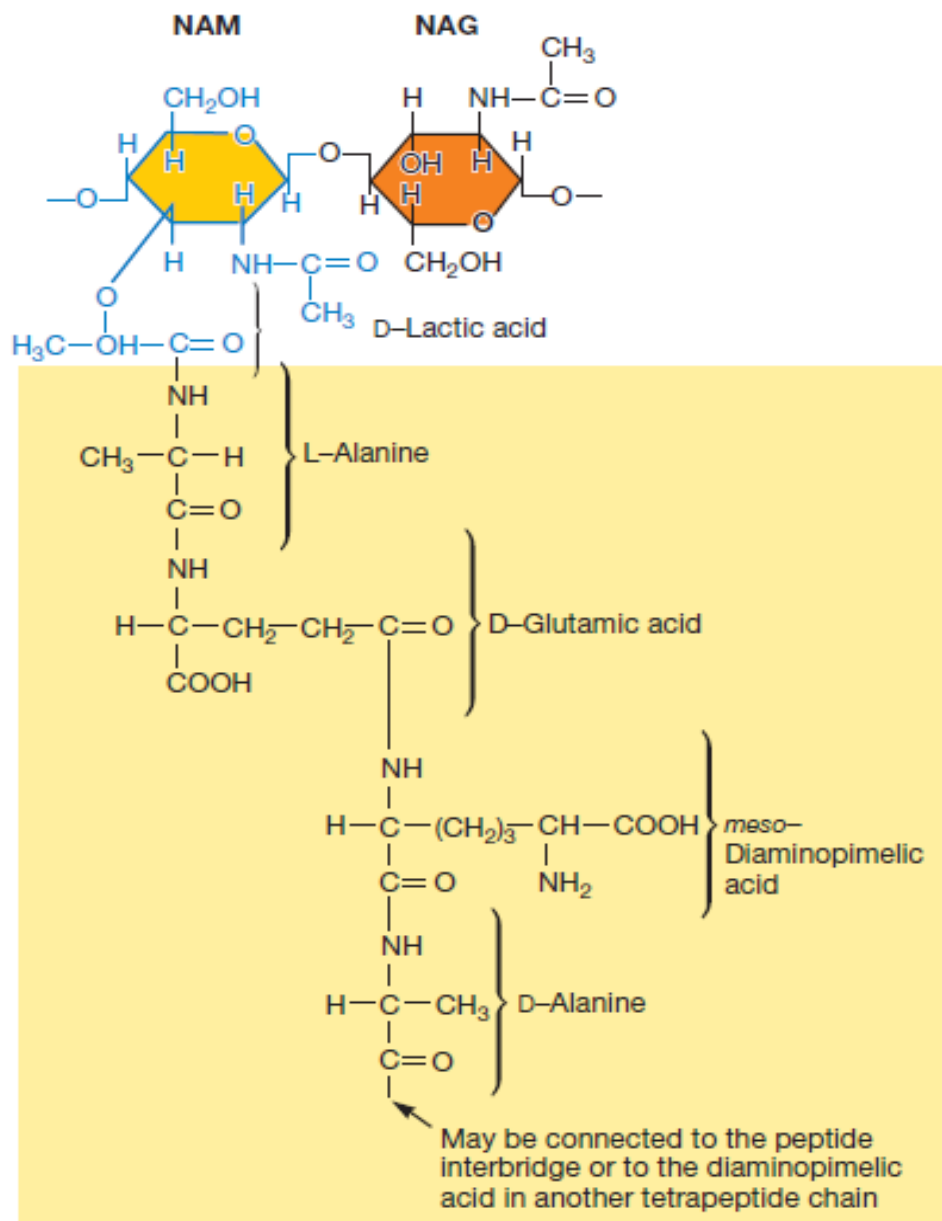


Fig. 18 Peptidoglycan Subunit Composition. The peptidoglycan subunit of *E. coli*, most other gram-negative bacteria, and many gram-positive bacteria. NAG is N-acetylglucosamine. NAM is N-acetylmuramic acid (NAG with lactic acid attached by an ether linkage). The tetrapeptide side chain is composed of alternating D- and L-amino acids since meso-diaminopimelic acid is connected through its L-carbon. NAM and the tetrapeptide chain attached to it are shown in different.

Lipopolysaccharides

Lipopolysaccharides (LPSs) are the most unusual constituents of the outer membrane. These large, complex molecules contain both lipid and carbohydrate, and consist of three parts: (1) lipid A, (2) the core polysaccharide, and (3) the O side chain. The LPS from *Salmonella* has been studied most, and its general structure is described here (fig.19). The lipid A region contains two glucosamine sugar derivatives, each with three fatty acids and phosphate or pyrophosphate attached. The fatty acids attach the lipid A to the outer membrane, while the remainder of the LPS molecule projects from the surface. The core polysaccharide is joined to lipid A. In *Salmonella* it is constructed of 10 sugars, many of them unusual in structure. The O side chain or O antigen is a polysaccharide chain extending outward from the core. It has several peculiar sugars and varies in composition between bacterial strains.

LPS has many important functions. Because the core polysaccharide usually contains charged sugars and phosphate (fig. 19), LPS contributes to the negative charge on the bacterial surface. As a major constituent of the exterior leaflet of the outer membrane, lipid A also helps stabilize outer membrane structure. LPS may contribute to bacterial attachment to surfaces and biofilm formation. A major function of LPS is that it aids in creating a permeability barrier. The geometry of LPS (figure 19b) and interactions between neighboring LPS molecules are thought to restrict the entry of bile salts, antibiotics, and other toxic substances that might kill or injure the bacterium. LPS also plays a role in protecting pathogenic gram-negative bacteria from host defenses. The O side chain of LPS is also called the O antigen because it elicits an immune response. This response involves the production of antibodies that bind the strain-specific form of LPS that elicited the response. However, many gram negative bacteria are able to rapidly change the antigenic nature of their O side chains, thus thwarting host defences. Importantly, the lipid A portion of LPS often is toxic; as a result, the LPS can act as an endotoxin and cause some of the symptoms that arise in gram-negative bacterial infections. If the bacterium enters the bloodstream, LPS endotoxin can cause a form of septic shock for which there is no direct treatment.

Despite the role of LPS in creating a permeability barrier, the outer

membrane is more permeable than the plasma membrane and permits the passage of small molecules like glucose and other monosaccharides. This is due to the presence of porin proteins.

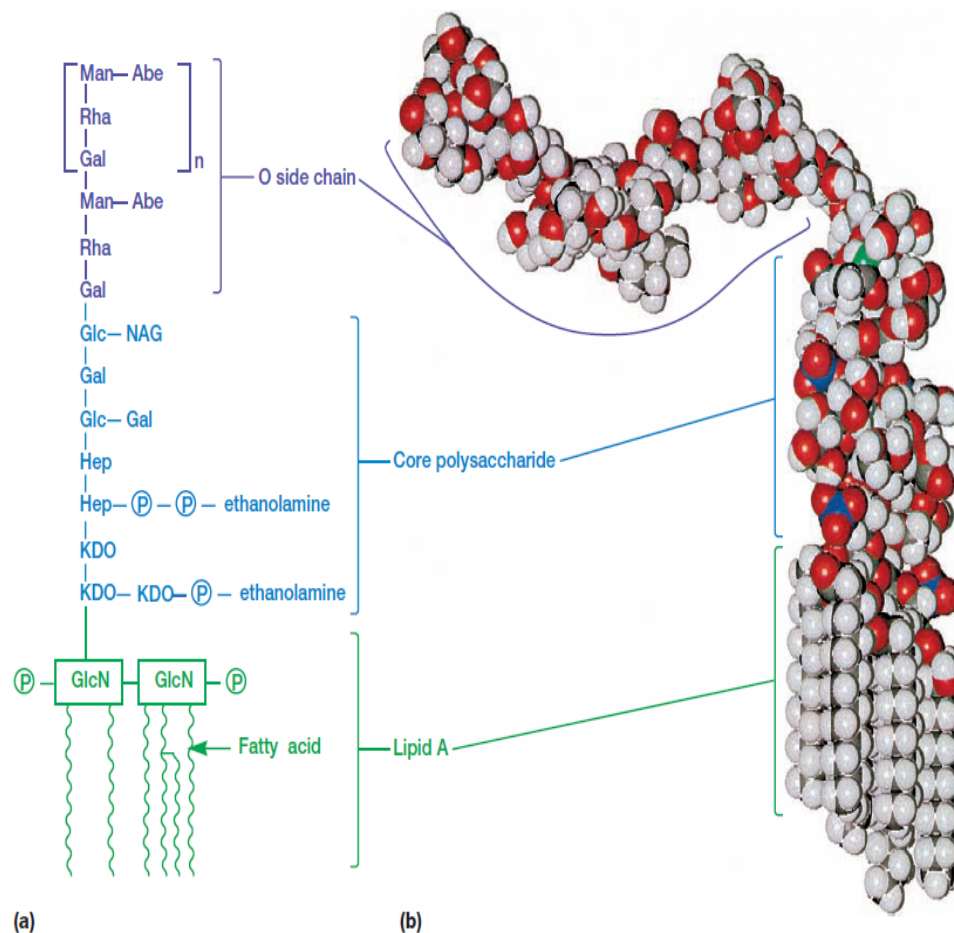


Fig. 19. Lipopolysaccharide Structure. (a) The lipopolysaccharide from Salmonella. This slightly simplified diagram illustrates one form of the LPS. Abbreviations: Abe, abequose; Gal, galactose; Glc, glucose; GlcN, glucosamine; Hep, heptulose; KDO, 2-keto-3-deoxyoctonate; Man, mannose; NAG, N-acetylglucosamine; P, phosphate; Rha, L-rhamnose. Lipid A is buried in the outer membrane. (b) Molecular model of an Escherichia coli lipopolysaccharide. The lipid A and core polysaccharide are straight; the O side chain is bent at an angle in this model.

Lipids

Biological lipids are a chemically diverse group of compounds, the common and defining feature of which is their insolubility in water. The biological functions of the lipids are as diverse as their chemistry. Fats and oils are the principal stored forms of energy in many organisms. Phospholipids and sterols are major structural elements of biological membranes. Other lipids, although present in relatively small quantities, play crucial roles as enzyme cofactors, electron carriers, and light absorbing pigments, hydrophobic anchors for proteins, “chaperones” to help membrane proteins fold, and emulsifying agents in the digestive tract, hormones, and intracellular messengers.

Storage Lipids

The fats and oils used almost universally as stored forms of energy in living organisms are derivatives of fatty acids. The fatty acids are hydrocarbon derivatives, at about the same low oxidation state (that is, as highly reduced) as the hydrocarbons in fossil fuels. The cellular oxidation of fatty acids (to CO₂ and H₂O), like the controlled, rapid burning of fossil fuels in internal combustion engines, is highly exergonic. We introduce here the structures and nomenclature of the fatty acids most commonly found in living organisms. Two types of fatty acid-containing compounds, triacylglycerols and waxes, are described to illustrate the diversity of structure and physical properties in this family of compounds.

Fatty Acids

Fatty acids are carboxylic acids with hydrocarbon chains ranging from 4 to 36 carbons long (C₄ to C₃₆). In some fatty acids, this chain is unbranched and fully saturated (contains no double bonds); in others the

chain contains one or more double bonds (Table 10–1). A few contain three-carbon rings, hydroxyl groups, or methyl group branches. A simplified nomenclature for these compounds specifies the chain length and number of double bonds, separated by a colon; for example, the 16-carbon saturated palmitic acid is abbreviated 16:0, and the 18-carbon oleic acid, with one double bond, is 18:1. The positions of any double bonds are specified by superscript numbers following Δ (delta); a 20-carbon fatty acid with one double bond between C-9 and C-10 (C-1 being the carboxyl carbon) and another between C-12 and C-13 is designated 20:2 ($\Delta^{9,12}$). The most commonly occurring fatty acids have even numbers of carbon atoms in an unbranched chain of 12 to 24 carbons (Table 10–1).

There is also a common pattern in the location of double bonds; in most monounsaturated fatty acids the double bond is between C-9 and C-10 (9), and the other double bonds of polyunsaturated fatty acids are generally 12 and 15. (Arachidonic acid is an exception to this generalization.) The double bonds of polyunsaturated fatty acids are almost never conjugated (alternating single and double bonds, as in OCHUCHOCHUCHO), but are separated by a methylene group: OCHUCHOCH₂OCHUCHO. In nearly all naturally occurring unsaturated fatty acids, the double bonds are in the cis-configuration. Trans fatty acids are produced by fermentation in the rumen of dairy animals and are obtained from dairy products and meat.

TABLE 10-1 Some Naturally Occurring Fatty Acids: Structure, Properties, and Nomenclature

Carbon skeleton	Structure*	Systematic name [†]	Common name (derivation)	Melting point (°C)	Solubility at 30 °C (mg/g solvent)	
					Water	Benzene
12:0	$\text{CH}_3(\text{CH}_2)_{10}\text{COOH}$	<i>n</i> -Dodecanolic acid	Lauric acid (Latin <i>laurus</i> , "laurel plant")	44.2	0.063	2,600
14:0	$\text{CH}_3(\text{CH}_2)_{12}\text{COOH}$	<i>n</i> -Tetradecanolic acid	Myristic acid (Latin <i>Myristica</i> , nutmeg genus)	53.9	0.024	874
16:0	$\text{CH}_3(\text{CH}_2)_{14}\text{COOH}$	<i>n</i> -Hexadecanolic acid	Palmitic acid (Latin <i>palma</i> , "palm tree")	63.1	0.0083	348
18:0	$\text{CH}_3(\text{CH}_2)_{16}\text{COOH}$	<i>n</i> -Octadecanolic acid	Stearic acid (Greek <i>stear</i> , "hard fat")	69.6	0.0034	124
20:0	$\text{CH}_3(\text{CH}_2)_{18}\text{COOH}$	<i>n</i> -Eicosanolic acid	Arachidic acid (Latin <i>Arachis</i> , legume genus)	76.5		
24:0	$\text{CH}_3(\text{CH}_2)_{22}\text{COOH}$	<i>n</i> -Tetracosanolic acid	Lignoceric acid (Latin <i>lignum</i> , "wood" + <i>cera</i> , "wax")	86.0		
16:1(Δ^9)	$\text{CH}_3(\text{CH}_2)_5\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	<i>cis</i> -9-Hexadecenolic acid	Palmitoleic acid	1-0.5		
18:1(Δ^9)	$\text{CH}_3(\text{CH}_2)_7\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	<i>cis</i> -9-Octadecenolic acid	Oleic acid (Latin <i>oleum</i> , "oil")	13.4		
18:2($\Delta^{9,12}$)	$\text{CH}_3(\text{CH}_2)_4\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	<i>cis</i> -, <i>cis</i> -9,12-Octadecadienolic acid	Linoleic acid (Greek <i>linon</i> , "flax")	1-5		
18:3($\Delta^{9,12,15}$)	$\text{CH}_3\text{CH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	<i>cis</i> -, <i>cis</i> -, <i>cis</i> -9,12,15-Octadecatrienolic acid	α -Linolenic acid	-11		
20:4($\Delta^{5,8,11,14}$)	$\text{CH}_3(\text{CH}_2)_4\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_3\text{COOH}$	<i>cis</i> -, <i>cis</i> -, <i>cis</i> -, <i>cis</i> -5,8,11,14-Icosatetraenolic acid	Arachidonic acid	-49.5		

The physical properties of the fatty acids, and of compounds that contain them, are largely determined by the length and degree of unsaturation of the hydrocarbon chain. The nonpolar hydrocarbon chain accounts for the poor solubility of fatty acids in water. Lauric acid (12:0, Mr 200), for example, has a solubility in water of 0.063 mg/g—much less than that of glucose (Mr 180), which is 1,100 mg/g. The longer the fatty acyl chain and the fewer the double bonds, the lower is the solubility in water. The carboxylic acid group is polar (and ionized at neutral pH) and accounts for the slight solubility of short-chain fatty acids in water.

Melting points are also strongly influenced by the length and degree of unsaturation of the hydrocarbon chain. At room temperature (25 °C), the saturated fatty acids from 12:0 to 24:0 have a waxy consistency, whereas unsaturated fatty acids of these lengths are oily liquids. This difference in melting points is due to different degrees of

packing of the fatty acid molecules (Fig. 10–1). In the fully saturated compounds, free rotation around each carbon–carbon bond gives the hydrocarbon chain great flexibility; the most stable conformation is the fully extended form, in which the steric hindrance of neighboring atoms is minimized. These molecules can pack together tightly in nearly crystalline arrays, with atoms all along their lengths in van der Waals contact with the atoms of neighboring molecules. In unsaturated fatty acids, a cis-double bond forces a kink in the hydrocarbon chain.

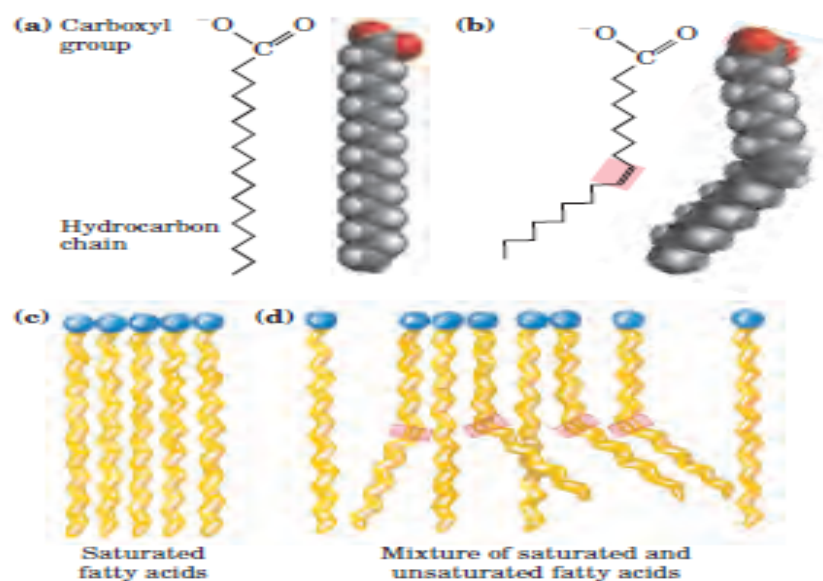


Fig. 10–1. The packing of fatty acids into stable aggregates. The extent of packing depends on the degree of saturation. (a) Two representations of the fully saturated acid stearic acid (stearate at pH 7) in its usual extended conformation. Each line segment of the zigzag represents a single bond between adjacent carbons. (b) The cis double bond (shaded) in oleic acid (oleate) does not permit rotation and introduces a rigid bend in the hydrocarbon tail. All other bonds in the chain are free to rotate. (c) Fully saturated fatty acids in the extended form pack into nearly crystalline arrays, stabilized by many hydrophobic interactions. (d) The presence of one or more cis-double bonds interferes with this tight packing and results in less stable aggregates.

Triacylglycerols

Triacylglycerols are fatty acid esters of glycerol. The simplest lipids constructed from fatty acids are the triacylglycerols, also referred to as triglycerides, fats, or neutral fats. Triacylglycerols are composed of three fatty acids each in ester linkage with a single glycerol (Fig. 10–2). Those containing the same kind of fatty acid in all three positions are

called simple triacylglycerols and are named after the fatty acid they contain. Simple triacylglycerols of 16:0, 18:0, and 18:1, for example, are tristearin, tripalmitin, and triolein, respectively. Most naturally occurring triacylglycerols are mixed; they contain two or more different fatty acids. To name these compounds unambiguously, the name and position of each fatty acid must be specified.

Because the polar hydroxyls of glycerol and the polar carboxylates of the fatty acids are bound in ester linkages, triacylglycerols are nonpolar, hydrophobic molecules, essentially insoluble in water. Lipids have lower specific gravities than water, which explains why mixtures of oil and water (oil-and-vinegar salad dressing, for example) have two phases: oil, with the lower specific gravity, floats on the aqueous phase.

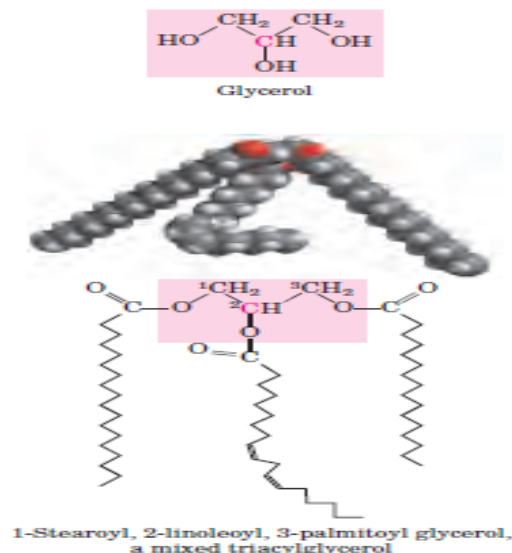


FIGURE 10–2 Glycerol and a triacylglycerol. The mixed triacylglycerol shown here has three different fatty acids attached to the glycerol backbone. When glycerol has two different fatty acids at C-1 and C-3, the C-2 is a chiral center.

In most eukaryotic cells, triacylglycerols form a separate phase of microscopic, oily droplets in the aqueous cytosol, serving as depots of metabolic fuel. In vertebrates, specialized cells called adipocytes, or fat cells, store large amounts of triacylglycerols as fat droplets that nearly fill the cell. Triacylglycerols are also stored as oils in the seeds of many types of plants, providing energy and biosynthetic precursors during seed germination. Adipocytes and germinating seeds contain lipases, enzymes that catalyze the hydrolysis of stored triacylglycerols, releasing fatty acids for export to sites where they are required as fuel. There are two significant advantages to using triacylglycerols as stored fuels, rather than polysaccharides such as glycogen and starch. First, because

the carbon atoms of fatty acids are more reduced than those of sugars oxidation of triacylglycerols yields more than twice as much energy, gram for gram, as the oxidation of carbohydrates. Second, because triacylglycerols are hydrophobic and therefore unhydrated, the organism that carries fat as fuel does not have to carry the extra weight of water of hydration that is associated with stored polysaccharides (2 g per gram of polysaccharide). Humans have fat tissue (composed primarily of adipocytes) under the skin, in the abdominal cavity, and in the mammary glands. Moderately obese people with 15 to 20 kg of triacylglycerols deposited in their adipocytes could meet their energy needs for months by drawing on their fat stores. In contrast, the human body can store less than a day's energy supply in the form of glycogen. Carbohydrates such as glucose and glycogen do offer certain advantages as quick sources of metabolic energy, one of which is their ready solubility in water.

Types of Lipids

There are five general types of membrane lipids: glycerophospholipids, in which the hydrophobic regions are composed of two fatty acids joined to glycerol; galactolipids and sulfolipids, which also contain two fatty acids esterified to glycerol, but lack the characteristic phosphate of phospholipids; archaebacterial tetraether lipids, in which two very long alkyl chains are ether-linked to glycerol at both ends; sphingolipids, in which a single fatty acid is joined to a fatty amine, sphingosine; and sterols, compounds characterized by a rigid system of four fused hydrocarbon rings.

The hydrophilic moieties in these amphipathic compounds may be as simple as a single OOH group at one end of the sterol ring system, or they may be much more complex. In glycerophospholipids and some sphingolipids, a polar head group is joined to the hydrophobic moiety by a phosphodiester linkage; these are the phospholipids. Other sphingolipids lack phosphate but have a simple sugar or complex oligosaccharide at their polar ends; these are the glycolipids (Fig. 10–6). Within these groups of membrane lipids, enormous diversity results from various combinations of fatty acid “tails” and polar “heads.”

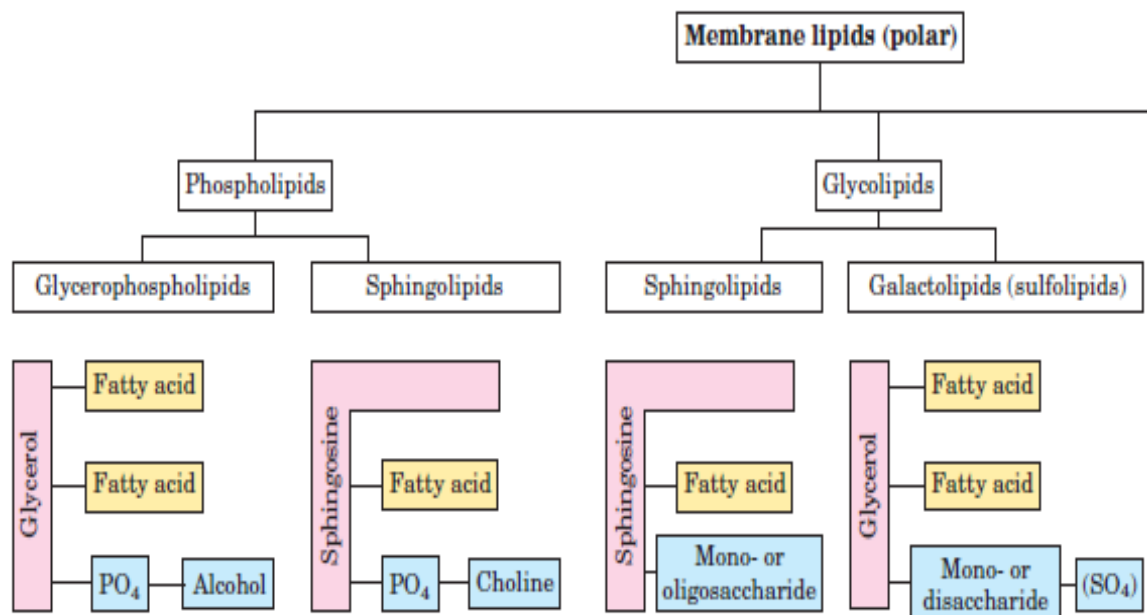


FIGURE 10–6 Some common types of storage and membrane lipids. All the lipid types shown here have either glycerol or sphingosine as the backbone (pink screen), to which are attached one or more longchain alkyl groups (yellow) and a polar head group (blue). In triacylglycerols, glycerophospholipids, galactolipids, and sulfolipids, the alkyl groups are fatty acids in ester linkage. Sphingolipids contain a single fatty acid, in amide linkage to the sphingosine backbone. The membrane lipids of archaebacteria are variable; that shown here has two very long, branched alkyl chains, each end in ether linkage with a glycerol moiety. In phospholipids the polar head group is joined through a phosphodiester, whereas glycolipids have a direct glycosidic linkage between the head-group sugar and the backbone glycerol.

Glycerophospholipids:

Glycerophospholipids, also called phosphoglycerides, are membrane lipids in which two fatty acids are attached in ester linkage to the first and second carbons of glycerol, and a highly polar or charged group is attached through a phosphodiester linkage to the third carbon. Glycerol is prochiral; it has no asymmetric carbons, but attachment of phosphate at one end converts it into a chiral compound, which can be correctly named either L-glycerol 3-phosphate, D-glycerol 1-phosphate, or *sn* glycerol 3-phosphate (Fig. 10–7). Glycerophospholipids are named as derivatives of the parent compound, phosphatidic acid (Fig. 10–8), according to the polar alcohol in the head group. Phosphatidylcholine and phosphatidylethanolamine have choline and ethanolamine in their polar head groups, for example. In all these compounds, the head group is joined to glycerol through a

phosphodiester bond, in which the phosphate group bears a negative charge at neutral pH. The polar alcohol may be negatively charged (as in phosphatidylinositol 4,5-bisphosphate), neutral (phosphatidylserine), or positively charged (phosphatidylcholine, phosphatidylethanolamine). As we shall see in Chapter 11, these charges contribute greatly to the surface properties of membranes. The fatty acids in glycerophospholipids can be any of a wide variety, so a given phospholipid (phosphatidylcholine, for example) may consist of a number of molecular species, each with its unique complement of fatty acids. The distribution of molecular species is specific for different organisms, different tissues of the same organism, and different glycerophospholipids in the same cell or tissue. In general, glycerophospholipids contain a C16 or C18 saturated fatty acid at C-1 and a C18 to C20 unsaturated fatty acid at C-2. With few exceptions, the biological significance of the variation in fatty acids and head groups is not yet understood.

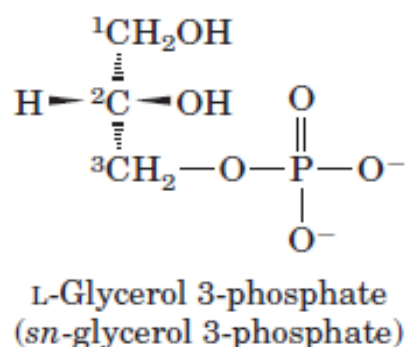
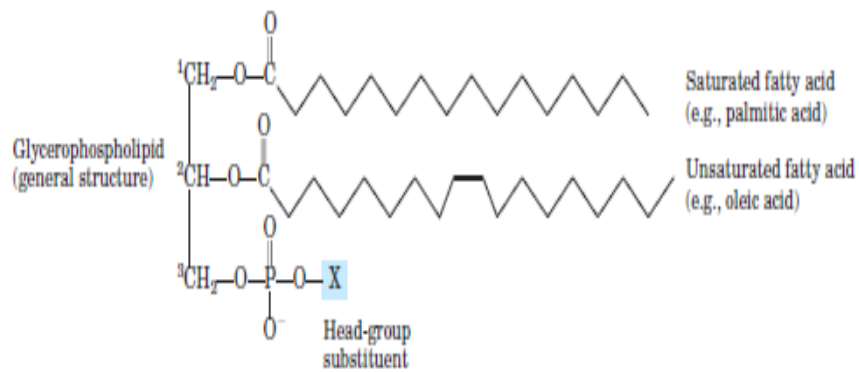


FIGURE 10-7 L-Glycerol 3-phosphate, the backbone of phospholipids. Glycerol itself is not chiral, as it has a plane of symmetry through C-2. However, glycerol can be converted to a chiral compound by adding a substituent such as phosphate to either of the OCH₂OH groups; that is, glycerol is prochiral. One unambiguous nomenclature for glycerol phosphate is the DL system (described on p. 77), in which the isomers are named according to their stereochemical relationships to glyceraldehyde isomers. By this system, the stereoisomer of glycerol phosphate found in most lipids is correctly named either L-glycerol 3-phosphate or D-glycerol 1-phosphate. Another way to specify stereoisomers is the stereospecific numbering (sn) system, in which C-1 is, by definition, that group of the prochiral compound that occupies the pro-S position. The common form of glycerol phosphate in phospholipids is, by this system, sn-glycerol 3-phosphate.

Some animal tissues and some unicellular organisms are rich in ether lipids, in which one of the two acyl chains is attached to glycerol in ether, rather than ester, linkage. The ether-linked chain may be saturated, as in the alkyl ether lipids, or may contain a double bond between C-1 and C-2, as in plasmalogens (Fig. 10-9). Vertebrate heart tissue is uniquely enriched in ether lipids; about half of the heart phospholipids

are plasmalogens.



Name of glycerophospholipid	Name of X	Formula of X	Net charge (at pH 7)
Phosphatidic acid	—	—H	-1
Phosphatidylethanolamine	Ethanolamine	—CH ₂ —CH ₂ —NH ₃ ⁺	0
Phosphatidylcholine	Choline	—CH ₂ —CH ₂ —N ⁺ (CH ₃) ₃	0
Phosphatidylserine	Serine	—CH ₂ —CH(NH ₃ ⁺)—COO ⁻	-1
Phosphatidylglycerol	Glycerol	—CH ₂ —CH(OH)—CH ₂ —OH	-1
Phosphatidylinositol 4,5-bisphosphate	myo-Inositol 4,5-bisphosphate		-4
Cardiolipin	Phosphatidyl-glycerol		-2

FIGURE 10–8 Glycerophospholipids. The common glycerophospholipids are diacylglycerols linked to head-group alcohols through a phosphodiester bond. Phosphatidic acid, a phosphomonoester, is the parent compound. Each derivative is named for the head-group alcohol (X), with the prefix “phosphatidyl-.” In cardiolipin, two phosphatidic acids share a single glycerol.

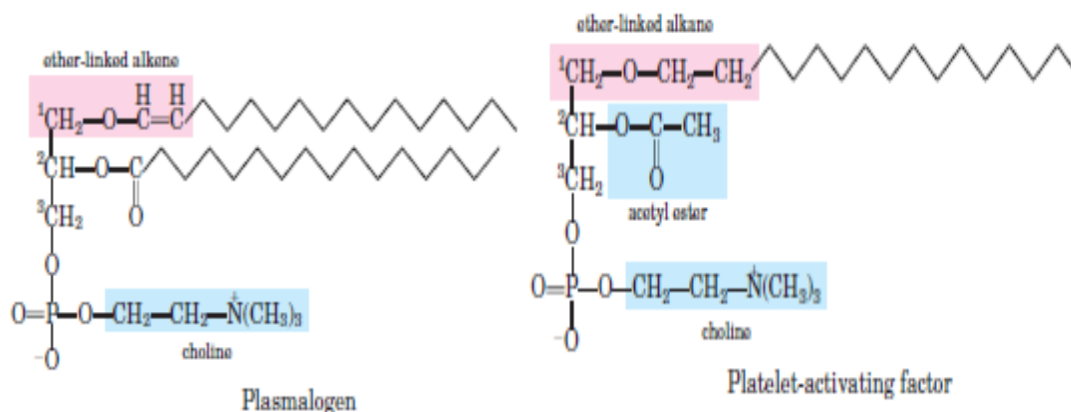


FIGURE 10–9 Ether lipids. Plasmalogens have an ether-linked alkenyl chain where most glycerophospholipids have an ester-linked fatty acid (compare Fig. 10–8). Platelet-activating factor has a long ether-linked alkyl chain at C-1 of glycerol, but C-2 is ester-linked to acetic acid, which makes the compound much more water-soluble than most glycerophospholipids and plasmalogens. The head-group alcohol is choline in plasmalogens and in platelet-activating factor