

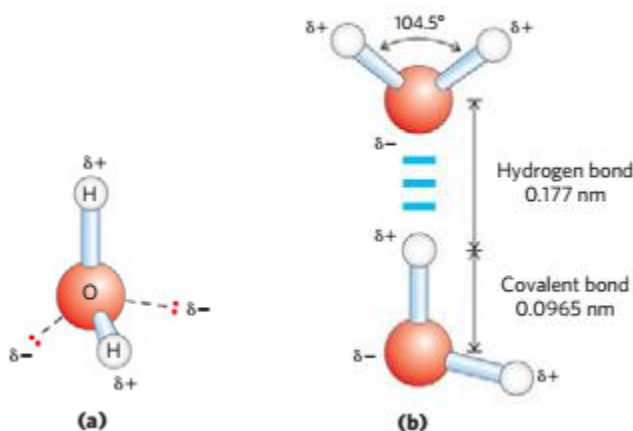
# Water, pH and Acid-base balance

Water is the most abundant substance in living systems, making up 70% or more of the weight of most organisms.

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.

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.

Water has a higher melting point, boiling point, and heat of vaporization than most other common solvents. Each hydrogen atom of a water molecule shares an electron pair with the central oxygen atom. The H–O–H bond angle is 104.5°.



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 ( $\delta^+$ ) and the oxygen atom bears a partial negative charge. As a result, there is an electrostatic attraction between the oxygen atom of one water molecule and the hydrogen of another called a **hydrogen bond**.

Hydrogen bonds account for the relatively high melting point of water, because much thermal energy is required to break a sufficient proportion of hydrogen bonds to destabilize the crystal lattice of ice.

When ice melts or water evaporates, heat is taken up by the system:



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 COH bond is only very weakly polar. Uncharged but polar biomolecules such as sugars dissolve readily in water because of the stabilizing effect of hydrogen bonds. Alcohols, aldehydes, ketones, and compounds containing N–H bonds all form hydrogen bonds with water molecules and tend to be soluble in water.

The attraction between the partial electric charges is greatest when the three atoms involved (in this case O, H, and O) lie in a straight line.

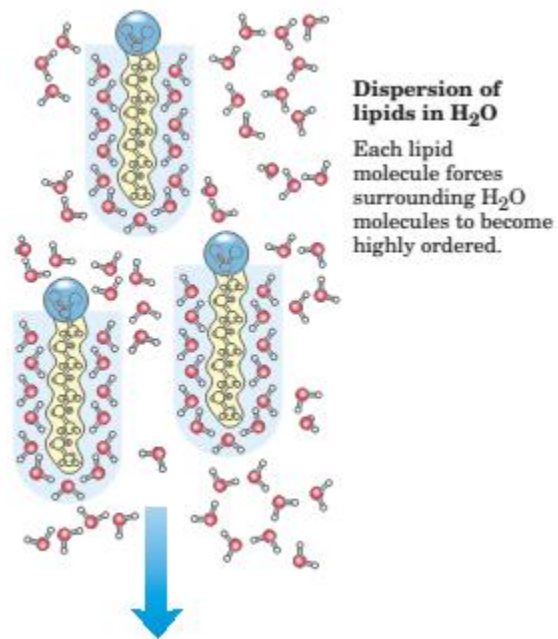
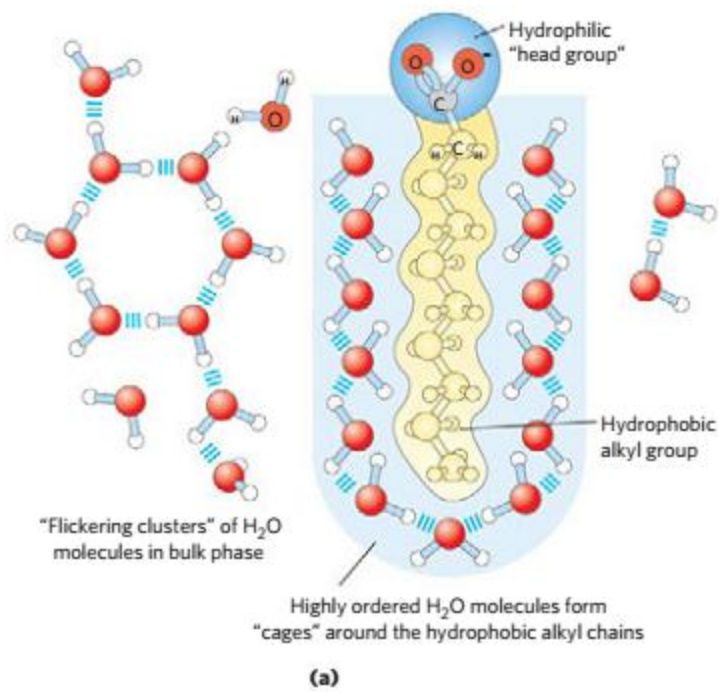
Water is a polar solvent. It readily dissolves most biomolecules, which are generally charged or polar compounds; compounds that dissolve easily in water are **hydrophilic** (Greek, “water-loving”). In contrast, nonpolar solvents such as chloroform and benzene are poor solvents for polar biomolecules but easily dissolve those that are **hydrophobic**—nonpolar molecules such as lipids and waxes.

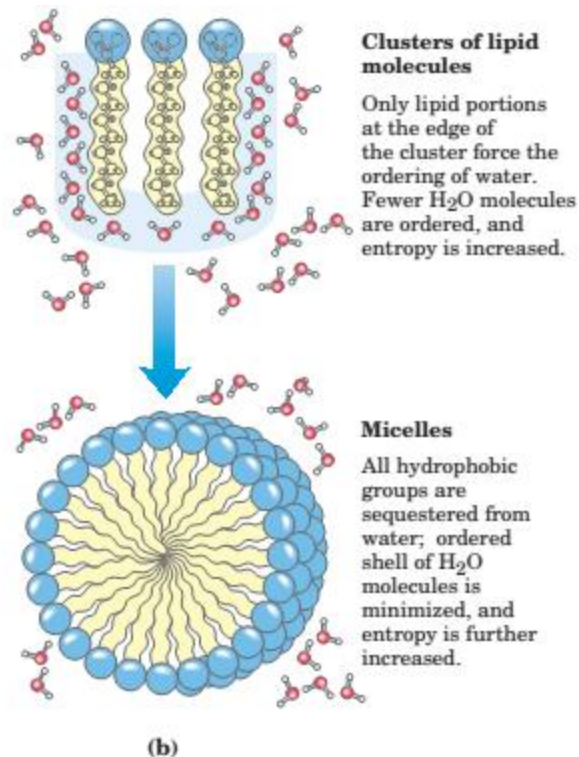
Gases  $\text{CO}_2$ ,  $\text{O}_2$ , and  $\text{N}_2$  are nonpolar. In  $\text{CO}_2$ , each C=O bond is polar, but the two dipoles are oppositely directed and cancel each other. Note that molecules dissolve far better even at low temperatures than do nonpolar molecules at relatively high temperatures.

### **Amphipathic:**

Compounds contain regions that are polar (or charged) and regions that are nonpolar.

Hydrophilic region interacts favorably with the solvent and tends to dissolve, but the nonpolar, hydrophobic region tends to avoid contact with the water. The nonpolar regions of the molecules cluster together. These stable structures of amphipathic compounds in water, called **micelles**.





When two uncharged atoms are brought very close together, electrons around one nucleus may create a transient electric dipole, which induces a transient, opposite electric dipole in the nearby atom. The two dipoles weakly attract each other, bringing the two nuclei closer. **These weak attractions are called van der Waals interactions.** As the two nuclei draw closer together, their electron clouds begin to repel each other. At the point where the van der Waals attraction exactly balances this repulsive force, the nuclei are said to be in van der Waals contact. Each atom has a characteristic **van der Waals radius, a measure of how close that atom will allow another to approach.**

When two atoms are joined covalently, the atomic radii at the point of bonding are less than the van der Waals radii, because the joined atoms are pulled together by the shared electron pair.

Solutes of all kinds alter certain physical properties of the solvent, water: its **vapor pressure, boiling point, melting point (freezing point), and osmotic pressure.** These are called **colligative ("tied together") properties,** because the effect of solutes on all four properties has the same basis: the concentration of water is lower in solutions than in pure water. Independent of the chemical properties of the solute; it depends only on the *number* of solute particles (molecules, ions) in a given amount of water. A compound such as NaCl, which dissociates in solution, has twice the effect on osmotic pressure, for example, as does an equal number of moles of a nondissociating solute such as glucose.

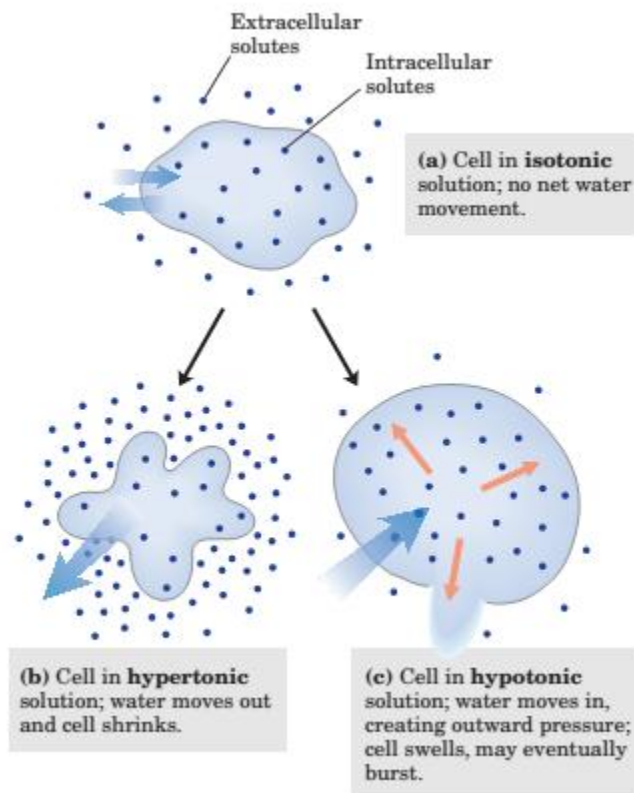
When two different aqueous solutions are separated by a semipermeable membrane (one that allows the passage of water but not solute molecules), water molecules diffusing from the

region of higher water concentration to that of lower water concentration produce osmotic pressure.

**Osmosis** is the water movement across a semipermeable membrane driven by differences in osmotic pressure. Solutions of equal osmolarity are said to be **isotonic**. Surrounded by an isotonic solution, a cell neither gains nor loses water. In a **hypertonic** solution, one with higher osmolarity than the cytosol, the cell shrinks as water flows out. In a **hypotonic** solution, with lower osmolarity than the cytosol, the cell swells as water enters.

The high concentration of albumin and other proteins in blood plasma contributes to its osmolarity.

Because the effect of solutes on osmolarity depends on the *number* of dissolved particles, not their *mass*, macromolecules (proteins, nucleic acids, polysaccharides) have far less effect on the osmolarity of a solution than would an equal mass of their monomeric components. For example, a *gram* of a polysaccharide composed of 1,000 glucose units has the same effect on osmolarity as a *milligram* of glucose. One effect of storing fuel as polysaccharides (starch or glycogen) rather than as glucose or other simple sugars is prevention of an enormous increase in osmotic pressure within the storage cell.



Although many of the solvent properties of water can be explained in terms of the uncharged H<sub>2</sub>O molecule, the small degree of ionization of water to hydrogen ions (H<sup>+</sup>) and hydroxide ions (OH<sup>-</sup>) must also be taken into account.

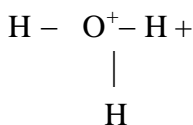
When weak acids are dissolved in water, they contribute H<sup>+</sup> by ionizing; weak bases consume H<sup>+</sup> by becoming protonated.

The total hydrogen ion concentration from all sources is experimentally measurable and is expressed as the pH of the solution.

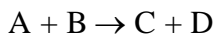
Water molecules have a slight tendency to undergo reversible ionization to yield a hydrogen ion (a proton) and a hydroxide ion, giving the equilibrium



Free protons do not exist in solution; hydrogen ions formed in water are immediately hydrated to **hydronium ions** (H<sub>3</sub>O<sup>+</sup>)



The position of equilibrium of any chemical reaction is given by its **equilibrium constant, K<sub>eq</sub>** (sometimes expressed simply as *K*). For the generalized reaction.



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

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

Equilibrium constants are dimensionless, 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.

The degree of ionization of water at equilibrium is small; at 25 °C only about two of every 10<sup>9</sup> molecules in pure water are ionized at any instant. The equilibrium constant for the reversible ionization of water is

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

In pure water at 25 °C, the concentration of water is 55.5 M –[grams of H<sub>2</sub>O in 1 L divided by its gram molecular weight: (1,000 g/L)/(18.015 g/mol)]– and is essentially constant in relation to the very low concentrations of H<sup>+</sup> and OH<sup>-</sup>, namely, 1x10<sup>-7</sup> M. Accordingly, we can substitute 55.5 M in the equilibrium constant expression to yield

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

On rearranging, this becomes

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

Where *K<sub>w</sub>* designates the product (55.5M) (*K<sub>eq</sub>*), the ion product of water at 25 °C.

The value for *K<sub>eq</sub>*, determined by electrical conductivity measurements of pure water is 18 × 10<sup>-16</sup> M at 25 °C.

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

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  $\text{H}^+$  and  $\text{OH}^-$ , as in pure water, the solution is said to be at neutral pH. At this pH, the concentration of  $\text{H}^+$  and  $\text{OH}^-$  can be calculated from the ion product of water as follows:

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

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

$$[\text{H}^+] = \sqrt{K_w} = \sqrt{1 \times 10^{-14} \text{ M}^2}$$

$$[\text{H}^+] = [\text{OH}^-] = 10^{-7} \text{ M}$$

The ion product of water,  $K_w$ , is the basis for the pH scale. 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:

$$\text{pH} = \log \frac{1}{1.0 \times 10^{-7}} = 7.0$$

Note that the concentration of  $\text{H}^+$  must be expressed in molar (M) terms.

Solutions having a pH greater than 7 are alkaline or basic; the concentration of  $\text{OH}^-$  is greater than that of  $\text{H}^+$ . Conversely, solutions having a pH less than 7 are acidic.

Keep in mind 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  $\text{H}^+$  concentration of the other, but it does not tell us the absolute magnitude of the difference.

The pH of an aqueous solution can be approximately measured using various indicator dyes, including litmus, phenolphthalein, and phenol red.

## The pH Scale

$[H^+]$ (M)	pH	$[OH^-]$ (M)	pOH*
$10^0$ (1)	0	$10^{-14}$	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  $pOH = -\log[OH^-]$ , which is analogous to the expression for pH. Note that in all cases,  $pH + pOH = 14$ .

Measurement of pH is one of the most important and frequently used procedures in biochemistry. The pH affects the structure and activity of biological macromolecules; for example, the catalytic activity of enzymes is strongly dependent on pH. The pH of the blood plasma of people with severe, uncontrolled diabetes, for example, is often below the normal value of 7.4; this condition is called acidosis. In certain other diseases the pH of the blood is higher than normal, a condition known as alkalosis. Extreme acidosis or alkalosis can be life-threatening.

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.

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. 2-16). Acetic acid ( $CH_3COOH$ ), a proton donor, and the acetate anion ( $CH_3COO^-$ ), the corresponding proton acceptor, constitute a conjugate acidbase 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 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.



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



Called ionization or **dissociation constants**, often designated  $K_a$ . Stronger acids, such as phosphoric and carbonic acids, have larger dissociation constants; weaker acids, such as monohydrogen phosphate ( $\text{HPO}_4^{2-}$ ), have smaller dissociation constants. 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.

Titration is used to determine the amount of an acid in a given solution. A measured volume of the acid is titrated with a solution of a strong base, usually sodium hydroxide (NaOH), of known concentration. The NaOH is added in small increments until the acid is consumed (neutralized), as determined with an indicator dye or a pH meter. The concentration of the acid in the original solution can be calculated from the volume and concentration of NaOH added.

A plot of pH against the amount of NaOH added (a **titration curve**) reveals the pKa of the weak acid.

The titration curve of a weak acid shows graphically that a weak acid and its anion –a conjugate acid-base pair–can act as a buffer.

Almost every biological process is pH dependent; a small change in pH produces a large change in the rate of the process.

Cells and organisms maintain a specific and constant cytosolic pH, keeping biomolecules in their optimal ionic state, usually near pH 7. In multicellular organisms, the pH of extracellular fluids is also tightly regulated. Constancy of pH is achieved primarily by biological buffers: mixtures of weak acids and their conjugate bases.

**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). As an example, a mixture of equal concentrations of acetic acid and acetate ion, is a buffer system. The titration curve of acetic acid has a relatively flat zone extending about 1 pH unit on either side of its midpoint pH of 4.76. In this zone, an amount of H or OH added to the system has much less effect on pH than the same amount added outside the buffer range. This relatively flat zone is the buffering region of the acetic acid–acetate buffer pair. The buffering power of the system is maximal; the pH at this point in the titration curve of acetic acid is equal to its pKa. The pH of the acetate buffer system does change slightly when a small amount of H or OH is added, but this change is very small compared with the pH change that would result if the same amount of H or OH were added to pure water or to a solution of the salt of a strong acid and strong base, such as NaCl, which has no buffering power.

Buffering results from two reversible reaction equilibria occurring in a solution of nearly equal concentrations of a proton donor and its conjugate proton acceptor.

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.

Each conjugate acid-base pair has a characteristic pH zone in which it is an effective buffer. The  $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$  pair has a  $\text{pK}_a$  of 6.86 and thus can serve as an effective buffer system between approximately pH 5.9 and pH 7.9; the  $\text{NH}_4^+/\text{NH}_3$  pair, with a  $\text{pK}_a$  of 9.25, can act as a buffer between approximately pH 8.3 and pH 10.3.

The shape of the titration curve of any weak acid is described by the **Henderson-Hasselbalch equation**.

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

The Henderson-Hasselbalch equation also allows us to (1) calculate  $\text{pK}_a$ , given pH and the molar ratio of proton donor and acceptor; (2) calculate pH, given  $\text{pK}_a$  and the molar ratio of proton donor and acceptor; and (3) calculate the molar ratio of proton donor and acceptor, given pH and  $\text{pK}_a$ .

Two especially important biological buffers are the phosphate and bicarbonate systems. The phosphate buffer system, which acts in the cytoplasm of all cells, consists of  $\text{H}_2\text{PO}_4^-$  as proton donor and  $\text{HPO}_4^{2-}$  as proton acceptor:



The phosphate buffer system is maximally effective at a pH close to its  $\text{pK}_a$  of 6.86 and thus tends to resist pH changes in the range between about 5.9 and 7.9.

Blood plasma is buffered in part by the bicarbonate system, consisting of carbonic acid ( $\text{H}_2\text{CO}_3$ ) as proton donor and bicarbonate ( $\text{HCO}_3^-$ ) as proton acceptor.

The bicarbonate buffer system is an effective physiological buffer near pH 7.4, because the  $\text{H}_2\text{CO}_3$  of blood plasma is in equilibrium with a large reserve capacity of  $\text{CO}_2(\text{g})$  in the air space of the lungs.

Human blood plasma normally has a pH between 7.35 and 7.45, and many of the enzymes that function in the blood have evolved to have maximal activity in that pH range. Enzymes typically show maximal catalytic activity at a characteristic pH, called the pH optimum. On either side of this optimum pH, catalytic activity often declines sharply.

In individuals with untreated diabetes mellitus, the lack of insulin, or insensitivity to insulin, disrupts the uptake of glucose from blood into the tissues and forces the tissues to use stored fatty acids as their primary fuel. This dependence on fatty acids results in the accumulation of high concentrations of two carboxylic acids,  $\beta$ -hydroxybutyric acid and acetoacetic acid. Dissociation of these acids lowers the pH of blood plasma to less than 7.45, causing acidosis.

Water is not just the solvent in which the chemical reactions of living cells occur; it is very often a direct participant in those reactions. The formation of ATP from ADP and inorganic phosphate is an example of a **condensation reaction** in which the elements of water are eliminated. The reverse of this reaction—cleavage accompanied by the addition of the elements of water—is a **hydrolysis reaction**. Hydrolysis reactions are also responsible for the enzymatic depolymerization of proteins, carbohydrates, and nucleic acids.

# Amino Acids, Peptides, and Proteins

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.) Twenty different amino acids are commonly found in proteins.

All 20 of the common amino acids are alpha  $\alpha$ -amino acids. They have a carboxyl group and an amino group bonded to the same carbon atom (the alpha  $\alpha$  carbon). They differ from each other in their side chains, or **R groups**. In addition to these 20 amino acids there are many less common ones.

The three-letter code is transparent, the abbreviations generally consisting of the first three letters of the amino acid name. For six amino acids (CHIMSV), the first letter of the amino acid name is unique and thus is used as the symbol. For others (AGLPT), the first letter is not unique but is assigned to the amino acid that is most common in proteins (for example, leucine is more common than lysine). For another four, the letter used is phonetically suggestive (RFYW: aRginine, Fenyl-alanine, tYrosine, tWiPe-tophan). The rest were harder to assign. Four (DNEQ) were assigned letters found within or suggested by their names (asparDic, asparagiNe, glutamEke, Q-tamine). That left lysine. Only a few letters were left in the alphabet, and K was chosen because it was the closest to L.

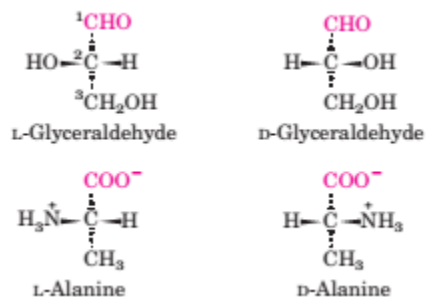
Amino Acid	Abbreviation / Symbol
<b>Nonpolar, aliphatic R groups</b>	
Glycine	Gly G
Alanine	Ala A
Proline	Pro P
Valine	Val V
Leucine	Leu L
Isoleucine	Ile I
Methionine	Met M
<b>Aromatic R groups</b>	
Phenylalanine	Phe F
Tyrosine	Tyr Y
Tryptophan	Trp W
<b>Polar, uncharged R groups</b>	
Serine	Ser S
Threonine	Thr T
Cysteine	Cys C

Asparagine	Asn N
Glutamine	Gln Q
<b>Positively charged R groups</b>	
Lysine	Lys K
Histidine	His H
Arginine	Arg R
<b>Negatively charged R groups</b>	
Aspartate	Asp D
Glutamate	Glu E

For all the common amino acids except glycine, the carbon is bonded to four different groups: a carboxyl group, an amino group, an R group, and a hydrogen atom. The  $\alpha$ -carbon atom is thus a **chiral center**. Thus amino acids have two possible stereoisomers. Since they are nonsuperposable mirror images of each other, the two forms represent a class of stereoisomers called **enantiomers**. All molecules with a chiral center are also **optically active**—that is, they rotate plane-polarized light.

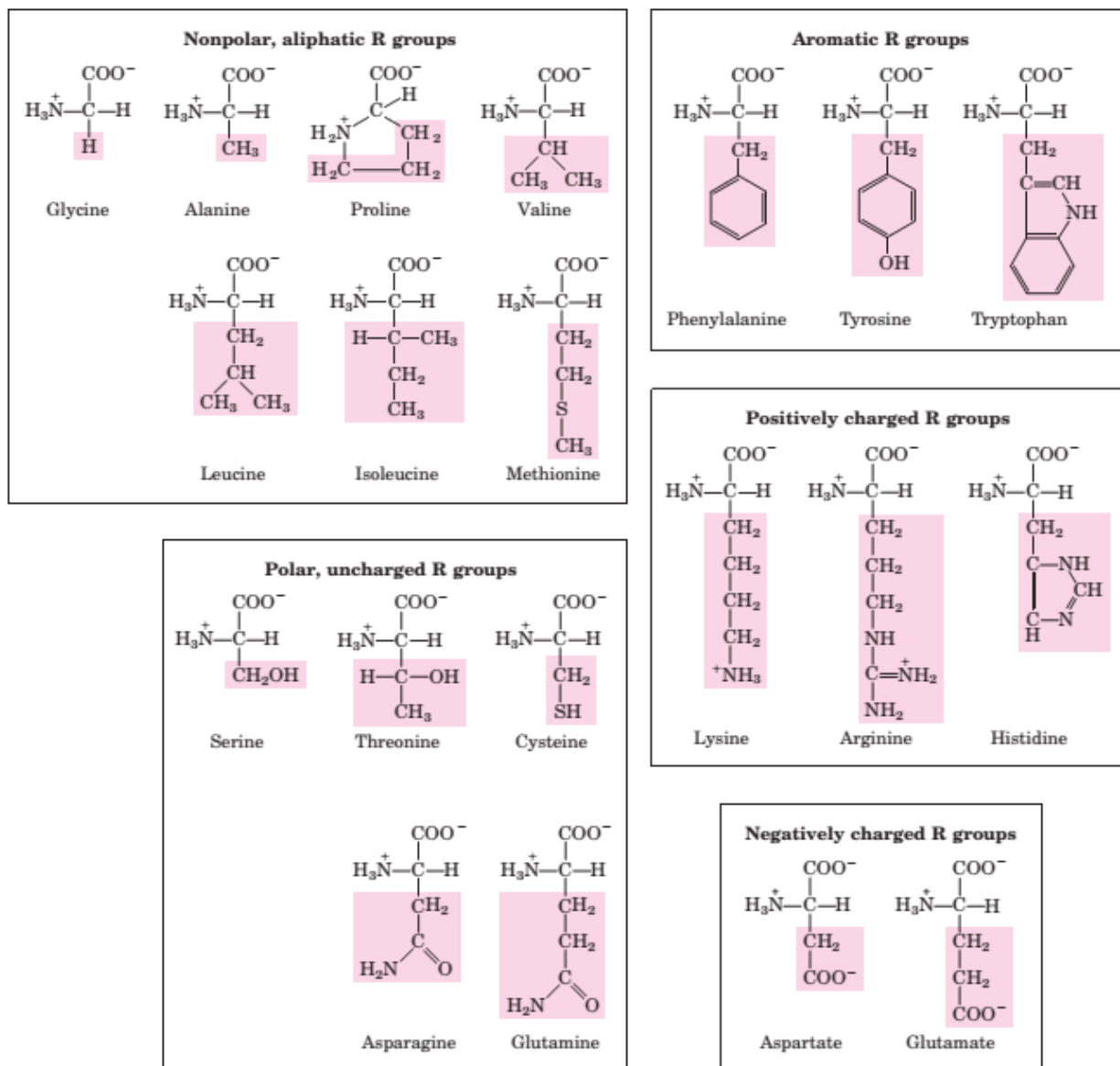
The additional carbons in an R group are commonly designated  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and so forth, proceeding out from the  $\alpha$  carbon. For most other organic molecules, carbon atoms are simply numbered from one end, giving highest priority (C-1) to the carbon with the substituent containing the atom of highest atomic number. Within this latter convention, the carboxyl carbon of an amino acid would be C-1 and the  $\alpha$  carbon would be C-2.

Special nomenclature has been developed to specify the **absolute configuration** of the four substituents of asymmetric carbon atoms. The absolute configurations of simple sugars and amino acids are specified by the **D, L system**, based on the absolute configuration of the three-carbon sugar glyceraldehyde, a convention proposed by Emil Fischer in 1891. For all chiral compounds, stereoisomers having a configuration related to that of L-glyceraldehyde are designated L, and stereoisomers related to D-glyceraldehyde are designated D. L and D refer *only* to the absolute configuration of the four substituents around the chiral carbon, not to optical properties of the molecule.



Nearly all biological compounds with a chiral center occur naturally in only one stereoisomeric form, either D or L. The amino acid residues in protein molecules are exclusively L stereoisomers. D-Amino acid residues have been found only in a few, generally small peptides, including some peptides of bacterial cell walls and certain peptide antibiotics.

Grouping the amino acids 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).



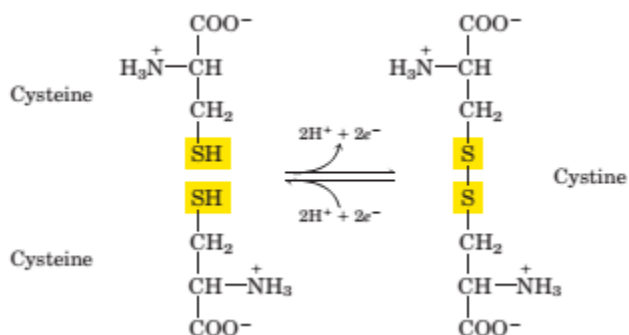
## NON POLAR AMINO ACIDS:

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.

**Aromatic R Groups:** **Phenylalanine**, **tyrosine**, and **tryptophan**, with their aromatic side chains, are relatively nonpolar (hydrophobic). Tyrosine and tryptophan are significantly more polar than phenylalanine, because of the tyrosine hydroxyl group and the nitrogen of the tryptophan indole ring.

**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**.

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. 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.

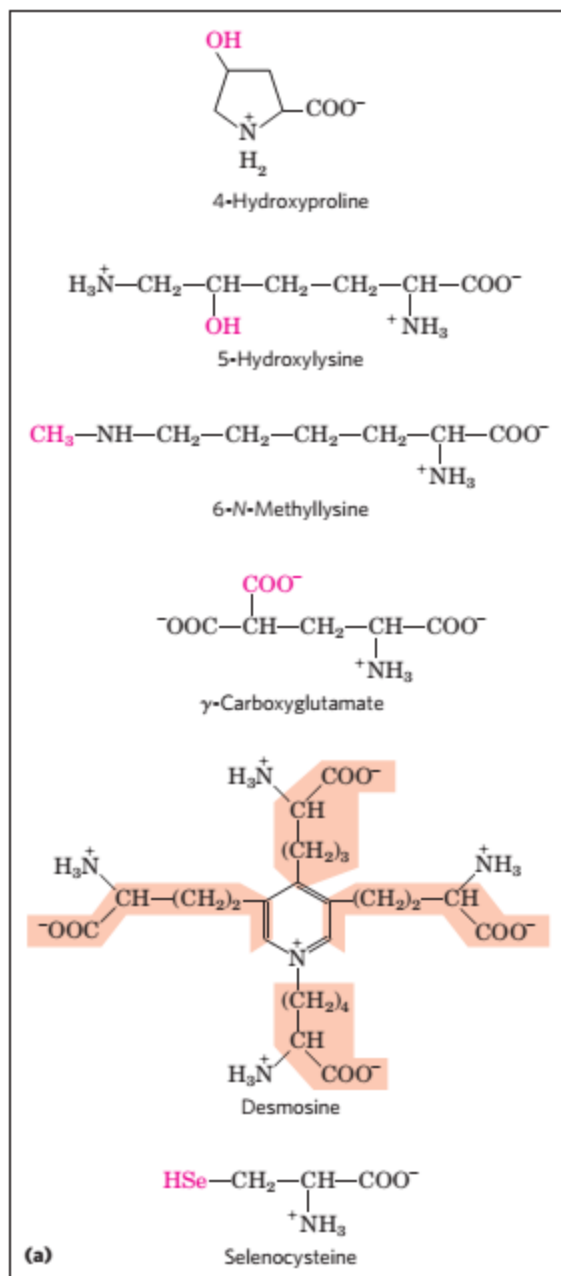


**Positively Charged (Basic) R Groups:** The most hydrophilic R groups are those that are either positively or negatively charged. The amino acids in which the R groups have significant positive charge at pH 7.0 are **lysine**, which has a second primary amino group at the position on its aliphatic chain;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

**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.

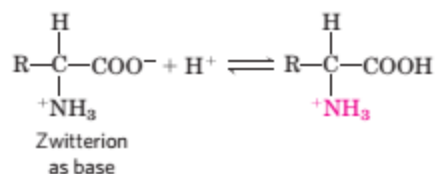
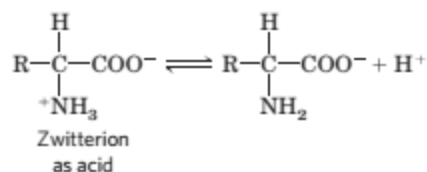
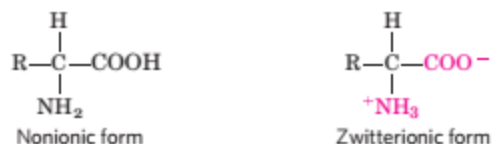
In addition to the 20 common amino acids, proteins may contain residues created by modification of common residues already incorporated into a polypeptide. Among these uncommon amino acids are **4-hydroxyproline**, a derivative of proline, and **5-hydroxylysine**, derived from lysine. The former is found in plant cell wall proteins, and both are found in collagen, a fibrous protein of connective tissues. **6-NMethyllysine** is a constituent of myosin, a contractile protein of muscle. Another important uncommon amino acid is **carboxyglutamate**, found in the bloodclotting protein prothrombin and in certain other proteins that bind  $\text{Ca}^{2+}$  as part of their biological function. More complex is **desmosine**, a derivative of four Lys residues, which is found in the fibrous protein elastin. **Selenocysteine** is a special case.

Ornithine and citrulline deserve special note because they are key intermediates (metabolites) in the biosynthesis of arginine and in the urea cycle.

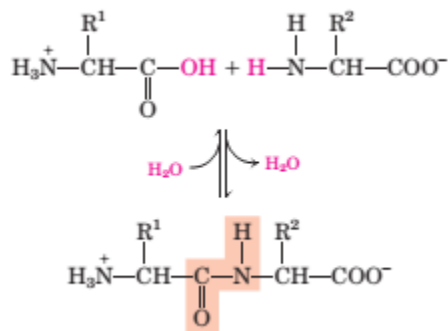


The amino and carboxyl groups of amino acids, along with the ionizable R groups of some amino acids, function as weak acids and bases. When an amino acid lacking an ionizable R groups is dissolved in water at neutral pH, it exists in solution as the dipolar ion, or **zwitterion** (German for “hybrid ion”), which can act as either an acid or a base. Substances having this dual (acid-base) nature are amphoteric and are often called ampholytes (from “amphoteric electrolytes”).





The characteristic pH at which the net electric charge is zero is called the **isoelectric point** or **isoelectric pH**, designated **pI**. 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 -carboxyl group of one amino acid and the -amino group of another. Peptide bond formation is an example of a condensation reaction.



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.

In a peptide, the amino acid residue at the end with a free -amino group is the **amino-terminal** (or *N*-terminal) residue; the residue at the other end, which has a free carboxyl group, is the **carboxyl-terminal** (*C*-terminal) residue.

When an amino acid sequence of a peptide, polypeptide, or protein is displayed, the aminoterminal end is placed on the left, the carboxyl-terminal end on the right. The sequence is read left to right, beginning with the amino-terminal end.

Peptides contain only one free -amino group and one free -carboxyl group, at opposite ends of the chain. These groups ionize as they do in free amino acids, although the ionization constants are different because an oppositely charged group is no longer linked to the carbon. The -amino and -carboxyl groups of all nonterminal amino acids are covalently joined in the peptide bonds, which do not ionize and thus do not contribute to the total acid-base behavior of peptides. However, the R groups of some amino acids can ionize, and in a peptide these contribute to the overall acid-base properties of the molecule.

Some proteins consist of a single polypeptide chain, but others, called **multisubunit** proteins, have two or more polypeptides associated noncovalently. The individual polypeptide chains in a multisubunit protein may be identical or different. If at least two are identical the protein is said to be **oligomeric**, and the identical units (consisting of one or more polypeptide chains) are referred to as **protomers**. Hemoglobin, for example, has four polypeptide subunits: two identical chains and two identical chains, all four held together by noncovalent interactions. Each subunit is paired in an identical way with a subunit within the structure of this multisubunit protein, so that hemoglobin can be considered either a tetramer of four polypeptide subunits or a dimer of  $\alpha\beta$  protomers.

A few proteins contain two or more polypeptide chains linked covalently. For example, the two polypeptide chains of insulin are linked by disulfide bonds. In such cases, the individual polypeptides are not considered subunits but are commonly referred to simply as chains.

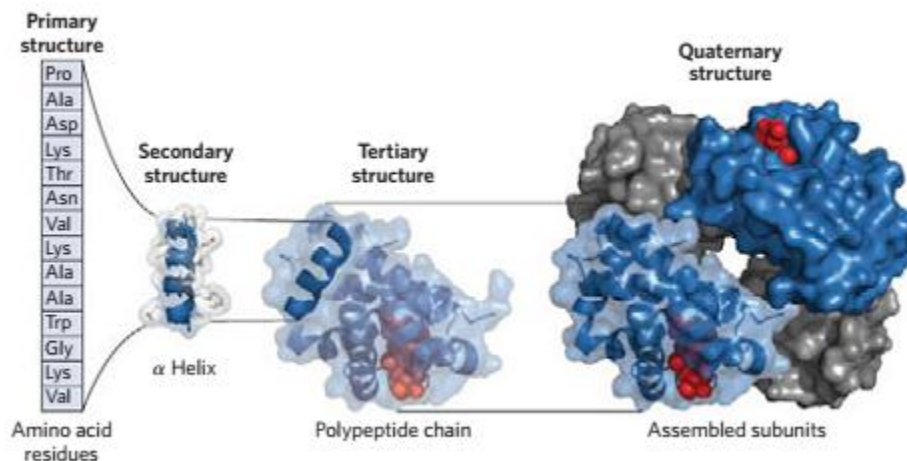
Many proteins, for example the enzymes ribonuclease A and chymotrypsinogen, contain only amino acid residues and no other chemical constituents; these are considered simple proteins. However, some proteins contain permanently associated chemical components in addition to amino acids; these are called **conjugated proteins**. The non-amino acid part of a conjugated protein is usually called its **prosthetic group**. Conjugated proteins are classified on the basis of the chemical nature of their prosthetic groups; for example, **lipoproteins** contain **lipids**, **glycoproteins** contain sugar groups, and **metalloproteins** contain a specific metal.

## Conjugated Proteins

Class	Prosthetic group	Example
Lipoproteins	Lipids	B1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin

Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

Four levels of protein structure are commonly defined. 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. The primary structure of a protein determines how it folds up into its unique three-dimensional structure, and this in turn determines the function of the protein.



# The Three-Dimensional Structure of Proteins

Proteins are big molecules. The covalent backbone of a typical protein contains hundreds of individual bonds. Because free rotation is possible around many of these bonds, the protein can assume an unlimited number of conformations. However, each protein has a specific chemical or structural function, strongly suggesting that each has a unique three-dimensional structure.

The spatial arrangement of atoms in a protein is called its **conformation**. The possible conformations of a protein include any structural state that can be achieved without breaking covalent bonds. A change in conformation could occur, for example, by rotation about single bonds.

Proteins in any of their functional, folded conformations are called **native** proteins.

In the context of protein structure, the term **stability** can be defined as the tendency to maintain a native conformation.

The chemical interactions that counteract these effects and stabilize the native conformation include disulfide bonds and the weak (noncovalent) interactions hydrogen bonds, and hydrophobic and ionic interactions.

About 200 to 460 kJ/mol are required to break a single covalent bond, whereas weak interactions can be disrupted by a mere 4 to 30 kJ/mol. Individual covalent bonds that contribute to the native conformations of proteins, such as disulfide bonds linking separate parts of a single polypeptide chain, are clearly much stronger than individual weak interactions. Yet, because they are so numerous, it is weak interactions that predominate as a stabilizing force in protein structure.

On carefully examining the contribution of weak interactions to protein stability, we find that hydrophobic interactions generally predominate. Hydrophobic amino acid side chains therefore tend to cluster in a protein's interior, away from water, forming a hydrophobic protein core.

Hydrogen bonds between groups in a protein form cooperatively (formation of one makes the next one more likely) in repeating secondary structures that optimize hydrogen bonding

The interaction of oppositely charged groups that form an ion pair, or salt bridge, can have either a stabilizing or destabilizing effect on protein structure. Ionic interaction also limit structural flexibility and confer a uniqueness to a particular protein structure.

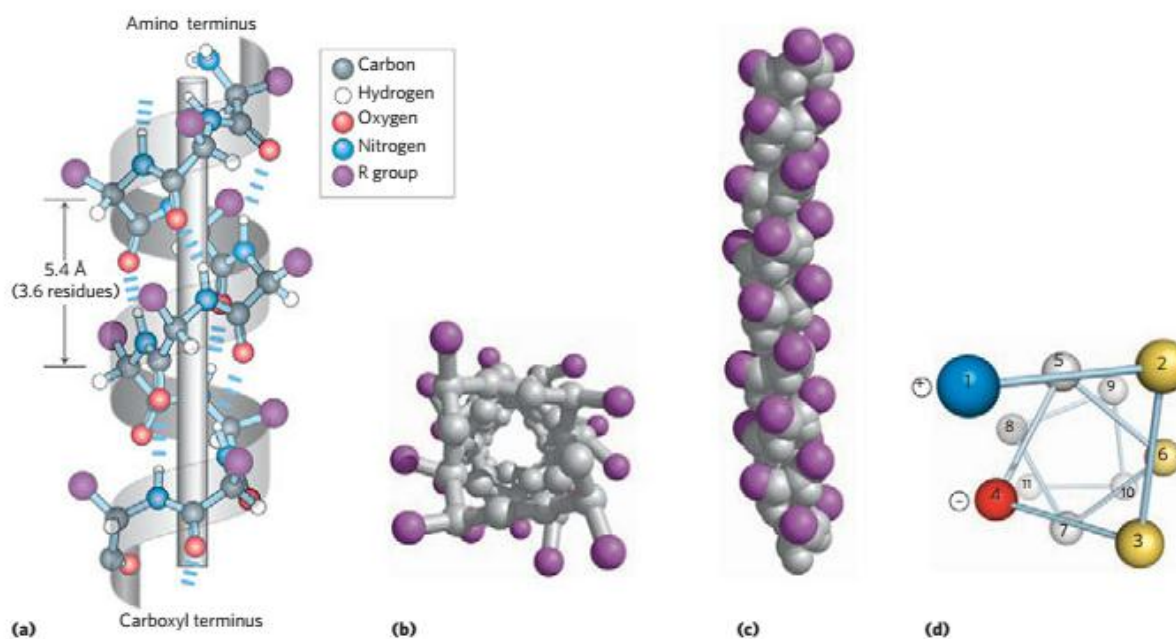
In the tightly packed atomic environment of a protein, one more type of weak interaction can have a significant effect—van der Waals interactions. Van der Waals interactions are dipole-dipole interactions involving the permanent electric dipoles in groups such as carbonyls.

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. The peptide C-N bonds are unable to rotate freely. The rigid peptide bonds limit the range of conformations that can be assumed by a polypeptide chain.

The term **secondary structure** refers to any chosen segment of a polypeptide chain and describes the local spatial arrangement of its main-chain atoms, without regard to the positioning of its side chains or its relationship to other segments. The most prominent are the  $\alpha$  helix and  $\beta$  conformations; another common type is the  $\beta$  turn.

### The $\alpha$ helix:

It is a helical structure, the  $\alpha$  helix. In this structure, the polypeptide backbone is tightly wound around an imaginary axis drawn longitudinally through the middle of the helix, and the R groups of the amino acid residues protrude outward from the helical backbone. The repeating unit is a single turn of the helix, which extends about 5.4 Å along the long axis. The right-handed  $\alpha$  helix is the common form. Extended left-handed  $\alpha$  helices are theoretically less stable and have not been observed in proteins. The  $\alpha$  helix proved to be the predominant structure in  $\alpha$ -keratins. The structure is stabilized by a hydrogen bond between the hydrogen atom attached to the electronegative nitrogen atom of a peptide linkage and the electronegative carbonyl oxygen atom of the fourth amino acid on the amino-terminal side of that peptide bond.



Not all polypeptides can form a stable  $\alpha$  helix. Each amino acid residue in a polypeptide has an intrinsic propensity to form an  $\alpha$  helix, reflecting the properties of the R group. Alanine shows the greatest tendency to form  $\alpha$  helices in most experimental model systems.

The bulk and shape of Asn, Ser, Thr, and Cys residues can also destabilize an  $\alpha$  helix if they are close together in the chain.

A constraint on the formation of the  $\alpha$  helix is the presence of Pro or Gly residues. In proline, the nitrogen atom is part of a rigid ring. Thus, a Pro residue introduces a destabilizing kink in an  $\alpha$  helix. In addition, the nitrogen atom of a Pro residue in peptide linkage has no

substituent hydrogen to participate in hydrogen bonds with other residues. For these reasons, proline is only rarely found within an  $\alpha$  helix. Glycine occurs infrequently in  $\alpha$  helices for a different reason: it has more conformational flexibility than the other amino acid residues. Polymers of glycine tend to take up coiled structures quite different from an  $\alpha$  helix.

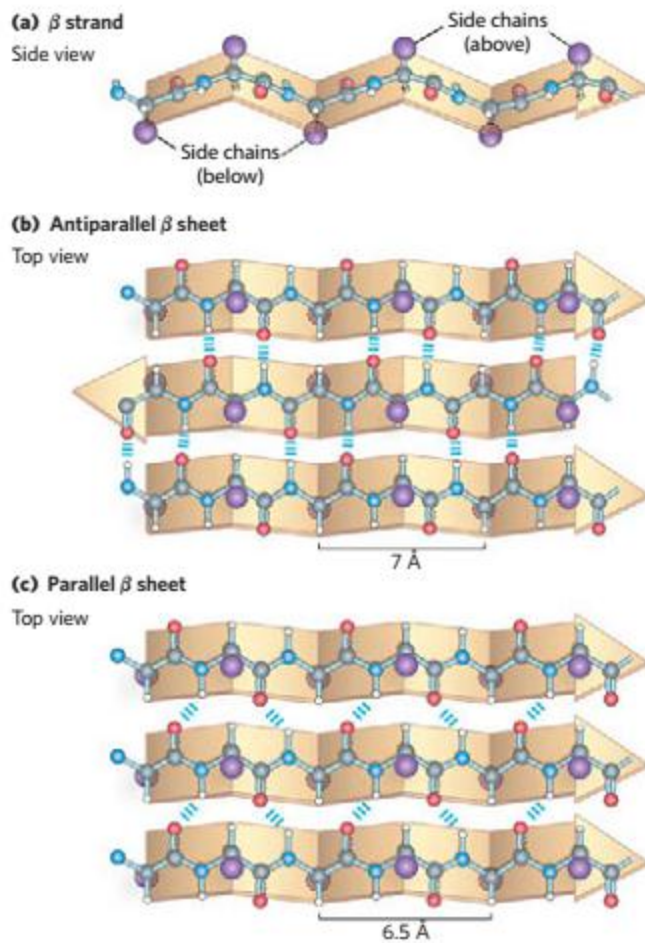
Negatively charged amino acids are often found near the amino terminus of the helical segment, where they have a stabilizing interaction with the positive charge of the helix dipole; a positively charged amino acid at the amino-terminal end is destabilizing. The opposite is true at the carboxyl-terminal end of the helical segment.

Thus, five different kinds of constraints affect the stability of an  $\alpha$  helix: (1) the electrostatic repulsion (or attraction) between successive amino acid residues with charged R groups, (2) the bulkiness of adjacent R groups, (3) the interactions between R groups spaced three (or four) residues apart, (4) the occurrence of Pro and Gly residues, and (5) the interaction between amino acid residues at the ends of the helical segment and the electric dipole inherent to the  $\alpha$  helix.

### **The $\beta$ sheet:**

Conformation, the backbone of the polypeptide chain is extended into a zigzag rather than helical structure. The arrangement of several segments side by side, all of which are in the  $\beta$  conformation, is called a  **$\beta$  sheet**. The zigzag structure of the individual polypeptide segments gives rise to a pleated appearance of the overall sheet. Hydrogen bonds form between adjacent segments of polypeptide chain within the sheet.

The adjacent polypeptide chains in a  $\beta$  sheet can be either parallel or antiparallel (having the same or opposite amino-to-carboxyl orientations, respectively).



### $\beta$ turns:

Some amino acid residues are in turns or loops where the polypeptide chain reverses direction. These are the connecting elements that link successive runs of  $\alpha$  helix or  $\beta$  conformation. Particularly common are  **$\beta$  turns** that connect the ends of two adjacent segments of an antiparallel  $\beta$  sheet. The structure is a  $180^\circ$  turn involving four amino acid residues, with the carbonyl oxygen of the first residue forming a hydrogen bond with the amino-group hydrogen of the fourth. Gly and Pro residues often occur in  **$\beta$  turns**.

The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's **tertiary structure**. Whereas the term “secondary structure” refers to the spatial arrangement of amino acid residues that are adjacent in the primary structure, tertiary structure includes *longer-range* aspects of amino acid sequence.

Some proteins contain two or more separate polypeptide chains, or subunits, which may be identical or different. The arrangement of these protein subunits in three-dimensional complexes constitutes **quaternary structure**.

In considering these higher levels of structure, it is useful to classify proteins into two major groups: **fibrous proteins**, having polypeptide chains arranged in long strands or sheets,

and **globular proteins**, having polypeptide chains folded into a spherical or globular shape. The two groups are structurally distinct: fibrous proteins usually consist largely of a single type of secondary structure; globular proteins often contain several types of secondary structure. The two groups differ functionally in that the structures that provide support, shape, and external protection to vertebrates are made of fibrous proteins, whereas most enzymes and regulatory proteins are globular proteins.

Fibrous proteins share properties that give strength and/or flexibility to the structures in which they occur. All fibrous proteins are insoluble in water, a property conferred by a high concentration of hydrophobic amino acid residues both in the interior of the protein and on its surface.

The -keratins have evolved for strength. Found in mammals, these proteins constitute almost the entire dry weight of hair, wool, nails, claws, quills, horns, hooves, and much of the outer layer of skin.  $\alpha$ -keratin is rich in the hydrophobic residues Ala, Val, Leu, Ile, Met, and Phe.

In  $\alpha$ -keratins, the cross-links stabilizing quaternary structure are disulfide bonds.

Like the -keratins, collagen has evolved to provide strength. It is found in connective tissue such as tendons, cartilage, the organic matrix of bone, and the cornea of the eye. The collagen helix is a unique secondary structure quite distinct from the  $\alpha$  helix. It is left-handed and has three amino acid residues per turn. Three separate polypeptides, called  $\alpha$  chains (not to be confused with  $\alpha$  helices), are supertwisted about each other.

The food product gelatin is derived from collagen; it has little nutritional value as a protein, because collagen is extremely low in many amino acids that are essential in the human diet.

The amino acid sequence in collagen is generally a repeating tripeptide unit, Gly-X-Y, where X is often Pro, and Y is often 4-Hyp. Only Gly residues can be accommodated at the very tight junctions between the individual chains; The Pro and 4-Hyp residues permit the sharp twisting of the collagen helix.

Scurvy is caused by lack of vitamin C, or ascorbic acid (ascorbate). Vitamin C is required for, among other things, the hydroxylation of proline and lysine in collagen; scurvy is a deficiency disease characterized by general degeneration of connective tissue. Manifestations of advanced scurvy include numerous small hemorrhages caused by fragile blood vessels, tooth loss, poor wound healing and the reopening of old wounds, bone pain and degeneration, and eventually heart failure. Despondency and oversensitivity to stimuli of many kinds are also observed. Milder cases of vitamin C deficiency are accompanied by fatigue, irritability, and an increased severity of respiratory tract infections. Most animals make large amounts of vitamin C, converting glucose to ascorbate in four enzymatic steps. But in the course of evolution, humans and some other animals—gorillas, guinea pigs, and fruit bats—have lost the last enzyme in this pathway and must obtain ascorbate in their diet. Vitamin C is available in a wide range of fruits and vegetables.

Collagen is constructed of the repeating tripeptide unit Gly-X-Y, where X and Y are generally Pro or 4-Hyp—the proline derivative (4T)-L-hydroxyproline, which plays an essential



role in the folding of collagen and in maintaining its structure. In the absence of vitamin C, cells cannot hydroxylate the Pro at the Y positions. This leads to collagen instability and the connective tissue problems seen in scurvy.

### **Globular Protein Structure:**

In a globular protein, different segments of the polypeptide chain (or multiple polypeptide chains) fold back on each other, generating a more compact shape than is seen in the fibrous proteins. Globular proteins include enzymes, transport proteins, motor proteins, regulatory proteins, immunoglobulins, and proteins with many other functions.

To understand a complete three-dimensional structure, we need to analyze its folding patterns. We begin by defining two important terms that describe protein structural patterns or elements in a polypeptide chain.

The first term is motif, also called a fold or (more rarely) supersecondary structure. A *motif or fold is a recognizable folding pattern involving two or more elements of secondary structure and the connection (s) between them.* A motif can be very simple, such as two elements of secondary structure folded against each other, and represent only a small part of a protein. An example is a  $\beta$ - $\alpha$ - $\beta$  loop. The segment defined as a motif or fold may or may not be independently stable. The distinctive arrangement of eight  $\alpha$  helices in myoglobin is replicated in all globins and is called the globin fold. Note that a motif is not a hierarchical structural element falling between secondary and tertiary structure. It is simply a folding pattern.

The second term for describing structural patterns is domain. A domain is a part of a polypeptide chain that is independently stable or could undergo movements as a single entity with respect to the entire protein. In many cases, a domain from a large protein will retain its native three-dimensional structure even when separated (for example, by proteolytic cleavage) from the remainder of the polypeptide chain. Different domains often have distinct functions.

Proteins can be denatured not only by heat but by extremes of pH, by certain miscible organic solvents such as alcohol or acetone, by certain solutes such as urea and guanidine hydrochloride, or by detergents. Each of these denaturing agents represents a relatively mild treatment in the sense that no covalent bonds in the polypeptide chain are broken. Organic solvents, urea, and detergents act primarily by disrupting the hydrophobic interactions that make up the stable core of globular proteins; extremes of pH alter the net charge on the protein, causing electrostatic repulsion and the disruption of some hydrogen bonding.

The tertiary structure of a globular protein is determined by its amino acid sequence. Certain globular proteins denatured by heat, extremes of pH, or denaturing reagents will regain their native structure and their biological activity if returned to conditions in which the native conformation is stable. This process is called **renaturation**.

A classic example is the denaturation and renaturation of ribonuclease A.

Not all proteins fold spontaneously as they are synthesized in the cell. Folding for many proteins is facilitated by the action of specialized proteins. **Molecular chaperones** are proteins that interact with partially folded or improperly folded polypeptides, facilitating correct folding pathways or providing microenvironments in which folding can occur.

Two major families of chaperones, both well studied, are the **Hsp70** family and the chaperonins.

The Hsp70 family of proteins generally have a molecular weight near 70,000 and are more abundant in cells stressed by elevated temperatures (hence, *heat shock proteins* of *M*, 70,000, or Hsp70). Hsp70 proteins bind to regions of unfolded polypeptides that are rich in hydrophobic residues, preventing inappropriate aggregation. These chaperones thus “protect” proteins that have been denatured by heat and peptides that are being synthesized (and are not yet folded). Hsp70 proteins also block the folding of certain proteins that must remain unfolded until they have been translocated across membrane.

Chaperonins are elaborate protein complexes required for the folding of a number of cellular proteins that do not fold spontaneously.

# Protein Function

Knowing the three-dimensional structure of a protein is an important part of understanding how the protein functions. Proteins are dynamic molecules whose functions almost invariably depend on interactions with other molecules.

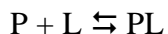
The functions of many proteins involve the reversible binding of other molecules. A molecule bound reversibly by a protein is called a **ligand**. A ligand may be any kind of molecule, including another protein.

A ligand binds at a site on the protein called the **binding site**, which is complementary to the ligand in size, shape, charge, and hydrophobic or hydrophilic character. Furthermore, the interaction is specific: the protein can discriminate among the thousands of different molecules in its environment and selectively bind only one or a few.

The binding of a protein and ligand is often coupled to a conformational change in the protein that makes the binding site more complementary to the ligand, permitting tighter binding. The structural adaptation that occurs between protein and ligand is called **induced fit**.

## QUANTITATIVE PROTEIN LIGAND INTERACTIONS:

In general, the reversible binding of a protein (P) to a ligand (L) can be described by a simple equilibrium expression:



The reaction is characterized by an equilibrium constant,  $K_a$ , such that

$$K_a = \frac{[PL]}{[P][L]} = \frac{k_a}{k_d}$$

Where  $k_a$  and  $k_d$  are rate constants. The term  $K_a$  is an association constant, that describes the equilibrium between the complex and the unbound components of the complex. The association constant provides a measure of the affinity of the ligand L for the protein.  $K_a$  has units of  $M^{-1}$ ; a higher value of  $K_a$  corresponds to a higher affinity of the ligand for the protein.

Consider the binding equilibrium from the standpoint of the fraction, ( $\theta$ ), of ligand binding sites on the protein that are occupied by ligand:

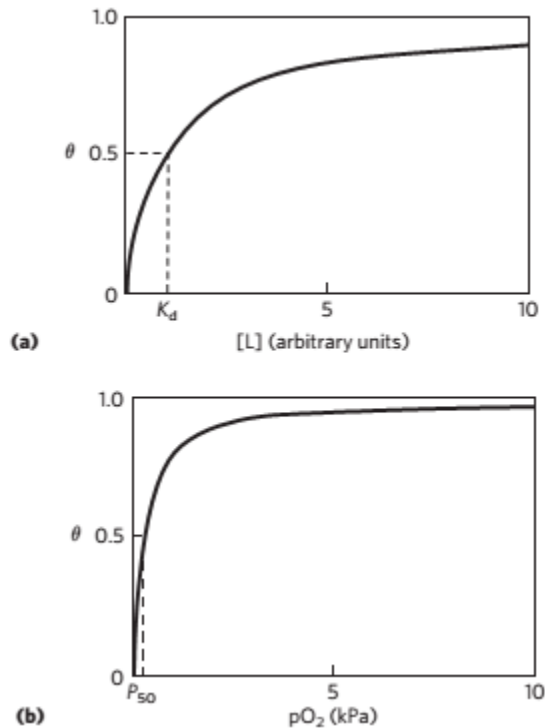
$$\theta = \frac{\text{binding sites occupied}}{\text{total binding sites}} = \frac{[PL]}{[PL] + [P]}$$

$\theta$  is thus found to be a hyperbolic function of  $[L]$ .

Dissociation constant,  $K_d$ , which is the reciprocal of  $K_a$  ( $K_d = 1/K_a$ ) and is given in units of molar concentration (M).  $K_d$  is the equilibrium constant.

Lower value of  $K_d$  corresponds to a higher affinity of ligand for the protein. The mathematics can be reduced to simple statements:  $K_d$  is equivalent to the molar concentration of

ligand at which half of the available ligand-binding sites are occupied. At this point, the protein is said to have reached half-saturation with respect to ligand binding. The more tightly a protein binds a ligand, the lower the concentration of ligand required for half the binding sites to be occupied, and thus the lower the value of  $K_d$ .



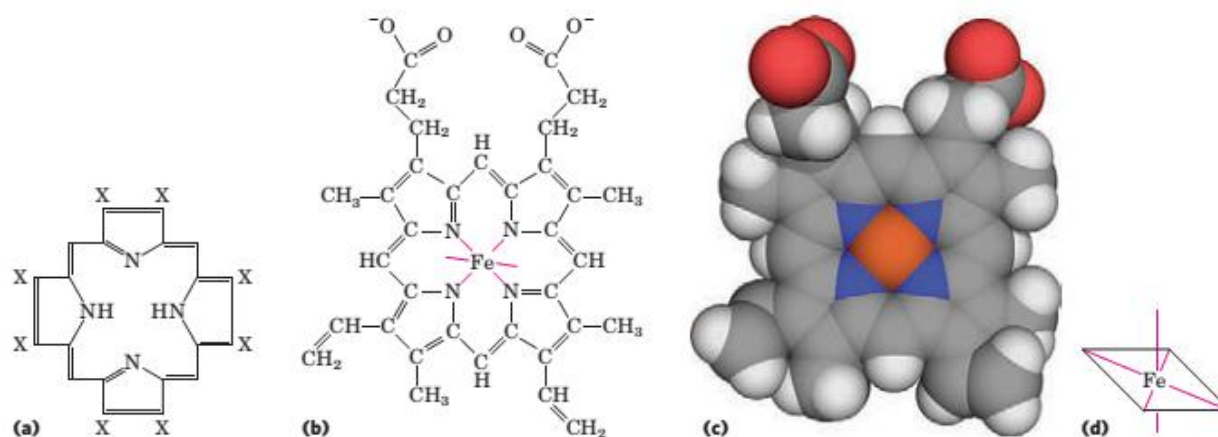
## OXYGEN BINDING PROTEINS:

Oxygen is poorly soluble in aqueous solutions and cannot be carried to tissues in sufficient quantity if it is simply dissolved in blood serum. Diffusion of oxygen through tissues is also ineffective over distances greater than a few millimeters. The evolution of larger, multicellular animals depended on the evolution of proteins that could transport and store oxygen. However, none of the amino acid side chains in proteins is suited for the reversible binding of oxygen molecules. This role is filled by certain transition metals, among them iron and copper, that have a strong tendency to bind oxygen.

Free iron promotes the formation of highly reactive oxygen species such as hydroxyl radicals that can damage DNA and other macromolecules. Iron used in cells is therefore bound in forms that sequester it and/or make it less reactive. Iron is often incorporated into a protein-bound prosthetic group called **heme**.

Heme (or haem) consists of a complex organic ring structure, **protoporphyrin**, to which is bound a single iron atom in its ferrous ( $Fe^{2+}$ ) state. The iron atom has six coordination bonds,

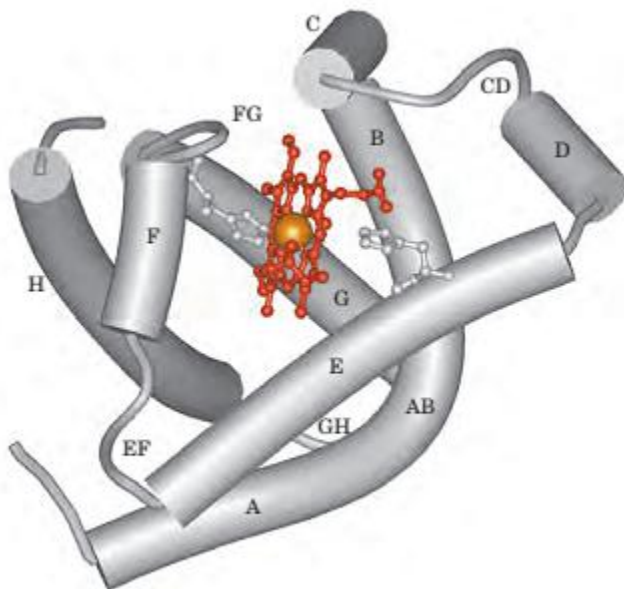
four to nitrogen atoms that are part of the flat **porphyrin ring** system and two perpendicular to the porphyrin. The coordinated nitrogen atoms (which have an electron-donating character) help prevent conversion of the heme iron to the ferric ( $\text{Fe}^{3+}$ ) state. Iron in the  $\text{Fe}^{2+}$  state binds oxygen reversibly; in the  $\text{Fe}^{3+}$  state it does not bind oxygen. Heme is found in a number of oxygen-transporting proteins, as well as in some proteins, such as the cytochromes, that participate in oxidation-reduction (electron-transfer) reactions.



Free heme molecules (heme not bound to protein), reaction of oxygen at one of the two “open” coordination bonds of iron (perpendicular to the plane of the porphyrin molecule, above and below) can result in irreversible conversion of  $\text{Fe}^{2+}$  to  $\text{Fe}^{3+}$ . In heme containing proteins, this reaction is prevented by sequestering of the heme deep within the protein structure where access to the two open coordination bonds is restricted. One of these two coordination bonds is occupied by a side-chain nitrogen of a His residue. The other is the binding site for molecular oxygen ( $\text{O}_2$ ). When oxygen binds, the electronic properties of heme iron change; this accounts for the change in color from the dark purple of oxygen-depleted venous blood to the bright red of oxygen-rich arterial blood. Some small molecules, such as carbon monoxide ( $\text{CO}$ ) and nitric oxide ( $\text{NO}$ ), coordinate to heme iron with greater affinity than does  $\text{O}_2$ . When a molecule of  $\text{CO}$  is bound to heme,  $\text{O}_2$  is excluded, which is why  $\text{CO}$  is highly toxic to aerobic organisms (a topic). By surrounding and sequestering heme, oxygen-binding proteins regulate the access of  $\text{CO}$  and other small molecules to the heme iron.

The globins are a widespread family of proteins, all having similar primary and tertiary structures. Most function in oxygen transport or storage, although some play a role in the sensing of oxygen, nitric oxide, or carbon monoxide. The monomeric myoglobin facilitates oxygen diffusion in muscle tissue. Myoglobin is particularly abundant in the muscles of diving marine mammals such as seals and whales, where it also has an oxygen-storage function for prolonged excursions undersea. The tetrameric hemoglobin is responsible for oxygen transport in the blood stream.

Myoglobin ( $M_r$  16,700, abbreviated Mb) is a single polypeptide of 153 amino acid residues with one molecule of heme. As is typical for a globin polypeptide, myoglobin is made up of eight  $\alpha$ -helical segments connected by bends. The helical segments are named A through H.



Myoglobin, with its hyperbolic binding curve for oxygen, is relatively insensitive to small changes in the concentration of dissolved oxygen and so functions well as an oxygen-storage protein. Hemoglobin, with its multiple subunits and  $O_2$ -binding sites, is better suited to oxygen transport.

Hemoglobin ( $M_r$  64,500; abbreviated Hb) is roughly spherical, with a diameter of nearly 5.5 nm. It is a tetrameric protein containing four heme prosthetic groups, one associated with each polypeptide chain. Adult hemoglobin contains two types of globin, two chains (141 residues each) and two chains (146 residues each).

Two major conformations of hemoglobin: the **R state** and the **T state**. Although oxygen binds to hemoglobin in either state, it has a significantly higher affinity for hemoglobin in the R state. Oxygen binding stabilizes the R state. When oxygen is absent experimentally, the T state is more stable and is thus the predominant conformation of **deoxyhemoglobin**. T and R originally denoted “tense” and “relaxed,” respectively.

The binding of  $O_2$  to a hemoglobin subunit in the T state triggers a change in conformation to the R state.

Hemoglobin must bind oxygen efficiently in the lungs, where the  $pO_2$  is about 13.3 kPa, and release oxygen in the tissues, where the  $pO_2$  is about 4 kPa. Myoglobin, or any protein that binds oxygen with a hyperbolic binding curve, would be ill-suited to this function.

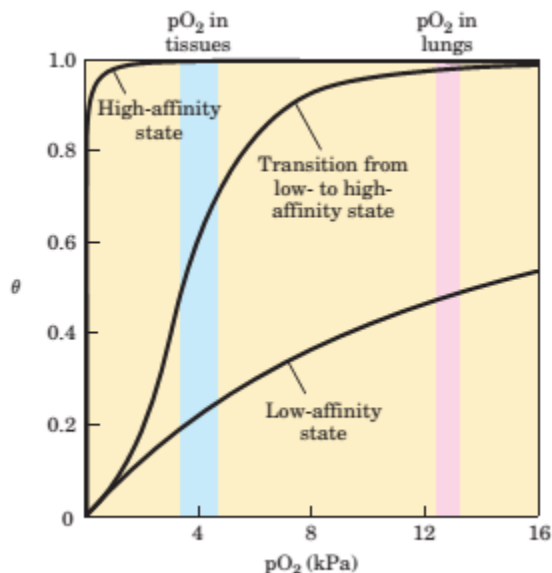
A protein that bound  $O_2$  with high affinity would bind it efficiently in the lungs but would not release much of it in the tissues. If the protein bound oxygen with a sufficiently low affinity to release it in the tissues, it would not pick up much oxygen in the lungs.

Hemoglobin solves the problem by undergoing a transition from a low-affinity state (the T state) to a high-affinity state (the R state) as more  $O_2$  molecules are bound. As a result, hemoglobin has a hybrid S-shaped, or sigmoid, binding curve for oxygen. A single-subunit protein with a single ligand binding site cannot produce a sigmoid binding curve— even if binding elicits a conformational change— because each molecule of ligand binds independently and cannot affect the binding of another molecule. In contrast,  $O_2$  binding to individual subunits of hemoglobin can alter the affinity for  $O_2$  in adjacent subunits. The first molecule of  $O_2$  that interacts with deoxyhemoglobin binds weakly, because it binds to a subunit in the T state. Its binding, however, leads to conformational changes that are communicated to adjacent subunits, making it easier for additional molecules of  $O_2$  to bind. In effect, the  $T \rightarrow R$  transition occurs more readily in the second subunit once  $O_2$  is bound to the first subunit. The last (fourth)  $O_2$  molecule binds to a heme in a subunit that is already in the R state, and hence it binds with much higher affinity than the first molecule.

An **allosteric protein** is one in which the binding of a ligand to one site affects the binding properties of another site on the same protein. The term “allosteric” derives from the Greek *allos*, “other,” and *stereos*, “solid” or “shape.” Allosteric proteins are those having “other shapes,” or conformations, induced by the binding of ligands referred to as modulators.

Cooperative binding of a ligand to a multimeric protein, such as we observe with the binding of  $O_2$  to hemoglobin, is a form of allosteric binding

There is only one binding site for  $O_2$  on each subunit, so the allosteric effects giving rise to cooperativity are mediated by conformational changes transmitted from one subunit to another by subunit-subunit interactions. A sigmoid binding curve is diagnostic of cooperative binding.



**A sigmoid (cooperative) binding curve.** A sigmoid binding curve can be viewed as a hybrid curve reflecting a transition from a low-affinity to a high-affinity state. Cooperative binding, as manifested by a sigmoid binding curve, renders hemoglobin more sensitive to the small differences in O<sub>2</sub> concentration between the tissues and the lungs, allowing hemoglobin to bind oxygen in the lungs (where pO<sub>2</sub> is high) and release it in the tissues (where pO<sub>2</sub> is low).

In addition to carrying nearly all the oxygen required by cells from the lungs to the tissues, hemoglobin carries two end products of cellular respiration—H and CO<sub>2</sub>— from the tissues to the lungs and the kidneys. The CO<sub>2</sub>, produced by oxidation of organic fuels in mitochondria, is hydrated to form bicarbonate:



This reaction is catalyzed by **carbonic anhydrase**. As you can see from the equation, the hydration of CO<sub>2</sub> results in an increase in the H concentration (a decrease in pH) in the tissues. The binding of oxygen by hemoglobin is profoundly influenced by pH and CO<sub>2</sub> concentration, so the interconversion of CO<sub>2</sub> and bicarbonate is of great importance to the regulation of oxygen binding and release in the blood.

The binding of H and CO<sub>2</sub> is inversely related to the binding of oxygen. At the relatively low pH and high CO<sub>2</sub> concentration of peripheral tissues, the affinity of hemoglobin for oxygen decreases as H and CO<sub>2</sub> are bound, and O<sub>2</sub> is released to the tissues. Conversely, in the capillaries of the lung, as CO<sub>2</sub> is excreted and the blood pH consequently rises, the affinity of hemoglobin for oxygen increases and the protein binds more O<sub>2</sub> for transport to the peripheral tissues. This effect of pH and CO<sub>2</sub> concentration on the binding and release of oxygen by hemoglobin is called the **Bohr effect**.

When the oxygen concentration is high, as in the lungs, hemoglobin binds O<sub>2</sub> and releases protons. When the oxygen concentration is low, as in the peripheral tissues, H is bound and O<sub>2</sub> is released.

The interaction of **2,3-bisphosphoglycerate (BPG)** with hemoglobin provides an example of heterotropic allosteric modulation. BPG is present in relatively high concentrations in erythrocytes. 2, 3-Bisphosphoglycerate is known to greatly reduce the affinity of hemoglobin for oxygen—there is an inverse relationship between the binding of O<sub>2</sub> and the binding of BPG.

BPG binds at a site distant from the oxygen-binding site and regulates the O<sub>2</sub>-binding affinity of hemoglobin in relation to the pO<sub>2</sub> in the lungs. BPG plays an important role in the physiological adaptation to the lower pO<sub>2</sub> available at high altitudes. For a healthy human strolling by the ocean, the binding of O<sub>2</sub> to hemoglobin is regulated such that the amount of O<sub>2</sub> delivered to the tissues is equivalent to nearly 40% of the maximum that could be carried by the blood. Imagine that this person is quickly transported to a mountainside at an altitude of 4,500 meters, where the pO<sub>2</sub> is considerably lower. The delivery of O<sub>2</sub> to the tissues is now reduced. However, after just a few hours at the higher altitude, the BPG concentration in the blood has begun to rise, leading to a decrease in the affinity of hemoglobin for oxygen.

As a result, the delivery of oxygen to the tissues is restored to nearly 40% of that which can be transported by the blood. The situation is reversed when the person returns to sea level.



Unlike O<sub>2</sub>, only one molecule of BPG is bound to each hemoglobin tetramer. BPG lowers hemoglobin's affinity for oxygen by stabilizing the T state.

The erythrocytes of these individuals are fewer and also abnormal. In addition to an unusually large number of immature cells, the blood contains many long, thin, crescent-shaped erythrocytes that look like the blade of a sickle. When hemoglobin from sickle cells (called hemoglobin S) is deoxygenated, it becomes insoluble and forms polymers that aggregate into tubular fibers. Normal hemoglobin (hemoglobin A) remains soluble on deoxygenation. The insoluble fibers of deoxygenated hemoglobin S are responsible for the deformed sickle shape of the erythrocytes, and the proportion of sickled cells increases greatly as blood is deoxygenated.

The altered properties of hemoglobin S result from a single amino acid substitution, a Val instead of a Glu residue at position 6 in the two  $\beta$  chains. Replacement of the Glu residue by Val creates a "sticky" hydrophobic contact point at position 6 of the  $\beta$  chain, which is on the outer surface of the molecule. These sticky spots cause deoxyhemoglobin S molecules to associate abnormally with each other, forming the long, fibrous aggregates characteristic of this disorder.

People with sickle-cell anemia suffer from repeated crises brought on by physical exertion. They become weak, dizzy, and short of breath, and they also experience heart murmurs and an increased pulse rate.

# Enzymes

There are two fundamental conditions for life. First, the living entity must be able to self-replicate; second, the organism must be able to catalyze chemical reactions efficiently and selectively. Catalysis is essential for life. Many of us, for example, consume substantial amounts of sucrose—common table sugar—as a kind of fuel, whether in the form of sweetened foods and drinks or as sugar itself. The conversion of sucrose to CO<sub>2</sub> and H<sub>2</sub>O in the presence of oxygen is a highly exergonic process, releasing free energy that we can use to think, move, taste, and see. However, a bag of sugar can remain on the shelf for years without any obvious conversion to CO<sub>2</sub> and H<sub>2</sub>O. Although this chemical process is thermodynamically favorable, it is very slow! Yet when sucrose is consumed by a human (or almost any other organism), it releases its chemical energy in seconds. The difference is catalysis. Without catalysis, chemical reactions such as sucrose oxidation could not occur on a useful time scale, and thus could not sustain life. In living organisms, enzymes promote catalysis.

With the exception of a small group of catalytic RNA molecules, all enzymes are proteins. Their catalytic activity depends on the integrity of their native protein conformation. If an enzyme is denatured or dissociated into its subunits, catalytic activity is usually lost.

Some enzymes require no chemical groups for activity other than their amino acid residues. Others require an additional chemical component called a **cofactor**—either one or more inorganic ions, such as Fe<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup>, or Zn<sup>2+</sup>, or a complex organic or metalloorganic molecule called a **coenzyme**. Coenzymes act as transient carriers of specific functional groups. A coenzyme or metal ion that is very tightly or even covalently bound to the enzyme protein is called a **prosthetic group**. A complete, catalytically active enzyme together with its bound coenzyme and/or metal ions is called a **holoenzyme**. The protein part of such an enzyme is called the **apoenzyme** or **apoprotein**.

## Some Inorganic Ions That Serve as Cofactors for Enzymes

Ions	Enzymes
Cu <sup>2+</sup>	Cytochrome oxidase
Fe <sup>2+</sup> or Fe <sup>3+</sup>	Cytochrome oxidase, catalase, peroxidase
K <sup>+</sup>	Pyruvate kinase
Mg <sup>2+</sup>	Hexokinase, glucose 6-phosphatase, pyruvate kinase
Mn <sup>2+</sup>	Arginase, ribonucleotide reductase
Mo	Dinitrogenase
Ni <sup>2+</sup>	Urease
Zn <sup>2+</sup>	Carbonic anhydrase, alcohol dehydrogenase, carboxypeptidases A and B

Many enzymes have been named by adding the suffix “-ase” to the name of their substrate or to a word or phrase describing their activity. Thus urease catalyzes hydrolysis of urea, and DNA polymerase catalyzes the polymerization of nucleotides to form DNA.

Biochemists, by international agreement, have adopted a system for naming and classifying enzymes. This system divides enzymes into six classes, each with subclasses, based on the type of reaction catalyzed.

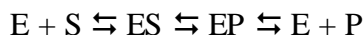
#### International Classification of Enzymes

Class No.	Class Name	Type of reaction catalyzed
1	Oxidoreductases	Transfer of electrons (hydride ions or H atoms)
2	Transferases	Group transfer reactions
3	Hydrolases	Hydrolysis reactions (transfer of functional groups to water)
4	Lyases	Addition of groups to double bonds, or formation of double bonds by removal of groups
5	Isomerases	Transfer of groups within molecules to yield isomeric forms
6	Ligases	Formation of COC, COS, COO, and CON bonds by condensation reactions coupled to ATP cleavage

Under biologically relevant conditions, uncatalyzed reactions tend to be slow. Furthermore, many common reactions in biochemistry entail chemical events that are unfavorable or unlikely in the cellular environment.

An enzyme circumvents these problems by providing a specific environment within which a given reaction can occur more rapidly. The distinguishing feature of an enzyme-catalyzed reaction is that it takes place within the confines of a pocket on the enzyme called the **active site**. The molecule that is bound in the active site and acted upon by the enzyme is called the **substrate**.

The enzyme substrate complex is central to the action of enzymes. A simple enzymatic reaction might be written.



Where E, S, and P represent the enzyme, substrate, and product; ES and EP are transient complexes of the enzyme with the substrate and with the product.

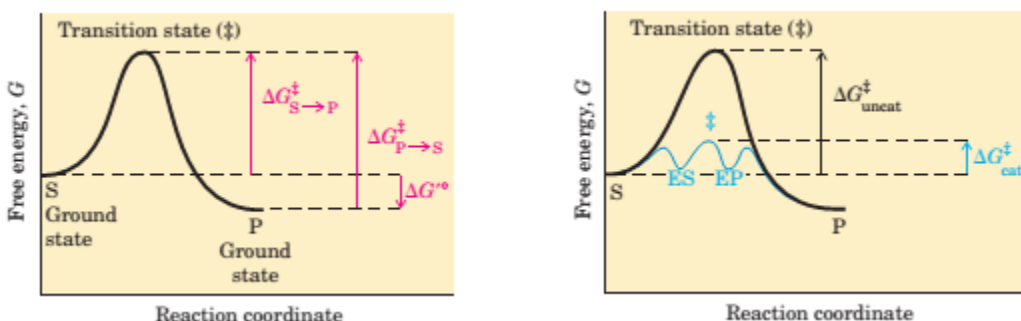
To understand catalysis, we must first appreciate the important distinction between reaction equilibria and reaction rates. The function of a catalyst is to increase the *rate* of a reaction. Catalysts do not affect reaction *equilibria*. (Recall that a reaction is at equilibrium when there is no net change in the concentrations of reactants or products). Energy in biological systems is described in terms of free energy, G.

To describe the free-energy changes for reactions, chemists define a standard set of conditions (temperature 298 K; partial pressure of each gas 1 atm, or 101.3 kPa; concentration of each solute 1 M) and express the free-energy change for this reacting system  $\Delta G'^{\circ}$ , the **standard free energy change**. Because biochemical systems commonly involve H

concentrations far below 1 M, biochemists define a **biochemical standard free-energy change**,  $\Delta G^{\circ}$  \_\_\_\_\_, the standard free-energy change *at pH 7.0*.

A favorable equilibrium does not mean that the  $S \rightarrow P$  conversion will occur at a detectable rate. The *rate* of a reaction is dependent on an entirely different parameter. There is an energy barrier between S and P. To undergo reaction, the molecules must overcome this barrier and therefore must be raised to a higher energy level. At the top of the energy hill is a point at which decay to the S or P state is equally probable (it is downhill either way). This is called the **transition state**. The transition state is not a chemical species with any significant stability and should not be confused with a reaction intermediate (such as ES or EP).

The difference between the energy levels of the ground state and the transition state is the **activation energy**,  $\Delta G$ . The rate of a reaction reflects this activation energy: a higher activation energy corresponds to a slower reaction. Reaction rates can be increased by raising the temperature, thereby increasing the number of molecules with sufficient energy to overcome the energy barrier. Alternatively, the activation energy can be lowered by adding a catalyst. *Catalysts enhance reaction rates by lowering activation energies.*



Enzymes are no exception to the rule that catalysts do not affect reaction equilibria. Any enzyme that catalyzes the reaction  $S \rightarrow P$  also catalyzes the reaction  $P \rightarrow S$ . The role of enzymes is to *accelerate* the interconversion of S and P. The enzyme is not used up in the process, and the equilibrium point is unaffected. However, the reaction reaches equilibrium much faster when the appropriate enzyme is present, because the rate of the reaction is increased.

When several steps occur in a reaction, the overall rate is determined by the step (or steps) with the highest activation energy; this is called the **rate-limiting step**.

Activation energies are energy barriers to chemical reactions. Without such energy barriers, complex macromolecules would revert spontaneously to much simpler molecular forms.

Reaction *equilibria* are inextricably linked to the standard free-energy change for the reaction,  $\Delta G^{\circ}$ , and reaction *rates* are linked to the activation energy,  $\Delta G^{\ddagger}$ .

An equilibrium such as  $S \rightleftharpoons P$  is described by an **equilibrium constant**,  $K_{eq}$

The rate of any reaction is determined by the concentration of the reactant (or reactants) and by a **rate constant**, usually denoted by  $k$ . For the unimolecular reaction  $S \rightarrow P$ , the rate (or

velocity) of the reaction,  $V$ —representing the amount of  $S$  that reacts per unit time—is expressed by a **rate equation**:

$$V = k [S]$$

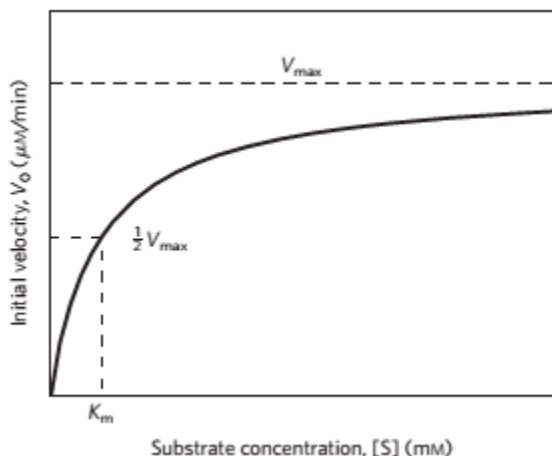
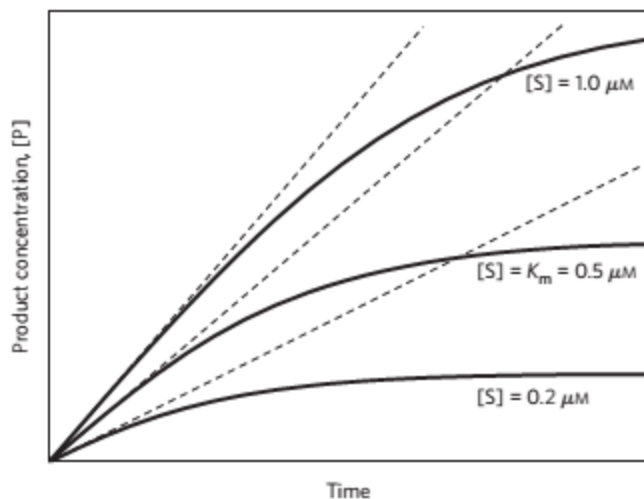
In this reaction, the rate depends only on the concentration of  $S$ . This is called a first-order reaction. Here,  $k$  is a first-order rate constant and has units of reciprocal time, such as  $S^{-1}$ . If a reaction rate depends on the concentration of two different compounds, or if the reaction is between two molecules of the same compound, the reaction is second order and  $k$  is a second-order rate constant, with units of  $M^{-1}S^{-1}$ . The rate equation then becomes.

$$V = k [S_1][S_2]$$

### **ENZYME KINETICS:**

The oldest approach to understanding enzyme mechanisms that remains the most important, is to determine the *rate* of the reaction and how it changes in response to changes in experimental parameters, a discipline known as **enzyme kinetics**.

A key factor affecting the rate of a reaction catalyzed by an enzyme is the concentration of substrate,  $[S]$ . However, studying the effects of substrate concentration is complicated by the fact that  $[S]$  changes during the course of an in vitro reaction as substrate is converted to product. One simplifying approach in kinetics experiments is to measure the **initial rate** (or **initial velocity**), designated  **$V_0$** , when  $[S]$  is much greater than the concentration of enzyme,  $[E]$ . In a typical reaction, the enzyme may be present in nanomolar quantities, whereas  $[S]$  may be five or six orders of magnitude higher. If only the beginning of the reaction is monitored (often the first 60 seconds or less), changes in  $[S]$  can be limited to a few percent, and  $[S]$  can be regarded as constant. At relatively low concentrations of substrate,  $V_0$  increases almost linearly with an increase in  $[S]$ . At higher substrate concentrations,  $V_0$  increases by smaller and smaller amounts in response to increases in  $[S]$ . Finally, a point is reached beyond which increases in  $V_0$  are vanishingly small as  $[S]$  increases. This plateau-like  $V_0$  region is close to the **maximum velocity**,  **$V_{max}$** .



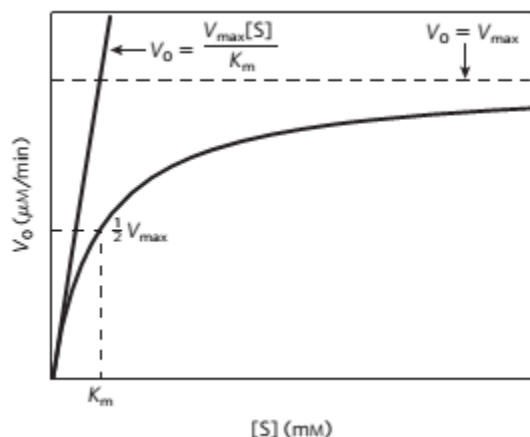
**Effect of substrate concentration on the initial velocity of an enzyme-catalyzed reaction.**  $V_{\text{max}}$  is extrapolated from the plot, because  $V_0$  approaches but never quite reaches  $V_{\text{max}}$ . The substrate concentration at which  $V_0$  is half maximal is  $K_m$ , the Michaelis constant. The concentration of enzyme in an experiment such as this is generally so low that  $[S] \gg [E]$  even when  $[S]$  is described as low or relatively low. The units shown are typical for enzyme-catalyzed reactions and are given only to help illustrate the meaning of  $V_0$  and  $[S]$ . (Note that the curve describes *part* of a rectangular hyperbola, with one asymptote at  $V_{\text{max}}$ . If the curve were continued below  $[S] = 0$ , it would approach a vertical asymptote at  $[S] = -K_m$ .)

At low  $[S]$ , most of the enzyme is in the uncombined form  $E$ . Here, the rate is proportional to  $[S]$ . The maximum initial rate of the catalyzed reaction ( $V_{\text{max}}$ ) is observed when virtually all the enzyme is present as the  $ES$  complex and  $[E]$  is vanishingly small. Under these conditions, the enzyme is “saturated” with its substrate, so that further increases in  $[S]$  have no effect on rate. This condition exists when  $[S]$  is sufficiently high.

The curve expressing the relationship between  $[S]$  and  $V_0$  has the same general shape for most enzymes (it approaches a rectangular hyperbola), which can be expressed algebraically by the Michaelis-Menten equation.

$$V_0 = \frac{V_{\max}[S]}{K_m + [S]}$$

All these terms  $[S]$ ,  $V_0$ ,  $V_{\max}$ , and a constant called the Michaelis constant, are readily measured experimentally.



**Dependence of initial velocity on substrate concentration.** This graph shows the kinetic parameters that define the limits of the curve at high and low  $[S]$ . At low  $[S]$ ,  $K_m \gg [S]$  and the  $[S]$  term in the denominator of the Michaelis-Menten equation becomes insignificant. The equation simplifies to  $V_0 = V_{\max}[S]/K_m$  and  $V_0$  exhibits a linear dependence on  $[S]$ , as observed here. At high  $[S]$ , where  $[S] \gg K_m$ , the  $K_m$  term in the denominator of the Michaelis-Menten equation becomes insignificant and the equation simplifies to  $V_0 = V_{\max}$ ; this is consistent with the plateau observed at high  $[S]$ . The Michaelis-Menten equation is therefore consistent with the observed dependence of  $V_0$  on  $[S]$ , and the shape of the curve is defined by the terms  $V_{\max}/K_m$  at low  $[S]$  and  $V_{\max}$  at high  $[S]$ .

#### DEFINITION OF $K_M$ :

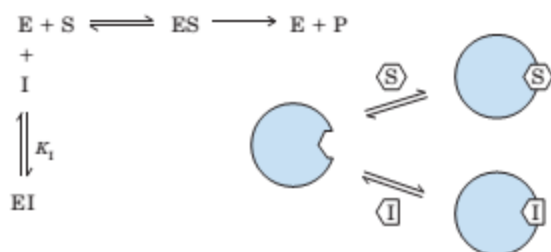
$K_m [S]$ , when  $V_0 = \frac{1}{2} V_{\max}$

This is a very useful, practical definition of  $K_m$ :  $K_m$  is equivalent to the substrate concentration at which  $V_0$  is one-half  $V_{\max}$ .

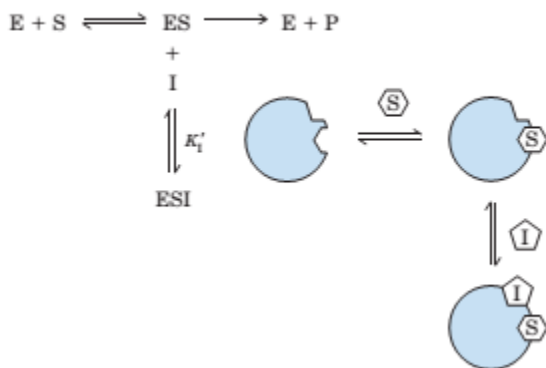
#### ENZYME INHIBITION:

Enzyme inhibitors are molecular agents that interfere with catalysis, slowing or halting enzymatic reactions. For example, aspirin (acetylsalicylate) inhibits the enzyme that catalyzes the first step in the synthesis of prostaglandins, compounds involved in many processes, including some that produce pain.

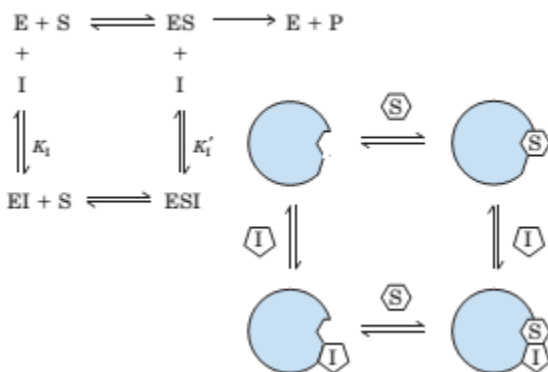
**(a) Competitive inhibition**



**(b) Uncompetitive inhibition**



**(c) Mixed inhibition**



*Reversible Inhibition* One common type of **reversible inhibition** is called competitive. A **competitive inhibitor** competes with the substrate for the active site of an enzyme. While the inhibitor (I) occupies the active site it prevents binding of the substrate to the enzyme. Many competitive inhibitors are compounds that resemble the substrate and combine with the enzyme to form an EI complex, but without leading to catalysis.

Because the inhibitor binds reversibly to the enzyme, the competition can be biased to favor the substrate simply by adding more substrate. When  $[S]$  far exceeds  $[I]$ , the probability that an inhibitor molecule will bind to the enzyme is minimized and the reaction exhibits a normal  $V_{\max}$ .

Two other types of reversible inhibition, uncompetitive and mixed, though often defined in terms of one substrate enzymes, are in practice observed only with enzymes having two or



more substrates. An **uncompetitive inhibitor** binds at a site distinct from the substrate active site and, unlike a competitive inhibitor, binds only to the ES complex.

A **mixed inhibitor** also binds at a site distinct from the substrate active site, but it binds to either E or ES.

*Irreversible Inhibition* The **irreversible inhibitors** are those that bind covalently with or destroy a functional group on an enzyme that is essential for the enzyme's activity, or those that form a particularly stable noncovalent association.

# Carbohydrates and Glycobiology

Carbohydrates are the most abundant biomolecules on Earth. Certain carbohydrates (sugar and starch) are a dietary staple in most parts of the world. Insoluble carbohydrate polymers serve as structural and protective elements in the cell walls of bacteria and plants and in the connective tissues of animals.

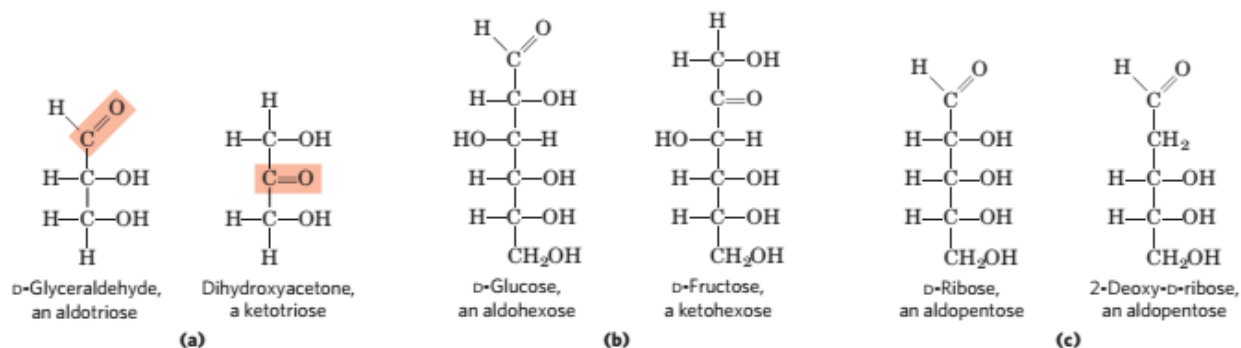
Carbohydrates are polyhydroxy aldehydes or ketones, or substances that yield such compounds on hydrolysis. Many, but not all, carbohydrates have the empirical formula  $(\text{CH}_2\text{O})_n$ ; There are three major size classes of carbohydrates: monosaccharides, oligosaccharides, and polysaccharides (the word “saccharide” is derived from the Greek *sakcharon*, meaning “sugar”). **Monosaccharides**, or simple sugars, consist of a single polyhydroxy aldehyde or ketone unit.

**Oligosaccharides** consist of short chains of monosaccharide units, or residues, joined by characteristic linkages called glycosidic bonds. The most abundant are the **disaccharides**, with two monosaccharide units. Typical is sucrose (cane sugar), which consists of the six-carbon sugars D-glucose and D-fructose.

The **polysaccharides** are sugar polymers containing more than 20 or so monosaccharide units, such as cellulose, Glycogen.

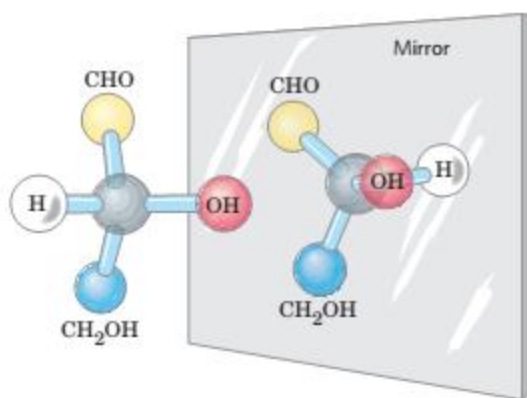
Many of the carbon atoms to which hydroxyl groups are attached are chiral centers, which give rise to the many sugar stereoisomers found in nature. Stereoisomerism in sugars is biologically significant because the enzymes that act on sugars are strictly stereospecific. It is as difficult to fit the wrong sugar stereoisomer into an enzyme's binding site as it is to put your left glove on your right hand.

Monosaccharides are colorless, crystalline solids that are freely soluble in water but insoluble in nonpolar solvents. Most have a sweet taste. In the open-chain form, one of the carbon atoms is double-bonded to an oxygen atom to form a carbonyl group; each of the other carbon atoms has a hydroxyl group. If the carbonyl group is at an end of the carbon chain (that is, in an aldehyde group) the monosaccharide is an **aldose**; if the carbonyl group is at any other position (in a ketone group) the monosaccharide is a **ketose**. The simplest monosaccharides are the two three-carbon trioses: glyceraldehyde, an aldotriose, and dihydroxyacetone, a ketotriose. The aldopentoses D-ribose and 2-deoxy-D-ribose are components of nucleotides and nucleic acids.

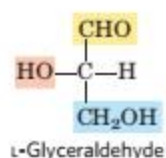
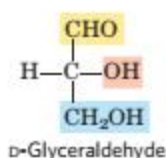


All the monosaccharides except dihydroxyacetone contain one or more asymmetric (chiral) carbon atoms and thus occur in optically active isomeric forms. The simplest aldose, glyceraldehyde, contains one chiral center (the middle carbon atom) and therefore has two different optical isomers, or enantiomers.

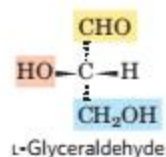
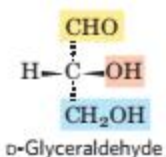
One of the two enantiomers of glyceraldehyde is, by convention, one of these two forms is designated the D isomer, the other the L isomer. As for other biomolecules with chiral centers, the absolute configurations of sugars are known from x-ray crystallography. To represent three-dimensional sugar structures on paper, we often use **Fischer projection formulas**. 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.



Ball-and-stick models

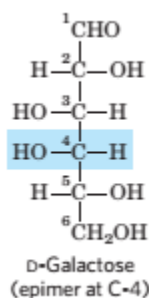
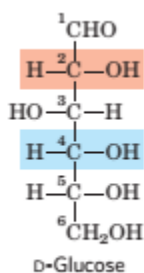
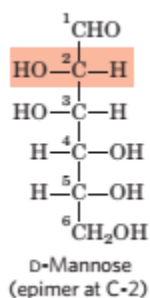


Fischer projection formulas



Perspective formulas

The carbons of a sugar are numbered beginning at the end of the chain nearest the carbonyl group. 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).

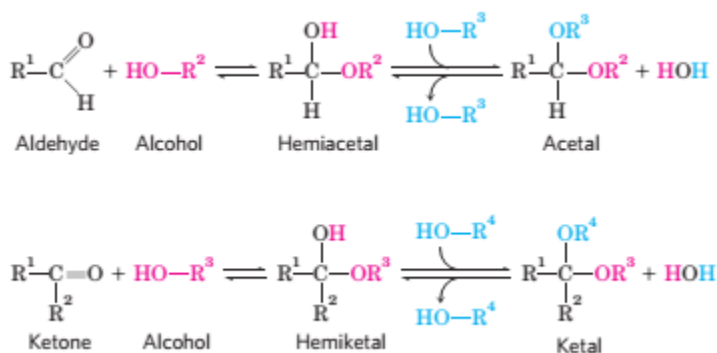


Some sugars occur naturally in their L form; examples are L-arabinose.

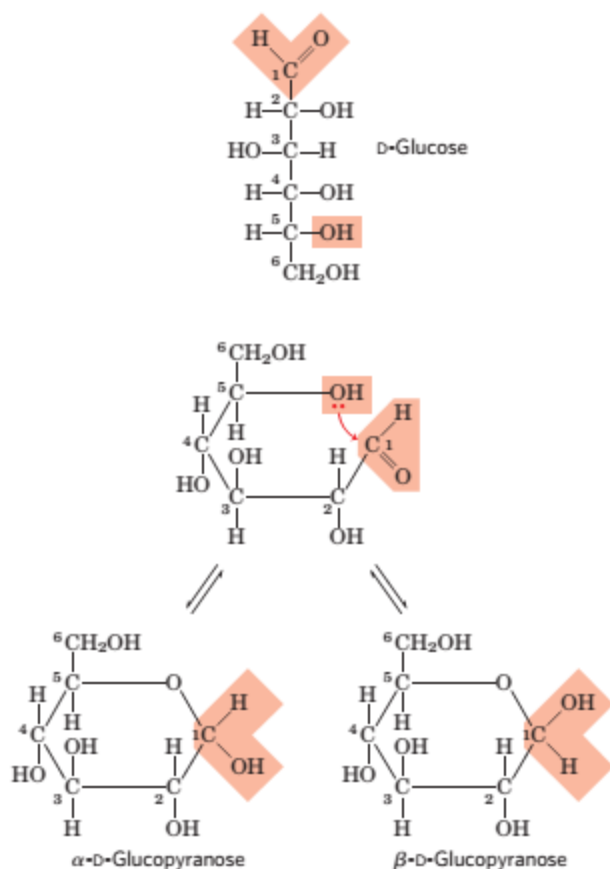
In aqueous solution, aldotetroses and all monosaccharides with five or more carbon atoms in the backbone occur predominantly as cyclic (ring) structures in which the carbonyl

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**. Two molecules of an alcohol can add to a carbonyl carbon; the product of the first addition is a hemiacetal (for addition to an aldose) or a hemiketal (for addition to a ketose).

Addition of the second molecule of alcohol produces the full acetal or ketal, and the bond formed is a glycosidic linkage.

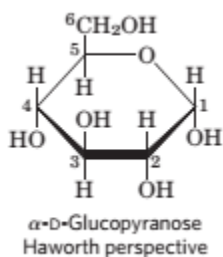


The reaction can produce either of two stereoisomeric configurations, denoted  $\alpha$  and  $\beta$ . Isomeric forms of monosaccharides that differ only in their configuration about the hemiacetal or hemiketal carbon atom are called anomers, and the carbonyl carbon atom is called the anomeric carbon.



Six-membered ring compounds are called pyranoses because they resemble the six-membered ring compounds pyran. The systematic names for the two ring forms of D-glucose are  $\alpha$ -D-glucopyranose and  $\beta$ -D-glucopyranose. Ketohehexoses (such as fructose) also occur as cyclic compounds with  $\alpha$  and  $\beta$  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.

Cyclic sugar structures are more accurately represented in Haworth perspective formulas. In Haworth projections the six-membered ring is tilted to make its plane almost perpendicular to that of the paper, with the bonds closest to the reader drawn thicker than those farther away.



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.

**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  $\alpha$  and  $\beta$  anomers, which differ only in the stereochemistry around the hemiacetal carbon.

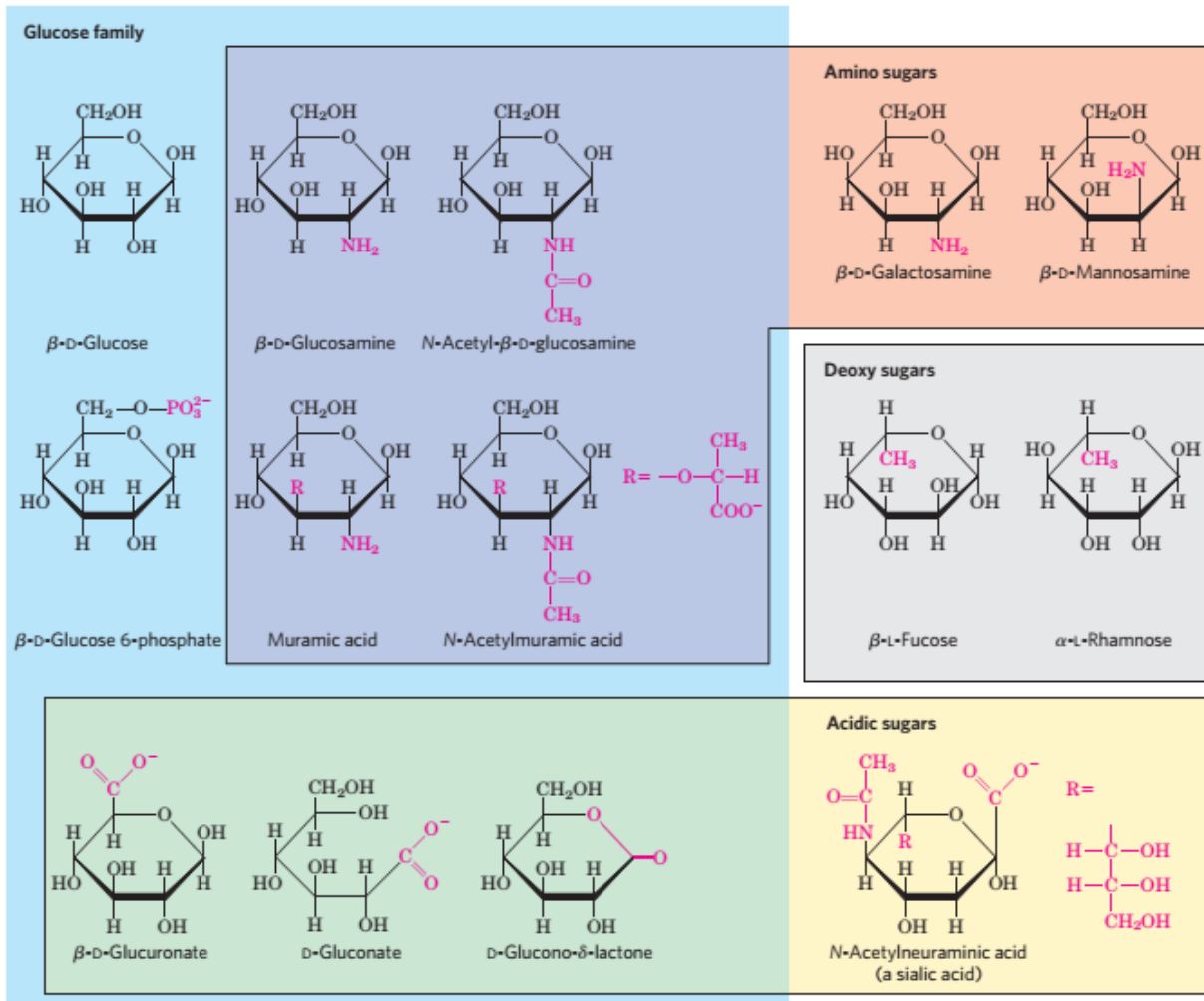
The terminal  $\text{--CH}_2\text{OH}$  group projects upward for the d-enantiomer, downward for the L-enantiomer. The hydroxyl on the anomeric carbon can point up or down. When the anomeric hydroxyl of a D-hexose is on the same side of the ring as C-6, the structure is by definition  $\beta$ ; when it is on the opposite side from C-6, the structure is  $\alpha$ .

The  $\alpha$  and  $\beta$  anomers of D-glucose interconvert in aqueous solution by a process called **mutarotation**, in which one ring form (say, the  $\alpha$  anomer) opens briefly into the linear form, then closes again to produce the  $\beta$  anomer.

Two *conformations* of a molecule are interconvertible without the breakage of covalent bonds, whereas two *configurations* can be interconverted only by breaking a covalent bond. To interconvert  $\alpha$  and  $\beta$  configurations, the bond involving the ring oxygen atom would have to be broken.

### **SUGAR DERIVATIVES:**

There are a number of sugar derivatives in which a hydroxyl group in the parent compound is replaced with another substituent, or a carbon atom is oxidized to a carboxyl group. (1) In glucosamine, galactosamine, and mannosamine, the hydroxyl at C-2 of the parent compound is replaced with an amino group. The amino group is nearly always condensed with acetic acid as in *N*-acetylglucosamine. (2) The substitution of a hydrogen for the hydroxyl group at C-6 of L-galactose or L-mannose produces L-fucose or L-rhamnose, respectively.



(3) Oxidation of the carbonyl (aldehyde) carbon of glucose to the carboxyl level produces gluconic acid; other aldoses yield other **aldonic acids**. (4) Oxidation of the carbon at the other end of the carbon chain—C-6 of glucose, galactose, or mannose—forms the corresponding **uronic acid**: glucuronic, galacturonic, or mannuronic acid. (5) Both aldonic and uronic acids form stable intramolecular esters called lactones.

(6) In the synthesis and metabolism of carbohydrates, the intermediates are very often not the sugars themselves but their phosphorylated derivatives. Condensation of phosphoric acid with one of the hydroxyl groups of a sugar forms a phosphate ester, as in glucose 6-phosphate.

Sugar phosphates are relatively stable at neutral pH and bear a negative charge. One effect of sugar phosphorylation within cells is to trap the sugar inside the cell; most cells do not have plasma membrane transporters for phosphorylated sugars. (7) Cupric ion oxidizes glucose and certain other sugars to a complex mixture of carboxylic acids.

### GLYCOSIDIC BOND:

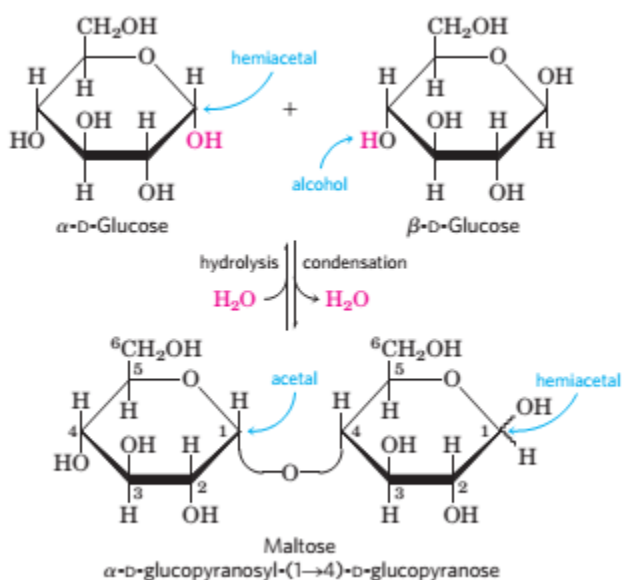
Disaccharides (such as maltose, lactose, and sucrose) consist of two monosaccharides joined covalently by an **O-glycosidic bond**, which is formed when a hydroxyl group of one sugar



reacts with the anomeric carbon of the other. 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). Glycosidic bonds are readily hydrolyzed by acid but resist cleavage by base. *N*-glycosyl bonds join the anomeric carbon of a sugar to a nitrogen atom in glycoproteins and nucleotides.

The oxidation of a sugar by cupric ion occurs only with the linear form. When the anomeric carbon is involved in a glycosidic bond, the easy interconversion of linear and cyclic forms is prevented. Because the carbonyl carbon can be oxidized only when the sugar is in its linear form, formation of a glycosidic bond renders a sugar nonreducing.

The disaccharide maltose contains two D-glucose residues joined by a glycosidic linkage. Because the disaccharide retains a free anomeric carbon, maltose is a reducing sugar. Sucrose is formed by plants but not by animals. In contrast to maltose and lactose, sucrose contains no free anomeric carbon atom; the anomeric carbons of both monosaccharide units are involved in the glycosidic bond. Sucrose is therefore a nonreducing sugar.



Trehalose, a disaccharide of D-glucose that, like sucrose, is a nonreducing sugar—is a major constituent of the circulating fluid (hemolymph) of insects, serving as an energy-storage compound.

Polysaccharides, also called **glycans**, differ from each other in the identity of their recurring monosaccharide units, in the length of their chains, in the types of bonds linking the units, and in the degree of branching. **Homopolysaccharides** contain only a single type of monomer for example – starch, glycogen, cellulose; **heteropolysaccharides** contain two or more different kinds. The bacterial cell envelope (the peptidoglycan) is composed in part of a heteropolysaccharide.

Unlike proteins, polysaccharides generally do not have definite molecular weights. Proteins are synthesized on a template for polysaccharide synthesis there is no template.

The most important storage polysaccharides are starch in plant cells and glycogen in animal cells.

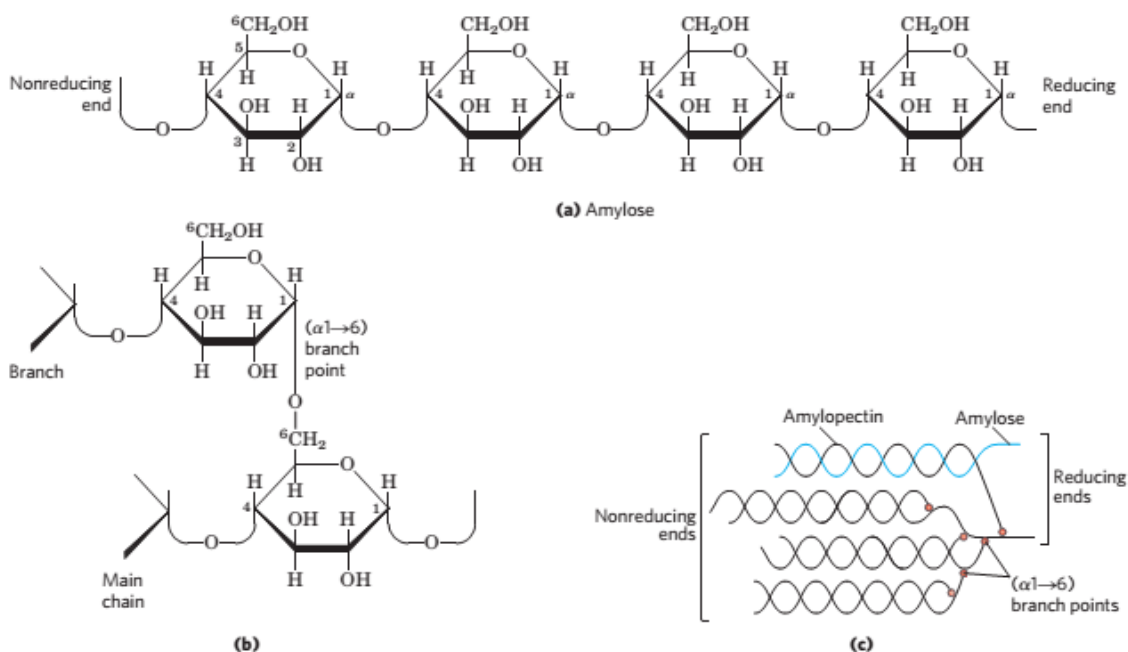
**Starch** contains two types of glucose polymer, amylose and amylopectin. The former consists of long, unbranched chains of D-glucose residues connected by (1→4) linkages.

Amylopectin, unlike amylose is highly branched. The glycosidic linkages joining successive glucose residues in amylopectin chains are (1→4); the branch points (occurring every 24 to 30 residues) are ( $\alpha$ 1→6) linkages.

**Glycogen** is the main storage polysaccharide of animal cells. Like amylopectin, glycogen is a polymer of ( $\alpha$ 1→4)-linked subunits of glucose, with ( $\alpha$ 1→6)-linked branches, but glycogen is more extensively branched (on average, every 8 to 12 residues) and more compact.

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. When glycogen is used as an energy source, glucose units are removed one at a time from the nonreducing ends.

Why not store glucose in its monomeric form? Glycogen is insoluble and contributes little to the osmolarity of the cytosol. If the cytosol contained 0.4 M glucose, the osmolarity would be threateningly elevated, leading to osmotic entry of water that might rupture the cell.



**Dextrans** are bacterial and yeast polysaccharides made up of (1→6) -linked poly-D-glucose; all have (1→3) branches, and some also have (1→2) or (1→4) branches. Dental plaque, is rich in dextrans.

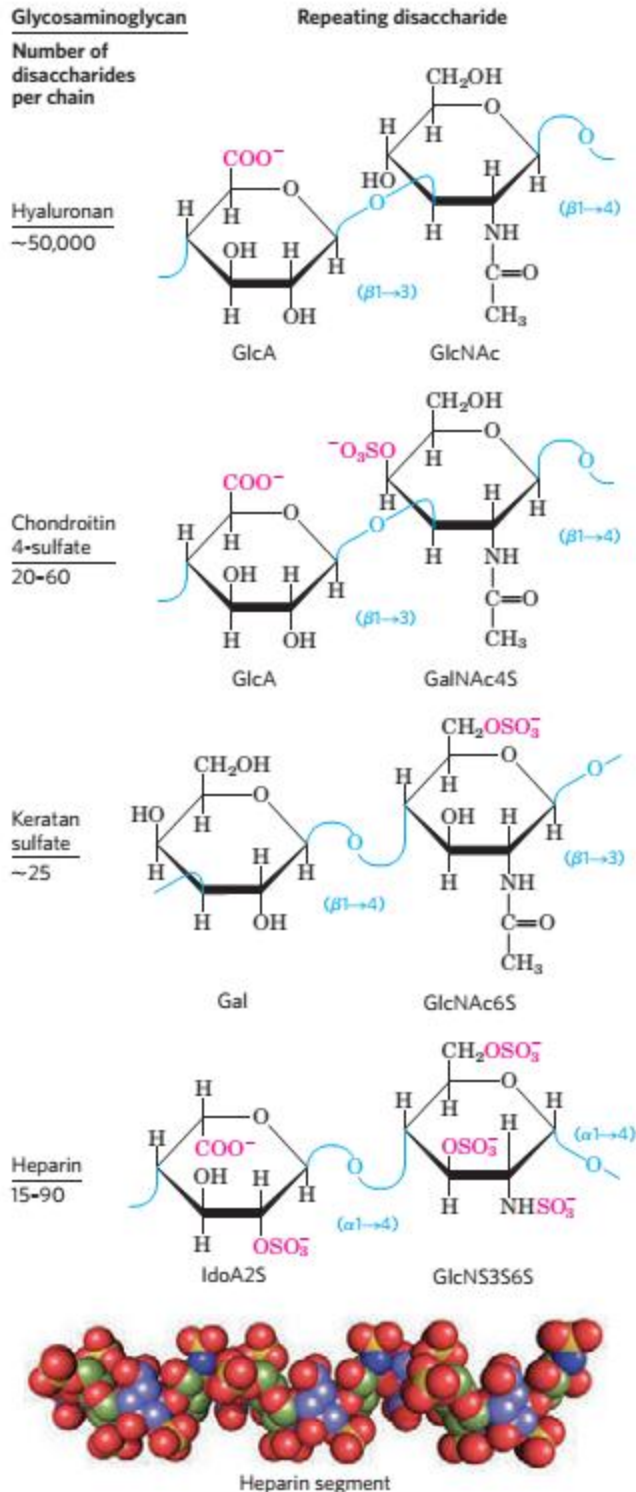
Cellulose, a fibrous, tough, water-insoluble substance, is found in the cell walls of plants, cotton is almost pure cellulose. Like amylose and the main chains of amylopectin and glycogen, the cellulose molecule is a linear, unbranched homopolysaccharide, consisting of 10,000 to 15,000 D-glucose units. But there is a very important difference: in cellulose the glucose residues have the  $\beta$  configuration, whereas in amylose, amylopectin, and glycogen the glucose is in the  $\alpha$  configuration. The glucose residues in cellulose are linked by (1 $\rightarrow$ 4) glycosidic bonds, in contrast to the (1 $\rightarrow$ 4) bonds of amylose, starch, and glycogen. This difference gives cellulose and amylose very different structures and physical properties.

Glycogen and starch ingested in the diet are hydrolyzed by  $\alpha$ -amylases. Most animals cannot use cellulose as a fuel source, because they lack an enzyme to hydrolyze the (1 $\rightarrow$ 4) linkages.

**Chitin** is a linear homopolysaccharide composed of *N*-acetylglucosamine residues in ( $\beta$ 1 $\rightarrow$ 4) linkage. The only chemical difference from cellulose is the replacement of the hydroxyl group at C-2 with an acetylated amino group. Chitin is the principal component of the hard exoskeletons of insects.

The rigid component of bacterial cell walls is a heteropolymer of alternating ( $\beta$ 1 $\rightarrow$ 4)-linked *N* acetylglucosamine and *N*-acetylmuramic acid residues, which prevents cellular swelling and lysis due to the osmotic entry of water. The enzyme lysozyme kills bacteria by hydrolyzing the (1 $\rightarrow$ 4) glycosidic bond between *N*-acetylglucosamine and *N*-acetylmuramic acid.

Heteropolysaccharides, the **glycosaminoglycans**, are a family of linear polymers composed of repeating disaccharide units. One of the two monosaccharides is always either *N*-acetylglucosamine or *N*-acetylgalactosamine; the other is in most cases a uronic acid, usually D-glucuronic or L-iduronic acid. In some glycosaminoglycans, one or more of the hydroxyls of the amino sugar are esterified with sulfate. The combination of sulfate groups and the carboxylate groups of the uronic acid residues gives glycosaminoglycans a very high density of negative charge.



The glycosaminoglycan **hyaluronic acid** (hyaluronate at physiological pH) contains alternating residues of D-glucuronic acid and N-acetylglucosamine they form clear, highly viscous solutions that serve as lubricants in the synovial fluid of joints and give the vitreous

humor of the vertebrate eye its jellylike consistency (the Greek *hyalos* means “glass”; Hyaluronate is also an essential component of the extracellular matrix of cartilage and tendons, to which it contributes tensile strength and elasticity.

Other glycosaminoglycans differ from hyaluronate in two respects: they are generally much shorter polymers and they are covalently linked to specific proteins (proteoglycans). Chondroitin sulfate (Greek *chondros*, “cartilage”) contributes to the tensile strength of cartilage, tendons, ligaments, and the walls of the aorta. Dermatan sulfate (Greek *derma*, “skin”) contributes to the pliability of skin and is also present in blood vessels and heart valves. In this polymer, many of the glucuronate (GlcA) residues present in chondroitin sulfate are replaced by their epimer, iduronate (IdoA).

Keratan sulfates (Greek *keras*, “horn”) have no uronic acid and their sulfate content is variable. They are present in cornea, cartilage, bone, and a variety of horny structures formed of dead cells: horn, hair, hoofs, nails, and claws. Heparin (Greek *he -par*, “liver”); it contains variable arrangements of sulfated and nonsulfated sugars. The sulfated segments of the chain allow it to interact with a large number of proteins. Heparin is a fractionated form of heparin sulfate derived mostly from mast cells, Inhibits coagulation through its capacity to bind the protease inhibitor antithrombin. Heparin binding causes antithrombin to bind to and inhibit thrombin, heparin has the highest negative charge density of any known biological macromolecule.

### **Structures and Roles of Some Polysaccharides**

Polymer	Type*	Repeating unit <sup>†</sup>	Size (number of monosaccharide units)	Roles/significance
Starch				Energy storage: in plants
Amylose	Homo-	( $\alpha$ 1→4)Glc, linear	50–5,000	
Amylopectin	Homo-	( $\alpha$ 1→4)Glc, with ( $\alpha$ 1→6)Glc branches every 24–30 residues	Up to $10^6$	
Glycogen	Homo-	( $\alpha$ 1→4)Glc, with ( $\alpha$ 1→6)Glc branches every 8–12 residues	Up to 50,000	Energy storage: in bacteria and animal cells
Cellulose	Homo-	( $\beta$ 1→4)Glc	Up to 15,000	Structural: in plants, gives rigidity and strength to cell walls
Chitin	Homo-	( $\beta$ 1→4)GlcNAc	Very large	Structural: in insects, spiders, crustaceans, gives rigidity and strength to exoskeletons
Dextran	Homo-	( $\alpha$ 1→6)Glc, with ( $\alpha$ 1→3) branches	Wide range	Structural: in bacteria, extracellular adhesive
Peptidoglycan	Hetero-; peptides attached	4)Mur2Ac( $\beta$ 1→4)GlcNAc( $\beta$ 1	Very large	Structural: in bacteria, gives rigidity and strength to cell envelope
Agarose	Hetero-	3)D-Gal( $\beta$ 1→4)3,6-anhydro-L-Gal( $\alpha$ 1	1,000	Structural: in algae, cell wall material
Hyaluronan (a glycosaminoglycan)	Hetero-; acidic	4)GlcA( $\beta$ 1→3)GlcNAc( $\beta$ 1	Up to 100,000	Structural: in vertebrates, extracellular matrix of skin and connective tissue; viscosity and lubrication in joints

\*Each polymer is classified as a homopolysaccharide (homo-) or heteropolysaccharide (hetero-).

<sup>†</sup>The abbreviated names for the peptidoglycan, agarose, and hyaluronan repeating units indicate that the polymer contains repeats of this disaccharide unit. For example, in peptidoglycan, the GlcNAc of one disaccharide unit is ( $\beta$ 1→4)-linked to the first residue of the next disaccharide unit.

In addition to their important roles as stored fuels (starch, glycogen, dextran) and as structural materials (cellulose, chitin, peptidoglycans), polysaccharides and oligosaccharides are information carriers. On almost every eukaryotic cell, specific oligosaccharide chains attached to components of the plasma membrane form a carbohydrate layer (the glycocalyx). Carbohydrate is covalently joined to a protein or a lipid to form a glycoconjugate which includes Proteoglycans, Glycoproteins and Glycolipids.

**Proteoglycans** are macromolecules of the cell surface or extracellular matrix in which one or more glycosaminoglycan chains are joined covalently to a membrane protein or a secreted protein.

**Glycoproteins** have one or several oligosaccharides of varying complexity joined covalently to a protein.

**Glycolipids** are membrane lipids in which the hydrophilic head groups are oligosaccharides. The brain and neurons are rich in glycosphingolipids, which help in nerve conduction and myelin formation.

# 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.

The fats and oils used almost universally as stored forms of energy in living organisms are derivatives of **fatty acids**.

Fatty acids are carboxylic acids with hydrocarbon chains ranging from 4 to 36 carbons long (C<sub>4</sub> to C<sub>36</sub>). 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.

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 (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(9,12).

The most commonly occurring fatty acids have even numbers of carbon atoms in an unbranched chain of 12 to 24 carbons.

## Some Naturally Occuring atty Acids: Structure, Properties, and Nomenclature

Carbon skeleton	Structure*	Systematic name <sup>†</sup>	Common name (derivation)	Melting point (°C)	Solubility at 30 °C (mg/g solvent)	
					Water	Benzene
12:0	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>10</sub> COOH	<i>n</i> -Dodecanoic acid	Lauric acid (Latin <i>laurus</i> , "laurel plant")	44.2	0.063	2,600
14:0	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>12</sub> COOH	<i>n</i> -Tetradecanoic acid	Myristic acid (Latin <i>Myristica</i> , nutmeg genus)	53.9	0.024	874
16:0	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>14</sub> COOH	<i>n</i> -Hexadecanoic acid	Palmitic acid (Latin <i>palma</i> , "palm tree")	63.1	0.0083	348
18:0	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>16</sub> COOH	<i>n</i> -Octadecanoic acid	Stearic acid (Greek <i>stear</i> , "hard fat")	69.6	0.0034	124
20:0	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>18</sub> COOH	<i>n</i> -Eicosanoic acid	Arachidic acid (Latin <i>Arachis</i> , legume genus)	76.5		
24:0	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>22</sub> COOH	<i>n</i> -Tetracosanoic acid	Lignoceric acid (Latin <i>lignum</i> , "wood" + <i>cera</i> , "wax")	86.0		
16:1(Δ <sup>9</sup> )	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>5</sub> CH=CH(CH <sub>2</sub> ) <sub>7</sub> COOH	<i>cis</i> -9-Hexadecenoic acid	Palmitoleic acid	1 to -0.5		
18:1(Δ <sup>9</sup> )	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>7</sub> CH=CH(CH <sub>2</sub> ) <sub>7</sub> COOH	<i>cis</i> -9-Octadecenoic acid	Oleic acid (Latin <i>oleum</i> , "oil")	13.4		
18:2(Δ <sup>9,12</sup> )	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> CH=CHCH <sub>2</sub> CH=CH(CH <sub>2</sub> ) <sub>7</sub> COOH	<i>cis</i> -, <i>cis</i> -9,12-Octadecadienoic acid	Linoleic acid (Greek <i>linon</i> , "flax")	1-5		
18:3(Δ <sup>9,12,15</sup> )	CH <sub>3</sub> CH <sub>2</sub> CH=CHCH <sub>2</sub> CH=CHCH <sub>2</sub> CH=CH(CH <sub>2</sub> ) <sub>7</sub> COOH	<i>cis</i> -, <i>cis</i> -, <i>cis</i> -9,12,15-Octadecatrienoic acid	α-Linolenic acid	-11		
20:4(Δ <sup>5,8,11,14</sup> )	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> CH=CHCH <sub>2</sub> CH=CHCH <sub>2</sub> CH=CHCH <sub>2</sub> CH=CH(CH <sub>2</sub> ) <sub>3</sub> COOH	<i>cis</i> -, <i>cis</i> -, <i>cis</i> -, <i>cis</i> -5,8,11,14-Icosatetraenoic acid	Arachidonic acid	-49.5		

\*All acids are shown in their nonionized form. At pH 7, all free fatty acids have an ionized carboxylate. Note that numbering of carbon atoms begins at the carboxyl carbon.

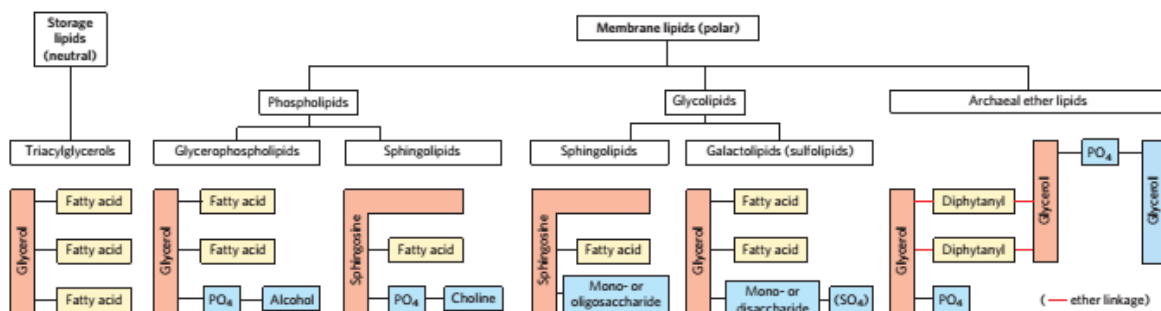
<sup>†</sup>The prefix *n*- indicates the "normal" unbranched structure. For instance, "dodecanoic" simply indicates 12 carbon atoms, which could be arranged in a variety of branched forms; "*n*-dodecanoic" specifies the linear, unbranched form. For unsaturated fatty acids, the configuration of each double bond is indicated; in biological fatty acids the configuration is almost always *cis*.

The family of polyunsaturated fatty acids (PUFAs) with a double bond between the third and fourth carbon from the methyl end of the chain are of special importance in human nutrition. Because the physiological role of PUFAs is related more to the position of the first double bond near the *methyl* end of the chain than to the carboxyl end, an alternative nomenclature is sometimes used for these fatty acids. The carbon of the methyl group—that is, the carbon most distant from the carboxyl group—is called the ω (omega) carbon and is given the number 1. In this convention, PUFAs with a double bond between C-3 and C-4 are called omega-3 (ω-3) fatty acids, and those with a double bond between C-6 and C-7 are omega-6 (ω-6) fatty acids.

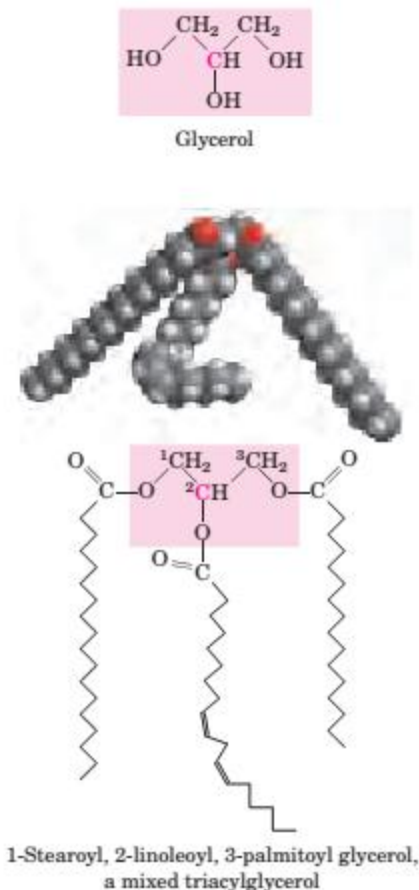


Humans require but do not have the enzymatic capacity to synthesize the omega-3 PUFA  $\alpha$ -linolenic acid (ALA; 18:3 ( $\Delta^{9,12,15}$ ), in the standard convention), and must therefore obtain it in the diet. From ALA, humans can synthesize two other omega-3 PUFAs important in cellular function: eicosapentaenoic acid and docosahexaenoic acid. An imbalance of omega-6 and omega-3 PUFAs in the diet is associated with an increased risk of cardiovascular disease. The optional dietary ratio of omega-6 to omega-3 PUFAs is between 1:1 and 4:1.

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. In unsaturated fatty acids, a *cis* double bond forces a kink in the hydrocarbon chain. Fatty acids with one or several such kinks cannot pack together as tightly as fully saturated fatty acids.



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. 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.



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).

In some animals, triacylglycerols stored under the skin serve not only as energy stores but as insulation against low temperatures.

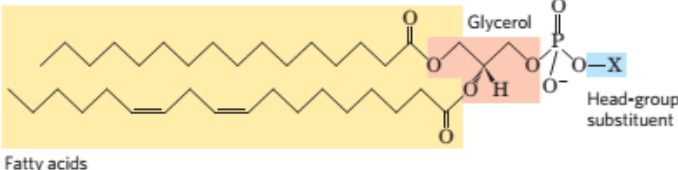
When lipid-rich foods are exposed too long to the oxygen in air, they may spoil and become rancid. The unpleasant taste and smell associated with rancidity result from the oxidative cleavage of the double bonds in unsaturated fatty acids, which produces aldehydes and carboxylic acids of shorter chain length and therefore higher volatility; these compounds pass readily through the air to your nose. To improve the shelf life of vegetable oils are prepared by partial hydrogenation. This process converts many of the cis double bonds in the fatty acids to single bonds and increases the melting temperature of the oils so that they are more nearly solid at room temperature (margarine is produced from vegetable oil in this way). Partial hydrogenation has another, undesirable, effect: some cis double bonds are converted to trans

double bonds. There is now strong evidence that dietary intake of trans fatty acids (often referred to simply as “trans fats”) leads to a higher incidence of cardiovascular disease.

The central architectural feature of biological membranes is a double layer of lipids, which acts as a barrier to the passage of polar molecules and ions. Membrane lipids are amphipathic: one end of the molecule is hydrophobic, the other hydrophilic. Their hydrophobic interactions with each other and their hydrophilic interactions with water direct their packing into sheets called membrane bilayers. In this section we describe 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; archaeobacterial 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.

**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.

Glycerophospholipids are named as derivatives of the parent compound, phosphatidic acid, according to the polar alcohol in the head group. Phosphatidylcholine and phosphatidylethanolamine have choline and ethanolamine in their polar head groups, for example.

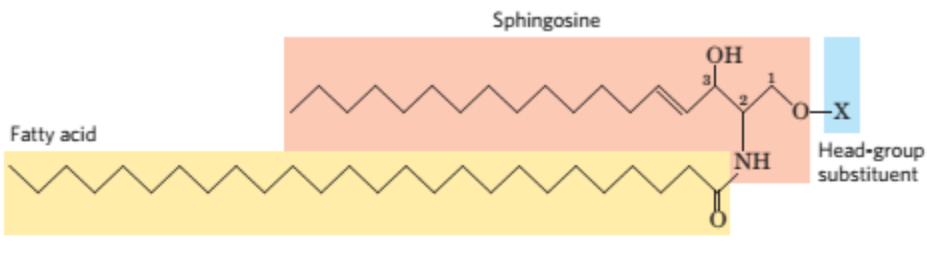
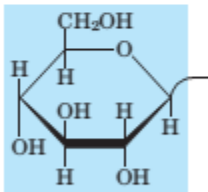

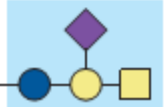
<div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"> <p>Saturated fatty acid (e.g., palmitic acid)</p> <p>Unsaturated fatty acid (e.g., linoleic acid)</p> <p>Fatty acids</p> </div>  </div>			
Name of glycerophospholipid	Name of X—O	Formula of X	Net charge (at pH 7)
Phosphatidic acid	—	— H	−2
Phosphatidylethanolamine	Ethanolamine		0
Phosphatidylcholine	Choline		0
Phosphatidylserine	Serine		−1
Phosphatidylglycerol	Glycerol		−1
Phosphatidylinositol 4,5-bisphosphate	<i>myo</i> -Inositol 4,5-bisphosphate		−4*
Cardiolipin	Phosphatidyl-glycerol		−2

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**. Vertebrate heart tissue is uniquely enriched in ether lipids; about half of the heart phospholipids are plasmalogens.

The second group of membrane lipids are those that predominate in plant cells: the **galactolipids**, in which one or two galactose residues are connected by a glycosidic linkage to C-3 of a 1,2-diacylglycerol. **sulfolipids**, in which a sulfonated glucose residue is joined to a diacylglycerol in glycosidic linkage.

**Sphingolipids**, the fourth large class of membrane lipids, also have a polar head group and two nonpolar tails, but unlike glycerophospholipids and galactolipids they contain no

glycerol. Sphingolipids are composed of one molecule of the long-chain amino alcohol sphingosine. When a fatty acid is attached in amide linkage to the ONH<sub>2</sub> on C-2, the resulting compound is a **ceramide**, which is structurally similar to a diacylglycerol. Ceramide is the structural parent of all sphingolipids.

		
Name of sphingolipid	Name of X—O	Formula of X
Ceramide	—	—H
Sphingomyelin	Phosphocholine	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{P}-\text{O}-\text{CH}_2-\text{CH}_2-\text{N}^+(\text{CH}_3)_3 \\   \\ \text{O}^- \end{array}$
Neutral glycolipids Glucosylcerebroside	Glucose	
Lactosylceramide (a globoside)	Di-, tri-, or tetrasaccharide	
Ganglioside GM2	Complex oligosaccharide	

There are three subclasses of sphingolipids, all derivatives of ceramide but differing in their head groups: sphingomyelins, neutral (uncharged) glycolipids, and gangliosides. **Sphingomyelins** contain phosphocholine or phosphoethanolamine as their polar head group and are therefore classified along with glycerophospholipids as phospholipids. Sphingomyelins are present in the plasma membranes of animal cells and are especially prominent in myelin, a membranous sheath that surrounds and insulates the axons of some neurons—thus the name “sphingomyelins.”

**Glycosphingolipids**, which occur largely in the outer face of plasma membranes, have head groups with one or more sugars connected directly to the OOH at C-1 of the ceramide moiety; they do not contain phosphate. **Cerebrosides** have a single sugar linked to ceramide;

those with galactose are characteristically found in the plasma membranes of cells in neural tissue, and those with glucose in the plasma membranes of cells in nonneural tissues. **Globosides** are neutral (uncharged) glycosphingolipids with two or more sugars, usually Dglucose, D-galactose, or *N*-acetyl-D-galactosamine. Cerebrosides and globosides are sometimes called **neutral glycolipids**, as they have no charge at pH 7.

**Gangliosides**, the most complex sphingolipids, have oligosaccharides as their polar head groups and one or more residues of *N*-acetylneuraminic acid (Neu5Ac), a sialic acid (often simply called “sialic acid”), at the termini.

The carbohydrate moieties of certain sphingolipids define the human blood groups and therefore determine the type of blood that individuals can safely receive in blood transfusions.

**Sterols** are structural lipids present in the membranes of most eukaryotic cells. The characteristic structure of this fifth group of membrane lipids is the steroid nucleus, consisting of four fused rings, three with six carbons and one with five. The steroid nucleus is almost planar and is relatively rigid; the fused rings do not allow rotation about COC bonds. **Cholesterol**, the major sterol in animal tissues, is amphipathic, with a polar head group (the hydroxyl group at C-3) and a nonpolar hydrocarbon body (the steroid nucleus and the hydrocarbon side chain at C-17), about as long as a 16- carbon fatty acid in its extended form. Similar sterols are found in other eukaryotes: stigmasterol in plants and ergosterol in fungi, for example.

# Biosignaling

The signal represents *information* that is detected by specific receptors and converted to a cellular response, which always involves a *chemical* process. This conversion of information into a chemical change, **signal transduction**, is a universal property of living cells.

## Features:

Signal transductions are remarkably specific and exquisitely sensitive. **Specificity** is achieved by precise molecular complementarity between the signal and receptor molecules. Multicellular organisms have an additional level of specificity, because the receptors for a given signal, or the intracellular targets of a given signal pathway, are present only in certain cell types.

Sensitivity of signal transducers: the high affinity of receptors for signal molecules, cooperativity and amplification of the signal by enzyme cascades.

**Cooperativity:** (recall the effect of cooperativity on oxygen binding to hemoglobin). **Amplification** results when an enzyme associated with a signal receptor is activated and, in turn, catalyzes the activation of many molecules of a second enzyme, each of which activates many molecules of a third enzyme, and so on, in a so-called enzyme cascade.

The sensitivity of receptor systems is subject to modification. When a signal is present continuously, **desensitization** of the receptor system results; when the stimulus falls below a certain threshold, the system again becomes sensitive.

**Integration**, the ability of the system to receive multiple signals and produce a unified response appropriate to the needs of the cell or organism.

1. G protein-coupled receptors that indirectly activate (through GTP-binding proteins, or G proteins) enzymes that generate intracellular second messengers. This type of receptors is illustrated by the  $\beta$ -adrenergic receptor system that detects epinephrine (adrenaline).
2. Receptor tyrosine kinases, plasma membrane receptors that are also enzymes, When one of these receptors is activated by its extracellular ligand, it catalyzes the phosphorylation of several cytosolic or plasma membrane proteins. The insulin receptor is one example.
3. Receptor guanylyl cyclases, which are also plasma membrane receptors with an enzymatic cytoplasmic domain. The intracellular second messenger for these receptors, cyclic guanosine monophosphate (cGMP), activated a cytosolic protein kinase that phosphorylates cellular proteins.
4. Gated ion channels of the plasma membrane that open and close (hence the term “gated”) in response to the binding of chemical ligands or changes in transmembrane potential. The acetylcholine receptor ion channel is an example.
5. Adhesion receptors that interact with macromolecular components of the extracellular matrix (such as collagen) and convey, instructions to the cytoskeletal system about cell migration or adherence to the matrix.

6. Nuclear receptors that bind specific ligands (such as the hormone estrogen) and alter the rate at which specific genes are transcribed and translated into cellular proteins.

G protein-coupled receptors (GPCRs) are receptors that are closely associated with a member of the guanosine nucleotide-binding protein (G protein) family. Three essential components define signal transduction through GPCRs. A plasma membrane receptor with seven transmembrane helical segments a G protein that cycles between active (GTP-bound) and inactive (GDP-bound) forms and an effector enzyme (or ion channel). The G protein, stimulated by the activated receptor, changes bound GDP for GTP, then dissociates from the occupied receptor and binds to the nearby effector enzyme, altering its activity. The activated enzyme then generates a second messenger.

Which mediates the effects of epinephrine, is the target of the “beta blockers”, “orphan receptors” their natural ligands are not yet identified.

**Updated (31/05/2016)**