

Bioenergetics

CONTENTS

- Coupling
- Concept of Energy
- Thermodynamic Principles
- Relationship Between Standard Free Energy Change and Equilibrium Constant
- Standard Free Energy Changes at pH 7.0 or ΔG°
- ATP as Universal Currency of Free Energy in Biological Systems
- Free Energy of Hydrolysis of ATP and other Organophosphates
- Structural Basis of the High Group Transfer Potential of ATP
- ATP Hydrolysis and Equilibria of Coupled Reactions
- Role of High Energy Phosphates as the 'Energy Currency' of the Cell
- Interconversion of Adenine Nucleotides



The potential energy of the water at the top of a waterfall is transformed into kinetic energy in spectacular fashion.

CHAPTER

20

Bioenergetics

COUPLING

Thermodynamics is the branch of physical chemistry that deals with the energy changes. And **biochemical thermodynamics** (or **biochemical energetics** or **bioenergetics**, as it is also called) is the field of biochemistry concerned with the transformation and use of energy by living cells. The chemical reactions occurring in living beings (or biochemical reactions) are associated with the liberation of energy as the reacting system moves from a higher to a lower energy level. Most often, the energy is liberated in the form of heat. In nonbiologic systems, heat energy may be transformed into mechanical or electrical energy. Since the biologic systems are isothermic, the heat energy cannot be used to drive the vital processes (such as synthesis, active transport, nerve conduction, muscular contraction etc.) obtain energy by chemical linkage (or coupling) to oxidation reactions. The simplest type of coupling (Fig. 20-1) may be represented by the equation.

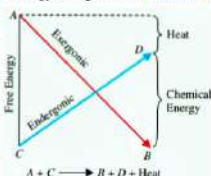


Fig. 20-1. Coupling of an exergonic to an endergonic reaction

$A + C \longrightarrow B + D + \text{Heat}$



BIOENERGETICS 435

The conversion of metabolite A to metabolite B occurs with the release of energy. It is coupled to another reaction, wherein energy is required to convert metabolite C to metabolite D.

CONCEPT OF ENERGY

Energy is defined as the capacity to do work, which is the product of a given force acting through a given distance:

$$\text{Work} = \text{Force} \times \text{Distance}$$

It is one of the fundamental components of any system. Energy exists in a variety of forms, such as electrical, mechanical, chemical, heat and light energy. These different forms of energy are interconvertible, for example, an electric motor converts electric into mechanical energy, a battery changes chemical into electrical energy and a steam engine transforms heat into mechanical energy. Some of the common examples of transformations of energy in biological systems are given in Fig. 20-2. Besides, these various forms of energy are also interrelated quantitatively, for example,

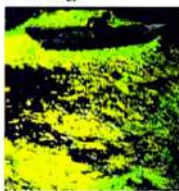
$$1.0 \text{ calorie of heat energy} = 4.185 \times 10^7 \text{ ergs of mechanical energy}$$



(a)



(b)



(c)

Fig. 20-2. Transformation of energy in biological systems

- (a) The running horse represents conversion of chemical energy to mechanical energy.
 (b) The electric fish (*Torpedinidae*) converts chemical energy to electrical energy.
 (c) The phosphorescent bacteria convert chemical energy into light energy.

Furthermore, during conversion of one form of energy to the other, there is always some loss. As an example, when an electric motor transforms electric into mechanical energy, the output of useful energy is always less than the input. This is due to friction in the motor which generates heat. The heat, in turn, is dissipated in the environment and is no longer useful. Thus, when a work is done or when one form of energy is changed to the other, there is a loss of useful energy.

THERMODYNAMIC PRINCIPLES

Many quantitative observations made on the interconversion of various forms of energy have led scientists to the formulation of two fundamental laws of thermodynamics, the first and second. These laws help us understanding:

- the direction of a reaction, whether from left to right or vice versa,
- the accomplishment of work, whether useful or not, and
- whether the energy for driving a reaction must be delivered from an external source.

The First Law : Principle of Conservation of Energy

In thermodynamics, a *system* is a matter within a defined region. The matter in the rest of the universe is called the *surroundings*. Thus, the system plus the surroundings constitute the universe (Fig. 20-3), which literally includes the entire earth, rather even the outer space. Some physical

or chemical processes can be made to take place in isolated systems which are unable to exchange energy with their surroundings. But in the biological world, the reacting system do exchange energy, and often matter also with their surroundings.

The principle of conservation of energy was first formulated by V. Mayer in 1841 as a result of a study of energy transformation in the inanimate world, but it is equally applicable to living systems. The first law of thermodynamics states that *the total amount of energy in the universe (i.e., the system + surroundings) remains constant*. Paraphrased, it says that energy cannot be created or destroyed. To date, there is no known exception to this law. Thus, whenever energy is used to do work or is converted from one to the other form, the total amount of energy is unchanged. The mathematical expression of the first law is :

$$\Delta E = E_B - E_A = Q - W$$

where,

ΔE = change in internal energy

E_A = energy of a system at the start of a process

E_B = energy of a system at the end of a process

Q = heat absorbed by the system

W = work done by the system.

A noteworthy point about equation 1 is that the change in energy of a system depends only on the initial and the final stages and not on the path of transformation.

The Second Law

The first law of thermodynamics cannot be used to predict whether a reaction can occur spontaneously, although ΔE is positive. In such cases, the system absorbs heat from its surroundings so that the sum of the energies of the system and its surroundings remains the constant. It is evident that a function other than ΔE is required. One such function is *entropy*, which is denoted by the symbol S . *Entropy is energy in a state of randomness or disorder*. It is unavailable, useless energy. In other words, entropy is a measure of the degree of randomness or disorder of a system. The entropy of a system increases (i.e., ΔS is positive) when it becomes more disordered. Entropy becomes maximum in a system as it approaches equilibrium. When equilibrium is attained, no further change can occur spontaneously unless additional energy is supplied from outside the system.

Thus, according to the second law, at each transfer or transformation of energy, part of that energy assumes a form that cannot be passed on any further because it is randomly dispersed, often as heat. And it is this randomly dispersed energy which is called *entropy*. Ultimately closed systems run down. Open systems, such as ecosystems, however, with their constant input and outflow, maintain a steady state in spite of the second law of thermodynamics. Biological systems, thus, do not seem to conform to the second law, for the tendency of life is to produce order out of disorder, to decrease rather than increase entropy.

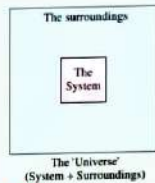


Fig. 20-3. Schematic representation of a reacting system and its surroundings
Energy can be exchanged between the system and the surroundings but only according to the laws of thermodynamics.

... (1)

The term **entropy** was first used in 1851 by Rudolf Clausius of Germany, one of the propounders of the second law of thermodynamics. The importance of entropy was pronounced by John A. Wheeler in following words : "No one is considered scientifically literate today who does not know what a Gaussian distribution is, or the meaning and scope of the concept of entropy."

BIOENERGETICS 437

Although it is true that energy is conserved in the universe, the first law of the thermodynamics does not account for many physical and chemical processes. A simple example may help make this clear. Suppose a full gas cylinder is connected to an empty cylinder by a tube with valve (Fig 20-4). If the valve is opened, gas flows from the full to the empty cylinder until the gas pressure is equal on both sides. Energy has not only been redistributed but also conserved. The expansion of gas is explained by the second law of thermodynamics and a condition of matter called *entropy* which, as already stated, is a measure of the randomness or disorder of a system. The greater the disorder of a system, the greater is its entropy. The second law states that physical and chemical processes proceed in such a way that the randomness or disorder of the universe (the system and its surroundings) increases to the maximum possible. Gas will always expand into an empty cylinder.

The second law of thermodynamics states that a process can occur spontaneously only if the sum of the entropies of the system and its surroundings increases. This can be represented as:

$$(\Delta S_{\text{system}} + \Delta S_{\text{surroundings}}) > 0 \quad \text{for a spontaneous process} \quad \dots (2)$$

Thus, the total entropy of a system must increase if a process is to occur spontaneously. However, the entropy of a system can decrease even during a spontaneous process, provided the entropy of the surroundings increases to such extent that their sum is positive. When water freezes, its entropy decreases because the water molecules of ice exist in a more ordered state with less freedom of movement than in a liquid state. The decrease of entropy is particularly apparent in the formation of a snowflake (Fig. 20-5).

The difference between the first and second laws of thermodynamics is that the *first law* is concerned with the accounting of the various kinds of energy involved in a given process, whereas the *second law* is concerned with the availability of the energy of a given system for doing useful work.

Combining the Two Laws

Since the entropy changes of chemical reactions are not readily measurable, the entropy is not used as a criterion whether a biochemical process can occur spontaneously or not. Furthermore, for spontaneity, both the entropy changes (that of the surroundings and of the system) should be known. These difficulties are alleviated by using a different thermodynamic function, called the *free energy*, which is denoted by the symbol G (In older literature, however, the symbol F is used in its place).

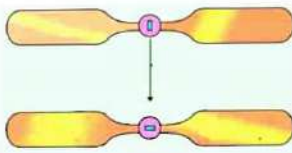


Fig. 20-4. A second law process

The expansion of gas into an empty cylinder simply redistributes the gas molecules until equilibrium is reached. The total number of molecules remains unchanged.



Fig. 20-5. When water freezes, its entropy decreases which becomes particularly apparent in the formation of a snowflake.

(Courtesy : Nurdiansy and Perennou)

438 FUNDAMENTALS OF BIOCHEMISTRY

In 1878, Gibbs created the free energy function by combining the first and second laws of thermodynamics in the form of following equation :

$$\Delta G = \Delta H - T\Delta S \quad \dots (3)$$

where, ΔG = the change in free energy of a reacting system,

ΔH = the change in heat content or enthalpy of this system,

T = the absolute temperature at which the process is taking

JOSIAH WILLARD GIBBS
(LT, 1839-1903)

Gibbs was an American physical chemist, who single-handedly created a large portion of chemical thermodynamics. The symbol G is given to free energy in his honour. His work is the basis of biochemical thermodynamics, and he is considered by some to have been the greatest scientist born in the United States.



(Courtesy: The Bettman Archive)

place, and

ΔS = the change in entropy of this system.

In fact, this equation represents quantitative interrelationship between the changes in free energy, heat and entropy in chemical reactions at constant temperature (T) and pressure (P), the conditions prevailing in biological systems.

The term $T\Delta S$ is that fraction of ΔH which cannot be put to useful work. The ΔG indicates the free energy change or the theoretically available useful work. Naturally, in most cases the system is 'inefficient' and not all of the theoretically available work can be utilized. However, the properties of the surroundings do not enter into this equation.

The enthalpy change, ΔH is given by the following equation :

$$\Delta H = \Delta E + P\Delta V \quad \dots (4)$$

where,

ΔE = the change in internal energy of a reaction, and

ΔV = the change in volume of this reaction.

As the volume change, ΔV is small for nearly all biochemical reactions, hence ΔH is nearly equal to the change in internal energy, ΔE . Therefore, equation 3 modifies as follows :

$$\Delta G = \Delta E - T\Delta S \quad \dots (5)$$

Thus, the change in free energy of a reaction, ΔG depends both on the change in internal energy and on the change in entropy of the system. The ΔG is a valuable criterion in determining whether a reaction can occur spontaneously. Thus,

- If ΔG is negative in sign, the reaction proceeds spontaneously with loss of free energy, i.e., it is **exergonic**. If, in addition, ΔG is of great magnitude, the reaction goes virtually to completion and is essentially irreversible.
- If, however, ΔG is positive, the reaction proceeds only if free energy can be gained, i.e., it is **endergonic**. If, in addition, ΔG is of high magnitude, the system is stable with little or no tendency for a reaction to occur.
- If ΔG is zero, the reaction system is at equilibrium and no net change takes place.

With regard to the free energy change, ΔG of a reacting system two more points need to be emphasized :

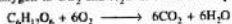
- Firstly, the ΔG of a reaction depends only on the free energy of the products minus that of the reactants. The ΔG of a reaction is independent of the path of transformation. Obviously, the mechanism of a reaction has no effect on ΔG . As an instance, the value

BIOENERGETICS 439

of ΔG is the same for the oxidation of glucose to CO_2 and H_2O whether it takes place by combustion or by a series of enzyme-catalyzed reactions.

- Secondly, the value of ΔG provides no information about the rate of a reaction. A negative ΔG indicates that a reaction can occur spontaneously, but it does not signify that it will occur at a perceptible rate. As already pointed out, the rate of a reaction rather depends on the free energy of activation ($\Delta G^\ddagger$), which is unrelated to ΔG .

To get a feeling for the magnitude of changes in various forms of energy, an actual example of aerobic oxidation of glucose may be cited. The living cells carry out oxidation of glucose in the presence of oxygen to CO_2 and H_2O at constant temperature and pressure.



Assuming that the temperature is 25°C (or 298 K) and the pressure is 1.0 atm (or 760 mm Hg), which are standard conditions in thermodynamic calculations, the following energy changes take place per molecule of glucose oxidized :

$$\Delta G = -686,000\text{ cal/mol} \quad (\text{i.e., the free energy of the reacting molecules has decreased})$$

$$\Delta H = -673,000\text{ cal/mol} \quad (\text{i.e., the reacting molecules have released heat})$$

The equation 3 may also be written as :

$$\Delta S = \frac{\Delta H - \Delta G}{T} \quad \dots (6)$$

Substituting the above values in equation 6, we get :

$$\begin{aligned} \Delta S &= \frac{-673,000 - (-686,000)}{298} \\ &= 44\text{ cal/deg} \quad (\text{i.e., the entropy of the universe has increased}) \end{aligned}$$

RELATIONSHIP BETWEEN STANDARD FREE ENERGY CHANGE AND EQUILIBRIUM CONSTANT

In a model reaction,



The free energy change, ΔG of this reaction is given by

$$\Delta G = \Delta G^\circ + RT \log \frac{[C][D]}{[A][B]} \quad \dots (8)$$

where,

ΔG° = Standard free energy change,

R = Gas constant,

T = Absolute temperature, and

$[A]$, $[B]$, $[C]$ and $[D]$ = Molar concentrations (i.e., activities) of the reactants

To elaborate, ΔG° is the free energy change for this reaction under standard conditions, i.e., when each of the reactants A , B , C and D is present at a concentration of 1.0 M . Thus, the ΔG of a reaction depends on the nature of the reactants (expressed in ΔG° term) and on their concentrations (expressed in logarithmic terms), as shown in equation 8.

At equilibrium,

$\Delta G = 0$. Equation 8 then becomes,

$$O = \Delta G^\circ + RT \log_e \frac{[C][D]}{[A][B]} \quad \dots(9)$$

$$\text{or} \quad \Delta G^\circ = -RT \log_e \frac{[C][D]}{[A][B]} \quad \dots(10)$$

The equilibrium constant under standard conditions, K'_{eq} for the reaction $A + B \rightleftharpoons C + D$, is given by

$$K'_{eq} = \frac{[C][D]}{[A][B]} \quad \dots(11)$$

However, in reactions in which more than one molecule of any reactant or product participates, the general reaction



where a , b , c and d are the number of molecules of A, B, C and D participating, the equilibrium constant is given by

$$K'_{eq} = \frac{[C]^c [D]^d}{[A]^a [B]^b} \quad \dots(12)$$

Substituting equation 11 into equation 10 gives

$$\Delta G^\circ = -RT \log_e K'_{eq} \quad \dots(13)$$

$$\text{or} \quad \Delta G^\circ = -2.303 RT \log_{10} K'_{eq} \quad \dots(14)$$

$$\text{or} \quad K'_{eq} = 10^{\Delta G^\circ / (2.303 RT)} \quad \dots(15)$$

Substituting $R = 1.98 \times 10^{-3} \text{ kcal mol}^{-1} \text{ degree}^{-1}$ and $T = 298^\circ \text{K}$ (corresponding to 25°C) gives

$$K'_{eq} = 10^{\Delta G^\circ / 1.36} \quad \dots(16)$$

when ΔG° is expressed in kcal/mol. Thus, the standard free energy, ΔG° and the equilibrium constant, K'_{eq} are related by a simple expression. For example, a change in equilibrium constant by a factor of 10 results in a change in standard free energy of -1.36 kcal/mol at 25°C (refer Table 20-1). At 37°C , however, the change in standard free energy would be of 1.42 kcal/mol . Values of ΔG° may be expressed in joules or calories per mole.

Table 20-1. Numerical relationship between equilibrium constants of chemical reactions and their standard free energy changes at 25°C

K'_{eq}	ΔG° (kcal/mol)
10^{-3} or 0.0001	+ 6.82
10^{-4} or 0.0001	+ 5.46
10^{-5} or 0.001	+ 4.09
10^{-6} or 0.01	+ 2.73
10^{-1} or 0.1	+ 1.36
1	0
10^1 or 10	- 1.36
10^2 or 100	- 2.73
10^3 or 1,000	- 4.09
10^4 or 10,000	- 5.46
10^5 or 1,00,000	- 6.82

ΔG° may also be defined as the difference between the free energy content of the reactants and that of the products under standard conditions (*i.e.*, 298 K pressure and 1.0 atmospheric pressure) when the reactants and products are present in their standard concentrations, namely 1.0 M. When ΔG° is negative (*i.e.*, the products contain less free energy than the reactants), the reaction will proceed to form the products under standard conditions, since all chemical reactions will tend to go in that direction resulting in decrease in the free energy of the system. When ΔG° is positive (*i.e.*, the products of the reaction contain more free energy than the reactants), the reaction will tend to go in the reverse direction if we start with 1 M concentration of all components. To be more explicit, reactions with a negative ΔG° proceed forward in the direction written when they start with all reactants and products at 1.0 M, until they reach equilibrium. Reactions with a positive ΔG° will proceed in the reverse of the direction written when they start with all components at 1.0 M. Table 20-2 summarizes these relationships.

Table 20-2. Relationship among K'_{eq} , ΔG° and the direction of chemical reactions under standard conditions

When K'_{eq} is :	ΔG° is :	Starting with 1 M components, the reaction :
> 1.0	Negative	proceeds forward
1.0	Zero	remains at equilibrium
< 1.0	Positive	proceeds backward

Two more points need be emphasized :

- Biochemical reactions take place near pH 7.0. Hence, pH 7.0 is conventionally designated the standard pH in biochemical energetics. The standard free energy change at pH 7.0 in biochemical energetics is designated by the symbol, $\Delta G'^\circ$, which shall be used in further discussions.
- The basic SI (Système International) unit of energy is joule (J). However, in biology and medicine, heat and energy measurements are equally expressed in calories (cal). The interrelationship between calories and joules is :

$$1,000 \text{ cal} = 4.184 \text{ J}$$

A **calorie (cal)** is equivalent to the amount of heat needed to raise the temperature of 1 gram of water from 14.5 to 15.5°C . A **kilocalorie (kcal)** is equal to 1,000 cal. Also, 1 kcal = 4.184 kJ.



JAMES PRESCOTT JOULE
(LT, 1818-1889)

James Prescott Joule was a British engineer who experimentally proved the first law of thermodynamics. He determined the mechanical equivalent of heat in 1843. The unit joule is named after him. A **joule (J)** is the amount of energy required to apply a 1 newton force over a distance of 1 metre. A **kilojoule (kJ)** is equal to 1,000 J. Faraday's belief in the unity of nature was vindicated in work by Joule, Thomson, Helmholtz, Clausius and Maxwell.

STANDARD FREE ENERGY CHANGES AT pH 7.0 OR $\Delta G'^\circ$

To calculate $\Delta G'^\circ$, an example of isomerization of dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (G-3-P) may be taken. This is one of the reactions of glycolysis.



At equilibrium, the ratio of glyceraldehyde 3-phosphate to dihydroxyacetone phosphate is 0.0475 at 25°C (298°K) and pH 7.0. Hence,

$$K'_{eq} = 0.0475$$

The standard free energy change for this reaction is, then, calculated from equation 14, as follows :

$$\begin{aligned}
 \Delta G^\circ &= -2.303 RT \log_{10} K'_{eq} \\
 &= -2.303 \times 1.98 \times 10^{-3} \times 298 \times \log_{10} (0.0475) \\
 &= +1.8 \text{ kcal/mol}
 \end{aligned}$$

When the initial concentration of DHAP is $2 \times 10^{-4} M$ and the initial concentration of G-3-P is $3 \times 10^{-6} M$, the ΔG can be calculated from equation 8, as follows :

$$\begin{aligned}
 \Delta G &= \Delta G^\circ + RT \log_e \frac{[C][D]}{[A][B]} \\
 &= 1.8 \text{ kcal/mol} + 2.303 RT \log_{10} \frac{3 \times 10^{-6} M}{2 \times 10^{-4} M} \\
 &= 1.8 \text{ kcal/mol} - 2.5 \text{ kcal/mol} \\
 &= -0.7 \text{ kcal/mol}
 \end{aligned}$$

The negative value for ΔG points out that isomerization of dihydroxyacetone phosphate to glyceraldehyde 3-phosphate can occur spontaneously, when these compounds are present at concentrations mentioned above. Although ΔG is negative for this reaction, the value of ΔG° is positive. It is important to note that the magnitude of ΔG for a reaction (whether smaller, larger or the same as ΔG°) depends on the concentrations of the reactants. The criterion of spontaneity for a reaction is ΔG and not ΔG° . Table 20-3 lists the standard free energy changes for some chemical reactions.

A perusal of the table (given on the next page) indicates that hydrolysis of esters, amides, peptides and glycosides, as well as rearrangements, and elimination proceed with relatively small standard free energy changes. Hydrolysis of acid anhydrides proceeds with relatively large decreases in standard free energy, whereas oxidation of organic compounds to CO_2 and H_2O takes place with rather huge decreases in standard free energy.

Difference Between ΔG and ΔG°

The difference between the free energy change, ΔG and the standard free energy change, ΔG° of a chemical reaction should be clearly understood. The actual free energy change, ΔG of a chemical reaction is a function of the conditions of concentration, pH and temperature under which the reaction is taking place. Moreover, the ΔG of an ongoing chemical process is always negative, becomes smaller (i.e., less negative) as the reaction proceeds and is zero at the point of equilibrium, indicating that no more work can be done by the reaction. The value of ΔG declines with time as the reaction proceeds because the actual concentrations of the reactants will be getting smaller and those of the resultants getting larger.

Table 20-3. Standard free energy changes for representative chemical reactions under standard conditions (i.e., at 25°C and pH 7.0)

Reaction Types and Reactions	ΔG° (in kcal/mol)
Elimination of Water	
Malate $\longrightarrow$ Fumarate + H_2O	+ 0.75
Rearrangements	
Glucose 1-phosphate $\longrightarrow$ Glucose 6-phosphate	-1.74
Fructose 6-phosphate $\longrightarrow$ Glucose 6-phosphate	-0.40
Hydrolysis	
Esters:	
Ethyl acetate + H_2O $\longrightarrow$ Ethanol + Acetate	-4.7
Glucose 6-phosphate + H_2O $\longrightarrow$ Glucose + Phosphate	-3.3
Amides and peptides:	
Glutamine + H_2O $\longrightarrow$ Glutamate + NH_4	-3.4
Glycylglycine + H_2O $\longrightarrow$ 2 Glycine	-2.2
Glycosides:	
Maltose + H_2O $\longrightarrow$ 2 Glucose	-3.7
Lactose + H_2O $\longrightarrow$ Glucose + Galactose	-3.8
Acid anhydrides:	
Acetic anhydride + H_2O $\longrightarrow$ 2 Acetate	-21.8
ATP + H_2O $\longrightarrow$ ADP + Phosphate	-7.3
Oxidations with Molecular Oxygen	
Glucose + 6O_2 $\longrightarrow$ 6CO_2 + $6\text{H}_2\text{O}$	-886
Palmitic acid + 23O_2 $\longrightarrow$ 16CO_2 + $16\text{H}_2\text{O}$	-2338

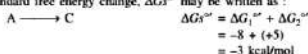
Thus, in every spontaneous chemical or physical process, the free energy of the reacting system always decreases, i.e., ΔG is negative. On the contrary, the value of standard free energy change, ΔG° for a chemical reaction is characteristic and unchanging, and may be positive, negative or zero, depending on the equilibrium constant of the reaction. ΔG° is, thus, an immutable constant and tells us in which direction and how far a given reaction will go in order to reach equilibrium when it occurs under standard conditions, i.e., when the initial concentration of all components is 1.0 M, pH is 7.0 and the temperature is 25°C.

Standard Free Energy Value of Chemical Reactions are Additive.

A noteworthy thermodynamic fact is that the overall free energy change for a series of reactions is equal to the sum of the free energy changes of the individual steps. Consider the two consecutive reactions where each reaction has its own equilibrium constant and a characteristic standard free energy change, ΔG_1° and ΔG_2° .



Since the two reactions are sequential, the intermediate product B cancels out and the overall reaction with its standard free energy change, ΔG_3° may be written as :

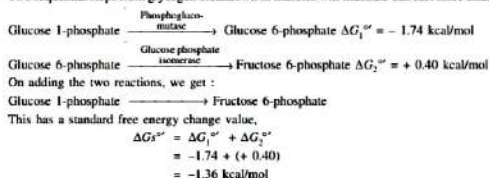


Under standard conditions, A can be spontaneously converted into B because ΔG is negative.

444 FUNDAMENTALS OF BIOCHEMISTRY

However, the conversion of *B* into *C*, under standard conditions is thermodynamically not feasible. But as the free energy changes are additive, the conversion of *A* into *C* has a ΔG° value of -3 kcal/mol, which obviously means that *A* can be converted into *C* spontaneously under standard conditions. Thus, the above two sequential reactions are coupled by the intermediate product, *B*. In other words, a *thermodynamically unfavourable reaction can be driven by a thermodynamically favourable reaction*.

Two sequential steps from glycogen breakdown in muscles will illustrate this fact more clearly.



Because ΔG_s° is negative, glucose 1-phosphate is converted into fructose 6-phosphate in the muscles.

ATP AS UNIVERSAL CURRENCY OF FREE ENERGY IN BIOLOGICAL SYSTEMS

The living objects require a continuous supply of free energy mainly for the following 4 purposes :

- to synthesize macromolecules from simpler and smaller precursors,
- to transport molecules and ions across membranes against gradients,
- to perform mechanical work, as in the muscle contraction, and
- to ensure fidelity of information transfer.

The free energy in these processes is derived from the environment. The phototrophs obtain this energy by trapping light energy from the sun. On the other hand, the chemotrophs obtain it by the oxidation of foodstuffs. This free energy (derived from light or from the oxidation of foodstuffs) is partly transformed into a special form before it is used for biosynthesis, transport, motion and fidelity. This special carrier of free energy is adenosine triphosphate (ATP). ATP plays a central role in the transference of free energy from the exergonic (= *energy-yielding*) to the endergonic (= *energy-requiring*) processes in the cells. During breakdown of energy-rich foodstuffs or fuel molecules, some of the free energy is harnessed to make ATP from adenosine diphosphate (ADP) and inorganic phosphate (P_i), a process that requires input of free energy. ATP then donates much of its chemical energy to energy-requiring processes (biosynthesis, transport etc.) by undergoing a breakdown to ADP and P_i (Fig. 20-6).

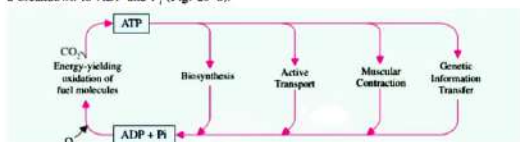


Fig. 20-6. The ATP cycle in cells

BIOENERGETICS 445

FRITZ ALBERT LIPMANN (LT, 1899–1986)



Lipmann was a German-born American biochemist. After obtaining M.D. degree, he started research career in 1927 as an unpaid graduate student under the famous biochemist Otto Meyerhof. He was awarded 1953 Nobel Prize in Medicine or Physiology for his discovery of coenzyme A and revealing its importance in intermediary metabolism, along with Sir Hans Adolf Krebs of Great Britain. Lipmann is popularly called as the 'father' of ATP cycle. He introduced the 'squiggle' notation (—) to designate the energy-rich bonds of biomolecules such as ATP and ADP. He was professionally active at Rockefeller University until his death at age 87. His famous quote (1949) reads as follows :
"It seems that in the field of biosynthesis we have a rare example of progress leading to simplification."

Adenosine triphosphate (ATP) was discovered in extracts of skeletal muscles by Karl Lohmann in Germany and by Cyrus Fiske and Yellapragada Subbarow in the United States, simultaneously in 1929. Later, it was found to be present in all types of cells— animal, plant and microbial. Although first thought to be concerned with muscular contraction only, ATP has now been

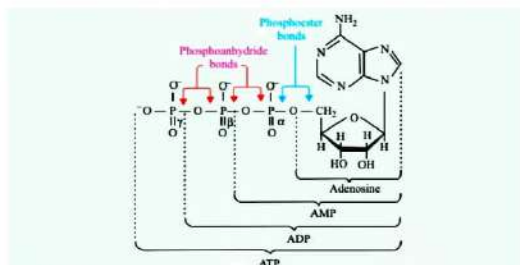


Fig. 20-7. The structure of ATP, indicating its relationship to ADP, AMP and adenosine

The phosphoryl groups of ATP, starting with that on AMP, are designated as α , β and γ phosphates. At pH 7.0, the phosphate groups are almost completely ionized, hence so shown here. The terminal phosphate group (*i.e.*, γ) can be enzymatically transferred to various phosphate acceptors. Note the differences between phosphoester and phosphoanhydride bonds.

assigned many cell activities. In 1941, Fritz A. Lipmann postulated that ATP is the primary and universal carrier of chemical energy in cells. He adequately stated that the small chemical units from which the material of organisms is built "are glued together by an enormously versatile condensing reagent, ATP." He also first proposed the ATP cycle, shown in a present-day form in Fig. 20-6.

Adenosine triphosphate (ATP) and its successive hydrolysis products, adenosine diphosphate (ADP) and adenosine monophosphate (AMP) are nucleotides, consisting of an adenine, a ribose and a 3, 2 or 1 phosphate group(s) respectively (Fig. 20-7). ATP, ADP and AMP occur not only in cell cytosol but also in mitochondria and the nucleus. In normal respiring cells, ATP makes up about 75% or more of the sum of all 3 adenine ribonucleotides (refer Table 20-4).

Table 20-4. Concentrations (in mM) of 3 adenine ribonucleotides, phosphate (Pi) and phosphocreatine (PCr) in some cells*

Cells type	ATP	ADP	AMP	Pi	PCr
Rat muscle	8.05	0.93	0.04	8.05	28
Rat liver	3.38	1.32	0.29	4.80	0
Rat brain	2.59	0.73	0.06	2.72	4.7
Escherichia coli	7.90	1.04	0.82	7.90	0
Human erythrocytes	2.25	0.25	0.02	1.65	0

* These data are for the entire cell contents except in the case of erythrocytes where the concentrations are those of cytosol, since erythrocytes are absent from nucleus and mitochondria both.

(Adapted from Lehninger AL, 1984)

At pH 7.0, the phosphate group(s) of ATP, ADP and AMP are almost fully ionized, so that they occur as multiple charged anions ATP^{4-} , ADP^{3-} and AMP^{2-} . But as the cell fluid contains high concentrations of Mg^{2+} , the ATP and ADP both exist largely as MgATP^{2-} and MgADP^{2-} complexes (Fig. 20-8). In fact, in phosphate transfer reactions, ATP participates as its complex form. However, ATP can also form complex with Mn^{2+} .

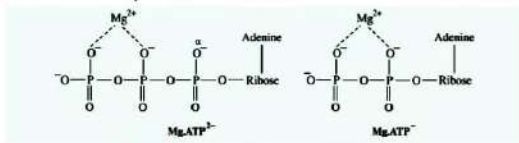


Fig. 20-8. Magnesium ion complexes of ATP and ADP

ATP serves as the principal immediate donor of free energy in biological systems rather than as a storage form of energy. In a typical cell, an ATP molecule is consumed within a minute of its formation. *The turnover of ATP is very high.* For instance, a resting human consumes about 40 kg ATP in a day. During strenuous labour, the ATP is consumed at the rate of even 0.5 kg per minute. The endergonic processes such as biosynthesis, active transport etc., can occur only if ATP is continuously regenerated from ADP (refer Fig. 20-9). Phototrophs harvest the free energy in light to regenerate ATP whereas chemotrophs form ATP by the oxidation of foodstuffs.



Fig. 20-9. The ATP-ADP cycle

The cycle is the fundamental mode of energy exchange in biological systems.

FREE ENERGY OF HYDROLYSIS OF ATP AND OTHER ORGANOPHOSPHATES

When ATP is hydrolyzed, it loses its terminal γ phosphate group to form ADP and orthophosphate or inorganic phosphate (Pi).



The standard free energy change, ΔG° for this reaction is -7.3 kcal/mol. Standard free energy changes have also been determined for the hydrolysis of other phosphorylated compounds or organophosphates (Table 20-5). Some phosphates yield more and some yield less free energy than ATP upon hydrolysis, under standard conditions. This intermediate position enables ATP to function efficiently as a carrier of phosphoryl groups. Thus, with respect to the ΔG° value of hydrolysis of ATP, two classes of organo phosphates are recognized : high energy phosphates exemplified by enolphosphates (e.g., phosphoenolpyruvate), phosphoguanidines (e.g., creatine phosphate and arginine phosphate) etc., which have ΔG° values larger than that of ATP and low energy phosphates, exemplified by ester phosphates found in the intermediates of glycolysis, which have ΔG° values smaller than that of ATP. However, the designations 'high' and 'low' do not clearly indicate that there are 3 classes of phosphates. Such phosphates as phosphoenolpyruvate, creatine phosphate etc., whose ΔG° values are higher than that of ATP, should better be designated as 'super' high energy phosphates, the ATP then bedesignatedas 'high' energyphosphate and theester phosphatesas 'low' energy phosphates.

Table 20-5. Standard free energy of hydrolysis of some phosphorylated compounds

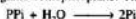
Compound	ΔG° (kcal/mol)
1. Phosphoenolpyruvate*	-14.8
2. Carbanoyl phosphate	-12.3
3. 3-phosphoglyceroyl phosphate or 1, 3-diphosphoglycerate ($\rightarrow$ 3-phosphoglycerate + P_i)	-11.8
4. Creatine phosphate or phosphocreatine	-10.3
5. Acetyl phosphate	-10.1
6. Arginine phosphate	-8.0
7. ATP ($\rightarrow$ AMP + P_i)	-7.7
8. ATP ($\rightarrow$ ADP + P_i)	-7.3
9. ADP ($\rightarrow$ AMP + P_i)	-7.3
10. Pyrophosphate ($\rightarrow$ 2 P_i)	-6.9
11. Glucose 1-phosphate	-5.0
12. Fructose 6-phosphate	-3.8
13. AMP ($\rightarrow$ Adenosine + P_i)	-3.4
14. Glucose 6-phosphate	-3.3
15. Glycerol 3-phosphate	-2.2

*Note that phosphoenolpyruvate has the highest phosphate group transfer potential of the compounds listed.

A perusal of Table 20-5 indicates that when ADP is hydrolyzed to AMP and inorganic phosphate, the ΔG° value of this reaction is the same as that of the reaction, $\text{ATP} \longrightarrow \text{ADP} + \text{P}_i$, that is -7.3 kcal/mol. Thus, the two terminal phosphate groups of ATP (β and γ) are both high energy groups. On the contrary, the ΔG° value of hydrolysis of AMP to yield adenosine and phosphate is much lower, that is only -3.4 kcal/mol. Thus, the phosphate group of AMP (i.e., the α phosphate group of ATP) is in the low energy class. The hydrolysis of ATP to yield AMP plus P_i proceeds with ΔG° -7.7 kcal/mol, slightly greater than the ΔG° for the hydrolysis of the terminal or γ phosphate bond.

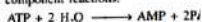


The inorganic pyrophosphate is subsequently hydrolyzed by the enzyme pyrophosphatase to yield 2 moles of inorganic orthophosphate. The ΔG° value of this reaction is -6.9 kcal/mol.



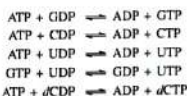
448 FUNDAMENTALS OF BIOCHEMISTRY

The overall reaction has a ΔG° value -14.6 kcal/mol which is the sum of the ΔG° values of the two sequential component reactions.



It is noteworthy that the ΔG° of the overall reaction is exactly twice the ΔG° of the terminal phosphate groups of ATP and ADP.

Some biosynthetic reactions are, however, driven by nucleotides that are analogous to ATP, namely guanosine triphosphate (GTP), cytidine triphosphate (CTP) and uridine triphosphate (UTP). They are present in all cells but in much lower concentrations than ATP. Also found in the cells in low concentrations are the corresponding deoxyribonucleoside 5'-triphosphates denoted as dATP, dGTP, dCTP and dTTP. Although ATP is the mainstream carrier of phosphate groups in the cell, the other types of nucleotides serve certain specific biosynthetic pathways. They acquire their terminal phosphate groups mainly from ATP in reactions catalyzed by Mg^{2+} -dependent enzymes called *nucleoside diphosphokinases*. These enzymes promote the following types of reversible reactions:



The energy contained in ATP may be channelized into different biosynthetic pathways via different nucleoside and deoxynucleoside triphosphates (refer Fig. 20-10).

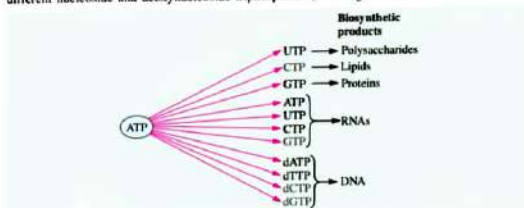


Fig. 20-10. Channeling of ATP energy into various biosynthetic pathways.

STRUCTURAL BASIS OF THE HIGH GROUP TRANSFER POTENTIAL OF ATP

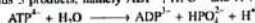
The ATP molecules deliver considerably more free energy on hydrolysis of the terminal phosphate group than the hydrolysis of ester phosphates, say glucose 6-phosphate. The ΔG° values of both these reactions are -7.3 kcal/mol and -3.3 kcal/mol respectively (refer Table 20-5). This means that ATP has a stronger tendency to transfer its terminal phosphate group to water than does glucose 6-phosphate. In other words, *ATP has a higher phosphate group transfer potential than glucose 6-phosphate*.

The standard free energy change is a measure of the difference in the free energy content of the reactants and the products. Therefore, the reason for the high ΔG° value for ATP should be sought in the properties of both the substrate and products. There are 3 main reasons of high ΔG°

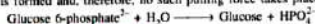
BIOENERGETICS 449

value (or high group transfer potential) for ATP:

1. **Degree of ionization.** At pH 7.0, ATP is almost completely ionized as the ATP^{4-} ion. This, on hydrolysis, yields 3 products, namely ADP^{3-} , HPO_4^{2-} and H^+ .



At pH 7.0 (the standard pH for ΔG° hydrolysis), the hydrogen ion concentration is only 10^{-7} M. This means that by the law of mass action, the equilibrium for ATP hydrolysis tends to be pulled far to the right, since $[\text{H}^+]$ at pH 7.0 is very small as compared with the standard concentration of 1.0 M. In contrast, when glucose 6-phosphate is hydrolyzed at pH 7.0, no significant extra H^+ ion is formed and, therefore, no such pulling force takes place.



2. **Electrostatic repulsion.** At pH 7.0, the ATP molecule carries 4 negative charges. These charges repel each other strongly because they are in close proximity. However, when ATP is hydrolyzed, the electrostatic repulsion between these negatively-charged groups is reduced because of the separation of the negatively-charged products, ADP^{3-} and HPO_4^{2-} . These products have relatively little tendency to approach each other to react in the backward direction to form ATP again. On the contrary, when glucose 6-phosphate is hydrolyzed, two products glucose and HPO_4^{2-} are formed. Because glucose has no charge, the two products do not repel each other and tend to recombine more readily.

3. **Resonance stabilization.** The two products of ATP hydrolysis, ADP^{3-} and HPO_4^{2-} are resonance hybrids, i.e., special stable forms in which certain electrons are in a configuration having much less energy than in their original positions in the ATP molecule. When ATP is hydrolyzed, the electrons in the two products (ADP^{3-} and HPO_4^{2-}) can sink to lower energy levels than in the nonhydrolyzed ATP. This causes ADP^{3-} and HPO_4^{2-} ions to contain less free energy than when they were still combined as ATP^{4-} .

ADP and P_i enjoy greater resonance stabilization than does ATP. For example, orthophosphate (P_i) has a number of resonance forms of similar energy (refer Fig. 20-11), whereas the terminal phosphate portion of ATP has fewer resonance forms per phosphate group.

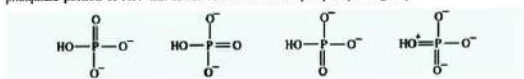


Fig. 20-11. Significant resonance forms of orthophosphate

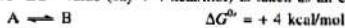
To indicate the presence of high energy phosphate group, Lipmann introduced the symbol $\sim\text{P}$ indicating high energy phosphate bond. Although in long usage among biochemists, the term '*high energy phosphate bond*' is misleading since it wrongly suggests that the bond itself contains high amount of energy. What the symbol $\sim\text{P}$ indicates is that the group attached to the bond, on transfer to an appropriate acceptor, results in transfer of large amount of free energy. Hence, the term '*group transfer potential*' is preferred to '*high energy bond*'. In fact, the free energy released by hydrolysis of phosphate esters does not come from the specific bond that is broken but results from the fact that the products of the reaction have a smaller free energy content than the reactants. Therefore, the use of the term '*high energy phosphate compound*' be preferred when referring to ATP and other phosphate compounds having a large negative ΔG° value of hydrolysis. Thus, whereas ATP contains two high energy phosphate groups and one low energy phosphate group, ADP contains one high energy and one low energy phosphate groups. AMP, however, contains only one phosphate group of low energy type (Fig. 20-12).

Adenosine -P-P-P ATP	Adenosine -P-P ADP	Adenosine -P AMP
-------------------------	-----------------------	---------------------

Fig. 20-12. Structure of adenine nucleotides, showing the position and number of high energy bonds (~)

ATP HYDROLYSIS AND EQUILIBRIA OF COUPLED REACTIONS

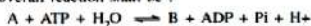
For a better understanding of the role of ATP in energy coupling, the following reaction with a positive $\Delta G^{\circ'}$ value (say +4 kcal/mol) is taken as an example :



This reaction is thermodynamically unfavourable without an input of free energy. The equilibrium constant, K'_{eq} of this reaction at 25°C as per equation 16 is :

$$\frac{[B]_{eq}}{[A]_{eq}} = K'_{eq} = 10^{-\Delta G^{\circ'}/1.36} = 1.15 \times 10^{-3}$$

Thus, A cannot be spontaneously converted into B when the molar ratio of B/A is equal to or greater than 1.15×10^{-3} . However, if the reaction is coupled to the hydrolysis of ATP (whose $\Delta G^{\circ'}$ value is -7.3 kcal/mol), A can be converted into B when $[B]/[A]$ ratio is higher than 1.15×10^{-3} . The new overall reaction shall be :



Because the $\Delta G^{\circ'}$ values of the sequential reactions are additive, the value of standard free energy change of the above reaction shall be :

$$\begin{aligned} \Delta G^{\circ'} &= \Delta G_1^{\circ'} + \Delta G_2^{\circ'} \\ &= +4 + (-7.3) \\ &= -3.3 \text{ kcal/mol} \end{aligned}$$

The equilibrium constant of this coupled reaction will be :

$$\begin{aligned} K'_{eq} &= \frac{[B]_{eq}}{[A]_{eq}} \times \frac{[\text{ADP}]_{eq} [\text{P}_i]_{eq}}{[\text{ATP}]_{eq}} \\ &= 10^{3.3/1.36} \\ &= 2.67 \times 10^2 \end{aligned}$$

At equilibrium, the ratio $[B]/[A]$ will be :

$$\frac{[B]_{eq}}{[A]_{eq}} = K'_{eq} \times \frac{[\text{ATP}]_{eq}}{[\text{ADP}]_{eq} [\text{P}_i]_{eq}}$$

The $[\text{ATP}]/[\text{ADP}] [\text{P}_i]$ ratio in the cells generating ATP is maintained at a high level, typically of the order of 500. Thus,

$$\begin{aligned} \frac{[B]_{eq}}{[A]_{eq}} &= 2.67 \times 10^2 \times 500 \\ &= 1.34 \times 10^5 \\ &= 10^5 \text{ approx.} \end{aligned}$$

This means that the hydrolysis of ATP enables A to be converted to B until the $[B]/[A]$ ratio reaches a value of 1.34×10^5 . This equilibrium is quite different from the value of 1.15×10^{-3} that does not include ATP hydrolysis. In other words, the coupled hydrolysis of ATP has changed the equilibrium ratio of B to A by a factor of about 10^8 . Thus, the cells maintain a high level of ATP by using light or substrates as sources of free energy. The hydrolysis of one ATP molecule

Contents

BIOENERGETICS 451

then changes the equilibrium ratio of products to reactants of a coupled reaction by a very large factor of about 10^8 . More generally, the hydrolysis of n ATP molecules changes the equilibrium ratio of a coupled reaction (or sequence of reactions) by factor of 10^{8n} . For example, the hydrolysis of 3 ATP molecules in a coupled reaction changes the equilibrium ratio by a factor of 10^{24} or 10^{25} . Thus, a thermodynamically unfavourable reaction sequence can be converted into a favourable one by coupling it to the hydrolysis of a sufficient number of ATP molecules.

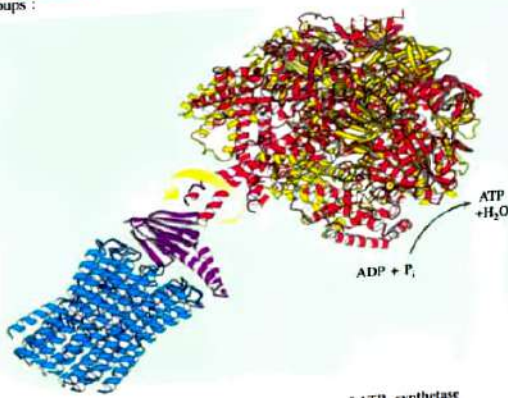
ROLE OF HIGH ENERGY PHOSPHATES AS THE 'ENERGY CURRENCY' OF THE CELL

ATP has a position midway down the list of standard free energies of hydrolysis of organophosphates (refer Table 20-5). Therefore, ATP acts as a donor of high energy phosphates to the compounds below it in the table.

Likewise, ADP can accept high energy phosphate, in the presence of enzyme, to form ATP from the compounds above ATP in the table.

Thus, an ATP/ADP cycle connects these two types of processes, i.e., those which generate -P and those which utilize -P.

A. Processes which generate -P. The processes that feed -P into ATP/ADP cycle fall mainly within 4 groups :



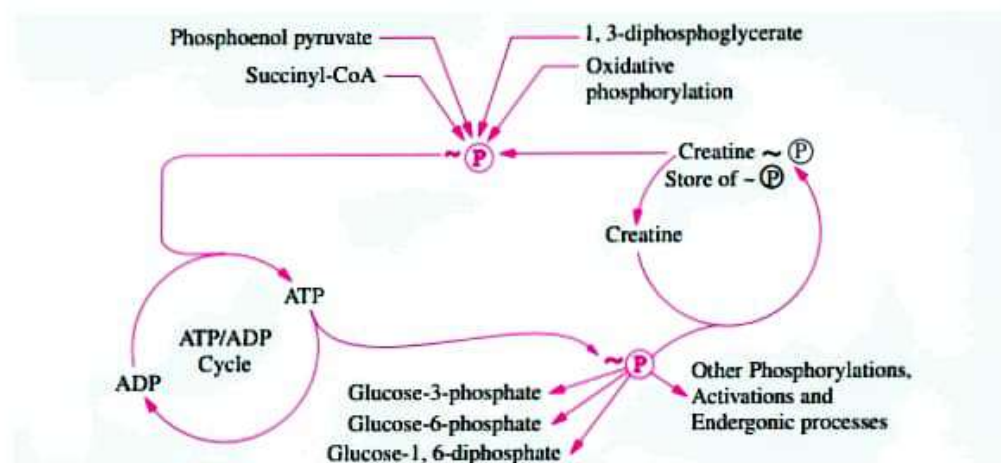


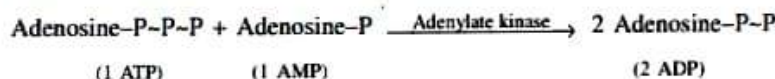
Fig. 20-14. Role of ATP/ADP cycle in transfer of high energy phosphate

An important feature of this cycle is that all super high-energy phosphate compounds must pass their phosphate groups via ATP to various acceptor molecules to form their low energy phosphate derivatives.

(Adapted from Harper, Rodwell and Mayes, 1979)

INTERCONVERSION OF ADENINE NUCLEOTIDES

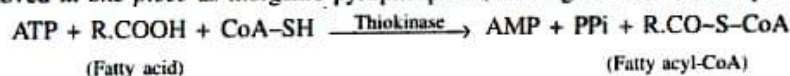
The enzyme adenylate kinase (= myokinase) is present in most cells. It catalyzes the interconversion of ATP and AMP on the one hand and ADP on the other.



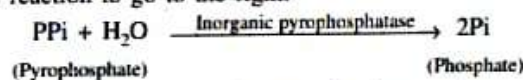
This reaction performs 3 main functions :

- It permits the high energy phosphate of ADP to be used in the formation of ATP.
- It is a means by which AMP can be rephosphorylated to form ADP.
- Under stress of ATP depletion, AMP concentration increases which acts as a metabolic signal to increase the rate of catabolic reactions so that more ATP molecules are regenerated.

When ATP reacts enzymatically to form AMP, inorganic pyrophosphate (symbolized PPi) is formed. An example is the enzymatic activation of a fatty acid to form its coenzyme A ester, a reaction that 'energizes' the fatty acid and converts it into an activated precursor, fatty acyl-CoA, for the biosynthesis of lipids. In this reaction, the two terminal groups of ATP (*i.e.*, β and γ groups) are removed in *one piece* as inorganic pyrophosphate, leaving adenosine monophosphate.



Thus, this activation reaction proceeds by a pyrophosphate cleavage of ATP, in contrast to the usual orthophosphate cleavage (refer page 362). This reaction, whose $\Delta G^{\circ'}$ value is + 0.2 kcal/mol, results in a loss of free energy as heat which ensures that the reaction will proceed to the right as written. Further, the hydrolytic splitting of PPi has also a large $\Delta G^{\circ'}$ value of -6.9 kcal/mol. This also aids the above reaction to go to the right.



It is noteworthy that activations via the pyrophosphate pathway result in the loss of 2 high energy phosphates ($\sim\text{P}$) rather than one, as happens when ADP and P_i are formed. The various interrelationships are represented cyclically in Fig. 20-15.

Carbohydrates

- Importance
- Nomenclature and Definition
- Classification
- Asymmetry
- Isomerism
- Kiliani Cyanohydrin Synthesis
- Optical Isomerism
- Formulation of Monosaccharides

Carbohydrates-I

General Structure of Monosaccharides

IMPORTANCE



Crystals of **glucose**, a key molecule in carbohydrate metabolism, viewed under polarised light.

The carbohydrates, often termed as sugars, are the 'staff of life' for most organisms. On the basis of mass, they are the most abundant class of biomolecules in nature. Carbohydrates are also known as *saccharides* (*sakcharon*^G = sugar or sweetness) since many of those of relatively small molecular weight have a sweet taste, although this is not true of those with large molecules. They are widely distributed molecules (*moles*^L = mass) in both plant and animal tissues. They are indispensable for living organisms, serving as skeletal structures in plants and also in insects and crustaceans. They also occur as food reserves in the storage organs of plants and in the liver and muscles of animals. In addition, they are an important source of energy required for the various metabolic activities of the living organisms; the energy being derived as a result of their oxidation. They also serve to lubricate skeletal joints, to provide adhesion between cells and to confer biological specificity on the surface of animal cells.

Plants are considerably richer in carbohydrates in comparison to the animals. In fact, animal and plant tissues differ widely in the relative abundance of the various major classes of constituent chemicals (Table 5-1).

74 FUNDAMENTALS OF BIOCHEMISTRY

Table 5-1. Percentage composition of various major classes of constituent chemicals in plants and animals

Organism	Water	Carbohydrates	Proteins	Lipids	Ash
Animal	60	1	20	15	4
Plant	60	30	5	1	4

NOMENCLATURE AND DEFINITION

The term 'carbohydrate' was originally coined for this class of compounds as most of them were 'hydrates of carbon' or could be represented by the general formula, $C_n(H_2O)_n$. Later, it was found that some of them, such as deoxyribose ($C_5H_{10}O_4$) and rhamnose ($C_6H_{12}O_5$) do not have the required ratio of hydrogen to oxygen. In addition, certain other carbohydrates are now known to possess nitrogen (e.g., glucosamine, $C_6H_{13}O_5N$), phosphorus or sulfur also and obviously do not coincide with the above general formula. Furthermore, formaldehyde (H_2CHO or CH_2O), acetic acid (CH_3COOH or $C_2H_4O_2$) and lactic acid ($CH_3CHOH.COOH$ or $C_3H_6O_3$) which have C, H and O and the ratio of H : O is also the same as in water, but are not carbohydrates. Hence, the continued usage of the term 'carbohydrate' is for convenience rather than exactness.

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

CLASSIFICATION

Carbohydrates are usually classified in 3 groups :

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

The monosaccharides may be subdivided into trioses, tetroses, pentoses, hexoses, heptoses etc., depending upon the number of carbon atoms they possess; and as aldoses or ketoses, depending upon whether they contain aldehyde or ketone group. Some important examples are :

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

Both these characters (i.e., the number of carbon atoms and the nature of functional group present) may also be combined into one. Thus, for example, **glycerose** (= glyceraldehyde) is an aldotriose; **ribulose**, a ketopentose and **glucose**, an aldohexose. It is noteworthy that, *except fructose, ketoses are not as common as aldoses*. The most abundant monosaccharide in nature is the 6-carbon sugar, D-glucose.

Sometimes, a distinction in naming between aldoses and ketoses is also maintained. The suffix

-oses is kept reserved for the aldoses and the suffix -uloses is used for ketoses. Thus, glucose is a hexose and fructose, a hexulose. However, a few ketoses are named otherwise, such as fructose (*fructus* = fruit) as fruits are a good source of this sugar.

B. Oligosaccharides or Oligosaccharoses (*oligo*^G = few). These are compound sugars that yield 2 to 10 molecules of the same or different monosaccharides on hydrolysis. Accordingly, an oligosaccharide yielding 2 molecules of monosaccharide on hydrolysis is designated as a disaccharide, and the one yielding 3 molecules of monosaccharide as a trisaccharide and so on. The general formula of disaccharides is $C_n(H_2O)_{n-1}$ and that of trisaccharides is $C_n(H_2O)_{n-2}$ and so on. A few examples are :

- | | |
|-------------------------|--|
| Disaccharides | - Sucrose, Lactose, Maltose, Cellobiose, Trehalose, Gentiobiose, Melibiose |
| Trisaccharides | - Rhamnose, Gentianose, Raffinose (= Melitose), Rabinose, Melezitose |
| Tetrasaccharides | - Stachyose, Scorodose |
| Pentasaccharide | - Verbascose |

The molecular composition of the 3 legume oligosaccharides (*viz.*, raffinose, stachyose, verbascose) is shown below :

α -Galactose (1-6) α -Glucose (1-2) β -Fructose

α -Galactose (1-6) α -Galactose (1-6) α -Glucose (1-2) β -Fructose

α -Galactose (1-6) α -Galactose (1-6) α -Glucose (1-6) α -Glucose (1-2) β -Fructose Verbascose

C. Polysaccharides or Polysaccharoses (*poly*^G = many). These are also compound sugars and yield more than 10 molecules of monosaccharides on hydrolysis. These may be further classified depending on whether the monosaccharide molecules produced as a result of the hydrolysis of polysaccharides are of the same type (homopolysaccharides) or of different types (heteropolysaccharides). Their general formula is $(C_6H_{10}O_5)_n$. Some common examples are :

- | | |
|------------------------------|---|
| Homopolysaccharides | - Starch, Glycogen, Inulin, Cellulose, Pectin, Chitin |
| Heteropolysaccharides | - "Specific soluble sugar" of pneumococcus type III, Hyaluronic acid, Chondroitin |

ASYMMETRY

Jacobus H. van't Hoff, the first Nobel Laureate in Chemistry (1901) and Joseph A. Le Bel, in 1894, introduced the concept of tetrahedral carbon atom.

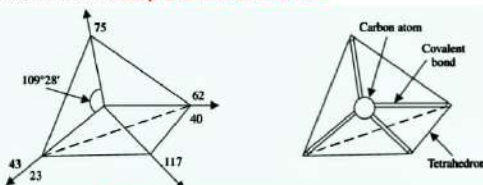


Fig. 5-1. Spatial arrangement of the valences of carbon

100 / 1254

Contents

76 FUNDAMENTALS OF BIOCHEMISTRY

JACOBUS HENRICUS VAN'T HOFF

(LT, 1852-1911)



A Dutch chemist. Best known for his hypothesis about the structure of carbon atom. Made notable contributions in the field of Physical Chemistry, particularly on the theory of dilute solutions. Won the first-ever Noble Prize in Chemistry in 1901 for work on rates of reactions, chemical equilibrium and osmotic pressure.

JOSEPH A. LE BEL

(LT, 1847-1930)



A French chemist. Best known for his stereochemical theory of carbon and its compounds. Co-discoverer of the cause of optical activity with van't Hoff. Synthesised the first optically-active compound with asymmetric nitrogen atom.

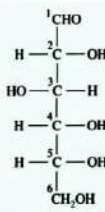
It is now recognized that the carbon atom has the shape of a tetrahedron in which the carbon nucleus resides in the centre of this tetrahedron and the 4 covalent bonds extend out to the corners of it. The angle between any two covalent bonds is $109^{\circ}28'$ (Fig. 5-1).

The 5th Incarnation of Carbon

Carbon is one of the most important elements in the universe since *life itself*, as we know it, is *carbon-based*. The versatility of water to form a huge variety of compounds and their stability even under extreme conditions are the secret of carbon acting as the vehicle of life. Versatility of carbon compounds apart, the element itself has become a centre of great attention in recent years. For long, we knew carbon could manifest itself in vividly contrasting styles. It was known to assume a soft and black form as **graphite**; it was known equally well in its hard and sparkling - white incarnation of **diamond** which is the hardest material on earth. Then it was discovered (in 1985) to have had the shape of a soccer ball (called **fullerenes**), and of a cylindrical tube (called **nanotubes**). And now an announcement has been made that carbon can also assume a foamy shape, called **nanof foam**.

This new spongy form of carbon is also a modification of graphitic form. It is extremely light-weighted form and was created when some researchers, at the Australian National University in Canberra, bombarded a piece of graphite with a high-power laser beam. The graphite piece vapourised as temperature rose to more than $10,000^{\circ}\text{C}$. On cooling, the vapourised carbon atoms formed a web of tubes. But these tubes instead of being aligned in any particular direction (as happens in the case of nanotubes), were found to be highly dispersed, criss-crossing with one another to form a highly-porous mass. This web of nanotubes is seen as a nanof foam. Surprisingly, the nanof foam, unlike all the other forms of carbon, is *magnetic*. Although the magnetism is ephemeral at room temperature, it is an entirely new and unexpected property in the case of carbon. Because of light weight, foamy nature and magnetic property, nanof foam is being seen as useful for cancer and brain scans through magnetic resonance imaging (MRI). Besides, it might also be useful in the treatment of tumours. Being foam, it would be a bad conductor of heat and if injected into tumours, and heated, it would retain the heat locally and kill tumours almost without affecting other body parts. But it needs confirmation through a series of studies.

In the formula given above for glucose, it may be observed that a different group is attached to each of the 4 valences of carbon atoms 2 to 5. For example, the 4 groups attached at carbon atom 2 are :



Glucose

[The carbon atoms of a sugar are numbered from the end closest to the aldehyde or ketone group.]

Thus, a carbon atom to which 4 different atoms or groups of atoms are attached is said to be **asymmetric** or **chiral** (*chir*^G = hand).

ISOMERISM

The term **isomer** (*isos*^G = equal ; *meros*^G = part) was originally applied by Jöns Jacob Berzelius, in 1827, to different compounds with the same molecular formula, and the phenomenon was called **isomerism**. The presence of asymmetric carbon atoms in the carbohydrates makes possible the formation of isomers in them. It shall, hence, be beneficial at this stage to discuss in brief about isomers.

The isomers are of 2 types : structural isomers and stereoisomers. **Structural isomers** have the same molecular formula but possess different structures. The difference in structure may be exhibited either in the length of the carbon chain (**chain isomers**) or in the position of a substituent group (**positional isomers**) or in possessing different functional groups (**functional-group isomers**).

Stereoisomers, on the other hand, have the same molecular formula and the same structure but differ only in spatial configuration. Stereoisomers are of 2 types : geometrical and optical. **Geometrical** or **cis-trans** (*cis*^T = same side; *trans*^T = across) **isomers** arise from peculiar geometry of compounds having a double bond within the carbon chain. It may be illustrated by *cis-trans* pair, maleic and fumaric acids.



Optical isomers (= enantiomers) differ from each other in the disposition of the various atoms or groups of atoms in space around the asymmetric carbon atom. These are, in fact, the mirror image of each other. These may also be likened to left- and right-handed gloves. For example, the glyceraldehyde has only one asymmetric carbon atom (numbered as 2) and it can, therefore, exist in 2 isomeric forms :



One form in which H atom at carbon 2 is projected to the left side and OH group to the right is designated as **D-form** and the other form where H atom is projected to the right side and OH group to the left is called as **L-form** (note the use of small capital letters D and L). Similarly, a compound having 2 asymmetric carbon atoms (e.g., a tetrose) might exist in 4 optical isomeric forms.

The number of possible isomers of any given compound, thus, depends upon the number of asymmetric carbon atoms present in its molecule. According to the **rule of n** (also called as

78 FUNDAMENTALS OF BIOCHEMISTRY

Le Bel-van't Hoff rule), the total number of optical isomers of a compound will be equal to 2^n , where n represents the number of asymmetric carbon atoms present in the molecule.

In fact, in sugars with 2 or more asymmetric carbon atoms, the designation of D- and L- forms depends upon the orientation of the H and OH groups around the lowermost asymmetric carbon atom (i.e., the asymmetric carbon furthest from the aldehyde or keto group). The distribution of the H and OH groups on the other carbon atoms in the molecule is of no importance in this connection. It is interesting to note that *majority of the monosaccharides found in the human body are of D type*. However, some sugars do occur naturally in their L form such as L-arabinose, a pentaldose.

KILIANI CYANOHYDRIN SYNTHESIS

A method for the synthesis of monosaccharides was first proposed by Heinrich Kiliani in 1886. It is, in fact, a method by which the chain length of a carbohydrate may be increased. The application of Kiliani synthesis to D-glyceraldehyde resulting in the production of 2 tetroses, D-erythrose and D-threose is shown in Fig. 5-2.

The process is based upon the addition of HCN to the carbonyl group of aldehydes (or ketones) of the sugars forming cyanohydrin (Reaction-1). This reaction creates a new asymmetric carbon atom, marked by an asterisk in the figure. Thus, two compounds differing in conformation about the newly-formed asymmetric carbon atom are formed. These two cyanohydrins are, then, hydrolyzed to produce carboxylic acids (Reaction-2) which are later converted to γ -lactones or inner esters (Reaction-3). Finally, the lactones are reduced to the corresponding aldoses, containing one carbon atom more than their parent sugar (Reaction-4).

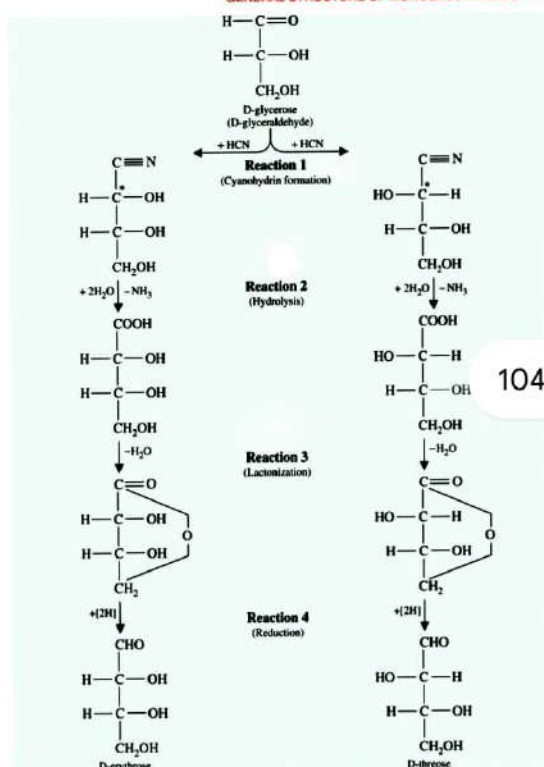
The process can be repeated and the 4 isomeric D-pentoses may be produced; and from these the 8 isomeric D-hexoses would also result.

The structural relationships of the monosaccharides (the aldoses, for example) may be visualized by the formulae given in Fig. 5-3. It is easy to remember the configurations of the aldoses by recalling the **Rosanoff scheme**, drawn out on the chart and mnemonic—ET; RAXL; AGMGIGT.

Any two sugars which differ from each other only in the configuration around a single asymmetric carbon atom other than the carbonyl carbon atom are called **epimers**. Glucose and galactose, for example, form an epimeric pair as they differ with respect to carbon 4 only. Similarly, glucose is also epimeric with mannose (differing in C_2 configuration) and allose with altrose (differing also in C_2 configuration).

The two D- and L-forms of a compound constitute a pair of **enantiomers** (*enantios*^G = opposite) or **enantiomorphs**. Thus, D-erythrose and L-erythrose are enantiomers; also D-glucose and L-glucose. The enantiomers (also called optical isomers or double image isomers) are nonsuperimposable mirror images of each other but are chemically identical in their reactions. They agree in their melting points, solubility etc., but differ in their ability to rotate the plane of polarized light in a polarimeter; a solution of one of the two enantiomers rotates the plane to the right, and a solution of the other to the left.

The term '**conformation**' was first introduced by Walter Norman Haworth (1929), an English chemist. In its broadest sense, conformation has been used to describe different spatial arrangements of a molecule which are not superimposable. This means that, in effect, the terms **conformation** and **configuration** are equivalent. In the classical sense, the definition of conformation does not include the internal forces acting on the molecule. The term **conformation**, however, is the spatial arrangement of the molecule when all the internal forces acting on the molecule are taken into account. In more restricted sense, the term **conformation** is used to denote different spatial arrangements arising due to twisting or rotation of bonds of a 'given' configuration (used in the classical sense). The terms **rotational isomers** and **constellations** have also been used in the same sense as conformations.



104 / 1254

Fig. 5-2. Application of the cyanohydrin synthesis to D-glyceraldehyde

The conversion of D-glyceraldehyde into an aldotetrose yields D-erythrose and D-threose. These are called **diastereoisomers** i.e., isomers but not mirror images of each other. In fact, diastereoisomers are different forms of a compound with two asymmetric centres. It may, hence,

Contents

80 FUNDAMENTALS OF BIOCHEMISTRY

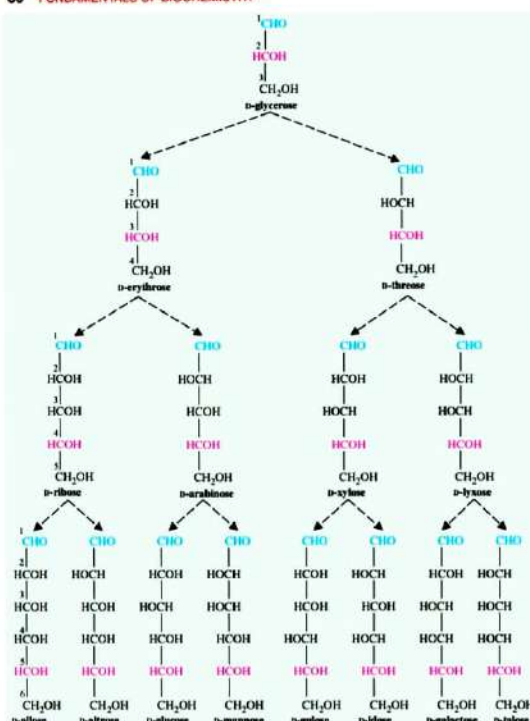


Fig. 5-3. D. Aldoses containing three, four, five and six carbon atoms

Aldoses contain an aldehyde group (shown in blue) and have the absolute configuration of D-glyceraldehyde at the asymmetric centre (shown in red) farthest from the aldehyde group. The numbers indicate the standard designations for each carbon atom. The formulae of L-aldoses are in each case the mirror images of these structures.

be inferred that if two optical isomers are not enantiomers, they are related as diastereoisomers. They differ from each other in their melting points, solubilities and the chemical properties, in general.

OPTICAL ISOMERISM

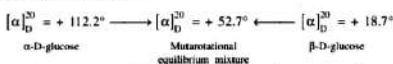
The presence of asymmetric carbon atoms in the compound also confers optical activity to it. When a beam of polarized light is passed through an optically active solution, it may be rotated either to the right or to the left, depending upon the type of optical isomer present. A compound rotating the plane of polarized light to the right is called as *dextrorotatory* (*dexter*² = right, on the right side; *rotatus*² = pp of *rotare*², to turn) or clockwise and is designated as *d* or (+) type. If the compound causes rotation of polarized light to the left, it is said to be *levorotatory* (*laevus*² = left, on the left side) or anticlockwise and is denoted as *l* or (–) type. The direction in which the light is rotated (or in other words, the optical rotation) is a specific property of the molecule. It should, however, be emphasized that the optical rotation is not at all related with the two *D* and *L* forms of a compound. Thus, *D*-glucose is dextrorotatory and *D*-fructose is levorotatory. If it is desired to indicate the direction of rotation, the two names may be written as *D*(+)-glucose and *D*(–)-fructose. Similarly, *D*-erythrose is levorotatory and *D*-threose, dextrorotatory.

When equal amounts of extorotatory and levorotatory isomers are present, the resulting mixture becomes optically inactive because the optical activities of each isomer cancel. Such a mixture is called a **racemic** or **dl-mixture** or (*±*)-**conglomerate** and this process is known as **racemisation**. An optically active compound into the racemic modification is known of **racemisation** (e.g., grape). Compounds produced synthetically are invariably racemic in nature. There are chances exist for the formation of two types of optical isomers or the *d* and *l* antipodes.

It may be concluded that all monosaccharides are optically active (except aldohexose and ketotriose) since they contain one or more asymmetric carbon atoms in their molecules.

Mutarotation

When a monosaccharide is dissolved in water, the optical rotatory power of the solution gradually changes until it reaches a constant value (Dubrunfaut, 1846). A freshly-prepared aqueous solution of α -*D*-glucose, for instance, has a specific rotation, $[\alpha]_D^{20}$ of +112.2°. And when this solution is allowed to stand, the rotation falls to +52.7° and remains constant at this value. The final stage can be attained more quickly either by heating the solution or by adding some catalyst which may be an acid or an alkali. This gradual change in specific rotation is known as **mutarotation** or changing rotation. In fact, this term reflects the discovery of the phenomenon by way of changes in the optical rotation of certain carbohydrates. The terms **multirotation** and **birotation** have sometimes also been used for mutarotation. The value of mutarotation for α -*D*-glucose is (+112.2°) – (+52.7°) or +59.5°. A fresh solution of β -*D*-glucose, on the other hand, has a rotation value of +18.7°; on standing, it also changes to the same value, +52.7°. All reducing sugars (except a few ketoses) undergo mutarotation. The specific rotations for the anomers of glucose and for the aqueous equilibrium mixture are :



These substances (open and ring forms) interconvert due to a dynamic equilibrium. Although the ring form predominates, each molecule spends some time in the open form, when asymmetry is lost. When the open-chain form passes back to the ring form, it is possible for it to give rise to either anomer. Thus, a pure sample of an anomer only remains so when in the solid state. In solution, particularly at alkaline pH, a pure sample of one anomer will change to an equilibrium mixture of the two forms. There need not be equal quantities of the two forms at equilibrium because there are other centres for asymmetry and so anomers (like *allo* and *threo* pairs of amino acids) do not have identical physical properties, including ΔG° of formation. Equilibrium between

106 / 1254

82 FUNDAMENTALS OF BIOCHEMISTRY

the anomers is called mutarotation and it can be followed by observing optical rotation (Fig. 5-4).

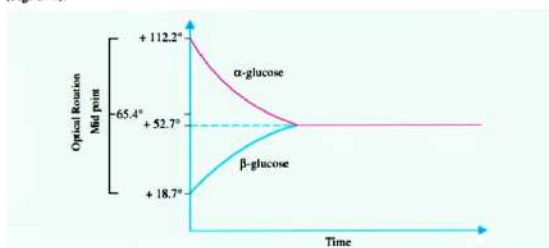


Fig. 5-4. The experimental detection of mutarotation

Mechanism of mutarotation. The phenomenon of mutarotation appears to be due to the changes of α - to β -forms and vice versa via the straight chain aldo or keto form (refer to the subsequent section). When an equilibrium is attained, the characteristic rotation is reached. According to Lowry (1925), mutarotation can take place only in the presence of an amphiprotic solvent (*i.e.*, a solvent which can act both as an acid and a base). Water is one such solvent. Thus, Lowry and Faulkner (1925) showed that mutarotation is checked in cresol solution (an acidic solvent) and in pyridine solution (a basic solvent). It is assumed that when mutarotation takes place, the ring opens and then recloses either in the inverted position or in the original position. The existence of the open chain form in mutarotation is proved by certain evidences. For example, the absorption spectra of fructose and sorbose in aqueous solution indicate the presence of open chain forms; the aldoses, however, give negative results (Bednarczyk *et al.*, 1938).

FORMULATION OF MONOSACCHARIDES

Formula of Glucose—Linear Form

An organic chemist intends to incorporate as many characteristics as possible in writing the formulae (Fig. 5-5). When the molecular formula of glucose, $C_6H_{12}O_6$ is written in a form as propounded by Fittig and Baeyer, it tells us the presence of an aldehyde and 5 hydroxyl groups. Replacement of this formula by Fischer projection formula informs the reader about some more details. For example, the presence of 4 asymmetric carbon atoms in the molecule indicates towards the existence of 2^4 or 16 stereoisomers.

Ring Form

However, even Fischer formula fails to describe certain other properties of glucose. Hence, another method of representing the structure was searched for. The aldehyde or ketone group of a sugar can react with hydroxyl groups of alcohols forming **hemiacetals** or **hemiketals**, respectively. For the larger sugars ($n > 4$), this happens *within the same molecule* to form a 5- or 6-membered ring.

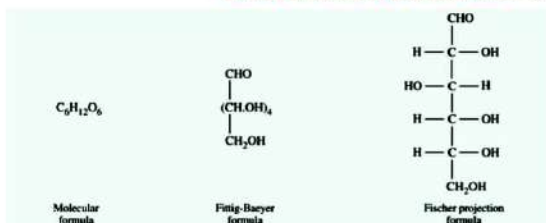
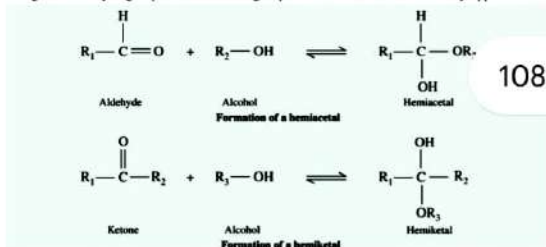


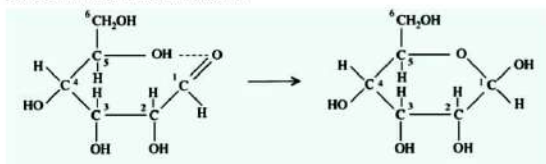
Fig. 5-5. Formulation of glucose

In fact, the angles of the tetrahedral carbon atom tend to bend the glucose molecule forming a ring. The aldehyde group and the alcohol group of carbon 5 or carbon 4 readily approach each



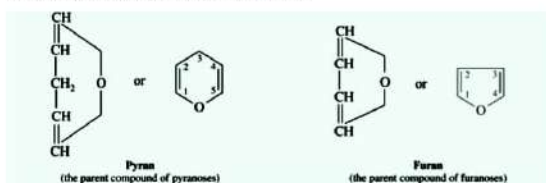
108 / 1254

other forming intramolecular *hemiacetals*. Similarly, the keto group of a ketose can also approach the alcohol group of either carbon 5 or carbon 4 forming intramolecular *hemiketals*. This results in the formation of either a 6-membered ring (**pyranose form**) or a 5-membered ring (**furanose form**). A 7-membered ring, however, becomes too strained to allow participation of the OH group of carbon 6 of aldohexoses with the CHO group in ring formation. The pyranose form has a lower ΔG° of formation than the furanose form.



84 FUNDAMENTALS OF BIOCHEMISTRY

The terminology of such structures is, in fact, based on the 2 simple organic compounds, exhibiting a similar ring structure. These are **pyran** and **furan**. The pyranose forms of the sugars are more stable than the furanose forms in solution.



The formation of a ring structure in the glucose molecule creates a new centre of asymmetry, i.e., the carbon 1 (carbon 2 in ketoses). One may, therefore, find that the ring form of glucose can exist as either of the 2 isomeric forms, termed as α and β isomers. These two isomers are, in fact, diastereoisomers rather than enantiomers because the α -form of D-glucose is not the mirror image of the β -form. They are also known as **anomers** (*ano*^G = upper) as they differ in configuration only around the hemiacetal or anomeric carbon atom (i.e., carbon 1). The designation of α - and β -forms is based on the suggestion of Freudenberg (1933) who stated that those anomers having the same configuration at both the anomeric and penultimate carbon atoms should be called as α -form while in β -form the configuration would be different at both these carbon atoms. In other words, the configuration of H atoms at both these carbons (anomeric and penultimate) is **erythro** (written on the same side of the structure) in α -form and **threo** (written on opposite sides) in β -form. The α - and β -forms of D-glucose in both pyranose and furanose rings along with the open chain formula are shown in Fig. 5-6.

As shown in the above figure, the hemiacetal formation is of reversible nature. Thus, D-glucose in aqueous solution exists as an equilibrium mixture of 5 compounds. The pyranose form, however, predominates in solution. An equilibrium mixture of glucose contains about one-third (37%) α anomer, two-thirds (62%) β anomer and very little (< 1%) of the open-chain form. Likewise, the α and β anomers of both the pyranose and the furanose forms of fructose interconvert through the open-chain form. In fact, under physiologic conditions in solution, monosaccharides with 5 or more carbons are typically more than 99% in the ring forms. The distribution between pyranose and furanose forms depends on : (a) the particular sugar structure, (b) the pH, (c) the solvent composition, and (d) the temperature. Representative data obtained from nuclear magnetic resonance (NMR) studies are shown in Table 5-2. When the monomers are incorporated into polysaccharides, the structure of the polymer may also influence the ring form chosen. As an instance, as Table 5-2 depicts, D-ribose exists in solution as a mixture of the two ring forms. But in biological polysaccharides, specific forms are stabilized. Ribonucleic acid, for example, contains exclusively ribofuranose whereas some plant cell wall polysaccharides have pentoses entirely in the pyranose form.

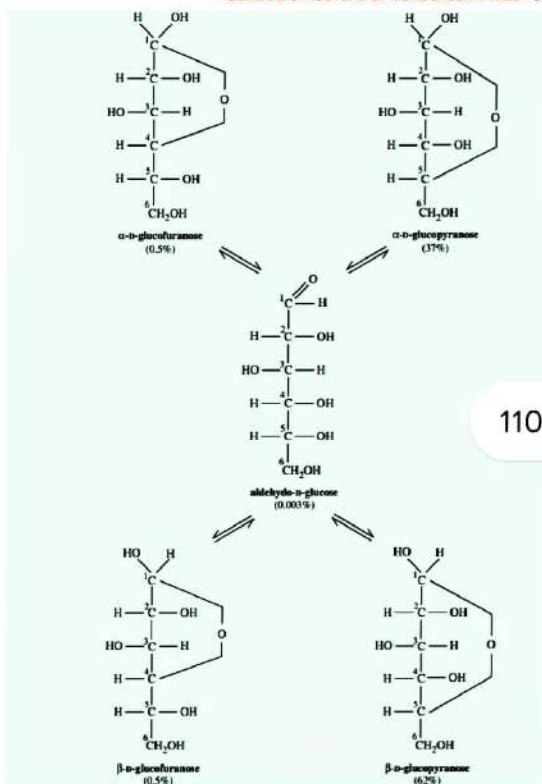


Fig. 5-4 Various forms of D-glucose present in an aqueous solution

[Note that in α -form, the OH group at anomeric carbon 1 is also on the right side like that at the penultimate carbon atom 5; whereas in β -form, the OH group at carbon 1 is on the left and that on carbon 5 is on the right side. Figures in parentheses indicate the molecular percentage present in an equilibrium mixture at pH 7.]

110 / 1254

Contents

86 FUNDAMENTALS OF BIOCHEMISTRY

Table 5-2. Relative amounts of tautomeric forms for some monosaccharide sugars at equilibrium in water at 40°C

Monosaccharide	Relative Amount (in%)				
	α pyranose	β pyranose	α furanose	β furanose	Total furanose
Ribose	20	56	06	18	24
Lyxose	71	29	— ^a	— ^a	<1
Altrose	27	40	20	13	33
Glucose	36	64	— ^a	— ^a	<1
Mannose	67	33	— ^a	— ^a	<1
Fructose	03	57	09	31	40

Note that in all cases, the open-chain form is much less than 1%.

^a designates much less than 1%.

Like other kinds of stereoisomers, these α and β forms rotate the plane of polarized light differently and can be distinguished that way. The monosaccharides can undergo interconversion between the α and β forms, using the open-chain structure as an intermediate. This process is referred to as **mutarotation**. A purified anomer, dissolved in aqueous solution, will approach the equilibrium mixture, with an accompanying change in the optical rotation of the solution. Enzymes called **mutarotases** catalyze the interconversion of anomeric sugars *in vivo*.

Haworth Perspective Formula

Although Fischer formulae are useful in indicating configurational differences among sugars, they are tedious to write. Haworth (1929), an English chemist, devised another method of representing their structure, resembling a hexagon (pyranose form) or a pentagon (furanose form). In such a projection, the carbon atoms in the ring are not explicitly shown. The approximate plane of the ring is perpendicular to the plane of the paper, with the heavy line on the ring projecting toward the reader. Haworth projections more clearly approximate the predominant "chair" structure of glucose (Fig. 5-7) and other hexose sugars that exist in solutions.

WALTER NORMAN HAWORTH

(LT, 1883–1950)

Walter Norman Haworth, a Briton, had been a student of Otto Wallach and W.H. Perkin at the Universities of Manchester and Göttingen. He was awarded 1937 Nobel Prize in Chemistry for his work on the constitution of carbohydrates and vitamin C, along with Paul Karrer of Switzerland. Haworth's contributions include: methylation of sugars, pyranose and furanose structures of monosaccharides, structure of starch, cellulose and amygdalin, and structure and synthesis of ascorbic acid.

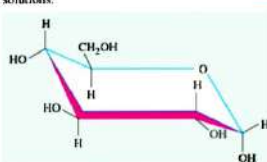
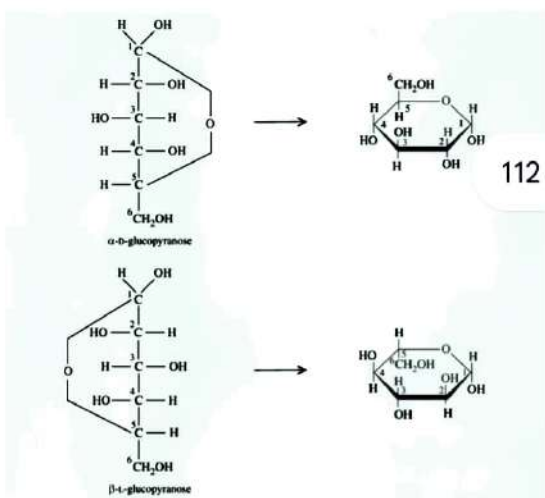


Fig. 5-7. "Chair" form of α -D-glucopyranose

In the pyranose form, carbon atoms 1 to 5 of the aldohexoses and the O atom of the ring are represented in the form of a hexagon and the plane of this ring projects from the paper. Bonds that are nearer to the reader are thickened in this type of formula and the carbon atoms are numbered clockwise. Fischer formula may be converted to Haworth formula by twisting it around imaginatively or by constructing models. Certain 'assumptions' have, however, been followed for conversion. These are:

1. Any group to the right of the carbon chain in Fischer formula is written *down* the plane of the ring in the Haworth formula; those to the left are written *up*. A notable exception to this assumption is carbon atom 5 where H atom, present on the left side, is shown below the plane of the ring instead of showing it above the plane of the ring. This is as a result of torsion needed to bring about ring closure. The carbon atom 5 rotates so that the O atom of the OH group on the anomeric carbon comes in the plane of the first five carbon atoms. Consequently, the H atom on carbon 5 is shifted to the other side of the carbon chain by rotating through more than 90° .

2. Another important assumption is that the carbon atom (or atoms) not involved in ring formation (e.g., carbon 6 in aldopyranoses) will be shown above the plane of the ring, if the oxide ring in Fischer formula is to the right. Conversely, if the oxide ring is to the left, the extra carbon atom (or atoms) would be shown below the plane of the ring in Haworth projection. Two examples (α -D-glucopyranose and β -L-glucopyranose) will illustrate the mode of conversion (Fig. 5-8).



112 / 1254

Fig. 5-8. Mode of conversion of two hexose sugars from Fischer projection formula to Haworth perspective formula

For the sake of simplicity, while writing Haworth formulae, the H and OH groups are not written and the bond indicating the single hydrogen atom is not shown. Thickened line in lower half of the hexagon is also omitted. The extra carbon atoms are, however, written as such. Thus, the above two compounds may be simply represented as in Fig. 5-9.

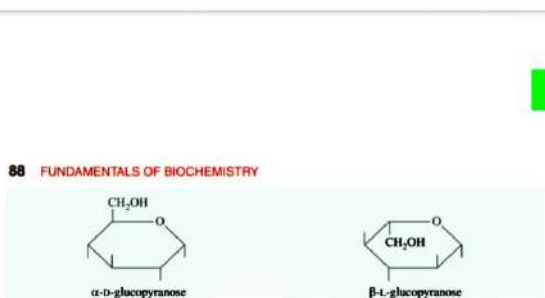


Fig. 5-9. Simplified Haworth formulae for two hexose sugars

The 6-membered pyranose ring, like cyclohexane, cannot be planar because of the tetrahedral geometry of its saturated carbon atoms. Instead, pyranose rings adopt *chair* and *boat* conformations (Fig. 5-10). The substituents on the ring carbon atoms have two orientations: *axial* and *equatorial*. Axial bonds are nearly perpendicular to the average plane of the ring, whereas equatorial bonds are almost parallel to this plane. Axial substituents emerge above and below the average plane of the ring and sterically hinder each other if they emerge on the same side of the ring. By contrast, equatorial substituents emerge at the periphery and are less crowded. The *chair* form of β -D-glucopyranose predominates because all axial positions are occupied by hydrogen atoms. The bulkier —OH and —CH₂OH groups emerge at the less hindered periphery. On the contrary, the boat form of glucose is highly disfavoured because it is much hindered sterically.

More generally, variations in the 3-D structure of biomolecules are described in terms of configuration and conformation. However, great care should be exercised in the use of these terms as they are not synonyms. **Conformation** refers to the spatial arrangement of substituent groups that are free to assume different positions in space, without breaking any bonds, because of the freedom of bond rotation. In ethane (C₂H₆), for example, there is nearly complete freedom of rotation around the single C—C bond. Many different interconvertible conformations of the ethane molecule are, therefore, possible, depending upon the degree of rotation. Two conformations are of special interest: the staggered conformation, which is more stable than others and thus predominates, and the eclipsed form, which is least stable. It is not possible to isolate either of these conformational forms because they are freely interconvertible without the breakage of any bonds and are in equilibrium with each other.

On the contrary, **configuration** denotes the spatial arrangement of an organic molecule that is conferred by the presence of (a) double bonds, around which there is no freedom of rotation, or (b) chiral centres, around which substituent groups are arranged in a specific sequence. The diagnostic feature of configurational isomers is that they can be interconverted only by breaking one or more covalent bonds. For example, maleic acid is a conformational isomer of fumaric acid.

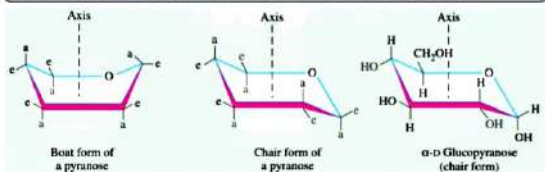


Fig. 5-10. Conformational formulae of the boat and chair forms of the pyranose ring

Substituents on the ring carbons may be either axial (a), projecting almost parallel with the vertical axis through the ring or equatorial (e) projecting almost perpendicular to this axis. For sugars with large equatorial substituents, the chair form is energetically more favourable because it is less hindered. The boat conformation is uncommon except in derivatives with very bulky substituents.

GENERAL STRUCTURE OF MONOSACCHARIDES 89

The 5-membered furanose rings, like pyranose rings, are not planar. They can be puckered so that four atoms are nearly coplanar and the fifth is about $0.5\text{\AA}$ away from this plane. This conformation is called *envelope form* because the structure resembles an opened envelope with the back flap raised (Fig. 5-11). In the ribose moiety of most biomolecules, either C-2 or C-3 is out of plane on the same side as C-5. These conformations are called $C_2\text{-endo}$ and $C_3\text{-endo}$. [Note that the sugars in DNA double helix are in the $C_2\text{-endo}$ form whereas the sugars in RNA are in the $C_3\text{-endo}$ form.] Furanose rings can interconvert rapidly between different conformational states. *Furanose rings are more flexible than pyranose rings* and possibly this may be the reason for their selection as components of DNA and RNA.

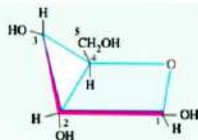


Fig. 5-11. An envelope form of β -D-ribose

The $C_3\text{-endo}$ conformation is shown. C-3 is out of plane on the same side as C-5.

114 / 1254

REFERENCES

See list following Chapter 8.

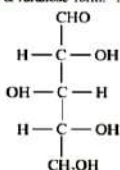
PROBLEMS

- Account for the origin of the term *carbohydrate*.
- Indicate whether each of the following pairs of sugars consists of anomers, epimers, or an aldose-ketose pair :
 - D-glyceraldehyde and dihydroxyacetone
 - D-glucose and D-mannose
 - D-glucose and D-fructose
 - α -D-glucose and β -D-glucose
 - D-ribose and D-ribulose
 - D-galactose and D-glucose
- Glucose and other aldoses are oxidized by an aqueous solution of a silver-ammonia complex. What are the reaction products ?
- The specific rotations of the α and β anomers of D-glucose are $+112$ degrees and $+18.7$ degrees, respectively. Specific rotation, $[\alpha]_D^{25}$, is defined as the observed rotation of light of wavelength 589 nm (the D line of a sodium lamp) passing through 10 cm of a 1 g ml^{-1} solution of a sample. When a crystalline sample of α -D-glucopyranose is dissolved in water, the specific rotation decreases from 112 degrees to an equilibrium value of 52.7 degrees. On the basis of this result, what are the proportions of the α and β anomers at equilibrium ? Assume that the concentration of the open-chain form is negligible.
- Glucose reacts slowly with hemoglobin and other proteins to form covalent compounds. Why is glucose reactive ? What is the nature of the adduct formed ?
- Compounds containing hydroxyl groups on adjacent carbon atoms undergo carbon-carbon

90 FUNDAMENTALS OF BIOCHEMISTRY

bond cleavage when treated with periodate ion (IO_4^-). How can this reaction be used to distinguish between pyranosides and furanosides ?

- Does the oxygen atom attached to C-1 in methyl α -D-glucopyranoside come from glucose or methanol ?
- Fructose in its β -D-pyranose form accounts for the powerful sweetness of honey. The β -D-furanose form, although sweet, is not as sweet as the pyranose form. The furanose form is the more stable form. Draw the two forms and explain why it may not always be wise to cook with honey.
- (a) Compare the number of reducing ends to nonreducing ends in a molecule of glycogen.
(b) Glycogen is an important fuel storage form that is rapidly mobilized. At which end—the reducing or nonreducing—would you expect most metabolism to take place ?
- Draw Haworth projections for the following :
 - in α -furanose form. Name the sugar.



- The L isomer of (a)
 - α -D-GlcNAc
 - α -D-fructofuranose
- α -D-galactopyranose rotates the plane of polarized light, but the product of its reduction with sodium borohydride (galactitol) does not. Explain the difference.
 - Provide an explanation for the fact that α -D-mannose is more stable than β -D-mannose, whereas the opposite is true for glucose.
 - What is the natural polysaccharide whose repeating structure can be symbolized by $\text{GlcUA}\beta(1\rightarrow3)\text{GlcNAc}$, with these units connected by $\beta(1\rightarrow4)$ links ?
 - A freshly prepared solution of the α form of D-galactose (1 g/mL in a 10 cm cell) shows an optical rotation of $+150.7^\circ$. When the solution is allowed to stand for a prolonged period of time, the observed rotation gradually decreases and reaches an equilibrium value of $+80.2^\circ$. In contrast, a freshly prepared solution (1 g/mL) of the β form shows an optical rotation of only $+52.8^\circ$. Moreover, when the solution is allowed to stand for several hours, the rotation increases to an equilibrium value of $+80.2^\circ$, identical to the equilibrium value reached by α -D-galactose.
 - Draw the Haworth perspective formulas of the α and β forms of galactose. What

CONTENTS

- Reaction of Glycosidic OH Group
- Reaction Involving Both Glycosidic and Alcoholic OH Groups
- Reaction of Alcoholic OH Group
- Reaction Involving Both Alcoholic OH and CHO/CO Groups
- Reactions of CHO/CO Group



Carbohydrates in food as important sources of energy.

Starch, found in plant-derived food such as pasta, consists of chains of linked glucose molecules. These chains are broken down into individual glucose molecules for eventual use in generation of ATP and building blocks for other molecules.

CHAPTER

6

Carbohydrates-II

Properties of Monosaccharides

Carbohydrates possess active groups which are responsible for their chemical behavior (or CO). Most of the reactions of mutarotation place in aqueous solution so that all the 3 f. available for the reaction.

116 / 1254

α -form $\rightleftharpoons$ aldehyde/keto form $\rightleftharpoons$ β -form

The above equilibrium is shifted in favour of the group which reacts. Some reagents, however, react with more than one group.

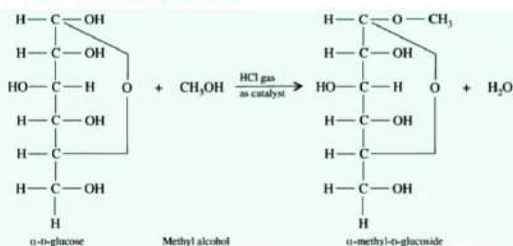
REACTION OF GLYCOSIDIC OH GROUP

Reaction with alcohol. The glycosidic OH group of mutarotating sugars react with alcohols to form α - and β -glycosides or acetals. Thus, glucose forms glucosides and fructose forms fructosides.

The glucosides, or glycosides in general, do not exhibit mutarotation as the aldehyde group in them is converted to the acetal group. A number of glycosides occur in nature, e.g., *phlorhizin* (glucose + phloretin) in rose bark; *digitonin* (4 galactose + xylose + digitogenin) in foxglove leaves; *amygdalin* (2 glucose + 2 mandelonitril) from bitter almonds and saponin (sugar + saponin) from soapwort. They are useful as medicaments.

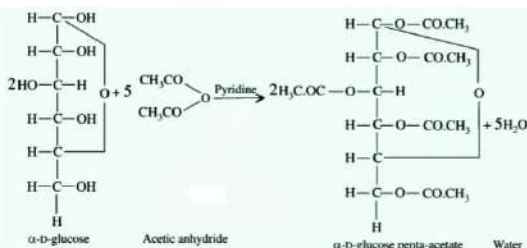
91

92 FUNDAMENTALS OF BIOCHEMISTRY



REACTION INVOLVING BOTH GLYCOSIDIC AND ALCOHOLIC OH GROUPS

Reaction with acetic anhydride (Esterification). The glycosidic and alcoholic OH groups of monosaccharides and disaccharides react with acetylating agents (like acetic anhydride in pyridine) to form acetate derivatives called **esters**.

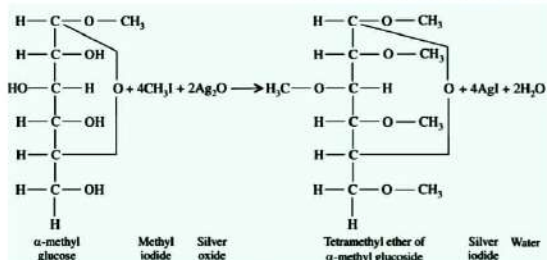


The ability of sugars to form esters indicates the presence of alcohol groups in their molecule. As glucose upon acetylation yields a penta-acetate derivative, it obviously contains 5 OH groups.

REACTION OF ALCOHOLIC OH GROUP

Reaction with methyl iodide (Etherification). The alcoholic OH groups of monosaccharides and disaccharides are converted to ether groups upon treatment with methylating agents.

[The glycosidic OH group of the sugars is, however, first protected with an R group by converting these sugars to glycosides which are, then, allowed to react with the methylating agent.]

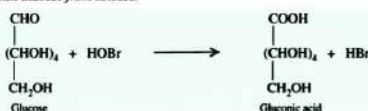


This reaction has been extensively used to ascertain which OH groups in a sugar are free and available for reaction. Thus, α -methyl glucoside is methylated at carbon 2, 3, 4 and 6 but not at carbon 5 which is involved in hemiacetal link.

REACTION INVOLVING BOTH ALCOHOLIC OH AND CHO/CO GROUPS

Oxidation with acids. The alcoholic OH group and CHO group (or CO group) or carboxyl groups by certain oxidizing agents. The oxidation may be brought about under vigorous oxidizing conditions.

(a) With mild oxidants (like HBr) – Only the aldehyde group is oxidized to produce **monocarboxylic acids**. Ketoses, however, do not respond to this reaction. Hence, this reaction is used to distinguish aldoses from ketoses.



(b) With strong oxidants (like conc. HNO_3) – Both the aldehyde group (or ketone group) and the primary alcohol group are oxidized to yield **dicarboxylic acids**. With aldoses, acids with same number of carbon atoms are obtained whereas ketoses react to produce acids with fewer number of carbon atoms.

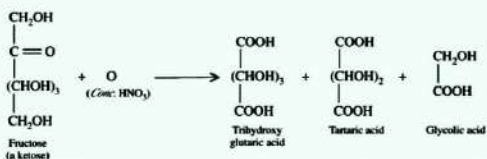
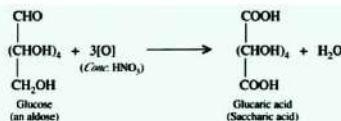
REACTIONS OF CHO/CO GROUP

1. **Oxidation with metal hydroxides.** Metal hydroxides like $\text{Cu}(\text{OH})_2$, AgOH and $\text{Bi}(\text{OH})_3$ oxidize the free aldehyde (or ketone) group of mutarotating mono- and di-saccharides and are, at the same time, themselves reduced to the lower oxides or to the free metals. Thus, $\text{Cu}(\text{OH})_2$ is reduced to Cu_2O and AgOH and $\text{Bi}(\text{OH})_3$ are reduced to the free metal, Ag and Bi. The reaction with cupric hydroxide is given below:



Contents

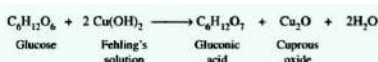
94 FUNDAMENTALS OF BIOCHEMISTRY



Cupric ion (Cu^{++}) is, in fact, the most common oxidizing agent of these. It is the active ingredient in Fehling's, Benedict's and Barfoed's reagents.

Fehling's reagent consists of 2 solutions: solution A contains 7% CuSO_4 and solution B contains 25% KOH and 35% sodium potassium tartarate. When the two solutions are mixed in equal amounts, a clear blue solution results for the tartarate forms a soluble complex with the copper hydroxide produced. This solution is widely used as an oxidizing agent preferably in quantitative sugar determinations.

With Fehling's reagent, the overall reaction occurs as follows:



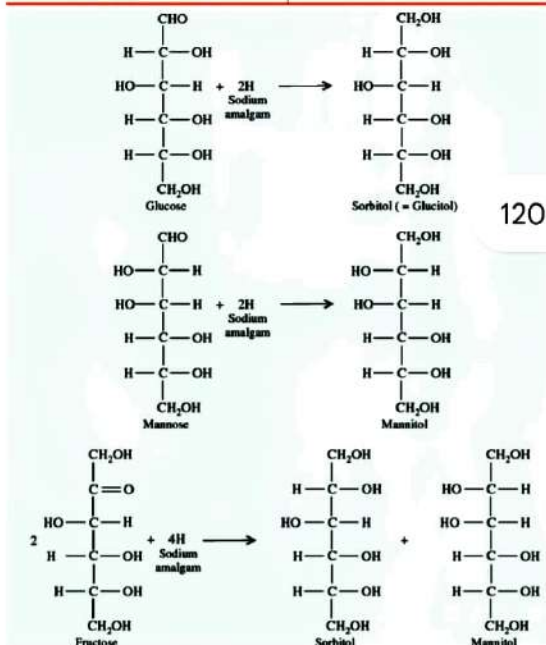
In clinical laboratories, however, a modification of Fehling's reagent is used for the qualitative determination of glucose in urine. This is known as **Benedict's reagent** and consists of a single solution containing CuSO_4 , Na_2CO_3 (in place of KOH) and sodium citrate (in place of potassium tartarate). Sodium carbonate reduces alkalinity to such an extent that, unlike the Fehling's reagent, the two solutions (CuSO_4 and sodium citrate) here are kept mixed indefinitely as one solution.

Barfoed's reagent utilizes a weakly acid solution for oxidation. It contains 7% cupric acetate and 1% acetic acid. This reagent can be used to distinguish monosaccharides from disaccharides as the former give a cuprous oxide precipitate in dilute acid more quickly than the latter. Monosaccharides are, thus, more active reducing agents than the disaccharides.

Carbohydrates with free carbonyl groups or in hemiacetal form give positive tests to these reagents without having been hydrolyzed first and are referred to as '**reducing**' sugars; others (i.e., the acetal types) are then '**nonreducing**' sugars (Table 6-1).

Table 6-1. Differences between reducing and nonreducing sugars

Reducing sugar	Nonreducing sugar
1. Carbohydrates with a free aldehyde (at C-1) or a free ketone (at C-2) group. 2. They are in hemiacetal or hemiketal form. 3. Do exhibit mutarotation. 4. Do form osazones with phenyl hydrazine. 5. Do form oximes with hydroxylamine. <i>Examples – Glucose, Fructose, Lactose, Maltose, Cellobiose</i>	1. Aldehyde or ketone group is not free but instead utilized in bond formation. 2. They are in acetal or ketal form. 3. Do not exhibit mutarotation. 4. Do not form osazones. 5. Do not form oximes. <i>Examples – Sucrose, Glycogen, Inulin</i>



Contents

96 FUNDAMENTALS OF BIOCHEMISTRY

The **Molisch test** employs sulfuric acid for dehydration and α -naphthol as the phenol. The acid also hydrolyzes acetal groups present and thus makes this test probably the most general of all the tests for carbohydrates.

Since the alkali present in some of the above tests causes fragmentation of the sugar molecule, a mixture of hydroxy acids of different chain length is obtained upon oxidation of carbohydrates. Hence, these tests, although widely employed in analytical methods for sugars, are of little value in preparative chemistry.

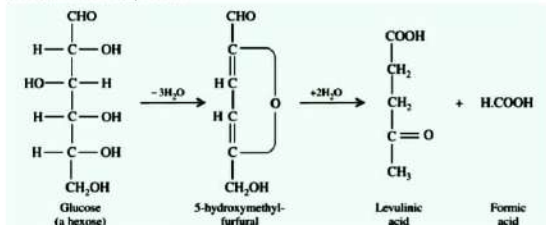
2. **Reduction.** The sugars may be reduced in various ways depending upon the type of reducing agent used.

(a) With sodium amalgam – The monosaccharides are reduced to their corresponding **alcohols** by treating them with reducing agents like Na-amalgam. Thus, glucose yields sorbitol (= glucitol), mannose yields mannitol, galactose yields dulcitol, fructose yields a mixture of sorbitol and mannitol, and glyceraldehyde yields glycerol.

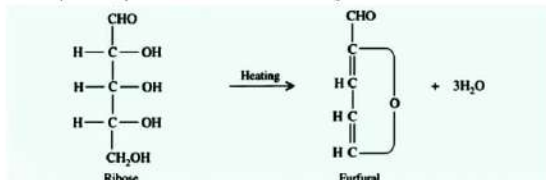
Sorbitol and mannitol both are important and are used in the manufacture of surface-active agents and explosives, respectively.

(b) With strong mineral acids – With **hexoses**, the reduction leads to the formation of 5-hydroxymethylfurfural, which on further heating is transformed to levulinic acid.

This reaction forms the basis of some colour tests (e.g., Molisch test and Bial's orcinol test) for sugars as the aldehyde products of these reactions condense with certain organic phenols to give characteristic coloured products.



With **pentoses**, a product **furfural** is obtained on heating.



Tetroses and **trioses**, however, do not undergo this reaction as they do not possess the necessary minimum 5 carbon atoms for furfural formation.

(c) With dilute alkalis – Glucose, fructose and mannose are interconvertible in weak alkaline solutions such as $Ca(OH)_2$ and $Ba(OH)_2$ at low temperatures. This reaction is known as **Lobry de**

Bruyn-Alberda van Ekenstein transformation. This reaction, proposed in 1890, is of special significance as a similar reaction is supposed to take place in the human body. The mechanism of this reaction involves **enolization**, the migration of a proton from a carbon atom onto the oxygen of an adjacent carbonyl group, resulting in the formation of an unsaturated alcohol called **enol** (*ene* = unsaturation; *ol* = alcohol). The interconversion between glucose, fructose and mannose may, in fact, be visualized as occurring through a common enediol form (Fig. 6-1).

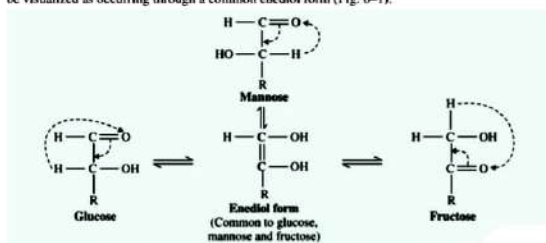
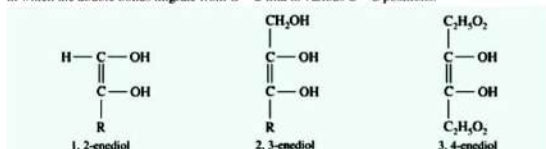


Fig. 6-1. Interconversion of glucose, mannose and fructose (Enolization)

122 / 1254

The above mechanism emphasizes that the same mixture (of glucose, fructose and . . . obtained even if, instead of D(+)-glucose, the starting material is D(-)-fructose or D(+)-mannose. Using deuterium oxide, Topper et al, in 1951, have supported this enolization mechanism but have concluded that fructose and mannose cannot arise from the same enediol and that there are two geometric isomeric enediol intermediates, both forms capable of changing into fructose.

If the mixture is, however, heated to 37°C, the acidity increases and a series of enols are produced in which the double bonds migrate from O—C link to various C—C positions.

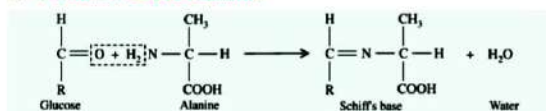


In concentrated alkali, sugar caramelizes and many decomposition products are formed. Certain colored pigments develop and salts may also be formed.

3. **Reaction with hydrogen cyanide (Kiliani synthesis).** As already discussed (see page 73), the addition of HCN to mutarotating sugars creates a new asymmetric carbon atom, producing two new compounds namely, cyanohydrin A and cyanohydrin B. These are, then hydrolyzed with dilute mineral acids to form hydroxy acids which are later converted to γ -lactones and finally reduced to two new sugars containing one more carbon atom than the parent sugar.

4. **Reaction with alanine.** The aldehyde group of carbohydrates condenses with the amino group of alanine to form Schiff's base.

98 FUNDAMENTALS OF BIOCHEMISTRY



The browning reaction (Maillard's reaction) occurring during the baking of bread and other mixtures of carbohydrates and proteins is believed to be due to the formation of Schiff's base between the amino groups of proteins and the aldehyde groups of carbohydrates (Mertz, 1960).

5. **Reaction with phenyl hydrazine.** Reaction with phenyl hydrazine involves only 2 carbon atoms viz., the carbonyl (i.e., the aldehyde or ketone group) and the adjacent one. One mole of phenyl hydrazine reacts with one mole of aldose (or ketose) to form a mole of hydrazone. With a second mole of phenyl hydrazine, the hydrazone is oxidized to **aldohydrazone** and the phenyl hydrazine itself is reduced to aniline and ammonia. Finally, a third mole of phenyl hydrazine reacts with the aldohydrazone to produce **osazone**. The hydrazone may, in fact, be regarded as a special type of Schiff's base and the osazone as a double Schiff's base (Fig. 6-2).

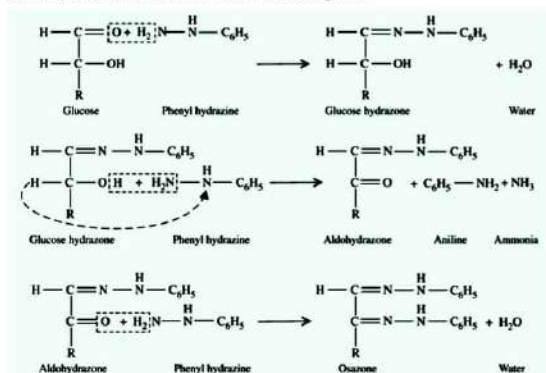


Fig. 6-2. Osazone formation

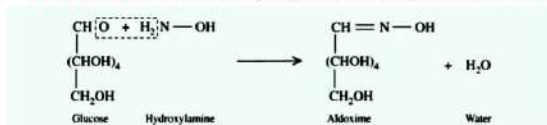
The reaction with ketose will also take place in a likewise manner.

The osazones of reducing (mutarotating) sugars are crystalline, yellow and usually insoluble compounds and hence may also be recovered. These have characteristic crystal structure and melting points and are, therefore, frequently used in the identification of sugars. It may, however, be noted that glucose, fructose, and mannose would yield the same osazone owing to their similar structure in the unaffected part of the molecule (from C_3 to C_6). Moreover, the asymmetry at C_2 is destroyed during the reaction. Obviously, galactose would, then, form a different osazone as it has a different configuration below C_2 .

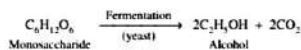
PROPERTIES OF MONOSACCHARIDES 99

Osazone formation is not restricted to monosaccharides but is also shown by most disaccharides. Lactosazone crystals have a typical hedgehog shape and are readily identifiable. Maltosazone crystals have a characteristic petal-like appearance and a cluster of them looks like the sunflower. Sucrose, however, does not form osazone since it does not contain a functional carbonyl group.

6. Reaction with hydroxylamine. Simple sugars react with hydroxylamine to yield **oximes**.



7. Fermentation. Monosaccharides such as glucose, fructose and mannose are readily fermented by yeast. The process of yeast fermentation is very complex. During this process, sugar phosphate and sugar acid phosphate are formed. Ordinarily, this process results in the formation of alcohol and carbon dioxide.



REFERENCES

See list following Chapter 8.

124 / 1254

PROBLEMS

- Sucrose is commonly used to preserve fruits. Why is glucose not well suited for preserving fruits?
- Which of the following carbohydrates is most rapidly fermented by brewer's yeast (*Saccharomyces cerevisiae*)?
 - glycogen
 - maltose
 - lactose
 - glucose
 - starch

CONTENTS

- Conformation of Pyranose and Furanose Rings
- Sucrose or Saccharose
- The Artificial or Synthetic Sweeteners
- Lactose
- Maltose
- Cellobiose
- Isomaltose
- Trehalose



Space-filling model of **aspartame**, a noncaloric sweetener, widely used in the drink and food industry.

CHAPTER

7

Carbohydrates-III

Oligosaccharides

The oligosaccharides yield 2 to 10 monosaccharide molecules on hydrolysis. *Disaccharides are the most common oligosaccharides found in nature.* Most of the naturally-occurring representatives occur in plant rather than in animal sources. The three most important disaccharides found free in appreciable quantities are sucrose, lactose and maltose; a few others are cellobiose and trehalose.

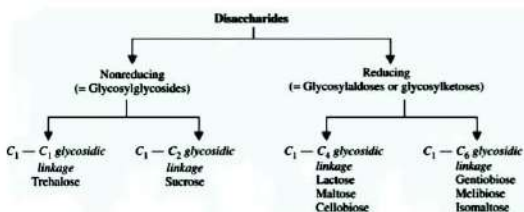
CONFORMATIONS OF PYRANOSE AND FURANOSE RINGS

To name disaccharides, certain rules are followed. First, the compound is written with its nonreducing end to the left. An *O* precedes the name of the first (left) monosaccharide unit, as a reminder that the sugar-sugar linkage is through an oxygen atom. The configuration at the anomeric carbon joining the first (left) monosaccharide unit to the second is then given (α or β). To distinguish 5- and 6-membered ring structures, "furanosyl" or "pyranosyl" is inserted in the name of each monosaccharide unit. The two carbon atoms joined by the glycosidic bond are then shown in parentheses, with an arrow connecting the two numbers; for example, (1 $\rightarrow$ 4) shows that C-1 of first sugar residue is joined to C-4 of the second. If there is a third residue, the second glycosidic bond is described next, by the same conventions. To shorten the description of a complex **polysaccharide**, 3-letter abbreviations for each monosaccharide are often used.

Following this convention for naming oligosaccharides, sucrose is therefore named *O*- α -D- glucopyranosyl-

(1 $\rightarrow$ 2)- β -D-fructofuranoside. Because most sugars are the D-enantiomers, and the pyranose form predominates, a shorthand version of the formal name of such compounds is used, giving the configuration of the anomeric carbon and naming the carbons joined by the glycosidic bond. In this abbreviated nomenclature, sucrose is Glc (α 1 $\rightarrow$ 2) Fru.

Based on the type of linkage present in their molecule, the disaccharides may be classified as follows:



SUCROSE OR SACCHAROSE

(Table sugar, Cane sugar, Beet sugar)

126 / 1254

Occurrence. Sucrose is the common sugar of commerce and kitchen (hence, 'household' sugar) and is widely distributed in all photosynthetic plants. It is the chief α -sugarcane (*Saccharum officinarum*), beet (*Beta vulgaris*) and maple (*Acer saccharina*) and is also present in pineapple (*Ananas sativus*) and carrot (*Daucus carota*). Sucrose occurs in varying amounts in different plant organs such as fruits, seeds, flowers and roots. Nectar of flowers is particularly rich in sucrose. It is the raw material for honey. *Sucrose is probably the only foodstuff used in the crystalline form.*

Sucrose is something of a riddle in plant biochemistry. Although D-glucose is the major building block of both starch and cellulose, sucrose is a major intermediate product of photosynthesis. It is the predominant form in which sugar is transported in most plants from the leaves to other organs of plants via their vascular system.

Chemistry. It is derived commercially from either sugarcane or beet. Refinement removes the yellow-brown pigments of unrefined sugar to produce the white crystal form of table sugar, while leaving a syrup which does not crystallize. This syrup, known as *molasses*, is also a commercial product and is the least refined form of sucrose. In plants, sucrose is formed by elimination of a molecule of water from glycosidic hydroxyl groups (marked by an asterisk) or α -D-glucose and β -D-fructose (Fig. 7-1.)

It is noteworthy that fructose possesses the furanose ring structure in the sucrose molecule, although the pyranose ring is the dominant and more stable form in the free ketohexose. The linkage between the two disaccharide moieties (= units) of sucrose is a glycosidic one between C₁ of glucose and C₂ of fructose.

Properties. Sucrose is a white crystalline solid, soluble in water and with a melting point 180°C. When heated above its melting point, it forms a brown substance known as *caramel*. Concentrated sulfuric acid chars sucrose, the product being almost pure carbon. It is dextrorotatory and has a

102 FUNDAMENTALS OF BIOCHEMISTRY

specific rotation of +66.7°. It is by far the sweetest of the 3 common disaccharides (sucrose, lactose, maltose). It is also sweeter than glucose (refer Table 7-1). It crystallizes in colourless crystals.

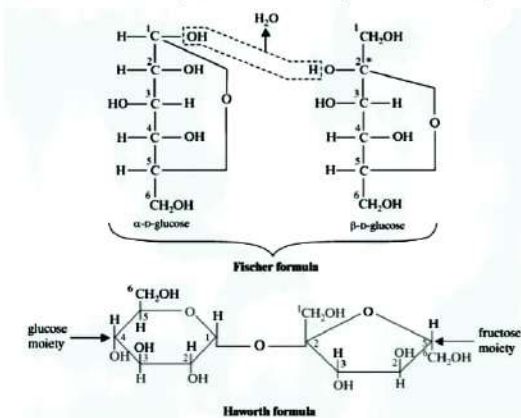


Fig. 7-1. Structure of sucrose

O- α -D-glucopyranosyl-(1 $\rightarrow$ 2)- β -D-fructofuranoside or Glc (α 1 $\rightarrow$ 2) Fru

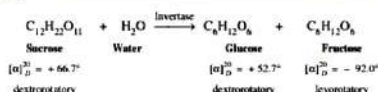
It can be seen from the formula of sucrose that both the carbonyl groups (marked by an asterisk) are involved in the formation of glycosidic bond. Consequently, sucrose contains no active group and, therefore, does not exhibit those properties which depend upon the presence of this group. Obviously, sucrose does not exhibit mutarotation and is *not* a reducing sugar. It also does not form an osazone or an oxime.

Hydrolysis. Upon hydrolysis, sucrose yields equimolar mixture of glucose and fructose which is often called *invert sugar*. This mixture readily reduces Fehling's solution and other alkali reagents. The name 'invert sugar' is given to this mixture because the levorotatory fructose, thus produced, changes (or inverts) the previous dextrorotatory action of the sucrose. In fact, sucrose (which is dextrorotatory with a specific rotation, $[\alpha]_D^{20}$ of +66.7°), upon hydrolysis, gives a mixture of equimolar quantities of D(+) glucose (dextrorotatory: $[\alpha]_D^{20}$ of +52.7°) and D(-) fructose (levorotatory: $[\alpha]_D^{20}$ of -92.0°). And as the levorotation of fructose is greater than the dextrorotation of glucose, the mixture so obtained is levorotatory, contrary to the initial dextrorotatory sucrose.

Table 7-1. Sweetness index* of some sugars and noncaloric sweeteners

Sugar / Sweetener	Relative sweetness
Sugars	
Lactose	0.16
Raffinose	0.20
D-galactose	0.32
Maltose	0.32
D-xylose	0.40
Glycol	0.50
D-glucose	0.74
Sucrose	1.00
Invert sugar	1.23
D-fructose	1.73
Sweeteners	
Sucaryl sodium	30
Aspartame	180
Acesulfame-K	200
Stevia extract	300
Saccharin	400
Neohesperidine dihydrochalcone	1,000
Monellin	2,000

* Sweetness index is ascertained with respect to sucrose whose sweetness level is arbitrarily taken as one.



Mixture of glucose and fructose is levorotatory with $[\alpha]_D^{20}$ value of $(92.0 - 52.7) = -39.3^\circ$.

This reaction which is called "inversion of sucrose" is catalyzed by the enzyme invertase and also by H^+ ions. Animals cannot absorb sucrose as such, but it is made available for absorption by the enzyme sucrase present in the intestinal mucosal cells. As already pointed out, this enzyme catalyzes the hydrolysis of sucrose to D-glucose and D-fructose, which are readily absorbed in the bloodstream.

It is interesting to note that the invert sugar is more sweeter than the sucrose itself owing to the presence of fructose in invert sugar which is sweetest of all the sugars. This also explains why honey which contains a large proportion of invert sugar is sweeter than sucrose. The invert sugar is used in candy both because of its sweetness and also because the monosaccharides do not crystallize.

Invertase is also called **sucrase**, **saccharase** or **6-D-fructosidase** the first name being the most preferred one. This enzyme has been found to occur in green leaves, fruits, grains, stems, potato tubers, some roots, pollen and such lower plants as fungi and bacteria. It is especially abundant in yeast. Salts of heavy metals (Ag, Cu, Hg) inhibit its action. Maximum activity is obtained with low concentrations of sucrose (5–10%). Sucrase also hydrolyzes the sugars gentianose, raffinose, stachyose and, to some extent, inulin.

128 / 1254

104 FUNDAMENTALS OF BIOCHEMISTRY

THE ARTIFICIAL OR SYNTHETIC SWEETENERS

Sweeteners are customarily grouped into two categories, based on the difference in the amount of energy provided by them. These two groupings are:

A. Nutritive or Caloric Sweeteners. These contain calories, hence have nutritive value. These include sugar sweeteners (e.g., refined sugars, high fructose corn syrup, crystalline fructose, glucose, dextrose, corn sweeteners, honey, lactose, maltose, invert sugars, concentrated fruit juice) and sugar alcohols.

B. Nonnutritive or Noncaloric Sweeteners. These contain no calories, hence have no nutritive value. These have been developed artificially, hence also called as **artificial sweeteners**. As they sweeten with little volume, they may also be referred to as **high-intensity sweeteners**. They have been developed especially for obese or diabetics, for whom sugar consumption is harmful. Artificial sweeteners stimulate the taste buds of the tongue that are stimulated by sugars but, unlike sugars, they are of no food value and may have harmful effects, if used to excess. Some sweeteners are considered 'Generally Recognized As Safe' (GRAS) ingredients and others are considered food additives. The safety limit of food additives or conditions of use is expressed as the acceptable daily intake (ADI), that is, the estimated amount per kilogram body weight that a person can safely consume every day over a lifetime without risk. ADI is a conservative level – it usually reflects an amount 100 times less than the maximum level at which no observed adverse effects occur in animals.

A description of some important artificial sweeteners (Fig. 7.2) follows:

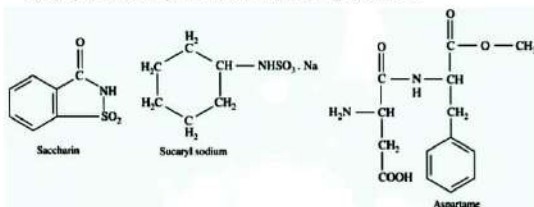


Fig. 7-2. Some artificial sweeteners

1. Saccharin. Saccharin is the first artificial sweetener which was discovered accidentally by Constantine Fahlberg in 1879. It is the most widely used artificial sweetener. In fact, saccharin was not used extensively as a sweetener until the advent of World War I when the use of sugar was strictly rationed. When World War II hit, sugar rationing started again, leading to another significant rise in saccharin use. Saccharin is sold under the tradename Sweetex™. It has a molecular formula $\text{C}_7\text{H}_5\text{SO}_2\text{NH}_2$ and its chemical structure is different from that of a carbohydrate. It possesses no OH group but rather a sulfonamide. It is manufactured from toluene, $\text{C}_6\text{H}_5\text{CH}_3$. Saccharin is a white, crystalline, sparingly soluble solid with a m.p. 227°C . It is heat-stable and is 400 times sweeter than sucrose by weight in aqueous solutions (Table 7-1). It is advised that pregnant women

While working on new food preservatives, Fahlberg accidentally spilled a compound he had synthesized. When he ate his dinner that night, he noticed that the intense residual sweetness of the chemical still lingered on his hands. He named the compound saccharin after the Latin *saccharum* which means sugar, referring sweetness in this context.

avoid saccharin.

2. **Sucaryl sodium.** It is also a sulfonamide and is 30 times sweeter than sucrose.

3. **Monellin.** This protein sweetener is present in the sap of serendipity berries, fruit of a western African plant, *Dioscoreophyllum cumminsii*. This sweet protein is composed of two noncovalently-associated polypeptide chains (one having 42 amino acid residues and the other 50 residues) and is 2,000 times as sweet as sucrose. Monellin's sweetness requires the undissociated protein since neither protomer has a sweet taste. Presently, monellin is being studied as a nonfattening, nontoxic food sweetener for human use. In case of the African plant *Katemfe*, its intense sweetness is attributed to two proteins, which together are 1,600 times sweeter than sucrose.

4. **Aspartame.** This is an essentially noncaloric, sugar-free sweetener being commercially used (Technically, aspartame is nutritive as it contains 4 calories per gram, but since it is added in such tiny doses that it really makes little, if any, contribution to the total caloric intake). Aspartame is a synthesized dipeptide, containing aspartic acid (Asp) and the methyl ester of phenylalanine (Phe).

Aspartame is marketed under the tradenames Nutrasweet™ (when included in food) and Equal™ (when sold as a powder). It is about 180 times sweeter than sucrose by weight (i.e., about half as sweet as saccharin) and does not have the bitter aftertaste, often associated with saccharin. It, thus, has a better taste profile than saccharin. Interestingly, (S-S)-aspartame is a sweetener whereas its enantiomer, (S-R)-aspartame is bitter in taste. However, aspartame has a shelf life of about 6 months, after which it breaks down into its components and loses its sweetening power. Also, it splits at high temperatures (which means that it cannot be used in hot or baked items). Intestinal enzymes hydrolyze aspartame to phenylalanine, aspartic acid and methanol in the ratio 5 : 4 : 1.

It is now widely used in the drink and food industry, e.g., soft drinks, cold cereal, gelatin desserts, puddings, toppings and fillings for baked goods, and cookies. As an additional advantage as it does not promote tooth decay, hence also used in gum and users have reported suffered from adverse reactions to aspartame such as headache, dizziness and nausea. Such persons must abstain from using it. Because aspartame is made from phenylalanine, people with phenylketonuria (who cannot metabolize Phe) should avoid aspartame. Also, the pregnant women are advised to refrain from aspartame. Aspartame is approved for use in more than 100 nations.

5. **Alitame.** Alitame, also a peptide, is composed of *l*-aspartic acid, *d*-alanine and a novel C-terminal amide moiety. It is 2,000 times sweeter than sucrose without the bitter and metallic qualities of noncaloric sweeteners. Curiously, although alitame tastes sweet to most people, it tastes bitter to a minuscule part of population. This may be due to a genetic difference in taste perception. Alitame blends with other artificial sweeteners to maximize the quality of sweetness.

6. **Neohesperidine dihydrochalcone.** It is a carboxylic derivative product from bitter components of the grape fruit's skin. It is about 1,500–1,800 times sweeter than sucrose at threshold limit concentration but the intensity depends on many factors such as pH and the product to which it is added. It takes some time to reach maximum sweetness perception and a very light sweet methanol-like aftertaste. It has a synergistic sweetening effect when combined with sugar alcohols, and with aspartame, saccharin, acesulfame-K and cyclamate. This characteristic has many advantages :

- (a) reduction in cost.
- (b) reduction in the daily intake of any particular sweetener
- (c) a more satisfying sucrose-like taste.

It may be used in beverages, juices, dairy products and alcoholic drinks such as beer. It is also used to sweeten diet and low-calorie food.

7. **Sugar alcohols.** Polyalcohols (or polyols) are part of a group of compounds called sugar alcohols. These include sorbitol, manitol and xylitol. Polyols can also be categorized as 'sugar

130 / 1254

106 FUNDAMENTALS OF BIOCHEMISTRY

'replacers' because they can replace sugar sweeteners, usually on a one-to-one basis. They offer less energy and offer potential health benefits (e.g., reduced glycemic response, reduced dental caries risk, etc). Commercially, these sweeteners are synthesized and not extracted from natural sources. Sugar alcohols contain 2-3 calories per gram but are only slowly metabolized to glucose. These are found in sugar-free candy, gum, cake mix and syrup, as well as some medications. Like aspartame, they also do not promote tooth decay. However, sugar alcohols can cause gastrointestinal symptoms ranging from mild stomach discomfort to severe diarrhea. Hence, those products likely to cause ingestion of 50 g or more of sorbitol are required to carry a level reading.

"Excess consumption may have a laxative effect".

8. **Acesulfame potassium (=Acesulfame-K).** It is the newest alternative sweetener approved in the US and is marketed under the tradenames Sunette™ and SweetOne™. It was formulated in the late 1960s and was evaluated for safety in 1983. Food and Drugs Administration (FDA) approved acesulfame-K in 1988. Acesulfame-K is about 200 times sweeter than sucrose. Blends of acesulfame-K with other nutritive and nonnutritive sweeteners can synergize the sweetness potential and the bitter taste. It has a shelf life of about 3-4 years as it is heat-stable (hence can be used for baked items). It does not provide any energy as it is not metabolized by the body and is excreted in the urine unchanged. Acesulfame-K finds applications in gum, powdered drink mixes, gelatins, puddings, candy, throat lozenges, tabletop sweeteners, yogurt and nondairy creamers. It is also used as a sweetener in soft drinks, chewing gum, coffee, tea, flavouring, salted food and dairy items.

9. **Stevia extract.** The plant sweet honey leaf (*Stevia rebaudiana*) is a herbaceous perennial. It is native to Brazil, Venezuela, Columbia and Paraguay where the native Guaraní have used it for over 1,500 years to sweeten otherwise unpalatable medicinal drinks. It is a natural herbal sweetener. The sweetness of *Stevia* is attributed to 2 compounds, *stevioside* and *rebaudioside*, which can be up to 300 times sweeter than sucrose. *Stevia* has an advantage over artificial sweeteners in that it is stable at high temperatures and a pH range of 3 to 9. *Stevia* extract is used as sweetener or flavour enhancer in many countries such as China, Japan, Korea, Israel, Brazil and Paraguay. It is also used in soft drinks, ice creams, cookies, pickles, chewing gum, tea and skin care products. *Stevia* plant and its extract both are used in weight-loss programmes because of their ability to reduce the cravings for sweet and fatty foods. Sweet honey leaf is, however, banned in food products in the US because a derivative of stevioside may be harmful to humans.

Thus, mankind's search for the ultimate sweetener goes on. In fact, sweetness is rather subjective measure of taste. The sensation of sweetness is believed to be due to the interaction between molecules of the sweetening agent and receptor sites of the tongue. Furthermore, perception of sweet taste can be influenced by genetic behaviour, health status and aging. Robert Margolske of the Mount Sinai of Medicine in New York hopes to have definite proof that gene TIR3 is truly the "sweet tooth" receptor gene. He feels that the present artificial sweeteners were developed at random and by chance. However, by having the receptor protein that normally binds sugars and sweeteners, one should be able to target more specifically better-designed molecules that will bind and activate the receptor.

LACTOSE (Milk sugar)

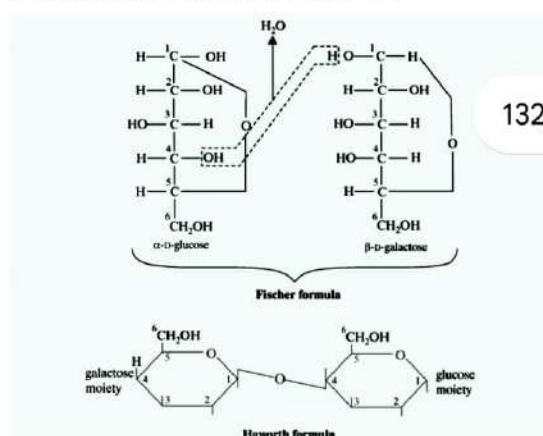
Occurrence. Lactose is solely of animal origin and is found in the milk of mammals. Human milk contains about 6-8 % (and 0.3% of higher oligosaccharides) ; cows' milk, about 4.8% (refer Table 7-2); hence, mother's milk is about 1.5 times sweeter than cow's milk. It is an unique product of mammary glands and is not found in other parts of the animal body. A minute quantity of lactose is also present in the milk as neuramin lactose. However, during pregnancy, it may be found in urine. Interestingly and curiously enough, there is no lactose in the milk of seals and other close relatives.

Kuhn (1949) has, however, found lactose to be present in the pollen of *Forsythia* plant.

Table 7-2. Percentage composition of milk

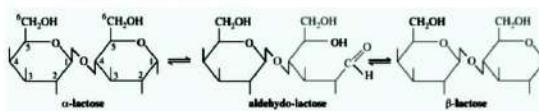
Contents	Human milk	Cow's milk
Water	88.5	87.0
Fat	3.3	3.5
Lactose	6.8	4.8
Casein	0.9	2.7
Lactalbumin + other proteins	0.4	0.7
Ash (containing minerals)	0.2	0.7

Chemistry. Lactose is prepared commercially from whey by evaporation to crystallization. Whey is obtained as a by-product in the manufacture of cheese. Lactose may be formed by elimination of a molecule of water from the glycosidic OH group of β -D-galactose and the alcoholic OH group on carbon atom 4 of D-glucose (Fig. 7-3). As D-glucose exists in 3 forms (α , aldehyde and β), it is obvious to have 3 forms of lactose viz., α form, aldehyde form and β -form. The aldehyde form is present in traces only. The formula shown in Fig. 7-3 is that of α -lactose.

Fig. 7-3. Structure of α -lactose

Lactose, in solution, consists of an equilibrium mixture of all the 3 forms. Using the simplified scheme of representing the Haworth formula, the 3 forms may be shown as below :

108 FUNDAMENTALS OF BIOCHEMISTRY



It may be emphasized that the α - and β -forms of lactose differ from each other only in hemiacetal configuration; the linkage in both the forms being β -1, 4-galactoside.

Properties. Lactose is a white, crystalline solid with a melting point 203°C (with decomposition) and is also dextrorotatory. The α - and β -forms have a specific rotation of $+90^\circ$ and $+35^\circ$ respectively. The equilibrium mixture has a specific rotation $+52.5^\circ$. It is less soluble in water and much less sweet than sucrose. The α -diastereoisomer is less soluble than the β -form. Sometimes, the α -isomer crystallizes in ice-cream, making the product seem sandy in texture. The more soluble β isomer is used in diets for infants. Since lactose has a free carbonyl group on carbon atom 1 of glucose unit, it is a reducing sugar. Hence, it exhibits mutarotation, reduces Cu^{++} ions to Cu^+ ions and forms osazone and oxime. It crystallizes in rhombic prisms with a mole of water.

Hydrolysis. Upon hydrolysis with the enzyme, *lactase*, it yields an equimolar mixture of glucose and galactose. It should be noted that lactase is a β -glycosidase, i.e., splits β -glycosides (hence, it is identical with emulsin). Lactose itself cannot be absorbed from the intestine into the bloodstream unless it is first hydrolyzed into monosaccharide units. The enzyme lactase is actively secreted by the intestinal mucosal cells in suckling infants and therefore the lactose, in the form of milk suckled by them, is easily hydrolyzed to the component monosaccharide units, thus effecting absorption by the intestinal tract. Although milk is the universal food of newborn mammals and one of the most complete human foods, many adult humans cannot digest milk because they are deficient in intestinal lactase, the enzyme that hydrolyzes the milk sugar, lactose. Such individuals may, therefore, show **lactose intolerance**, which is genetically determined. It is characterized by abdominal bloating, cramps, flatulence, colic pains, abnormal intestinal flow, nausea and watery diarrhea. All these symptoms appear within 30 to 90 minutes after ingesting milk or its unfermented by-products (fermented dairy products such as yogurt and cheese create no intolerance problems). Most Europeans and their descendants, which include the majority of North American whites, are most tolerant of milk. Many other ethnic groups are generally intolerant to milk, including the Japanese, Chinese, Jews in Israel, Eskimos, South American Indians and most African blacks. Only about 30% of North American blacks are tolerant; those who are tolerant are mostly descendants of slaves brought from east and central Africa, where dairying is traditional and tolerance to lactose is high. Lactase deficiency appears to be inherited as an autosomal recessive trait and is usually first expressed in adolescence or young adulthood. The prevalence of lactase deficiency in human populations varies greatly. For example, 3% of Danes are deficient in lactase, compared with 97% of Thais.

Lactose intolerance should not be confused with the genetic galactosemia.

MALTOSE (Malt sugar)

Occurrence. Maltose does not occur abundantly in nature. However, its occurrence has been occasionally reported. It is the major product of enzymic hydrolysis of starch. Sprouting cereal grains

are rich in *amylases* which split the starch present to dextrins and maltose. Malt, prepared from sprouting barley, is an excellent source of maltose.

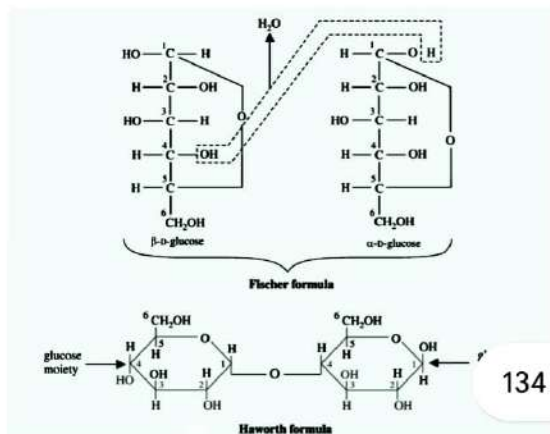
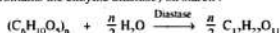


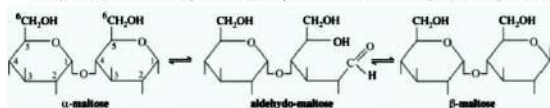
Fig. 7-4. Structure of β -maltose
O- α -D-glucopyranosyl-(1 $\rightarrow$ 4)- β -D-glucopyranose or G1c (α 1 $\rightarrow$ β 4)G1c

Chemistry. Maltose is the simplest of the disaccharides and is produced by the action of malt (which contains the enzyme *diastase*) on starch:



Maltose may be considered as originating by splitting out a molecule of water from the glycosidic OH group of α -D-glucose and the alcoholic OH group on carbon atom 4 of D-glucose (Fig. 7-4). Maltose, like lactose, has one free hemiacetal group. Consequently, it too exists in 3 forms: α , β and *aldehyde*. Maltose is usually found in β -form. Maltose, in solution, also exists as an equilibrium mixture of all the 3 forms.

It may be noted that the linkage in both α - and β -forms of maltose is always an α -1, 4-glucoside.



Contents

110 FUNDAMENTALS OF BIOCHEMISTRY

Properties. Maltose is a white crystalline solid, with a melting point 160–165 °C. It is soluble in water and is dextrorotatory. Because of the free aldehyde group, maltose is also a reducing sugar and forms osazone with phenyl hydrazine. It is also capable of exhibiting mutarotation.

Hydrolysis. Maltose is easily hydrolyzed into 2 identical units of glucose by dilute acids or by the enzyme, *maltase*, found in the intestine. It should also be noted that the enzyme maltase hydrolyzes or splits only α -glycosidic linkages and has no action upon β -glycosides. It has, hence, been used to ascertain the presence of this bond in disaccharides and polysaccharides. Hydrolysis may also be brought about by the enzyme, *diastase* (= *amylase*), found in sprouting barley.

CELLOBIOSE

Occurrence. It is probably present in only traces in nature. However, it is apparently released during the digestion of a polysaccharide, cellulose by the *cellulases* of microorganisms.

Chemistry. Cellobiose is identical with maltose except that the former has a β -1, 4-glucosidic linkage (Fig. 7-5) in contrast to the α -1, 4-glucosidic of the latter. It also exists in an equilibrium mixture of 3 forms: α , β and *aldehyde*. On hydrolysis, cellobiose yields glucose units only.

Properties. Cellobiose is a white crystalline solid with a melting point 225 °C. It is soluble in water and is dextrorotatory. Since cellobiose contains a free hemiacetal group, it is also a reducing sugar and undergoes mutarotation in aqueous solution and forms an osime or osazone. In fact, all the disaccharides with a free hemiacetal group (lactose, maltose, cellobiose, etc.) are reducing sugars and

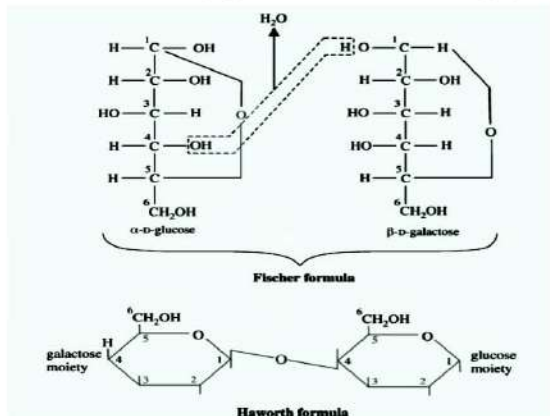


Fig. 7-5. Structure of α -cellobiose
O- β -D-glucopyranosyl-(1 $\rightarrow$ 4)- α -D-glucopyranose or G1c (β 1 $\rightarrow$ α 4)G1c

as such reduce cupric copper to cuprous oxide in the Fehling test and the silver ion to metallic silver in the silver mirror reaction. However, they are not as powerful reducing agents as are the

monosaccharides glucose, galactose and fructose, and do not readily reduce Barfoed's reagent to cuprous copper. The reason for this difference is that while C-1 of one of the two glucose residues is free to revert back to the carbonyl (= reducing) form, that on the second residue is stabilized in the cyclic form by the existence of the glycosidic linkage.

Hydrolysis. When hydrolyzed with dilute acids or by the enzyme *emulsin*, cellobiose yields identical units of D (+)-glucose. It may be noted that the enzyme *emulsin* splits only β -glycosidic linkages (*cf* maltase).

ISOMALTOSE

It is similar to maltose except that it has an α -1, 6-glucoside linkage (Fig. 7-6).

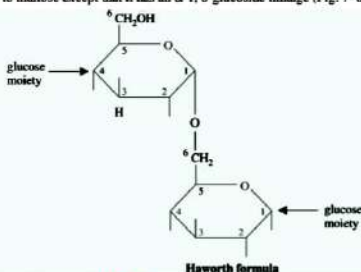


Fig. 7-6. Structure of isomaltose

O - α -D-glucopyranosyl-(1 $\rightarrow$ 6)- α -D-glucopyranose or Glc (α 1 $\rightarrow$ α 6)Glc

136 / 1254

TREHALOSE

Trehalose is a major constituent of the circulating fluid (hemolymph) of insects, in which it serves as an energy storage compound. It is also found in yeasts and other fungi. Here, the two anomeric carbon atoms of the two α -D-glucose moieties connect each other (Fig. 7-7). Consequently, it resembles sucrose in being a nonreducing sugar as it has no free aldehyde group. Trehalose does not form an osazone. On hydrolysis, it yields glucose. It is hydrolyzed by HCl to 2 moles of D-glucose.

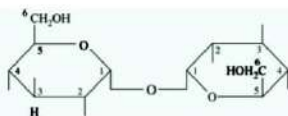


Fig. 7-7. Structure of trehalose

O - β -D-glucopyranosyl-(1 $\rightarrow$ 4)- α -D-glucopyranose or Glc (β 1 $\rightarrow$ α 4)Glc

Two isomeric forms of trehalose are recognized :

- Isotrehalose* (β -D-glucopyranosyl- β -D-glucopyranoside)
- Neotrehalose* (α -D-glucopyranosyl- β -D-glucopyranoside)

A summary chart of the important disaccharides is given on the next page.

Contents

112 FUNDAMENTALS OF BIOCHEMISTRY

Table 7-3. Summary of occurrence, structure and properties of some important disaccharides

Name	Occurrence	Constituent monosaccharides*	Oside linkage involved	Groups involved in linkage	Properties	Hydrolysis
Sucrose (Cane sugar)	Solely of plant origin : found in sugarcane, beet and maple	α -D-glucose and β -D-fructose	glycosidic linkage between C ₁ of glucose and C ₂ of fructose	glycosidic OH on C ₁ glycosidic OH on C ₂	Decarboxylatory, a non-reducing sugar, does not exhibit mutarotation; does not form osazone; Decarboxylatory; a reducing sugar; exhibits mutarotation; forms osazone; non-fermentable.	Yields an equimolar mixture of glucose and fructose on hydrolysis with invertase.
Lactose (Milk sugar)	Solely of animal origin; found in milk of mammals	β -D-galactose and D-glucose	β -1, 4-galactoside	glycosidic OH on C ₁ alcoholic OH on C ₄	Decarboxylatory; a reducing sugar; exhibits mutarotation; forms osazone; non-fermentable.	Yields an equimolar mixture of glucose and fructose on hydrolysis with lactase.
Maltose (Malt sugar)	Does not occur abundantly in nature; milk from sprouting barley; an excellent source	α -D-glucose and D-glucose	α -1, 4-glucoside	glycosidic OH on C ₁ alcoholic OH on C ₄	A reducing sugar; exhibits mutarotation; forms osazone; fermentable.	Yields identical units of glucose on hydrolysis with maltase or diastase.
Cellobiose	Probably present in only traces in nature.	β -D-glucose and D-glucose	β -1, 4-glucoside	glycosidic OH on C ₁ alcoholic OH on C ₄	A reducing sugar; exhibits mutarotation	Yields identical units of glucose on hydrolysis with β -glycosidases or with dilute acids.
Trehalose	Major sugar of insect hemolymph; also occurs in fungi and yeasts	α -D-glucose and α -D-glucose	α -1, 1-glucoside	glycosidic OH on C ₁ glycosidic OH on C ₁	A nonreducing sugar; does not exhibit mutarotation; does not form osazone	Yields glucose units on hydrolysis

* All the constituent monosaccharides are present in pyranose ring form except β -D-fructose which has a furanose ring structure.

CONTENTS

- Homopolysaccharides
 - Starch
 - Glycogen
 - Inulin
 - Cellulose
 - Pectin
 - Chitin
 - Hemicelluloses
 - Xylan
- Heteropolysaccharides
 - Hyaluronic Acid
 - Chondroitin
 - Chondroitin Sulfates
 - Dermatan Sulfate
 - Keratan sulfate
 - Heparin
- Analysis of Carbohydrates



The glistening outer skeleton of this grasshopper is primarily made up of chitin one of the most abundant biopolymer on the Earth.

CHAPTER

8

Carbohydrates-IV

Polysaccharides

139 / 1254

Polysaccharides (variously called as *glycans* or *polysaccharides* or *polysaccharides*) are high molecular weight carbohydrates which, on hydrolysis, yield mainly monosaccharides or product monosaccharides. They may also be regarded as

anhydrides of simple sugars. D-glucose is the component of polysaccharides. However, D- and L-galactose, D-mannose, D-xylose, L-arabinose as well as D-glucuronic, D-galacturonic, D-mannuronic acids, D-glucosamine, D-galactosamine and amino uronic acids also occur as constituents of polysaccharides. The various polysaccharides differ from one another not only in the composition of the constituent monosaccharide but also in the molecular weight, in the nature of the chain (whether linear or branched), in the type of glycosidic bond (whether α or β) and in the type of linkage (whether 1 $\rightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4, or 1 $\rightarrow$ 6) involved in the respective monosaccharide units. A great majority of carbohydrates of nature occur as polysaccharides. Chemically, the polysaccharides may be distinguished into **homopolysaccharides** (or **homoglycans**), which yield, on hydrolysis, a single monosaccharide and **heteropolysaccharides** (or

The small arrow shows the direction of the glycosidic bond which is always from the hemiacetal hydroxyl to some alcohol or hemiacetal group of the following sugar.

POLYSACCHARIDES 115

heteroglycans), which produce a mixture of monosaccharides on hydrolysis.

Based on their functional aspect, the polysaccharides may be grouped under two heads :

- (a) **Nutrient (or digestible) polysaccharides.** These act as metabolic reserve of monosaccharides in plants and animals, e.g., starch, glycogen and inulin.
- (b) **Structural (or indigestible) polysaccharides.** These serve as rigid mechanical structures in plants and animals, e.g., cellulose, pectin and chitin and also hyaluronic acid and chondroitin.

HOMOPOLYSACCHARIDES

These yield, on hydrolysis, a single monosaccharide. They serve as both storage (starch, glycogen, inulin) and structural (cellulose, pectin, chitin) polysaccharides.

Starch

Occurrence. It is the most important reserve food material of the higher plants and is found in cereals, legumes, potatoes and other vegetables. *More than half the carbohydrate ingested by humans is starch.* Sago starch is obtained from sago palm, *Metaxylon rumphii*; the arrowroot from *Maranta arundinacea* and the tapioca, a starchy food, from *Manihot utilissima*. It is usually present inside the plant cells as compact insoluble granules which may be spherical, lens-shaped or ovoid, and which have a distinctly layered structure.

Chemistry. Natural starches consist of two components : *amylose* (15–20%), a long unbranched straight-chain component and *amylopectin* (80–85%), a branched chain polysaccharide. Starch from waxy corn is notable as it consists practically of amylopectin component, there being no amylose.

α -amylose or simply amylose (Fig. 8-1) has a molecular weight range of 10,000 to 50,000. It may be formed in plant cells by elimination of a molecule of water from glycosidic OH group of one α -D-glucose molecule and alcoholic OH group on carbon 4 of the adjacent α -D-glucose molecule. The linkage in amylose is, thus, an α -1, 4-glucoside, like that in maltose. Enzymic hydrolysis of amylose with *amylase*, henceforth, yields maltose units mainly. Amylose may be considered as an anhydride of α -D-glucose units.

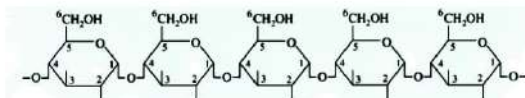


Fig. 8-1. Structure of amylose (= A-fraction)

Various types of starches differ in the amount of the amylose component present in them (Table 8-1).

β -amylose or isoamylose or amylopectin (Fig. 8-2) has a high molecular weight range of 50,000 to 1,000,000, thus indicating the presence of 300–5,500 glucose units per molecule. This possesses the same basic chain of α -1, 4-glucoside linkage like that of amylose but has, in addition, many side chains attached to the basic chain by α -1, 6-glucoside linkages, similar to those in isomaltose. It may, thus, be seen that the glucose unit, present at each point of branching, has substituents not only on carbon atoms 1 and 4 but also on carbon atom 6. In other words, these glucose units have 3 points of attachment to serve as branching points. The average chain length is about 24 glucose units. Amylopectin, upon incomplete hydrolysis, yields the disaccharide isomaltose.

Table 8-1. Amylose contents of some starches

Starch	Amylose component (in %)
Sugary mutant corn	70
Steadfast pea	67
Alderman pea	65
Buckwheat	28
Barley	27
Sorghum	27
Commercial corn	26
Wheat	25
White potato	23
Arrowroot	21
Sweet potato	20
Tapioca	18
Waxy barley	3
Waxy corn	0-6

(Adapted from Whistler RD and Smart CL, 1953.)

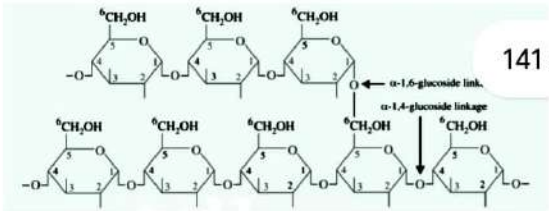


Fig. 8-2. Structure of amylopectin (= B-fraction)

The detailed structure of amylopectin (Fig. 8-3) is still speculative. The general consensus appears to be that amylopectin is composed of 3 types of chain, A, B and C; each chain consisting of 24 glucose units. A-chains are linked to B-chains in 1, 6, or 1, 3 manner. These B-chains are further linked to other B-chains, the terminal one of which is linked to a single C-chain, which characteristically possesses a free reducing group. There are 2 different types of combinations of these chains which explain reasonably well the properties of amylopectin: the laminated structure as proposed by Haworth in 1937 and the randomly highly-branched structure as given by Meyer in 1940.

Properties. Starch is a white soft amorphous powder and lacks sweetness. It is insoluble in water, alcohol and ether at ordinary temperature. The specific rotation of starch, $[\alpha]_D^{20}$ is $+196^\circ$. The microscopic form of the starch grains is characteristic of the source of starch. Starch, on partial hydrolysis by boiling with water under pressure at about 250°C , breaks down into large fragments called **dextrins**. The resulting dextrins then confer stiffness to clothes that have been starched and ironed. Starch molecule is highly hydrated since it contains many exposed hydroxyl groups. With the

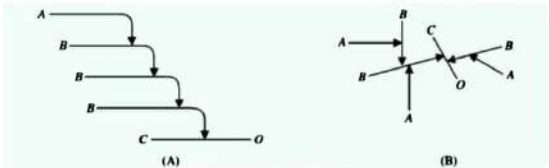


Fig. 8-3. Detailed structure of amylopectin

(A) Laminated structure (B) Randomly branched structure

result, starch, when extracted from granules with hot water, forms turbid colloidal solutions. In common with other polysaccharides, starch is a nonreducing carbohydrate since carbonyl groups of all units (except one of the two terminal ones) participate in the glycosidic linkages.

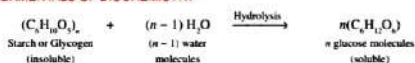
The two components of starch differ in following points :

- Amylose has a simpler structure and hence more soluble in water than amylopectin. Because of this difference in solubility, the two components may be separated partially by keeping starch in water for a prolonged period. When potatoes are boiled, amylose is extracted by the hot water turning it milky and opalescent. The amylopectin left behind accounts for most of the starch in boiled potatoes.
- Amylose is soluble without swelling in hot water whereas amylopectin swells in water.
- Amylose is readily dispersed in water but does not form the characteristic gel or starch paste. However, when dispersed in sufficient concentrations, the amylopectin forms the typical starch gel.
- Amylose produces a typical blue colour with iodine whereas amylopectin gives a purplish colour with iodine. In fact, amylose is an open, helical molecule. The inside diameter of the helix is just large enough to accommodate an iodine molecule and it is the consequent change in the light-absorbing properties of the halogen that is responsible for the blue colour given in the starch-iodine test. By contrast, most other polysaccharides, including amylopectin, give only dull, reddish-brown colours.

Hydrolysis. One enzyme, α -amylase, found in the digestive tract of animals (in saliva and the pancreatic juice) hydrolyzes the linear **amylose** chain by attacking α (1 $\rightarrow$ 4) linkages at random throughout the chain to produce a mixture of maltose and glucose. β -amylase, an enzyme found in plants, attacks the nonreducing end of amylose to yield successive units of maltose (The prefixes α and β used with the amylases do not refer to glycosidic linkage, but simply designate these two enzymes). **Amylopectin** can also be attacked by α - and β -amylases, but the α (1 $\rightarrow$ 4) glycosidic bonds near the branching point in amylopectin and the α (1 $\rightarrow$ 6) bond itself are not hydrolyzed by these enzymes. A separate "debranching" enzyme, an α (1 $\rightarrow$ 6) glucosidase can hydrolyze the bond at the branch point. Therefore, the combined action of α -amylase and the α (1 $\rightarrow$ 6) glucosidase will hydrolyze amylopectin ultimately to a mixture of glucose and maltose. Starch is readily hydrolyzed by mineral acid with the final production of glucose. The course of hydrolysis may be followed by the gradual change in colour when treated with iodine (blue $\longrightarrow$ purple $\longrightarrow$ red $\longrightarrow$ none) and also by the increase in concentration of reducing sugar. A compound which yields only glucose molecules on hydrolysis is called a **glucosan**. Hence, starch is a glucosan.

The suffix **-an** is used for designating polymers; for example, starch is a glucosan as it yields only glucose molecules on hydrolysis; similarly, inulin is a fructosan as it yields only fructose molecules on hydrolysis.

118 FUNDAMENTALS OF BIOCHEMISTRY



If hydrolysis of starch is interrupted early, dextrins are formed. The dextrins are white powders, and are used for making adhesives and confectionery, for sizing paper, etc. These materials possess adhesive properties and are used as adhesives on paper products.

Glycogen

Occurrence. It is the major reserve food in animals and is often called **'animal starch'**. Glycogen is stored in the liver and muscles of animals. It is especially abundant in the liver where it may attain up to 7% of the wet weight. It occurs in animal cells as particles much smaller than the starch grains.

Chemistry. Glycogen is a branched-chain polysaccharide and resembles amylopectin very much in structure, rather than amylose (for which there is no 'animal' equivalent), but has somewhat more glucose residues per molecule and about one-and-a-half times as many branching points. Also the chains are shorter (10–20 glucose units), and hence the molecule is even more highly branched and more compact. These differences, however, do not alter the functional behaviour of the molecule to any significant extent. Its molecular weight is $1-2 \times 10^7$.

Properties. It is a white powder and is more soluble in water than amylopectin. Hence, it readily forms suspension even in cold water. Glycogen is precipitated from aqueous solutions by addition of ethyl alcohol and is fairly stable in hot alkali. It is a nonreducing sugar and gives red iodine. The red colour, however, disappears on boiling and reappears on cooling.

Hydrolysis. On incomplete hydrolysis with α -1,4-glucanmaltohydrolase, glycogen

However, the acids completely hydrolyze it into glucose.

Although of animal origin, a glycogen-like polysaccharide has also been described in maize.

143 / 1254

Inulin

Occurrence. It is the storage form of carbohydrate in the members of the family—Compositae, such as dahlias (*Dahlia sp.*), artichokes (*Helianthus tuberosus*) and dandelions (*Taraxacum officinale*). Inulin is stored in the tubers of the dahlia and artichoke and in the roots of dandelion. It is also found in onion and garlic.

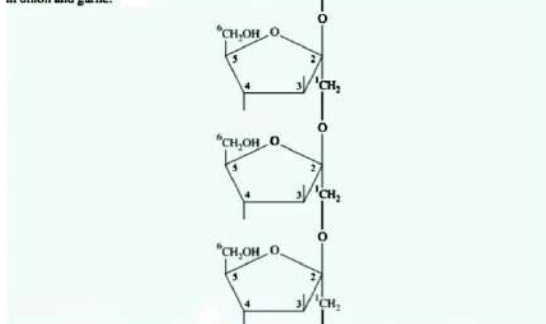


Fig. 8-4. Structure of inulin

[The two glucose units have not been shown for the sake of simplification.]

POLYSACCHARIDES 119

Chemistry. Inulin (Fig. 8-4) has a molecular weight of about 5,000 and consists of about 30–35 fructose units per molecule. It is formed in the plants by eliminating a molecule of water from the glycosidic OH group on carbon atom 2 of one β -D-fructose unit and the alcoholic OH group on carbon atom 1 of the adjacent β -D-fructose unit. In inulin, the fructose residues (in furanose form) are, thus, joined together in a straight chain by β -2, 1-fructosidic linkages. Inulin is, therefore, 2 $\rightarrow$ 1- β -polyfructose.

On hydrolysis, however, inulin also yields a small amount of glucose besides fructose molecules. It is now thought that there are 2 glucose units in the inulin molecule, one somewhere in the centre and the other at the reducing end of the chain.

Properties. It is a white tasteless powder and is insoluble in cold water. In warm water, it easily dissolves forming a colloidal solution which does not form a gel on cooling. Inulin solution does not give any colour with iodine. It is a nonreducing sugar.

Hydrolysis. Inulin, on complete hydrolysis by dilute acids, yields D-fructose and hence it is a **fructosan or levan**.

Cellulose

Occurrence. Cellulose is not only the most abundant extracellular structural polysaccharide of the plant world but is also *undoubtedly the most abundant of all biomolecules in the biosphere*. It is present in all land plants, but is completely lacking in meat, egg, fish and milk. It is, however, not metabolized by the human system. Some 10^{13} kg of cellulose is synthesized and degraded on earth each year! It shares as much as 50% of all the carbon in vegetation. It is the most widely distributed carbohydrate of the plants, although certain tunicates also possess it. Cellulose occurs in the cell walls of plants where it contributes in a major way to the structure of the organism. Lacking a skeleton of bone onto which organs and tissues may be organized, the higher plant relies on its cell walls to bear its own weight, whether it is the smallest *Wolffia* flower or the tallest *Sequoia* tree. Cellulose is often found associated with other structural substances such as lignin. Plant residues in soil consist of 40–70% cellulose. This high cellulose content emphasizes the importance of cellulolytic organisms in the mineralization process and in the carbon cycle. It occurs in nearly pure form in cotton (94%), flax (90%) and in the wood of many plants (60%). It is not usually found in bacterial cell walls, but it is the substance which keeps the cells of *Sarcina ventriculi* together in large packets or bunches.

Chemistry. The molecular weight of cellulose (Fig. 8-5) ranges between 200,000 and 2,000,000, thus corresponding to 1,250–12,500 glucose residues per molecule. It may be formed by taking out a molecule of water from the glycosidic OH group on carbon atom 1 of one β -D-glucose molecule and the alcoholic OH group on carbon atom 4 of the adjacent β -D-glucose molecule. It, thus, resembles in structure with amylose except that the glucose units are linked together by β -1, 4-glucosidic linkages. Cellulose may, henceforth, be regarded as an anhydride of β -D-glucose units.

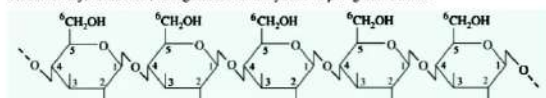


Fig. 8-5. Structure of cellulose

The correlation between the structure of cellulose molecule and the function it performs is illustrated in Fig. 8-6.

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

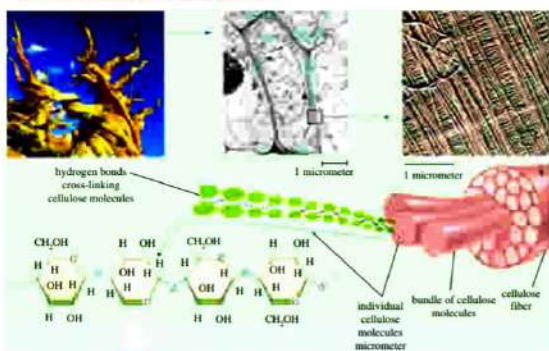


Fig. 8-6. Interrelation between cellulose structure and function

Cellulose, like starch, is composed of glucose monomers but the orientation of the bond between cellulose is different such that every other glucose molecule is "upside down." Unlike starch, cellulose structural strength, due partly to the difference in bonding and partly to the arrangement of parallel cellulose into long, cross-linked fibres. Plant cells often deposit cellulose fibres in layers that run each other, resulting in resistance to tearing in both directions. The final product can be incredibly tough, as this 3000-year-old bristlecone pine testifies.

145 / 1254

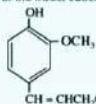
Because of the lack of chemical reactivity, the cellulose is of no nutritive value unlike starch. However, the same characteristics have made it so useful as fibres for paper and cloth. In man, the cellulose is not digested since it does not possess an enzyme capable of catalyzing the cleavage of β -glucoside bonds. Hence, it serves as an important source of "bulk" in the diet. However, the ruminants (cattle, sheep, goats, camels, giraffes) and certain wood-eating insects are able to digest it because the microorganisms, present in their digestive tract, do possess such digesting enzymes. Also certain wood-eating insects such as termites readily digest cellulose because their intestinal tract harbours a parasitic protozoan, *Trichonympha*. The latter secretes *cellulase*, a cellulose-hydrolyzing enzyme which enables termites to digest wood. Wood-rot fungi and bacteria also produce cellulase.

Hydrolysis. Cellulose is a relatively inert material and is completely degraded only under the most drastic conditions. For example, it may be hydrolyzed to glucose when treated with conc. H_2SO_4 or HCl or with conc. NaOH.

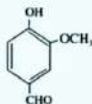
Biological importance. Huge quantities of cellulose are made annually by the plants of the globe. It is estimated that every day some 50 kg of cellulose is synthesized by the plant world for each human being on the earth. Cellulose is also a very useful substance. Wood, cotton, paper and cardboard are all largely cellulose. Dried wood contains about 65% cellulose, 30% lignin and 2% pentosans. In the manufacture of **paper**, wood is cooked with water, lime and SO_2 to remove lignin leaving behind the cellulose. These fibres are then spread into mats and rolled into sheets which may be coated to provide a smooth writing surface. When heavy paper is treated with H_2SO_4 , the translucent parchment paper is obtained.

The amount of **lignin** in wood varies in different species and with age in the same species. It makes up about 23-33% of softwoods, but only about 16-25% of hardwoods.

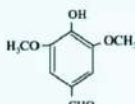
When wood passes from the green pliable stage, it is hardened by the interpenetration of a highly complex non-polysaccharide polymeric substances called **lignin**. This substance is formed by the cross-linking of a number of non-sugar aromatic compounds, such as coniferyl alcohol, vanillin, syringaldehyde etc. The process of lignification has some chemical resemblance to that which strengthens the proteins of the insect cuticle.



Coniferyl alcohol



Vanillin



Syringaldehyde

Fruits with seeds and lignified cells of pear are important sources of lignin. Woody nondigestible parts of vegetables too are rich in lignin. Lignin is water-insoluble and non-digestible. It adsorbs organic substances like bile acid from the intestinal lumen and thus facilitates its fecal excretion. Lignin is an intractable material that is difficult to remove from the wood.

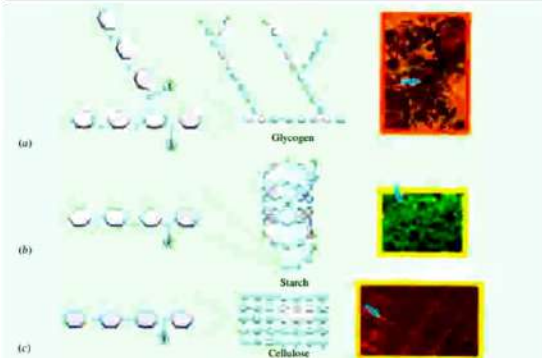


Fig. 8.7. Three polysaccharides with identical sugar monomers but dramatically different properties Glycogen (a), starch (b), and cellulose (c) are each composed entirely of glucose subunits, yet their physical and chemical properties are very different due to the distinct ways that the monomers are linked together (three different types of linkages are indicated by the circled numbers). Glycogen molecules are the most highly branched, starch molecules assume a helical arrangement, and cellulose molecules are unbranched and highly extended. Whereas glycogen and starch are energy stores, cellulose molecules are bundled together into tough fibres that are suited for their structural role. Colourized electron micrographs show glycogen granules in liver cell, starch grains (amyloplasts) in a plant seed, and cellulose fibres in a plant cell wall; each is indicated by an arrow.

[Photo insets : (Top) Don Fawcett; (Centre) Jeremy Burgess; (Bottom) Cubisco]

Cotton treated with alkali becomes somewhat translucent, acquires a silky lustre, and is called **mercerized cotton**. Rayon is made from cellulose by dissolving the material in NaOH and CS₂. The viscous material is forced through small jets into an acid solution of sodium bisulfite, regenerating the cellulose in small threads.

Cellulose nitrates are used in explosives, lacquers, celluloid and collodion. **Cellulose acetate** is widely used in photographic film and packing materials.

It may, thus, be easily visualized that the way the constituent monomeric units of various polysaccharides (glycogen, starch and cellulose, for example) are linked together has a profound bearing on the physical and chemical properties of these molecules, although the polysaccharides are made up of identical glucose units (Fig. 8-7).

Pectin

Occurrence. Pectins are found as intercellular substances in the tissues of young plants and are especially abundant in ripe fruits such as guava, apples and pears. They are components of the middle lamella found between the cell wall and adjacent cells.

Chemistry. Pectin is a polysaccharide of α -D-galacturonic acid where some of the free carboxyl groups are, either partly or completely, esterified with methyl alcohol and others are combined with calcium or magnesium ions. Chemically, they are called polygalacturonides (Fig. 8-8) and have high molecular weights.

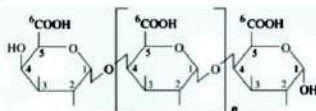


Fig. 8-8. Structure of α -polygalacturonic acid

Biological properties. The importance of pectin rests not so much on their quantity, rather on the functional part they play in plant stability and solidity. The characteristic property of pectins is the ability of their solutions to gelate, i.e., form jellies. Pectins, along with sugar, furnish the gelling characteristics of jellies and preserves made from fruit.

Chitin

Occurrence. Chitin is probably the most abundant polysaccharide of nature after cellulose. It is also one of the most abundant biopolymer on the earth. It is found in fungi but principally among the arthropods (crabs and insects). The armour of crabs and the exoskeleton (= cuticula) of insects consist mostly of chitin and some protein. The excellent mechanical properties of insect skeleton are due to chitin. The chitin framework of lobster and crabs shells is impregnated and hardened with calcium carbonate.

Chemistry. Chitin (Fig. 8-9) is closely related to cellulose. Here the alcoholic OH group on carbon atom 2 of β -D-glucose units is replaced by an N-acetyl amino group. It is, thus, a linear polymer of N-acetyl-D-glucosamine units joined together by β -1,4-glycosidic linkages. Chitin's pronounced stability is based on the hydrogen bonding of the N-acetyl side chains. Like cellulose, chitin is indigestible by vertebrate animals.

Hydrolysis. On hydrolysis with mineral acids, chitin yields 2 final end products, namely, glucosamine and acetic acid. Glucosamine is an important component of some glycoproteins (=mucoproteins) such as mucin of saliva. *Chitinases* (from the gastric juice of snails or from bacteria), however, decompose the chitin to N-acetylglucosamine.

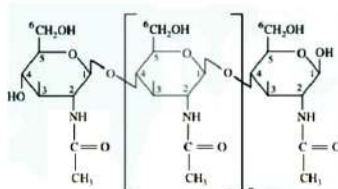


Fig. 8-9. Structure of chitin, a unique polysaccharide

Chitin has the same "alternating upside down" bonding of glucose molecules as cellulose has, but in chitin the glucose subunits are modified by the replacement of one of the hydroxyl groups with a nitrogen-containing functional group (yellow). Tough, slightly flexible chitin supports the otherwise soft bodies of arthropods (insects, spiders, and their relatives) and fungi.

Hemicelluloses

Xylan and other related substances are collectively referred to as hemicelluloses. They are not structurally related to cellulose, nor do they contain the same building blocks but they are, at least partially, soluble in water or alkali. The hemicelluloses consist of either pentoses (xylose, arabinose) or hexoses (glucose, mannose, galactose) as well as uronic acids. Unrefined cereals, some fruits and vegetables and whole wheat are rich sources of hemicelluloses. They adsorb water and are partially digestible. They function as storage and supporting substances in plants. The designation 'hemicelluloses' has been discontinued since a large number of similar polysaccharides were discovered in fungi and bacteria.

Xylan is representative of the group and is, therefore, being discussed below.

Xylan

Occurrence. Xylan is the next most abundant and widely distributed carbohydrate after cellulose. Straw and bark consist of upto 30% xylan; residues of sugarcane also contain about 30% xylan. It also occurs in conifer wood (7-12%) and deciduous wood (20-25%).

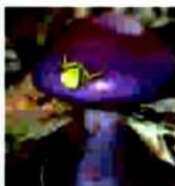
Chemistry. Xylan is a linear homopolymer of D-xylose (an aldopentose) in β -1 $\rightarrow$ 4 linkage, bearing side chains of 4-O-methylglucuronic acid and/or arabinose. Xylan can be derived from a cellulose chain by substituting hydrogen atoms for the —CH₂OH groups, but its polymer size (number of units per polymer) is considerably lower (30-100).

Properties. Xylan is more rapidly degraded by a large number of microorganisms than cellulose. Many cellulose-degrading organisms (such as *Sporocytophaga myxococcoides*) also produce xylanase. The ability to utilize xylans is very common among fungi. Xylan is even an excellent substrate for the cultivation of mushrooms. In bacteria, xylanase is formed constitutively in some organisms (e.g., Clostridia) and in others it is inducible by xylan.

HETEROPOLYSACCHARIDES

These yield, on hydrolysis, a mixture of monosaccharides. They are numerous in both plants and animals.

The 'specific soluble sugar' of pneumococcus type III (Fig. 8-10) is one of the simplest heteropolysaccharides. It contains repeating units of a mixed disaccharide consisting of glucose and glucuronic



124 FUNDAMENTALS OF BIOCHEMISTRY

acid. Apparently, on hydrolysis, it yields equimolar amounts of glucose and glucuronic acid.

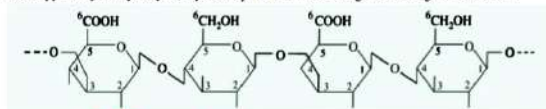


Fig. 8-10. Structure of 'specific soluble sugar' of pneumococcus type III

Mucopolysaccharides

Polysaccharides that are composed not only of a mixture of simple sugars but also of derivatives of sugars such as amino sugars and uronic sugars are called mucopolysaccharides. They are gelatinous substances of high molecular weights (up to 5×10^6). Most of these act as structural support material for connective tissue or mucous substances of the body. They serve both as a lubricant and a cementing substance.

Structurally, they have a common feature. They consist of disaccharide units in which a uronic acid is bound by a glycosidic bond to the C₃ of an acetylated amino acid (1 $\rightarrow$ 3 linkage). These disaccharide residues are polymerized by 1 $\rightarrow$ 4 linkages to give a linear macromolecule. The uronic and sulfuric acid residues impart a strong acidic character to these substances. Table 8-1 lists the structural features of some common mucopolysaccharides.

Table 8-2. Some common mucopolysaccharides

Mucopolysaccharide	Two components of the disaccharide units	Linkage	
Hyaluronic acid	D-glucuronic acid + N-acetyl-D-glucosamine	β -1 $\rightarrow$ 3	β -1 $\rightarrow$ 4
Chondroitin sulfate A	D-glucuronic acid + N-acetyl-D-galactosamine-4-sulfate	β -1 $\rightarrow$ 3	β -1 $\rightarrow$ 4
Chondroitin sulfate C	D-glucuronic acid + N-acetyl-D-galactosamine-6-sulfate	β -1 $\rightarrow$ 3	β -1 $\rightarrow$ 4
Dermatan sulfate (Chondroitin sulfate B)	L-iduronic acid + N-acetyl-D-galactosamine-4-sulfate	α -1 $\rightarrow$ 3	β -1 $\rightarrow$ 4
Keratosulfate	D-galactose + N-acetyl-D-glucosamine-6-sulfate	β -1 $\rightarrow$ 4	β -1 $\rightarrow$ 3

* Linkage of the first column represents the linkage involved between the two monosaccharides of the disaccharide unit whereas the linkage of the second column is the one involved between the repeating disaccharide units.

Hyaluronic acid

Occurrence. It is the most abundant member of mucopolysaccharides and is found in higher animals as a component of various tissues such as the vitreous body of the eye, the umbilical cord and the synovial fluid of joints. The high viscosity of the synovial fluid and its role as biological lubricant is largely due to the presence of its hyaluronic acid content (about 0.03%). Frequently, it is prepared from umbilical cord.

Chemistry. Hyaluronic acid (Fig. 8-11) has the least complicated structure among mucopolysaccharides. It is a straight-chain polymer of D-glucuronic acid and N-acetyl-D-glucosamine (NAG) alternating in the chain. Its molecular weight approaches approximately, 5,000,000. Here, apparently two linkages are involved, β -1 $\rightarrow$ 3 and β -1 $\rightarrow$ 4. Hyaluronic acid is an acidic substance,

149 / 1254

POLYSACCHARIDES 125

because the carboxyl groups are largely ionized at cellular pH.

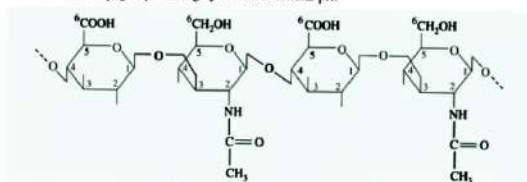


Fig. 8-11. Structure of hyaluronic acid

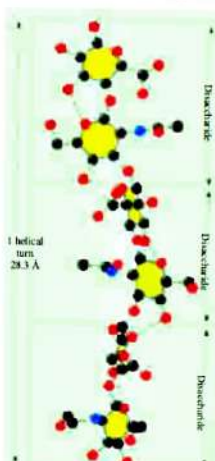


Fig. 8-12. The X-ray fibre structure of Ca^{2+} hyaluronate. The hyaluronate polyanion forms an extended left-handed single-stranded helix with three disaccharide units per turn that is stabilized by intramolecular hydrogen bonds (dashed lines). H atoms and Ca^{2+} ions are omitted for clarity.

The Ca^{2+} polyanionic form of hyaluronic acid is depicted in Fig. 8-12. It is an extended left-handed helix, which is stabilized by hydrogen bonds.

Hydrolysis. Hyaluronic acid, upon hydrolysis, yields an equimolar mixture of D-glucuronic acid, D-glucosamine and acetic acid.

It is noteworthy that hyaluronic acid is split swiftly by the enzyme *hyaluronidase*. The enzyme brings about depolymerization of hyaluronic acid (leading to a drop in its viscosity) and cleavage to smaller fragments. Hyaluronidase is a 'spreading factor' of skin and connective tissue. The depolymerization effect allows any foreign bodies (such as pigments, pathogenic bacteria) to penetrate the tissue, since the cementing substance is being dissolved. The enzyme also has a physiologic role in fertilization. The sperm is rich in hyaluronidase and hence can advance better in the cervical canal and finally fertilize the ovum.

Chondroitin

Occurrence. Chondroitin is of limited distribution. It is found in cartilage and is also a component of cell coats. It is a parent substance for two more widely distributed mucopolysaccharides, chondroitin sulfate A and chondroitin sulfate B.

Chemistry. Chondroitin is similar in structure to hyaluronic acid except that it contains galactosamine rather than glucosamine. It is, thus, a polymer of β -D-glucuronido-1, 3-N-acetyl-D-galactosamine joined by β -1 $\rightarrow$ 4 linkages.

Hydrolysis. On hydrolysis, chondroitin produces equimolar mixture of D-glucuronic acid, D-galactosamine and acetic acid.

Chondroitin Sulfates

Occurrence. The two chondroitin sulfate A and C are widely distributed and form major structural components of cartilage, tendons and bones. They are very often associated with collagen and probably with other proteins too.

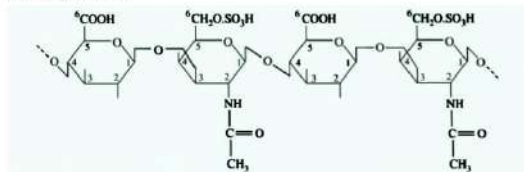


Fig. 8-13. Structure of chondroitin sulfate C

Chemistry. Chondroitin sulfates may be regarded as derivatives of chondroitin where, in the galactosamine moiety, a sulfate group is esterified either at carbon 4 as in chondroitin sulfate A or at carbon 6 as in chondroitin sulfate C (Fig. 8-13). The two linkages involved in both types of chondroitin sulfate would, obviously, be the same. These are β -1 $\rightarrow$ 3 and β -1 $\rightarrow$ 4.

Hydrolysis. On hydrolysis, chondroitin sulfates A and C yield approximately equivalent amounts of D-glucuronic acid, D-galactosamine, acetic acid and sulfuric acid.

Dermatan Sulfate

Dermatan sulfate is a mucopolysaccharide structurally similar to chondroitin sul. that the D-glucuronic acid is replaced by L-iduronic acid (the two uronic acids differ in co. only at C₅). The two linkages involved are α -1 $\rightarrow$ 3 and β -1 $\rightarrow$ 4. Dermatan sulfate is also known by its conventional name, **chondroitin sulfate B**. This is, however, a misnomer since dermatan sulfate differs from both chondroitin sulfate A and C in the composition of their repeating disaccharide unit.

Keratosulfate

Keratosulfate differs from other mucopolysaccharides in that the uronic acid component is replaced by D-galactose. Here, the second acetylated amino sugar component (which is N-acetyl-D-glucosamine in this case) is esterified by a sulfate group at carbon 6. Although, the two alternating linkages involved are β -1 $\rightarrow$ 4 and β -1 $\rightarrow$ 3, in this case the linkage between the repeating disaccharide units is β -1 $\rightarrow$ 3 rather than β -1 $\rightarrow$ 4.

Heparin

Occurrence. Related to the sulfated mucopolysaccharides is heparin. It is present in liver, lung, arterial walls and, indeed, wherever mast cells are found, possibly for the purpose of neutralizing biogenic amines (e.g., histamine).

Chemistry. Heparin (Fig. 8-14) is a heteropolysaccharide composed of D-glucuronic acid units, most of which (about 7 out of every 8) are esterified at C₂ and D-glucosamine-N-sulfate (= sulfonylaminoglucose) units with an additional O-sulfate group at C₆. Both the linkages of the polymer are alternating α -1 $\rightarrow$ 4. Thus, the sulfate content is very high and corresponds to about 5-6 molecules per tetrasaccharide repeating unit. The relative positions of the sulfate residues may also vary. Its molecular weight ranges between 17,000 and 20,000.

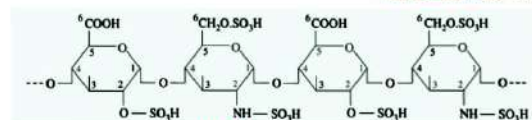


Fig. 8-14. Structure of heparin

Heparin acts as an anticoagulant. It prevents coagulation of blood by inhibiting the prothrombin-thrombin conversion. This eliminates the effect of thrombin on fibrinogen.

The heteropolysaccharides described above contain two different sugars in their repeating units or monomers. However, there are many heteropolysaccharides which contain more than 2 carbohydrates in their repeating units. Vegetable 'gums' and agar-agar are two notable examples.

Vegetable 'gums'. These contain as many as 4 different monosaccharide units. Most common of these are D-glucuronic acid, D-mannose, D-xylose (the second most abundant sugar in the biosphere) and L-rhamnose.

Agar-agar. It is a gelatinous polysaccharide produced by certain marine red algae (= rhodophyceae members) such as species of *Gelidium*, *Gracilaria*, *Gigartina*, *Eucheuma*, *Campylaea*, *Amphelitia*, *Hypnea*, etc. Japan is the major producer of agar-agar. It consists of D- and L-galactose, predominantly with 1 $\rightarrow$ 3 bonds and always contains some amount of sulfuric acid. On hydrolysis, it yields D- and L-galactose in a ratio of 9 : 1. It forms highly viscous gels and has melting point between 90 and 100°F. At lower temperature, it solidifies. It is insoluble in cold but soluble in hot water.

Agar-agar is of great value in the preparation of foodstuffs, particularly as articles of diet for invalids. It is used extensively in biological laboratories as the base material for the preparation of culture media, especially for bacteria and fungi. It is also used in the preparation of some medicines, and in cosmetics and leather industry. It is largely used as a solidifying agent in desserts, as a laxative and as a sizing material for textiles. It also finds application as an emulsifier in dairy products. Agar-agar is not utilized by man and hence adds to the bulk of the feces and helps in its propulsion.

A wide variety of heteropolysaccharides are found in animals and microbes. These, on hydrolysis, yield various types of monosaccharides. These more complex polysaccharides are found bound to proteins (**glycoproteins**) and lipids (**glycolipids**). Carbohydrates occur in them as an oligosaccharide moiety.

Analysis of Carbohydrates

Mixtures of carbohydrates can be resolved into their individual components (Fig. 8-15) by many techniques (such as differential centrifugation, ion exchange chromatography and gel filtration) which are also used in protein and amino acid separation. Each carbohydrate separated in the first stage of analysis is subjected to the following 3 analytic routes for complete characterization. The 3 routes are as follows :

1. Hydrolysis in strong acid yields a mixture of monosaccharides, which after conversion to suitable volatile derivatives may be separated, identified and quantified by gas-liquid chromatography to yield the overall composition of the polymer.
2. For simple linear polymers such as amylose, the position of the glycosidic bond between monosaccharides is determined by treating the intact polysaccharide with methyl iodide to convert all free hydroxyls to acid-stable methyl esters. When the methylated polysaccharide is hydrolyzed in, the only free hydroxyls present in the monosaccharides produced are those that were involved in glycosidic bonds.

3. To determine the stereochemistry at the anomeric carbon, the intact polymer is tested for sensitivity to purified glycosidases known to hydrolyze only α - or only β -glycosides. Total structure determination for complex heteropolysaccharides is much more difficult. Stepwise degradation with highly specific glycosidases, followed by isolation and identification of the products, is often helpful. Mass spectral analysis and high-resolution nuclear magnetic resonance (NMR) spectroscopy are highly powerful analytic tools for carbohydrates.

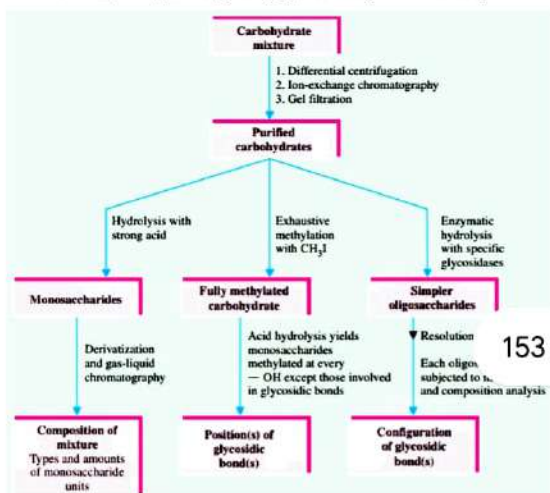


Fig. 8-15. Some common methods for the analysis of complex carbohydrates

Each carbohydrate separated in the first stage of the analysis requires all 3 analytic routes for complete characterization.

A summary chart of the important homopolysaccharides is given on the next page.

Contents

POLYSACCHARIDES 129

Table 8-3. Summary of occurrence, structure and properties of some important homopolysaccharides

Name	Occurrence	Constituent monosaccharide*	Oside linkage(s) involved	Groups involved in linkage	Properties	Hydrolysis
Sucrose	Major reserve food of higher plants; found in cereals, legumes, potatoes and other vegetables	α -D-glucose	(In amylose) α -1, 4-glycoside (In amylopectin) α -1, 4-glycoside and α -1, 6-glycoside Chains are shorter and much more branched than amylopectin	glycoside: OH on C_1 alcoholic: OH on C_2 glycoside: OH on C_1 alcoholic: OH on C_2	amylase gives blue colour with iodine; amylopectin gives purple colour with iodine; deacetylation	Yields glucose units only on complete hydrolysis, hence a glucose
Glycogen	Major reserve food of animals; found in liver and muscles of animals	α -D-glucose	α -1, 6-glycoside and α -1, 4-glycoside Chains are shorter and much more branched than amylopectin	glycoside: OH on C_1 alcoholic: OH on C_2 alcoholic: OH on C_2	a nonreducing sugar; gives red colour with iodine	Yields glucose units only on complete hydrolysis, hence a glucose
Inulin	Found in tubers and roots of certain composites (e.g., dandelion, artichokes and dahlias)	β -D-fructose	β -2, 1-fructoside	glycoside: OH on C_2 alcoholic: OH on C_1	a nonreducing sugar; gives no colour with iodine	Yields fructose units only on complete hydrolysis, hence a fructose
Cellulose	Most abundant organic compound found in world; widely distributed in plants such as cotton and flax	β -D-glucose	β -1, 4-glycoside	glycoside: OH on C_1 alcoholic: OH on C_4	gives no colour with iodine	Yields glucose units only on complete hydrolysis, hence a glucose
Pectin	Fruits of many plants	α -D-galacturonic acid	α -1, 4-galactoside	glycoside: OH on C_1 alcoholic: OH on C_4	a nonreducing sugar	Yields N-acetyl-glucosamine on hydrolysis with chitinase
Chitin	Most abundant in nature after cellulose; found in fungi and arthropods	N-acetyl-D-glucosamine	β -1, 4-glycoside	glycoside: OH on C_1 alcoholic: OH on C_4	—	—

* All the constituent monosaccharides are present in pyranose ring form except β -D-fructose which has a furanose ring structure.

Lipid

CONTENTS

- Importance
- Definition
- Alcohols
- Fatty Acids
 - Saturated Fatty Acids
 - Unsaturated Fatty Acids
 - Hydroxy or Oxygenated Fatty Acids
 - Cyclic Fatty Acids
- Biological Roles of Lipids

CHAPTER

12

Lipids—I

General Structure

IMPORTANCE



Triacylglycerols fuel the long migration flights of the golden plover (*Pluvialis dominica*).

(Courtesy: Gerard Fuhner)

The lipids are important constituents of the diet because of their high energy value and also because of the fat-soluble vitamins and the essential fatty acids found with the fat of the natural foodstuffs. In the body, the fats serve as efficient source of energy which is stored in the adipose tissues. They also serve as an insulating material in the subcutaneous tissues and around certain organs. Fats combined with proteins (lipoproteins) are important constituents of the cell membranes and mitochondria of the cell.

Proteins, polysaccharides, DNA and RNA are macromolecules. *Lipids are not generally classed as macromolecules*, even though they share some of their features: for example, most are synthesized as linear polymers of a smaller molecule (the acetyl group on acetyl-CoA), and they self-assemble into larger structures (membranes). It is noteworthy that water and protein comprise most of the mass of both mammalian and bacterial cells (Table 12-1).

STRUCTURE OF LIPIDS 231

Table 12-1. Approximate chemical composition of a typical bacterium and a typical mammalian cell

Component	Percent of total cell weight	
	<i>E. coli</i> bacterium	Mammalian cell
H ₂ O	70	70
Inorganic ions (Na ⁺ , K ⁺ , Mg ²⁺ , Ca ²⁺ , Cl ⁻ etc)	1	1
Miscellaneous small metabolites	3	3
Proteins	15	18
RNA	6	1.1
DNA	1	0.25
Phospholipids	2	3
Other lipids	—	2
Polysaccharides	2	2
Total cell volume:	$2 \times 10^{-12} \text{ cm}^3$	$4 \times 10^{-6} \text{ cm}^3$
Relative cell volume:	1	2,000

DEFINITION

The lipids are a heterogeneous group of compounds related to fatty acids and include fats, oils, waxes and other related substances. These are oily or greasy organic substances, relatively insoluble in water and considerably soluble in organic solvents like ether, chloroform and benzene. They are, thus, **hydrophobic** in nature. These are variously called as *lipins* or *lipoids*. The latter term is, however, sometimes used to refer "fat-like" substances which may not actually be related to the fatty acids. The term 'lipid' was first used by the German biochemist Bloor in 1943 for a major class of tissue components and foodstuffs. The structure of some common fats and fatty acids is presented in Fig. 12-1.

Chemical Abstracts and some biochemical journals use the spelling 'lipide'.

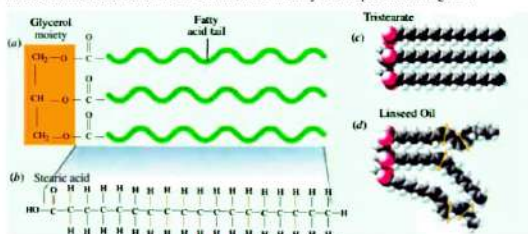
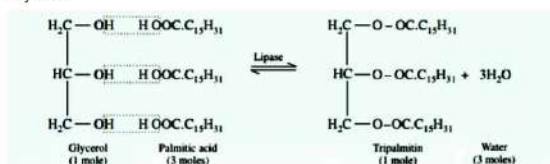


Fig. 12-1. Fats and fatty acids

(a) Basic structure of a triacylglycerol (also called a triglyceride or a neutral fat). The glycerol moiety, indicated in orange, is linked by three ester bonds to the carboxyl groups of three fatty acids whose tails are indicated in green. (b) Stearic acid. It is an 18-carbon saturated fatty acid that is common in animal fats. (c) Space-filling model of tristearate. It is a triacylglycerol containing three identical stearic acid chains. (d) Space-filling model of linseed oil. It is a triacylglycerol containing two different unsaturated fatty acids, derived from flax seeds. The sites of unsaturation, which produce links in the molecule, are indicated by the yellow-orange bars.

232 FUNDAMENTALS OF BIOCHEMISTRY

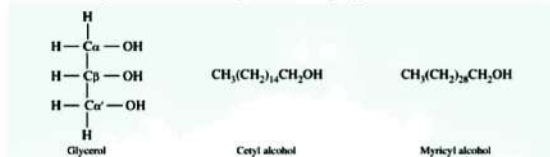
Chemically, the fats are defined as the esters of glycerol and fatty acids or as the triglycerides of fatty acids.



Both these components are described below for a better understanding of the lipids.

ALCOHOLS

Alcohols found in lipid molecules may be **saturated**. These commonly include glycerol, cholesterol and higher alcohols such as cetyl alcohol and myricyl alcohol.



In the structural formula of glycerol, the C atoms are designated by **stereospecific numbers (sn)**, 1, 2 and 3 from any end. As the C atoms 1 and 3 are identical, these three carbon atoms are better denoted by an older nomenclature, as α , β and α' .

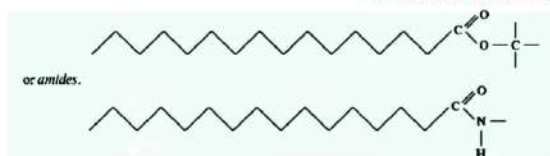
Among the **unsaturated** alcohols found in fats are included a number of pigments like phytol and lycophyll. The former is a constituent of chlorophyll, whereas the latter is found in tomatoes as a purple pigment.

FATTY ACIDS

Fatty acids are long-chain organic acids having usually from 4 to 30 carbon atoms; they have a single carboxyl group and a long, nonpolar hydrocarbon 'tail', which gives most lipids their hydrophobic and oily or greasy nature. Fatty acids do not occur in free or uncombined state in cells or tissues but are present in covalently bound form in different classes of lipids. Fatty acids which occur in natural fats are usually monocarboxylic and contain an even number of C atoms as these are synthesized from 2 carbon units. These are usually straight-chain derivatives. The chain may be **saturated** (containing only single bonds) or **unsaturated** (containing one or more double bonds). Some fatty acids may have hydroxyl group(s) in the chain (**hydroxy** or **oxygenated** fatty acids) and still others may possess ring structure (**cyclic** fatty acids). Fatty acids are stored as an energy reserve (fat) through an ester linkage to glycerol to form **triglycerides**. If free, the **carboxyl group** of a fatty acid will be ionized.

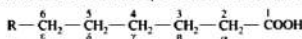


STRUCTURE OF LIPIDS 233



Nomenclature. The systematic nomenclature of the fatty acids is based on the **Genevan system**. According to this system, the fatty acid is named after the hydrocarbon with the same number of carbon atoms, the suffix **-oic** is written in place of the final letter **e** in the name of the hydrocarbon. The names of saturated fatty acids end with the suffix **-anoic** and those of unsaturated acids with the suffix **-enoic**.

The position of carbon atoms in the fatty acid chain is indicated either by **numbering** (in which case the carboxyl carbon is numbered as C1, the carbon adjacent to C1 as C2 and so on) or by the **use of Greek letters** (in which case C2 is denoted as α -carbon, C3 as β -carbon and so on). Thus :



A widely-used convention to indicate the number and position of the double bond(s) in the case of unsaturated fatty acids is to write the number of carbon atoms, the number of double bond(s) and the position of the double bond(s) below the name of the acid. For example, oleic acid having 18 carbon atoms and a double bond between carbon atoms 9 and 10 is written as 18:1; 9. Similarly, linoleic acid (18 carbon atoms and 2 double bonds at C 9 and C 12) is written as 18:2; 9, 12.

An alternative method to write the name of an unsaturated fatty acid is to write first the position of double bond(s) in numerals and then the total number of carbon atoms in Roman followed by the suffix **-enoic** acid. Thus, oleic acid may be written as 9-octadecenoic acid and linoleic acid as 9, 12-octadecadienoic acid.

A description of the various categories of fatty acids involved in lipid formation follows.

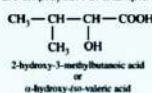
1. Saturated fatty acids

The general formula for these acids is $\text{C}_n\text{H}_{2n+1}\text{COOH}$. Table 12.2. lists some even-numbered straight chain saturated fatty acids, found distributed in both plant and animal worlds. In addition, lipids from all sources contain small amounts of saturated fatty acids with an odd number of carbon atoms (C 5 through C 17). Generally, these odd-carbon acids account for less than 1% of the total fatty acids.

In animal fats, palmitic and stearic acids (C_{16} and C_{18}) are the most abundantly found saturated fatty acids, next in order are shorter chain fatty acids (C_{14} and C_{12}) and longer chain fatty acids (C_{20} , C_{22} and C_{24}). Fatty acids of 10 carbon atoms or less are present in limited amounts in animal lipids except the milk fat which contains appreciable amounts of lower molecular weight fatty acids. The preponderance of these acids may, thus, be shown in the descending order as below :



Generally, the numbers are used with the International Union of Pure and Applied Chemistry (IUPAC) **system of nomenclature** and the Greek letters when the common names are employed. For example :



The **position of a double bond** is indicated by the lower number of the two carbon atoms involved in double bonding.

Table 12-2. Straight chain saturated fatty acids, commonly found in natural fats

Trivial name	Systematic name*	Carbon skeleton	Structure†	Common source
Butyric	n-Butanoic	4 : 0	$\text{CH}_3(\text{CH}_2)_1\text{COOH}$	Butter
Caproic	n-Hexanoic	6 : 0	$\text{CH}_3(\text{CH}_2)_4\text{COOH}$	Cocunut and palm oils
Caprylic	n-Octanoic	8 : 0	$\text{CH}_3(\text{CH}_2)_6\text{COOH}$	Cocunut and palm oils
Capric	n-Decanoic	10 : 0	$\text{CH}_3(\text{CH}_2)_8\text{COOH}$	Cocunut and palm oils
Lauric	n-Dodecanoic	12 : 0	$\text{CH}_3(\text{CH}_2)_{10}\text{COOH}$	Laurel oil, Spermacti
(<i>laurus</i> [‡] = laurel plant)				
Myristic	n-Tetradecanoic	14 : 0	$\text{CH}_3(\text{CH}_2)_{12}\text{COOH}$	Butter and wool fats
(<i>Myristica</i> [‡] = nutmeg genus)				
Palmitic	n-Hexadecanoic	16 : 0	$\text{CH}_3(\text{CH}_2)_{14}\text{COOH}$	Animal and plant fats
(<i>palma</i> [‡] = palm tree)				
Stearic	n-Octadecanoic	18 : 0	$\text{CH}_3(\text{CH}_2)_{16}\text{COOH}$	Animal and plant fats
(<i>stear</i> = hard fat)				
Arachidic	n-Eicosanoic	20 : 0	$\text{CH}_3(\text{CH}_2)_{18}\text{COOH}$	Groundnut oil
(<i>Arachis</i> [‡] = legume genus)				
Behenic	n-Docosanoic	22 : 0	$\text{CH}_3(\text{CH}_2)_{20}\text{COOH}$	Groundnut oil
Lignoceric	n-Tetracosanoic	24 : 0	$\text{CH}_3(\text{CH}_2)_{22}\text{COOH}$	Groundnut and Rapeseed oils
(<i>lignum</i> [‡] = wood; <i>cera</i> [‡] = wax)				
Cerotic	n-Hexacosanoic	26 : 0	$\text{CH}_3(\text{CH}_2)_{24}\text{COOH}$	Wool fat
Montanic	n-Octacosanoic	28 : 0	$\text{CH}_3(\text{CH}_2)_{26}\text{COOH}$	—

* The prefix *n*- indicates the "normal" unbranched structure. For instance, "octadecanoic" simply indicates 18 carbon atoms, which could be arranged in a variety of branched forms. Thus, *n*-octadecanoic specifies the linear, unbranched form.

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

In addition to the straight chain fatty acids, a number of branched chain fatty acids[‡] an even or an odd number of carbon atoms have been identified as minor component and oils. Table 12-3 lists a representative sample of these fatty acids.

259 / 1254

Table 12-3. Some branched chain saturated fatty acids, identified in natural fats

Trivial	Systematic atoms	No. of C atoms	Structure	Common source
Isopalmitic	Isobexadecanoic	16	$\begin{array}{c} \text{CH}_3 \\ \\ \text{CH}(\text{CH}_2)_{11}\text{COOH} \end{array}$	Wool fat
Anteiso-palmitic	14-methyl-hexadecanoic	17	$\begin{array}{c} \text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH}(\text{CH}_2)_{12}\text{COOH} \end{array}$	Wool fat
Tuberculo-stearic	D-(-) 10-methyloctadecanoic	19	$\begin{array}{c} \text{CH}_3 \\ \\ \text{CH}_4(\text{CH}_2)_7\text{CH}(\text{CH}_2)_9\text{COOH} \end{array}$	Bacteria

2. Unsaturated fatty acids

These may be classified, based on the degree of unsaturation.

A. Monoethenoid acids — These contain one double bond and conform to the general formula, $\text{C}_n\text{H}_{2n-1}\text{COOH}$. The common example is oleic acid.

B. Diethenoid acids — Two double bonds;

$\text{C}_n\text{H}_{2n-2}\text{COOH}$; Linoleic acid.

C. Triethenoid acids — Three double bonds;

$\text{C}_n\text{H}_{2n-5}\text{COOH}$; Linolenic acid.

D. Tetraethenoid acids — Four double bonds;

$\text{C}_n\text{H}_{2n-7}\text{COOH}$; Arachidonic acid

Monoethenoid acids are commonly called as *monounsaturated fatty acids (MUFAs)* and the remaining ones as *polyunsaturated fatty acids (PUFAs)*. A few important unsaturated acids are listed in Table 12-4. A perusal of the table indicates that in most of the unsaturated fatty acids, there is a double bond (designated Δ^9) between carbon atoms 9 and 10. This is particularly true of the unsaturated fatty acids, commonly found in the plant world. If there are additional bonds, they usually occur between Δ^9 and the methyl-terminal end of the chain. It may, however, be generalized that in mammals, polyunsaturated fatty acids can have up to 22 carbon atoms and 6 double bonds, but in plants these acids do not exceed 18 carbon atoms and 4 double bonds.

Human body can convert stearic acid to oleic acid by inserting a double bond but is incapable of inserting further double bonds so that the oleic acid cannot be converted to either linoleic, linolenic or arachidonic acid. For normal cell functioning esp., of skin tissues any one of these acids is needed. Since they cannot be synthesized by the cells, they must be obtained from diet. On account of the important physiological role, these 3 acids are collectively called as *essential fatty acids (EFA)*, a term introduced by Burr and Burr in 1930.

Vegetable oils contain two types of polyunsaturated fatty acids (PUFAs) — linoleic acid (*lin* with 2 double bonds) and α -linolenic acid (*len* with 3 double bonds). In chemical jargon, *lin* and longer chain fatty acids derived from it are referred to as *n-6 fatty acids*, since the first double bond in their molecule occurs on carbon no. 3. *Len* and fatty acids derived from it are referred to as *n-3 fatty acids* since the first double bond in their molecule occurs on carbon no. 3. *Lin* is more abundant in nature and is present in all vegetable oils, whereas *len* is present only in some vegetable oils such as mustard oil, rapeseed and soybean. Fish oils are, however, good sources of both *lin* and *len* fatty acids and hence are considered nutritionally quite rich. Hydrogenated fat or 'vanaspai' is manufactured from vegetable oils by the process of hydrogenation. It is a saturated fat and, though derived from vegetable oils, behaves like other saturated fats of animal origin.

All foods contain small quantities of bound fat or invisible fat. Such bound fat is found in green leafy vegetables and pulses in plenty and is a rich source of *n-3* fatty acids. About half of the daily human requirement of fat (40 g) can be derived through such bound fat. Thus, the minimum daily requirement of visible fat (oil etc) is only 20 g. For long, the 2 types of PUFAs, *n-6* and *n-3*, were thought to have similar effects. But recent study shows that *n-6* and *n-3* PUFAs have different types of effects on blood lipids and blood clotting and that they should be present in the diet in a certain proportion. Too much of *n-6* with very little *n-3* is not desirable. A healthy cooking practice, therefore, would be to blend vegetable oils containing both *lin* and *len*, instead of patronizing any one type of oil, including even the much-publicized unsaturated oils like safflower or rice bran which are very effective in reducing blood cholesterol.

A most unusual unsaturated fatty acid, *nemotinic acid*, is excreted in the growth medium by a

Linoleic acid was once known as **vitamin F**. Its function in this capacity is now discredited. It is a yellow oily liquid with a b.p. 229°C.

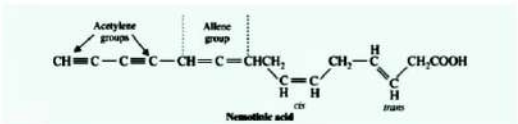
Table 12-4 Unsaturated fatty acids* commonly found in natural fats

Trivial name	Systematic name	Carbon skeleton	Structure	Common source
Crotonic	2-butenic	4 : 1 : 2	$\text{CH}_3\text{CH}=\text{CHCOOH}$	Croton oil
Myristic	9-tetradecenoic	14 : 1 : 9	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	Pyrambutan
Palmitic	9-hexadecenoic	16 : 1 : 9	$\text{CH}_3(\text{CH}_2)_5\text{CH}=\text{CH}(\text{CH}_2)_9\text{COOH}$	Animal and plant fats
Oleic (<i>oleum</i> = oil)	9-octadecenoic	18 : 1 : 9	$\text{CH}_3(\text{CH}_2)_7\text{CH}=\text{CH}(\text{CH}_2)_9\text{COOH}$	Animal and plant fats
Vaccenic	11-octadecenoic	18 : 1 : 11	$\text{CH}_3(\text{CH}_2)_5\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$	Bacterial fat
Linoleic (<i>linum</i> = flax)	9, 12-octadecadienoic	18 : 2 : 9, 12	$\text{CH}_3(\text{CH}_2)_4\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_3\text{COOH}$	Livseed and cottonseed oils
Eleostearic	9, 11, 13-octadecatrienoic	18 : 3 : 9, 11, 13	$\text{CH}_3(\text{CH}_2)_3\text{CH}=\text{CH}-\text{CH}=\text{CH}-\text{CH}=\text{CH}(\text{CH}_2)_4\text{COOH}$	Tung oil
Linolenic	9, 12, 15-octadecatrienoic	18 : 3 : 9, 12, 15	$\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_2\text{COOH}$	Livseed oil
Arachidonic	5, 8, 11, 14-eicosatetraenoic	20 : 4 : 5, 8, 11, 14	$\text{CH}_3(\text{CH}_2)_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_2\text{COOH}$	Animal fat
Nervonic	15-heteroenoic	24 : 1 : 15	$\text{CH}_3(\text{CH}_2)_9\text{CH}=\text{CH}(\text{CH}_2)_{13}\text{COOH}$	-

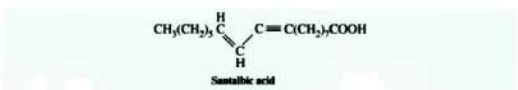
* All of the unsaturated fatty acids in this table are of the *cis* form with the exception of eleostearic acid, which is of the *cis,9-trans,11-trans,13-trans* form.

STRUCTURE OF LIPIDS 237

citrivorum mould. This fatty acid is unique in that it contains the single, double and triple C—C linkages. Nemotinic acid is one of the few naturally-occurring compounds containing the allene group.



The acetylene group has been detected in a number of unsaturated fatty acids, found in higher plants and microorganisms. For instance, *santalbic acid*, a major component of the seed oil of sandalwood contains 1 acetylene group as against 2 of nemotinic acid.



Geometric isomerism. On account of the presence of double bond(s), the unsaturated fatty acids exhibit geometric (or *cis-trans*) isomerism. Most unsaturated fatty acids are found as the **unstable *cis* isomer** rather than as the more stable *trans* isomer.



The hydrocarbon chain of saturated fatty acids (stearic acid, for example) has a zigzag configuration with the C—C bond forming a bond angle of 109° (Fig. 12-2).

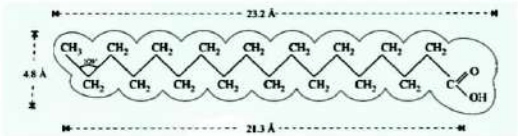


Fig. 12-2. Stearic acid

For the sake of simplicity, the long chains of CH_2 groups are represented by zigzag lines where each corner represents a C atom and the hydrogen atoms are left out. The zigzag line represents the most stable configuration of such carbon chains. The free end of this line, if not indicated otherwise, obviously represents the presence of a methyl group. The simplified formula for stearic acid would, thus, be as represented in Fig. 12-3.



Fig. 12-3. Stearic acid (simplified formula)

The double bond in the zigzag line is indicated by drawing an extra line in between the carbon atoms involved in double bond formation.

When a *cis* double bond is inserted (as in oleic acid), the molecule bends assuming the shape, shown in Fig. 12-4. This double bond is *rigid* and creates a *kink* in the chain. The rest of the chain is free to rotate about the other C—C bonds.

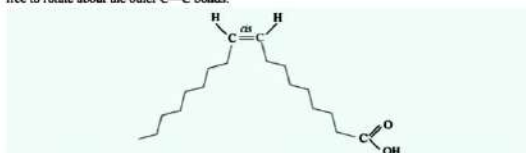


Fig. 12-4. Oleic acid

Nonconjugated double-bond system. Another structural peculiarity of naturally occurring polyunsaturated fatty acids is the presence of a **nonconjugated double-bond system**. It has a methylene group (—CH₂—) flanked by double bonds on both the sides as in linoleic, linolenic and arachidonic acids. The **conjugated double-bond system** is, however, rarely present. In it the methylene group is not found in between the double bonds which, henceforth, occur one after the other as in elostearic acid. This acid has valuable properties as a drying oil since it polymerizes readily. These two double-bond systems (Fig. 12-5) have different chemical reactivities.

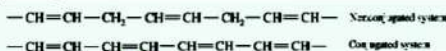
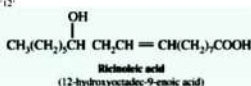


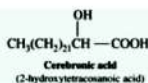
Fig. 12-5. Two double-bond systems

3. Hydroxy or oxygenated fatty acids

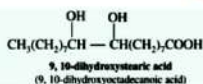
One of these found in castor oil is **ricinoleic acid** (87%). It is a C 18 acid with a double bond at C₉ and an OH group on C₁₂.



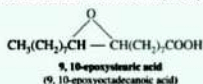
Cerebronic acid, a C 24 acid obtained from animal lipid, is another important hydroxy acid with an OH group on C₂.



A common oxygenated fatty acid, isolated from plants and bacterial lipids, is 9, 10-dihydroxystearic acid.



Similarly, 9, 10-epoxystearic acid is isolated from rust spore lipids (20%).



4. Cyclic fatty acids

These are of rare occurrence. Chaulmoogra oil, obtained from the plant *Hydnocarpus kurzii* and used in the treatment of leprosy, contains 2 such acids *hydnoicarpic* and *chaulmoogric*. Chaulmoogric acid has a cyclopentenyl ring in its 18-carbon structure.

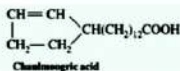
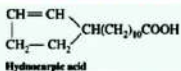
CLINICAL IMPLICATIONS

Leprosy, the oldest disease known to mankind, is loathsome, endemic disease. The myth that leprosy was hereditary and a 'Curse of God' was exploded in 1873, when Dr. C. A. Hansen in Norway discovered rod-shaped bacteria in cells from leprosy nodules. The leprosy bacillus (*Mycobacterium leprae*), also called Hansen's bacillus, is the slowest multiplying prokaryote known and takes about 2 weeks to double. Once infected, it takes a long time for the disease to develop: from less than a year to as long as 30 years, though normally 3-7 years. About 20% of the infected people develop the disease, on an average. In patients with high immune resistance, the disease is relatively mild (*tuberculoid form*); those with low resistance develop highly infectious and more severe conditions (*lepromatous leprosy*), affecting not only nerves and skin but also lymph nodes, eyes, nose, mouth, larynx, spleen etc. and the bacilli may appear in tears, nasal mucus, sputum, ulcerating nodules etc. *M. leprae* is the only bacterium that can enter and destroy the nerves. Because of sensory impairment due to nerve damage, the person does not feel pain. However, the primary deformities are pronounced on the face. The infected, untreated individual is the only source of infection. There is no known natural reservoir other than man. Leprosy is only weakly contagious: usually prolonged and close contact is considered necessary. The nose is the main portal of exit of the bacilli. Although leprosy affects individuals of both sexes, males are more susceptible than females, often in the ratio of 2 : 1. Although leprosy had assumed epidemic proportions in Europe in 13th and 14th century A.D., at present, Western Europe is virtually free from it. And these days, the leprosy 'hot spots' are India, Indonesia, Myanmar, Brazil, Columbia and few pockets of Central and Eastern Europe. Developed countries are, however, free from leprosy due to their improved socio-economic conditions. Of the estimated disease burden, Asia accounts for about 60%, followed by Africa with 35% and Latin America with 4%; the rest of the world accounting for about 1% of the disease load. India alone accounts for about 1/5th of the total lepers in the world.

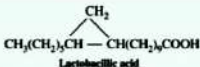
Under chemotherapy, the drugs most commonly used against leprosy were dapsone, rifampicin and clofazimine, the last one being the best alternative as it had the much-desired anti-inflammatory effect and produced least resistance. However, in 1982, the WHO recommended the multi-drug therapy (MDT), combining all the 3 above drugs, as it has a very high potential for control of leprosy. The duration of treatment in most cases is about 3 years or more. A single dose of MDT kills about 95% of bacilli in a patient and there are no significant side effects. However, the overall long duration of the treatment regimen in chemotherapy, has led to the use of immunotherapy, using leprosy vaccines. Under the aegis of WHO, a vaccine was developed combining *M. leprae* (inactivated) derived from armadillo, and the BCG vaccine. Trials of this vaccine have shown that its application can restore cell-mediated immunity in patients and thus it could remarkably suppress the disease in about 18 months only.

Leprosy eradication programme has made spectacular progress worldwide. From a long list of 122 endemic countries in 1982, the disease is now endemic only in 24 of them. Ninety per cent of leprosy is now confined to only 12 countries, including India, Brazil, Indonesia, Myanmar, Nepal, Mozambique, Guinea, and Angola.

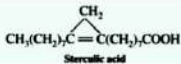
240 FUNDAMENTALS OF BIOCHEMISTRY



Lipids from the lactobacilli contain a fatty acid, *lactobacillic acid*, with a cyclopropyl group. This fatty acid may result from the addition of a methylene group across the double bond of vaccenic acid (18:1; 11).



Similarly, *stereolic acid* from plant sources has a comparable structure, with a suggested relationship to oleic acid. It may be derived from oleic acid by the addition of a methylene group across the double bond in a manner that the unsaturated nature is not altered, unlike the lactobacillic acid.



The percentage composition of various types of fatty acids in some typical fats and oils is presented in Table 12-5.

BIOLOGICAL ROLES OF LIPIDS

The lipids perform a wide variety of functions (Fig. 12-6). These are briefly listed below:

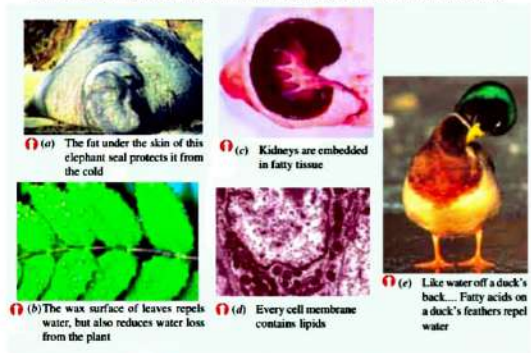


Fig. 12-6. Various functions performed by lipids

STRUCTURE OF LIPIDS 241

Table 12-5. Fatty acid content of some animal and plant fats and oils

Fat/Oil	Butyric	Caproic	Caprylic	Capric	Lauroic	Myristic	Palmitic	Stearic	Arachidic	Palmitoleic	Oleic	Linoleic	Linolenic	Others
Butter fat	3.6	2.3	0.5	1.0	2.5	11.1	29.0	9.2	2.4	4.6	26.7	3.6	-	-
Beef tallow	-	-	-	-	-	6.3	27.4	14.1	-	-	49.6	2.5	-	-
Human depot fat	-	-	-	-	-	2.7	24.0	8.4	-	-	46.9	10.2	-	-
Pork fat (lard)	-	-	-	-	-	1.3	28.3	11.9	-	2.7	47.5	6.0	-	-
Casor oil	-	-	-	-	-	Trace	Trace	-	-	-	7.4	3.1	-	88 [*]
Corn oil	-	-	-	-	-	1.4	10.2	3.0	-	1.5	49.6	34.3	-	-
Soybean oil	-	-	-	-	-	0.1	9.8	2.4	-	0.4	28.9	50.7	6.5	-
Linseed oil	-	-	-	-	-	-	6.8	2.5	0.5	-	19.0	24.1	47.4	-
Herring oil	-	-	-	-	-	7.3	13.0	Trace	-	4.9	-	-	20.7	30.1 [†] 23.2 [‡]
Whale oil	-	-	-	-	-	0.2	15.6	2.8	-	14.4	35.2	-	-	13.6 [§] 5.9 [¶]

* Not detected

† Ricinoleic acid

‡ C:20 polyunsaturated

§ C:22 polyunsaturated

242 FUNDAMENTALS OF BIOCHEMISTRY

1. **Food material.** Lipids provide food, highly rich in calorific value. One gram lipid produces 9.3 kilocalories of heat.

2. **Food reserve.** Lipids provide are insoluble in aqueous solutions and hence can be stored readily in the body as a food reserve.

3. **Structural component.** Lipids are an important constituent of the cell membrane.

4. **Heat insulation.** The fats are characterized for their high insulating capacity. Great quantities of fat are deposited in the subcutaneous layers in aquatic mammals such as whale and in animals living in cold climates.

5. **Fatty acid absorption.** Phospholipids play an important role in the absorption and transportation of fatty acids.

6. **Hormone synthesis.** The sex hormones, adrenocorticoids, cholic acids and also vitamin D are all synthesized from cholesterol, a steroidal lipid.

7. **Vitamin carriers.** Lipids act as carriers of natural fat-soluble vitamins such as vitamin A, D and E.

8. **Blood cholesterol lowering.** Chocolates and beef, especially the latter one, were believed to cause many heart diseases as they are rich in saturated fatty acids, which boost cholesterol levels in blood and clog the arterial passage (Fig. 12-7). But researches conducted at the University of Texas by Scott Grundy and Andrea Bonanome (1988) suggest that at least one saturated fatty acid stearic acid, a major component of cocoa butter and beef fat, does not raise blood cholesterol level at all. The researchers placed 11 men on three cholesterol-poor liquid diets for three weeks each in random order. One formula was rich in palmitic acid, a known cholesterol booster, the second in oleic acid; and the third in stearic acid. When compared with the diet rich in palmitic acid, blood cholesterol levels were 14% lower in subjects put on the stearic acid diet and 10% lower in those on the oleic acid diet.

9. **Antibiotic agent.** *Squalamine*, a steroid from the blood of sharks, has been shown to be an antibiotic and antifungal agent of intense activity. This seems to explain why sharks rarely contract infections and almost never get cancer.



Fig. 12-7. A cholesterol deposit, (known as atheroma) filling almost all the space inside the artery.

REFERENCES

See list following Chapter 14.

PROBLEMS

- Why are the most unsaturated fatty acids found in phospholipids in the *cis* rather than the *trans* conformation? Draw the structure of a 16-carbon fatty acid as saturated, *trans* monounsaturated, and *cis* monounsaturated.
- Given these molecular components—glycerol, fatty acid, phosphate, long-chain alcohol, and carbohydrate—answer the following:
 - Which two are present in both waxes and sphingomyelin?
 - Which two are present in both fats and phosphatidylcholine?
 - Which are present in a ganglioside but not in a fat?

STRUCTURE OF LIPIDS 243

- The melting points of a series of 18-carbon fatty acids are stearic acid, 69.6°C; oleic acid, 13.4°C; linoleic acid -5°C; and linolenic acid, -11°C. What structural aspect of these 18-carbon fatty acids can be correlated with the melting point? Provide a molecular explanation for the trend in melting points.
- How is the definition of "lipid" different from the definitions of other types of biomolecules that we have considered, such as amino acids, nucleic acids, and proteins?
- Johann Thudichum, who practiced medicine in London about 100 years ago, also dabbled in lipid chemistry in his spare time. He isolated a variety of lipids from neutral tissue, and characterized and named many of them. His carefully sealed and labeled vials of isolated lipids were rediscovered many years later. How would you confirm, using techniques available to you but not to him, that the vials he labeled "sphingomyelin" and "cerebroside" actually contain these compounds?
- How would you distinguish sphingomyelin from phosphatidylcholine by chemical, physical, or enzymatic tests?
- Do all kinds of fibre help lower one's cholesterol?
- How much fat does one need in his diet?
- Is a high-protein diet good for one's health?
- Why is tea oil better for frying than olive oil?

- Simple Lipids
 - Fats and Oils
 - Waxes
- Compound Lipids
 - Phospholipids
 - Glycolipids
- Derived Lipids
 - Steroids
 - Terpenes
 - Carotenoids

Lipids-II

Classification



(a)



(b)

Lipids

(a) **Fat.** A European brown bear ready to hibernate. Fat is an efficient way to store energy. If this bear stored the same amount of energy in carbohydrates instead of fat, she probably would be unable to walk! (b) **Wax.** Wax is a highly saturated lipid that remains very firm at normal outdoor temperatures. Its rigidity allows it to be used to form the strong but thin-walled hexagons of this honeycomb.

Blor (1943) has proposed the following classification of lipids based on their chemical composition.

A. Simple lipids or Homolipids. These are esters of fatty acid with various alcohols.

1. **Fats and oils (triglycerides, triacylglycerols).** These are esters of fatty acids with a trihydroxy alcohol, glycerol. A fat is solid at ordinary room temperature whereas an oil is liquid.

2. **Waxes.** These are esters of fatty acids with high molecular weight monohydroxy alcohols.

B. Compound lipids or Heterolipids. These are esters of fatty acids with alcohol and possess additional group(s) also.

1. **Phospholipids (phosphatids).** These are compounds containing, in addition to fatty acids and glycerol, a phosphoric acid, nitrogen bases and other substituents.

2. **Glycolipids (cerebrosides).** These are the compounds of fatty acids with carbohydrates and contain nitrogen but no phosphoric acid. The glycolipids also include certain structurally-related compounds comprising the groups, gangliosides, sulfolipids and sulfatides.

C. Derived lipids. These are the substances derived from simple and compound lipids by hydrolysis. These include fatty acids, alcohols, mono- and diglycerides, steroids, terpenes and carotenoids.

CLASSIFICATION OF LIPIDS 245

Glycerides and cholesterol esters, because of their uncharged nature, are also called *neutral lipids*. However, Conn and Stumpf (1976) have traditionally classified lipids into following 6 classes:

1. Acyl glycerols
2. Waxes
3. Phospholipids
4. Sphingolipids
5. Glycolipids
6. Terpenoid lipids including carotenoids and steroids

SIMPLE LIPIDS

FATS AND OILS

(= Triglycerides or Triacylglycerols)

The *triglycerides* are the most abundant of all lipids. They constitute about 98% of total dietary lipids; the remaining 2% consists of phospholipids and cholesterol and its esters. They are the major components of storage or depot fats in plant and animal cells but are not normally found in membranes. They are nonpolar, hydrophobic molecules since they contain no electrically charged or highly polar functional groups. In animals, the *fat cells* or *adipocytes* contain very large quantities of triglycerides in the form of fat droplets, which fill almost the entire cell volume (Fig. 13-1). Adipocytes are abundantly found under the skin, in the abdominal cavity and in the mammary glands. Triglycerides can be stored in quantities, sufficient to supply the energy needs of the body for many months, as in the case of obese persons. On



Fig 13-2. Insulating blubber of whales and the fat molecules

Fat such as the triacylglycerol molecule (lower) are widely used to store excess energy for later use and to fulfill other purposes, illustrated by the insulating blubber of whales. The natural tendency of fats to exist in nearly water-free forms makes these molecules well-suited for these roles.

[Courtesy: (Upper) Francois Cohier]

as well as glycerol. However, the milk of spiny anteater is an exception in that it comprises almost pure triolein. Human body contains enough fat to make 7 bars of soap!

In a normal man, weighing 70 kg, at least 10-20% of the body weight is lipid, the bulk of which is triacylglycerol (TAG). TAG is found in all organs of the human body, particularly in adipose

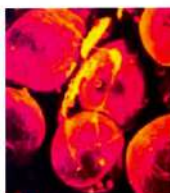


Fig 13-1. A scanning electron micrograph of adipocytes

Each adipocyte contains a fat globule that occupies nearly the entire cell.

(Courtesy: Fred E. Hoxler)

the contrary, the body can store the carbohydrate glycogen in meagre amounts, sufficient to supply energy need of a day only. *Triglycerides are much better adapted than glycogen to serve as storage form of energy.* They are not only stored in large amounts but also yield over twice as much energy as carbohydrates. Since fats tend to remain in the stomach longer than carbohydrates and are digested more slowly, they also have greater satiety value than carbohydrates. The arctic and antarctic animals (Fig. 13-2) such as whales, seals, walrus and penguins are amply padded with triglycerides to serve both as energy storage depots and as an insulation against very low temperatures. Most fats and oils, upon hydrolysis, yield several fatty acids

246 FUNDAMENTALS OF BIOCHEMISTRY

tissue, in which droplets of triacylglycerols may represent more than 90% of the cytoplasm of the cells. Body lipid is a reservoir of potential chemical energy. About 100 times more energy is stored as mobilizable lipid than as mobilizable carbohydrate in the normal human being. TAG is stored in a relatively water-free state in the tissue, in comparison to carbohydrate, which is heavily hydrated.

Chemically, triglycerides are esters of glycerol with 3 fatty acid molecules. Their generic formula is shown in Fig. 13-3.



Fig. 13-3. Generic formula of triglycerides

Obviously, when the groups attached to carbon 1 and 3 differ, a centre of asymmetry is created at C 2. The 2 optical isomers may, thus, be represented as shown in Fig. 13-4. *The naturally-occurring fats are of L-type.*

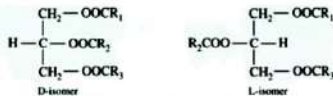


Fig. 13-4. Two optical isomers of triglycerides

A fat molecule contains 3 moles of fatty acids which may be similar or dissimilar. Those containing a single kind of fatty acid in all 3 positions (α , β , α') are called *simple* (or *symmetrical*) *triglycerides*; they are named after the fatty acids they contain (Fig. 13-5). Examples are tripalmitin, tristearin



Fig. 13-5. Two simple triglycerides

and triolein. They occur very infrequently in natural fats. Such simple triglycerides have been synthesized in the laboratory *e.g.*, tristearin and triolein.

Most of the triglycerides of nature are mixed (or asymmetrical) triglycerides, *i.e.*, those containing 2 or 3 different fatty acid units in the molecule. Representatives of such mixed triglycerides are oleodipalmitin and oleopalmitostearin (Fig. 13-6).

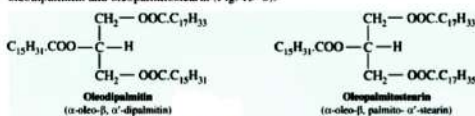


Fig. 13-6. Two mixed triglycerides

Most **animal fats** such as those from meat, milk and eggs are relatively rich in saturated fatty acids but contain a rather low content of polyunsaturated fatty acids (Table 13-1); two exceptions

271 / 1254

CLASSIFICATION OF LIPIDS 247

are chicken fat and fish fat. The large proportion of saturated fatty acids (*esp.*, palmitic and stearic) with high melting point confers solid state to the animal fats. The **plant fats**, on the other hand, contain a large proportion of unsaturated fatty acids (*esp.*, polyunsaturated). The unsaturated fatty acids have low melting point and confer liquid state to the plant fats.

Table 13-1. Fatty acid composition of important animal and plant fats

Fat/Oil	Percentage of total fatty acids		
	Saturated	Monounsaturated	Polyunsaturated
Animal fats			
Butter fat	60	36	4
Pork fat	59	39	2
Beef fat	53	44	2
Chicken fat	39	44	21
Plant fats			
Olive oil*	20	26	54
Corn oil	15	31	53
Soybean oil	14	24	53
Soft margarine	23	22	52

* data not confirmed.

The differences between the animal and plant fats are presented in Table 13-2.

Waxes

Far less spread but equally important are the waxes. The term 'wax' originates in the Old English word *weax*, meaning 'the material of the honeycomb', reminding of beeswax, the honeycomb is made of. These are esters of long-chain saturated and unsaturated fatty acids with long-chain monohydroxy alcohols. The fatty acids range in between C_{12} and C_{26} and the alcohols range from C_{18} to C_{26} . In vertebrates, waxes are secreted by cutaneous glands as a protective coating to keep the skin pliable, lubricated and water-proof. Hair, wool and fur are also coated with wax. Birds, particularly waterfowl, secrete waxes in their preen glands to make their feathers water-repellent. The leaves of many plants such as *Rhododendron*, *Calotropis* etc., are shiny because of the deposition of protective waxy coating. Waxes also serve as the chief storage form of fuel in planktons. Since marine organisms (whale, herring, salmon) consume planktons in large quantities, waxes act as major food and storage lipids in them.

Table 13-2. Differences between animal and plant fats

Animal fat	Plant fat
1. Relatively rich in saturated fatty acids, <i>esp.</i> C_{16} and C_{18} acids.	1. Relatively rich in unsaturated fatty acids, <i>esp.</i> polyunsaturated acids.
2. Solid at ordinary room temperature.	2. Liquid at ordinary room temperature.
3. These have usually low iodine number.	3. These have usually high iodine number.
4. These have usually high Reichert-Meissl number.	4. These have usually low Reichert-Meissl number.
5. These are stored mainly in liver and bone marrow.	5. These are stored mainly in seeds and fruits.
6. Oxidative rancidity is observed more frequently.	6. Oxidative rancidity is observed less frequently.
Examples —Butterfat, Beef fat, Tallow	Examples —Olive oil, Castor oil, Soybean oil, Corn oil



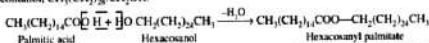
Jain - Fundamentals...



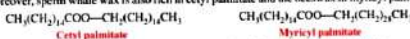
Contents

248 FUNDAMENTALS OF BIOCHEMISTRY

Most of the waxes are mixtures of esters. Thus, *sperm whale wax* (spermaceti) and *beeswax* are composed mainly of palmitic acid esterified with either hexacosanol, $\text{CH}_3(\text{CH}_2)_{24}\text{CH}_2\text{OH}$ or triacontanol, $\text{CH}_3(\text{CH}_2)_{28}\text{CH}_2\text{OH}$.

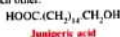


Moreover, sperm whale wax is also rich in cetyl palmitate and the beeswax in myricyl palmitate.

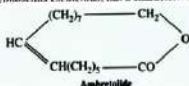


Carnauba wax, the hardest known wax, consists mainly of fatty acids esterified with tetracosanol, $\text{CH}_3(\text{CH}_2)_{22}\text{CH}_2\text{OH}$ and tetratriacontanol, $\text{CH}_3(\text{CH}_2)_{32}\text{CH}_2\text{OH}$.

The waxes from the conifers contain polymers formed by the ester-linking of many ω -hydroxy acids, such as juniperic acid, with each other.



There are certain waxes which have a characteristic odour. This is due to the presence of hydroxy acids in the form of lactones in them. For example, a wax named *ambretolide*, which is extracted from the seeds of lady's finger, *Abelmoschus esculentus*, has a characteristic musky smell.



Lipids of marine organisms such as starfish, squid and shark contain fatty alcohols that are long-chain alkyl ethers of glycerol. Three such *glyceryl ethers* (Fig. 13-7), as they are called have been isolated from shark oil.

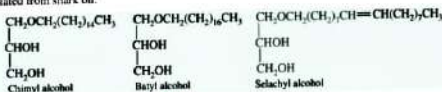


Fig. 13-7. Some glyceryl ethers

Waxes are unusually inert due to their saturated nature of the hydrocarbon chain. However, they can be split slowly with hot alcoholic KOH. They are insoluble in water and highly resistant to atmospheric oxidation. Hence, these are used in polishing furnitures and automobiles and also in wax-coated paper used to wrap perishable food products such as biscuits, cakes, etc. Lanolin (from lamb's wool), beeswax (from honeycomb), carnauba wax (from a Brazilian palm tree) and spermaceti oil (from sperm whales) are widely used in the manufacture of lotions, ointments and polishes.

COMPOUND LIPIDS

PHOSPHOLIPIDS

(= Phosphatids)

Phospholipids are the most abundant membrane lipids. They serve primarily as structural components of membranes and are never stored in large quantities. As their name implies, phospholipids contain phosphorus in the form of phosphoric acid groups. They differ from triglycerides in possessing usually one hydrophilic polar "head" group and usually two hydrophobic nonpolar "tails". For this reason, they are often called *polar lipids*. Thus, *phospholipids are amphipathic, whereas the storage*

Contents

CLASSIFICATION OF LIPIDS 249

lipids (triglycerides and waxes) are not. In phospholipids, two of the OH groups in glycerol are linked to fatty acids while the third OH group is linked to phosphoric acid. The phosphate is further linked to one of a variety of small polar head groups (alcohols). Folch and Sperry (1955) have classified phospholipids into phosphoglycerides, phosphoinositides and phosphosphingolipids.

A. Phosphoglycerides

These are the major phospholipids found in membranes and contain two fatty acid molecules or "tails" esterified to the first and second hydroxyl groups of glycerol. The third hydroxyl group of glycerol forms an ester linkage with phosphoric acid. In addition, phosphoglycerides contain a second alcohol, which is also esterified to the phosphoric acid. This is referred to as "head alcohol group" as it is present at one end ("head") of the long phosphoglyceride molecule.

The various phosphoglycerides differ in their head alcohol groups. However, all of them contain two nonpolar tails, each consisting of a long chain (usually C_{16} or C_{18}) fatty acid. Usually one of the fatty acids is saturated and the other unsaturated; the latter is always esterified to the middle or β -hydroxy group of glycerol. A noteworthy feature of the phosphoglycerides is that they contain an asymmetric carbon atom at position 2 in the glycerol part of their molecule. It has the L-configuration since it is related to L-glyceraldehyde. All phosphoglycerides have a negative charge on phosphoric group at pH 7. In addition, the head alcohol group may also have one or more electric charges at pH 7.

1. *Lecithins* (= *phosphatidyl choline*) - Lecithins (*likithos* = yolk) are widely distributed in nature. Various oils like soybean and the yeasts are important sources from plant world. In animals, the glandular and nervous tissues are rich in these lipids. The lecithins are required for the normal transport and utilization of as other lipids *esp.*, in the liver of animals. In their absence, accumulation of lipids occurs in the liver to as much as 30% against a normal value of 3-4%, giving rise to a condition called "*fatty liver*". This fatty infiltration may lead to fibrotic changes, characteristic of the liver disease *cirrhosis*.

In addition to glycerol and 2 moles of fatty acids, the lecithins (Fig. 13-8) also contain phosphoric acid and a nitrogen base choline at either the end or middle carbon atom of glycerol unit. Accordingly, two forms of lecithins, α and β are recognized.

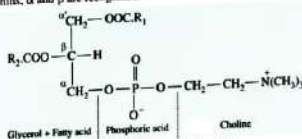


Fig. 13-8. Structure of α -lecithin

On complete hydrolysis, lecithin yields choline, phosphoric acid, glycerol and 2 moles of fatty acids. But partial hydrolysis of lecithins by *lecithinases* (active principles found in snake venoms) causes removal of only one fatty acid to yield substances called *lysolecithins* (Fig. 13-9). These, therefore, contain only one acyl radical. When subjected into the blood stream by sting as a result of snake bite or by needle, the lysolecithins cause rapid rupture (hemolysis) of the red blood corpuscles.

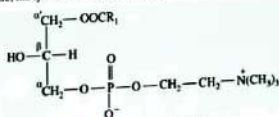


Fig. 13-9. Structure of α -lysolecithin



lipids (triglycerides and waxes) are not. In phospholipids, two of the OH groups in glycerol are linked to fatty acids while the third OH group is linked to phosphoric acid. The phosphate is further linked to one of a variety of small polar head groups (alcohols). Folch and Sperry (1955) have classified phospholipids into phosphoglycerides, phosphoinositides and phosphosphingosides.

A. Phosphoglycerides

These are the major phospholipids found in membranes and contain two fatty acid molecules or "tails" esterified to the first and second hydroxyl groups of glycerol. The third hydroxyl group of glycerol forms an ester linkage with phosphoric acid. In addition, phosphoglycerides contain a second alcohol, which is also esterified to the phosphoric acid. This is referred to as "head alcohol group" as it is present at one end ("head") of the long phosphoglyceride molecule.

The various phosphoglycerides differ in their head alcohol groups. However, all of them contain two nonpolar tails, each consisting of a long chain (usually C_{16} or C_{18}) fatty acid. Usually one of the fatty acids is saturated and the other unsaturated; the latter is always esterified to the middle or β -hydroxy group of glycerol. A noteworthy feature of the phosphoglycerides is that they contain an asymmetric carbon atom at position 2 in the glycerol part of their molecule. It has the L-configuration since it is related to L-glyceraldehyde. All phosphoglycerides have a negative charge on phosphoric group at pH 7. In addition, the head alcohol group may also have one or more electric charges at pH 7.

1. *Lecithins* (= *phosphatidyl choline*) – Lecithins (*likithos* = yolk) are widely distributed in nature. Various oil seeds like soybean and the yeasts are important sources from plant world. In animals, the glandular and nervous tissues are rich in these lipids. The lecithins are required for the normal transport and utilization of as other lipids *esp.*, in the liver of animals. In their absence, accumulation of lipids occurs in the liver to as much as 30% against a normal value of 3–4%, giving rise to a condition called "fatty liver". This fatty infiltration may lead to fibrotic changes, characteristic of the liver disease *cirrhosis*.

In addition to glycerol and 2 moles of fatty acids, the lecithins (Fig. 13–8) also contain phosphoric acid and a nitrogen base choline at either the end or middle carbon atom of glycerol unit. Accordingly, two forms of lecithins, α and β are recognized.

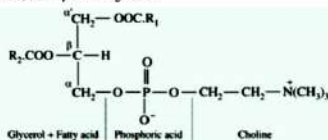


Fig. 13–8. Structure of α -lecithin

On complete hydrolysis, lecithin yields choline, phosphoric acid, glycerol and 2 moles of fatty acids. But partial hydrolysis of lecithins by *lecithinases* (active principles found in snake venoms) causes removal of only one fatty acid to yield substances called *lysolecithins* (Fig. 13–9). These, therefore, contain only one acyl radical. When subjected into the blood stream by sting as a result of snake bite or by needle, the lysolecithins cause rapid rupture (hemolysis) of the red blood corpuscles.

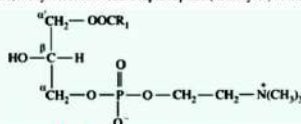


Fig. 13–9. Structure of α -lysolecithin

274 / 1254

250 FUNDAMENTALS OF BIOCHEMISTRY

2. *Cephalins* – The cephalins (*kephalos* = head) are closely associated with lecithins in animal tissues. These have also been identified from soybean oil. These are similar in structure to the lecithins except that the choline is replaced by either ethanolamine or serine. Serine is the biochemical precursor of ethanolamine.



Accordingly, two types of cephalins are recognized, phosphatidyl ethanolamine and phosphatidyl serine. Like lecithins, the cephalins (Fig. 13–10) also exist in 2 forms, α and β , depending upon the relative positions of the two substituent fatty acids.

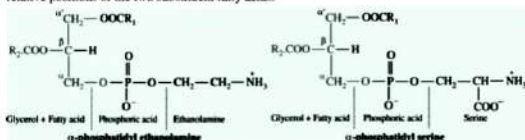


Fig. 13–10. The two cephalins

Since the primary amino group of ethanolamine is a weaker base than the quaternary ammonium group of choline, the cephalins are more acidic than lecithins. Moreover, the cephalins are comparatively less soluble in alcohol than lecithins.

Venoms containing lecithinases also split off fatty acids from cephalins, leaving hemolytic *lysoccephalins*.

3. *Plasmalogens* (= *Phosphoglyceracetal*) – Plasmalogens constitute about 10% of the phospholipids of the brain and muscle. These are apparently not found in significant quantities in plant tissues. Structurally, these resemble lecithins and cephalins but have one of the fatty acids replaced by an unsaturated ether. Since the nitrogen base can be choline, ethanolamine or serine, three types of plasmalogens (Fig. 13–11) are accordingly distinguished: phosphatidyl choline, phosphatidyl ethanolamine and phosphatidyl serine.

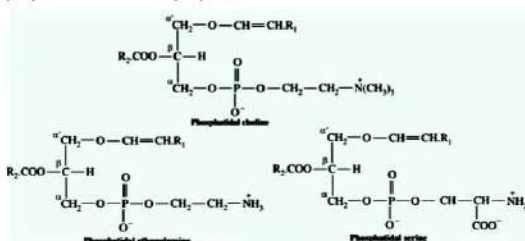


Fig. 13–11. The three plasmalogens

B. Phosphoinositides (=Phosphatidyl inositols)

Phosphoinositides (Fig. 13-12) have been found to occur in phospholipids of brain tissue and of soybeans and are of considerable importance because of their role in transport processes in cells. These are phospholipids where a cyclic hexahydroxy alcohol called inositol replaces base. The inositol is present as the stereoisomer, *myo*-inositol. On hydrolysis, the phosphoinositides yield 1 mole of glycerol, two moles of fatty acid, 1 mole of inositol and 1, 2, or 3 moles of phosphoric acid. Accordingly, mono-, di- or triphosphoinositides are found.

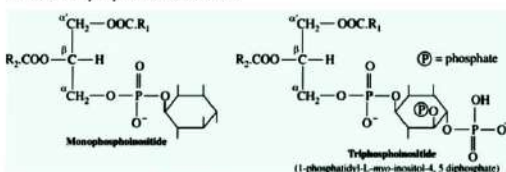


Fig. 13-12. The two phosphoinositides

Phosphoinositides are also classified as glycolipids, in as much as they contain carbohydrate residue.

C. Phosphosphingolipids (=Sphingomyelins)

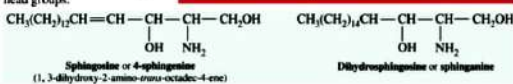
These compounds are commonly found in nerve tissue *esp.*, in the myelin sheath of the nerve (hence their name, sphingomyelins) and apparently lack in plants and the microorganisms. In a *syndrome* called *Niemann-Pick* disease, the sphingomyelins are stored in the brain in large quantities. These differ from other phospholipids in their lack of glycerol and the presence of another nitrogenous base sphingosine or a closely related dihydrosphingosine, besides choline, in place of glycerol. Sphingomyelins are electrically charged molecules and contain phosphocholine as their polar head groups.

A *syndrome* (*syn*^C = together; *dramen*^C = to run) is a group of symptoms associated with a particular disease or abnormality, although not every symptom is necessarily shown by every patient with the disease.

CLINICAL IMPLICATIONS

Niemann-Pick disease (= sphingomyelin lipidosis) is a rare genetic disorder, inherited as an autosomal recessive condition. The disease is caused by a deficiency of the enzyme sphingomyelinase, which cleaves ceramide-phosphocholine bond of sphingomyelin. As a result, sphingomyelin accumulates in large amounts in the reticuloendothelial system since their synthesis is *normal* in rate but their degradation is interrupted. Niemann-Pick cells (prototype of 'foam cells') are found in the bone marrow, spleen, lymphoid tissues, liver, lung and tissues, of virtually any organ. Clinically, the disease manifests itself in two forms:

- Early or infantile form.** This is characterized by hepatosplenomegaly, macular degeneration, cherry-red retinal spot (= macula) in nearly half of the cases, mental retardation and blindness. Death ensues by two years of age.
- Late or adult form.** This perhaps is the most common of the two forms and reveals intellectual impairment during late infancy, slow evolution of the disease, no changes in the ocular fundi and survival beyond the age of five. However, death of the patient occurs by the second decade.



252 FUNDAMENTALS OF BIOCHEMISTRY

On hydrolysis, the phosphosphingolipids yield equimolar amounts of fatty acid, phosphoric acid, choline and sphingosine or dihydrosphingosine but no glycerol. Thus, in these compounds the atomic ratio N/P is 2, in contrast to phosphoglycerides where this ratio equals unity. Phosphoinositides, however, do not contain a nitrogen base.

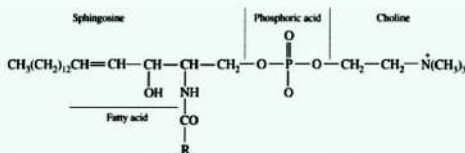


Fig. 13-13. Structure of sphingomyelin

It may be observed from the formula of sphingomyelin (Fig. 13-13) that sphingosine carries the phosphoric acid on its primary alcohol group and the fatty acid by amide linkage on its primary amino group.

GLYCOLIPIDS

(= Cerebrosides or Glycosphingolipids)

The cerebrosides, as the name suggests, are important constituent of brain where they amount to about 8% of the solid matter. These may also occur in tissues other than brain. Since the head group characteristically consists of one or more sugar units, cerebrosides are often called glycosphingolipids. Like phospholipids, glycolipids are composed of a hydrophobic region, containing 2 long hydrocarbon tails, and a polar region, which now contains one or more sugar residues and no phosphate. Both phospholipids and glycolipids form self-sealing lipid bilayers that are the basis for all cellular membranes. In Gaucher disease, the cerebrosides appear in relatively large amount in the liver and the spleen. They are also present in large amount in the brain in Niemann-Pick disease. These are present in much higher concentration in medullated than in nonmedullated nerve fibres. There is evidence that they also occur in some plant organs.

The structure of cerebrosides (Fig. 13-14) is somewhat similar to that of phosphosphingolipids. They contain a high molecular weight fatty acid, sphingosine and either galactose or glucose instead of choline but no phosphoric acid. They have no electric charge since their polar head groups are neutral. In general properties, they resemble sphingomyelins.

CLINICAL IMPLICATIONS

Gaucher disease (= glucocerebrosidosis) is a not-so-rare, autosomal recessive, hereditary syndrome, often familial and of Jewish origin. It is 30 times more prevalent among the Ashkenazic Jews, with 1 case in 2,500 births. The disease is due to a deficiency of the enzyme *glucosyl ceramide hydrolase* (= glucocerebrosidase), which cleaves ceramide from ceramide trihexoside, causing deposition of the later. Gaucher's cells (pale lipid-containing cells) are found in abundance in all the organs specifically in the lungs and bone marrow, resulting in marked hepatosplenomegaly. Thus, the disease affects the lungs and causes destruction of bones. Clinically, two patterns of the disease have been reported:

- Infantile or acute form.** In this, the symptoms appear early with hepatosplenomegaly, mental retardation, hypertonicity and signs of respiratory involvement. Most patients die during infancy.
- Adult or chronic form.** This is more common a form of this disease. There is slow evolution of the disease with hepatosplenomegaly, hypersplenism, pathological fractures and often no neurological deficit.

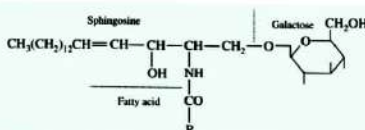


Fig. 13-14. Structure of a ceramide

The name **ceramide** is commonly used to designate the sphingosine–fatty acid (or N-acylsphingosine) portion of the cerebrosides. Here again the sphingosine carries the galactose by glycosidic linkage on its primary alcohol group and the fatty acid by an amide linkage on its primary amino group.

Individual cerebrosides are differentiated on the basis of their fatty acid component (Fig. 13-15). The various classes, so differentiated, are as follows:

- Keratin—contains saturated C 24 lignoceric acid.
- Phrenosin (cerebron)—contains a 2-hydroxy derivative of lignoceric acid called cerebronic acid.

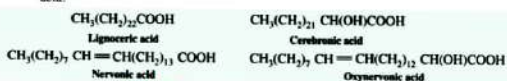


Fig. 13-15. Characteristic fatty acids of cerebrosides

- Nervon—contains an unsaturated homologue of lignoceric acid called nervonic acid.
- Oxynervon—contains a 2-hydroxy derivative of nervonic acid called oxynervonic acid.

Gangliosides

In 1955, Klenk isolated a new type of glycolipid from brain tissue and named it as **ganglioside**. These are found in significant concentrations in ganglion cells of nervous tissue (hence so named) and also in most parenchymatous tissues like spleen and erythrocytes. They make up about 6% of the membrane lipids in the gray matter of the brain. They are also found in lesser amounts in the membranes of most nonneural tissues. Gangliosides are thought to act as receptors for toxic agent pathogens, *Vibrio cholerae* influenza virus and tetanus toxin. They are also implicated in cell-cell interaction.

The structure of gangliosides (Fig. 13-16) is complex and related to that of cerebrosides. They contain a ceramide (N-acylsphingosine) linked to a carbohydrate (galactose or glucose).

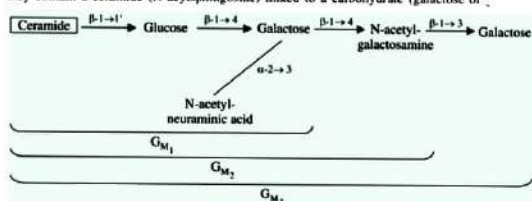


Fig. 13-16. The three gangliosides

278 / 1254

254 FUNDAMENTALS OF BIOCHEMISTRY

In addition to these, the gangliosides also generally contain 2 additional moles of carbohydrates, 1 mole of N-acetylgalactosamine and from 1 to 5 moles of N-acetylneuraminic acid (NANA). In Tay-Sachs disease, the gangliosides are stored in relatively large amounts in the brain and spleen.

CLINICAL IMPLICATIONS

Tay-Sachs disease (named after an English physician Warren Tay who first described the disease and an American neurologist Bernard Sachs who worked out its details), also called **G_{M2} gangliosidosis**, was first described in 1881. This is the oldest medically described lipid storage disease or lipidosis. It is a common hereditary disorder. Although the disease is rare in the population at large (1 in 3,00,000 births), it has a very high incidence (1 in 1,600 births) in Ashkenazic Jews, who make up about 90% of the Jewish population in America. One in 30 Ashkenazic Jews carries the defective gene in recessive form (i.e., the carrier rate is 1/30), whereas the carrier rate is 1/300 in non-Jewish Americans. Consequently, the incidence of the disease is about 100 times higher in Jewish Americans. When both parents are carrier, their children are likely to develop Tay-Sachs disease even though the parents show no symptoms of the disease. Because the disease is irreversible and needs intensive care, genetic counseling of parents has become important in averting its consequences. The disease is caused due to a deficiency of the enzyme N-acetyl-galactosaminidase (= hexosaminidase A), which cleaves a specific bond (beta-8 to 4) between an N-acetyl-D-galactosamine and a D-galactose residue in the polar head of the ganglioside, G_{M2}. In effect, G_{M2} is not degraded to G_{M1}. With the result, G_{M2}, which is also called as Tay-Sachs ganglioside, is accumulated in large amounts in the lysosomes, particularly in the brain cells to the tune of 100–300 times the normal G_{M2} content, causing degeneration of the nervous system. In Tay-Sachs disease, the symptoms are evident before an affected infant is a year old. Weakness, retardation in development and difficulty in eating are typical early symptoms. Mental retardation and blindness usually follow a few months later; death, between two and five years, is inevitable. Over 90% of the patients have a characteristic cherry-red spot in the retina. Tay-Sachs disease can be diagnosed during fetal development. Amniotic fluid is obtained by amniocentesis and assayed for b-N-acetyl-galactosaminidase activity. As mentioned, Tay-sachs disease shares the tragic element of Niemann-Pick and Gaucher diseases in that the affected child seldom lives beyond age three.

More than 15 different gangliosides have been characterized and their structures determined. These are commonly abbreviated by the letter G, with a subscript _{MF}, _D or _T to designate that they contain one, two or three sialic acid (N-acetylneuraminic or N-glycolyl neuraminic) residues respectively, and a number or letter to distinguish different members of a group from one another. In essence, they consist of an oligosaccharide chain attached to ceramide (N-acylsphingosine) by a sugar residue which is usually glucose. They lack phosphoric acid.

Sulfolipids.

A glycolipid that contains sulfur is widely distributed in plants. It is localized in the chloroplasts but is also found in the chromatophores of photosynthetic bacteria. As the sulfur in this compound is present as a sulfonic group in a hexose, this may be included under a class of compounds called sulfolipids. Its structure is given in Fig. 13-17.

Sulfatides.

A sulfate ester analogue of phrenosin, abundant in white matter of brain, is another sulfur-containing glycolipid. In it, the sulfate is present in ester linkage at C 3 of the galactose portion of the molecule. Members of this group of cerebroside-sulfuric esters have been designated as sulfatides.

The classification of compound lipids as presented here is now becoming obsolete. It is, however, more logical to classify them according to the main alcohol component (refer Table 13-3). According to this system, the compound lipids fall under 2 categories: one in which the main alcohol component is glycerol (**glycerophosphatides**) and the other in which sphingosine is the main alcohol component (**sphingolipids**). The further classification of these two categories depends on the nature of the alcohol components.

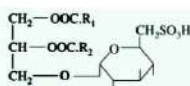


Fig. 13-17. Structure of a sulfolipid (6-sulfo-6-deoxy-α-glucosyl monoglyceride)

Table 13-3. Classification of compound lipids

Name	Main alcohol component(s)	Other alcohol component(s)*	N:P ratio
Glycerophosphatides			
Phosphatidic acids	Diglyceride (= glycerol diester)	—	0 : 1
Lecithins	Diglyceride (= glycerol diester)	Choline	1 : 1
Cephalins	Diglyceride (= glycerol diester)	Ethanolamine, serine	1 : 1
Inositides	Diglyceride (= glycerol diester)	Inositol	0 : 1
Plasmalogens	Monoglyceride (= glycerol ester) + Enol ether	Choline, ethanolamine, serine	1 : 1
Sphingolipids			
Sphingomyelins	N-acylsphingosine	Choline	2 : 1
Cerebrosides	N-acylsphingosine	Galactose, glucose	1 : 0
Sulfatides	N-acylsphingosine	Galactose (H ₂ SO ₄)	
Gangliosides	N-acylsphingosine	Galactose, glucose + N-acetylgalactosamine + N-acetylneuraminic acid	No P

* The other alcohol components in all the sphingolipids except sphingomyelins are not present as phosphoric esters, but rather in glycosidic linkages. For this reason, these sphingolipids are grouped under glycolipids.

DERIVED LIPIDS

The group derived lipids is a “catch all” group in Bloor’s classification. It includes the hydrolysis products of simple and compound lipids and also various other compounds such as steroids, terpenes, fatty acids, alcohols, fatty aldehydes, ketones etc.

STERIODS

The steroids (*stereos*^G = solid) are one of the most studied classes of biological compounds and are often found in association with fat. Since they contain no fatty acids, they are *non-saponifiable*, i.e., cannot be hydrolyzed by heating with alkali to yield soaps of their fatty acid components. Fats, on the other hand, are saponifiable and form soaps when hydrolyzed with alkali. The steroids may be separated from the fat after the latter is saponified since they occur in ‘non-saponifiable residue’. All steroids may be considered as derivatives of a fused and fully saturated ring system called **cyclopentanoperhydrophenanthrene** or **sterane**. This system consists of 3 cyclohexane rings (A, B and C) fused in nonlinear or phenanthrene manner and a terminal cyclopentane ring (D). The sterane nucleus along with the conventional numbering of the carbon atoms is shown in Fig. 13-18.

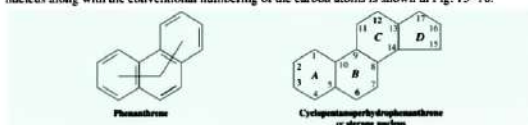


Fig. 13-18. Phenanthrene and sterane nucleus

256 FUNDAMENTALS OF BIOCHEMISTRY

It may, however, be emphasized that in steroids the hexagonal rings are not the benzene rings as in them the valences of C atoms are fully satisfied by hydrogen bonds, unless shown otherwise.

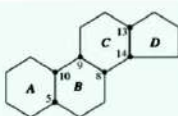


Fig. 13-19. Asymmetric centres in sterane (= gonane) nucleus (shown by solid black circles)

Stereochemical considerations of steroids. Without even considering the substituent groups, there are 6 possible asymmetric centres present in the steroid nucleus (Fig. 13-19). Consequently, the steroids have many potential stereoisomers. Each of 3 cyclohexane rings, on the basis of the tetrahedral theory, is capable of existing in 3-dimensional conformation either of a ‘chair’ or of a ‘boat’ (Fig. 13-20).

These 2 forms are aplanar and strainless and were first proposed by Sachse in 1890. The ‘chair’ form is *rigid* (in the sense that it resists distortion) and when it is changed to the ‘boat’ form, some angular deformation invariably takes place. The ‘chair’ conformation of cyclohexane itself is more *stable* than the ‘boat’ conformation. The ‘chair’ and ‘boat’ forms are both free from angle strain but because of differences in steric strain and bond opposition strain, the two forms differ in energy content (Finar, 1975).



Fig. 13-20. Two types of conformation of a cyclohexane ring

In naturally occurring steroids, all the rings are in ‘chair’ form. With respect to each other, the rings can be either *-cis* or *-trans* (Fig. 13-21). The junction between the rings A and B (A/B junction) in naturally-occurring steroids may be either *-cis* or *-trans*. But the B/C and C/D junctions are, with certain exceptions, always in *-trans* form. The idea of *-cis* and *-trans* forms with respect to A/B junction may be explained as below. The hydrogen atom on C 5 may be either on the same side of the plane of molecule as the methyl group on C 10 or on the opposite side. In the former case, rings A and B will be *-cis* to each other and the molecule is said to belong to *normal configuration*. If the hydrogen atom and the methyl group are on the opposite sides, the rings A and B are *-trans* to each other and the molecule is of *allo configuration*.

Steroids containing hydroxyl groups (hydroxy steroids) are often referred to as *sterols*. Broadly speaking, the various sterols fall under 2 groups based on their source : the animal sterols or *zoosterols* (cholesterol, cholestanol, coprostanol) and the plant sterols or *phytosterols* (ergosterol, stigmasterol, spinasterol). A third group of sterols, which are obtained from fungi including yeasts, is sometimes separately referred to as *mycosterols*. This classification, however, is not rigid as some sterols are obtained from more than one of these sources.

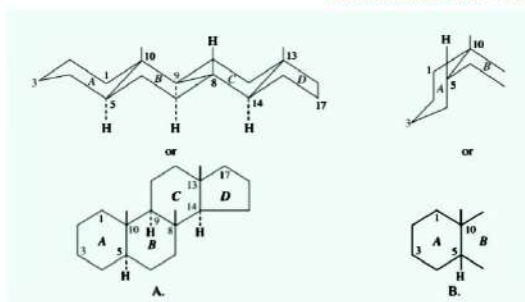


Fig. 13-21. Stereochemical aspects of steroid nucleus, A, all α -trans configuration between adjacent rings, B, all α -cis configuration between rings A and B

In the structural representation of the steroids, substituents that are above the plane of the ring (or β -oriented) are shown with solid lines, while those which are below the plane of the ring (or α -oriented) are represented with broken lines. The methyl groups at C 10 and C 13, the hydroxyl group on C 3 and the side chain at C 17 are all β -oriented.

Certain common steroids are described below.

C-29, C-28 and C-27 Steroids

1. Cholesterol (*chole*^G = bile). Cholesterol is undoubtedly the most publicized lipid in nature, because of the strong correlation between high levels of cholesterol in the blood and the incidence of diseases of the cardiovascular system in humans. It is not only an important component of some cell membranes and of plasma lipoproteins but also the precursor of many other biologically important steroids, such as bile acids and various steroid hormones. It is the principal sterol of higher animals and is especially abundant in nerve tissues and in gallstones. It occurs either free or as fatty esters in all animal cells. It was first isolated in 1784, from human gallstones which consist almost entirely of cholesterol and hence so named (cholesterol literally means 'solid alcohol from bile'). Its main sources are fish liver oils and the brain and spinal cord of cattle. White matter contains as much as 14%, gray matter 5%, spinal

Cholesterol is generally believed to be notorious as a major cause of heart disease. There are 2 types of cholesterol, the low-density lipoprotein cholesterol (LDL-C) and the high-density lipoprotein cholesterol (HDL-C). The LDL-C grows with intake of eggs and dairy products and triggers heart trouble. However, in a study conducted by Dr. Ian Craig of the Göteborg University, Sweden in the early 70s showed that patients with heart trouble such as myocardial infarction had low levels of HDL-C, as compared to other people. He pointed out that most of the preventive heart attack studies have so far been aimed at lowering the LDL-C levels through drugs and diet, but it seemed that normal variations in diet made little difference in HDL-C levels. Also, HDL-C levels were generally found higher in women than men. Perhaps this is the reason why fewer women than men have heart attacks. There are, thus, hopeful signs that incidence of heart disease may be lowered by increasing the HDL-C level in heart patients.

258 FUNDAMENTALS OF BIOCHEMISTRY

cord 12% and liver about 1% cholesterol. Cholesterol is, however, not found in plant fats. Its parent hydrocarbon is cholestane, $C_{27}H_{48}$. The structure of cholesterol was determined by the German chemist, Adolph Windaus (LT, 1879 – 1959), who received 1928 Nobel Prize in Chemistry. Cholesterol (Fig. 13-22) has a molecular formula, $C_{27}H_{46}OH$. In addition to an OH group at C3, there is a double bond at C5. The hydroxyl group constitutes its polar head, the rest of the molecule is hydrophobic. It is a white crystalline solid and is optically active, $[\alpha]_D^{25} + 39^\circ$. The crystals are rhombic plates with one of the angles broken. It has a melting point of $149^\circ C$. Konrad S. Bloch, a Germany-born American biochemist, elucidated the biosynthetic pathway of cholesterol synthesis, one of the most complex known, for which he received the coveted Nobel Prize in Chemistry along with Feodor Lynen, a German and John Cornforth, a Briton, in 1964. Bloch combined work and pleasure by traveling to Bimini for collection of shark livers.

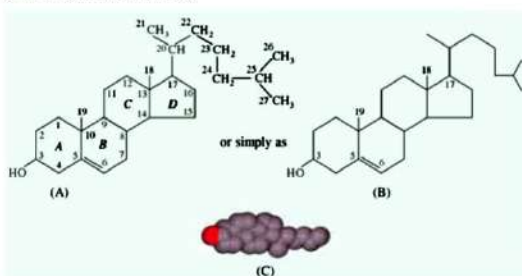


Fig. 13-22. Structure of cholesterol
(3-hydroxy-5, 6-cholestane)

- (A) : Molecular representation. The individual rings in the fused-ring system are designated A through D.
(B) : Simplified version. It is customary to write steroid structure in the manner shown here, without, however, the methyl substituents.
(C) : Space-filling model. The C is shown with grey, vinyl C with yellow, O with red, N with blue and P with green. H is not shown.

[Courtesy : (C) Richard Pastor and Richard Venable]

On receiving the Nobel Prize in 1985 for their work on cholesterol, Michael Brown and Joseph Goldstein recounted in their lecture the extraordinary history of cholesterol :

"Cholesterol is the most highly decorated small molecule in biology. **Thirteen Nobel Prizes** have been awarded to scientists who devoted major parts of their careers to cholesterol. Ever since it was isolated from gallstones in 1784, cholesterol has exerted an almost hypnotic fascination for scientists from the most diverse areas of science and medicine.... Cholesterol is a Janus-faced molecule. The very property that makes it useful in cell membranes, namely its absolute insolubility in water, also makes it lethal."

(Courtesy : The Nobel Foundation, 1985)

Table 13-4 provides data on fat and cholesterol contents of some notable animal foods.

Table 13-4. Fat and cholesterol content of some animal foods*

Food Item	Fat g/100 g	Saturated fatty acids g/100 g	Cholesterol mg/100 g
Butter	80	50	250
Ghee	100	65	300
Milk (cow)	4	2	14
Milk (buffalo)	8	4	16
Milk (skimmed)	0.1	—	2
Milk (condensed)	10	6	40
Cream	13	8	40
Cheese	25	15	100
Egg (whole)	11	4	400
Egg (yolk)	30	9	1120
Chicken (with skin)	18	6	100
Chicken (without skin)	4	1	60
Beef	16	8	70
Mutton	13	7	65
Pork	35	13	90
Organ meats			
Brain	6	2	2000
Heart	5	2	150
Kidney	2	1	370
Liver	9	3	300

* The data are based on per 100 g edible portion of the food item.

No vegetable oil contains any cholesterol; hence, it is misleading for oil manufacturers to advertise any oil as 'low cholesterol' oil. Saturated fats of animals origin (such as butter, ghee and lard) do contain cholesterol but *vanaspathi*, which is derived from vegetable oils, does not contain cholesterol. Only a little portion of the body cholesterol is derived from diet. The bulk of it is synthesized in the body. Unsaturated fatty acids lower blood cholesterol levels by controlling cholesterol's synthesis as well as its elimination from the body. Thus, vegetable oils when consumed within the recommended amounts help to maintain proper levels of lipids in the blood than do the saturated fats like butter, ghee, lard and *vanaspathi*. These saturated fats elevate blood lipids (cholesterol and triglycerides) and reduce the ratio of good to bad cholesterol even when consumed within the recommended limits. High intake of dietary cholesterol can probably raise blood cholesterol levels and hence should be avoided, esp., by the elderly people and those having high blood cholesterol.

The level of cholesterol in the blood is measured in milligrams per decilitre (mg/dl), which is equivalent to parts per 1,00,000. The levels range from less than 50 in infants to an average of 215 in adults to 1,200 or more in individuals suffering from a rare inherited disease called *familial hypercholesterolemia*. For those persons in the normal range, about two-thirds of their cholesterol is transported as LDLs. Most of the rest is carried by HDLs.

An often-posed question relates to the consumption of eggs and meat, and sometimes even milk, because of the relatively high content of cholesterol in these foods. Milk is no doubt a very good and safe food. Persons aspiring to control their cholesterol intake (and calories) must use defatted or skimmed milk. Egg and meat proteins are of high grade and desirable especially for the growing children. Those desirous of reducing their cholesterol intake should eat only the white of egg and lean on white meat like chicken and fish. Nutritionally, fish is a very good food since it has high-quality

Contents

260 FUNDAMENTALS OF BIOCHEMISTRY

protein (fish fat is rich in *n-3* PUFA) and its flesh has very little cholesterol. Organ meats like brain and liver are rich in cholesterol and should hence be avoided by those wanting to restrict cholesterol.

The notion that cholesterol is a poisonous substance in the body is misleading. It is the parent hydrocarbon from which many important hormones (corticosteroids and sex steroids, for example) are synthesized. The problem creates when blood cholesterol level rise and the level of LDL-cholesterol, which carries cholesterol to the tissues, goes below the desired level. HDL-cholesterol or good cholesterol removes cholesterol from the tissues and helps eliminating it from the body. High blood cholesterol results in the deposition of cholesterol in the arteries and narrows the passage, a condition called *atherosclerosis*. Also, the arterial surface roughens due to cholesterol deposition, which leads to frequent blood clotting or *thrombosis*. Both are major health hazards. Studies, however, point out that even very low levels of blood cholesterol are too not desirable. *The thumb rule, therefore, is moderation*. Both too much and too little of fat may be harmful.

In blood about two-thirds of the cholesterol is esterified mainly to unsaturated fatty acids, the remaining portion occurring as the free alcohol. Reduction of the double bond gives rise to 2 products, coprostanol and cholestanol.

2. **Coprostanol (= coprosterol)**. It occurs in feces and is produced in the intestine as a result of bacterial action on the double bond of cholesterol. The A/B junction is *-cis* in coprostanol in contrast to *-trans* in cholesterol.

3. **Cholestanol**. It occurs as a minor constituent of the sterols of blood and other tissues. Here the A/B junction is in *-trans* form.

4. **Ergosterol**. It is present in ergot (hence its nomenclature), yeast and the mould *Neurospora*. Its parent hydrocarbon is ergostane, $C_{28}H_{48}$. Ergosterol (Fig. 13-23) has a molecular formula, $C_{28}H_{44}OH$ with one OH group at C_3 and 3 double bonds at C_5 , C_7 and C_{22} . It is also optically active, $[\alpha]_D^{25} -135^\circ$.

Rupture of the ring B by UV radiation produces vitamin D_2 which is chemically known as ergocalciferol. A similar compound cholecalciferol (or vitamin D_3) is, however, obtained from 7-dehydrocholesterol on irradiation with UV light (refer Chapter 32).

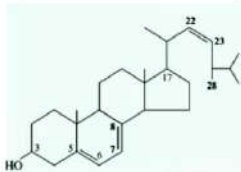


Fig. 13-23. Structure of ergosterol

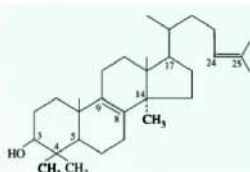


Fig. 13-24. Structure of lanosterol

5. **Lanosterol (= cryptosterol)**. It is a major constituent of wool fat and is also present in minor quantities in liver and yeast. Lanosterol (Fig. 13-24) is a C_{30} compound with twin methyl groups at C_4 and a third angular methyl group on C_{14} . There are 2 double bonds at C_5 and C_{24} . It is an intermediate in the biosynthesis of cholesterol.

A number of sterols have also been obtained from various plant and animal sources, for example **stigmasterol** (from soybean and wheat germ oils), **spinasterol** (spinach and cabbage), **sitosterol** (many higher plants), **ostreasterol** (oysters) and **chondrillasterol** (marine sponges).

C 24 Steroids or Bile Acids

The bile acids are the important end products of cholesterol metabolism in higher plants. About 20 natural bile acids have been characterized (Table 13-4). All these are derived from a C 24 parent steroid, cholanolic acid and resemble coprostanol in having rings A and B in *cis* form. The most abundant bile acids in human bile are cholic acid (25-60% of the total bile acids), chenodeoxycholic acid (30-50%) and deoxycholic acid (5-25%). Various bile acids differ from each other in the number and position of OH groups which are all in a configuration. In the bile acids, the number of OH group(s) may be 1, 2, or 3 and the position of OH group(s) may be any of the following : 3, 6, 7, 11, 12 and 23. The side chain is usually made up of 5 carbon atoms and bears the carboxyl group.

Table 13-5. Some important natural bile acids

Name	Formula	Melting point, °C	Position of OH group(s)	Source
Monohydroxycholanolic acid Lithocholic acid	$C_{23}H_{40}(OH)_1COOH$	186	3	Man, Ox
Dihydroxycholanolic acids Deoxycholic acid	$C_{23}H_{40}(OH)_2COOH$	172	3, 12	Man, Ox
Chenodeoxycholic acid		140	3, 7	Man, Ox, Hen
α -hydroxydeoxycholic acid		197	3, 6	Pig
Trihydroxycholanolic acid Cholanolic acid	$C_{23}H_{40}(OH)_3COOH$	195	3, 7, 12	Man, Ox

The structure of two principal bile acids is given in Fig. 13-25.

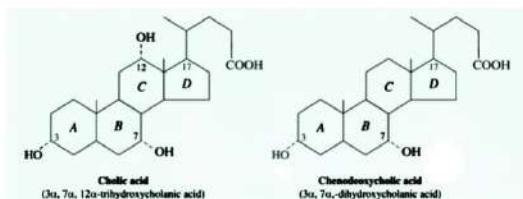


Fig. 13-25. Two principal bile acids

These acids are coupled in amide linkage to the amino acids glycine (NH_2-CH_2-COOH) and taurine ($NH_2-CH_2-CH_2-SO_3H$) to form glycocholic and taurocholic acids respectively (Fig. 13-26). The reaction is characteristic of COOH group and involves the side chain only. The salts of these conjugated acids have high surface activity and hence promote the intestinal absorption of lipids like cholesterol. They are water-soluble and are powerful detergents.

Contents

262 FUNDAMENTALS OF BIOCHEMISTRY

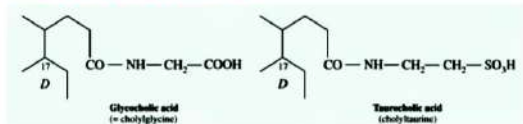


Fig. 13-26. Side chain of two bile-acid amides

C 21, C 19 and C 18 Steroids

On account of their hormonal role, these have been described in Chapter 30.

Several important poisons are based on the steroid structure. **Digitoxin** (Fig. 13-27 a, b) is present in the foxglove plant (*Digitalis purpurea*) and has a powerful heart stimulating action. Another cardiac glycoside, **Quabain**, (Fig. 13-27 c) is isolated from the East African Quabio tree and has the sugar, rhamnose, attached to a modified sterol nucleus. It is of interest to note that it is a powerful inhibitor of the 'sodium pump', a device which normally ensures that the cell content of potassium is higher and that of sodium is lower than in the circumbient fluid.

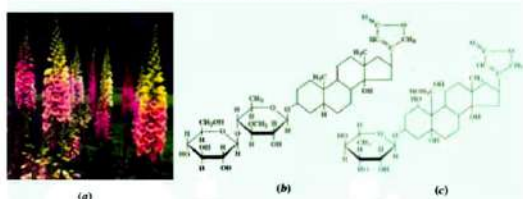


Fig. 13.27. The foxglove plant and two most-commonly prescribed cardiac drugs

(a) The foxglove plant, *Digitalis purpurea*. The leaves of purple foxglove plant are the source of the popular heart muscle stimulant, digitalis.

(b) Digitoxin. It is extracted from the dried leaves of foxglove plant and is the major component of digitalis, one of the most-widely used cardiac muscle stimulant.

(c) Quabain. It is a cardiac glycoside, isolated from the East African Quabio tree. (Courtesy : (a) Derek Fell)

Table 13-6 lists the occurrence and biological role of some important steroids.

Table 13-6. Some steroids and their biological roles

Group	Representative	Formula	Occurrence	Biological role
Sterols				
C ₃₀	Lanosterol	$C_{30}H_{54}O$	Wool fat	Biosynthesis of cholesterol
C ₂₈	Ergosterol	$C_{28}H_{44}O$	Yeast	Provitamin D
C ₂₇	Cholesterol	$C_{27}H_{46}O$	Animal fats	Structural component and precursor

CLASSIFICATION OF LIPIDS 263

	7-dehydro-cholesterol	$C_{27}H_{44}O$	Skin	Provitamin D
Bile acids	C_{24} Cholic acid	$C_{24}H_{46}O_6$	Intestine and gall-bladder	Absorption of fats
	Chenodeoxycholic acid	$C_{24}H_{46}O_6$	Intestine and gall-bladder	Absorption of fats
Hormones	C_{21} Progesterone	$C_{21}H_{32}O_2$	Corpus luteum	Hormone
	Deoxycorticosterone	$C_{21}H_{30}O_3$	Adrenal cortex	Hormone
	Cortisol	$C_{21}H_{30}O_5$	Adrenal cortex	Hormone
	C_{19} Testosterone	$C_{19}H_{28}O_2$	Testes	Hormone
	C_{18} Estradiol	$C_{18}H_{24}O_2$	Ovaries	Hormone

TERPENES

Among the nonsaponifiable lipids found in plants are many hydrocarbons known as terpenes (from *turpentine*). In general, these hydrocarbons and their oxygenated derivatives have less than 40 carbon atoms. The simplest terpenes are called *monoterpenes* and conform to the formula $C_{10}H_{16}$ (equivalent to 2 isoprene units, Fig. 13-28), those with the formula $C_{15}H_{24}$ are called as *sesquiterpenes*, with $C_{20}H_{32}$ as *diterpenes* and with $C_{30}H_{48}$ as *triterpenes*. Terpenes with 40 carbon atoms (or *tetraterpenes*) include compounds called carotenoids (refer Chapter 32).

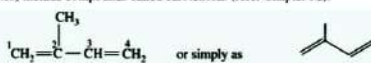
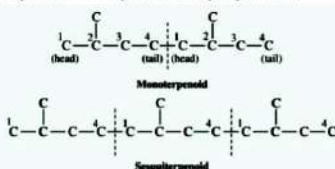


Fig. 13-28. Structure of isoprene (2-methyl-1,3-butadiene)

In fact, O. Wallach (1910 Nobel Laureate in Chemistry) was the first to point out, in 1887, that nearly all the terpenoids are made of varying number of repetitive units (C_5H_8), called isoprene units. His finding later came to be known **isoprene rule**. Structurally, isoprene is a 5-carbon diene. The carbon skeletons of open-chain monoterpenoids and sesquiterpenoids are :



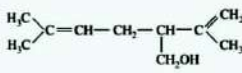
Even not only the presence of isoprene units but their special type of arrangement is found to be present in nearly all the terpenoids. Ingold, in 1925, formulated this observation under another rule, the **special isoprene rule**, according to which the isoprene units in terpenoids are usually joined in *head-to-tail linkages* or 1,4 linkages (the branched end of the isoprene unit was considered as the head).

288 / 1254

264 FUNDAMENTALS OF BIOCHEMISTRY



Cryptone



Lavandulol

Exceptions, however, do occur for these two isoprene rules. For example, *cryptone*, a natural terpenoid, contains only 9 carbon atoms instead of 10. Also, in *lavandulol*, the 2 isoprene units are not joined in head-to-tail manner, and in *carotenoids*, the two halves (each with 4 isoprene units) are linked with each other by *tail-to-tail* at the centre.

Monoterpenes and Sesquiterpenes

The fragrances of many plants arise from volatile C_{10} and C_{15} terpenes. These and their oxygenated derivatives occur as components of the essential oils, some of which are used as perfumes. Important monoterpenes (Fig. 13-29) are *myrcene* (from oil of bay), *geraniol* (from rose oil), *limonene* (from lemon oil) and *menthol* (from peppermint oil).

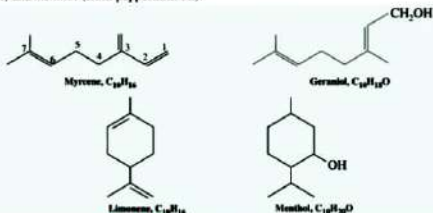


Fig. 13-29. Four important monoterpenes

Among the sesquiterpenes, farnesol (Fig. 13-30) is widely distributed but is present in essential oils in small amounts.



Fig. 13-30. Structure of farnesol

Diterpenes

These are usually found as substituents of the resins and balsams. The two resin acids, **abietic** and **sapietic** (Fig. 13-31) are the best known tricyclic diterpenes.

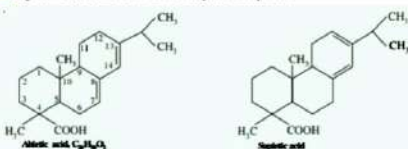


Fig. 13-31. Two best known diterpenes

Vitamins A₁ and A₂ and their aldehydes called retinenes are important monocyclic derivatives of diterpenes.

Phytol

(C₂₀H₄₀O; b.p. 145°C) is an acyclic diterpene and is obtained from the hydrolysis of chlorophyll. It was isolated from nettles by Paul Karrer *et al* in 1943. The phytol molecule (Fig. 13-32) has 2 chiral (= asymmetric) centres, 7 and 11. Natural phytol is very weakly dextrorotatory.

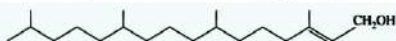


Fig. 13-32. Structure of phytol

Triterpenes

These are not widespread in nature but are significant in that some of them are intermediates in the biosynthesis of cholesterol. Two such compounds are: a tetracyclic alcohol, **lanosterol** (see page 260) and an acyclic hydrocarbon, **squalene**. Squalene (C₃₀H₅₀; b.p. 240–242°C) was first isolated from the liver of sharks (genus *Squalus*), hence so named. The olive oil and several other vegetable oils are other sources. It has also been detected in the leaves. Each squalene molecule (Fig. 13-33) has 6 double bonds. The double-bond system present is of nonconjugated type. The conjugated double bonds are, however, absent from the molecule.



Fig. 13-33. Structure of squalene

Polyterpene

Mention, however, should also be made of **rubber** (MW = ca 3,00,000), a polyterpene present in the latex of many tropical plants. A molecule of rubber (Fig. 13-34) is composed of about 500 to 5,000 isoprene units joined in a long straight chain. When the bark of the rubber tree is cut, latex slowly exudes from the cut. Addition of acetic acid coagulates the rubber, which is then separated from the liquor and either processed into blocks or rolled into sheets, and finally dried in a current of warm air, or smoked.

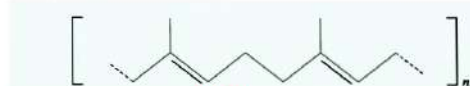


Fig. 13-34. Structure of rubber
(*cis* configuration about double bonds)

Carotenoids

The carotenoids are tetraterpenes. These are widely distributed in both the plant and animal kingdoms but are exclusively of plant origin. These occur in unsaponifiable residue of plant and animal lipids. They are isoprene derivatives with a high degree of unsaturation. Because of the presence of many conjugated double bonds, they are coloured red or yellow. As an example, the pigment of

266 FUNDAMENTALS OF BIOCHEMISTRY

tomato (*lycopene*) and that of carrot (α - and β -*carotene*) are red while many oxygen-containing carotenoids are yellow (*xanthophylls*). Since the olefinic (= double) bonds permit *cis-trans* isomerism, numerous forms are possible. *Most carotenoids, however, exist in all-trans form*. The presence of long hydrocarbon chain in carotenoids makes them lipid-soluble; they are hence also called *lipochromes* or *chromolipids*.

Lycopene

Lycopene (C₄₀H₅₆) is the main pigment of tomato, paprika and many other fruits. It is a highly unsaturated, unbranched, long chain hydrocarbon (a polyene) and is composed of two identical units (C₂₀H₂₈), joined by a double bond between C₁₅ and C_{15'}. Each of these units may, in turn, be considered to have been derived from 4 isoprene units (C₅H₈). A molecule of lycopene (Fig. 13-35) in all contains as many as 13 double bonds, of which 11 are conjugated.

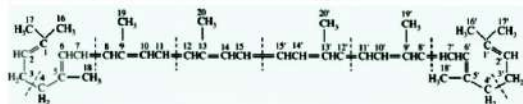


Fig. 13-35. Structure of lycopene

(Dotted lines mark the positions where various isoprene-like segments are joined with each other.)

Carotene

Another group of naturally occurring carotenoids, with the same molecular formula as that of lycopene, is **carotene**. Carotene was first isolated by Wackenroder (1831) from carrots (this was the origin of the name *carotin*, which was later on modified to *carotene*). The 3 types of carotenes are:

1. α -carotene—violet crystals; m.p. 187°C; optically active (dextrorotatory).
2. β -carotene—red crystals; m.p. 183°C; optically inactive.
3. γ -carotene—dark red crystals; m.p. 152–154°C; optically inactive.

(The prefixes α , β and γ in the carotenes were assigned not on the basis of structural properties but rather as a result of observations of their elution patterns on a CaCO₃ chromatographic column, eluted with petroleum ether).

The carotenes are obtained commercially by chromatography. The two best sources are carrots and alfalfa. Their relative proportion, however, varies with the source; for example carrots contain 15% α -, 85% β - and 0.1% γ -form. These are unstable to air, heat, acids and alkalis.

It is to be noted that the ends of the lycopene chain can easily close up to form rings, with the disappearance of the double bonds. Ring closure on only one end results in the production of

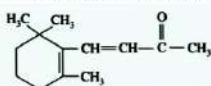
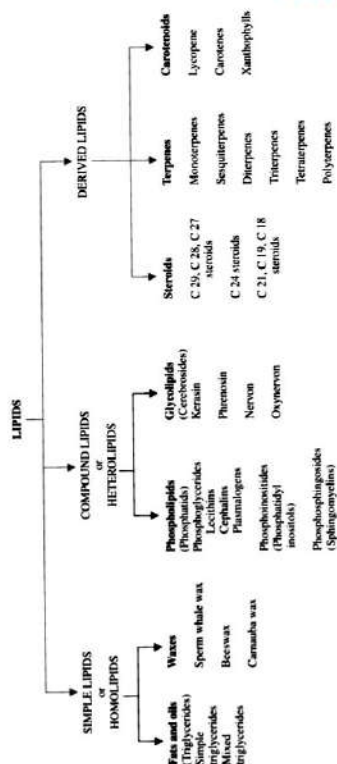


Fig. 13-36. Structure of β -ionone

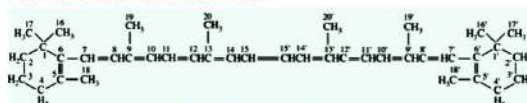
γ -carotene while closure of the ring at both ends produces β - and α -carotenes which differ from each other in the position of the double bonds in the rings. The ring is structurally related to the ionones. (Fig. 13-36). In β -carotene, both annular double bonds are in conjugation with the system of double bonds of the long chain (β -ionone structure). In α -carotene, one of the annular double bonds is removed from the system of conjugation by one position (α -ionone structure). Of particular interest is β -carotene (Fig. 13-37), which is precursor of vitamin A.



292 / 1254

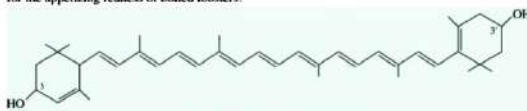


268 FUNDAMENTALS OF BIOCHEMISTRY

Fig. 13-37. Structure of β -carotene

Xanthophylls

Xanthophylls are characterized by the presence of hydroxyl groups in the ionone rings of carotenes, in the para position to the long chain. The xanthophyll of the leaf (*lutein*, $C_{40}H_{56}O_2$, Fig. 13-38) is derived from α -carotene while that of corn (*zeaxanthine*) from β -carotene. The xanthophyll of crustaceans (*astaxanthine*) is, however, more rich in oxygen. Astaxanthine is responsible for the appetizing redness of boiled lobsters.

Fig. 13-38. Structure of *lutein*
(3, 3'-dihydroxy- α -carotene)

The classification of lipids described above may be summarized as depicted on page 252.

REFERENCES

See list following Chapter 14.

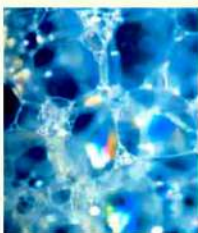
PROBLEMS

- A common structural feature of membrane lipid molecules is their amphipathic nature. For example, in phosphatidylcholine, the two fatty acid chains are hydrophobic and the phosphocholine head group is hydrophilic. For each of the following membrane lipids, name the components that serve as the hydrophobic and hydrophilic units:
 - phosphatidylethanolamine
 - sphingomyelin
 - galactosylcerebroside
 - cholesterol
- Which of the following could not be classified as a lipid?
 - steroids
 - fats and oils
 - waxes
 - photosynthetic pigments
 - anthocyanin pigments

- Physical Properties
 - State • Colour, Odour and Taste • Solubility • Melting Point • Specific Gravity • Geometric Isomerism • Insulation • Emulsification • Surface Tension
- Chemical Properties
 - Reactions Involving COOH Group • Reactions Involving Double Bond • Reactions Involving OH Group
- Quantitative Tests
- Fats-Facts and Fantasies

Lipids-III

Properties of Fats and Oils



The surface of a soap bubble is a bilayer formed by detergent molecules.

The polar heads (red) pack together leaving the hydrophobic groups (green) in contact with air on the inside and outside of the bubble. Other bilayer structures define the boundary of a cell.

[Courtesy : (Upper) Photonica]

The fats and oils owe the manifestation of their properties to the fatty acids and alcohols, the two component units.

PHYSICAL PROPERTIES

1. State. Fats containing saturated fatty acids are solid at ordinary room temperature. The *animal fats* belong to this category. Most *plant fats*, on the contrary, possess unsaturated fatty acids and are, henceforth, liquid at room temperature.

2. Colour, Odour and Taste. When pure, the fats are colourless, virtually odourless and possess an extremely bland taste. They are capable of absorbing a variety of odours and hence flavour during storage. For the same reason, a housewife knows that the flavour of an onion quickly permeates butter that is stored with it in a refrigerator. In some cases, however, this absorbing property of fats is of advantage. For example, the perfumes of some flowers can be isolated by placing their petals in contact with the fat for a certain period, then extracting the fat with alcohol and concentrating the essence.

3. Solubility. The fats are, however, only sparingly soluble in water. These are, therefore, described as *hydrophobic* in contrast to the water-soluble or *hydrophilic* substances like many carbohydrates and proteins.

However, these are freely soluble in organic solvents like chloroform, ether, acetone and benzene. These solvents, as they dissolve fats in them, are also known as '*fat solvents*'. The solubility of the fatty acids in organic

294 / 1254

Contents

270 FUNDAMENTALS OF BIOCHEMISTRY

solvents, in fact, decreases with the increase of chain length. The introduction of hydroxyl groups, however, increases solubility.

4. Melting point. The melting point of fats depends on the chain length of the constituent fatty acid and the degree of unsaturation. Fats containing saturated fatty acids from C 4 to C 8 are liquid at room temperature but those containing C 10 or higher saturated fatty acids are solid and their melting points increase with increasing chain length. With the introduction of double bond in the fat molecule, the melting point lowers considerably. It may be stated, in general, that *greater the degree of unsaturation (or higher the number of double bonds) of the constituent fatty acid, the lower is the melting point of the fat*. This may be easily visualized in terms of constituent fatty acids from the Table 14-1. In fact, short chain length and unsaturation enhance the fluidity of fatty acids and of their derivatives.

Table 14-1. Melting points of the common fatty acids*

Fatty acid	No. of carbon atoms	No. of double bonds	M.P. (in °C)
Saturated			
Caprylic	8	0	16.0
Capric	10	0	31.0
Lauroic	12	0	44.2
Myristic	14	0	53.9
Palmitic	16	0	63.1
Stearic	18	0	69.6
Arachidic	20	0	76.5
Behenic	22	0	79.9
Lignoceric	24	0	86.0
Unsaturated			
Palmitoleic	16	1	11.0
Oleic	18	1	13.4
Linoleic	18	2	-5.0
Linolenic	18	3	-11.0
Arachidonic	20	4	-49.5

* Here mention has been made only of even numbered fatty acids as they alone are one of the two constituents of the saturated fats. If, however, the entire homologous series of the saturated fatty acids is taken into consideration, the fact emerges that this series shows an alternation or oscillation of melting points, i.e., the melting point of an 'even' acid is higher than that of the 'odd' acid immediately below and above it. Thus, the fatty acid series follows what is known as the *odd-even rule*. However, in many other homologous series the melting points of the *n*-members rise continuously, tending towards a maximum value.

5. Specific gravity. The specific gravity of the fats is less than 1 (about 0.86) and, therefore, they float on water surface. Solid fats are lighter than the liquid fats. Oils spread on water to form thin monomolecular layers. In general, either unsaturation of the fatty acid chains increase or increase in chain length of the fatty acid residues tend to increase the specific gravity.

6. Geometric isomerism. As stated earlier, the presence of double bond (*s*) in the unsaturated fatty acid part of the fat molecule produces geometric (or *cis-trans*) isomerism.

Contents

7. Insulation. The fats possess *high insulating power*, i.e., they are bad conductor of heat. A layer of fat below the skin provides a sort of blanket for warm-blooded animals (or *homiotherms*). This is especially important for whales and seals which have to maintain a high temperature in cold waters. The fishes are cold-blooded animals (or *poikilotherms*) and, therefore, do not require maintenance of high temperature and so have very little subcutaneous fat.

8. Emulsification. It is the process by which a lipid mass is converted into a number of small lipid droplets. The fats may be emulsified by shaking either with water or with emulsifying agents like soaps, gums, proteins etc. An emulsifying agent helps in the production of a finely divided suspension of a fat in an aqueous medium. The hydrocarbon portions of the two (the emulsifier and the fat) tend to aggregate. This leaves the water-soluble group of the emulsifier projecting into the aqueous phase. A fat droplet will associate with a number of molecules of the emulsifier, thus producing a new water-soluble surface. Water molecules, henceforth, tend to be held in a layer or 'cloud' around each droplet, thus disallowing the aggregation of the fat droplets (Fig. 14-1).

The process of emulsification is of great metabolic significance. In fact, *the fats have to be emulsified before they can be absorbed by the intestinal wall*. The process is accomplished by the bile juice secreted from liver.

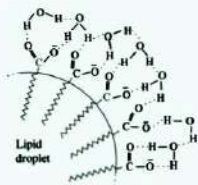


Fig. 14-1. A portion of the surface of an emulsified fat globule

[Here, the emulsifier selected is a simple anion of a fatty acid i.e., a soap. The projecting COOH groups are linked with water molecules by H_2 bonds. Additional H_2 bonding between the water molecules leads to the further stability of the resulting water 'cloud' around the fat droplet.]

(Adapted from Fairley JL and Kilgour GL, 1966)

9. Surface tension. The force with which the surface molecules are held together is called the surface tension. When liquid fat is poured on water, it spreads uniformly over the surface of water in the form of a unimolecular layer and thus reduces the surface tension of water.

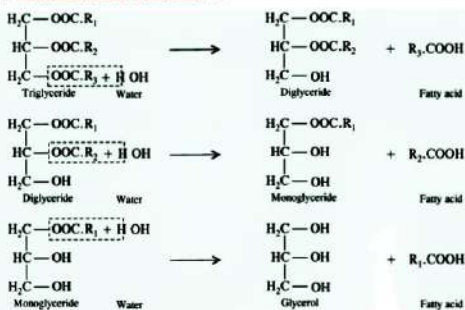
CHEMICAL PROPERTIES

The chemical reactions of the fats reflect the reactivities of the ester linkage and the degree of unsaturation of the hydrocarbon chain.

REACTIONS INVOLVING COOH GROUP

1. Hydrolysis. The fats are hydrolyzed by the enzymes *lipases* to yield fatty acids and glycerol. The lipases catalyze this reaction at a slightly alkaline pH (7.5–8.5) in a stepwise manner. The fats first split to produce *diglycerides*, part of these are then split to *monoglycerides*. Finally, part of the monoglycerides split to yield *fatty acid* and *glycerol*.

272 FUNDAMENTALS OF BIOCHEMISTRY



In the intestine, the absorption of mono-, di- and triglycerides is so rapid that very little free glycerol is formed.

2. Saponification. The hydrolysis of fats by alkali is called *saponification*. This reaction results in the formation of glycerol and salts of fatty acids which are called *soaps* (Fig. 14-2). The soaps are of

two types : hard and soft. *Hard soaps* such as the common bar

soaps are the sodium salts of the higher fatty acids. *Soft soaps* are the potassium salts of higher fatty acids and are marketed as semisolids or pastes. The fatty acid salts of calcium, magnesium, zinc and lead are, however, insoluble in water. Calcium soaps are used industrially as lubricating greases. Zinc soaps are employed in the manufacture of talcum powder and other cosmetics. Lead and magnesium soaps are used in paints industry to hasten the process of drying.

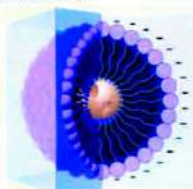
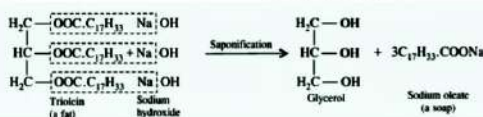


Fig. 14-2. Soaps consist of fatty acids

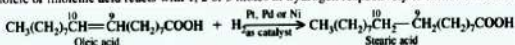
In this schematic drawing of a soap micelle, the nonpolar tails of the fatty acids are directed inward, where they interact with the greasy matter to be dissolved. The negatively-charged heads are located at the surface of the micelle, where they interact with the surrounding water. Membrane proteins, which also tend to be insoluble in water, can also be solubilized in this way by extraction of membranes with detergents.

Soaps are important cleansing agents. Their cleansing property is due to their *emulsifying action* (i.e., capacity to render more prolonged the mixing of oil and water). This is accomplished by means of negative charge the soap anion confers on oil droplets. The electrostatic repulsion then prevents the coalescence of soap and oil droplets into an oil phase.

3. Hydrolytic rancidity. When butter or other fats are stored, they often become rancid and hence unpalatable. Rancidity is caused by the growth of microorganisms which secrete enzymes like *lipases*. These split the fats into glycerol and free fatty acids. The fatty acids impart unpleasant odour and flavour to the fat. However, butter may be prevented from becoming rancid by refrigeration or by exclusion of water.

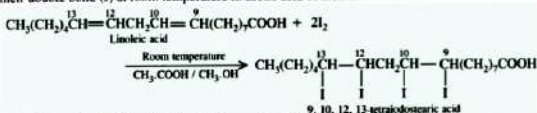
REACTIONS INVOLVING DOUBLE BOND

1. Hydrogenation. Unsaturated fatty acids, either free or combined in lipids, react with gaseous hydrogen to yield the *saturated fatty acids*. The reaction is catalyzed by platinum, palladium or nickel. The addition of hydrogen takes place at the C—C double bond(s). Thus, 1 mole of oleic, linoleic or linolenic acid reacts with 1, 2 or 3 moles of hydrogen respectively to form stearic acid.



This reaction is of great commercial importance since it permits transformation of inexpensive and unsaturated liquid vegetable fats into solid fats. The latter are used in the manufacture of candles, vegetable shortenings and oleomargarine.

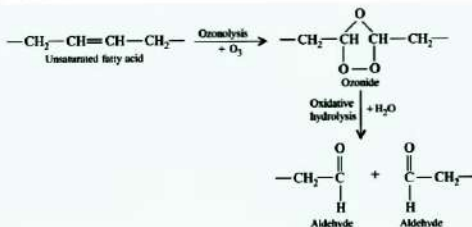
2. Halogenation. Unsaturated fatty acids and their esters can take up halogens like Br_2 and I_2 at their double bond (s) at room temperature in acetic acid or methanol solution.



This reaction is the basis of the 'iodine number determination'.

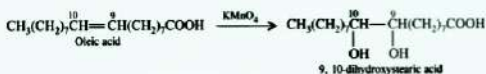
3. Oxidation. Unsaturated fatty acids are susceptible to oxidation at their double bonds. Oxidation may be carried with ozone or KMnO_4 .

(a) With ozone – An unstable *ozonide* is formed which later cleaves by water to give rise to 2 *aldehydic groups*.

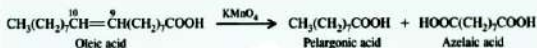


274 FUNDAMENTALS OF BIOCHEMISTRY

(b) With KMnO_4 – Under mild conditions, the glycols are formed at the sites of double bonds.



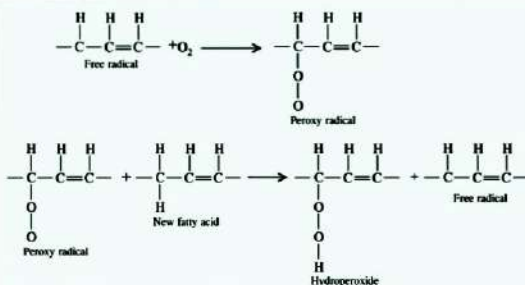
Under vigorous conditions, the same reagent cleaves the molecule at the double bond and oxidizes the terminal portions to the *carboxyl group*.



The oxidation reactions have been extensively used in establishing the position of double bond(s) in the fatty acid chain. This gives important clues regarding lipid structure.

4. Oxidative rancidity. Oils containing highly unsaturated fatty acids are spontaneously oxidized by atmospheric oxygen at ordinary temperatures. The oxidation takes place slowly and results in the formation of *short chain fatty acids* (C_4 to C_{10}) and *aldehydes* which give a rancid taste and odour to the fats. This type of rancidity or rancidification is called 'oxidative rancidity' and is due to a reaction called 'autooxidation'. Autooxidation proceeds by a free radical mechanism in which the α -methylene group is primarily attacked. A hydrogen atom is removed from an α -methylene group. This initiates a chain of reactions leading to oxidation (Holman, 1954).

Oxidative rancidity is observed more frequently in animal fats than in vegetable fats. This is due to the presence, in the vegetable oils, of natural 'antioxidants' such as tocopherols (= vitamin E), phenols, naphthols etc., which check autoxidation. Vitamin E is, therefore, some-times added to foods to prevent rancidity. Animal shortenings such as lard are nowadays protected against oxidative rancidity by the addition of synthetic antioxidants such as nordihydroguaiaric acid (NDGA), tertiary butyl hydroxy anisole (BHA) etc. Linseed oil, a plant oil used as a base for paints, is highly rich in unsaturated fatty acids. It undergoes autoxidation when exposed to air, followed by polymerization to a hard, resinous coating as it 'dries' or oxidizes.



Protein

CONTENTS

- Importance
- Amino Acids
- Nonstandard Protein Amino Acids
- Nonprotein Amino Acids
- Peptides
- Chemical Bonds Involved in Protein Structure
- Characteristics of Chemical Bonds
- Protein Configuration
- Dynamics of Globular Proteins Structure
- Prediction of Secondary and Tertiary Protein Structures
- Cleavage of a Protein
- Biological Roles of Proteins



Two Examples of the Thousands of Biological Structures Composed Predominantly of Protein

(a) Lenses of eyes, as in this net-casting spider, which focus light rays, and (b) *Mantis wildlife Films* (b) feathers, which are adaptations in birds for thermal insulation, flight, and sex recognition. Courtesy: (a) Frans Larling

CHAPTER

9

Proteins-I

General Structure

IMPORTANCE

157 / 1254

The name protein (*protein*^G = pre-*emine*), was first suggested, in 1838, by a Swedish chemist Berzelius to a Dutch chemist Mulder, who referred it

Jöns Jacob Berzelius (LT, 1779-1848) was Professor of Chemistry at Stockholm, Sweden. He is famous for developing the modern system of chemical symbols. He isolated several elements and is also credited with the discovery of selenium (Se). He developed the concept of atomic weight. With Humphrey Davy, he laid the foundations of electrochemistry. He is also credited with publishing the earliest book on Biochemistry entitled 'Lectures in Animal Chemistry' in 1806.

Gerardus Johannes Mulder's (LT, 1802-1880) comment about proteins reads as follows:

"There is present in plants and animals a substance which... is without doubt the most important of the known substances in living matter, and without it, life would be impossible on our planet. This material has been named Protein."

to the complex organic nitrogenous substances found in the cells of the living beings. They are most abundant intracellular macro-molecules and constitute over half the dry weight of most organisms. "Proteins occupy a central position in the architecture and functioning of living matter. They are

132

GENERAL STRUCTURE OF PROTEINS 133

intimately connected with all phases of chemical and physical activity, that constitute the life of the cell. Some proteins serve as important structural elements of the body, for example, as hair, wool and collagen, an important constituent of connective tissue; other proteins may be enzymes, hormones or oxygen-carriers. Still other proteins participate in muscular contraction, and some are associated with the genes, the hereditary factors." (White, Handler and Smith, 1964).

They are, therefore, essential to cell structure and cell function. The proteins with catalytic activity (enzymes) are largely responsible for determining the phenotype or properties of a cell in a particular environment. The total hereditary material of the cell or genotype dictates which type of protein the cell can produce. In fact, the proteins have built into their structure the information that instructs them in "what to do" (catalytic activity), "where to go" in the cell (intracellular organization) and "when and how to perform" (control of function through bacterial interactions with other proteins, small activators or inhibitors).

The bacterium *Escherichia coli* is estimated to contain about 5,000 different types of compounds which include some 3,000 different kinds of proteins and 1,000 nucleic acids. In humans, there may be 1,00,000 different kinds of proteins each with a unique structure. None of the proteins of *E. coli* is identical with any of the human proteins. Thus, in about 1.5 million species of living organisms, there are probably 10^{10} to 10^{12} different kinds of protein molecules and about 10^{10} different kinds of nucleic acids.

The constituent elements of proteins are carbon, hydrogen, oxygen, nitrogen and very rarely sulfur also. In certain complex proteins, phosphorus occurs as well. The elemental composition of proteins in plants and animals presents a great deal of variation (Table 9-1).

Table 9-1. Elemental composition of plant and animal proteins

Protein	Carbon	Hydrogen	Oxygen	Nitrogen
Green plants	54	7	38	0.003
Mammals	21	10	62	3.0

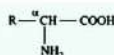
Most animal proteins contain from 0.5 to 2.0 per cent sulfur. Insulin is, however, a notable exception to this in possessing about 3.4% sulfur.

AMINO ACIDS

Structure

All proteins are macromolecules because of their very high molecular weights. These are the polymers, i.e., chain-like molecules produced by joining a number of small units of amino acids called monomers. The amino acids are, therefore, regarded as 'building blocks of proteins'. The general formula of an amino acid along with its space-filling model is presented in Fig. 9-1.

The term **macromolecule** was coined by Schaudinger in 1920s



(a) General formula



(b) Space-filling model.

The large green sphere represents the R group.

134 FUNDAMENTALS OF BIOCHEMISTRY

Each amino acid is a nitrogenous compound having both an acidic carboxyl (—COOH) and a basic amino (—NH_2) group. R stands for the side chains that are different for each amino acid. R can be as simple as a hydrogen atom (H) or a methyl group (—CH_3) or a more complex structure. The first carbon is the part of the carboxyl group. The second carbon, to which is attached the amino group, is called the α -carbon. The α -carbon of most amino acids is joined by covalent bonds to 4 different groups. Thus, the α -carbon in all the amino acids is asymmetric except in glycine where the α -carbon is symmetric. Because of this asymmetry, the amino acids (of course, except glycine) exist in two optically active forms: those having —NH_2 group to the right are designated as D-forms and those having —NH_2 group to the left as L-forms (Fig. 9-2). However, the two amino acids, threonine and isoleucine have two asymmetric carbon atoms each and thus have $2^n = 2^2 = 4$ optical isomers. At pH 7.0, both the carboxyl and amino groups are ionized.



Fig. 9-2. Two optical isomers of amino acids

Specific Rotation

It is interesting to note that the amino acids found in the proteins belong to the L-series. Many of the naturally occurring L-amino acids rotate the plane of polarized light to the left (*i.e.*, they are levorotatory) while others rotate the plane of polarized light to the right (*i.e.*, they are dextrorotatory). Table 9-2. shows the specific rotation of some L-amino acids.

Table 9-2. Specific rotation of some amino acids isolated from proteins

Amino acid	Specific rotation, $[\alpha]_D^{25}$
L-alanine	-86.2
L-histidine	-38.5
L-phenylalanine	-34.5
L-threonine	-28.5
L-serine	-7.5
L-alanine	+1.8
L-glutamic acid	+12.0
L-isoleucine	+12.4
L-arginine	+12.5
L-lysine	+13.5

Thus, it is evident that the symbols D and L do not identify the property of light rotation, *i.e.*, D-isomers can be either dextrorotatory (*d*) or levorotatory (*l*); similarly, L-isomers can be either (*d*) or (*l*). However, to minimize confusion, the symbols *d* and *l* are usually not used nowadays. Moreover, the **DL nomenclature** has limitations because it describes the asymmetry of only one carbon atom in a compound and many biomolecules contain two or more asymmetric carbon atoms. The R and S classification or **RS notation** of isomers, introduced in 1956 by Robert Cahen, Christopher Ingold and Vladimir Prelog and being currently used in chemistry, is more useful for defining the asymmetry

GENERAL STRUCTURE OF PROTEINS 135

of biomolecules because it accounts for all asymmetric carbons in an isomer. If any atom (other than H) or group on the asymmetric carbon is on the right side, that asymmetric carbon is designated as R (from *rectus* = right); conversely, if any atom (other than H) or group is on the left side, the asymmetric carbon is then designated as S (from *sinister* = left). As an illustration, isoleucine (with two asymmetric carbon atoms, C-2 and C-3) will have four stereoisomers. And the configuration of the biological isomer, L-isoleucine is more completely described as (2S), (3S)-isoleucine (Fig. 9-3).

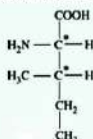


Fig. 9-3. Fisher projection formula of L-isoleucine of (2S), (3S)-isoleucine

The two asymmetric carbon atoms are indicated by an asterisk each.

Distribution in Proteins

The distribution of the 20 amino acids is not uniform in all proteins. Nearly 40% by weight of fibroin and 25% by weight of collagen are accounted for by glycine. Fibroin is all rich in alanine (30% by weight). Serine and threonine predominate in casein and phospholipids. Collagen (in connective tissue), gliadin (in wheat) and zein (in corn) are rich in proline. Human serum albumin with 585 amino acid residues has only one tryptophan moiety.

The **pulse** are notable as they lack S-containing amino acid, methionine (Met) but contain good amount of the basic amino acid, lysine (Lys); whereas **cereals** lack lysine but have sufficient quantity of methionine. When combined, these make good the deficiency of each other through mutual supplementation and are therefore better utilized in human body.

Location in Proteins

Amino acids with uncharged polar side chains are relatively hydrophilic and are usually on the outside of the proteins, while the side chains on nonpolar amino acids tend to cluster together on the inside. Amino acids with acidic or basic side chains are very polar, and they are nearly always found on the outside of the protein molecules.

Physical Properties

Regarding their physical characteristics, the amino acids are colourless crystalline substances. The crystal form may vary from slender needles (tyrosine) to thick hexagonal plates (cystine). They may be either tasteless (tyrosine), sweet (glycine and alanine) or bitter (arginine). It is interesting to note that (S)-phenylalanine is bitter in taste whereas its enantiomer (R)-phenylalanine is sweet. They have high melting points (above 200°C) and often result in decomposition. Amino acids are soluble in polar solvents such as water and ethanol but they are insoluble in nonpolar solvents such as benzene and ether. Some amino acids like tryptophan, tyrosine, histidine and phenylalanine absorb ultraviolet rays at 260–290 mμ. This property enables the identification of not only these amino acids but also the proteins which contain them.

The order of hydrophilicity and hydrophobicity of various amino acids is as follows:

136 FUNDAMENTALS OF BIOCHEMISTRY

Hydrophilic (= polar) amino acids :

Tyr > Ser > Asp > Glu > Asn > Gln > Arg

Hydrophobic (= apolar) amino acids :

Phe > Ala > Val > Gly > Leu > Cys

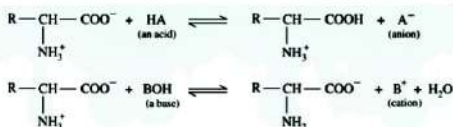
Although over 100 amino acids have been shown to be present in various plants and animals, only 20 of them (L-isomers) are found as constituent of most proteins. These 20 amino acids of proteins are often referred to as **standard, primary or normal amino acids**, to distinguish them from others. The almost universal use of these 20 amino acids for the synthesis of various protein molecules is "one of nature's enigmatic rules." These have, therefore, rightly been called as the "**magic 20**", a phrase coined by Francis Crick.

Electrochemical Properties. All molecules possessing both acidic and basic groups might exist as uncharged molecules or in ionic form or as a mixture of both. Calculations have revealed that an aqueous solution of most amino acids, glycine for example, can have only one part of uncharged molecules in one lakh parts of the ionic form. Hence, it is more accurate to write the structure of glycine as



In the ionic form, the proton migrates from the carboxyl to amino group, thus producing (COO⁻) and ammonium (NH₃⁺) ions.

Amino acids react with both acids and bases. Hence, they are **amphoteric** in nature.



Thus, in acid solution, the COO⁻ ion acquires a proton and the amino acid becomes an ammonium salt of the acid. Conversely, in alkaline solution, the NH₃⁺ ion loses a proton and the amino acid becomes the anion of a salt. These reactions are of reversible nature and depend on the pH of the medium. Henceforth, amino acids serve as buffers and tend to prevent pH change when an acid or a base is added.

The α-COOH and α-NH₂ groups of amino acids are ionized in solutions at physiological pH, with the deprotonated carboxyl group (—COO⁻) bearing a negative charge and the protonated amino group (—NH₃⁺) a positive charge. An amino acid in its dipolar state is called a **zwitterion**. The dissociable α-COOH and α-NH₃⁺ groups are responsible for the two characteristic pK values (pK₁ for —COOH and pK₂ for —NH₃⁺) of α-amino acids. An amino acid with a third dissociable group in its side chain (*i.e.*, R group) exhibits an additional pK_a value (refer Table 9-3). However, pK values depend on temperature, ionic strength and the microenvironment of the ionizable group.

161 / 1254

GENERAL STRUCTURE OF PROTEINS 137

Table 9-3. Some properties associated with the standard amino acids

Amino acid	Mr	pK ₁ (—COOH)	pK ₂ (—NH ₃ ⁺)	pK _R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Nonpolar R groups							
Alanine	89	2.34	9.69		6.01	1.8	9.0
Valine	117	2.32	9.62		5.97	4.2	6.9
Leucine	131	2.36	9.60		5.98	3.8	7.5
Isoleucine	131	2.36	9.68		6.02	4.5	4.6
Proline	115	1.99	10.96		6.48	-1.6	4.6
Phenylalanine	165	1.83	9.13		5.48	2.8	3.5
Tryptophan	204	2.38	9.39		5.89	-0.9	1.1
Methionine	149	2.28	9.21		5.74	1.9	1.7
Polar, uncharged R groups							
Glycine	75	2.34	9.60		5.97	-0.4	7.5
Serine	105	2.21	9.15	13.60	5.68	-0.8	7.1
Threonine	119	2.11	9.62	13.60	5.87	-0.7	6.0
Tyrosine	181	2.20	9.11	10.07	5.66	-1.3	3.5
Cysteine	121	1.96	8.18	10.28	5.07	2.5	2.8
Asparagine	132	2.02	8.80		5.41	-3.5	4.4
Glutamine	146	2.17	9.13		5.65	-3.5	3.9
Negatively-charged R groups							
Aspartate	133	1.88	9.60	3.65	2.77	-3.5	5.5
Glutamate	147	2.19	9.67	4.25	3.22	-3.5	6.2
Positively-charged R groups							
Lysine	146	2.18	8.95	10.53	9.74	-3.9	7.0
Arginine	174	2.17	9.04	12.48	10.76	-4.5	4.7
Histidine	154	1.82	9.17	6.00	7.59	-3.2	2.1

* A scale combining hydrophobicity and hydrophilicity; can be used to predict which amino acid will be found in an aqueous environment (-values) and which will be found in a hydrophobic environment (+ values).

† Average occurrence in over 200 proteins

Note that in nature, the commonest amino acid is alanine and the rarest amino acid is tryptophan.

(Adapted from Klapper MH, 1977)

As an instance, ionization of the amino acid, **alanine** (Fig. 9-4) is discussed.

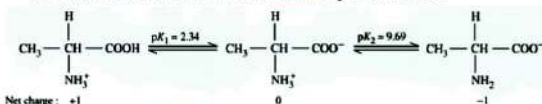


Fig. 9-4. The two dissociations of L-alanine

138 FUNDAMENTALS OF BIOCHEMISTRY

Titration involves the gradual addition or removal of protons. Fig. 9-5 shows the titration curve of the diprotic form of alanine. Each molecule of added base results in the net removal of one proton from one molecule of amino acid. The plot has two distinct stages, each corresponding to the removal of one proton from alanine. At low pH values, molecules of alanine bear a net charge of +1 because both functional groups are protonated; for example, at pH 0.35, 99% molecules are positively-charged and are present as $^+H_3N-CH(CH_3)-COOH$. At the midpoint in the first stage of titration, in which the $-COOH$ group of alanine loses its proton, equimolar concentrations of proton-donor, $^+H_3N-CH(CH_3)-COOH$ and proton-acceptor, $^+H_3N-CH(CH_3)-COO^-$ species are present. At the midpoint of a titration, the pH is equal to the pK_a of the protonated group being titrated. For alanine, the pH at the midpoint is 2.34, thus its $-COOH$ group has a pK_a of 2.34.

The pH and pK_a are simply convenient notations for proton concentration and equilibrium constant for ionization, respectively. The pK_a is a measure of the tendency of a group to give up a proton, with that tendency decreasing tenfold as the pK_a increases by one unit.

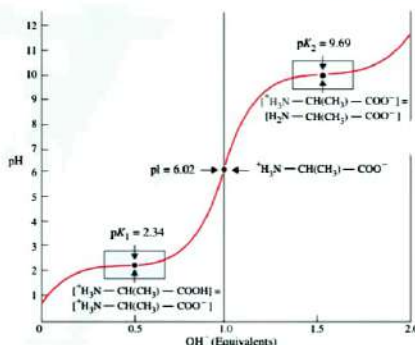


Fig. 9-5. The titration curve of alanine

The ionic species predominating at key points in the titration are shown above the graph. The shaped boxes, centred about $pK_1 = 2.34$ and $pK_2 = 9.69$, indicate the regions of maximum buffering power.

As the titration proceeds, another important point is reached at pH 6.02. Here there is a point of inflexion, at which removal of the first proton is essentially complete, and removal of the second has just begun. At this pH, the alanine is present as its dipolar form, fully ionized ($^+H_3N-CH(CH_3)-COO^-$) but with no net electric charge.

The second stage of titration corresponds to the removal of a proton from the $-NH_3^+$ group of alanine. The pH at the midpoint of this stage is 9.69 which is equal to the pK_a for the $-NH_3^+$ group. The titration is complete at a pH of 11.69, at which point the molecules of alanine bear a net charge of -1 because both functional groups are deprotonated. And 99% of the molecules are negatively charged and are present as $H_2N-CH(CH_3)-COO^-$. It may, however, be noted that the characteristic pH value of 6.02, at which the average net charge is zero (called the isoelectric point or isoelectric pH), is midway between the two pK_a values [(2.35 + 9.69)/2].

163 / 1254

GENERAL STRUCTURE OF PROTEINS 139

The study of titration curve provides three useful informations:

1. It gives a quantitative measure of the pK_a of each of the two ionizing groups, 2.34 for the $-COOH$ group and 9.69 for the $-NH_3^+$ group in case of alanine.
2. It gives information that alanine, say for example, has two regions of buffering power: one centred around the first pK_a of 2.34 and the second centred around pH 9.69.
3. It also provides information regarding the relationship between the net electric charge of the amino acid and the pH of the solution.

Classification

Three systems of classifying amino acids are in vogue.

A. On the basis of the composition of the side chain or R group. There are 20 different amino acids which regularly appear in proteins. These possess a side chain which is the only variable feature present in their molecules. The other features such as α -carbon, carboxyl group and amino group are common to all the amino acids. The common component of an amino acid appears in Fig. 9-6.

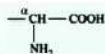


Fig. 9-6. Common component of an amino acid

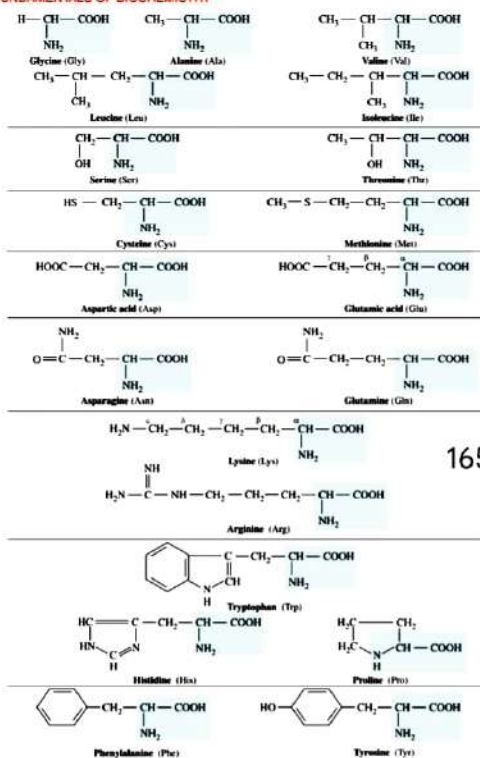
Based on the composition of the side chain, the twenty amino acids, whose structure is shown in Fig. 9-7, may be grouped into following 8 categories (Fairley and Kilgour, 1966):

- (i) **Simple amino acids.** These have no functional group in the side chain, e.g., glycine, alanine, valine, leucine and isoleucine.
- (ii) **Hydroxy amino acids.** These contain a hydroxyl group in their side chain, e.g., serine and threonine.
- (iii) **Sulfur-containing amino acids.** These possess a sulfur atom in the side chain, e.g., cysteine and methionine.
- (iv) **Acidic amino acids.** These have a carboxyl group in the side chain, e.g., aspartic acid and glutamic acid.
- (v) **Amino acid amides.** These are derivatives of acidic amino acids in which one of the carboxyl group has been transformed into an amide group ($-\text{CONH}_2$), e.g., asparagine and glutamine.
- (vi) **Basic amino acids.** These possess an amino group in the side chain, e.g., lysine and arginine.
- (vii) **Heterocyclic amino acids.** These amino acids have in their side chain a ring which possesses at least one atom other than the carbon, e.g., tryptophan, histidine and proline.
- (viii) **Aromatic amino acids.** These have a benzene ring in the side chain, e.g., phenylalanine and tyrosine.

The classification given above is only a practical one and can conveniently be followed. It does not, however, strictly delimit the various categories. For example, tryptophan may also be included under aromatic amino acids and similarly, histidine under basic amino acids.

B. On the basis of the number of amino and carboxylic groups. McGilvery and Goldstein (1979) have classified various amino acids as follows:

1. Monoamino-monocarboxylic amino acids:
 1. Unsubstituted — Glycine Alanine, Valine, Leucine, Isoleucine
 2. Heterocyclic — Proline
 3. Aromatic — Phenylalanine, Tyrosine, Tryptophan
 4. Thioether — Methionine



165 / 1254

Fig. 9-7. The twenty amino acids ("Magic 20") found in proteins
The standard three-letter abbreviation for each amino acid is written in bracket.
The common component is shown in a blue enclosure.

GENERAL STRUCTURE OF PROTEINS 141

5. Hydroxy — Serine, Threonine
6. Mercapto — Cysteine
7. Carboxamide — Asparagine, Glutamine

II. Monoamino-dicarboxylic amino acids : Aspartic acid, Glutamic acid

III. Diamino-monocarboxylic amino acids : Lysine, Arginine, Histidine

C. On the basis of polarity of the side chain or R group. A more meaningful classification of amino acids is, however, based on the polarity of the R groups present in their molecules, i.e., their tendency to interact with water at biological pH (near pH 7.0). The R groups of the amino acids vary widely with respect to their polarity from totally nonpolar or hydrophobic (water-hating) R groups to highly polar or hydrophilic (water-loving) R groups. This classification of amino acids emphasizes the possible functional roles which they perform in proteins. The system recognizes following 4 categories :

I. Amino acids with nonpolar R groups. The R groups in this category of amino acids are hydrocarbon in nature and thus hydrophobic. This group includes five amino acids with aliphatic R groups (alanine, valine, leucine isoleucine, proline), two with aromatic rings (phenylalanine, tryptophan) and one containing sulfur (methionine).

1. **Alanine** (α -aminopropionate). It was first isolated in 1888 from silk fibroin where it occurs in abundance, along with glycine and serine. It is the parent substance of all the amino acids except glycine. The various amino acids may be derived from alanine by replacement of one or two H atoms of the methyl group present on α -carbon atom. Alanine is the least hydrophobic of the 8 nonpolar (=hydrophobic) amino acids because of its small methyl side chain.

2. **Valine** (α -aminoisovalerate). It is widely distributed but rarely occurs in amounts exceeding 10%. It is a branched chain amino acid and can be derived from alanine by the introduction of two methyl groups in place of two H atoms of the methyl group present on α -carbon atom.

3. **Leucine** (α -aminoisocaproate). Its presence in proteins was shown, first of all the amino acids, by Prout in 1819. It was first isolated from cheese, but later was obtained in purer form from hydrolysates of wool. Leucine is generally more prevalent than valine or isoleucine in proteins. It is also a branched chain amino acid and is the next higher homologue of valine. As such, it has much in common with valine from the viewpoint of reactivity and function. It is also one of the few amino acids that are sparingly soluble in water.

4. **Isoleucine** (α -amino- β -methylvalerate). It is an isomer of leucine and is also a branched chain amino acid. It was discovered somewhat late (in 1904) by Paul Ehrlich (LT, 1854-1915) because it was difficult to separate from leucine owing to their similar chemical composition and properties. This amino acid is one of the 5 common ones that have more than one asymmetric carbon atoms. It has 2 asymmetric carbon atoms and thus occurs in 4 stereoisomeric forms, but the natural form obtained from protein hydrolysates is the L-erythro isomer, where erythro signifies that the orientation of the methyl and amino groups is similar to that found in erythrose, with the groups concerned on the same side of the carbon chain.

Valine, leucine and isoleucine are quite alike chemically and possess a branched carbon chain. As most animals cannot synthesize the branched carbon chain, these amino acids are, therefore, indispensable or essential in the diet.

5. **Proline** (2-pyrrolidinecarboxylate). It is present in almost all proteins that have been studied, but zein from corn and gelatin are relatively high in proline content. It is a cyclized derivative of glutamic acid. Its α -amino group is not free but is substituted by a portion of its R group to yield a cyclic structure. The secondary amino group in proline is held in a rigid conformation. This reduces the structural flexibility of the protein because of the maximal steric hindrance of its side chain. Unlike other amino acids, proline is very soluble in ethanol. Moreover, it does not give many of the characteristic amino acid tests which are generally based on the presence of the unsubstituted α -amino group.

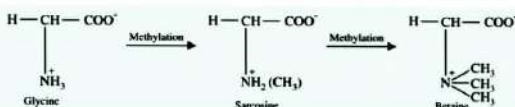
9. *Glycine or glycoall* (α -aminoacetate). It is the simplest amino acid and the only one lacking an asymmetric α -carbon atom. Where it is present in a protein, the minimal steric hindrance of the glycine side chain allows much more structural flexibility than the other amino acids. Glycine represents the opposite structural extreme in comparison to proline. It is, therefore, unique in being optically inactive. It is one of the first to be isolated from proteins and has a characteristic sweet taste, hence so named (*glykos*² = sweet). Glycine is present in abundance in scleroproteins. This occurs in especially large amounts in the protein gelatin (25% of the total amino acids) and in silk fibroin (40%). In addition, many nonprotein compounds also contain glycine as a component part, for instance glutathione (a common tripeptide), sarcosine (= N-methylglycine) and hippuric acid (= benzoylglycine). Methylation of glycine yields sarcosine which is found in many peptides. Further methylation produces betaine, a substance which exists exclusively in the zwitterionic or salt form (for the same reason, zwitterionic formulae are sometimes called as betaine structures).

Some authors use the term "im" for proline and hydroxyproline correct, because an imine is cl by a $>CH = NH$ and not a $C=O$.

167 / 1254

II. Amino acids with polar but uncharged R groups.

9. *Glycine or glycoall* (α -aminoacetate). It is the simplest amino acid and the only one lacking an asymmetric α -carbon atom. Where it is present in a protein, the minimal steric hindrance of the glycine side chain allows much more structural flexibility than the other amino acids. Glycine represents the opposite structural extreme in comparison to proline. It is, therefore, unique in being optically inactive. It is one of the first to be isolated from proteins and has a characteristic sweet taste, hence so named (*glykos*² = sweet). Glycine is present in abundance in scleroproteins. This occurs in especially large amounts in the protein gelatin (25% of the total amino acids) and in silk fibroin (40%). In addition, many nonprotein compounds also contain glycine as a component part, for instance glutathione (a common tripeptide), sarcosine (= N-methylglycine) and hippuric acid (= benzoylglycine). Methylation of glycine yields sarcosine which is found in many peptides. Further methylation produces betaine, a substance which exists exclusively in the zwitterionic or salt form (for the same reason, zwitterionic formulae are sometimes called as betaine structures).



III. Amino acids with negatively charged (= acidic) R groups. These are monoamino-dicarboxylic acids. In other words, their side chain contains an extra carboxyl group with a dissociable proton. The resulting additional negative charge accounts for the electrochemical behaviour of proteins. The two amino acids which belong to this category are aspartic and glutamic.

Cystine, from which the name 'cysteine' was later derived, was isolated in 1843 from urinary stones and was hence so named (cystos^G = bladder).

144 FUNDAMENTALS OF BIOCHEMISTRY

16. **Aspartic acid** (α -aminosuccinate). Its presence in protein was discovered by Ritthausen in 1868. It is the parent compound of asparagine.

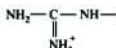
17. **Glutamic acid** (α -aminoglutarate). It is the homologue of aspartic acid. Its presence in protein was discovered also by Ritthausen in 1866. It was found in wheat gluten, hence so named. It is the parent compound of glutamine and occurs widely in proteins.

Both aspartic and glutamic acids in almost all proteins are linked to other amino acids only through their α -carboxyl and α -amino groups, not through their side chain carboxyl groups. Aspartic and glutamic acids are usually called aspartate and glutamate respectively to emphasize that their side chains are nearly always negatively charged at physiological pH.

IV. Amino acids with positively charged (=basic) R groups. These are diamino-monocarboxylic acids. In other words, their side chain contains an extra amino group which imparts basic properties to them. Lysine, arginine and histidine belong to this category.

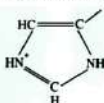
18. **Lysine** (α , ϵ -diaminopropate). It has a second amino group at ϵ -position. Lysine is generally abundant in animal proteins but present in limited amounts in plant proteins, such as those of corn and wheat. Ionic charges are provided by protonation of the amino group of the ϵ -carbon. It is nutritionally important since it is not synthesized by higher animals. Pulses contain good amount of lysine whereas cereals lack in it.

19. **Arginine** (α -amino- δ -guanidinovaleate). It is abundant in highly basic proteins of the cell nucleus (histones) and in sperm proteins, such as the protamines from salmon and herring sperm. In the latter, arginine may be as much as 80% of the total amino acids. It is found generally in all proteins, although in lesser amounts. Arginine is unique in possessing the guanidinium group, and due to which it is more strongly basic than lysine. Protonation of the guanidinium group imparts ionic charges.



Guanidinium group (charged form) of arginine

20. **Histidine** (α -amino- β -imidazolopropionate). **Histidine is the last entry in the list of "magic 20"**. It contains a weakly basic imidazolium R group with pK' value of 6.0, and, therefore, is less than 10% protonated at pH 7. In many enzyme proteins, it functions as a proton donor or acceptor. Histidine is the only amino acid which has a proton that dissociates in the neutral pH range. It is this property which allows certain histidine residues to play an important role in the catalytic activities of



Imidazolium group (charged form) of histidine

some enzymes. Histidine occurs in limited quantities in most proteins, but hemoglobin, protamines and histones contain relatively large amounts. It is also a component in the simple peptides carnosine and anserine. Histidine's basic properties are clearly marginal. The two ring nitrogen atoms have a relatively weak affinity for an H^+ and are only partly positive at neutral pH.

Carnosine ($\text{C}_4\text{H}_8\text{O}_2\text{N}_2$) is an optically active crystalline dipeptide; m.p. 260°C ; found in muscle tissue.

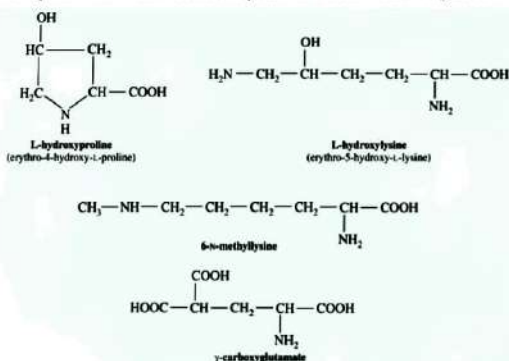
NONSTANDARD PROTEIN AMINO ACIDS

In addition to the above-mentioned twenty standard amino acids which are building blocks of

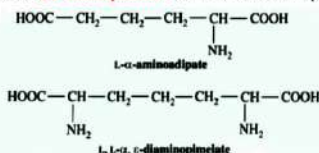
169 / 1254

GENERAL STRUCTURE OF PROTEINS 145

proteins and have a wide range of distribution in proteins, several other amino acids exist. These have a limited distribution but may be present in high amounts in a few proteins and hence deserve mention. As an example, **hydroxyproline** has a limited distribution in nature but constitutes as much as 12% of the composition of collagen, an important structural protein of animals. Similarly, **hydroxylysine** is also a component of collagen, where it accounts for about 1% of the total amino acids. **N-methyllysine** is found in myosin, a contractile protein of muscle. Another important nonstandard or less common amino acid is **γ -carboxyglutamate**, which is found in the blood-clotting protein, prothrombin as well as in certain other proteins that bind Ca^{2+} in their biological function.



In protein from corn, **α -aminoadipate** has been detected. It has one CH_2 group more than glutamic acid. Another amino acid, **α - ϵ -diaminopimelate** has been found in bacterial protein.



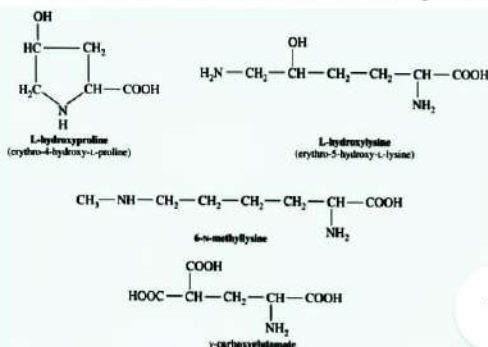
NONPROTEIN AMINO ACIDS

There are some 300 additional amino acids which are never found as constituents of proteins but which either play metabolic roles or occur as natural products.

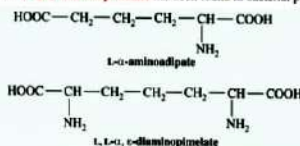
Among the important nonprotein amino acids, which play metabolic roles, are L-ornithine, L-citrulline, β -alanine, creatine and γ -aminobutyrate. **L-ornithine** and **L-citrulline** occur in free state in the animal tissues and are metabolic intermediates in the urea cycle. L-ornithine possesses one CH_2 group less than its homologue, lysine. **β -alanine**, an isomer of alanine, occurs free in nature and also

GENERAL STRUCTURE OF PROTEINS 145

proteins and have a wide range of distribution in proteins, several other amino acids exists. These have a limited distribution but may be present in high amounts in a few proteins and hence deserve mention. As an example, **hydroxyproline** has a limited distribution in nature but constitutes as much as 12% of the composition of collagen, an important structural protein of animals. Similarly, **hydroxylysine** is also a component of collagen, where it accounts for about 1% of the total amino acids. **N-methyllysine** is found in myosin, a contractile protein of muscle. Another important nonstandard or less common amino acid is **γ-carboxyglutamate**, which is found in the blood-clotting protein, prothrombin as well as in certain other proteins that bind Ca^{2+} in their biological function.



In protein from corn, **α-aminoadipate** has been detected. It has one CH_2 group more than glutamic acid. Another amino acid, **α-ε-diaminopimelate** has been found in bacterial protein.



NONPROTEIN AMINO ACIDS

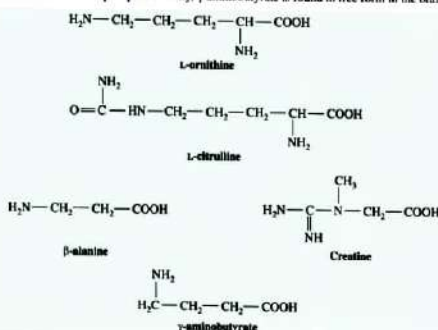
There are some 300 additional amino acids which are never found as constituents of proteins but which either play metabolic roles or occur as natural products.

Among the important nonprotein amino acids, which play metabolic roles, are L-ornithine, L-citrulline, β-alanine, creatine and γ-aminobutyrate. **L-ornithine** and **L-citrulline** occur in free state in the animal tissues and are metabolic intermediates in the urea cycle. L-ornithine possesses one CH_2 group less than its homologue, lysine. **β-alanine**, an isomer of alanine, occurs free in nature and also

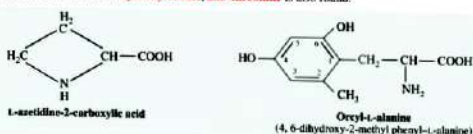
170 / 1254

146 FUNDAMENTALS OF BIOCHEMISTRY

as a constituent of an important vitamin pantothenic acid and of coenzyme A. It is also found in the naturally occurring peptides, carnosine and anserine. The quaternary amine **creatine**, a derivative of glycine, plays an important role in the energy storage process in vertebrates where it is phosphorylated and converted to creatine phosphate. Lastly, γ-aminobutyrate is found in free form in the brain.



Higher plants are especially rich in nonprotein amino acids. These nonprotein amino acids are usually related to the protein amino acids as homologues or substituted derivatives. They have a limited distribution, sometimes to a single species even. Thus, **L-azetidine-2-carboxylic acid**, a homologue of proline, accounts for 50% of the nitrogen present in the rhizome of Solomon's seal, *Polygonatum multiflorum*. **Orcylalanine** is found in the seed of corncockle, *Agrostemma githago*. It may be considered as a substituted phenylalanine. Furthermore, in the toxic polypeptides of *Amanita phalloides*, in addition to **hydroxy-leucine**, **allo-threonine** is also found.



PEPTIDES

Peptide Bond

The amino acid units are linked together through the carboxyl and amino groups to produce the primary structure of the protein chain. The bond between two adjacent amino acids is a special type of amide bond, in which the hydrogen atom of amino ($-\text{NH}_2$) group is replaced by an R radical. Such a substituted amide bond is known as the **peptide bond**. And the chain, thus formed, by linking



together of many amino acid units is called a **peptide chain** (Fig. 9-8).

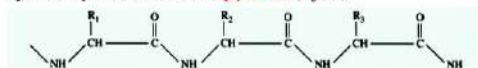


Fig. 9-8. Peptide chain

The characteristic structure of the peptide bond is shown in Fig. 9-9.

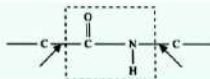


Fig. 9-9. Peptide bond

The peptide bond is shown enclosed in the dashed box. The four atoms (C, O, N, H) of the peptide bond form a rigid planar unit. There is no freedom of rotation about the C—N bond. On the contrary, the 2 single bonds (shown with arrows) on either side of the rigid peptide unit, exhibit a high degree of rotational freedom.

Each peptide chain is of considerable length and may possess from 50 to millions of amino acid units. Depending on the number of amino acid molecules composing a chain, the peptides are termed as a **dipeptide** (containing 2 amino acid units), a **tripeptide** (containing 3 amino acid units), and so on. If a peptide is made up of not more than 10 amino acids, it is called an **oligopeptide**; beyond that it is a **polypeptide**. Polypeptides when they are made up of over 100 amino acids are, sometimes, called as **macropeptides**.

Strictly speaking, the proteins are polypeptides with more than 100 amino acids. All naturally-occurring important peptides, however, possess a shorter individual name, such as glutathione etc. Proteins differ widely in amino acid content. Various types of proteins in an organism may have varied amounts of a particular amino acid. Some amino acids are in abundance in one protein, may be in meagre amounts in others and may even be lacking in the rest. Tryptophan, for instance, lacks in certain proteins. However, *most of the proteins contain all the 20 amino acids.* As the number and manner in which the amino acids are grouped is highly variable, the number of proteins approaches almost to infinity.

It is analogous to the indefinite number of words that can be formed with the 26 letters of English alphabet. While the words have to be restricted in length, there is no such restriction regarding the number of amino acids that may form a protein. According to an estimate given by Erlene B. Cunningham (1978), if each protein molecule were to consist of only 250 amino acid residues, the utilization of all the 20 different monomers would permit the formation of 10325 different protein molecules! In reality, protein molecules often contain more than 250 amino acid units, and hence there is possibility for an even greater number of different protein molecules.

N- and C-terminals

Each amino acid in the chain is termed a **residue**. The two ends of the peptide chain are named as amino terminal and carboxyl terminal or simply as an N-terminal and C-terminal respectively. These two terminal groups, one basic and another acidic, are the only ionizable groups of any peptide chain

Sometimes, the word '**proteins**' is used for short polypeptides containing up to 18 amino acids (Fried GH, 1990).

172 / 1254

148 FUNDAMENTALS OF BIOCHEMISTRY

except those present in the side chain. The terminal amino acid with the free amino group is called as the **N-terminal amino acid** and the one with the free carboxyl group at the other end as **C-terminal amino acid**.

Representation of Peptide Chain

To fix the convention for representation of peptide structures in mind, it is helpful to imagine a rattlesnake moving from left to right across the page. The C-terminal residue forms its fangs and the N-terminal residue its rattle (Fig. 9-10).



Fig. 9-10. Mnemonic device for peptide chain

Naming of Peptide Chain

In naming a polypeptide, the convention is that the N-terminal residue (which is shown at the left hand part of the structure) is written first and the C-terminal residue in the formation of each peptide

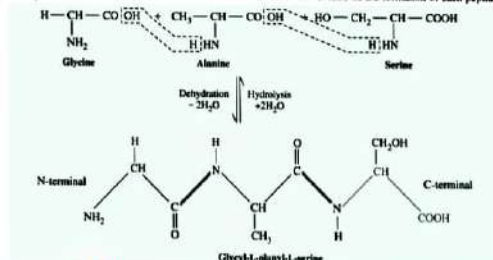


Fig. 9-11. Construction of a tripeptide chain from three different amino acids

The formation of each peptide bond (indicated by bold lines) involved splitting of a molecule of water.

Thus, the process of synthesis of proteins may be deemed as essentially a dehydration synthesis. The names of various intermediary amino acid residues are written in the same sequence as they are placed. Further, the names of all the amino acid residues, except the last one, are written by adding the suffix *-yl* because all these are the acyl groups. The name of the last amino acid, however, is written as such. For example, a tripeptide containing glycine, alanine and serine (structure shown in Fig. 9-11) is named as glycyl-L-alanyl-L-serine and abbreviated as Gly-Ala-Ser. If the sequence of amino acid in such a tripeptide is not known, the abbreviation would be (Gly, Ala, Ser), the parenthesis and commas indicating that only the composition of the tripeptide is known.

Similarly, glutathione or GSH (a tripeptide containing glutamic acid, cysteine and glycine) is named as γ-glutamyl-cysteyl-glycine and abbreviated as Glu-Cys-Gly. Glutathione is a naturally occurring and widely distributed polypeptide.

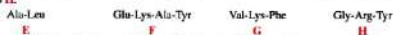
Determination of the Amino Acid Sequence of a Polypeptide

This can be explained by taking the example of a **dodecapeptide** whose composition was found to be Ala₂, Arg, Gly, Glu, Leu, Lys₂, Phe, Tyr₂, Val. It was determined that the N-terminal amino acid

of the dodecapeptide was valine and the C-terminal amino acid, leucine. Hydrolysis of the dodecapeptide by trypsin yielded four peptides whose structures were determined and found to be those given in A to D.



Since valine was the N-terminus and leucine the C-terminus of the dodecapeptide, it is apparent that peptide C must represent the amino acid sequence at the N-terminal end, and peptide D the amino acid sequence at the C-terminal end of the dodecapeptide. To establish the order of the A and B peptides in the interior of the dodecapeptide, another sample of the dodecapeptide was hydrolyzed by *chymotrypsin*, the four peptides formed were sequenced, and their structures were found to be those given in E to H.



The sequences Gly-Arg-Tyr in peptide H and Glu-Lys-Ala-Tyr in peptide F clearly establish that peptide B must precede peptide A in the dodecapeptide. Hence, the structure of the dodecapeptide is unambiguously determined to be that as shown below:



Stereochemistry of Peptide Chains

All proteins are made of amino acids of L-configuration. This fixes the steric arrangement at the α -carbon atom. The dimensions of the peptide chain are known exactly. These have been depicted in Fig. 9-12.

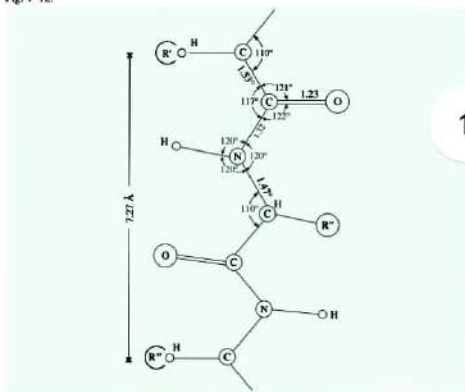


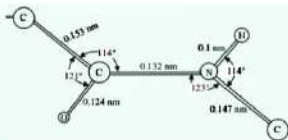
Fig. 9-12. Dimensions within the peptide chain

174 / 1254

Contents

150 FUNDAMENTALS OF BIOCHEMISTRY

The peptide bond, which is an imide (substituted amide) bond, has a planar structure. The 6 atoms within the plane are related to each other by bond lengths and angles that vary little from amino acids residue to amino acid residue. Only 3 of these bonds are part of the peptide chain *per se*: the α -carbon to carbonyl carbon bond, the C-N bond, and the imide nitrogen to α -carbon bond. Since the double bond character of the C-N bond limits rotation about it, only the first and the last allow rotation. The rotation angles ϕ and ψ establish the relative positions of any 2 successive amide planes along the polypeptide chain. The α -carbon atoms can be thought of as shivel centres for the adjacent amide planes.



Biological Roles

Peptides participate in a number of biological activities.

1. They serve as intermediates in the formation of **proteins**.
2. They appear as constituents in a group of compounds called **alkaloids**. Majority of these have been isolated from fungi, although they are also found in higher plants. Ergotamine is a peptide alkaloid from rye ergot and has pronounced pharmacological properties. The four components of this alkaloid are lysergic acid, alanine, proline and phenylalanine.
3. Many of them possess **antibacterial activities** and are usually present in fungi and bacteria. Penicillin G with 3 components (valine, cysteine and phenylacetic acid) is a common antibiotics.
4. Certain other peptides serve as **growth factors**. Folic acid, a water-soluble vitamin, is a noteworthy example of it (see Chapter 34). Another group of peptides serving as growth factor for a variety of microorganisms is streptogenins.
5. Higher animals do synthesize certain peptides serving as **hormones** (see Chapter 31).
6. Certain peptides like glutathione participate in controlling the **oxidation-reduction potential** of the cell. This may also serve as a key intermediate in electron-transfer systems.
7. A direct correlation has been found to exist between the amount of peptides in the urine of the patients and their **mental state of disturbance**. A group of Norwegian doctors have found an excess of peptides in urine specimens from patients with psychiatric disturbances. The peptides have been shown to induce in animals some of the conditions for the development of psychiatric disorders which lead to mania, depression or schizophrenia. The urine tests can, thus, indicate if a person is suffering from mental illness. The cause of this hypersecretion of peptides is not yet well established. It may exist from birth in organic genetic derangements or may be induced in healthy people by environmental factors.

CHEMICAL BONDS INVOLVED IN PROTEIN STRUCTURE

Given a full assortment of amino acids, a cell can synthesize all of its protein components. *Protein synthesis is a multiple dehydration process* (refer Fig. 9-11). The net structure of a protein becomes possible as a result of linking together of various amino acid units. The union of these amino acids to each other forming a chain and also among various amino acid residues of different chains involves various types of chemical bonds (Fig. 9-13). These are described below.



Contents

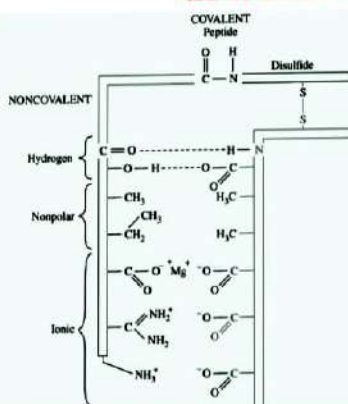


Fig. 9-13. Types of chemical bonds involved in protein structure

(After Hartman and Siskin)

176 / 1254

A. Primary Bond

The principal linkage found in all proteins is the covalent **peptide bond**, $-\text{CO}-\text{NH}-$ (Fischer, 1906). It is a specialized amide linkage where C atom of $-\text{COOH}$ group of one amino acid is linked with the N atom of $-\text{NH}_2$ group of the adjacent amino acid. **Peptide bond is, in fact, the backbone of the protein chain.**

B. Secondary Bonds

Many of the properties of proteins, however, do not coincide with the linear chain structure, thus indicating that a variety of bonds other than the peptide exist in them. These secondary bonds, as they are called, hold the chain in its natural configuration. Some of the secondary bonds commonly found in proteins are listed below:

1. **Disulfide Bond ($-\text{S}-\text{S}-$).** In addition to the peptide bond, a second type of covalent bond found between amino acid residues in proteins and polypeptides is the disulfide bond, which is formed by the oxidation of the thiol or sulfhydryl ($-\text{SH}$) groups of two cysteine residues to yield a mole of **cystine**, an amino acid with a disulfide bridge (Fig. 9-15). In generalized form, the above reaction may be written as:



A disulfide bond is characterized by a bond strength of approximately 50 kilocalories per mole and a bond length of about 2Å between the two sulfur atoms. Hence, disulfide bond formation between 2 cysteine residues located some distance apart in the polypeptide chain requires that the polypeptide chain be folded back on itself to bring the sulfur groups close together. Although disulfide bridges are very strong, when compared to the strength of noncovalent bonds they are very short-range since, as

152 FUNDAMENTALS OF BIOCHEMISTRY

with all covalent bonds, even a slight extension breaks them completely. They, therefore, only stabilize the tertiary structure when it has reached something approximating to its final form.

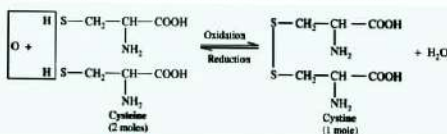


Fig. 9-14. Formation of cystine (= dicysteine)

Oxytocin (Fig. 9-15), a hormone stimulating the contraction of smooth muscles especially during childbirth, is an example where an internal disulfide bond is present between two cysteine units separated from each other in the peptide chain by 4 other amino acid units.

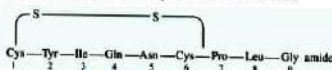
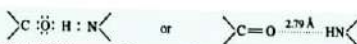


Fig. 9-15. Structure of oxytocin

Insulin (Fig. 9-16) is another excellent example where two peptide chains are linked together by 2 disulfide bonds. The presence of an internal disulfide bond in the glycyl (or A) chain between residues 6 and 11 is noteworthy.

These two chemical bonds, namely, peptide and disulfide, are relatively stable. Both these bonds, collectively or individually, maintain the linear form (or the *primary structure*) of the protein molecule.

2. **Hydrogen Bond ($>\text{CO}\cdots\text{HN}<$).** When a group containing a hydrogen atom, that is covalently-bonded to an electronegative atom, such as oxygen or nitrogen, is in the vicinity of a second group containing an electronegative atom, an energetically favourable interaction occurs which is referred to as a hydrogen bond. The formation of a hydrogen bond is due to the tendency of hydrogen atom to share electrons with two neighbouring atoms, *esp.*, O and N. For example, the carbonyl oxygen of one peptide bond shares its electrons with the hydrogen atom of another peptide bond. Thus,



An interaction sets in between a $\text{C}=\text{O}$ group and the proton of an NH or OH group if these groups come within a distance of about 2.8 Å. This secondary valence bond is symbolized by a dotted line, $\cdots$ (The solid line, however, represents the normal covalent bond). The strength of the hydrogen bond is only 5 to 8 kilocalories per mole and is maximal when the bond is linear (Fig. 9-17). Hydrogen bonding between amides or peptides of the type depicted in Fig. 9-17 (c) plays an important role in stabilizing some conformations of the polypeptide chain.

Silk fibroin, composed mainly of glycine, alanine and serine units, is an example of the presence of hydrogen bonds involving the imide ($>\text{NH}$) and carbonyl ($>\text{C}=\text{O}$) groups of the peptide bonds. Here the hydrogen bonds link the vicinal peptide chains (refer Fig. 9-26).

In other proteins like **keratin of wool**, however, the hydrogen bonds link the side chains so that a single peptide chain is held in a coiled or helical form (refer Fig. 9-24).

Since the binding energy of a hydrogen bond amounts to only 1/10th of that of a primary valence, the hydrogen bonds are relatively weak linkages but many such bonds collectively exert considerable

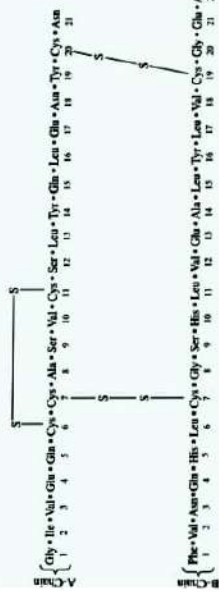


Fig. 9-16. Structure of bovine insulin

Numbers refer to specific amino acid residues. Glycine and phenylalanine are the N-terminal amino acids and the asparagine and alanine, the C-terminal amino acids. The A chain of the insulin of man, cow, dog, rabbit and sperm whale are identical. The B chains of the cow, pig, dog, goat and horse are identical. The amino acid replacements in the A-chain in positions 8, 9 and 10 (for details, refer Chapter 10: Animal Hormones)

178 / 1254

Contents

154 FUNDAMENTALS OF BIOCHEMISTRY

force and help in maintaining the helical structure (or the *secondary structure*), characteristic of many proteins (Mirsky and Pauling, 1936).

3. Nonpolar or Hydrophobic Bond. Many amino acids (like alanine, valine, leucine, isoleucine, methionine, tryptophan, phenylalanine and tyrosine) have the side chains or R groups which are essentially hydrophobic, *i.e.*, they have little attraction for water molecules in comparison to the strong hydrogen bonding between water molecules. Such R groups can unite among themselves with elimination of water to form linkages between various segments of a chain or between different chains. This is very much like the coalescence of oil droplets suspended in water.

The association of various R groups in this manner leads to a relatively strong bonding. It also serves to bring together groups that can form hydrogen bonds or ionic bonds in the absence of water. Each type linkage, thus, helps in the formation of the other; the hydrophobic bonds being most efficient in this aspect. The hydrophobic bonds also play important role in other protein interactions, for example, the formation of enzyme-substrate complexes and antibody-antigen interactions.

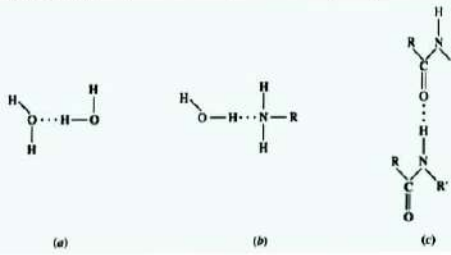


Fig. 9-17. Some common examples of hydrogen bonding

The hydrogen bonds (represented by a dotted line) have been shown between (a) two water molecules, (b) water and an amine, and (c) two amide groups.

4. Ionic or Electrostatic Bond or Salt linkage or Salt bridge. Ions possessing similar charge repel each other whereas the ions having dissimilar charge attract each other. For example, divalent cations like magnesium may form electrostatic bonds with 2 acidic side chains.

Another instance of ionic bonding may be the interaction between the acidic and basic groups of the constituent amino acids shown at the bottom of Fig. 9-13. The R groups of glutamic acid and aspartic acid contain negatively charged carboxylate groups, and the basic amino acids (arginine, histidine, lysine) contain positively charged amino groups in the physiological pH range. Thus, these amino acids contribute negatively charged and positively charged side chains to the polypeptide backbone. When two oppositely charged groups are brought close together, electrostatic interactions lead to a strong attraction, resulting in the formation of an electrostatic bond. In a long polypeptide chain containing a large number of charged side chains, there are many opportunities for electrostatic interaction. Intramolecular ionic bonds are rather infrequently used in the stabilization of protein structure but when they are so used, it is often with great effect. In fact, ionized groups are more frequently found stabilizing interactions between protein and other molecules. Thus, ionic bonds between positively charged groups (side chains of lysine, arginine and histidine) and negatively charged groups (COO^- group of side chain of aspartic and glutamic acids) do occur.

These ionic bonds, although weaker than the hydrogen bonds, are regarded as responsible for maintaining the folded structure (or the *tertiary structure*) of the globular proteins.

CHARACTERISTICS OF CHEMICAL BONDS

Table 9-4 lists some characteristics of the 2 types of chemical bonds : covalent and non covalent. The strength of a bond or the **bond strength** can be measured by the energy required to break it, here in the table given in kilocalories per mole (kcal/mole). One kilocalorie is the quantity of energy needed to raise the temperature of 1,000 g of water by 1°C. An alternative unit in wide use is the kilojoule, kJ which is equal to 0.24 kcal. Individual bonds vary a great deal in strength, depending on the atoms involved and their precise environment, so that *the values are only a rough guide*. The **bond length** is the centre-to-centre distance between the two interacting atoms.

Table 9-4. Characteristics of chemical bonds

Bond Type	Length (nm)*	Strength (kcal/mole)†	
		In vacuum	In water
Covalent	0.15	90	90
Noncovalent			
Ionic	0.25	80	3
Hydrogen	0.30	4	1
van der Waal's attraction (per atom)	0.35	0.1	0.1

* The length given here for a hydrogen bond is that between its two nonhydrogen atoms.
† Note that the aqueous environment in a cell will greatly weaken both the ionic and the hydrogen bonds between nonwater molecules.

180 / 1254

PROTEIN CONFIGURATION

To describe a complicated macromolecule like protein, the biochemists have, for convenience, recognized 4 **basic structural levels of organization** of proteins based on the degree of complexity of their molecule (Fig. 9-18). These structural levels were first defined by Linderström-Lang and are often referred to as primary, secondary, tertiary and quaternary. Three of these structural levels (primary, secondary and tertiary) can exist in molecules composed of a single polypeptide chain, whereas the fourth (*i.e.*, quaternary) involves interactions of polypeptides within a multichained protein molecule. In mathematical term, these are also depicted as 1°, 2°, 3° and 4° respectively. The basic *primary structure* of a protein is relatively simple and consists of one or more linear chains of a number of amino acid units. This linear, unfolded structure or the polypeptide chain often assumes a helical shape to produce the *secondary structure*. This, in turn, may fold in certain specific patterns to produce the twisted three-dimensional or the *tertiary structure* of the protein molecule. Finally, certain other proteins are made up of subunits of similar or

IRVING GEIS
(LT, 1908 - 1997)



Irving Geis is well known for his lucid visualizations of molecular structures, particularly proteins and nucleic acids. These have appeared in *Scientific American* for the past thirty five years and in major chemistry, biology, and biochemistry textbooks. He is a co-author with R.E. Dickerson, Director of the Molecular Biology Institute of UCLA, of 3 books 1. Chemistry, Matter and the Universe 2. The Structure and Action of Proteins 10 Hemoglobin: Structure, Function, Evolution and Pathology.
In addition to drawing, painting, and writing, Irving Geis is a frequent lecturer at universities and medical schools on protein structure and function.
A recent Guggenheim fellowship made possible the assembly and cataloging of his drawings and paintings into The Geis Archives of molecular structure.

156 FUNDAMENTALS OF BIOCHEMISTRY

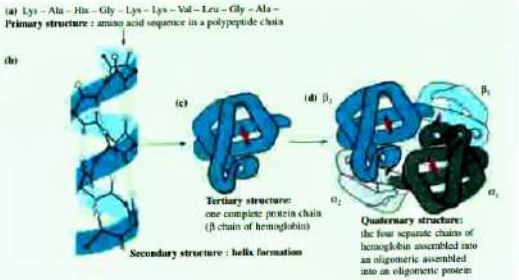


Fig. 9-18. The structural hierarchy in proteins

dissimilar types of the polypeptide chains. These subunits interact with each other in a specific manner to give rise to the so-called *quaternary structure* of the protein. This, in fact, defines the degree of polymerization of a protein unit.

1. Primary Structure : Amino Acid Sequence

The *primary structure* of a protein refers to the number and sequence of amino acids, the constituent units of the polypeptide chain. The main mode of linkage of the amino acids in proteins is the peptide bond which links the α -carboxyl group of one amino acid residue to the α -amino group of the other. The proteins may consist either of one or of more peptide chains.

Rigid and Planar Peptide Bond

Linus Pauling and Robert Corey, in the late 1930s, demonstrated that the α -carbons of adjacent

LINUS CARL PAULING (LT, 1901-1994)

Linus Pauling (with α -his helix ball-and-stick model in the photograph), the son of a German father and English mother, was an American physical chemist. Although, he had initially an undistinguished school education, Pauling obtained his doctorate degree in 1925 from California Institute of Tech (Caltech). He was one of the exceptional men of his times and earned the message of Quantum Mechanics to the New World. His pioneering application of principles of Quantum Physics to explain chemical properties, his investigations of molecular structures of proteins and, above all, his work on the nature of chemical bond rank among the outstanding pieces of chemical research of the century. To the graduate students of Chemistry in the 1940s and 50s, his book, *'The Nature of Chemical Bond'* (1939) became a Bible. At age 30, Pauling had become a full professor in 1931 and Chairman of Chemistry and Chemical Engineering some 6 years later. With Robert Corey (LT, 1867-1971), he investigated amino acids and polypeptides. Around 1940s, he and his colleagues conducted researches on antibody-antigen reactions. Pauling was the first one to devise a scale for comparing the electronegativity of different elements.



Pauling had his fair share of awards, which included, besides **Nobel Chemistry Prize (1954)**, the prestigious Davy Medal of the Royal Society (1947), the Willard Gibbs Medal (1946) and the Presidential

Award for Merit for his distinguished work during World War II in the Explosives Division of the Natural Research Commission (USA).

Soon after war, however, Pauling spoke bitterly against the nuclear arms race and advocated multilateral disarmament and an end to atomic testing. Following the publication of his famous book, 'No More War' (1958), he then sent, in Jan '58, a petition signed by 11,021 scientists to the United Nations, urging an end to the testing of nuclear weapons. For all this, he earned the **1962-Nobel Peace Prize**, which was presented to him on October 10, 1961, the day a US-Soviet partial nuclear test ban treaty came into force. He also won the International Lenin Peace Prize. His political activities took a heavy toll of his time and energy. The result was that he did not produce scientific work of significance after 1951. Nevertheless, during his career in science spanning more than 60 years, Pauling published several books and more than 1,000 scientific papers. He also established **Linus Pauling Institute of Science and Medicine** at Palo Alto.

However, Pauling's fame had been tarnished in his later years by his strong advocacy of vitamin C (ascorbic acid) as a 'wonder drug' and helpful in maintaining youthfulness for a longer period. He also published a book, 'Vitamin C and the Common Cold' (1971) wherein he claimed that by ingesting 1,000 mg daily of vitamin C (instead of the 60 mg daily minimum as recommended by the U.S. National Research Council), a person would catch 45% fewer colds and suffer 60% fewer days of illness. His findings were bitterly criticized experimentally by Terence Anderson of the University of Toronto in Canada. But this should not minimize the magnitude of his earlier contributions. Till today, he holds the distinction of being **the only person to have won two Nobel Prizes on his own**. His life fell into 3 distinct phases: the first evoked *reverences*, the second *love* and the third *ridicule*. Truly, Pauling ranks among the most versatile scientists of more than one generation of 20th century.

amino acids are separated by three covalent bonds, arranged $C_\alpha-C-N-C_\alpha$. They also demonstrated that the amide $C-N$ bond in a peptide is somewhat shorter (1.32— or 0.132 nm) than the $C-N$ bond in a simple amine (1.49 Å or 0.149 nm) and that the atoms associated with the bond are coplanar. This indicated a **resonance** or partial sharing of two pairs of electrons between the carbonyl oxygen and the amide nitrogen [Fig. 9-19 (a)]. The oxygen has a partial negative charge and the nitrogen partial positive charge, setting up a small electric dipole. The 4 atoms of the peptide group ($C=O$ and $N-H$) lie in a single plane, in such a way that the oxygen atom of the carbonyl group and the hydrogen of the amide nitrogen are *trans* to each other. **Virtually, all peptide bonds in proteins occur in the *trans* configuration.** From these studies, Pauling and Corey concluded that the amide $C-N$ bonds are to rotate freely because of their **partial double-bond character**. The backbone of a polypeptide chain can thus be separated by substituted methylene groups $-CH(R)-$ [Fig. 9-19 (c)]. The rigid peptide bonds limit the number of conformations that can be assumed by a polypeptide chain.

182 / 1254

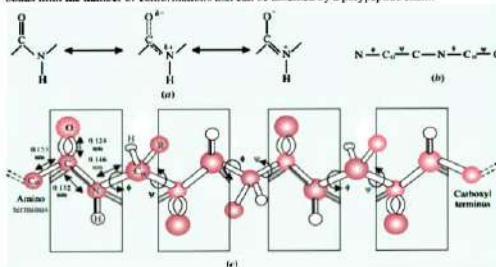


Fig. 9-19. The details of the planar peptide bond

- The planar peptide group. Note that the oxygen and hydrogen atoms are on opposite sides of the $C-N$ bond. This is *trans* configuration.
- The three bonds between the sequential C_α carbons in a polypeptide chain. The $N-C_\alpha$ and $C_\alpha-C$ bonds can rotate, with bond angles designated ϕ and ψ respectively.
- Limited rotation around two ($N-C_\alpha$ and $C_\alpha-C$) of the three types of bond in a polypeptide chain. The third type *i.e.*, $C-N$ bonds in the planar polypeptide groups, which make one-third of all the backbone bonds, are not free to rotate.
- The two coplanar peptide bonds flanking an α carbon. By convention, ϕ and ψ are both defined as 0° when the two peptide bonds flanking an α carbon are in the same plane. In a fully stretched out polypeptide chain, $\phi = \psi = 180^\circ$.

However, rotation is permitted about the bond between the nitrogen and α -carbon atoms of the main chain ($N-C_\alpha$) and between the α carbon and carbonyl carbon atoms ($C_\alpha-C$). By convention, the degree of rotation at the $N-C_\alpha$ bond is called *phi* (ϕ) and that between $C_\alpha-C$ bond is called *psi* (ψ). Again, by convention, both ϕ and ψ are defined as 0° in the conformation in which the two peptide bonds connected to a single carbon are in the same plane [Fig. 9-19 (d)]. In principle, ϕ and ψ can have any value between -180° and $+180^\circ$, but many values of ϕ and ψ are prohibited by steric interference between atoms in the polypeptide backbone and the amino acid side chains. The conformation in which ϕ and ψ are both 0° is prohibited for this reason.

Ramachandran Plot

The conformation of the main polypeptide chain can be completely determined if the values ϕ and ψ for each amino acid residue in the chain are known. In a fully stretched polypeptide chain, $\phi = \psi = 180^\circ$. G.N. Ramachandran (1963) recognized that an amino acid residue in a polypeptide chain cannot have just any pair of values of ϕ and ψ . By assuming that atoms behave as hard spheres, allowed ranges of ϕ and ψ can be predicted and visualized in a steric contour diagram called **Ramachandran plots**. Such a plot for poly-L-alanine (or any amino acid except glycine and proline) shows three separate allowed regions (the screened regions in Fig. 9-19). One of them contains $\phi-\psi$ values that generate the antiparallel β sheet, the parallel β sheet and the collagen helix. A second region has $\phi-\psi$ values that produce the right-handed α helix: a third, the left-handed α helix. Though sterically allowed, left-handed α helices are not found in proteins because they are energetically much less favoured.

Gopalasamudram Narayanan Ramachandran (1922–) is an Indian biophysicist and crystallographer who, along with Copinath Kartha, worked out the triple helical structure of collagen.

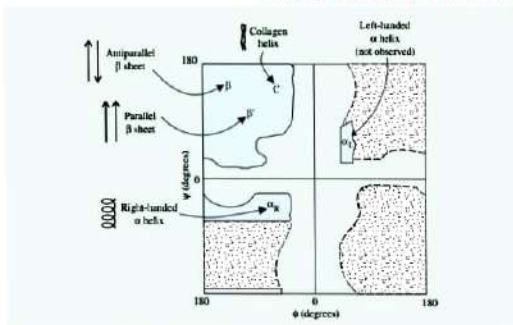
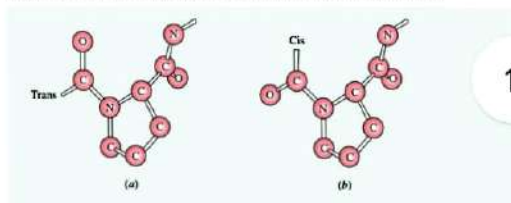


Fig. 9-20. A Ramachandran plot

The screened regions show allowed values of ϕ and ψ for L-alanine residues. Additional conformations are accessible to glycine (dotted regions) because it has a very small side chain.

(After Lubert Stryer, 1995)

For **glycine**, these three allowed regions are larger, and a fourth appears (shown as dotted in Fig. 9-19) because a hydrogen atom causes less steric hindrance than a methyl group. *Glycine enables the polypeptide backbone to make turns that would not be possible with another residue.*

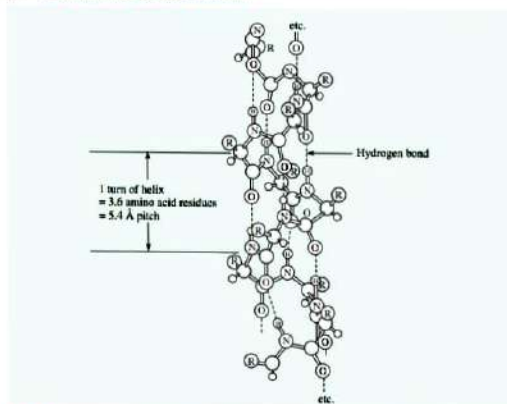
Fig. 9-20(a) Formation of *trans* (A) and *cis* (B) peptide bonds between proline and its preceding residue in the polypeptide chain

The presence of a 5-membered ring in proline, which locks ϕ (the angle between N and C_{α}) at about -65° , is its another distinctive feature.

Proline, too, is special [Fig. 9-20(a)]. The 5-membered ring of proline prevents rotation about the $N-C_{\alpha}$ bond, which fixes ϕ at about -65° . Hence, a proline residue has a markedly restricted range of allowed conformations. The residue on the N-terminal side of a proline is also constrained because

184 / 1254

160 FUNDAMENTALS OF BIOCHEMISTRY

Fig. 9-21. Portion of a right-handed α helix showing its three-dimensional (or '3-D') structure

All the shaded balls represent carbon atoms or residues (R) of amino acids.

(Redrawn from Linus Pauling and Robert Corey, 1955)

of steric hindrance imposed by the 5-membered ring. Proline also disfavors α -helix formation because it lacks an amide H atom for hydrogen bonding.

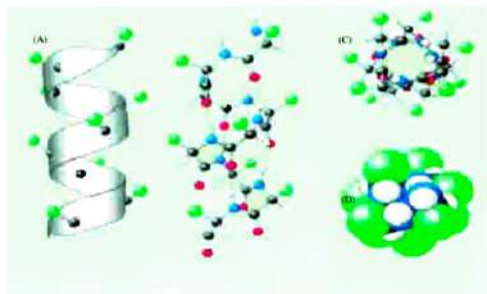
2. Secondary Structure : Helix Formation or Local Folding

If the peptide bonds were the only type of linkage present in proteins, these molecules would have behaved as irregularly coiled peptide chains of considerable length. But the globular proteins, however, do show some regular characteristic properties, indicating the presence of a regular coiled structure in these molecules. This involves the folding of the chain which is mainly due to the presence of hydrogen bonds. Thus, folding and hydrogen bonding between neighbouring amino acids results in the formation of a rigid and tubular structure called **α helix**. This constitutes the *secondary structure* of proteins, which refers to the steric or spatial relationship of amino acids that are near to each other in the amino acid sequence.

Based on the nature of hydrogen bonding (whether intramolecular or intermolecular), Pauling and Corey (1951) identified two regular types of secondary structure in proteins : α helix (α -helix) and beta pleated sheet (β -pleated sheet).

α -Helix

Early x-ray diffraction studies conducted by William Astbury in 1930s, of fibrous proteins such as hair and wool, showed a major **periodicity** or repeat unit of 5.0 to 5.5 Å, indicating some regularity in the structure of these proteins. A minor repeat unit of 1.5 Å was also observed. With x-ray diffraction studies, Pauling and Corey (1951) found that a polypeptide chain with planar peptide bonds would

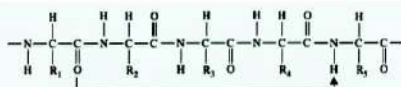
Fig. 9-22. Structure of the α helix (A)

A ribbon depiction with the α -carbon atoms and side chains (green) shown. (B) A side view of a ball-and-stick version depicts the hydrogen bonds (dashed lines) between NH and CO groups. For (C), consult box figure (D) A space-filling view of part C shows the tightly packed interior core of the helix.

186 / 1254

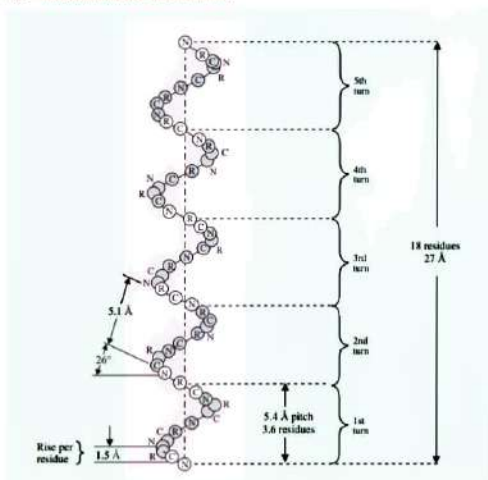
form a right-handed helical structure by simple twists about the α -carbon-to-nitrogen α -carbon-to-carboxyl carbon bonds. They called this helical structure as α -helix. The helix, named because of the mobility of α -carbon atoms.

The α -helix is a rodlike structure. The tightly coiled polypeptide main chain forms the inner part of the rod, and the side chains extend outward in a helical array (Figs. 9-21 and 9-22). The α -helix is stabilized by hydrogen bonds between the NH and CO groups of the main chain. The CO group of each amino acid is hydrogen-bonded to the NH group of the amino acid that is situated four residues ahead in the linear sequence (Fig. 9-23). Thus, all the main chain CO and NH groups are hydrogen-bonded.

Fig. 9-23. Diagram showing that in the α -helix, the CO group of residue n is hydrogen-bonded to the NH group of residue $(n + 4)$

It is, thus, apparent that the α -helical structure depends on the intramolecular (= intrachain) hydrogen bonding between the NH and CO groups of peptide bonds. The hydrogen bonding occurs spontaneously and, as a result, a polypeptide can assume a rod-like structure with well-defined dimensions (Fig. 9-24). The α -helix (or α -conformation, as it is also called) has a pitch of 5.4 Å (= 0.54 nm) and contains 3.6 amino acids per turn of the helix, thereby giving a rise per residue of 5.4/3.6 = 1.5 Å (= 0.15 nm), which is the *identity* period of α -helix. The amino acid residues in an α -helix have conformations with $\phi = -60^\circ$ and $\psi = -45^\circ$ to -50° .

162 FUNDAMENTALS OF BIOCHEMISTRY

Fig. 9-24. Average dimensions of an α -helix

Note the average dimensions of an α -helix. The letters C and N represent the carbon and nitrogen atoms respectively in the peptide bonds. The letter R represents an α -carbon with a side chain attached. The shaded circles represent atoms back of the plane of the paper while the white circles indicate atoms above the plane.

(Redrawn from Laurs Pauling and Robert Corey, 1955)



Fig. 9-25. Two types of helices. Right- and left-handed helices are related to each other in the same way as right and left hands.

A helix can be right-handed (clockwise) or left-handed (anticlockwise) (refer Fig 9-25); α -helices of known polypeptides (i.e., α -amino acids) are right-handed. Biologically functional proteins do not usually exhibit cent per cent α -helical structures. Some have a high percentage of their residues in α -helical structures, e.g., myoglobin and hemoglobin; others have a low percentage, e.g., chymotrypsin and cytochrome C. A very interesting feature of the α -helix, besides the periodicity, is the fact that the carbonyl group ($>C=O$) of every peptide bond is in a position to form a hydrogen bond with the $>N-H$ group of the peptide bond in the next turn of the helix, thereby contributing to the *stability* of the α -

helix. Long polypeptides chains in the α -helical conformation, forming structures are reminiscent of a multistranded rope. The microfibril of hair, for example, is 20 Å in diameter and appears to be made up of 3 right-handed helices wound about each other to form a left-handed supercoil.

The model-building experiments have shown that an α -helix can form with either D- or L-amino acids. However, the residues must all be of one stereoisomer; a D-amino acid will disrupt any regular structure consisting of L-amino acids, and vice versa. The preferential formation of such a helix over many other possible conformations in nature is due, in part, to the fact that it (α -helix) makes optimal use of internal hydrogen bonds.

Many globular proteins contain short regions of such α -helices, and those portions of a transmembrane protein that cross the lipid bilayer are usually α -helices because of the constraints imposed by the hydrophobic lipid environment. In aqueous environment, an isolated α -helix is usually not stable on its own. Two identical α -helices that have a repeating arrangement of nonpolar side chains, however, will twist around each other gradually to form a particularly stable structure known as a coiled coil. Long rod-like coiled coils are found in many fibrous proteins, such as intracellular α -keratin fibres that reinforce skin and its appendages.

α -helix occurs in the protein α -keratin, found in skin and its appendages such as hair, nails and feathers and constitutes almost the entire dry weight of hair, wool, feathers, nails, claws, quills, scales, horns, hooves, tortoise shell, and much of the outer layer of skin. The basic structural unit of α -keratin usually consists of 3 right-handed helical polypeptides in a left-handed coil that is stabilized by crosslinking disulfide bonds (Fig. 9-26).

Destabilization of α -helical conformation by certain amino acid residues (Table 9-5) can occur in various ways:

Table 9-5. Amino acids affecting α -helical structure

Destabilize α -helix		Create bends in α -helix
Aspartic acid	Glycine	Proline
Glutamic acid	Serine	Hydroxyproline
Lysine	Isoleucine	
Arginine	Threonine	



Fig. 9-26. Schematic of the 3 α -helical coils in keratin

1. A prolyl residue has its α -N atom in a rigid ring system and cannot participate in α -helical structure; instead, it creates a sharp bend in the helix.
2. A sequence of aspartyl and/or glutamyl residues can destabilize α -helical structure because the negatively-charged side chains repel one another (electrostatic repulsion), and the forces of repulsion are greater than those of hydrogen bonding.
3. A cluster of isoleucyl residues, because of steric hindrance imposed by their bulky R groups, also disrupts helical conformation.
4. Glycine, with a small hydrogen atom as an R group, is another destabilizer. The lack of a side chain on glycine allows for a great degree of rotation about the amino acid's α carbon; hence, conformations other than a helical bond angles are possible.

188 / 1254

164 FUNDAMENTALS OF BIOCHEMISTRY

β -Pleated Sheet

Pauling and Corey (1953) identified a second type of repetitive, minimum-energy or stable conformation, which they named **β -pleated sheet** (β because it was the second structure they elucidated, the α -helix having been the first). The formation of β -pleated sheets depends on **intermolecular** (= interchain) hydrogen bonding, although intramolecular hydrogen bonds are also present. The pleated sheet structure is formed by the parallel alignment of a number of polypeptide chains in a plane, with hydrogen bonds between the $>C=O$ and $-N-H$ groups of adjacent chains. The R groups of the constituent amino acids in one polypeptide chain alternately project above and below the plane of the sheet, leading to a two-residue repeat unit. The β sheet structures are quite common in nature and are favoured by the presence of amino acids, glycine and alanine. Silk and certain synthetic fibres such as nylon and orlon are composed of β -structures.

The β -pleated sheet differs markedly from the rodlike α -helix:

1. A polypeptide chain in a β -pleated sheet, called a β -strand, has **fully extended conformation**, rather than being tightly coiled as in the α -helix.
2. The **axial distance** between adjacent amino acids in β -pleated sheets is 3.5 Å, in contrast with 1.5 Å for the α -helix.
3. β -sheet is stabilized by hydrogen bonds between NH and CO groups in **different** polypeptide strands, whereas in the α -helix, the hydrogen bonds are between NH and CO groups in the same strand.

There are two types of β -pleated sheet structures. If the N-terminal ends of all the participating polypeptide chains lie on the same edge of the sheet, with all C-terminal ends on the opposite edge, the structure is known as a **parallel β -pleated sheet**. In contrast, if the direction of the chains alternates so that the alternating chains have their N-terminal ends on the same side of the sheet, while their C-terminal ends lie on the opposite edge, the structure is known as the **antiparallel β -pleated sheet**. In other words, the hydrogen-bonded neighbouring polypeptides are aligned in the same N-to-C terminus direction in the parallel pleated sheets and in the opposite N-to-C direction in the antiparallel pleated sheets. Both parallel and antiparallel β -pleated sheets have similar structures, although the repeat period is shorter (6.5 Å or 0.65 nm) for the parallel conformation in comparison to antiparallel conformation (7 Å or 0.7 nm). Pleated-sheet structures can also be formed from a single polypeptide chain if the chain repeatedly folds back on itself. Although a β -pleated sheet structure is usually associated with structural proteins, it is also known to occur in the 3-dimensional structures of certain globular proteins, e.g., the enzymes lysozyme and carboxypeptidase A.

Silk fibroin is one example of a protein that has the antiparallel pleated sheet structure. It is a member of a class of fibrillar proteins called α -keratins. Silk fibroin (Fig. 9-27) is composed mainly

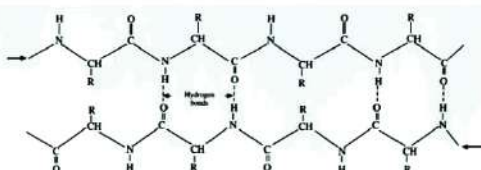
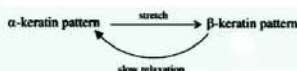


Fig. 9-27. A portion of two chains of silk fibroin

Note that the two chains run in opposite direction.

of glycine, alanine and serine units linked together by the peptide bonds. Glycine constitutes approximately 45% of the total amino acid residues, and alanine plus serine compose another 42%. Hence, the R groups extending above and below the plane of the pleated sheet are small and allow the pleated sheets to stack. Here the two chains which run in opposite directions are linked by hydrogen bonds (shown by dotted lines). It may be noted that each of the two chains can also form hydrogen bonds with still other protein and so on. Very large protein aggregates are, thus, formed provided the R groups are relatively small. If the R groups are too large, the hydrogen bonding may not occur as the chains are held far apart for union. An interesting phenomenon occurs when hair or wool (α -keratins) is treated with moist heat and stretched because β -keratin structures with parallel β pleated sheets are produced.



Random Coil

Besides the α -helix and β -pleated sheet structures of proteins, which were recognized by Pauling and Corey in 1950s, there also exist a third type of secondary structure in proteins called **random coil**. When a polypeptide contains adjacent bulky residues such as isoleucine or charged residues such as glutamic acid and aspartic acid, repulsion between these groups causes the polypeptide to assume a random coil configuration (Fig. 9-28).

Thus, we see that the R groups distributed along the polypeptide backbone determine the secondary structure adopted by different portions of the polypeptide (α -helix or β -pleated sheet), or the lack of a well-defined structure (random coil).

Other Secondary Structures

The α -helix and β -conformation are the major repetitive secondary structures easily recognized in a wide variety of proteins. The γ -helix, however, is an example of a highly hydrogen-bonded structure that is insufficiently stable to be used in proteins because of the lack of interatomic contacts. In addition to the α -helix and β -pleated sheet structures, other repetitive structures also exist in one or a few specialized proteins. Two such structures (β -turn and collagen triple helix) are described below :

β Turn or β Bend or Hairpin Bend

Most proteins have compact, globular shapes owing to reversals in the direction of their polypeptide chain. β turn (Fig. 9-29) is often found where a polypeptide chain abruptly reverses direction. β turns usually connect the ends of two adjacent segments of an antiparallel β pleated sheet, hence their name. The structure is a tight turn ($\sim 180^\circ$) involving 4 amino acids. The essence of this hairpin turn is that the CO

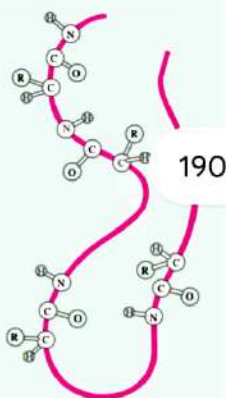


Fig. 9-28. A random coil

The backbone of the polypeptide is drawn, but only a few amino acid residues are shown.

(Adapted from Francisco J. Ayala and John A. Kiger Jr., 1980)

190 / 1254

166 FUNDAMENTALS OF BIOCHEMISTRY

group of residue n of a polypeptide is hydrogen-bonded to the NH group of residue $n + 3$. In other words, the hydrogen bond between the peptide groups of the first and fourth amino acid residues is involved in bend formation. Glycine and proline residues often occur in β turns : the former because it is small and flexible and the latter because peptide bonds involving the imino nitrogen of proline readily assume the *cis* configuration, a form that is particularly amenable to a tight turn. β turns are often found near the surface of a protein. They are also called as **reverse turns**.

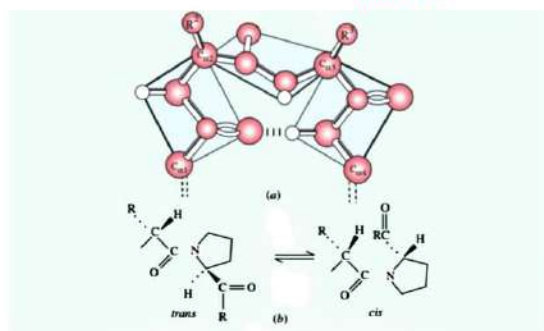


Fig. 9-29. Structure of a β turn or hairpin bend

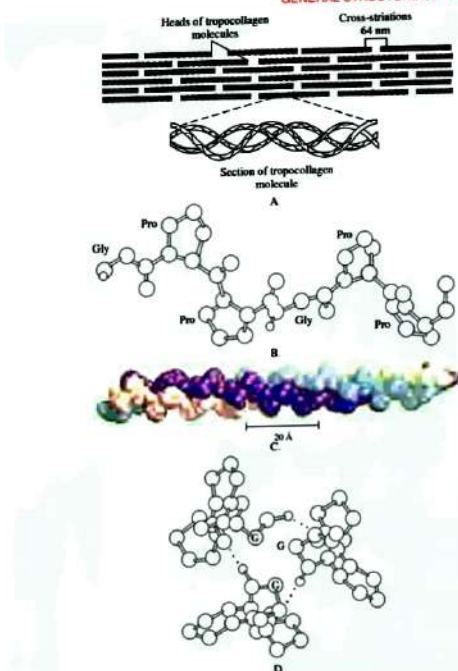
(a) Hydrogen bonding in a tetrapeptide involving first and fourth amino acid residues. Here the NH and CO groups of residue 1 of the tetrapeptide are hydrogen-bonded, respectively, to the CO and NH groups of residue 4.

(b) The *trans* and *cis* isomers of a peptide bond involving the imino nitrogen of proline.

Collagen Triple Helix

Collagen is the most abundant protein of mammals. It is the principal structural element of the human body and makes up 25–33% of all the body protein. It is found in connective tissues such as tendons, cartilage, the organic matrix of bones and the cornea of the eye. This extracellular protein contains three helical polypeptide chains, each nearly 1,000 residues long. The amino acid sequence of collagen is remarkably regular : *nearly every third residue is glycine*. Proline is also present to a much greater extent than in most other proteins. Furthermore, collagen contains 4-hydroxyproline (Hyp), which is rarely found elsewhere. The percentage composition of predominant amino acids found in collagen is : Gly (35%), Ala (11%) and Pro + Hyp (25%). The amino acid sequence in collagen is generally a repeating tripeptide unit, Gly–X–Pro or Gly–X–Hyp, where X can be any amino acid. This repeating tripeptide sequence adopts a left-handed helical structure with 3 residues per turn.

Collagen is a rod-shaped molecule, about 3,000 Å (= 300 nm) long and only 15 Å (= 1.5 nm) in diameter. The helical motif of its 3 chains is entirely different from that of the α helix. Hydrogen bonds within a strand (*i.e.*, intrachain hydrogen bonds) are absent. Instead, each of these 3 helices is stabilized by steric repulsion of the pyrrolidone rings of the proline and hydroxyproline residues. The 3 strands wind around each other in a cable fashion (Fig. 9-30), to form a **superhelix**. The superhelical



192 / 1254

Fig. 9-30. The structure of collagen

- Collagen fibres and a section of tropocollagen molecule. The basic collagen monomer is a triple helix, composed of 3 helical α chains. The collagen monomers before aligned in rows in which the molecules in one row are staggered relative to those in the neighbouring rows.
- Conformation of a single strand of the collagen triple helix. The sequence shown here is Gly-Pro-Gly-Pro-Gly-Pro.
- Space-filling model of the collagen triple helix. The three strands are shown in different shades.
- Cross section of a model of the triple-stranded helix of collagen. Each strand is hydrogen-bonded to other two (... denotes a hydrogen bond). The α -carbon atom of a glycine residue in each strand is labelled G. Every third residue must be glycine because there is no space near the helix axis (centre) for a larger amino acid residue. Note that the pyrrolidine rings are on the outer side.

168 FUNDAMENTALS OF BIOCHEMISTRY

twisting is right-handed (cf α -keratin). The rise (axial distance) per residue in this superhelix is 2.9 Å, and there are nearly 3 residues per turn. The three strands are stabilized by the interchain formation of hydrogen bonds between the $>C=O$ group of one chain and the $>N-H$ group of another chain. Proline does not have a hydrogen atom attached to its nitrogen when participating in a peptide bond and therefore cannot participate in this interchain hydrogen-bonding. This 3-stranded superhelix is known as the **collagen triple helix**. The amino acid residue on either side of glycine is located on the outside of the cable, where there is room for the bulky rings of Pro and Hyp residues. The tight wrapping of the collagen triple helix provides great tensile strength with no capacity to stretch. Collagen fibres can support up to 10,000 times their own weight and are said to have greater tensile strength than a steel wire of equal cross section.

The collagen fibrils consist of recurring 3-stranded polypeptide units called **tropocollagen** (MW 3,00,000), arranged head-to-tail in parallel bundles (Fig. 9-29). In some collagens, all three chains are identical in amino acid sequence and thus are homotrimers; while in others, two chains are identical and the third differs and thus are heterotrimers. The heads of adjacent molecules are staggered, and the alignment of the head groups of every fourth molecule produces characteristic cross-striations 640 Å (= 64 nm) apart. A series of complex covalent cross-links are formed within and between the tropocollagen molecules in the fibril, leading to the formation of strong mature collagen. The rigid, brittle character of the connective tissue in older people is the result of an accumulation of covalent cross-links in collagen as they age. The electron micrograph of human collagen fibres is given in Fig 9-31.

The food product **gelatin** is derived from collagen. Although it is a protein, it has little nutritional value because collagen lacks significant amounts of many amino acids that are essential in the human diet.

Human genetic defects involving collagen illustrate the close relationship between amino acid sequence and three-dimensional structure in this protein. **Osteogenesis imperfecta** results in abnormal bone formation in human babies. **Ehlers-Danlos syndrome** is characterized by loose joints. Both can be lethal and both result if a different glycine residue in each case is replaced by a cysteine (in case of osteogenesis imperfecta) or a serine (in case of Ehlers-Danlos syndrome) residue. These seemingly small replacements have a catastrophic effect on collagen function because they disrupt the Gly-X-Pro (or Hyp) repeat that gives collagen its unique helical structure.

Table 9-6 summarizes the important differences between the structures of α -keratin of hair and collagen of bones.



Fig. 9-31. An electron micrograph 0.3 μ m of human collagen fibers showing their characteristic banding pattern. The bands repeat along the fibre with a periodicity of 64 to 70 nm.

(Courtesy : Jerome Gross and Francis O. Schmit)

Table 9-6. Structural differences between α -keratin and collagen

α -keratin	Collagen
1. Found in skin and its appendages such as hair, nails and feathers.	1. Found in connective tissues such as tendons, cartilage, bones and the cornea of the eye.
2. The polypeptide chain is a right-handed helix.	2. The unique polypeptide chain is a left-handed helix.
3. Hydrophobic amino acids (Phe, Ile, Val, Met and Ala) predominate in the helix.	3. Gly, Pro, Hyp and Ala predominate in the helix.
4. The 3 helical strands wrap together into a superhelix, called protofibril.	4. The 3 helical strands wrap together into a repeating superhelical structure, called tropocollagen.
5. The superhelical twisting is left-handed.	5. The superhelical twisting is right-handed.
6. The covalent cross-links, between polypeptide chains within the triple-helical ropes and between adjacent ones, are contributed by disulfide bonds.	6. The covalent links are contributed by an unusual type of covalent link between two Lys residues that creates a nonstandard amino acid residue called lysinonorleucine.

Elastin

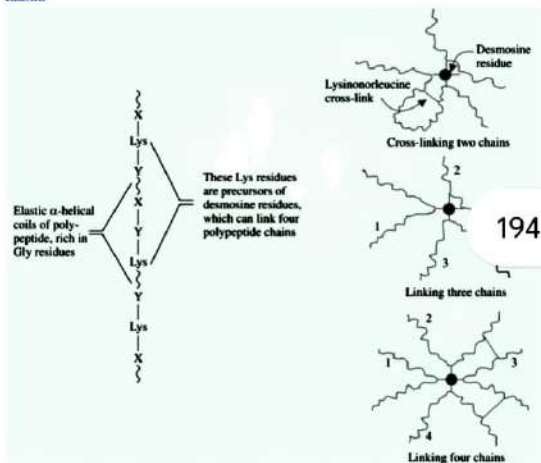


Fig. 9-32. Tropoelastin molecules and their linkage to form a network of polypeptide chains in elastin. As shown, desmosine residues can link two, three or four tropoelastin molecules. In addition, other types of linkages such as lysinonorleucine are also present.

(Adapted from Lehninger, Nelson and Cox, 1993)

Contents

170 FUNDAMENTALS OF BIOCHEMISTRY

Elastin is found in elastic connective tissues such as ligaments and blood vessels. It resembles collagen in some of its properties, but differs in other. The polypeptide subunit of elastin fibrils is **tropoelastin** (MW 72,000), containing about 800 amino acid residues. Tropoelastin differs from tropocollagen in having many lysine but few proline residues. Also, the helix which it forms is quite different from a helix and the collagen helix.

Tropoelastin consists of α helical portions of polypeptides rich in Gly residues, separated by short regions containing Lys and Ala residues. The helical regions stretch on applying tension but regain their original length when tension is released. The regions containing Lys residues form covalent cross-links. Four lysine side chains come together and are enzymatically converted into **desmosine** and a related compound, **isodesmosine**; these amino acids are found exclusively in elastin. Like collagen, elastin also contains lysinonorleucine. These nonstandard amino acids are capable of joining tropoelastin chains into arrays that can be stretched reversibly in all directions (Fig. 9-32).

Hydrogen-bonding Potentiality of Proteins

All reversible molecular interactions in biological systems are mediated by 3 kinds of forces: electrostatic bonds, hydrogen bonds and van der Waals bonds. We have already seen that the hydrogen bonds between main-chain NH and CO groups work in forming α helices, β sheets and collagen fibrils. In fact, side chains of 11 of the 20 fundamental amino acids also can participate in hydrogen-bonding. These have been separated into 3 groups, according to their hydrogen-bonding potentialities:

1. **Tryptophan and Arginine.** Their side chains can serve as *hydrogen bond donors only*.



2. **Asparagine, Glutamine, Serine and Threonine.** Like the peptide unit itself, their side chains can serve as *hydrogen bond donors and acceptors*.

3. **Lysine, Aspartic acid, Glutamic acid, Tyrosine and Histidine.** The hydrogen-bonding capabilities of lysine (and the terminal amino group), aspartic and glutamic acids (and the terminal carboxyl group), tyrosine and histidine vary with pH. These groups can serve as *both acceptors and donors* within a certain pH range and as *acceptors or donors* (but not both) at other pH values (Fig. 9-33).

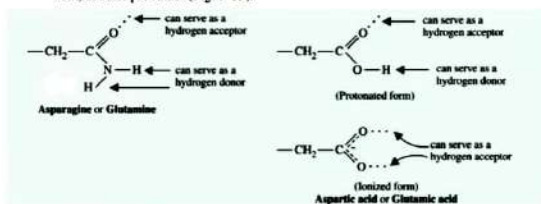


Fig. 9-33. Hydrogen-bonding groups of several side chains in proteins

3. Tertiary Structure : Folding of the Chain or Overall Folding

If the globular proteins consisted only of a small helix, these molecules would have been elongated structures with considerable length and a small cross-sectional area (*i.e.*, a large axial ratio). But as we now know about the existence of globular proteins, the helix must, therefore, possess many other types of bonds placed at regular intervals. These additional bonds include disulfide, hydrogen, hydrophobic and ionic. In such globular proteins (including enzymes, transport proteins, some peptide hormones and immunoglobulins), polar groups because of their hydrophobicity are most often located on the molecule's exterior and nonpolar R groups in the interior, where their interactions create a hydrophobic environment. The *tertiary structure*, thus, involves the folding of the helices of globular proteins. It refers to the spatial arrangement of amino acids that are far apart in linear sequence and to the pattern of disulfide bonds. The dividing line between secondary and tertiary structure is, hence, a matter of taste. X-ray crystallographic studies have revealed the detailed 3-'D' structures of more than 300 proteins.

Myoglobin (Mb)

Myoglobin (*myo*^G = muscle ; *globin*^G = a type of protein) is a relatively small, oxygen-binding heme protein, found in muscle cells. It has the distinction of being the first globular protein to have its 3-'D' structure elucidated by x-ray diffraction studies. This was accomplished by John C. Kendrew at a resolution of 6 Å in 1957, 2 Å in 1959 and 1.4 Å in 1962 (Fig. 9-34). Myoglobin molecule (Fig. 9-35) contains a single polypeptide chain of 153 amino acid residues and a single prosthetic iron-porphyrin (or heme) group, identical with that of hemoglobin. The heme group is responsible for the deep red-brown colour of myoglobin (and also of hemoglobin). Myoglobin is especially abundant in the muscles of diving mammals such as the whale, seal and porpoise, whose muscles are so rich in this protein that they are brown. Storage of oxygen by muscle myoglobin permits these animals to remain submerged for long periods. The function of myoglobin is to bind oxygen in the muscles and to enhance its transport to the mitochondria, which consume oxygen during respiration.

Between 1912 and 1915, William Henry Bragg and his son, William Lawrence Bragg, developed the technique of x-ray diffraction by determining the crystalline structure of NaCl. As the joint 1915 Nobel Prize recipients in physics, the Braggs became the only father-son combination to receive the award and W. Lawrence, who was 25 years old at the time, the youngest scientist so honoured.

Myoglobin (MW = 16,700) is an extremely compact macromolecule with oblate, spheroid shape and leaves little empty space in its interior. Its overall molecular dimensions

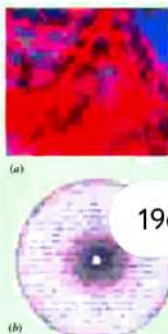


Fig. 9-34. Sperm whale myoglobin crystal and x-ray photograph

(a) Crystal of myoglobin. (b) An X-ray diffraction pattern (or photograph) of a single crystal of sperm whale myoglobin. The pattern of spots is produced as a beam of x-rays is diffracted by the atoms in the protein crystal, causing the x-rays to strike the film at specific sites. Information derived from the position and intensity (darkness) of the spots can be used to calculate the positions of the atoms in the protein that diffracted the beam. The intensity of each diffraction maximum (the darkness of each spot) is a function of the myoglobin crystal's electron density. The photograph contains only a small fraction of the total diffraction information available from a myoglobin crystal.

172 FUNDAMENTALS OF BIOCHEMISTRY

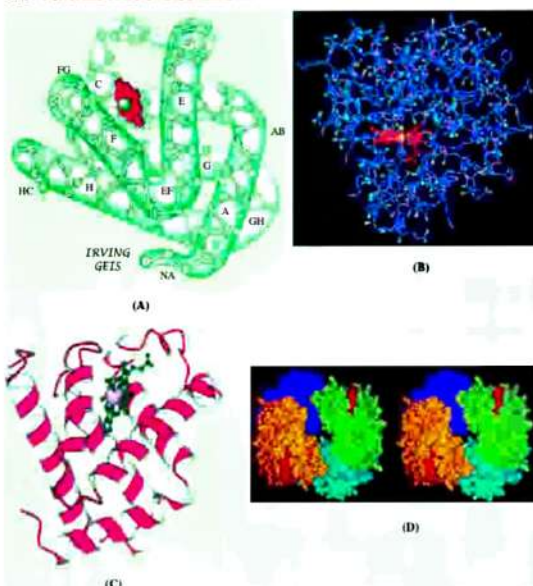


Fig. 9-35. The structure of sperm whale myoglobin

(A) The tertiary structure. Its 153 C_α positions are numbered from the N-terminus and its eight helices are sequentially labeled A through H. The last half of the EF corner is now regarded as a turn of helix and is therefore designated the F' helix, the heme group is shown in red. Most of the amino acids are part of α helices, the nonhelical regions occur primarily as turns, where the polypeptide chain changes direction. The position of the heme is indicated in red.

(B) The ball- and stick model. The three-dimensional structure of myoglobin shows the position of all of the molecule's atoms other than hydrogen and reveals many interactions between the amino acids. The heme group is indicated in red.

(C) Ribbon model. A schematic view shows that the protein consists largely of α helices. The heme group is shown in black and the iron atom is shown as a purple sphere

(D) A stereo, space-filling model showing the quaternary structure of hemoglobin. The model shows the α₁, α₂, β₁, and β₂ subunits as colored yellow, green, light blue, and dark blue, respectively. Heme groups are red. The protein is viewed along its molecular twofold rotation axis which relates the α₁β₁ protomer to the α₂β₂ protomer.

(Courtesy : (A) Irving Geis (B) Ken Eward)

are $45 \times 35 \times 25 \text{ \AA}$. The sausage-like outline of the tertiary structure of myoglobin chain is folded into an odd, irregular form. The backbone of the molecule is made up of 8 almost-straight α -helical segments, designated from the N-terminus as A through H (Fig 9-36). The first residue in helix A is designated A_1 , the second A_2 and so forth. Interspersed among the helical segments are 5 nonhelical regions, each identified by the two helical segments it joins, e.g., AB is located between helical sections A and B. There are also two other nonhelical regions: two residues at the N-terminus (named NA_1 and NA_2) and five residues at the C-terminus (named HC_1 through HC_5). The longest helical segment has 23 amino acid residues and the shortest only 7. All the helical segments are of α -type and right-handed; there being no β -structure (refer Table 9-7).

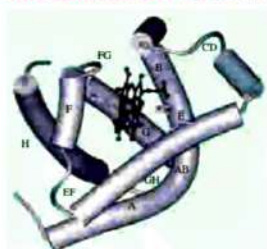
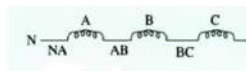


Fig. 9-36. The eight α helical segments of myoglobin molecule.

The eight α helical segments (shown here as cylinders) are labeled A through H. Nonhelical residues in the bends that connect them are labeled AB, CD, EF, and so forth, indicating the segments they interconnect. A few bends, including BC and DE, are abrupt and do not contain any residues; these are not normally labeled. (The short segment visible between D and E is an artifact of the computer representation) is bound in a pocket made up largely of helices, although amino acid residues of the protein also participate.

Table 9-7. Approximate amounts of α helix and β -structure in some single-chain proteins*

Protein	Total residues	Residues, %	
		α -helix	β -structure
Cytochrome C	104	39	0
Ribonuclease	124	26	35
Lysozyme	129	40	12
Myoglobin	153	78	0
Chymotrypsin	247	14	45
Carboxypeptidase	307	38	17

* Portions of polypeptide chains, that are not accounted for by α -helix or β -structure, consist of bends, reverse turns and irregularly-coiled stretches.

(Adapted from Cantor and Schimmel, 1980)

Of the 153 amino acid residues, 121 (i.e., 79%) are present on the helical regions and the remaining 32 amino acid residues are distributed over the nonhelical areas. The nonhelical areas possess various types of bonding such as hydrogen and nonpolar linkages. The flat heme group is tightly but noncovalently bound to the polypeptide chain. A notable feature of whale myoglobin is the absence of a disulfide bridge, since both cysteine and cystine residues are lacking. The electron density map of sperm whale myoglobin is presented in Fig. 9-37.

174 FUNDAMENTALS OF BIOCHEMISTRY

Other important features of the myoglobin molecule are listed below:

1. The molecule is very compact and leaves so little space in its interior as to accommodate only 4 water molecules.
2. All the polar R groups of the molecule, except two, are located on its outer surface and all of them are hydrated.
3. Most of the hydrophobic R groups are located in the interior of the molecule. Hydrophobic R groups of helices E and F form the sides of a pocket into which the hydrophobic heme group fits. The porphyrin ring of the heme is largely hydrophobic, except for the 2 propionic acid side chains which stick out of the pocket and into the aqueous environment.
4. Each of the 4 proline residues occurs at a turn. Other turns or bends contain serine, threonine and asparagine.
5. All the peptide bonds of the polypeptide chain are planar, with the carbonyl and amide groups being *trans* to each other.
6. The heme group is flat and rests in a crevice in the molecule. The iron atom

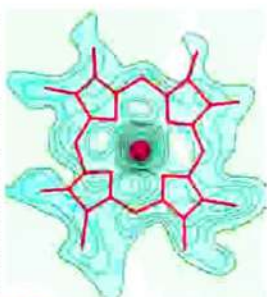


Fig. 9-37. A section through the 2.0- $\text{\AA}$ -resolution electron density map of sperm whale myoglobin.

The heme group is represented by red. The large peak in the center of the map represents the electron-dense Fe atom.

(After Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC and Shore VC,

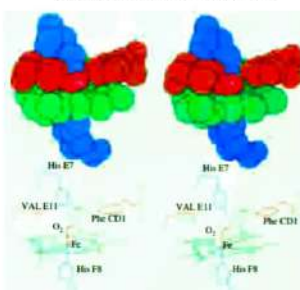


Fig. 9-38. Stereo drawings of the heme complex in oxyhemoglobin.

In the upper drawing, atoms are represented as spheres of van der Waals radii; the lower drawing shows the corresponding skeletal model with a dashed line representing the hydrogen bond between the distal His and the bound O_2 .

(After Phillips SEV, 1980)

in the centre of the heme group has two coordination bonds that are perpendicular to the plane of the heme group. One bond is attached to the R group of histidine (93), whereas the other bond is the site to which an O_2 molecule is bound. Upon oxygenation, the iron atom descends into the heme plane and the oxygen bound to myoglobin is stabilized by hydrogen-bonding to the imidazole ring of His E₇ (Fig. 9-38). Reversible oxygenation requires that the iron atom be in the ferrous state (Fe^{2+}) and myoglobin with or without oxygen bound to the Fe^{2+} of heme is called oxyhemoglobin and deoxyhemoglobin respectively.

7. The inside and outside are well defined. There is little empty space inside. The interior consists almost entirely of nonpolar residues such as Leu, Val, Met, and Phe. On the contrary, Glu, Asp, Gln, Asn, Lys and Arg are absent from the interior of

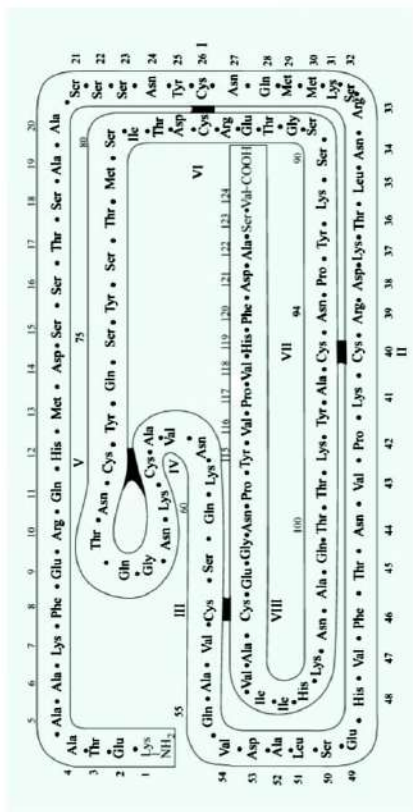


Fig. 9-39. The amino acid sequence of the enzyme ribonuclease from bovine. This enzyme molecule consists of 124 amino acids. The 4 disulfide bonds or bridges, shown in the diagrammatic, the polypeptide is actually folded to a 3-dimensional configuration. (After Stryer, D.G., Serin, W.H. and Moore, S., 1963.)

200 / 1254

176 FUNDAMENTALS OF BIOCHEMISTRY

protein. Residues with polar and nonpolar parts within them (Thr, Tyr, Trp) are so oriented that their nonpolar portions point inward. The only polar residues inside myoglobin are two histidines which have a critical function at the binding site. The exterior of the protein has both polar and nonpolar residues.

- Myoglobin without its heme prosthetic group is designated *apomyoglobin*. The main function of apomyoglobin molecule is to provide a hydrophobic environment for the heme group and a properly oriented imidazole group to occupy the 5th coordination position of the iron. The presence of the hydrophobic environment and the proximal histidine enable the heme to combine reversibly with O_2 and prevent the oxidation of Fe^{2+} to Fe^{3+} .

Ribonuclease (RNase)

Ribonuclease (MW = 13,700) is another noteworthy example of a protein with tertiary structure. Two groups of workers (Christian Anfinsen and his associates, 1950, at the National Institute of Health; William Stein and Stanford Moore, 1958, at Rockefeller Institute) have elucidated the complete structure of this pancreatic protein. Ribonuclease molecule (Fig. 9-39) consists of 124 amino acid units contained in a single polypeptide chain with 4 disulfide bridges. In establishing the amino acid sequence, the RNase was subjected to hydrolysis either by pepsin or by trypsin (for further details, refer Fig. 17-10 on page 300).

Carboxypeptidases

Carboxypeptidases are made by the exocrine cells of the pancreas as their inactive zymogens, *procarboxypeptidases*. Synthesis of these enzymes protects the exocrine cells from destructive proteolytic attack. Carboxypeptidases are exopeptidase enzymes that catalyze the hydrolysis of (or cleave) proteins either from N- or C-terminus, i.e., inwardly from only one end of

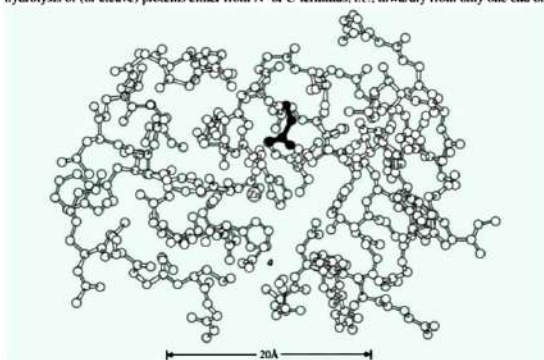


Fig. 9-40. The structure of carboxypeptidase A

The three shaded amino acids are Glu 270 (upper), Tyr 248 (middle) and Arg 145 (lower). Also shown is the zinc ion (Zn^{2+}) located in a groove near the surface of the molecule.

(After William N. Lipscomb, 1971)

the chain. (By contrast, endopeptidases are enzymes that cleave internal peptide bonds.). Different forms of carboxypeptidases are usually *substrate-specific*. *Carboxypeptidase A*, the enzyme used initially in C-terminal residue determinations, readily hydrolyzes (or cleaves) a C-terminal residue with an aromatic or a bulky aliphatic side chain. *Carboxypeptidase B* cleaves C-terminal lysine or arginine residues while *carboxypeptidase C* cleaves a C-terminal proline residue. However, *carboxypeptidase Y* has the capacity to cleave all C-terminal amino acids.

Carboxypeptidase A (MW = 34,600) is a zinc-containing proteolytic enzyme. Its 3-D structure (Fig. 9-40) was accomplished at a resolution of 2 Å by William Lipscomb in 1967. This enzyme is a globular protein which contains 307 amino acid residues in a single polypeptide chain. Carboxypeptidase A molecule is a compact ellipsoid with dimensions of $50 \times 42 \times 38$ Å. This molecule is considerably larger than myoglobin and shows more variety in its secondary structures. Of the 307 amino acid residues, 38% are involved in the formation of 8 helical segments (designated A through H) and 17% are involved in β pleated sheet structure that extends through the molecule, forming a core. Helix D is connected to helix E by a segment containing 52 amino acids. The folding of this segment is very complex and contains no recognizable regular secondary structure. A tightly bound zinc ion (which is essential for enzymatic activity) is located in a groove near the surface of the molecule, where it coordinates with 2 histidine side chains, a glutamate side chain and a water molecule. A large pocket near the zinc ion accommodates the side chain of the terminal residue of the peptide substrate.

Fig. 9-41 gives the ribbon model of the structure of bovine carboxypeptidase A, along with its central β sheet.



Fig. 9-41. Bovine carboxypeptidase A showing its central β sheet

(Photo Courtesy : Irving Geis, 1986)

Two aspects of catalytic mechanism of carboxypeptidase A are noteworthy :

(A) **Binding of substrate** : For instance, the binding of glycyltyrosine (a slowly-hydrolyzed substrate) is accompanied by a large structural rearrangement of the active site. The phenolic hydroxyl group of tyrosine 248 moves a distance of 12 Å from the surface of the molecule to the vicinity of the terminal carboxylate of the substrate. Consequently, the active site cavity closes so that the water molecule extrudes from it.

202 / 1254

178 FUNDAMENTALS OF BIOCHEMISTRY

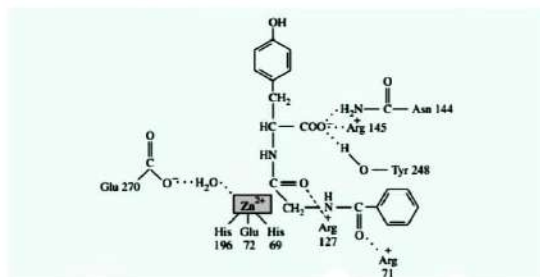


Fig. 9-42. Mode of binding of N-benzoylglycyltyrosine to carboxypeptidase A

(After Christianson DW and Lipscomb WN, 1989)

How does carboxypeptidase A recognize the terminal carboxylate group of a peptide substrate? Three interactions explain this (Fig. 9-42). The induced-fit movement of Tyr 248 enables its phenolic OH group to become hydrogen-bonded to the terminal carboxylate group. Further, the guanidium of Arg 145 moves 2 Å in this transition so as to enable it to form a salt bridge with the terminal COO^- , which is hydrogen-bonded to a side chain Asn 144. Besides, the terminal side chain of the substrate sits in a hydrophobic pocket of the enzyme (this is the reason why carboxypeptidase A requires an aromatic or bulky nonpolar residues at this position). Two interactions involving the penultimate amino acid residue are also noteworthy. The carbonyl oxygen of the scissile peptide bond interacts with the guanidium group of Arg 127. Finally, Tyr 248 is also hydrogen-bonded to the peptide NH of penultimate residue.

(B) **Activation of bound water** : Proteins and peptides are stable at neutral pH in the absence of a protease because water does not readily attack peptide bonds. In fact, carboxypeptidase A also activates a water molecule. This is accomplished by the bound zinc ion with the help of adjacent carboxylate group of Glu 270. Thus, the *zinc-bound water behaves much like an OH^- ion*. The activation of water by zinc ion involves two steps :

1. First step in catalysis is the attack of activated water molecule on the carbonyl group of the scissile peptide bond (Fig. 9-43). Specifically, the nucleophilic oxygen atom of activated water attacks the carbonyl carbon atom. Concurrently, Glu 270 accepts a proton from the water. A negatively-charged tetrahedral intermediate is, thus, formed. This intermediate is stabilized by electrostatic interactions with Zn^{2+} and the positively-charged side chain of Arg 127.

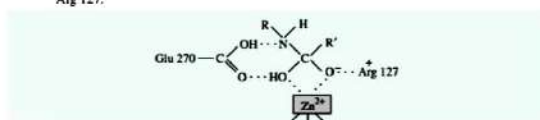


Fig. 9-43. Proposed tetrahedral transition state in the peptide bond hydrolysis by carboxypeptidase A

2. The final step is the transfer of a proton from the COOH group of Glu 270 to the peptide NH group. The peptide bond is simultaneously cleaved and the reaction products diffuse away from the active site. In fact, the bound substrate is surrounded on all sides by catalytic groups of the enzyme. And the substrate-induced structural changes in the active site promote catalysis in 3 ways :
 - (a) activation of H_2O by Zn^{2+} ,
 - (b) proton abstraction and donation by Glu 270, and
 - (c) electrostatic stabilization by Arg 127.

It is, thus, apparent that a substrate could not enter such an array of catalytic groups (nor could a product leave) unless the enzyme were flexible.

4. Quaternary Structure : Protein-Protein Interactions or Multichain Association

A fourth degree of complexity in protein structure has recently been recognized to be of great value in many proteins. Some globular proteins consist of 2 or more interacting peptide chains. Each peptide chain in such a protein is called a *subunit*. These chains may be identical or different in their primary structure. This specific association of a number of subunits into complex large-sized molecules is referred to as the *quaternary structure*. In other words, quaternary structure refers to the spatial arrangement of subunits and the nature of their contact. The same forces (disulfide, hydrogen, hydrophobic and ionic bonds) involved in the formation of tertiary structure of proteins are also involved here to link the various polypeptide chains.

Tobacco Mosaic Virus (TMV)

TMV, with 158 amino acid residues, is an instance of the protein-protein interactions. TMV particles are rod like, tubular structures measuring 3,000 Å length and 180 Å diameter. The tube is made up of a single RNA molecule (30,000 Å long) which is coiled round the central axis. The RNA molecule is a right-handed helical filament and is embedded in a protein matrix or the coat which is made up of 2,130 identical molecules of a single protein that interact to form a cylinder RNA genome. The amino acid sequence of this coat protein (Fig. 9-44) was elucidated independently by H. Fraenkel-Conrat and H. Wittmann. The coat protects the RNA

204 / 1254

Ac-Ser-Tyr-Ser-Ile-Thr-Thr-Pro-Ser-Gln-Phe-Val-Phe-Leu-Ser-Ser-Ala-Trp-Ala-Asp-Pro-				
1	5	10	15	20
Ile-Glu-Leu-Ile-Asn-Leu-Cys-Thr-Asn-Ala-Leu-Gly-Asn-Gln-Phe-Gln-Thr-Gln-Gln-Ala				
25	30	35	40	
Arg-Thr-Val-Val-Gln-Arg-Gln-Phe-Ser-Glu-Val-Trp-Lys-Pro-Ser-Pro-Gln-Val-Thr-Val-				
45	50	55	60	
Arg-Phe-Pro-Asp-Ser-Asp-Phe-Lys-Val-Tyr-Arg-Tyr-Asn-Ala-Val-Leu-Asp-Pro-Leu-Val-				
65	70	75	80	
Thr-Ala-Leu-Leu-Gly-Ala-Phe-Asp-Thr-Arg-Asn-Arg-Ile-Ile-Glu-Val-Glu-Asn-Gln-Ala-				
85	90	95	100	
Asn-Pro-Thr-Thr-Ala-Glu-Thr-Leu-Asp-Ala-Thr-Arg-Arg-Val-Asp-Asp-Ala-Thr-Val-Ala				
105	110	115	120	
Ile-Arg-Ser-Ala-Ile-Asn-Asn-Leu-Ile-Val-Glu-Leu-Ile-Arg-Gly-Thr-Gly-Ser-Tyr-Asn-				
125	130	135	140	
Arg-Ser-Ser-Phe-Glu-Ser-Ser-Ser-Gly-Leu-Val-Trp-Thr-Ser-Gly-Pro-Ala-Thr				
145	150	155	158	

Fig. 9-44. Amino acid sequence of the coat protein or tobacco mosaic virus (TMV)

180 FUNDAMENTALS OF BIOCHEMISTRY

Note that the serine in position 1 is acetylated (AcSer). Also note that there are no extended repetitions of an amino acid sequence, that there are no clusters of either polar or nonpolar amino acids, and that the frequencies of occurrence of different amino acids vary widely.

Hemoglobin (Hb)

Hemoglobin ($hemo^G = blood$; $globin^G = a$ protein, belonging to the myoglobin-hemoglobin family), the oxygen transporter in erythrocytes, constitutes about 90% of the protein of red blood cells. It is a tetrameric protein, i.e., it contains 4 polypeptide chains. The four chains (of which two are of one kind and two of another) are held together by noncovalent interactions. Each chain contains a heme group and a single oxygen-binding site. Hemoglobin A, the principal hemoglobin in adults, consists of two alpha (α) chains and two beta (β) chains. Adults also have a minor hemoglobin (~2% of the total hemoglobin) called hemoglobin A₂, which contains two delta (δ) chains in place of two β chains of hemoglobin A. Thus, the subunit composition of hemoglobin A is $\alpha_2\beta_2$, and that of hemoglobin A₂ is $\alpha_2\delta_2$. HbA and HbA₂ are the postnatal forms of hemoglobin.

Prior to birth, embryonic and fetal hemoglobins are used as oxygen carrier. The first hemoglobin to appear in embryonic development has the subunit composition $\zeta_2\epsilon_2$; the two zeta (ζ) chains are analogous to the α chain and the two epsilon (ϵ) chains are analogous to the β chains. When, after about 6 weeks, ζ chain production ceases, the tetramer $\alpha_2\epsilon_2$ appears, i.e., ζ is replaced by α . A third embryonic hemoglobin, $\zeta_2\gamma_2$ has also been identified where ϵ is replaced by γ . These last two hemoglobins represent transition phases leading to the appearance of fetal hemoglobin (HbF), whose tetrameric composition is $\alpha_2\gamma_2$. The α and ζ chains contain 141 residues each. The β , γ and δ chains contain 146 residues each and have homologous (similar but not identical) amino acid sequences. The γ and δ chains differ from the β chain at 39 and 10 amino acid residues respectively.

Hemoglobin (MW = 64,500) provides an example of the interaction of unidentical protein subunits. Since hemoglobin is 4 times as large as myoglobin, much more time and effort were required to solve its 3-D structure, finally achieved by Max Perutz, John C. Kendrew and their colleagues in 1959. They determined the 3-D structure of adult hemoglobin (HbA) of horse, which is very similar to that of human hemoglobin (HbA). The x-ray analysis has revealed that the hemoglobin molecule is nearly spherical with a diameter of 55 Å (5.5 nm).

Human hemoglobin protein (Fig. 9-45) consists of 4 polypeptide chains of two types, two α -chains and two β -chains. The polypeptide portion is collectively called as *globin*. The α -chain has valine at the N-terminal and arginine at the C-terminal whereas in the β -chain, valine is situated at the N-terminal and histidine at the C-terminal. Each α chain is in contact with both β chains. In contrast, there are few interactions

The designations α and β , which are common labels in Biochemistry, are used simply to identify two different polypeptide chains and are not to be confused with α and β secondary structures or α and β amino acids.

between the two α -chains or between the two β -chains. Each chain has a heme prosthetic group in a crevice near the exterior of the molecule. The heme groups are involved in the binding of oxygen. The α -chain has 141 residues and the β -chain, which is more acidic, has 146 residues. The protein, thus, has a total $(141 \times 2 + 146) = 574$ amino acid residues. Each of the 4 chains has a characteristic tertiary structure, in which the chain is folded. Like myoglobin, the α - and β -chains of hemoglobin contain several segments of α -helix, separated by bends. The α - and β -chains are held together as a pair by ionic and hydrogen bonds. The 2 pairs are then joined to each other by additional ionic bonds, hydrogen bonds and the hydrophobic forces. Thus, the 4 polypeptide chains fit together almost tetrahedrally to produce the

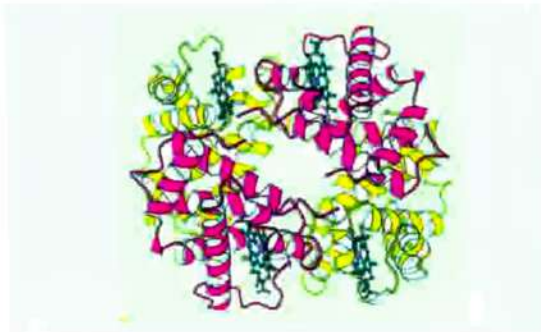


Fig. 9-45. Quaternary structure of hemoglobin molecule, showing interaction of 4 polypeptide chains. Hemoglobin, which is composed of 2 α chains and 2 β chains, functions as a pair of $\alpha\beta$ dimers. The structure of the two identical α subunits (red) is similar to but not identical with that of the two identical β subunits (yellow). The molecule contains four heme groups (black with the iron atom shown in purple).

characteristic *quaternary structure*. The hemes are 2.5 nm apart from each other and tilted at different angles. Each heme is partly buried in a pocket lined with hydrophobic R groups. It is bound to its polypeptide chain through a coordination bond of the iron atom to the N^{δ} of a histidine residue. The sixth coordination bond of the iron atom of each heme binds a molecule of oxygen. Myoglobin and the α and β chains of hemoglobin have the same tertiary structure. Both have well over 70% α -helical nature, both have similar α -helical segments and the bends have about the same angles.

206 / 1254

Similarity in 3- $^{\circ}$ Structure of Hemoglobin and Myoglobin

The 3- $^{\circ}$ structures of myoglobin and the α and β chains of human hemoglobin are strikingly similar (Fig. 9-46). The 8 helices in each chain of hemoglobin are virtually superposable on those of myoglobin. This close resemblance in the folding of their main chains was unexpected because their amino acid sequences are rather different. In fact, these 3 chains are identical at only 24 or 141 positions.

The amino acid sequence of hemoglobins from more than 60 species, ranging from lamprey eels to humans, are known. A comparison of these sequences shows considerable variations at most positions. However, 9 positions have the same residues in most species studied so far (Table 9-7). These highly conserved 9 residues are especially important for the functioning of the hemoglobin molecule. For example, the invariant **F8 histidine** is directly bonded to the heme iron; proline **C2** may be essential because it defines one end of the C helix; and tyrosine **HC2** stabilizes the structure by forming a hydrogen bond between the H and F helices.

182 FUNDAMENTALS OF BIOCHEMISTRY

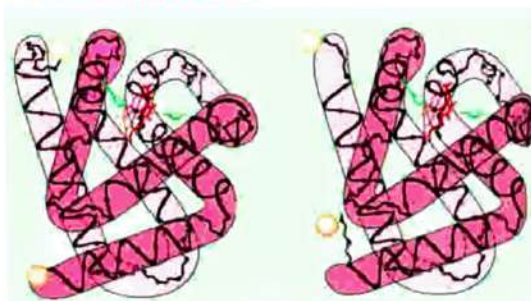


Fig. 9-39. Comparison of the conformation of the main chain of myoglobin and the β chain of hemoglobin. A_1

Proximal His F8, distal His E7 and Val E11 side chains are shown. Other amino acids of polypeptide chain are represented by α -carbon positions only; the letters M, V and P refer to the methyl, vinyl and propionate side chains of the heme. The overall structures are very similar, except at NH_2 -terminal and $COOH$ -terminal ends.

(Courtesy: Fersht A, 1977)

Table 9-8. The nine highly conserved (or invariant) amino acid residues in hemoglobins

Position	Amino acid	Role
F8	Histidine	Proximal heme-linked histidine
E7	Histidine	Distal histidine near the heme
CD1	Phenylalanine	Heme contact
F4	Leucine	Heme contact
B6	Glycine	Allows close approach between B and E helices
C2	Proline	Helix formation
HC2	Tyrosine	Cross-links the F and H helices
C4	Threonine	Uncertain
H10	Lysine	Uncertain

(Adapted from Lubert Stryer, 1993)

The amino acid residues in the interior of hemoglobin are strikingly nonpolar but vary considerably. However, the change is always of one nonpolar residue for another as from alanine to isoleucine. The nonpolar core functions in binding the heme group and in stabilizing the 3- $^{\circ}$ structure of each subunit. On the contrary, the surface amino acid residues are highly variable. Indeed, few are consistently positively or negatively charged.

Hemoglobin as an Allosteric Protein

The influence that the binding of an O_2 molecule to hemoglobin has on the subsequent oxygenation of the other heme groups of the molecule is an example of an interaction at one site on a protein affecting another site located in a distinctly different region of the same molecule. Such interactions are generally referred to as *allosteric interactions*. In addition to hemoglobin, many enzymes are allosteric proteins. A quaternary structure is usually characteristic of allosteric proteins.

Hemoglobin is a much more intricate and sentient molecule than is myoglobin. Hemoglobin, in addition to transporting oxygen, also transports CO_2 , a waste product of metabolism, to the lungs to be respired. The ability of hemoglobin to bind H^+ (another waste product of metabolism) is also an important physiological function of the macromolecule, since it is essential for the maintenance of physiological pH. Also, the oxygen-binding properties of hemoglobin are regulated by interactions between separate, nonadjacent sites. Hemoglobin is an allosteric protein, whereas myoglobin is not. This difference is expressed in 3 ways.

1. The binding of O_2 to hemoglobin enhances the binding of extra O_2 to the same hemoglobin molecule. In other words, *oxygen binds cooperatively to hemoglobin*. In contrast, the binding of O_2 to myoglobin is not cooperative.
2. The affinity of hemoglobin for O_2 depends on pH (i.e., pH-dependent), whereas that of myoglobin is independent of pH.
3. The oxygen affinity of hemoglobin is further regulated by organic phosphates such as 2,3-bisphosphoglycerate (BPG), whereas that of myoglobin is not. Thus, *hemoglobin has a lower affinity for oxygen than does myoglobin*.

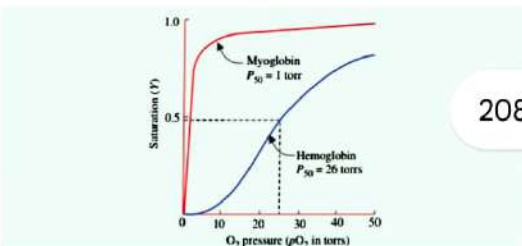


Fig. 9-47. Oxygen dissociation curves of myoglobin and hemoglobin in whole blood

Cooperativity in Oxygen-binding of Hemoglobin

The differences between hemoglobin and myoglobin as oxygen carriers are readily apparent when their oxygen dissociation curves are compared. An **oxygen dissociation curve** is a plot of the binding sites i.e., heme groups with oxygen (expressed as Y) versus the partial pressure of oxygen (expressed as pO_2 in torr*) to which the protein solution is exposed, i.e., $Y = [HbO_2] / [HbO_2] + [Hb]$. The oxygen dissociation curves of

Torr is a unit of pressure equal to that exerted by a column of mercury 1 mm high at 0°C and standard gravity; named after Evangelista Torricelli (I.T., 1608-1647), the inventor of the mercury barometer.

208 / 1254

Contents

184 FUNDAMENTALS OF BIOCHEMISTRY

myoglobin and hemoglobin (Fig. 9-47) differ in two ways:

- A. For any given pO_2 , Y is higher for myoglobin than for hemoglobin. This means that *myoglobin has a higher affinity for O_2 than does hemoglobin*. Oxygen affinity can be characterized by a quantity P_{50} , which is the partial pressure of oxygen at which 50% of sites are filled (i.e., at which $Y = 50$). For myoglobin, P_{50} is typically 1 torr, whereas for hemoglobin, P_{50} is 26 torr.
- B. Secondly, *the oxygen dissociation curve of myoglobin is hyperbolic, whereas that of hemoglobin is sigmoidal (S-shaped)*.

Because myoglobin has only a single heme group, the phenomenon of cooperative binding is not possible, as evidenced by its hyperbolic curve. It has been calculated that the cooperative binding of oxygen by hemoglobin enables it to deliver 1.83 times as much O_2 under typical physiological conditions as it would if the sites were independent.

The Bohr Effect : H^+ and CO_2 promote the release of O_2 .

Hemoglobin not only furnishes O_2 to tissues but also transports the waste products of metabolism ($H^+ + CO_2$). Since the same biomolecule is responsible for both transport systems, it is not surprising that there is a regulatory interplay between the two functions. It has been observed that increased concentrations of CO_2 and H^+ (i.e., lowering of pH) decrease the O_2 affinity of hemoglobin; conversely, increased concentrations of O_2 lower the affinity for CO_2 and H^+ (Fig. 9-48). By contrast, increased concentrations of CO_2 and H^+ have almost no effect on myoglobin's O_2 affinity. Increasing the concentrations of CO_2 (at constant pH) also lowers the O_2 affinity. In actively metabolizing tissue, such as contracting muscle, much CO_2 and acid are produced. The presence of higher levels of H^+ and CO_2 in the capillaries of such metabolically active tissue promotes the release of O_2 from oxyhemoglobin. This important mechanism for meeting the higher O_2 needs of metabolically active tissues was discovered by Christian Bohr in 1904. This metabolic interdependence of the two transport functions of hemoglobin is known as Bohr effect.

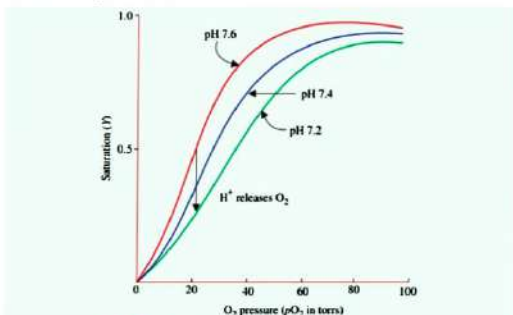


Fig. 9-48. Effect of pH on the oxygen affinity of hemoglobin

The three O_2 dissociation curves of hemoglobin at different pH have been shown. Lowering the pH from 7.5 to 7.2 results in the release of O_2 from oxyhemoglobin.

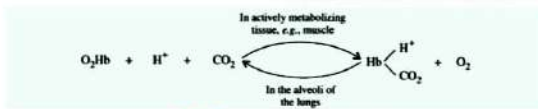
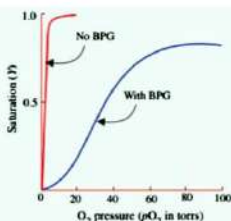


Fig. 9-49. Summary of the Bohr effect

J.B.S. Haldane, in 1914, discovered the reciprocal effect which occurs in the alveolar capillaries of the lungs. The high concentration of O_2 there unloads H^+ and CO_2 from hemoglobin, just as the high concentrations of H^+ and CO_2 in active tissues drives off O_2 (Fig. 9-49).

BPG as a Hemoglobin Regulator

Joseph Bancroft, as early as 1921, observed that the O_2 affinity of hemoglobin within erythrocytes is lower than that of hemoglobin in free solution. Much later in 1967, Reinhold Benesch and Ruth Benesch showed that an anionic organic phosphate present in human red cells, that is 2, 3-bisphosphoglycerate, BPG (also known as 2,3-diphosphoglycerate, DPG) binds to hemoglobin and thereby lowering the O_2 affinity of hemoglobin. In the absence of BPG the P_{50} of hemoglobin is 1 torr, like that of myoglobin. In its presence, P_{50} becomes 26 torr (Fig. 9-50). Thus, BPG lowers the O_2 affinity of hemoglobin by a factor of 26, which is essential in enabling hemoglobin to unload oxygen in tissue capillaries. BPG reduces the oxygen affinity of hemoglobin by binding in the central cavity of deoxyhemoglobin and not to the oxygenated form. As oxygenation occurs, the accompanying conformational changes in hemoglobin make the central cavity too small to accommodate BPG, which is then expelled.

Fig. 9-50. BPG regulation of O_2 affinity of hemoglobin

Note that 2,3-bisphosphoglycerate (BPG) decreases the oxygen affinity of hemoglobin molecule.

Oxygen Affinity of Fetal Hemoglobin

Hemoglobin F has a higher O_2 affinity under physiological conditions than does hemoglobin A (Fig. 9-51). The higher O_2 affinity of hemoglobin F optimizes the transfer of oxygen from the maternal to the fetal circulation. Hemoglobin F is oxygenated at the expense of hemoglobin A on the other side of the placental circulation. In fact, hemoglobin F does not bind BPG as strongly as does hemoglobin A because the γ chain (analogous to the β chain of hemoglobin A) has a seryl residue in the H21 position instead of a positively charged histidine. Because of this particular amino acid difference, hemoglobin F has a higher affinity for oxygen than does hemoglobin A. In the absence of BPG the O_2 affinity of hemoglobin F is actually lower than that of hemoglobin A. Thus, we see that different

210 / 1254

186 FUNDAMENTALS OF BIOCHEMISTRY

forms of a protein, called **isoforms** or **isotypes**, in different tissues have a clear-cut biological advantage, as beautifully illustrated by diverse forms of hemoglobin.

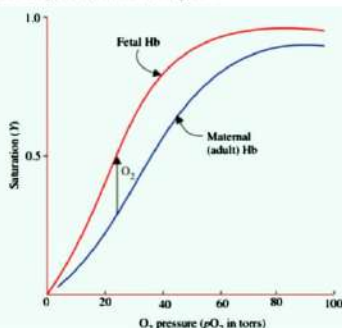


Fig. 9-51. Oxygen dissociation curves for fetal hemoglobin (HbF) and adult hemoglobin (HbA)

In the presence of BPG the O_2 affinity of fetal hemoglobin is higher than that of maternal hemoglobin. The arrow represents transfer of oxygen from maternal oxyhemoglobin to fetal deoxyhemoglobin.

Quaternary Structural Changes in Hemoglobin (= Oxygenation of Hemoglobin)

Hemoglobin can be dissociated into its constituent chains. The isolated α chain has a high O_2 affinity, a hyperbolic dissociation curve and O_2 -binding property which is indifferent to pH, CO_2 concentration and BPG level. β chains by themselves readily associate to form a tetramer (B_4). Like the α chain and myoglobin molecule, B_4 lacks the allosteric properties of hemoglobin and has a high oxygen affinity. In short, the allosteric properties of hemoglobin arise from interactions between its subunits.

Hemoglobin undergoes a major conformational change on binding oxygen, and as such oxy- and deoxyhemoglobin differ markedly in quaternary structure. In the quaternary structure of deoxyhemoglobin, there are 8 additional electrostatic interactions (salt linkages), not found in oxyhemoglobin, making deoxyhemoglobin the more rigid molecule of the two (Fig. 9-52). Six of these 8 interactions are between chains. The C-terminal residue of the 4 chains of deoxy-hemoglobin are also involved in salt linkages. As a consequence of oxygenation, a hemoglobin molecule undergoes conformational changes which disrupt 8 salt linkages and the cooperative binding observed with oxygen is the result of these structural changes.

In the oxyhemoglobin molecule, the distance between the iron atoms of the β chains decreases from 40 to 33 Å, thus making the molecule more compact. During the transition phase from oxy- to deoxyhemoglobin, large structural changes take place at two of the four contact regions (the $\alpha_1\beta_2$ contact and the identical $\alpha_2\beta_1$ contact) but not at the others (the $\alpha_1\beta_1$ contact and the identical $\alpha_2\beta_2$ contact). In fact, the $\alpha_1\beta_2$ contact region is designated to act as a switch between two alternative

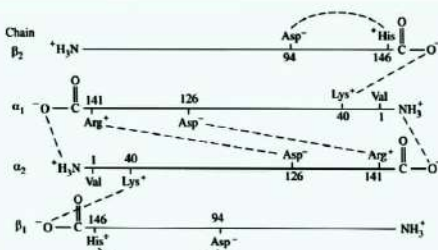


Fig. 9-52. Schematic representation of 8 electrostatic interactions (shown by dotted lines) that occur between α and β chains of deoxyhemoglobin

(Adapted from Frank B. Armstrong, 1989)

structures. In oxyhemoglobin, the C-terminal residues of all 4 chains have almost complete freedom of rotation. By contrast, in deoxyhemoglobin, these terminal groups are anchored. Deoxyhemoglobin is a tautomer, more constrained molecule than oxyhemoglobin because of the presence of additional salt links. The quaternary structure of deoxyhemoglobin is termed the T (tense or taut) form; that of oxyhemoglobin, the R (relaxed) form.

The above-mentioned conformational changes

take place at some distance from the heme. But changes also take place at the heme group on oxygenation. In deoxyhemoglobin, the iron atom is about 0.4 Å out of the porphyrin plane proximal histidine, so that the heme group domes (convex) in the same direction (F oxygenation, the iron atom moves into the plane of the porphyrin to form a strong bond, the heme becomes more planar. Besides, His F8, because of its bonding to the iron atom, toward the heme plane. This slight displacement of the histidine residue initiates a sequential series of conformational changes that ultimately results in the disruption of some of the subunit interactions in the quaternary structure of deoxyhemoglobin.

The designations **R** and **T** are generally used to describe alternative quaternary structures of an allosteric protein, the T form having a lower affinity for the substrate.

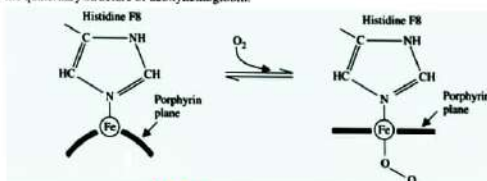


Fig. 9-53. Effect of oxygenation on heme group

The iron atom moves into the plane of the heme group on oxygenation. The proximal histidine (F8) is pulled along with the iron atom and becomes less tilted.

212 / 1254

188 FUNDAMENTALS OF BIOCHEMISTRY

Mechanism of Allosteric Interactions of Hemoglobin

The structures of fully deoxygenated and fully oxygenated hemoglobin provide insight into how the binding of O_2 , CO_2 , H^+ and BPG influence each other. How does allosteric mechanism take place? How does hemoglobin switch from the deoxy to the oxy structure when it binds successive O_2 molecules? Two models – the sequential model and the concerted model – have been proposed to explain the allosteric changes in multi-subunit protein. Both these models are based upon the ideas that the proteins are flexible in conformation and that a protein subunit can exist in only two states.

A. Sequential Model. Developed by Daniel E. Koshland Jr. in 1966, the sequential model, in its simplest form, makes 3 assumptions:

1. Only two conformational states, T and R, are accessible to any subunit.
2. The binding of the ligand switches the conformation of the subunit to which it is bound but not that of its neighbours.
3. The conformational change elicited by the binding of substrate in one subunit can increase or decrease the binding affinity of the other subunit in the same molecule. A T subunit with an R neighbour has higher affinity for ligand than does a T subunit with a T neighbour because the subunit interfaces are different in TR and TT.

The sequential model (Fig. 9-54), in its essence, proposes that *symmetry is not conserved in allosteric transitions and that the subunits change conformation one at a time*. The T state is symbolized by a square, and the R state by a circle. Deoxyhemoglobin is in the T_4 state. The binding

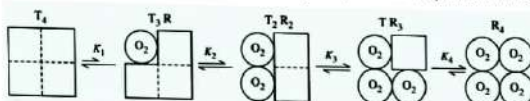


Fig. 9-54. Simple sequential model for a tetrameric allosteric protein

The binding of a ligand to a subunit changes the conformation of that particular subunit from the T (square) to the R (circle) form. This transition affects the affinity of the other subunits for the ligand.

of O_2 to one of the subunits changes its conformation from T to R, but leaves the other subunits in the T form. The oxygen-binding affinity of unoccupied sites in RT_3 is higher than in T_4 because some salt links have been broken on binding the first O_2 . R_2T_2 and RT_3 , which have higher oxygen affinities than does RT_2 , are formed when the second and third O_2 bind respectively. Finally, R_4 is produced on binding the fourth O_2 . Thus, according to the sequential model, the affinity of hemoglobin for successive oxygen molecules increases because fewer salt bridges need be broken. The model can be better understood in terms of a *postage-stamp analogy* (refer Fig. 9-55). Two perforated edges (i.e., sets of salt bridges) must be torn to remove the first stamp. Only one perforated edge must be torn to remove the second stamp, and one edge again to remove the third stamp. The fourth stamp is then free.



Fig. 9-55. Postage-stamp analogy of the simple sequential model

B. Concerted Model (or Symmetry Model). Developed by Jacques Monod, Jeffries Wyman and Jean-Pierre Changeux (hence, also called **MWC model**) in 1965, the concerted model takes a different view of allosteric interactions. The essence of this elegant and incisive model (Fig. 9–56) is that *symmetry is conserved in allosteric transitions* and that *all the subunits change conformation together*. This model also is based on 3 *assumptions*:

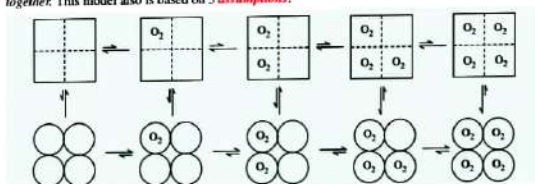


Fig. 9–56. Concerted model (or MWC model) for a tetrameric allosteric protein

The squares denote the T form, and the circles denote the R form.

1. The protein interconverts between two conformations, T and R. All the subunits of a particular molecule must be in the T form, or all must be in the R form. Hybrids such as TR are forbidden.
2. Ligands bind with low affinity to the T form, and with high affinity to the R form.
3. The binding of each ligand increases the probability that all subunits in that molecule are in the R form. The allosteric transition is said to be *concerted* because subunits change in unison from T to R or vice versa. Stated in terms of the postage-stamp analogy, either all four perforated edges or none are broken.

It is, thus, apparent that the two models offer contrasting views regarding the mode of cooperative interactions in multi-subunit protein. The models differ in many points through:

1. The symmetry model assumes that the active and inactive states are in *equilibrium*; sequential model assumes that the transition from inactive state to active state *214 / 1254*
2. The symmetry model states that if one subunit changes its form, the other subunit changes to that form also because, unless symmetry is preserved, the two subunits cannot interact. By contrast, the sequential model allows interaction between subunits in *different* states.
3. The symmetry model assumes that binding of the first molecule necessarily enhances binding of a second molecule of the same type, whereas the sequential model allows *either an increase or decrease* of affinity for the second molecule.
4. The symmetry model has few intermediate states, whereas the sequential model has *more* intermediate states.

In fact, the actual allosteric mechanism of hemoglobin is more complex than visualized by either the sequential or the concerted model. These models should be regarded as limiting cases. Actual allosteric processes combine, of course in varying degrees, elements of both.

The *protein-protein interactions*, thus, serve many important functions :

- (a) protect nucleic acids of virus particles from destruction,
- (b) conserve genetic information, and
- (c) construct the various enzyme-complexes found in the cells.

The foregoing discussion and also the evidences presently available confirm the view that "the primary structure of the protein, which dictates the limitations of secondary and tertiary structure, also delineates the correct quaternary structure necessary for biological activity of the molecule." (Hartman and Suskind, 1969).

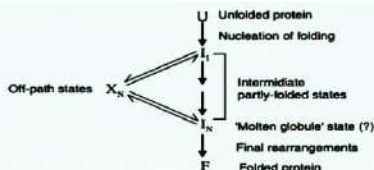
190 FUNDAMENTALS OF BIOCHEMISTRY

DYNAMICS OF GLOBULAR PROTEIN STRUCTURE

The precise pictures produced by x-ray diffraction give the impression of a rigid and inflexible structure for the globular proteins. Recent researches, however, indicate that globular proteins fold via complex kinetic pathways and that even the folded structure, once acquired, is a dynamic structure. Some of the dynamic aspects of globular proteins are described below :

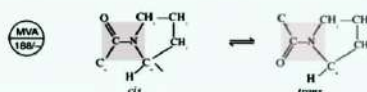
1. Kinetics of Protein Folding

The folding of globular proteins from their denatured states (or conformations) is a remarkably rapid process and often completes in less than a second. This viewpoint was first expressed in 'Levinthal's paradox', first propounded by Cyrus Levinthal in 1968. A rough estimate reveals that about 10^{30} conformations are possible for a polypeptide chain such as ribonuclease (124 residues). And if the molecule tries a new conformation every 10^{-11} second, it would still take about 10^{18} years to try a significant fraction of them. In actuality however, ribonuclease folds experimentally *in vitro* in about 1 minute. To explain this anomaly, it has been suggested that folding takes place through a series of intermediate states. These observations led to the **path way model of folding**, as depicted below :



The nucleation step is critical because it is much more difficult to begin an α helix than to extend it (note that at least 4 residues must fold properly to make the first stabilizing H bond). On the contrary, it is now recognized that nucleation may begin at many places and that all of these partially folded structures will be 'funneled' by energy minimizations toward the final state. The funneling model proposes that there is not just one but *many* possible paths from the denatured state to the folded state, and each path leads downhill in energy. During the descent toward the free energy minimum, there may be pauses corresponding to metastable intermediates, as incorporated in the pathway model. An important folding intermediary for many proteins appears to be what is called as the "molten globule" state. It is a compact structure in which much of the secondary and tertiary folding has occurred, but the internal hydrophobic residues have not yet settled into their final packing.

Evidence suggest for the presence of "off-path" states— those in which some key element is incorrectly folded. However, the cell has ways to assist incorrectly-folded proteins to find their proper conformations. Larger complex proteins may never achieve 100% folding success *in vivo* but the cell can identify those which are incorrect and dispose of them. One such common folding error occurs via the incorrect *cis-trans* isomerization of the amide bond adjacent to a proline residue :



Unlike other peptide bonds in proteins, for which the *trans* isomer is highly favoured (by a factor of

about 1,000), proline residues favour the *trans* form in the preceding bond by a factor of only 4. Hence, there are more chances that the 'wrong' isomer (in terms of the functional protein conformation) will form first; this will later be converted to the correct configuration involving chain rearrangement enzymatically to speed up *in vivo* folding.

2. Kinetics of Disulfide Bond Formation

If a protein is folded from a state in which disulfide bonds have been disrupted and must hence be re-formed, the process is more complicated and slower—often taking many minutes. Some disulfide bonds that are missing in the native structure are formed in intermediate stages of the folding. Obviously, the protein can utilize a number of alternative pathways to fold but ultimately finds both its proper tertiary structure and the correct set of disulfide bonds. This process is aided *in vivo* by enzymatic catalysis of —S—S— bond rearrangement.

3. Chaperonins

It has been discovered that protein folding and assembly *in vivo* sometimes requires the aid of special proteins called *chaperonins* or *molecular chaperones*. As the name signifies, the function of these chaperones is to keep the newly-formed protein away from improper folding or aggregation. Improper folding may correspond to being trapped in a deep local minimum on the energy scenario. Aggregation is often a danger because the protein, released from the ribosome in an unfolded state, will have groups exposed. These will be tucked inside in normal folding, but when exposed they stand the chance of making hydrophobic interactions with other polypeptide strands and thereby aggregating.

Many chaperonin systems have been discovered but the GroEL-ES complex from *E. coli* is the best-studied one. The structure of this enormous complex has recently been elucidated by x-ray diffraction [Fig. 9-57(a)]. It consists of 2 basic portions—GroEL and GroES. GroEL is made up of 2 rings, each consisting of 7 protein molecules; the centre of each ring is an open cavity, accessible to the solvent at the ends. Either cavity can be 'capped' with GroES, which is again a 7-

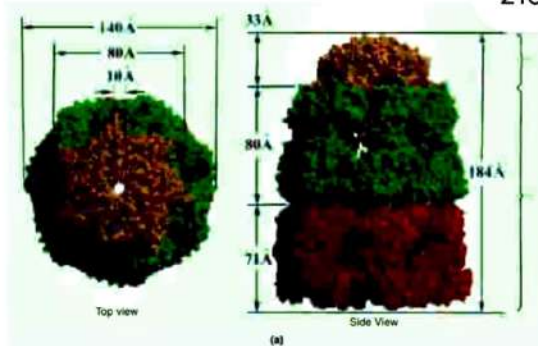


Fig. 9-57. The GroEL-GroES chaperonin complex

(a) X-ray diffraction structure of the EL-ES (ADP) complex.

(b) A schematic of the function of GroEL-ES complex.

The unfolded protein enters into a sort of cavity lined with hydrophobic residues. Then the cavity changes, presenting a hydrophilic lining. This releases the protein from the walls and it folds and is then released. Note that ATP is required, probably to drive the process in one direction.

(Adapted from (a) Xu Z, Horwich AL and Sigler PB, 1997 and (b) Netzer and Hartl, 1998)

It is argued that basically, the cavities provide 'shelters' in which nascent protein chains can be 'incubated' until they have folded properly. The GroEL-ES complex does not stipulate the folding pattern *i.e.*, upto the protein itself to do. But insulation from the environment prevents chances of aggregation or misfolding. The cycle experienced by a protein molecule is schematically shown in [Fig. 9-57(b).]

The conformational changes in GroEL are diagrammatically represented in (Fig 9-58).

The reconstructions of the GroEL and GroEL-GroES complexes, based on high resolution electron micrographs, have been presented in (Fig. 9-59).

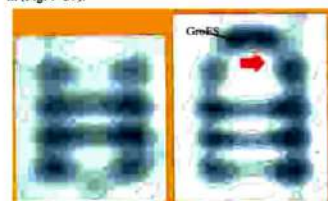


Fig. 9-59. GroEL-GroES complexes

Reconstructions of the GroEL and GroEL-GroES complexes, based on high resolution electron micrographs, taken of specimens that had been frozen in liquid ethane and examined at -178°C . The GroEL complex with GroES appears as a dome on one end of the cylinder. It is evident that the binding of the GroES is accompanied by a marked change in conformation of the apical end of the proteins that make up the top GroEL ring (arrow), which results in a marked enlargement of the upper chamber.

(Adapted from Chens *et al.*, 1994)

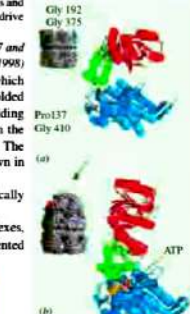


Fig. 9-58. Conformational change in GroEL

(a) The drawing on the left shows a surface view of the two rings that make up the GroEL chaperonin. The drawing on the right shows the tertiary structure of one of the subunits of the top GroEL ring. The polypeptide chain can be seen to fold into three domains.

(b) When a GroES ring (arrow) binds to the GroEL cylinder, the apical domain of each GroEL subunit of the adjacent ring undergoes a dramatic rotation of approximately 60° with the intermediate domain (shown in green) acting like a hinge. The effect of this shift in parts of the polypeptide is a marked elevation of the GroEL wall and enlargement of the enclosed chamber.

(Adapted from Xu Z, Horwich AL and Sigler PB, 1997)

However, processing *via* chaperonins is not a universal phenomenon as only a small fraction of the proteins made in *E. coli* (or any cell) are processed *via* chaperonins. Some proteins are too large to be accommodated within the cavity whereas others fold safely on their own.

4. Motion within Globular Proteins

Evidences accumulated indicate that the folded globular protein molecules are not static and they continually undergo a wide variety of internal motions. These motions are due to the interactions of protein molecules with their environment. The resulting motions can be roughly grouped into 3 classes (Table 9-9).

Table 9-9. Motions within globular protein molecules

Class	Type of Motion	Approximate Amplitude (nm)	Range Time(s)
1	Vibrations and oscillations of individual atoms and groups	0.2	10^{-15} – 10^{-12}
2	Constructed motions of structural elements, like α helices and groups of residues	0.2–1	10^{-12} – 10^{-8}
3	Motions of whole domains; opening and closing of clefts	1–10	$\geq 10^{-8}$

Class 1 motions occur even within protein molecules in crystals and account, in part, for the limits of resolution obtainable in x-ray diffraction studies. In **class 2** and **class 3**, the motions are larger in magnitude and slower in rate. These are more likely to occur in solution. Some of the motions, like the opening and closing of clefts in molecules, are probably involved in the enzymatic functions. Binding or release of a small molecule from a protein depends on the time required to open or close a shaft. Likewise, the protein 'gates' that pass molecules and ions through membranes *rap* from open to closed states.

5. Prions

Until very recently, scientists believed that the diseases could be transmitted from one to another *via* viruses or microorganisms. After all, DNA and RNA were the carriers of genetic information. But now evidences have gathered to point out that some diseases are transmitted by a protein and nothing more. Table 9-10 lists some such diseases (called **prion diseases**) along with their host and geographic distribution.

In 1900, a fatal neurodegenerative disease called **Kuru** was reported only in a tribe of Papua New Guinea. Carleton Gajdusek, a virologist at the US National Institute of Health, discovered and understood the basis of the disease and the mode of its transmission. He pointed out that a peculiar ritual of the tribal people of eating the brain of dead relatives was the main mode of transmission of this disease. The frequency of Kuru gradually declined since the practice was abandoned. Gajdusek was awarded 1976 Nobel prize in medicine or physiology for this basic work, along with Baruch Blumberg, a virologist of the Institute of Cancer Research, Philadelphia.

But Stanley B. Prusiner, an American biochemist, was not convinced about the nature of the infectious agent which was described by Gajdusek and others as 'unconventional slow virus'. However, Prusiner was against the so-called virus theory because of the following 3 reasons:

- The infective agent is extremely resistant to UV and ionizing radiations whereas all viruses are sensitive.
- The infectivity of the scrapie agent is not changed by treatment with nucleases (DNAase and RNAase enzymes). That means the infective agent does not carry DNA or RNA as

218 / 1254

194 FUNDAMENTALS OF BIOCHEMISTRY

Table 9-10. Prion disease in different species

Host	Disease	Geographic distribution	First clinical observation
Sheep	Scrapie	Cosmopolitan except Australia, New Zealand and some European countries	1730
Gout	Scrap	—	—
Man	Kuru ('Laughing death') Creutzfeldt-Jacob disease Gertsmann-Straussler-Scheinker syndrome Fatal familial insomnia	Papua New Guinea Cosmopolitan Cosmopolitan	1900 1920 1926
Mink	Transmissible mink encephalopathy	North America; Europe	1947
Mule deer	Chronic wasting disease	North America	1967
Cattle	Bovine spongiform encephalopathy, BSE (= Mad cow disease)	U.K., Ireland and some other European countries	1985

genetic material which is the most essential component of a virus particle.

- Absence of any agent-specific antibody titre, strong tendency to aggregate and hydrophobicity go against the nature of a true virus.

Backed by these evidences, Prusiner and his colleagues in 1982, extracted the infectious material from hamster brains, and they gave the first blow to the virus theory. Prusiner suggested that the infectious agent, causing certain degenerative disorders of central nervous system (CNS) in animals and more rarely in humans, is a small proteinaceous infectious particle, which he called **prion** (pronounced as 'preon') and the protein believed to be responsible for infection was called **prion-related protein** or **PrP**. Soon he found that there were 2 isoforms of PrP. The first one is the normal cellular **prion-related protein**, **PrP^C** (the superscript 'c' denotes cellular), which is the nonpathological form; the PrP^C may act as an acetyl-choline receptor inducer and plays an important role in the transmission of nerve signal. The other isoform of PrP is infective and called **prion-related protein scrapie**, **PrP^{Sc}** (the superscript 'sc' denotes scrapie, which is now used to refer to all infectious form of prion causing scrapie-like diseases in animals and humans). It is this form, in which the disordered N-terminal portion appears to fold into a β -sheet, that wrecks havoc with the nervous system. It is postulated that when PrP^{Sc} is ingested into the body, it induces the conversion of PrP^C in the recipient to PrP^{Sc}; thus the disease is transmitted. It is something like a bad guy, who converts a good guy into one of its 'bad' form. How this conversion is catalyzed is unknown, but it strongly suggests that PrP^C represents an especially stable "off-path" folding of the type hypothesized in the preceding section. Fig. 9-60 presents the computer-generated images of the two forms of human prion protein.

In 1988, Prusiner and his team reported that human prion diseases can certainly be inherited, i.e., they could be heritable and communicable. Later studies on transmission of human prion to transgenic mice threw more light on prion diseases in 1994. When researchers tried to infect transgenic mice (carrying the human PrP gene) with human prion, there was no development of CNS dysfunction as expected. The mice could become susceptible to human prions only after removal of mouse PrP gene which is called '**gene knockout**' in genetics terminology.

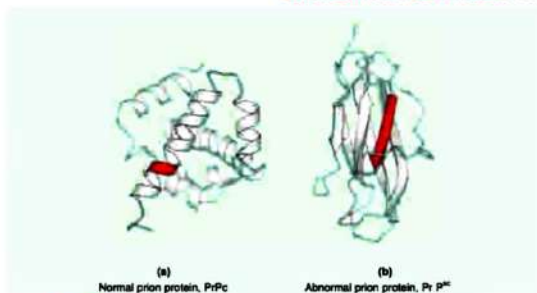


Fig. 9-60. Computer-generated images of the two forms of human prion protein.

Unfortunately, prion diseases have already played havoc among cattle population in some European countries. **Bovine spongiform encephalopathy (BSE)**, or more commonly known as **mad cow disease**, took its toll in cattle population in Britain and elsewhere. Epidemiological studies revealed that the most probable cause of the epidemic was the inclusion of ruminant-derived meal and bone meal (which utilizes sheep brain, spinal cord etc.) in the cattle feed. Things became worse as reports of 12 persons afflicted with Creutzfeldt-Jakob disease came in. It was reported that these new cases occurred due to consumption of BSE-infected beef. These reports sent a shock wave throughout the world, apprehending the possible impact of prion diseases.

The recognition of the relationship of PrP to diseases won Prusiner the Nobel prize in physiology or medicine in 1997. His discovery of a new biological principle of infection goes by the name '**prion theory**'.

PREDICTION OF SECONDARY AND TERTIARY PROTEIN STRUCTURES

A. Prediction of Secondary Structure

The protein amino acids arrange in innumerable ways, with the help of a variety of bonds, to produce a definite pattern of secondary structure in proteins, which may be either an α helix, or a β sheet or a turn. Acquisition of any one, of these 3 forms (α helix, β sheet, turn) by a protein depends upon the frequency of occurrence of particular amino acid residues in these secondary structures. Table 9-11, lists the relative frequencies (P_{α} , P_{β} , P_t) of amino acids in producing secondary structures, i.e., α helix, β sheet and turn, respectively.

A perusal of the table (on the next page) reveals that Met, Glu, Ala, and Leu residues tend to be present in α helices, whereas Val, Ile, Phe and Tyr tend to be present in β strands. Pro, Gly, Asp, and Ser have a propensity (or inclination) to lie in turns.

220 / 1254

Contents

196 FUNDAMENTALS OF BIOCHEMISTRY

Table 9-11. Relative frequencies of amino acid residue occurrence in secondary structures of different globular proteins.

Amino acid	α helix (P_{α})	β sheet (P_{β})	Turn (P_t)	
Alanine (Ala)	1.29	0.90	0.78	Favour α -helices
Cysteine (Cys)	1.11	0.74	0.80	
Leucine (Leu)	1.30	1.02	0.59	
Methionine (Met)	1.47	0.97	0.39	
Glutamic acid (Glu)	1.44	0.75	1.00	
Glutamine (Gln)	1.27	0.80	0.97	
Histidine (His)	1.22	1.08	0.69	
Lysine (Lys)	1.23	0.77	0.96	Favour β sheets
Valine (Val)	0.91	1.49	0.47	
Isoleucine (Ile)	0.97	1.45	0.51	
Phenylalanine (Phe)	1.07	1.32	0.58	
Tyrosine (Tyr)	0.72	1.25	1.05	
Tryptophan (Trp)	0.99	1.14	0.75	
Threonine (Thr)	0.82	1.21	1.03	
Glycine (Gly)	0.56	0.92	1.64	Favour turns
Serine (Ser)	0.82	0.95	1.33	
Aspartic acid (Asp)	1.04	0.72	1.41	
Asparagine (Asn)	0.90	0.76	1.28	
Proline (Pro)	0.52	0.64	1.91	
Arginine (Arg)	0.96	0.99	0.88	

* Note that arginine shows no significant preference for any of the structures.

(Adapted from Creighton TE, 1992)

There are some obvious reasons for these preferences :

- The α helix can be regarded as the default conformation. Branching at the β carbon atom (as in **Val, Thr and Ile**) tends to destabilize α helices because of steric clashes. These residues are easily accommodated in β strands, where their side chains project out of the plain containing the main chain.
- Ser, Asp and Asn** tend to disintegrate α helices because their side chains contain H bond donors or acceptors in close proximity to the main chain, where they compete for the main-chain NH and CO groups.
- Pro** tends to disrupt both α helices and β strands because it lacks an NH group and because its ring structure restricts its ϕ value to near -60 degrees.
- Gly** readily fits into all structures and hence does not favour helix formation, in particular.

It is worth mentioning that the conformational preferences of amino acid residues are not topped over all the way to one structure, as seen in Table 9-11. As an instance, glutamic acid, one of the strongest helix formers, prefers α helix to β strand by only a factor of (1.44/0.75) about 2. The preference ratios of most other amino acid residues are smaller; for example, for methionine it is

(1.47/0.97) about 1.5 and for alanine it is (1.29/0.90) about 1.25. Indeed, some penta- and hexapeptide sequences have been found to adopt one structure in one protein and an entirely different structure in another protein (refer Fig. 9-51). Hence, some amino acid sequences do not uniquely determine secondary structure of proteins. Tertiary interactions between residues that are far apart in the sequence may be decisive in specifying the secondary structure of some segments.

It is now possible to predict the protein secondary structure with moderate accuracy. Based on the relative frequency values of the constituent amino acids, P.Y. Chou and G.D. Fasman (1974) have framed certain rules for the prediction of globular protein secondary structures. Following are the 3

Chou-Fasman rules for prediction :

1. Any segment of 6 residues or more, with $(P_{\alpha}) \geq 1.03$, as well as $(P_{\alpha}) > (P_{\beta})$, and not including Pro, is predicted to be α helix.
2. Any segment of 5 residues or more, with $(P_{\beta}) \geq 1.05$, and $(P_{\beta}) > P_{\alpha}$, is predicted to be β sheet.
3. Examine the sequence for tetrapeptides with $(P_{\alpha}) < 0.9$, $(P_{\beta}) > (P_{\alpha})$. They have a good chance of being turns. The actual rules for predicting β turns are more complex, but this method will work in most cases.

B. Prediction of Tertiary Structure

Prediction of tertiary structure of proteins is much more difficult because the higher-order folding depends so critically on specific side chain interactions, often between residues far removed from one another in the sequence. However, recent recognition of overall patterns in tertiary folding has given some success. For example, predictions of secondary structure have been used to predict an α/β barrel structure for the enzyme **tryptophan synthase**, that is in excellent harmony with x-ray result.

The method for predicting the tertiary structure of proteins depends on the fact that, in their spontaneous folding, proteins are seeking a free energy minimum. A random-coil chain is allowed, in computer simulation, to undergo a large number of small permutations in its configuration, through rotation about individual bonds. The computer programme keeps track of the total energy, in terms of possible interactions, and seeks an energy minimum. This approach is still in its infancy.

CLEAVAGE OF A PROTEIN

Certain proteolytic enzymes and chemical reagents cleave protein between specific residues (Table 9-12). The enzyme **trypsin**, for example, cuts on the carboxyl side of arginine, whereas the chemical **cyanogen bromide** splits peptide bonds next to methionine. Since these enzymes and chemicals cleave at relatively few sites in a protein molecule, they ~~can~~ create relatively large and relatively few peptides. If such a mixture of peptides is separated by biochemical procedures (chromatography, electrophoresis etc), a characteristic pattern or **peptide map** would be obtained which will be diagnostic of the protein from which the peptides were generated. This peptide map is, sometimes, referred to as the **protein's fingerprint**.

Table 9-12: Some reagents frequently used to cleave peptide bonds in proteins*

Reagent	Amino acid 1	Amino acid 2
Enzymes		
Trypsin	Lys or Arg	Any
Chymotrypsin	Phe, Trp or Tyr	Any
V8 protease	Glu	Any
Chemicals		
Cyanogen bromide	Met	Any
2-nitro-5-thiocyanobenzoate	Any	Cys

* The specificity for the amino acids on either side of the cleaved bond is indicated. The carboxyl group of amino acid 1 is released by the cleavage; this amino acid is to the left of the peptide bond as normally written.

222 / 1254

BIOLOGICAL ROLES OF PROTEINS

Proteins are of utmost significance to biological systems. These are most critical to life and perform various functions. Some of their roles are given below.

1. Many proteins act as **catalysts**, thus usually enhancing the rate of chemical reactions to such extents as needed by the living cells.

2. The fibrous proteins serve as **components of the tissues** holding the skeletal elements together. Collagen is a structural unit of connective tissues.

3. The nucleoproteins serve as **carriers of genetic** characters and hence govern inheritance of traits.

4. Proteins also perform **transport functions**. Many compounds enter the cells and accumulate inside at much higher concentrations than expected from diffusion alone. These changes require the input of energy and are usually termed active transport. The mechanism of active transport involves proteins either as catalysts or as adsorbents or as both.

5. Various protein **hormones** are known. These regulate the growth of plants and animals, besides controlling many other physiological functions.

6. Under conditions of non-digestion and no chances for denaturation, the proteins accumulate inside the cells and produce **toxicity**. Venoms of snakes and insects are injected by biting into the blood. Certain of the foreign proteins present in venom are actually enzymes. These enzymes attack body tissue causing destruction of blood cells leading ultimately to death.

In other cases, the toxicity from foreign proteins results from responses by the affected animals. Allergic reactions exemplify this category. These reactions occur when an animal is exposed to a foreign protein to which it has been sensitized by prior exposure. The toxic disorders include skin blisters, swelling of limbs and respiratory congestion, leading sometimes to death.

7. Blood plasma, which is obtained after removal of the blood cells by centrifugal action, is essentially a solution of proteins in water. It is used for the **treatment of shock** produced by serious injuries and operations.

8. **Interferon (IF or IFN)** is a generic term which applies to a number of (about over 20) related low molecular weight, regulatory glycoproteins produced by many eukaryotic cells in response to numerous inducers: a virus infection, double-stranded RNA, endotoxins, antigenic stimuli, mitogenic agents, and many parasitic organisms capable of intracellular growth (*Listeria monocytogenes*, chlamydiae, rickettsias, protozoa). They are effective in treating viral diseases and cancer, and in eliminating its side effects. The most widely-studied property of interferons is their ability to 'interfere' (hence, their nomenclature) with the replication of viruses. They are usually species-specific but virus-nonspecific. Interferon was discovered in 1957 in London by the British virologist Alick Issacs and a visiting scientist from Switzerland Jean Lindenmann, both of the National Institute for Medical Research, London. They found an agent responsible for such viral interference; a protein released by cells exposed to a virus, that enables other cells to resist viral infection. They called it interferon (IF). Since then, several different classes of interferons have been identified. Human interferons have been classified into 3 classes, depending on which type of they are produced by:

- (a) Fibroblasts (F) or alpha interferons (IFN- α)
- (b) Leukocytes (L) or beta interferons (IFN- β)
- (c) Immune type (I) or gamma interferons (IFN- γ)

In 1978, a Tokyo metropolitan medical team has cultivated this "wonder substance" using cells from the placenta taken at the time of birth (or parturition). Large-scale production of this protein is possible because the human cells reproduce in tissue culture. Because of their antiviral and antiproliferative activities, interferons have been considered important therapeutically in treating viral

diseases such as hepatitis, encephalitis, cancer and even common cold. The IFs may not be a magic bullet like penicillin against all viruses, yet it may be a very useful addition to the armamentarium already available against viruses in general, and cancer in particular.

9. Peptides from humans called *defensins* have been found to be **antibiotic** in nature. Produced by the immune system, these cells smother and kill the invading pathogens. They are secreted by the epithelial cells lining the moist body surface of mammals and serve as the body's own disinfectants.

10. Another group of peptides called *endorphins* are found in the brain and are involved in the **suppression of pain**, creation of euphoric highs and feelings of joy. Candace Pert, a codiscoverer of the body's opiates, says "They reveal the harmony as well as the economy inherent in nature. Nothing is wasted". She adds "Once a molecule is found effective, it tends to be used over and over again in the ladder of life, in newer and ever more exciting configurations."

11. The frog's secretion *adenoregulin*, a 33-amino-acid-long peptide, works on the receptors in the brain which handle adenosine. Adenosine is a fundamental component in all human cell fuel. And the frog peptide seems to **enhance the binding of adenosine** by subtly altering the receptor in the brain.

12. The American scientists at the Chicago Medical School and the Harvard Medical School, in 1982, have isolated about 30 g of a glycopeptide, known as *S factor*, from 4.5 tones of urine from healthy males. The *S factor* is composed of alanine, glutamic acid, diaminopimelic acid, and muramic acid. The *S factor* acts as a **soporific**, i.e., as a sleep-promoting agent and has little side effects. The researchers infused a very light concentration of the *S factor* into the brain of the rabbit and it was found that it induced a 50% increase in what is known as *slow wave sleep*—a deep dream-free sleep that occurs in animals and humans after sleep deprivation and is normal as judged by various criteria.

REFERENCES

See list following Chapter 11.

PROBLEMS

- (a) Tropomyosin, a 70-kd muscle protein, is a two-stranded α -helical coiled coil. What is the length of the molecule? (b) Suppose that a 40-residue segment of a protein (a' stranded antiparallel β structure with a 4-residue hairpin turn. What is the long of this motif?
- Glycine is a highly conserved amino acid residue in the evolution of proteins. Why?
- Identify the groups in protein that can form hydrogen bonds or electrostatic bonds with an arginine side chain at pH 7.
- The shape of hair is determined in part by the pattern of disulfide bonds in keratin, its major protein. How can curls be induced?
- Proteins are quite stable. The lifetime of a peptide bond in aqueous solution is nearly 1000 years. However, the ΔG° of hydrolysis of proteins is negative and quite large. How can you account for the stability of the peptide bond in light of the fact that hydrolysis releases much energy?
- For an amino acid such as alanine, the major species in solution at pH 7 is the zwitterionic form. Assume a pK_a value of 8 for the amino group and a pK_a value of 3 for the carboxylic acid and estimate the ratio of the concentration of neutral amino acid species (with the carboxylic acid protonated and the amino group neutral) to that of the zwitterionic species at pH 7.
- All L amino acids have an *S* absolute configuration except L-cysteine, which has the *R*

224 / 1254

200 FUNDAMENTALS OF BIOCHEMISTRY

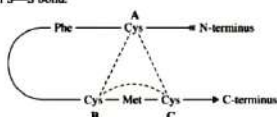
configuration. Explain why L-cysteine is designated as the *R* absolute configuration.

- Translate the following amino acid sequence into one-letter code: Leu-Glu-Ala-Arg-Asn-Ile-Asn-Gly-Ser-Cys-Ile-Glu-Cys-Glu-Ile-Ser-Gly-Arg-Glu-Ala-Thr.
- Would you expect Pro-X peptide bonds to tend to have *cis* conformations like those of X-Pro bonds? Why or why not?
- A protein was purified to homogeneity. Determination of the molecular weight by molecular exclusion chromatography yields 60 kd. Chromatography in the presence of 6 M urea yields a 30-kd species. When the chromatography is repeated in the presence of 6 M urea and 10 mM β -mercaptoethanol, a single molecular species of 15 kd results. Describe the structure of the molecule.
- The three-dimensional structure of biomolecules is more conserved evolutionarily than is sequence. Why is this the case?
- The sequences of three proteins (A, B, and C) are compared with one another, yielding the following levels of identity:

	A	B	C
A	100%	65%	15%
B	65%	100%	55%
C	15%	55%	100%

Assume that the sequence matches are distributed relatively uniformly along each aligned sequence pair. Would you expect protein A and protein C to have similar three-dimensional structures? Explain.

- Would a homopolymer of alanine be more likely to form an α helix in water or in a hydrophobic medium? Explain.
- Using the data in Table 9-3, calculate the *average* amino acid residue weight in a protein of typical composition. This is a useful number to know for approximate calculations.
- The melanocyte-stimulating peptide hormone α -melanotropin has the following sequence: Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val
(a) Write the sequence using the one-letter abbreviations.
(b) Calculate the molecular weight of α -melanotropin, using data in Table 9-3. Why is this result not *exactly* correct at neutral pH?
- A protein has been sequenced after destruction of —S—S— bonds. It is known to contain 3 Cys residues, located as shown below. However, only one of these is a free —SH; two are involved in an S—S bond.



The only methionine and the only aromatic amino acid (Phe) in this protein are in the positions indicated. Cleavage of the *intact* protein (with —S—S— bridge intact) by either cyanogen bromide or chymotrypsin does not break the protein into two peptides. Where is the —S—S— bridge (AB, BC, or AC)?

- Apanine* is a small protein toxin present in the venom of the honeybee. It has the sequence

- Classification based on the source of protein molecule
- Classification based on the shape of protein molecule
- Classification based on composition and solubility
- Classification based on biological function

Proteins Classification



THE HORSE, LIKE ALL ANIMALS, IS POWERED BY THE MOLECULAR MOTOR PROTEIN, MYOSIN

A portion of myosin moves dramatically (as shown above) in response to ATP binding, hydrolysis, and release, propelling myosin along an actin filament. This molecular movement is translated into movement of the entire animal, excitingly depicted in da Vinci's rearing horse. (Left) Leonardo da Vinci's "Study of a rearing horse" for the battle of Anghiari (c. 1504) from the Royal Collection © Her Royal Majesty Queen Elizabeth II.)

The proteins are all remarkably similar in structure insofar as they contain amino acids. But as little is known so far regarding their structure, a classification based totally on this criterion is not possible. However, a few systems based on one or the other criterion are given below.

CLASSIFICATION BASED ON THE SOURCE OF PROTEIN MOLECULE

Since long, the proteins have been traditionally divided into two well-defined groups: animal proteins and plant proteins. **Animal proteins** are the proteins derived from animal sources such as eggs, milk, meat and fish usually called *higher-quality proteins* because (and hence supply) adequate amounts of all amino acids. On the other hand, **plant proteins** are called *lower-quality proteins* since they have a low content (limiting amount) of one or more of the essential amino acids. The four most common limiting amino acids are methionine, lysine, threonine and tryptophan (Table 10-1). Although plant proteins have limiting amounts of some (but not all) amino acids, it should not be construed that they are poor protein sources.

229 / 1254

[Contents](#)

CLASSIFICATION OF PROTEINS

205

Table 10-1. Limiting amino acids in some plant proteins

Food	Amino acid(s)
Cereal grains and millets	Lysine, Threonine
Rice and soybeans	Methionine
Legumes (peas and beans)	Methionine, Tryptophan
Groundnuts	Methionine, Lysine, Threonine
Sunflower seeds	Lysine
Green leafy vegetables	Methionine

CLASSIFICATION BASED ON THE SHAPE OF PROTEIN MOLECULE

On the basis of the shape of protein molecule, the proteins have been grouped under two categories: globular and fibrous.

1. Globular or Corpuscular Proteins.

These have an axial ratio (length : width) of less than 10 (usually not over 3 or 4) and, henceforth, possess a relatively spherical or ovoid shape. These are usually soluble in water or in aqueous media containing acids, bases, salts or alcohol, and diffuse readily. As a class, *globular proteins are more complex in conformation than fibrous proteins, have a far greater variety of biological functions and are dynamic rather than static in their activities*. Tertiary and quaternary structures are usually associated with this class of proteins. Nearly all enzymes are globular proteins, as are protein hormones, blood transport proteins, antibodies and nutrient storage proteins.

A simple functional classification of globular proteins is not possible because of 2 reasons:

(a) Firstly, these proteins perform a variety of different functions.

(b) Secondly, many widely-differing globular proteins perform almost similar functions.

However, Conn and Stumpf (1976) have classified globular proteins as follows:

Globular Proteins	Cytochrome C
	Blood proteins
	Serum albumin
	Glycoproteins
	Antibodies (= Immunoglobulins)
	Hemoglobin
	Hormones
	Enzymes
	Nutrient proteins

2. Fibrous or Fibrillar Proteins.

These have axial ratios greater than 10 and, henceforth, resemble long ribbons or fibres in shape. These are mainly of animal origin and are insoluble in all common solvents such as water, dilute acids, alkalis and salts and also in organic solvents. Most fibrous proteins serve in a structural or protective role.

The fibrous proteins are extremely strong and possess two important properties which are characteristic of the elastomers. These are:

(a) They can stretch and later recoil to their original length.

(b) They have a tendency to creep, i.e., if stretched for a long time, their basic length increases and equals the stretched length but, if the tension on the two ends of the fibril is relaxed, they creep to their shorter and shorter length. A large scar, for example, creeps to a smaller size if there is no

[Contents](#)

tension on the scar. On the contrary, if the scar is in a region of high tension, the scar becomes larger and larger as happens in the skin of a person gradually becoming obese.

It is a heterogeneous group and includes the proteins of connective tissues, bones, blood vessels, skin, hair, nails, horns, hoofs, wool and silk. The important examples are:

I. **Collagens.** These are of mesenchymal origin and form the major proteins of white connective tissues (tendons*, cartilage) and of bone. *More than half the total protein in mammalian body is collagen*; acted upon by boiling in water, (NH₄)₂SO₄ or alkalis to produce the soluble gelatins; *unique in containing high contents (12%) of hydroxyproline*; poor in sulfur since cysteine and cystine are lacking.

II. **Elastins.** Also of mesenchymal origin; form the major constituents of yellow elastic tissues (ligaments, blood vessels); differ from collagens in not being converted to soluble gelatins.

III. **Keratins.** These are of ectodermal origin; form the major constituents of epithelial tissues (skin, hair, feathers, horns, hoofs, nails); *usually contain large amounts of sulfur in the form of cystine*—human hair has about 14% cystine.

IV. **Fibroin.** It is the principal constituent of the fibres of silk; composed mainly of glycine, alanine and serine units.

Tendons connect muscle to bone, whereas ligaments attach one bone to another. Tendons are nonelastic but ligaments can be stretched, because they contain, in addition to collagen fibres, the protein elastin.

CLASSIFICATION BASED ON COMPOSITION AND SOLUBILITY

This is nowadays the most accepted system of classification and is based on the proposals made by the committees of *British Physiological Society* (1907) and the *American Physiological Society* (1908). The system divides the proteins into 3 major groups, based on their composition *viz.*, simple, conjugated and derived.

A. Simple Proteins or Holoproteins.

These are of globular type except for scleroproteins which are fibrous in nature. This group includes proteins containing only amino acids, as structural components. On decomposition with acids, these liberate the constituent amino acids.

These are further classified mainly on their solubility basis as follows:

1. **Protamines and histones.** These are basic proteins and occur almost entirely in animals, mainly in sperm cells; possess simplest structure and lowest molecular weight (approximately 6,000), soluble in water; unlike most other proteins, not coagulated by heat; strongly basic in char to high content of basic amino acids (lysine, arginine); form salts with mineral acid proteins. Protamines are virtually devoid of sulfur and aromatic amino acids. Histones are weaker bases and are, therefore, insoluble in NH₄OH solution, whereas the protamines are

e.g., protamines—*clupeine* from herring sperm, *salmine* from salmon sperm, *sturtine* from sturgeon and *cypripine* from carp.

histones—*nucleohistones* of nuclei; *globin* of hemoglobin.

2. **Albumins.** These are widely distributed in nature but more abundant in seeds; soluble in water and dilute solutions of acids, bases and salts; precipitated with a saturated solution of an acid salt like (NH₄)₂SO₄ or a neutral salt like Na₂SO₄; coagulated by heat.

e.g., *leucosine* in cereals, *legumeline* in legumes, *ovalbumin* from white of egg, *serum albumin* from blood plasma, *myosin* of muscles and *lactalbumin* of milk whey.

3. **Globulins.** These are of two types—pseudoglobulins and euglobulins*. Euglobulins are more widely distributed in nature than the pseudoglobulins; either soluble

Pseudoglobulins and euglobulins differentiated on the basis of solubility behaviour. A pseudoglobulin is soluble at very low ionic strength, whereas euglobulin remains sparingly soluble until the ionic strength is raised. Euglobulins remain soluble at the isoelectric point (pI), whereas pseudoglobulins do not.

231 / 1254

Contents

CLASSIFICATION OF PROTEINS 207

(pseudoglobulins) or insoluble (euglobulins) in water; precipitated with half saturated solution of (NH₄)₂SO₄; coagulated by heat.

e.g., pseudoglobulins—*pseudoglobulin* of milk whey.

euglobulins—*serum globulin* from blood plasma, *ovoglobulin* from eggwhite; *myosinogen* from muscle; globulins of various plant seeds like hemp (*edestin*), soybeans (*glycinine*), peas (*legumine*), peach (*amandine*), oranges (*pomeline*); also potato (*tuberin*).

4. **Glutelins.** These have been isolated only from plant seeds; insoluble in water, dilute salt solutions and alcohol solutions but soluble in dilute acids and alkalis; coagulated by heat.

e.g., *glutenin* from wheat, *glutelin* from corn, *oryzenin* from rice, etc.

5. **Prolamines.** These have also been isolated only from plant seeds; insoluble in water and dilute salt solutions but soluble in dilute acids and alkalis and also in 60–80% alcohol solutions; not coagulated by heat

e.g., *gliadin* from wheat, *zein* from corn, *hordein* from oat, etc.

Some biochemists like Karlson (1968) are of the viewpoint that glutelins and prolamines should not be granted the status of exclusive classes since they are small groups of vegetable proteins occurring in grain kernels.

6. **Scleroproteins or Albuminoids.** These occur almost entirely in animals and are, therefore, commonly known as the 'animal skeleton proteins'; insoluble in water, dilute solution of acids, bases and salts and also in 60–80% alcohol solutions; not attacked by enzymes.

e.g., collagen of bones, *elastin* in ligaments, *keratin* in hair and horny tissues and *fibroin* of silk.

B. Conjugated or Complex Proteins or Heteroproteins.

These are also of globular type except for the pigment in chicken feathers which is probably of fibrous nature. These are the proteins linked with a separable nonprotein portion called *prosthetic group*. The prosthetic group may be either a metal or a compound. On decomposition with acids, these liberate the constituent amino acids as well as the prosthetic group.

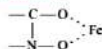
Their further classification is based on the nature of the prosthetic group present. The various divisions are metalloproteins, chromoproteins, glycoproteins, phosphoproteins, lipoproteins and nucleoproteins. (Instead of metalloproteins, chromoproteins etc., the terms metalloproteinoids, chromoproteinoids etc., are sometimes used.)

1. **Metalloproteins.** These are the proteins linked with various metals. These may be of stable nature or may be more or less labile. Based on their reactivity with metal ions, the metalloproteins may be classified into 3 groups:

I. **Metals strongly bound by proteins.** Some heavy metals (Hg, Ag, Cu, Zn) become strongly bound with proteins like collagen, albumin, casein etc., through the —SH radicals of the side chains. Some other proteins have strong binding affinities for Fe (*siderophilin**) and Cu (*ceruloplasmin***). In these cases, the following pattern of binding may be present:

Siderophilin, also called as *transferrin*, is an important metalloprotein and constitutes about 30% of the total plasma protein. It has a molecular weight of about 90,000 and is capable of binding 2 atoms of iron per mole. It facilitates iron transport.

Ceruloplasmin is an important blue copper-binding protein in the blood of humans and other vertebrates. This protein contains about 90% of copper in serum. It has a molecular weight of about 150,000 and contains 8 atoms of copper per mole. Ceruloplasmin is only one of the many sialoglycoproteins whose removal from the bloodstream is triggered by the loss of sialic acid units. It probably functions by reversibly releasing and binding copper at various sites in the body, whereby regulating copper absorption. In its deficiency, the Wilson's disease develops in man which is characterized by hepatolenticular degeneration.



II. *Metals bound weakly by proteins.* Ca belongs to this category. Here the binding takes place with the help of radicals possessing the electron charge.

III. *Metals which do not couple with proteins.* Na and K belong to this group. These form compounds with nucleic acids where apparently electrostatic bonds are present.

2. **Chromoproteins.** These are proteins coupled with a coloured pigment. Such pigments have also been found among the enzymes like catalase, peroxidase and flavoenzymes. Similarly, chlorophyll is present in leaf cells in the form of a protein, the chloroplastin. The chloroplastin dissolves in water as a colloid and is readily denatured.

e.g., myoglobin, hemoglobin, hemocyanin, hemoerythrin, cytochromes, flavoproteins, catalase, etc.

3. **Glycoproteins and Mucoproteins.** These are the proteins containing carbohydrate as prosthetic group. Glycoproteins contain small amounts of carbohydrates (less than 4%), whereas mucoproteins contain comparatively higher amounts (more than 4%).

e.g., glycoproteins—egg albumin, elastase certain serum globulins and also certain serum albumins.

mucoproteins—ovomucoid from egg-white, mucin from saliva and *Dioscorea* tubers, osseomucoid from bone and tendomucoid from tendon.

4. **Phosphoproteins.** These are proteins linked with phosphoric acid; mainly acidic.

e.g., casein from milk and ovovitellin from egg yolk.

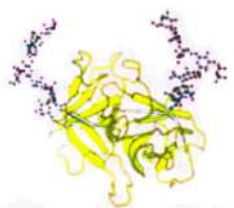
5. **Lipoproteins.** Proteins forming complexes with lipids (cephalin, lecithin, cho. called lipoproteins; soluble in water but insoluble in organic solvents).

e.g., lipovitellin and lipovitellenin from egg yolk; lipoproteins of blood.

The lipoproteins are in reality the temporary intermediates in the process of transfer of lipids from the site of absorption to the site of utilization. The classification of lipoproteins is frequently based on an operational definition, i.e., the migration of the fraction in a density gradient separation. On this basis, the lipoproteins have been classified into following 4 categories (Table 10-2):

- Very high density lipoproteins (VHDLs). These have densities greater than 1.21.
- High density lipoproteins (HDLs). These possess density range of 1.063 to 1.21.
- Low density lipoproteins (LDLs). Their densities range between 1.05 and 1.063.
- Very low density lipoproteins (VLDLs). Their density range is from 0.93 to 1.05.

Table 10-2 also lists the compositions of the fractions floated at the respective densities, along with proposed functions.



Elastase, a secreted glycoprotein, showing linked carbohydrates on its surface. Elastase is a protease found in serum. Note that the oligosaccharide chains have substantial size even for this protein, which has a relatively low level of glycosylation.

233 / 1254

Contents

CLASSIFICATION OF PROTEINS 209

Table 10-2. Classification of lipoproteins based on density

Name	Density range	Composition by weight (in blood plasma)	Half-life	Function
Very high density lipoprotein	> 1.21	62% protein 28% lipid	> 5 days	Phospholipid transport (contains 8–15% of serum phospholipids)
High density lipoprotein	1.063 – 1.21	50% protein 50% lipid	3 – 5 days	Cholesterol transport Lipid transport
Low density lipoprotein	1.05 – 1.063	22% protein 12% triglyceride 20% phospholipids 46% cholesterol	3 – 5 days	Lipid transport (congenital absence accompanied by intestinal malabsorption and distorted shape of erythrocyte)
Very low density lipoprotein	0.93 – 1.05	Largely protein, triglyceride and phospholipid; composition variable	3 – 4 hours	Triglyceride transport (related to chylomicron units)

(Adapted from Mallette MF, Clagett CO, Phillips AT and McCari RL, 1979)

6. **Nucleoproteins.** These are compounds containing nucleic acid and protein, esp., protamines and histones. These are usually the salt-like compounds of proteins since the two components have opposite charges and are bound to each other by electrostatic forces. They are present in nuclear substances as well as in the cytoplasm. These may be considered as the sites for the synthesis of proteins and enzymes.

e.g., nucleoproteins from yeast and thymus and also viruses which may be regarded as large molecules of nucleoproteins; nucleohistones from nuclei-rich material like glandular tissues; nuclein.

C. Derived Proteins.

These are derivatives of proteins resulting from the action of heat, enzymes or chemical reagents. This group also includes the artificially-produced polypeptides.

I. **Primary derived proteins.** These are derivatives of proteins in which the size of protein molecule is not altered materially.

1. **Proteans.** Insoluble in water; appear as first product produced by the action of acids, enzymes or water on proteins.

e.g., edestin derived from edestin and myosin derived from myosin.

2. **Metaproteins or Intraproteins.** Insoluble in water but soluble in dilute acids or alkalies; produced by further action of acid or alkali on proteins at about 30–60°C.

e.g., acid and alkali metaproteins.

3. **Coagulated Proteins.** Insoluble in water; produced by the action of heat or alcohol on proteins.

e.g., coagulated eggwhite.

II. **Secondary derived proteins.** These are derivatives of proteins in which the hydrolysis has certainly occurred. The molecules are, as a rule, smaller than the original proteins.

- 1. **Proteoses.** Soluble in water; coagulable by heat; produced when hydrolysis proceeds beyond the level of metaproteins; *primary proteoses* are salted out by half saturation with $(\text{NH}_4)_2\text{SO}_4$ and precipitated by HNO_3 and picric acid; *secondary proteoses* are salted out only by complete saturation with $(\text{NH}_4)_2\text{SO}_4$ but are not precipitated by HNO_3 or picric acid.
e.g., albumose from albumin; *globulose* from globulin.
 - 2. **Peptones.** Soluble in water; noncoagulable by heat; produced by the action of dilute acids or enzymes when hydrolysis proceeds beyond proteoses; neither salted out by $(\text{NH}_4)_2\text{SO}_4$ nor precipitated by HNO_3 or picric acid.
 - 3. **Polypeptides.** These are combinations of two or more amino acid units. In fact, the *proteins* are essentially long chain polypeptides.
- Drawbacks.* Although widely accepted, the system outlined above has certain discrepancies:
- 1. The classification is arbitrary.
 - 2. The criterion of solubility is not well demarcated as some globulins (pseudoglobulins) are also soluble in water.
 - 3. Protamines and histones should have been kept under derived proteins.
 - 4. The group metaproteins is an artificial assemblage.
- The two (second and third) systems of classification, described above, may be merged into one, as shown on page 197.

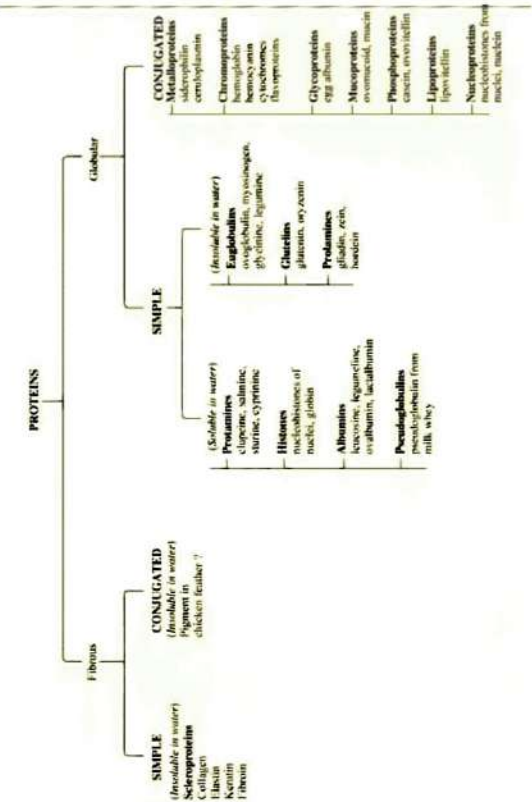
CLASSIFICATION BASED ON BIOLOGICAL FUNCTION

Depending upon their physical and chemical structure and location inside the cell, different proteins perform various functions. As such diverse proteins may be grouped under following categories, based on the metabolic functions they perform (Table 10-3):

Table 10-3. Classification of proteins on the basis of their biological functions

Class of protein	Function	Examples
Enzymic proteins	Biological catalysts	Urease, Amylase, Catalase Cytochrome C, Alcohol dehydrogenase.
Structural proteins	Strengthening or protecting biological structures	Collagen, Elastin, Kera
Transport or carrier proteins	Transport of ions or molecules in the body	Myoglobin, Hemoglobin, Ceruloplasmin, Lipoproteins
Nutrient and storage proteins	Provide nutrition to growing embryos and store ions	Ovalbumin, Casein, Ferritin
Contractile or motile proteins	Function in the contractile system	Actin, Myosin, Tubulin
Defense proteins	Defend against other organisms	Antibodies, Fibrinogen, Thrombin
Regulatory proteins	Regulate cellular or metabolic activities	Insulin, G proteins, Growth hormone
Toxic proteins	Hydrolyze (or degrade) enzymes	Snake venom, Ricin

- 1. **Enzymic proteins.** The most varied and most highly specialized proteins are those with catalytic activity—the enzymes. Virtually, all the chemical reactions of organic biomolecules are catalyzed by the enzymes. *Nearly all enzymes are globular proteins.* Chemically, some enzymes are simple proteins.



212 FUNDAMENTALS OF BIOCHEMISTRY

containing only amino acid residues; others are complex proteins, containing a major protein part (apoenzyme) and a small nonprotein part (prosthetic group) associated with the protein unit. Enzymes catalyze a variety of reactions. *Urease, amylase, catalase, cytochrome C, alcohol dehydrogenase* are some of the examples of enzymic proteins.

2. **Structural proteins.** The structural proteins are usually inert to biochemical reactions. They maintain the native form and position of the organs. The cell wall and primary fibrous constituents of the cell have structural proteins. *Collagen*, which has very high tensile strength, is the most abundant protein of animals. It is found in connective tissue such as tendons, cartilage, matrix of bones and cornea of the eye. *Leather is almost pure collagen*. Ligaments contain *elastin*, a structural protein capable of stretching in two dimensions. α *keratin* constitutes almost the entire dry weight of hair, wool, feathers, nails, claws, quills, scales, horns, hooves, tortoise shell and much of the outer layer of skin. The major component of silk fibres and spider web is *fibrin*. The wing hinges of some insects are made of *resilin*, which has nearly perfect elastic properties.

3. **Transport or carrier proteins.** Certain proteins, specially in the animals, are involved in the transport of many essential biological factors to various parts of the organisms. *Hemoglobin* of erythrocytes binds O_2 as the blood passes through the lungs, carries it to the peripheral tissues, and there releases it to participate in the oxidation of nutrients. The blood plasma contains *lipoproteins*, which carry lipids from the liver to other organs. Other kinds of transport proteins are present in the plasma membranes and intracellular membranes of all organisms. *Ceruloplasmin* transports copper in blood.

4. **Nutrient and storage proteins.** *Ovalbumin* is the major protein of eggwhite. The milk protein, *casein* stores amino acids. The seeds of many plants store nutrient proteins, required for the growth of the germinating seedlings. *Ferritin*, found in some bacteria and in plant and animal tissues.

5. **Contractile or motile proteins.** Some proteins endow cells and organisms with contract, to change shape, or to move about. *Actin* and *myosin* function in the contract. skeletal muscle and also in many nonmuscle cells. *Tubulin* is the protein from which microtubules are built.

6. **Defense proteins.** Many proteins defend organism against invasion by other species or protect them from injury. The *antibodies* (or immunoglobulins), the specialized proteins made by the lymphocytes of vertebrates, can precipitate or neutralize invading bacteria, viruses or foreign proteins from another species. *Fibrinogen* and *thrombin*, although enzymic, are blood-clotting proteins that prevent loss of blood when the vascular system is injured.

7. **Regulatory proteins.** Some proteins help regulate cellular or physiological activity. Among them are many hormones, such as *insulin* which regulates sugar metabolism, and *growth hormone* which is required for bone growth. The cellular response to many hormonal signal is often mediated by a class of GTP-binding proteins called *G proteins*. Other regulatory proteins bind to DNA and regulate the biosynthesis of enzymes and RNA molecules involved in cell division.

8. **Toxic proteins.** Some proteins act as toxic substances, such as *snake venom*, *bacterial toxins* and toxic plant proteins like *ricin*. These toxic proteins also have defensive functions.

EGG PROTEINS

The eggs and milk are consumed as food by an appreciable portion of the global population. Hence, the egg and milk proteins deserve a special mention. Eggs contain 2 types of proteins: the egg white protein and egg yolk protein. The various components present in the egg white protein are listed in Table 10-4.

237 / 1254

CLASSIFICATION OF PROTEINS

213

Table 10-4. Protein composition of egg white

Protein	Percentage
Total protein	{ 10-11% (on wet basis) 82.8% (on dry basis)
Ovalbumin	70% of total proteins
Conalbumin	9%
Ovomucoid	13%
Globulins	
Lysozyme (G_1)	2.6%
(G_2) }	
(G_3) }	7%
Mucin	2%
Avidin	0.06%

(Adapted from Fevold HL, 1951)

When egg yolk is diluted with water, proteins precipitate. When the yolk is heated, the proteins undergo heat denaturation and precipitate. Egg yolk contains 2 phosphoproteins, lipovitellin and lipovitellin, which differ from each other in the following characteristics:

Lipovitellin

Lipid concentration: 17-18%
Phosphorus concentration: 1%
Protein present: Vitellin

Lipovitellin

36-41% (mainly lecithins)
0.29%
Vitellin

Egg yolk also contains water-soluble protein that does not precipitate on dilution of the yolk. This fraction is called *livet*. The enzymatic activity of yolk lies in this fraction.

Whole egg is an excellent food because it is a very rich source not only of protein and lipid but also of most of the vitamins (except vitamin C) and most of the required minerals (except calcium).

MILK PROTEINS



Milk, a source of nourishment for all mammals, is composed, in part, of a variety of proteins.

Milk (Fig. 10-2) contains about 0.6-0.7% protein which is not precipitated on acidification to pH 4.7. This represents about 20% of the protein contained in skim milk. These *why proteins* are separated into 2 fractions: *lactalbumin* and *lactoglobulin*.

The name *casein* is assigned to the fraction precipitated by acidifying milk to a pH of 4.7. It is present in cow's milk (3.3-5%) and human milk (0.3-0.6%). Casein may be further purified by redissolving and precipitating again. It is of 3 types: α , β and γ . These differ one another in their molecular weights, their rate of migration in an electric field and their phosphorus content.

REFERENCES

See list following Chapter 11.

CONTENTS

- Physical Properties
 - Colour and Taste
 - Shape and Size
 - Molecular Weight
 - Colloidal Nature
 - Denaturation
 - Amphoteric Nature
 - Ion Binding Capacity
 - Solubility
 - Optical Activity
- Chemical Properties
 - Hydrolysis
 - Reactions Involving COOH Group
 - Reactions Involving NH₂ Group
 - Reactions Involving Both COOH and NH₂ Groups
 - Reactions Involving R Group or Side Chain
 - Reactions Involving SH Group

Proteins—III

General Properties

The general properties of proteins are reminiscent of those of the amino acids.

PHYSICAL PROPERTIES

1. **Colour and Taste.** Proteins are colourless and usually tasteless. These are homogeneous and crystalline.

2. **Shape and Size.** As already discussed, the proteins range in shape from simple crystalloid spherical structures to long fibrillar structures. Two distinct patterns of shape have been recognized :

A. Globular proteins—These are spherical in shape and occur mainly in plants, esp., in seeds and in leaf cells. These are bundles formed by folding and crumpling of protein chains.

e.g., pepsin, edestin, insulin, ribonuclease

B. Fibrillar proteins—These are the ellipsoidal in shape and occur generally in animals. Most of the studies regarding protein structure, conducted using these proteins.

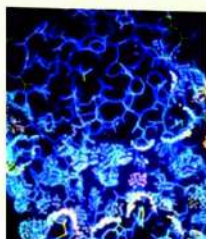
e.g., fibrinogen, myosin etc.

Each protein molecule is characterized by its specific size (Fig. 11-1). For example :

(a) Hemoglobin has a diameter of 55 Å.

(b) Edestin has a diameter of 80 Å.

(c) Catalase has dimensions of 80 × 64 × 54 Å (of the axes).



Computer-generated model of the protein pepsin

This enzyme protein helps in the digestion of food ingested by living beings.

239 / 1254

Contents

GENERAL PROPERTIES 215

(d) Human fibrinogen has a diameter of 38 Å and a length of 700 Å.

(e) Collagen is one of the longest proteins, with a length of 3,000 Å.

In general, the protein molecules are always very large, as can be seen in the following examples :

(a) Gliadin (from wheat)—C₆₈₅H₁₀₆₈N₁₉₆O₂₁₁S₅

(b) Zein (from corn)—C₇₃₆H₁₁₆₁N₁₈₁O₂₃₈S₃

(c) Casein (from milk)—C₇₀₈H₁₁₃₀N₁₈₀O₂₂₄S₄P₄

(d) Beta-lactoglobulin (from milk)—C₁₈₄₂H₂₈₀₂O₄₈₁N₄₂₀S₁₈

3. **Molecular Weight.** The extraordinary size, poor stability, specific solubility conditions and high reactivity have rendered the determination of molecular weight of proteins as a difficult task (Edsall, 1953). However, the proteins generally have large molecular weights ranging between 5×10^5 and 1×10^6 (see Table 11-1). It might be noted that the values of molecular weights of many proteins lie close to or multiples of 35,000 and 70,000. Previously, this was interpreted as a regularity under the name *Svedberg's rule*. Also, it was then assumed that proteins are composed of units of molecular weight 17,500. This corresponds to about 145–150 amino acid residues, since the average molecular weight of an amino acid residue amounts to about 115–120. The discovery in recent times of too many exceptions to this rule, however, finally forced its abandonment.

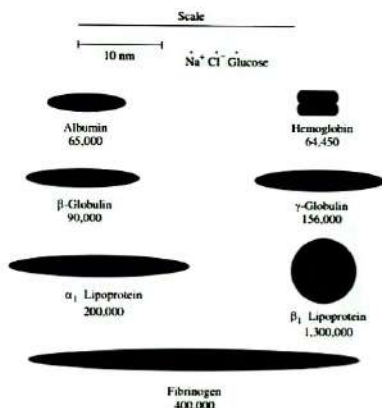


Fig. 11-1. Relative dimensions and molecular weights of some of the protein molecules in the blood

(After Oncley JL, 1949)

216 FUNDAMENTALS OF BIOCHEMISTRY

The approximate number of amino acid residues in a simple protein having no prosthetic group can be calculated by dividing its molecular weight by 110. The average molecular weight of the 20 amino acids is about 138. But as the smaller amino acids predominate in most proteins, hence the average molecular weight of an amino acid is nearer 128. Since a molecule of water ($MW = 18$) is eliminated to produce each peptide bond, the average molecular weight of the amino acid residue is about $128 - 18 = 110$. Table 11-1 also gives the number of amino acid residues present in different proteins.

Table 11-1. Exact / Approximate molecular weights and the isoelectric points of some important proteins

Protein	Molecular weight (MW)	Number of residues	Number of chains	Isoelectric point (pI)
1. Insulin (bovine)	5,733	51	2	5.4
2. Cytochrome C	12,500	104	—	9.8
3. Ribonuclease (bovine)	14,000	124	1	7.8/9.5
4. Lysozyme (eggwhite)	14,600	129	1	11.0
5. Myoglobin (horse)	16,700	153	1	7.0
6. Chymotrypsin (bovine)	22,600	241	3	—
7. Pepsin	35,500	—	—	2.7
8. Ovalbumin (hen)	40,000	—	—	4.6
9. Zein	40,000	—	—	—
10. Hemoglobin (human)	64,500	574	4	—
11. Serum albumin (human)	68,500	~ 550	1	—
12. Hexokinase (yeast)	96,000	~ 800	4	—
13. γ -globulin (horse)	149,900	~ 1,250	4	—
14. Catalase	250,000	—	—	—
15. Edestin	300,000	—	—	6.9
16. Fibrinogen	450,000	—	—	5.5
17. Urease	480,000	—	—	5.0
18. Glutamate dehydrogenase (bovine)	1,000,000	~ 8,300	~ 40	—
19. Virus protein of TMV	60,000,000	—	—	—

241 / 1254

4. Colloidal Nature. Because of their giant size, the proteins exhibit many colloidal properties, such as :

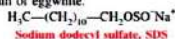
- I. Their diffusion rates are extremely slow.
- II. They may produce considerable light-scattering in solution, thus resulting in visible turbidity (Tyndall effect).

5. Denaturation. Denaturation refers to the changes in the properties of a protein. In other words, it is the loss of biologic activity. In many instances the process of denaturation is followed by **coagulation**—a process where denatured protein molecules tend to form large aggregates and to precipitate from solution.

Denaturation may be brought about by a variety of agents, both physical and chemical. The **physical agents** include mechanical action (like shaking), heat treatment cooling and freezing

GENERAL PROPERTIES 217

operations, rubbing, high hydrostatic pressures, (5,000 to 10,000 atm.), ultraviolet rays, etc. The **chemical agents**, that cause denaturation, are many ionizing radiations (like X-rays, radioactive and ultrasonic radiations), organic solvents (acetone, alcohol), aromatic anions (salicylates), some anionic detergents (like sodium dodecyl sulfate), etc. A common example of protein easily denatured by shaking or heat is the albumin of eggwhite.



It was suggested by Wu (1931) that denaturation leads mainly to the unfolding of the peptide chain, thus causing disorganization of the internal structure of protein (Fig. 11-2). This is evidenced by the fact that the denatured proteins are more easily hydrolyzed (Mirsky, 1935).

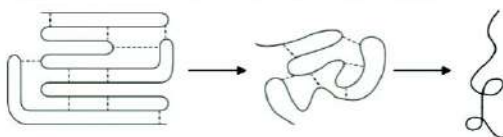


Fig. 11-2. Representation of denaturation of a protein

The process of denaturation breaks bonds.

When the peptide chains or the protein molecules are unrolled, certain bonds split and new sites of bundles are exposed to the action of certain proteolytic enzymes causing hydrolysis. Thus, the H-bonds linking the 2 peptide chains are partly freed and the disulfide ($-\text{S}-\text{S}-$) bonds also linking the two peptide chains split open to yield the free sulfhydryl ($-\text{SH}$) groups.

According to Putnam (1953), the proteins, on denaturation, undergo following changes :

1. Decrease in their solubility.
2. Cessation of their biochemical activity as enzymes or hormones.
3. Decrease in size and shape of the molecule.
4. Increased activity of some radicals present in the molecule such as $-\text{SH}$ group of cysteine, $-\text{S}-\text{S}-$ bond of cysteine and phenolic group of tyrosine.

Further, on denaturation, new ionizable groups become available for acid-base titration (Steinhardt and Zaiser, 1955) and also there occurs a change in optical rotation in the direction of increased levorotation (Simpson and Kauzmann, 1953). Denaturation also leads to alteration in surface tension and loss of antigenicity.

Some proteins, when denatured, cannot be brought back to their original state. In that case denaturation is described as of 'irreversible' type. On the other hand, denaturation in other proteins is of 'reversible' type. For example, if trypsin is exposed to a temperature of $80-90^\circ\text{C}$, it denatures and when this solution is cooled at 37°C , the solubility and the activity of this protein-enzyme is regained. The process of regaining normal protein properties by a denatured protein is called **renaturation** or **refolding**. During renaturation, certain antibodies may cause a re-rolling of the protein bundles so that most of the original bonds are recovered (Pauling, 1940). The recovery of the renatured protein is, however, never complete.

6. Amphoteric Nature. Like amino acids, the proteins are **amphoteric**, i.e., they act as acids and alkalis both (Fig. 11-3). These migrate in an electric field and the direction of migration depends upon the net charge possessed by the molecule. The net charge is influenced by the pH value. Each protein has a fixed value of isoelectric point (pI) at which it will move in

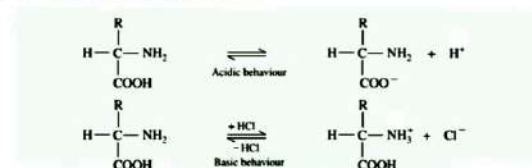


Fig. 11-3. Amphoteric nature of amino acids (and of proteins)

an electric field (refer Table 11-1). **Isoelectric point** (or **isoelectric point**) is the pH value at which the number of cations is equal to that of anions. Thus, at isoelectric point, the net electric charge of a protein is always zero. But the total charge on the protein molecule (sum of positive and negative charges) at this point is always maximum. Thus, the proteins are **dipolar ions** or **internal salts** or **zwitterions** (German for 'ion of both kinds'; amphoteric ions) at pI and exist in solution as:



This general structure of amino acids, represented as an inner salt, was originally proposed by N. Bjerrum in 1923. The formula depicts the carboxyl group as being dissociated while the amino group is protonated.

At pH values lower than pI , the protein will have a net positive charge and, as a cation, will migrate towards negative pole (cathode). Similarly, at pH values higher than pI , the protein will have a net negative charge and, as an anion, will move towards positive pole (anode).

It may be stated in general that those proteins having an excess of carbonyl groups will tend to have a lower pI while those having an excess of amino groups will tend to have a higher pI .

The osmotic pressure and viscosity of the protein solution are a minimum at the pI point. Also at the isoelectric point, proteins are found to be least soluble and can be most easily.

7. Ion Binding Capacity. Being amphoteric in nature, the proteins can form cations and anions based on their net charge. In fact, a mixture of different proteins (pH except at pI) will include cations and anions both and the salts of protein-protein combinations will be formed. This occurs in tissues since both acidic and basic proteins are present.

Many ions form insoluble salts with proteins and serve as excellent precipitating agents for proteins. For example, anions of some acids like phosphotungstic, trichloroacetic, picric etc., form insoluble salts with proteins when the latter behave as cations (acid side of their pI).

Heavy metals are used for precipitating proteins on the alkaline side of their pI , the proteins behaving as anions. Ions of Hg, Cu, Ag, Zn etc., are frequently used for this purpose. Many acid dyes find practical use for colouring the insoluble proteins like silk and wool.

8. Solubility. The solubility of proteins is markedly influenced by pH. Solubility is lowest at isoelectric point and increases with increasing acidity or alkalinity. This is because when the protein molecules exist as either cations or anions, repulsive forces between ions are high, since all the molecules possess excess charges of the same sign. Thus, they will be more soluble than in the isoelectric state.

A. 'Salting-in' effect. Globulins are sparingly soluble in water but their solubility is greatly increased by the addition of neutral salts like NaCl. This phenomenon is commonly described as

243 / 1254

'salting-in' effect.

B. 'Salting-out' effect. Proteins are precipitated from aqueous solution by high concentrations of neutral salts. This is the 'salting-out' process. Divalent and trivalent ions are more effective than univalent ions. The salts commonly used for this purpose are Na_2SO_4 , $(NH_4)_2SO_4$, magnesium salts and phosphates.

C. **Isoelectric precipitation.** Some proteins like casein of milk, however, are readily precipitated at or near their isoelectric point. This process is, therefore, described as isoelectric precipitation.

9. Optical Activity. All protein solutions rotate the plane of polarized light to the left, i.e., these are *levorotatory*. For example, the specific rotation $[\alpha]_D$ for ovalbumin is near -30° over the pH range between 3.5 and 11. However, at lower or higher pH values the rotation becomes more negative, e.g., at pH 13, the $[\alpha]_D$ is about -60° . The rotation is further increased by subjecting proteins to high temperatures.

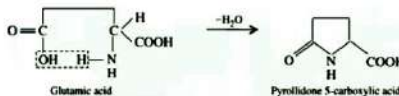
CHEMICAL PROPERTIES

A. HYDROLYSIS

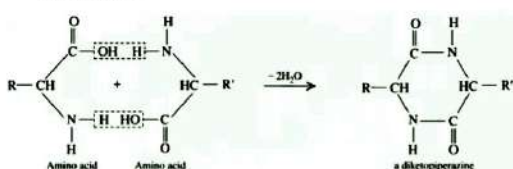
Proteins are hydrolyzed by a variety of hydrolytic agents.

1. By acidic agents. Proteins, upon hydrolysis with conc. HCl (6-12N) at 100-110°C for 6 to 20 hrs, yield amino acids in the form of their hydrochlorides. Undesirable side-effects of acid hydrolysis include the following:

- Tryptophan, serine and threonine are destroyed during acid hydrolysis and as such this reaction is not used for their isolation.
- Asparagine and glutamine are deamidated to aspartate and glutamate respectively.
- Glutamic acid undergoes intramolecular dehydration to pyrrolidone 5-carboxylic acid.



(d) Other amino acids may undergo intermolecular dehydration forming cyclic anhydrides or diketopiperazines.



2. By alkaline agents. Proteins may also be hydrolyzed with 2N NaOH. Alkaline hydrolysis is, however, less used as it is highly disadvantageous:

- It leads to the destruction of certain amino acids like arginine, cysteine, cystine, serine, threonine etc.

220 FUNDAMENTALS OF BIOCHEMISTRY

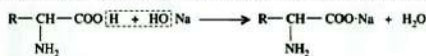
(b) It also causes loss of optical activity (or racemization) of the amino acids.

3. **By proteolytic enzymes.** Under relatively mild conditions of temperature and acidity, certain proteolytic enzymes like pepsin and trypsin hydrolyze the proteins. Enzyme hydrolysis is used for the isolation of certain amino acids like tryptophan. Two important drawbacks with this type of hydrolysis are:

- It requires prolonged incubation.
- Hydrolysis is incomplete.

B. REACTIONS INVOLVING COOH GROUP

1. **Reaction with alkalis (Salt formation).** The carboxylic group of amino acids can release a H^+ ion with the formation of carboxylate (COO^-) ions. These may be neutralised by cations like Na^+ and Ca^{2+} to form salts. Thus, amino acids react with alkalis to form salts.

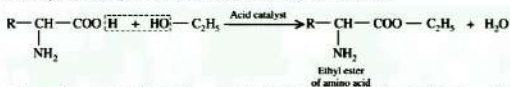


Sodium salt of glutamic acid (monosodium glutamate) is used commercially as a flavouring agent. It imparts a meat-like flavour to soups, for example.

Monosodium glutamate (MSG) or 'ajinomoto' in common parlance, or the 'taste powder' as some people call it, has a long history of use in China. Like sugar or salt, it is regarded as a fairly standard part of cuisine. It makes a great difference to the way the food tastes. MSG actually represents a new taste - the meaty, broth-like taste, common to foods as diverse as steak, lobster and tomato. Virtually, all proteins contain glutamates, including mother's milk. Most glutamates are 'bound' and have no effect on the flavour but 'free' glutamates increase food's palatability. Every 100 g of Parmesan cheese, for instance, contains 9,847 mg of bound and 1,200 mg of free glutamates. Tomatoes, peas, meat stock and canned foods and snacks also have it. Human body carries about 2 kg of glutamate in muscles, brain, organs and tissues. Oriental cooks have long added a large seaweed, *Laminaria japonica*, to their soup stocks to enhance that brothy flavour. The seaweed contains glutamate. MSG is used as a flavour enhancer not only by food processors but also by consumers as well as culinary experts in restaurants and hotels. In fact, MSG is one of the most ubiquitous of all food additives and is a popular ingredient of Chinese cuisine. Consumption containing too much MSG has precipitated attacks of sweating, headaches and *gi* disorders in individuals, sensitive to the chemical. The disease characterized by the above is called **Kwok's disease** or the **'Chinese Restaurant Syndrome' (CRS)**. Kwok, a patte described people experiencing numbness of the back of the neck, general weakness and after eating in Chinese restaurants (hence, its nomenclature). People with high blood pressure, however, abstain foods added with MSG.

245 / 1254

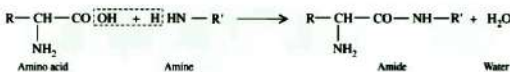
2. **Reaction with alcohols (Esterification).** With alcohols, corresponding esters are produced. The esters, so obtained, are volatile in contrast to the free amino acids.



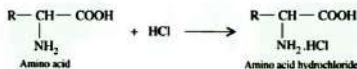
The reaction was, for the first time, used by Emil Fischer for the isolation of amino acids in pure form from protein hydrolysates by the fractional distillation in vacuum of their ethyl esters.

3. **Reaction with amines.** Amino acids react with amines to form amides.

GENERAL PROPERTIES 221

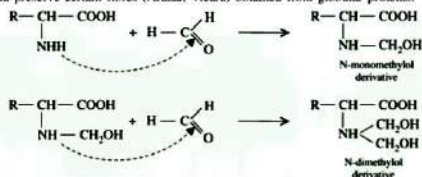
C. REACTIONS INVOLVING NH_2 GROUP

1. **Reaction with mineral acids (Salt formation).** When either free amino acids or proteins are treated with mineral acids like HCl, the acid salts are formed.



The basic amino acids, arginine and lysine react with CO_2 in the presence of air to form carbonate salts. Because of this property, these are usually stored and also sold in the form of their monochlorides.

2. **Reaction with formaldehyde.** With formaldehyde, the *hydroxy-methyl* derivatives are formed. These derivatives are insoluble in water and resistant to attack by microorganisms. Because of this action, formaldehyde is the principal reagent in embalming fluids and is used to harden and preserve certain fibres (Aralac, Vicara) obtained from globular proteins.

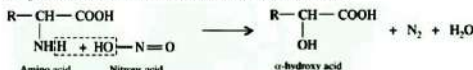


This reaction is the basis of the Sorensen titration method for determining the purity of the individual amino acids.

3. **Reaction with benzaldehyde.** Schiff's bases are formed.



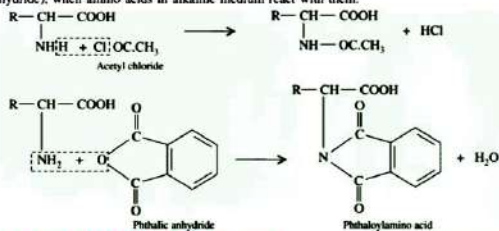
4. **Reaction with nitrous acid (Van Slyke reaction).** The amino acids react with HNO_2 to liberate N_2 gas and to produce the corresponding α -hydroxy acids.



This reaction is characteristic of aliphatic primary amines and has been utilized by Van Slyke (1912) as the basis for his 'nitrous acid' method for the estimation of amino acids by measuring the volume of N_2 gas liberated. The imino acids proline and hydroxyproline, however, do not respond to this reaction.

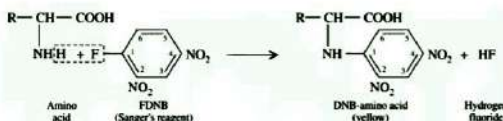
222 FUNDAMENTALS OF BIOCHEMISTRY

5. Reaction with acylating agents (Acylation). Acylation is brought about by many acid chlorides (CH_3COCl , $\text{C}_6\text{H}_5\text{COCl}$) and acid anhydrides ($\text{CH}_3\text{CO}-\text{O}-\text{OC}\cdot\text{CH}_3$, phthalic anhydride), when amino acids in alkaline medium react with them.



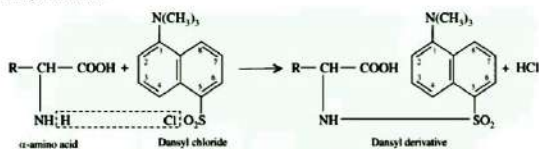
6. Reaction with FDNB* or Sanger's reagent. In mildly alkaline solution, FDNB (1-fluoro-2, 4-dinitrobenzene) reacts with α -amino acids to produce yellow coloured derivative, *DNB-amino acid*.

FDNB is also called as **DNFB** (2, 4-dinitrofluorobenzene).



This reaction is valuable in elucidation of protein structure and has been successful by Sanger in England in determining the sequence of amino acids in insulin.

7. Reaction with dansyl chloride. The N-terminal amino acid of the protein reacts with 1-dimethylaminonaphthalene-5-sulfonyl chloride (or dansyl chloride) to form a *dansyl derivative*.

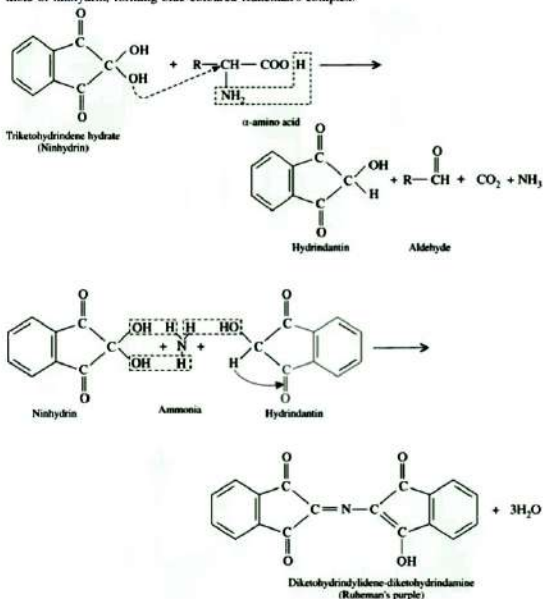
D. REACTIONS INVOLVING BOTH COOH AND NH₂ GROUPS

1. Reaction with triketohydrindene hydrate (Ninhydrin reaction). Ninhydrin (= indane-1, 2, 3, -trione hydrate) is a powerful oxidizing agent and causes oxidative decarboxylation of α -amino acids producing CO_2 , NH_3 and an aldehyde with one less carbon atom than the parent

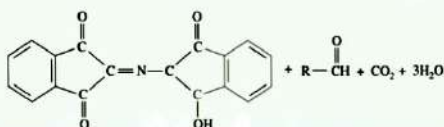
247 / 1254

GENERAL PROPERTIES 223

amino acid. The reduced ninhydrin (or hydrindantin) then reacts with the liberated NH_3 and a mole of ninhydrin, forming blue-coloured Ruheman's complex.



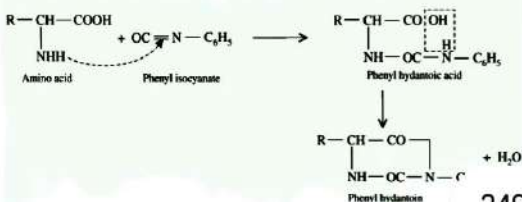
The net equation may, thus, be written as follows:



This reaction has been used by Van Slyke as a basis for quantitative estimation of α -amino acids. Here the CO_2 produced may be measured manometrically. It is more specific than the 'nitrous acid' method. This reaction is extremely sensitive and gives reliable results with small amounts of material.

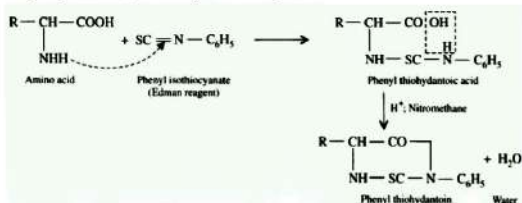
Amines, other than α -amino acids, also react with ninhydrin forming a blue complex but without evolving CO_2 . The evolution of CO_2 is, thus, indicative of the presence of an α -amino acid. Proline and hydroxyproline, however, produce yellow complexes rather than blue with ninhydrin.

2. Reaction with phenyl isocyanate. With phenyl isocyanate, *hydantoic acid* is formed which in turn can be converted to *hydantoin*.



249 / 1254

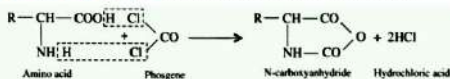
3. Reaction with phenyl isothiocyanate or Edman reagent. Phenyl isothio, reacts similarly with amino acids to produce *thiohydantoic acid*. On treatment with acids in nonhydroxylic solvents, the latter cyclize to *thiohydantoin*.



GENERAL PROPERTIES 225

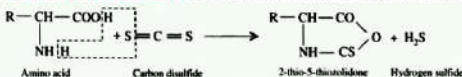
This reaction has proved useful in the studies of protein structure.

4. Reaction with phosgene. With phosgene, *N-carboxyanhydride* is formed.



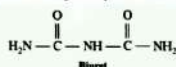
Phosgene (synonyms: carbonyl chloride, carbon oxychloride, carbonic dichloride), an acidic chloride, is a colourless gas and has a musty odour, resembling fresh mown hay or green corn. It liquefies at 8°C . Since it is heavier (3.43 times) than air, it was used extensively in gas warfare in World War I, and caused 80% of the deaths by gas in that conflict. Luckily, gas warfare was sparingly used in World War II, thus reducing the number of casualties. Phosgene does not occur in nature. It was first synthesized by Sir Humphrey Davy in 1812 by passing CO and chlorine through charcoal; it is deadlier than both CO and Cl_2 as it kills rapidly in as low a concentration as 550 ppm. Phosgene was also implicated in the Bhopal tragedy that occurred in India in December, 1984. However, phosgene is also an important industrial chemical being used in the synthesis or manufacture of isocyanates, polyurethane, polycarbonate resins, aniline dyes, pharmaceuticals, plastics and insecticides.

5. Reaction with carbon disulfide. With carbon disulfide, 2-thio-5-thiozolidone is produced.

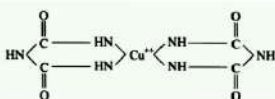


E. REACTIONS INVOLVING R GROUP OR SIDE CHAIN

1. Biuret test. Compounds containing peptide bonds produce a characteristic purple colour when treated with an alkaline 0.2% copper sulfate solution (or biuret reagent). This reaction is termed as 'biuret reaction' since it is also given by the substance biuret.



The colour deepens as the number of peptide bonds is increased and the proteins produce a deep blue-violet colour due to the probable formation of a coordination complex whose structure is given below:

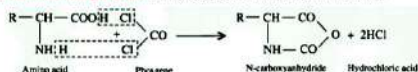


The test is, in fact, given by biuret as well as any similar structure having 2 amide or peptide bonds linked directly or through an intermediate carbon atom. The required unit is shown below between the two broken lines:

GENERAL PROPERTIES 225

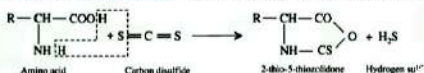
This reaction has proved useful in the studies of protein structure.

4. Reaction with phosgene. With phosgene, *N*-carboxyanhydride is formed.



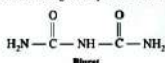
Phosgene (synonyms: carbonyl chloride, carbon oxychloride, carbonic dichloride), an acidic chloride, is a colourless gas and has a musty odour, resembling fresh mown hay or green corn. It liquefies at 8°C. Since it is heavier (3.43 times) than air, it was used extensively in gas warfare in World War I, and caused 80% of the deaths by gas in that conflict. Luckily, gas warfare was sparingly used in World War II, thus reducing the number of casualties. Phosgene does not occur in nature. It was first synthesized by Sir Humphrey Davy in 1812 by passing CO and chlorine through charcoal; it is deadlier than both CO and Cl₂ as it kills rapidly in as low a concentration as 550 ppm. Phosgene was also implicated in the Bhopal tragedy that occurred in India in December, 1984. However, phosgene is also an important industrial chemical being used in the synthesis or manufacture of isocyanates, polyurethane, polycarbonate resins, aniline dyes, pharmaceuticals, plastics and insecticides.

5. Reaction with carbon disulfide. With carbon disulfide, 2-thio-5-thiazolidone is produced.

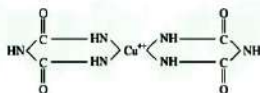


E. REACTIONS INVOLVING R GROUP OR SIDE CHAIN

1. Biuret test. Compounds containing peptide bonds produce a characteristic purple colour when treated with an alkaline 0.2% copper sulfate solution (or biuret reagent). This reaction is termed as 'biuret reaction' since it is also given by the substance biuret.



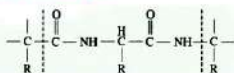
The colour deepens as the number of peptide bonds is increased and the proteins produce a deep blue-violet colour due to the probable formation of a coordination complex whose structure is given below:



The test is, in fact, given by biuret as well as any similar structure having 2 amide or peptide bonds linked directly or through an intermediate carbon atom. The required unit is shown below between the two broken lines:

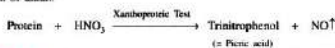
250 / 1254

226 FUNDAMENTALS OF BIOCHEMISTRY



All proteins except dipeptides, therefore, respond to this reaction. This reaction is widely used both as a qualitative test for the detection of proteins and as a quantitative measure of protein concentration.

2. Xanthoproteic test. Yellow colour develops on boiling proteins with conc. HNO₃ due to the presence of benzene ring. This reaction is due to the nitration of the phenyl rings (of tyrosine, tryptophan and phenylalanine) to yield yellow substitution products, which turn orange upon addition of alkali.



3. Millon's test. Red colour develops when proteins are heated with Hg(NO₃)₂ in HNO₃. The reaction is specific for tyrosine and takes place between mercuric and mercurous nitrates and tyrosine residues of the protein. Tryptophan also responds to this reaction.

4. Hopkins-Cole test or Glyoxylic acid test. Violet ring develops on addition of conc. H₂SO₄ (36 N) at the junction of protein and glyoxylic acid solutions. The test is specific for tryptophan.

5. Folin's test. Blue colour develops with phosphomolybdotungstic acid in alkaline solution due to the presence of phenol group. The test is specific for tyrosine.

6. Sakaguchi test. Red colour develops with α-naphthol and sodium hypochlorite. The test is applied for the detection of arginine.

7. Pauly test. Red colour develops with diazotized sulfanilic acid in alkaline solution. The reaction is specific for tyrosine and histidine.

8. Ehrlich test. With *p*-dimethylaminobenzaldehyde in 12 N HCl, tryptophan develops a blue colour.

F. REACTIONS INVOLVING SH GROUP

1. Nitroprusside test. Red colour develops with sodium nitroprusside in dilute NH₄OH. The test is specific for cysteine.

2. Sullivan test. Cystine develops red colour in the presence of sodium 1, 2-naphthoquinone-4-sulfonate and sodium hydrosulfite.

The various colour reactions for different amino acids are summarized in Table 11-2.

Table 11-2. Colour reactions for specific amino acids

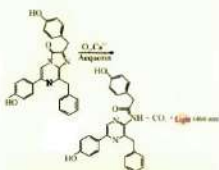
S.N.	Test	Reagent	Colour	Amino acid(s)
1.	Biuret test	Alkaline CuSO ₄ soln.	Violet	Peptide bonds
2.	Xanthoproteic test	Conc. HNO ₃	Yellow	Tyrosine Tryptophan Phenylalanine
3.	Millon's test	Hg(NO ₃) ₂ in HNO ₃ ; heat	Red	Tyrosine Tryptophan

Enzyme

- Importance
- Historical Resume
- Nomenclature and Classification
- Isoenzymes or Isozymes
- Multienzyme Systems
- Biological Roles of Enzymes

Enzymes-I

Nomenclature and Classification



The activity of an enzyme is responsible for the glow of the luminescent jellyfish at left shown above.

The enzyme aequorin catalyzes the oxidation of a compound by oxygen in the presence of calcium to release CO₂ and light.

[Fred Bavendam/Peter Arnold]

IMPORTANCE

Life is an intricate meshwork involving a perfect coordination of a vast majority of chemical reactions. Some of these reactions result in synthesizing large molecules, others in cleaving large molecules and all of them either utilize energy or liberate energy. All these reactions occur very slowly at the low temperatures and the atmospheric pressures—the conditions under which living cells carry on their life processes. Yet in the living cells these reactions proceed at extremely high rates. This is due to the presence of some catalysts produced and synthesized inside the body of the organisms. The term 'enzyme' was coined in 1878 by Friedrich Wilhelm Kuhne to designate these 'biological catalysts' that had previously been called 'ferments'. As they quicken most of the chemical reactions occurring in the body, the enzymes have been designated as the 'nature's impatience'.

In fact, Kuhne intended for the name **enzyme** to apply to both yeast ferment and the extracellular catalysts of more complex animals, e.g., *pepsin* and *trypsin*. Prophetically, his definition specifically implied that lower and higher forms of life are not so fundamentally different.

358 / 1254

The name '**enzyme**' (*en*^G = in ; *zyme*^G = yeast) merely means 'in yeast'. This was referred to denote one of the most noteworthy reactions wherein the production of ethyl alcohol and carbon dioxide through the agency of an enzyme, the

333

Contents

334 FUNDAMENTALS OF BIOCHEMISTRY

zymase, present in yeast takes place. This reaction is most popularly known as *alcoholic fermentation*. Sumner and Myrbäck (1950) have beautifully defined the enzymes as 'simple or combined proteins acting as specific catalysts'. They are soluble, colloidal molecules which are produced by living cells. All enzymes are globular proteins with a complex 3-D structure, capable of binding substrate molecules to a part of their surface. They affect the life of an organism to such an extent that life has aptly been called as "an orderly function of enzymes."

As an analogy, if one considers a living cell as a factory, the individual enzymes might be compared to the machines that work together to cause the transformation of a raw material (like steel) into parts of a finished product (like an automobile).

Difference from catalysts. Like catalysts, the enzymes do not alter the chemical equilibrium point of a reversible reaction but only the speed of the reaction is changed. They, however, differ from catalysts in being the biological products, i.e., produced from the living cells. Moreover, the enzymes are all protein and, unlike catalysts, cannot last indefinitely in a reaction system since they, being colloidal in nature, often become damaged or inactivated by the reactions they catalyze. Henceforth, they must be replaced constantly by further synthesis in the body. Furthermore, unlike catalysts, most individual enzymes are very specific in that they act either on a single or at the most on some structurally-related substrates.

Difference from vitamins. All enzymes are proteinaceous in nature and differ from vitamins in the fact that the latter are not synthesized by the animal cells, in contrast to the former.

Endoenzymes and exoenzymes. Most of the enzymes usually act within the cells in which they are produced and hence are called **intracellular enzymes** or **endoenzymes**, e.g. most of the plant enzymes. As these enzymes catalyze the metabolic reactions of the cell, they are also referred to as **metabolic enzymes**. On the other hand, certain enzymes which are liberated by living cells, catalyze useful reactions outside the cell in its environment and hence are known as **extracellular enzymes** or **exoenzymes**, e.g., enzymes found in bacteria, fungi and some insectivores like *Drosera* and *Nepenthes*. They act chiefly as digestive enzymes, catalyzing the breakdown of complex substances to simpler ones which can readily be absorbed by the cell.

LOUIS PASTEUR (1822-1895)

Louis Pasteur, a French chemist and biologist, was born on December 27, 1822 in Dole France into a family of poor tanners. His father Joseph Pasteur was a sergent-major in Napoleon's army and his valour had earned him the Legion of Honour. Louis did not seem to be particularly outstanding at school but was an excellent artist and made numerous portraits of his family. He studied at the Paris school, Ecole Normale Supérieure, a school founded by Napoleon to train professors, and graduated as a Ph.D. in Sciences. In 1848, he achieved distinction in organic chemistry for his discovery that tartaric acid, a C-4 compound, forms two different types of crystals and successfully separated the two forms while looking through a microscope. He formulated a fundamental law : asymmetry differentiates the organic world from the mineral world. His work became the basis of a new science—**stereochemistry**. In January 1849, he became an Assistant Professor of Chemistry at the University of Strasbourg and soon after in May, 1849 married Marie Laurent, daughter of the University principal. She often discussed his work, spurred his thinking and was one of the best scientific collaborators. The couple had 5 children, but only 2 survived to adulthood. And in September 1854, at the age of 32, he was appointed full-fledged Professor of Chemistry and Dean at the University of Lille in Northern France.

Three years later (i.e., in 1857), he was called back and appointed as Research Director at the Ecole Normal on the Rue d'Ulm in Paris. As a researcher, he showed



such as :

- (a) Trypsin from beef pancreas by John H. Northrop and Kunitz (1936)
- (b) Catalase from beef liver by Sumner and Dounce (1937)
- (c) RNase from beef pancreas by Kunitz (1940)
- (d) Pepsin from swine stomach by John H. Northrop (1946)
- (e) DNase from beef pancreas by Kunitz (1950)

NOMENCLATURE AND CLASSIFICATION

With the continuous increase in our knowledge of enzymology, various systems have evolved to name and classify the enzymes, using one or the other criterion as the basis. However, many of the enzymes were known before these systems of naming enzymes were adopted. The names of such enzymes were not changed under the new systems. In this category belong : bromelin, chymotrypsin, diastase, emulsin, papain, pepsin, ptyalin, rennin, trypsin etc.

A. Substrate acted upon by the enzyme. The substance upon which an enzyme acts is called the substrate. Duclaux (1883) named the enzymes by adding the suffix *-ase* in the name of the substrate catalyzed. For example, enzymes acting upon carbohydrates were named as *carbohydrases*, upon proteins as *proteinases*, upon lipids as *lipases*, upon nucleic acids as *nucleases* and so on. A few of the names were even more specific like *maltase* (acting upon maltose), *sucrase* (upon sucrose), *urease* (upon urea), *lecithinase* (upon lecithin), *tyrosinase* (upon tyrosine) etc.

B. Type of reaction catalyzed. The enzymes are highly specific as to the reaction they catalyze. Hence, this has necessitated their naming by adding the suffix *-ase* in the name of the reaction; for example *hydrolases* (catalyzing hydrolysis), *isomerases* (isomerization), *oxidases* (oxidation), *dehydrogenases* (dehydrogenation), *transaminases* (transamination), *transaldolases* (transaldolization), *transketolases* (transketolization), *phosphorylases* (phosphorylation) etc.

Although these two systems are quite simple and easy to follow, there are certain discrepancies present in them. The former system does not take into account the type of the reaction catalyzed, whereas in the latter system no idea can be derived regarding the nature of the substrate acted upon by the enzyme.

C. Substrate acted upon and type of reaction catalyzed. The names of some enzymes give clue of both the substrate utilized and the type of reaction catalyzed. For example, the enzyme *succinic dehydrogenase* catalyzes the dehydrogenation of the substrate succinic acid. Similarly, *L-glutamic dehydrogenase* indicates an enzyme catalyzing a dehydrogenation reaction involving *L-glutamic acid*.

D. Substance that is synthesized. A few enzymes have been named by adding the suffix *-ase* to the name of the substance synthesized, viz., *rhodanase* that forms rhodanate irreversibly from hydrocyanic acid and sodium thiosulphate, and also *fumarase* that forms fumarate irreversibly from *L-malate*.

In fact, the consensus at that time in Europe was that enzymes were not proteins, because of the findings of the influential German chemist **Richard Willstätter** (Nobel Laureate, 1915), who, in 1920s, reported that he could not detect protein in purified enzyme preparations from yeast. With hindsight, it is realized that the protein assays used in that era were not sensitive enough to detect the small amounts present in Willstätter's purified preparations. The nonprotein nature of enzyme was so entrenched in scientific thinking that decades passed before their polypeptidyl composition was unequivocally accepted.

363 / 1254

Contents

NOMENCLATURE AND CLASSIFICATION OF ENZYMES 339

E. Chemical composition of the enzyme. Based on their chemical composition, the enzymes have been classified into following three categories :

1. Enzyme molecule consisting of protein only—e.g., pepsin, trypsin, urease, papain, amylase etc.
2. Enzyme molecule containing a protein and a cation—e.g., carbonic anhydrase (containing Zn^{2+} as cation), arginase (Mn^{2+}), tyrosinase (Cu^{2+}) etc.
3. Enzyme molecule containing a protein and a nonprotein organic compound known as prosthetic group—Tauber (1950) has further subdivided them, on the basis of the nature of prosthetic group involved :
 - (a) Iron prophyrin enzymes—catalase, cytochrome c peroxidase I and II.
 - (b) Flavoprotein enzymes—glycine oxidase, pyruvate oxidase, histamine.
 - (c) Diphosphothiamin enzymes— β -carboxylase, pyruvate mutase.
 - (d) Enzymes requiring other coenzymes—phosphorylase, amino acid decarboxylase.

F. Substance hydrolyzed and the group involved.

1. Carbohydrate-hydrolyzing enzymes
 - (a) Glycosidases—cellulase, amylase, sucrase, lactase, maltase
 - (b) β -glucuronidase
2. Protein-hydrolyzing enzymes
 - (a) Peptide bonds
 - I. Endopeptidases
Animals—pepsin, trypsin, rennin
Plants—papain, ficin, bromelin
 - II. Exopeptidases—dipeptidase, tripeptidase
 - (b) Nonpeptide C—N linkages (amidases)
urease, arginase, glutaminase
3. Lipid-hydrolyzing enzymes
lipases, esterases, lecithinases
4. Other ester-hydrolyzing enzymes
 - (a) Phosphatases
 - (b) Cholinesterases
 - (c) Chlorophyllases
 - (d) Sulfatases
 - (e) Pectinesterases
 - (f) Methylases
5. Oxidation-reduction enzymes
hydrases, mutases, oxidases, dehydrogenases, peroxidases
6. Miscellaneous enzymes
catalase, carboxylase, carbonic anhydrase, thiaminase, transpeptidase

G. Over-all chemical reaction taken into consideration.

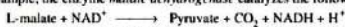
The chemical reaction catalyzed is the specific property which distinguishes one enzyme from another. In 1961, *International Union of Biochemistry (I.U.B.)* used this criterion as a basis for

The I.U.B. system is based on the report of the Commission on Enzymes, International Union of Biochemistry Symposium, Vol. 20, Pergamon Press, New York, 1961. In fact, the international symbolic language of chemistry was originally developed by **Jöns Jacob Berzelius**, who proposed that an element be identified by the initial letter or first two letters of its Latin name.

the classification and naming of enzymes. Although complicated, the I.U.B. system is precise, descriptive and informative.

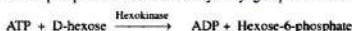
The major features of this system of classification of enzymes are as follows :

- The reactions and the enzymes catalyzing them are divided into 6 major classes, each with 4 to 13 subclasses.
- Each enzyme name has 2 parts—the first part is the name of the substrate(s) and the second part which ends in the suffix *-ase*, indicates the type of reaction catalyzed.
- Additional information regarding the nature of the reaction, if needed, is given in parenthesis. For example, the enzyme *malate dehydrogenase* catalyzes the following reaction :



This enzyme has now been designated as L-malate : NAD oxidoreductase (decarboxylating).

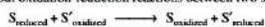
- Each enzyme has been allotted a systemic code number called Enzyme Commission (E.C.) number. The E.C. number for each enzyme consists of a series of numbers at 4 places : the first place numbers representing the major class to which the enzyme belongs, the two median numbers denoting the subclass and the sub-subclass of the enzyme within the major class. The last place number or the fourth digit represents the serial number of the enzyme within the sub-subclass. Thus E.C. 2.7.1.1 represents class 2 (a transferase), subclass 7 (transfer of phosphate), sub-subclass 1 (an alcohol group as phosphate acceptor). The final digit denotes the enzyme, *hexokinase* or *ATP: D-hexose-6-phosphotransferase*. This enzyme catalyzes the transfer of phosphate from ATP to the hydroxyl group on carbon 6 of glucose.



- Where no specific category has been created for an enzyme, it is listed with a final figure of 99 in order to leave space for new subdivisions. For example, 4.2.99 refers to "other carbon-oxygen lyases."

The 6 major classes of enzymes with some important examples from some subclasses are described below :

1. Oxidoreductases. This class comprises the enzymes which were earlier called dehydrogenases, oxidases, peroxidases, hydroxylases, oxygenases etc. The group, in fact, includes those enzymes which bring about oxidation-reduction reactions between two substrates, S and S'.



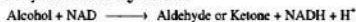
More precisely, they catalyze electron transfer reactions. In this class are included the enzymes catalyzing oxidoreductions of CH-OH , C=O , CH-CH , CH-NH_2 and CH=NH groups. Some important subclasses are :

1.1 Enzymes acting on CH-OH group of electron donor. For example :

1.1.1.1 Alcohol : NAD oxidoreductase

[common or trivial name, Alcohol dehydrogenase]

This enzyme catalyzes the following reaction :



1.3 Enzymes acting on CH-CH group of electron donor. For example :

1.3.2.2 Acyl-CoA : cytochrome c oxidoreductase

[Acyl-CoA dehydrogenase]

Acyl-CoA + oxidized cytochrome c $\longrightarrow$ 2,3-dihydroacyl-CoA + reduced cytochrome c

1.9 Enzymes acting on the heme groups of electron donors. For example :

1.9.3.1 Cytochrome c : O_2 oxidoreductase

[Cytochrome oxidase]

365 / 1254

[Contents](#)

NOMENCLATURE AND CLASSIFICATION OF ENZYMES 341

4. reduced cytochrome + $\text{O}_2 + 4\text{H}^+ \longrightarrow$ 4 oxidized cytochrome c + $2\text{H}_2\text{O}$

1.11 Enzymes acting on H_2O_2 as electron acceptor. For example :

1.11.1.6 H_2O_2 : H_2O_2 oxidoreductase

[Catalase]



2. Transferases. Enzymes which catalyze the transfer of a group, G (other than hydrogen) between a pair of substrates, S and S' are called transferases.



In these are included the enzymes catalyzing the transfer of one-carbon groups, aldehydic or ketonic residues and acyl, glycosyl, alkyl, phosphorus or sulfur-containing groups. Some important subclasses are :

2.3 Acyltransferases. For example :

2.3.1.6 Acetyl-CoA : choline O-acetyltransferase

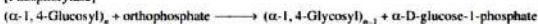
[Choline acetyltransferase]



2.4 Glycosyltransferases. For example :

2.4.1.1 α -1, 4-Glucan : orthophosphate glycosyl transferase

[Phosphorylase]



2.7 Enzymes catalyzing the transfer of phosphorus-containing groups. For example :

2.7.1.1 ATP : D-hexose-6-phosphotransferase

[Hexokinase]



3. Hydrolases. These catalyze the hydrolysis of their substrates by adding constituents of water across the bond they split. The substrates include ester, glycosyl, ether, peptide, acid-anhydride, C—C, halide and P—N bonds. Representative subclasses are :

3.1 Enzymes acting on ester bonds. For example :

3.1.1.3 Glycerol ester hydrolase

[Lipase]



3.2 Enzymes acting on glycosyl compounds. For example :

3.2.1.23 β -D-galactoside galactohydrolase

[β -galactosidase]



3.4. Enzymes acting on peptide bonds

Here the classical trivial names (pepsin, trypsin, thrombin, plasmin etc.) have been largely retained due to their consistent long usage and also due to dubious specificities which make systematic nomenclature almost impractical at this time.

3.5 Enzymes acting on C—N bonds, other than peptide bonds. For example :

3.5.3.1 L-arginine ureohydrolase

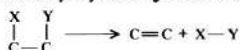
[Arginase]



Although not an oxidizing enzyme, **catalase** is usually classified with oxidases because its action is closely connected with physiological oxidation.

342 FUNDAMENTALS OF BIOCHEMISTRY

4. Lyases (= Desmolases). These are those enzymes which catalyze the removal of groups from substrates by mechanisms other than hydrolysis, leaving double bonds.



In these are included the enzymes acting on C—C, C—O, C—N, C—S and C—halide bonds. Important subclasses include:

4.1 Carbon-carbon lyases. For example:

4.1.2.7 Ketose-1-phosphate aldehyde-lyase
[Aldolase]

A ketose-1-phosphate $\longrightarrow$ Dihydroxyacetone phosphate + an aldehyde

4.2 Carbon-oxygen lyases. For example:

4.2.1.2 L-malate hydro-lyase
[Fumarase]

L-malate $\longrightarrow$ Fumarate + H₂O

4.3 Carbon-nitrogen lyases. For example:

4.3.1.3 L-histidine ammonia-lyase
[Histidase]

L-histidine $\longrightarrow$ Urocanate + NH₃

5. Isomerases. These catalyze interconversions of optical, geometric or positional isomers by intramolecular rearrangement of atoms or groups. Important subclasses are:

5.1 Racemases and epimerases. For example:

5.1.1.1 Alanine racemase

L-alanine $\longrightarrow$ D-alanine

5.2 *Cis-trans* isomerases. For example:

5.2.1.3 *All trans*-retinene 11-*cis-trans* isomerase
[Retinene isomerase]

All *trans*-retinene $\longrightarrow$ 11-*cis*-retinene

5.3 Intramolecular oxidoreductases. For example:

5.3.1.9 D-glucose-6-phosphate keto-isomerase
[Glucosephosphate isomerase]

D-glucose-6-phosphate $\longrightarrow$ D-fructose-6-phosphate

6. Ligases (ligare^t = to bind) or Synthetases. These are the enzymes catalyzing the linking together of two compounds utilizing the energy made available due to simultaneous breaking of a pyrophosphate bond in ATP or a similar compound. This category includes enzymes *participating* in reactions forming C—O, C—S, C—N and C—C bonds. Important subclasses are:

6.2 Enzymes catalyzing formation of C—S bonds. For example:

6.2.1.1 Acetate : CoA ligase (AMP)

[Acetyl-CoA synthetase]

ATP + acetate + CoA $\longrightarrow$ AMP + pyrophosphate + acetyl-CoA

6.3 Enzymes catalyzing formation of C—N bonds. For example:

6.3.1.2 L-glutamate : ammonia ligase (ADP)

[Glutamine synthetase]

ATP + L-glutamate + NH₃ $\longrightarrow$ ADP + orthophosphate + L-glutamine

367 / 1254



NOMENCLATURE AND CLASSIFICATION OF ENZYMES 343

6.4 Enzymes catalyzing formation of C—C bonds. For example:

6.4.1.2 Acetyl-CoA : CO₂ ligase (ADP)

[Acetyl-CoA carboxylase]

ATP + acetyl-CoA + CO₂ + H₂O $\longrightarrow$ ADP + orthophosphate + malonyl-CoA

To date, over 2,000 different enzymes are known, of which the oxidoreductases, transferases and hydrolases predominate. Because official names are often lengthy, the trivial names of enzymes are generally used after initial identification.

ISOENZYMES OR ISOZYMES

Many enzymes occur in more than one molecular form in the same species, in the same tissue or even in the same cell. In such cases, the different forms of the enzyme catalyze the same reaction but since they possess different kinetic properties and different amino acid composition, they can be separated by appropriate techniques such as electrophoresis.

Such multiple forms of the enzymes are called **isoenzymes** or **isozymes**. Isozymes are of widespread nature. Over a hundred enzymes are now known to be of isozymic nature and consequently occur in two or more molecular forms. *Lactic dehydrogenase*, LDH (E.C. No 1.1.1.27), for example, is an enzyme which exists in 5 possible forms in various organs of most vertebrates. LDH catalyzes the reversible oxidation-reduction reaction:



Two basically-different types of LDH are found:

(a) *heart LDH*. This predominates in the heart and is active at low levels of pyruvate. This has 4 identical subunits called *H* subunits (H for heart).

(b) *muscle LDH*. This is characteristic of many skeletal muscles and maintains its activity in much higher concentrations of pyruvate. This also has 4 identical subunits called *M* subunits (*M* for muscle) which are enzymatically inactive.

The two types of subunits, H and M, have the same molecular weight (35,000) but differ in amino acid composition and in immunological properties. The two subunits are produced by two separate genes. LDH can be formed from two types of polypeptide chains designated as H and M subunits to yield a pure H tetramer and a pure M tetramer. Combinations of H and M subunits will, however, produce 3 additional types of hybrid enzymes, thus making the total number of possible forms as five (refer Fig. 16-1). This is confirmed by the fact that when the 2 subunits are mixed in equal proportions, a sequence of 5 bands is obtained by electrophoresis. The 5 different LDH enzyme forms are designated as H₄, H₃M, H₂M₂, HM₃ and M₄. The various isozyme forms of LDH differ significantly in the maximum activities V_{max} in the Michaelis constant Km for their substrates, especially for pyruvate and in the degree of their allosteric inhibition by pyruvate.

The relative concentrations of the LDH enzymes differ from one type of tissue to another, and the relative composition is characteristic of the tissue. For example, LDH-1 plus LDH-2 makes up about 60% of the LDH of cardiac muscle, while LDH-4 plus LDH-5 constitutes about 80% of the LDH of liver. Thus, a diagnosis based on elevated serum LDH can be made more specific with respect to the organ involved by an electrophoretic analysis of the LDH isozymes. For example, the elevated serum level of LDH in a myocardial infarct is primarily due to an increase in LDH-1 and LDH-2; while in liver disease the elevated serum LDH is primarily due to increases in LDH-4 and LDH-5.

Isoenzymes is in fact, the preferable spelling according to the International Union of Biochemistry.

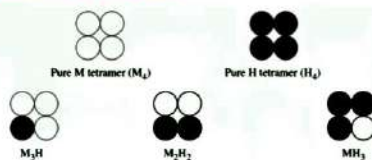


Fig. 16-1. Possible combinations of H and M subunits of lactic dehydrogenase

Damages to tissues and cell death both result in a release of intracellular enzymes from damaged cells into the blood. Therefore, the concentrations of these enzymes increase in the blood serum and can serve as a valuable diagnostic aid in a number of diseases, including myocardial infarction, pancreatitis, liver disease and prostatic cancer.

The serum concentrations of 3 enzymes are often used in diagnosis of myocardial infarction (MI). These are creatinine phosphokinase (CPK), glutamate oxaloacetate transaminase (GOT), and lactate dehydrogenase (LDH). In more than 95% of MI patients, serum levels of GOT rise rapidly and then return to normal in 4–5 days. The serum levels of CPK characteristically rise and fall even more rapidly than GOT levels. However, the serum levels of LDH typically rise and fall more slowly, returning to normal levels in about 10 days.

There seems to be an interesting relation between LDH and the physiological role of flight in birds. In those birds grouped as *short flyers*, over 90% muscle LDH is present in breast muscle. In *long flyers*, breast muscle contains over 95% heart LDH which suggests a possible relation to sustained heart contraction. Those birds classed as *intermediate flyers* contain mixtures of both heart LDH and muscle LDH.

Other noteworthy examples of isozymes are *malic dehydrogenase (MDH)*, *hexokinase*, *esterase* and *glycol dehydrogenase*.

MULTIENZYME SYSTEMS

A few examples of complex enzyme systems are known to exist. These are not independent molecules but occur as aggregates in a mosaic pattern involving several different enzymes. *Pyruvic acid dehydrogenase* of *E. coli* is one such example. This complex molecule has a molecular weight 4,800,000 and consists of 3 enzymes : 24 moles of pyruvate decarboxylase (90,000), 24 moles of pyruvate decarboxylase (90,000), 24 moles of dihydrolipoic dehydrogenase (55,000) and 8 subunits of lipoyl reductase transacetylase (120,000). Each component of this complex enzyme is so arranged as to provide an efficient coupling of the individual reactions catalyzed by these enzymes. In other words, the product of the first enzyme becomes the substrate of the second and so on.

BIOLOGICAL ROLES OF ENZYMES

The enzymes find many applications in our daily life. Enzymatic processes such as baking, brewing and tanning have been known from antiquity. The manifold applications of the enzymes are described below :

1. **Wine manufacturing.** Much of the early interest in enzymology was developed like Pasteur, Payen and Persoz, who were associated with food, wine, and Pasteur was perhaps best known to the French nation as the "saviour of the " because his pasteurization process salvaged an ailing industry beset with problems of microbial contamination. *Papain* is used in brewing industry as a stabilizer for chill-proof beer, because it removes small amounts of protein that cause turbidity in chilled beer.

369 / 1254

Contents

NOMENCLATURE AND CLASSIFICATION OF ENZYMES 345

2. **Cheese making.** Since long the animal rennin (or *rennet*) is employed in making cheese. The enzyme rennet is obtained on a commercial scale from the fourth or true stomach of the unweaned calves which are specifically slaughtered for this purpose. One calf produces only 5 to 10 gm of rennet. The enzyme helps in coagulating the casein of milk. Certain preservatives (boric acid, benzoic acid or sodium chloride) are, sometimes, added to prevent decomposition of the enzyme preparations by bacteria. An enzyme *lipase* is added to cheese for imparting flavour to it. Many vegetarians are unaware that the cheese made in India contains animal rennet. However, an international charitable trust concerned with the welfare of animals, the Beauty Without Cruelty (BWC) has, with the help of Aurey Dairy, Mumbai, undertaken successful experimental trials in cheese making using nonanimal rennet.
3. **Candy making.** An enzyme, *invertase* helps preventing granulation of sugars in soft-centred candies. Another enzyme, *lactase* prevents formation of lactose crystals in ice cream which would otherwise not allow the product seem sandy in texture.
4. **Bread whitening.** *Lipoxygenase* is used for whitening the bread.
5. **Clarifying fruit-juices.** The enzymes are being used in processing of fruit juices such as apple juice and grape juice. The juices are clarified by adding a mixture of *pectic enzymes* which hydrolyze the pectic substances causing turbidity.
6. **Tenderizing meat.** Because hydroxyprolyl residues create bends in collagen helices, which contribute to the tough and rubbery texture often associated with cooked meat, treating the meat with a protease (*bromelain* or *papain*) prior to its cooking hydrolyzes peptide bonds, and thus tenderizes it.
7. **Desizing fabrics.** The woven fabrics are sized by applying starch to the warp (lengthwise) threads to strengthen the yarn before weaving. But when these fabrics are printed or dyed, the sizing should be removed. Desizing may be done by acids, alkalis or enzymes. Enzymatic desizing is, however, preferred as it does not weaken the fabrics. Enzymes for this purpose are obtained from a variety of sources including bacteria, fungi and malt.
8. **Destaining fabrics.** In drycleaning, the stains due to glue, gelatin or starch are removed by employing certain enzymes, such as *alcalase*.
9. **Dehairing hide.** In the manufacture of leather, the hide is made free from hair. This is done by employing *pancreatic enzymes* which hydrolyze the proteins of the hair follicles, thus freeing the hair so that it may be easily scraped off from the hide.
10. **Recovering silver.** *Pepsin* is used to digest gelatin in the process of recovering silver from photographic films.
11. **Correcting digestion.** When the enzymes are present insufficiently in the body, certain digestive disorders come up. These may be cured by supplying the lacking enzymes. *Pepsin*, *papain* and *amylases* aid digestion in the stomach while *pancreatic enzymes* act in the duodenum.
12. **Wound healing.** *Proteolytic enzymes* from pig pancreas are used to alleviate skin diseases, bed sores and sloughing wounds. These enzymes act by destroying proteolytic enzymes of man, that prevent the healing of such wounds. The enzymes commonly used for wound debridement are the proteases such as *streptodornase*, *ficin* and *trypsin*.

Rennin is a milk-curdling enzyme found in the stomach of ruminant mammals and is probably widely distributed among other mammals. By clotting and precipitating milk proteins, it apparently slows the movement of milk through the stomach. Human infants, who lack rennin, digest milk proteins with acidic pepsin, just as adults do. Rennin should not be confused with another phonetically similar enzyme called *renin* which is produced by the kidney. Renin catalyzes the synthesis of angiotensins which cause vasoconstriction in the kidneys, thereby causing electrolyte and water retention in the body.

346 FUNDAMENTALS OF BIOCHEMISTRY

13. **Analyzing biochemicals.** Certain enzymes are used in clinical analysis. For example, *uricase* and *urease* are employed in the determination of uric acid and urea respectively in blood. Besides, sucrose and raffinose contents in sugar mixtures are determined by polarimetry before and after treating the solutions with the enzymes, *sucrase* and *melibiase*.
14. **Dissolving blood clot.** The enzyme *urokinase*, which is manufactured from urine, is being used effectively in Japan in the treatment of blood clot in brain, artery and other circulatory diseases. A team of Soviet scientists led by Yevgeni Chazov, Director of the National Cardiological Research Centre, Moscow, have, in 1982, developed an effective enzyme *streptokinase*, which can dissolve blood clots in vessels. The new enzyme is particularly useful in preventing heart attacks as clots are responsible for 9 out of 10 fatal cases of cardiac arrests.
15. **Changing the blood type.** In 1981, Prof. Ken Furukawa and his associates of the Gunma University's Medical School, Japan have successfully employed several types of specific enzymes in an epoch-making experiment to freely change human blood types. They found that the composition of polysaccharide on the surface of blood corpuscles determines each person's type of blood. Different kinds of sugar characteristics of each blood type form on the surface of RBCs due to the function of a synthetic enzyme. If the sugar is separated from the surface of RBCs by using a specific decomposition enzyme, type A blood and type B blood can be reverted to type O, the prototype of the two blood types. If this breakthrough can be put to practical use, it will fulfill a long-cherished dream of doctors to administer blood transfusions irrespective of the type of blood a patient has by merely changing the patient's blood type to match the blood available.
16. **Diagnosing hypertension.** A new method called radio immunoassay procedure for diagnosing cases of hypertension has been developed by Bhabha Atomic Research Centre (BARC). In it, the activity of *renin*, a proteolytic enzyme secreted by the kidneys, is calculated indirectly by measuring angiotensin-I which is formed by the action of renin. Renin acts as part of a complex feedback mechanism for regulating blood volume and pressure.
17. **Augmenting surgery.** A technique using the enzyme *trypsin* as an adjunct to cataract surgery has been developed in 1980 by Dr. Joseph Spina Jr. of Philadelphia. With older techniques it required an incision about 2.5 cm long in the white of the eye to remove the clouded lens. Modern microsurgery has, however, reduced this cut to only 0.3 cm. But Dr. Spina's method involves a still smaller cut wide enough for a needle 0.025 cm wide. The hollow needle is used to inject a microscopic amount of trypsin, a digestive enzyme secreted by the pancreas. Trypsin digests and liquefies the semisolid interior of the lens without harming other parts of the eye. Once the enzyme liquefies the lens—which takes from a few hours to overnight—the lens is removed by suction through the same hollow needle. This eliminates the necessity of intervention in the eye, the constant passing in and out of the instruments and suturing. The lesser the tissue is wounded, the quicker it recovers. This enzyme surgery for cataracts could be done as an outpatient operation. The patients would come in one day to have the enzyme injected and return the next day to have the cataract removed.
18. **Breaking down chemicals.** Recently, in 1993, a group of scientists from the Netherlands led by Han G. Brunner have found a tiny genetic defect that appears to predispose some men toward aggression, impulsiveness and violence. The afflicted persons often react to the most

CLINICAL IMPLICATIONS

A **cataract** is a clouding of eye lens which can cause blindness. It is the most common eye disease and comes with age. The lens of the eye is made of protein and as time goes by, there occurs denaturation process and the protein gets cloudy and one cannot see. India, at present, has 18 million blind people, out of which 12 million are cataract patients.

NOMENCLATURE AND CLASSIFICATION OF ENZYMES 347

mildly stressful occasions with aggressive outbursts, cursing or assaulting the persons they deem a threat. The researchers have linked the abnormal behaviours to mutations in the gene responsible for the body's production of *monoamine oxidase-a*, an enzyme critical for breaking down chemicals that allow brain cells to communicate. It is proposed that lacking the enzyme, the brains of afflicted men end up with excess deposits of potential signalling molecules like serotonin, dopamine and noradrenaline. Those surplus neurotransmitters, in turn, stimulate often hostile conduct. The erratic behaviour is due to point mutation. The gene is on the X chromosome, which explains why only males, with their single copy of the X chromosome, can suffer from the enzyme deficiency. Women can serve as carriers of the genetic defect, but are themselves protected from its symptoms by their possession of a second, good copy of the gene, sitting on their second X chromosome. Although the number of persons afflicted with this disease is not known but based on other types of hereditary disorders, the researchers estimate that the illness is likely to be quite rare in the general population, i.e., no more than one in 1,00,000 people.

19. **Destroying acids.** Sprouts are the natural health boosters. They are basically the young new plants and are, in fact, the organic answer to simple natural health. Almost any edible seed can be sprouted. Far from being the invention of food faddists, sprouting dates back to 2939 B.C. in China. Sprouts are found in all shapes and colours and it is best to choose these from legume plants. Sprouting greatly improves the safety and nutritional quality of all pulses, seeds and grains. The enzymes which go into action during sprouting not only neutralize trypsin-inhibiting factors but also destroy harmful acids like phytic acid. Phytic acid, an integral constituent of grains, tends to bind minerals, making them unavailable to the body.
20. **Syrup manufacturing.** An *immobilized enzyme* is one that is physically entrapped or covalently-bonded by chemical means to an insoluble matrix, e.g., glass beads, polyacrylamide or cellulose. Immobilization of an enzyme often greatly enhances its stability, which makes its prolonged catalytic life a valuable industrial trait. These days, *immobilized glucose isomerase* is being successfully used in the production of high-fructose corn syrup, esp., in the United States.

Table 16-1 outlines some of industrial applications of enzymes. This is a rapidly changing field, and new applications of enzyme technology appear all the time.

Table 16.1 Some Applications of Enzymes

Enzyme	Reaction	Source of enzyme	Application
Industrial applications			
α -amylase	breaks down starch	bacteria	converts starch to glucose in the food industry
Glucose isomerase	converts glucose to fructose	fungi	production of high fructose syrups
Proteases	digests protein	bacteria	washing powder
Rennin	clots milk protein	animal stomach linings; bacteria	cheese making
Catalase	splits hydrogen peroxide into $H_2O + O_2$	bacteria; animal livers	turns latex into foam rubber by producing gas
β -galactosidase	hydrolyses lactose	fungi	in dairy industry, hydrolyses lactose in milk or whey

seem a threat. The researchers have linked the abnormal behaviours to mutations in the gene responsible for the body's production of *monamine oxidase-a*, an enzyme critical for breaking down chemicals that allow brain cells to communicate. It is proposed that lacking the enzyme, the brains of afflicted men end up with excess deposits of potential signalling molecules like serotonin, dopamine and noradrenaline. Those surplus neurotransmitters, in turn, stimulate often hostile conduct. The erratic behaviour is due to point mutation. The gene is on the X chromosome, which explains why only males, with their single copy of the X chromosome, can suffer from the enzyme deficiency. Women can serve as carriers of the genetic defect, but are themselves protected from its symptoms by their possession of a second, good copy of the gene, sitting on their second X chromosome. Although the number of persons afflicted with this disease is not known but based on other types of hereditary disorders, the researchers estimate that the illness is likely to be quite rare in the general population, *i.e.*, no more than one in 1,00,000 people.

19. **Destroying acids.** Sprouts are the natural health boosters. They are basically the young new plants and are, in fact, the organic answer to simple natural health. Almost any edible seed can be sprouted. Far from being the invention of food faddists, sprouting dates back to 2939 B.C. in China. Sprouts are found in all shapes and colours and it is best to choose these from legume plants. Sprouting greatly improves the safety and nutritional quality of all pulses, seeds and grains. The enzymes which go into action during sprouting not only neutralize trypsin-inhibiting factors but also destroy harmful acids like phytic acid. Phytic acid, an integral constituent of grains, tends to bind minerals, making them unavailable to the body.
20. **Syrup manufacturing.** An *immobilized enzyme* is one that is physically entrapped or covalently bonded by chemical means to an insoluble matrix, *e.g.*, glass beads, polyacrylamide or cellulose. Immobilization of an enzyme often greatly enhances its stability, which makes its prolonged catalytic life a valuable industrial trait. These days, *immobilized glucose isomerase* is being successfully used in the production of high-fructose corn syrup, *esp.*, in the United States.

Table 16-1 outlines some of industrial applications of enzymes. This is a rapidly changing field, and new applications of enzyme technology appear all the time.

Table 16.1 Some Applications of Enzymes

Enzyme	Reaction	Source of enzyme	Application
Industrial applications			
α -amylase	breaks down starch	bacteria	converts starch to glucose in the food industry
Glucose isomerase	converts glucose to fructose	fungi	production of high fructose syrups
Proteases	digests protein	bacteria	washing powder
Rennin	clots milk protein	animal stomach linings; bacteria	cheese making
Catalase	splits hydrogen peroxide into $H_2O + O_2$	bacteria; animal livers	turns latex into foam rubber by producing gas
β -galactosidase	hydrolyses lactose	fungi	in dairy industry, hydrolyses lactose in milk or whey

372 / 1254

Contents

348 FUNDAMENTALS OF BIOCHEMISTRY

Medical applications

L-asparaginase	removes L-asparagine from tissues - this nutrient is needed for tumour growth	bacteria (<i>E. coli</i>)	cancer chemotherapy - particularly leukaemia
Urokinase	breaks down blood clots	human urine	removes blood clots, <i>e.g.</i> , in heart disease patients

Analytical applications

Glucose oxidase	oxidises glucose	fungi	used to test for blood glucose, <i>e.g.</i> , in Clinistix TM diabetes
Luciferase	produces light	marine bacteria; fireflies	binds to particular chemicals indicating their presence, <i>e.g.</i> , used to detect bacterial contamination of food

Manipulative applications

Lysozyme	breaks 1-4 glycosidic bonds	hen egg white	disrupts bacterial cell walls
Endonucleases	breaks DNA into fragments	bacteria	used in genetic manipulation techniques, <i>e.g.</i> , gene transfer, DNA Finger printing.

REFERENCES

See list following Chapter 18.

PROBLEM

1. Which of the reactions listed below are catalyzed by an isomerase, lyase, hydrolase or transferase?
 - (a) Protein $\longrightarrow$ Amino acids
 - (b) Histidine $\longrightarrow$ Histamine + CO_2
 - (c) Glucose + ATP $\longrightarrow$ Glucose-6-phosphate + ADP
 - (d) Glucose-6-phosphate $\longrightarrow$ Glucose + H_2PO_4
 - (e) $CH_3COCOCH_2COOH + 2H^+ \longrightarrow CH_3CHOHCOCH_2COOH$
 - (f) $CH_3COCOCH_2COOH \longrightarrow CH_3CHO + CO_2$
 - (g) 3-phosphoglycerate $\longrightarrow$ 2-phosphoglycerate
 - (h) Tryptophan $\longrightarrow$ Tryptamine + CO_2
 - (i) L-lysine $\longrightarrow$ D-lysine
 - (j) Acetylcholine $\longrightarrow$ Acetic acid + choline
 - (k)
$$\begin{array}{c} OH \\ | \\ H_3C-C-COOH \\ | \\ H \end{array} \longrightarrow \begin{array}{c} H \\ | \\ H_3C-C-COOH \\ | \\ OH \end{array}$$
 - (l) Glutaric acid + oxaloacetate $\longrightarrow$ α -ketoglutarate + aspartic acid

- Chemical Nature
- Characteristics
 - Colloidal Nature
 - Catalytic Nature
 - Specificity of Enzyme Action
 - Thermolability
 - Reversibility
 - pH Sensitivity
- Three Dimensional Structure of the Enzymes
 - Ribonuclease
 - Lysozyme
 - Chymotrypsin
 - Trypsin

Enzymes-II

Characteristics and 3 'D' Structure

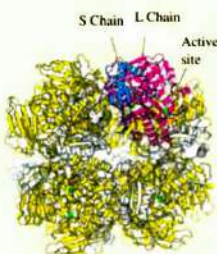
CHEMICAL NATURE

All the enzymes are essentially proteins and possess properties characteristic to these. Dixon and Webb (1964) have stressed the protein nature of an enzyme by defining it as "a protein with catalytic properties due to its power of specific activation".

Evidences Proving the Protein Nature of the Enzymes

- Elementary composition.** In their elementary composition, the enzymes show the usual proportion of C, H, N and S, as found in the proteins. Some crystalline enzymes, however, also contain minute quantities of P or metal ions such as Cu^{2+} , Mg^{2+} , Zn^{2+} , etc. On hydrolysis, the crystalline enzymes yield the amino acids.
- Identical action of some enzymes over other enzymes and the proteins.** Enzymes are subjected to the action of those enzymes which are specifically meant for the breakdown of peptide bonds of proteins.
- Amphoteric nature.** Like other proteins, the enzymes behave as ampholytes in an electric field. The isoelectric point (pI) for various enzymes has also been determined.
- Denaturation.** Enzymes, like other π undergo denaturation. If the crystal 'chymotrypsin' is subjected to an un 374 / 1254 some part of protein becomes de. percentage of denatured protein is usually be equal to the per cent loss in enzymic activity, thus

349



Ribbon model of the tertiary structure of ribulose 1,5-bisphosphate carboxylase/oxygenase (rubisco)

The enzyme comprises eight large subunits (one shown in the red and the others in yellow) and eight small subunits (one shown in blue and the others in white). The active sites lie in the large subunits.

Contents

350 FUNDAMENTALS OF BIOCHEMISTRY

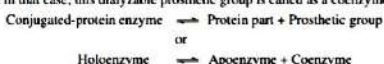
proving a sort of correlation between the enzymes and the proteins.

- Formation of antibodies.** Many purified enzymes, on injection into animal body, produce the specific antibodies. Since many nonprotein materials have been shown to serve as antigens, this cannot be treated as an evidence in support of the protein nature of enzymes but simply a further support to it.

Chemically, the enzymes may be divided into 2 categories :

- Simple-protein enzymes.** These contain simple proteins only e.g., *urease*, *amylase*, *papain* etc.
- Complex-protein enzymes.** These contain conjugated proteins i.e., they have a protein part called *apoenzyme* (*apo* = away from) and a nonprotein part called *prosthetic group* associated with the protein unit. The two parts constitute what is called a *holoenzyme*, e.g., *catalase*, *cytochrome c* etc.

The activity of an enzyme depends on the fact that the non-proteinaceous prosthetic group is intimately associated with the proteinaceous apoenzyme. But sometimes the prosthetic group is loosely bound to the protein unit and can be separated by dialysis and yet indispensable for the enzyme activity. In that case, this dialyzable prosthetic group is called as a *coenzyme* or *cofactor*. Thus :



Coenzymes are thermostable, dialyzable organic compounds. They may be either attached to the protein molecules or may be present in the cytoplasm. The coenzyme accounts for about 1% of the entire enzyme molecule. Sometimes, a distinction is made between coenzymes and cofactors ; the former includes the organic prosthetic groups and the latter the metal ions (Fairley and Kilgour, 1966).

CHARACTERISTICS

The enzymes possess many outstanding characteristics. These are enumerated below :

- Colloidal Nature.** Enzyme molecules are of giant size. Their molecular weights range from 12,000 to over 1 million. They are, therefore, very large compared with the substrates or functional group they act upon (Fig. 17-1).

It has been observed that the molecular weights of many enzymes prove to be approximately an n -fold multiple (where n is an integer) of 17,500 which is found to be an unit in most proteins (Table. 17-1).

On account of their large size, the enzyme molecules possess extremely low rates of diffusion and form colloidal systems in water. Being colloidal in nature, the enzymes are nondialyzable although some contain dialyzable or dissociable component in the form of coenzyme.

- Catalytic Nature or Effectiveness.** An universal feature of all enzymatic reactions is the virtual absence of any side products. Therefore, just as hemoglobin is precisely tailored to transport

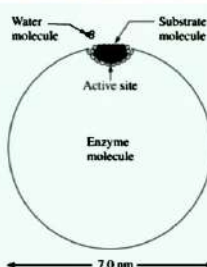


Fig. 17-1. Relative dimensions of a medium-sized enzyme molecule (MW 1,00,000 ; diameter 7 nm) and a typical substrate molecule (MW 250 ; diameter 0.8 nm)

[Note that the active site occupies only a small fraction of the surface area of the enzyme molecule. Also shown for comparison is a water molecule.]

oxygen, an enzyme is precisely adapted to catalyze a particular reaction. They act catalytically and accelerate the rate of chemical reactions occurring in plant and animal tissues. They do not normally participate in these reactions or if they do so, at the end of the reaction, they are recovered as such without undergoing any qualitative or quantitative change. This is the reason why they, in very small amounts, are capable of catalyzing the transformation of a large quantity of substrate. Thus, *the catalytic potency of enzymes is exceedingly great.*

Table 17-1. Molecular weight of some enzymes

Enzyme	Molecular weight	n°
Pepsin	35,500	2
Catalase	250,000	14
Urease	480,000	27

* n = an integer, which is a multiple of 17,500

The catalytic power of an enzyme is measured by the "turnover number" (a term devised by Weechszahl) or **molecular activity** (a term devised by Norman Arthur Edwards and Kenneth Arnold Hassall, 1980) which is defined as—the number of substrate molecules converted into product per unit time, when the enzyme is fully saturated with substrate. For example, a single molecule of *catalase* can convert 50,000,000 H_2O_2 molecules into H_2O and O_2 in a minute (Sumner and Somers, 1947). The value of turnover number varies with different enzymes and depends upon the conditions in which the reaction is taking place. However, for most enzymes, the turnover numbers fall between 1 to 10^6 per second (ref. Table 17–2). The turnover number of 600,000 sec^{-1} for carbonic anhydrase is one of the largest known. Carbonic anhydrase (Fig. 17–2) catalyzes the hydration of carbon dioxide to produce 3,600,000 molecules of carbonic acid per minute. This catalyzed reaction is 6×10^6 times faster than the uncatalyzed one.

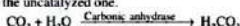


Fig. 17-2. Ribbon model of the tertiary structure of human carbonic anhydrase

The α helices are represented as cylinders and each strand of β sheet is drawn as an arrow pointing towards the polypeptide's C-terminus. The grey ball in the middle represents a Zn^{2+} ion that is coordinated by three His side chains (blue). Note that the C-terminus is tucked through the plane of a surrounding loop of polypeptide chain so that carbonic anhydrase is one of the rare native proteins in which a polypeptide chain forms a knot.

(Courtesy: Kannan KK et al. 1971)

Table 17-2. Maximum turnover numbers of some enzymes

Enzyme	Turnover (per sec.)
1. Lysozyme	0.5
2. Tryptophan synthetase	2
3. DNA polymerase I	15
4. Phosphohexokinase	20.5

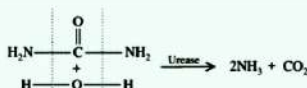
376 / 1254

352 FUNDAMENTALS OF BIOCHEMISTRY

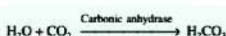
5. Chymotrypsin	100
6. β -galactosidase	208
7. Lactate dehydrogenase	1,000
8. Penicillinase	2,000
9. β -amylase	18,333
10. Acetylcholinesterase	25,000
11. Carbonic anhydrase	600,000

3. Specificity of Enzyme Action. With few exceptions, the enzymes are specific in their action. Their specificity lies in the fact that they may act (a) on one specific type of substrate molecule or (b) on a group of structurally-related compounds or (c) on only one of the two optical isomers of a compound or (d) on only one of the two geometrical isomers. Accordingly, four patterns of enzyme specificity have been recognized :

A. Absolute specificity. Some enzymes are capable of acting on only one substrate. For example, *urease* acts only on urea to produce ammonia and carbon dioxide.



Similarly, *carbonic anhydrase* brings about the union of carbon dioxide with water to form carbonic acid.



B. Group specificity. Some other enzymes are capable of catalyzing the reaction of a structurally-related group of compounds. For example, *lactic dehydrogenase* (LDH) catalyzes the interconversion of pyruvic and lactic acids and also of a number of other structurally-related compounds.

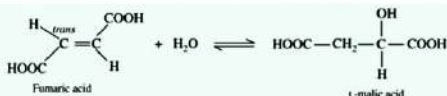


C. Optical specificity. The most striking aspect of specificity of enzymes is that a particular enzyme will react with only one of the two optical isomers. For example, *arginase* acts only on L-arginine and not on its D-isomer. Similarly, *D-amino acid oxidase* oxidizes the D-amino acids only to the corresponding keto acids.

Although, the enzymes exhibit optical specificity, some enzymes, however, interconvert the two optical isomers of a compound. For example, *alanine racemase* catalyzes the interconversion between L- and D-alanine.

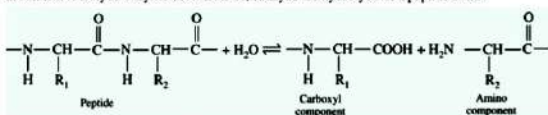


D. Geometrical specificity. Some enzymes exhibit specificity towards the *cis* and *trans* forms. As an example, *fumarase* catalyzes the interconversion of fumaric and malic acids :

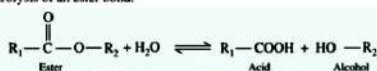


It does not react with maleic acid which is the *cis* isomer of fumaric acid or with D-malic acid.

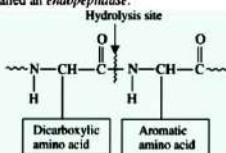
The degree of specificity of the enzymes for substrate is usually high and sometimes virtually absolute. Proteolytic enzymes, for instance, catalyze the hydrolysis of a peptide bond:



Many proteolytic enzymes (pepsin, trypsin, chymotrypsin) catalyze a different but related reaction, namely the hydrolysis of an ester bond.

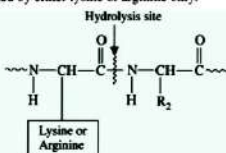


These enzymes vary markedly in their degree of specificity. For example, *subtilisin*, a bacterial enzyme, does not discriminate the nature of the side chains adjacent to the peptide bond to be cleaved. Another enzyme *pepsin* prefers bonds involving the carboxyl and amino groups of dicarboxylic and aromatic amino acids respectively. Since the bonds attached are usually located in the interior of the protein substrate, pepsin is called an *endopeptidase*.



Specificity of pepsin

Trypsin, likewise, is an endopeptidase but is quite specific in that it splits peptide bonds in which carboxyl group is contributed by either lysine or arginine only.



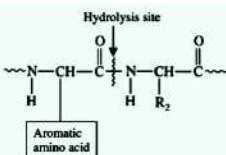
Specificity of trypsin

378 / 1254

Contents

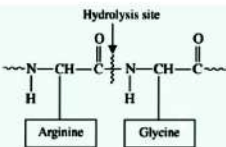
354 FUNDAMENTALS OF BIOCHEMISTRY

Chymotrypsin preferentially splits peptide bonds in which the carboxyl group is from an aromatic amino acid.



Specificity of chymotrypsin

Thrombin, an enzyme involved in blood coagulation, is even more specific in that the side chain on the carboxyl side of the susceptible peptide bond must be arginine whereas the one on the amino side must be glycine.



Specificity of thrombin

Alteration of enzyme specificity – The specificity of some enzymes is altered by physiological behaviour. *Lactose synthetase* (Fig. 17-3), for example, catalyzes the synthesis of lactose (a sugar consisting of a galactose and a glucose residue) in the mammary glands. It consists of a catalytic subunit and a modifier subunit. The *catalytic subunit* alone cannot synthesize lactose. Instead, it has a different role of catalyzing the attachment of galactose to proteins that contain a covalently linked carbohydrate chain. The *modifier subunit* alters the specificity of the catalytic subunit so that it links galactose to glucose to form lactose. The level of modifier subunit is under hormonal control. During pregnancy, the catalytic subunit is formed in the mammary glands and very little modifier subunit is formed. But at the time of childbirth (= parturition), the hormonal levels change significantly and the modifier subunit is synthesized in great quantities, thus resulting in the production of large amounts of lactose.

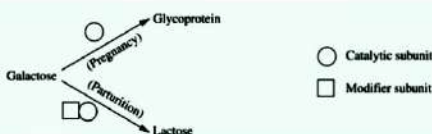


Fig. 17-3. Alteration in enzyme specificity of lactose synthetase

There are, however, instances where one enzyme acts on more than one substrate and conversely a substrate may also be catalyzed by more than one enzyme (Fig. 17-4). For example,

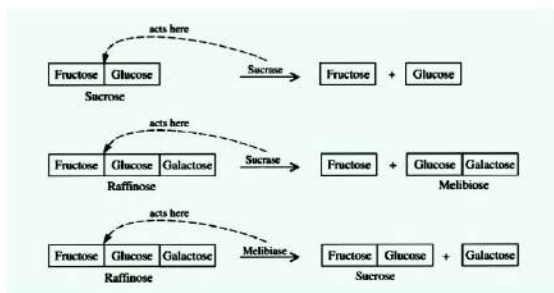


Fig. 17-4. Enzyme specificity of sucrase and melibiase

sucrase acts on both sucrose (a disaccharide sugar, containing one mole of glucose and fructose each) and raffinose (a trisaccharide sugar, containing one mole each of glucose, fructose and galactose). But in both these cases, the enzyme sucrase attacks only the glucose-fructose linkage resulting in the production of fructose and glucose (in the case of sucrose) or fructose and melibiose (in the case of raffinose). However, raffinose is also acted upon by another enzyme, the *melibiase*. But this enzyme, unlike sucrase, breaks up glucose-galactose linkage so that at the end of the reaction sucrose and galactose are produced.

The coenzymes possess much less specificity. For example, among the *hydrolases*, NAD^+ and NADP^+ act as common coenzymes. The relative nonspecificity of the coenzymes, in contrast to the absolute specificity of the enzymes, can be visualized by comparing coenzyme to a common hammer, used equally by various apoenzyme workers (ironworker, watchmaker, shoemaker or electrician). Although some differences may occur in the nature of hammers (e.g., between NAD^+ and NADP^+), the apoenzyme worker is strictly specific, corresponding to the nature of the substrate concerned.

4. Thermolability (= Heat sensitivity).

Being proteinaceous in nature, the enzymes are very sensitive to heat. The rate of an enzyme action increases with rise in temperature, the rate being frequently increased 2 to 3 times for a rise in temperature of 10°C , i.e., the value of temperature quotient or Q_{10} is 2 to 3. But at higher temperatures, the value of coefficient does not remain constant and decreases rapidly. Above 60°C , the enzymes coagulate and thus become inactivated, because there occurs an irreversible change in their chemical structure. The enzymes of dry tissues like seeds and spores, however, can endure still higher temperatures of about 100° to 120°C .

A temperature difference of 10°C has become a standard that is used to measure the temperature sensitivity of a biological function. This value, called the **temperature quotient (Q_{10})**, is determined (for temperature intervals of exactly 10°C) by dividing the value of a rate function at the higher temperature by the value at the lower temperature. Purely physical processes, such as diffusion, have much lower Q_{10} values, usually close to 1.

380 / 1254

356 FUNDAMENTALS OF BIOCHEMISTRY

The observed effect of temperature on enzyme action is the net result of the effect of temperature on the rate of enzyme action and their destruction as well. There will, thus, be obtained an optimum temperature for the enzyme action (Fig. 17-5). The curve AB represents the effect of temperature on action alone and the curve CD, the temperature effect on enzyme destruction. The observed relation between the temperature and the rate of enzyme action will then be represented by the curve AE.

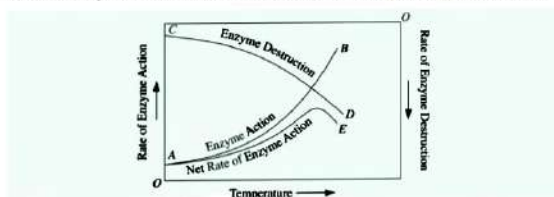


Fig. 17-5. Graph representing the relation between temperature and the rate of enzyme activity (After Duclaux, 1883)

If, however, the effect of increasing temperature (in terms of three ill-demarcated categories of low, medium and high) on enzyme activity is studied (Fig. 17-6), it may be observed that the initial velocity of the reaction (given by the shape of the curves at $t = 0$) steadily increases with temperature. However, after a certain temperature is passed, the cessation of activity comes earlier and earlier with the result that less product is formed. There is, thus, a somewhat ill-defined optimum region of temperature which is that at which the two factors of increased initial rate and decreased active life of the enzyme are balanced to produce the most product in a reasonable time. It is not easy to determine the exact value for the optimum temperature because it is somewhat vague concept, and will depend on the length of time over which the measurements are made. However, the approximate values obtained often show a distinct correlation with the body temperatures of the organisms from which the enzyme came. Thus, mammalian enzymes often have optimum temperatures in the range 35 – 45°C , while the enzymes from the bacteria that live in volcanic hot springs may have optima of 80°C .

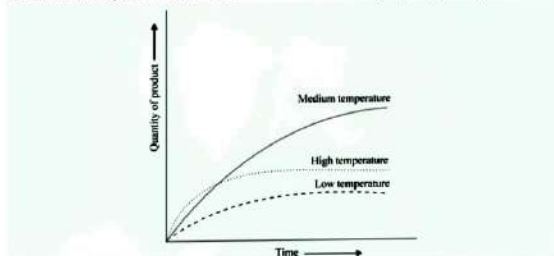


Fig. 17-6. Effect of temperature (coupled with time) on enzyme activity

In general, the optimum temperature range for most enzymes varies between 30 and 45°C . For

instance, it is 30°C for *catalase*. Because enzymes are globular proteins, most are thermolabile and begin to denature (indicated by loss of enzyme activity) at temperatures between 45° and 50°C (Fig. 17-7).

At low temperatures, the catalytic activity of the enzyme predominates, although some thermal denaturation takes place during this period. Decreasing temperatures to near or below 0°C although inactivate the enzyme but this is a reversible type of change and the enzyme regains its catalyzing power upon increasing the temperature to optimum. At higher temperature, although the catalytic activity of the enzyme increases, yet its denaturation predominates. Henceforth, all the enzyme is denatured in a very short time. The enhanced enzyme activity with the rise in temperature is due to the fact that the energy of molecule becomes greater which, in turn, enhances the inherent reactivity of the molecules and the frequency of their collisions. It is because of the high rate of enzyme destruction at increasing temperatures that an enzyme is stable for weeks at 0°C, for days at 10°C, for hours at 30°C but for fraction of seconds at 70°C.

The effect of heat also manifests itself in the preservation of enzyme activity during storage. The best preservation of enzyme preparations is by refrigeration or quick freezing. This has been shown by Nord (1932) in the case of *zymase*.

5. Reversibility of a Reaction. The enzymes are capable of bringing about reversion in a chemical reaction. The digestive enzymes catalyze the hydrolytic reactions which are reversible. For instance, *lipase*, which catalyzes the synthesis of fat from glycerol and fatty acid, can also hydrolyze them into their component units.

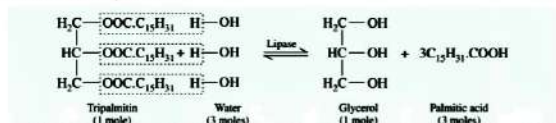


Fig. 17-8 shows the results of experiments on the action of *lipase* from castor on a fat, triolein. The final equilibrium mixture is the same whether one starts with the ester or with its individual components.

The direction in which the reaction proceeds depends upon many factors like –

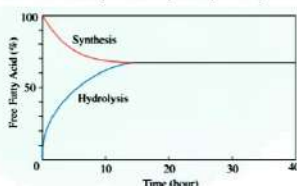


Fig. 17-8. Reversibility of action of *lipase* (from castor) on triolein

382 / 1254

358 FUNDAMENTALS OF BIOCHEMISTRY

- the pH of the cell sap,
- the presence of reacting substances, and
- the accumulation of end products.

It does not, however, necessarily follow that the same enzyme invariably catalyzes both the synthesis and degradation of a given kind of molecule. For instance, urea is synthesized from arginine by the action of the enzyme, *arginase* but is hydrolyzed by action of another enzyme, *urease* to produce ammonia and carbon dioxide.

6. pH Sensitivity. The pH value or the H^+ ion concentration of the medium controls the activity of an enzyme to a great extent. This is mainly related to the degree of dissociation, to the electric charge of the enzyme and, through this, to the formation of the enzyme-substrate complex (a discussion of which will follow in the succeeding chapter). Each enzyme, thus, acts best in a certain pH value which is specific to it and its activity slows down with any appreciable change (increase or decrease) in the H^+ ion concentration. In fact, the pH will affect the efficiency of an enzyme and usually there will be a pH at which the activity is at a maximum. The activity will fall off on either side of this value. Fig. 17-9 depicts the effect of pH on an enzyme-catalyzed reaction.

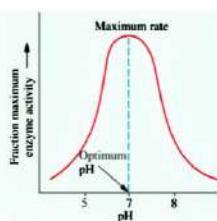


Fig. 17-9. Hypothetical pH activity profile of an enzyme

Table 17-3. pH optima for various enzymes

S.No.	Enzyme	Optimum pH of the medium	Nature of the medium
1.	Pepsin	1.5-1.6	Highly acidic
2.	Invertase	4.5	Acidic
3.	Lipase (stomach)	4.0-5.0	Acidic
4.	Lipase (castor oil)	4.7	Acidic
5.	Lipase (pancreas)	8.0	Alkaline
6.	Amylase (malt)	4.6-5.2	Acidic
7.	Amylase (pancreas)	6.7-7.0	Acidic-neutral
8.	Cellobiase	5.0	Acidic
9.	Maltase	6.1-6.8	Acidic
10.	Sucrase	6.2	Acidic
11.	Catalase	7.0	Neutral
12.	Urease	7.0	Neutral
13.	Cholinesterase	7.0	Neutral
14.	Ribonuclease	7.0-7.5	Neutral
15.	Fumarase	7.8	Alkaline
16.	Trypsin	7.8-8.7	Alkaline
17.	Adenosine triphosphatase	9.0	Alkaline
18.	Arginase	10.0	Highly alkaline

Some optimum pH values for various enzymes are given in Table. 17-3. A perusal of the table indicates that the approximate optimum pH value for most enzymes lies near neutrality. This value depends on many factors such as :

- the nature of buffer system,
- the presence of other colloids, activators or inhibitors,
- the age of the cell tissue, and
- the nature of the substrate.

Usually maximum enzyme activity is obtained at or near the isoelectric point (pI) of the enzymes. Thus *trypsin*, whose pI value is 10.1, shows maximum activity at pH range between 7 and 9.

The correlation between the enzymic activity and the pH value for 3 enzymes has been graphically represented in Fig. 17-10.

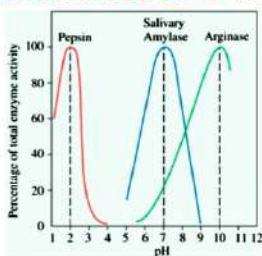


Fig. 17-10. Effect of pH on enzyme action (Modified from Fruton and Simmonds, 1958)

THREE DIMENSIONAL STRUCTURE OF THE ENZYMES

A single crystal of protein or the protein fibres will deflect x-rays and the resultant image formed on a photographic plate can give certain important clues regarding the structure of the crystal or the fibres. This technique is called **x-ray crystallography** and has been widely used for the elucidation of protein structure at micro level. X-ray crystallography has so far revealed the structure of many enzymes, namely, *ribonuclease*, *lysozyme*, *chymotrypsin*, *trypsin* etc. The structure of four of them is described below :

1. Ribonuclease (RNase)

Ribonuclease (Fig. 17-11), a small globular protein, is an enzyme secreted by the pancreas into the small intestine, where it catalyzes the hydrolysis of certain bonds in ribonucleic acids present in ingested food.

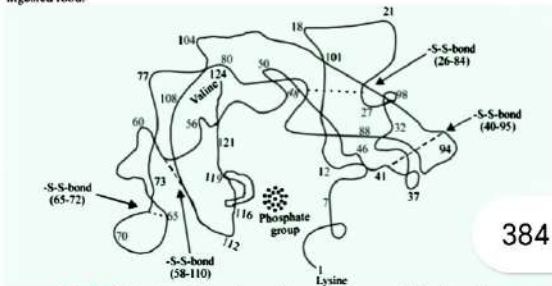


Fig. 17-11. Structure of ribonuclease as determined from x-ray diffraction studies. Numbers refer to specific amino acid residues.

(From Harper and Rodwell, 1973)

384 / 1254

360 FUNDAMENTALS OF BIOCHEMISTRY

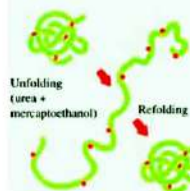


Fig. 17-12. Denaturation and refolding of ribonuclease

A native ribonuclease molecule (with intramolecular disulfide bonds indicated) is reduced and unfolds with β -mercaptoethanol and 8 M urea. After removal of these reagents, the protein undergoes spontaneous refolding. (From: CJ Epstein, RF Goldberger and CB Anfinsen, 1963).

It is interesting to note that there is remarkable similarity in structure between the ribonucleases from cows and human beings (Fig. 17-13). The structural similarity is often followed by functional similarity.

The bacterial enzyme, *subtilisin*, cleaves the chain into 2 inactive fragments : the shorter one (S-peptide) consisting of first 21 amino acids from N-terminal and the longer one (S-protein) with remaining residues. On reunion of the two segments, the enzyme molecule regains its full activity upon treatment with 8 M urea and mercaptoethanol, the native ribonuclease molecule become reduced and unfolded. After removal of these reagents, the enzyme undergoes spontaneous refolding (Fig. 17-12).

It is interesting to note that there is remarkable similarity in structure between the ribonucleases from cows and human beings (Fig. 17-13). The structural similarity is often followed by functional similarity.

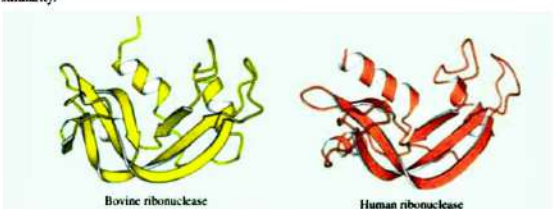


Fig. 17-13. Ribbon diagrams of the structure of ribonucleases from cows and human beings. Structural similarity often follows functional similarity.

2. Lysozyme (= Muramidase)

Lysozyme (Figs. 17-12 and 17-13), another small globular protein, is an enzyme present in tears, nasal mucus, gastric secretions, milk and egg white. Lysozyme is so named because it can

364 FUNDAMENTALS OF BIOCHEMISTRY

A molecule of chymotrypsin (MW = 25,000) consists of 3 short polypeptide chains (A, B, and C) of 11, 131 and 97 amino acid residues respectively, connected by two inter-chain disulfide bonds between 1-122 and 136-201 and three intra-chain disulfide bonds between 42-58, 168-182 and 191-220 amino acid residues (Fig. 17-18). The 3-dimensional structure of the enzyme was elucidated at 2 Å resolution by the x-ray crystallographic studies of David Blow (Fig. 17-19). The molecule is a compact ellipsoid of dimensions $51 \times 40 \times 40$ Å. Chymotrypsin consists of several antiparallel β pleated sheet regions and, unlike myoglobin and hemoglobin, has little α helical structure. All charged

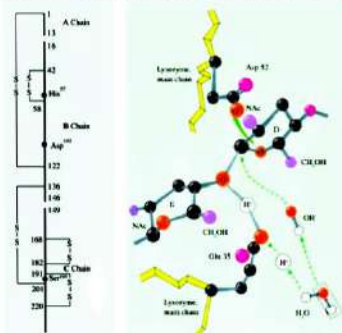


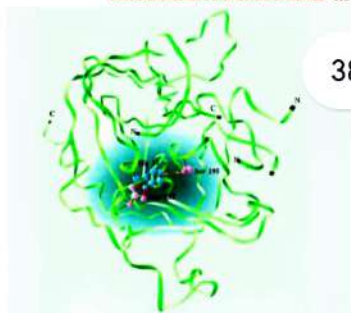
Fig. 17-18. Representation of primary structure of chymotrypsin.

Note the presence of 3 polypeptide chains with 5 disulfide bonds, of which 2 are inter-chain and 3 are intra-chain. Located 3 amino acid residues forming catalytic triad is shown. The active-site amino acids are found grouped together in the 3-D structure.

17-18. The mechanism for lysine action as proposed by Phillips. The bond between the fourth (D) and fifth (E) sugar of the hexacarbohydrate residing in the cleft of the lysine molecule is cleaved by acid hydrolysis using a proton donated by the carboxyl group of the closely applied glutamic acid residue (Glu 35). The formation of the positively charged oxocarbenium ion at the C1 position of sugar D is facilitated by the donation of the sugar shown in the figure and stabilized by the nearby aspartic acid residue (Asp 52) of the enzyme. In the final step, the oxocarbenium ion reacts with an OH group from the solvent.

(From D. Blow and G. Blow, 1995)

CHARACTERISTICS AND 3-D STRUCTURE OF ENZYMES 365



17-20. Ribbon model of the tertiary structure of chymotrypsin.

The amino- and carboxyl-termini of the 3 constituent chains are labelled as N and C respectively. The tertiary structure of chymotrypsin places the essential amino acid residues close to one another. They are shown as ball-and-stick representations.

(From D. Blow and G. Blow, 1995)

groups are on the surface of the molecule except for three (His⁵⁷, Asp¹⁰² and Ser¹⁹⁵) that play a critical role in catalysis. A tertiary (or 3-dimensional) structure of chymotrypsin molecule in ribbon form is shown in Fig. 17-20.

Double Displacement Mechanism. Chymotrypsin, like many proteases, hydrolyzes ester bonds, in addition to peptide bonds. The hydrolysis of peptide or ester bonds takes place by a two-step displacement with an amine being produced first, followed by production of an acid. The two steps of this double displacement mechanism are:

First step: Acylation: Formation of the acetyl-enzyme complex.

p-nitrophenylacetate (p-NPA) combines with chymotrypsin to form an enzyme-substrate (ES) complex. The ester bond of the substrate then cleaves. One of the products, *p*-nitrophenol is released from the enzyme, whereas the acetyl group of the substrate becomes covalently attached to the enzyme.

Second step: Deacylation: Hydrolysis of the acetyl-enzyme complex.

366 FUNDAMENTALS OF BIOCHEMISTRY

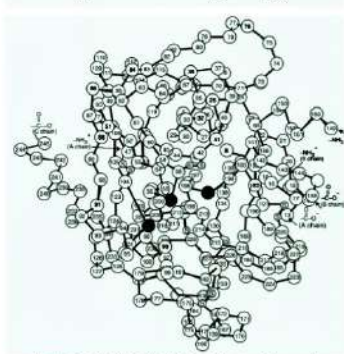
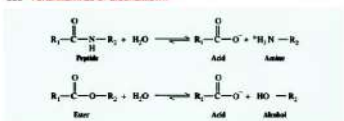


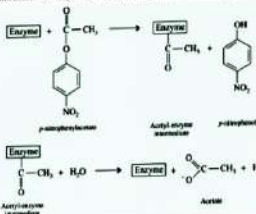
Fig. 17-21. Ball-and-stick model of the three-dimensional structure of chymotrypsin.

Only the α carbon atoms are shown. Residues of the catalytic triad (His⁵⁷, Asp¹⁰² and Ser¹⁹⁵) are labelled. The hydrophobic pocket of the substrate is indicated by the dark residues.

(After R.E. Dickerson and J. Drenth, 1969)

CHARACTERISTICS AND 3-D STRUCTURE OF ENZYMES 367

Water then attacks the acetyl-enzyme complex to yield acetate ion and regenerate the enzyme



The second step (deacylation) is much slower than the first step (acylation), so that it determines the overall rate of hydrolysis of esters by chymotrypsin. The acetyl-enzyme complex is sufficiently stable to be isolated under proper conditions. The catalytic mechanism of chymotrypsin can, thus, be represented by:



where P_1 is the amine (or alcohol) component of the substrate, $E-P_1$ is the covalent intermediate, and P_2 is the acid component of the substrate. A distinct feature of this mechanism is the appearance of a covalent intermediate. In the first-step reaction, an acetyl group is covalently bonded to the enzyme and the group attached to chymotrypsin at $E-P_1$ stage is an acyl group. Thus, $E-P_1$ is an acyl-enzyme intermediate.

Catalytic Triad. Proteolytic enzymes containing a highly reactive serine residue are known as **serine proteases**. These enzymes are readily identifiable by their rapid inactivation by DIFP. Chymotrypsin, trypsin and elastase are noteworthy examples of this class.

Chymotrypsin contains 28 acyl residues but only one of them (Ser¹⁹⁵) is a strong nucleophile. This is due to a specific spatial relationship between three amino acid residues (His⁵⁷, Asp¹⁰² and Ser¹⁹⁵) which constitute a **catalytic triad** and is based on hydrogen bonding (Fig. 17-26). The hydrogen bonding, that occurs between the buried Asp¹⁰² and His⁵⁷ and between His⁵⁷ and Ser¹⁹⁵, establishes an equilibrium that allows for the loss of the proton of the OH group of Ser¹⁹⁵ (at the catalytic site) to His⁵⁷. This loss makes the oxygen atom of Ser¹⁹⁵ residue a strong nucleophile, i.e., makes serine¹⁹⁵ an active serine. The proton gained by His⁵⁷ converts the side chain of that residue into a positive imidazolium ion that forms an **ion pair** with the negative carboxylate ion of Asp¹⁰². Thus, loss of a proton by Ser¹⁹⁵ and formation of an ion pair by His⁵⁷ and Asp¹⁰² explain the mechanism behind the functioning of catalytic triad.

368 FUNDAMENTALS OF BIOCHEMISTRY

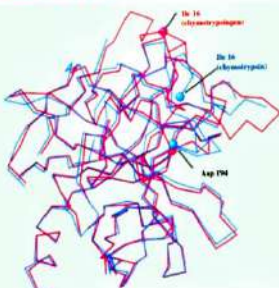


Fig. 17-22. Conformations of chymotrypsinogen (red) and chymotrypsin (blue).

The electrostatic interaction between the carboxylate of aspartate 194 and the π atom group of histidine 56, essential for the reaction of active chymotrypsin, is possible only in chymotrypsin.

4. Trypsin

Trypsin (and another enzyme elastase) are homologues of chymotrypsin. These two have catalytic triads similar to that discovered in chymotrypsin. The catalytic triad in trypsin consists of His 57, Asp 102 and Ser 195 (Fig. 17-25). Their sequences are approximately 80% identical with that of chymotrypsin, and their overall structures are the same (Fig. 17-24). However, they have very different substrate specificities. Trypsin cleaves at the peptide bond after residues with long, positively charged side chains namely, arginine and lysine (Fig. 17-25), whereas elastase, cleaves at the peptide bond after amino acids with small side chains such as alanine and serine. In trypsin, an aspartate residue (Asp 102) is present at the bottom of the S_1 pocket (in place of a serine residue in chymotrypsin). The aspartate residue attracts and stabilizes a positively charged arginine or lysine residue in the substrate.

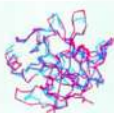


Fig. 17-25. Structural similarity of trypsin and chymotrypsin.

An overlay of the structures of chymotrypsin (red) on that of trypsin (blue) shows the high degree of similarity. Only α -carbon atom positions are shown. The mean deviation in position between corresponding α -carbon atoms is 1.7 Å.

CHARACTERISTICS AND 3-D STRUCTURE OF ENZYMES 369

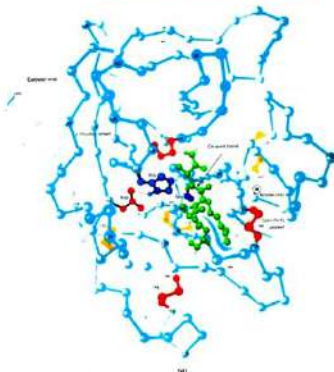


Fig. 17-24. The x-ray structure of bovine trypsin.

- The ball-and-stick model. Each circle represents a single amino acid residue. Numbers refer to specific amino acid residues. The drawing of the enzyme with a polypeptide substrate (green) that has its Arg side chain occupying the enzyme's specificity pocket (toppling). The C_{α} backbone of the enzyme is shown together with its disulfide bonds and the side chains of the catalytic triad: Ser 195, His 57, and Asp 102.
- (over page) The ribbon model. This diagram highlights its secondary structure and indicates the arrangement of its catalytic triad.
- (over page) A computer-generated model. It shows the surface of trypsin (blue) superimposed on its polypeptide backbone (purple). The side chains of the catalytic triad are shown in green.

FIGURE 17-24 and 17-25 have approximately the same orientation.

(Courtesy: (a) Irving Getz, (b) & (c) Arthur Olson)

(b) (see 12 page) The ribbon model. This diagram highlights its secondary structure and indicates the arrangement of its catalytic triad.
(c) (see 12 page) A computer-generated model. It shows the surface of trypsin (blue) superimposed on its polypeptide backbone (purple). The side chains of the catalytic triad are shown in green.
Parts a, b, and c have approximately the same orientation.
(Courtesy : (a) Irving Geis, (b) & (c) Arthur Olson)

Contents

370 FUNDAMENTALS OF BIOCHEMISTRY

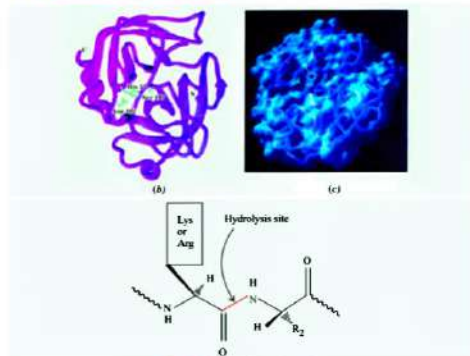


Fig. 17-25. Specificity of trypsin

Trypsin cleaves on the carboxyl side of arginine and lysine residues.

To produce active trypsin, the cells that line the duodenum secrete an enzyme, *enteropeptidase* that hydrolyzes a unique lysine-isoleucine peptide bond trypsinogen as the precursor enters the duodenum from the pancreas. The small amount of trypsin produced in this way activates more trypsinogen and the other zymogens. Thus, the formation of trypsin by enteropeptidase is the master activation step.

It is to be noted that in all the three enzymes described above, the active site is present in the groove or depression. The groove is, in fact, the ideal place for the active site as it provides a nonpolar microenvironment in which alone the van der Waal's forces and the hydrogen bond formation can operate between the polar groups of the active site and the substrate. In the case of an 'exposed' site, on the contrary, the water molecules will interfere with such activity.

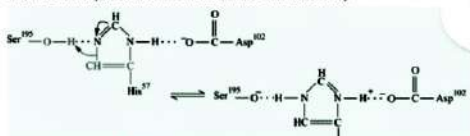


Fig. 17-26. Role of the catalytic triad in chymotrypsin

Note that the catalytic triad is created by the hydrogen-bonding of Ser¹⁹⁵, His⁵⁷ and Asp¹⁰².

395 / 1254

Contents

CHARACTERISTICS AND 3-D STRUCTURE OF ENZYMES 371

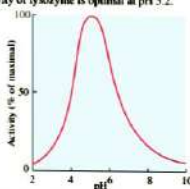
From the foregoing discussion, it may be concluded that the chemical nature of an enzyme depends much on the active site of the enzyme molecule. Apart from the various kinds of constituent amino acid residues comprising an enzyme, its properties further depend on the pattern of '3-D' structure. The typical folding of the polypeptide chain, for example, ensures that the constituent amino acids of the active site are brought closer in the grooves.

REFERENCES

See list following Chapter 18.

PROBLEMS

- The sweet taste of freshly picked corn is due to the high level of sugar in the kernels. Store-bought corn (several days after picking) is not as sweet because about 50% of the free sugar of corn is converted into starch within one day of picking. To preserve the sweetness of fresh corn, the husked ears are immersed in boiling water for a few minutes ("blanched") and then cooled in cold water. Corn processed in this way and stored in a freezer maintains its sweetness. What is the biochemical basis for this procedure?
- The enzymatic activity of lysozyme is optimal at pH 5.2.



The active site of lysozyme contains two amino acid residues essential for catalysis: Glu³⁵ and Asp⁵². The pK_a values of the carboxyl side chains of these two residues are 5.9 and 4.5, respectively. What is the ionization state (protonated or deprotonated) of each residue at the pH optimum of lysozyme? How can the ionization states of these two amino acid residues explain the pH-activity profile of lysozyme shown above?

- Which of the following statements is not universally applicable to enzymes :
(a) they generally work very rapidly.
(b) they can catalyze a reaction in both directions
(c) they are not used up during a reaction.
(d) they will bind one substrate only.
(e) they are proteins.
- Why does a wound on tongue heal faster than those on other parts of the body?

CONTENTS

- Energy Mechanics of Enzymatic Reactions
- Michaelis-Menten Hypothesis
- Active Site
 - Fisher's Lock and Key Model
 - Koshland's Induced Fit Model
- Enzyme Reaction Rates
- Modifiers of Enzyme Activity
 - Inorganic Modifiers
 - (+) Enzyme Activation
 - (-) Enzyme Inhibition
 - Biobioactive Reactions
 - Terminology
 - Types of B/B Reactions
 - Kinetics of B/B Reactions
- Allosteric Enzymes
 - Simple Sequential Model
 - Concerted or Symmetry Model

CHAPTER

18

Enzymes-III

Mechanism of Enzyme Action



Bovine pancreatic ribonuclease B in complex with a non-hydrolyzable substrate analogue. The proximity of the phosphate group.

ENERGY MECHANICS OF ENZYMATIC REACTIONS

A simple enzymatic reaction might be written:



where, E, S and P represent enzyme, substrate and product, respectively. ES and EP are complexes of the enzyme with substrate and with the product, respectively.

For a better understanding of the kinetics of a chemical reaction, the two terms, reaction equilibria and reaction rates must be differentiated. The function of a catalyst is to increase the rate of a reaction. Catalysts do not affect reaction equilibria. Any reaction, such as $S \rightleftharpoons P$, can be described by a **reaction coordinate diagram** (Fig. 18-1). Energy in biological systems is described in terms of free energy, G. In the coordinate diagram, the free energy of the system is plotted against the progress of the reaction. Energy diagrams describe the energetic course of the reaction (vertical axis) and the progressive chemical changes (horizontal axis) as the substrate is converted to product.

S = free energy of the substrate

P = free energy of the product

$\Delta G^\ddagger$ = transition state

$\Delta G^\ddagger$ = activation energies for the two reactions,



ΔG° = overall standard free-energy change in moving from S to P

372

MECHANISM OF ENZYME ACTION 373

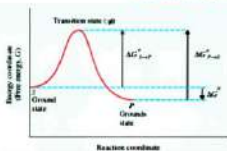


Fig. 18-1. Free energy diagram for a simple chemical reaction, $S \rightarrow P$.

described in terms of free energy, G. In the coordinate diagram, the free energy of the system is plotted against the progress of the reaction (or reaction coordinate). In its normal stable form or **ground state**, any molecule contains a specific amount of free energy. Chemists express the free energy change for this reaction system under standard set of conditions (temperature, 298 K; partial pressure of gases, each 1 atm or 101.3 kPa; pH = 0; concentration of solutes, each 1M) and call this as **standard free-energy change, ΔG°** . Because biochemical systems commonly involve 1M concentrations far from 1M, biochemists define a constant ΔG° , the standard free-energy change at pH 7.0.

The equilibrium between S and P reflects the difference in the free energy of their ground states. In the example shown in Fig., the free energy of the ground state of P is lower than that of S, hence ΔG° for the reaction is negative and the equilibrium favours P. This equilibrium is *not* affected by any catalyst. A favourable equilibrium, however, does not mean that $S \rightarrow P$ conversion is fast. The rate of a reaction, in fact, depends on an entirely different parameter. There exists an energetic barrier between S and P that represents the energy required for alignment of reacting groups, bond rearrangements and other changes needed for the reaction to occur in either direction. To undergo reaction, the molecules must overcome this barrier or "energetic hill" and therefore must be raised to a higher energy level. At the top of energy hill is a point at which decay to the S or P state is equally probable, which is downhill either way. This is called the **transition state** and should not be confused with a reaction intermediate. It is simply a moment of fleeting molecules in which certain events (bond breakage, bond formation, charge development etc) have proceeded to decide the future course of the reaction, i.e., a collapse to either substrate or product. The difference between the energy levels of the ground state and the transition state is called the **Gibbs free energy of activation** or simply **activation energy** and is symbolized by $\Delta G^\ddagger$. The double dagger ($\ddagger$) denotes a thermodynamic quantity of a transition state. The rate of 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 higher energy to overcome this energy barrier. As an alternate, the activation energy can be lowered by adding a catalyst (Fig. 18-2). Thus, catalysts enhance reaction rates by lowering activation energies.

The ES and EP intermediates occupy minima in the energy process curve of the enzyme-catalyzed reaction. The terms $\Delta G_{ES}^\ddagger$ and $\Delta G_{EP}^\ddagger$ correspond to the activation energies for the nonenzymatic and enzymatic reactions. The activation energy for the overall process is lower when the enzyme catalyzes the reaction. Note that for both reactions, the free energy for S and P, called the ΔG° of the reaction, is the same.

The catalysts affect the reaction rates, not the reaction equilibria. And so are the enzymes, the bidirectional arrows put in the equation on page make the point clear: any enzyme that catalyzes the reaction, $S \rightarrow P$ also catalyzes the reverse reaction, $P \rightarrow S$. Its only role is to accelerate the

397-398 / 125

374 FUNDAMENTALS OF BIOCHEMISTRY

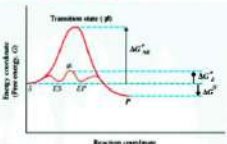


Fig. 18-2. Energy diagram, comparing the nonenzymatic and enzymatic reactions, $S \rightarrow P$. Interconversion of S and P. The enzyme is not consumed in the process, and the equilibrium point remains unaffected. However, the reaction reaches equilibrium much faster when the appropriate enzyme is present because the rate of the reaction is increased.

Practically, any reaction may have several steps involving the formation and decay of transient (unstable) chemical compounds called **reaction intermediates**. When the $S \rightarrow P$ reaction is catalyzed by an enzyme, the ES and EP complexes are intermediates and they occupy valleys in the reaction coordinate diagram (Fig. 18-2). When several steps occur in a reaction, the overall rate is determined by the step(s) with the maximum activation energy. This is called the **rate limiting step**. In practice, the rate limiting step can vary with reaction conditions. It may, however, be inferred from the above discussion that reaction equilibria are intimately linked with ΔG° and reaction rates are linked with $\Delta G^\ddagger$.

The energy of activation is a measure of the energy needed for the conversion of molecules to the reactive state. An inorganic catalyst lowers the activation energy and the organic catalyst (i.e., an enzyme) further decreases this. For example, decomposition of H_2O_2 requires 18,000 cal/mol, this is lowered to 11,700 when colloidal platinum acts as catalyst and is further lowered to 2,000 when an enzyme catalase catalyzes the reaction. It is, thus, evident that enzymic catalysts are far more efficient in lowering the energy of activation than the nonenzymic catalysts for the same reaction. A perusal of Table 18-1 supports the above statement.

Table 18-1. Activation energy for enzymic and nonenzymic catalysts

S.No.	Reaction	Catalyst	Activation Energy (expressed as cal/mol)
1.	Decomposition of hydrogen peroxide	None	18,000
		Colloidal platinum	11,700
		Catalase	2,000
2.	Hydrolysis of ethyl butyrate	Hydrogen ion	16,800
		Hydrazyl ion	10,200
		Lipase	4,500
3.	Hydrolysis of sucrose	Hydrogen ion	25,600
		Invertase	10,000
4.	Hydrolysis of casein	Hydrogen ion	20,600
		Trypsin	12,000

(Adapted from White, Handley and Smith, 1964.)

MICHAELIS-MENTEN HYPOTHESIS

Leonor Michaelis and Maud L. Menten (1913), while studying the hydrolysis of sucrose catalysed by the enzyme *invertase*, proposed this theory. Their theory is, however, based on the following assumptions:

1. Only a single substrate and a single product are involved.
2. The process proceeds essentially to completion.
3. The concentration of the substrate is much greater than that of the enzyme in the system.
4. An intermediate enzyme-substrate complex is formed.
5. The rate of decomposition of the substrate is proportional to the concentration of the enzyme-substrate complex.

The theory postulates that the enzyme (E) forms a weakly bonded complex (ES) with the substrate (S). This enzyme-substrate complex, on hydrolysis, decomposes to yield the reaction product (P) and the free enzyme (E). These reactions may be symbolically represented as follows:



Although one may not feel any difficulty (at least theoretical) in describing the kinetics of these reactions, yet the difficulty is encountered when one starts to determine the concentration of ES or even S practically. The same difficulty was experienced by Michaelis and Menten, who devised an equation where these immeasurable quantities were replaced by those which could be easily measured experimentally.

Following symbols may be used for deriving Michaelis-Menten equation:

(E_t) = total concentration of enzyme

(S) = total concentration of substrate

(ES) = concentration of enzyme-substrate complex

$(E_f) - (ES)$ = concentration of free enzyme

The rate of appearance of products (i.e., velocity, V) is proportional to the concentration of the enzyme-substrate complex.

$$V = k \times (ES) \quad \text{---(1)}$$

The maximum reaction rate, V_m , will occur at a point where the total enzyme E_t is bound to the substrate. Then the maximum concentration of ES will be equal to the total enzyme concentration, E_t . Thus:

$$V_m = k \times (E_t) \quad \text{---(2)}$$

Dividing equation (1) by (2) we get:

$$\frac{V}{V_m} = \frac{(ES)}{(E_t)} \quad \text{---(3)}$$

With the help of this equation, one can easily measure the immeasurable quantities, (ES) and (E_t) , in terms of the reaction rates experimentally.

Now coming back to the reversible reaction, $E + S \rightleftharpoons ES$, one can write the equilibrium constant for dissociation of ES as K_m which is equal to:

$$K_m = \frac{(E_f) \times (S)}{(ES)} \quad \text{---(4)}$$

$$\text{or } (ES) \times K_m = (E_f) \times (S) - (ES) \times (S)$$

400-401 / 1254

376 FUNDAMENTALS OF BIOCHEMISTRY

$$\text{or } (ES) \times K_m + (ES) \times (S) = (E_f) \times (S)$$

$$\text{or } (ES) \times [K_m + (S)] = (E_f) \times (S)$$

$$\text{or } \frac{(ES)}{(E_f)} = \frac{(S)}{K_m + (S)}$$

Substituting the value of $\frac{(ES)}{(E_f)}$ from equation (3) to equation (5), we get:

$$\frac{V}{V_m} = \frac{S}{K_m + S}$$

$$\text{or } V = \frac{V_m \times (S)}{K_m + (S)} \quad \text{---(6)}$$

$$\text{or } K_m = (S) \left[\frac{V_m}{V} - 1 \right] \quad \text{---(7)}$$

Equation (6) is called as **Michaelis-Menten equation**. This can be used to calculate K_m after experimentally determining the reaction rates at various substrate concentrations. This equilibrium constant, K_m , is usually called **Michaelis constant**. It is a measure of the affinity of an enzyme for its substrate. Referring to the equation (4), the greater the concentration of ES complex, the lower is the concentration of free enzyme and consequently the lower is the value of K_m .

For experimental determination of K_m , the velocity of the reaction (relative activity of the enzyme) is measured as a function of substrate concentration. When $V = \frac{1}{2}V_m$, it will be seen from equation (6) that K_m is numerically equal to the substrate concentration or in other words K_m is equal to the concentration of the substrate which gives half the numerical maximal velocity, V_m . Thus, it is possible to determine K_m . The K_m shown in Fig. 18-3, is indicated to be 0.017 M. It is noteworthy that for any enzyme-substrate system, K_m has a characteristic value which is independent of the enzyme concentration.

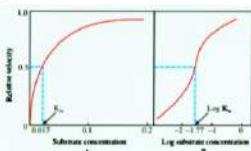


Fig. 18-3. Relative initial velocity as a function of substrate concentration (A) and as a function of the logarithm of the substrate concentration (B) for the action of yeast invertase on sucrose.

(Adapted from White, Handley and Smith, 1964)

Lineweaver-Burk equation. The method described above for the determination of K_m is somewhat complex and, therefore, simpler methods have been devised. Two such methods (Fig. 18-4) are given below:

First method. A convenient means of evaluating K_m and V_m is to plot kinetic data as the reciprocals

of V and (S) . Such a double reciprocal plot was proposed by Hans Lineweaver and Dean Burk in 1934. If one takes the reciprocal of Michaelis-Menten equation, the following equation is obtained:

$$\frac{1}{V} = \frac{K_m + (S)}{V_m \times (S)}$$

$$\text{or } \frac{1}{V} = \frac{K_m}{V_m} \times \frac{1}{(S)} + \frac{1}{V_m} \quad \text{---(8)}$$

This is known as **Lineweaver-Burk equation**. This equation is of the form, $y = mx + b$, if one considers the variables to be $\frac{1}{V}$ and $\frac{1}{(S)}$. When one plots the graph against these two variables, a straight line is obtained. The slope of this line corresponds to $\frac{K_m}{V_m}$ and the $\frac{1}{V_m}$ intercept corresponds to $\frac{1}{V_m}$. Since V_m can be determined from the intercept, the K_m may also be calculated.

Second Method. Another graphical method for the measurement of K_m from experimental data on V as a measure of (S) makes use of the above Lineweaver-Burk equation. Multiplication of both sides of this equation by (S) gives:

$$\frac{(S)}{V} = \frac{K_m}{V_m} + \frac{(S)}{V_m} \quad \text{---(9)}$$

A plot of $\frac{(S)}{V}$ versus (S) gives a straight line. The intercept of the line on $\frac{(S)}{V}$ axis is $\frac{K_m}{V_m}$ and the slope is $\frac{1}{V_m}$.

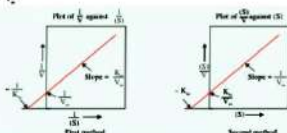


Fig. 18-4. Two methods of obtaining linear plots involving measurement of initial velocity, V and substrate concentration, (S) .

V_m and K_m can be obtained from the intercept on the slope and the abscissa.

(Adapted from Fruton and Amis, 1960)

A Lineweaver-Burk plot provides a quick test for adherence to Michaelis-Menten kinetics and allows easy evaluation of the critical constants. As we shall see, it also allows discrimination between different kinds of enzyme inhibition and regulation. A disadvantage of Lineweaver-Burk plot is that a long extrapolation is often required to determine K_m with corresponding uncertainty in the result. Consequently, other ways of plotting the data are sometimes used. One such alternative is to rearrange equation (6) into the form:

$$V = V_m - K_m \frac{V}{(S)} \quad \text{---(10)}$$

and graph V versus $V/(S)$. This yields what is called an **Eadie-Hofstee plot** (Fig. 18-5).

378 FUNDAMENTALS OF BIOCHEMISTRY

Significance of K_m and V_m Values

The **Michaelis constant** (K_m) values of the enzymes differ greatly. However, for most of the enzymes, the general range is between 10^{-3} and 10^{-6} M (refer Table 18-2). The K_m value depends on the particular substrate and on the environmental conditions such as temperature and ionic concentration.

The K_m value signifies two meanings:

- (a) K_m is the concentration of substrate at which half the active sites are occupied. If K_m is known, the fraction of the sites filled, f_{ES} , at any substrate concentration, can be calculated from:

$$f_{ES} = \frac{V}{V_m + (S)} \quad \dots(11)$$

- (b) K_m is related to the rate constant of the individual steps in the catalytic scheme.

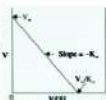


Fig. 18-5 An Enthalp-Haas plot. Graphing V versus $V/[S]$, one obtains V_m at $(V/[S]) = 0$ and K_m from the slope of the line.

Table 18-2 Values of Michaelis constant (K_m) of some enzymes

Enzyme	Substrate	K_m^a
1. Lysozyme	Hexa-N-acetylglucosamine	6×10^{-5} M
2. Penicillinase	Benzylpenicillin	5×10^{-4} M
3. β -galactosidase	Lactose	4×10^{-3} M
4. Chymotrypsin	Acetyl-L-tryptophanamide	5×10^{-3} M
5. Carbonic anhydrase	Carbon dioxide	8×10^{-5} M
6. Pyruvate carboxylase	Pyruvate	4×10^{-3} M
7. Arginine (RNA) cyclotase	ATP	6×10^{-3} M
	RNA	4×10^{-3} M
	Arginine	3×10^{-4} M
8. Hexokinase	ATP	3×10^{-4} M
	D-glucose	4×10^{-3} M
	D-fructose	5×10^{-3} M
9. Sucrose	Sucrose	15×10^{-3} M
10. α -glucosidase	Methyl- α -D-glucoside	0.018 – 0.04 M
11. β -glucosidase	Methyl- β -D-glucoside	0.037 – 0.075 M
12. Pepsin	Ovalbumin	0.060 – 1.12 M
13. Trypsin	Cazim	4.5%
14. Urease	Urea	2.0%
15. Catalase	Hydrogen peroxide	0.025 M
16. Lipase (Esterase)	Ethyl butyrate	0.025 M
17. Phosphatase	Glycerophosphate	< 0.003 M
18. Xanthine oxidase	Xanthine	4×10^{-2} M
19. Succinate dehydrogenase	Succinate	5×10^{-2} M
20. Amylase	Starch	0.8 – 0.25%
21. Diaphorase	Glycylglycine	0.02 – 0.07 M

^a K_m values are not absolute constants but depend on the temperature, substrate and source of enzyme.

MECHANISM OF ENZYME ACTION 379



where k_1 , k_2 and k_3 are the rate constants for the 3 reactions. If k_3 is much greater than k_2 , the dissociation of ES complex is more rapid than formation of P and regenerates

$$K_m = \frac{k_2}{k_1}$$

The dissociation constant of ES complex is given by:

$$K_m = \frac{(E)(S)}{(ES)} = \frac{k_2}{k_1} \quad \dots(13)$$

In other words, K_m is equal to the dissociation constant of the ES complex if k_3 is much smaller than k_2 ($k_3 \ll k_2$). When this condition is reached, K_m is a measure of the strength of ES complex. The high K_m value indicates weak binding whereas the low K_m value signifies strong binding. It may, however, be emphasized that K_m indicates the affinity of the ES complex only when $k_3 \gg k_2$.

The **maximal rate** (V_m) represents the turnover number of an enzyme, if the concentration of the active sites (E_t) is known, since:

$$V_m = k_3(E_t) \quad \dots(14)$$

A 10^{-6} M solution of carbonic anhydrase, for instance, catalyzes the formation of 0.6 M carbonic acid per second when it is fully saturated with substrate. Hence, k_3 is $6 \times 10^5 \text{ sec}^{-1}$. The kinetic constant, k_3 is known as the **turnover number**.

HERMANN EMIL FISCHER

(1852-1919)

Emil Fischer, a German, was the son of a wealthy merchant. He graduated from the Gymnasium of Bonn in 1868. After an abortive foray into the business world, he entered the University of Bonn in 1871 to study chemistry under Frederick August Kekulé (1818-1896), master of structural chemistry. After receiving his doctoral degree from the University of Strasbourg in 1874, he taught at the Universities of Erlangen and Würzburg, eventually becoming Professor of Chemistry at the University of Berlin in 1912 while succeeding Hofmann. He published his discovery of phenylhydrazine in 1873. Later in 1890, he turned to a detailed study of proteins. It was in the study of structural chemistry to the study of proteins that Fischer saw the possible future collaboration of biology and chemistry. He developed the theory of the 'peptide bond', the chemical linkage by which all amino acids are joined together to form their respective proteins. He also chemically synthesized an octadecapeptide, composed of 3 leucine + 15 glycine residues. Frequently referred to as the **father of biochemistry**, Fischer received the second-most Nobel Prize in Chemistry (1902) for his work on the synthesis of purines and sugars. Fischer's collection of 9,600 reference compounds is housed in the Department of Biochemistry, the University of California at Berkeley. The private collection was a gift of H.O.L. Fischer, a biochemist and the only one of Fischer's three sons to survive World War I. Following the deaths of his two sons in World War I, Emil Fischer committed suicide.



ACTIVE SITE

As the substrate molecules are comparatively much smaller than the enzyme molecules, there should be some specific regions or sites on the enzyme for binding with the substrate. Such sites of attachment are variously called as '**active sites**' or '**catalytic sites**' or '**substrate sites**'.

Although the enzymes differ widely in their properties, the active site present in their molecule possesses some common features (Fig. 18-6). These are listed below:

380 FUNDAMENTALS OF BIOCHEMISTRY

- The active site occupies a relatively small portion of the enzyme molecule.
- The active site is neither a point nor a line or even a plane but is a 3-dimensional entity. It is made up of groups that come from different parts of the linear amino acid sequence. For example, as already stated, lysozyme (refer Fig. 17-16) has 6 substrates in the active site. The amino acid residues located at the active site are 35, 52, 59, 62, 63 and 107.
- Usually the arrangement of atoms in the active site is well defined, resulting in a marked specificity of the enzymes. Although cases are known where the active site changes its configuration in order to bind a substance which is only slightly different in structure from its own substrate.
- The active site binds the substrate molecule by relatively weak forces.
- The active sites in the enzyme molecules are grooves or crevices from which water is largely excluded. It contains amino acids such as aspartic acid, glutamic acid, lysine serine etc. The side chain groups like $-\text{COOH}$, $-\text{NH}_2$, $-\text{CH}_2\text{OH}$ etc., serve as catalytic groups in the active site. Besides, the crevice creates a micro-environment in which certain polar residues acquire special properties which are essential for catalysis.

Fischer's Lock and Key Model

Previously, the interaction of substrate and enzyme was visualized in terms of a lock and key model (also known as *zoster model*), proposed by Emil Fischer in 1898. According to this model, the union between the substrate and the enzyme takes place at the active site more or less in a manner in which a key fits a lock and results in the formation of an enzyme-substrate complex (Fig. 18-7).

To quote Fischer (1904) himself, "There is a relation between the unknown structure of an active enzyme and that of substrate, they are complementary and the one may be said to fit the other as a key fits a lock."



Fig. 18-7. Formation of an enzyme-substrate complex according to Fischer's lock and key model.

In fact, the enzyme-substrate union depends on a reciprocal fit between the molecular structure of the enzyme and the substrate. And as the two molecules (that of the substrate and the enzyme) are involved, this hypothesis is also known as the **concept of intermolecular fit**. The enzyme-substrate complex (Fig. 18-8) is highly unstable and almost immediately this complex decomposes to produce

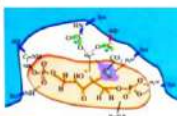


Fig. 18-8. Diagrammatic representation of the active site of the enzyme ribonuclease (Babcock) showing the various sites of interaction between the bound substrates (RuBP and UTP) and various amino acid side chains of the enzyme.

In addition to determining the substrate-binding properties of the active site, these noncovalent interactions alter the properties of the substrate in ways that accelerate its conversion to products.

(After Harris DA, 1995)

- The active site binds the substrate molecule by relatively weak forces.
- The active sites in the enzyme molecules are grooves or crevices from which water is largely excluded. It contains amino acids such as aspartic acid, glutamic acid, lysine serine etc. The side chain groups like $-\text{COOH}$, $-\text{NH}_2$, $-\text{CH}_2\text{OH}$ etc., serve as catalytic groups in the active site. Besides, the crevice creates a micro-environment in which certain polar residues acquire special properties which are essential for catalysis.

Fischer's Lock and Key Model

Previously, the interaction of substrate and enzyme was visualized in terms of a lock and key model (also known as *zoster model*), proposed by Emil Fischer in 1898. According to this model, the union between the substrate and the enzyme takes place at the active site more or less in a manner in which a key fits a lock and results in the formation of an enzyme-substrate complex (Fig. 18-7).

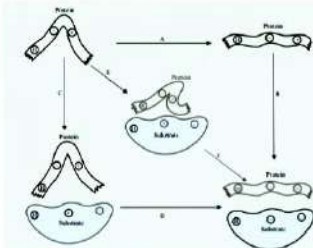


Fig. 18-12. Alternative pathways for a substrate-induced conformational change.

(Adapted from Eschenberg and Nier, 1968)

Kinetics are measured zero and first order in substrate concentration. Reversible enzyme assays are designed to follow zero-order kinetics to avoid the influence of substrate concentration on reaction velocity.

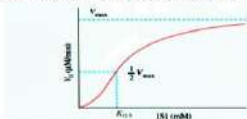


Fig. 18-13. Effect of substrate concentration on the velocity of an enzyme-catalyzed reaction.

Under such conditions, measured rates are directly proportional to the concentration of the enzyme itself. In contrast, *unsubstituted reactions* do not show this saturation effect.

MODIFIERS OF ENZYME ACTIVITY

The catalytic activity of certain enzymes is reversibly altered by certain inorganic and organic molecules called **modifiers**. Those molecules which increase the enzyme activity are called **positive modifiers** or **activators** and those which decrease the enzyme activity are **negative modifiers** or **inhibitors**.

MECHANISM OF ENZYME ACTION 365

Many metals act both as positive and negative modifiers, whereas certain organic molecules retard enzyme activity, thus acting as negative modifiers.

Inorganic Modifiers

(Enzyme Activators)

Certain enzymes, apart from a requirement of a coenzyme, also need a metal ion for full activity. Removal of the metal often results in partial or total loss of enzyme activity. The activity may, however, be restored by replacing the original or a similar metal ion. Some such metal ions (or cations) are K^+ , Ca^{2+} , Fe^{2+} , Mg^{2+} , Mn^{2+} , Cu^{2+} , Zn^{2+} , Pb^{2+} , etc. Mg participates in phosphoryl-transfer reactions and Fe , Cu and Ni are required in oxidoreduction reactions. Certain mechanisms as to how the metal ions bring about activation are given below.

Metals are the inorganic modifiers controlling the enzymatically reactions, metallo type of such reactions.

1. **Direct participation in catalysis.** Certain metals may directly participate in the oxidation-reduction reactions by undergoing a valence change and thus function as electron transport system. Fe , for example, functions similarly in cytochromes or in catalase.

2. **Formation of a metallo-substrate.** Sometimes a metal combines with the substrate to form a metallo-substrate (MS) which, in fact, is the true substrate for the enzyme and forms an enzyme-metal-substrate (EMS) complex. This complex, later on, decomposes to produce the reaction product (P) and regenerate the enzyme and the metal.



3. **Formation of a metallo-enzyme.** A metal ion may first combine with an enzyme to form a metallo-enzyme (ME) which then combines with the substrate forming an enzyme-metal-substrate (EMS) complex.



4. **Alteration of equilibrium constant.** Metals may also change the nature of the reactants so that the apparent equilibrium constant of the reaction is also altered.

5. **Conformational change in the enzyme.** Metal ions may also bring about conformational change in the enzyme molecule, converting it into an active form. In such a case, the metal may be linked at a point far remote from the substrate and may serve to maintain an active tertiary or quaternary structure.

Regulation of Enzyme Activity

The activity of the enzymes is controlled by certain mechanisms which are mentioned below :

1. **Zymogen activation.** Some enzymes (e.g., digestive and coagulating ones) are synthesized in inactive forms. These inactive forms of enzymes are called **zymogens** or **prozymogens**. These zymogens are precursors of enzymes. They are not active on their own but can be active after conversion. Generally, prefix *pro-* or *in-* suffix *-ogen* is added to enzyme's name to denote its zymogen. Examples are prothrombin, pepsinogen, trypsinogen, pepsinogen, fibrinogen, etc. It is difficult to argue why cells should make precursors for some enzymes and not others, rather than merely forming the active enzymes in all cases. Perhaps, the precursors are evolutionary mechanisms or else they are required at times or places where the active enzymes would injure cells.

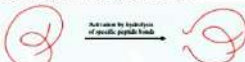


Fig. 18-15. Zymogen activation by hydrolysis of specific peptide bonds.

409-410 / 1254



The precursors should be converted to the active forms before they can catalyze a chemical reaction. This type of activation usually involves cleavage of the peptide bond (Fig. 18-15). Often, a part of the zymogen is also removed. Examples of some common precursor systems are given below :

(A) Chymotrypsin is synthesized by the exocrine cells of the pancreas in its precursor form, chymotrypsinogen. Hydrolysis by trypsin converts chymotrypsinogen into the active form, chymotrypsin.

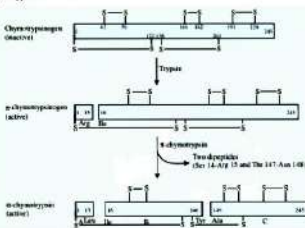
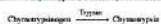


Fig. 18-16. Proteolytic activation of chymotrypsinogen.

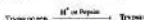
The 3 chains of α -chymotrypsin are linked by two interchain disulfide bonds (A to B and B to C).



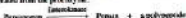
Chymotrypsinogen is composed of a single polypeptide chain with 245 amino acid residues (representing all the 20 standard amino acids) and 5 interchain disulfide bridges (Fig. 18-16). It is virtually devoid of enzymatic activity and is converted into a fully active enzyme when the peptide bond linking Arg^{15} and Ile^{16} is cleaved by trypsin. The resulting enzyme, called α -chymotrypsin, then acts on other α -chymotrypsin molecules. Two dipeptides between residues 14 and 15 and residues 147 and 148 are removed to yield α -chymotrypsin, the stable form of the active enzyme. The 3 resulting chains (designated A, B and C) in α -chymotrypsin remain covalently linked to each other by two interchain disulfide bonds, although 3 more interchain disulfide bonds, one in chain B and two in chain C, are also present in the molecule. The additional cleavages made in converting α -chymotrypsin into α -chymotrypsin are superfluous, because α -chymotrypsin is already fully active. The striking feature of this activation process is that cleavage of a single specific peptide bond transforms the protein from a catalytically inactive form into one that is fully active.

(B) An intestinal enzyme termed enterokinase converts trypsinogen, a proenzyme also secreted

by pancreas, to trypsin. Trypsin acts upon peptide linkages involving the carboxyl group of arginine and lysine.



(C) Another instance is the proenzyme of the gastric juice called as pepsin. This is derived from its proenzyme *pepsinogen*, secreted by gastric mucosa. Pepsinogen is converted to the active form, pepsin, both by the acidity of the gastric juice and by pepsin itself. This process is, therefore, called as **autocatalysis**. During this process, a polypeptide is liberated from the proenzyme.



Autocatalytic activity of an enzyme precursor may be graphically represented. If a graph of time versus enzyme activity is plotted, a sigmoid S-shaped curve is obtained (Fig. 18-15).



Fig. 18-15. Autocatalytic activation of an enzyme precursor

2. **Covalent modification.** The activity of serine-containing enzymes is regulated by the covalent insertion of a small group on an enzyme molecule. For example, the activity of the enzymes that synthesise and cleave glycogen, is regulated by the attachment of a phosphoryl group to a specific serine residue on these enzymes. This modification can be reversed by hydrolysis.



3. **Feedback inhibition.** Some enzymes catalyse the synthesis of small molecules (such as amino acids) in a number of steps. The enzyme catalysing the first step in this biosynthesis is inhibited by the end product of the reaction. Such type of regulatory mechanism, which is called feedback inhibition (Fig. 18-16), is beautifully illustrated by the biosynthesis of isoleucine from threonine.



Fig. 18-16. Feedback inhibition of the first enzyme in a multistep reaction by reversible binding of the end product

The reaction completes in 5 steps. The first step reaction is catalyzed by the enzyme *threonine deaminase*. The activity of this enzyme is inhibited upon accumulation of high quantities of isoleucine. Isoleucine binds to a different site from threonine. This is called **allosteric interaction**. However, when the level of isoleucine drops sufficiently, the enzyme reactivates and isoleucine is released.

Organic Modifiers

(Enzyme Inhibitors)

Compounds which convert the enzymes into inactive substances and thus adversely affect the

388 FUNDAMENTALS OF BIOCHEMISTRY

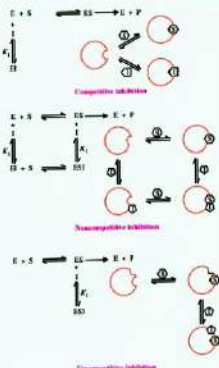


Fig. 18-17. Three types of reversible enzyme inhibition

Competitive inhibitors bind to the enzyme's active site. Noncompetitive inhibitors generally bind to a separate site termed **allosteric site**. Uncompetitive inhibitors also bind to an allosteric site but they bind only to the ES complex.

rate of enzymatically-catalyzed reaction are called as enzyme inhibitors. Such a process is known as **enzyme inhibition**. Two broad classes of enzyme inhibitors are generally recognised – reversible and non-reversible, depending on whether the enzyme-inhibitor (EI) complex dissociates rapidly or very slowly.

Reversible Enzyme Inhibition

A reversible inhibitor dissociates very rapidly from its target enzyme because it becomes very loosely bound with the enzyme. Three general types of reversible inhibition (Fig. 18-17) are

MECHANISM OF ENZYME ACTION 389

distinguished – competitive, noncompetitive and uncompetitive, depending on three factors:

- whether the inhibition is or is not overcome by increasing the concentration of the substrate,
- whether the inhibitor binds at the active site or at allosteric site, and
- whether the inhibitor binds either with the free enzyme only, or with the enzyme-substrate complex only or with either of the two.

A. Competitive or Substrate analogue inhibition

This type of competition occurs at the active site (Fig. 18-18). Here the structure of the inhibitor (I) closely resembles with that of the substrate (S). It may, thus, combine with the enzyme (E), forming an enzyme-inhibitor (EI) complex rather than ES complex. The inhibitor, thus, competes with the substrate to combine with the enzyme. The degree of inhibition depends upon the relative concentrations of the substrate and the inhibitor. Thus, by increasing the substrate concentration and keeping the inhibitor concentration constant, the amount of inhibition decreases and conversely a decrease in substrate concentration results in an increased inhibition. It may, however, be noted that in competitive inhibition, the enzyme can bind substrate (forming an ES complex) or inhibitor (EI), but not both (ESI). Thus, we see that a competitive inhibitor diminishes the rate of the reaction by reducing the proportion of the enzyme molecules bound to a substrate.



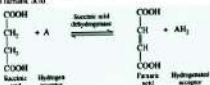
Fig. 18-18. Competitive or substrate analogue inhibition

A competitive inhibitor fits into the active site of the enzyme, preventing the real substrate from passing as free. The inhibitor cannot be converted to the products of the reaction and so the overall rate of reaction is slowed down. If an inhibitor is present in equal concentration to the substrate, and if both types of molecules bind to the active site equally well, the enzyme can only work at half its normal rate.

Competitive inhibition can be analysed quantitatively by steady-state kinetics. Because the inhibitor binds reversibly to the enzyme, the competition can be biased to favour the substrate only simply by adding more substrate. When more substrate is present, the probability that an inhibitor molecule will bind is diminished, and the reaction exhibits a normal V_{max} . However, the $[S]$ at which $V_p = 1/2 V_{max}$, the K_m will increase in the presence of inhibitor.

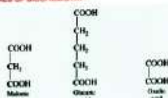
Some well-known examples of competitive inhibition are given below:

1. An enzyme, **succinate acid dehydrogenase** (= succinyl dehydrogenase) catalyses the conversion of succinate acid to fumaric acid.

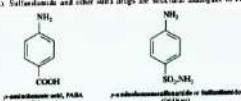


Many organic compounds, which are structurally related to succinic acid, combine with the enzyme, thus inhibiting the reaction. A few inhibitors of this reaction are:

Malonic acid is most efficient of all these inhibitors. When the inhibitor : enzyme ratio is 1 : 50, the enzyme is inhibited 50%. Malonic acid differs from succinic acid in having one rather than two methylene groups.



2. Many microorganisms, like bacteria, synthesise the vitamin folic acid from para-aminobenzoic acid (PABA). Sulfasalazide and other sulfa drugs are structural analogues of PABA. Hence,

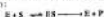


sulfa drugs act as enzyme inhibitor and occupy the active site of some bacterial enzymes catalysing this reaction. Failure of PABA to combine with the bacterial enzymes at the active site results in blocking off the folic acid synthesis. The resulting deficiency of this vitamin is fatal to those microorganisms. Since man lacks the enzymes necessary for folic acid synthesis from PABA, folic acid is needed as a vitamin in the diet. Thus, the sulfa drugs inhibit growth of the bacteria in man by competing with PABA for the active centres of the bacterial enzymes.

3. Competitive inhibition is used therapeutically to treat patients who have ingested methanol (CH_3OH), a solvent found in gas-line antifreeze. Methanol is converted to formaldehyde by the action of the enzyme alcohol dehydrogenase (E.C. No. 1.1.1.1). Formaldehyde damages many tissues esp., the optic ones, causing blindness. Ethanol competes effectively with methanol as a substrate for alcohol dehydrogenase. Ethanol, thus, acts as an inhibitor for the substrate methanol and competes with it to occupy the active site of the enzyme. Methanol poisoning may, then, be cured by an intravenous infusion of ethanol to the patients so that formaldehyde formation is considerably lowered. Most of the methanol can be excreted harmlessly in the urine.

4. A physiologically important example of competitive inhibition is found in the formation of 3-bisphosphoglycerate (BPG) from 1,3-bisphosphoglycerate. Bis-phosphoglycerate mutase, the enzyme catalysing this isomerisation reaction, is completely inhibited by even low levels of 2,3-bisphosphoglycerate. In fact, it is not uncommon for an enzyme to be completely inhibited by its own product because of the product's structural resemblance to the substrate. However, increasing the concentration of substrate checks the inhibitory effect.

Michaelis-Menten equation may also be applied to the competitive inhibition of enzymes. Here, besides the normal equation:



one must also consider the equilibrium state between the enzyme and the inhibitor, I as follows:



MECHANISM OF ENZYME ACTION 391

In the presence of the competitive inhibitor, henceforth, the concentration of the free enzyme would be expressed as:

$$[E] = [E_0] - ([ES] + [EI])$$

The dissociation of the enzyme-inhibitor compound, K_i , would then be defined as:

$$K_i = \frac{[E][I] - [EI]}{[EI]}$$

Deriving the Michaelis-Menten equation for this case of inhibition, one obtains as follows:

$$V = \frac{V_m \times (S) \times K_i}{K_m K_i + K_m (S) + K_i (S)} \quad (15)$$

Reversal of this equation gives the modified form of Lineweaver-Burk equation:

$$\frac{1}{V} = \frac{K_m K_i + K_m (S) + K_i (S)}{V_m \times (S) \times K_i}$$

$$\text{or} \quad \frac{1}{V} = \frac{K_m}{V_m} \left(1 + \frac{[I]}{K_i} \right) \times \frac{1}{[S]} + \frac{1}{V_m} \quad (16)$$

When $\frac{1}{V}$ is plotted against $\frac{1}{[S]}$ (Fig. 18-15), the intercept $\frac{1}{V_m}$ remains the same as in the case of non-inhibited reaction but the slope, which is now $\frac{K_m}{V_m} \left(1 + \frac{[I]}{K_i} \right)$, is increased by the

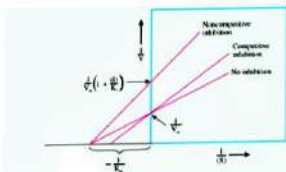


Fig. 18-15. Double reciprocal graph (Lineweaver-Burk plot)

A plot of $\frac{1}{V}$ versus $\frac{1}{[S]}$ for an enzyme reaction with (i) no inhibitor (ii) a competitive inhibitor and (iii) a noncompetitive inhibitor

(Modified from Fawcett and Kofoid, 1966)

factor, $1 + \frac{[I]}{K_i}$. Thus, if the substrate concentration is large enough, the effect of the competitive inhibitor can be overcome and V_m may be reached.

The phenomenon of competitive inhibition is of practical value. It may be employed to prevent the growth of one organism in the presence of other, for example: (a) bacterial growth inhibited in the

presence of an animal (A) growth of the insect inhibited on a fruit tree.

B. Noncompetitive inhibition

Here no competition occurs between the substrate, S and the inhibitor, I (Fig. 18-20). The inhibitor has little or no structural resemblance with the substrate and it binds with the enzyme at a place other than the active site (i.e., at the allosteric sites). Since I and S may combine at different sites, formation of both EI and ESI complexes takes place. Both ES and ESI may break down to produce the reaction product (P). It may, however, be noted that in noncompetitive inhibition, the inhibitor and substrate can bind simultaneously to an enzyme molecule since their binding sites are different and hence do not overlap. The enzyme is inactivated when inhibitor is bound, whether or not substrate is also present. Thus, it is apparent that a noncompetitive inhibitor acts by lowering the turnover number rather than by decreasing the proportion of enzyme molecules that are bound to the substrate. Noncompetitive inhibition, in contrast to competitive inhibition, cannot be overcome by increasing substrate concentration. The inhibitor effectively lowers the concentration of active enzyme and hence lowers V_m . There is almost no effect on K_m , however.

Certain noteworthy examples of noncompetitive inhibition are as follows:

1. Various heavy metals ions (Ag^+ , Hg^{2+} , Pb^{2+}) inhibit the activity of a variety of enzymes. Urease, for example, is highly sensitive to any of these ions in traces. Heavy metals form mercaptides with sulfhydryl ($-\text{SH}$) groups of enzymes:



The established equilibrium inactivates enzymes that require a $-\text{SH}$ group for activity. Because of the reversibility of mercaptide formation, the inhibition can be relieved by removal of the heavy metal ion. In the treatment of lead poisoning, advantage is taken of the metal's affinity for $-\text{SH}$ groups. Therefore, the sulfhydryl compounds are administered to interact with the metal in the circulatory system and form mercaptides, which are then excreted.

2. Similarly, cyanide and hydrogen sulfide strongly inhibit the action of iron-containing enzymes like carboxylase and peroxidase.

The Lineweaver-Burk equation for this type of inhibition would be:

$$\frac{1}{V} = \left[1 + \frac{[I]}{K_i} \right] \left[\frac{1}{V_m} + \left(\frac{K_m}{V_m} \right) \times \frac{1}{[S]} \right] \quad (17)$$

Thus, here both the slope and the intercept are altered, rather increased by the factor $1 + \frac{[I]}{K_i}$ as contrast to the competitive inhibition where only the slope is changed. Moreover, the maximal velocity attained is less than that found in noninhibited case.

C. Uncompetitive inhibition. An uncompetitive inhibitor also binds at an allosteric site (like the noncompetitive inhibitor) but the binding takes place only with the enzyme-substrate (ES) complex, and not the free enzyme molecule.

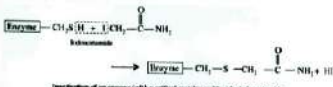
Irreversible Enzyme Inhibition

Although irreversible inhibition was once categorised and listed as noncompetitive inhibition, it is now recognised as a distinct type of inhibition. Irreversible inhibitors are those that combine with or destroy a functional group on the enzyme that is essential for its activity. In fact, an irreversible inhibitor dissociates very slowly from its target enzyme because it becomes very tightly bound to its

active site, thus inactivating the enzyme molecule. The bonding between the inhibitor and enzyme may be covalent or non-covalent.

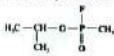
Two common examples of irreversible inhibition are discussed below.

1. Alkylating reagents, such as **iodoacetamide**, irreversibly inhibit the catalytic activity of some enzymes by modifying cysteine and other side chains. Iodoacetamide is a widely used agent for the detection of sulfhydryl group.

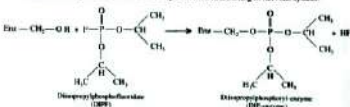


2. Organophosphorus compounds, such as diisopropyl phosphorofluoridate, DFP, are potent irreversible inhibitors of enzymes that have active seryl residues at their catalytic sites. DFP is closely

DFP is also called **diisopropyl fluorophosphate**, DFP (A. Lehninger, DL Nelson and MM Cox, 1993)



related chemically to nerve gas, whose lethality is due to the inactivation of *acetylcholinesterase*, an enzyme critical for the transmission of nerve impulses. *Acetylcholinesterase* (Fig. 14-21) cleaves the neurotransmitter *acetylcholine*, an essential step in normal functioning of nervous system.



Inactivation of *acetylcholinesterase* by diisopropyl phosphorofluoridate, DFP

A special class of irreversible inhibitors are the **suicide inhibitors**. These compounds are relatively unreactive until they bind to the active site of a specific enzyme. A suicide inhibitor is designed to carry out the first few chemical steps of the normal enzyme reaction. Instead of being transformed into the normal product, however, the inhibitor is converted to a very reactive compound that combines irreversibly with the enzyme. These are also called **mechanism-based inactivators**, because they utilize the normal enzyme reaction mechanism to inactivate the enzyme.

BISUBSTRATE REACTIONS

We have heretofore been concerned with reactions involving enzymes that require only one substrate. Yet, enzymatic reactions involving two substrates and yielding two products account for ~60% of known biochemical reactions.

394 FUNDAMENTALS OF BIOCHEMISTRY

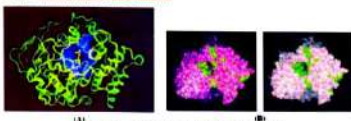


Fig. 14-21. The 3-D structure of acetylcholinesterase (AChE).
(A) The ribbon model. The aromatic side chains lining its active site gorge (purple) are shown in stick form surrounded by their Van der Waals dot surface. The ACh substrate, which was modeled into the active site after the enzyme was crystallized in its phosphate form, is shown in ball-and-stick form with its atoms gold and its bonds light blue. The entrance to the gorge is at the top of the figure.
(B) Space-filling stereo view looking down into the active site. Aromatic residues are in green. Ser²⁰³ is red. His⁴⁴⁰ is cyan, and other residues are grey.

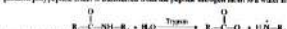
[Continued: (A) Joel Sussman, *The Weizmann Institute of Science, Israel* and (B) Joel Sussman et al., 1992]



Almost all of these so-called **bisubstrate reactions** are **intertransfer reactions** in which the enzyme catalyzes the transfer of a specific functional group, X, from one of the substrates to the other:



or **oxidation-reduction reactions** in which reducing equivalents are transferred between the two substrates. For example, in the peptide hydrolysis reaction catalyzed by trypsin, the peptide carbonyl group with its pendant polypeptide chain is transferred from the peptide nitrogen atom to a water molecule:



Similarly, in the alcohol dehydrogenase reaction, a hydride ion is formally transferred from ethanol to NAD⁺:



A. Terminology

In such reactions, the sequence of binding of the substrate to the enzyme molecule may be of different types. Depending on the sequence of binding of the substrate to the enzyme, W.W. Cleland (1967) proposed 4 types of mechanisms and explained them with the help of what are commonly called '**Cleland diagrams**' (Fig. 14-22). Cleland's nomenclature concept is based on the following 3 conventions:

1. The substrates are designated by the letters A, B, C and D in the order that they add to the enzyme.
2. The products are denoted by the letters P, Q, R and S in the order that they leave the enzyme.
3. The enzyme is indicated by the letter E. The intermediary stable forms of enzymes, that are produced and then disappear during the course of reaction, are represented by the letters F, G, H etc.
4. The enzyme-substrate complexes are shown in parentheses.

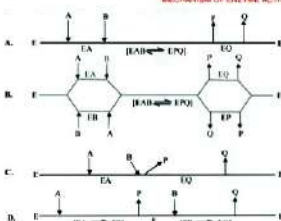


Fig. 14-22. Four substrate binding patterns in bisubstrate enzyme reactions (Cleland diagrams)

- A. The ordered sequential mechanism
- B. The random sequential mechanism
- C. Theorist-Chance mechanism
- D. The ping-pong mechanism

For describing the enzymatic reactions using Cleland's shorthand notation, certain conventions are followed. The enzyme is represented by a horizontal line and successive additions of substrates and release of products are denoted by vertical arrows. Enzyme forms are placed under the line and rate constants, if given, are to the left of the arrow or on top of the line for forward reactions.

5. The number of reactants and products in a given reaction are specified, in order by the terms *Dis* (one), *Bi* (two), *Tri* (three), and *Quad* (four). For example, a reaction requiring one substance and yielding three products is designated a *1 to 3* reaction.

B. Types of Bi Bi reactions

In this section, we shall deal with reactions that require two substrates and yield two products, that is **Bi Bi reactions**. It may, however, be remembered that there are numerous examples of even more complex reactions. Enzyme-catalyzed group-transfer Bi Bi reactions fall under two major mechanistic classifications.

1. Sequential reactions

Reactions in which all substrates combine with the enzyme before a reaction can occur and products be released are known as **sequential reactions**. In such reactions, the group being transferred, X, is directly passed from A (= P-X) to B, yielding P and Q (= B-X). Hence, such reactions are also called **single-displacement reactions**.

Sequential reactions can be further classified into 2 types:

- (a) those with a compulsory order of substrate addition to the enzyme, which are said to have an **ordered mechanism**, and
- (b) those with no preference for the order of substrate addition, which are described as having a **random mechanism**.

In the ordered mechanism, the binding of the first substrate is apparently required for the enzyme

to form the binding site for the second substrate, whereas for the random mechanism, both binding sites are present on the free enzyme.

2. Ping pong reactions

Reactions in which one or more products are released before all substrates have been added are known as *ping pong reactions*. In such a reaction, a functional group X of the first substrate A ($A \rightarrow P + X$) is displaced from the substrate E to yield the first product P and a stable enzyme form $E' (= E - X)$ in which X is tightly (often covalently) bound to the enzyme (ping). In the second stage of the reaction, X is displaced from the enzyme by the second substrate B to yield the second product Q ($B + X \rightarrow Q + E$), thereby regenerating the original form of the enzyme, E (pong). Such reactions are, hence, also known as *double-displacement reactions*. A notable feature of ping pong reactions is that the substrates A and B do not encounter one another on the surface of the enzyme.

The salient features of the mechanisms underlying these different reaction types are described hereunder:

1. **The ordered sequential mechanism.** This mechanism envisages that the first substrate (A) binds with the enzyme (E), followed by the second substrate (B). The substrate A which combines first is called the *leading substrate* and the substrate B which combines later is termed the *following substrate*. A ternary complex (EAB) between enzyme and the two substrates is formed. In this complex, both substrates are converted to products (P and Q) and then liberated in the same sequence their corresponding substrate had combined, i.e., the product P is released first followed by Q. Many NAD^+ - and $NADP^+$ -requiring dehydrogenases follow an ordered Bi Bi mechanism in which the coenzyme is the leading substrate.

2. **The random sequential mechanism.** In this mechanism, there is no definite sequence of substrate association to the enzyme. They can bind in any order and the products from the ternary complex can also be released in any order. Some dehydrogenases and kinases operate through this mechanism.

3. **The ordered-Chance mechanism.** It is a variant of the ordered sequential mechanism which the ternary complex (EAB) is not at all formed. This mechanism operates in the reaction catalysed by alcohol dehydrogenase (E.C. No. 1.1.1.1).

4. **The ping pong mechanism.** This mechanism envisages that one substrate binds and one product is released before the second substrate can bind to the enzyme and may release the second substrate. Thus, here the enzyme acts as a board and the substrates act as ping pong balls. Many enzymes, including glutamate dehydrogenase (E.C. No. 1.4.1.2), chymotrypsin, trypsin, and some flavoenzymes obey ping pong mechanisms.

C. Kinetics of Bi Bi Reactions

Steady state kinetic measurements can be utilized to distinguish among the foregoing bisubstrate mechanisms. For doing so, one must first derive their rate equations. This can be done in much the same way as for single-substrate enzymes, i.e., solving a set of simultaneous linear equations consisting of an equation expressing the steady state condition for each kinetically distinct enzyme complex plus one equation representing the conservation condition for the enzyme. This, indeed, is a more complex situation for bisubstrate enzymes than it is for single-substrate or monosubstrate enzymes.

The rate equations for the above described bisubstrate mechanisms in the absence of products are given below in double reciprocal form.

For ordered Bi Bi reactions:

$$\frac{1}{v_0} = \frac{1}{V_m} + \frac{K_m^A}{V_m[A]} + \frac{K_m^B}{V_m[B]} + \frac{K_m^A K_m^B}{V_m[A][B]} \quad (18)$$

For random Bi Bi reactions:

The rate equation for the general random Bi Bi reaction is quite complicated. However, in the special case that both substrates are in rapid and independent equilibrium with the enzyme; that is, the EAB-BPQ interconversion is rate determining, the actual rate equation reduces to the following relatively simple form. This mechanism is known as the *rapid equilibrium random Bi Bi mechanism*:

$$\frac{1}{v_0} = \frac{1}{V_m} + \frac{K_m^A K_m^B}{V_m K_m^A [A]} + \frac{K_m^B}{V_m [B]} + \frac{K_m^A K_m^B}{V_m [A][B]} \quad (19)$$

For ping pong Bi Bi reactions:

$$\frac{1}{v_0} = \frac{K_m^A}{V_m [A]} + \frac{K_m^B}{V_m [B]} + \frac{1}{V_m} \quad (20)$$

Physical Significance of the Bisubstrate Kinetic Parameters

The parameters in the equations describing bisubstrate reactions have meanings similar to those for single-substrate reactions. V_m is the maximal velocity of the enzyme obtained when both A and B are present at saturating concentrations. K_m^A and K_m^B are the respective concentrations of A and B, necessary to achieve $\frac{1}{2} V_m$ in the presence of a saturating concentration of the other; and K_m^A and K_m^B are the respective dissociation constants of A and B from the enzyme, E.

ALLOSTERIC ENZYMES

A **modulator** is a metabolite which, when bound to the allosteric site of an enzyme, alters its kinetic characteristics. The modulators for allosteric enzymes may be either stimulatory or inhibitory. A stimulator is often the substrate itself. The regulatory enzymes for which substrate and modulator are identical are called

The term **allosteric** (altē-st-er; *altēra* = 'other'; *stēra* = 'space' or 'site') has been introduced by the two Nobel Laureates, Monod and Jacob, to denote an enzyme site, different from the active site, which noncompetitively binds molecules other than the substrate and may influence the enzyme activity.

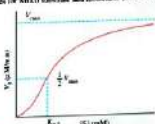


Fig. 18-23. Schematic of an allosteric enzyme's activity by its substrate.

Note the sigmoid curve given by an allosteric enzyme, as which the substrate also serves as a positive (stimulatory) modulator. The analogous curve for the nonregulatory enzymes is hyperbolic, as also predicted by the Michaelis-Menten equation. The allosteric enzymes, however, do not obey the Michaelis-Menten kinetics. Also note that a relatively small increase in $[S]$ in the steep part of the curve can cause a very large increase in V_0 .

homotropic. When the modulator has a structure different than the substrate, the enzyme is called **heterotropic**. Some enzymes have more than one modulators.

The allosteric enzymes are structurally different from the simple nonregulatory enzymes. Besides the presence of an active or catalytic site, the allosteric enzymes also have one or more regulatory or allosteric sites for binding the modulator. Just as an enzyme's active site is specific for

FRANÇOIS JACOB
(born, 1920)



Both Jacob and Monod served heroically during World War II, Jacob as member of the Free French Forces, squared seriously in Normandy during August 1944, and Monod as a leader of the Paris Resistance. Monod had already obtained his PhD, Jacob had to wait until 1947 to gain his M.D. from the University of Paris. Both spent the major parts of their careers at the Pasteur Institute, the famous Paris research centre set up around Louis Pasteur in the late nineteenth century and still one of the most influential European laboratories. Jacob contributed to some of the early work on bacteriophages, with André Lwoff and Elie Wollman, before beginning the collaboration with Monod that led to the **operon theory** and the **concept of messenger RNA** in 1961. They shared the 1965 Nobel Prize with Lwoff. Both have published important books: *Chance and Necessity* by Monod is a powerful vindication of natural selection, *The Logic of Life* by Jacob is an illuminating account of changes in scientific thinking over the past three centuries. Jacob's autobiography, *The Statue Within*, was published in 1987.

JACQUES MONOD
(d. 1910-1976)



its substrate, the allosteric site is specific for its modulator. Enzymes with several modulators generally have different specific binding sites for each. By contrast, in homotropic enzymes, the active site and allosteric site are the same. Most of them have two or more polypeptide chains or subunits. For example, *aspartate transcarbamoylase* (Fig. 18-23) has 12 subunits.

Other differences between nonregulatory enzymes and allosteric enzymes involve kinetic properties. The allosteric enzymes often display sigmoidal plots (Fig. 18-24) of the reaction velocity, V versus substrate concentration, $[S]$, rather than the hyperbolic plots predicted by Michaelis-Menten equation. Recall that the oxygen-binding curve of myoglobin is hyperbolic, whereas that of hemoglobin is sigmoidal. The binding of enzymes to substrates is analogous. In allosteric enzymes, the binding of substrate to one active site can affect the properties of other active sites in the same enzyme molecule. A possible outcome of this interaction between subunits is that the binding of substrate becomes cooperative. In addition, the activity of an allosteric enzyme may be altered by regulatory molecules that are bound to the allosteric sites, just as oxygen binding to hemoglobin is affected by H^+ , CO_2 and BPG.

With heterotropic enzymes, however, it is difficult to generalize about the shape of the substrate-saturation curve (Fig. 18-25). An activator may cause the substrate-saturation curve to become more nearly hyperbolic with a decrease in K_m but no change in V_m , thus resulting in an increased reaction velocity at a fixed substrate concentration. On the contrary, a negative modulator (i.e. an inhibitor) may produce a more sigmoid substrate-saturation curve, with an increase in K_m . In other words, allosteric activation shift the substrate-saturation curve to the left (to higher saturation), whereas

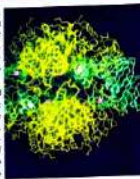


Fig. 18-24. The structure of the enzyme aspartate transcarbamoylase (ATCase) in the T from Catalytic subunits are shown in yellowed and regulatory subunits in green.

allosteric inhibitors shift it to the right (to lower saturation).

[Although one can find a value of $[S]$ on the sigmoid-saturation curve at which V_0 is half-maximal, one cannot put the designation K_m to it because the enzyme does not follow the hyperbolic Michaelis-Menten relationship. Instead, the symbol $[S]_{0.5}$ or $K_{0.5}$ is often used to represent the substrate concentration giving half-maximal velocity of the reaction catalyzed by an allosteric enzyme.]

Mechanism of Kinetic Behaviour of Allosteric Enzymes

Two general models for the interconversion of inactive and active forms of allosteric enzymes have been proposed.

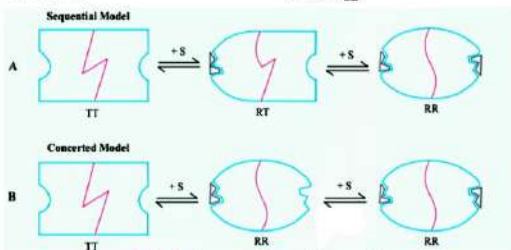


Fig. 18-26. Comparison of the two allosteric models

A. Simple sequential model

B. Concerted or symmetry model

(Adapted from Lubert Stryer, 1995)

A. Simple sequential model

Proposed by Daniel E. Koshland Jr. in 1966, the model envisages that the allosteric enzyme can exist in only two conformational change individually. Binding of substrate increases the probability of the conformational change. A conformational change in one subunit makes a similar change in an adjacent subunit, as well as the binding of a second substrate molecule, more likely. There are more potential intermediate states in this model than in the other (symmetry) model.

Consider an allosteric enzyme consisting of two identical subunits, each containing an active site (Fig. 18-26A). The T (tense) form has low affinity and the R (relaxed) form has high affinity for substrate. In this model, the binding of substrate to one of the subunits induces a $T \rightarrow R$ transition in that subunit but not in the other. The affinity of the other subunit for substrate is increased because the subunit interface has been altered by the binding of the first substrate molecule.

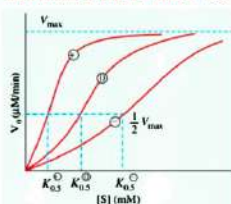


Fig. 18-25. Effects of an activator A , and inhibitor I and no modulator N on an allosteric enzyme in which $K_{0.5}$ is modulated without a change in V_{max}

400 FUNDAMENTALS OF BIOCHEMISTRY



Fig. 18-27. Effect of allosteric activator and allosteric inhibitor on substrate binding

Note that in the concerted model, an allosteric inhibitor stabilizes the T state, whereas an allosteric activator stabilizes the R state.

B. Concerted or Symmetry model

Proposed by Jacques Monod and colleagues in 1965, the model envisages that an allosteric enzyme can exist in still two conformations, active and inactive. All subunits are in the active form or all are inactive. Every substrate molecule that binds increases the probability of transition from the inactive to the active state.

In the concerted model, the binding of substrate to one of the subunits increases the probability that both switch from the T to the R form (Fig. 18-26B). Thus, symmetry is conserved in the concerted model but not in the sequential model.

The effects of allosteric activators and inhibitors can be explained quite easily by the concerted model. An allosteric inhibitor binds preferentially to the T form, whereas an allosteric activator binds preferentially to the R form (Fig. 18-27). Consequently, an allosteric inhibitor shifts the $R \rightarrow T$ conformational equilibrium toward T, whereas an allosteric activator shifts it toward R. The result is that an allosteric activator increases the binding of substrate to the enzyme, whereas an allosteric inhibitor decreases substrate binding.

REFERENCES

1. Abelson JN, Simon MI (editors) : Methods in Enzymology. Academic Press Inc., New York. 1992.
2. Annual Review : Advances in Enzymology. Interscience Publishers, Inc., New York. 1941-Current.
3. Atkinson DE : Regulation of Enzyme Action. *Ann Rev. Biochem.* **35** : 85-124, 1966.
4. Bell JE, Bell ET : Proteins and Enzymes. Prentice-Hall Inc., New Jersey. 1988.
5. Bernhard S : The Structure and Functions of Enzymes. W.A. Benjamin Inc., New York. 1968.
6. Boyer PD, Lardy H, Myrback K (editors) : The Enzymes. 3rd ed. 16 vols. Academic Press, Inc., New York. 1970-1983.
7. Brookhaven Symposia in Biology : Enzyme Models and Enzyme Structure. No. 15, Brookhaven National Laboratory, Upton, New York. 1962.
8. Cech TR : RNA as an enzyme. *Sci. Amer.* **255** (5) : 64-75, 1986.
9. Chipman DM, Sharon N : Mechanism of Lysozyme Action. *Science* **165** : 454-465, 1969.
10. Cleland WW : Enzyme Kinetics. *Ann. Rev. Biochem.* **36** : 77, 1967.
11. Colowick SP, Kaplan NO (editors) : Methods in Enzymology. 52 vols. Academic Press, Inc., New York. 1955-1978.