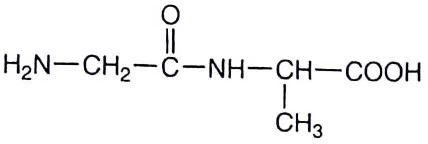
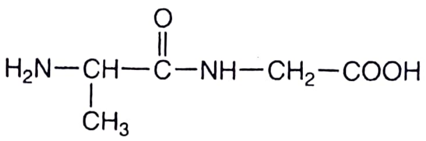
VARIATIONS IN PEPTIDE AND PROTEIN STRUCTURES

There are two ways in which two amino acids can combine to form a dipeptide. For example glycine and alanine may combine to give the dipeptide *Gly-Ala* or *Ala-Gly*. In the first, glycine is the *N*-terminal and alanine is the *C*-terminal. In the second, alanine is the *N*-terminal and glycine is the *C*-terminal.



$H_2N-CH_2-\overset{\overset{O}{\|}}{C}-NH-\underset{CH_3}{\underset{|}{CH}}-COOH$

Gly-Ala
(Glycylalanine)



$H_2N-\underset{CH_3}{\underset{|}{CH}}-\overset{\overset{O}{\|}}{C}-NH-CH_2-COOH$

Ala-Gly
(Alanylglycine)

Three amino acids can combine in six different ways to form six tripeptides. For example, glycine alanine, and phenylalanine may combine to give the following tripeptides.

Gly-Ala-Phe	*Ala-Gly-Phe*	*Ala-Phe-Gly*
Phe-Ala-Gly	*Phe-Gly-Ala*	*Gly-Phe-Ala*

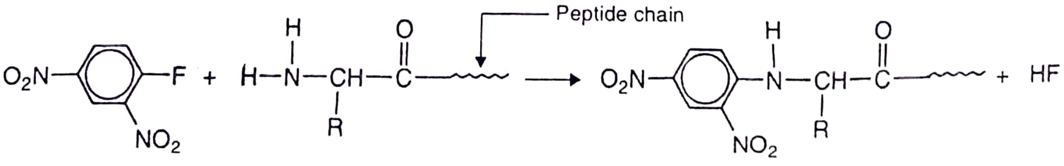
As the size of the peptide increases, the possibilities of variation in structure increase at an almost unbelievable rate. The different ways of combining the amino acids give rise to the fantastic number of variations possible in protein structure. For example, a decapeptide contains 10 amino acid units and has a molecular weight of something over 1000. It is far smaller than a protein. Even so, the number of different decapeptides that can be constructed (using each amino acid no more than once in each structure) is over 4,000,000,000,000. Most proteins have molecular weights of several thousands and some many millions. Proteins can use the same amino acid more than once in a molecule and may also form cross-linked structures. The number of variations is beyond comprehension.

DETERMINATION OF STRUCTURE OF PEPTIDES

To determine the structure of a peptide (or a protein), we normally begin by carrying out its complete hydrolysis. The peptide is refluxed with dilute hydrochloric acid so that all the peptide linkages are broken. Analysis of the resulting solution tells us the kind and the amounts of the amino acids present in the peptide. However, complete hydrolysis tells us nothing about the sequence of amino acids in the peptide.

END-GROUP ANALYSIS

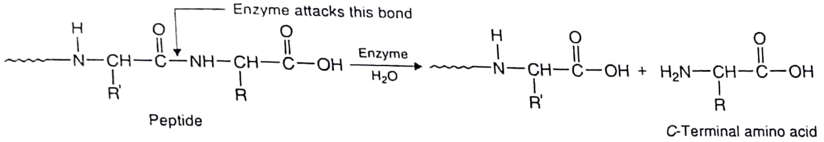
Identification of the *N*-terminal amino acid of the peptides can be accomplished by the **Sanger's method**. This involves the reaction of 2,4-dinitrofluorobenzene (DNFB) with the free amino group of the *N*-terminal residue before the peptide is hydrolyzed.



DNFB Peptide DNP-Peptide

The dinitrophenyl (DNP) derivative of the *N*-terminal amino acid is colored and can easily be isolated and identified.

Identification of the *C*-terminal amino acid can be accomplished by the enzyme *carboxypeptidase*. This enzyme selectively cleaves the *C*-terminal amino acid, which can be isolated and identified.



The sequence of internal amino acids of the peptide can be determined from data obtained from partial hydrolysis of the peptide. Partial hydrolysis of the peptide cleaves it into *di-*, *tri-*, *tetra-*, and higher peptides. The tripeptide fragments are separated from the other and the *N*-terminal and *C*-terminal amino acids and the central amino acid in each tripeptide are identified. Once the structure of all the tripeptides are known, the information can be fitted together like a jigsaw puzzle, and the structure of the original peptide can be deduced.

As an example, let us take the case of an unknown tetrapeptide. Complete hydrolysis, separation, and analysis tells us that it contains alanine, isoleucine, and valine. We now subject the unknown peptide to end-group analysis. Treatment with DNFB followed by hydrolysis, gives us the DNP-derivative of alanine. Therefore, the *N*-terminal amino acid must be alanine. Treatment with *carboxypeptidase* gives us valine. Partial hydrolysis of the unknown tetrapeptide gives us two different tripeptides. These tripeptides are in turn subjected to complete hydrolysis and end-group analysis. By overlapping the pieces, we can determine the sequence of amino acids in the original tetrapeptide.

(1) Tetrapeptide $\xrightarrow[H_2O]{H^+}$ *Ala—Gly—Ile* + *Gly—Ile—Val*

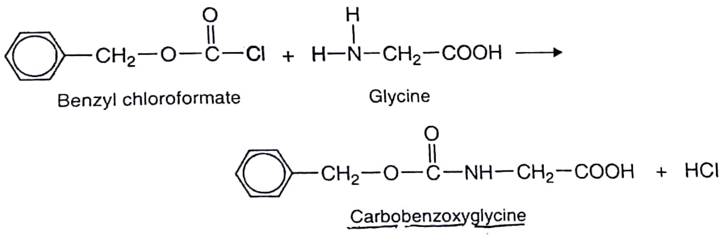
(2) *Ala—Gly—Ile*
Gly—Ile—Val

(3) Structure : *Ala—Gly—Ile—Val*

SYNTHESIS OF PEPTIDES

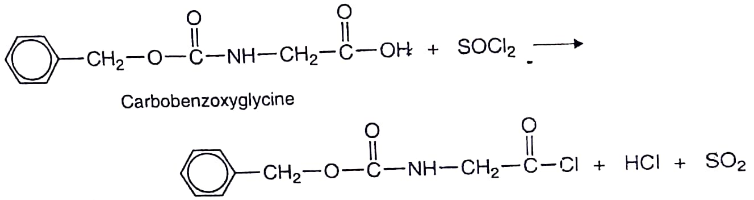
Specific peptides can be obtained by the following steps. Glycylalanine (*Gly—Ala*) is taken as an example.

Step 1. The amino group of glycine is protected by treatment with benzyl chloroformate.

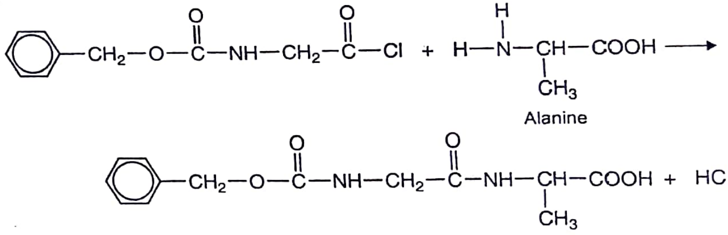


Step 2. The protected glycine is converted to the corresponding acid chloride by treatment with thionyl chloride.

15

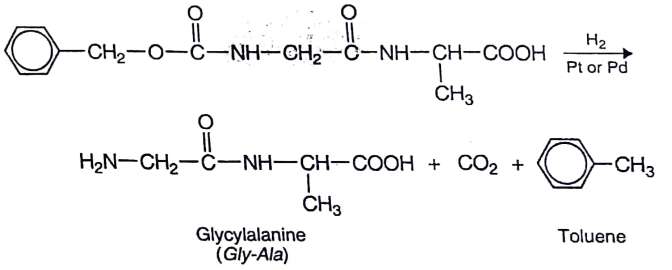


Step 3. The acid chloride is condensed with alanine.



If the amino group of glycine is not protected as in Step (1) and the reaction is carried straight with glycine acid chloride, two reactions will occur. Glycine acid chloride will react with the amino group of another molecule of glycine acid chloride as well as with alanine as in this step.

Step 4. The protecting group of glycine is removed by reduction.



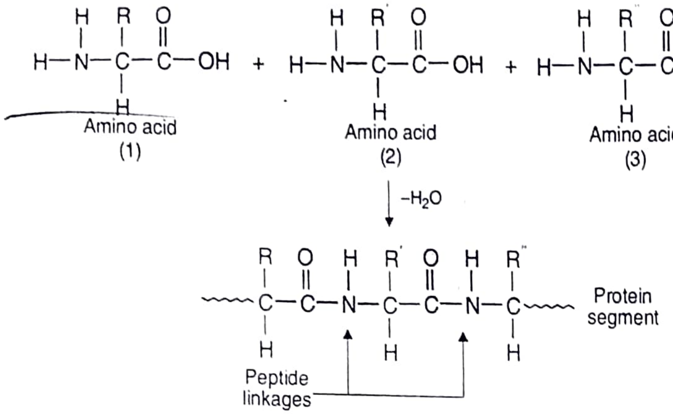
If higher peptides are to be synthesized, the protecting group is left on the amino group and the carboxyl group of each new peptide is reacted with $SOCl_2$ and a new amino acid. Although the process of polypeptide synthesis is laborious, small proteins like ribonuclease with 124 amino acid units have been successfully synthesized.

III. PROTEINS

Proteins are polymers of α-amino acids bonded by peptide linkages. Their molecular weights range from 5000 to many millions. They occur in all living cells. Without proteins life would not be possible.

11

T1



CLASSIFICATION

There are two methods for classifying proteins. One method classifies them according to composition as either *simple proteins* or *conjugated proteins*. The second method classifies them according to their physiological functions.

I. CLASSIFICATION ACCORDING TO COMPOSITION

(A) Simple Proteins. Simple proteins are those which yield only α-amino acids upon hydrolysis. They are further subdivided according to their solubility in various solvents and also whether they are coagulated by heat.

(*a*) **Albumins.** Albumins are water-soluble proteins, which are coagulated by heat. They are found in plants and animals. Examples are egg-albumin and serum-albumin.

(*b*) **Globulins.** Globulins are insoluble in water but soluble in dilute salt solutions, which are coagulated by heat. They are found in plants and animals. Examples are serum-globulin and vegetable-globulin.

(*c*) **Scleroproteins** (*Albuminoids*). Scleroproteins are insoluble in water and most other solvents. They are found only in animals. Example is keratin in hair and fingernails.

(*d*) **Glutenins.** Glutenins are insoluble in water but soluble in dilute acids and alkalis, which are coagulated by heat. Example is glutenin from wheat.

(*e*) **Histones.** Histones are soluble in water but insoluble in dilute ammonium hydroxide, which are not coagulated by heat. They are found in animals. Example is globin in haemoglobin.

(*f*) **Prolamines.** Prolamines are insoluble in water but soluble in 70 per cent ethanol, which are not coagulated by heat. Examples are *zein* from corn and *gliadin* from wheat.

(*g*) **Protamines.** Protamines are soluble in water and dilute ammonium hydroxide, which are not coagulated by heat. Examples are *salmine* from salmon and *sturine* from sturgeon.

(B) Conjugated Proteins. Conjugated proteins are those which yield α-amino acids *plus* a nonprotein material upon hydrolysis. The nonprotein material is called the *prosthetic* group (Gr. *prosthesis*, an addition). The conjugated proteins are also subdivided into several classes according to the nature of the prosthetic group.

(*a*) **Glycoproteins.** Glycoproteins contain a carbohydrate derivative as their prosthetic group. *Mucin*, a constituent of saliva, is a glycoprotein.

(*b*) **Phosphoproteins.** Phosphoproteins are proteins which contain α-amino acids linked to phosphoric acid. Casein, which is found in milk, is an example of this class.

(*c*) **Chromoproteins.** Chromoproteins consist of a pigmented prosthetic group combined with a simple protein. Haemoglobin and cytochromes are examples of chromoproteins.

(*d*) **Nucleoproteins.** Nucleoproteins are complex substances that occur abundantly in the nuclei of plant and animal cells. The prosthetic groups are nucleic acids. Examples of nucleoproteins are the *nuclein* and *nucleohistones* of glandular tissues, *yeast*, *chromosomes*, and other materials rich in cell nuclei.

(*e*) **Lipoproteins.** Lipoproteins consist of cholesterol esters and phospholipids attached to protein molecules. They are found in brain and nerve tissue and are an integral part of cell membranes.

II. CLASSIFICATION ACCORDING TO FUNCTIONS

The functional classification of proteins includes the following groups :

(1) **Structural Proteins.** These are fibrous proteins, such as collagen which comprises half of man's total protein in the form of skin, cartilage and bone.

(2) **Contractile Proteins.** Contractile proteins are found in muscles. Examples are *myosin*, *actin*.

(3) **Hormones.** Many proteins function as *hormones*, that is, as communication links between different parts of organism. *Insulin* is a common example of a protein hormone.

(4) **Enzymes.** Proteins of this group serve as catalyst for the chemical reactions in living organisms, rendering specificity and control to these reactions. *Pepsin*, and *trypsin* are examples of the class.

(5) **Antibodies.** When the body is invaded by infectious species that release foreign proteins or antigens, antibodies are produced to remove the invading species from the system. *Gamma globulins* present in the blood are examples of antibodies.

(6) **Blood Proteins.** The three major protein constituents of the blood are *albumins*, *hemoglobin* and *fibrinogen*. Their presence contributes to the maintenance of osmotic pressure, oxygen transport and blood coagulation respectively.

STRUCTURE OF PROTEINS

Proteins have definite three-dimensional structure. There are a number of factors which determine the exact shape of a protein. These are considered in terms of four levels of structural organization called the primary, secondary, tertiary, and quaternary structures of the protein. Each succeeding level of organization is more complex than the previous one and is a direct result of the chemical features of the previous levels.

(1) **Primary Structure.** The primary structure of proteins refers to the sequence of amino acids held together by peptide linkages. The amino acid sequence in proteins can be determined by the methods used for peptides. These methods have been applied to a number of proteins and several generalizations regarding the primary structure of proteins have been made. They are : (*a*) Proteins are made up of L-amino acids only ; (*b*) A protein may contain more than one amino acid chain. If so, the chains are usually bonded to each other at specific points by disulfide —S—S— linkages; (*c*) Sequence of amino acids along the protein chains is essentially random. Repeating sequences within a protein molecule are not common ; and (*d*) Small variations in the sequence of amino acids have pronounced effects on the chemical and physical properties of protein.

(2) **Secondary Structure.** The secondary structure of a protein refers to the shape in which the long amino acid chain exists. Many proteins consist of amino acid chain coiled into a spiral known as a **helix**. Such a helix may be either right- or left-handed, as in the case of screws. The right-handed helix is known as the α-helix, and the left-handed helix is known as the β-helix. It has been found that an α-helix constitutes the more stable arrangement. The spiral is held together by hydrogen bonds between N–H and C=O groups vertically adjacent to one another in the helix (Fig.27.7). X-Ray studies have shown that there are approximately 3.6 amino acid units for each turn in the helix.

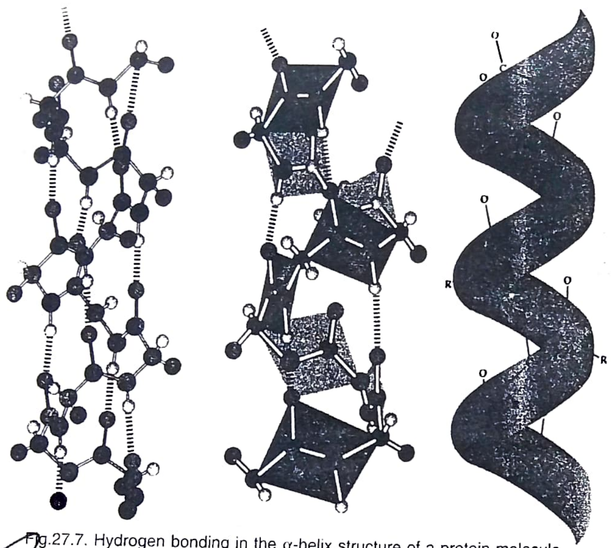


Fig.27.7. Hydrogen bonding in the α-helix structure of a protein molecule.

(3) **Tertiary Structure.** An α-helix may be considered to be a piece of a rope which is free to bend, twist, and fold. The tertiary structure of a protein refers to the final three-dimensional shape that results from the twisting, bending, and folding of the protein helix.

(4) **Quaternary Structure.** Complex proteins are often formed from two or more amino acid chains rather than a single amino acid chain. Each chain is a complete protein with a characteristic primary, secondary, and tertiary structure. The quaternary structure refers to the way in which these amino acid chains of a complex protein are associated with each other.

PROPERTIES OF PROTEINS

Some of the general properties and reactions of proteins are described below.

(1) Most proteins are colorless amorphous substances with no definite melting points or boiling points. On heating they undergo decomposition. They are amphoteric in nature and, like amino acids, exist as *Dipolar ions* or *Zwitterions*. Most proteins are levorotary.

(2) **Colloidal Nature.** Proteins form colloidal dispersions in water. Protein, being colloidal, will pass through a filter paper but not through a membrane. The inability of the protein to pass through a membrane is of great importance in the body. Proteins present in the bloodstream cannot pass through the cell membranes and should remain in the bloodstream. Since proteins cannot pass through membranes, there should be no protein material present in the urine. The presence of protein in the urine indicates damage to the membranes in the kidneys.

(3) **Isoelectric Points.** Proteins have isoelectric points in the same way as do amino acids. This is because of the presence on the protein chain of additional acidic or basic groups that are not involved in peptide linkages. The isoelectric points of some proteins are given in Table 27.4. Most proteins show minimum solubility and stability at their isoelectric points.

12

ISOELECTRIC POINTS OF SOME PROTEINS

Protein	Isoelectric Point
Casein	4.60
Gelatin	4.80 - 4.85
Serum albumin	4.88
Insulin	5.30 - 5.35
Serum globulin	5.50
Haemoglobin	6.79 - 6.83

(4) **Precipitation.** Proteins are easily precipitated (or coagulated) by certain agents. Many of the normal functions in the body are essentially precipitation reactions ; for example, the clotting of blood or the precipitation of casein during digestion. Since animal tissues are chiefly protein in nature, reagents that precipitate protein will have a marked toxic effect if introduced into the body. Bacteria, which are mainly protein, are effectively destroyed when treated with suitable precipitants. Many of the common poisons and disinfectants act in this way. Precipitation of proteins is an irreversible change and the precipitated protein is said to be **Denatured**. Some of the common methods of protein precipitation are described below :

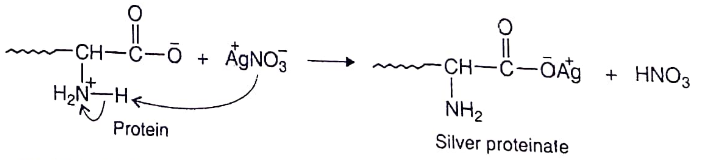
(*a*) **By Heat.** Heat coagulates almost all protein solutions. Egg-white, a substance containing a high percentage of protein, coagulates on heating. Heat coagulates and destroys protein present in bacteria. Hence sterilization of instruments and clothing for use in operating rooms of hospitals requires the use of high heats. Routine examinations of urine specimens for protein are made by heating the urine in a test tube to coagulate any protein that might be present.

(*b*) **By Alcohol.** Alcohol coagulates all types of protein except prolamines. A 70 per cent solution of ethanol is used as a disinfectant because of its ability to coagulate the protein present in bacteria.

(*c*) **By Concentrated Inorganic Acids.** Proteins are precipitated from their solutions by acids such as hydrochloric, sulfuric, and nitric acids. This forms the basis of **Heller's Ring Test** which is used to detect the presence of albumin in urine. Concentrated nitric acid is added slowly down the sides of a test tube containing the urine sample. If albumin is present, it will precipitate out as a white ring at the interface of the two liquids.

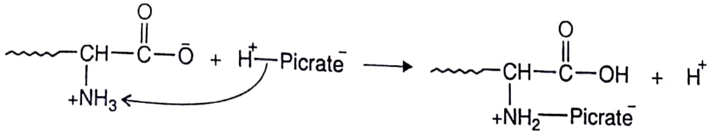
(*d*) **By Salts of Heavy Metals.** Heavy metal salts such as mercuric chloride or silver nitrate coagulate proteins. They are very poisonous if taken internally because they coagulate and destroy protein present in the body. The antidote for the mercuric chloride or silver nitrate, when these poisons are taken internally, is egg-white. The heavy metal salts react with the egg-white and precipitate out. The precipitates thus formed must be removed from the stomach by an emetic or else the stomach will digest the egg-white and return the poisonous material to the system.

Dilute silver nitrate solution is used as a disinfectant the eyes of newborn infants. Strong solutions of silver nitrate are used to causterize fissures and destroy excessive granulation of tissues. The reaction of a protein with silver nitrate may be illustrated as follows.



$$\sim\!\sim\text{CH}(\overset{+}{\text{N}}\text{H}_3)-\text{C}(=\text{O})-\bar{\text{O}} + \overset{+}{\text{Ag}}\text{NO}_3^- \longrightarrow \sim\!\sim\text{CH}(\text{NH}_2)-\text{C}(=\text{O})-\bar{\text{O}}\overset{+}{\text{Ag}} + \text{HNO}_3$$

(*e*) **By Alkaloidal Reagents.** Alkaloidal reagents such as tannic acid and picric acid precipitate proteins from solutions. The reaction of a protein with picric acid may be illustrated as follows.



Protein picrate

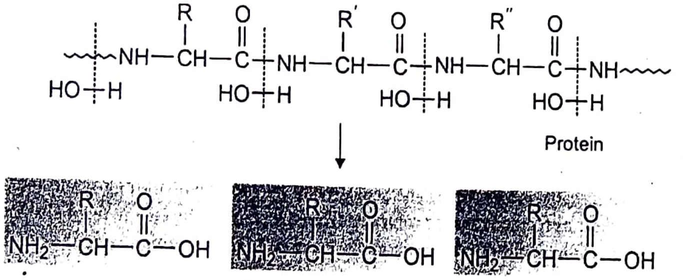
Tannic acid has been used extensively in the treatment of burns. When this substance is applied to a burn area, it causes the protein to precipitate as a tough covering, thus reducing the amount of water loss from the area. It also reduces exposure to air.

Newer drugs have taken the place of tannic acid for burns, but an old-fashioned remedy still in use for emergencies involves the use of strong tea (which contains tannic acid).

(*f*) **By Salting Out.** Most proteins are insoluble in concentrated salt solutions and precipitate out unchanged. To separate a protein from a mixture of other substances, the mixture is placed in a concentrated salt solution. The protein precipitates out and is removed by filtration. The protein is then purified from the remaining salt solution by the process of dialysis. Dialysis is the separation of solution particles from colloidal particles by means of a membrane.

(*g*) **By Radiation.** Ultraviolet light or X-rays cause precipitation of proteins. The radiations provide kinetic energy to cause excessive vibration of atoms in protein molecules. As a result, hydrogen bonds and salt linkages are broken and coagulation occurs. Thus ultraviolet radiation destroys bacteria by denaturing some of their vital enzymes.

(5) **Hydrolysis.** Proteins can be hydrolyzed by acids (HCl and H_2SO_4), alkalis (NaOH), or enzymes (*proteases*). The final products of hydrolysis are the amino acids, together with the prosthetic groups from any conjugated proteins present. Hydrolysis by alkalis causes racemization of the optically active amino acids, whereas hydrolysis by acids does not. Hydrolysis by enzymes is slow and often incomplete but does not cause racemization or decomposition of the sensitive amino acids.



Protein

Mixture of Amino acids

6 x 9 x

(6) **Oxidation.** Proteins are oxidized on burning and putrefaction. The products include nitrogen, amines, carbon dioxide and water. The bad smell produced during putrefaction of dead animals is largely due to the formation of amines and other nitrogen products by the bacterial oxidation of body proteins.

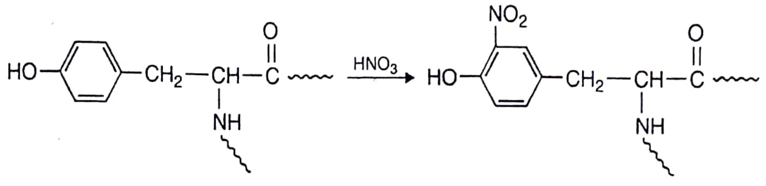
COLOR TESTS OF PROTEINS

Proteins give a number of color reactions which serve as their test. In fact, these reactions are due to the presence of certain specific groups in the protein structure. Therefore, a given protein may not respond to all the general tests listed below.

(1) **Xanthoproteic Test.** When a protein is warmed with conc. HNO_3, a yellow color is produced.

Protein + Conc. HNO_3 = Yellow Color

This test is given by a protein that consists of amino acids containing a benzene ring, such as tyrosine and tryptophan. The aromatic rings undergo nitration to give yellow products.



Tyrosine

Yellow

(2) **Biuret Test.** A small amount of NaOH solution is added to the protein. After mixing the two reagents, copper sulfate solution is added. Violet color is produced.

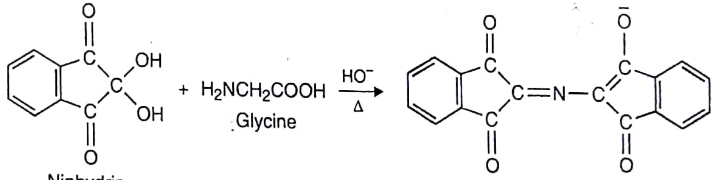
Protein + NaOH + Copper sulfate = Violet Color

This test is given by all proteins that contain two or more peptide linkages.

(3) **Ninhydrin Test.** When proteins are boiled with a dilute aqueous solution of ninhydrin, a violet color is produced.

Protein + Ninhydrin = Violet Color

This test is given by α-amino acids, proteins, and dipeptides.



Ninhydrin (2 molecules)

Glycine

Violet-complex

(4) **Millon's Test.** The Million's reagent consists of mercury dissolved in nitric acid (forming a mixture of mercuric and mercurous nitrates and nitrites). When Millon's reagent is added to a protein, a white precipitate is formed.

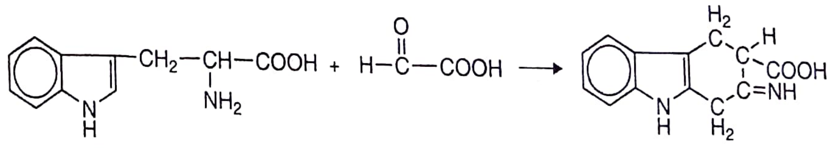
Protein + Millon's Reagent = White Precipitate

This test is given by proteins which yield tyrosine on hydrolysis.

(5) **Hopkins-Cole Test.** Concentrated sulfuric acid is added down the side of a test-tube containing a solution of protein and glyoxalic acid, to form a layer. A violet ring appears between the two layers.

Protein + Glyoxalic acid + conc. H_2SO_4 = Violet Ring

This test is given by proteins which have the amino acid tryptophan in its structure.



Tryptophan

Glyoxalic acid

Violet complex

(6) **Sulfur Test.** When a sulfur-containing protein is heated with a mixture of sodium hydroxide and lead acetate solutions, a black precipitate of lead sulfide is formed.

ISOLATION AND PURIFICATION OF PROTEINS

Most proteins occur as mixtures with other proteins. Their separation is difficult because proteins have similar properties. Column chromatography is commonly used to separate protein mixtures. Sometimes careful control of the pH of a solution of proteins also results in one being precipitated.

It is also difficult to determine whether a protein is pure or not. This is because most proteins do not have sharp melting points and decompose on heating. The most common method of testing the purity of a protein is by the use of *ultracentrifuge*. In this method, the protein is spun at a very high speed. Due to centrifugal force, the protein moves to the outer end of the spinning cell at a rate which depends upon its size. A special optical system enables the solution to be photographed during this process and reveals the moving protein. If impurities of different molecular weights are present, they will travel at different rates and consequently can be detected in the resulting photographs.

DETERMINATION OF PROTEINS

It is often desirable to know the protein content of various foods and biological materials. The analysis of the protein content of such a material is based on its nitrogen content. Since the average nitrogen content of proteins is 16 per cent, the protein content of a substance may be obtained by multiplying its nitrogen value by the factor 100/16 = 6.25. For example, if a certain food contains 2 per cent nitrogen, on analysis its protein content would equal 2 times 6.25, or 12.5 per cent.

The total nitrogen content of proteins and peptides may be determined by the *Dumas method* or the *Kjeldahl method*. The aromatic amino acids in proteins absorb ultraviolet light at a wavelength of 2800 Å. The measurement of light absorption at 2800Å is a convenient method for determining the amount of protein in solution.

INDUSTRIAL IMPORTANCE OF PROTEINS

Industrially, proteins have great importance. We are familiar with the use of leather made by tanning of hides. This is essentially a precipitation of the protein by tannic acid. Gelatin is obtained by heating bones, skin, and tendons in water. Gelatin is used in desserts, salads, candies, bakery goods, etc. Wool and silk are also protein materials.

Casein is another protein that has been used industrially for a long time. Casein plastics are used in the manufacture of buttons and buckles. Casein is also used in sizing of paper and in making casein glues, cosmetics, hard rubber, insecticide sprays, linoleum, paint, plywood, safety glass, and veneer. Casein forms the basis of artificial wool fiber known as *Lanital*.

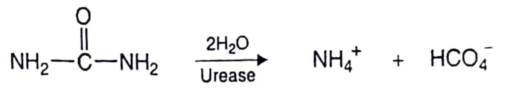
The proteins of soybeans are used in the manufacture of plastics and can be converted into filaments which may be spun and dyed.

ENZYMES

Enzymes are usually **globular proteins** that catalyze biochemical reactions. That is, they enable biochemical processes to occur at useful rates under mild conditions. Enzymes, and all catalysts, lower the activation energy and thus speed up reactions that are spontaneous (have negative ΔH values) but that are slow in the absence of the enzyme. All biochemical processes are catalyzed by enzymes, and since thousands of biochemical processes occur, thousands of enzymes are synthesized in the cell. They aid in the synthesis of all molecules, including enzyme molecules, as well as in hydrolysis, energy production, and other vital cell processes. A very precise 2° and 3° structure is needed in order for an enzyme to function as a catalyst, and because the structures of the proteins are pH and solvent dependent, a small change in cellular pH from the normal 7.4 can destroy enzyme function and spell disaster for the cell. In fact, a cell cannot function if its pH changes by more than a few tenths of a unit. The presence of alcohol or other organic solvents can also change the enzyme structure and stop catalytic activity.

(14)

The efficiency of these amazing catalysts can be illustrated by the reaction of urea to form an ammonium ion (NH_4^+) and a bicarbonate ion (HCO_3^-).



$$NH_2-\overset{O}{\overset{\|}{C}}-NH_2 \xrightarrow[\text{Urease}]{2H_2O} NH_4^+ + HCO_4^-$$

Without a catalyst, at room temperature and a pH of 7, this reaction is very slow–only 10^{-6} moles react in a year. In the presence of urease (an enzyme) the reaction occurs at the rate of 10^4 moles per second.

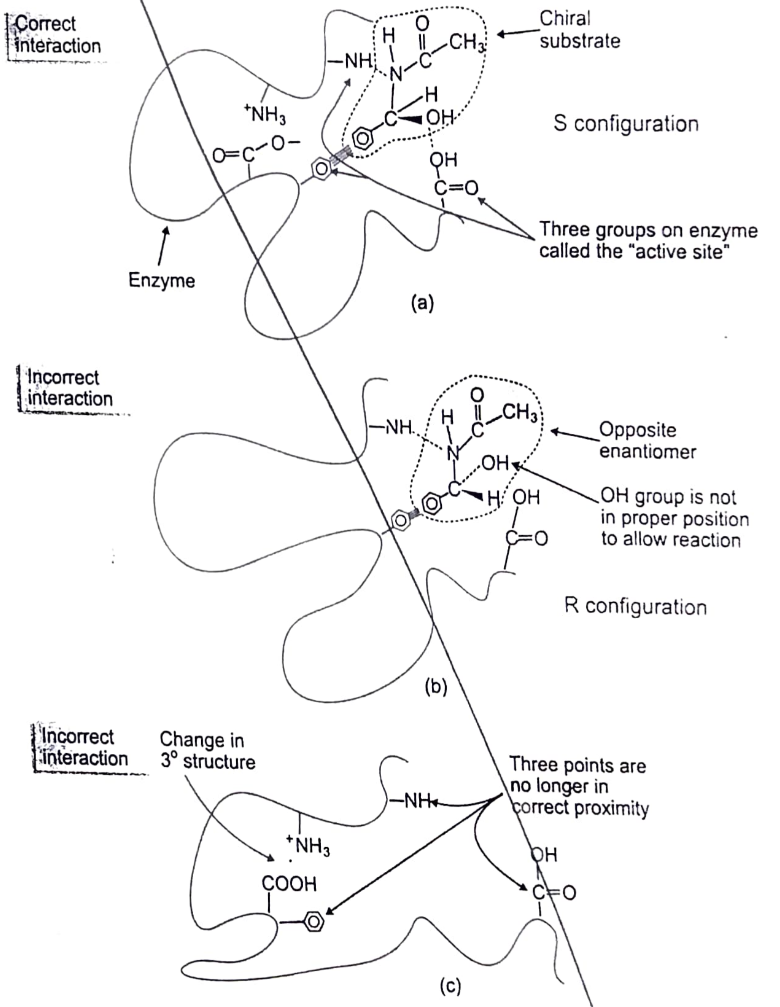


Fig.27.8. General enzyme-substrate interaction.