CHAPTER
4

Proteins:
Structure and Function

CHAPTER AT A GLANCE

The reader will be able to answer questions on the following topics:

1. Peptide bonds
2. Primary structure of proteins
3. Secondary structure
4. Tertiary structure
5. Quaternary structure
6. Sequence analysis (study of primary structure)
7. Iso-electric pH of proteins
8. Precipitation reactions of proteins
9. Classification of proteins
10. Quantitative estimation of proteins

The word protein is derived from Greek word, "proteios" which means primary. As the name shows, the proteins are of paramount importance for biological systems. Out of the total dry body weight, 3/4ths are made up of proteins. Proteins are used for body building; all the major structural and functional aspects of the body are carried out by protein molecules. Abnormality in protein structure will lead to molecular diseases with profound alterations in metabolic functions.

Proteins contain Carbon, Hydrogen, Oxygen and Nitrogen as the major components while Sulphur and Phosphorus are minor constituents. Nitrogen is characteristic of proteins. **On an average, the nitrogen content of ordinary proteins is 16% by weight.** All proteins are polymers of amino acids.

Amino Acids are Linked by Peptide Bonds

Alpha carboxyl group of one amino acid reacts with alpha amino group of another amino acid to form a peptide bond or CO-NH bridge (Fig. 4.1A).

Proteins are made by polymerisation of amino acids through peptide bonds. Two amino acids are combined to form a **dipeptide**; three amino acids form a **tripeptide**; four will make a **tetrapeptide**; a few amino acids together will make an **oligopeptide**; and combination of 10 to 50 amino acids is called as a **polypeptide.** By convention, big polypeptide chains containing more than 50 amino acids are called **proteins**.

In a tripeptide, there are 3 amino acids, but these 3 can be any of the total 20 amino acids. Thus 20^3 = 8000 different permutations and combinations are possible in a tripeptide. An ordinary protein having about 100 amino acids, will have 20^{100} different possibilities. This number is more than the total number of atoms present in the whole universe. Thus, even though there are only 20 amino acids, by changing the sequence of combination of these amino acids, nature produces enormous number of markedly different proteins.

STRUCTURE OF PROTEINS
(Organisation of Proteins)

Proteins have different levels of structural organisation; primary, secondary, tertiary and quaternary.

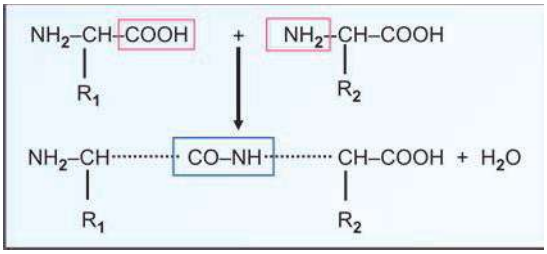

Fig. 4.1A. Peptide bond formation

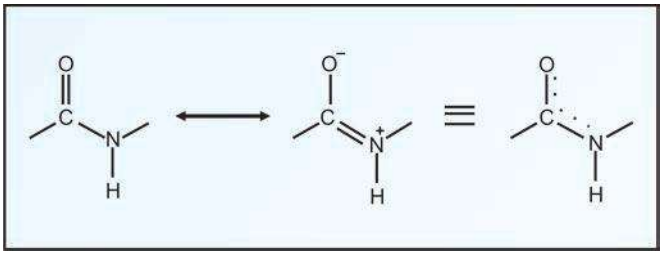

Fig. 4.1B. Peptide bond is a partial double bond

1. Primary Structure

1-A. Sequence of amino acids in proteins

Protein structure is studied as the primary, secondary, tertiary and quaternary levels (Box 4.1). **Primary structure denotes the number and sequence of amino acids in the protein.** The higher levels of organisation are decided by the primary structure. Each polypeptide chain has a unique amino acid sequence decided by the genes. The primary structure is maintained by the covalent peptide bonds (Fig. 4.1A).

Students should have a clear concept of the term "sequence". See the following example:

Gly - Ala - Val (1)

Gly - Val - Ala (2)

Both the tripeptides shown above contain the same amino acids; but their sequence is altered. When the sequence is changed, the peptide is also different.

1-B. Characteristics of a Peptide Bond

i. The peptide bond is a **partial double bond**.
ii. The **C–N bond is 'trans'** in nature and there is no freedom of rotation because of the partial double bond character. (Fig. 4.1B)
iii. The distance is 1.32Å which is midway between single bond (1.49 Å) and double bond (1.27Å).
iv. The side chains are free to rotate on either side of the peptide bond.
v. The angles of rotation known as **Ramachandran angles**, therefore determine the spatial orientation of the peptide chain (Fig. 4.2). (Dr GN Ramachandran did pioneering work on the structural aspects of proteins during 1950s and 1960s).

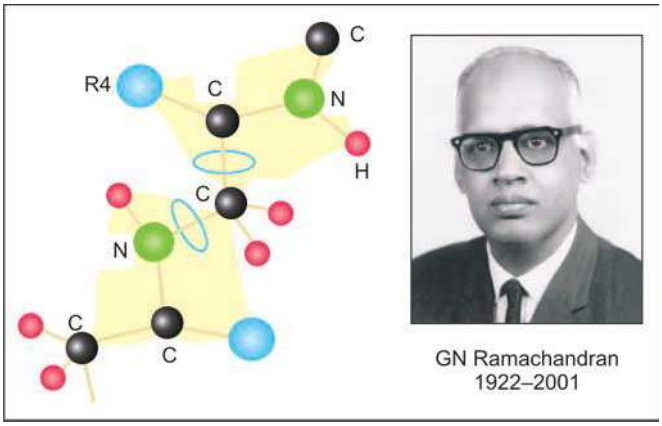

Fig. 4.2. Angles of rotation in a peptide bond

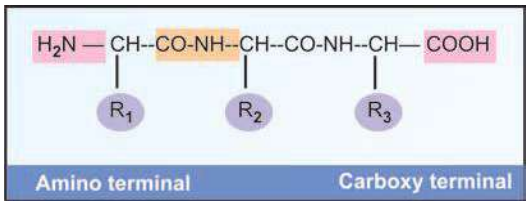

Fig. 4.3. End groups of polypeptide chain

1-C. Numbering of Amino Acids in Proteins

i. In a polypeptide chain, at one end there will be one free alpha amino group. This end is called the **amino terminal (N-terminal) end** and the amino acid contributing the alpha-amino group is named as the **first amino acid**. (Fig. 4.3).
ii. Usually the N-terminal amino acid is written on the left hand side when the sequence of the protein is denoted. Incidentally, the bio-synthesis of the protein also starts from the amino terminal end.
iii. The other end of the polypeptide chain is the **carboxy terminal end (C-terminal)**, where there is a free alpha carboxyl group which is contributed by the **last amino acid** (Fig. 4.3). All other alpha amino and alpha carboxyl groups are involved in peptide bond formation.
iv. Amino acid residues in polypeptides are named by changing the suffix "-ine" to "-yl", for example, Glycine to Glycyl. Thus, peptide bonds formed by carboxyl group of glycine with amino group of Alanine, and then carboxyl group of Alanine with amino group of Valine and is called glycyl-alanyl-valine and abbreviated as
NH_2-Gly-Ala-Val-COOH
or Gly-Ala-Val
or simply as GAV

Box 4.1. Definitions of Levels of Organization

1. **Primary structure** of protein means the order of amino acids in the polypeptide chain and the location of disulfide bonds, if any.
2. **Secondary structure** is the steric relationship of amino acids, close to each other.
3. **Tertiary structure** denotes the overall arrangement and interrelationship of the various regions, or domains of a single polypeptide chain.
4. **Quaternary structure** results when the proteins consist of two or more polypeptide chains held together by non-covalent forces.

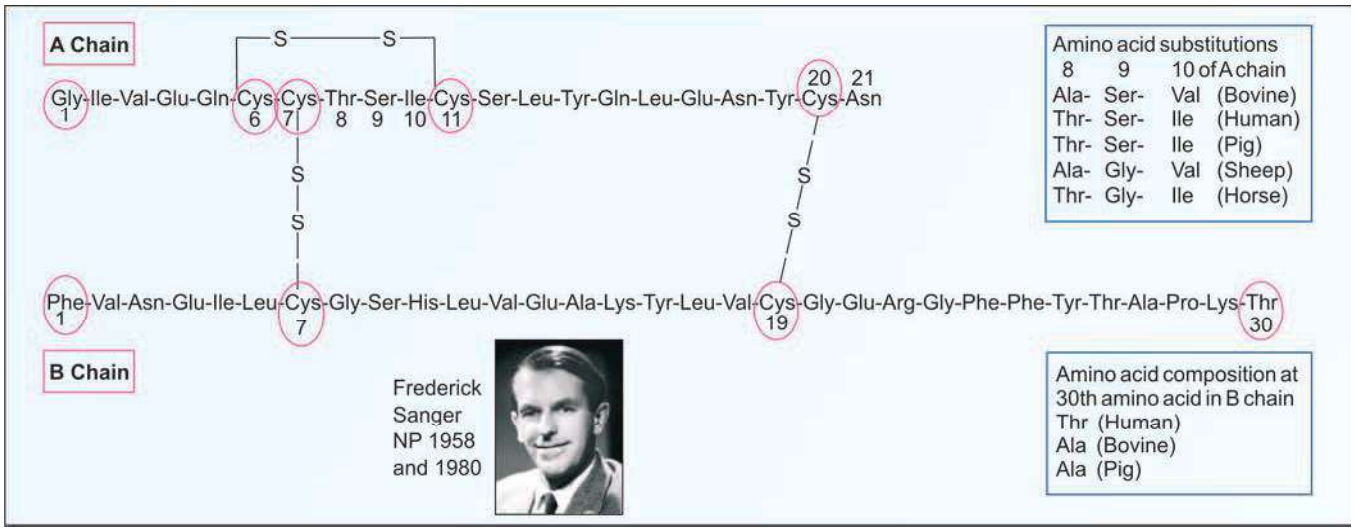

Fig. 4.4. Primary structure of human insulin

1-D. Branched and Circular Proteins

i. Generally, the polypeptide chains are linear. However, branching points in the chains may be produced by interchain **disulphide bridges**. The covalent disulphide bonds between different polypeptide chains in the same protein (interchain) or portions of the same polypeptide chain (intrachain) are also part of the primary structure.

ii. Rarely, instead of the alpha COOH group the **gamma carboxyl** group of glutamic acid may enter into peptide bond formation, e.g. Glutathione (gamma-glutamyl-cysteinyl-glycine) (Fig.15.19).

The term **pseudopeptide** (or isopeptide) is used to denote such a peptide bond formed by carboxyl group, other than that present in alpha position.

iii. Very rarely, protein may be in a circular form, e.g. Gramicidin.

1-E. Primary Structure of Insulin

As an example of the primary structure of a protein, that of insulin is shown in Fig. 4.4. This was originally described by Sanger in 1955 who received the Nobel prize in 1958.

i. Insulin has **two polypeptide chains.** The A chain (**Glycine** chain) has 21 amino acids and B (**Phenyl alanine**) chain has 30 amino acids.

ii. They are held together by **two interchain disulphide bonds** (Fig. 4.4). A chain 7th cysteine and B chain 7th cysteine are connected. Similarly A chain 20th cysteine and B chain 19th cysteine are connected. There is another **intrachain** disulphide bond between 6th and 11th cysteine residues of A chain.

iii. The species variation is restricted to amino acids in position 8, 9 and 10 in A chain and in C-terminal of B chain (Fig. 4.4). The amino acid sequence has been conserved to a great extent during evolution.

iv. The porcine insulin and human insulin are structurally similar, except the terminal amino acid in B chain (Thr → Ala) (Fig. 4.4). Bovine Insulin may produce antibodies in humans by repeated injections. But de-alaninated porcine

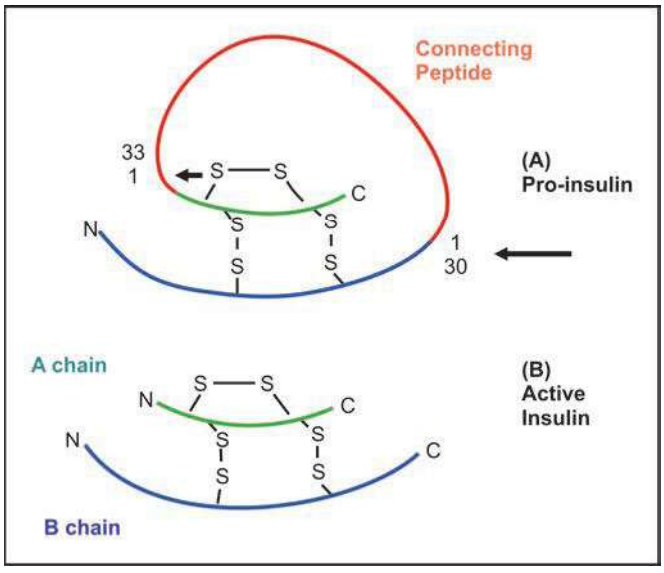

Fig. 4.5. Conversion of Pro-insulin to active insulin. Arrows = site of action of proteolytic enzymes

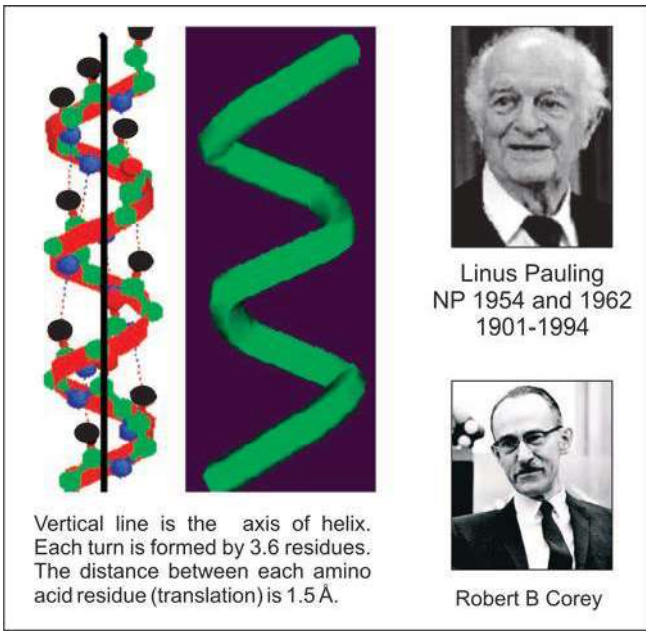

Fig. 4.6. Structure of alpha helix

Insulin, bearing no antigenic difference from human Insulin will not produce antibodies in diabetic patients even after a long-term use. Nowadays human Insulin is being produced by recombinant DNA technology.

1-F. Pro-insulin

Beta cells of pancreas synthesize insulin as a prohormone. Proinsulin is a **single polypeptide chain** with 86 amino acids. Biologically active insulin (2 chains) is formed by removal of the central portion of the pro-insulin before release. The **C-peptide** (connecting peptide) is also released into the circulation (Fig. 4.5).

1-G. Primary Structure Determines Biological Activity

A protein with a specific primary structure, when put in solution, will automatically form its natural three dimensional shape. So the higher levels of organization are dependent on the primary structure.

Even a single amino acid change (**mutation**) in the linear sequence may have profound biological effects on the function. For example, in HbA (normal hemoglobin) the 6th amino acid in the beta chain is glutamic acid; it is changed to valine in HbS (**sickle cell anemia**).

Box 4.2. Configuration and Conformation

Configuration of a protein denotes the spatial relationship between particular atoms, e.g. L and D amino acids. Conformation means the spatial relationship of every atom in a molecule, e.g. rotation of a portion of the molecule

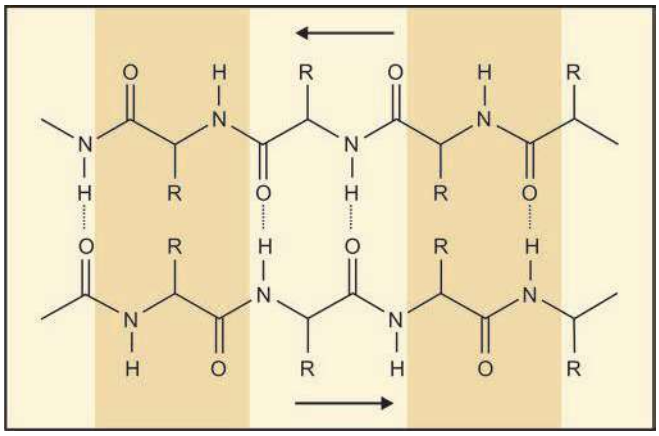

Fig. 4.7. Structure of beta-pleated sheet

2. Secondary Structure of Proteins

The term "secondary structure" denotes the configurational relationship between residues which are about 3–4 amino acids apart in the linear sequence (Box 4.2). Secondary and tertiary levels of protein structure are preserved by **noncovalent forces** or bonds like hydrogen bonds, electrostatic bonds, hydrophobic interactions and van der Waals forces. These forces are described in Chapter 1.

i. **A hydrogen bond** is a weak electrostatic attraction between one electronegative atom like O or N and a hydrogen atom covalently linked to a second electronegative atom. Hydrogen atoms can be donated by -NH (imidazole, indole, peptide); -OH (serine, threonine) and -NH_2 (arginine, lysine). Hydrogen accepting groups are COO^- (aspartic, glutamic) C=O (peptide); and S–S (disulphide).

ii. **Electrostatic bonds (ionic bonds)**: Positive charges are donated by epsilon amino group of lysine, guanidinium group of arginine and imidazolium group of histidine. Negative charges are provided by beta and gamma carboxyl groups of aspartic and glutamic acids.

iii. **Hydrophobic bonds** are formed by interactions between nonpolar hydrophobic side chains by eliminating water molecules. This serves to hold lipophilic side chains together.

iv. The **van der Waals forces** are very weak, but collectively contribute maximum towards the stability of protein structure.

2-A. Alpha helix

Pauling (Nobel prize, 1954) and Corey described the alpha-helix and beta-pleated sheet structures of polypeptide chains in 1951.

i. **The alpha-helix is the most common and stable conformation** for a polypeptide chain. In proteins like hemoglobin and myoglobin, the alpha-helix is abundant, whereas it is virtually absent in chymotrypsin.
ii. The alpha helix is a **spiral structure** (Fig. 4.6). The polypeptide bonds form the back-bone and the side chains of amino acids extend outward.
iii. The structure is stabilized by hydrogen bonds between NH and C=O groups of the main chain.
iv. Each turn is formed by 3.6 residues. The distance between each amino acid residue (translation) is 1.5 Å.
v. The alpha-helix is generally **right handed**. Left handed alpha helix is rare, because amino acids found in proteins are of L-variety, which exclude left handedness. Proline and hydroxy proline will not allow the formation of alpha-helix.

2-B. Beta-pleated sheet

i. The polypeptide chains in beta-pleated sheet is almost fully extended. The distance between adjacent amino acids is 3.5Å.
ii. It is stabilized by hydrogen bonds between NH and C=O groups of neighboring polypeptide segments.

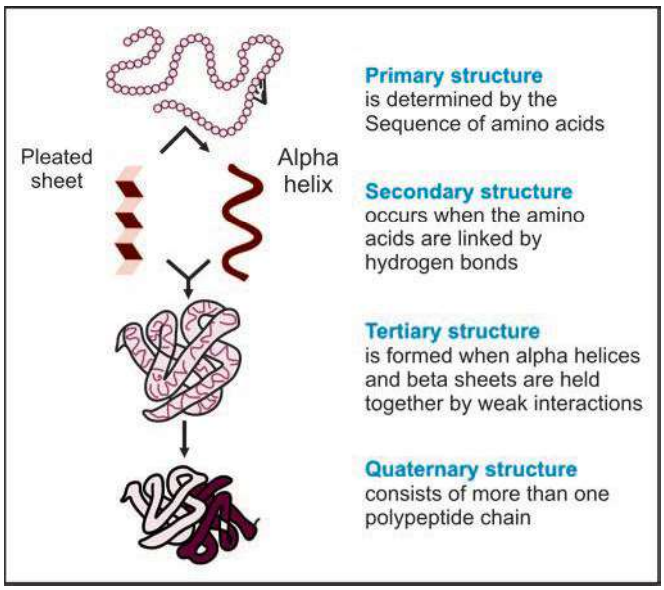

Fig. 4.8. Levels of organizations of proteins

iii. Adjacent strands in a sheet can run in the same direction with regard to the amino and carboxy terminal ends of the polypeptide chain (parallel) or in opposite direction (anti parallel beta sheet) (Fig. 4.7). Beta-pleated sheet is the major structural motif in proteins like silk Fibroin (anti parallel), Flavodoxin (parallel) and Carbonic anhydrase (both).
iv. Beta bends may be formed in many proteins by the abrupt U-turn folding of the chain. Intrachain disulfide bridges stabilize these bends.

2-C. Collagen helix

It is a triple helical structure found in collagen (details in Chapter 52).

3. Tertiary Structure

i. Secondary structure denotes the configurational relationship between residues which are about 3-4 amino acids apart; or secondary level defines the organization at immediate vicinity of amino acids. The tertiary structure denotes three dimensional structure of the whole protein (Box 4.1 and Fig. 4.8). The tertiary structure defines the steric relationship of amino acids which are far apart from each other in the linear sequence, but are close in the three-dimensional aspect.
ii. The tertiary structure is maintained by **non-covalent** interactions such as hydrophobic bonds, electrostatic bonds and van der Waals forces. The tertiary structure acquired by native protein is always thermodynamically most stable.
iii. **Domain** is the term used to denote a compact globular functional unit of a protein. A domain is a relatively independent region of the protein, and may represent a functional unit. The domains are usually connected with relatively flexible areas of protein (see immunoglobulins, Chapter 49). Phenyl alanine hydroxylase enzyme contains 3 domains, one regulatory, one catalytic and one protein-protein interaction domains.

4. Quaternary Structure

i. Certain polypeptides will **aggregate to form one functional protein** (Box 4.1 and Fig. 4.8). This is referred to as the quaternary structure.
ii. The protein will lose its function when the subunits are dissociated.

iii. The forces that keep the quaternary structure are hydrogen bonds, electrostatic bonds, hydrophobic bonds and van der Waals forces.
iv. Depending on the number of polypeptide chain, the protein may be termed as monomer (1 chain), dimer (2 chains), tetramer (4 chains) and so on. Each polypeptide chain is termed as **subunit** or **monomer. Homodimer** contains two copies of the same polypeptide chain. **Heterodimer** contains two different types of polypeptides as a functional unit.
v. For example, 2 alpha-chains and 2 beta-chains form the **Hemoglobin** molecule. Similarly, 2 heavy chains and 2 light chains form one molecule of **immunoglobulin** G. Creatine kinase (CK) is a dimer. Lactate dehydrogenase (LDH) is a tetramer.

Structure-Function Relationship

The functions of proteins are maintained because of their ability to recognize and interact with a variety of molecules. The three dimensional structural conformation provides and maintains the functional characteristics. The three dimensional structure, in turn, is dependent on the primary structure. So, any difference in the primary structure may produce a protein which cannot serve its function. To illustrate the structure-function relationship, the following three proteins are considered; each belongs to a different class in the functional classification.

Enzymes: The first step in enzymatic catalysis is the binding of the enzyme to the substrate. This, in turn, depends on the structural conformation of the active site of the enzyme, which is precisely oriented for substrate binding (see Chapter 5). Carbonic anhydrase catalyses the reversible hydration of carbon dioxide. This enzyme makes it possible for the precise positioning of the CO_2 molecule and the hydroxyl (OH^-) ion for the formation of bicarbonate ion. The zinc ion is located at a deep cleft coordinated to histidine residues. The CO_2 binding residues are very near to the zinc ion. Water binds to zinc ion, gets ionized to hydroxyl ion and it binds to the CO_2 which is proximally located. The substrates are brought in close proximity for the reaction to proceed.

Transport proteins: Hemoglobin, the transporter of oxygen is a tetrameric protein (alpha2, beta2), with each monomer having a heme unit (see Chapter 22). Binding of oxygen to one heme facilitates oxygen binding by other subunits. Binding of H^+ and CO_2 promotes release of O_2 from hemoglobin. This allosteric interaction is physiologically important, and is termed as Bohr effect. Even a single amino acid substitution alters the structure and thereby the function. For example, in sickle cell anemia (HbS), the 6th amino acid in the beta chain is altered, leading to profound clinical manifestations.

Structural proteins: Collagen is the most abundant protein in mammals and is the main fibrous component of skin, bone, tendon, cartilage and teeth. Collagen forms a superhelical cable where the 3 polypeptide chains are wound around itself (Chapter 52). In collagen, every 3rd residue is a glycine. The only amino acid that can fit into the triple stranded helix is glycine. The triple helix of collagen is stabilized by the steric repulsion of the rings of hydroxyproline and also by the hydrogen bonds between them. In vitamin C deficiency, failure of hydroxylation of proline/lysine leads to reduced hydrogen bonding and consequent weakness of collagen (Chapter 52). The **quarter staggered triple helical structure** of collagen is responsible for its tensile strength.

STUDY OF PROTEIN STRUCTURE

The first protein to be sequenced was insulin by Sanger in 1955 (Nobel prize in 1958). Before studying the structure, first a pure sample of the protein has to be available. The proteins are extracted and purified by various chromatography techniques (ion exchange, adsorption, partition, size exclusion, affinity, HPLC). The purity of the protein thus isolated is studied by electrophoresis (agar, PAGE, iso electric focussing). Further, molecular weight is determined by mass spectroscopy or by MALDI. Principles of all the above-said techniques are discussed in Chapter 54.

A. Steps for Determining the Primary Structure

1. Determination of the number of polypeptide chains in a protein. This is ascertained by treating them with **Dansyl chloride,** which combines with the N-terminal amino acid (Fig. 4.10). The tagged polypeptide chains are subjected to **complete hydrolysis** by boiling with 6 N HCl at 110°C for 18–36 hours under anerobic conditions to give a mixture of amino acids. The number and nature of the dansyl amino acids can be determined and will indicate the number of polypeptide chains in the protein.

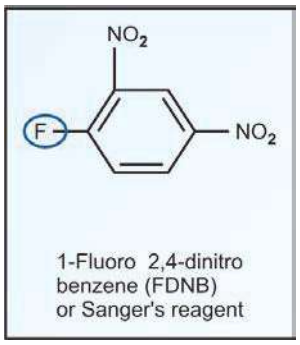

Fig. 4.9 FDNB

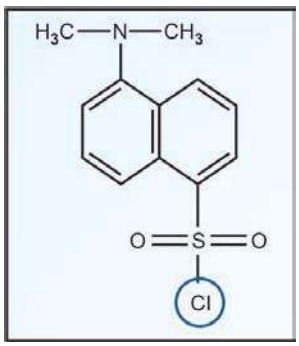

Fig. 4.10. Dansyl chloride

For example, if there are two different polypeptide chains in a protein, two different dansyl amino acids can be identified.

2. Determination of the amino acid composition by complete hydrolysis of the polypeptide chains, by chromatographic separation and quantitation.
3. Identification of N-terminal and C-terminal amino acids (Fig. 4.11).
4. Site specific hydrolysis of the polypeptide chain using specific enzymes to get a mixture of overlapping peptides.
5. Separation and purification of each of these peptides, and then analysing the amino acid sequence of each of the small peptides, and then deciphering the sequence of the whole protein.

1-B. End Group Analysis

i. The N-terminal amino acid has already been identified by treatment with dansyl chloride (Fig. 4.10). Originally Sanger used fluorodinitro benzene (FDNB, **Sanger's reagent**) for identification of N-terminal amino acid (Fig. 4.9).

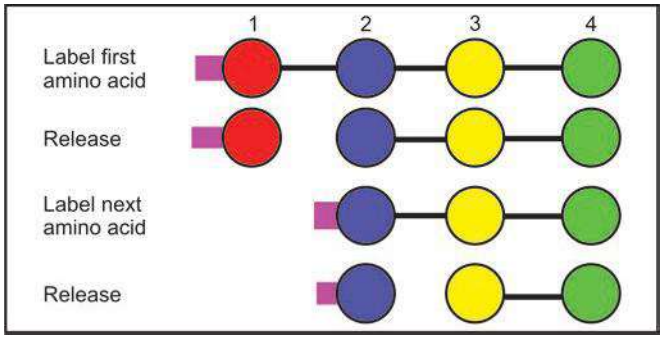

Fig. 4.11. Steps in Edman's degradation process. The numbers show the amino acid sequence. Amino terminal amino acid reacts with Edman's reagent; then it is hydrolyzed. This cycle repeats

ii. The **C-terminal amino acid** may be identified by Carboxypeptidase A and B.

These enzymes specifically hydrolyse and release the C-terminal amino acid from the polypeptide chain. Continued action of the enzyme would release amino acids sequentially from the C-terminal end. Carboxypeptidase A will not act if the C-terminal residue is Arginine, Proline or Lysine. Carboxypeptidase B will act only if the penultimate residue is proline.

1-C. Sequencing

The purified individual polypeptide chains are then sequenced using **Edman's degradation technique.** Edman's reagent is phenyl-isothiocyanate. It forms a covalent bond to the N-terminal amino acid of any peptide chain (Fig. 4.11). This can be identified. The Edman's reagent would then react with the second amino acid which now has the alpha amino group. The degradation is useful in sequencing first 10-30 amino acids.

1-D. Partial Hydrolysis

For very long chain proteins, the chain is broken into many small peptides of overlapping sequences. This is done by subjecting the polypeptide chain to hydrolysis by two or more different site specific enzymes. Each of these small peptides can be purified and subjected to Edman's degradation and sequenced.

Trypsin hydrolyses peptide bonds formed by alpha carboxyl group of Lysine and Arginine.

Chymotrypsin preferentially acts on peptide bonds formed by carboxyl group of amino acids Phe, Tyr, Trp, or Leu.

Cyanogen bromide (CNBr) attacks C-side of methionine residue and breaks the peptide bond.

Pehr Victor
Edman
1916-1977

Robert Merrifield
NP, 1984
1921-2006

Vernon M Ingram
1924-2006

Box 4.3. Significance of Iso-electric pH (pI)

1. The amino acid composition will determine the iso-electric pH (pI) of protein. The alpha amino group and carboxyl group are utilized for peptide bond formation, and hence are not ionisable. All other ionisable groups present in the protein will influence the pI of the protein.
2. At the iso-electric point, the number of anions and cations present on the protein molecule will be equal and the **net charge is zero**.
3. At the pI value, the proteins **will not migrate** in an electrical field. At the pI, solubility, buffering capacity and viscosity will be minimum; and **precipitation** will be maximum.
4. On the acidic side of pI, the proteins are cations and on alkaline side, they are anions in nature.
5. The pI of pepsin is 1.1; casein 4.6; human albumin 4.7; human insulin 5.4; human globulins 6.4; human hemoglobin 6.7; myoglobin 7; and lysozyme 11.
6. **Acidic dyes** such as eosin will dissociate into H^+ + dye^-, which will then attach with protein–$NH3^+$ (protein cations). **Basic dyes** such as hematoxylin and methylene blue are dissociated to OH^- + dye^+, which will then stain Protein-COO^- (anions). Thus the staining characteristic of a protein is determined by the pI of that protein.

Each peptide is then analyzed and the whole sequence of the polypeptide is determined as if fitting in the parts of a jigsaw puzzle. The position of **disulphide bonds** can be determined by cleaving the native protein sample to get fragments with intact S–S bonds. These fragments are then identified.

Finger Printing Method (Ingram's technique)

This method was developed by Vernon Ingram in 1956. It helps to easily identify any qualitative abnormalities in protein structure. Here the protein is digested into many small peptides by trypsin. The mixtures of peptides are separated by chromatography (peptide mapping). The position of the peptide containing the altered amino acid is found to be different when compared with the peptide map of the normal polypeptide, e.g. beta chain of hemoglobin in HbA and HbS.

Automated Sequencing

Using the Edman's degradation technique, sequencing can be completed within a few hours by automatic sequencers.

Study of Higher Levels of Protein Structure

The higher levels of protein structure may be studied by techniques using X-ray diffraction, ultraviolet light spectroscopy, optical rotatory dispersion, circular dichroism, nuclear magnetic resonance (NMR), etc.

NMR spectroscopy measures the absorbance of radio frequency of atomic nuclei. By studying the frequency at which a particular nucleus absorbs energy, we could get an idea of the functional group available in the molecule. Two dimensional NMR permits a three dimensional representation of the protein in solution. It also helps to study the alterations in conformation of a protein during binding with another ligand.

A beam of X-ray is diffracted by the electrons around each atom and the intensity of diffracted beam is detected by a photographic plate or collected by an electronic device. This **X-ray diffraction study** is possible only on crystallized proteins.

Nowadays, **DNA sequencing** is used to determine the amino acid sequence. In this method, at first, a rough sequencing of protein is done by Edman's method. Based on this knowledge, small length oligonucleotide primers are made. These are used to amplify the appropriate gene by polymerase chain (PCR) reaction (See Chapter 55) and correct DNA clone is obtained. The sequence of that part of DNA is done. Using the knowledge of the genetic code (Chapter 41), the sequence of the encoded protein is identified.

Chemical Synthesis of Peptides

Peptides are artificially synthesized for the following purposes:

1. To check whether the sequence analysis is correct or not.
2. The primary structure of a peptide is altered by one or two amino acids, so as to determine the biologically important area or the active center.
3. To get pure preparations for medical or diagnostic purpose. For example, HIV antibody in the blood of AIDS patients is detected by ELISA method . For this, pure antigen from HIV is to be coated in the test tubes. Preparation of enough quantity of antigen from the virus is tedious and hazardous. The best way is to synthesize the antigenic part of the protein. Commercially it is cheaper to synthesize small peptides, than isolating them from biological sources.

Emil Fischer in 1890 developed the basic mechanism to protect or activate reactive groups of amino acids. Robert Merrifield in 1961 introduced the **solid phase** peptide synthesis (Nobel prize, 1984). He simplified the process by adding the carboxy terminal end amino acid to insoluble polystyrene beads, so that washing and purification processes become rapid . In principle, the carboxyl group of the last amino acid is fixed on the resin; and other amino acids are added sequentially. Insulin was the first major protein chemically synthesized. In 1964, Panayotis Katsoyannis in USA and Helmut Zahn from Germany, indpendently synthesized insulin.

PHYSICAL PROPERTIES OF PROTEINS

1. Protein solutions exhibit colloidal properties and therefore scatter light and exert **osmotic pressure**. Osmotic pressure exerted by plasma proteins is clinically important (Chapter 28).
2. **Molecular weights** of some of the proteins are: Insulin (5,700); Hemoglobin (68,000); Albumin (69,000); Immunoglobulins (1,50,000); Rabbit Papilloma Virus Protein (4,70,00,000).
3. **Shape** of the proteins also vary. Thus, Insulin is globular, Albumin is oval in shape, while

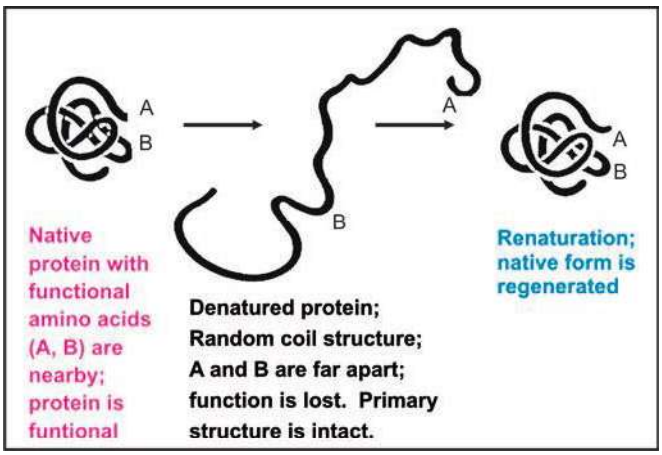

Fig. 4.12. Denaturation of protein

Fibrinogen molecule is elongated. Bigger and elongated molecules will increase the viscosity of the solution.

4. **Iso-electric pH** of amino acids has been described in Chapter 3. Since proteins are made of amino acids, the pI of all the constituent amino acids will influence the pI of the protein. Application of pI is shown in Box 4.3.

PRECIPITATION REACTIONS OF PROTEINS

Purification of enzymes and other proteins usually start with precipitating them from solution. The stability of proteins in solution will depend mainly on the charge and hydration. Polar groups of the proteins (-NH_2, COOH, OH groups) tend to attract water molecules around them to produce a shell of hydration. Any factor which **neutralises the charge or removes water of hydration** will therefore cause precipitation of proteins. The following procedures are used for protein precipitation:

1. Salting Out

When a neutral salt such as ammonium sulphate or sodium sulphate is added to protein solution, the shell of hydration is removed and the protein is precipitated. This is called salting out. As a general rule, higher the molecular weight of a protein, the salt required for precipitation is lesser. Thus **globulins are precipitated with half saturation of ammonium sulphate; but albumin will need full saturation** with ammonium sulphate for complete precipitation.

Fig. 4.13. Heat coagulation
On heating, liquid white portion of egg becomes solid white coagulum

2. Iso-electric Precipitation

Proteins are least soluble at their iso-electric pH. Some proteins are precipitated immediately when adjusted to their iso-electric pH. The best example is **Casein** which forms a flocculent

Box 4.4. Denaturation of Proteins

1. Mild heating, treating with urea, salicylate, X-ray, ultraviolet rays, high pressure, vigorous shaking and similar physico-chemical agents produce denaturation.
2. There will be non-specific alterations in secondary, tertiary and quaternary structures of protein molecules. **Primary structure is not altered** during denaturation (Fig. 4.12).
3. In general, during the process the solubility is decreased while precipitability of the protein is increased. It often causes **loss of biological activity**.
4. Native proteins are often resistant to proteolytic enzymes, but denatured proteins will have more exposed sites for enzyme action. Since cooking leads to denaturation of proteins, cooked foods are more easily digested.
5. Denatured proteins are sometimes re-natured when the physical agent is removed. Ribonuclease is a good example for such **reversible denaturation**. Immunoglobulin chains are dissociated when treated with urea. When the urea is removed by dialysis, the subunits are reassociated and biological activity of immunoglobulin is regained.
6. But many proteins undergo irreversible denaturation. For example, albumin once heated, cannot be renatured by cooling.

Box 4.5. Significance of Heat Coagulation

When **heated at iso-electric point**, some proteins will denature **irreversibly** to produce thick floating conglomerates called coagulum. This process is called heat coagulation. **Albumin** is easily coagulated, and globulins to a lesser extent. (See Fig. 4.13). This is the basis of '**Heat and Acetic Acid test**', very commonly employed to detect the presence of albumin in urine (See Chapter 27).

precipitate at pH 4.6 and redissolves in highly acidic or alkaline solutions. When milk is curdled, the casein forms the white curd, because lactic acid produced by the fermentation process lowers the pH to the iso-electric point of casein.

3. Precipitation by Organic Solvents

When an organic solvent is added to the protein solution, water molecules available for proteins are reduced, and precipitation occurs. Organic solvents reduce the dielectric constant of the medium which also favors protein precipitation. Hence, **alcohol** is a powerful protein precipitating agent. This may explain the disinfectant effect of alcohol.

4. Precipitation by Heavy Metal Ions

In alkaline medium, proteins have net negative charge, or are anions. To such a solution, if salts of heavy metals are added, positively charged metal ions can complex with protein molecules and metal proteinates are precipitated. Salts of **Copper, Zinc, Lead, Cadmium and Mercury** are toxic, because they tend to precipitate normal proteins of the gastro-intestinal wall. Based on this principle, **raw egg** is sometimes used as an **antidote** for mercury poisoning.

5. Precipitation by Alkaloidal Reagents

Tungstic acid, Phosphotungstic acid, **Trichloro acetic acid,** Picric acid, Sulphosalicylic acid and **Tannic acid** are powerful protein precipitating agents. These acids lower the pH of medium, when proteins carry net positive charges. These protein cations are complexed with negatively charged ions to form protein-tungstate, protein-picrate, etc. and thick flocculent precipitate is formed. In clinical laboratory phospho-tungstic or trichloro acetic acid are usually used for precipitating proteins. **Tanning** in leather processing is based on the protein precipitating effect of tannic acid. Under certain conditions, proteins undergo **denaturation**, which is a mild form of precipitation reaction (Box 4.4). **Heat coagulation** is an irreversible precipitation process (Box 4.5 and Fig. 4.3).

CLASSIFICATION OF PROTEINS

It is almost impossible to correctly classify all proteins. The following classifications are given only to introduce a broader idea to the students.

Table 4.1. Examples of conjugated proteins

Conjugated Protein	Protein part	Prosthetic group
Hemoglobin	Globin	Heme
Nucleoprotein	Histones	DNA
Rhodopsin	Opsin	11-cis-retinal
Succinate dehydrogenase	Protein	Riboflavin as FAD
Ferritin	Apoferritin	Iron
Ceruloplasmin	Apocerulo-plasmin	Copper

A. Classification based on functions

1. Catalytic proteins, e.g. enzymes
2. Structural proteins, e.g. collagen, elastin
3. Contractile proteins, e.g. myosin, actin.
4. Transport proteins, e.g. hemoglobin, myoglobin, albumin, transferrin
5. Regulatory proteins or hormones, e.g. ACTH, insulin, growth hormone
6. Genetic proteins, e.g. histones
7. Protective proteins, e.g. immunoglobulins, interferons, clotting factors.

B. Classification based on Composition and Solubility

B-1. Simple Proteins

According to definition, they contain only amino acids.

i. **Albumins:** They are **soluble in water** and coagulated by heat. Human serum albumin has a molecular weight of 69,000. Other examples are lactalbumin of milk and egg albumin.

ii. **Globulins:** These are insoluble in pure water, but soluble in **dilute salt solutions**. They are also coagulated by heat. Examples are egg globulin, serum globulins, legumin of peas.

iii. **Protamines:** These are soluble in water, **dilute acids and alkalies.** They are not coagulated by heating. They contain large number of arginine and lysine residues, and so are strongly basic. Hence, they can combine with other acidic proteins. Protamine zinc insulinate is a common commercial preparation of insulin.

iv. **Prolamines:** They are soluble in 70-80% **alcohol**, but insoluble in pure water. They are rich in proline but lack in lysine. Examples are zein from corn, gliadin of wheat, hordein of barley.

v. **Lectins:** Lectins are precipitated by 30–60% **ammonium sulphate**. They are proteins

having high affinity to sugar groups. Lectin from *Dolichos biflorus* will agglutinate human blood group A1 RBCs. Phytohemagglutinin (PHA), a lectin from *Phaseolus vulgaris* (red kidney bean) agglutinates all RBCs and WBCs. Concanavalin-A (ConA) from legumes will specifically attach to mannose and glucose.

vi. **Scleroproteins:** They are insoluble in water, salt solutions and organic solvents and soluble only in **hot strong acids**. They form supporting tissues. Examples are collagen of bone, cartilage and tendon; keratin of hair, horn, nail and hoof.

B-2. Conjugated Proteins

They are combinations of protein with a non-protein part, called **prosthetic group** (Table 4.1). Conjugated proteins may be classified as follows:

i. **Glycoproteins:** These are proteins combined with carbohydrates. Hydroxyl groups of serine or threonine and amide groups of asparagine and glutamine form linkages with carbohydrate residues. **Blood group antigens** and many serum proteins are glycoproteins. When the carbohydrate content is more than 10% of the molecule, the viscosity is correspondingly increased; they are sometimes known as **mucoproteins** or proteoglycans.

ii. **Lipoproteins:** These are proteins loosely combined with lipid components. They occur in blood and on cell membranes. Serum lipoproteins are described in Chapter 12.

iii. **Nucleoproteins:** These are proteins attached to nucleic acids, e.g. Histones. The DNA carries negative charges, which combines with positively charged proteins.

iv. **Chromoproteins:** These are proteins with colored prosthetic groups. Hemoglobin (Heme, red); Flavoproteins (Riboflavin, yellow), Visual purple (Vitamin A, purple) are some examples of chromoproteins.

v. **Phosphoproteins:** These contain phosphorus. **Casein** of milk and **vitellin** of egg yolk are examples. The phosphoric acid is esterified to the hydroxyl groups of serine and threonine residues of proteins.

vi. **Metalloproteins:** They contain metal ions. Examples are Hemoglobin (Iron), Cytochrome (Iron), Tyrosinase (Copper) and Carbonic anhydrase (Zinc).

B-3. Derived Proteins

They are degradation products of native proteins. Progressive hydrolysis of protein results in smaller and smaller chains: Protein → peptones → peptides → amino acids.

C. Classification Based on the Shape

C-1. Globular Proteins

They are spherical or oval in shape. They are easily soluble. Examples are albumins, globulins and protamines.

C-2. Fibrous Proteins

The molecules are elongated or needle shaped. Their solubility is minimum. They resist digestion. Collagen, elastin and keratins are examples.

D. Classification Based on Nutritional Value

D-1. Nutritionally Rich Proteins

They are also called as **complete proteins or first class proteins**. They contain all the essential amino acids in the required proportion. On supplying these proteins in the diet, children will grow satisfactorily. A good example is **casein** of milk.

D-2. Incomplete Proteins

They **lack one essential amino acid.** They cannot promote body growth in children; but may be able to sustain the body weight in adults. Proteins from **pulses are deficient in methionine**, while proteins of **cereals lack in lysine**. If both of them are combined in the diet, adequate growth may be obtained. (See mutual supplementation, Chapter 36).

D-3. Poor Proteins

They **lack in many essential amino acids** and a diet based on these proteins will not even sustain the original body weight. Zein from corn lacks tryptophan and lysine.

Biologically Important Peptides

When 10 or less number of amino acids are joined together, it is called an oligopeptide. Some of them are biologically active. A few examples are given below:

i. **Thyrotropin releasing hormone** (TRH) is a tripeptide with the sequence of Glu-His-Pro; but the Glu is modified to form pyroglutamic acid.

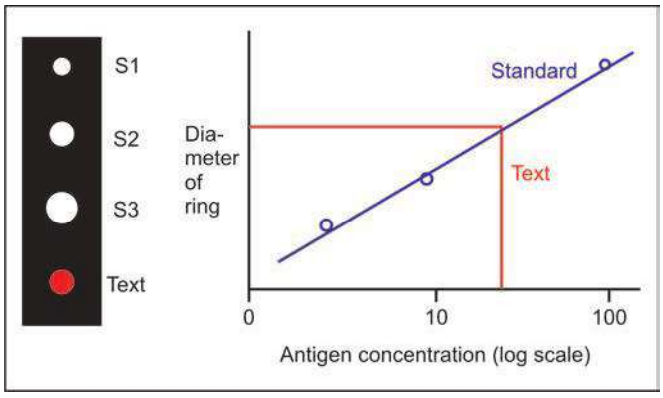

Fig. 4.14. Radial immuno diffusion

ii. **Glutathione** is a tripeptide. It is gamma glutamyl cysteinyl glycine (See Fig.15.19). It is involved in erythrocyte membrane integrity and is important in keeping enzymes in active state.

iii. **Oxytocin and Vasopressin** (ADH) are nanopeptides; with 9 amino acids. They are secreted by posterior pituitary.

iv. **Angiotensin I** has 10 amino acids and Angiotensin II has 8 amino acids. They cause hypertension (Chapter 30)

v. **Gramicidin S**, an antibiotic produced by *Bacillus brevis*, contains 10 amino acids. It is circular and contains D-phenyl alanine (usual proteins contain only L-amino acids).

Polypeptide hormones (more than 10 amino acids) are described in Chapters 24 and 45.

QUANTITATIVE ESTIMATION

1. Kjeldahl's Procedure

i. The protein sample is digested by boiling (360°C) with concentrated sulphuric acid in presence of copper sulphate and sodium sulphate as catalysts.

ii. The nitrogen present in the protein is reduced to ammonia which is absorbed by acid medium to become ammonium sulphate. After cooling, the digest is made alkaline by adding excess alkali. Now ammonia (NH_3) is liberated, which is absorbed by a known quantity of standard acid kept in a vessel. The excess acid present in the vessel is back-titrated with a standard base, from which the liberated ammonia can be calculated.

iii. Then the quantity of nitrogen originally present in the protein is assessed. Since **proteins,** on an average contain **16% nitrogen**, the weight of nitrogen x 100/16, or nitrogen x 6.25 will give the value of proteins present in the original sample.

iv. **Advantage:** This is the most accurate and precise method. It is generally used for standardising a particular protein; that protein is then used for calibrating other proteins employing other easier methods.

v. **Disadvantage:** It takes many days to get the result, and is unsuitable for routine clinical work.


Johan Kjeldahl
1849-1900


OH Lowry


Richard A Zsigmondy
NP1925
1865-1929

2. Biuret Method

i. Cupric ions chelate with peptide bonds of proteins in alkaline medium to produce a pink or violet color. The intensity of the color is proportional to the number of peptide bonds. The color is then compared with a standard protein solution treated with the biuret reagent, and estimated colorimetrically. The principle of colorimetry is discussed in Chapter 54.

ii. **Advantage:** The biuret method is a simple one step process, and is the most widely used method for plasma protein estimations.

iii. **Disadvantage:** The sensitivity of the method is less and is unsuitable for estimation of proteins in milligram or microgram quantities.

3. Lowry's Method

i. This is based on the reduction of Folin-Ciocalteau phenol reagent (phosphomolybdic acid and phosphotungstic acid) by the tyrosine and tryptophan residues of protein. A blue color is developed which is compared with that produced by a known standard.

ii. **Advantage:** This method is very sensitive and protein content in microgram range can be measured. If the tyrosine and tryptophan content of the proteins of test and standard vary widely, then the accuracy is lost; this is a minor disadvantage of this method.

OH Lowry published the paper on protein estimation in 1951. This is the most cited article in the scientific literature.

4. Spectrophotometric Estimation

i. Proteins will absorb ultraviolet light at **280 nm**. This is due to the tyrosine and tryptophan residues in the protein. Quantitation is done by comparing the absorbance of the test solution with a known standard.

ii. **Advantage and disadvantage:** The method is accurate, simple and highly sensitive upto microgram quantities. Since color reaction using other chemicals is not employed, the protein is not wasted in this method. However, the instrument is costly. Please see discussion in Chapter 54.

5. Radial Immuno Diffusion (Mancini's technique)

Please note that the name is "radial" and not "radio". There is no radiation applied in this process. As the precipitation arc is moving radially outward from the point of application, the name "radial" is given. Here specific antiserum is incorporated in the liquid agar, and then allowed to solidify on a glass plate. Then small wells (1 mm dia) are cut in the agar, and antigen (protein solution, patient's sera, etc.) is added in the well. The plate is incubated at 4°C for 1 to 3 days. The antigen molecules diffuse radially around the wells and react with the specific antibody molecules present all over the agar. A white ring of precipitate is seen, where equimolecular concentration (1:1 ratio) of antigen and antibody is attained. The **diameter of the precipitation ring will be proportional to the log of antigen concentration**. If known standards of different concentration is included along with the test, a standard graph is plotted, from which quantity of test substance can be obtained (Fig. 4.14).

Advantage: The procedure is simple and sensitive enough to quantitate mg or microgram quantities of proteins. Since the method is based on antibodies, the test is exquisitely specific. Serum levels of immunoglobulins, complement components, etc. are routinely assayed by this method.

6. Nephelometry

It is based on the measurement of scattering of light by colloids, originally studied by Richard Zsigmondy (Nobel prize, 1925). Scattering of light by antigen-antibody complexes was used as a quantitation method by Libby in 1938. Nephelometry is defined as the **detection of light scattered by turbid particles in solution**. If albumin is to be estimated, specific antibody against albumin is added. The resultant antigen-antibody complex will form turbidity of the solution. A beam of light (preferably laser beam) is passed through the solution. The particles in the solution will scatter light. The light turning at 30° to 90° angle (generally 60°)is collected and passed into a detector system. The emergent scattered light will be proportional to the turbidity of the solution, which in turn will be proportional to the antigen.

Advantage and disadvantage: This is a very rapid method and suitable for automated programs. Microgram quantities can be accurately estimated. The instrument and reagents are costly and needs careful standardization.

7. Turbidimetry

Proteins in biological fluids like urine and CSF can be estimated by adding protein precipitating agents (sulphosalicylic acid or specific antibody); the turbidity thus produced is measured. The method is simple, but less accurate. Nephelometry and turbidimetry are based on the same principle of scattering of light by colloidal particles. In nephelometry, emergent light scattered at 60° angle is observed; while in turbidimetry, light emerging at 180° angle is detected. Turbidimetry is comparatively cheaper. Newer techniques are based on immunoturbidimetry.

8. RIA and ELISA Tests

If proteins of nanogram and picogram quantities are to be estimated, radioimmunoassay (RIA) or enzyme linked immuno sorbent assay (ELISA) techniques are to be employed. These are described in detail in Chapter 54.

Proteomics

Proteomics is the study of the entire galaxy of proteins produced by a cell under different conditions. At a particular time, a gene is "on" in a particular cell; but it will be "off" in another cell. Expression of proteins during growth and development will be different from the resting cell. Proteins produced by a gastrointestinal cell and a neuronal cell will be entirely different. Many proteins are getting post-translational modification, that too, at different levels at various organs. But genes are the same in all cells at all times. Therefore, study of genes (genomics) will give only a partial picture of what is going on in nature. Human body contains hundreds of different cells, which express thousands of proteins, at different times and under the influence of different stimuli. Proteomics attempt to study all these multifaceted picture in toto.