BIOMOLECULES
SEM-5, CC-12
PART-6, PPT-26

Part-6: Peptides-III, Proteins

CONTENTS

1. *Partial Hydrolysis of Peptides*
2. *Site Specific Cleavage of Peptides*
3. *Use of CNBr*
4. *Protein Subunits, Cyclic Structures, and Disulfide Bonds*
5. *Summary of Primary Structure Determination*
6. *General Nature of Proteins*
7. *Classification of Proteins*

Dr. Kalyan Kumar Mandal
Associate Professor
St. Paul's C. M. College
Kolkata

BIOMOLECULES (PART-6, PPT-26)

Peptides-III, Proteins

Partial Hydrolysis of Peptides

The 'overlapping procedure' is generally used for the elucidation of the amino acid sequence, therefore, methods are necessary to bring about partial hydrolysis of peptides (or proteins). Different hydrolytic paths are possible because of the different susceptibilities of the various peptide bonds (-CONH-) to attack by hydrolytic reagents.

Application of the overlapping procedure to sequence peptides:

Let us consider a hexapeptide whose amino acids have been shown to be (A, B, C, D, E, F) and whose *N*-terminal amino acid has been shown to be C. The hexapeptide may, therefore, be written as C (A, B, D, E, F). Now suppose that on partial hydrolysis the small peptides obtained were (as shown by amino acid analysis):

C-A, (B, E), (B, D), C. (A. E), (B, D, F) and (B, D, E, F).

Since the sequenced dipeptide is C-A, C.(A. E) must be C-A-E. Hence, (B, E) is E-B, and now, a tetrapeptide can be constructed as C-A-E-B. Therefore, (B, D) is B-D, and so the pentapeptide is C-A-E-B-D. Finally, since only F is missing, the hexapeptide must be C-A-E-B-D-F. This explains fragment (B, D, E F). These results may be tabulated as shown in the Figure 1. Thus *N*-T-AA analysis and partial hydrolysis of the peptide gives useful information regarding the sequence of the chosen hexapeptide.

N-T-AA		*C*-T-AA			
C	A				
C	A	E			
		E	B		
			B	D	
			B	D	F
		E	B	D	F
C	A	E	B	D	F

Figure 1: Overlapping of fragments to form a hexapeptide

As the peptide becomes more complex, the amino acids usually occur more than once. This will increase the difficulty in elucidating the amino acid sequence. For example, both hexapeptides C-A-E-B-D-E and C-A-E-E-B-D might possibly give rise to fragments C-A, (B, E), (B, D), C. (A. E), (B, D, E) and (B, D, E, E). Since two (B, D, E) are possible, it will necessary to use end-group analysis to decide which is correct: overlapping is of no help here.

Suppose it is found it to be E. (B, D). This could still have been derived from either hexapeptide. If, however, B (D, E) is found, then the hexapeptide cannot be C-A-E-E-B-D. It

also follows that (B, E) is E-B and so B (D, E) is B-D-E. The order in (B, E) would be confirmed by end group analysis, and this would also show (B, D, E, E) is E (B, D, E). Hence, the hexapeptide is C-A-E-B-D-E. Figure 2 illustrates the joining of the fragments.

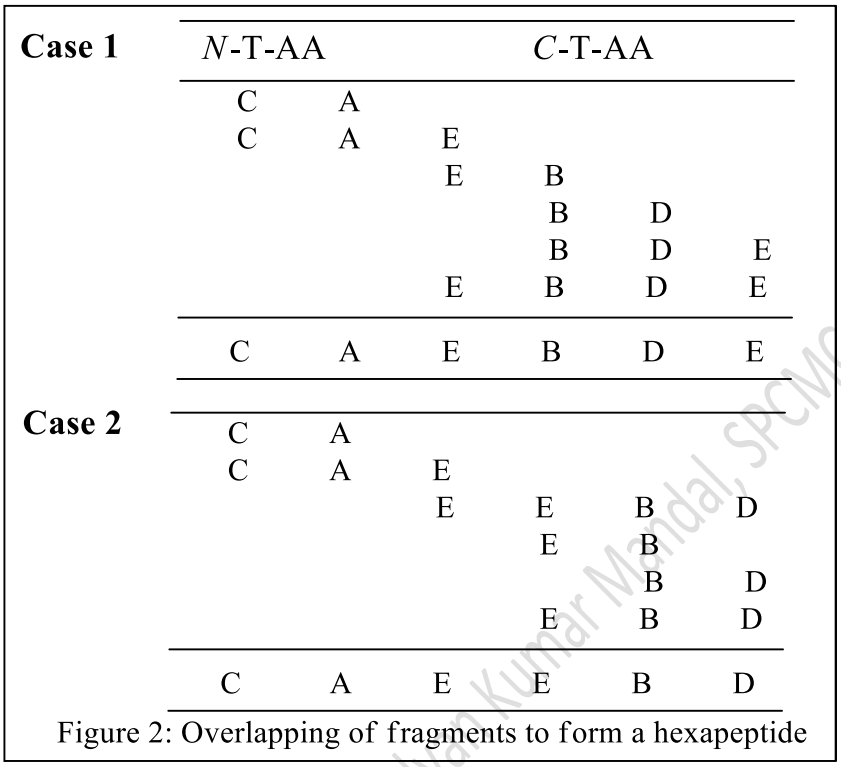

Figure 2: Overlapping of fragments to form a hexapeptide

From this example it can be seen that it will usually be necessary to carry out end group analysis on each fragment-peptide obtained. Partial hydrolysis with acids is generally unsatisfactory, since bond-breaking tends to occur randomly and also results in a large number of small peptides which may be very difficult to separate. Even so, this approach is often successful for relatively small peptides. As an example, let us consider a pentapeptide which gives Gly-Ala, Leu-Phe, Leu-Leu and Ala-Leu upon partial hydrolysis and *N*-phenylthiohydantoin of glycine upon Edman degradation.

Analysis: Edman degradation involves the determination of *N*-T-AA residue of a peptide. Hence, glycine is the *N*-T-AA of the required pentapeptide. By matching the overlapping regions of the fragments obtained on partial hydrolysis, the sequence of the pentapeptide can be established.

Gly-Ala + Ala-Leu + Leu-Leu + Leu-Phe → Gly-Ala-Leu-Leu-Phe

Enzymic hydrolysis is extremely useful because each enzyme hydrolyses only certain types of peptide bond. Trypsin splits peptide bonds in which the carbonyl group is part of a lysine or an arginine residue. Chymotrypsin splits peptide bonds in which the carbonyl group is part of a phenylalanine, tyrosine, or tryptophan residue. Hence, the separate use of these enzymes will result in splitting of the peptide (or protein) in different ways to give relatively large fragments.

Other enzymes are also available, e.g., pepsin (NH group of Leu, Asp, Glu, etc.), papain (CO group of Gly, Arg, Lys, etc.). Since the specificity of the various enzymes differs considerably, the less specific enzymes are generally more satisfactory when used with relatively small peptides. Large peptides require the use of the specific enzymes. Otherwise, a large number of fragment-peptides will be obtained.

Complete Sequence Analysis

Sequential analysis using the Edman degradation or other methods becomes impractical with large proteins and polypeptides. However, there are techniques to cleave peptides into fragments that are of manageable size. Partial hydrolysis with dilute acid, for example, generates a family of peptides cleaved in random locations and with varying lengths. Sequencing these cleavage peptides and looking for points of overlap allows the sequence of the entire peptide to be pieced together.

Let us consider a pentapeptide known to contain valine (two residues), leucine (one residue), histidine (one residue), and phenylalanine (one residue), as determined by hydrolysis and automatic amino acid analysis. With this information the "molecular formula" of the required pentapeptide can be written in the following way. Abbreviated form of the given amino acids is written using commas to indicate that the sequence is unknown:

(2V, L, H, F)

Then, let us assume that by using FDNB and carboxypeptidase it is ascertained that valine and leucine are the *N*- and *C*-terminal amino acid residues, respectively. So the partly sequenced structure for the pentapeptide would be the following:

V (V, H, F) L

But the sequence of the three nonterminal amino acids is still unknown. The pentapeptide is then subjected to partial acid hydrolysis and obtain the following dipeptides in addition to the individual amino acids and larger pieces, i.e., tripeptides and tetrapeptides:

VH + HV + VF + FL

The points of overlap of the dipeptides (i.e., H, V, and F) indicate that the original pentapeptide must have been V-H-V-F-L.

Site-Specific Cleavage

Site-specific cleavage of peptide bonds is possible with enzymes and specialized reagents as well, and these methods are now more widely used than partial hydrolysis. For example, the enzyme trypsin preferentially catalyzes hydrolysis of peptide bonds on the *C*-terminal side of arginine and lysine. Chemical cleavage of peptides (or proteins) at specific sites can be done with cyanogen bromide (CNBr) in aqueous formic acid at room temperature, which cleaves peptide bonds on the *C*-terminal side of methionine residues.

Using these site-selective cleavage methods on separate samples of a given polypeptide results in fragments that have overlapping sequences. After sequencing the individual fragments, aligning them with each other on the basis of their overlapping sections results in a sequence for the intact protein.

Cleavage at Methionine with Cyanogen Bromide

When a peptide reacts with cyanogen bromide (NC-Br or BrCN) in aqueous HCl, a peptide bond is cleaved specifically at the carboxy side of each methionine residue. Only peptides in which the CO group is that of a methionine residue are split, the products being a homoserine lactone and the 'rest' of the peptide.

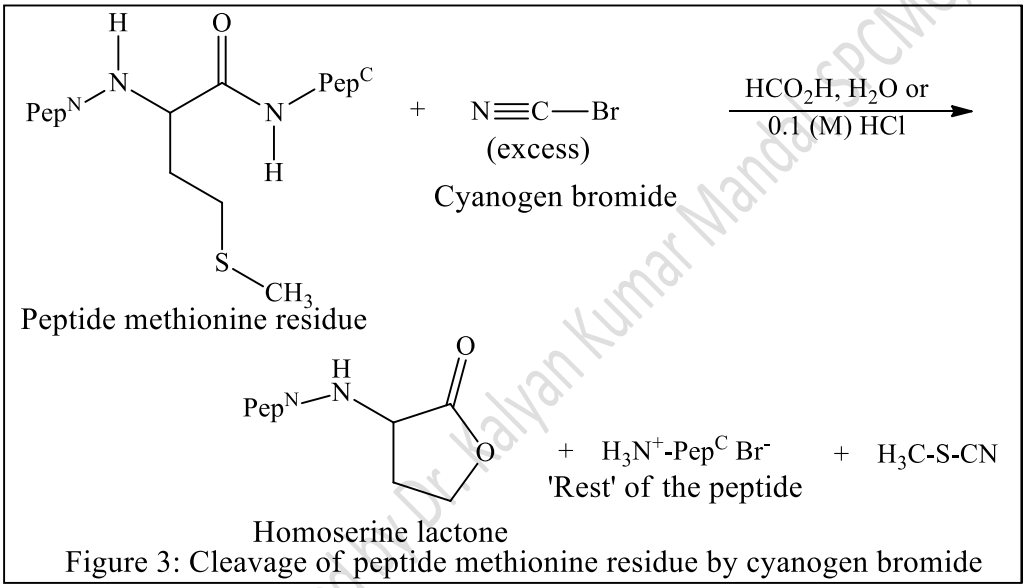

Figure 3: Cleavage of peptide methionine residue by cyanogen bromide

Mechanism of cleavage with Cyanogen Bromide

The sulfur in the methionine side chain acts as a nucleophile, displacing bromide from NC-Br to give sulfonium ion. The sulfonium ion, with its electron withdrawing cyanide, is an excellent leaving group and is displaced by the oxygen of the neighbouring amide bond to form a five membered ring. The last mechanistic step in the cleavage is hydrolysis of the ion formed.

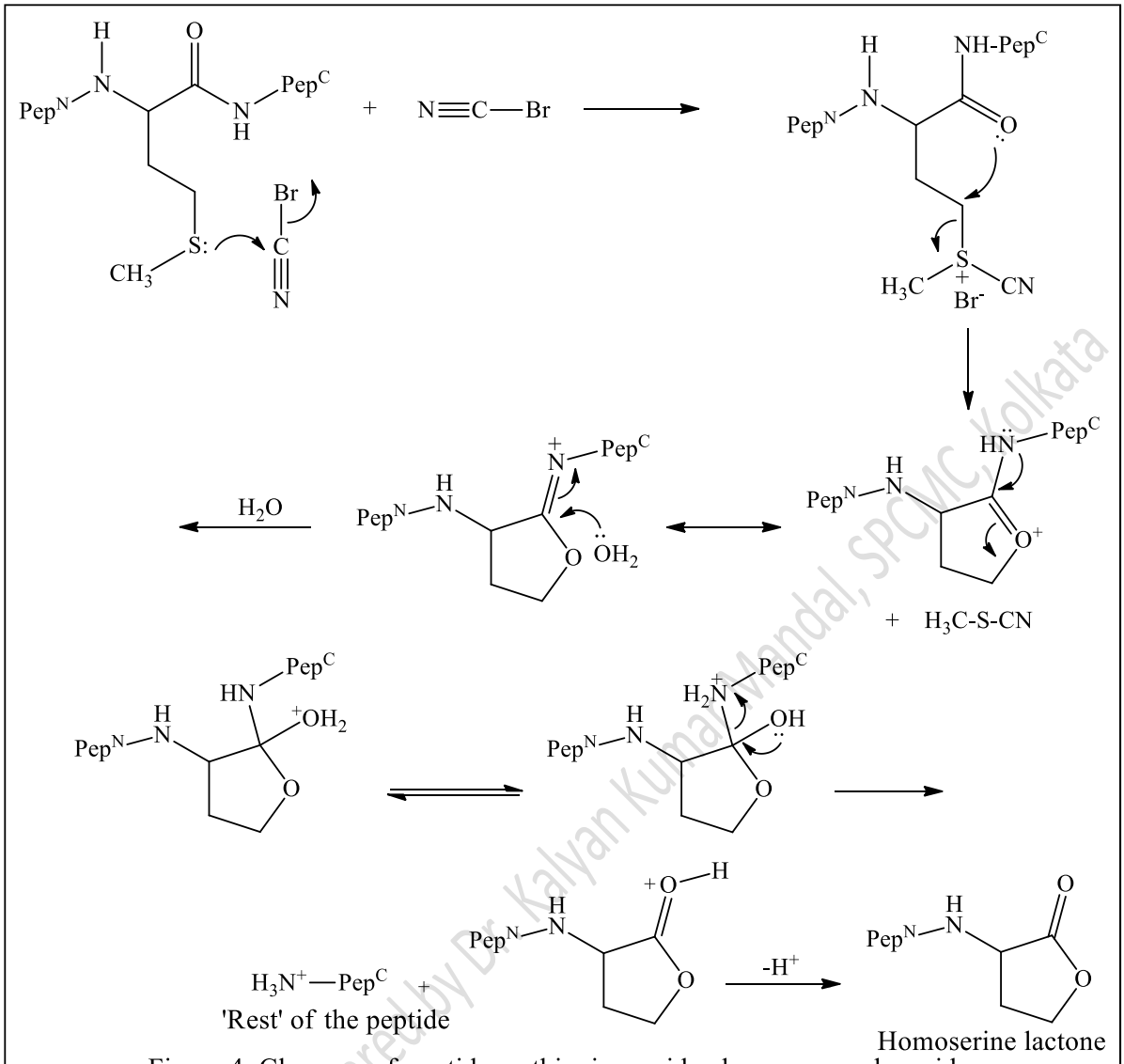

Figure 4: Cleavage of peptide methionine residue by cyanogen bromide

Protein Subunits, Cyclic Structures, and Disulfide Bonds

If the protein consists of a single linear peptide chain, then the methods described for the determination of amino acid sequence of peptide (or protein) may be readily applied. Thus, end group analysis will show the presence of a single *N*-T-AA or of a single *C*-T-AA (or one of each if both end group analyses are performed). In many cases, however, the results indicate the presence of *n* end groups (*N* or *C*). This usually means that the protein is composed of *n* peptide chains. If there are no end group (i.e., no free α-NH_2 or α-CO_2H), this is strong evidence that the protein has a cyclic structure. Other evidence that may be used to show the presence of a cyclic peptide is the neutrality towards electrophoresis and the failure to give the ninhydrin test.

If when the subunits in a protein are not cross-linked by covalent bonds, they are readily dissociated into the individual units, e.g., dissolving in a urea solution. These units are then separated, purified, and examined by different chemical methods. On the other hand, if

cystine is present the subunits may be held together by the disulfide bond and /or a single chain may contain an intramolecular disulfide ring. These bonds are usually split before the primary structure is determined.

One common method is oxidation of the molecule with performic acid to give cysteic acid (the sulfonic acid). Alternatively, the disulfide bond may be reduced to thiol by means of, e.g., sodium borohydride, and the products treated with iodoacetic acid as shown in Figure 5.

Determination of Disulfide Bonds

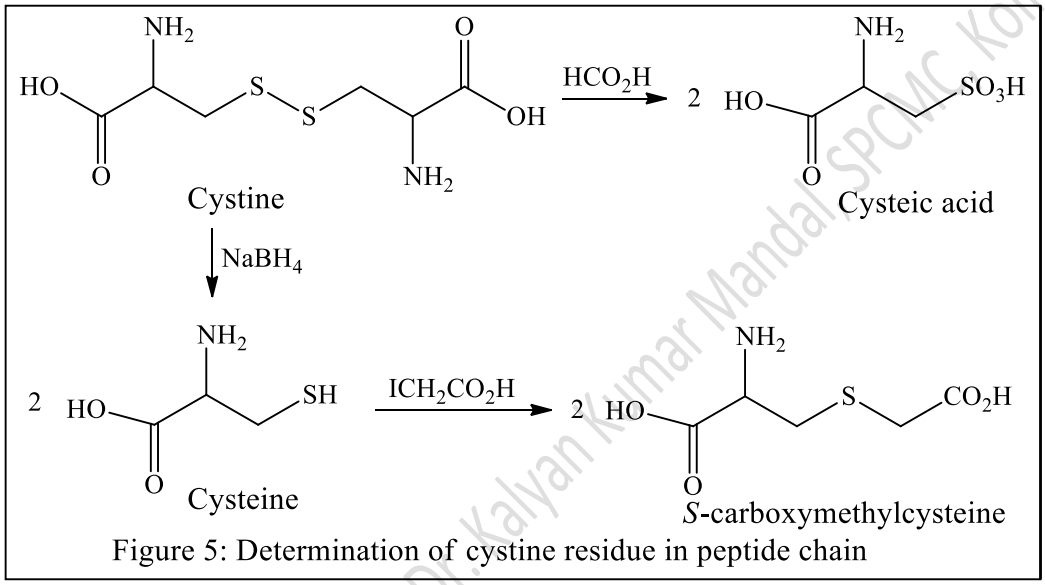

Figure 5: Determination of cystine residue in peptide chain

The primary structures of these products are then determined and the positions of the disulfide linkages are deduced from the positions (in the sequence) of the cysteic acid residues or of the carboxymethylcysteine residues. It should also be noted that performic acid oxidizes methionine to the corresponding sulfone and also destroys tryptophan.

Summary of Primary Structure Determination

The strategy adopted to determine the primary structure of peptides and proteins is the following:

1. The peptide or protein must be isolated in a pure state.
2. It is first necessary to ascertain whether the protein consists of a single peptide chain or whether it is composed of a number of subunits. If the later, then the subunits are separated and each chain is examined separately.
3. The protein is completely hydrolyzed into its constituent amino acids and their nature and amounts are determined.

4. The minimum molecular weight is determined from the amino acid parentage composition and the molecular weight is also determined by a physical method.
5. The end group analysis is carried out to determine the nature of the *N*- and *C*-terminal groups.
6. The amino acid sequence may be determined by the Edman automated *N*-terminal method where possible. Alternatively, if the protein is relatively small, it may be subjected to controlled hydrolysis to a number of simple peptides. These are isolated and purified, and end group analysis is then applied to these and the amino acid sequence in the protein may be deduced by the overlapping procedure.

   When the protein is relatively large, partial hydrolysis is effected in at least different ways. The amino acid sequence in each purified fragment is determined and the amino acid sequence in the protein is deduced by the overlapping procedure.

General Nature of Proteins

The name protein was introduced by Mulder (1839), who derived it from the Greek word *proteios* (meaning *first*). Proteins are nitrogenous substances which occur in the protoplasm of all animal and plant cells. Proteins can be broken down into smaller and smaller fragments until the final products are amino acids. The sequence may be written as

$$\text{Protein} \rightarrow \text{polypeptides} \rightarrow \text{peptides} \rightarrow \text{amino acids}$$

There is no sharp dividing line between peptides, polypeptides and proteins. One arbitrary convention designates proteins as those molecules with a molecular weight above ~ 10000 and peptides (polypeptides) as those molecules with a molecular weight below ~ 10000.

Proteins are amphoteric, their behaviours as an anion or a cation depending on the pH of the solution. At some definite pH, characteristic of each protein, the positive and negative charges are exactly balanced, i.e., there is no net charge on the protein molecule, and the molecules will not migrate in an electric field. In this condition the protein is said to be at its isoelectric point, and at this pH the protein has its least solubility, i.e., it is most readily precipitated.

The osmotic pressure and viscosity of the protein solution are also a minimum at the isoelectric point. The amphoteric nature of proteins is due to the presence of a large number of free acidic and basic groups arising from the amino acid units in the molecule. These groups can be titrated with alkali or acid, and by this means it has been possible to identify acidic and basic groups belonging to the various amino acid units.

All proteins are optically active, and may be coagulated and precipitated from aqueous solution by heat, the addition of acids, alkalis, salts, organic solvents miscible with water, etc. Proteins in this precipitated state are said to be *denatured*, and the process of reaching this state, *denaturation*, occurs most readily near the isoelectric point. Denaturation is now believed to be the result of changes in *conformation* or *unfolding* of the protein molecule.

Associated with denaturation are changes in optical rotation and (usually) the loss of biological activity, e.g., enzymes (all are proteins) become inactive when denatured.

Denaturation is generally irreversible, but many examples are now known where the process has been reversed. This reversal of denaturation has been called *renaturation* or *refolding*. When denaturation is effected by heat, renaturation does not usually result on rapid cooling. If, however, cooling is carried out very slowly, renaturation often occurs. In these circumstances the process is renaturation has been referred to as *annealing*. Proteins exhibit a variety of colour reactions. Some of them are given below.

1. **Biuret reaction:** Addition of a very dilute solution of copper sulphate to an alkaline solution of a protein produces a red colour. This reaction is due to the presence of the grouping -CO-NH-CHR-CO-NH-. At least two peptide linkages (-CONH-) must be present. Therefore, the dipeptides do not give the test.

2. **Xanthoproteic reaction:** Proteins usually produce a yellow colour when warmed with concentrated nitric acid, and the colour becomes orange when the solution is made alkaline. This reaction is due to the nitration of the benzene ring in phenylalanine, tyrosine and tryptophan.

3. **Millon's reaction:** Millon's reagent (mercuric nitrate in nitric acid containing a traces of nitrous acid) usually produces on addition to a protein solution a white precipitate which turns red on heating. This reaction is characteristic of phenols, and so is given by proteins containing tyrosine (this is the only phenolic amino acid that occurs in proteins).

4. **Ninhydrin test:** Proteins (and peptides) give the test, but the colours are different from that of the amino acids. Whilst ninhydrin detected amino acids with high sensitivity it was much less sensitive towards a variety of common proteins.

This is because most proteins have only one free amino group at the *N*-terminal tail and although dibasic amino acids, i.e., lysine and arginine, occur in most proteins, steric hindrance limits the ability of ninhydrin to react with them. The molecular weights of proteins have been determined by means of ultracentrifugal sedimentation, osmotic pressure measurements, X-ray diffraction, light scattering effects, molecular sieves (gel filtration), and by chemical analysis.

Chemical methods are based on the estimation of a particular amino acid. The values of molecular weights recorded for proteins vary considerably, ranging from about 5000 to many millions.

Classification of Proteins

Several arbitrary classifications of the proteins are in use. One method divides the proteins into two groups, *fibrous proteins*, which are insoluble in common solvents, but are soluble in concentrated acids and alkalis, and *globular proteins*, which are soluble in water and in dilute acids, alkalis and salts.

A more common method of classification is the division of proteins into the three main groups: simple, conjugated and derived proteins. Each group is subdivided into a number of classes designated by general names. Each class contains sub-classes of proteins of similar but not identical physical and chemical properties.

A. **Simple Proteins:** These give only amino acids or their derivatives on hydrolysis. Common examples include:

a) ***Albumins:*** These are soluble in water (and in acids and in alkalis), and are coagulated by heat. They are precipitated by saturating their solutions with ammonium sulfate.

b) ***Globulins:*** These are insoluble in water, but are soluble in dilute solutions of strong inorganic acids and alkalis. They are precipitated by half-saturating their solutions with ammonium sulfate, and they are coagulated by heat.

c) ***Collagens*** (in skin, tendons and bones): These form gelatin (a water-soluble protein) when boiled with water. Collagens are attacked by pepsin or trypsin.

B. **Conjugated Proteins:** These are proteins which contain a non-protein group (i.e., a compound not containing amino acid residues) attached to the protein part. The non-protein group is known as the prosthetic group, and it may be separated from the protein part by careful hydrolysis. Therefore, the conjugated protein is named according to the nature of the prosthetic group.

a) ***Nucleoproteins:*** The prosthetic group is a nucleic acid.

b) ***Chromoproteins:*** These are characterized by the presence of a coloured prosthetic group. Examples: chlorophyll and haemoglobin.

c) ***Glycoproteins:*** In these the prosthetic group contains a carbohydrate or a derivative of the carbohydrates. They are also known as *mucoproteins*.

d) ***Lipoproteins:*** In these the prosthetic group is lecithin, kephalin.

e) ***Phosphoproteins:*** These are conjugated proteins in which the prosthetic group contains phosphoric acid in some form other than in the nucleic acids or in the lipoproteins.

f) ***Metalloproteins:*** These all contain a metal which is an integral part of the structure. Many metals occur, e.g., iron, magnesium, copper, manganese, Examples are haemoglobin, chlorophyll which may also be classed as chromoproteins.

g) **Derived Proteins:** These are degradation products obtained by the action of acid, alkalis or enzymes on proteins.

Protein → *Denatured proteins*; insoluble proteins formed by the action of heat, etc., on proteins.

↓

Primary proteoses (*metaproteins*); insoluble in water or dilute salt, but are soluble in acids or alkalis. They are precipitated by half-saturation with ammonium sulfate.

↓

Secondary proteoses; soluble in water, not coagulated by heat, and are precipitated by saturation with ammonium sulfate.

↓

Protones

↓

Polypeptides

↓

Simple peptides

} These are soluble in water, not coagulated by heat, and are not precipitated by saturation with ammonium sulfate.

↓

Amino acids