BIOMOLECULES
SEM-5, CC-12
PART-5, PPT-25

Part-5: Peptides-II, Reactions, End group Analysis

CONTENTS

1. *The Primary Structure of Peptide/Determination of Peptide sequence*
2. *Hydrolysis of Peptide to Amino Acids*
3. *Separation, Identification and Quantification of Amino Acids*
4. *End-Group Analysis*

A. Determination of N-terminal Amino Acids

a) Edman Degradation

b) Sanger N-Terminal Analysis

c) DANSYL Method

d) Enzymic Method

B. Determination of C-Terminal Amino Acids

a) *Hydrazinolysis Method*

b) *Reduction Method Enzymic Method*

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

BIOMOLECULES (PART-5, PPT-25)

Peptides-II, Reactions, End group Analysis

The Primary Structure of Peptide
Determination of Peptide sequence

The sequence of amino acid residues in a polypeptide or protein is called its primary structure. A simple peptide composed of three amino acids (a tripeptide) can have 6 different amino acid sequences. A tetrapeptide can have as many as 24 different sequences. For a protein composed of 20 different amino acids in a single chain of 100 residues, there are $2^{100} = 1.27 \times 10^{130}$ possible peptide sequences, a number much greater than the number of atoms estimated to be in the universe (9×10^{78})!

Clearly, one of the most important things to determine about a protein is the sequence of its amino acids. Fortunately, there are a variety of methods available to determine the sequence of amino acids in a polypeptide. The steps performed to establish the sequence of a peptide (or protein) by chemical analysis are usually the following:

- Hydrolysis of peptide to the corresponding amino acids.
- Separation, identification and quantification of amino acids.
- End-group analysis, i.e., to determine *N*-T-AA and *C*-T-AA residues present in the peptide.
- Partial hydrolysis of the peptide, i.e., the breakdown of the peptide into smaller fragments.
- Sequential degradation of peptides.
- Specific cleavage of peptide.
- Ultimately, combining together all the information obtained from various sources give the whole sequence of the peptide under study.

Hydrolysis of Peptide to Amino Acids

When a protein or polypeptide is refluxed with 6*M* hydrochloric acid for 24 h at 100 °C, hydrolysis of all the amide linkages usually takes place, liberating its constituent amino acids as a mixture. Tryptophan residues are largely destroyed during acidic hydrolysis. In addition, the amide side chains of asparagine, Asp-NH_2, [$H_2NCOCH_2CH(NH_2)CO_2H$] and that in glutamine, Glu-NH_2, [$H_2NCOCH_2CH_2CH(NH_2)CO_2H$] are hydrolyzed to acidic side chains producing aspartic acid [$HO_2CCH_2CH(NH_2)CO_2H$] and glutamic acid [$HO_2CCH_2CH_2CH(NH_2)CO_2H$], respectively.

Sulphur containing amino acids, such as, cysteine and cystine residues are unstable under the condition of the hydrolysis.

Separation, Identification and Quantification of Amino Acids

When a peptide is hydrolyzed, different amino acids are obtained. These amino acids are then separated, identified, and quantified employing different techniques. In some cases, the separation is based on acid-base properties. The water solubility of most amino acids or peptides is at a minimum when they are in zwitterionic form. Consequently, amino acids as well as peptides are precipitated from water when the pH of the medium is adjusted to their isoelectric point (p*I*). Amino acids and peptides are more soluble in water in their ionic forms. Therefore, they are more soluble in aqueous media at pH values far removed from their isoelectric points.

Another useful analytical technique for separating amino acids and peptides is electrophoresis which also operates on the basis of their isoelectric points. After the hydrolysis of peptide as many as 22 different amino acids may be present in the mixture. Therefore, isolation, identification and quantification of the individual amino acids are extremely difficult. Chromatographic separation and quantitative analysis of the resulting mixture can then be used to determine which amino acids composed the intact polypeptide and their relative amounts.

One chromatographic method for separation of a mixture of amino acids is based on the use of *cation-exchange resins* (Figure 1), which are insoluble polymers containing sulfonate groups. If an acidic solution containing a mixture of amino acids is passed through a column packed with a cation-exchange resin, the amino acids will be adsorbed by the resin because of attractive forces between the negatively charged sulfonate groups and the positively charged amino acids.

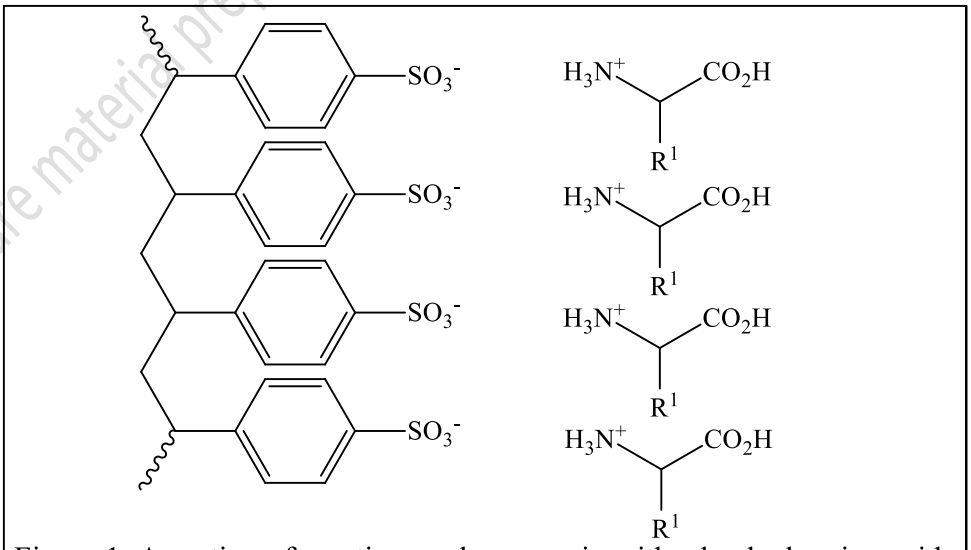

Figure 1: A section of a cation-exchange resin with adsorbed amino acids

The strength of the adsorption varies with the basicity of the individual amino acids. The most basic amino acids are held most strongly. The column is then washed with a buffered solution at a given pH, under this condition the individual amino acids move down the column at different rates and ultimately become separated. The rate at which each amino acid emerges from the column under very carefully defined conditions is accurately known from the standards.

In an automated version of this analysis developed at Rockefeller University in 1950, the eluate is allowed to mix with ninhydrin, a reagent that reacts with most amino acids to give a derivative with an intense purple color (λ_{max} 570 nm). The intensity of the resulting colour is proportional to the amount of amino acids present in the effluent.

The colour intensity, and therefore, the amount of each amino acid, is recorded as a function of time. The amino acid analyzer is designed so that it can measure the absorbance of the eluate (at 570 nm) continuously and record this absorbance as a function of the volume of the effluent. Thus, by hydrolysis of the peptide, separation of individual amino acids through column chromatography (using cation exchange resin) as a function of time, followed by reaction with ninhydrin, and quantification of the colour produced identifies the amino acids and the relative amounts of different amino acids present in a peptide can be determined.

End-Group Analysis

An amino acid sequence is ambiguous unless the direction in which to read it – left to read, or right to left is known. It is very important to know which end is the *N* terminus and which is the *C* terminus in a peptide. In determining the amino acid sequence of a peptide, it is often desirable to know what amino acids are present at the ends of the peptide chain. In particular, methods for identifying the amino acid terminal residue of the peptide have particular importance in peptide chemistry.

The terminal residue analysis techniques are, therefore, used to identify the *N*- and *C*-terminal amino acids in a peptide (or protein). Identification of the *N*-terminal and *C*-terminal amino acid units of a peptide chain provides helpful information.

Determination of *N*-terminal Amino Acids

Several chemical methods have been devised for identifying the *N*-terminal amino acid (*N*-T-AA) in a peptide (or protein). They all take advantage of the fact that the *N*-terminal amino group is free and can act as a nucleophile. The *α*-amino groups of all the other amino acids present in the peptide chain /molecule are part of amide linkages, are not free, and are much less nucleophilic.

These methods are based upon the following strategy. The amino group at the *N*-terminus of the peptide is ‘labeled’ by a reaction with a suitable reagent that is specific for the amines. The resultant peptide is then hydrolyzed to its constituent amino acids. After hydrolysis, the *N*-terminal amino acid retains the label so that it can be readily recognized and identified.

N-Terminal Amino Acid Determination

Edman Degradation

The most widely used procedure for identifying the *N*-terminal amino acid in a peptide is the Edman degradation method (1950). The chemistry of the Edman degradation is based on a labeling reaction between the *N*-terminal amino group of the peptide (or protein) and phenyl isothiocyanate, C_6H_5 N=C=S, in the presence of dilute alkali (Figure 2). A free amine function, usually in equilibrium with zwitterion species, is necessary for the initial bonding to the phenyl isothiocyanate reagent.

Phenyl isothiocyanate reacts with the *N*-terminal amino group of the peptide to form a phenylthiocarbamyl (PTC) derivative, a substituted thiourea. This moiety is conformationally mobile, and the sulfur atom may approach the carboxyl carbonyl of the *N*-terminal amino acid. Phenylthiocarbamyl (PTC) derivative is then cleaved from the peptide chain by acid (hydrochloric or trichloroacetic acid). Anhydrous HCl protonates the amide oxygen atom, activating it to nucleophilic attack by sulfur.

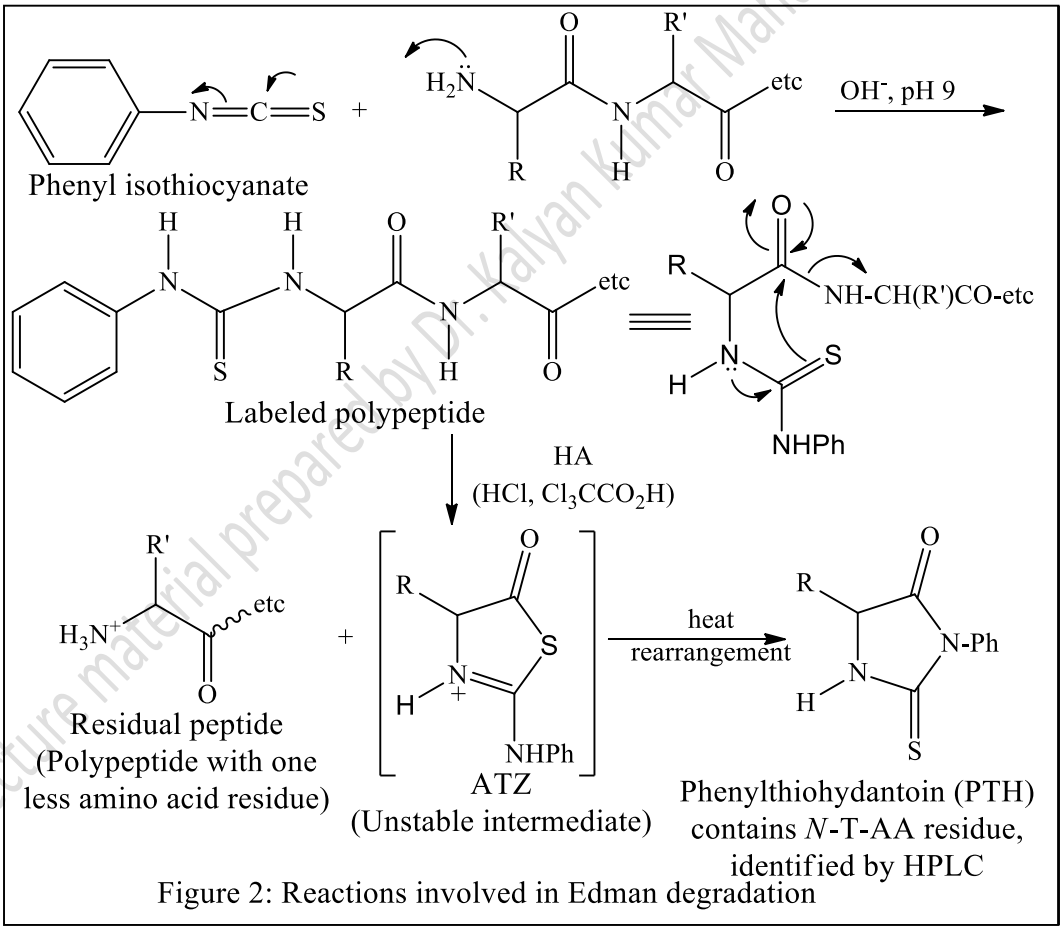

Figure 2: Reactions involved in Edman degradation

Next, a thiazolinone heterocycle incorporating the *N*-terminal unit is cleaved, leaving behind a shorter peptide chain which now has lost the *N*-T-AA of the original of the peptide. This heterocyclic ring, anilinothioazolinone (ATZ), is unstable and undergoes acid-catalyzed rearrangement to the isomeric more stable phenyl thiohydantoin (PTH) derivative of the

amino acid. Characteristic hydantoin derivatives of all the amino acids have been made and catalogued, so identification of the terminal unit is accomplished easily by comparison.

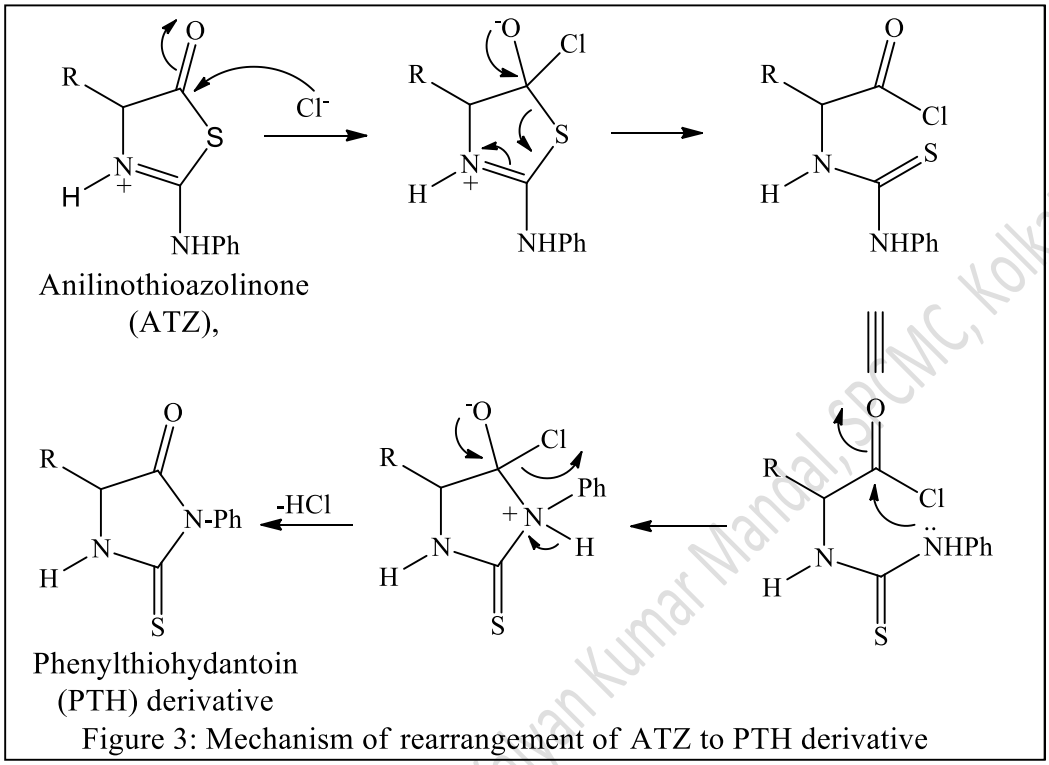

Figure 3: Mechanism of rearrangement of ATZ to PTH derivative

The PTH may be separated and identified by paper chromatography and the process can now be repeated on the degraded peptide. Used repetitively, the Edman degradation method can be used to sequence peptides up to about 60 residues in length. The process works so well that machines called amino acid sequencers have been developed to carry out the Edman degradation process in automated cycles and as a consequence, it can be used to determine the amino acid sequence in polypeptides, i.e., the step involving the splitting of polypeptides (or proteins) into smaller fragments.

In the automated process, the PTH derivative is introduced directly to a high-performance liquid chromatograph (HPLC) and identified by comparison of its retention time with known amino acid PTH derivatives. The cycle is then repeated for the next *N*-terminal amino acid.

A major advantage of the Edman procedure is that the remaining peptide chain is not further degraded by the reaction. This means that the *N*-terminal analysis may be repeated several times, thus providing the sequence of the first three to five amino acids in the chain.

A disadvantage of the procedure is that is peptides larger than 30 to 40 units do not give reliable results.

Sanger *N*-Terminal Analysis: DNP Method

Sanger's *N*-terminal analysis (DNP method, 1945) for the sequencing of peptides (or proteins), based on the use of 1-fluoro-2,4-dinitrofluorobenzene (FDNB). When a polypeptide is treated with FDNB in mildly basic solution, a nucleophilic aromatic substitution reaction (S_NAr) takes place involving the free amino group of the *N*-terminal residue. Subsequent hydrolysis of the polypeptide gives a mixture of amino acids in which the *N*-terminal amino acid is labeled with a 2,4-dinitrophenyl group. After separating this amino acid from the mixture, it can be identified by comparison with known standards.

1-Fluoro-2,4-dinitrofluorobenzene (FDNB) very readily reacts with amino groups in the presence of sodium hydrogen carbonate solution at room temperature to form 2,4-dinitrophenyl (DNP) derivatives which are stable to acids. Hence, hydrolysis with acid of the DNP peptide produces the DNP-amino acid and a mixture of free amino acids. Only the *N*-terminal amino acid residue of a peptide will bear the 2,4-dinitrophenyl group at its α-amino group. The overall reaction is outlined in Figure 4.

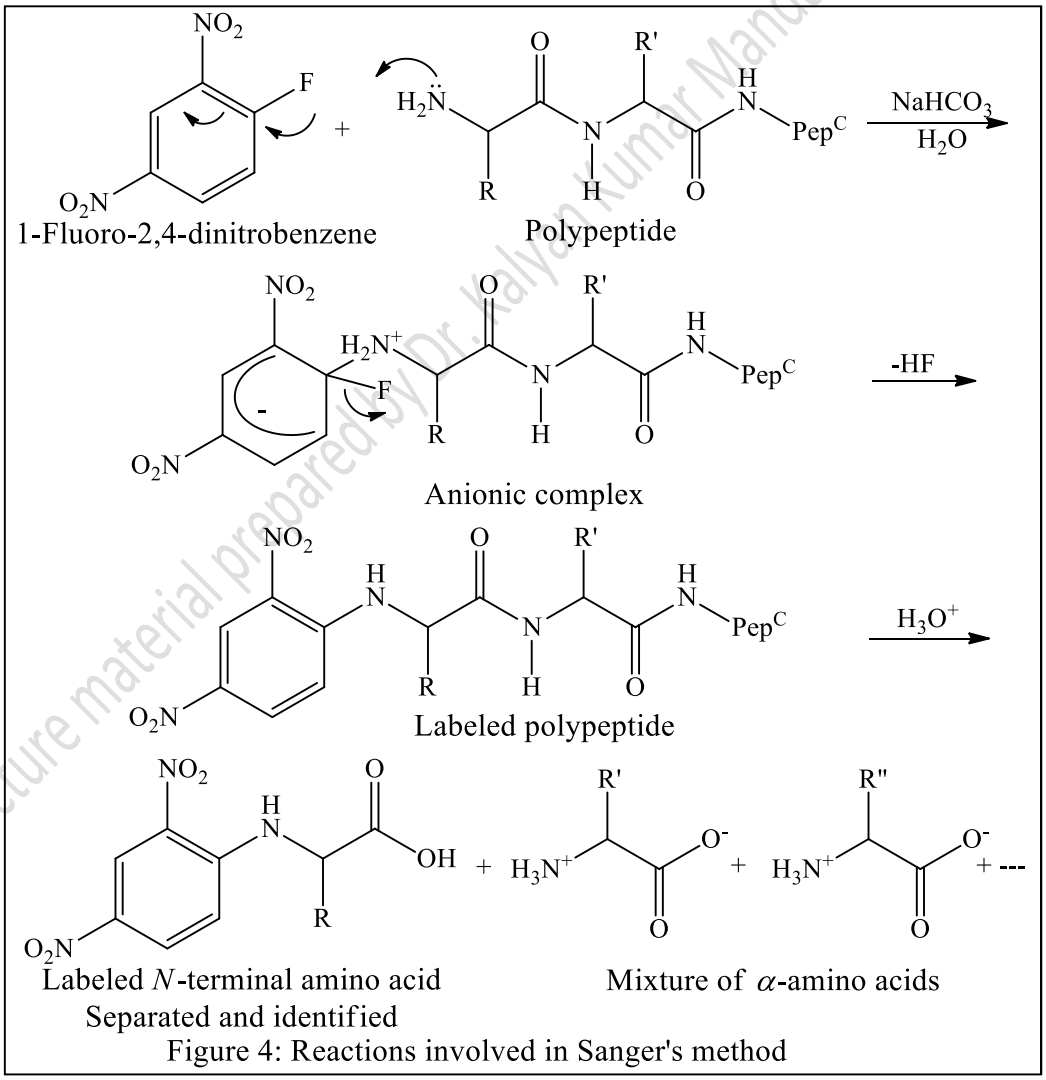

Figure 4: Reactions involved in Sanger's method

The electron-withdrawing property of the 2,4-dinitrophenyl group makes separation of the labeled amino acid very easy. The tagged amino acid (i.e., DNP derivative) being yellow

coloured and can be easily identified spectrophotometrically or identified by chromatographic (particularly TLC) comparison with known DNP-derivatives of the different amino acids. This method cannot be used repetitively, since its use requires complete hydrolysis of the DNP-derivative.

DNP-derivatives are formed with any free amino groups. Thus 2,4-Dinitrofluorobenzene will react with any free amino group in a polypeptide, including the ε-amino group of lysine. The hydroxyl group of tyrosine, the thiol group of cysteine, and the imidazole nucleus of histidine also react (although more slowly than an amino group). Hence, the DNP method may give rise to a number of DNP derivatives, and this fact complicates Sanger analyses. These, however, may be readily isolated and identified by chromatography (particularly TLC).

If the basic amino acid is not *N*-terminal, then it will form the mono-DNP derivative, and if *N*-terminal, the di-DNP derivative. The DNP derivatives of most of the amino acids have been prepared and characterized. Nevertheless, the Edman method of *N*-terminal analysis is much more widely used. The electron-withdrawing property of the 2,4-dinitrophenyl group makes separation of the labeled amino acid very easy.

The method is specific for FDNB. The reaction is a bimolecular nucleophilic substitution process, facilitated by the electron withdrawing nitro groups *ortho* and *para* to the site of nucleophilic attack. This results from the very powerfully electron withdrawing fluorine speeding up the reaction:

1. by making the nuclear carbon to which it is attached more positive and hence more readily attacked by the terminal -NH_2 group and
2. by helping to stabilize the anionic intermediate

The cheaper reagent 1-Chloro-2,4-dinitrobenzene is not used and is less effective in this role. This is because here the reaction is very slow. The reaction is an example of activated nucleophilic substitution which follows the order: Aryl fluoride > Aryl chloride > Aryl bromide > Aryl iodide

The DNP method was introduced by Frederick Sanger of Cambridge University in 1945. Sanger made extensive use of this procedure in his determination of the amino acid sequence of insulin and won the Nobel Prize in Chemistry for the work in 1958.

Determination of *N*-Terminal Amino Acids DANSYL Method

A modification of the sanger's DNP method is the use of 5-dimethylamino-1-sulfonyl chloride, 'dansyl' chloride (DNS-Cl), in place of FDNB. This modification is called the "dansyl'' method, illustrated in Figure 5, its use is similar to that of the DNP method. The dansylated amino acid is separated from the unlabeled amino acids and identified. The amino acid that is dansylated must correspond to the residue at the amino terminus of the original peptide. The dansyl method is now being widely used because the dansyl group, being highly fluorescent, permits the detection and estimation of dansyl amino acids in minute amounts by fluorometric methods.

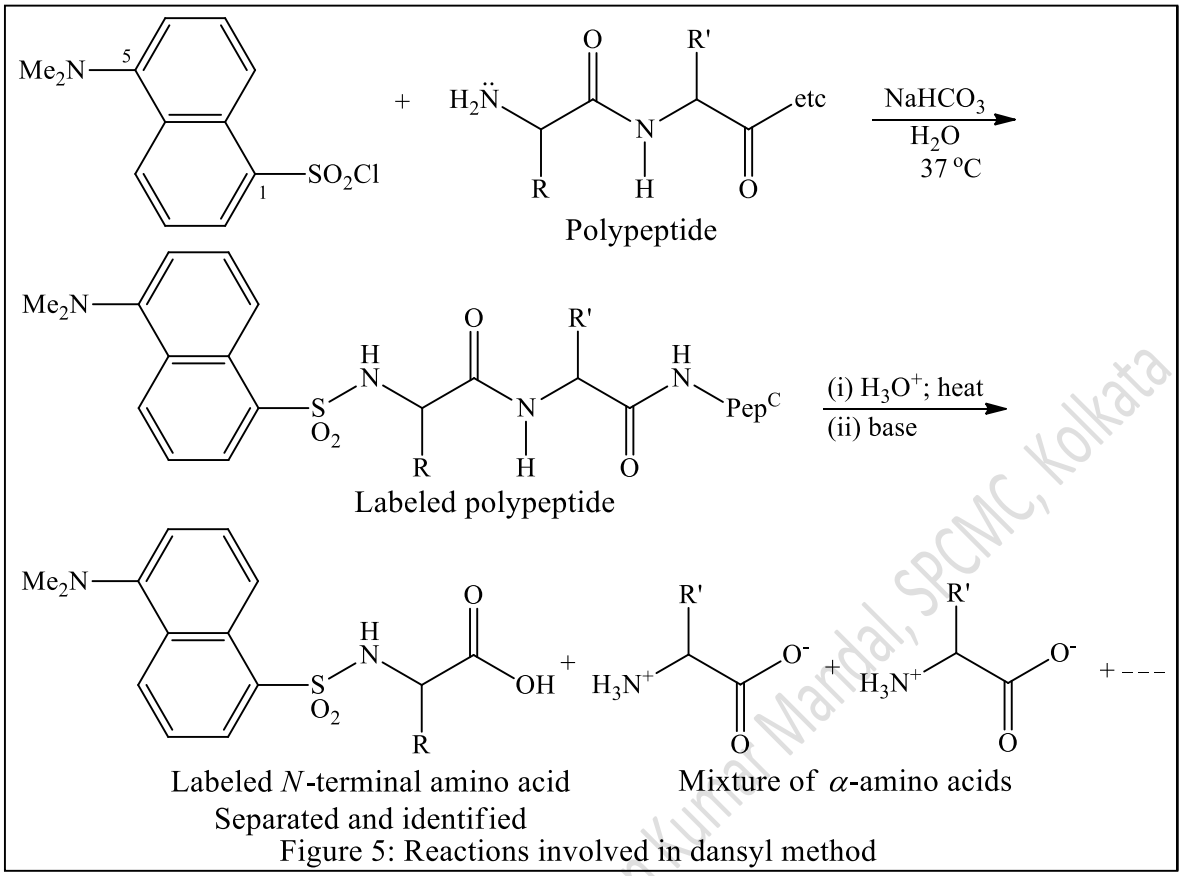

Figure 5: Reactions involved in dansyl method

Determination of *N*-Terminal Amino Acids Enzymic Method

Apart from some other chemical methods, an enzymic method is also available for *N*-terminal amino acid determination of a peptide (or protein). The enzyme leucine aminopeptidase attacks peptides (or proteins) only at the end which contains the free amino group and proceeds to liberate, in succession, each new terminal amino acid. Thus, in the peptide X. Y. Z---, after a given time of hydrolysis, a number of 'successive' amino acids will have been liberated, but in amounts X > Y > Z> ---. Hence, after a given time of hydrolysis, estimation of the amounts of free amino acids, their identification and quantitative determination will give the sequence.

Determination of *C*-Terminal Amino Acids

Hydrazinolysis Method

The most widely used method for the *C*-terminal amino acid residue in a peptide (or protein) is that of hydrazinolysis (1956). The peptide (or protein) is heated with anhydrous hydrazine at 100 °C (Figure 6). This converts all amino acid residues except the *C*-terminal one into amino acid hydrazides.

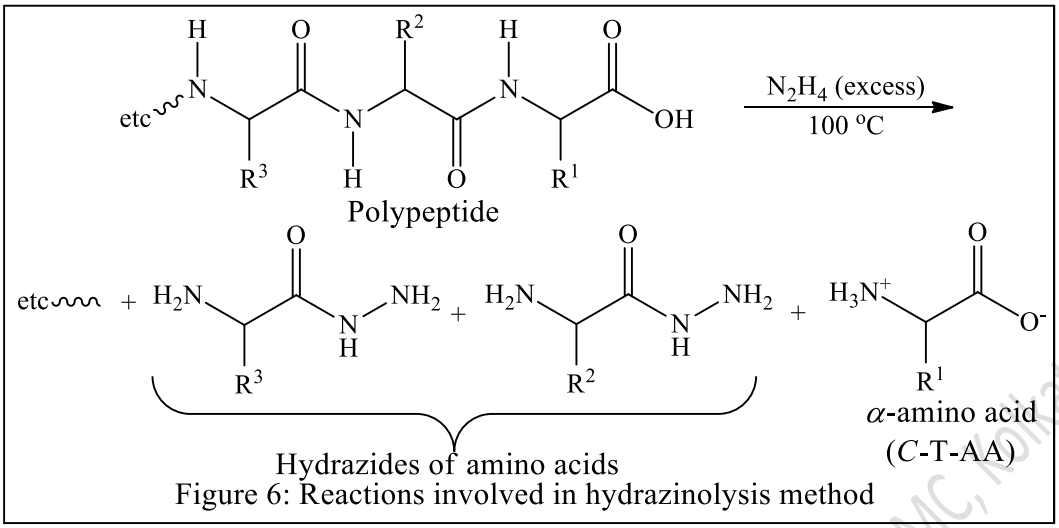

Figure 6: Reactions involved in hydrazinolysis method

The mixture of products is subjected to chromatography on a column of a strong cation-exchange resin. On elution the strongly basic hydrazides are retained, but the free amino acid is eluted and can be identified by paper chromatography with the known samples.

Determination of *C*-Terminal Amino Acids
Reduction Method

An important method for the determination of *C*-terminal amino acid residue involves the reduction of the peptide (or protein) with lithium borohydride ($LiBH_4$) or lithium aluminium hydride ($LiAlH_4$) (Figure 7). This converts the free terminal carboxyl group to a primary alcoholic group. Hydrolysis produces a mixture of amino acids and an amino alcohol, the latter being separated and identified by paper chromatography.

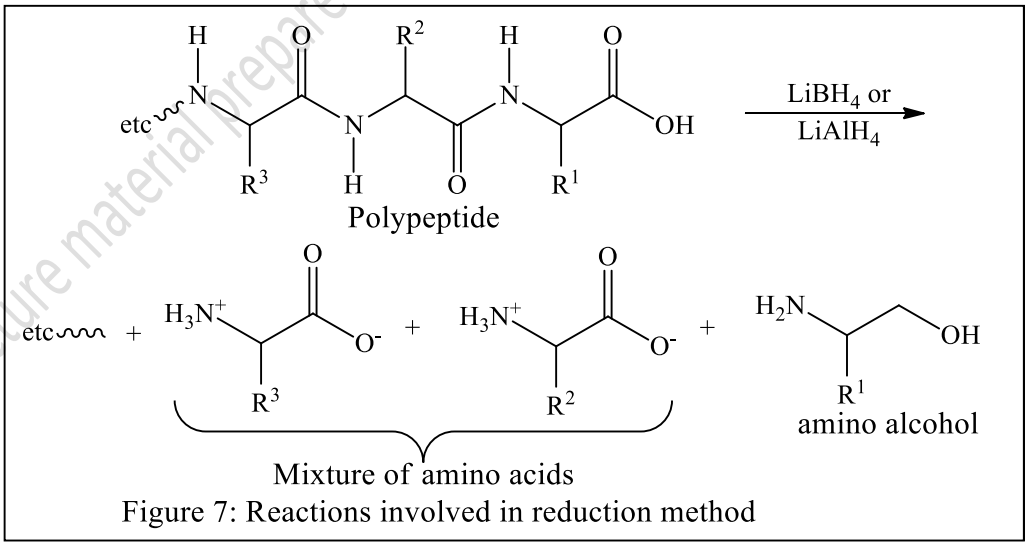

Figure 7: Reactions involved in reduction method

Determination of *C*-Terminal Amino Acids
Enzymic Method

A widely used method makes use of the enzyme carboxypeptidase for the determination of *C*-terminal amino acid of a peptide (or protein). This enzyme attacks peptides (or proteins) only at the end which contains the free α-carboxy group. When this terminal amino acid residue is liberated, the new terminal free carboxyl group is attacked by the enzyme. Thus, in the peptide --- A. B. C, after a given time of hydrolysis, a number of 'successive' amino acids will have been liberated, but in amounts C > B > A > ---. Hence, by identification and quantitative determination of the amino acids, their sequence can be established.