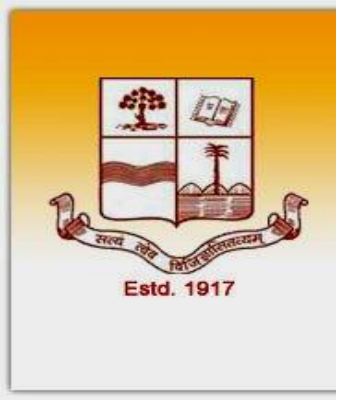
PROTEIN HYDROLYSIS AND AMINO ACID SEQUENCING



DEPARTMENT OF CHEMISTRY
PATNA UNIVERSITY, PATNA

Dr. Mithilesh Kumar Singh

Associate Professor

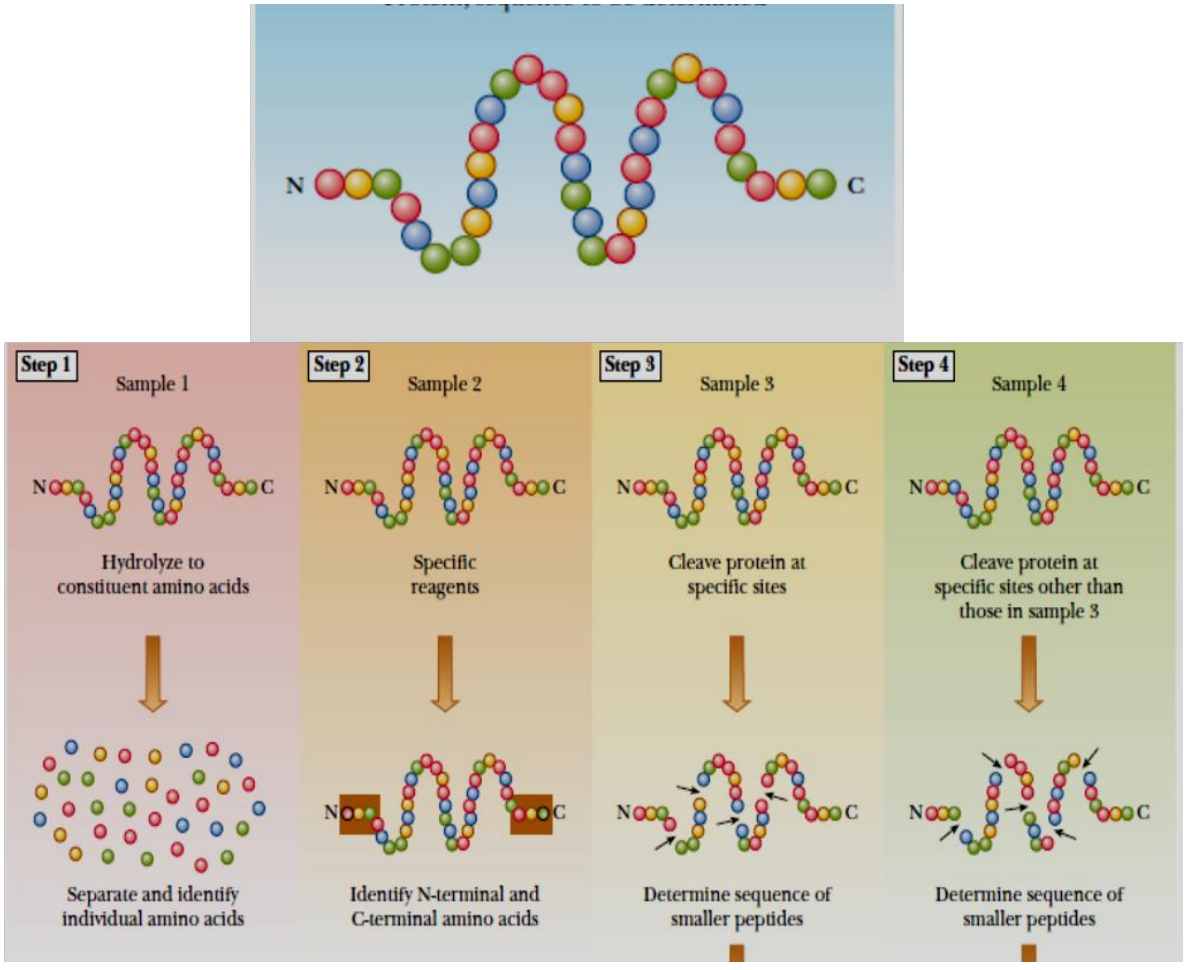
Exploring the Primary Structure of a Protein or Polypeptide

Determining the sequence of amino acids in a protein is carried out in several steps as mentioned below.

Step 1 : Determining Composition of Protein by Hydrolysis:

In determining the primary structure of a protein is to establish which amino acids are present and in what proportions.

(a) Degradation : Breaking down a protein to its component amino acids is relatively easy: heating a solution of the protein in acid, usually 6 *M* HCl, at 100°C to 110°C for 12 to 36 hours to hydrolyze the peptide bonds.



Common strategy for determining the primary structure of a given protein.

Amino acids in hydrolysates can be separated by ionexchange chromatography on columns of sulfonated polystyrene. The identity of the amino acid is revealed by its elution volume, which is the volume of buffer used to remove the amino acid from the column, and quantified by reaction with ninhydrin. Amino acids treated with ninhydrin give an intense blue color, except for proline, which gives a yellow color because it contains a secondary amino group. The concentration of an amino acid in a solution, after heating with ninhydrin, is proportional to the optical absorbance of the solution. This technique can detect a microgram (10 nmol) of an amino acid, which is about the amount present in a thumbprint. As little as a nanogram (10 pmol) of an amino acid can be detected by replacing ninhydrin with fluorescamine, which reacts with the α -amino group to form a highly fluorescent product. A comparison of the chromatographic patterns of our sample hydrolysate with that of a standard mixture of amino acids would show that the amino acid composition of the peptide.

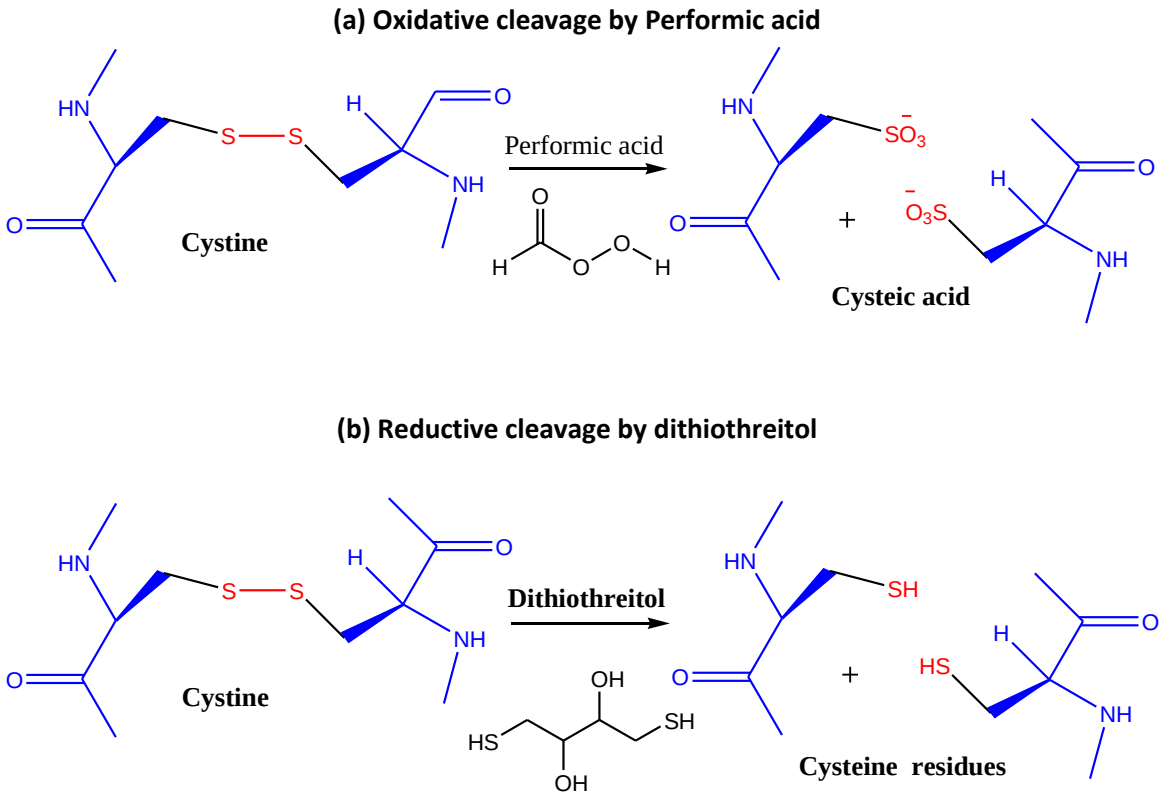
(b)Separation and Identification : Separation and identification of the products are somewhat more difficult and are best done by an amino acid analyzer. This automated instrument gives both qualitative information about the identities of the amino acids present and quantitative information about the relative amounts of those amino acids. The next step is often to identify the N-terminal amino acid by labeling it with a compound that forms a stable covalent bond. Fluorodinitrobenzene (FDNB) was first used for this purpose by Frederick Sanger. Dabsyl chloride is now commonly used because it forms fluorescent derivatives that can be detected with high sensitivity. It reacts with an uncharged a -NH_2 group to form a sulfonamide

derivative that is stable under conditions that hydrolyze peptide bonds. Hydrolysis of our sample dabsyl-peptide in 6 N HCl would yield a dabsyl-amino acid, which could be identified as dabsyl-alanine by its chromatographic properties. Dansyl chloride, too, is a valuable labeling reagent because it forms fluorescent sulfonamides. Although the dabsyl method for determining the amino-terminal residue is sensitive and powerful, it cannot be used repeatedly on the same peptide, because the peptide is totally degraded in the acid-hydrolysis step and thus all sequence information is lost. Pehr Edman devised a method for labeling the amino-terminal residue and cleaving it from the peptide without disrupting the peptide bonds between the other amino acid residues.

The amino acid can be determined by four different analyses performed on four separate samples of the same protein. An amino acid analyzer separates the mixture of amino acids either by ion-exchange chromatography or by high performance liquid chromatography (HPLC).

Step-2: Cleavage of Disulphide Bonds :

If any disulfide bonds are present, they must be broken. Each fragment is purified, then sequenced by the Edman procedure. Finally, the order in which the fragments appear in the original protein is determined and disulfide bonds (if any) are located. Following two schemes for oxidative and reductive cleavage of disulphide bonds are underlined below.



Step 3: Terminal Identification : The identities of the N-terminal and C-terminal amino acids in a protein sequence are determined.

Steps 4 and 5: Sequential Analysis of Amino acid Residues :

The protein is cleaved into smaller fragments, and the amino acid sequence is determined. Automated instruments can perform a stepwise modification starting from the N-terminal end, followed by cleavage of each amino acid in the sequence and the subsequent identification of each modified amino acid as it is removed. This process is called the Edman degradation.

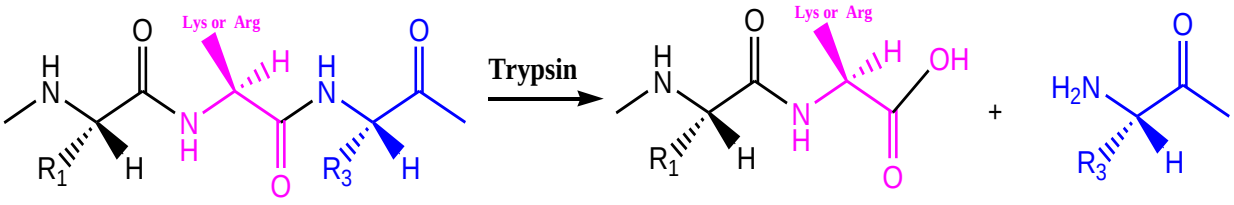
Why are the proteins cleaved into small fragments for sequencing?

The Edman degradation method becomes more difficult as the number of amino acids increases. In most proteins, the chain is more than 100 residues long. For sequencing, it is usually necessary to break a long polypeptide chain into fragments, ranging from 20 to 50 residues for reasons that will be explained later.

Cleavage of the Protein: Chemical and Enzymatic Methods :

Proteins can be cleaved at specific sites by enzymes or by chemical reagents.

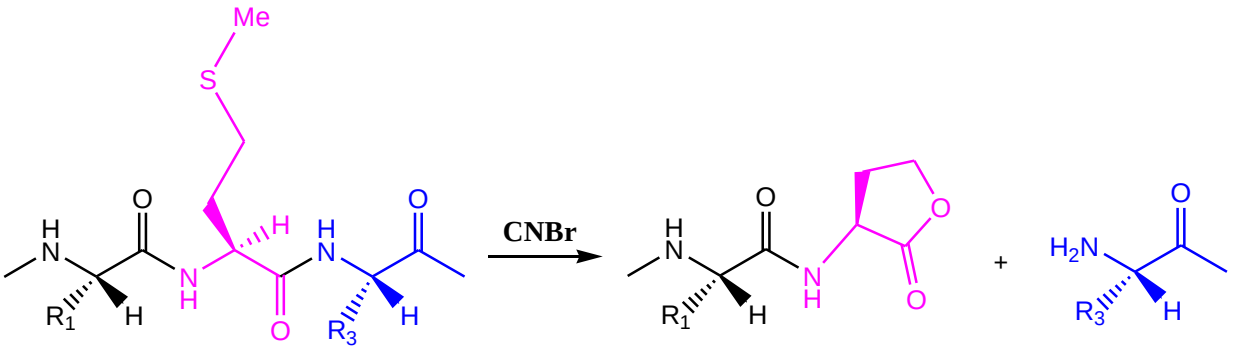
TRYPSIN : The enzyme trypsin cleaves peptide bonds preferentially at amino acids that have positively charged R groups, such as lysine and arginine. The cleavage takes place in such a way that the amino acid with the charged side chain ends up at the C-terminal end of one of the peptides produced by the reaction. The C-terminal amino acid of the original protein can be any one of the 20 amino acids and is not necessarily one at which cleavage takes place.



CHYMOTRYPSIN : Another enzyme, chymotrypsin, cleaves peptide bonds preferentially at the aromatic amino acids: tyrosine, tryptophan, and

phenylalanine. The aromatic amino acid ends up at the C-terminal ends of the peptides produced by the reaction.

CYANOGEN BROMIDE : In the case of the chemical reagent cyanogen bromide (CNBr), the sites of cleavage are at internal methionine residues. The sulfur of the methionine reacts with the carbon of the cyanogen bromide to produce a homoserine lactone at the C-terminal end of the fragment .



Methionene **Homoserine lactone**

Chemical and Enzymatic Cleavage of Peptides and Proteins :

S.No	Reagent	Cleavage Site
CHEMICAL CLEAVAGE		
1	Cyanogen bromide	Carboxyl side of methionine residues
2	*O*-Iodosobenzoate	Carboxyl side of tryptophan residues
3	Hydroxylamine	Asparagine-glycine bonds
4	2-Nitro-5-thiocyanobenzoate	Amino side of cysteine residues

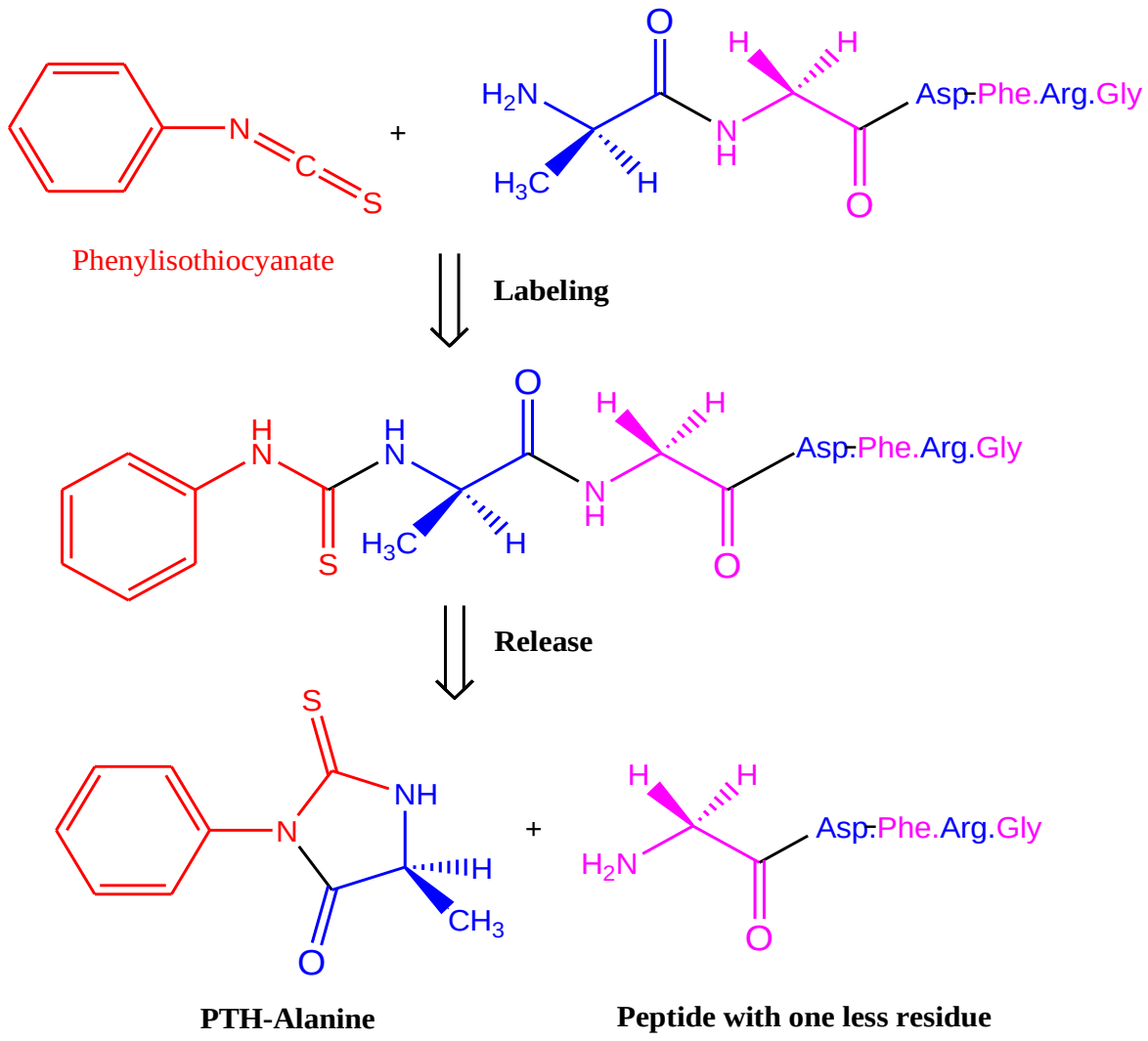
ENZYMATIC CLEAVAGE		
1	Trypsin	Carboxyl side of lysine and arginine residues
2	Clostripain	Carboxyl side of arginine residues
3	Chymotrypsin :	Carboxyl side of tyrosine, tryptophan, phenylalanine, leucine, and methionine
4	Carboxypeptidase A	Amino side of C-terminal amino acid (not arginine, lysine, or proline)

Cleavage of a protein by any of these reagents produces a mixture of peptides, which are then separated by high-performance liquid chromatography.

Degradation Outcomes for an Oligopeptide Chain Underlined Below	
$H_3\overset{+}{N}$ - Leu - Asn - Asp - Phe - His- Met - Thr - Met - Ala- Trp - Val - Lys- $\overset{-}{COO}$	
Chymotrypsin	$H_3\overset{+}{N}$ - Leu - Asn - Asp - Phe
Cyanogen bromide	$H_3\overset{+}{N}$ - Leu - Asn - Asp - Phe - His- Met
Chymotrypsin	His- Met - Thr - Met - Ala- Trp
Cyanogen bromide	Thr - Met
Cyanogen bromide	Ala- Trp - Val - Lys- $\overset{-}{COO}$
Chymotrypsin	Val - Lys- $\overset{-}{COO}$

The sequences of a set of peptides produced by one reagent overlap the sequences produced by another reagent. As a result, the peptides can be arranged in the proper order after their own sequences have been determined.

The Edman Degradation : Disintegrating Longer Polypeptide Chains



The Edman degradation sequentially removes one residue at a time from the amino end of a peptide Phenyl isothiocyanate reacts with the uncharged terminal amino group of the peptide to form a phenylthiocarbamoyl derivative.

Then, under mildly acidic conditions, a cyclic derivative of the terminal amino acid is liberated, which leaves an intact peptide shortened by one amino acid. The cyclic compound is a phenylthiohydantoin (PTH)-amino acid, which can be identified by chromatographic procedures. The Edman procedure can then be repeated on the shortened peptide, yielding another PTH amino acid, which can again be identified by chromatography. Three more rounds of the Edman degradation will reveal the complete sequence of the original peptide pentapeptide.

The development of automated sequencers has markedly decreased the time required to determine protein sequences. One cycle of the Edman degradation the cleavage of an amino acid from a peptide and its identification is carried out in less than 1 hour. The labeled amino-terminal residue (PTH-alanine in the first round) can be released without hydrolyzing the rest of the peptide. Hence, the amino-terminal residue of the shortened peptide (GlyAsp-Phe-Arg-Gly) can be determined in the second round. Three more rounds of the Edman degradation reveal the complete sequence of the original peptide. By repeated degradations, the amino acid sequence of some 50 residues in a protein can be determined. High-pressure liquid chromatography provides a sensitive means of distinguishing the various amino acids. Gas-phase sequenators can analyze picomole quantities of peptides and proteins. This high sensitivity makes it feasible to analyze the sequence of a protein sample eluted from a single band of an SDS-polyacrylamide gel.

Recombinant DNA Technology Has Revolutionized Protein Sequencing:

Hundreds of proteins have been sequenced by Edman degradation of peptides derived from specific cleavages. Nevertheless, too much effort is required to elucidate the sequence of large proteins, those with more than 1000 residues. For sequencing such proteins, a complementary experimental approach based on recombinant DNA technology is often more efficient. Long stretches of DNA can be cloned and sequenced, and the nucleotide sequence directly reveals the amino acid sequence of the protein encoded by the gene. Recombinant DNA technology is producing a wealth of amino acid sequence information at a remarkable rate. Even with the use of the DNA base sequence to determine primary structure, there is still a need to work with isolated proteins. The amino acid sequence deduced by reading the DNA sequence is that of the nascent protein. Many proteins are modified after synthesis. Some have their ends trimmed, and others arise by cleavage of a larger initial polypeptide chain. Cysteine residues in some proteins are oxidized to form disulfide links, connecting either parts within a chain or separate polypeptide chains. Specific side chains of some proteins are altered. Amino acid sequences derived from DNA sequences are rich in information, but they do not disclose such posttranslational modifications. Chemical analyses of proteins in their final form are needed to delineate the nature of these changes, which are critical for the biological activities of most proteins.

Thus, genomic and proteomic analyses are complementary approaches to elucidating the structural basis of protein function.