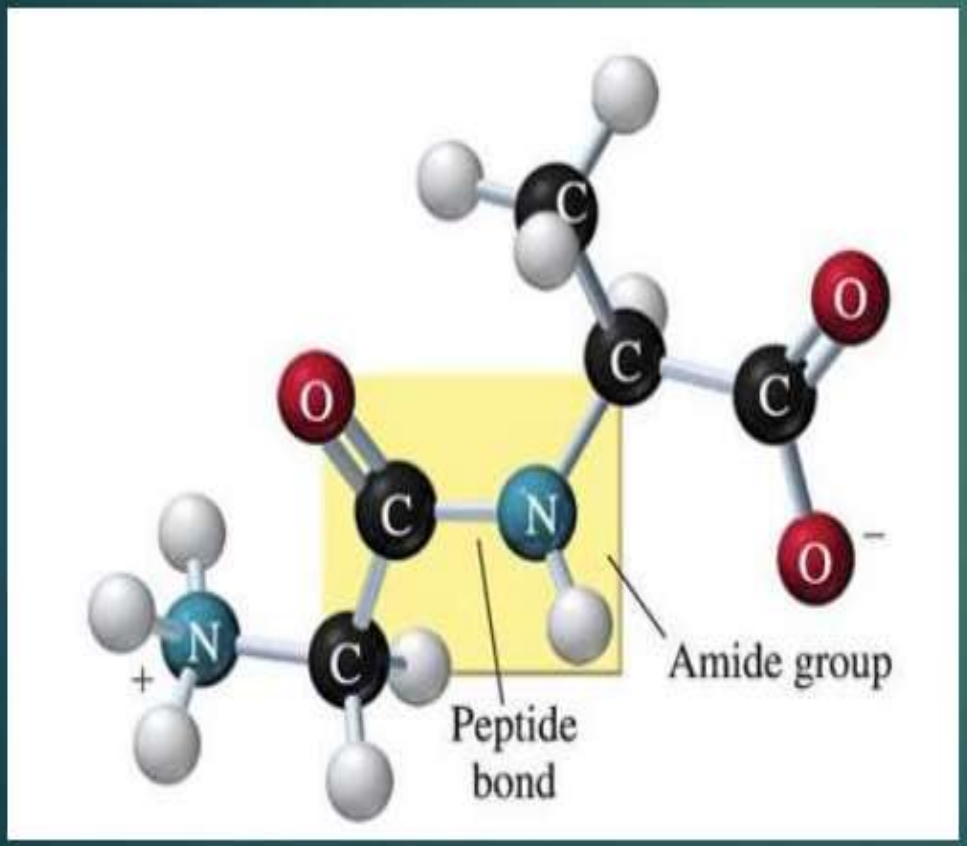
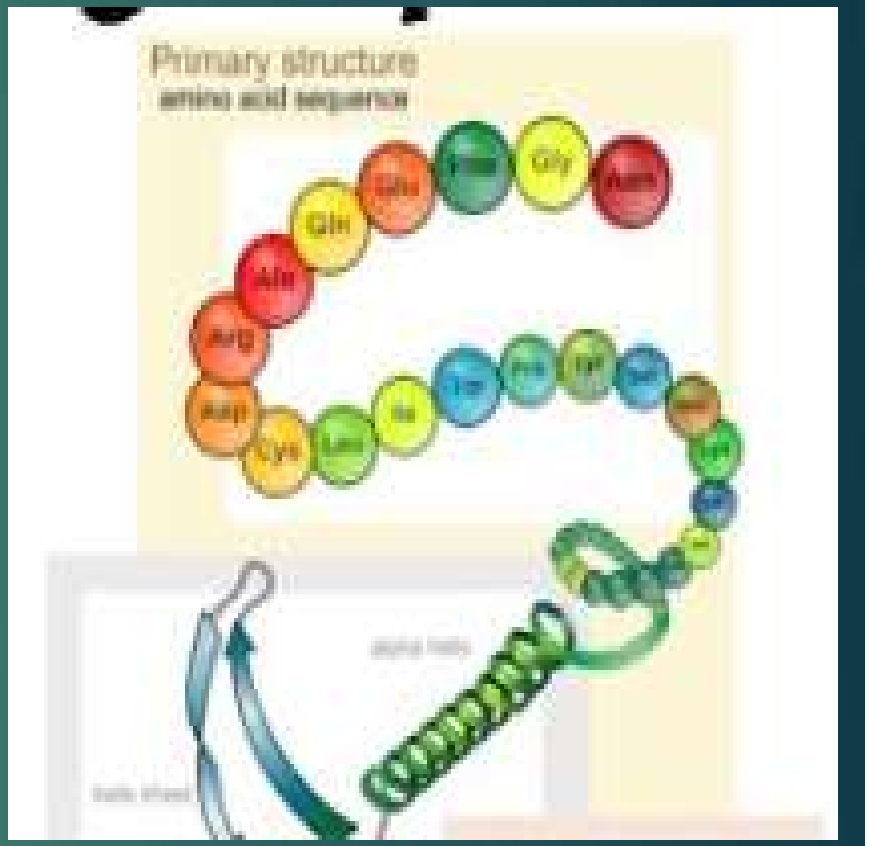
DETERMINATION OF PRIMARY STRUCTURE OF PROTEINS





Determination of primary structure

- The primary structure comprises the identification of amino acids with regard to their quality, quantity and sequence in a protein structure.

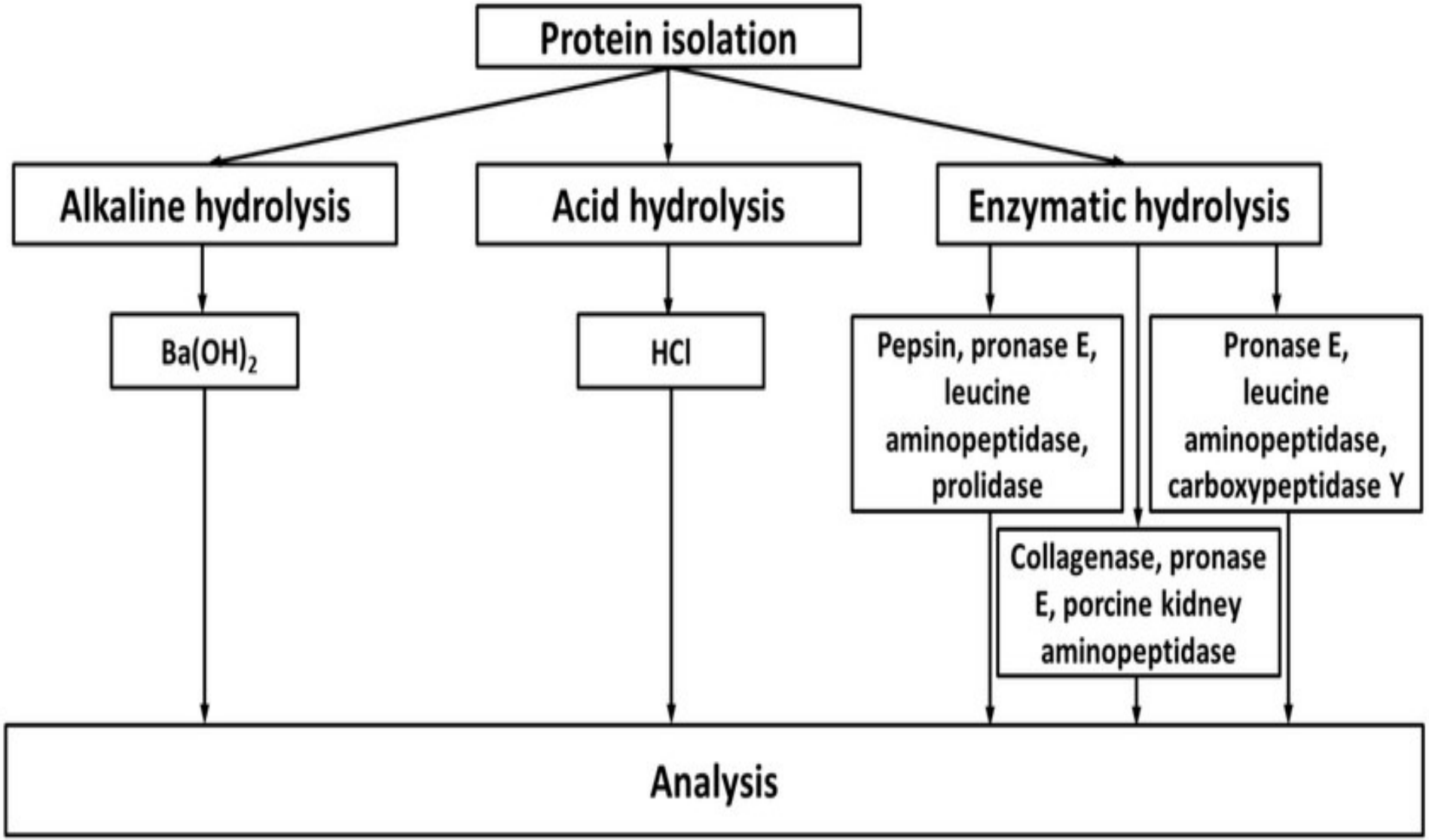
Determination of primary structure involves 3 stages:

- Determination of amino acid composition.
- Degradation of protein or polypeptide into smaller fragments.
- Determination of the amino acid sequence

Determination of amino acid composition

- The protein or polypeptide is completely hydrolyzed to liberate the amino acids
- The hydrolysis may be carried out either by acid or alkali treatment or by enzyme hydrolysis.
- *Pronase* is a mixture of non-specific proteolytic enzymes that causes complete hydrolysis of proteins.

Hydrolysis of proteins



I.1

Amino acid composition

Hydrolysis peptide
(HCl 6 M, 110 °C, 24 h)

↓

Seperate amino acids
(HPLC, ion-exchange chromatography, etc)

↓

Analysis results
(comperation of the chromatographic pattern of our sample with that of a standard mixture of amino acids)

↓

Amino acid composition
(elements and concentration)

Ala-Gly-Asp-Phe-Arg-Gly

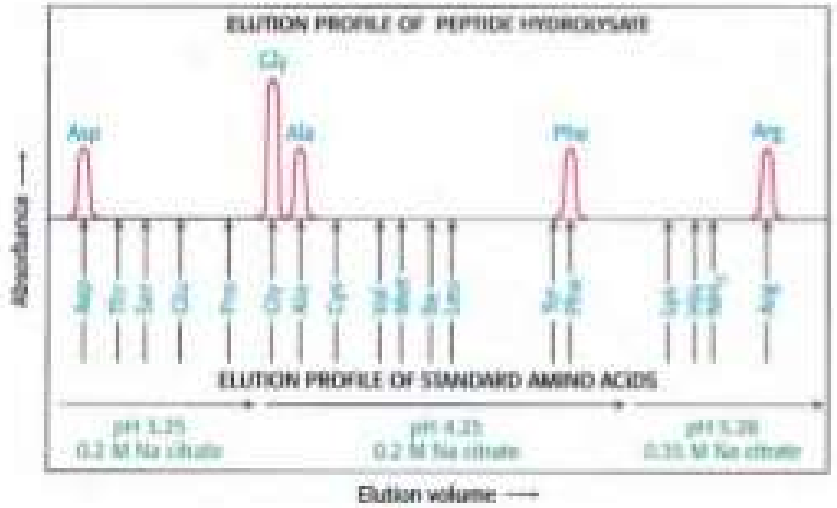
↓

Ala,Gly,Asp,Phe,Arg,Gly

↓

ion-exchange chromatography, ninhydrin reaction

↓



↓

Ala, Arg, Asp, 2Gly, Phe

Strategy for protein sequencing

- Determination of sub-units in polypeptides
- Breakage of disulphide bonds (inter and intra-chain)

- Disruption of tertiary and quaternary structure (Denaturation)
- Cleavage of polypeptide chain into smaller polypeptides
- Utilization of different enzymes/chemicals for generating several overlapping peptides
- Sequencing using Edman's degradation method (specifically identify N-terminal amino acid)
- Complete the sequence from overlaps of different peptide frangments in order to determine accurate amino acid sequence in peptides and polypeptides

Determination of number of polypeptides

- Number of polypeptides: The number of polypeptide chains can be identified by treatment of protein with *dansyl chloride.*
- It specifically binds with N-terminal amino acids to form dansyl polypeptides which on hydrolysis yield N-terminal dansyl amino acid.
- The number of dansyl amino acids produced is equal to the number of polypeptide chains in a protein.

Reduction/breakage of disulphide bonds

Reducing agents: Performic acid, 2-Mercaptoethanol, Dithiothereitol (DTT)

- **Breaks disulphide bonds (–S-S- bonds)**

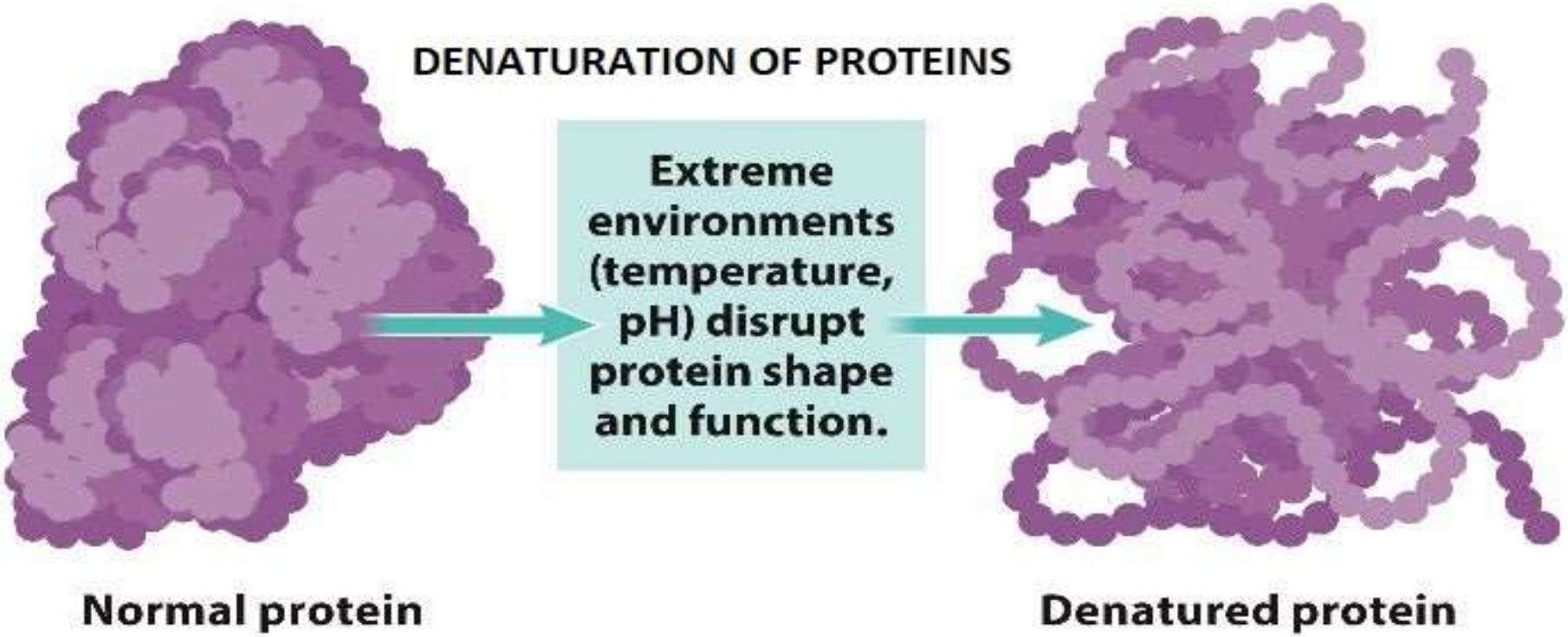
Denaturation of Protein

-The term denaturation denotes a reversible or irreversible change of native conformation (tertiary structure) without cleavage of covalent bonds (except for disulfide bridges).

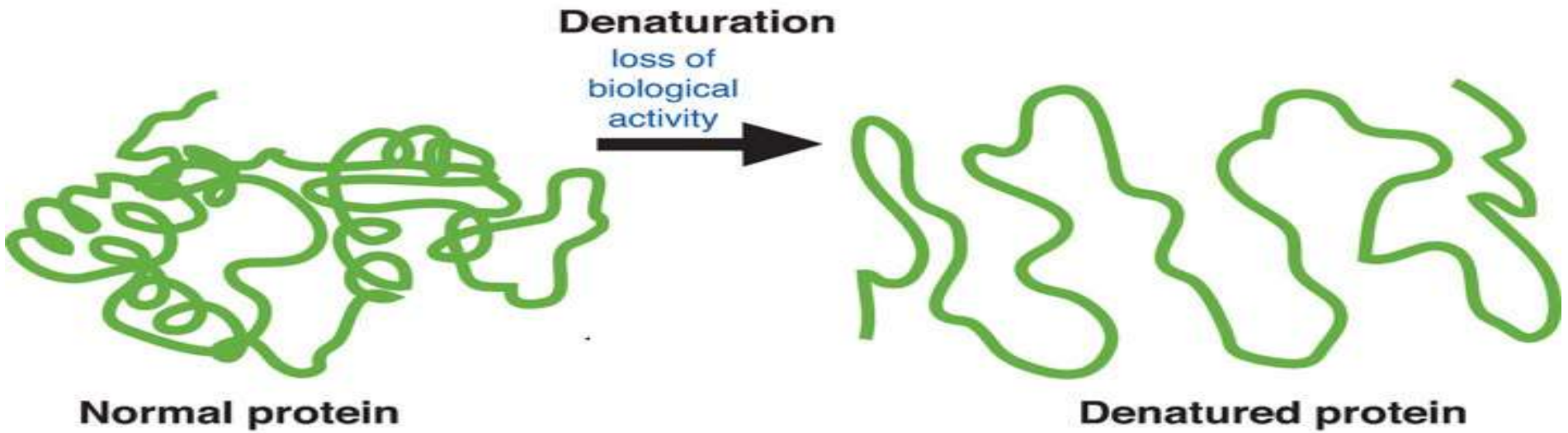
- The primary structure of the protein is not changed because the peptide bonds are not affected

Denaturing agents include:

1. Heat
2. Changes in pH (concentrated acids or alkali)
3. Ultraviolet rays
4. X ray
5. High salt concentration Ex: Detergents- Sodium dodecyl sulphate (SDS)
6. Heavy metals. Ex: Hg^{2+}, Pb^{2+}
7. Urea (6M)- disrupts the hydrogen bonds



agents: pH, temp, ionic strength, solubility

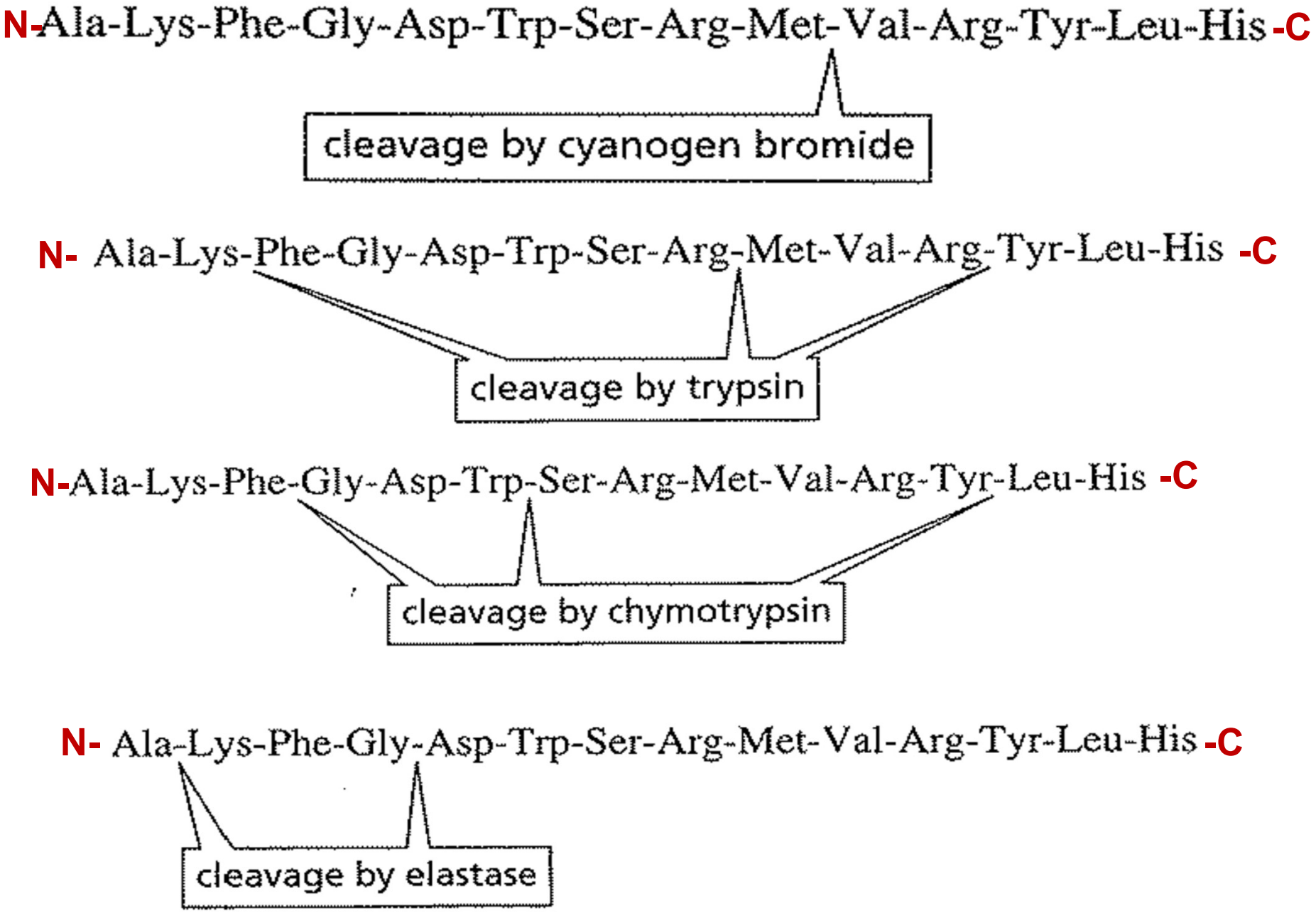


Breakdown of polypeptides into fragments

- Polypeptides are degraded into smaller peptides by enzymatic or chemical methods
- Enzymatic cleavage: The proteolytic enzymes such as *trypsin, chymotrypsin, pepsin and elastage* exhibit specificity in cleaving the peptide bonds.
- Among these enzymes, trypsin is commonly used.

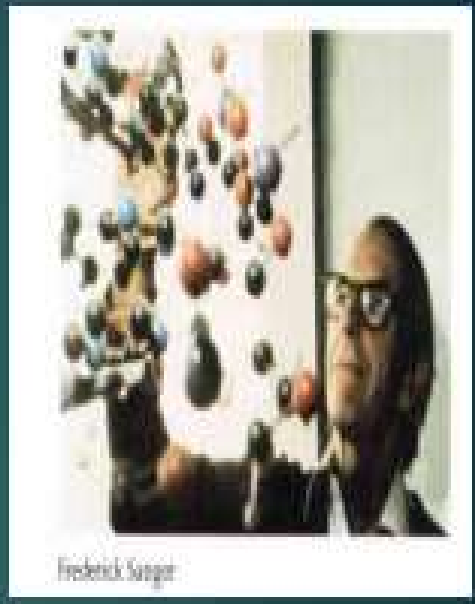
After denaturation of proteins, various types of chemical treatments or enzymatic treatment can be used to breaking down to smaller fragments and determination of the order/sequence of amino acids in protein.

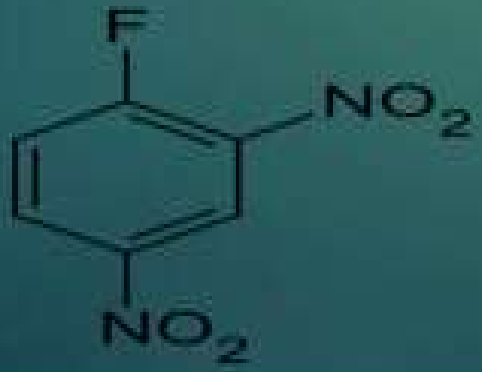
Classification	Reagent	Specificity
Chemical Reagents	Edman's Reagent or phenylisothiocyanate	Removes N-terminal AA
	Cyanogen bromide	Hydrolyzes C-side of Met
Use of Enzymes	Specific Carboxypeptidase A	Removes C-terminal AA (NOT Arg or Lys) • A for aliphatic and aromatic AA
	Specific Carboxypeptidase B	Removes C-terminal AA (ONLY Arg or Lys) • B for basic AA
	Carboxypeptidase (non-specific)	Removes C-terminal AA
	Trypsin	Hydrolyzes C-side of Arg & Lys
	Chymotrypsin	Hydrolyzes C-side of aromatic AA (Phe, Tyr, Trp)
	Elastase	Hydrolyzes C-side of small AAs (Gly, Ala)


N-Ala-Lys-Phe-Gly-Asp-Trp-Ser-Arg-Met-Val-Arg-Tyr-Leu-His-C
cleavage by cyanogen bromide
N- Ala-Lys-Phe-Gly-Asp-Trp-Ser-Arg-Met-Val-Arg-Tyr-Leu-His -C
cleavage by trypsin
N-Ala-Lys-Phe-Gly-Asp-Trp-Ser-Arg-Met-Val-Arg-Tyr-Leu-His -C
cleavage by chymotrypsin
N- Ala-Lys-Phe-Gly-Asp-Trp-Ser-Arg-Met-Val-Arg-Tyr-Leu-His -C
cleavage by elastase

Sanger's Method

- This was the first technique to determine the sequence of proteins.
- Sanger's Reagent – **2,4- Dinitroflouro Benzene**.
- Sanger' s Reagent derivatizes the amino terminal residues.
- The first proteins to be sequenced by the method is Insulin by **Fredrick Sanger**. He got Noble Prize in 1958.
- **Only dipeptides or tripeptides can be sequenced.**


Frederick Sanger

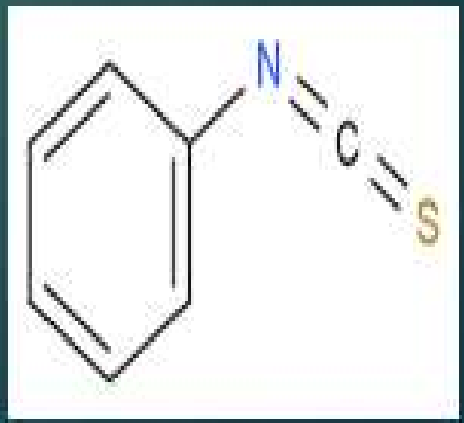


Steps:

- 2,4- dinitro fluorobenzene is reacted with amino group of a peptide or a proteins to form 2,4- dinitrophenyl derivative of N- terminal amino acid which is yellow in colour.
- The treated peptide is then subjected to acid hydrolysis which cleaves all the peptide bonds except the bond between 2,4- DNF and NH_2 group which is resistant to acid hydrolysis .
- Separated by Chromatography.

Edman's Degradation Method

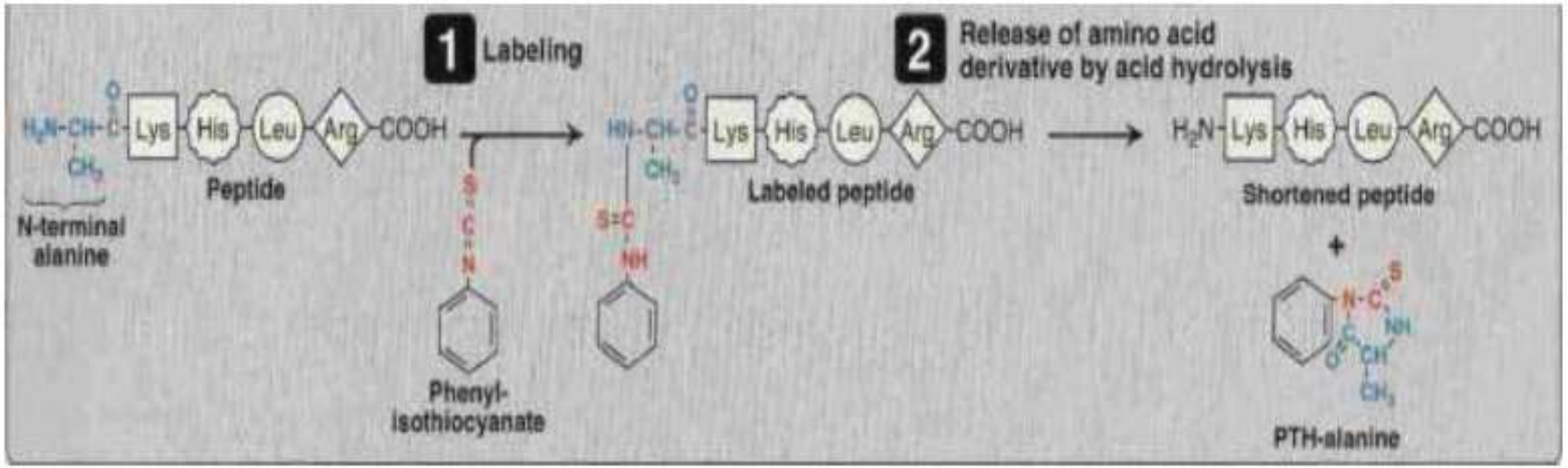
- This process was developed by Pehr Edmen.
- Edman 's Reagent – Phenyl isothiocyanate
- It is a technique for identifying specific amino acids at each position in the peptide chain beginning at the amino terminal end.
- Edman's technique can sequence many residues (5-40)of a single polypeptide sample.



Edmans degradation method

- Edmans reagent: Phenyl isothiocyanate is Edmans reagent.
- It reacts with the N-terminal amino acid of peptide to form a phenyl thiocarbamyl derivativve.
- On treatment with mild acid, phenyl thiohydrantoin (PTH) –amino acid, a cyclic compound is liberated.

- PTH amino acid can be identified by chromatography
- Edmans reagent has an advantage, a peptide can be sequentially degraded liberating N-terminal amino acids one after another which can be identified.
- This is due to the fact that the peptide as a whole is not hydrolyzed but only releases PTH- amino acids.



Overlapping peptides

- In the determination of primary structure of protein, several methods are simultaneously employed.
- This results in the formation of overlapping peptides
- Overlapping peptides are very useful in determining the amino acid sequence.

Protein Sequencing: Overlapping Sequences

Polypeptide

Procedure

hydrolyze; separate amino acids

Result

A	5	H	2	R	1
C	2	I	3	S	2
D	4	K	2	T	1
E	2	L	2	V	1
F	1	M	2	Y	2
G	3	P	3		

Conclusion

Polypeptide has 38 amino acid residues. Trypsin will cleave three times (at one R (Arg) and two K (Lys)) to give four fragments. Cyanogen bromide will cleave at two M (Met) to give three fragments.

react with FDNB; hydrolyze; separate amino acids

2,4-Dinitrophenylglutamate detected

E (Glu) is amino-terminal residue.

reduce disulfide bonds (if present)

cleave with trypsin; separate fragments; sequence by Edman degradation

- T-1 GASMALIK
- T-2 EGAAYHDFEPIDPR
- T-3 DCVHSD
- T-4 YLIACGPMTK

T-2 placed at amino terminus because it begins with E (Glu).

T-3 placed at carboxyl terminus because it does not end with R (Arg) or K (Lys).

cleave with cyanogen bromide; separate fragments; sequence by Edman degradation

- C-1 EGAAYHDFEPIDPRGASM
- C-2 TKDCVHSD
- C-3 ALIKYLIACGPM

C-3 overlaps with T-1 and T-4, allowing them to be ordered.

establish sequence

Amino terminus → EGAAYHDFEPIDPRGASMALIKYLIACGPMTKDCVHSD ← Carboxyl terminus

T-2 T-1 T-4 T-3

C-1 C-3 C-2

Sequenator

- This is an automatic machine to determine the amino acid sequence in a polypeptide
- It is based on the principle of Edman's degradation .
- Amino acids are determined sequentially from N-terminal end
- The PTH-amino acid liberated is identified by HPLC.
- Sequenator takes about 2 hours to determine each amino acid.

Reverse sequencing technique

- It is the genetic material (DNA) which ultimately determines the sequence of amino acids in a polypeptide chain
- By analyzing the nucleotide sequence of DNA that codes for protein, it is possible to translate the nucleotide sequence into amino acid sequence.
- This technique, however, fails to identify the disulfide bonds and changes that occur in the amino acids after the protein is synthesized.

• Steps for Determining the Primary Structure

Steps	Technique
I. Determination of amino acid composition in a protein	
Complete hydrolysis of protein	Pronase enzyme
Separation and estimation of amino acids	Chromatographic technique
II. Degradation of protein into smaller fragments	
Liberation of polypeptides	Urea or guanidine hydrochloride
Determination of the number of polypeptide chains in a protein	Dansyl chloride
Breakdown of polypeptides into fragments	Enzymatic cleavage Chemical cleavage
III. Determination of amino acid sequence	Sanger's reagent Edman's reagent
Nowadays, DNA sequencing is used to determine the amino acid sequence.	

Sample questions

Question 1: A pentapeptide was found to have the composition Ala Arg Gly Pro Trp. Reaction of the pentapeptide with Sanger's reagent, followed by hydrolysis, gave the DNP derivative of proline. Treatment of the pentapeptide with carboxypeptidase initially produced alanine. Treatment of the pentapeptide with trypsin gave a tetrapeptide which, when treated with chymotrypsin, produced a tripeptide. Which one of the following is the sequence of the pentapeptide? (note: trypsin cleaves after basic amino acids)

A) Pro Gly Arg Trp Ala
B) Pro Arg Gly Trp Ala
C) Pro Trp Arg Gly Ala
D) Pro Gly Trp Arg Ala
E) Pro Trp Gly Arg Ala

Question 2: Oxytocin, a hormone peptide of nine amino acids, is widely used in obstetrics to induce uterine contractions. There is an intramolecular disulfide bond which must be reduced before sequencing. Reduced oxytocin has the composition Asn Cys2 Gln Gly Ile Leu Pro Tyr. Partial hydrolysis of reduced oxytocin led to the following fragments.

Asn-Cys
Cys-Tyr
Tyr-Ile-Gln
Cys-Pro-Leu
Ile-Gln
Leu-Gly
Gln-Asn-Cys

- Reaction of reduced oxytocin with carboxypeptidase showed glycine as the first liberated amino acid. Which one of the following is the sequence of oxytocin?

A) Asn Cys Tyr Ile Gln Cys Pro Leu Gly
B) Gln Asn Cys Tyr Ile Cys Pro Leu Gly
C) Cys Pro Gln Asn Cys Tyr Ile Leu Gly
D) Cys Tyr Ile Gln Asn Cys Pro Leu Gly
E) Gln Asn Cys Tyr Ile Cys Pro Leu Gly