PROTEIN SEQUENCING

LECTURE DELIVERED

BY

NIRAJ KUMAR

DEPARTMENT OF BIOCHEMISTRY
PATNA UNIVERSITY

SEQUENCING OF PROTEINS

- Introduction
- What is protein ?
- What is protein sequencing ?
- History
- Determination of amino acid composition
  - -Hydrolysis
  - -Separation
  - -Quantitative Analysis
- Mechanism of protein sequencing
  - -Sanger's Method
  - -Edman Degradation
- Limitation of Edman reaction
- Application of protein sequencing
- Institute related to protein sequence
  - A. India
  - B. Abroad
- Conclusion
- References

INTRODUCTION

SEQUENCING OF PROTEINS

- ✓Protein sequencing is a technique to determine the amino acid sequence of a protein.
- ✓It is a method to understand the structure and function of proteins in living organism.
- ✓Amino acid sequence determines the eventual three dimensitional structure of the protein.

INTRODUCTION

SEQUENCING OF PROTEINS

What is protein?

All proteins are polymers of amino acid and all these amino acids excepts two have one amino group attached to the carboxyl group .

What is protein Sequencing ?

Protein sequencing is the technique to determine the amino acid sequence of a protein .

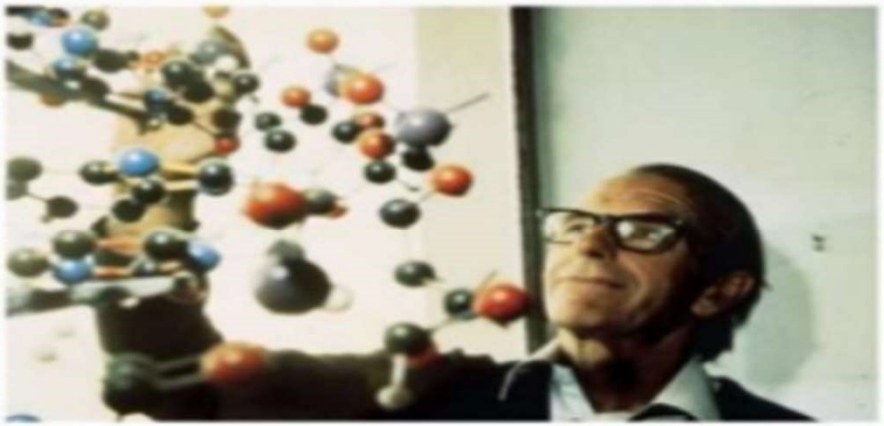
SEQUENCING OF PROTEINS

HISTORY

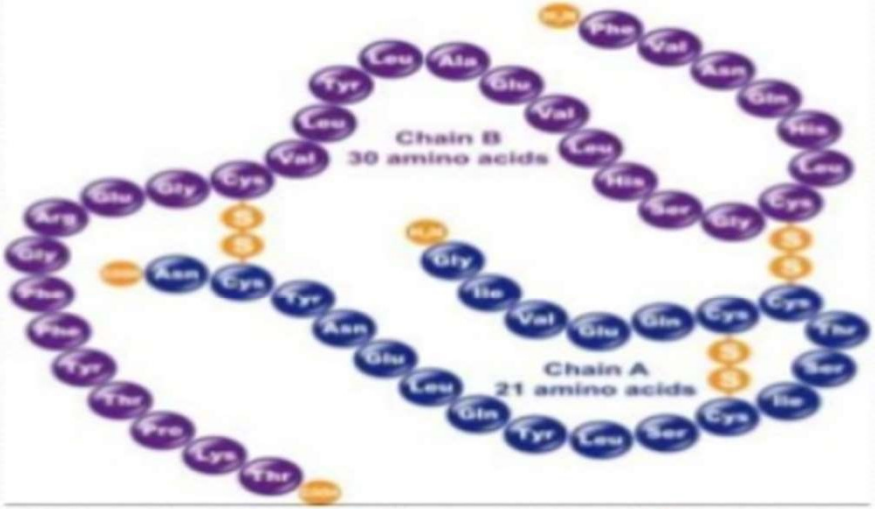
•**Pehr Edman**	**1947**	He found the method to decode the amino acid sequence of a protein using chemicals
•**Fredrick Sanger**	**1955**	He was able to present the complete sequence of insulin.

INTRODUCTION:

- **Protein sequencing** is the practical process of determining the amino acid sequence of all or part of a protein or peptide.
- In 1953, **Frederick Sanger** of Cambridge University in England reported the amino acid sequences of the two polypeptide chains composing the protein insulin.

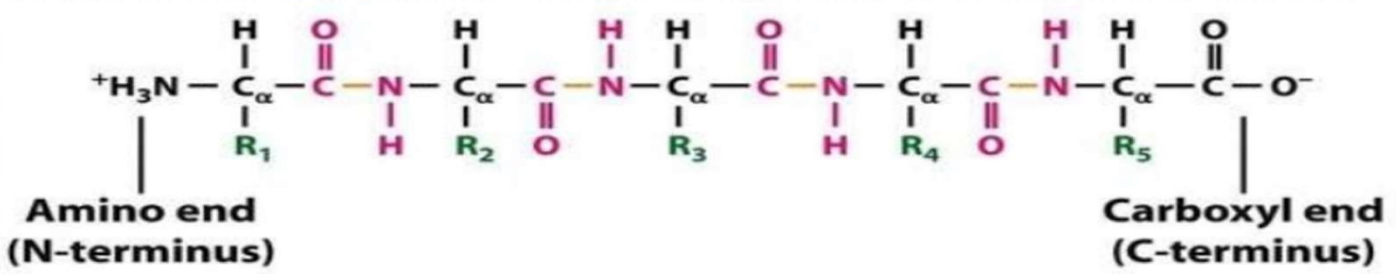


Frederick Sanger



Human Insulin

- Sanger's results clearly established that all of the molecules of a given protein have a fixed amino acid **composition**, a defined amino acid **sequence**, and therefore an invariant molecular weight. In short, proteins are well defined chemically.
- In 1958, Sanger was awarded the **Nobel Prize** for his contributions to determine the structure of insulin and other proteins.
- An alternative approach was that described by **Pehr Edman** (1950). This allowed determination of extended sequences of peptides or whole proteins, and has been used widely up to the present day.



PROTEIN SEQUENCING STRATEGY:

- The usual strategy for determining the amino acid sequence of a protein involves **eight basic steps**:

1. If the protein contains more than one polypeptide chain, the chains are separated and purified.
2. Intrachain (within a chain) S--S (disulfide) cross-bridges between cysteine residues in the polypeptide chain are cleaved. If these disulfides are interchain linkages, then step 2 precedes step 1.
3. The amino acid composition of each polypeptide chain is determined.
4. The N-terminal and C-terminal residues are identified.
5. Each polypeptide chain is cleaved into smaller fragments.
6. Sequence determination of peptide fragments.
7. The overall amino acid sequence of the protein is reconstructed from the sequences in overlapping fragments.
8. The positions of S--S cross-bridges formed between cysteine residues are located.

N-TERMINAL AMINO ACID ANALYSIS:

- A generalised method for *N*-terminal amino acid analysis follows:

i. React the peptide with a **reagent** that will selectively label the terminal amino acid.

ii. **Hydrolyse** the protein.

iii. Determine the amino acid by **chromatography** and comparison with standards.

- Determining which amino acid forms the *N*-terminus of a peptide chain is useful for the **ordering** of individual peptide fragments' sequences into a whole chain.
- Because the first round of Edman degradation is often contaminated by impurities and therefore does not give an accurate determination of the *N*-terminal amino acid.

C-TERMINAL AMINO ACID ANALYSIS:

- The number of methods available for C-terminal amino acid analysis is much smaller than the number of available methods of N-terminal analysis.
- The most common method is to add **carboxypeptidases** to a solution of the protein, take samples at regular intervals, and determine the terminal amino acid by analysing a plot of amino acid concentrations against time.
- This method will be very useful in the case of polypeptides and protein-blocked N termini.
- C-terminal sequencing would greatly help in verifying the **primary structures** of proteins predicted from DNA sequences and to detect any postranslational (after protein biosynthesis) processing of gene products from known codon sequences.

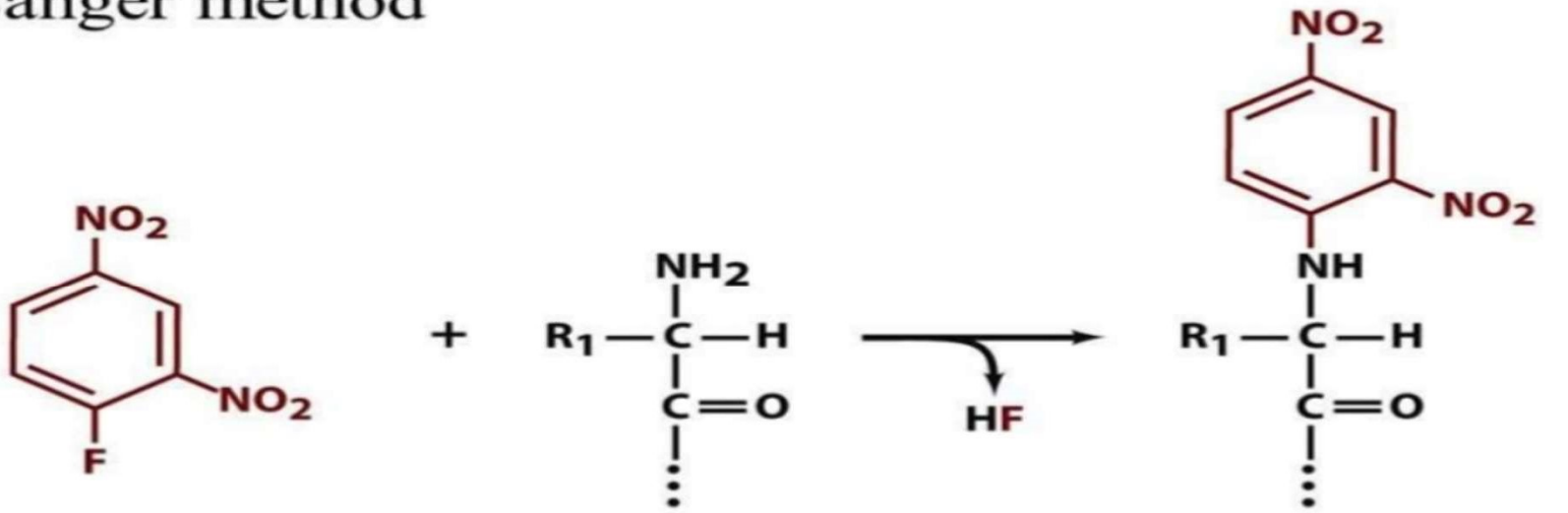
SANGER'S METHOD:

- Frederick Sanger had demonstrated a method to determine the amino acid residue located on the N-terminal end of a polypeptide chain by using the reagent **fluorodinitrobenzene (FDNB).**
- Sanger showed for the first time that amino acids are **covalently** together by α-amino and α-carboxyl group.
- He suggested stepwise release and identification of amino acid starting from N-terminal.
- He used a reagent fluorodinitrobenzene (FDNB)which is commonly called **Sanger's reagent**.
- The FDNB reacts with **free NH_2 group** of N-terminus.
- Upon hydrolysis a yellow coloured **dinitrophenol (DNP)** derivative of N-terminal amino acid is produced.

- The DNP amino acid is identified comparing it with a known standard DNP-amino acid by using **gel chromatography**.
- While it was thought, that at most, this method could only provide the sequences found on the N-terminal, Sanger was able to take the method one step further.
- By using several **proteolytic enzymes** (proteases, chymotrypsin and trypsin), **partial hydrolysis** and early version of chromatography, Sanger was able to cleave the protein into fragments and piece together the residues like a **jigsaw puzzle**.
- It wasn't until 1955 that Sanger was able to present the complete sequence of insulin which led to him being awarded a Nobel Prize in Chemistry in 1958.

❖ **Hydrolysis:** The peptide is hydrolyzed into its constituent amino acids by heating it with 6N HCL at 110º c for 24 hours.

Protein sequencing: Sanger method



2,4-Dinitrofluoro-benzene (DNFB)

Polypeptide

DNP Polypeptide

Upon hydrolysis, the N-terminal amino acid has a chromophore attached, which can be separated and identified

EDMAN DEGERADATION:

- The Edman degradation is a very important reaction for protein sequencing, because it allows the **ordered** amino acid composition of a protein to be discovered.
- **Automated Edman sequencers** are now in widespread use, and are able to sequence peptides up to approximately 50 amino acids long.
- The Edman Degradation method is based on the principle that single amino acid residues can be modified chemically such that they can be cleaved from the chain without disrupting the bonds between any other residues.
- The procedure can be achieved with very minute amounts of peptide, usually amounts on the order of 10-100 picomoles will allow for successful completion.

STEPS:

- Break any disulfide bridges in the protein with a **reducing agent** like 2-mercaptoethanol. A protecting group such as **iodoacetic acid** may be necessary to prevent the bonds from re-forming.
- Separate and purify the individual chains of the protein complex, if there are more than one.
- Determine the amino acid composition of each chain.
- Determine the terminal amino acids of each chain.
- Break each chain into fragments under 50 amino acids long.
- Separate and purify the fragments.
- Determine the sequence of each fragment.
- Repeat with a different pattern of cleavage.
- Construct the sequence of the overall protein.

- **Digestion into peptide fragments:**

  Digestion is done either by **endopeptidases** such as **trypsin** or **pepsin** or by chemical reagents such as **cyanogen bromide**. Different enzymes give different cleavage patterns, and the overlap between fragments can be used to construct an overall sequence.

- **Reaction:**

- The peptide to be sequenced is **adsorbed** onto a solid surface. One common substrate is glass fibre coated with **polybrene**, a cationic polymer.
- The Edman reagent, **phenylisothiocyanate (PITC)**, is added to the adsorbed peptide, together with a mildly basic buffer solution of 12% **trimethylamine.** This reacts with the amine group of the N-terminal amino acid.
- The terminal amino acid can then be selectively detached by the addition of **anhydrous** (contains no water) **acid**.

- The derivative then isomerises to give a substituted **phenylthiohydantoin**, which can be washed off and identified by chromatography, and the cycle can be repeated. The efficiency of each step is about 98%, which allows about 50 amino acids to be reliably determined.

- **Protein sequenator:**

- A **protein sequenator** is a machine that performs Edman degradation in an automated manner. A sample of the protein or peptide is **immobilized** (restrict the movements) in the reaction vessel of the protein sequenator and the Edman degradation is performed.
- Each cycle releases and derivatises one amino acid from the protein or peptide's *N*-terminus and the released amino-acid derivative is then identified by **HPLC**. The sequencing process is done repetitively for the whole polypeptide until the entire measurable sequence is established or for a pre-determined number of cycles.

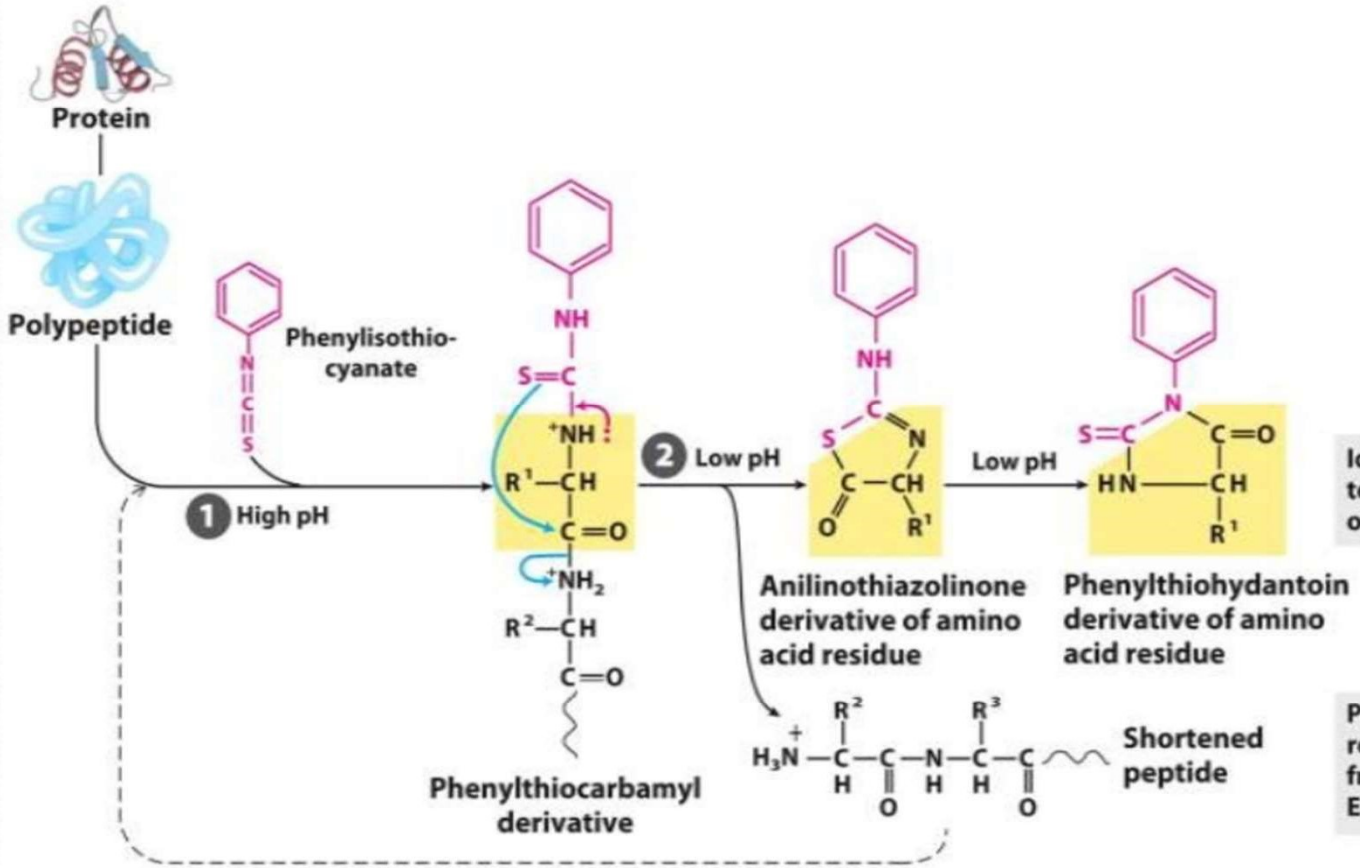


Figure 3-27
Lehninger Principles of Biochemistry, Sixth Edition

The peptide bond nearest to the amino terminus of the protein or polypeptide is cleaved in two steps. The two steps are carried out under very different reaction conditions (*basic* conditions in step 1, *acidic* in step 2), allowing one step to proceed to completion before the second is initiated.

LIMITATIONS:

- Because the Edman degradation proceeds from the N-terminus of the protein, it will not work if the N-terminus has been chemically modified (e.g. by acetylation or formation of **pyroglutamic acid**).
- Sequencing will stop if a non-α-amino acid is encountered (e.g. isoaspartic acid), since the favored five-membered ring intermediate is unable to be formed.
- Edman degradation is generally not useful to determine the positions of disulfide bridges.
- It also requires peptide amounts of 1 picomole or above for discernible results.

SEQUENCING OF PROTEINS

DETERMINATIONS

Determination of amino acid composition

- ✓ Amino acid composition and purity must be known before starting sequencing.
- ✓ The polypeptide chains of multimeric proteins should be separated and molecular weight of each chain should be measured.
- ✓ The determination of amino acid is done by hydrolysis, separation & quantitative analysis

DETERMINATIONS

SEQUENCING OF PROTEINS

Hydrolysis

The peptide is hydrolyzed into its constituent amino acids by heating it in 6N HCL at 110° C for 24 hours.

Separation

Separation of protein is done by chromatography, dialysis etc.

DETERMINATIONS

SEQUENCING OF PROTEINS

Quantitative Analysis

The amino acid residue of peptides reacts quantitatively with ninhydrin.

On heating, an α-amino acids reacts with two molecules of ninhydrin to yield an intensely coloured product.

Purple colour is given in this test by all amino acids and peptide having a free α-amino group where as proline gives yellow colour.

MECHANISM

SEQUENCING OF PROTEINS

Mechanism of protein sequencing

- Sanger's method
- Edman's method

Sanger's method

✓Fredrick sanger was a Nobel Laureate.

✓He showed for the first time that amino acids are covalently together by α-amino and α-carboxyl group.

✓He suggested stepwise release and identification of amino acid starting from N-terminal.

SEQUENCING OF PROTEINS

MECHANISM

- ✓He used a reagent fluro-dinitro benzene (FDB)which is commonly called sanger's reagent.
- ✓The FDB reacts with free NH_2 group of N-terminus.
- ✓Upon hydrolysis a yellow coloured dinitrophenol (DNP)-derivative of N- terminal amino acid is produced.
- ✓The DNP amino acid is identified comparing it with a known standard DNP-amino acid by using gel chromatography.

SEQUENCING OF PROTEINS

MECHANISM

Densyl Chloride

- ✓The Densyl chloride is commonly used because it forms an intensively coloured derivatives.
- ✓That can be detected with high sensitivity that the dinitrophenyl compound .
- ✓It reacts with an uncharged C and N terminal to form a sulfoamide derivatives that is the stable under condition that by hydrolyzer peptide bonds.
- ✓Although the densyl method for determining the amino terminal residue is sensitive and powerful.

SEQUENCING OF PROTEINS

DISADVANTAGE

Disadvantage

- ✓ It cannot be used repeatedly on the same peptide because the peptide is totally degraded in the acid hydrolysis step.

SEQUENCING OF PROTEINS

MECHANISM

Edman degradation

- ✓In 1950 Pehr Edman developed a method of protein sequencing.
- ✓It involves sequential identification of amino acids from N to C termini.
- ✓Phenyl isothyocynate (PTC) reagent is used for the Edman degradation.
- ✓The amino terminal N-terminal residue of a protein can be identifies by reaction the protein with the PTC that forms a stable covalent link with the free α-amino group prior to hydrolysis with 6M HCl.

SEQUENCING OF PROTEINS

MECHANISM

Edman degradation

- ✓Phenyl isothiocynate reacts with the uncharged N-terminal amino group of the peptide to form a phenylthiocarbonyl derivation.
- ✓The labeled N-terminal amino acid can be identified by comparison of its chromatographic properties with stanrdard fluorodinitro benzene and dansyl chloride.
- ✓Then, under mild acidic the cyclic (PTH) Phenylthio hydantion of the terminal amino acid is librated which leaves an intact peptide shorted by one amino acid.

SEQUENCING OF PROTEINS

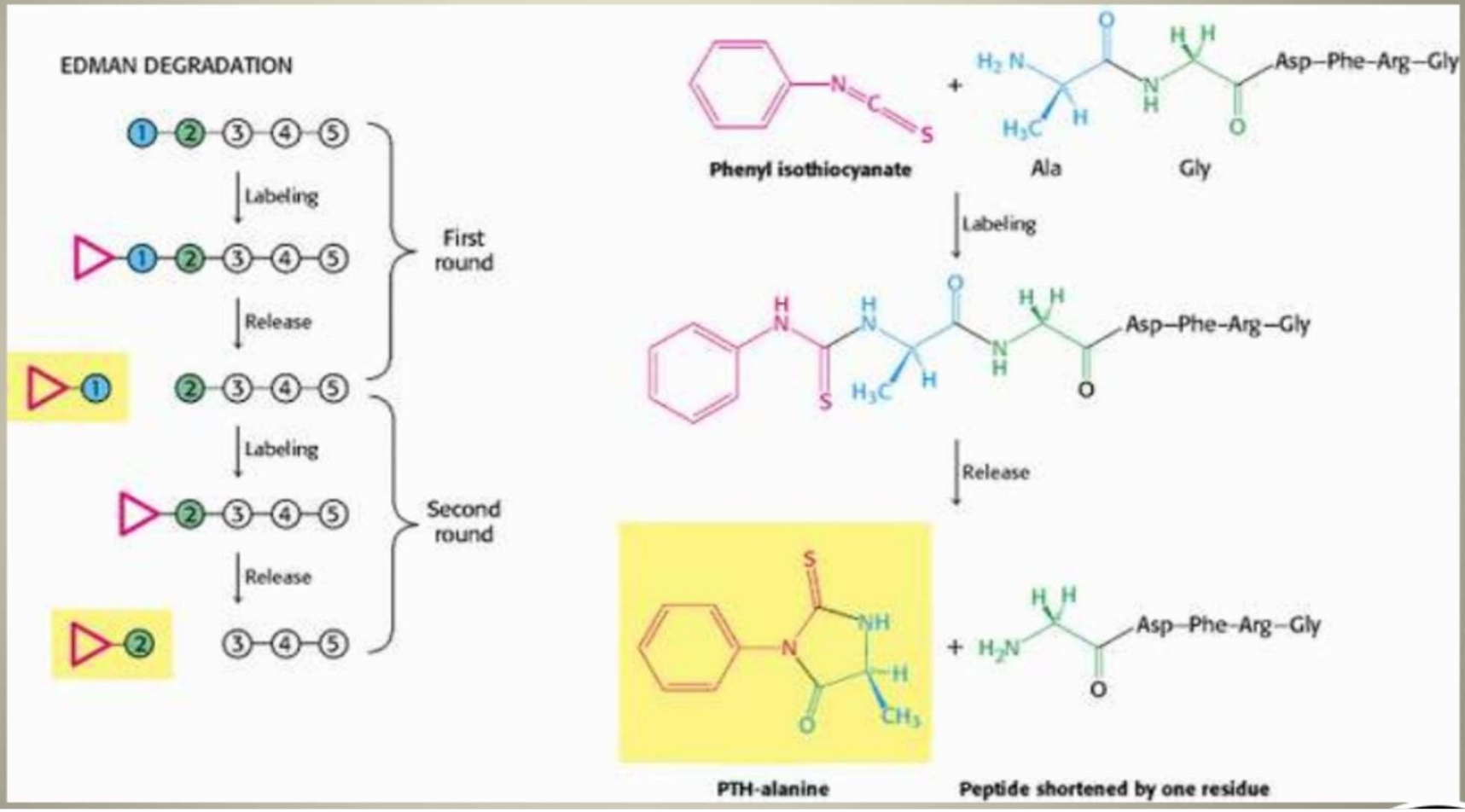
MECHANISM

Edman degradation

- ✓The released PTH amino acid is identified by high performance liquid chromatography.
- ✓The sequencing technique has been automated and refined so that upwards of 50 residues from the N-termines of a protein can be sequenced from picomolequantities of material.

SEQUENCING OF PROTEINS

MECHANISM



SEQUENCING OF PROTEINS

LIMITATION

Limitations of Edman reaction

- ✓The Edman degradation proceeds from the N-terminus of the protein it will not work it the N-terminal amino acid has been chemically modified.

- ✓It also required the use of either guess work or a separate-procedure to determine the position of disulfide bridges.

APPLICATION

SEQUENCING OF PROTEINS

Applications of protein sequencing-

- ✓Knowledge of the sequence of amino acids in a protein can offer insights into its three dimensional structure and its function in cellular location.
- ✓Certain amino acid sequences serve as signals that determine the cellular location chemical modification on an half life of a protein.
- ✓Protein sequences can elucidate the history of life on Earth.

CONCLUSION

SEQUENCING OF PROTEINS

- **Protein sequencing** is a technique to determine the amino acid sequence of a protein
- it also gives information regarding which conformation the protein adopts
- Discovering the structures and functions of proteins in living organisms is an important tool for understanding cellular processes
- It allows drugs that target specific metabolic pathways to be invented more easily.
- The first protein sequencing was achieved by Frederic Sanger in 1953. He determined the amino acid sequence of bovine insulin
- Sanger was awarded the Nobel Prize in 1958.

MASS SPECTROMETRY:

- Mass spectrometry is quickly becoming the **gold standard** by which to identify protein sequences due to its ease of automation and extreme accuracy.
- The use of mass spectroscopy now dominates the process of sequencing proteins .
- The two most popular methods to identify protein sequences using Mass Spectrometry are:

I. **Peptide Mass Fingerprinting/ Protein fingerprinting** - Cleaving an unknown protein into smaller fragments so that these smaller fragments can then be accurately measured with a mass spectrometer.

II. **Tandem Mass spectrometry**- It describes the partitioning of mass spectroscopy into separate steps where fragmentation occurs in between these steps.

Digested the peptide bonds .

By protease enzyme

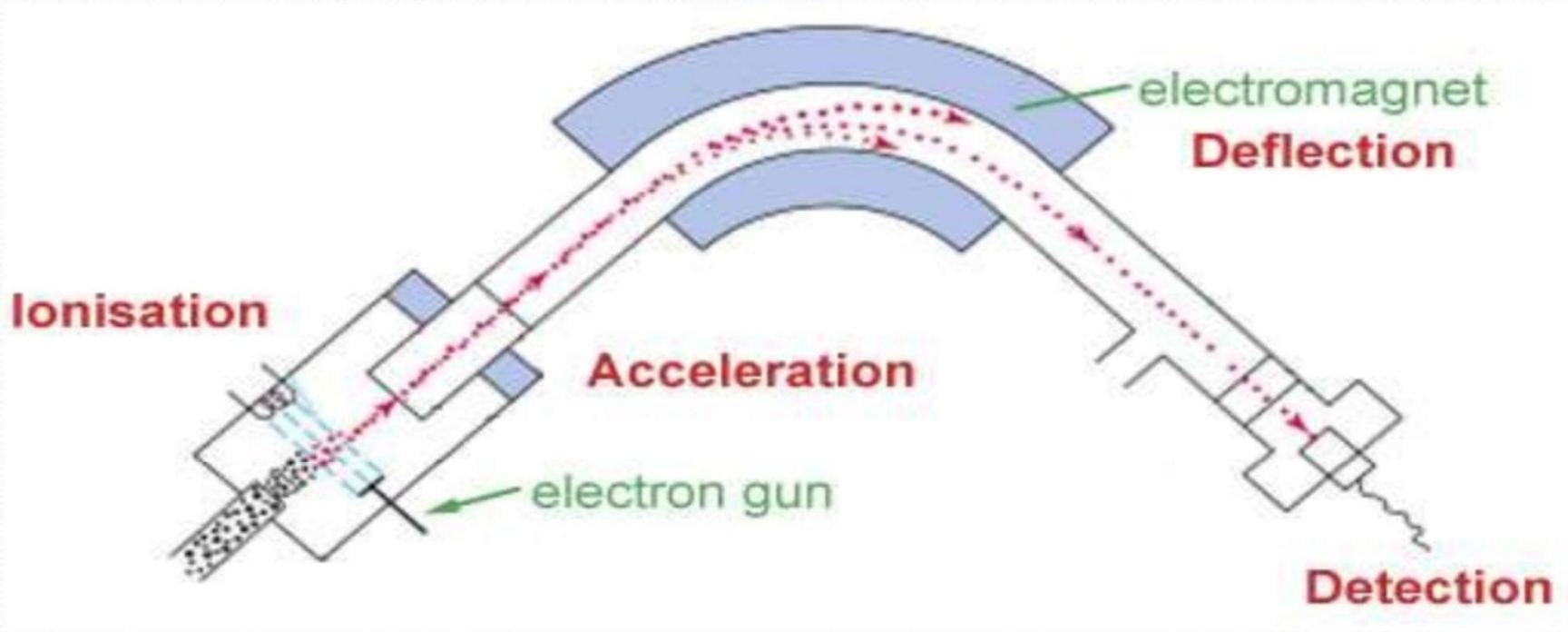
Ionization by MALDI(matrix assisted laser desorption/ionization) or electro spray.

Then ionized molecules are introduced into mass- analyzer.

Using the mass spectrometer the masses of each of these ionized peptides is calculated.

Using the masses it is now possible to know the amino acid sequence of these protein.

Mass spectrometry



A STEPWISE OVERVIEW:

Break and then block disulfide bonds.

Cleave protein into smaller peptides.

Separate the peptides.

Sequence the peptides.

Align the peptide sequences.


THANK YOU