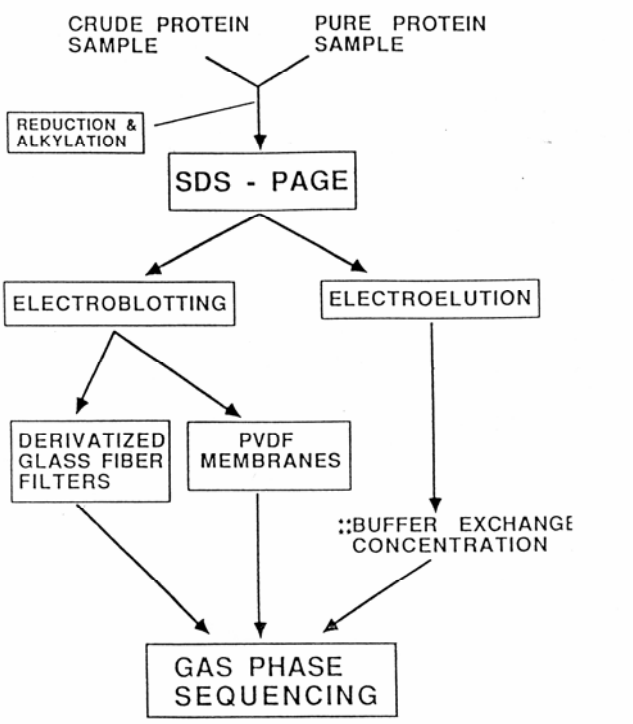
Chapter 4 : Protein and Peptide Purification for Microsequencing

The information of protein sequencing is valuable for the following reasons:

1. Knowledge of a protein's amino acid sequence is prerequisite for determining its **three-dimensional structure**.
2. Sequence comparisons among analogous proteins from different species yield insights into **protein function** and reveal **evolutionary relationships** among the proteins and the organisms that produce them.
3. Many inherited diseases are caused by mutations leading to an amino acid change in a protein. Amino acid sequence analysis can assist in the **development of diagnostic tests and effective therapies**.

Protein versus nucleic acid sequencing

1. Direct protein sequencing can reveal the locations of disulfide bonds in proteins.
2. Many proteins are modified after they are synthesized. For instance, the secretory proteins are usually processed into mature form of proteins before they enter into the lumen area of endoplasmic reticulum (ER). Only protein sequencing can confirm whether and where they actually occur.
3. The inadvertent insertion or deletion of a single nuleotide, a common error, shifts the reading frame of the gene and thereby alters the predicted amino acid sequence from that point on. Therefore, the protein sequence makes it possible to verify the accuracy of the nucleic acid sequence.
4. Even partial sequences of a protein can help to develop oligosynthetic probes to identfy its gene. The partial sequence can also be used to produce antibodies.



Protein sequencing: The Edman reaction

Proteins are sequenced by degradation from their N-terminus using the **Edman reagent**, **phenyl isothiocyanate (PITC)**. The reaction is divided into three steps: coupling, cleavage, and conversion. Today, most laboratories perform the Edman degradation with automated gas-phase sequenators. With these instruments, certain reagents used for the coupling and cleavage steps are delieved as gases. In gas-phase sequenators, the sample is adsorbed to a polybrene-coated glass-fiber disk or electroeluted onto a specially treated glass-fiber filter or a porous polyvinylidene difluoride (PVDF) membrane. The peptide remains bound to the support during the coupling and cleavage steps, which occur in a temperature-controlled reaction chamber.

(1) coupling step:

PITC chemically modifies the x-amino groups. Because an α-amino group is protonated in this pH range, the side chain of a lysine residue is not modified. Coupling is inhibited by N-terminal modifications like acetylation, formylation, and cyclization.

(2) cleavage step:

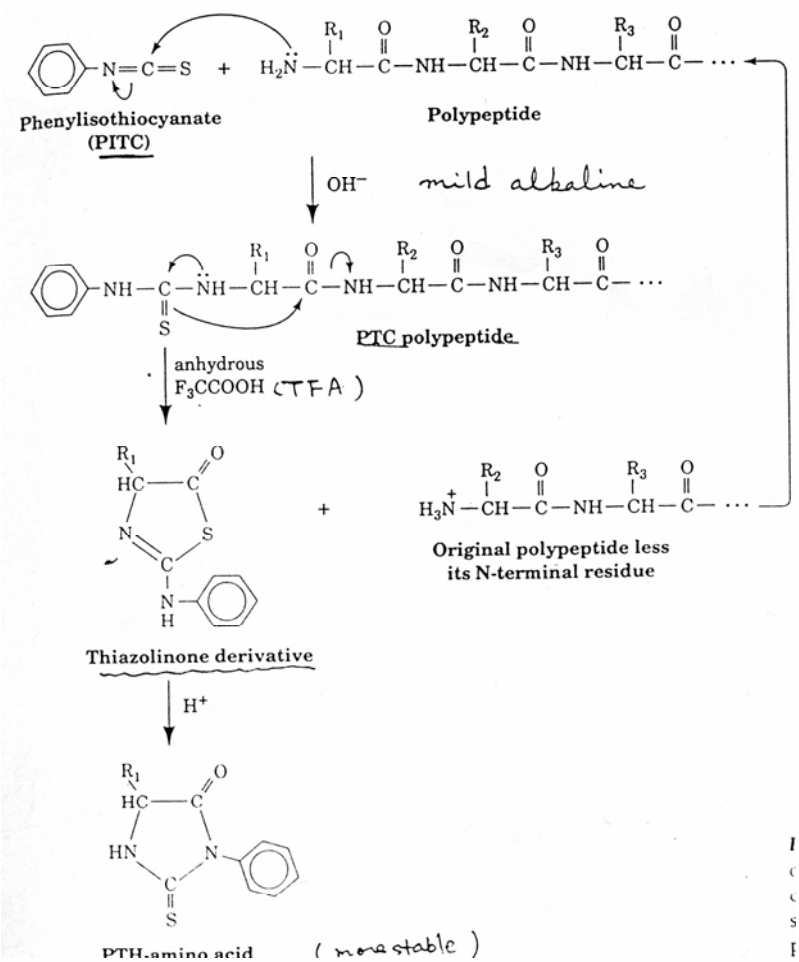
The PTC-N terminal residue is rapidly cleaved from the polypeptide chain under anhydrous, acidic conditions to liberate two products, an ATZ-amino acid and the n-1 polypeptide. The n-1 polypeptide has a reactive N-terminus and can undergo another cycle of coupling and cleavage steps.

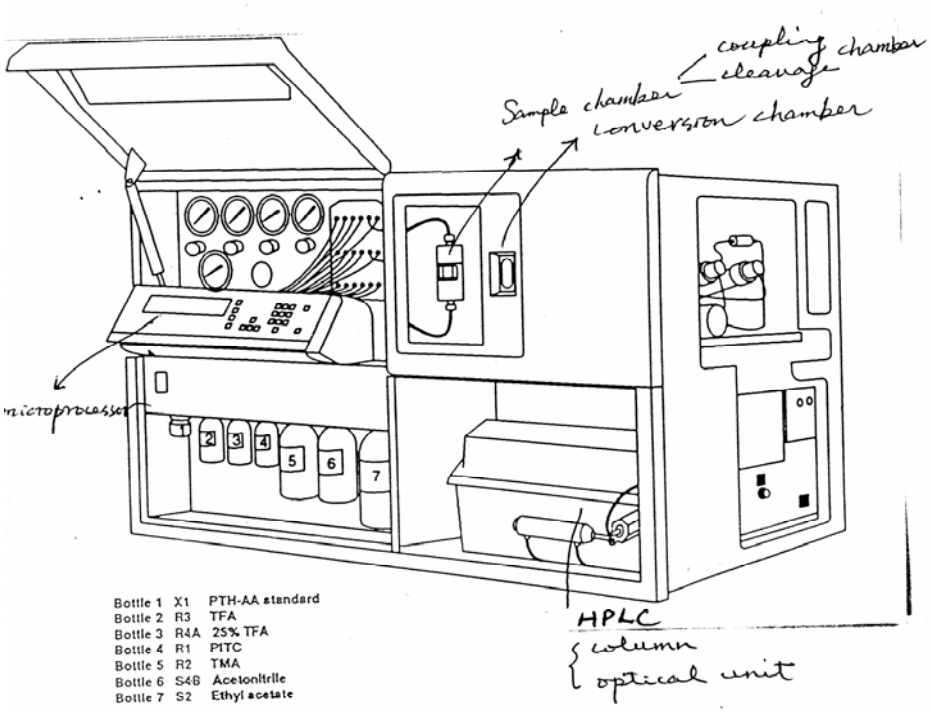
(3) conversion step:

The ATZ-amino acid is extracted from the support and delivered to a small flask. ATZ-derivatives are unstable but can be converted in the third step of the cycle into stable PTH-derivatives. Conversion is a two-step reaction and occurs in aqueous, acidic solutions. First, the ATZ-amino acid is rapidly hydrolyzed to a PTC-amino acid. In the second step, the PTC-amino acid cyclizes to a stable PTH-amino acid.

Identification of PTH-amino acids

The most convenient method for identifying the PTH-amino acids generated during each sequencing cycle is by UV absorbance and HPLC chromatography. All 20 amino acids are easily resolved by gradient elution from a reverse-phase HPLC support. Each amino acid is detected by its UV absorbance at 269 nm and is identified by its characteristic retention time.





Thus, the device is primarily composed of two compartments { Edman degradation
HPLC.

Computer → data analysis

Instrument limitations

The performance of the instrument is measured by

1. The sensitivity limit of the instrument-- the lowest amount of readable sequence (1-10 pmol).
2. The initial yield-- a measure of the percent of the sample loaded on the sequencer that is sequenceable (usually 50-80% unless the N terminus is blocked).
3. The repetitive yield-- the percent of the sequence detected after each cycle of the Edman reaction. This value is measured by a linear regression fit to the amount of PTH-amino acid at each cycle. Extrapolation to cycle zero gives the initial amount of sequenceable peptide. The repetitive yield varies between 80 and 98% and depends on the sequence and size of the peptide and the mode of attachment to the support.

*Longer sequences can be obtained from larger amounts of sample and higher repetitive yields. Theoretically, a 10-pmol peptide can be sequenced with a 95% repetitive yield for approximately 40 cycles before a limit of 1 pmol of sequenceable peptide remains. The same limit is reached for 100 pmol of peptide in 40 cycles with a 90% repetitive yield. However, in practice, shorter lengths of sequence are obtained because background peaks generated by spurious cleavage of the peptide chain and by lagging sequences generated by incomplete cleavage or coupling reactions will obscure the peak derived from the N-terminal residue. **One could reasonably expect 10-15 cycles of sequence in most cases from 1-5 µg of sample**.

Chemical limitations

To avoid problems with the Edman chemistry, the gas-phase samples should be free of the following reagents:

1. Buffers and primary amines. Tris buffer is commonly used for protein purification. Tris and glycine are common in samples recovered from SDS-PAGE. Therefore, CAPS buffer is used for electroblotting instead of Tris/Glycine buffer.
2. Glycerol or sucrose. These reagents are often added to buffers designed for the storage and handling of proteins. These compounds are not volatile and leave a highly viscous residue.
3. Nonionic detergents. Triton X-100, Brij. and Tween solutions often contain aldehydes, oxidants, and other contaminants that can inhibit the Edman degradation.
4. SDS. Large quantities of SDS can cause instrument malfunctions and may lead to the loss of sample from the filter.
5. Cyanate and aldehydes are common contaminants that may react with N-termini of proteins. It should be noted that ammonium cyanate is generated in significant

quantities in the concentrated alkaline-urea solutions that are often used in protein purification.

6. All reagents and solvents must be of the highest purity (HPLC-grade solvents, sequencing-grade, and electrophoresis-grade reagents), available in order to avoid nonvolatile contaminants that may leave interfering residues upon drying. One should **avoid "molecular biology"-grade reagents**, since these often contain insoluble material and UV-absorbing contaminants.
7. Dialysis tubing is often a source of contaminants and other interfering substances. Avoid dialysis as a last step in sample preparation or use thoroughly cleaned, high-quality tubing. Do not dialyze the protein against water because the contaminants will remain absorbed to the protein and dialysis membrane in the absence of salt or acid.

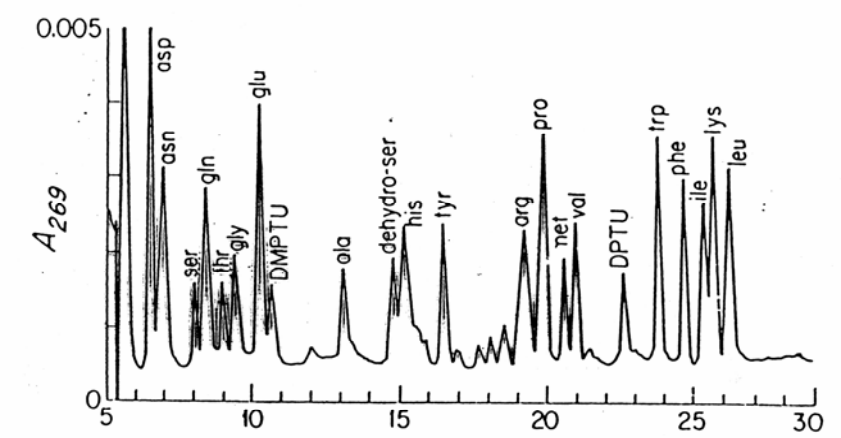
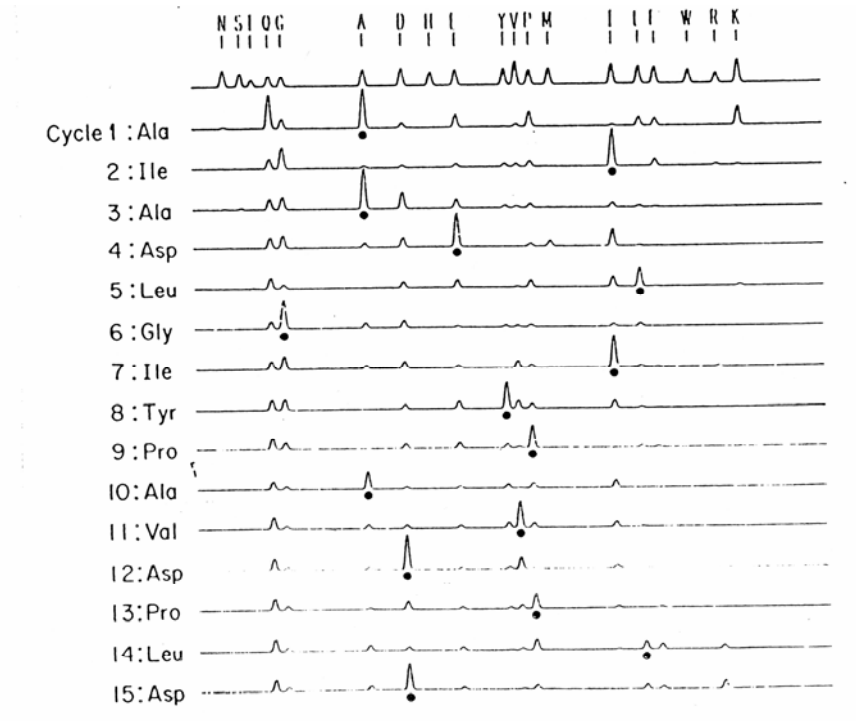


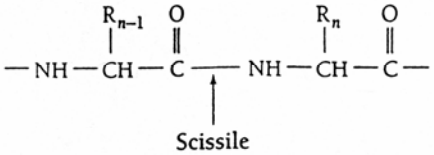
Figure 3 The elution pattern of the PTH-amino acid (15 pmol) standards. DMPTU and DPTU are by-products of the sequencing reaction and serve as useful reference peaks in the chromatograms.



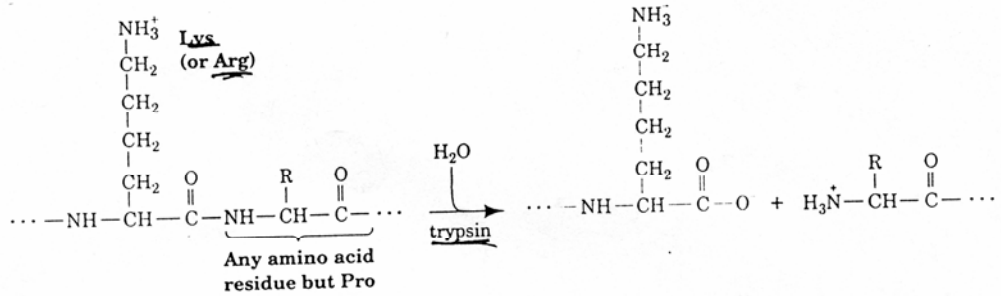
Polypeptide cleavage

* When sequencing of a polypeptide is blocked, the protein must be cleaved, either enzymatically or chemically, to specific fragments before sequencing.
* Various endopeptidases can be used to fragment polypeptides. These include trypsin, chymotrypsin, endopeptidase V8...etc.
* Several chemical reagents promote peptide bond cleavage process at specific residues. The most useful of these, cyanogen bromide (CNBr), cleaves on the C side of Met residues.

Table 5-5. **Specificities of Various Endopeptidases**



Enzyme	*Source*	*Specificity*	*Comments*
Trypsin	Bovine pancreas	R_{n-1} = positively charged residues: Arg, Lys; $R_n \neq$ Pro	Highly specific
Chymotrypsin	Bovine pancreas	R_{n-1} = bulky hydrophobic residues: Phe, Trp, Tyr; $R_n \neq$ Pro	Cleaves more slowly for R_{n-1} = Asn, His, Met, Leu
Elastase	Bovine pancreas	R_{n-1} = small neutral residues: Ala, Gly, Ser, Val; $R_n \neq$ Pro	
Thermolysin	*Bacillus thermoproteolyticus*	R_n = Ile, Met, Phe, Trp, Tyr, Val; $R_{n-1} \neq$ Pro	Occasionally cleaves at R_n = Ala, Asp, His, Thr; heat stable
Pepsin	Bovine gastric mucosa	R_n = Leu, Phe, Trp, Tyr; $R_{n-1} \neq$ Pro	Also others; quite nonspecific: pH optimum = 2
Endopeptidase V8	*Staphylococcus aureus*	R_{n-1} = Glu	



Reconstructing the protein's sequence

* This is accomplished by conducting more than one round of protein cleavage with a reagent of different specificity and then comparing the amino acid sequences of the overlapping sets of peptide fragments.

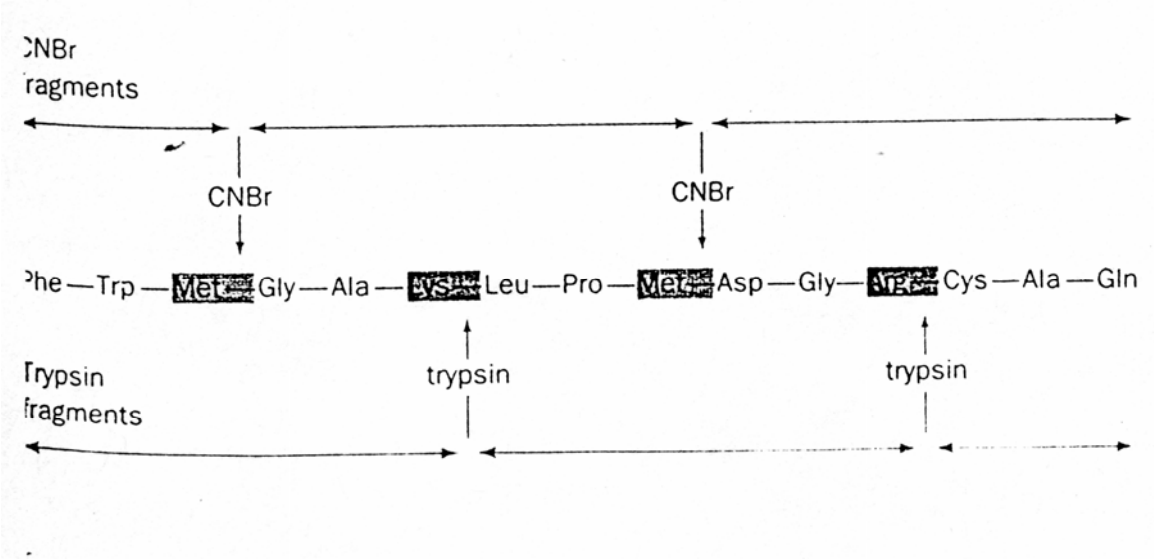


Figure 5-16. **Generating overlapping fragments to determine the amino acid sequence of a polypeptide.** In this example, two sets of overlapping peptide fragments are made by using trypsin to cleave the polypeptide after all its Arg and Lys residues and, in a separate reaction, using CNBr to cleave it after all its Met residues.