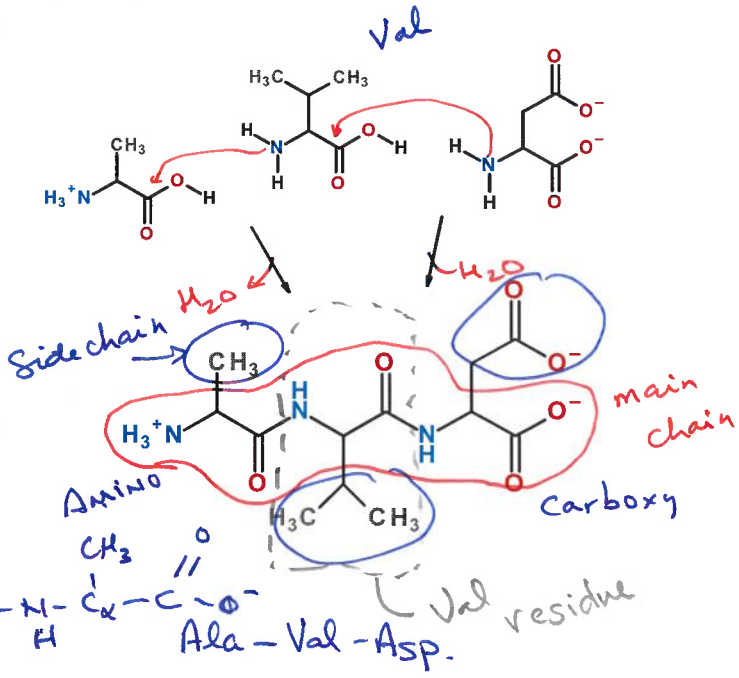
Lecture 6: Primary Structure & Protein Sequencing.

1. **Primary structure (1°):** The amino acid sequence.
2. **Secondary structure (2°):** Configuration of mainchain atoms – local structure.
3. **Tertiary structure (3°):** Complete structure of one chain – main and sidechain atoms.
4. **Quaternary structure (4°):** Association of subunits.

Expectations:

- Draw chemical structure given the sequence.
- Determine the seq. from chem. structure.
- Distinguish/identify:
  - Mainchain atoms,
  - Sidechain atoms,
  - Residue = aa in polymer,
  - Amino terminus,
  - Carboxy terminus,
  - Peptide bond(s).
- Determine seq. from mass spec. data.
- Determine seq. from Edman degradation data.



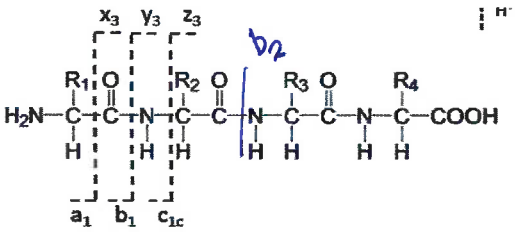
Example: Draw the structure of Gly-Ala-Ala

i) draw mainchain atoms first.

ii) add sidechains

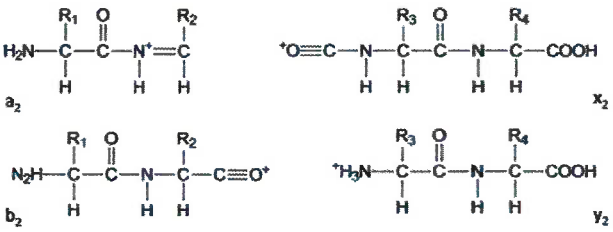
Protein sequence can be obtained by:

- DNA, mRNA sequencing
- Mass spectrometry (fragmentation)
- Chemical degradation (Edman).



Mass Spectrometry:

- peptides are fragmented by bombardment with ions.
- a, b, and y ions are the most commonly produced
- the masses of each fragment can be accurately determined (to within one Dalton) and can be used to determine the sequence, due to differences in the mass of the sidechain. Using the b ions as an example



b_1 = mass of 1st residue + 1 (amino proton)
b_2 = mass of 1st + 2nd residue + 1
b_3 = mass of 1st + 2nd + 3rd residue + 1

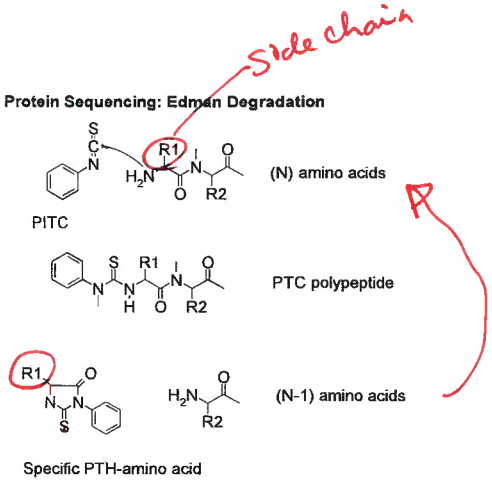
- approximately 20 residues can be sequenced.

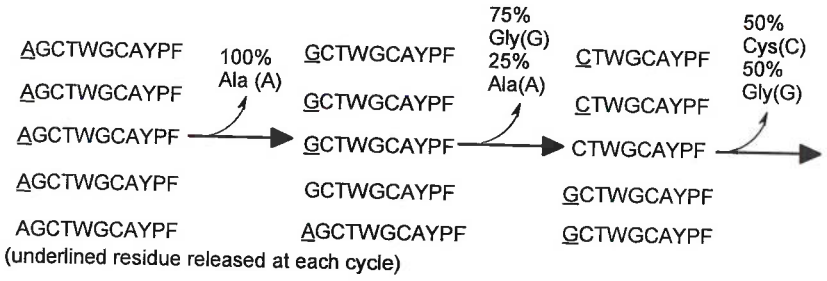
Edman Degradation: Cleaves single amino acids from the amino-terminus. Producing:

i) The PTH derivative of the released amino acid can be identified.

ii) A peptide that is one residue shorter is produced, this can be treated with PITC to obtain the next residue, etc.

iii) Errors accumulate because the release of the PTH-amino acid is not 100%, limiting the length of sequence to ~75 residues:





70 - 100 Cycles Typical

Cleavage/Fragmentation: Because it is not possible to sequence proteins larger than ~75 residues it is necessary to fragment the protein to extend the sequence information if the protein is longer than 75 residues. After cleavage, the individual peptide fragments are separated from each other and each is subject to N-terminal sequencing using the Edman degradation method.

You need to know these cleavage patterns: Common

Chem.

a. Cyanogen bromide (CNBr) cleaves the peptide bond after Methionine residues. You do not need to know the mechanism, just the specificity of the reaction.

$$Ser-Met-Gly-Ala-Phe-Arg-Leu-Ile \xrightarrow{CNBr} Ser-Met \quad + \quad Gly-Ala-Phe-Arg-Leu-Ile$$

Enz

b. Chymotrypsin hydrolyzes the peptide bonds that follow large hydrophobic residues, *e.g.* Phenylalanine, Tyrosine, Tryptophan. You should remember this cleavage pattern.

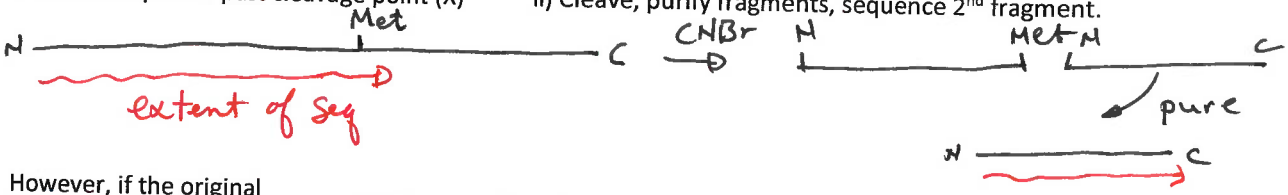
$$Ser-Met-Gly-Ala-Phe-Arg-Leu-Ile \xrightarrow{Chymotrypsin} Ser-Met-Gly-Ala-Phe \quad + \quad Arg-Leu-Ile$$

c. Trypsin hydrolyzes the peptide bonds that follow positively charged residues, *e.g.* Lysine and Arginine. You should remember this cleavage pattern as well.

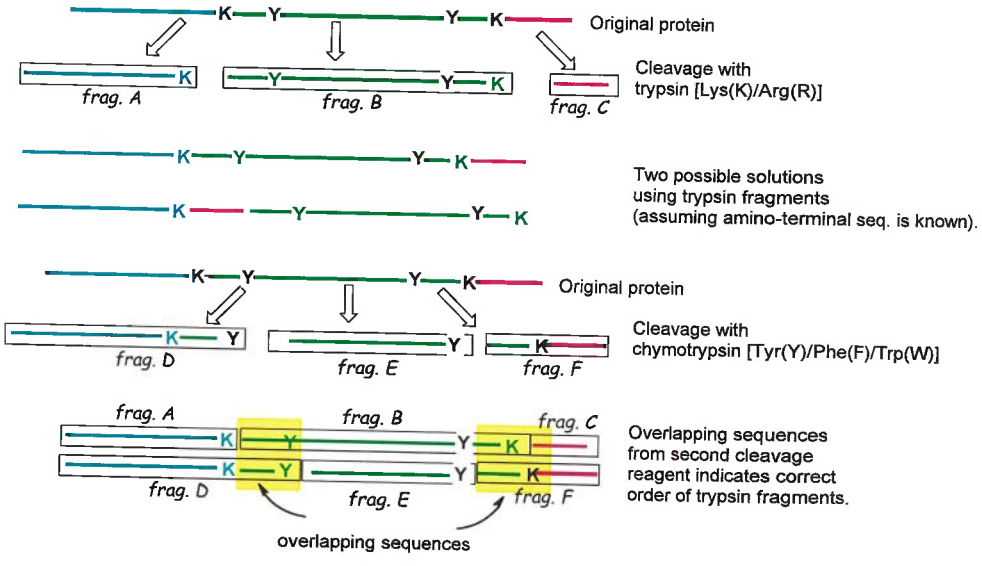
$$Ser-Met-Gly-Ala-Phe-Arg-Leu-Ile \xrightarrow{Trypsin} Ser-Met-Gly-Ala-Phe-Arg \quad + \quad Leu-Ile$$

Sequence Strategies: If only two fragments are produced by the cleavage reaction, then it is straightforward to reconstruct the sequence using the known amino terminal sequence of the original protein to determine which fragment is first.

i) Obtain sequence past cleavage point (X) ii) Cleave, purify fragments, sequence 2nd fragment.



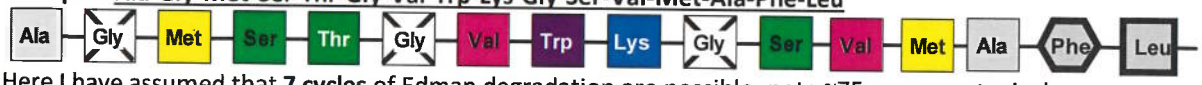
However, if the original protein is cleaved into three or more fragments, or they cannot be sequenced in their entire length, then it is not possible to determine the correct order of fragments using a single cleavage agent. Multiple overlapping fragments have to be used to determine the correct ordering:



Sequence alignment strategy: Find overlaps between fragments obtained with different cleavage reagents & use these overlaps to correctly *order* the peptides obtained from one sequencing reaction.

The overlaps can be readily identified by finding a cleavage site in a peptide that would be cut by another cleavage reagent and then identifying the correct fragment based on the expected amino-terminal sequence.

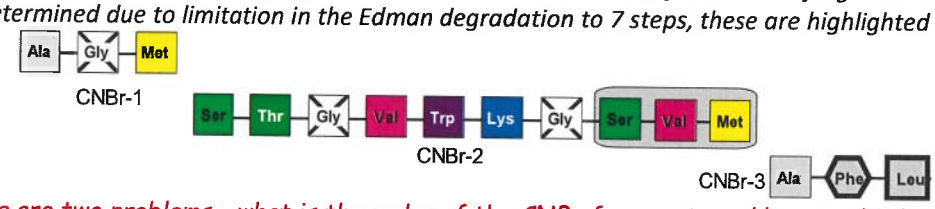
Example: Ala-Gly-Met-Ser-Thr-Gly-Val-Trp-Lys-Gly-Ser-Val-Met-Ala-Phe-Leu



Here I have assumed that **7 cycles** of Edman degradation are possible; note ~75 are more typical.

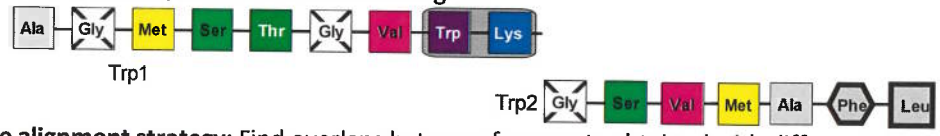
A. The first 7 cycles of Edman sequencing on the intact peptide released the following amino acids, in this order: Ala, Gly, Met, Ser, Thr, Gly, Val, therefore , the initial, amino-terminal sequence of the peptide is:

B. A **new** sample of the intact peptide was treated with *CNBr*. The three peptides (CNBr-1, CNBr-2, CNBr-3) were produced. These were isolated and *each* was subject to sequencing in separate reactions, giving the following sequences: *The last three residues of the second fragment could not be determined due to limitation in the Edman degradation to 7 steps, these are highlighted gray.*



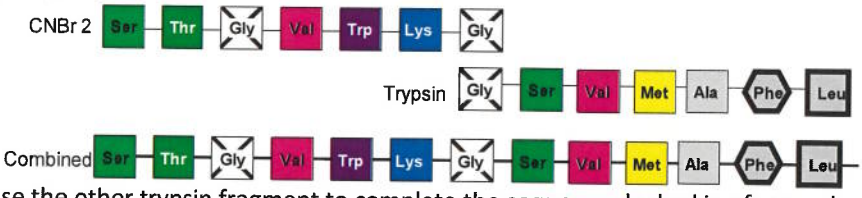
There are two problems - what is the order of the CNBr fragments and how to obtain the missing sequence in the 2nd CNBr fragment (Ser-Val-Met). Both can be solved by obtaining sequence data using another cleavage reagent and overlapping common sequences.

C. A new sample of the peptide was treated with *Trypsin*. The two peptides (Trp1, Trp2) that were produced were isolated and *each* was sequenced, giving only the first 7 residues of the first peptide and all of the second, since it is 7 residues long.

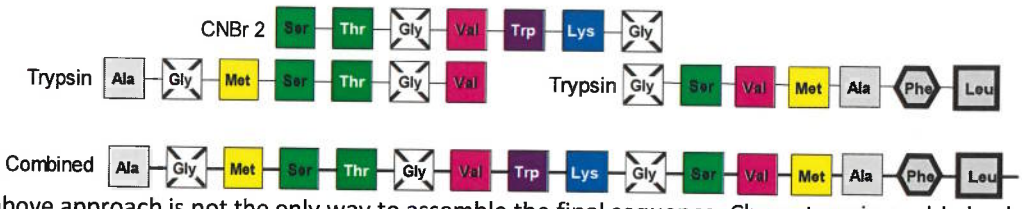


Sequence alignment strategy: Find overlaps between fragments obtained with different cleavage reagents & use these overlaps to correctly *order* the peptides obtained from one sequencing reaction. ***The overlaps can be readily identified by finding a cleavage site in a peptide that would be cut by another cleavage reagent and then identifying the correct fragment based on the expected amino-terminal sequence.***

A) The second CNBr fragment has a lysine residue, indicating that there should be a trypsin fragment that begins with Gly and this would overlap the two CNBr fragments, giving the correct order of the CNBr fragments.



B) We can use the other trypsin fragment to complete the sequence, by looking for overlaps between it and the CNBr-2 fragment:



The above approach is not the only way to assemble the final sequence. Chymotrypsin could also be used, taking advantage of the Trp residue in the 2nd CNBr fragment.

You are trying to determine the sequence of a small protein using Edman degradation.

For the entire, intact protein, you obtain.

Ala-Gly-Thr-Ser-Ala-Glu-Met-Asp-Asn-Ala

You treat the protein with cyanogen bromide, purify the individual fragments, and sequence each fragment to obtain the following:

CNBr 1: Ala-Gly-Thr-Ser-Ala-Glu-Met

CNBr 2: Asp-Asn-Ala-Leu-Val-Ser-Thr Lys-Ala-Gly

CNBr 3: Ala-Gly-Ser-Thr-Val-Leu-Trp-Leu-Gly-Pro

You treat a new sample of the intact protein with trypsin, purify the individual fragments, and sequence each fragment to obtain the following:

T1: Ala-Gly-Thr-Ser-Ala-Glu-Met-Asp-Asn-Ala

T2: Ala-Gly-Met-Ala-Gly-Ser-Thr-Val-Leu-Trp

You treat a new sample of the intact protein with chymotrypsin, purify the individual fragments, and sequence each fragment to obtain the following:

C1: Ala-Gly-Thr-Ser-Ala-Glu-Met-Asp-Asn-Ala

C2: Leu-Gly-Pro-Gly-Ile-Asp-Glu-Gln

Reconstruct as much of the sequence as possible (Hint: Find overlaps by finding cleavage sites for other reagents within a fragment, e.g. use the chymotrypsin cleavage sites within the CNBr fragments to align the chymotrypsin fragments with the CNBr fragments.).

CNBr 3: Ala-Gly-Ser-Thr-Val-Leu-Trp-Leu-Gly-Pro

C2: Leu-Gly-Pro-Gly-Ile-Asp-Glu-Gln

T2: Ala-Gly-Met-Ala-Gly-Ser-Thr-Val-Leu-Trp

CNBr 2: Asp-Asn-Ala-Leu-Val-Ser-Thr-Lys-Ala-Gly

Ala-Gly-Thr-Ser-Ala-Glu-Met-Asp-Asn-Ala *(from sequencing the entire peptide)*

Final sequence, assembled from the sequences of fragments.

Ala-Gly-Thr-Ser-Ala-Glu-Met-Asp-Asn-Ala-Leu-Val-Ser-Thr Lys-Ala-Gly-Met-Ala-Gly-Ser-Thr-Val-Leu-Trp-Leu-Gly-Pro-Gly-Ile-Asp-Glu-Gln