C. Ionization Properties of Amino Acids.

Ionizable Amino Acids/Protonated - deprotonated forms

- Carboxy-Terminus \( pK_a = 2 \)
- Amino-Terminus \( pK_a = 9 \)

Protonated (pH < \( pK_a \))

- Aspartic Acid (Asp) \( pK_a = 4 \)
- Glutamic Acid (Glu) \( pK_a = 4 \)

Deprotonated (pH > \( pK_a \))

- Histidine (His) \( pK_a = 6 \)
- Lysine (Lys) \( pK_a = 9 \)
- Arginine (Arg) \( pK_a = 12 \)

Basic amino acids

Other sidechain ionizations that are less important:
- Tyr-OH \( pK_a = 10 \), Cys-SH, \( pK_a = 8 \)

\[ \text{pH} = 7.0 \]

Which groups don't ionize at physiological pH ranges?

- Ser-OH
- Thr-OH

Hydroxyl

Amino acids with unique properties:

- Cysteine (Cys)
  - \( pK_a = 8 \)
- Proline (Pro)
  - no NH in protein
Lecture 6: UV Absorption of Amino Acids & Primary Structure

Goals:
- Calculate protein concentration from absorbance at 280 nm.
- Determine sequence of proteins by chemical methods and enzymatic cleavage.

UV Absorption Properties of Amino Acids
Three aromatic amino acids absorb light in the ultraviolet range (UV).

The extinction coefficients (or molar absorption coefficients) of these amino acids are:

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Extinction Coefficient $\varepsilon_{\lambda_{\text{MAX}}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp</td>
<td>5,500 M$^{-1}$ cm$^{-1}$ (280 nm)</td>
</tr>
<tr>
<td>Tyr</td>
<td>1,490 M$^{-1}$ cm$^{-1}$ (274 nm - 280 nm)</td>
</tr>
<tr>
<td>Phe</td>
<td>$\approx$ 0 M$^{-1}$ cm$^{-1}$ (280 nm)</td>
</tr>
</tbody>
</table>

The amount of light absorbed by a solution of concentration $[X]$ is given by the Beer-Lambert Law:

$$ A = \log \frac{I_0}{I} = \varepsilon [X] l $$

where
- $A$ is the absorbance of the sample;
- $I_0$ is the intensity of the incident light;
- $I$ is the intensity of the light that leaves the sample.
- $\varepsilon$ is the molar extinction coefficient at a specific wavelength, e.g. at $\lambda_{\text{MAX}}$.
- $[X]$ is the concentration of the absorbing species
- $l$ is the path length (usually 1 cm).

Therefore, given an extinction coefficient it is possible to measure the concentration of a protein.

**Calculation of molar extinction coefficients**: If a molecule contains a mixture of $N$ different chromophores, the molar extinction coefficient can generally be calculated as the sum of the molar extinction coefficient for each absorbing group in the protein:

$$ \varepsilon_{\text{Protein}} = \sum_{i=1}^{N} \left( n_i \varepsilon_{\text{Trp}} \cdot \varepsilon_{\text{Trp}} + \left( n_i \varepsilon_{\text{Tyr}} \cdot \varepsilon_{\text{Tyr}} \right) \right) $$

Therefore, the molar extinction coefficient for a protein can be calculated from its amino acid composition.

**Example**:

i) A protein has two Tryptophan (Trp) residues and one Tyrosine (Tyr) what is its extinction coefficient?

$$ \varepsilon_{\text{protein}} = 2 \times 5500 \, \text{M}^{-1} \, \text{cm}^{-1} + 1 \times 1490 \, \text{M}^{-1} \, \text{cm}^{-1} = 12490 \, \text{M}^{-1} \, \text{cm}^{-1} $$

ii) The absorption of a solution of the protein is 0.5 with a path length of 1 cm. What is the concentration of the protein?

$$ A = \varepsilon [X] l $$

$$ 0.5 = 12490 \, \text{M}^{-1} \, \text{cm}^{-1} \times [X] \times 1 \, \text{cm} $$

$$ [X] = \frac{0.5}{12490} \approx 0.4 \times 10^{-5} \, \text{M} $$
Protein Structural Hierarchy:
1. **Primary structure** (1°): The amino acid sequence.

**Primary Structure**
- Mainchain atoms
- Sidechain atoms
- Amino terminus
- Carboxy terminus
- Peptide bond(s)
- Sequence (N→C)

**Example:** Draw the structure of Gly-Ala-Ala
   i) draw mainchain atoms first.
   ii) add sidechains

**Sequencing Proteins** - The protein sequence can be obtained by:
- DNA sequencing
- mRNA sequencing
- Mass spectrometry (fragmentation)
- Chemical degradation (Edman).

We will focus on N-terminal sequencing using Edman degradation coupled with fragmentation of the peptide in the case of larger proteins.

**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:
**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. After cleavage, the individual peptide fragments are separated from each other and each is subject to N-terminal sequencing using the Edman degradation method.

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.

\[
\text{Ser} - \text{Met} - \text{Gly} - \text{Ala} - \text{Phe} - \text{Arg} - \text{Leu} - \text{Ile} \xrightarrow{\text{CNBr}} \text{Ser} - \text{Met} + \text{Gly} - \text{Ala} - \text{Phe} - \text{Arg} - \text{Leu} - \text{Ile}
\]

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

\[
\text{Ser} - \text{Met} - \text{Gly} - \text{Ala} - \text{Phe} - \text{Arg} - \text{Leu} - \text{Ile} \xrightarrow{\text{Chymotrypsin}} \text{Ser} - \text{Met} - \text{Gly} - \text{Ala} - \text{Phe} + \text{Arg} - \text{Leu} - \text{Ile}
\]

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

\[
\text{Ser} - \text{Met} - \text{Gly} - \text{Ala} - \text{Phe} - \text{Arg} - \text{Leu} - \text{Ile} \xrightarrow{\text{Trypsin}} \text{Ser} - \text{Met} - \text{Gly} - \text{Ala} - \text{Phe} - \text{Arg} + \text{Leu} - \text{Ile}
\]

**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.
An 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 just the first 7 residues:

B. A new sample of the intact peptide was treated with CNBr. The three peptides (CNBr-1, CNBr-2, CNBr-3) that were produced were isolated and each was subject to sequencing, 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 (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.

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. Also note that the amino-terminal sequence is enough to actually order the two CNBr fragments (since it crosses a Met residue), but that information was not used in the above example.