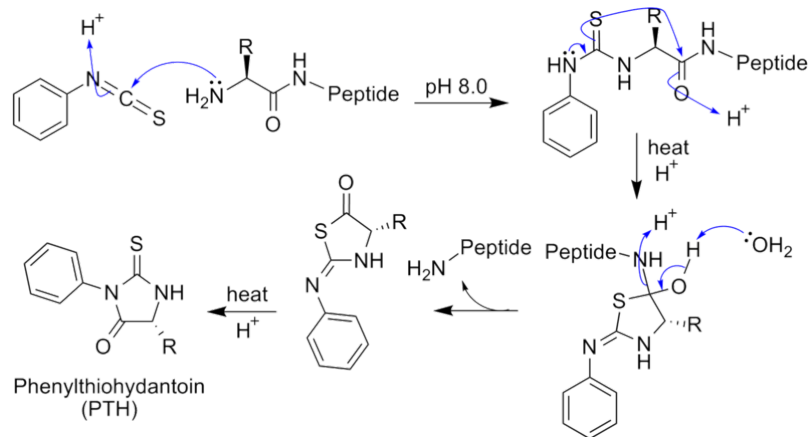


2.2: Protein Sequencing

Edman degradation is the method of sequencing amino acids in a peptide by sequentially removing one residue at a time from the amino end of a peptide. To solve the problem of damaging the protein by hydrolyzing conditions, Pehr Edman created a new way of labeling and cleaving the peptide. Edman thought of a way of removing only one residue at a time, which did not damage the overall sequencing. This was done by adding **Phenyl isothiocyanate (Edman's reagent)**, which creates a phenylthiocarbamoyl derivative with the N-terminal. The N-terminal is then cleaved under less harsh acidic conditions, creating a cyclic compound of phenylthiohydantoin PTH-amino acid. This does not damage the protein and leaves two constituents of the peptide. This method can be repeated for the rest of the residues, separating one residue at a time.



Advantage

Sensitive, simple, and inexpensive method of sequencing amino acids.

Edman degradation is very useful because it does not damage the protein.

The technique allows sequencing of the protein to be done in less time.

Limitations

The technique lacks high-throughput capabilities as sequencing proceeds on samples of single proteins only.

Additional Reading

Sequencing Larger Proteins

Larger proteins cannot be sequenced by the Edman sequencing because of the less than perfect efficiency of the method. A strategy called divide and conquer successfully cleaves the larger protein into smaller, practical amino acids. This is done by using a certain chemical or enzyme which can cleave the protein at specific amino acid residues. The separated peptides can be isolated by chromatography. Then they can be sequenced using the Edman method, because of their smaller size.

In order to put together all the sequences of the different peptides, a method of overlapping peptides is used. The strategy of divide and conquer followed by Edman sequencing is used again a second time, but using a different enzyme or chemical to cleave it into different residues. This allows two different sets of amino acid sequences of the same protein, but at different points. By comparing these two sequences and examining for any overlap between the two, the sequence can be known for the original protein.

For example, trypsin can be used on the initial peptide to cleave it at the carboxyl side of arginine and lysine residues. Using trypsin to cleave the protein and sequencing them individually with Edman degradation will yield many different individual results. Although the sequence of each individual cleaved amino acid segment is known, the order is scrambled. Chymotrypsin, which cleaves on the carboxyl side of aromatic and other bulky nonpolar residues, can be used. The sequence of these segments overlap with those of the trypsin. They can be overlapped to find the original sequence of the initial protein. However, this method is

limited in analyzing larger sized proteins (more than 100 amino acids) because of secondary hydrogen bond interference. Other weak intermolecular bonding such as hydrophobic interactions cannot be properly predicted. Only the linear sequence of a protein can be properly predicted assuming the sequence is small enough.

Mass spectrometry has been replacing traditional methods to determine the molecular mass and structure of a protein. Its power comes from its exquisite sensitivity and modern computational methods to determine structure through comparisons of ion fragment data with computer databases of known protein structures.

In mass spectrometry, a molecule is **first ionized** in an ion source. Sample introduction into the ion source occurs through simple diffusion of gases and volatile liquids from a reservoir, by injection of a liquid sample containing the analyte by spraying a fine mist, or for very large proteins by desorbing a protein from a matrix using a laser. The charged particles are then accelerated by an electric field into a **mass analyzer** where they are subjected to an external magnetic field. The external magnetic field interacts with the magnetic field arising from the movement of the charged particles, causing them to deflect. The deflection is proportional to the mass to charge ratio, m/z . Ions then enter the **detector** which is usually a photomultiplier. Analysis of complex mixtures is done by coupling HPLC with mass spectrometry in a LCMS. (<http://www.chm.bris.ac.uk/ms/theory/...onisation.html>)

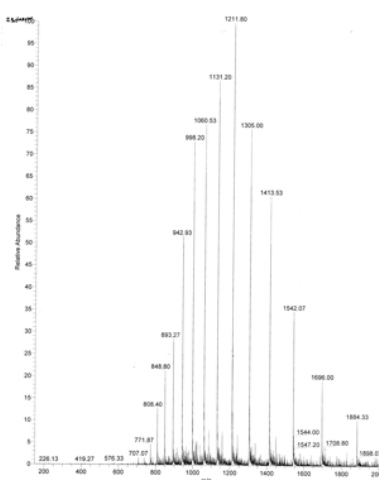
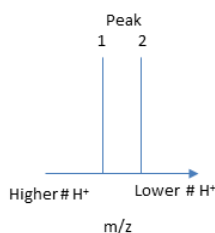


Figure 2.3.1:
ESI Mass Spectrum of Apo-Myoglobin

The molecular mass of the protein can be determined by analyzing two adjacent peaks, as shown in the figure below.



If M is the molecular mass of the analyte protein, and n is the number of positive charges on the protein represented in a given m/z peak, then the following equations give the molecular mass M of the protein for each peak:

$$M_{\text{peak2}} = n(m/z)_{\text{peak2}} - n(1.008) \quad (1)$$

$$M_{\text{peak1}} = (n+1)(m/z)_{\text{peak1}} - (n+1)(1.008) \quad (2)$$

where 1.008 is the atomic weight of H. Since there is only one value of M , the two equations can be set equal to each other, giving:

$$n(m/z)_{\text{peak2}} - n(1.008) = (n+1)(m/z)_{\text{peak1}} - (n+1)(1.008) \quad (3)$$

Solving for n gives:

$$n = [(m/z)_{\text{peak1}} - 1.008] / [(m/z)_{\text{peak2}} - (m/z)_{\text{peak1}}] \quad (4)$$

Knowing n , the molecular mass M the protein can be calculated for each m/z peak. The best value of M can then be determined by averaging the M values determined from each peak (16,956 from the above figure). For peaks from m/z of 893-1542, the calculated values of n ranged from +18 to +10.

Advantage

Protein sequencing by MS has become an invaluable tool in the field of **proteomics**. What we know today about protein structure, function, modification and global protein dynamics comes from the development of high-throughput MS proteomics workflows.

Contributors

- [Prof. Henry Jakubowski](#) ([College of St. Benedict/St. John's University](#))
- [Dr. Dietmar Kennepohl](#) FCIC (Professor of Chemistry, [Athabasca University](#))
- Wikibooks (Structural Biopchemistry). The content on this page is licensed under a CC-SA-BY 3.0 license.

2.2: Protein Sequencing is shared under a [CC BY-SA](#) license and was authored, remixed, and/or curated by LibreTexts.