

Biochemistry IN THE LAB

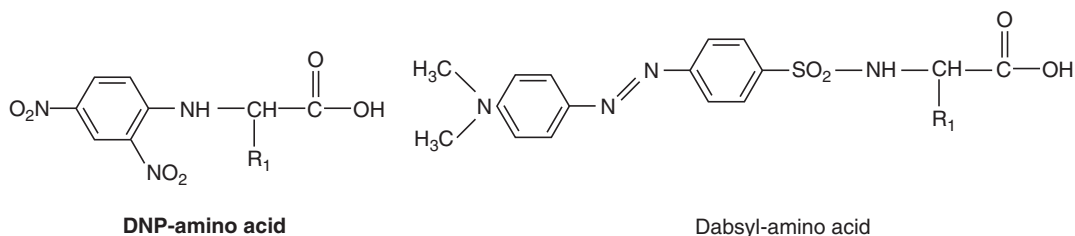
Protein Sequence Analysis

The first step in protein sequence analysis is to determine how many of each type of amino acid residue are present in the molecule. The process for obtaining this information, referred to as the *amino acid composition*, begins with the complete hydrolysis of all peptide bonds. Hydrolysis is typically accomplished with 6 N HCl for 10 to 100 hours. Long reaction times are required because of difficulties in the hydrolysis of three aliphatic amino acids (Leu, Ile, and Val). Hydrolysis is followed by analysis of the resulting amino acid mixture, referred to as the *hydrolysate*. Because of the vigorous conditions of acid hydrolysis, several amino acids (Trp, Gln, Ser, Thr, Tyr, and cystine) are degraded. The concentrations of these molecules in the protein are determined by alternate means.

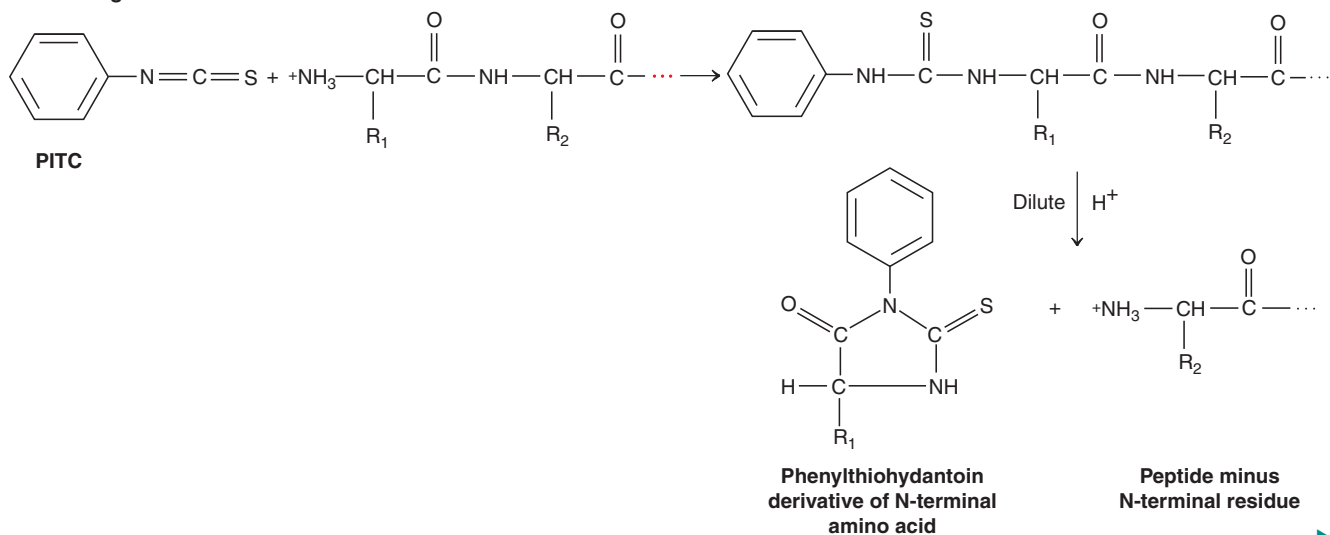
Currently, most protein hydrolysates are analyzed by automated high-pressure liquid chromatography (HPLC). In HPLC, after the hydrolysate has been treated with compounds such as Edman's reagent (described shortly), the products are forced at high pressure through a stainless steel column packed with a stationary phase. Each amino acid derivative is identified according to its retention time on the column. Amino acid analysis by HPLC takes about 1 hour.

Determining a protein's primary structure is similar to solving a complex puzzle. Several steps are involved in solving the amino acid sequence of any protein.

- 1. Cleavage of all disulfide bonds.** Oxidation with performic acid is commonly used.
- 2. Determination of the N-terminal and C-terminal amino acids.** Several methods are available to determine the N-terminal amino acid. In Sanger's method, the polypeptide chain is reacted with 1-fluoro-2,4-dinitrobenzene. The dinitrophenyl (DNP) derivative of the N-terminal amino acid can then be isolated. Alternatively, dabsyl chloride is used to determine N-terminal amino acid residues. Dabsylation is a very sensitive technique because the dabsyl group, a fluorescent marker, is easily detected in small amounts. The N-terminal derivatives of both of these reagents can then be identified by HPLC analysis. A group of enzymes called the carboxypeptidases are used to identify the C-terminal residue. Because these enzymes sequentially cleave peptide bonds starting at the C-terminal residue, the first amino acid liberated is the C-terminal residue.



Edman degradation



3. Cleavage of the polypeptide into fragments. The polypeptide is broken into smaller peptides because technical problems prevent the direct sequencing of long polypeptides. The use of several reagents, each of which cuts the chain at a different site, creates overlapping sets of fragments. After the amino acid sequence of each fragment has been determined, the investigator uses this information to work out the entire sequence of the polypeptide. Of all the enzymes commonly used, the pancreatic enzyme trypsin is the most reliable. It cleaves peptide bonds on the carboxy side of either lysine or arginine residues. The peptide fragments, referred to as *tryptic peptides*, have lysine or arginine carboxy terminal residues. Chymotrypsin, another pancreatic enzyme, is also often used. It breaks peptide bonds on the carboxyl side of phenylalanine, tyrosine, leucine, methionine, or tryptophan. Treating the polypeptide with the reagent cyanogen bromide also generates peptide fragments. Cyanogen bromide specifically cleaves peptide bonds on the carboxyl side of methionine residues.

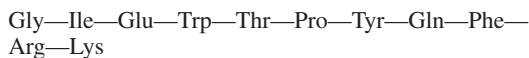
4. Determination of the sequences of the peptide fragments. Each fragment is sequenced through repeated cycles of a procedure called the *Edman degradation*. In this method phenylisothiocyanate (PITC), often referred to as Edman's reagent, reacts with the N-terminal residue of each fragment. Treatment of the product of this reaction with acid cleaves the N-terminal residue as a phenylthiohydantoin derivative. The derivative is then identified by comparing it with known standards, using electrophoresis or various chromatographic methods (most commonly HPLC). Because of the large number of steps involved in sequencing peptide fragments, Edman degradation is usually carried out by using a computer-programmed machine called a sequenator.

5. Ordering the peptide fragments. The amino acid sequence information derived from two or more sets of polypeptide fragments is next examined for overlapping segments. Such segments make it possible to piece together the overall sequence.

A typical primary sequence determination problem is given, along with its solution.

Problem

Consider the following peptide:



What amino acids and peptides are produced when this peptide is treated with each of the following reagents?

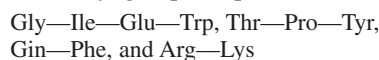
- Carboxypeptidase
- Chymotrypsin
- Trypsin
- DNFB (dinitrofluorobenzene)

Solution

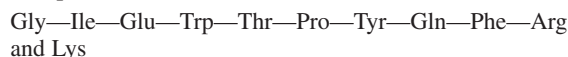
- Because carboxypeptidase cleaves at the carboxyl end of peptides, the products are



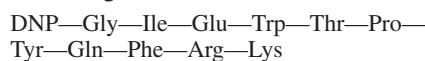
- Because chymotrypsin cleaves peptide bonds in which aromatic amino acids (i.e., Phe, Tyr, and Trp) contribute a carboxyl group, the products are



- Trypsin cleaves at the carboxyl end of lysine and arginine. The products are



- DNFB tags the amino-terminal amino acid. The product is



Hydrolysis then cleaves all the peptide bonds, and DNP—Gly can be identified by a chromatographic method.

Automation of the Edman degradation method has increased the speed and accuracy of the sequencing process. Computer-assisted devices called sequenators can determine the amino acid sequence of vanishingly small samples in a fraction of the time needed for the manual method.