

References: Lehninger, A.L.; Nelson, D.L.; Cox, M.M. *Principles of Biochemistry*, 2nd ed.; Worth Publishers: New York, 1993.

Berg, J.; Tymoczko, J.; Stryer, L. *Biochemistry*, 5th ed.; W. H. Freeman: New York, 2002.

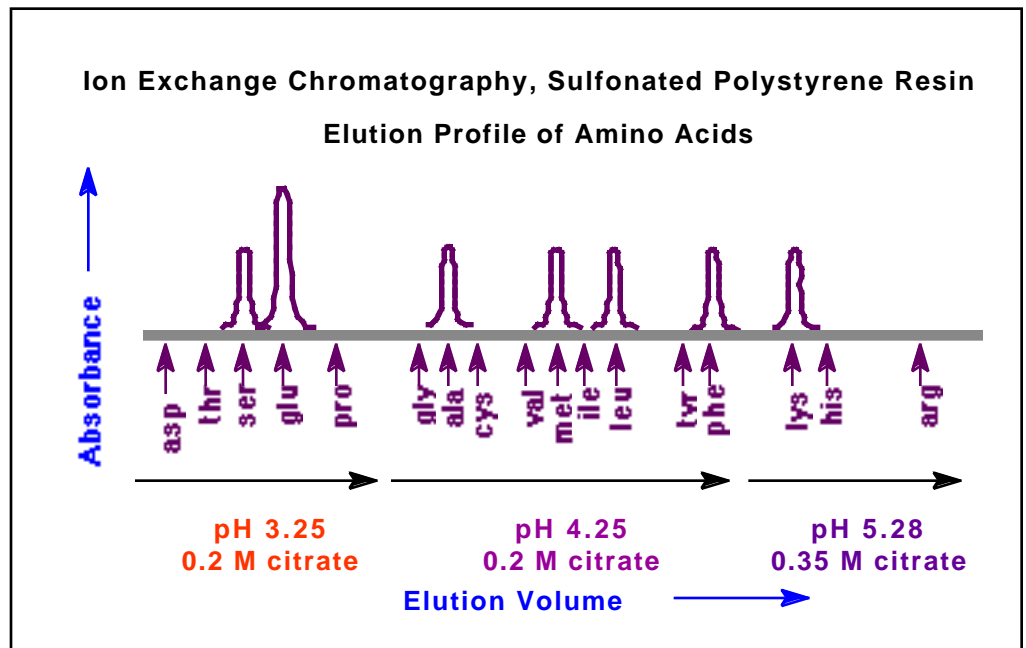
Introduction:

I. Determining Protein Amino Acid Sequence

Once the protein of interest has been extracted and purified, and its molar mass determined, the next step is to completely hydrolyze the protein (6 N HCl at 110°C for 24 hours) and determine its amino acid composition. Amino acids in the hydrolysate are separated and identified by ion exchange chromatography.

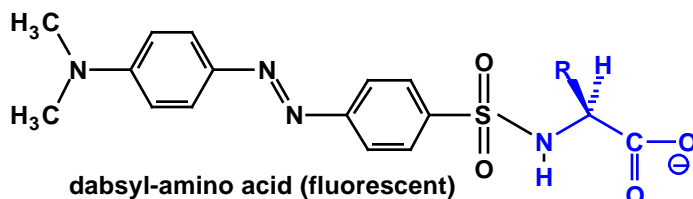
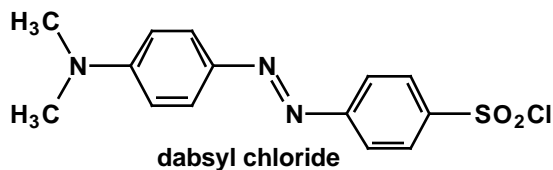
In most cases, the next step is to identify the first amino acid in the sequence, or in other words, the protein's amino terminal.

The amino terminal is typically reacted with a labelling reagent, such as dabsyl chloride, that forms a bond stable to hydrolysis. The peptide is hydrolyzed and the labeled amino acid identified.



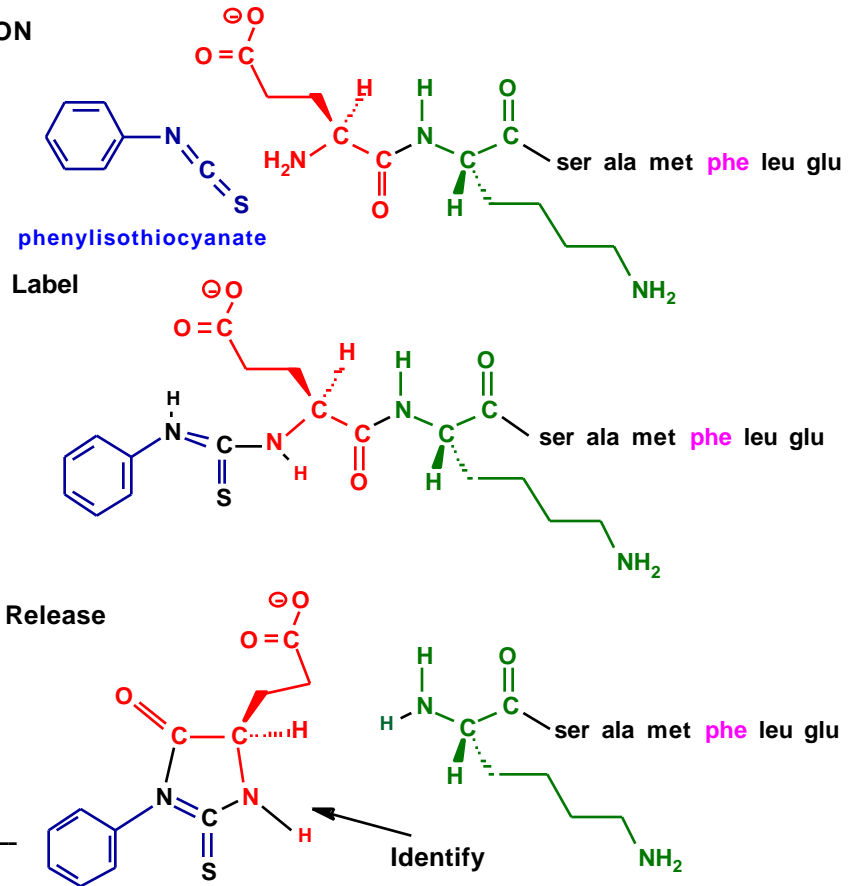
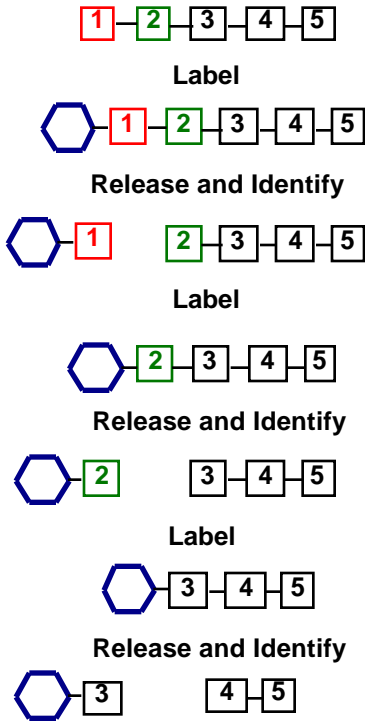
E K S A M F L E

glu lys ser ala met phe leu glu



Peptides up to 50 residues long can be sequenced by a cyclic procedure where the amino terminal is labelled, cleaved, identified and the process repeated on the shortened chain. This procedure, the Edman degradation method, has been automated. This machine, called a sequenator, performs the reactions, separations and identifications as well as recording all results.

EDMAN DEGRADATION



Endopeptidases and Sequencing by Fragment Overlap

trypsin
hydrolyzes peptide bonds on C side of lys, arg

resulting peptides

E K S A M F L E
glu lys ser ala met phe leu glu

glu lys ser ala met phe leu glu
E K S A M F L E

chymotrypsin
hydrolyzes peptide bonds on C side of phe, trp, tyr

resulting peptides

E K S A M F L E
glu lys ser ala met phe leu glu

glu lys ser ala met phe leu glu

Deduce sequence by comparing fragments

must be ←

glu lys ser ala met phe leu glu

glu lys ser ala met phe leu glu
trypsin
chymotrypsin

Proteins longer than 50 or so residues must be sequenced in an additional manner. The protein chain is broken into smaller fragments by site specific chemical reagents or endopeptidases such as trypsin and chymotrypsin. The resulting segments are separated and sequenced. The correct ordering of these sequences is made possible by repeating the procedure with chemical reagents or enzymes that cleave specifically at **different** sites than previously. This second analysis should yield a set of different fragments, which when compared to the first set and "overlapped", will allow the deduction of the complete protein amino acid sequence.

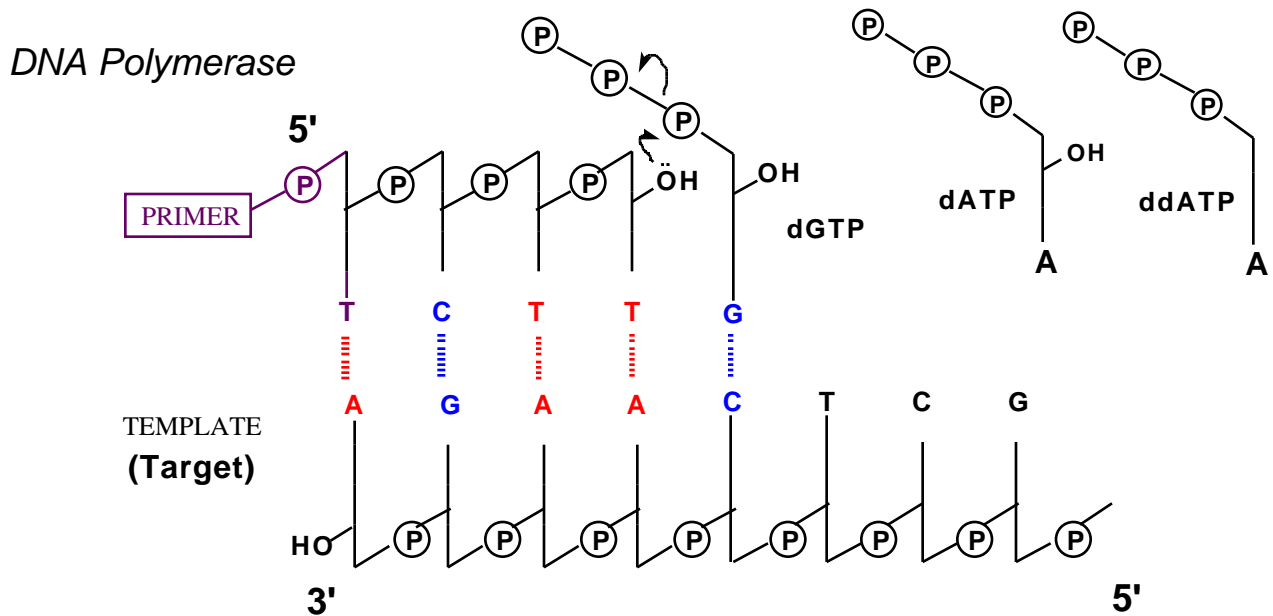
Recently, the development of recombinant gene technology and DNA sequencing methods make it possible to sequence proteins via an alternative strategy. If the gene for a given protein can be isolated, and the nucleic acid sequence determined, the protein amino acid sequence can be deduced through the genetic code.

II. Determining Nucleic Acid Nucleotide Sequence

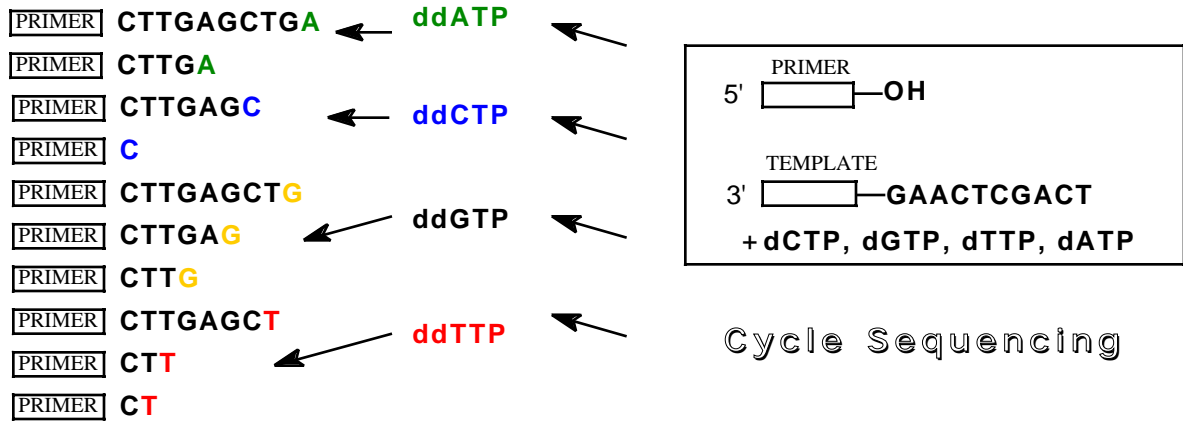
The development of a technique by Frederick Sanger has made it relatively easy to sequence large DNA molecule fragments. The method depends on the abilities to:

- 1) find two appropriate DNA primers that border the target DNA fragment
- 2) synthesize the complementary strand to the target utilizing the primers, DNA polymerase and deoxynucleotides

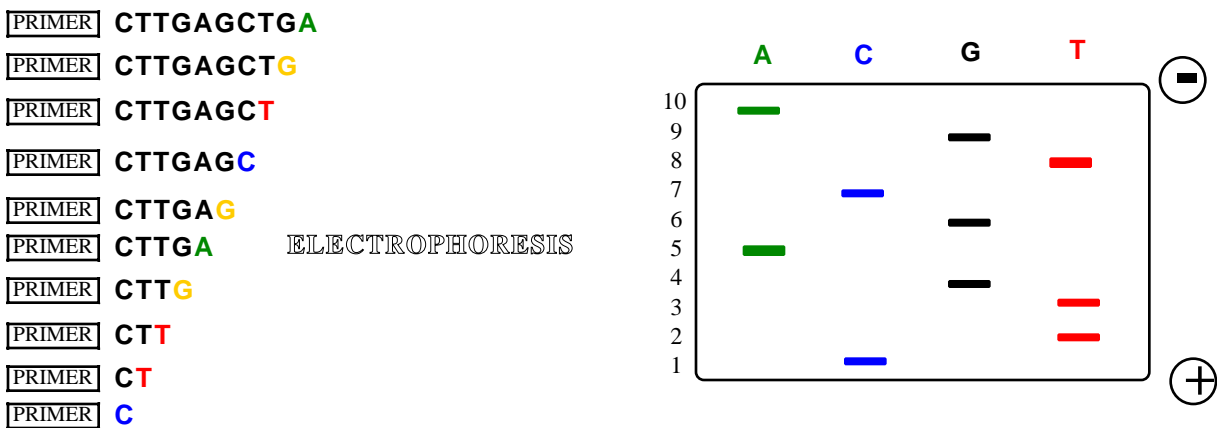
DNA SEQUENCING by the SANGER (dideoxy) METHOD



3) generate four sets of labelled complementary DNA fragments, with each set generated by a specific reaction including a *dideoxynucleotide*, and



4) separate labeled complementary DNA fragments, by electrophoresis, that differ in length by only **one** nucleotide



The four sets of labelled fragments are electrophoretically separated side-by-side, and the pattern of bands produced directly yields the nucleotide sequence.

For example above, the labelled complementary DNA fragment, pCTTGAGCTGA, can be produced from the target pTCAGCTCAAG by four different *dideoxynucleotide* reactions, each yielding a *dideoxynucleotide* at the 3' end of fragments that all begin at the same starting point.

The "*dideoxynucleotide* A" electrophoresis lane will show fragments 5 and 10 nucleotides long; the "*dideoxynucleotide* C" lane will have fragments 1 and 7 nucleotides long; the "*dideoxynucleotide* G" lane will have fragments 4, 6 and 9 nucleotides long; and the "*dideoxynucleotide* T" lane will have fragments 2, 3 and 8 nucleotides long.

The final result is that there will be one labelled complementary DNA fragment for each length up to the total number of bases, and each of these fragments ends with the base at that number position in the sequence of the labelled complementary DNA fragment. The sequence of the original target DNA fragment may now be deduced from complementary fragment using base pairing rules.