Protein Sequencing

Primary Structure of Proteins

Dr Shilpa D Kaistha
Department of Life Sciences & Biotechnology
School of Science
CSJM University Kanpur

Levels of Protein Structure

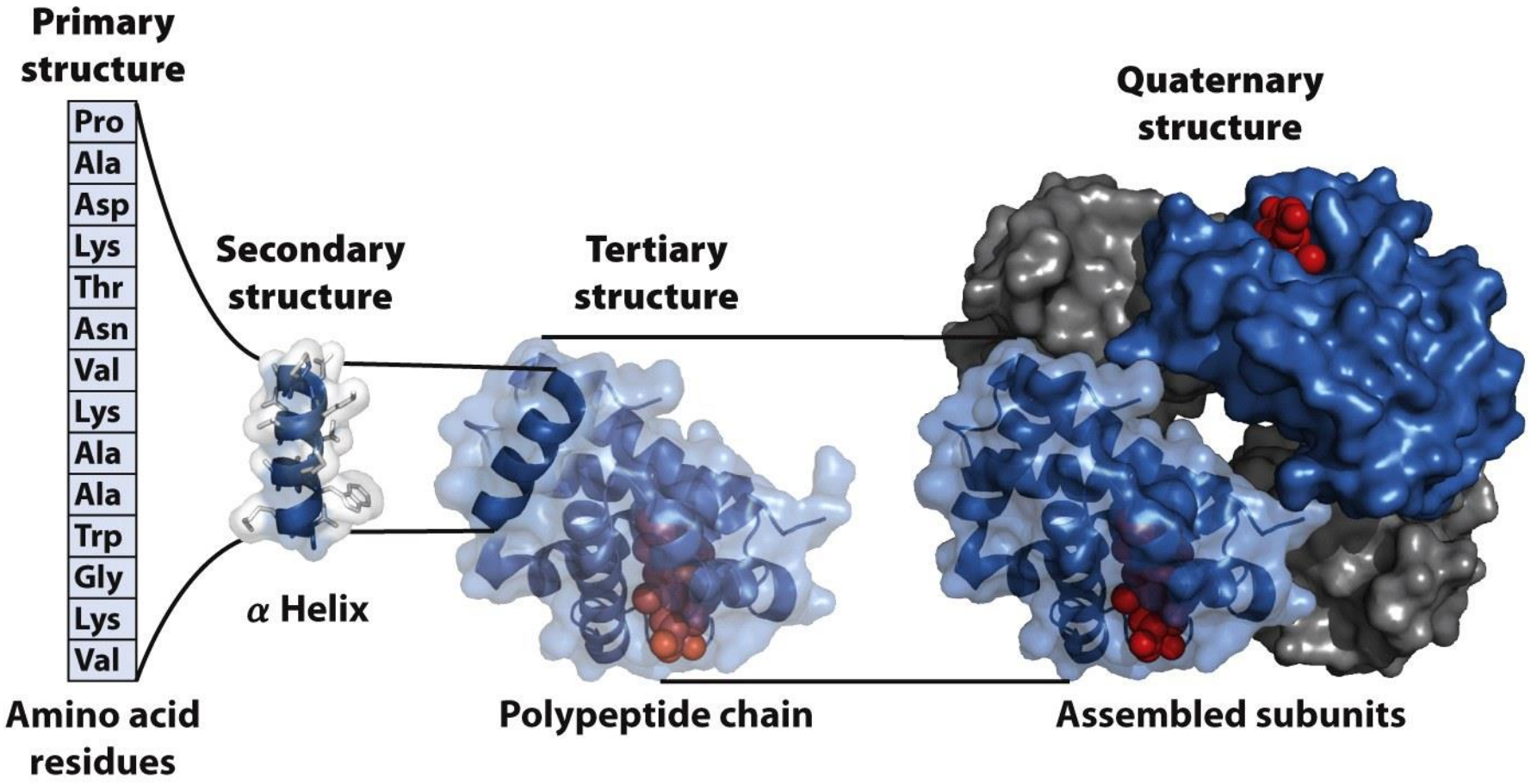
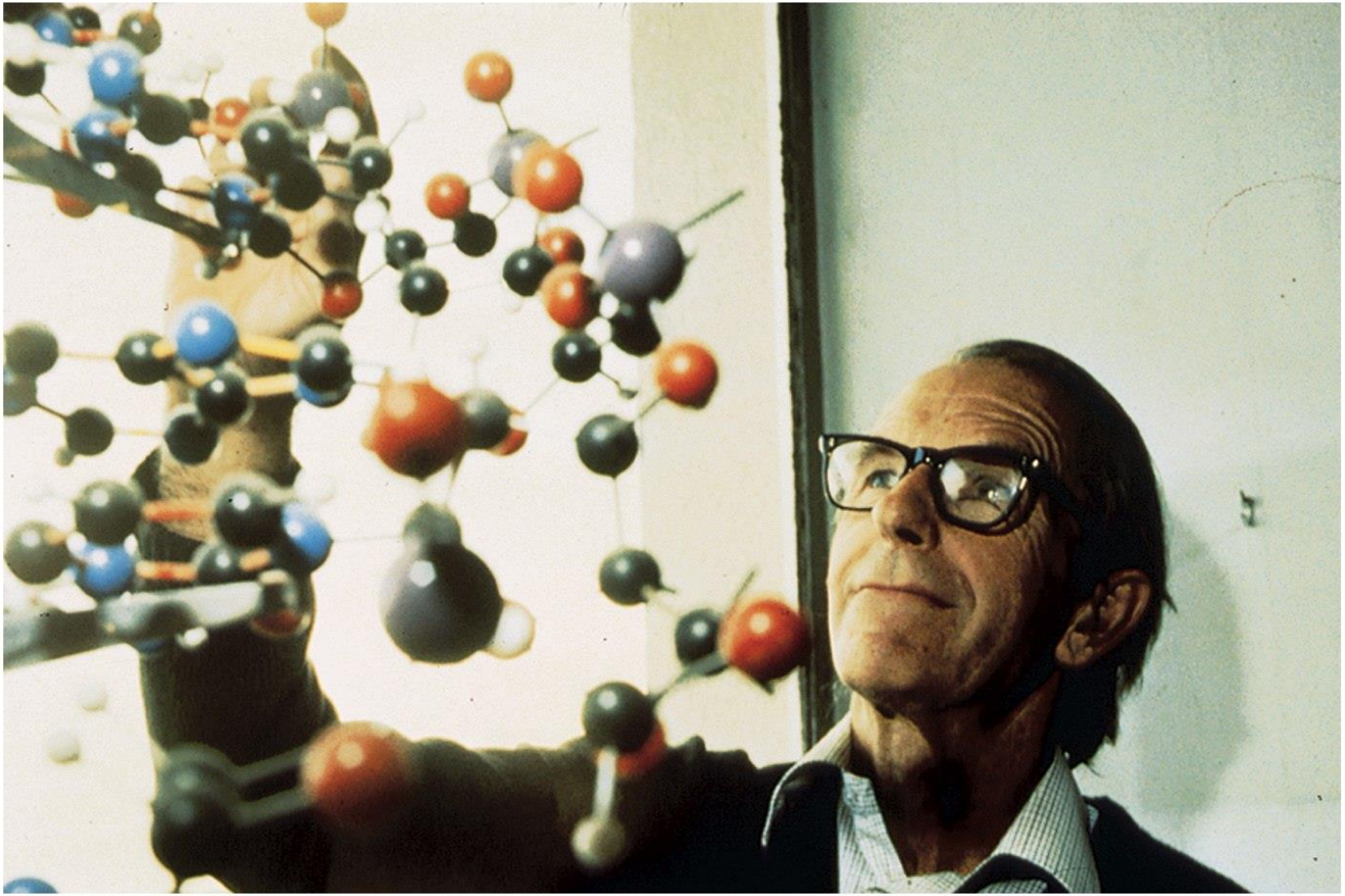


Figure 3-23
Lehninger Principles of Biochemistry, Fifth Edition



Fredrick Sanger – developed first N-terminal determination and early sequencing methods

Protein Sequencing

It is essential to further biochemical analysis that we know the sequence of the protein we are studying

Edman Degradation (Classical method)

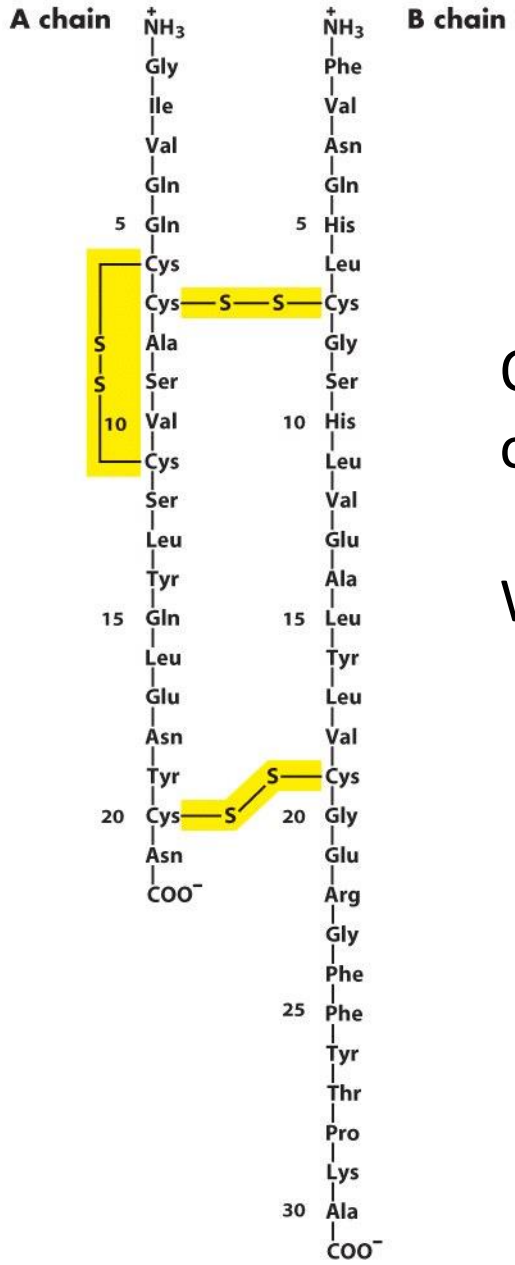
- Successive rounds of N-terminal modification, cleavage, and identification
- Can be used to identify protein with known sequence

Mass Spectrometry (Modern method)

- MALDI MS and ESI MS can precisely identify the mass of a peptide, and thus the amino acid sequence
- Can be used to determine post-translational modifications.
- Actual sequence can also be determined from DNA sequence

- Bioinformatics

Insulin – the First Protein Sequenced



Once an pure protein is obtained, then….

What needs to be done ?

R-S-S-R

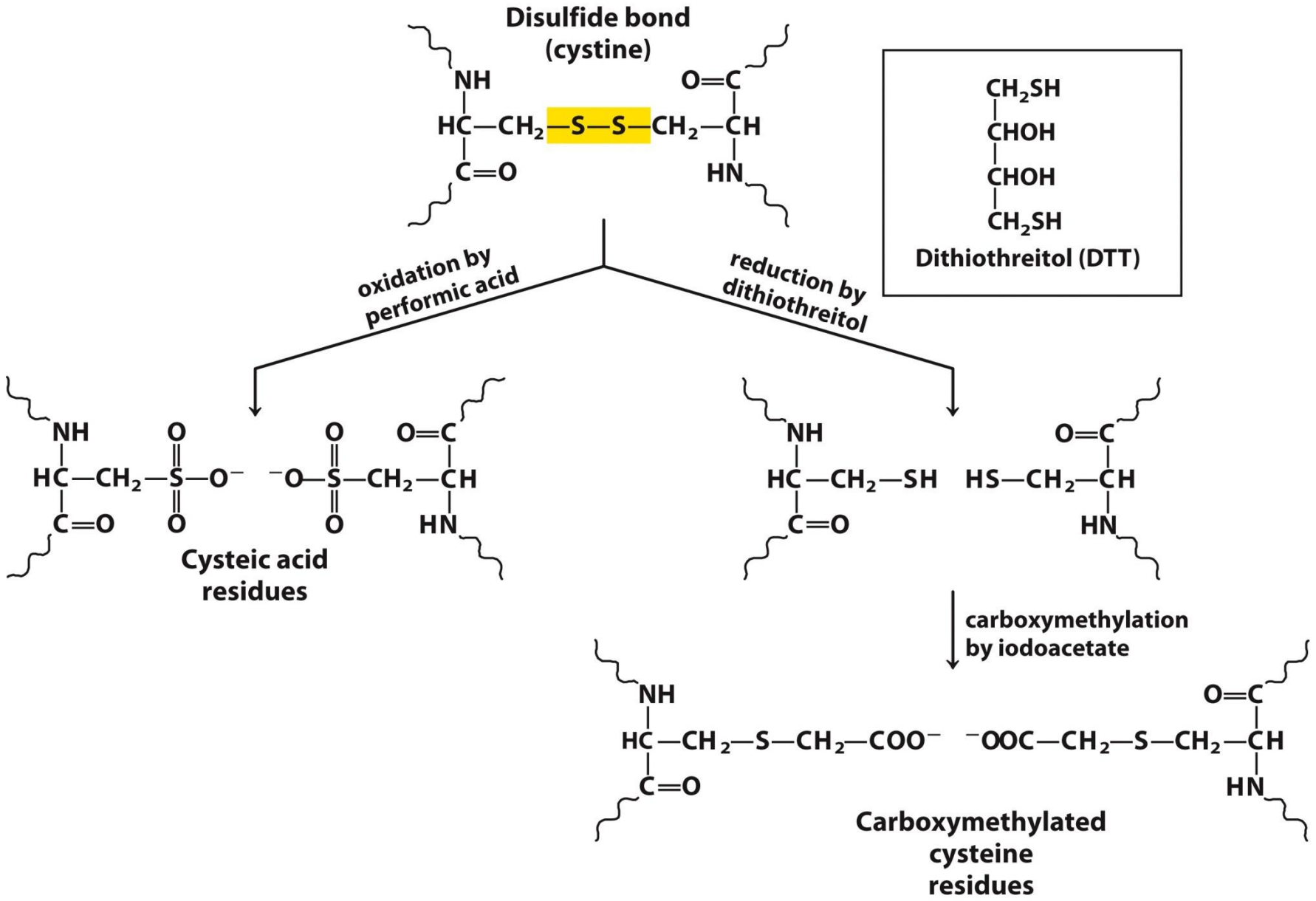


Figure 3-28
Lehninger Principles of Biochemistry, Sixth Edition

Acid Hydrolysis of Proteins

6 N HCl for several hours,100-105^{o}C,to hydrolyze Peptide bonds

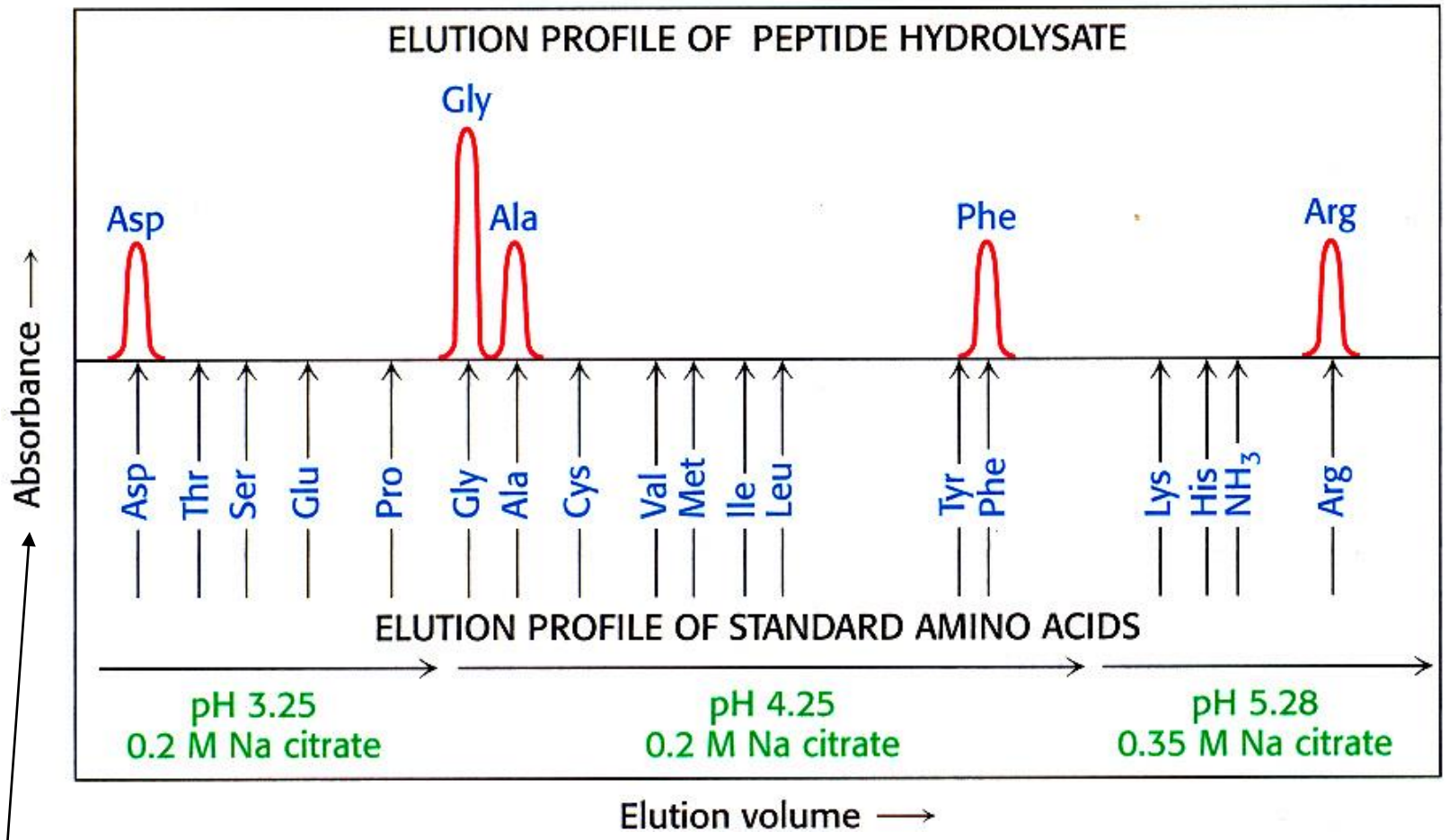
Four Problems to Deal With:

1. Destroys W.
2. Partially (slowly) destroys S and T.
3. Converts N ➔ D + NH_4^+ , and Q ➔ E + NH_4^+.
4. Slowly hydrolyzes peptide bonds between vicinal ile, leu, and val.

Dealt with by:

1. KOH hydrolysis to determine W.
2. HCl hydrolysis over 2 hr, 4hr, 6hr….for S, T, I, L, V.
3. Measure NH_4^+ to determine amount of N+Q ➔ D + E.

Amino Acid Analysis of AGDFRG

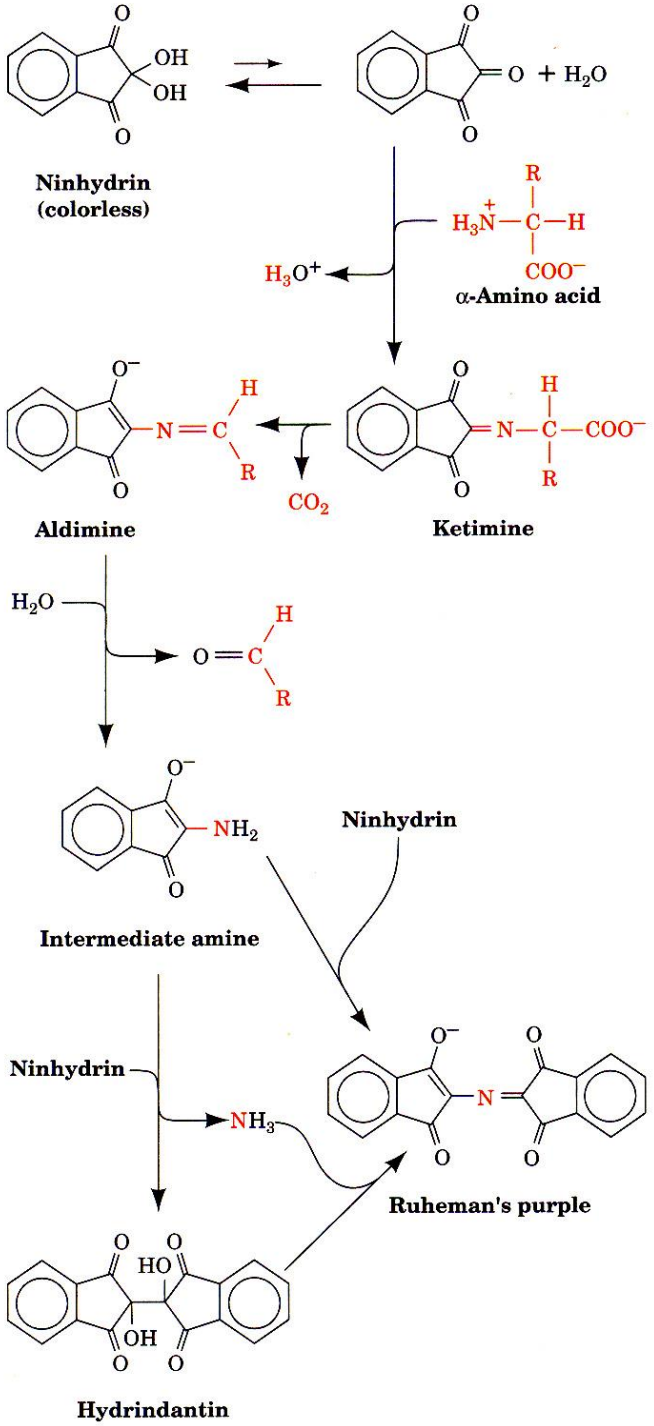


Based on Ninhydrin Reaction

Amino Acid Analysis of Acid Hydrolysate

HPLC using ion exchange or other chromatography – all automated equipment.

Amino acids, as they come off the column, reacted with **ninhydrin.**



N-terminal Reagents

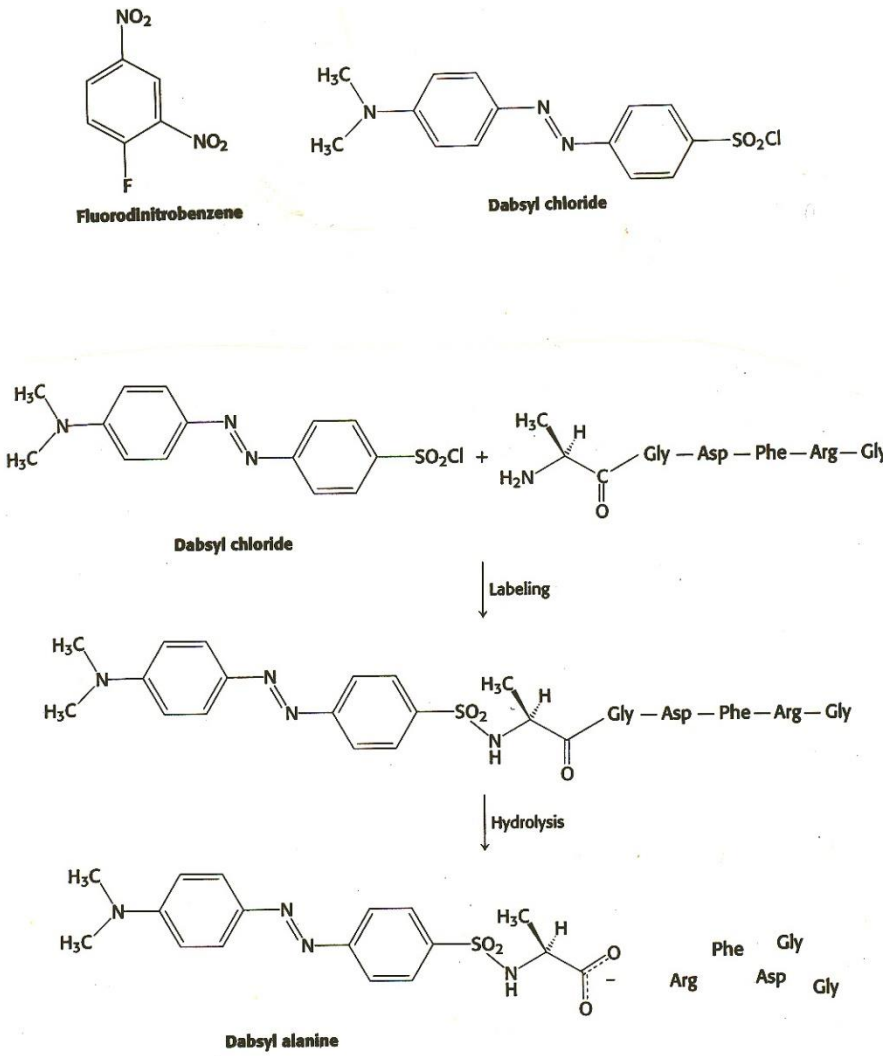


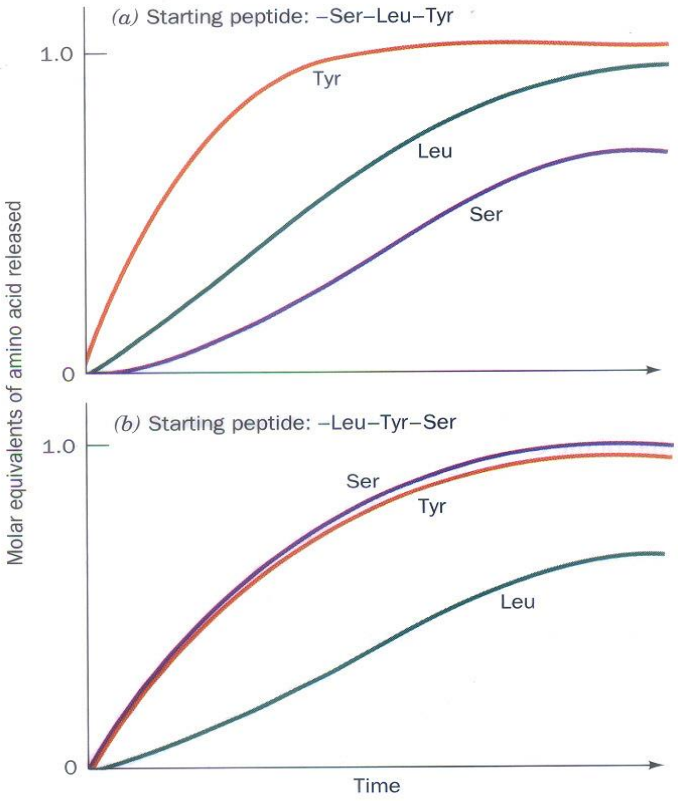
FIGURE 4.20 Determination of the amino-terminal residue of a peptide. Dabsyl chloride labels the peptide, which is then hydrolyzed with the use of hydrochloric acid. The dabsyl–amino acid (dabsyl–alanine in this example) is identified by its chromatographic characteristics.

C terminal Carboxypeptidases

Table 7-1 **Specificities of Various Exopeptidases**

Enzyme	Source	Specificity[a]
Carboxypeptidase A	Bovine pancreas	$R_n \neq$ Arg, Lys, Pro; $R_{n-1} \neq$ Pro
Carboxypeptidase B	Bovine pancreas	$R_n \neq$ Arg, Lys, $R_{n-1} \neq$ Pro
Carboxypeptidase C	Citrus leaves	All free C-terminal residues; pH optimum = 3.5
Carboxypeptidase Y	Yeast	All free C-terminal residues, but slowly with R_n = Gly
Leucine aminopeptidase	Porcine kidney	$R_1 \neq$ Pro
Aminopeptidase M	Porcine kidney	All free N-terminal residues

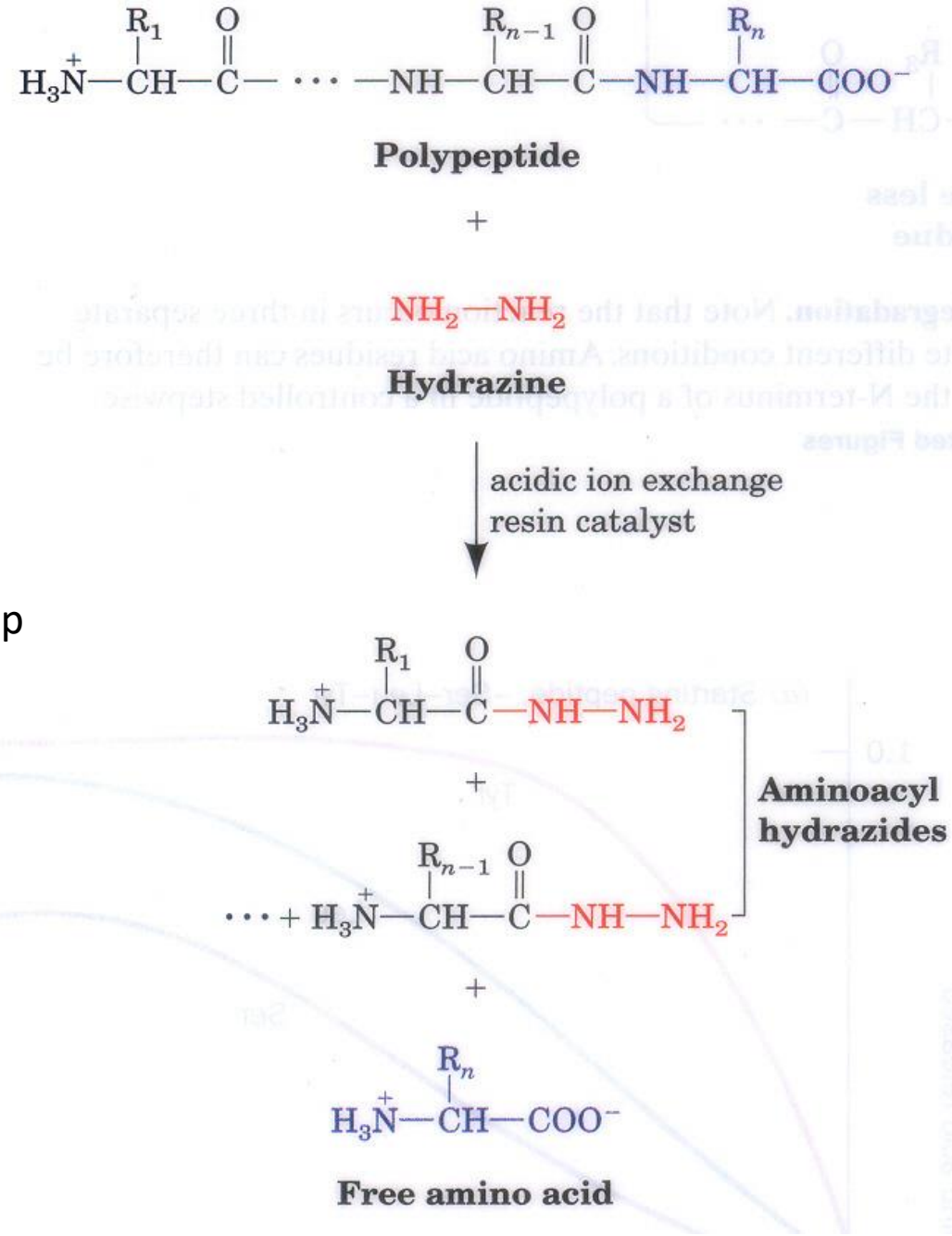
[a] R_1 = the N-terminal residue; R_n = the C-terminal residue.



Hydrazinolysis

Polypeptide + anhyd-Hydrazine at 90°C + mildly acidic ion exchange resin (catalyst) ➔ for 20-100 hrs.

Modifies pepetide carboxyl group to form acyl hydrazide



Edman Degradation

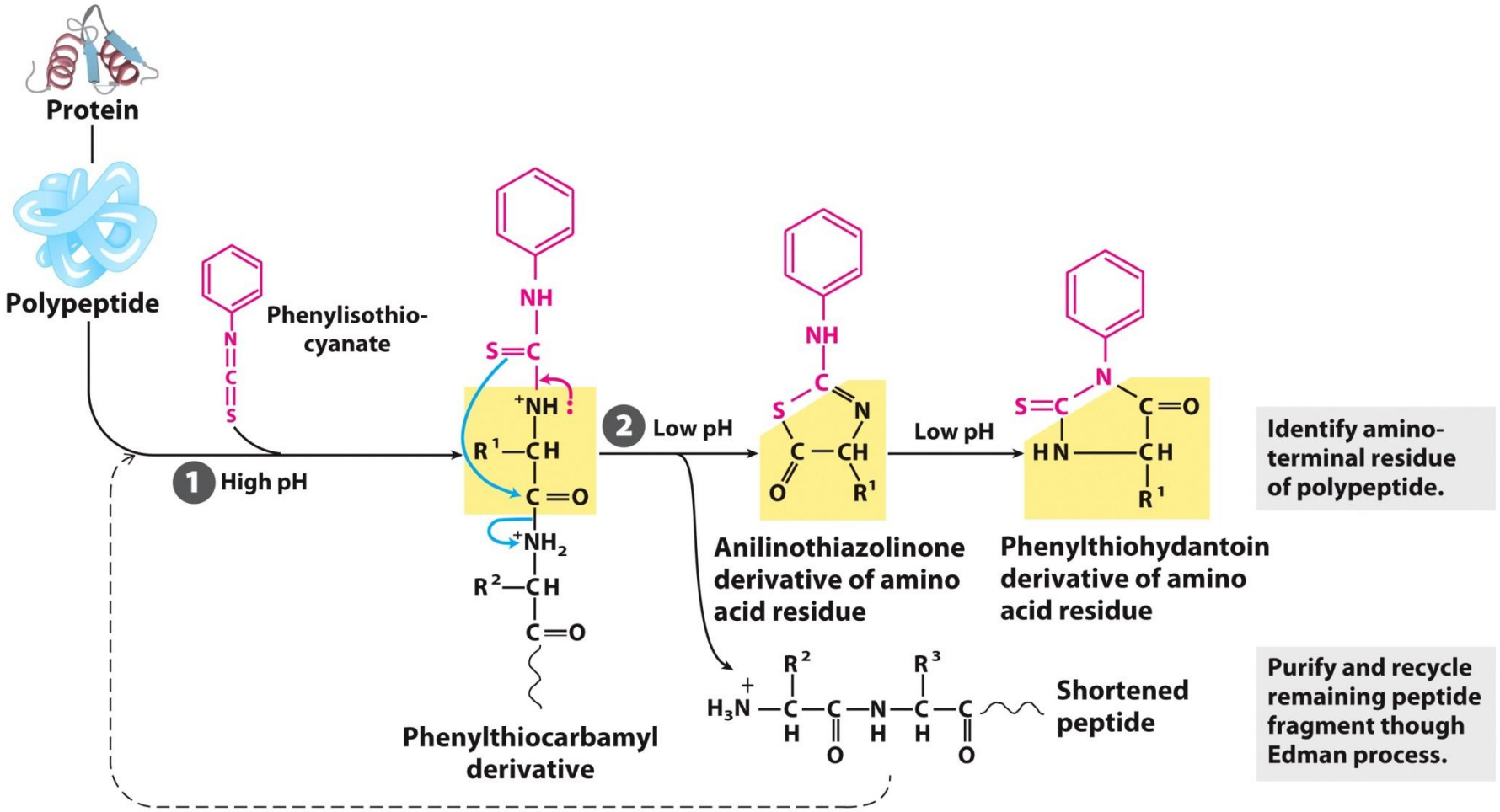
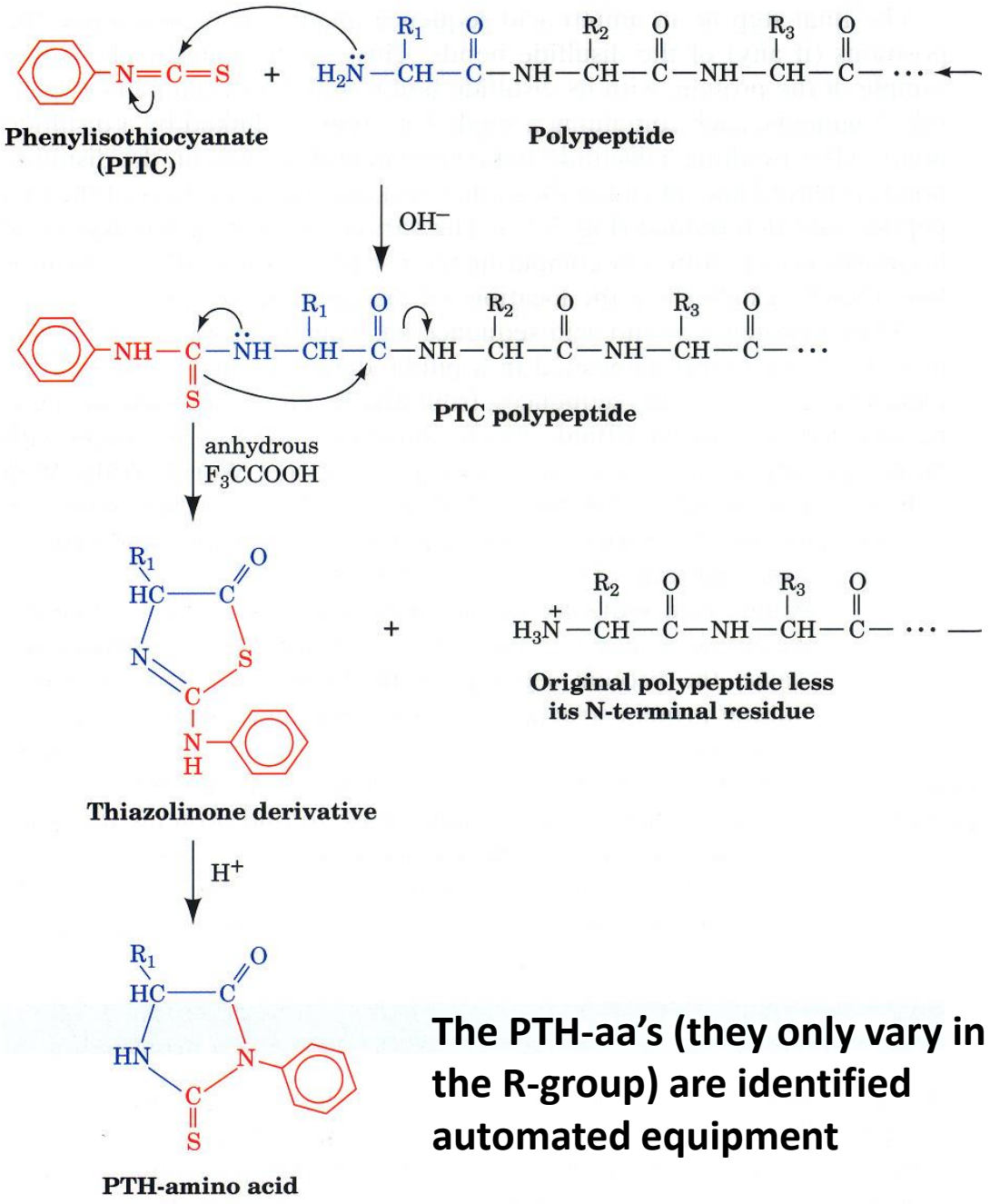


Figure 3-27
Lehninger Principles of Biochemistry, Sixth Edition

Edman Degradation = Amino Acid Sequencing



The PTH-aa's (they only vary in the R-group) are identified automated equipment

Protein Fragmentation Methods

TABLE 3–6 The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

Reagent (biological source)*	Cleavage points†
Trypsin (bovine pancreas)	Lys, Arg (C)
Submaxillary protease (mouse submaxillary gland)	Arg (C)
Chymotrypsin (bovine pancreas)	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease (bacterium *S. aureus*)	Asp, Glu (C)
Asp-*N*-protease (bacterium *Pseudomonas fragi*)	Asp, Glu (N)
Pepsin (porcine stomach)	Leu, Phe, Trp, Tyr (N)
Endoproteinase Lys C (bacterium *Lysobacter enzymogenes*)	Lys (C)
Cyanogen bromide	Met (C)

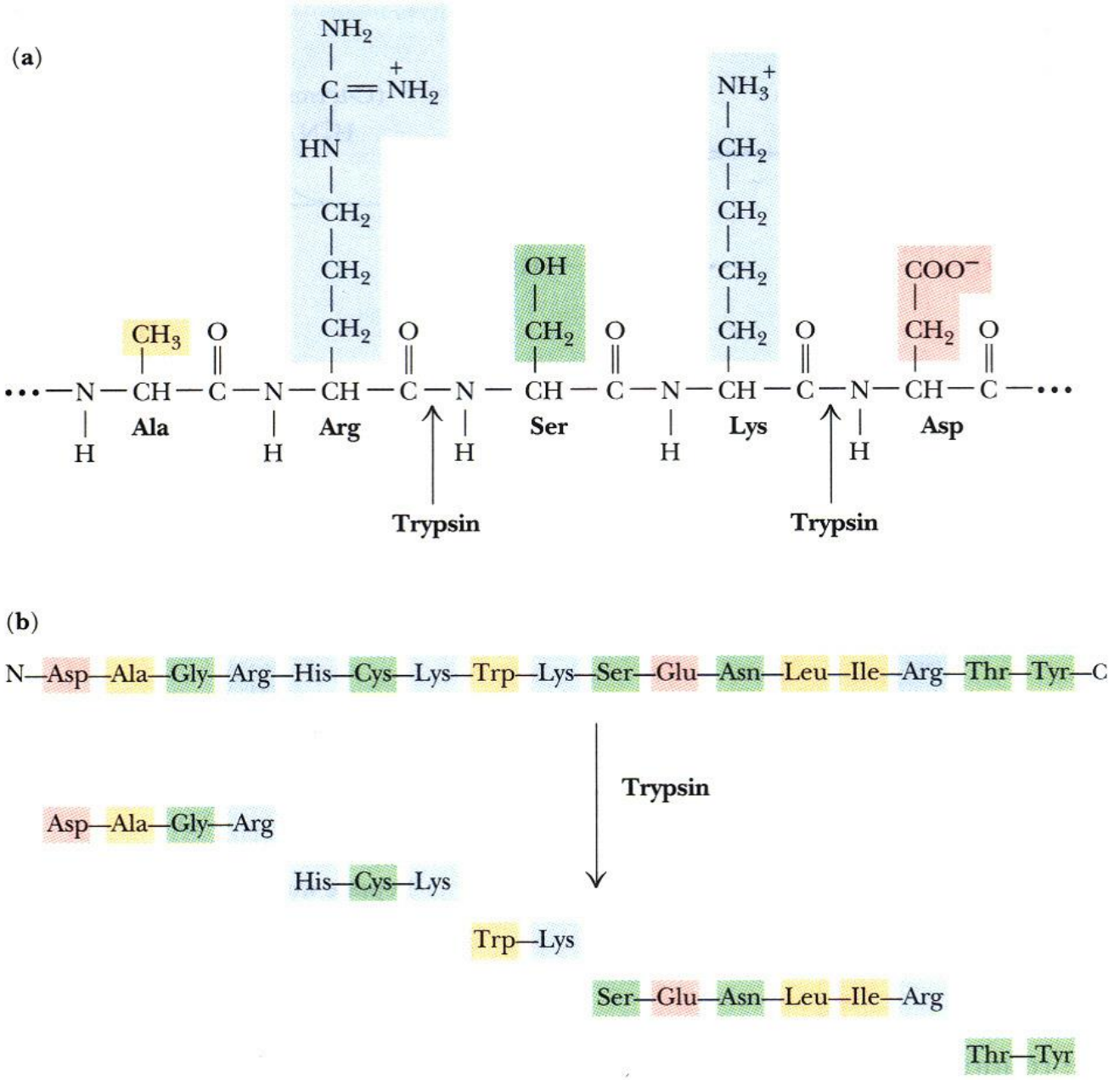
*All reagents except cyanogen bromide are proteases. All are available from commercial sources.

†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

Table 3-6
Lehninger Principles of Biochemistry, Sixth Edition

You need to know 4 of them: trypsin, chymotrypsin, pepsin and cyanogen bromide fragmentation methods.

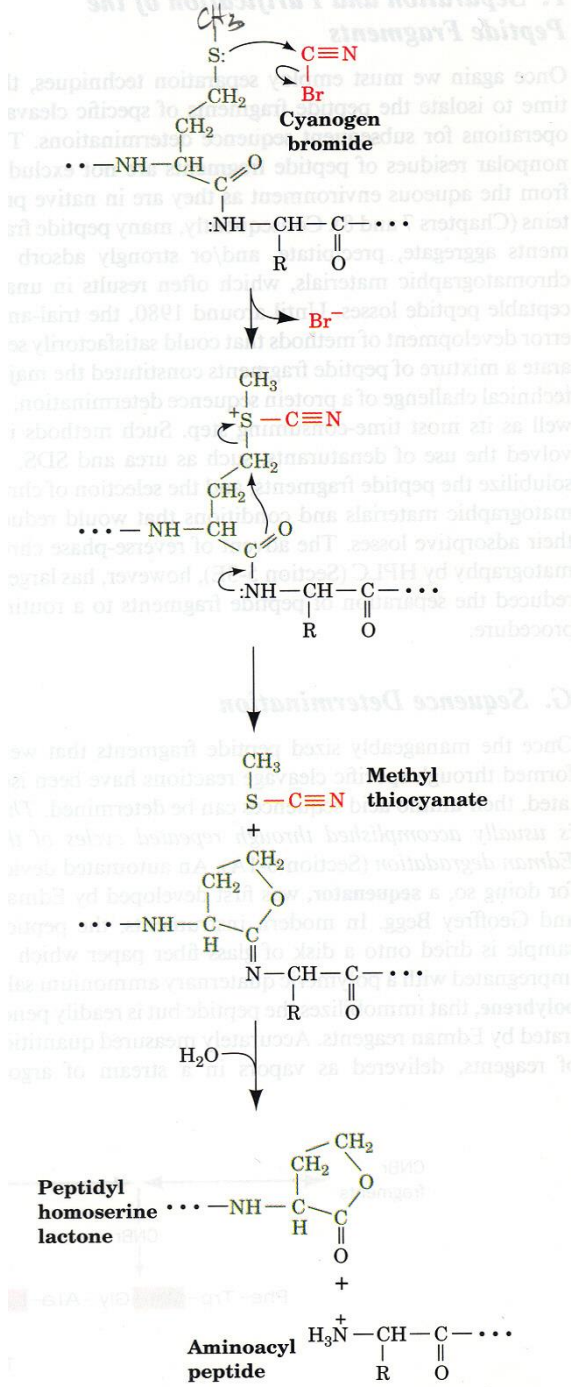
Trypsin Fragmentation



CNBr Fragmentation

Cleaves the C-terminal side of Met…

….and converts the Met to Homoserine Lactone



Protein Sequencing Overall Flow

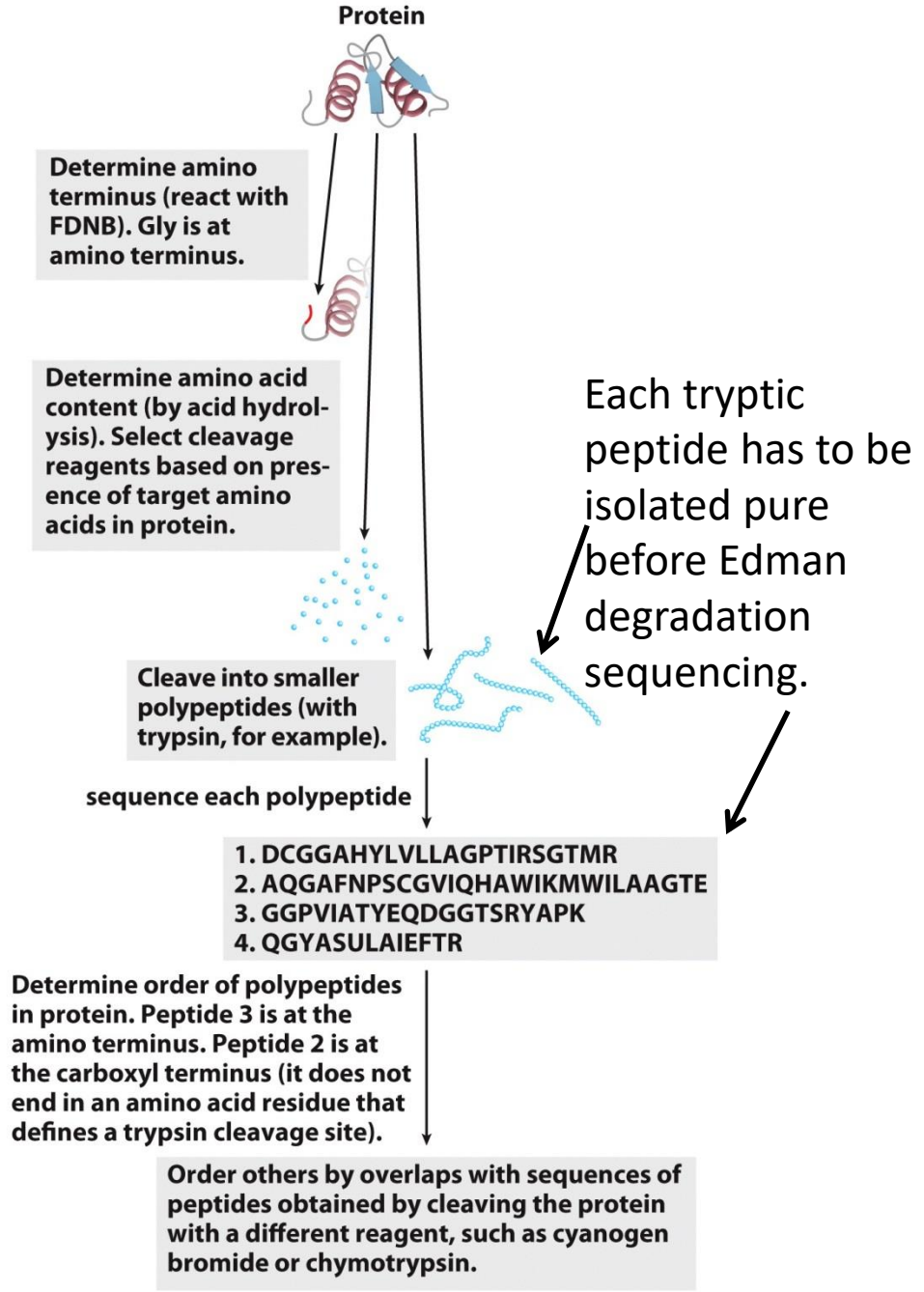
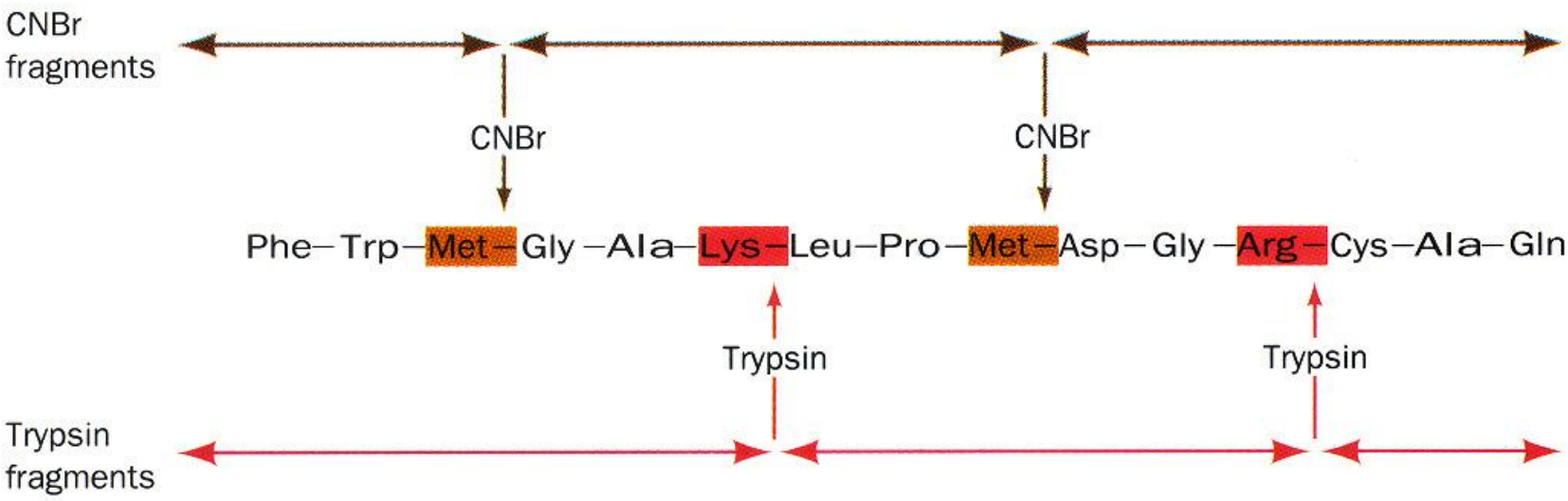


Figure 3-25
Lehninger Principles of Biochemistry, Sixth Edition

Trypsin and CNBr Fragmentation Each Done Separately



Separation of Protein Fragments
The Classic Paper Chromatography + Electrophoresis

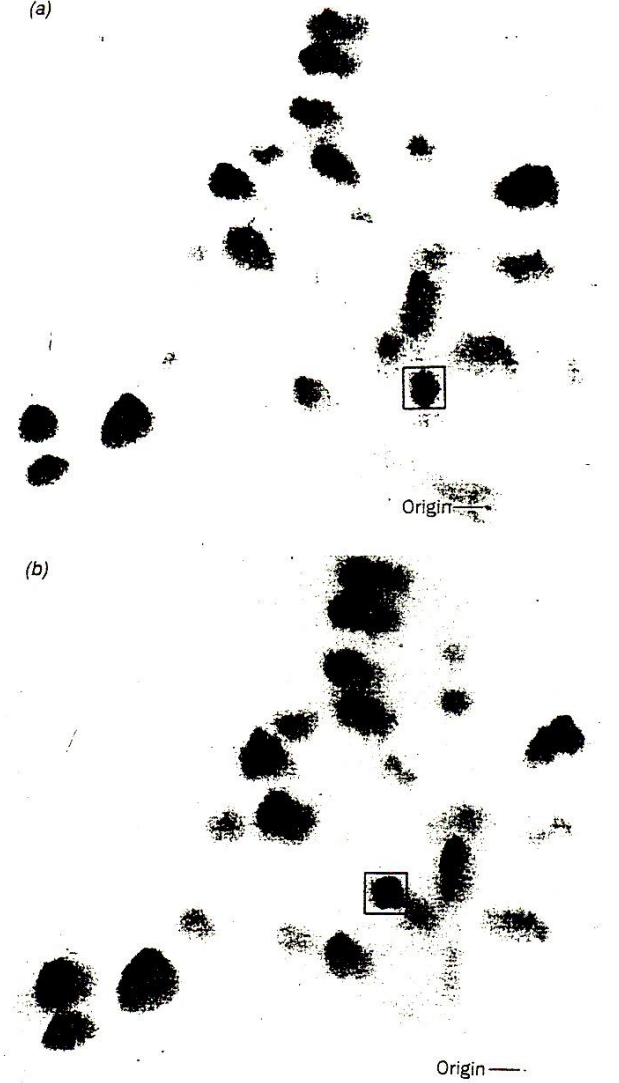
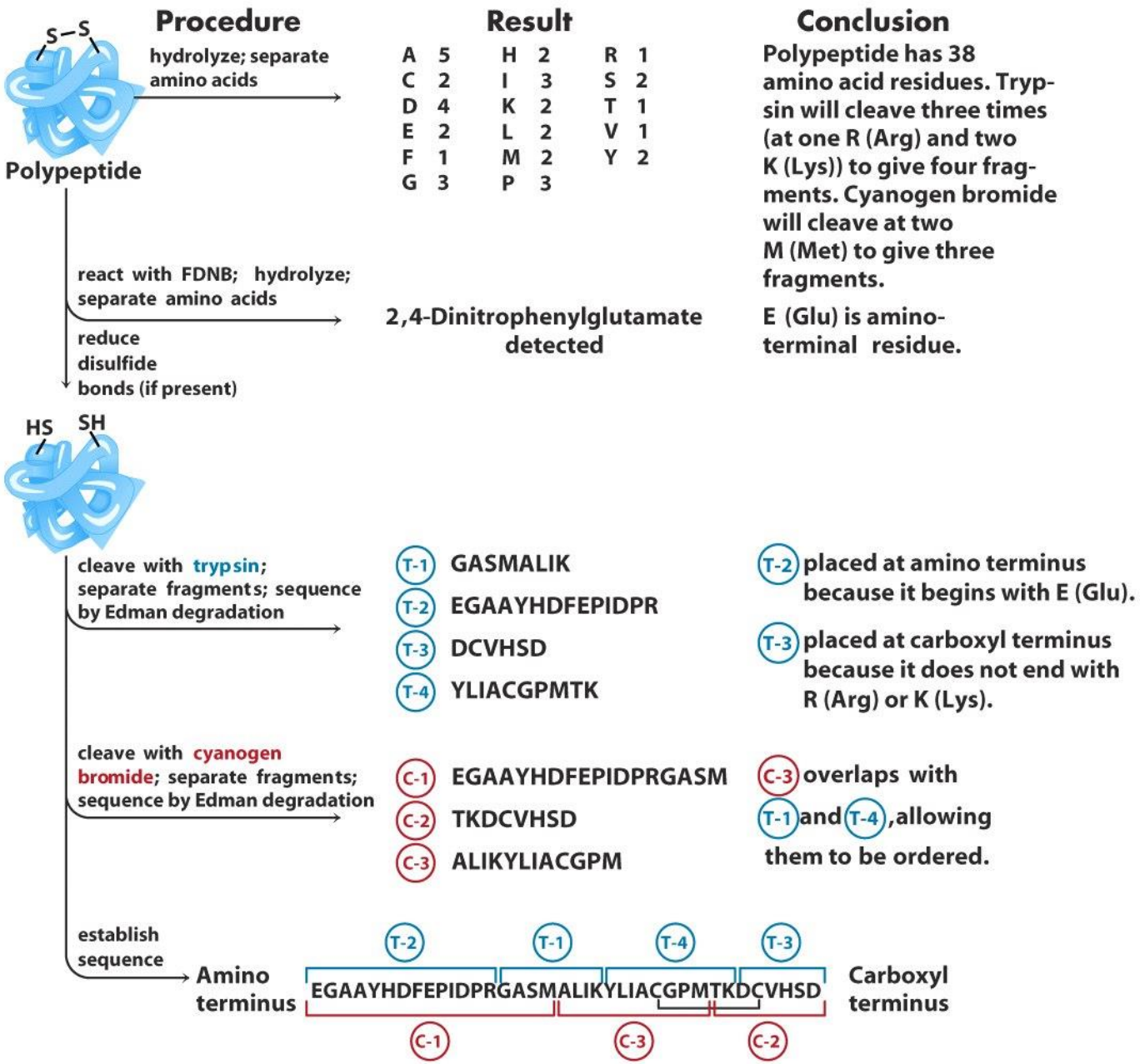


FIGURE 6-10. A comparison of the ninhydrin-stained fingerprints of trypsin-digested (*a*) hemoglobin A (HbA) and (*b*) hemoglobin S (HbS). The peptides that differ in these two forms of hemoglobin are boxed. These peptides constitute the eight N-terminal residues of the β subunit of hemoglobin. Their amino acid sequences are

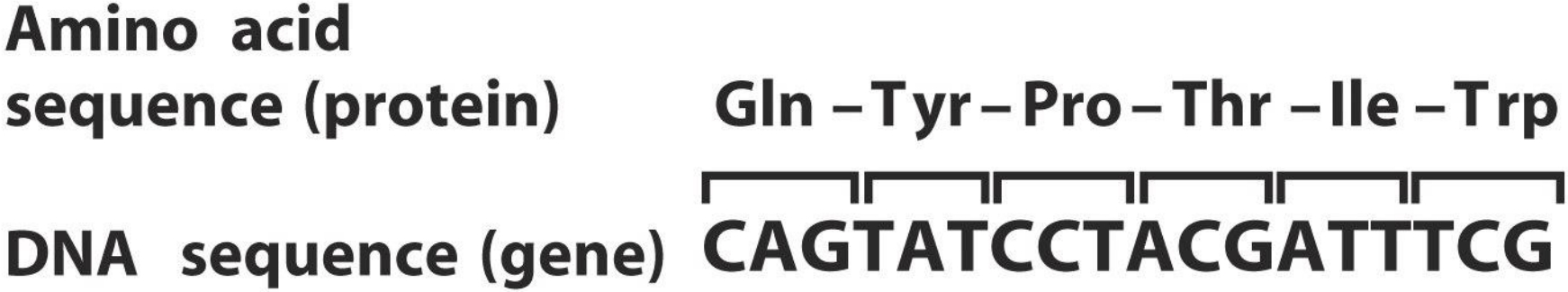
Hemoglobin A	Val-	His-	Leu-	Thr-	Pro-	**Glu-**	Glu-	Lys
Hemoglobin S	Val-	His-	Leu-	Thr-	Pro-	**Val-**	Glu-	Lys
-	β1	2	3	4	5	**6**	7	8

[Courtesy of Corrado Baglioni, State University of New York at Albany.]

Protein Sequencing: Overlapping Sequences



Protein Sequence from DNA Sequence



MS Procedures for Sequence IDs

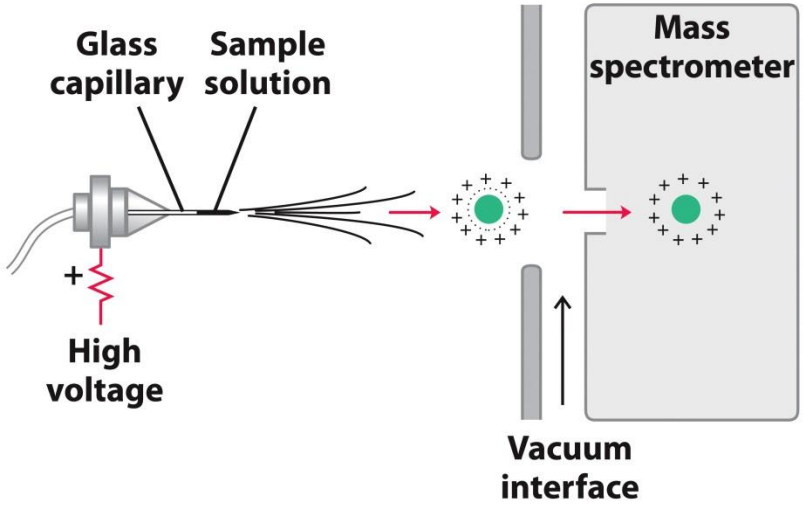


Figure 3-30a
Lehninger Principles of Biochemistry, Sixth Edition

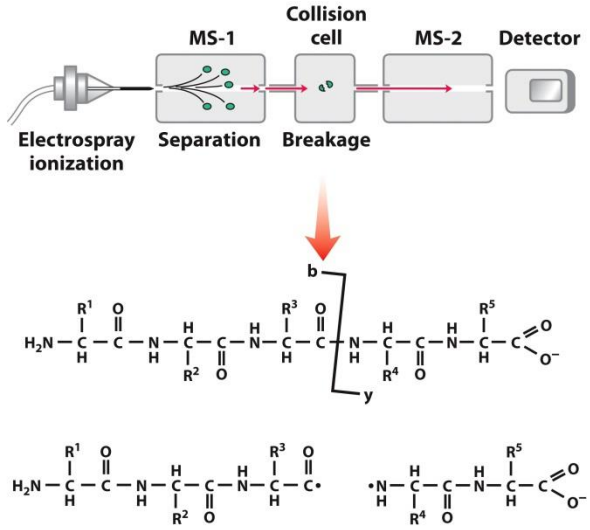


Figure 3-31a
Lehninger Principles of Biochemistry, Sixth Edition

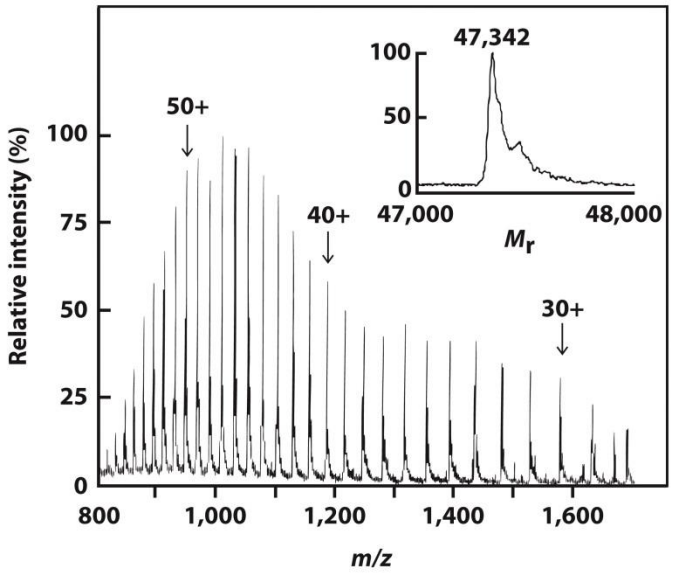


Figure 3-30b
Lehninger Principles of Biochemistry, Sixth Edition

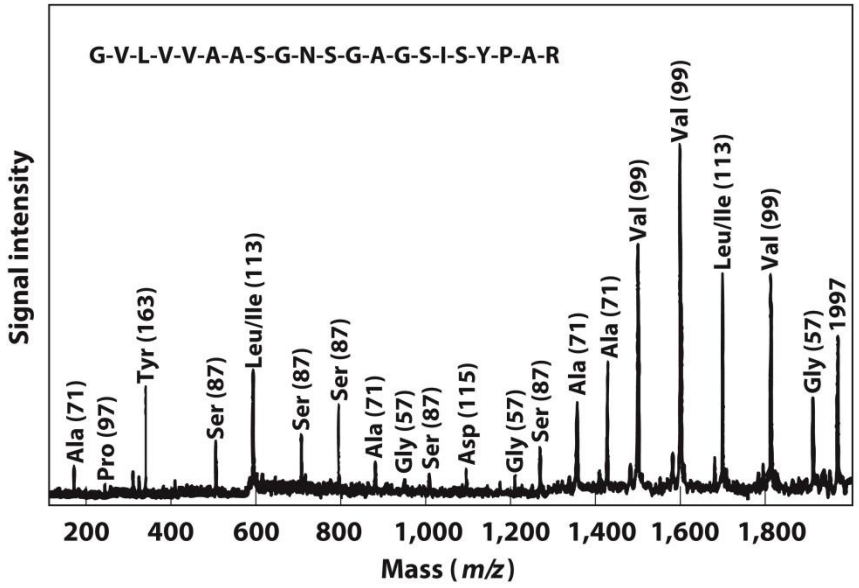
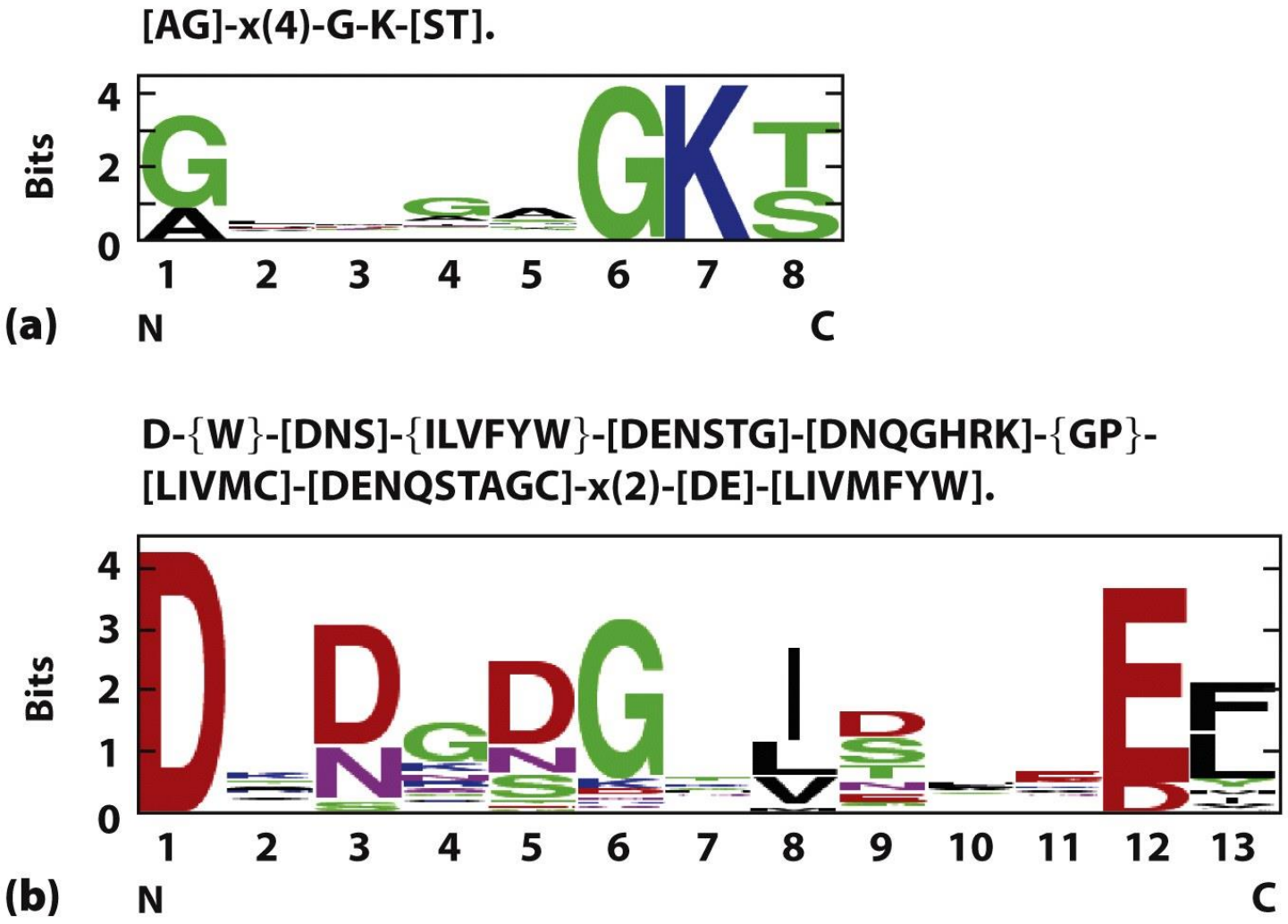
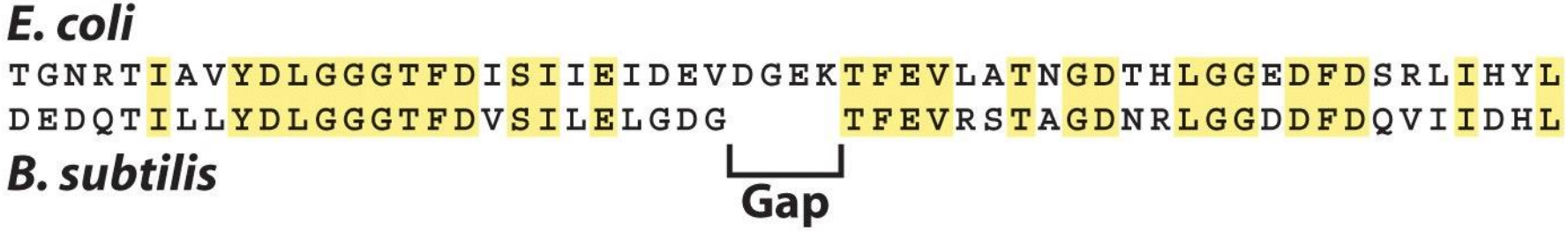


Figure 3-31b
Lehninger Principles of Biochemistry, Sixth Edition

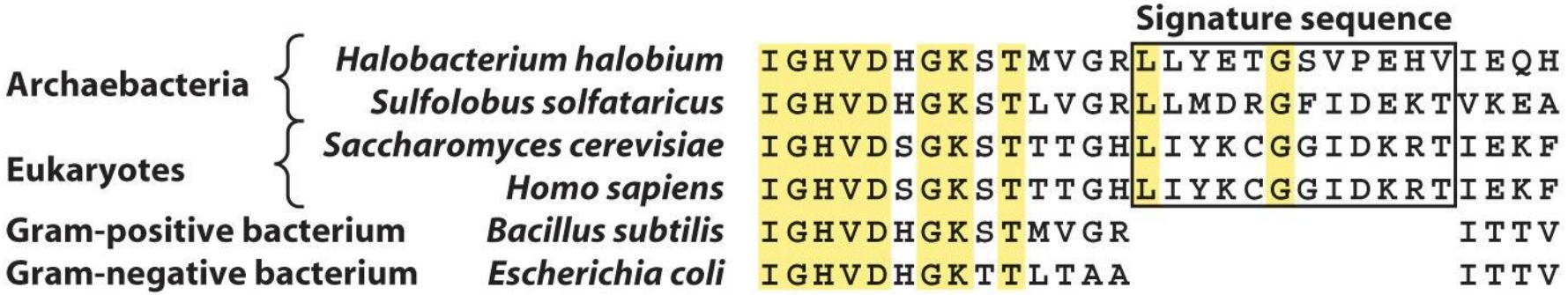
Consensus Sequences



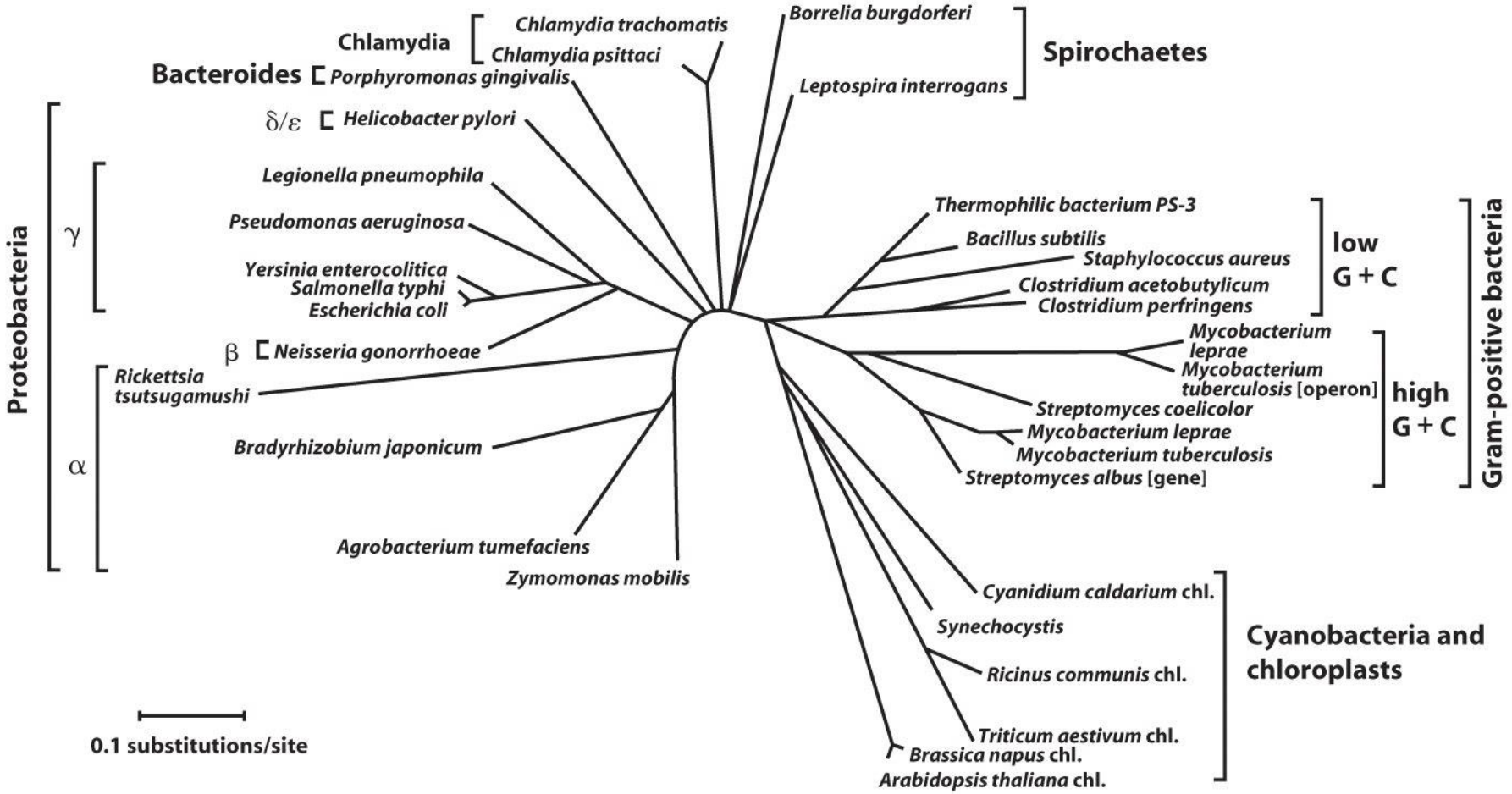
Partial Primary Structure of Elongation Factor Tu Sequences Aligned with Gaps



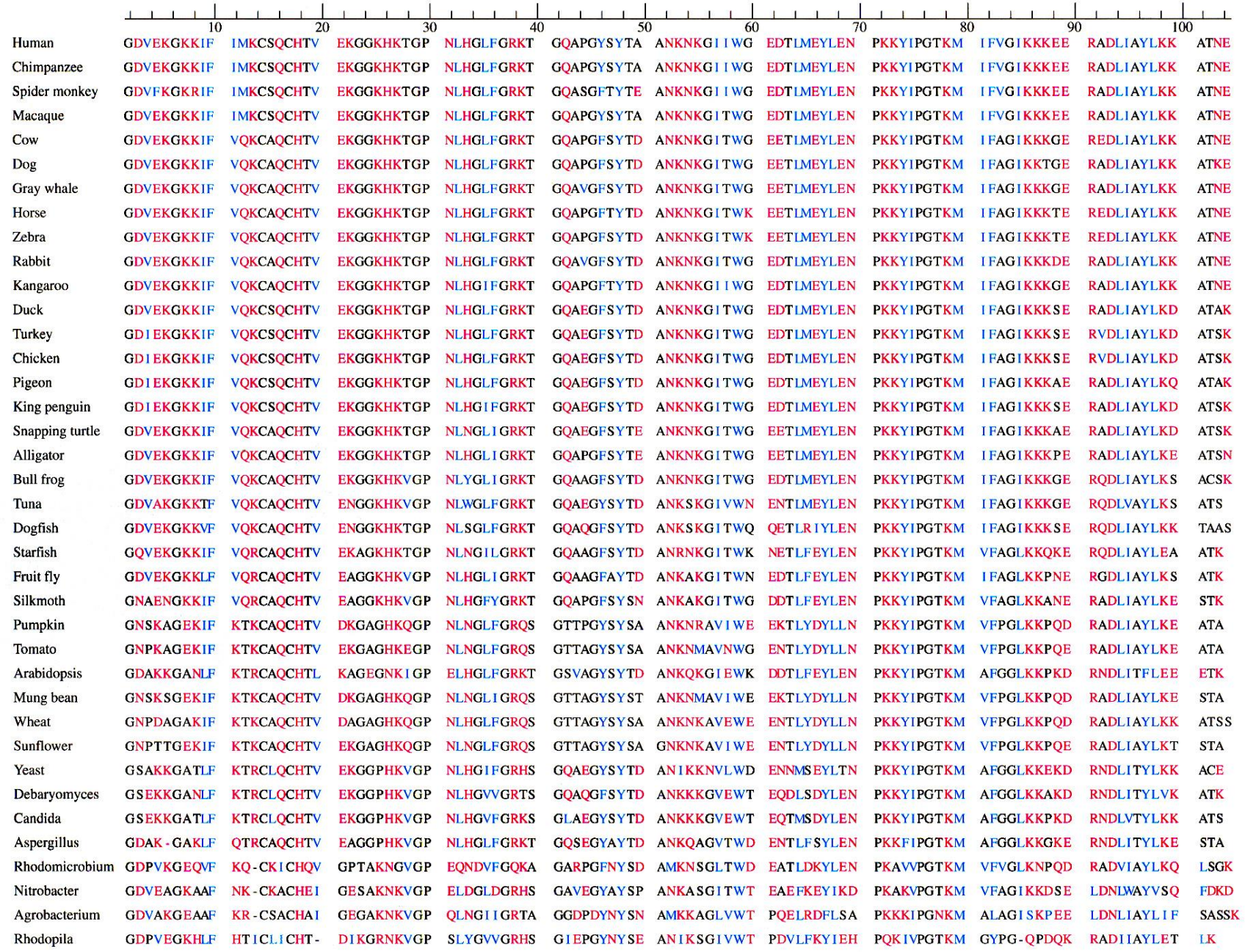
EF-Tu Signature Sequences



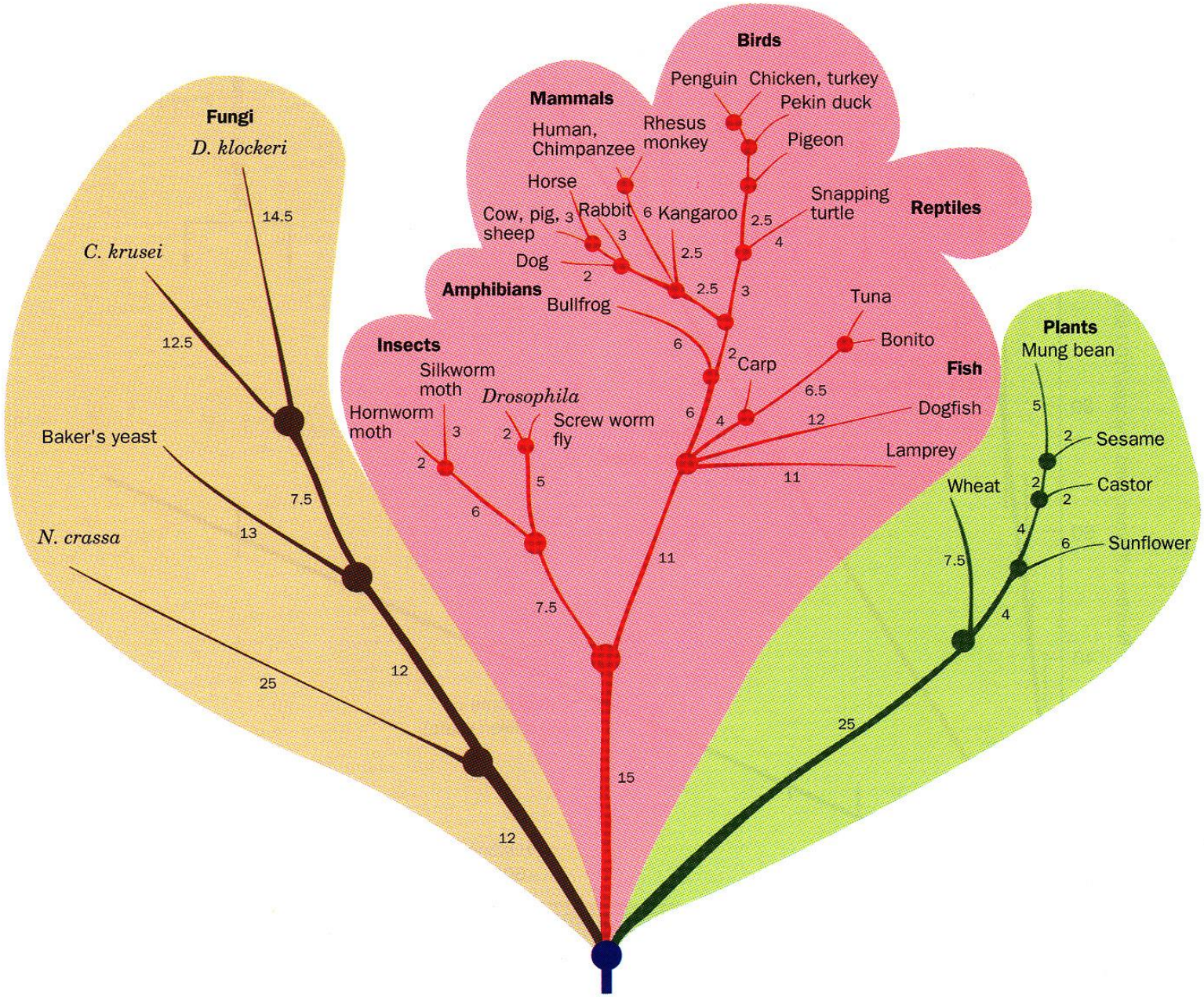
GroEL Phylogeny



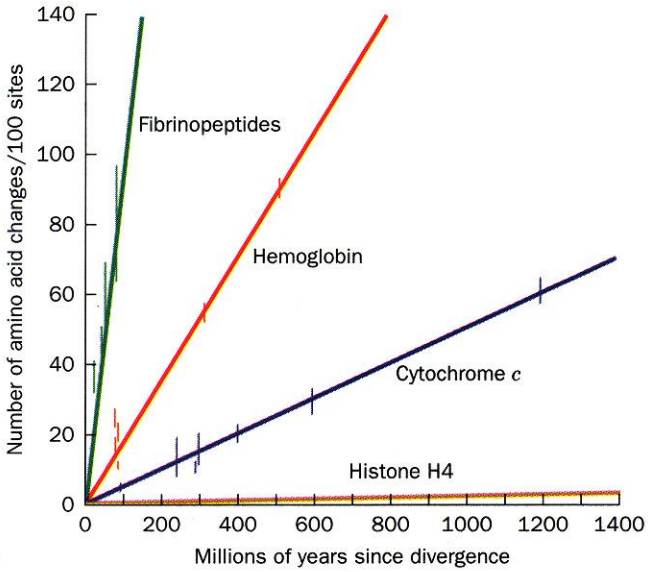
Cytochrome C



Cytochrome C Phyolgeny



Proteins Evolve at Different Rates



Simple Sequencing Problem: What Each Part Tells You

18. Sequence of Leucine Enkephalin, a brain opioid peptide.

a. Complete hydrolysis by 6M HCl at 110^{o}C followed by amino acid analysis indicated the presence of G, L, F, and Y in a 2:1:1:1 molar ratio.

This means the peptide could be 2:1:1:1, or 4:2:2:2, or….

Simple Sequencing Problem: What Each Part Tells You

18. Sequence of Leucine Enkephalin, a brain opioid peptide.

a. Complete hydrolysis by 6M HCl at 110^{o}C followed by amino acid analysis indicated the presence of G, L, F, and Y in a 2:1:1:1 molar ratio.

b. Treatment of the peptide with 1-fluoro-2,4, dinitrobenzene followed by complete hydrolysis and chromatography indicated the presence of 2,4-dinitrophenyl derivative of tyrosine. No free tyrosine could be found.

What does this tell you?

Simple Sequencing Problem: What Each Part Tells You

18. Sequence of Leucine Enkephalin, a brain opioid peptide.

a. Complete hydrolysis by 6M HCl at 110^{o}C followed by amino acid analysis indicated the presence of G, L, F, and Y in a 2:1:1:1 molar ratio.

b. Treatment of the peptide with 1-fluoro-2,4, dinitrobenzene followed by complete hydrolysis and chromatography indicated the presence of 2,4-dinitrophenyl derivative of tyrosine. No free tyrosine could be found.

What does this tell you?

Y is the N-terminal amino acid: so the peptide is Y _ _ _ _

Simple Sequencing Problem: What Each Part Tells You

18. Sequence of Leucine Enkephalin, a brain opioid peptide.

a. Complete hydrolysis by 6M HCl at 110^{o}C followed by amino acid analysis indicated the presence of G, L, F, and Y in a 2:1:1:1 molar ratio.

b. Treatment of the peptide with 1-fluoro-2,4, dinitrobenzene followed by complete hydrolysis and chromatography indicated the presence of 2,4-dinitrophenyl derivative of tyrosine. No free tyrosine could be found.

the peptide is Y _ _ _ _

c. Complete digestion of the peptide with chymotrypsin followed by chromatography yielded free tyrosine and leucine with a tripeptide containing Phe and Gly in a 1:2 ratio.

Simple Sequencing Problem: What Each Part Tells You

18. Sequence of Leucine Enkephalin, a brain opioid peptide.

a. Complete hydrolysis by 6M HCl at 110^{o}C followed by amino acid analysis indicated the presence of G, L, F, and Y in a 2:1:1:1 molar ratio.

b. Treatment of the peptide with 1-fluoro-2,4, dinitrobenzene followed by complete hydrolysis and chromatography indicated the presence of 2,4-dinitrophenyl derivative of tyrosine. No free tyrosine could be found.

the peptide is Y _ _ _ _

c. Complete digestion of the peptide with chymotrypsin followed by chromatography yielded free tyrosine and leucine with a tripeptide containing Phe and Gly in a 1:2 ratio.

so… it is YGGFL