

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

Primary structure

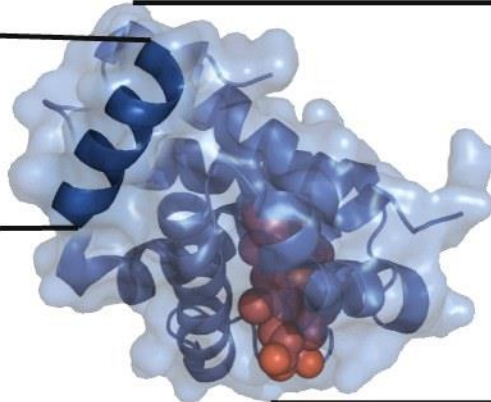
Pro
Ala
Asp
Lys
Thr
Asn
Val
Lys
Ala
Ala
Trp
Gly
Lys
Val

Secondary structure



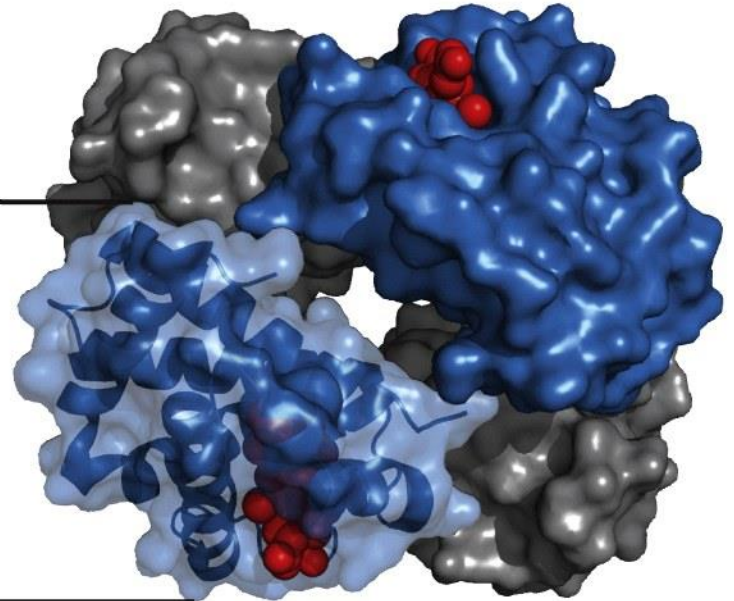
α Helix

Tertiary structure



Polypeptide chain

Quaternary structure



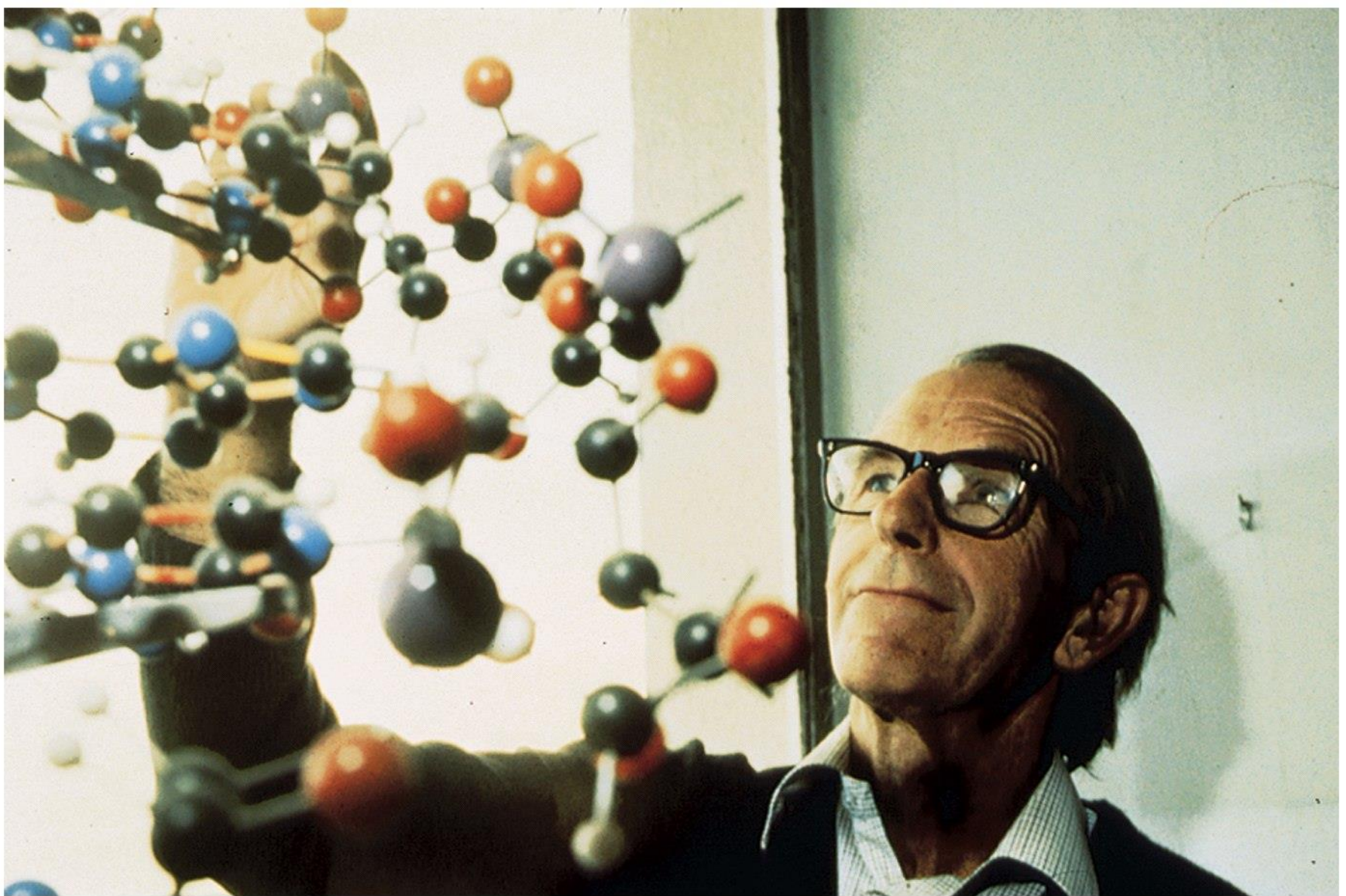
Assembled subunits

Amino acid residues

Figure 3-23

Lehninger Principles of Biochemistry, Fifth Edition

© 2008 W. H. Freeman and Company



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

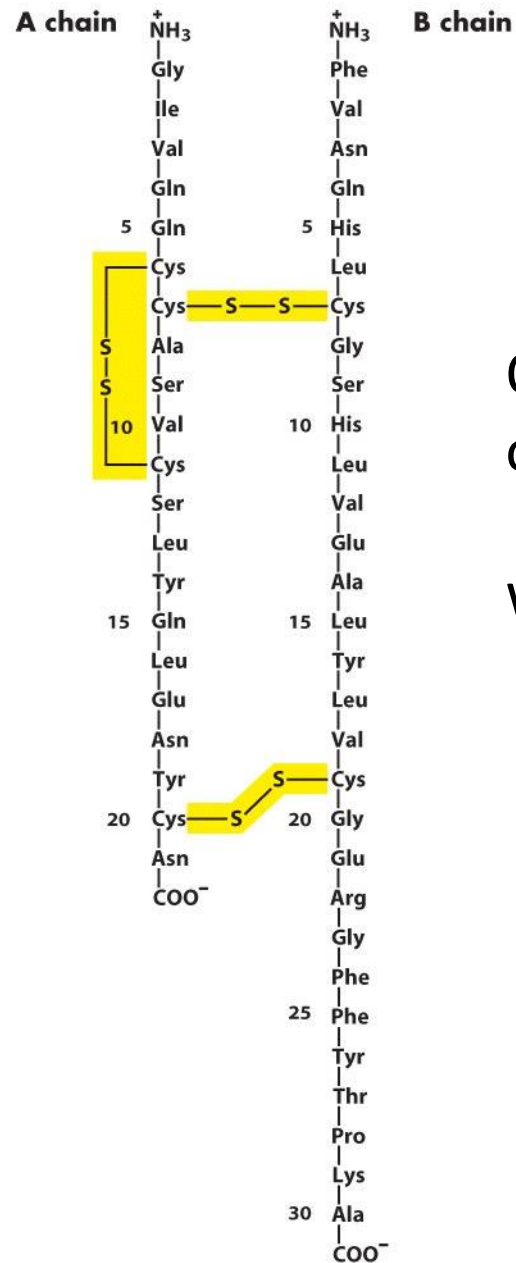
Edman Degradation (Classical method)

- 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 a pure protein is obtained, then....

What needs to be done ?

R-S-S-R

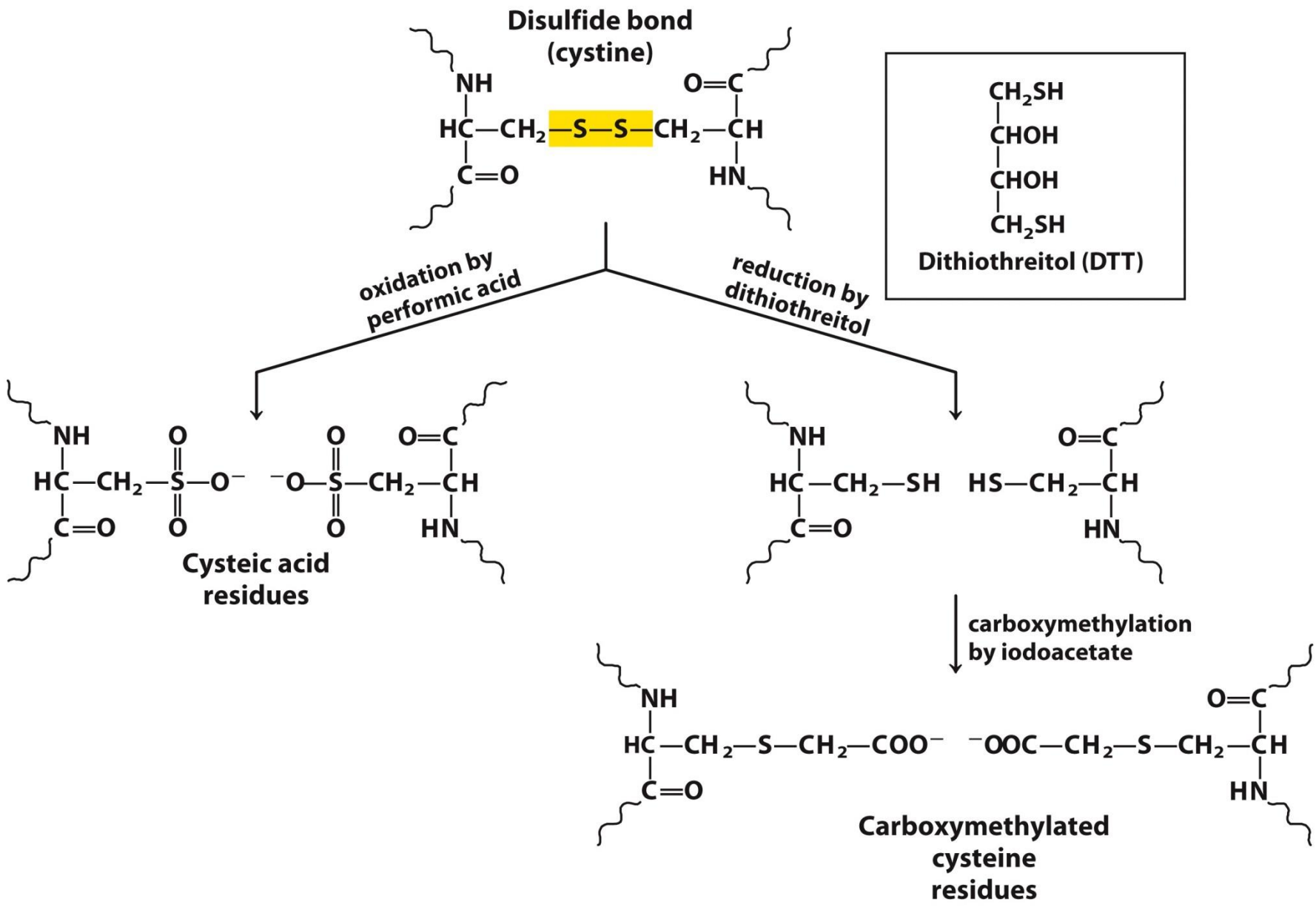


Figure 3-28
Lehninger Principles of Biochemistry, Sixth Edition
 © 2013 W. H. Freeman and Company

Acid Hydrolysis of Proteins

6 N HCl for several hours, 100-105°C, to hydrolyze Peptide bonds

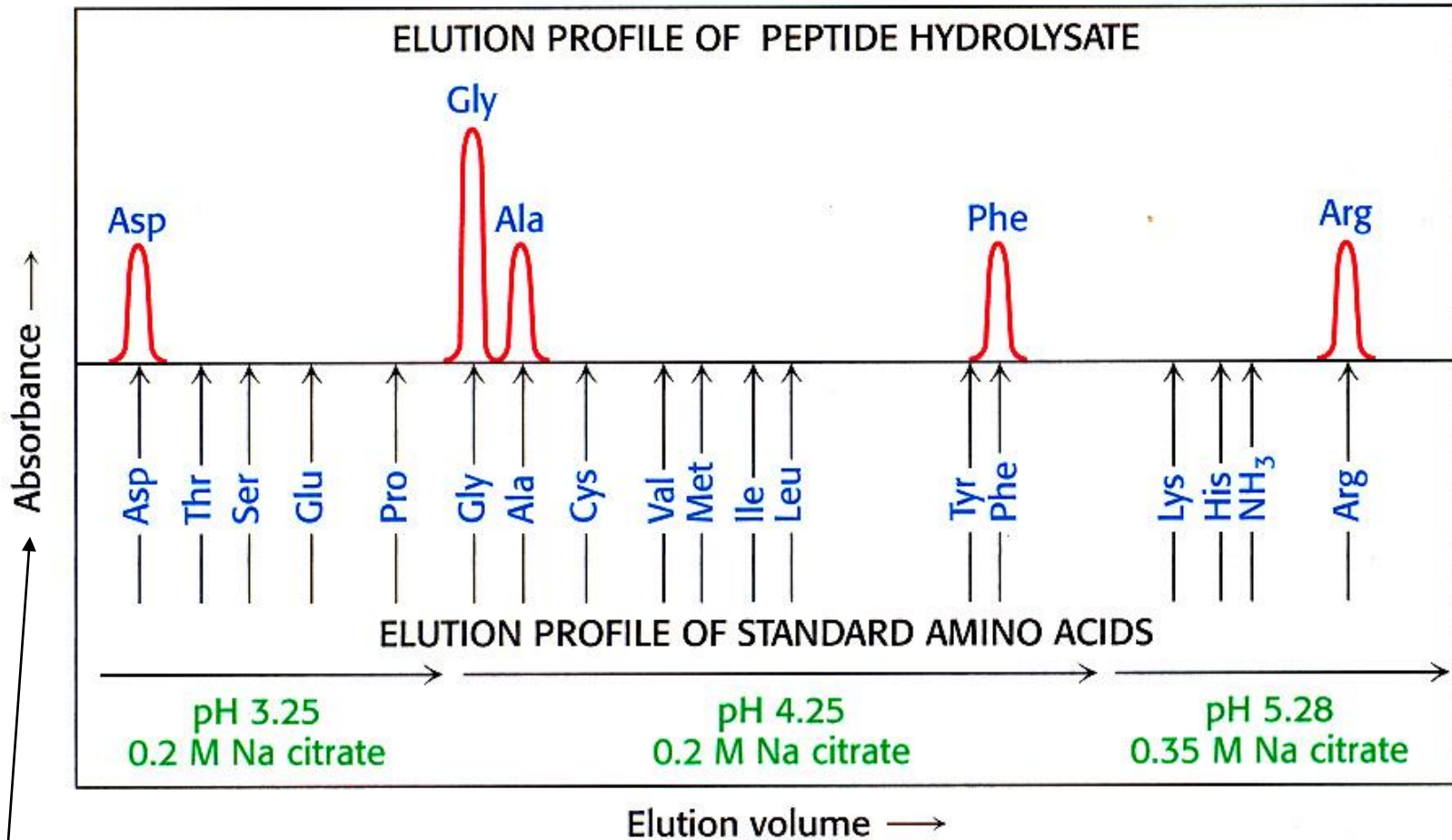
Four Problems to Deal With:

1. Destroys W.
2. Partially (slowly) destroys S and T.
3. Converts N \rightarrow D + NH_4^+ , and Q \rightarrow 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 \rightarrow D + E.

Amino Acid Analysis of AGDFRG

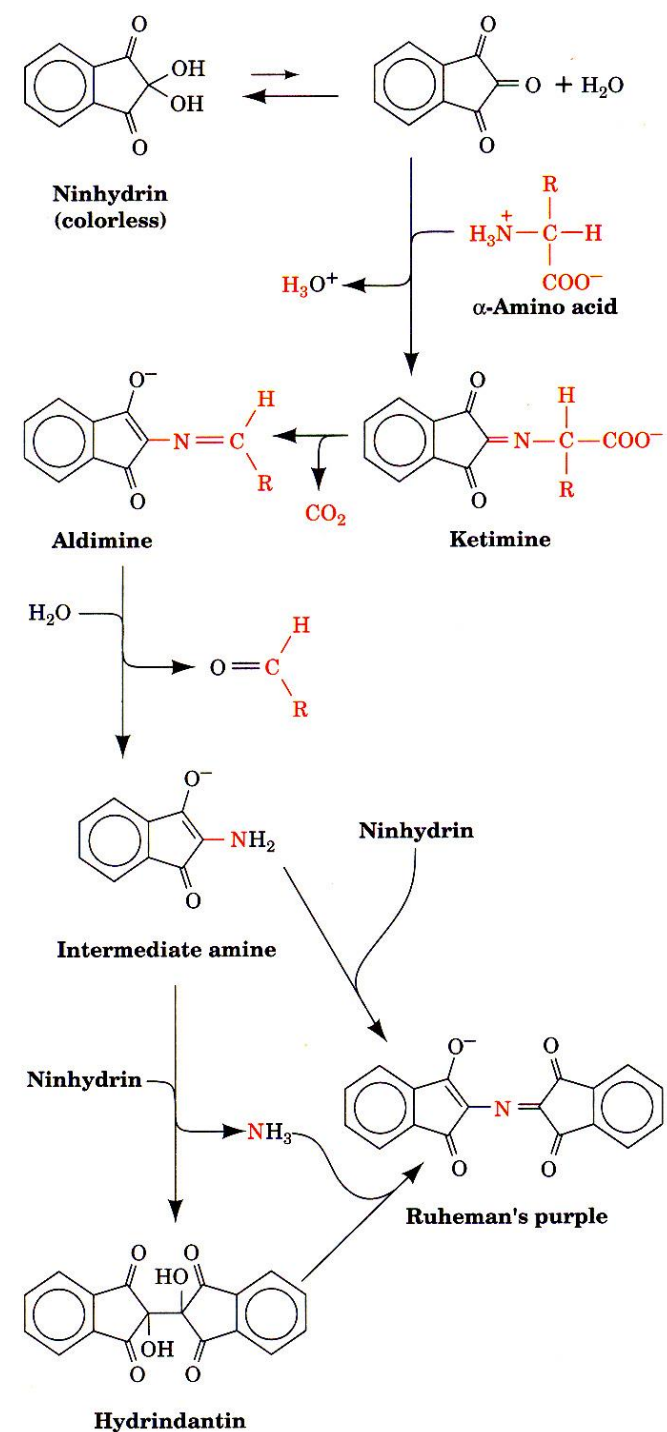


Based on Ninhydrin Reaction

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

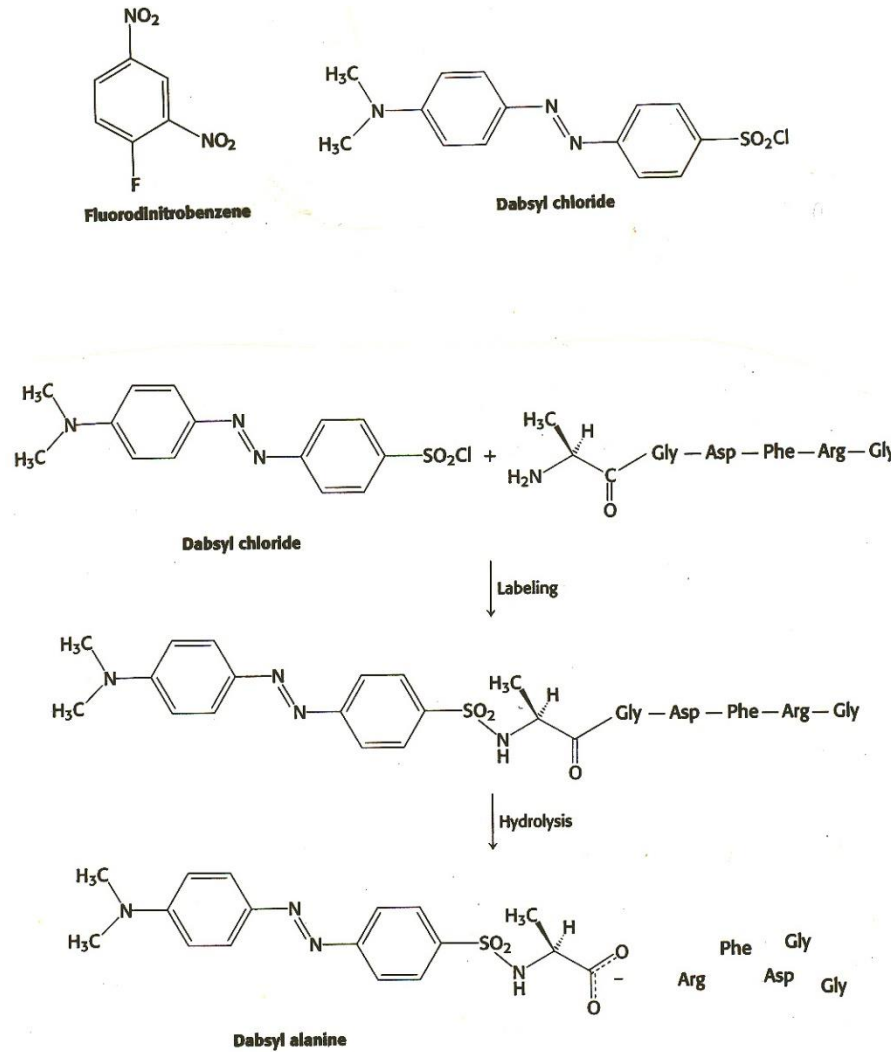


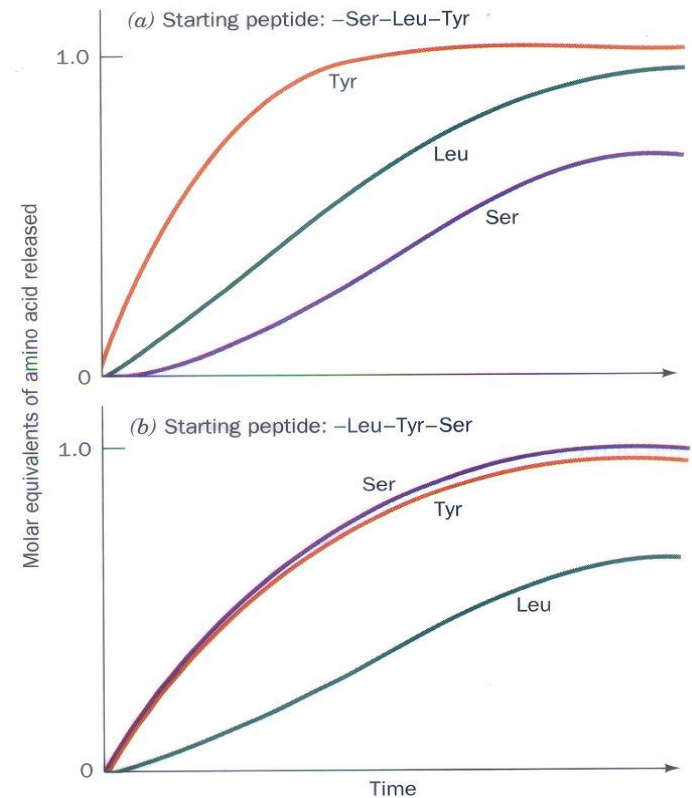
FIGURE 4.20 Determination of the amino-terminal residue of a peptide. Dabsyl chloride labels the peptide, which is then hydrolyzed with the use of hydrochloric acid. The dabsyl-amino acid (dabsyl-alanine in this example) is identified by its chromatographic characteristics.

C terminal Carboxypeptidases

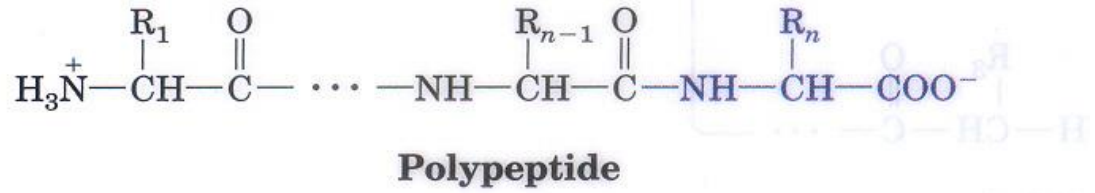
Table 7-1 Specificities of Various Exopeptidases

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

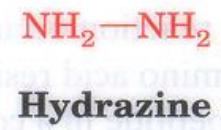
^a R_1 = the N-terminal residue; R_n = the C-terminal residue.



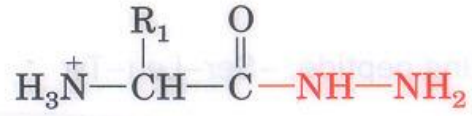
Hydrazinolysis



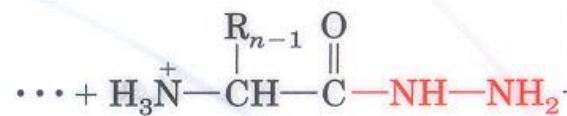
+



↓
acidic ion exchange
resin catalyst



+



+



Free amino acid

**Aminoacyl
hydrazides**

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

Modifies peptide carboxyl group
to form acyl hydrazide

Edman Degradation

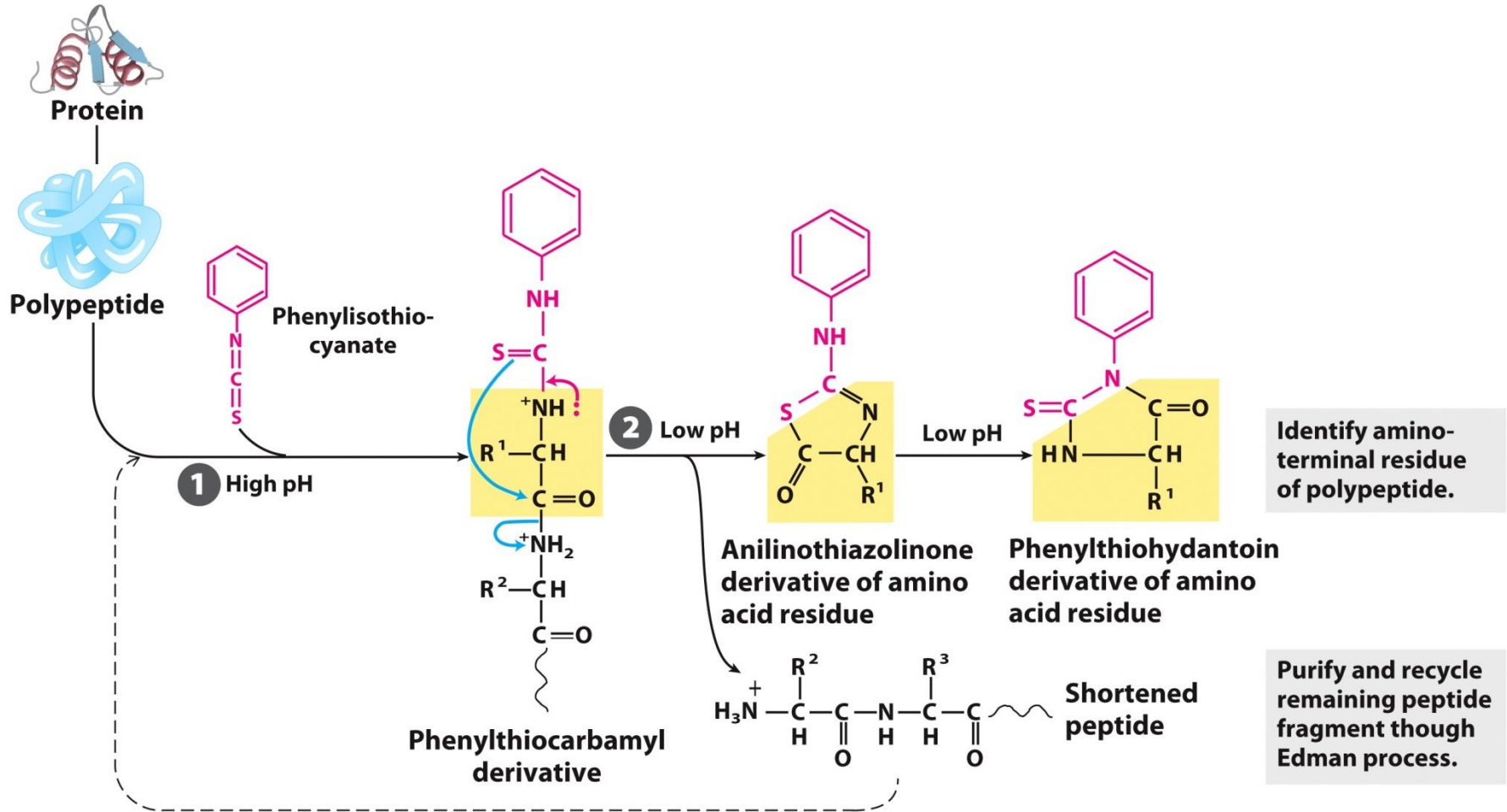
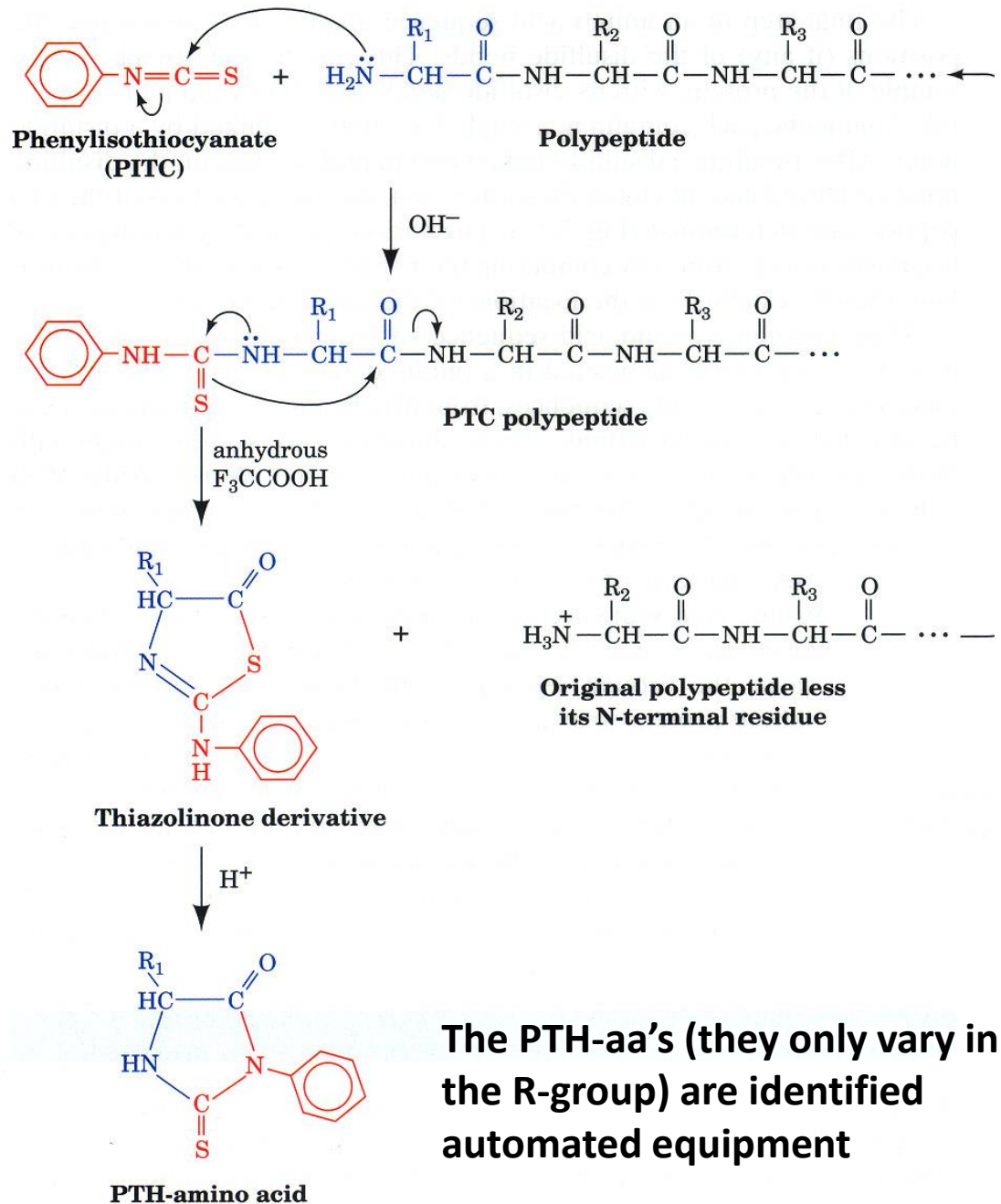


Figure 3-27
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Edman Degradation = Amino Acid Sequencing



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)
<i>Staphylococcus aureus</i> V8 protease (bacterium <i>S. aureus</i>)	Asp, Glu (C)
Asp-N-protease (bacterium <i>Pseudomonas fragi</i>)	Asp, Glu (N)
Pepsin (porcine stomach)	Leu, Phe, Trp, Tyr (N)
Endoproteinase Lys C (bacterium <i>Lysobacter enzymogenes</i>)	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

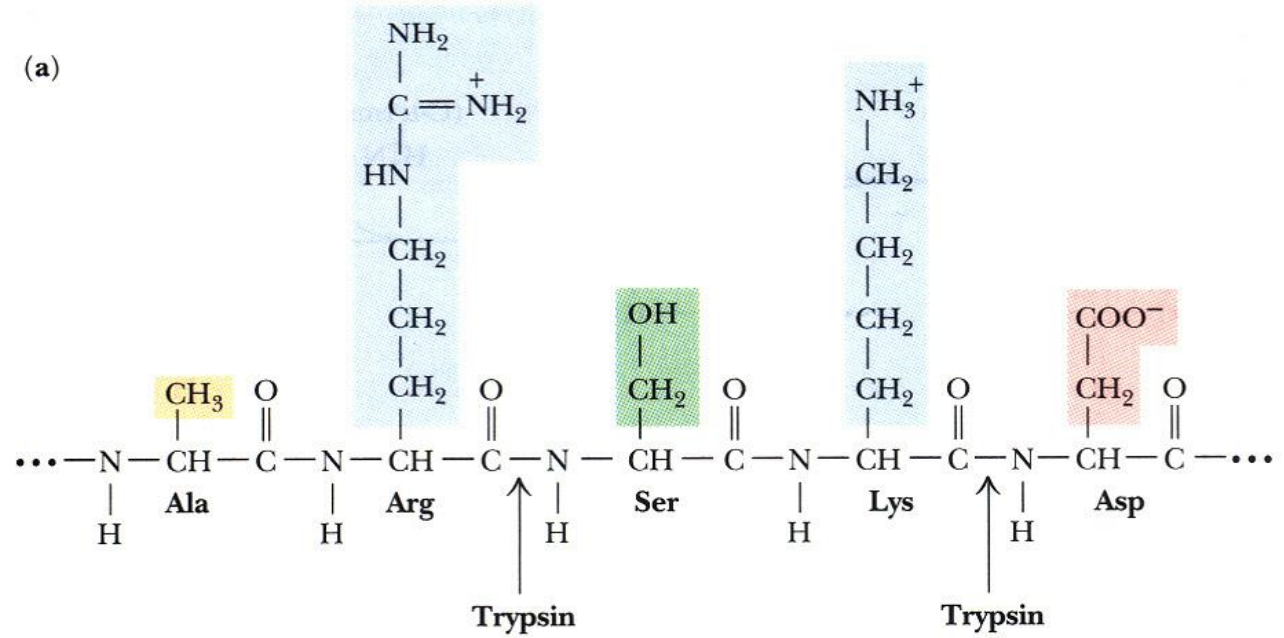
Lehninger Principles of Biochemistry, Sixth Edition

© 2013 W. H. Freeman and Company

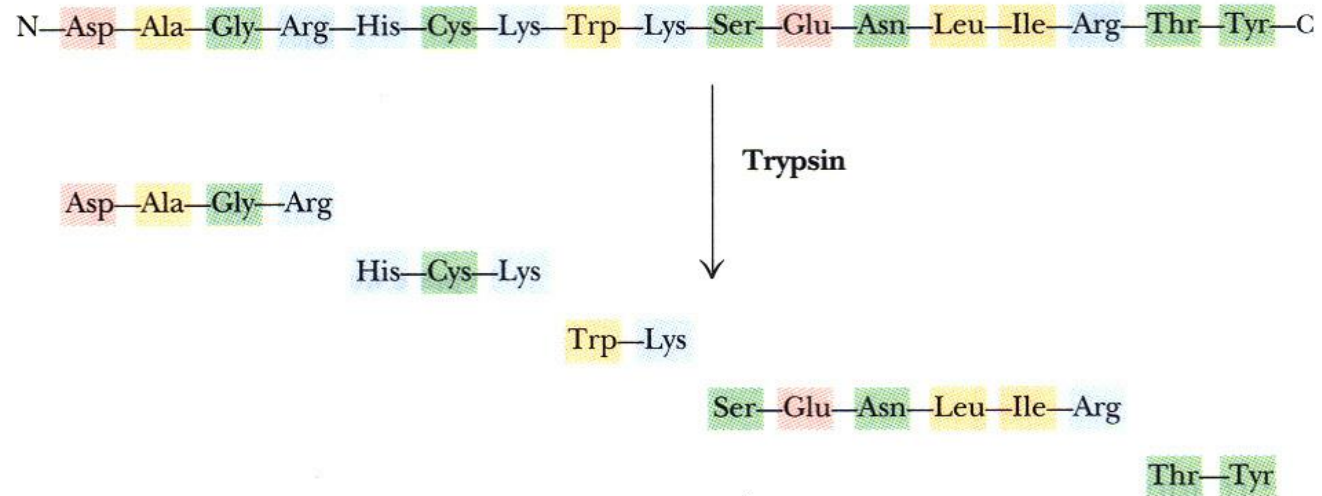
You need to know 4 of them: trypsin, chymotrypsin, pepsin and cyanogen bromide fragmentation methods.

Trypsin Fragmentation

(a)



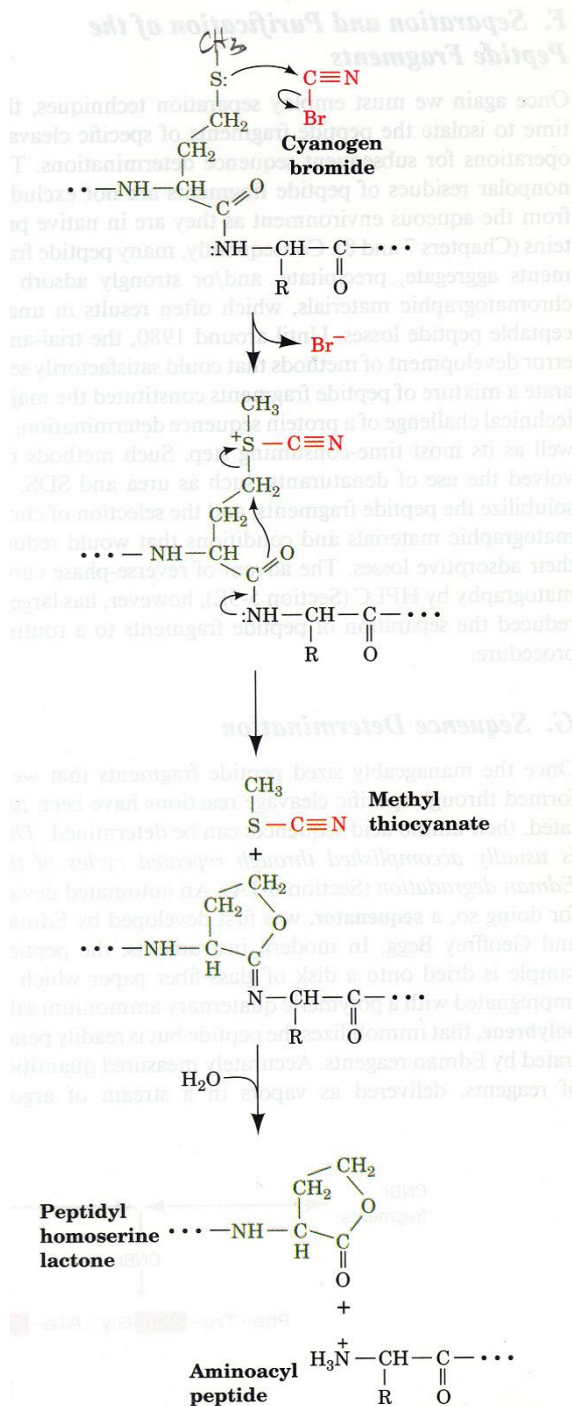
(b)



CNBr Fragmentation

Cleaves the C-terminal
side of Met...

....and converts the Met
to Homoserine Lactone



Protein Sequencing Overall Flow

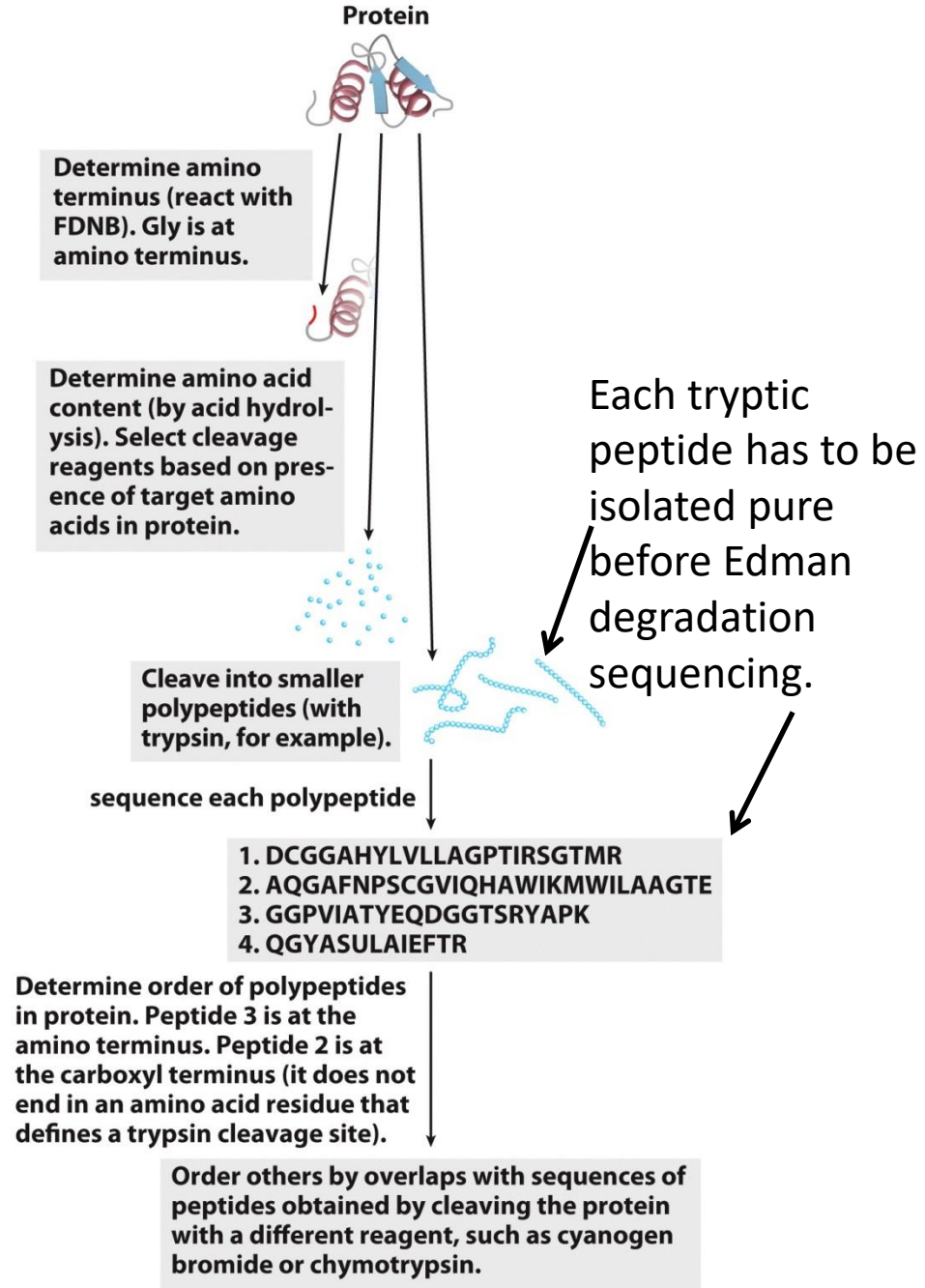
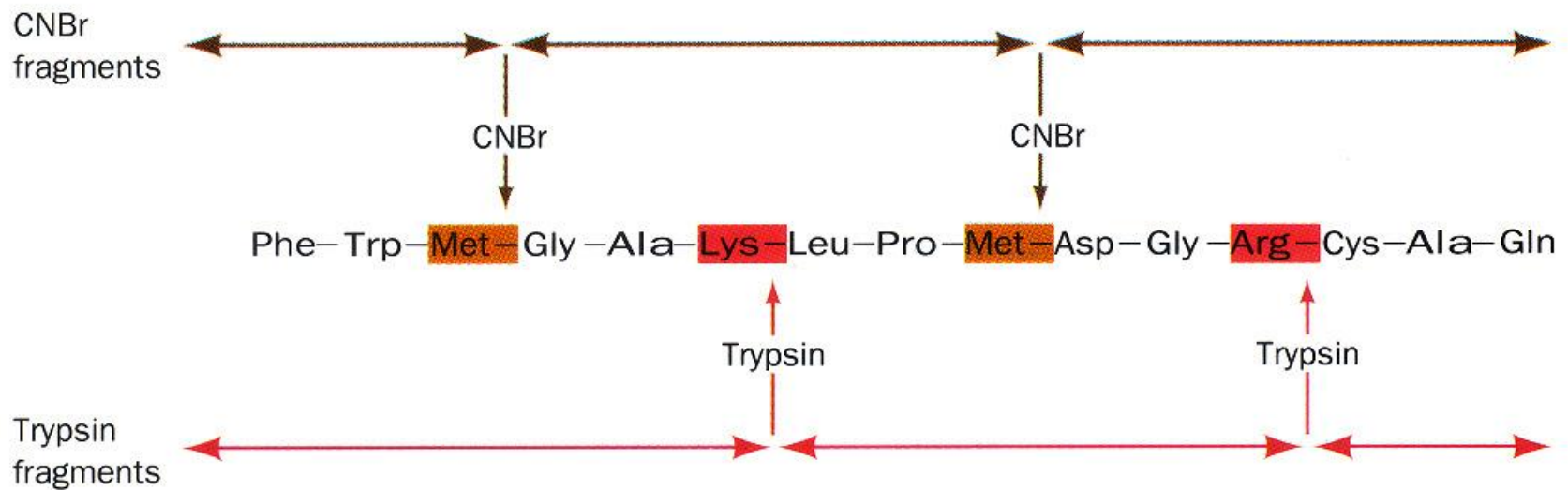


Figure 3-25

Lehninger Principles of Biochemistry, Sixth Edition

© 2013 W. H. Freeman and Company

Trypsin and CNBr Fragmentation Each Done Separately



Separation of Protein Fragments

The Classic Paper Chromatography + Electrophoresis

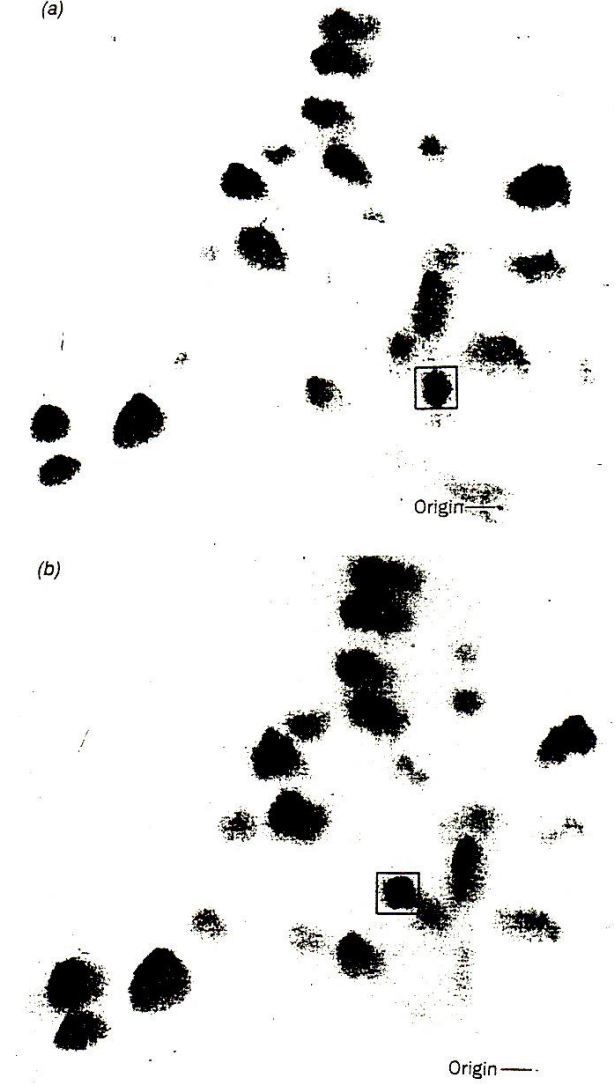
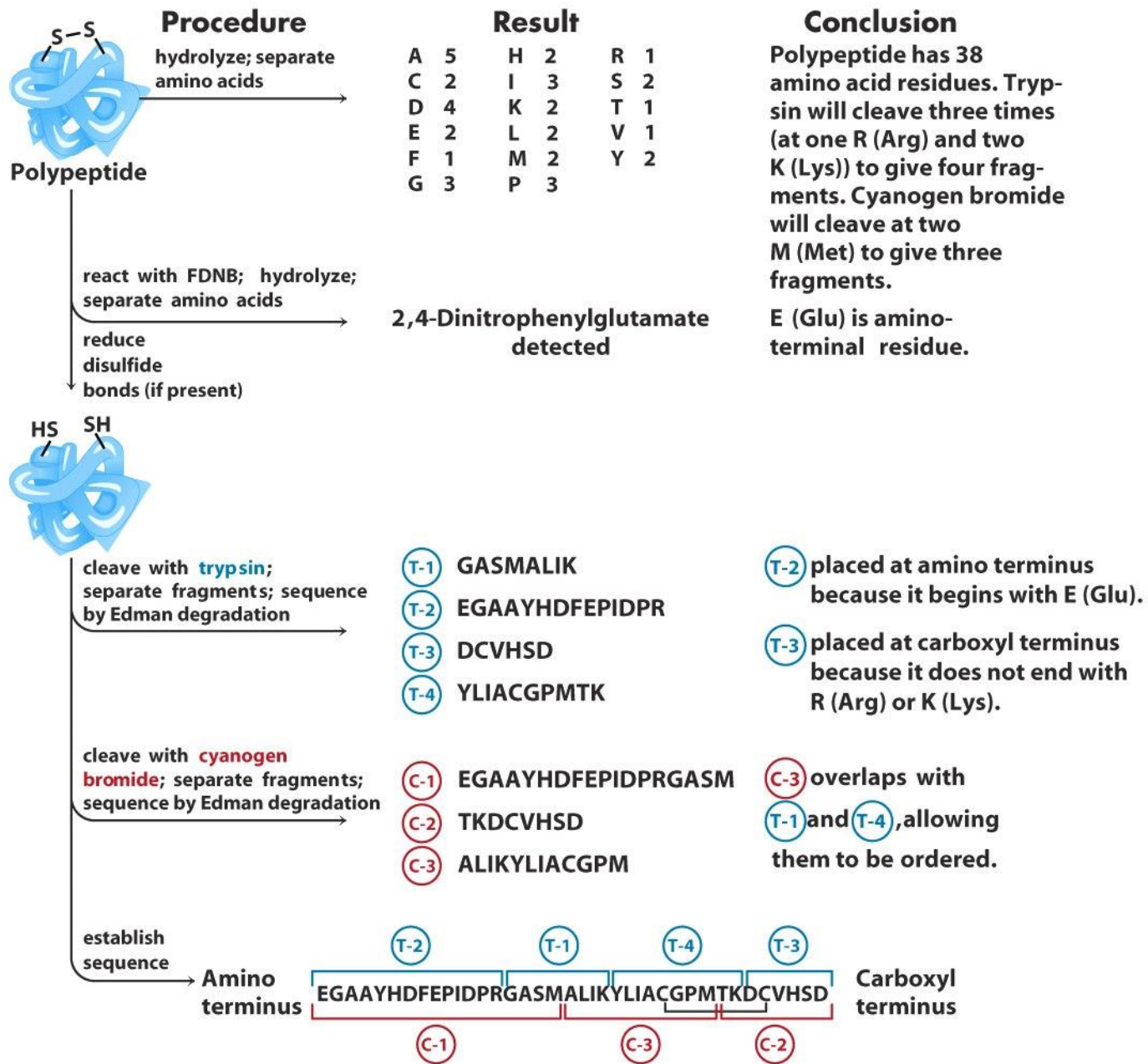


FIGURE 6-10. A comparison of the ninhydrin-stained fingerprints of trypsin-digested (a) hemoglobin A (HbA) and (b) hemoglobin S (HbS). The peptides that differ in these two forms of hemoglobin are boxed. These peptides constitute the eight N-terminal residues of the β subunit of hemoglobin. Their amino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

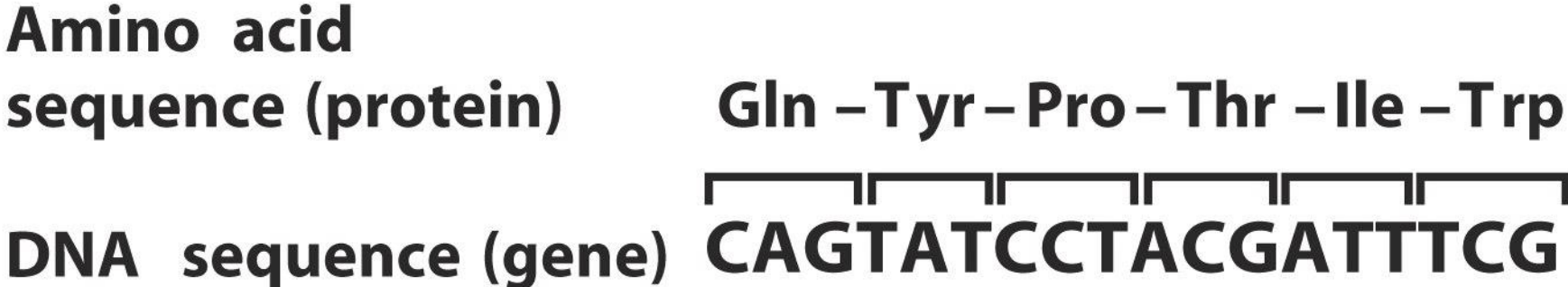
Hemoglobin A	Val-	His-	Leu-	Thr-	Pro-	Glu-	Glu-	Lys
Hemoglobin S	Val-	His-	Leu-	Thr-	Pro-	Val-	Glu-	Lys
	β 1	2	3	4	5	6	7	8

[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

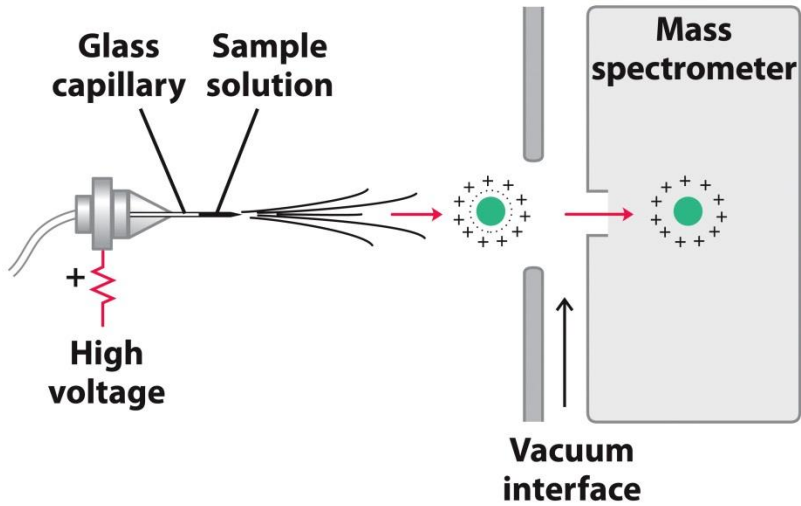


Figure 3-30a
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

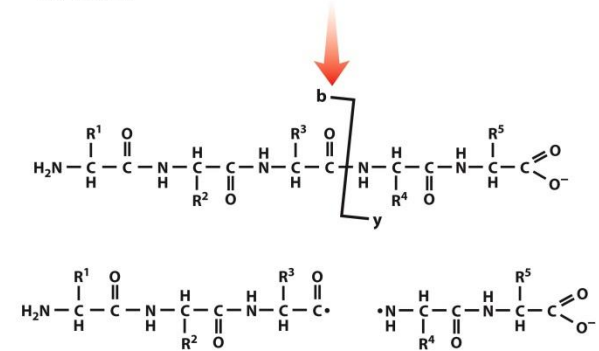
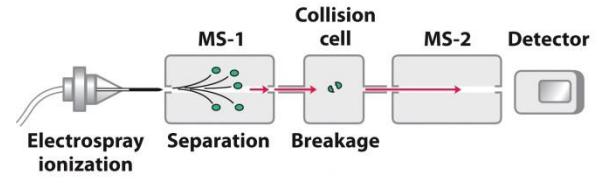


Figure 3-31a
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

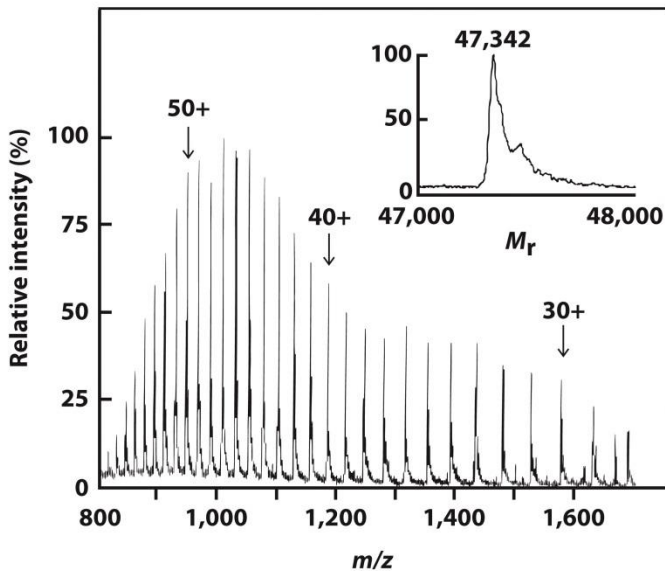


Figure 3-30b
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

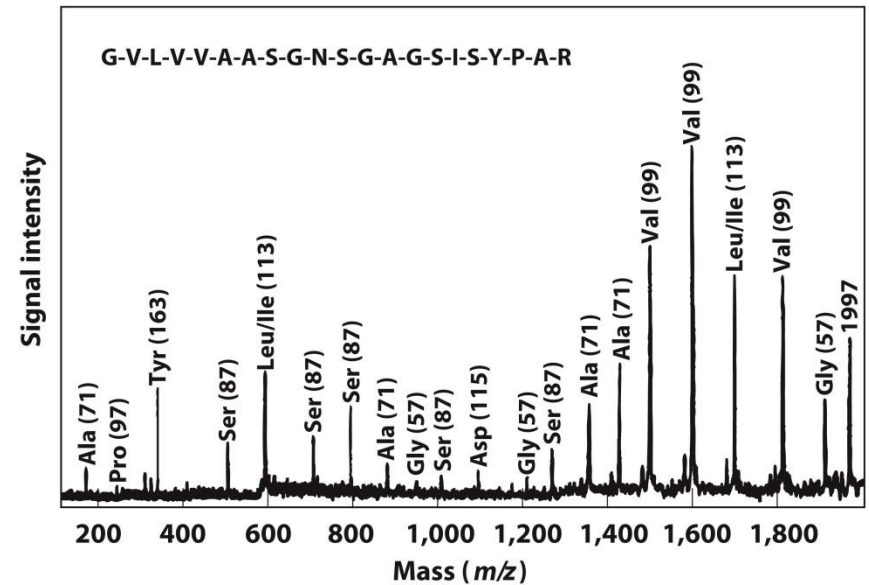
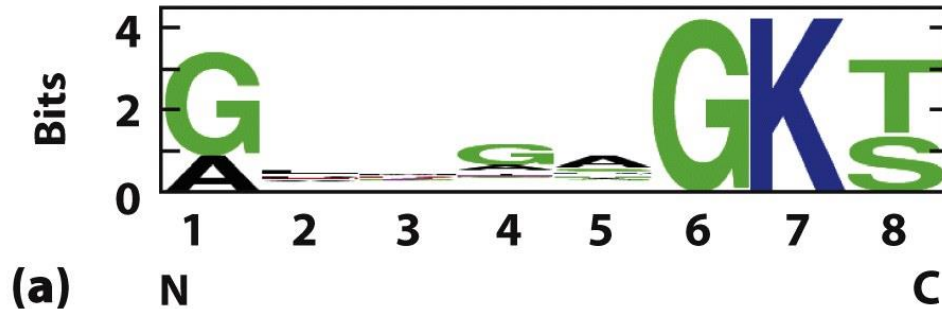


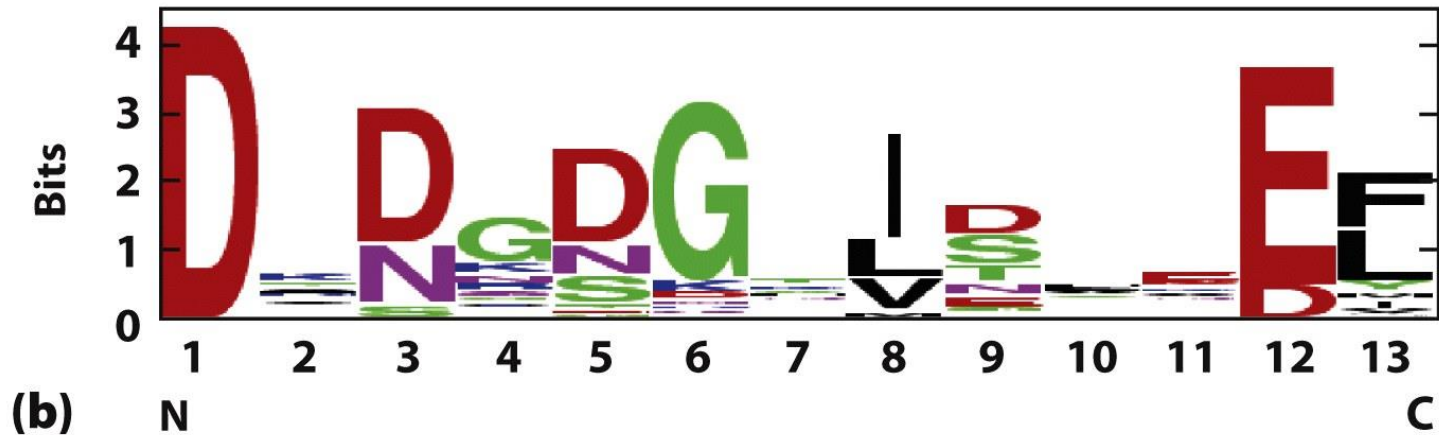
Figure 3-31b
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Consensus Sequences

[AG]-x(4)-G-K-[ST].



D-{W}-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHHRK]-{GP}-
[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW].



Partial Primary Structure of Elongation Factor Tu Sequences Aligned with Gaps

E. coli

TGNRTIAVYDLGGGTFDISIIEIDEVDGEKTFEVLATNGDTHLGGEDFDSRLIHYL
DEDQTILLYDLGGGTFDVSILELGDG TFEVRSTAGDNRLGGDDFDQVIIDHL

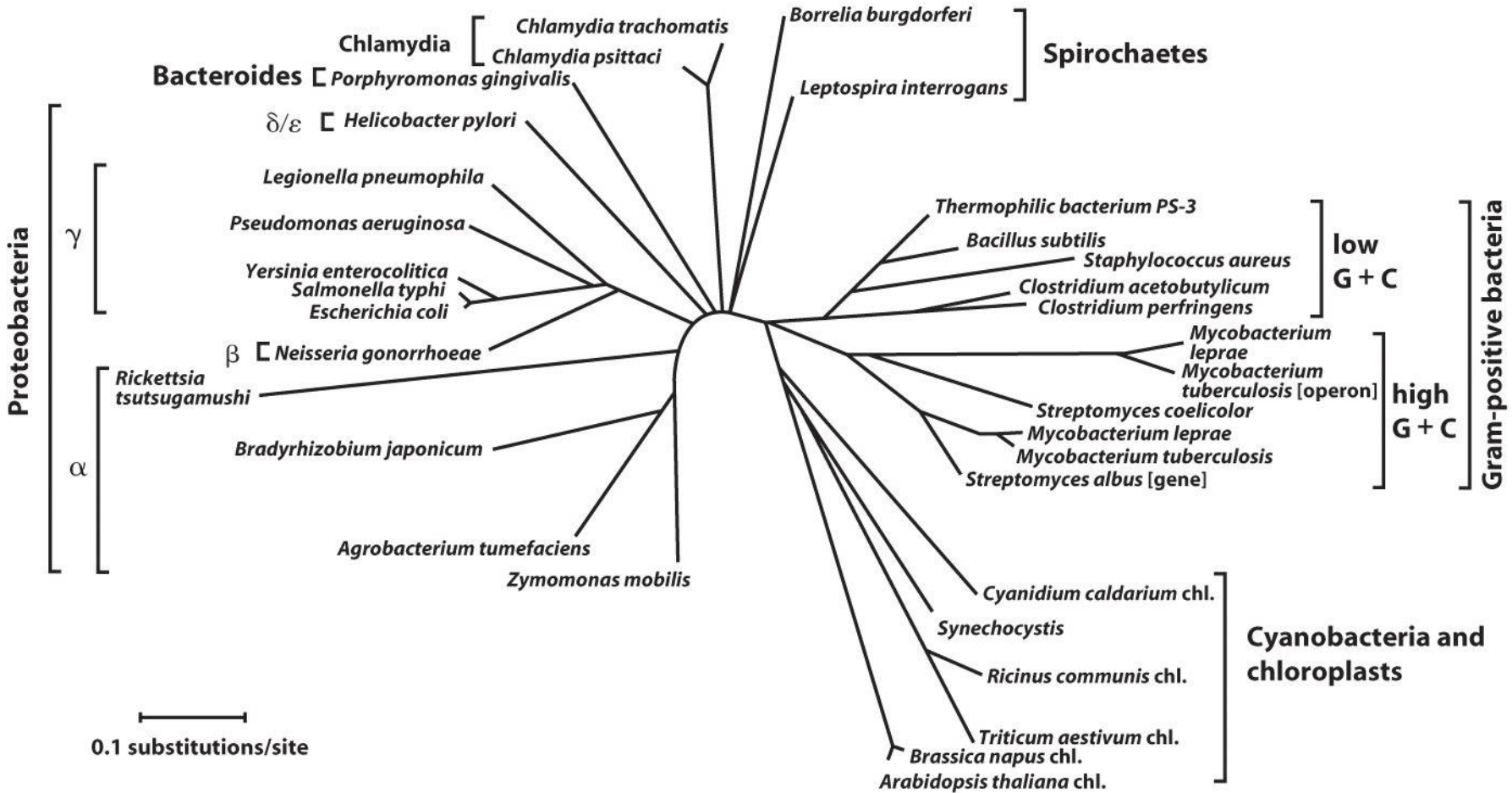
B. subtilis

┌
└
Gap

EF-Tu Signature Sequences

		Signature sequence																																					
Archaeobacteria	{	<i>Halobacterium halobium</i>	I	G	H	V	D	H	G	K	S	T	M	V	G	R	L	L	Y	E	T	G	S	V	P	E	H	V	I	E	Q	H							
		<i>Sulfolobus solfataricus</i>	I	G	H	V	D	H	G	K	S	T	L	V	G	R	L	L	M	D	R	G	F	I	D	E	K	T	V	K	E	A							
Eukaryotes	{	<i>Saccharomyces cerevisiae</i>	I	G	H	V	D	S	G	K	S	T	T	T	G	H	L	I	Y	K	C	G	G	I	D	K	R	T	I	E	K	F							
		<i>Homo sapiens</i>	I	G	H	V	D	S	G	K	S	T	T	T	G	H	L	I	Y	K	C	G	G	I	D	K	R	T	I	E	K	F							
Gram-positive bacterium		<i>Bacillus subtilis</i>	I	G	H	V	D	H	G	K	S	T	M	V	G	R																	I	T	T	V			
Gram-negative bacterium		<i>Escherichia coli</i>	I	G	H	V	D	H	G	K	T	T	L	T	A	A																				I	T	T	V

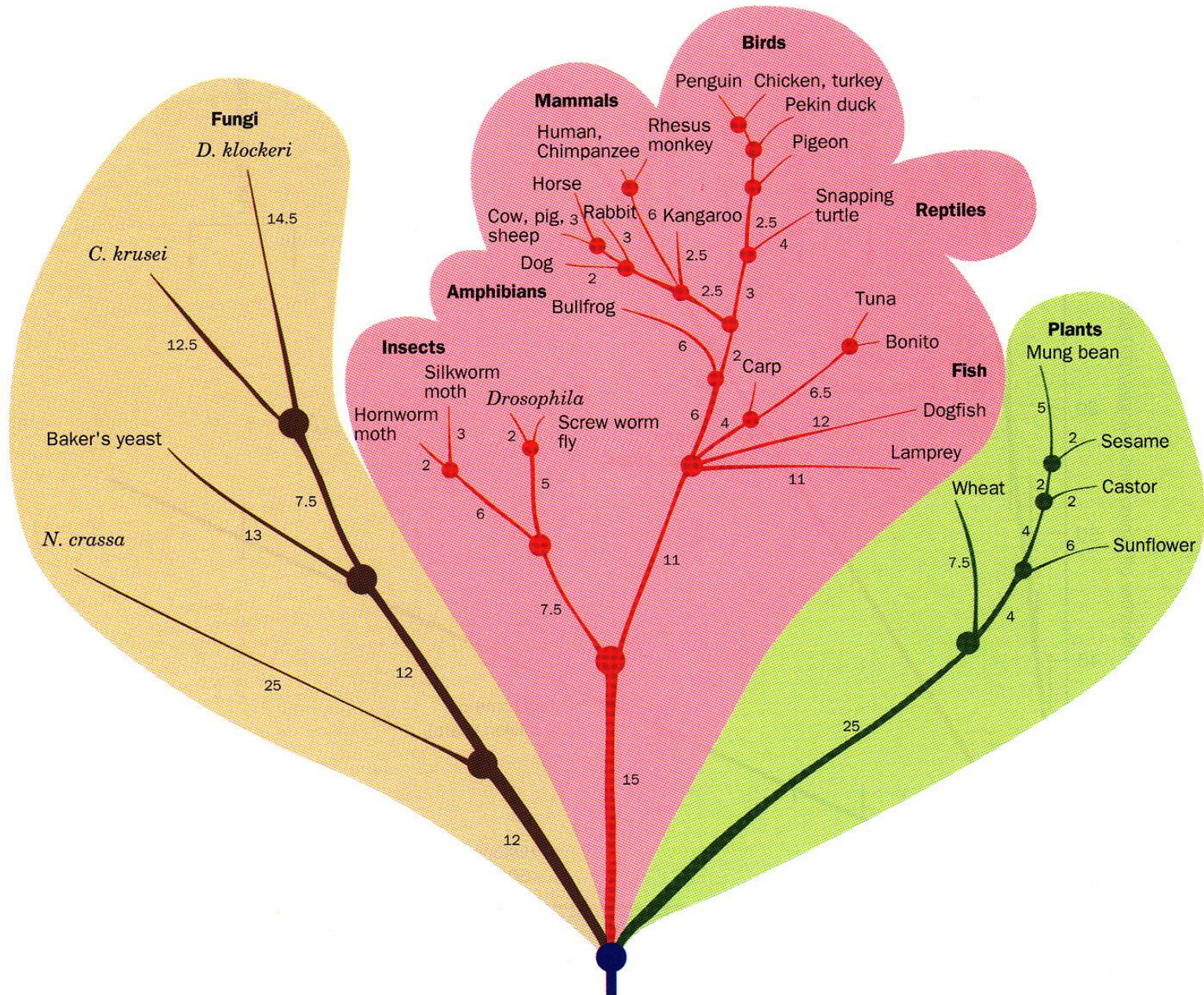
GroEL Phylogeny



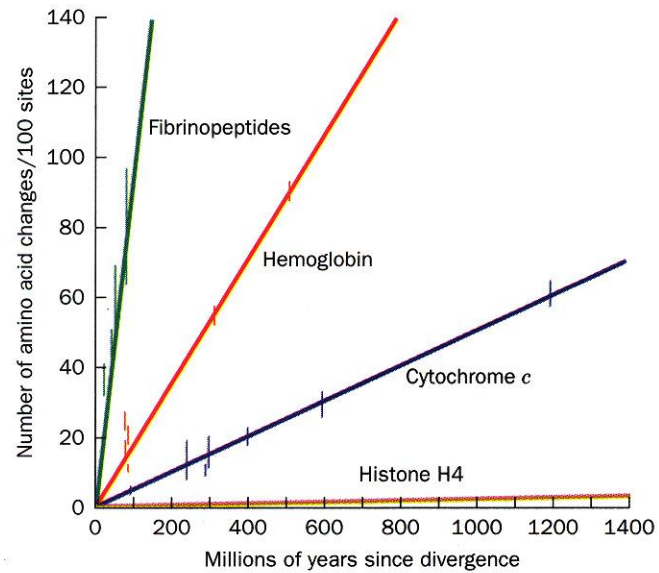
Cytochrome C

	10	20	30	40	50	60	70	80	90	100	
Human	GDVEKGGKKIF	IMKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAPGYSYTA	ANKNKG I I WG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKKEE	RADLIAYLKK	ATNE
Chimpanzee	GDVEKGGKKIF	IMKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAPGYSYTA	ANKNKG I I WG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKKEE	RADLIAYLKK	ATNE
Spider monkey	GDVFKGKRIF	IMKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQASGFYTYE	ANKNKG I I WG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKKEE	RADLIAYLKK	ATNE
Macaque	GDVEKGGKKIF	IMKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAPGYSYTA	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFVGIKKKKEE	RADLIAYLKK	ATNE
Cow	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAPGFSYTD	ANKNKG I TWG	EETLMEYLEN	PKKYIPGTKM	IFAGI KKKGE	REDLIAYLKK	ATNE
Dog	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAPGFSYTD	ANKNKG I TWG	EETLMEYLEN	PKKYIPGTKM	IFAGI KKTGE	RADLIAYLKK	ATKE
Gray whale	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAVGFSYTD	ANKNKG I TWG	EETLMEYLEN	PKKYIPGTKM	IFAGI KKKGE	RADLIAYLKK	ATNE
Horse	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAPGFYTYD	ANKNKG I TWK	EETLMEYLEN	PKKYIPGTKM	IFAGI KKKTE	REDLIAYLKK	ATNE
Zebra	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAPGFSYTD	ANKNKG I TWK	EETLMEYLEN	PKKYIPGTKM	IFAGI KKKTE	REDLIAYLKK	ATNE
Rabbit	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAVGFSYTD	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKDE	RADLIAYLKK	ATNE
Kangaroo	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGI FGRKT	GQAPGFYTYD	ANKNKG I I WG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKGE	RADLIAYLKK	ATNE
Duck	GDVEKGGKKIF	VQKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKSE	RADLIAYLKD	ATAK
Turkey	GD I EKGGKKIF	VQKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKSE	RVDLIAYLKD	ATSK
Chicken	GD I EKGGKKIF	VQKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKSE	RVDLIAYLKD	ATSK
Pigeon	GD I EKGGKKIF	VQKCSQCHTV	EKGGKHKHTGP	NLHGLFGRKT	GQAEGFSYTD	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKAE	RADLIAYLKQ	ATAK
King penguin	GD I EKGGKKIF	VQKCSQCHTV	EKGGKHKHTGP	NLHGI FGRKT	GQAEGFSYTD	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKSE	RADLIAYLKD	ATSK
Snapping turtle	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLNGLI GRKT	GQAEGFSYTE	ANKNKG I TWG	EETLMEYLEN	PKKYIPGTKM	IFAGI KKKAE	RADLIAYLKD	ATSK
Alligator	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKHTGP	NLHGLI GRKT	GQAPGFSYTE	ANKNKG I TWG	EETLMEYLEN	PKKYIPGTKM	IFAGI KKKPE	RADLIAYLKE	ATSN
Bull frog	GDVEKGGKKIF	VQKCAQCHTV	EKGGKHKVGP	NLYGLI GRKT	GQAAGFSYTD	ANKNKG I TWG	EDTLMEYLEN	PKKYIPGTKM	IFAGI KKKGE	RQDLIAYLKS	ACSK
Tuna	GDVAKGKKTIF	VQKCAQCHTV	ENGGKHKVGP	NLWGLFGRKT	GQABGYSYTD	ANKSKGI VWN	ENTLMEYLEN	PKKYIPGTKM	IFAGI KKKGE	RQDLVAYLKS	ATS
Dogfish	GDVEKGGKVF	VQKCAQCHTV	ENGGKHKVGP	NLSGLFGRKT	GQAQGFYSYTD	ANKSKGI TWQ	QETLR IYLEN	PKKYIPGTKM	IFAGI KKKSE	RQDLIAYLKK	TAAS
Starfish	GQVEKGGKKIF	VQRCAQCHTV	EKAGKHKHTGP	NLNGILGRKT	GQAAGFSYTD	ANRNKG I TWK	NETLFEYLEN	PKKYIPGTKM	VFAGLKKQKE	RQDLIAYLEA	ATK
Fruit fly	GDVEKGGKLF	VQRCAQCHTV	EAGGKHKVGP	NLHGLI GRKT	GQAAGFAYTD	ANKAKGI TWN	EDTLFEYLEN	PKKYIPGTKM	IFAGLKKPNE	RGDLIAYLKS	ATK
Silkmoth	GNAENGKKIF	VQRCAQCHTV	EAGGKHKVGP	NLHGFYGRKT	GQAPGFSYSN	ANKAKGI TWG	DDTLFEYLEN	PKKYIPGTKM	VFAGLKKANE	RADLIAYLKE	STK
Pumpkin	GNSKAGEKIF	KTKCAQCHTV	DKGAGHKQGP	NLNGLFGRQS	GTTPGYSYSA	ANKNRAVI WE	EKTLYDYLLN	PKKYIPGTKM	VFPGLKKPQD	RADLIAYLKE	ATA
Tomato	GNPKAGEKIF	KTKCAQCHTV	EKGAGHKEGP	NLNGLFGRQS	GTTAGYSYSA	ANKNMAVNWG	ENTLYDYLLN	PKKYIPGTKM	VFPGLKKPQE	RADLIAYLKE	ATA
Arabidopsis	GDAKKGANLF	KTRCAQCHTL	KAGEGNGI GP	ELHGLFGRKT	GSVAGYSYTD	ANKQKGI EWK	DDTLFEYLEN	PKKYIPGTKM	AFGGLKKPKD	RNDLI T FLEE	ETK
Mung bean	GNSKSGEKIF	KTKCAQCHTV	DKGAGHKQGP	NLNGLI GRQS	GTTAGYSYST	ANKNMAVI WE	EKTLYDYLLN	PKKYIPGTKM	VFPGLKKPQD	RADLIAYLKE	STA
Wheat	GNPDAGAKIF	KTKCAQCHTV	DAGAGHKQGP	NLHGLFGRQS	GTTAGYSYSA	ANKNKAVEWE	ENTLYDYLLN	PKKYIPGTKM	VFPGLKKPQD	RADLIAYLKK	ATSS
Sunflower	GNPTTGEKIF	KTKCAQCHTV	EKGAGHKQGP	NLNGLFGRQS	GTTAGYSYSA	GNKNKAVI WE	ENTLYDYLLN	PKKYIPGTKM	VFPGLKKPQE	RADLIAYLKT	STA
Yeast	GSAKKGATLF	KTRCLQCHTV	EKGPPHKVGP	NLHGI FGRHS	GQAEGYSYTD	ANI KKNVLD	ENNMS EYLTN	PKKYIPGTKM	AFGGLKKEKD	RNDLI TYLKK	ACE
Debaryomyces	GSEKKGANLF	KTRCLQCHTV	EKGPPHKVGP	NLHG VVGRTS	GQAQGFYSYTD	ANKKKGVEWT	EQDLSDYLEN	PKKYIPGTKM	AFGGLKKAAD	RNDLI TYLVK	ATK
Candida	GSEKKGATLF	KTRCLQCHTV	EKGPPHKVGP	NLHG V FGRKS	GLAEGYSYTD	ANKKKGVEWT	EQTMSDYLEN	PKKYIPGTKM	AFGGLKKPKD	RNDLV TYLKK	ATS
Aspergillus	GDAK - GAKLF	QTRCAQCHTV	EAGGPHKVGP	NLHGLFGRKT	GQSEGYAYTD	ANKQAGVTWD	ENTLF SYLEN	PKKFIPGTKM	AFGGLKKGKE	RNDLI TYLKE	STA
Rhodomicrobium	GDPVKGQVQF	KQ - CKI CHQV	GPTAKNGVGP	EQNDVFGQKA	GARPGFNYSD	AMKNSGLTWD	EATLDKYLEN	PKAVVPGTKM	VFVGLKNPQD	RADV IAYLKKQ	LSGK
Nitrobacter	GDVEAGKAAF	NK - CKACHEI	GESAKNKVGP	ELDGLDGRHS	GAVEGYAYSP	ANKASGI TWT	EAEFK EYIKD	PKAKVPGTKM	VFAGI KKDSE	LDNLWAYVSQ	FDKD
Agrobacterium	GDVAKGEAAF	KR - CSACHAI	GEGAKNKVGP	QLNGI I GRTA	GGDPDYNYSN	AMKKAGLVWT	PQELRDFLSA	PKKKI PGNKM	ALAGI SKPBE	LDNLIAYL I F	SASSK
Rhodopila	GDPVEGKHLF	HT I CL I CHT -	DIKGRNKVGP	SLYGVVGRHS	G I EPGNYSE	ANI KSGI VWT	PDVLFKYI BH	PQK I VPGTKM	GYPG - QPDQK	RADI IAYLET	LK

Cytochrome C Phylogeny



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°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°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°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°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°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