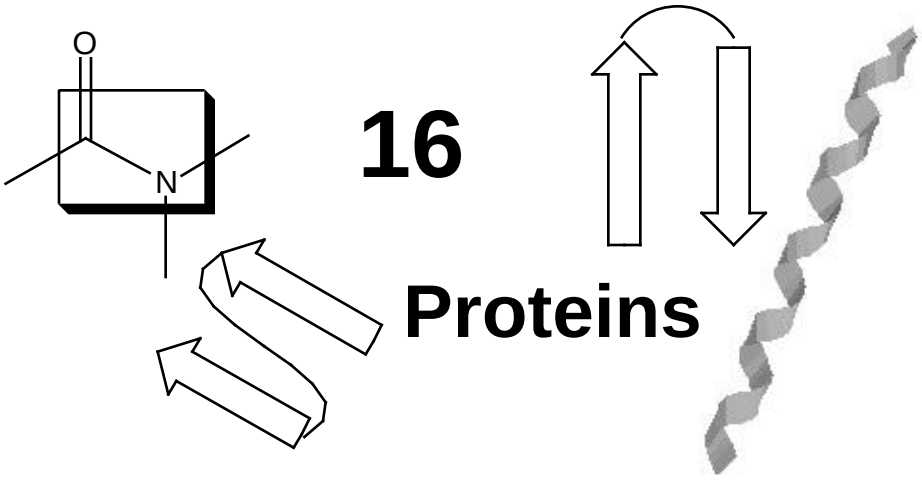


16 Proteins

CHAPTER SUMMARY

Proteins are polymers of amino acids. With **20 different fundamental amino acids** as building blocks, an extraordinarily large variety of proteins can be biosynthesized under the direction of the genetic code.

16.1 Structure of Amino Acids

A. Fundamental Structure - An Amine and An Acid

As the term amino acid describes, each monomer has an **amine group** and a **carboxylic acid group** attached to a prochiral carbon. In addition **side chains** can also be present. These range from a simple hydrogen to long carbon chains with functional groups.

B. Ionization of Amino Acids

The amine and carboxyl groups exhibit typical **acid-base behavior** which is pH-dependent. At low pH both groups are protonated: the amine group has a plus (+) charge and the carboxyl is neutral (0). As the pH rises the carboxyl loses its proton becoming negatively charged (-). At higher pH values the amine (+) deprotonates to produce a neutral amine (0). The result of this sequential deprotonation is a series of charged forms ranging from + to 0 to -. If the side chains are capable of acid-base reactions, **the number of possible charged forms depends upon the number and types of amino acids present, the pH, and the pK_a of each ionizable group**. This is true of proteins as well as amino acids. The pH at which the molecule has a net

charge of zero, the **zwitterion form**, is called the **pI** or isoelectric (isoelectronic) state. The pI can be calculated by taking the average of the two pK_a values on either side of the zwitterion form. At a pH lower than the pI the molecule will be in a net + charged form while at a pH greater than the pI it will be in a net - charged form. Charged forms can be separated in an electric field, a process known as **electrophoresis**.

C. The Common Amino Acids

There are 20 common amino acids which can be grouped by the nature of the R side chain. Our groups are acidic, basic, alkyl, polar, aromatic, sulfur-containing, and cyclic.

16.2 The Peptide Bond: Formation of Polypeptides and Proteins

Polypeptides and proteins are the products of amide, or **peptide, bond** formation between the amine group of one amino acid and the carboxyl of another.

16.3 The Hierarchy of Protein Structure

A. Primary Protein Structure - The Sequence of Amino Acids

The sequence of amino acids in the polymer, from the free amino- or N-terminus to the free carboxyl- or C-terminus, is called the **primary (1^0) structure** of a protein. This sequence is dictated by the genetic code.

B. Secondary Protein Structure - Helices and Pleated Sheets

A peptide bond has partial double bond character that makes it planar; the geometry is usually trans. As the polypeptide chain grows, the peptide bond can participate in hydrogen bonding - amide hydrogen to carbonyl oxygen. Because of the geometry of the peptide bond, this hydrogen bonding goes on between amino acids which are distant from each other. Organized, folded **secondary (2^0) structures** are formed. The **alpha helix** and **beta pleated sheet** are the two most common secondary structures. In the alpha helix hydrogen bonding usually occurs between the peptide bonds of four amino acids distant from each other. Beta structure involves the polypeptide chain in its fully extended form coming back on itself to hydrogen bond side-to-side. The two polypeptide strands in beta structures may be **parallel** or **antiparallel** to each other.

Secondary structures are, in turn, organized into **domains**, or supersecondary structures.

Collagen, which is the most abundant protein of the body, has unique primary and secondary structures. A high glycine and proline content leads to fairly rigid, kinked strands which can intertwine in a triple helical structure held together by hydrogen bonding between strands. The collagen helices aggregate to form skin, bone and connective tissue.

C. Protein Tertiary Structure

Side chains of the amino acids participate in **tertiary (3^0) structure**, that is, they stabilize the overall conformation of the protein molecule. The forces which hold tertiary structure together include covalent (**disulfide bridges**) and noncovalent (**hydrogen bonding, salt bridge, hydrophobic) interactions**. Shapes of tertiary structure subunits can be **globular** or **fibrous**.

D. Quaternary Protein Structure - Association of Subunits

Many proteins have more than one folded subunit, linked by the same types of noncovalent forces which hold 3^0 structure together. All of the subunits are needed for the protein to function properly. This is known as **quaternary (4^0) structure**.

E. Complex Proteins - Proteins Plus

All of the interactions mentioned above are integral parts of the **simple** structure of a protein. In addition proteins may have cofactors such as metal ions, carbohydrates or lipids, and/or organic molecules associated with them. This makes the proteins **complex**. **Myoglobin** and **hemoglobin** are examples of related complex proteins. Myoglobin has a single globular protein subunit complexed with an organic heterocyclic system known as **heme**. The heme in turn holds an iron (II) ion which can bind molecular oxygen, O_2. All of these components contribute to the function of myoglobin: the storage of oxygen in muscle tissue. Hemoglobin is related to myoglobin both structurally and functionally. It contains four myoglobin-type subunits each of which has an iron(II)-heme complex that can bind O_2. However, the four subunits interact cooperatively in order to transport oxygen in the blood from the lungs to the cells.

CONNECTIONS 16.1 Sickle Cell Anemia - A Biochemical Disease

F. Denaturation

The forces which hold a protein molecule together can be disrupted by changes in temperature and pH as well as by organic solvents and mechanical manipulation. This is known as **denaturation**.

CONNECTIONS 16.2 Mad Cow Disease

16.4 Functions of Proteins

With the great structural versatility available, proteins exhibit a phenomenal breadth of function. Catalysis, protection and regulation were but three discussed in this chapter.

A. Enzymes - Biological Catalysts

Enzymes are proteins which act as **catalysts** to the complex reactions that occur in the metabolism of living organisms. These reactions include oxidation-reduction, the formation and breaking of carbon-carbon, carbon-nitrogen, and other bonds, hydrolysis, synthesis, group transfer, and isomerization. An enzyme functions by presenting an interactive, three dimensional environment to the reactants (**substrates**). This allows the reaction to be **stereospecific, rapid, and selective**, that is, producing few, if any, spurious by-products. The **active site** of an enzyme has a **substrate binding subsite** and a group of amino acids which effect catalysis, the **catalytic site**. Nonprotein components are common partners in a cooperative catalytic process.

B. Enzyme Control

The actions of enzymes can be controlled and/or modified by species known as **inhibitors** or an enzyme may be activated/inactivated by **covalent modification**. Most enzymes have precursor forms which are inactive. These are known as **zymogens**.

C. Antibodies - Immune System Protection

The complex protective network of higher organisms is called the **immune system**. One part consists of **glycoproteins** (carbohydrate-protein) called **antibodies**. Antibodies bind to foreign substances, **antigens**, and help to mark and destroy the invader. This assault is a key component to the process of **immunization** in which the immune system is trained to respond aggressively to unwanted toxins, bacteria,

and viruses. The specificity of antibodies has proven invaluable in diagnostics and has high potential for targeted medications.

CONNECTIONS 16.3 Testing for Drugs, Pregnancy, and AIDS

D. Polypeptide and Protein Hormones - Metabolic Regulation

The regulation of metabolism is in part due to polypeptide and protein **hormones**, the products of the endocrine system. With the development of recombinant DNA techniques, specific protein hormones can now be made using bacteria and yeast. There has been ongoing discussion and controversy concerning the genetic manipulation of proteins for medical and commercial purposes.

CONNECTIONS 16.4 Growth Hormone

16.5 Determination of Protein Structure

There exists a general concensus that **the primary structure of a protein eventually determines its tertiary structure**. Therefore it is extremely important to be able to study a protein's primary structure.

A. Amino Acid Composition

Amino acid content is found by complete hydrolysis of the peptide bonds, separation of the constituent amino acids by column chromatography, and quantitation using reagents such as **ninhydrin** or **dansyl chloride**. However, this gives us no information about the N- to C-sequence.

B. Sequence of Amino Acids - Determination of Primary Structure

Sequential analysis can be accomplished by using the Edman technique. Treatment of an intact polypeptide with **phenylisothiocyanate** derivatizes the N- amino acid leaving the rest of the peptide intact for further **Edman degradation**. Large chains must be **fragmented** into shorter peptides, more easy to work with chemically. Cleavage of peptide bonds at specific amino acid residues is accomplished using enzymes such as **trypsin** (Lys, Arg), **chymotrypsin** (aromatics), and **carboxypeptidase** (C-terminus amino acids).

16.6 Organic Synthesis of Polypeptides

A. General Considerations

Polypeptides can be produced synthetically by reactions common to organic chemistry. Since both the amine and carboxyl groups are functionally active, a general procedure of functional group blocking, activation of other groups, and coupling of amino acids is carried out.

B. Solid-State Synthesis

An organized series of synthesis reactions can conveniently be carried out using the **solid state**, that is, columns to which the growing polypeptide chain is attached while various reagents are washed through.

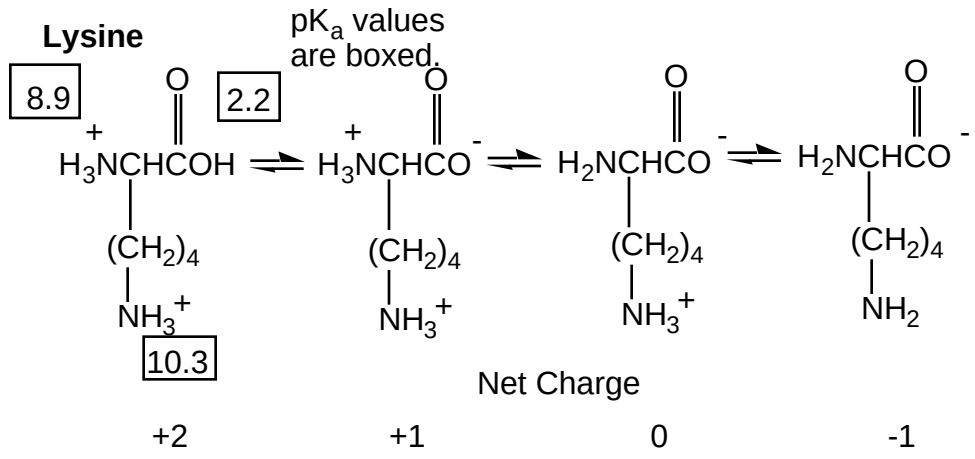
An understanding of proteins is essential for appreciating the link between organic chemistry and biochemistry.

SOLUTIONS TO PROBLEMS

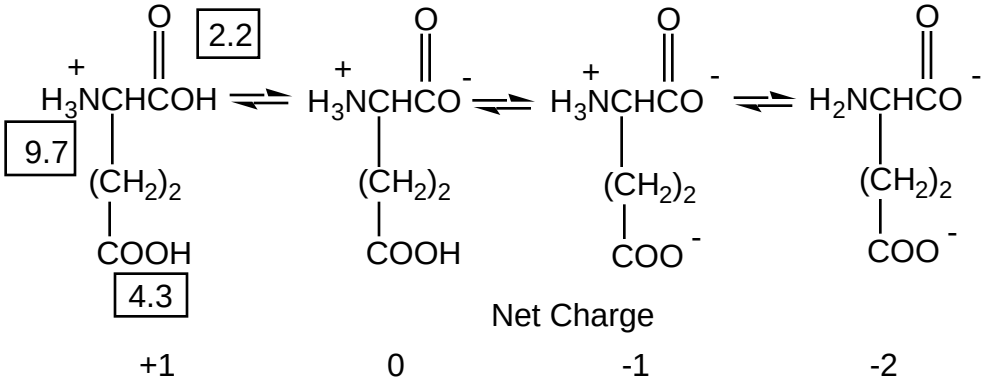
16.1 Amino Acid Structure: Ionization Section 16.1B

Arginine, lysine, and histidine have (+1) to (0) ionization transitions, while aspartic acid, glutamic acid, cysteine, and tyrosine have (0) to (-1) transitions.

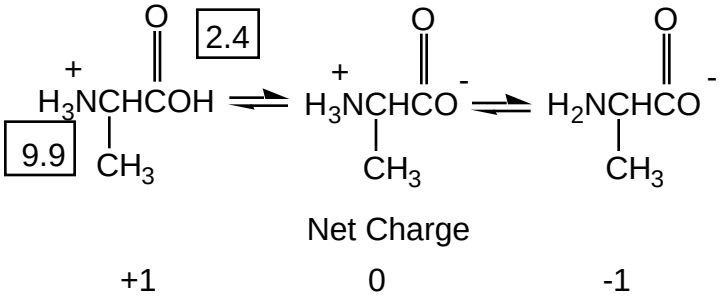
16.2 Ionized Forms of Amino Acids: Section 16.1



Glutamic Acid

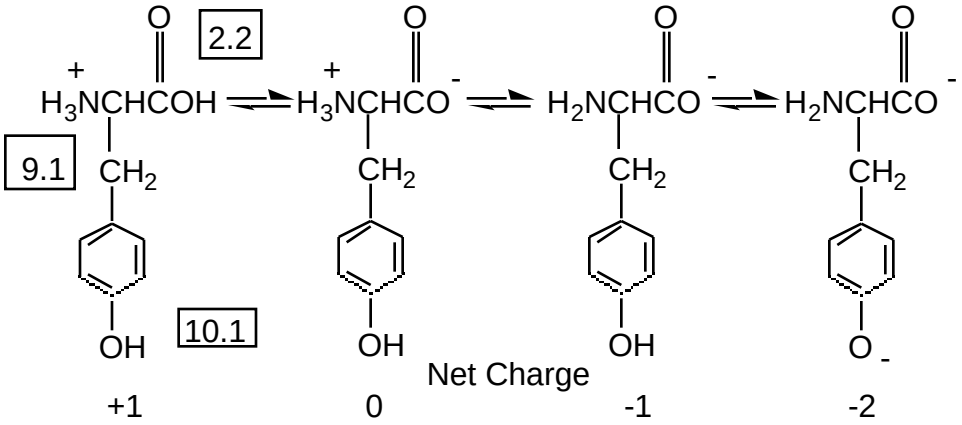


Alanine

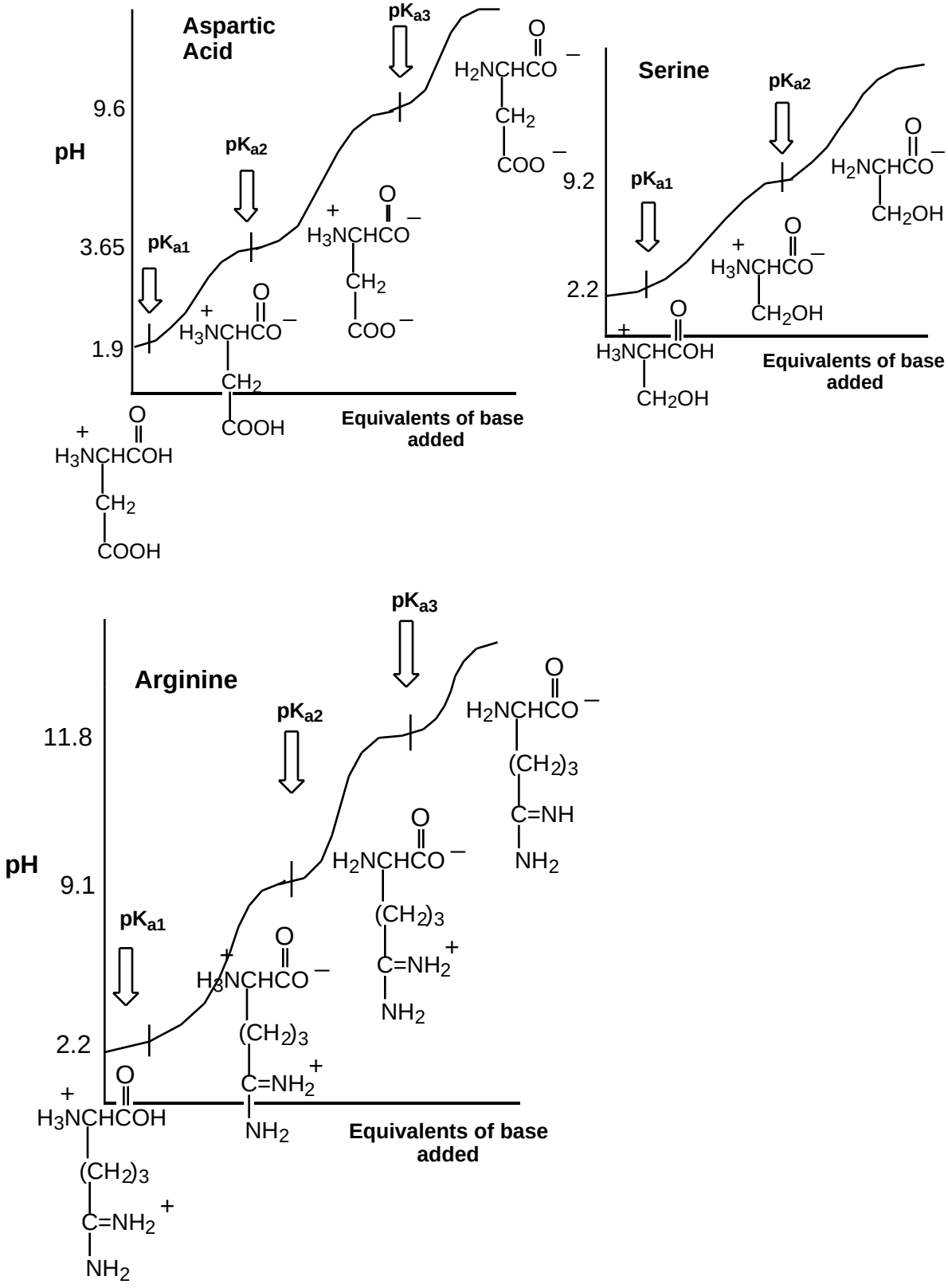


Tyrosine

pK_a values are boxed.



16.3 Acid-Base Behavior of Amino Acids: Section 16.1B

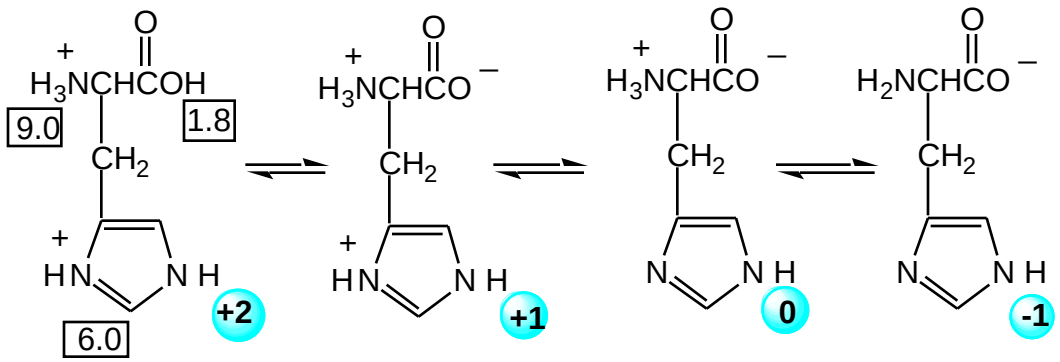


16.4 Ionization of Amino Acids: Section 16.1

group	pK_a	charge change	charge at pH 8.7	movement
glutamic acid			net -1	
α-COOH	2.2	0 → -1	-1	towards
α-NH_2	9.7	+1 → 0	+1	(+) pole
R	4.3	0 → -1	-1	
arginine			net +1	
α-COOH	2.2	0 → -1	-1	towards
α-NH_2	9.1	+1 → 0	+1	(-) pole
R	11.8	+1 → 0	+1	
threonine			net 0	
α-COOH	2.2	0 → -1	-1	no
α-NH_2	9.1	+1 → 0	+1	movement
R	-			
tyrosine			net 0	
α-COOH	2.2	0 → -1	-1	no
α-NH_2	9.1	+1 → 0	+1	movement
R	10.1	0 → -1	0	
histidine			net 0	
α-COOH	1.8	0 → -1	-1	no
α-NH_2	9.0	+1 → 0	+1	movement
R	6.0	+1 → 0	0	

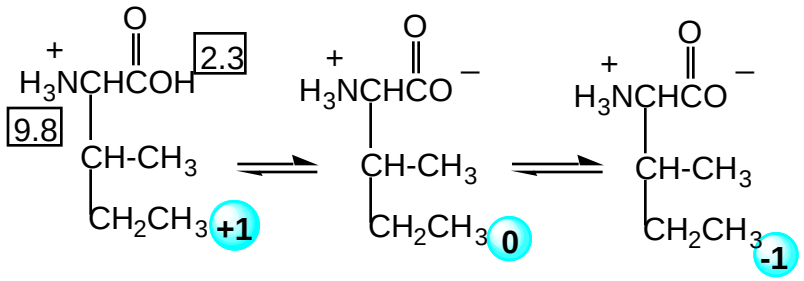
16.5 Ionization of Amino Acids: Section 16.1

histidine



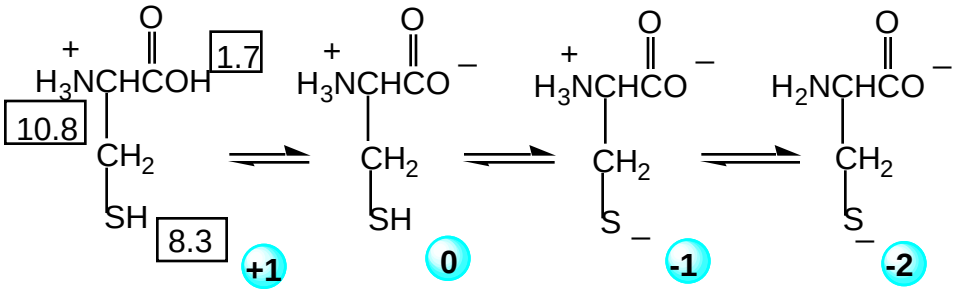
$$pI = \frac{6.0 + 9.0}{2} = 7.5$$

isoleucine



$$pI = \frac{2.3 + 9.8}{2} = 6.05$$

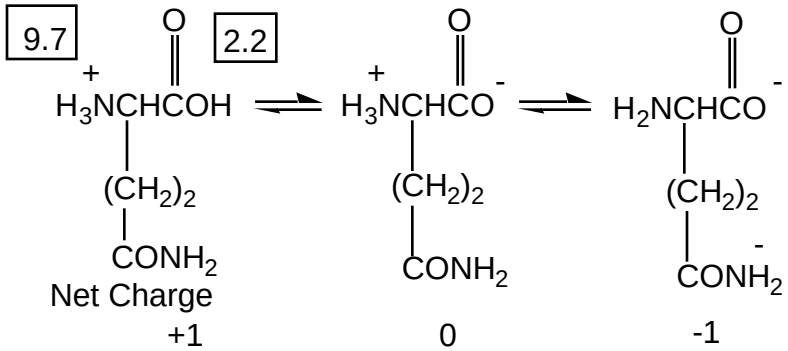
cysteine



$$pI = \frac{1.7 + 8.3}{2} = 5.0$$

16.6 Ionization ofAmino Acids: Section 16.1

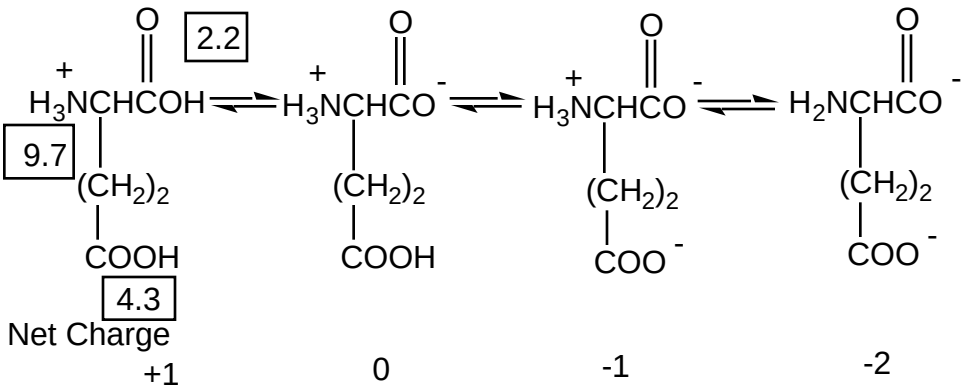
glutamine



$$pI = \frac{2.2 + 9.7}{2} = 5.95$$

The pI for Gln is higher than that for Glu due to the loss of ionizability of the side chain carboxyl group.

glutamic acid



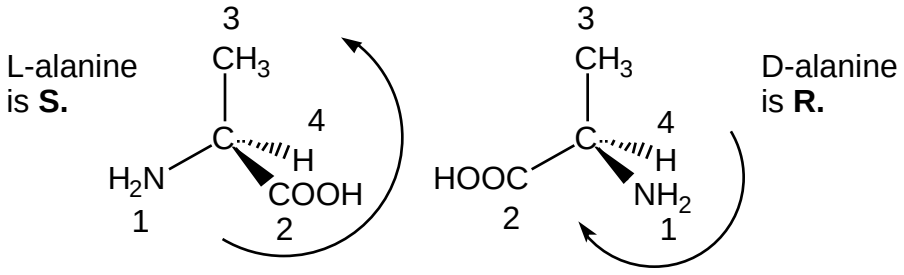
16.7 Ionization of Amino Acids: Section 16.1

See problems 16.4 and 16.5 for ionization information.

Histidine would most likely be in the 0 or zwitterion form at pH 6.8.

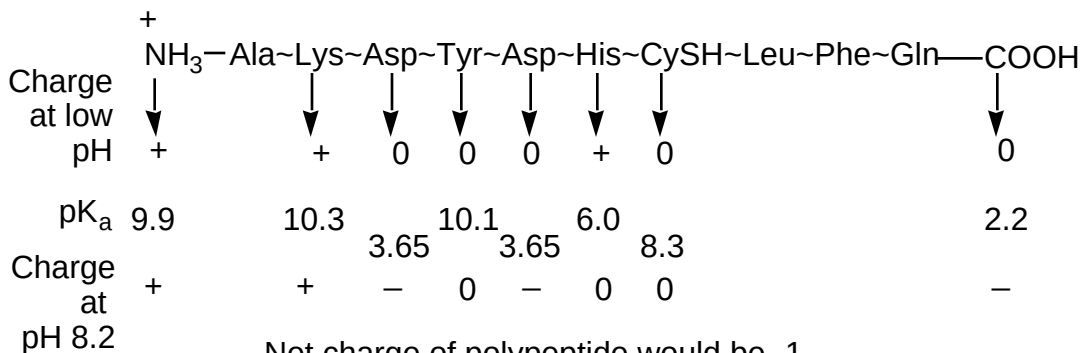
Tyrosine should be in its -2 form at pH 13.4.

16.8 Chirality of Amino Acids: Section 16.1



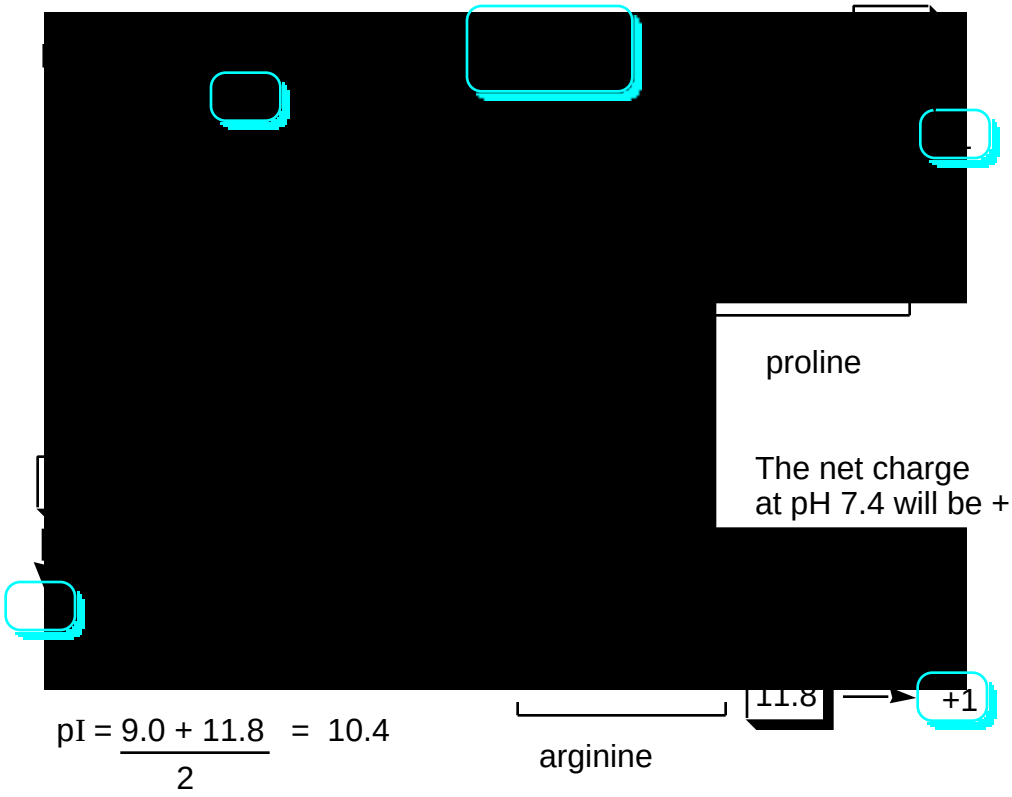
16.9 Chirality of Amino Acids: Section 16.1Glycine is optically inactive because it has two hydrogens on the alpha carbon (C-2). Four different groups are required for optical activity.

16.10 Polypeptides: Structure



Net charge of polypeptide would be -1.

16.11 Ionization of Polypeptides: Sections 16.1 and 16.2



$$pI = \frac{9.0 + 11.8}{2} = 10.4$$

16.12 Hierarchy of Protein Structure: Section 16.3B

At pH 7.4 polyaspartic acid would have a large net negative charge on its side chains while polylysine would have a large net positive charge. This would cause repulsion of the R groups and lead to helix destabilization.

16.13 Hierarchy of Protein Structure: Section 16.3B

Polythreonine has an alcohol group and a methyl group on the beta carbon. Polyisoleucine has a methyl and an ethyl group on this carbon. The presence of groups which can hydrogen bond or which introduce bulk close to the polypeptide backbone seem to be impediments to the formation of helical segments.

16.14 Hierarchy of Protein Structure: Section 16.3B

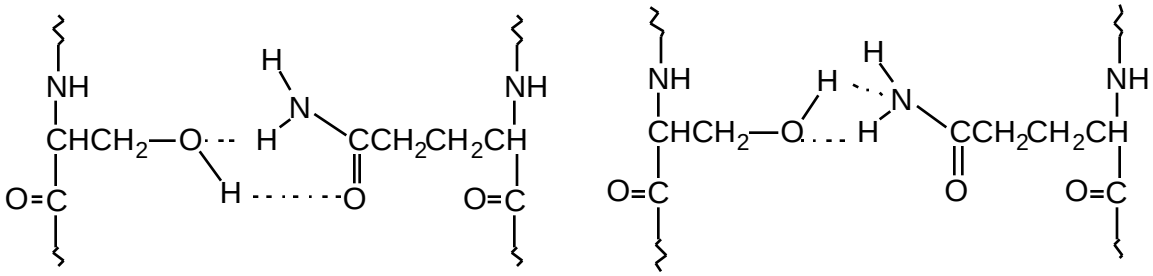
Leu, Ala, Ser, and Tyr would be "comfortable" in alpha helices because they have either small side chains (Ala and Ser) or extended alkyl groups (Leu) or a planar structure (Tyr).

Ala, Ser, and Gly could work in a beta sheet structure because of their small or nonexistent side chains which could allow the stacking of beta chains.

Pro with its ring structure would not fit into either of the conventional secondary structures but rather would be a place where one secondary structure could transition into another. Gly, with its ability for free rotation, could also be found at bends and breaks in regular secondary structure.

Lys has a charged, nitrogen-containing side chain under most pH conditions. It could exist in an alpha helix if there weren't any other positively charged groups in the area. Also at pHs above the pK_a of the R group, Lys would be "happy" in a helix.

16.15 Hierarchy of Protein Structure: Section 16.3C



16.16 Hierarchy of Protein Structure: Section 16.3C

a) Thr and H_2O - hydrogen bonding

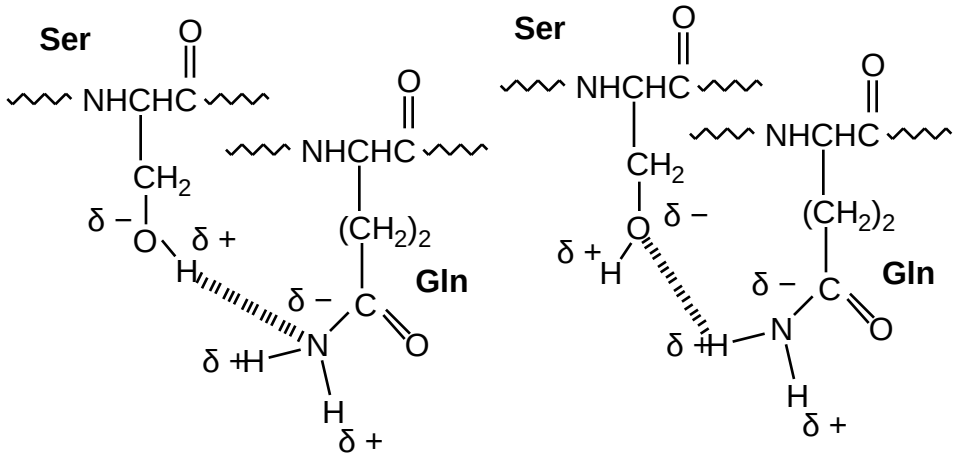
b) Asn and Trp - hydrogen bonding

c) Asp and Glu - repulsive forces

d) His and Val - hydrophobic interactions if above pH 6.0

16.17 Hierarchy of Protein Structure: Section 16.3F

Since the interior of a water soluble protein has a large degree of hydrophobicity or nonpolarity, nonpolar O_2 and N_2 could stabilize the denaturation of a protein by exposing the nonpolar interior to the air.

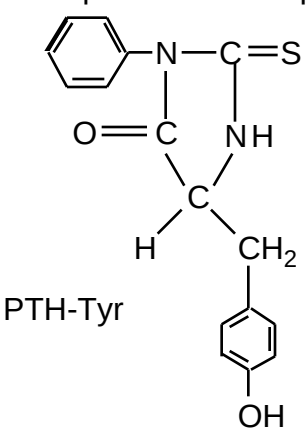


16.18 Hierarchy of Protein Structure: Section 16.3

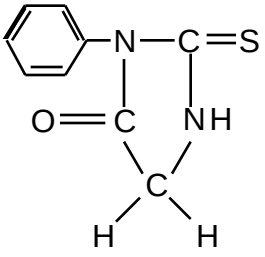
Salt bridges and ion-dipole interactions would be upset by lowering the pH of a protein solution.

16.19 Determination of Protein Structure: Section 16.5B

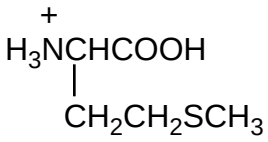
Two more cycles of degradation on the polypeptide remaining in Example 16.3 would produce PTH-Tyr, PTH-Gly and free Met.



PTH-Tyr



PTH-Gly



Met

16.20 Determination of Protein Structure: Section 16.5B

The theoretical yield for a five-step N-terminal sequential degradation would be

Step 1: 85%

Step 2: (0.85) * 85% = 72.25%

Step 3: (0.85) * 72.25% = 61.4%

Step 4: (0.85) * 61.4% = 52.2 %

Step 5: (0.85) * 52.2% = 44.4%

16.21 Determination of Protein Structure: Section 16.5B

Chymotrypsin digestion of the polypeptide in Example 16.4 would have produced the fragments: Gly ~ His ~ Lys ~ Gly ~ Phe and free Ile.

Trypsin digestion followed by chymotrypsin would produce the following three fragments: Gly ~ His ~ Lys, Gly ~ Phe and free Ile.

16.22 The Organic Synthesis of Polypeptides: Section 16.6

For the hypothetical amino acids - A, B, C, and D - 4! or 24 possible combinations exist.

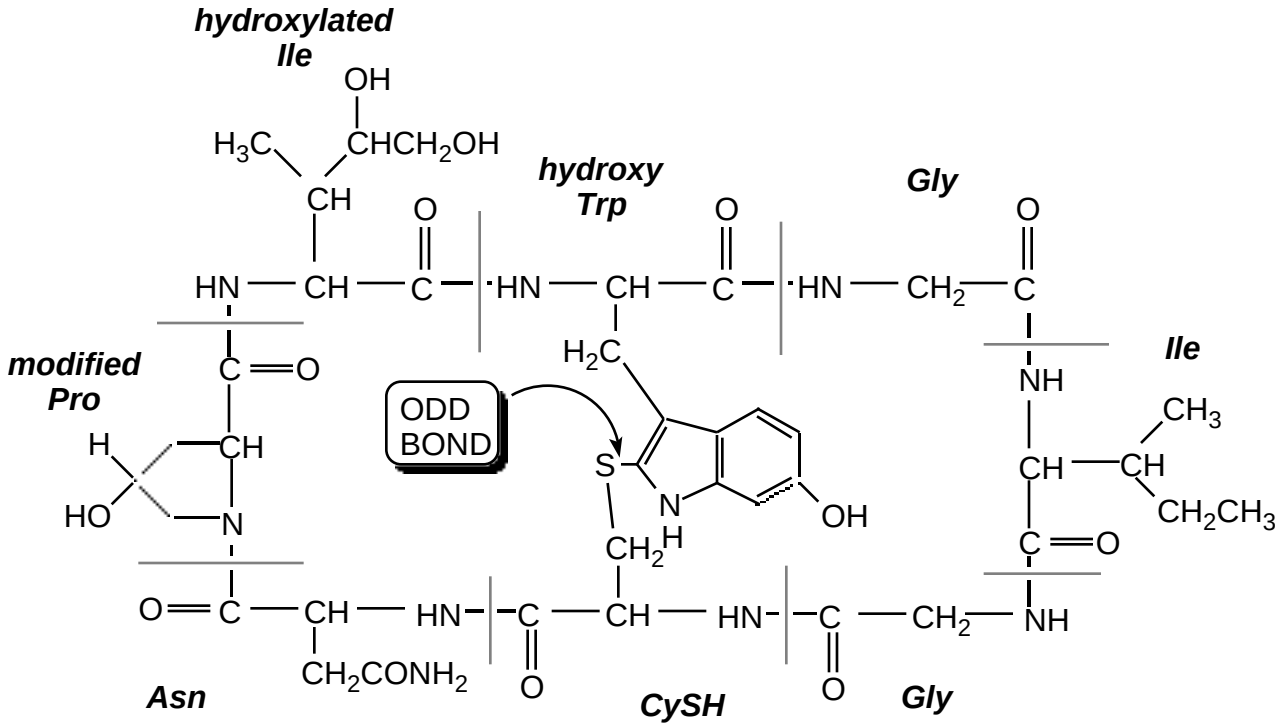
ABCD	ADBC	BCDA	BACD
ABDC	ADCB	BCAD	BADC
ACDB		BDAC	
ACBD		BDCA	

CDAB	DABC
CDBA	DACB
CABD	DBCA
CADB	DBAC
CBDA	DCAB
CBAD	DCBA

16. 23 Structure: Section 16.1

a) glycine
b) tyrosine
c) cysteine
d) all except Gly, Thr, Ile
e) proline
f) serine, threonine, asparagine, glutamine, histidine, tryptophan, tyrosine
g) threonine, isoleucine

16.24 Structure: Section 16.2



16.25 Structure: Sections 16.1 and 16.5

The amino acids, from N- to C-termini are: Glu, Ile, Thr, Lys.

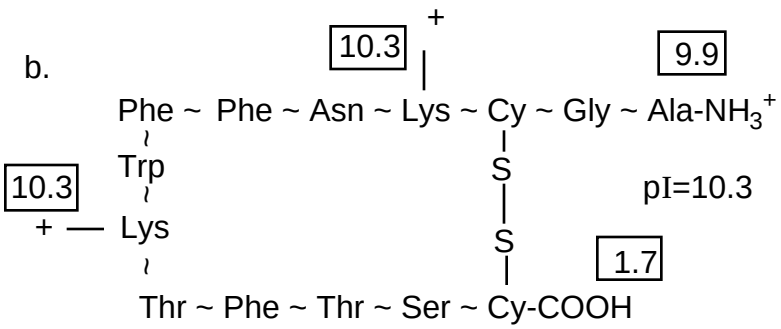
16.26 Structure: Section 16.1

a. H_3N^+—Tyr ~ Gly ~ Gly ~ Phe ~ Met—COOH

9.1 OH 10.1 2.3

$$pI = \frac{2.3 + 9.1}{2} = 5.7$$

b.



At low pH this polypeptide has a +3 charge.

At pH>1.7 it will be +2; at pH> 9.9 it will be +1.

The next two ionizable groups are both lysines. The average of their pK_as will be 10.3.

16.27 Structure: Sections 16.1 and 16.2

To associate with the negatively charged nucleic acids, histones would have a net positive charge, that is, they are basic. The basic amino acids are lysine and arginine with some contributions from histidine, depending upon the pH.

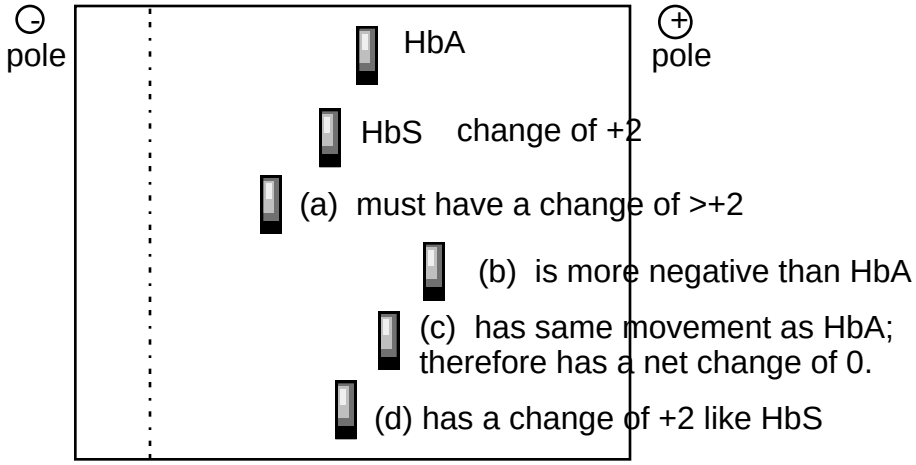
16.28 Structure: Sections 16.1, 16.2, and 16.4

Keep in mind that each hemoglobin molecule has two α and two β chains.

Using normal hemoglobin, HbA, as a starting point, find the change in charge which occurs with the change in amino acid.

Changes in Primary Sequence

Hb variant	chain	position from N-terminus	AA in HbA	AA in variant	Charge alteration
S	β	6	Glu	Val	change of +2
C	β	6	Glu	Lys	change of +4
Chesapeake	α	92	Arg	Leu	change of -2
Hasharon	α	47	Asp	His	change of +2
Koln	β	98	Val	Met	no change



a) is Hb C; b) is Hb $_{Chesapeake}$; c) is Hb $_{Koln}$; d) is Hb $_{Hasharon}$.

16.29 Hierarchy of Protein Structure: Section 16.3

a) 4^0 b) $3^0,4^0$ c) $2^0,3^0,4^0$ d) 1^0 e) $3^0,4^0$ f) 3^0

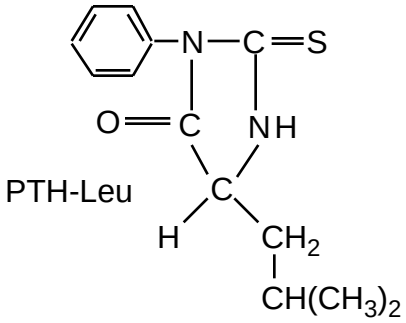
16.30 Hierarchy of Protein Structure: Section 16.3

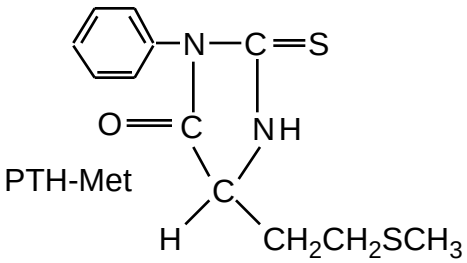
a) hydrogen bonding b) hydrophobic interactions

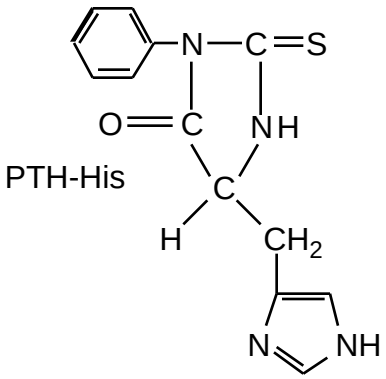
c) salt bridges d) none

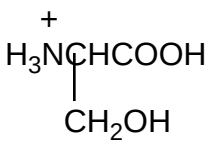
16.31 Determination of Protein Structure: Section 16.5

Three cycles of the Edman degradation would produce three PTH - amino acids and a free amino acid.

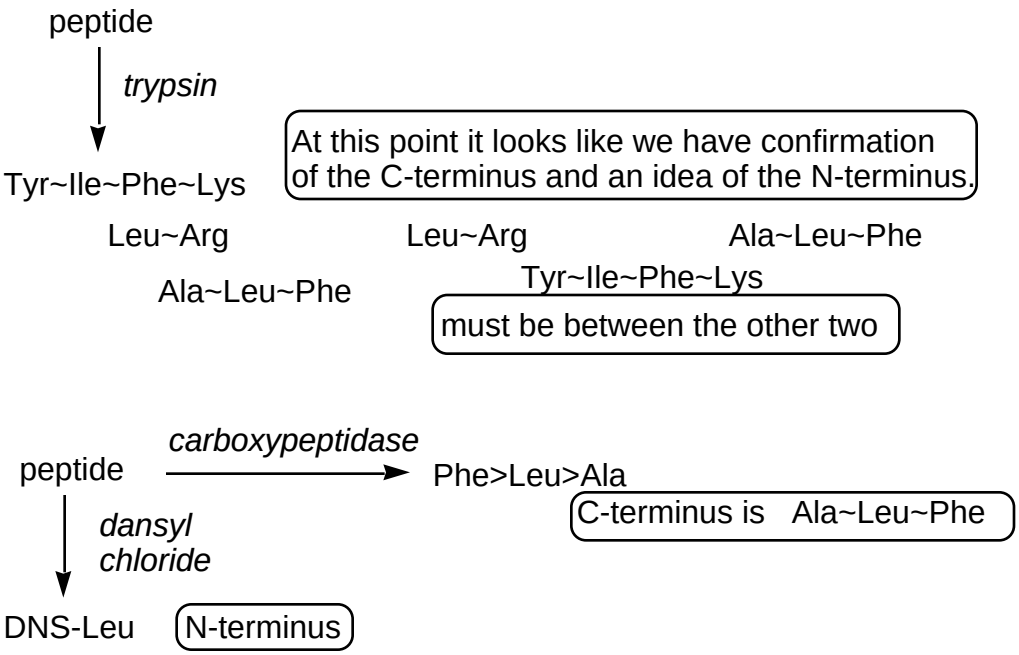




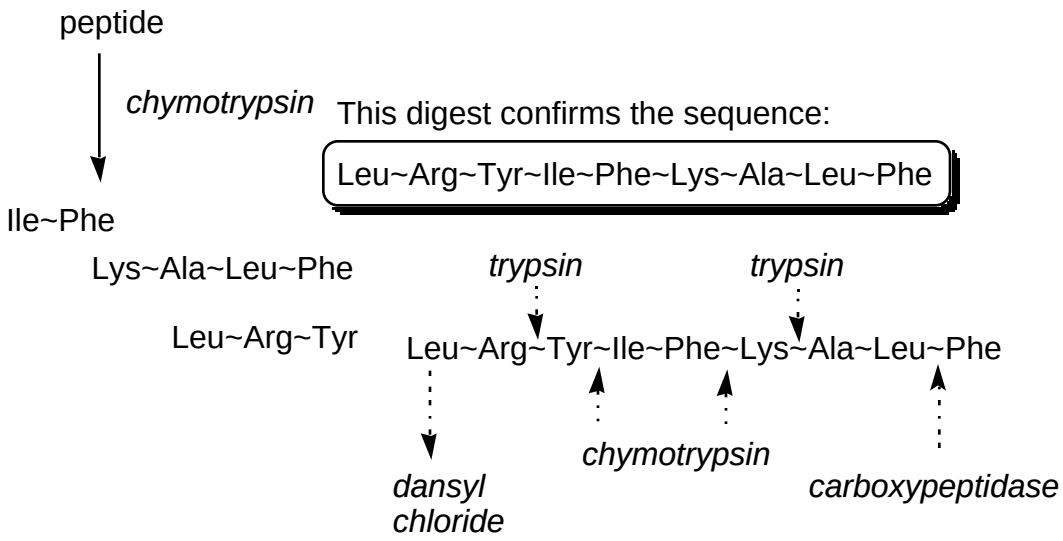


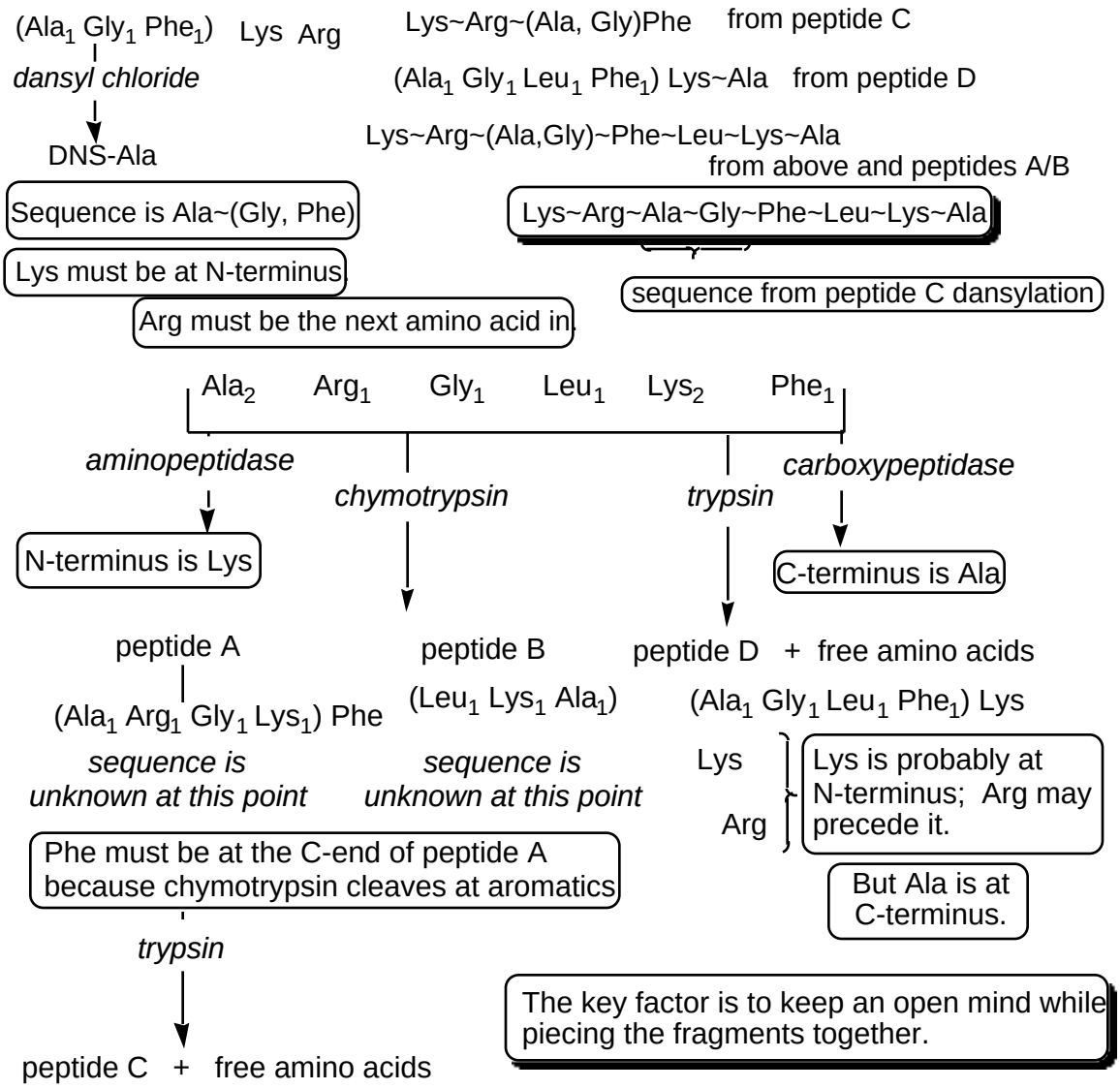


16.32 Determination of Protein Structure: Section 16.5



16.33 Determination of Protein Structure: Section 16.5




(Ala1 Gly1 Phe1) Lys Arg
dansyl chloride
DNS-Ala
Sequence is Ala~(Gly, Phe)
Lys must be at N-terminus
Arg must be the next amino acid in
Lys~Arg~(Ala, Gly)Phe from peptide C
(Ala1 Gly1 Leu1 Phe1) Lys~Ala from peptide D
Lys~Arg~(Ala,Gly)~Phe~Leu~Lys~Ala
from above and peptides A/B
Lys~Arg~Ala~Gly~Phe~Leu~Lys~Ala
sequence from peptide C dansylation
Ala2 Arg1 Gly1 Leu1 Lys2 Phe1
aminopeptidase
chymotrypsin
trypsin
carboxypeptidase
N-terminus is Lys
C-terminus is Ala
peptide A
(Ala1 Arg1 Gly1 Lys1) Phe
sequence is unknown at this point
peptide B
(Leu1 Lys1 Ala1)
sequence is unknown at this point
peptide D + free amino acids
(Ala1 Gly1 Leu1 Phe1) Lys
Lys
Arg
Lys is probably at N-terminus; Arg may precede it.
But Ala is at C-terminus.
Phe must be at the C-end of peptide A because chymotrypsin cleaves at aromatics
trypsin
peptide C + free amino acids
The key factor is to keep an open mind while piecing the fragments together.