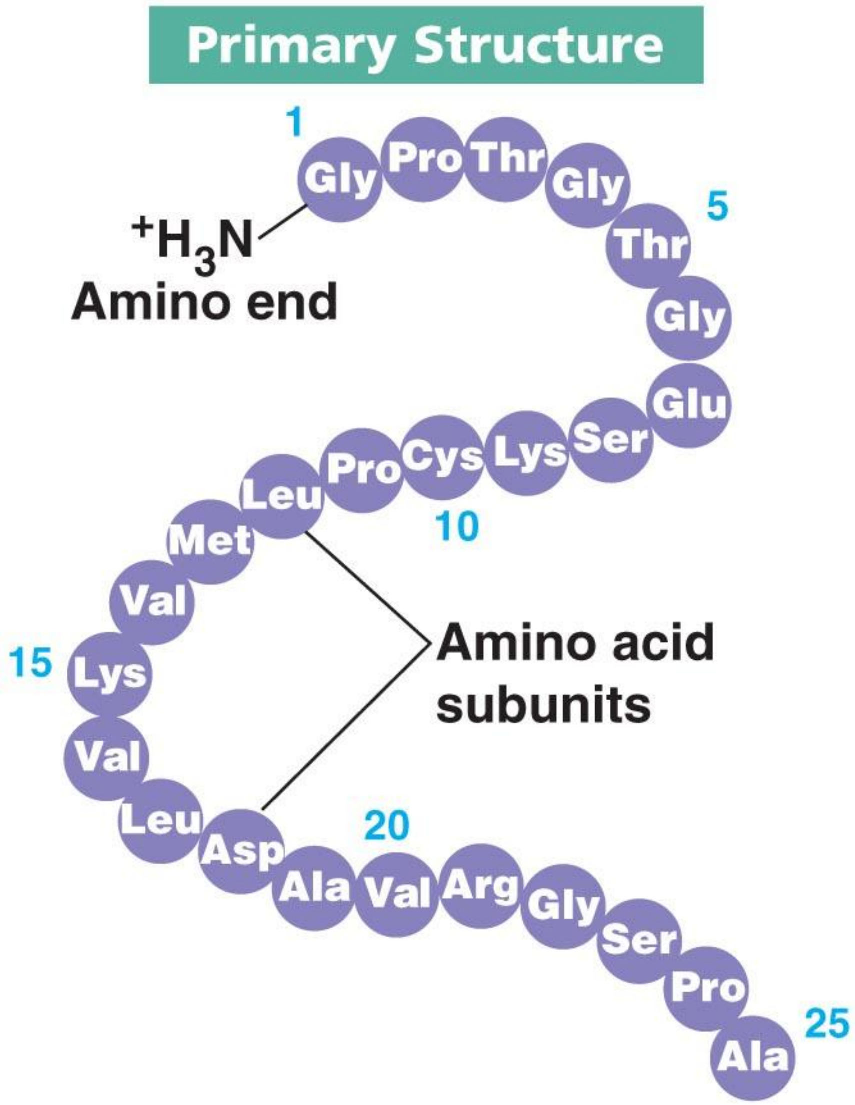
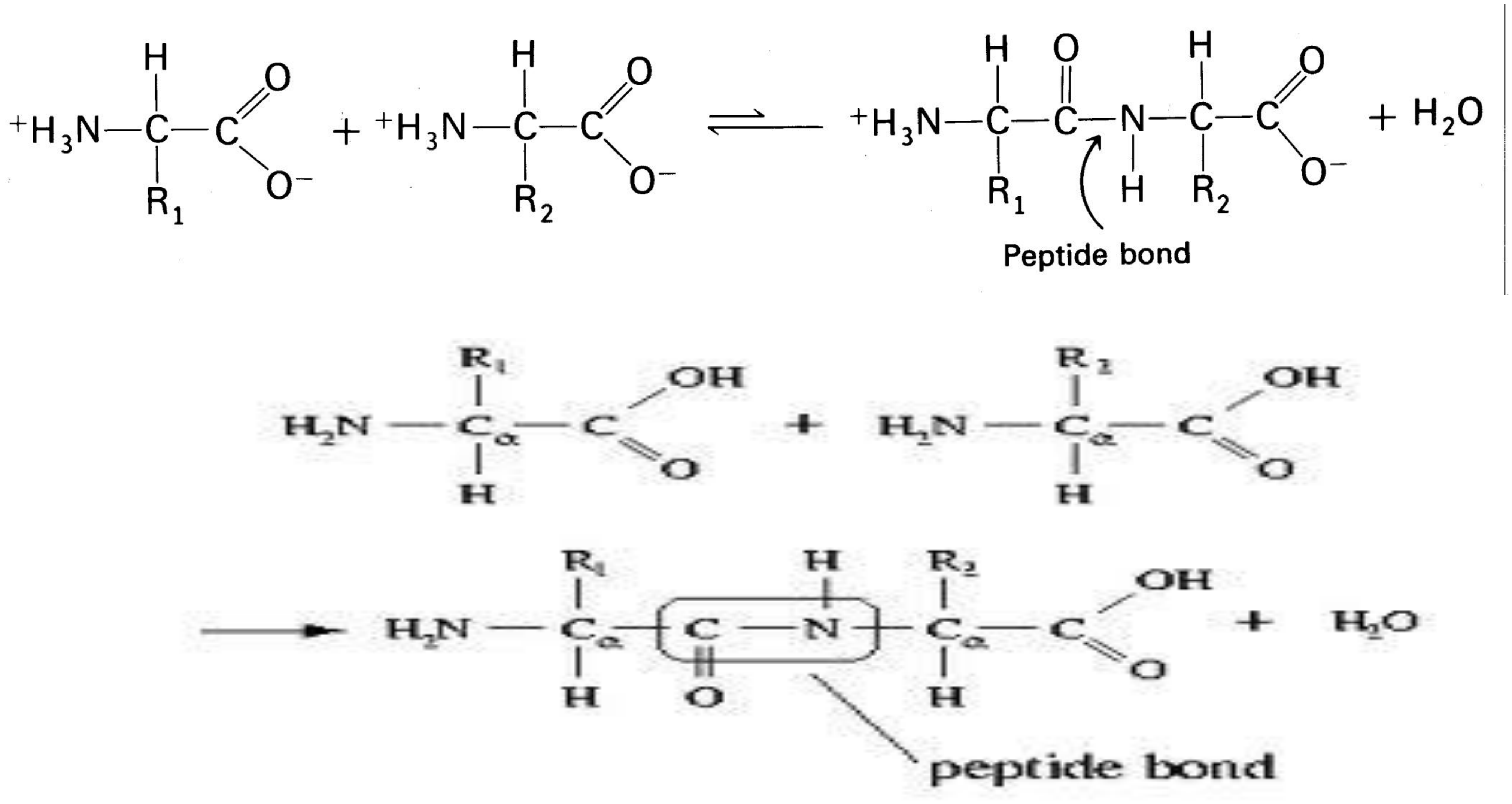
outline of protein principles

Asst. Prof. Dr. Ekhlass M. Taha

- Proteins are large organic compounds made of amino acids arranged in a linear chain and joined together by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues. Proteins are polyamides, and hydrolysis of a protein yields amino acids.


Primary Structure
^+H_3N
Amino end
1
Gly
Pro
Thr
Gly
5
Thr
Gly
Glu
Ser
Lys
Cys
Pro
10
Leu
Met
Val
15
Lys
Val
Leu
Asp
Ala
20
Val
Arg
Gly
Ser
Pro
25
Ala
Amino acid subunits

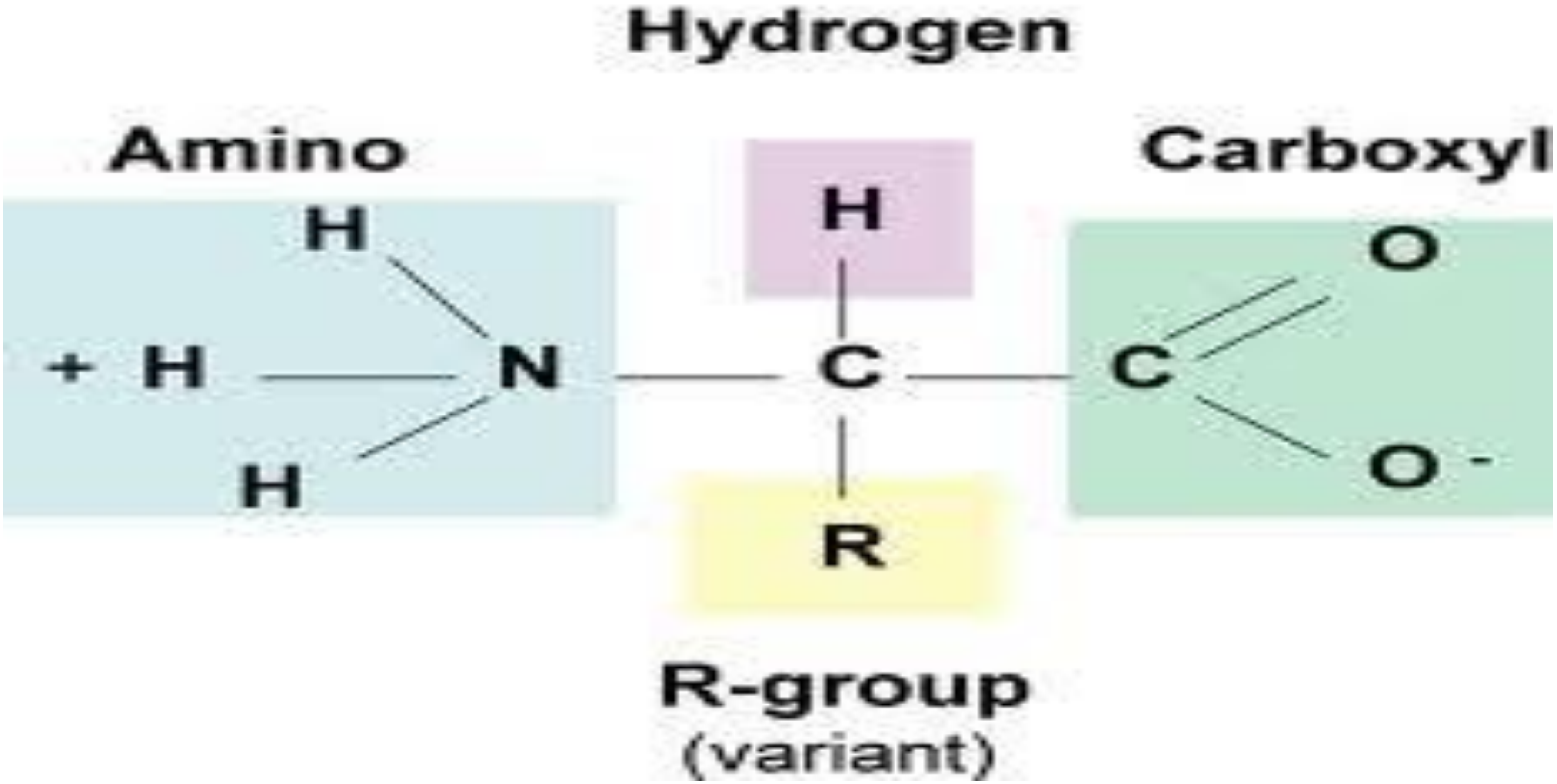

Peptide bond
peptide bond

- **Like other biological macromolecules such as polysaccharides and nucleic acids, proteins are essential parts of organisms and participate in every process within cells. Many proteins are enzymes that catalyze biochemical reactions and are vital to metabolism.**
-
- **Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are important in cell signaling, immune responses, cell adhesion, and the cell cycle.**
- **Proteins are also necessary in animals' diets, since animals cannot synthesize all the amino acids they need and must obtain essential amino acids from food. Through the process of digestion, animals break down ingested protein into free amino acids that are then used in metabolism.**
-
- **Proteins are linear polymers built from 20 different α-amino acids. These amino acids are commonly found in plant and animal proteins, yet these 20 amino acids can be combined in a variety of ways to form muscles, tendons, skin, fingernails, feathers, silk, hemoglobin, enzymes, antibodies, and many hormones.**

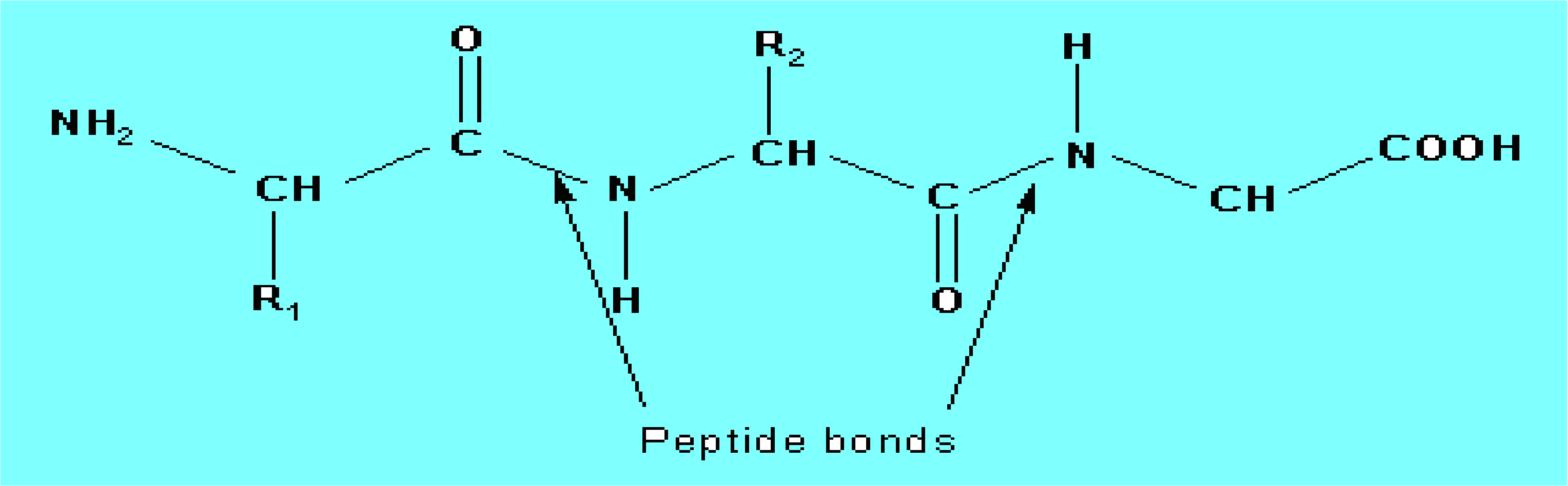
- **The structure of amino acids:**
- **The amino acids found in proteins are α-amino carboxylic acids. All amino acids possess common structural features, including an α-carbon to which an amino group, a carboxyl group, and a variable side chain are bonded. Only proline differs from this basic structure as it contains an unusual ring to the N-end amine group, which forces the CO–NH amide moiety into a fixed conformation.**
-
- **The side chains of the standard amino acids, detailed in the list of standard amino acids, have different chemical properties that produce three-dimensional protein structure and are therefore critical to protein function.**
-
- **The amino acids in a polypeptide chain are linked by peptide bonds formed in a dehydration reaction. Once linked in the protein chain, an individual amino acid is called a residue, and the linked series of carbon, nitrogen, and oxygen atoms are known as the main chain or protein backbone.**
-
- **Due to the chemical structure of the individual amino acids, the protein chain has directionality. The end of the protein with a free carboxyl group is known as the C-terminus or carboxy terminus, whereas the end with a free amino group is known as the N-terminus or amino terminus.**

- **The commonly occurring amino acids are of 20 different kinds which contain the same dipolar ion group $H_3N^+.CH.COO^-$. They all have in common a central carbon atom to which are attached a hydrogen atom, an amino group (NH_2) and a carboxyl group (COOH). The central carbon atom is called the C_{alpha}-atom and is a chiral centre. All amino acids found in proteins encoded by the genome have the L-configuration at this chiral centre.**

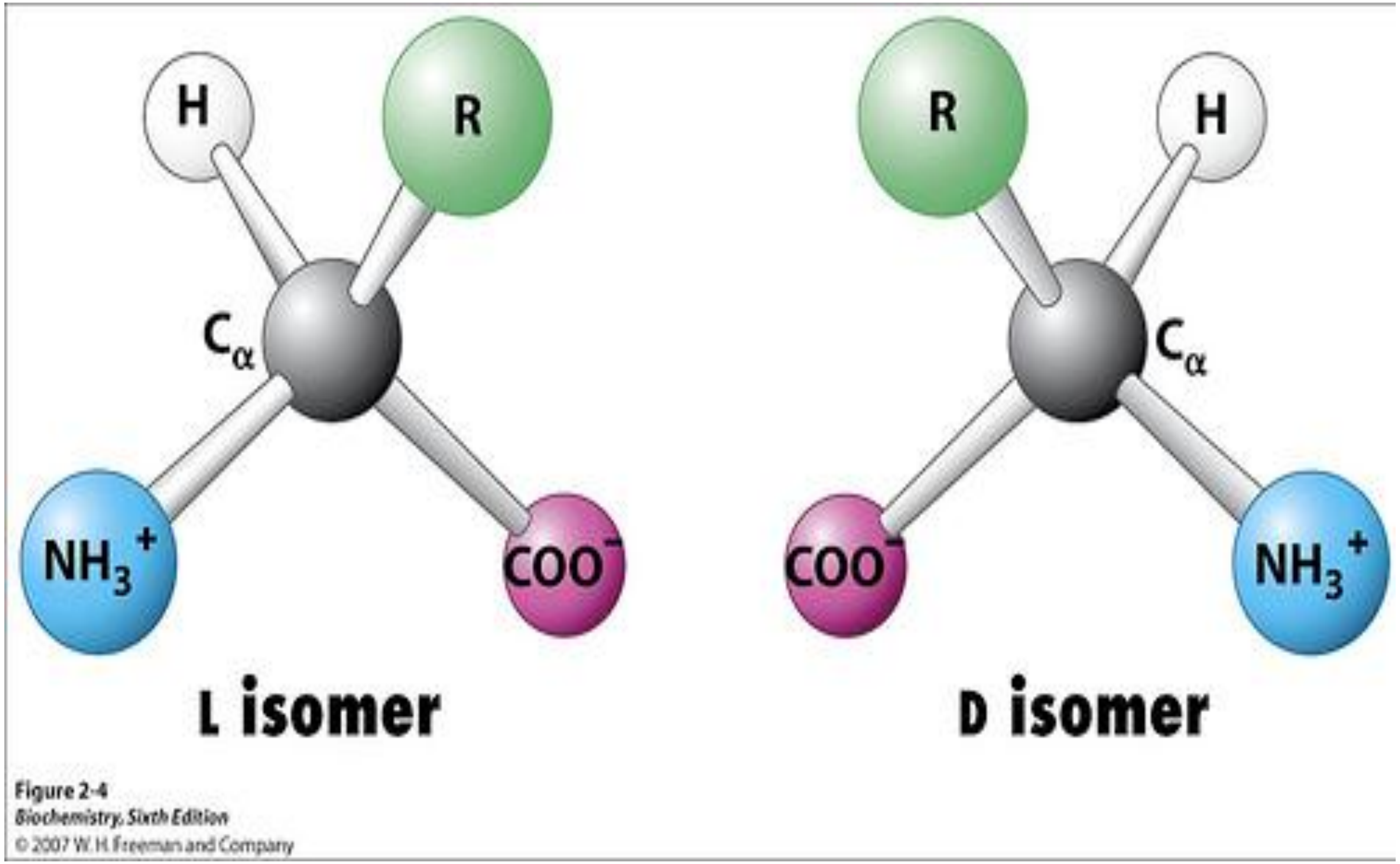
Amino Acid Structure



- **Amino acids in proteins (or polypeptides) are joined together by peptide bonds. The sequence of R-groups along the chain is called the primary structure.**



- The simplest amino acid is amino acetic acid ($H_2NCH_2CO_2H$), called **glycine**, which has no side chain and consequently does not contain a chiral carbon. All other amino acids have side chains, and therefore their α-carbons are chiral.
-
- Isomerism
- Of the standard α-amino acids, all but glycine can exist in either of two optical isomers, called L- or D-amino acids, which are mirror images of each other. While L-amino acids represent all of the amino acids found in proteins during translation in the ribosome, D-amino acids are found in some proteins produced by enzyme post-translational modifications after translation and translocation to the endoplasmic reticulum, as in exotic sea-dwelling organisms such as cone snails. They are also abundant components of the peptidoglycan cell walls of bacteria, and D-serine may act as a neurotransmitter in the brain.
-
- The L and D convention for amino acid configuration refers not to the optical activity of the amino acid itself, but rather to the optical activity of the isomer of glyceraldehyde from which that amino acid can theoretically be synthesized (D-glyceraldehyde is dextrorotary; L-glyceraldehyde is levorotary).



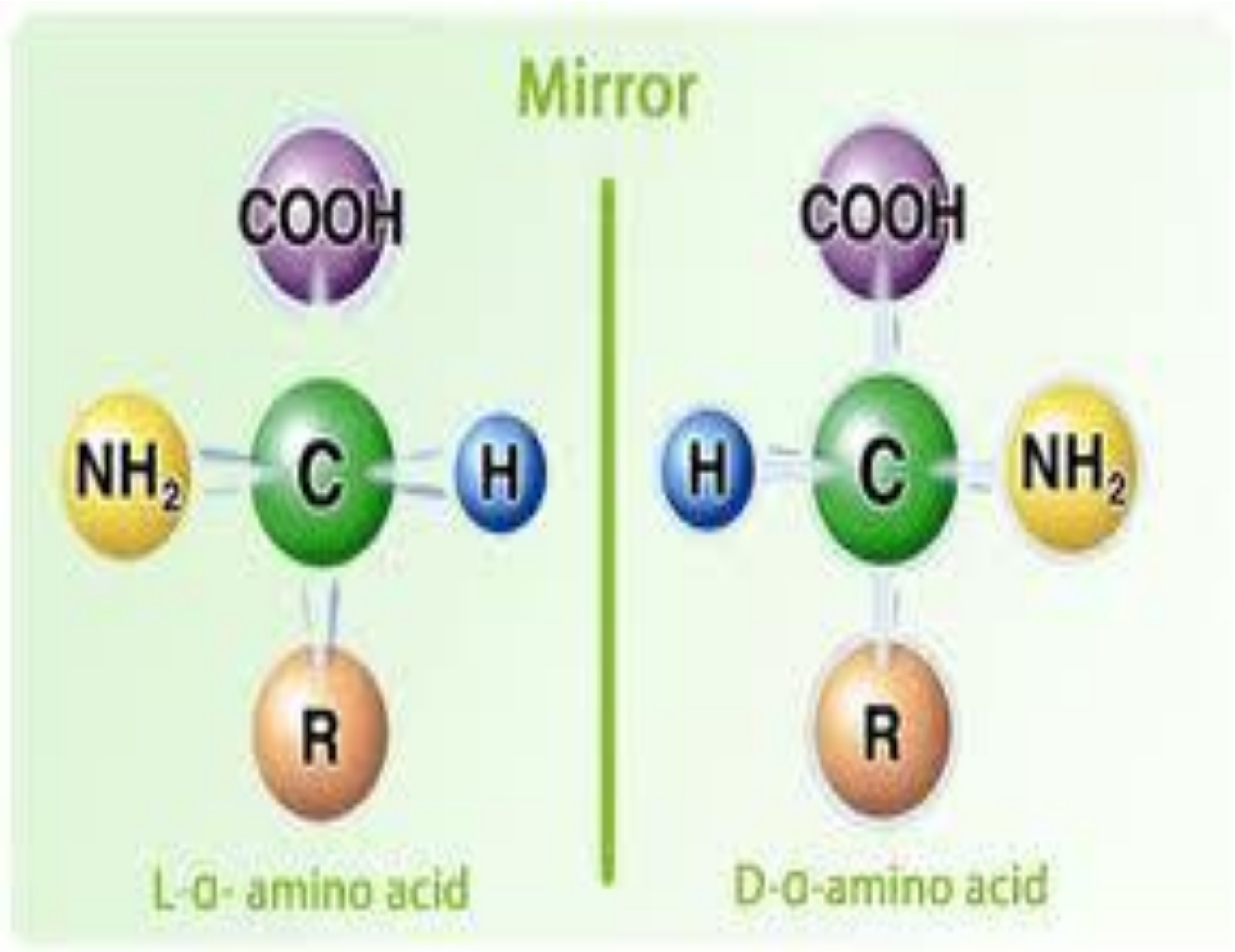
L isomer

D isomer

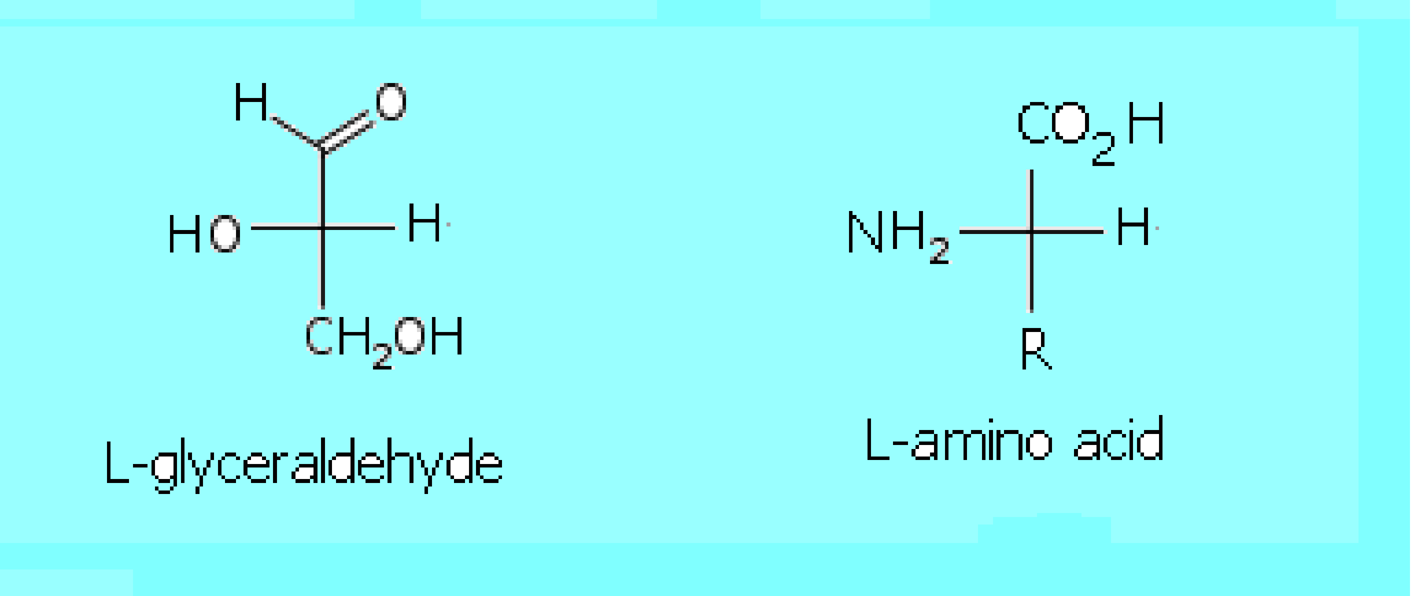
Figure 2-4

Biochemistry, Sixth Edition

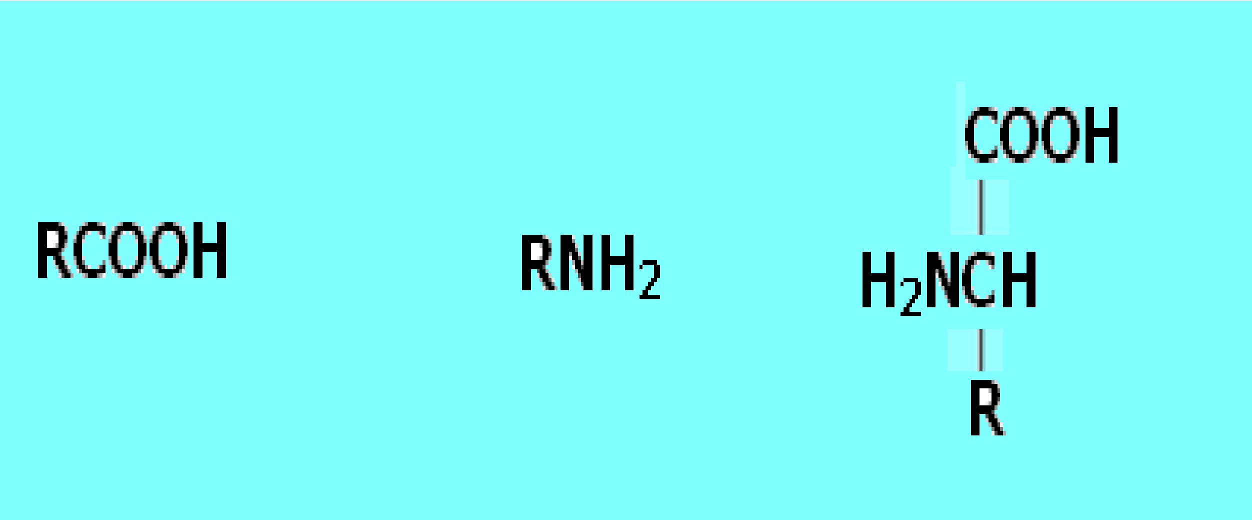
The two optical isomers


Mirror
COOH
NH2
C
H
R
L-α- amino acid
COOH
H
C
NH2
R
D-α-amino acid

- **Amino acids from proteins belong the L-series- that is, the groups around the α-carbon have the same configuration as L-glyceraldehyde.**



- **Amino acids do not always behave like organic compounds** **because:**



$RCOOH$

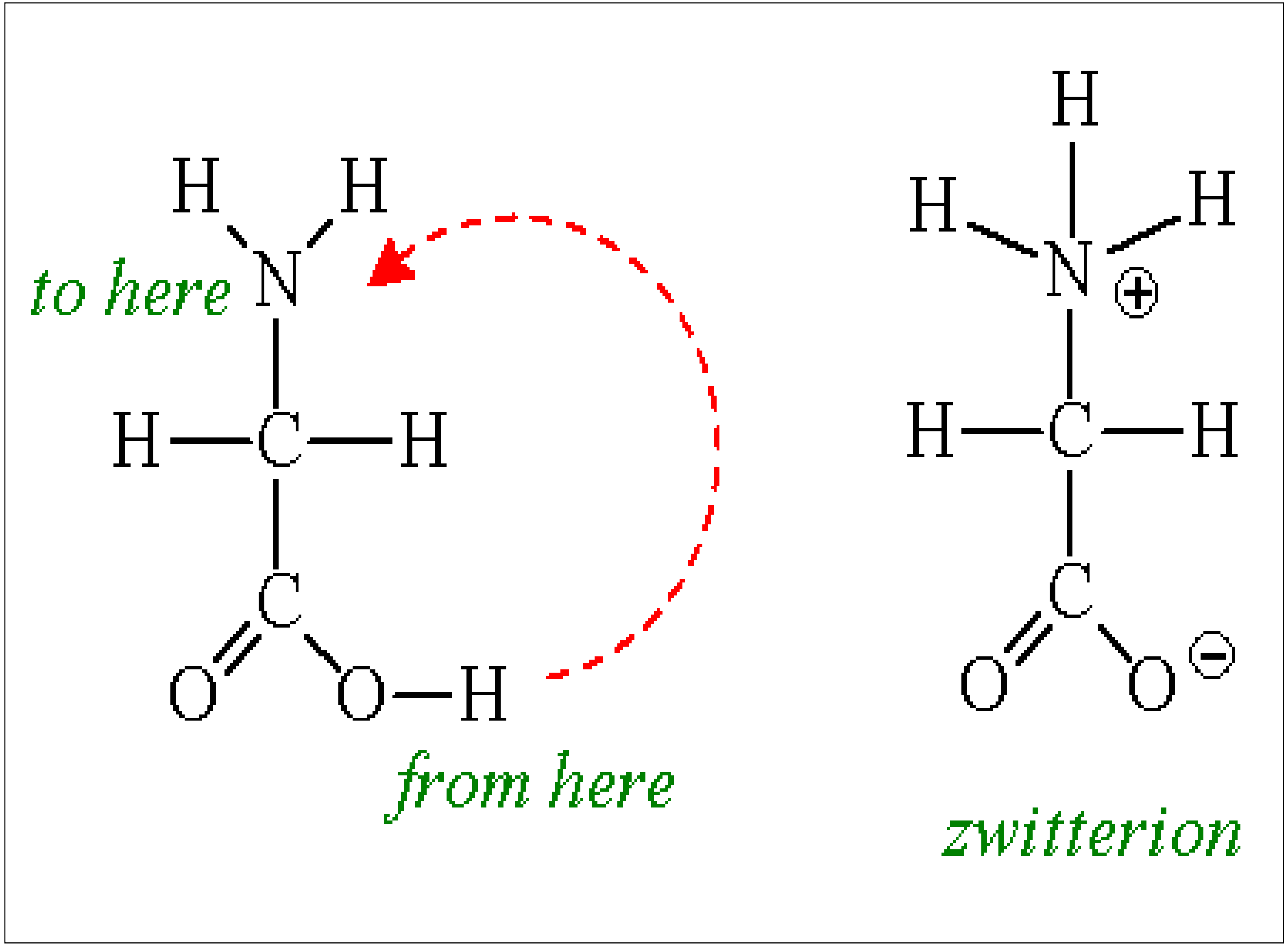
RNH_2

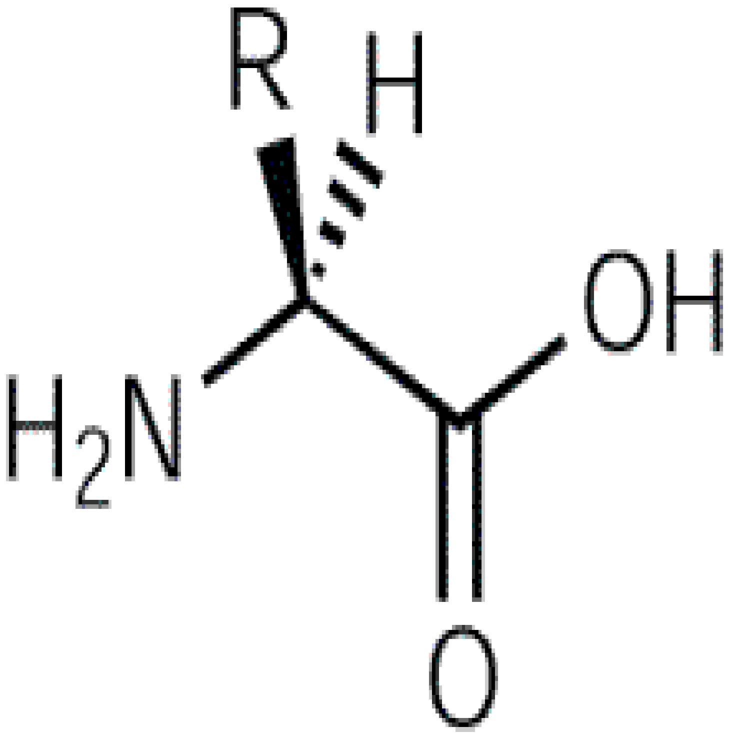
$$\begin{array}{c} \quad COOH \\ | \\ H_2NCH \\ | \\ R \end{array}$$

1. They have melting points of over 200° , whereas most organic compounds of similar molecular weight are liquids at room temperature.
2. Amino acids are soluble in water and other polar solvents, but insoluble in non-polar solvents such as diethyl ether or benzene.
3. Amino acids have large dipole moments.
4. They are less acidic than most carboxylic acids and less basic than most amines.

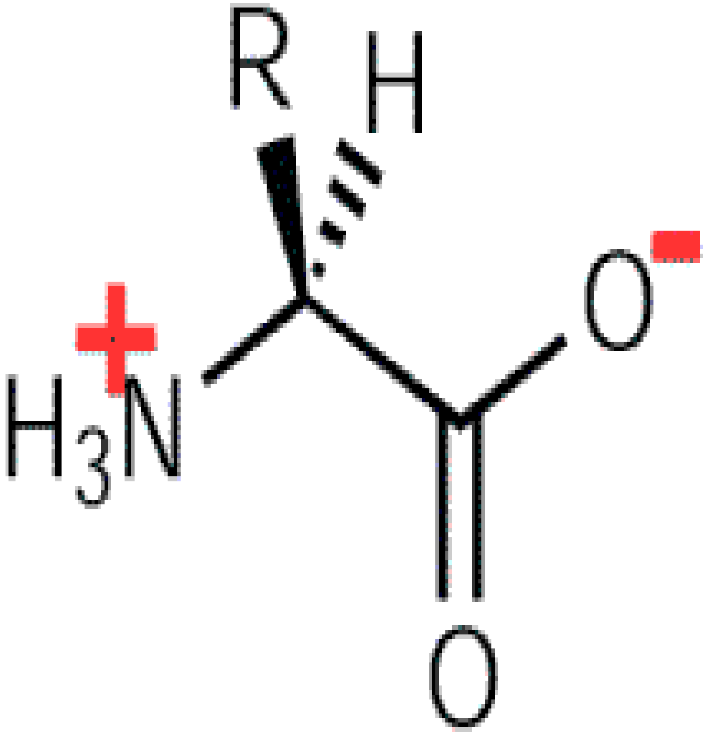
-

- Why do amino acids exhibit such unusual properties? The reason is that an amino acid contains a basic amino group and an acidic carboxyl group in the same molecule. An amino acid undergoes an internal acid-base reaction to yield a dipolar ion, also called a zwitter ion.
- Because of the resultant ionic charges, an amino acid has many properties of a salt. Furthermore, the pKa of an amino acid is not the pKa of a –COOH group, but that of an $-NH_3^+$ group. The pKb is not that of a basic amino group, but that of the very weakly basic $–COO^-$ group.


to here
from here
zwitterion

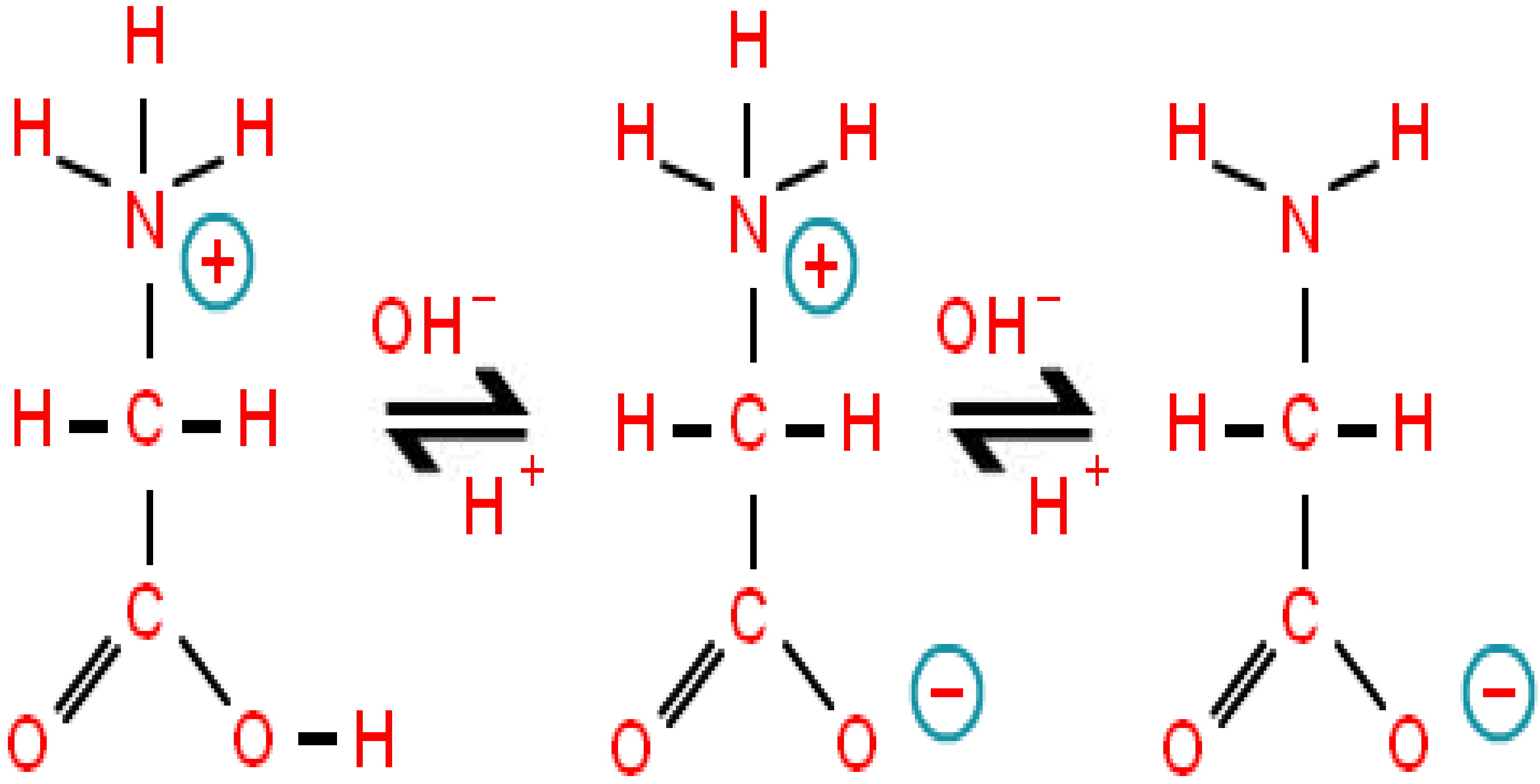


Nonzwitterionic



Zwitterionic

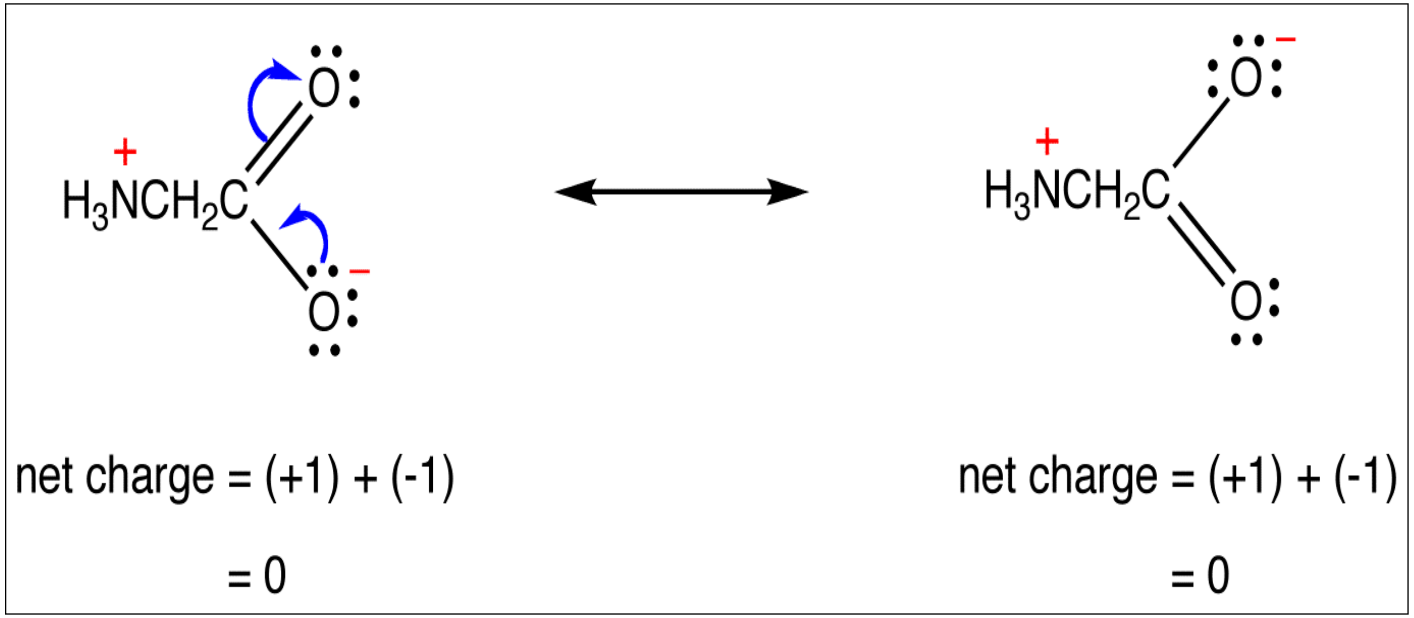
- **ISOELECTRIC POINT**
- **Amino acids have both amine and carboxylic acid functional groups and are therefore both an acid and a base at the same time. At a certain pH known as the isoelectric point an amino acid has no overall charge, since the number of protonated ammonia groups (positive charges) and deprotonated carboxylate groups (negative charges) are equal.**



ion at low pH

zwitterion
neutral pH

ion at high pH



net charge = (+1) + (-1)
= 0

net charge = (+1) + (-1)
= 0

The amino acids all have different isoelectric points. The ions produced at the isoelectric point have both positive and negative charges and are known as a *zwitter ion*, which comes from the German word *Zwitter* meaning "hermaphrodite" or "hybrid". Amino acids can exist as zwitter ions in solids and in polar solutions such as water, but not in the gas phase. Zwitter ions have minimal solubility at their isolectric point and an amino acid can be isolated by precipitating it from water by adjusting the pH.

Calculating the net charge on a polypeptide: example from lecture

Glu-Gly-Ala-Lys

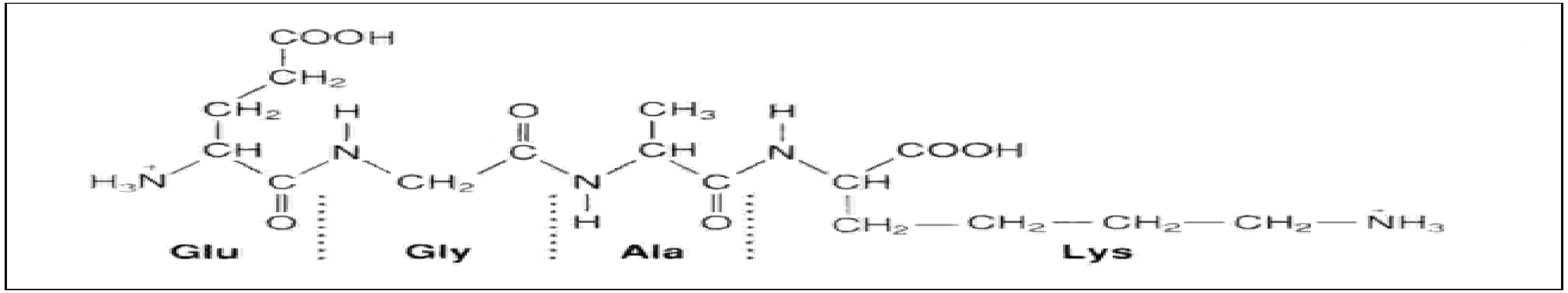
Step 1: Write out the structure with the charges that you would expect at pH 0

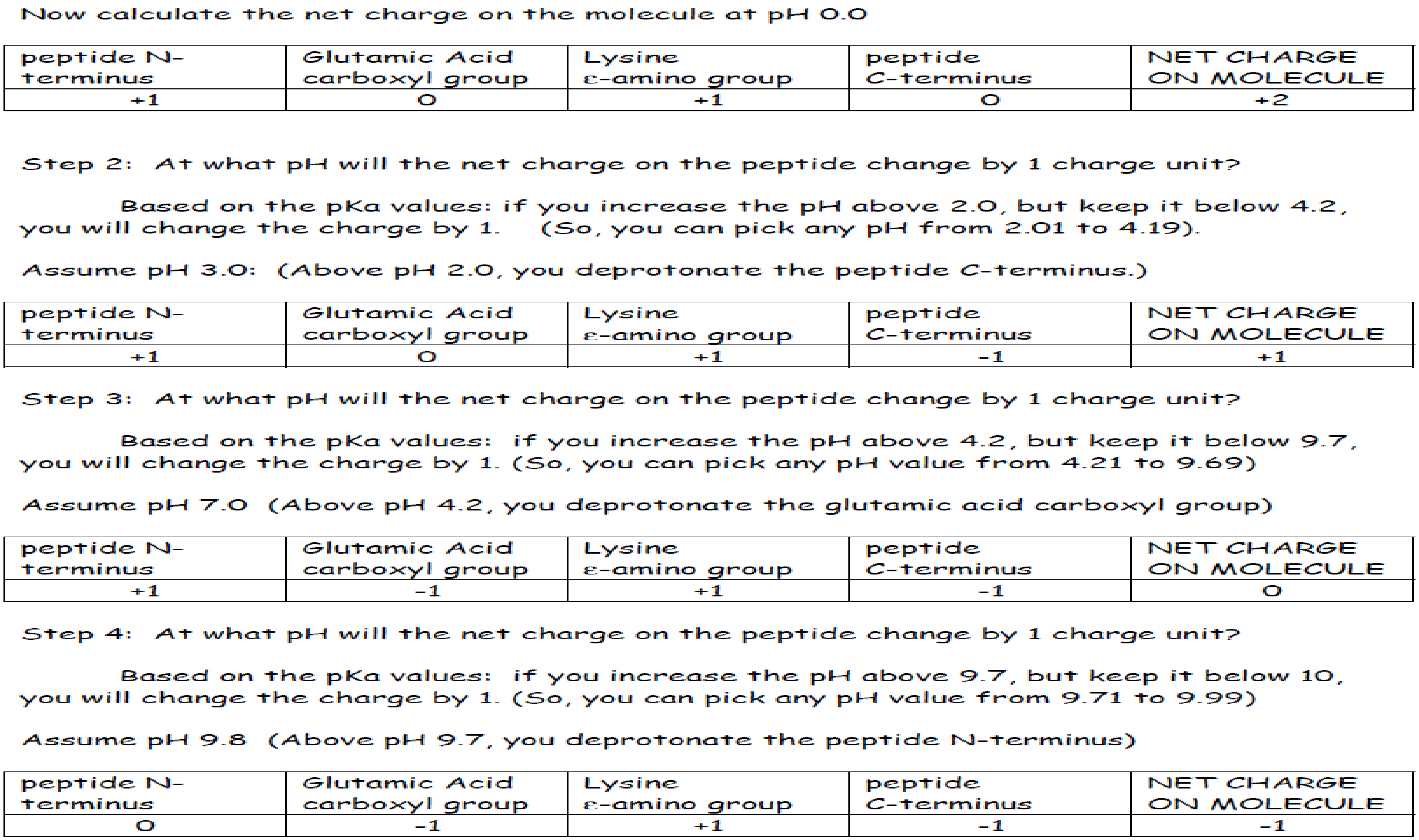
How do you know what charges would be expected at pH 0?
Based on the pKa values for all the ionizable groups.
What are ionizable - should be able to recognize these based on amino acid structures.
In this example - we have four ionizable groups (write out in order of increasing pKa values).

Ionizable Group	pKa
Carboxy-terminus (COOH)	~2.0
Glutamic acid carboxyl group (COOH)	~4.2
Amino-terminus (NH_3^+)	~9.7
Lysine ε-amino group (NH_3^+)	~10

Recall that below a pH represented by a particular pKa value, that group will be protonated! What is the definition of pKa? Recall that it is based on the dissociation of an acid and is the point at which 50% of the groups are ionized.

This means that at pH 0, all the titratable groups will be protonated, as drawn below.





Now calculate the net charge on the molecule at pH 0.0

peptide N-terminus	Glutamic Acid carboxyl group	Lysine ε-amino group	peptide C-terminus	NET CHARGE ON MOLECULE
+1	0	+1	0	+2

Step 2: At what pH will the net charge on the peptide change by 1 charge unit?

Based on the pKa values: if you increase the pH above 2.0, but keep it below 4.2, you will change the charge by 1. (So, you can pick any pH from 2.01 to 4.19).

Assume pH 3.0: (Above pH 2.0, you deprotonate the peptide C-terminus.)

peptide N-terminus	Glutamic Acid carboxyl group	Lysine ε-amino group	peptide C-terminus	NET CHARGE ON MOLECULE
+1	0	+1	-1	+1

Step 3: At what pH will the net charge on the peptide change by 1 charge unit?

Based on the pKa values: if you increase the pH above 4.2, but keep it below 9.7, you will change the charge by 1. (So, you can pick any pH value from 4.21 to 9.69)

Assume pH 7.0 (Above pH 4.2, you deprotonate the glutamic acid carboxyl group)

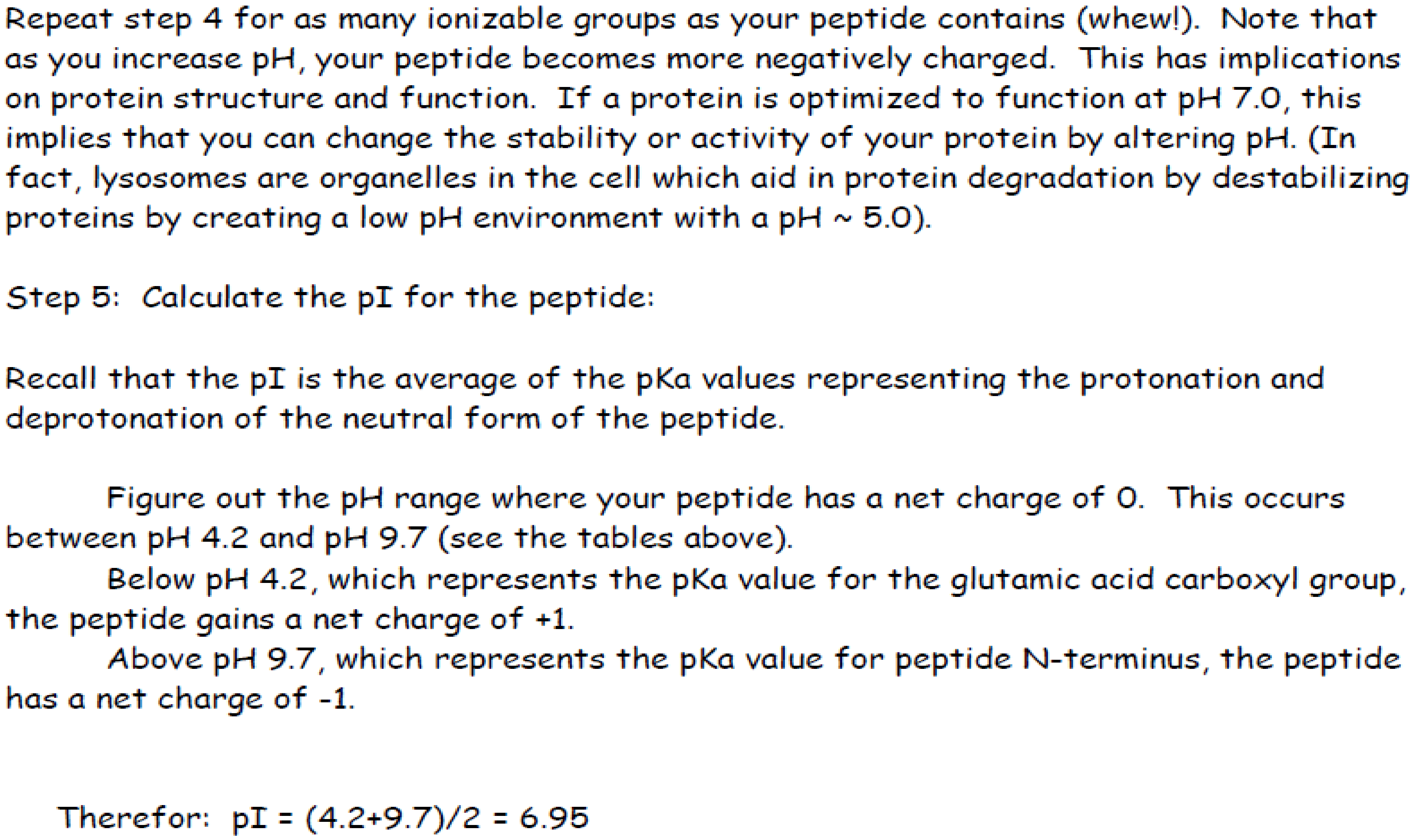
peptide N-terminus	Glutamic Acid carboxyl group	Lysine ε-amino group	peptide C-terminus	NET CHARGE ON MOLECULE
+1	-1	+1	-1	0

Step 4: At what pH will the net charge on the peptide change by 1 charge unit?

Based on the pKa values: if you increase the pH above 9.7, but keep it below 10, you will change the charge by 1. (So, you can pick any pH value from 9.71 to 9.99)

Assume pH 9.8 (Above pH 9.7, you deprotonate the peptide N-terminus)

peptide N-terminus	Glutamic Acid carboxyl group	Lysine ε-amino group	peptide C-terminus	NET CHARGE ON MOLECULE
0	-1	+1	-1	-1



Repeat step 4 for as many ionizable groups as your peptide contains (whew!). Note that as you increase pH, your peptide becomes more negatively charged. This has implications on protein structure and function. If a protein is optimized to function at pH 7.0, this implies that you can change the stability or activity of your protein by altering pH. (In fact, lysosomes are organelles in the cell which aid in protein degradation by destabilizing proteins by creating a low pH environment with a pH ~ 5.0).

Step 5: Calculate the pI for the peptide:

Recall that the pI is the average of the pKa values representing the protonation and deprotonation of the neutral form of the peptide.

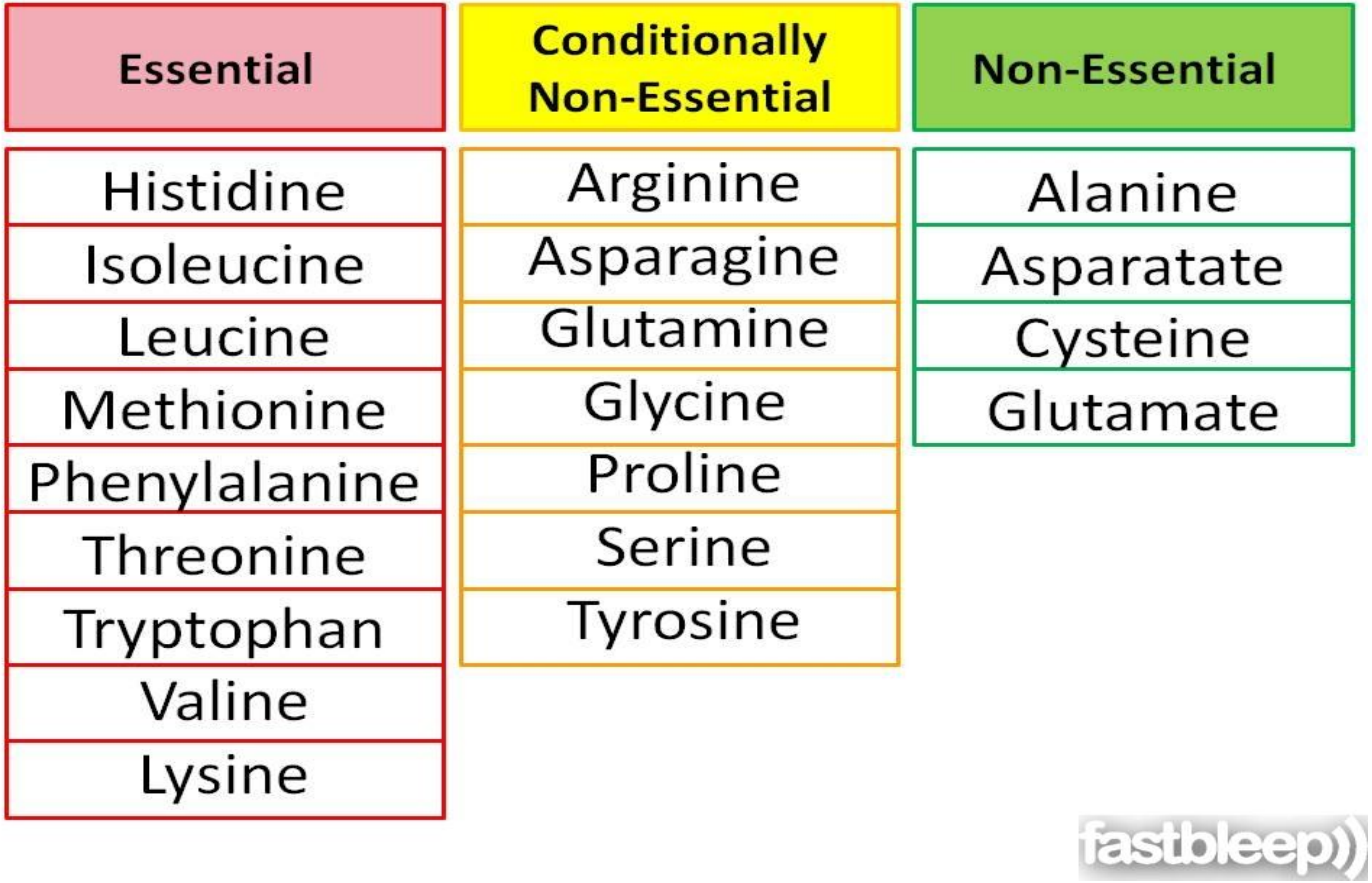
Figure out the pH range where your peptide has a net charge of 0. This occurs between pH 4.2 and pH 9.7 (see the tables above).

Below pH 4.2, which represents the pKa value for the glutamic acid carboxyl group, the peptide gains a net charge of +1.

Above pH 9.7, which represents the pKa value for peptide N-terminus, the peptide has a net charge of -1.

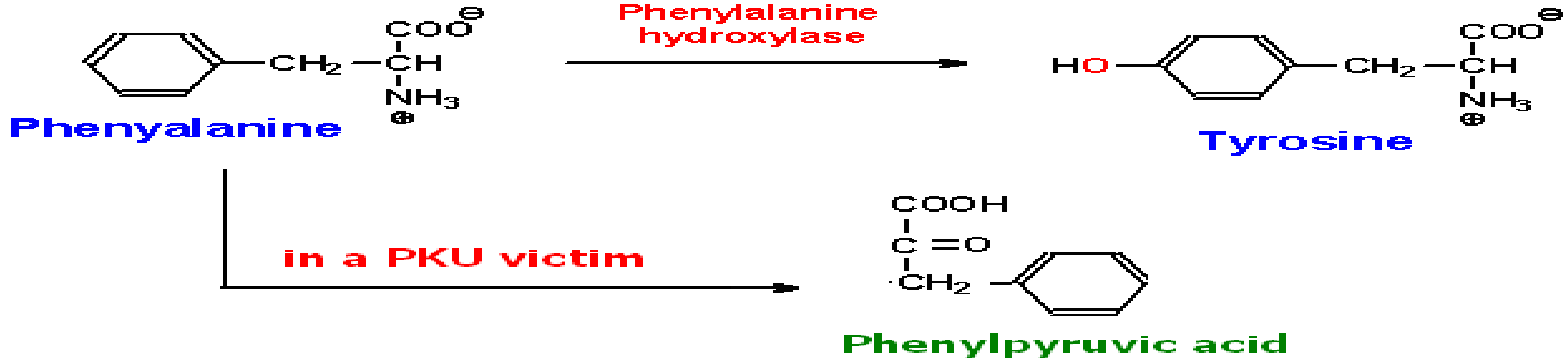
Therefor: pI = (4.2+9.7)/2 = 6.95

- **Essential and non-essential amino acids**
- **Amino acids are join together to form short polymer chains called peptides or longer chains called either polypeptides or proteins. These polymers are linear and un-branched, with each amino acid within the chain attached to two neighboring amino acids.**
-
- **When taken up into the human body from the diet, amino acids are either used to synthesize proteins and other biomolecules or oxidized to urea and carbon dioxide as a source of energy. The oxidation pathway starts with the removal of the amino group by a transaminase, the amino group is then fed into the urea cycle. The other product of transamidation is α-keto acid that enters the citric acid cycle.**
-
- **Of the twenty amino acids, eight are called essential amino acids because the human body cannot synthesize them from other compounds at the level needed for normal growth, so they must be obtained from food. However, the rest can be synthesized in our body and therefore named non-essential amino acids.**
-

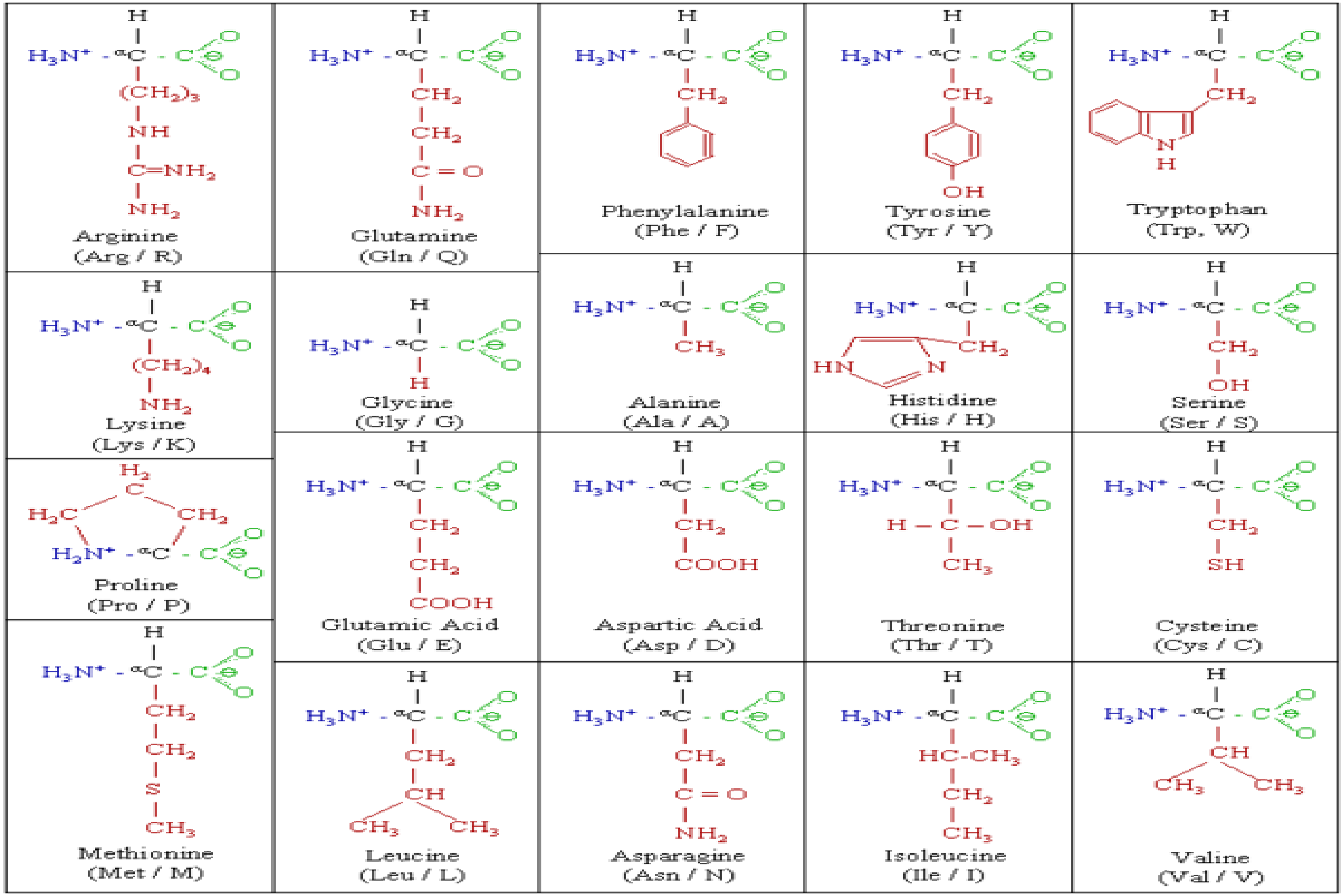


Essential	Conditionally Non-Essential	Non-Essential
Histidine	Arginine	Alanine
Isoleucine	Asparagine	Asparatate
Leucine	Glutamine	Cysteine
Methionine	Glycine	Glutamate
Phenylalanine	Proline	
Threonine	Serine	
Tryptophan	Tyrosine	
Valine		
Lysine		

- **Individual differences in requirements for amino acids:**
- **The amino acid tyrosine is conditionally non-essential where in most people, tyrosine can be synthesized from phenylalanine. A small percent of individuals who have inherited a condition known as phenylketonuria (PKU), which means "phenyl ketones in the urine," do not have the enzyme phenylalanine hydroxylase necessary for this conversion. In these people, the phenylalanine is converted to phenylpyruvic acid.**



- **The diet of a person with PKU must contain some tyrosine and must also be limited in its quantity of phenylalanine; otherwise, excessive amounts of phenylpyruvic acid accumulate in the brain, causing mental retardation.**
- **The structure of the common amino acids found in proteins are:**

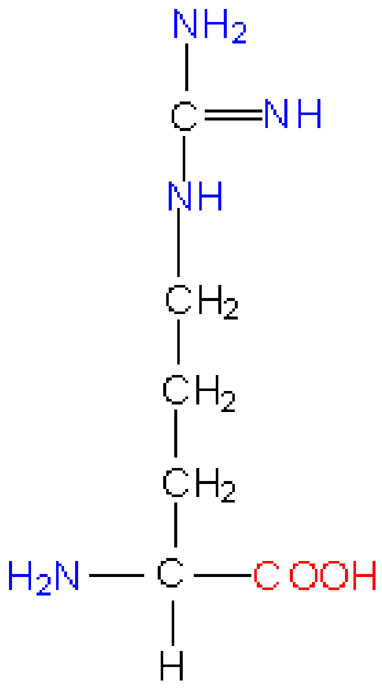

Arginine
(Arg / R)
Glutamine
(Gln / Q)
Phenylalanine
(Phe / F)
Tyrosine
(Tyr / Y)
Tryptophan
(Trp, W)
Lysine
(Lys / K)
Glycine
(Gly / G)
Alanine
(Ala / A)
Histidine
(His / H)
Serine
(Ser / S)
Proline
(Pro / P)
Glutamic Acid
(Glu / E)
Aspartic Acid
(Asp / D)
Threonine
(Thr / T)
Cysteine
(Cys / C)
Methionine
(Met / M)
Leucine
(Leu / L)
Asparagine
(Asn / N)
Isoleucine
(Ile / I)
Valine
(Val / V)

- **Importance of side chain:**
- **There are three types of amino acids:**

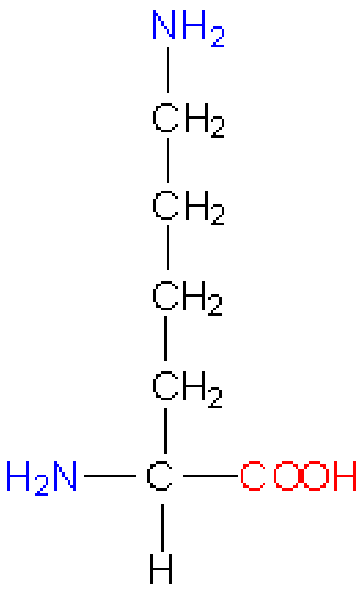
1. Acidic amino acids: They contain more carboxyl groups in their side chain.

aspartic acid - Asp; D glutamic acid - Glu; G

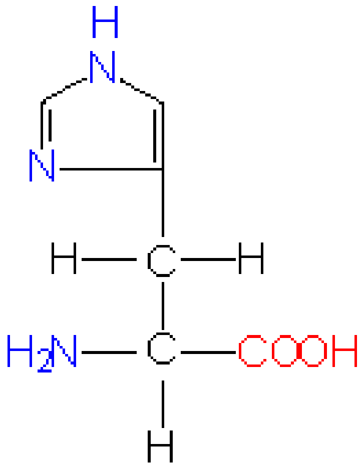
- **Basic amino acids: They contain more amino groups in their side chain.**



arginine - Arg; R

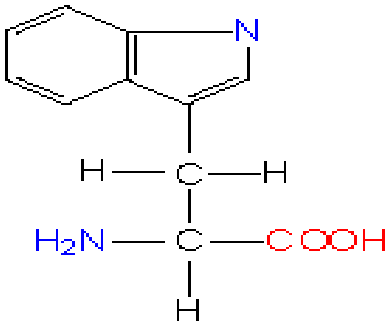


lysine -. Lys; K

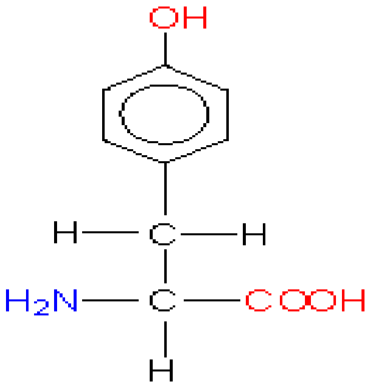


histidine - His; H

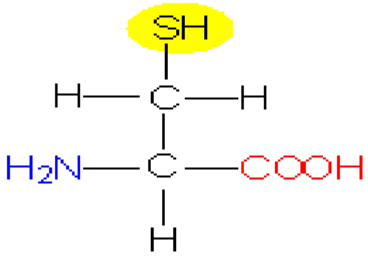
- **Neutral amino group: They contain groups other than acidic or amino in their side chains. They divided into two categories:**
- **a)-Polar neutral amino acids: They contain –OH, -SH, or other polar groups**



tryptophane - Trp; W

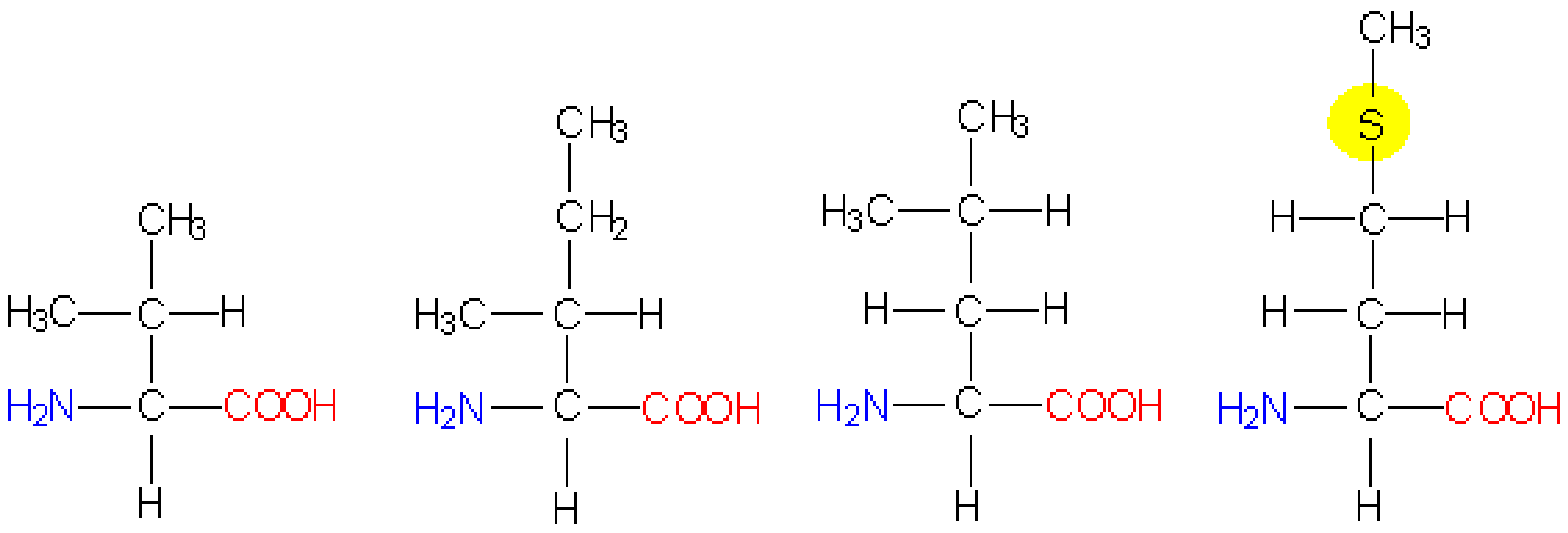


tyrosine - Tyr; Y



cysteine - Cys; C

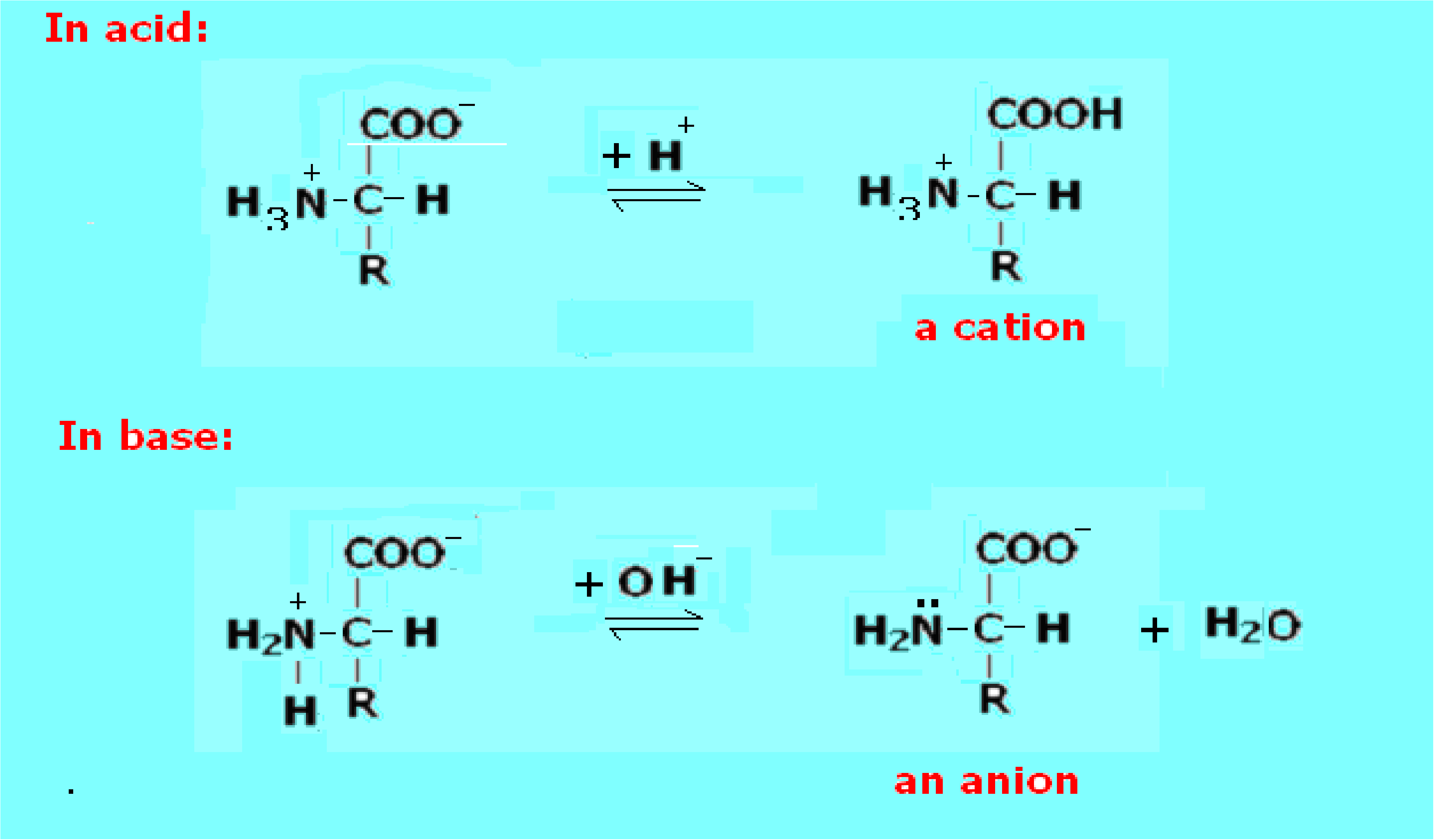
- **b)-Non-polar neutral amino acids: They contain non-polar groups.**



valine - Val; V isoleucine - Ile; I leucine - Leu; L methionine - Met; M

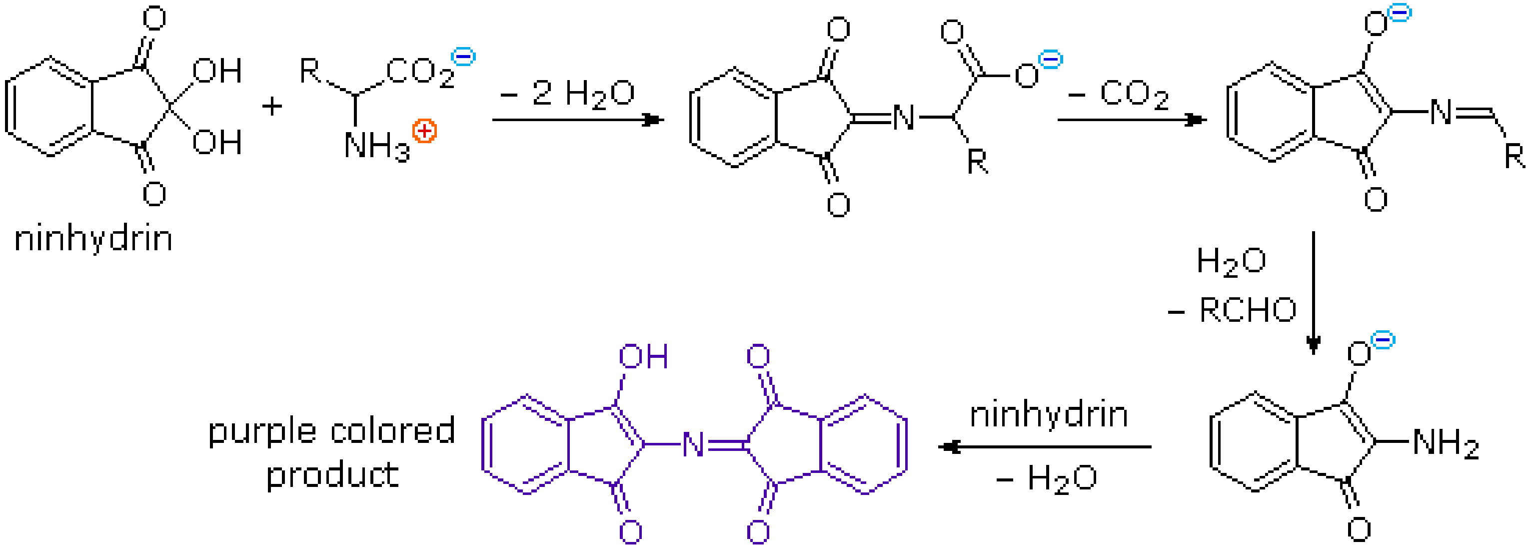
- **This side chains of amino acids determine the structure reactivity and the characteristic of the protein.**
- **The order in which amino acids are found in a protein molecule determines the relationship of the side chains to one another and consequently determines how the protein interacts with itself and with its environment.**
- **For example , a hormone or other water-soluble protein contains many amino acids with polar side chains, while an insoluble muscle protein contains a greater proportion of amino acids with nonpolar side chains.**
-

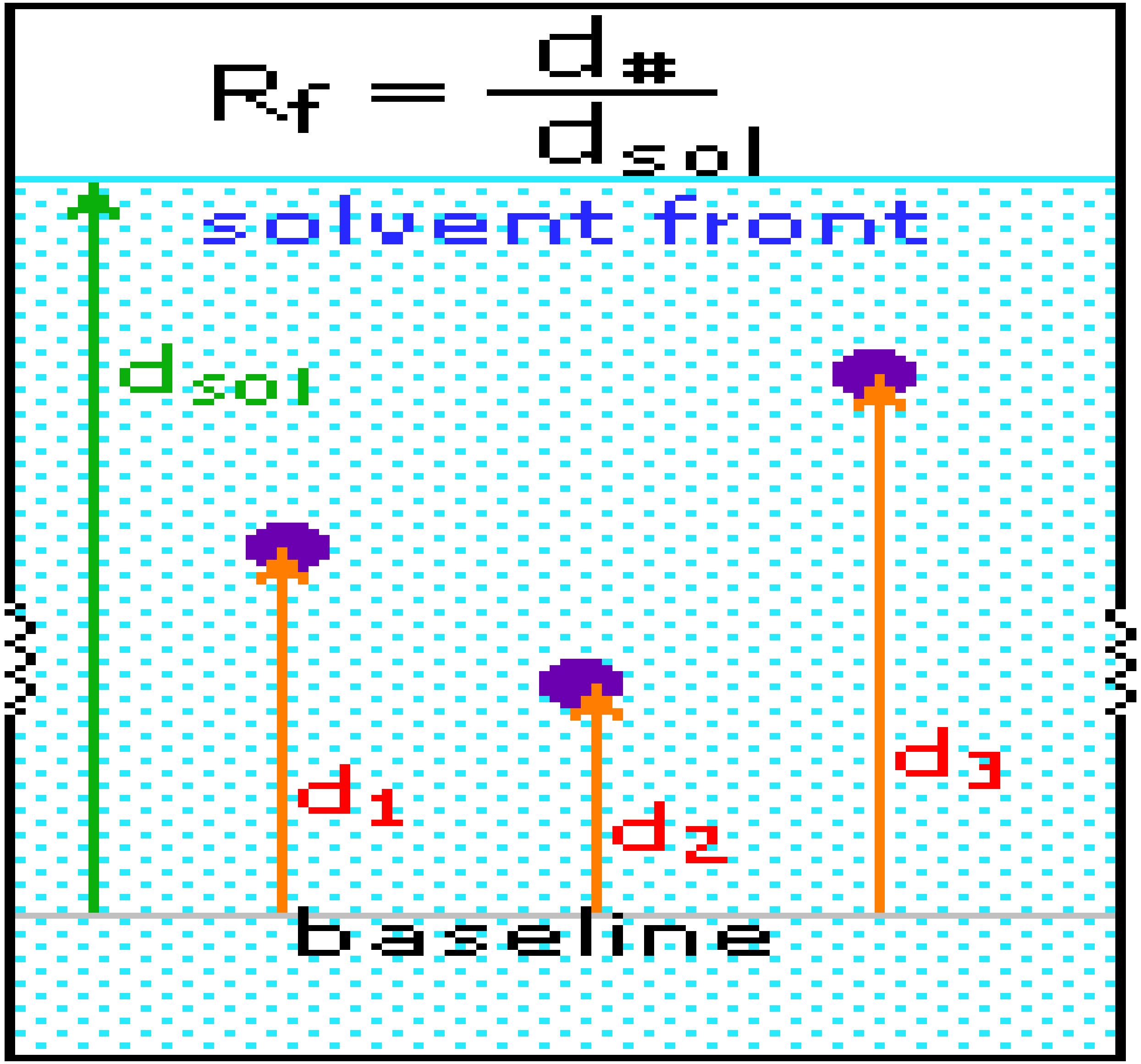
- **Reaction of amino acids:**
- **1. Amphoterism of amino acids:**
- **As amino acids have both a primary amine group and a primary carboxyl group, therefore, an amino acid is amphoteric: it can undergo reaction with either an acid or a base to yield a cation or an anion, respectively.**
-


In acid:
COO⁻
H₃N⁺-C-H
R
+ H⁺
COOH
H₃N⁺-C-H
R
a cation
In base:
COO⁻
H₂N⁺-C-H
H
R
+ OH⁻
COO⁻
H₂N̈-C-H
R
+ H₂O
an anion

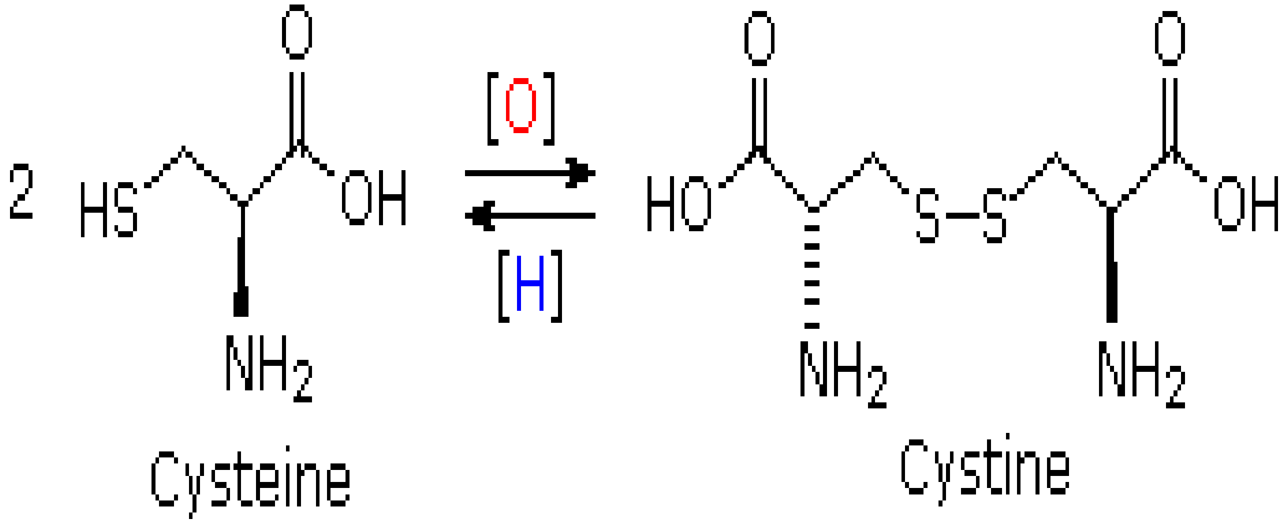
Reaction with ninhydrin:

Amino acids react with ninhydrin to form a blue-violet product called Ruhemann's purple. The reaction is commonly used as a spot test to detect the presence of amino acids on chromatography paper.



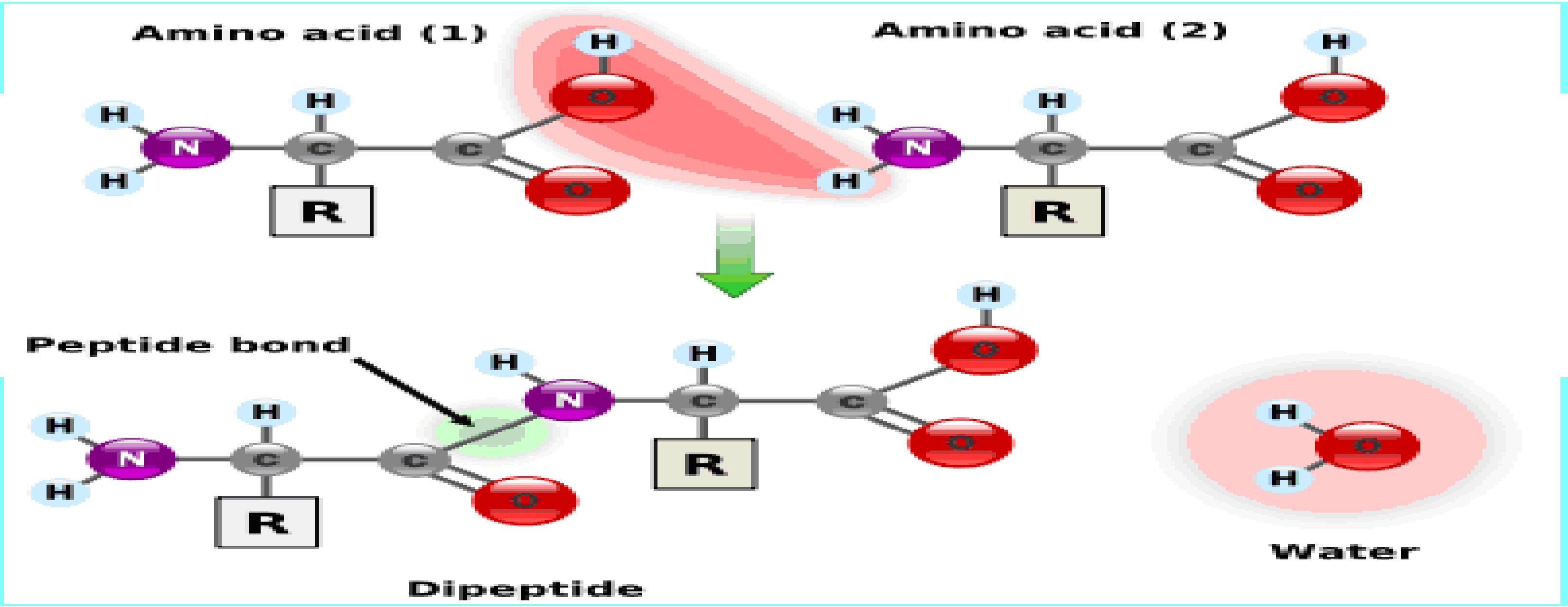

$R_f = \frac{d_{\#}}{d_{sol}}$
solvent front
d_{sol}
d_1
d_2
d_3
baseline

Oxidative coupling:



Peptide bond formation

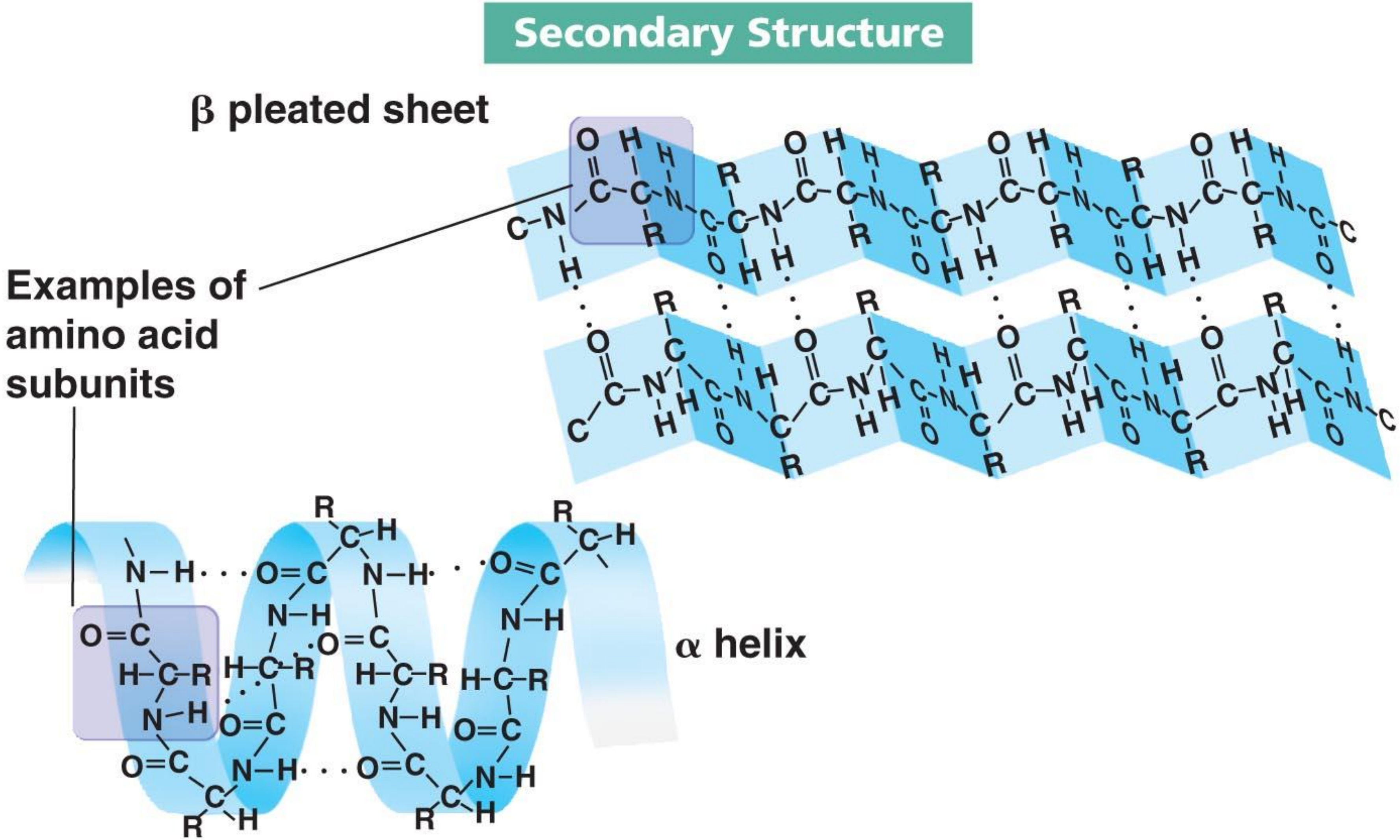
- The condensation of two amino acids to form a [peptide bond]() are carried as described below:

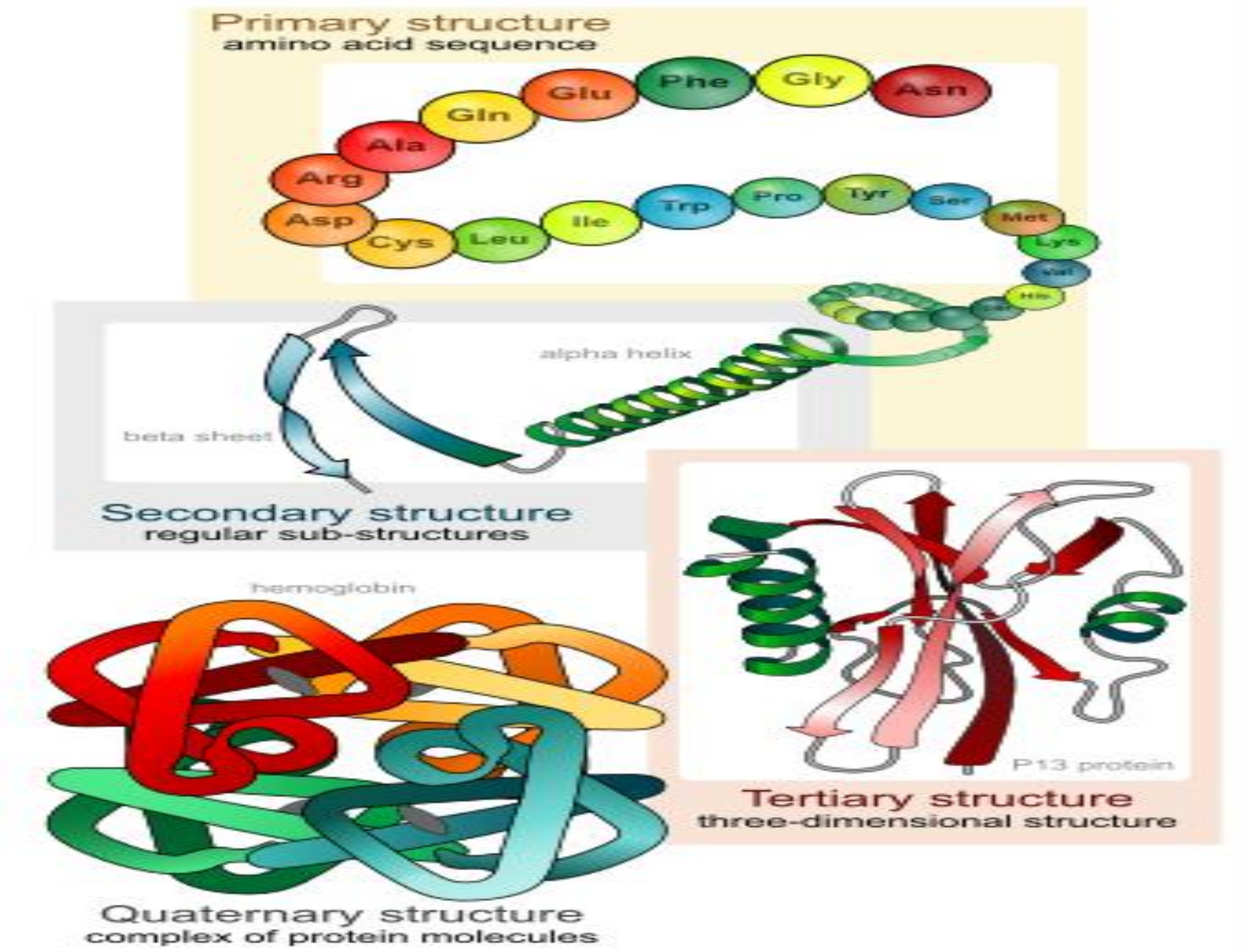


- **As both the amine and carboxylic acid groups of amino acids can react to form amide bonds, one amino acid molecule can react with another and become joined through an amide linkage. This polymerization of amino acids is what creates proteins. This condensation reaction yields the newly formed peptide bond and a molecule of water.**

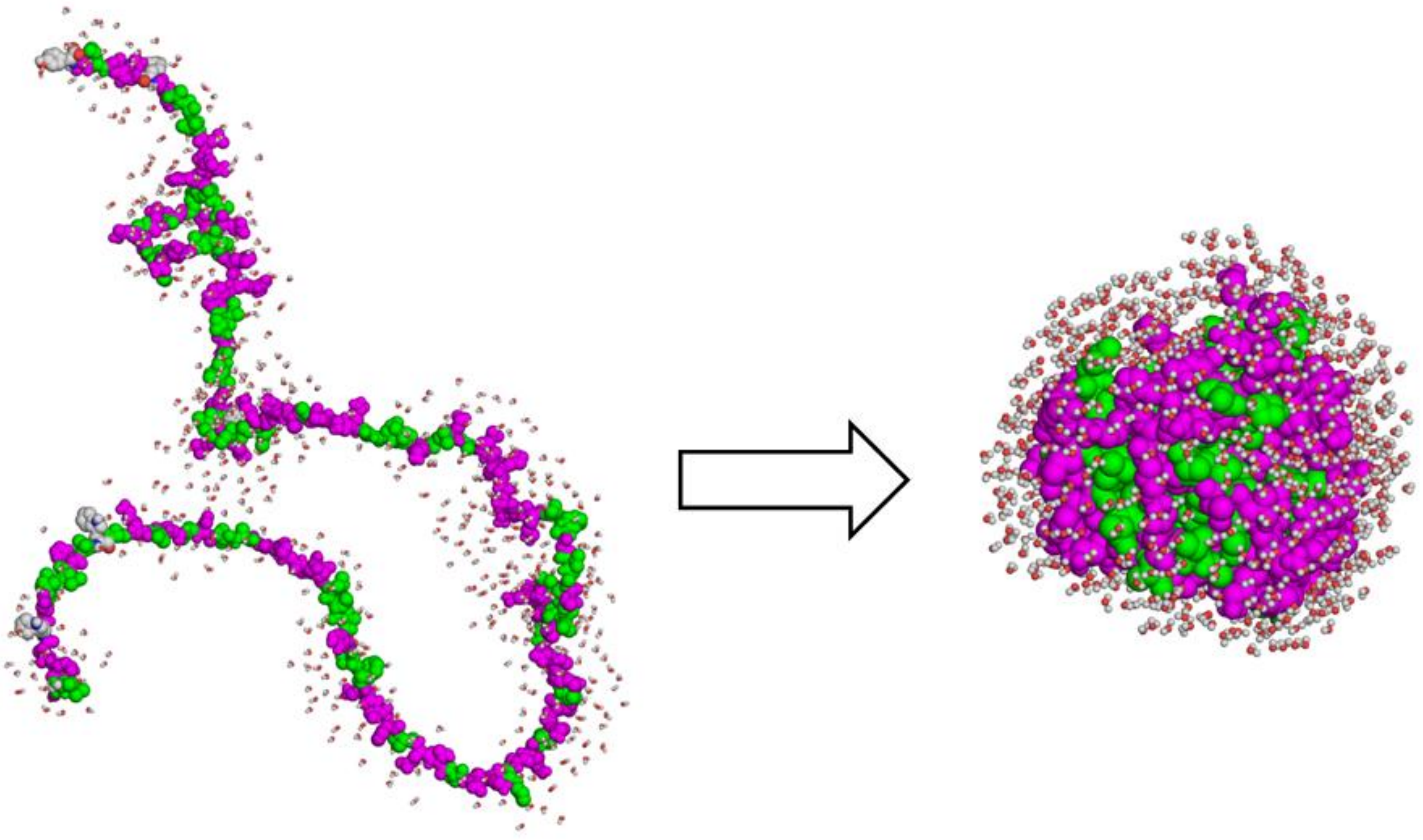
- **Chemical synthesis of proteins**
- **Short proteins can also be synthesized chemically by a family of methods known as peptide synthesis, which rely on organic synthesis techniques such as chemical legation to produce peptides in high yield. Chemical synthesis allows for the introduction of non-natural amino acids into polypeptide chains, such as attachment of fluorescent probes to amino acid side chains. These methods are useful in laboratory biochemistry and cell biology, though generally not for commercial applications. Chemical synthesis is inefficient for polypeptides longer than about 300 amino acids, and the synthesized proteins may not readily assume their native tertiary structure. Most chemical synthesis methods proceed from C-terminus to N-terminus, opposite the biological reaction.**

- **Structure of proteins**
- **Most proteins fold into unique 3-dimensional structures. The shape into which a protein naturally folds is known as its native state.**
-
- **Biochemists often refer to four distinct aspects of a protein's structure (Higher structure of proteins):**
- **1. Primary structure: The sequence of amino acids in a protein molecule is called the primary structure of the protein.**
- **The primary structure of a segment of a polypeptide chain or of a protein is the amino-acid sequence of the polypeptide chain(s), without regard to spatial arrangement (apart from configuration at the alpha-carbon atom).**
-
- **2. Secondary structure: regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the alpha helix and beta sheet. Because secondary structures are local, many regions of different secondary structure can be present in the same protein molecule.**
- **The common secondary structures in proteins are namely α-helix, β-sheets and turns.**
- **Anti-parallel beta-sheets are more often twisted than parallel sheets.**
- **Turns are the third of the three "classical" secondary structures that serve to reverse the direction of the polypeptide chain. They are located primarily on the protein surface and accordingly contain polar and charged residues.**


Secondary Structure
β pleated sheet
Examples of amino acid subunits
α helix


Primary structure
amino acid sequence
Glu
Phe
Gly
Asn
Gln
Ala
Arg
Asp
Cys
Leu
Ile
Trp
Pro
Tyr
Ser
Met
Lys
Val
His
alpha helix
beta sheet
Secondary structure
regular sub-structures
hemoglobin
P13 protein
Tertiary structure
three-dimensional structure
Quaternary structure
complex of protein molecules

- **3. Tertiary structure: the overall shape of a single protein molecule; the spatial relationship of the secondary structures to one another. Tertiary structure is generally stabilized by non-local interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen bonds, disulfide bonds, and even post-translational modifications. The term "tertiary structure" is often used as synonymous with the term fold.**
-
- **The tertiary structure of a protein molecule, or of a subunit of a protein molecule, is the arrangement of all its atoms in space, without regard to its relationship with neighboring molecules or subunits.**
- **The folding of the backbone upon itself to form a sphere, are called the tertiary structure.**



Unfolded

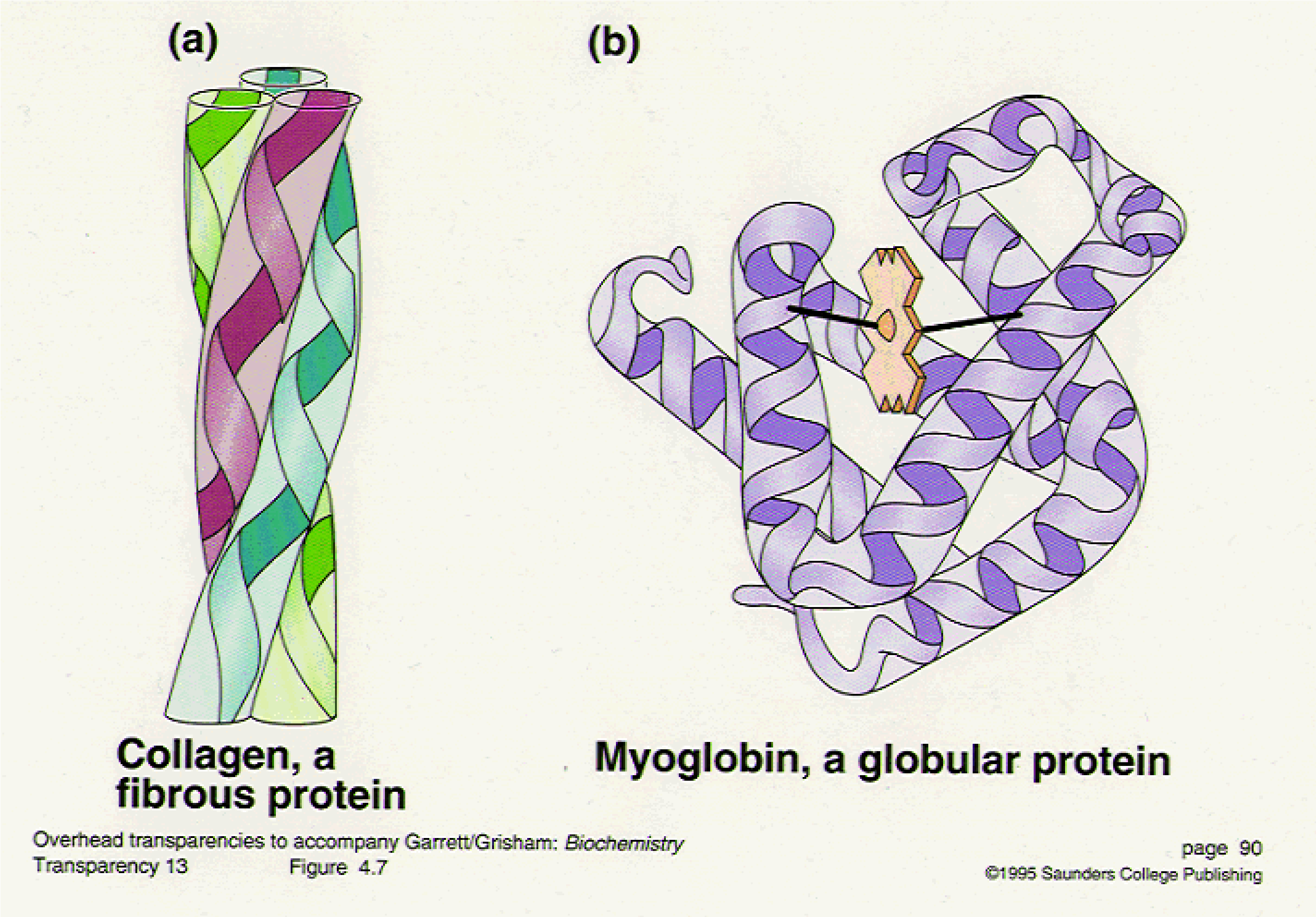
Folded

- **4. Quaternary structure: the shape or structure that results from the interaction of more than one protein molecule, usually called protein subunits in this context, which function as part of the larger assembly or protein complex.**
- **The quaternary structure of a protein molecule is the arrangement of its subunits in space and the ensemble of its intersubunit contacts and interactions, without regard to the internal geometry of the subunits. The subunits in a quaternary structure must be in non-covalent association. Haemoglobin contains four polypeptide chains held together non-covalently in a specific conformation as required for its function.**

- **Classes of Proteins:**

- **Proteins can be informally divided into three main classes, which correlate with typical tertiary structures:**

- **1.Globular proteins: These are small proteins, somewhat spherical in shape because of folding of the protein chains upon themselves. Globular proteins are water-soluble and perform various functions in an organism. For example, hemoglobin transports oxygen to the cells; insulin aids in carbohydrates metabolism; antibodies render foreign protein inactive; fibrinogen (soluble) can form insoluble fibers that result in blood clots; and hormones carry messages throughout the body. Examples of globular proteins are albumines, globulines and histones.**



Collagen, a fibrous protein

Myoglobin, a globular protein

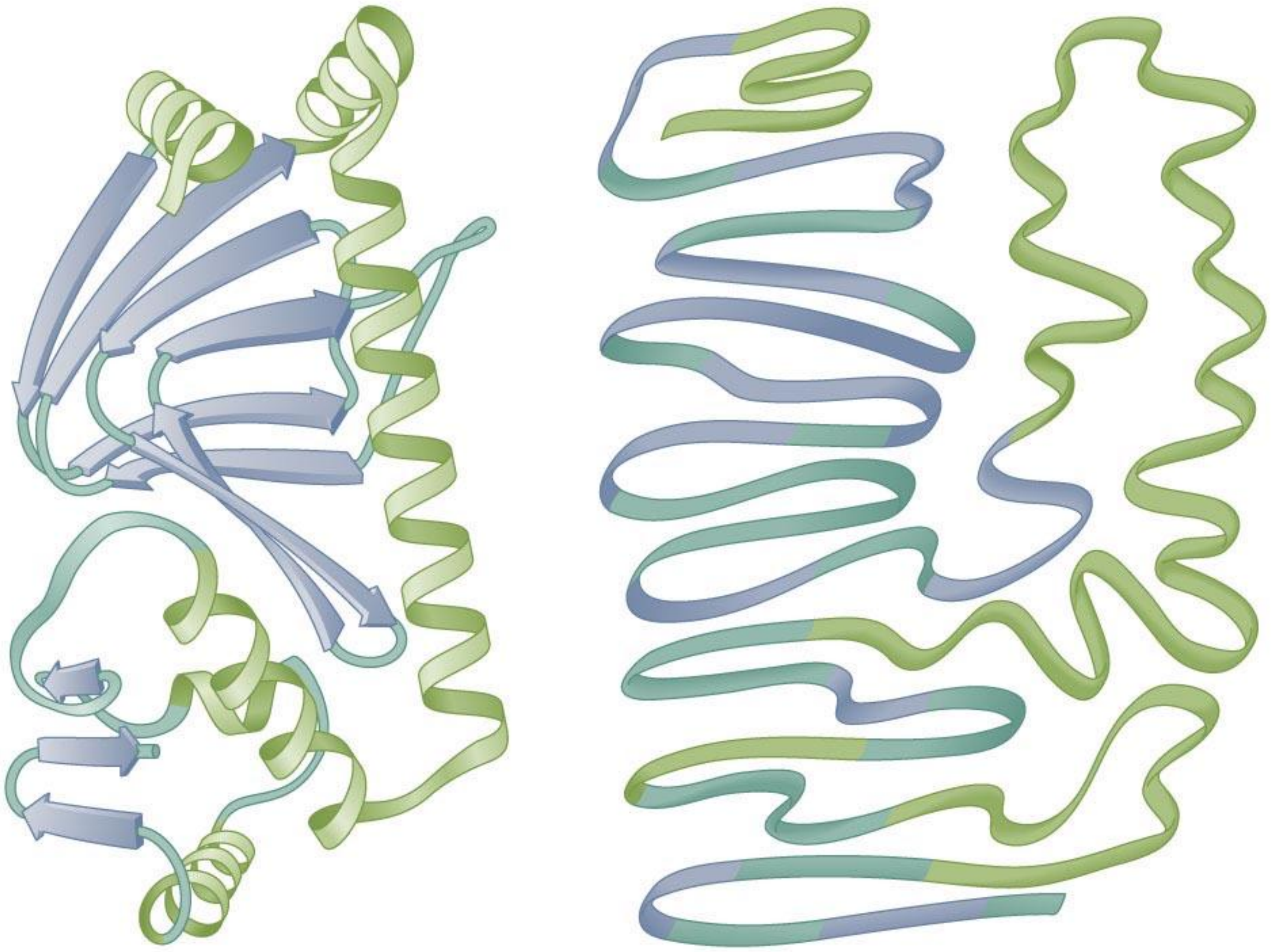
Overhead transparencies to accompany Garrett/Grisham: *Biochemistry*
Transparency 13 Figure 4.7

page 90

- **2.Fibrous proteins (structural proteins): They are composed of long thread-like molecules that are tough and insoluble. It forms skin, muscles, the walls of arteries and hair. Examples of fibrous proteins are collagens, elastins and keratins.**
-
- **3.Conjugated proteins (membrane proteins): Proteins connected to a non-protein moiety such as a sugar, perform various functions throughout the body. A common mode of action linkage between the protein and non-protein is by a functional side chain of the protein. For example, an acidic side chain of the protein can form an ester with an –OH group of a sugar molecule. Examples of conjugated proteins are nucleoproteins, mucoproteins, glycoproteins and lipoproteins.**
-
- **Almost all globular proteins are soluble and many are enzymes. Fibrous proteins are often structural; conjugated proteins often serve as receptors or provide channels for polar or charged molecules to pass through the cell membrane.**
-
- **A special case of intramolecular hydrogen bonds within proteins, poorly shielded from water attack and hence promoting their own dehydration, are called dehydrons.**
-

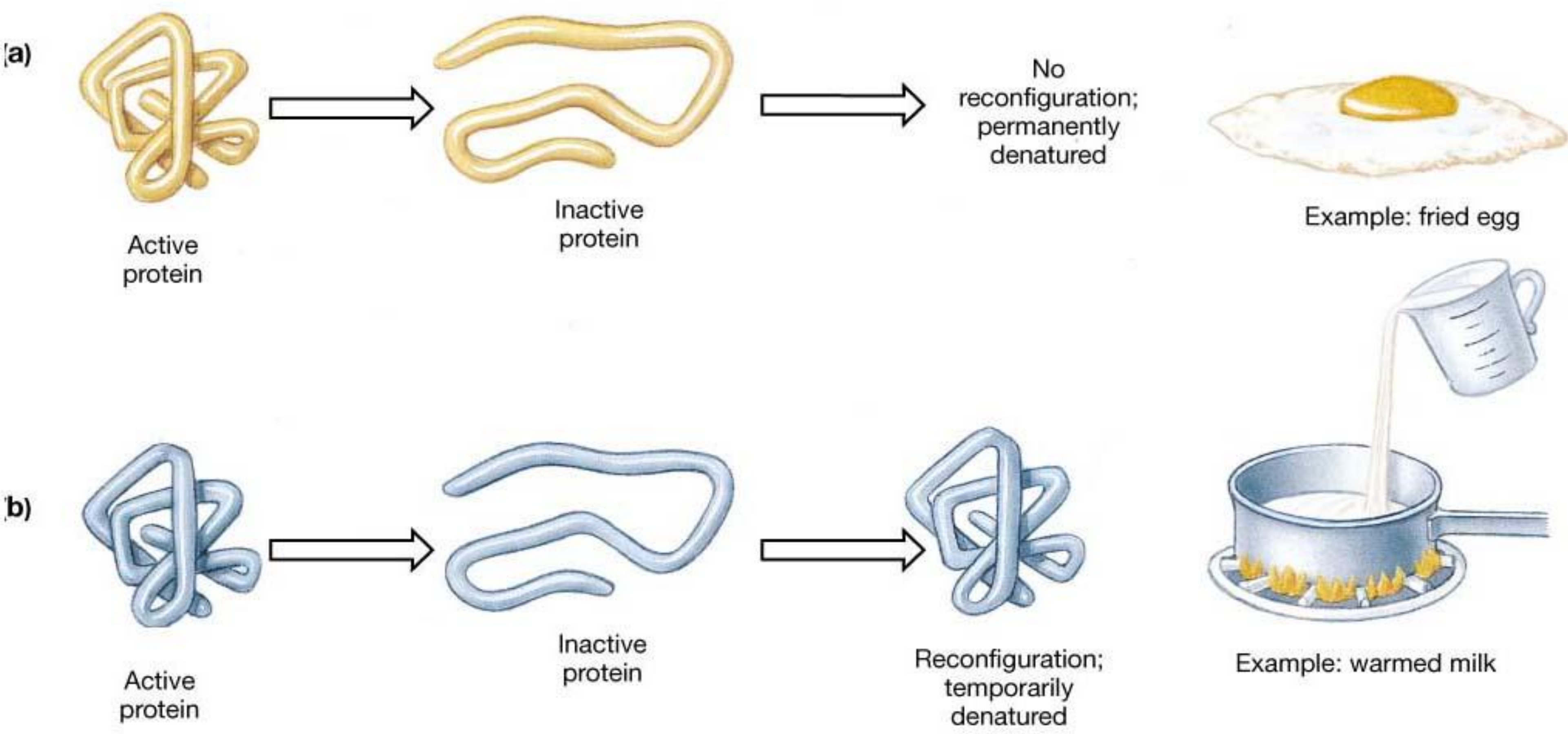
- **Cellular functions**
- **The best-known role of proteins in the cell is their duty as enzymes, which catalyze chemical reactions. Enzymes are usually highly specific catalysts that accelerate only one or a few chemical reactions. Enzymes carry out most of the reactions involved in metabolism and catabolism, as well as DNA replication, DNA repair, and RNA synthesis. Some enzymes act on other proteins to add or remove chemical groups in a process known as post-translational modification. About 4,000 reactions are known to be catalyzed by enzymes. The rate acceleration conferred by enzymatic catalysis is often enormous - as much as 10^{17}-fold increase in rate over the un-catalyzed reaction in the case of orotate decarboxylase (78 million years without the enzyme, 18 milliseconds with the enzyme).**
-
- **The molecules bound and acted upon by enzymes are known as substrates. Although enzymes can consist of hundreds of amino acids, it is usually only a small fraction of the residues that come in contact with the substrate, and an even smaller fraction - 3-4 residues on average - that are directly involved in catalysis. The region of the enzyme that binds the substrate and contains the catalytic residues is known as the active site.**

- **Protein Denaturation**
- **Protein denaturation has been defined in several ways, for example as a change in solubility or by simultaneous changes in chemical, physical and biological properties under some standard reference set of conditions. These changes in physical, and to a lesser extent chemical properties are manifestations of configurational changes taking place in the polypeptide chains.**
-
- **The denaturation process presumably involves an unfolding or at least an alteration in the nature of the folded structure. Most denaturation changes consist of changes in secondary bonds: ion-dipole, hydrogen and Van der Waals, and in the rotational positions about single bonds which are controlled by the secondary bond structure.**
-
- **The term denaturation denotes the response of the native protein to heat, acid, alkali, and a variety of other chemical and physical agents which cause marked changes in the protein structure. Rice et al. (1958) suggested denaturation to mean a class of reactions which lead to changes in the structure of the macromolecule with no change in molecular weight.**
-



Functional protein **Denatured protein**

- Timasheff and Gibbs (1957), pointed out that the approaches used to define the concept of denaturation can be classified into two types:
- molecular, in terms of actual structural changes taking place on the molecule, and
- operational, in terms of changes in measurable properties.
- The operational approach to denaturation has the advantage of being purely phenomenological, but it cannot lead to a precise definition, since the concept of property, even restricted to measurable property, is itself without finite boundary. On the other hand, the molecular point of view, although still involves a few assumptions despite the important advances in investigations of protein structure, will allow us to reach a definition which is very convenient, owing to its general validity. Whatever it is, changes can be quantitatively described by comparison with the native state.
-
- The denaturation process can be achieved by any one of the following methods:
- Increasing temperature.
- Changing pH.
- Using denaturants (i.e. urea, guanidine hydrochloride, beta-mercaptoethanol, dithiothreitol).
- Using inorganic salts (i.e lithium bromide, potassium thiocyanate, sodium iodide).
- Using organic solvents and (i.e. formamide, dimethylformamide, dichloro- and trichloroacetic acids and their salts).
- Using detergents (i.e. sodium dodecyl sulphate).
- High pressure.
- Ultrasonic homogenization.


(a)
Active protein
Inactive protein
No reconfiguration; permanently denatured
Example: fried egg
(b)
Active protein
Inactive protein
Reconfiguration; temporarily denatured
Example: warmed milk

Next lecture :

- Enzymes
- DNA and RNA
- Vitamins
- Hormones