BME 2901-BIOCHEMISTRY

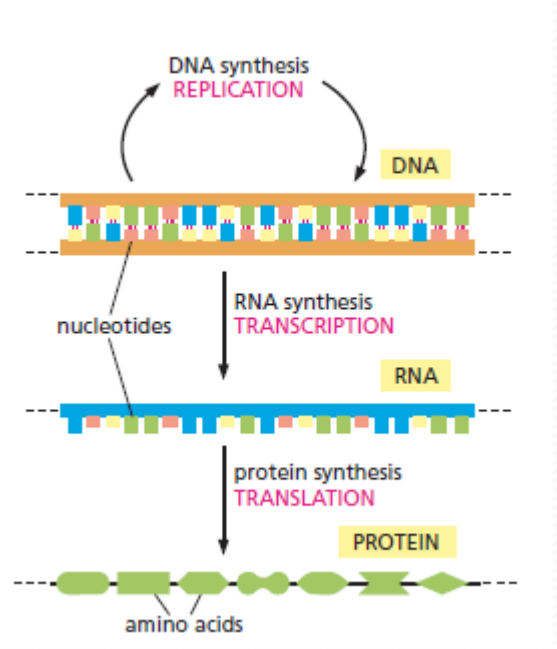
Aminoacids, Peptides and Proteins

by Assist. Prof. Görke Gürel Peközer

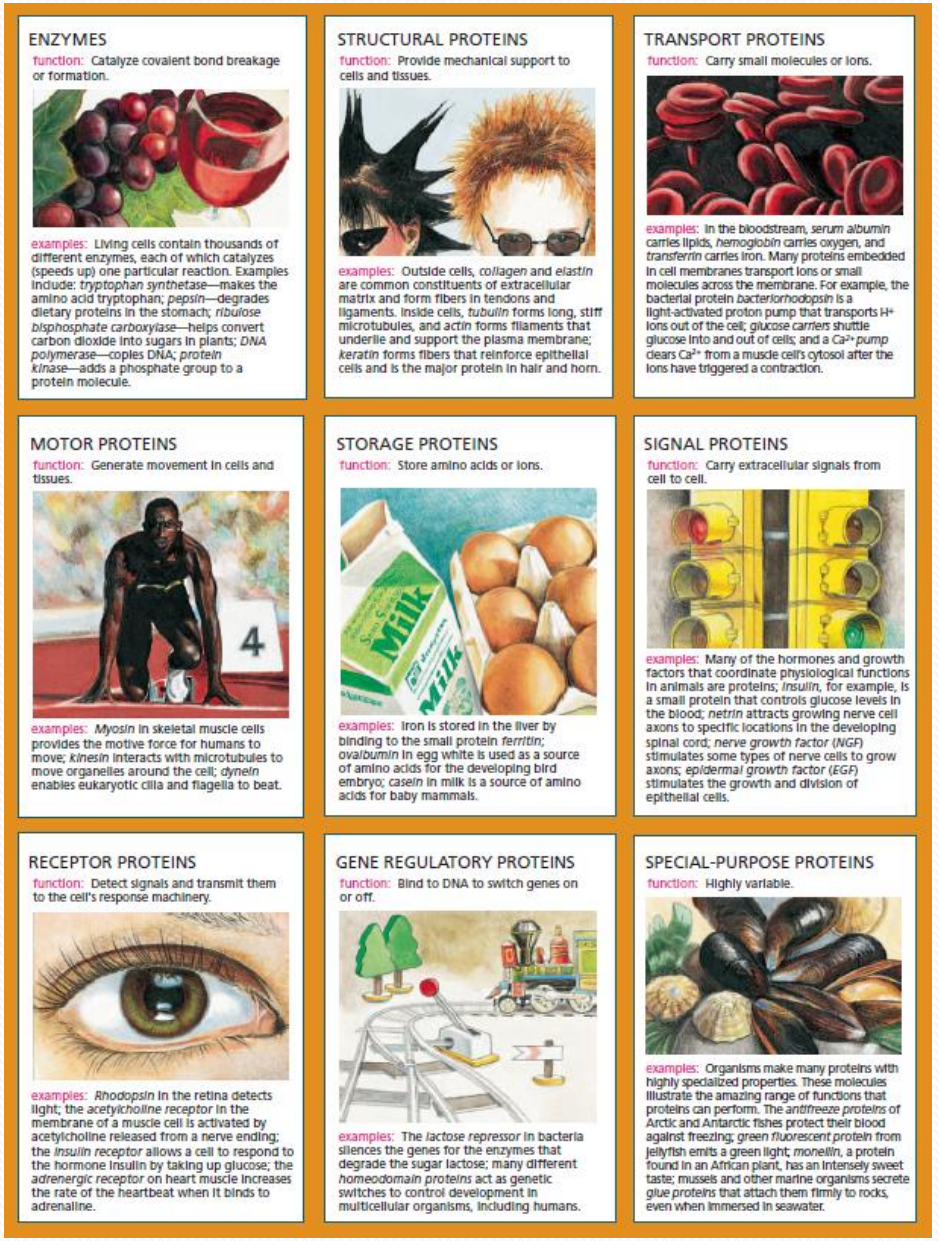
Yıldız Technical University
Biomedical Engineering Department
Fall 2019

Proteins

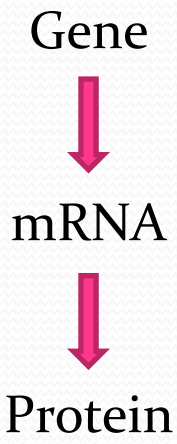
- Proteins are the main building blocks from which cells are assembled, and they constitute most of the cell's dry mass.
- They are the molecular instruments through which genetic information is expressed.
- The multiplicity of functions carried out by proteins arises from the huge number of different shapes they adopt.



- From a chemical point of view, proteins are by far the most structurally complex and functionally sophisticated molecules known.
- This is perhaps not surprising, considering that the structure and activity of each protein has been developed and fine-tuned over billions of years of evolution.



- How many proteins do you expect to be found in humans?



	Total	Code for proteins
Gene	60656	19954
RNA	229580	84485
Protein	92179	

Aminoacids

- All proteins, whether from the most ancient lines of bacteria or from the most complex forms of life, are constructed from the same ubiquitous set of 20 amino acids, covalently linked in characteristic linear sequences.
- Because each of these amino acids has a side chain with distinctive chemical properties, this group of 20 precursor molecules may be regarded as the alphabet in which the language of protein structure is written.

Protein Diversity

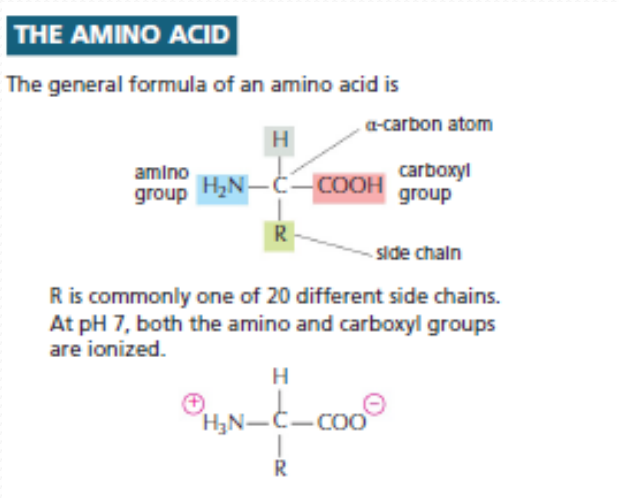
- Cells can produce proteins with strikingly different properties and activities by joining the same 20 amino acids in many different combinations and sequences.
- From these building blocks different organisms can make such widely diverse products as enzymes, hormones, antibodies, transporters, muscle fibers, the lens protein of the eye, feathers, spider webs, rhinoceros horn, milk proteins, antibiotics, mushroom poisons.

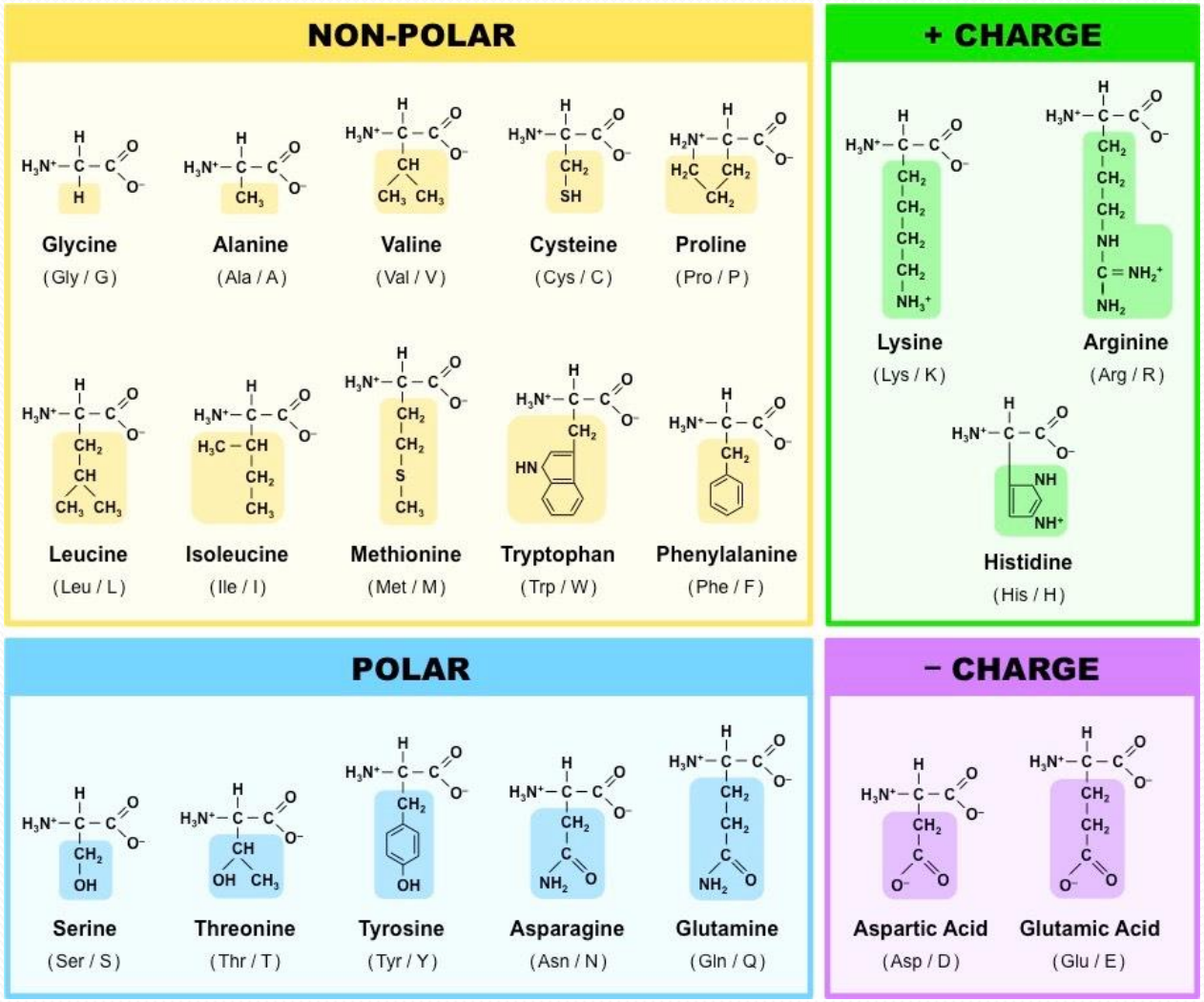
Proteins: Polymers of Aminoacids

- A protein molecule is made from a long chain of these **amino acids**, held together by covalent **peptide bonds**.
- Proteins are therefore referred to as **polypeptides**, and their amino acid chains are called **polypeptide chains**.
- In each type of protein, the amino acids are present in a unique order, called the **amino acid sequence**, which is exactly the same from one molecule of that protein to the next.
- One molecule of human insulin, for example, has the same amino acid sequence as every other molecule of human insulin.

Aminoacids

- All 20 of the common amino acids are α-amino acids.
- They have a carboxyl group, an amino group and H atom bonded to the same carbon atom (the α carbon).
- They differ from each other in their side chains, or **R groups**, which vary in structure, size, and electric charge, and influence the solubility of the amino acids in water.

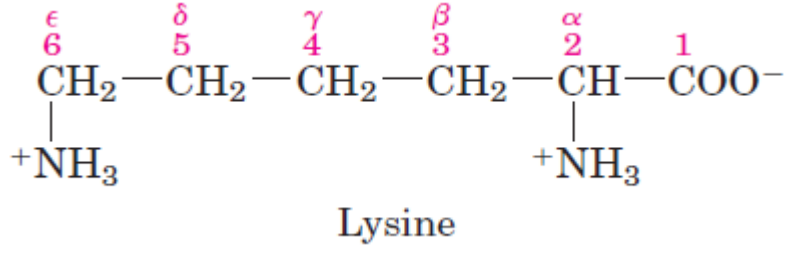




The common amino acids of proteins have been assigned **three-letter abbreviations** and **one-letter symbols**, which are used as shorthand to indicate the composition and sequence of amino acids polymerized in proteins.

Carbon numbering in aminoacids

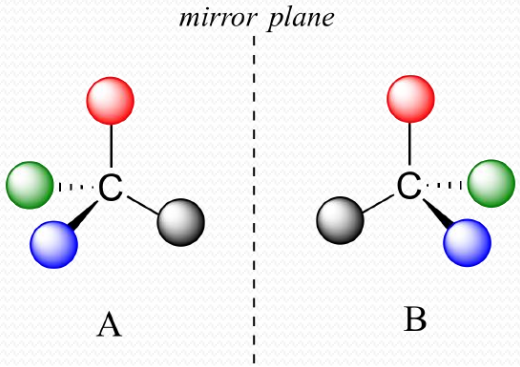
- Aminoacids have a carboxyl group, an amino group and H atom bonded to the same carbon atom (the α carbon).
- The additional carbons in an R group are commonly designated α, β, γ, δ, ε and so forth, proceeding out from the α carbon.



Lysine

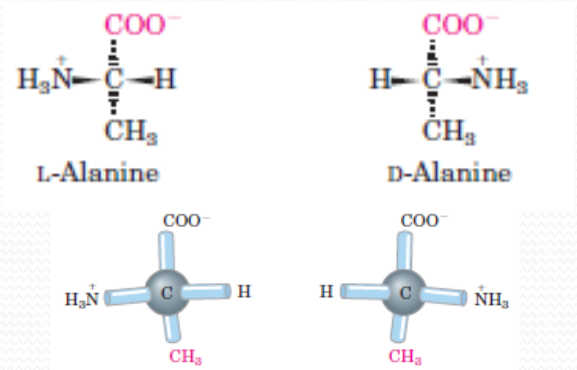
Revisit to Stereochemistry

- Chiral center: a tetrahedral atom (usually Carbon) that is bonded to 4 different groups.
- Chiral molecules are the ones that have at least 1 chiral center in their structure.
- Chiral molecules cannot be superimposed on its mirror image by any combination or rotations.



α-carbon atom is a chiral center

- For all amino acids except glycine, the carbon is bonded to four different groups: a carboxyl group, an amino group, an R group, and a hydrogen atom.
- The α-carbon atom is thus a **chiral center**. thus amino acids have two possible stereoisomers: D-aminoacids and L-aminoacids.

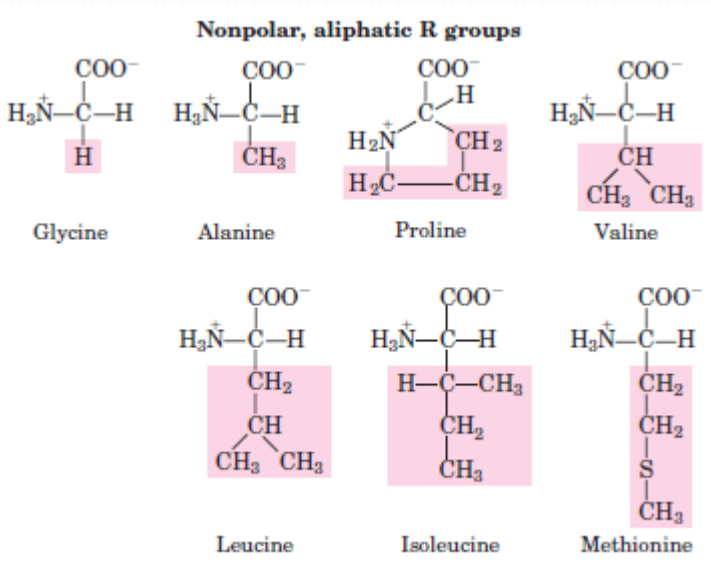


Aminoacids in proteins: L stereoisomers

- Nearly all biological compounds with a chiral center occur naturally in only one stereoisomeric form, either D or L. The amino acid residues in protein molecules are exclusively L stereoisomers.
- D-Amino acid residues have been found only in a few, generally small peptides, including some peptides of bacterial cell walls and certain peptide antibiotics.
- Cells are able to specifically synthesize the L isomers of amino acids because the active sites of enzymes are asymmetric, causing the substrate not to 'fit' in the active site of the enzyme and thus reactions they catalyze to be stereospecific.

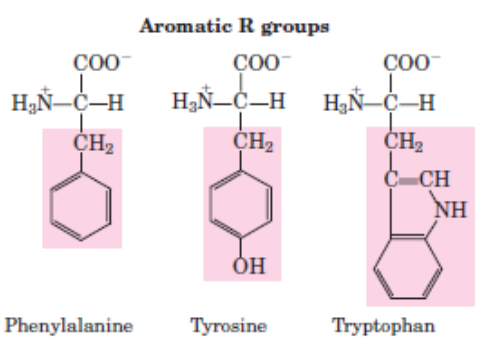
Aminoacids with Nonpolar, Aliphatic R Groups

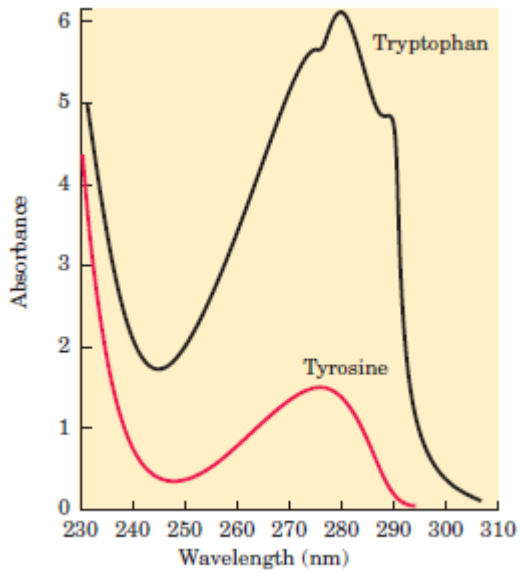
- Nonpolar aminoacids tend to cluster together within proteins, stabilizing protein structure by means of hydrophobic interactions.
- **Glycine** has the simplest structure. Although it is formally nonpolar, its very small side chain makes no real contribution to hydrophobic interactions.
- **Proline** has an aliphatic side chain with a distinctive cyclic structure. The secondary amino (imino) group of proline residues is held in a rigid conformation that reduces the structural flexibility of polypeptide regions containing proline.



Aminoacids with Aromatic R Groups

- The hydroxyl group of tyrosine can form hydrogen bonds, and it is an important functional group in some enzymes.
- Tyrosine and tryptophan are significantly more polar than phenylalanine, because of the hydroxyl group in tyrosine and the nitrogen of the tryptophan indole ring.
- Tryptophan and tyrosine, and to a much lesser extent phenylalanine, absorb ultraviolet light. This accounts for the characteristic strong absorbance of light by most proteins at a wavelength of 280 nm, a property exploited by researchers in the characterization of proteins.





Absorption of Light by Molecules: The Lambert-Beer Law

A wide range of biomolecules absorb light at characteristic wavelengths, just as tryptophan absorbs light at 280 nm (see Fig. 3–6). Measurement of light absorption by a spectrophotometer is used to detect and identify molecules and to measure their concentration in solution. The fraction of the incident light absorbed by a solution at a given wavelength is related to the thickness of the absorbing layer (path length) and the concentration of the absorbing species (Fig. 1). These two relationships are combined into the Lambert-Beer law,

$$\log \frac{I_0}{I} = \varepsilon c l$$

where I_0 is the intensity of the incident light, I is the intensity of the transmitted light, ε is the molar extinction coefficient (in units of liters per mole-centimeter), c is the concentration of the absorbing species (in moles per liter), and l is the path length of the light-absorbing sample (in centimeters). The Lambert-Beer law assumes that the incident light is parallel and monochromatic (of a single wavelength) and that the solvent and solute molecules are randomly oriented. The expression $\log (I_0/I)$ is called the **absorbance**, designated A.

It is important to note that each successive millimeter of path length of absorbing solution in a 1.0 cm cell absorbs not a constant amount but a constant fraction of the light that is incident upon it. However, with an absorbing layer of fixed path length, *the absorbance,* A, *is directly proportional to the concentration of the absorbing solute.*

The molar extinction coefficient varies with the nature of the absorbing compound, the solvent, and the wavelength, and also with pH if the light-absorbing species is in equilibrium with an ionization state that has different absorbance properties.

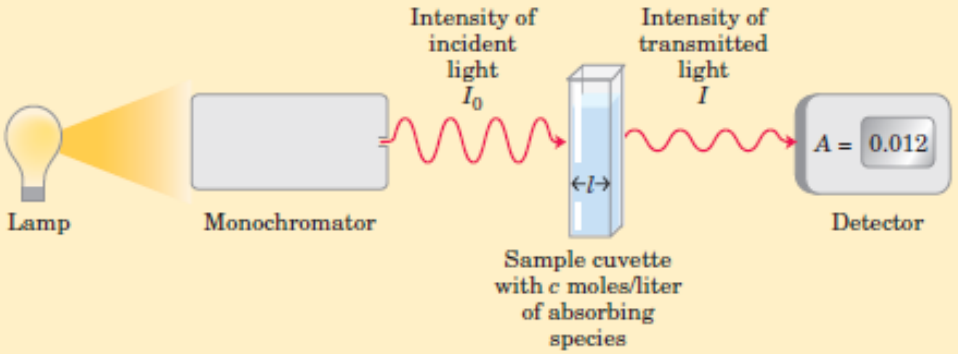
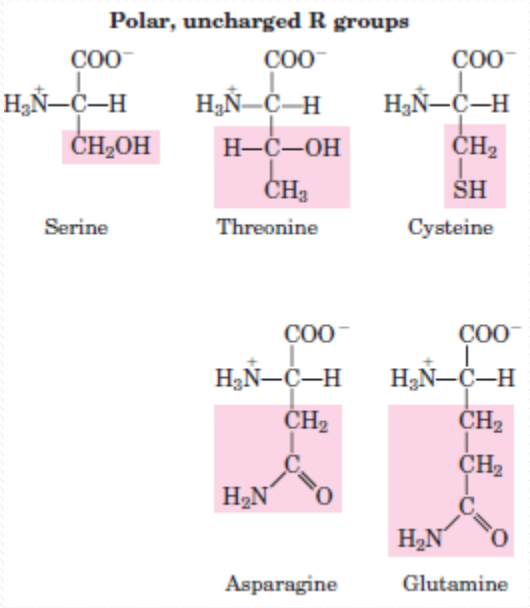


FIGURE 1 The principal components of a spectrophotometer. A light source emits light along a broad spectrum, then the monochromator selects and transmits light of a particular wavelength. The monochromatic light passes through the sample in a cuvette of path length l and is absorbed by the sample in proportion to the concentration of the absorbing species. The transmitted light is measured by a detector.

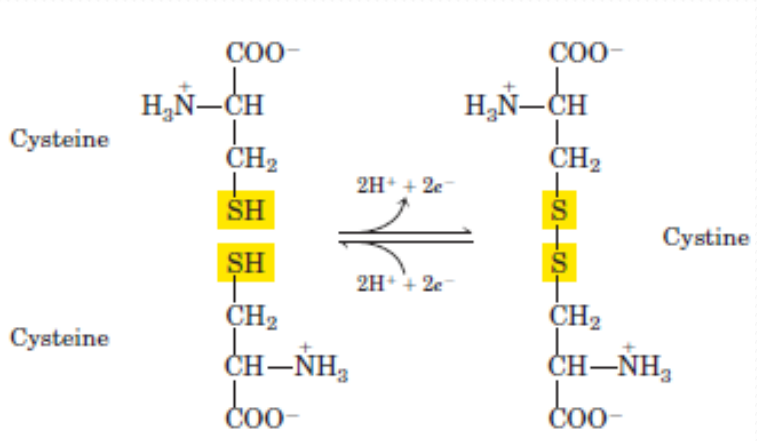
Aminoacids with Polar, Uncharged R Groups

- The R groups of these amino acids are more soluble in water, or more hydrophilic, than those of the nonpolar amino acids, because they contain functional groups that form hydrogen bonds with water.
- This class of amino acids includes **serine, threonine, cysteine, asparagine**, and **glutamine.**
- The polarity of serine and threonine is contributed by their hydroxyl groups; that of cysteine by its sulfhydryl group; and that of asparagine and glutamine by their amide groups.



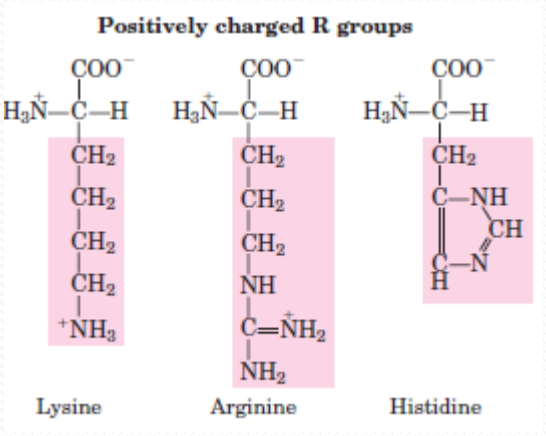
Disulfide Bond

- Cysteine is readily oxidized to form a covalently linked dimeric amino acid called **cystine**, in which two cysteine molecules or residues are joined by a disulfide bond.
- The disulfide-linked residues are strongly hydrophobic (nonpolar) and play a special role in the structures of many proteins by forming covalent links between parts of a protein molecule or between two different polypeptide chains.



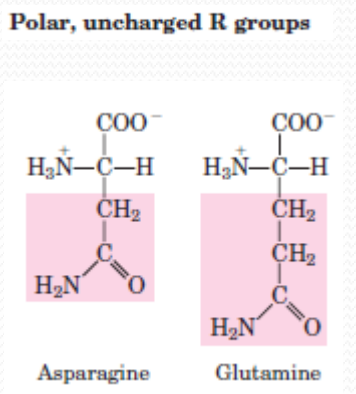
Aminoacids with Positively Charged R Groups

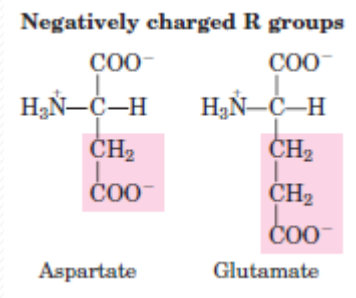
- The most hydrophilic R groups are those that are either positively or negatively charged.
- The amino acids in which the R groups have significant positive charge at pH 7.0 are **lysine,** which has a second primary amino group at the position on its aliphatic chain; **arginine,** which has a positively charged guanidino group; and **histidine,** which has an imidazole group.
- Histidine is the only common amino acid having an ionizable side chain with a p*K*a near neutrality (pKa=6.0).
- Below ph=6.0 N in the imidazole ring of His is protonated nad His has a net positive charge. At neutral pH His is neutral.
- In many enzyme-catalyzed reactions, a His residue facilitates the reaction by serving as a proton donor/acceptor.



Aminoacids with Negatively Charged R Groups

- Asparagine and glutamine (aminoacids with polar uncharged R groups) can be hydrolyzed to aspartate and glutamate by deamination reaction and become negatively charged aspartic acid and glutamic acid.
- Glutamic acid serves as a neurotransmitter in brain.





AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	negatively charged
Glutamic acid	Glu	E	negatively charged
Arginine	Arg	R	positively charged
Lysine	Lys	K	positively charged
Histidine	His	H	positively charged
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar

POLAR AMINO ACIDS

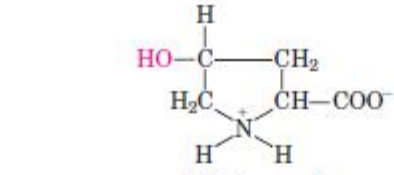
AMINO ACID			SIDE CHAIN
Alanine	Ala	A	nonpolar
Glycine	Gly	G	nonpolar
Valine	Val	V	nonpolar
Leucine	Leu	L	nonpolar
Isoleucine	Ile	I	nonpolar
Proline	Pro	P	nonpolar
Phenylalanine	Phe	F	nonpolar
Methionine	Met	M	nonpolar
Tryptophan	Trp	W	nonpolar
Cysteine	Cys	C	nonpolar

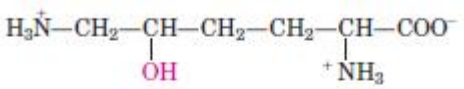
NONPOLAR AMINO ACIDS

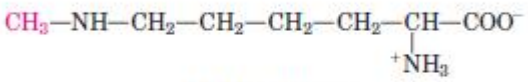
Uncommon Aminoacids

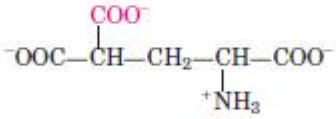
- In addition to the 20 common amino acids, proteins may contain residues created by modification of common residues already incorporated into a polypeptide.
- They can not be incorporated into the polypeptide chain residue readily since protein synthesis machinery does not recognize them. (There is no codon to code for them and there is no tRNA specific to those aminoacids.)
- Instead they are formed by the modification of common aminoacids in the polypeptide chain after protein synthesis (by post-translational modification).

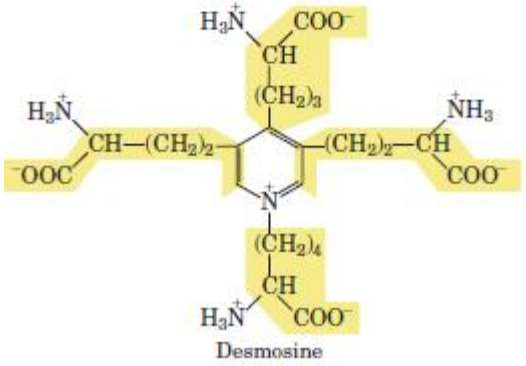
- **4-hydroxyproline:** found in plant cell wall proteins and collagen
- **5-hydroxylysine:** found in collagen
- **6-*N*methyllysine** is a constituent of myosin, a contractile protein of muscle
- **γ-carboxyglutamate,** found in the bloodclotting protein prothrombin and in certain other proteins that bind Ca^{2} as part of their biological function.
- **Desmosine,** a derivative of four Lys residues, which is found in the fibrous protein elastin.
- **Selenocysteine** is a special case. This rare amino acid residue is **introduced during protein synthesis rather than created through a post-translational modification**. It contains selenium rather than the sulfur of cysteine. Actually derived from serine, selenocysteine is a constituent of just a few known proteins.
- **Ornithine** and **citrulline** are the key intermediates in the biosynthesis of arginine in the urea cycle.

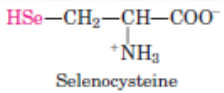

4-Hydroxyproline

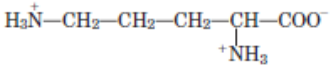

5-Hydroxylysine

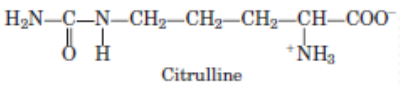

6-*N*-Methyllysine


γ-Carboxyglutamate


Desmosine

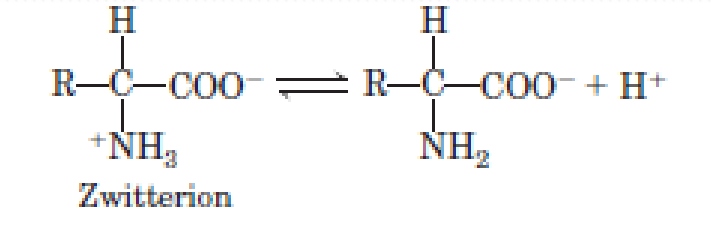

Selenocysteine

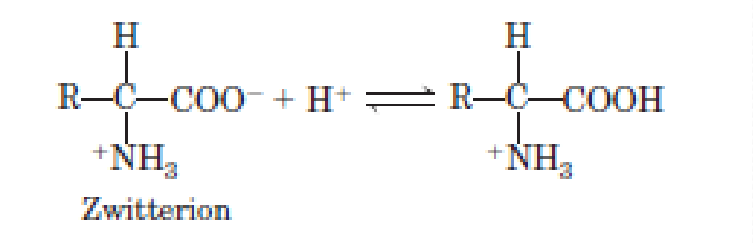

Ornithine


Citrulline

Aminoacids as Acid and Base

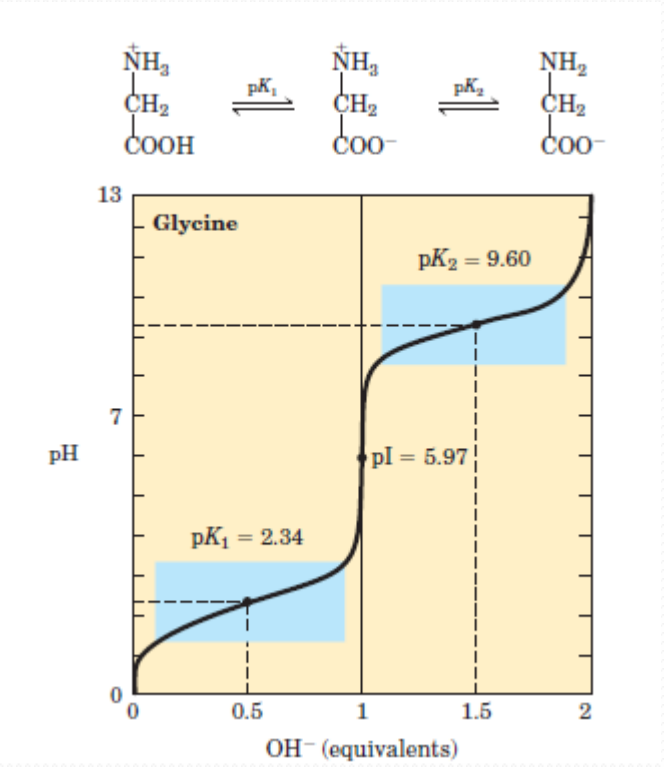
- When an amino acid is dissolved in water, it exists in solution as the dipolar ion, or **zwitterion** (German for "hybrid ion").
- A zwitterion can act as either an acid (proton donor) or a base (proton acceptor):





Titration curve of aminoacids

- From the titration curve of glycine we can derive several important pieces of information.
- First, it gives a quantitative measure of the p*K*a of each of the two ionizing groups: 2.34 for the -COOH group and 9.60 for the -NH_3^+ group.



- The second piece of information provided by the titration curve of glycine is that this amino acid has *two* regions of buffering power.
- One of these is the relatively flat portion of the curve, extending for approximately 1 pH unit on either side of the first p*K*a of 2.34, indicating that glycine is a good buffer near this pH.
- The other buffering zone is centered around pH 9.60.
- Note that glycine is not a good buffer at the pH of intracellular fluid or blood, about 7.4.
- **Which aminoacids are good buffers around physiological ph?**

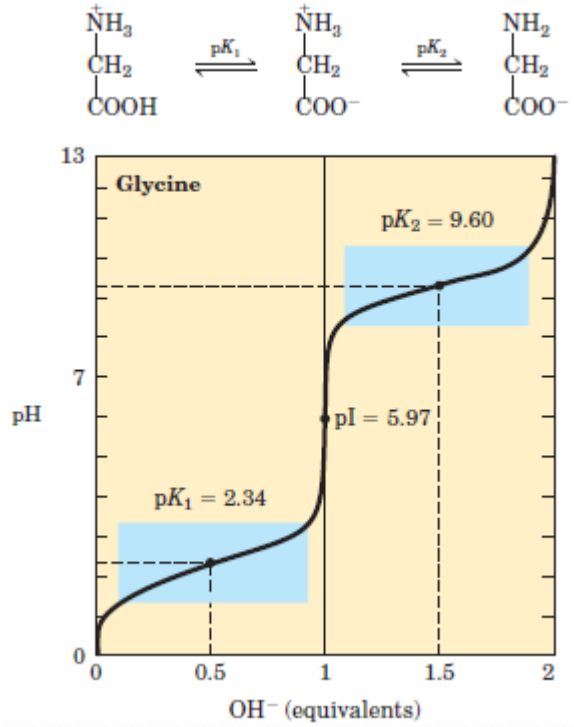
Aminoacid	pK_1(-COOH)	pK_2(-NH_3^+)	pK_R
Alanine	2,34	9,69	
Arginine	2,17	9,04	12,48
Asparagine	2,02	8,80	
Aspartic Acid	2,09	9,82	3,86
Cysteine	1,71	10,78	8,33
Glutamic Acid	2,19	9,67	4,25
Glutamine	2,17	9,13	
Glycine	2,34	9,60	
Histidine	1,82	9,17	6,00
Isoleucine	2,36	9,60	
Leucine	2,36	9,60	
Lysine	2,18	8,95	10,53
Methionine	2,28	9,21	
Phenylalanine	1,83	9,13	
Proline	1,99	10,60	
Serine	2,21	9,15	
Threonine	2,63	10,43	
Tryptophan	2,83	9,39	
Tyrosine	2,20	9,11	10,07
Valine	2,32	9,62	

Isoelectric Point

- The characteristic pH at which the net electric charge is zero is called the **isoelectric point** or **isoelectric pH**, designated **pI.**
- For glycine, which has no ionizable group in its side chain, the isoelectric point is simply the arithmetic mean of the two p*K*a values:

$$\mathrm{pI} = \frac{1}{2}(\mathrm{p}K_1 + \mathrm{p}K_2) = \frac{1}{2}(2.34 + 9.60) = 5.97$$

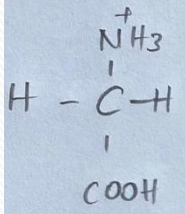
- Glycine has a net negative charge at any pH above its pI and will thus move toward the positive electrode (the anode) when placed in an electric field. At any pH below its pI, glycine has a net positive charge and will move toward the negative electrode (the cathode).
- The further the pH of a glycine solution is from its isoelectric point, the greater the net electric charge of the population of glycine molecules.



Glycine

pK_{COOH} : 2.34

pK_{NH_3} : 9.60



pH < 2.34

0 + 1 = +1

2.34 < pH < 9.60

−1 + 1 = 0

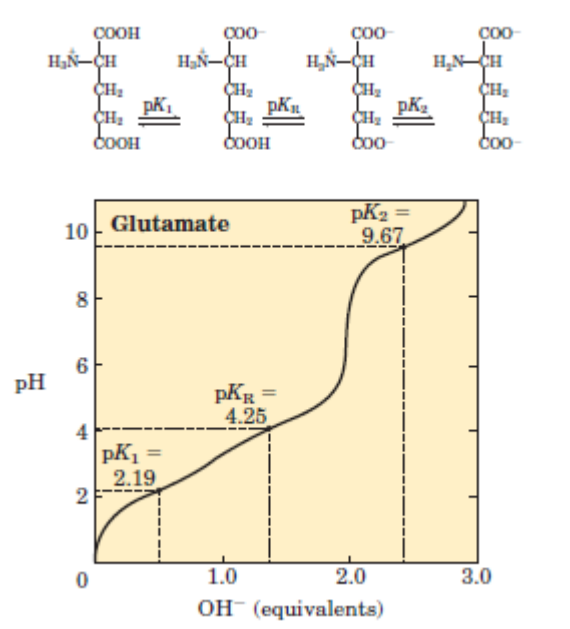
$pI = \frac{2.34 + 9.60}{2}$

pI = 5.97

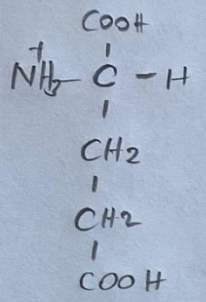
pH > 9.60

−1 + 0 = −1

Calculate pI values of glutamate



Glutamate :



$pK_{COOH_1} : 2.19$

$pK_{COOH_R} : 4.25$

$pK_{NH_3} : 9.67$

$pH < 2.19$

$0 + 0 + 1 = +1$

$2.19 < pH < 4.25$

$-1 + 0 + 1 = 0$

$pI = \frac{2.19 + 4.25}{2}$

$pI = 3.22$

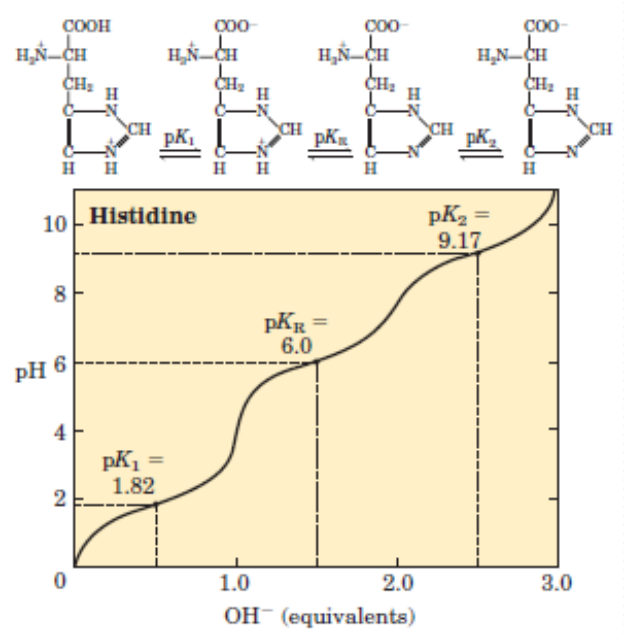
$4.25 < pH < 9.67$

$-1 - 1 + 1 = -1$

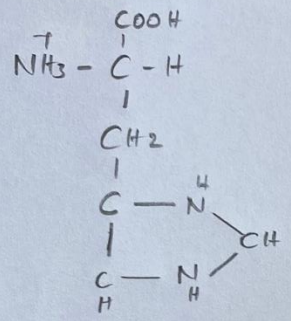
$pH > 9.67$

$-1 - 1 + 0 = -2$

Calculate the pI value of histidine



Histidine :



$pK_{COOH} : 1.82$

$pK_{COOH_R} : 6.0$

$pK_{NH_3^+} : 9.17$

$pH < 1.82$

$0 + 1 + 1 = +2$

$1.82 < pH < 6.0$

$-1 + 1 + 1 = +1$

$6.0 < pH < 9.17$

$-1 + 0 + 1 = 0$

$pI = \frac{6.0 + 9.1}{2}$

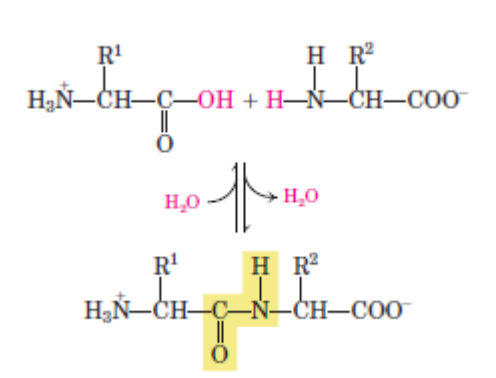
$pI = 7.58$

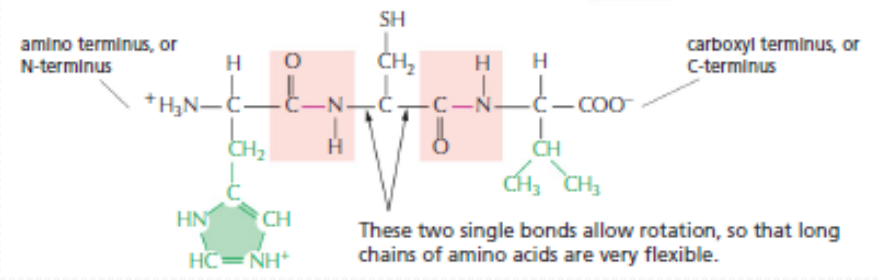
$pH > 9.17$

$-1 + 0 + 0 = -1$

Peptide Bond

- Two amino acid molecules can be covalently joined through a substituted amide linkage, termed a **peptide bond,** to yield a dipeptide.
- Such a linkage is formed between the -carboxyl group of one amino acid and the –amino group of another by forming water molecule (condensation reaction).

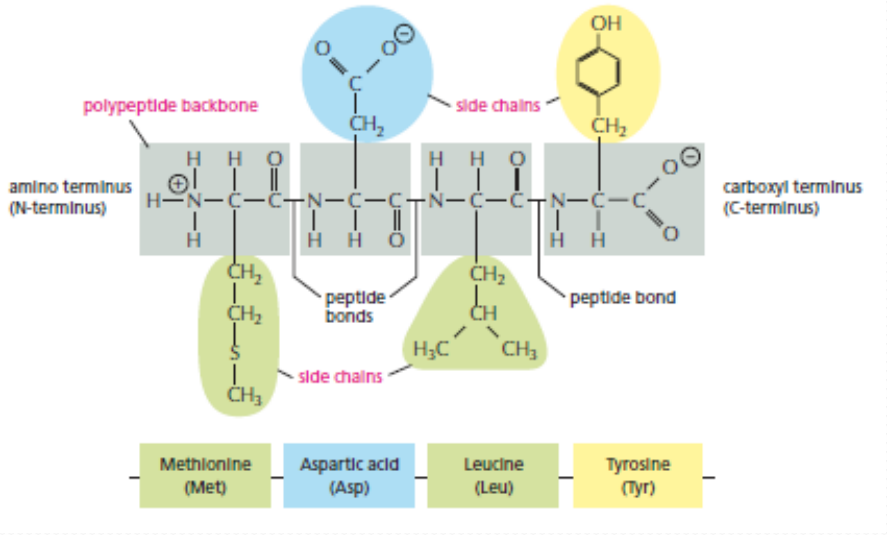




- Three amino acids can be joined by two peptide bonds to form a tripeptide; similarly, amino acids can be linked to form tetrapeptides, pentapeptides, and so forth.
- When a few amino acids are joined in this fashion, the structure is called an **oligopeptide.** When many amino acids are joined, the product is called a **polypeptide.**
- Proteins may have thousands of amino acid residues. Although the terms “protein” and “polypeptide” are sometimes used interchangeably, molecules referred to as polypeptides generally have molecular weights below 10,000, and those called proteins have higher molecular weights.

N- and C-terminals

- In a peptide, the amino acid residue at the end with a free -amino group is the **amino-terminal** (or *N*-terminal) residue; the residue at the other end, which has a free carboxyl group, is the **carboxyl-terminal** (*C*-terminal) residue.

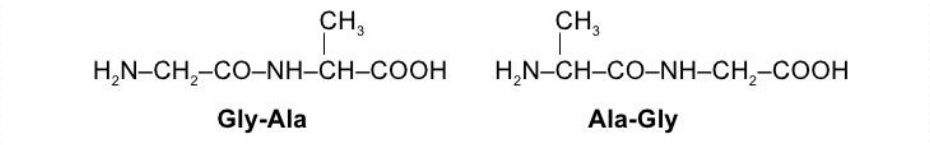


Peptide Nomenclature

N-terminal C-terminal

- Aminoacyl-aminoacid: Valyl-methionine
- Aminoacyl-aminoacyl-aminoacid: Valyl-methionyl-cysteine

N-terminal C-terminal

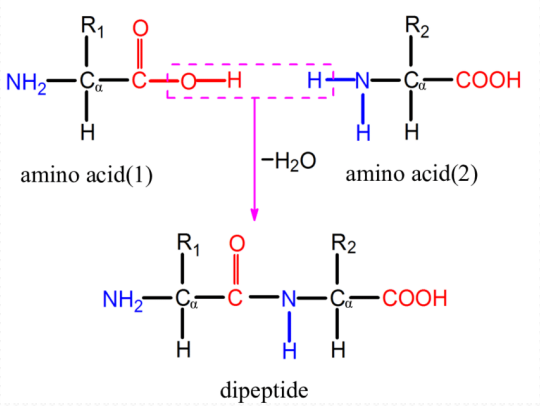


Glycyl-alanine Alanyl-glycine

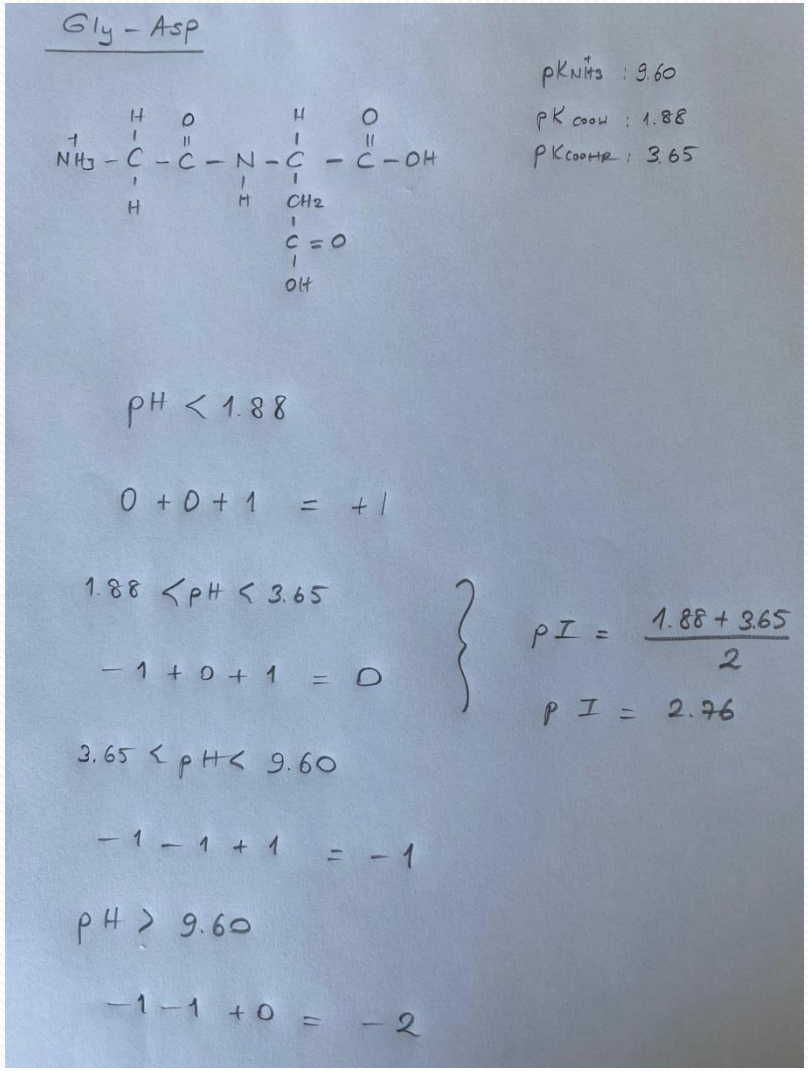
Not the same

Peptides Can Be Distinguished by Their Ionization Behavior

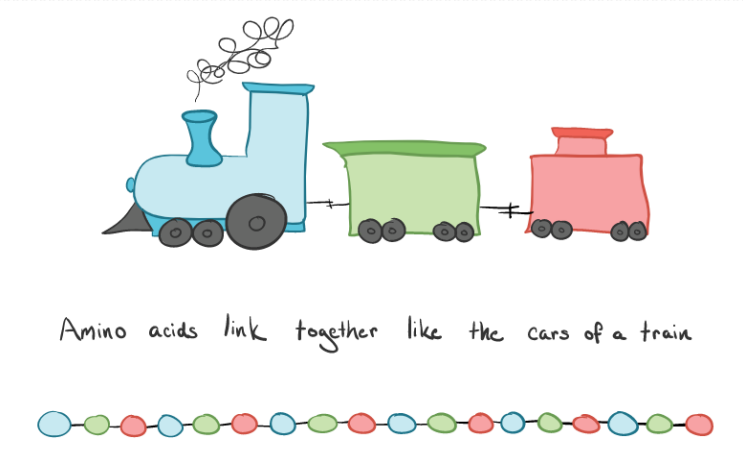
- Peptides contain only one free -amino group and one free -carboxyl group, at opposite ends of the chain. These groups ionize as they do in free amino acids, although the ionization constants are different because an oppositely charged group is no longer linked to the α-carbon.
- The α-amino and α-carboxyl groups of all nonterminal amino acids are covalently joined in the peptide bonds, which do not ionize and thus do not contribute to the total acid-base behavior of peptides.
- However, the R groups of some amino acids can ionize, and in a peptide these contribute to the overall acid-base properties of the molecule.
- Thus the acid-base behavior of a peptide can be predicted from its free α-amino and α-carboxyl groups as well as the nature and number of its ionizable R groups.
- Like free amino acids, peptides have characteristic titration curves and a characteristic isoelectric pH (pI).
- These properties are exploited in some of the techniques used to separate peptides and proteins.



Calculate the pI Gly - Asp

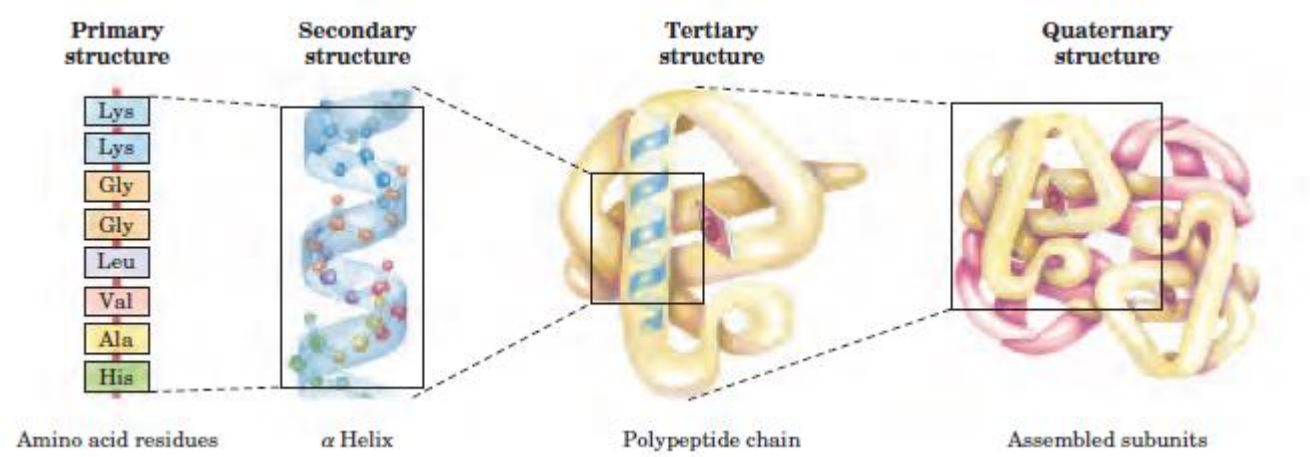


- Proteins are not the structures that simply bonded by peptide bonds together.
- Each type of protein has a particular three-dimensional structure, which is determined by the order of the amino acids in its polypeptide chain.



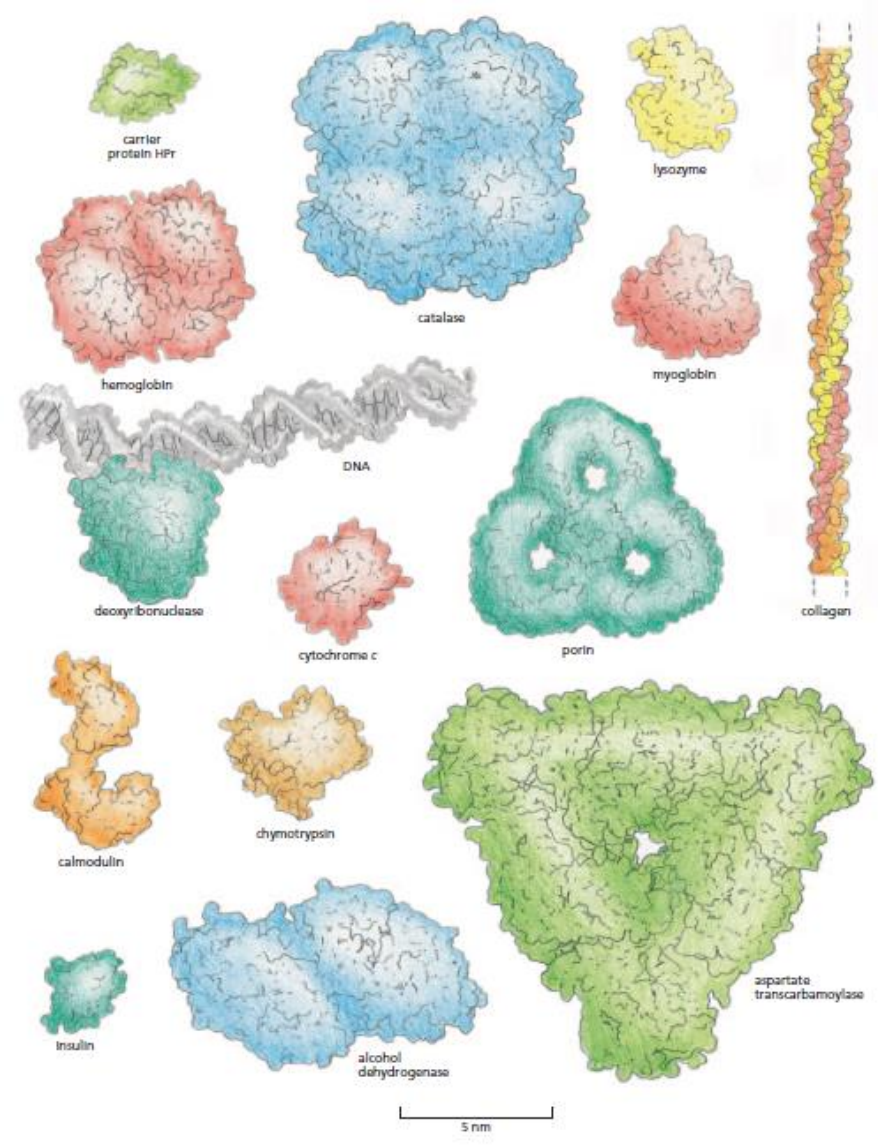
Levels of Structure in Proteins

- A description of all covalent bonds (mainly peptide bonds and disulfide bonds) linking amino acid residues in a polypeptide chain is its **primary structure.** The most important element of primary structure is the *sequence* of amino acid residues.
- **Secondary structure** refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns.
- **Tertiary structure** describes all aspects of the three-dimensional folding of a polypeptide.
- When a protein has two or more polypeptide subunits, their arrangement in space is referred to as **quaternary structure.**



Proteins Come in a Wide Variety of Shapes and Sizes

- Proteins can be globular or fibrous, and they can form filaments, sheets, rings, or spheres
- They range in size from about 30 amino acids to more than 10,000, the vast majority are between 50 and 2000 amino acids long.




water
reaction
structure
rate
enzyme
energy
living
molecules
substrate
properties
function
experiment
organisms
reactions
chemical
bond
information
functioning
unique
temperature
relate
affect
altering
activity
biological
catalytic
characteristics
effect
describe
explain
enzymes
identify
initial
logic
extract
found
active
carbohydrates
biomolecules
anabolism
concerning
descriptions
determined
concept diagrams
connections
compound
capacity
alter
abilities
adhesion
allosteric
cells
sabunch
catalysts
concentration
cohesion
covalent
coenzyme
control
deductive
denature
define
design
earth
different
demonstrate
dehydration
buffer
factors
enzymatic
free
group
heated
formation
hydrolysis
increasing
hydrophilic
hydrophobic
hydrogen
lab
levels
induced-fit
inhibition
influence
intervals
isolated
inferred
investigate
mixed
molecule
nucleic
oligosaccharide
ionic
organic
polypeptide
phospholipid
possible
progress
relationship
nature
place
model
new
nomenclature
interaction
proteins
select
speeds
synthesis
work
tissue
system
triglyceride
substance
sterol
repeating
responsible
related
protein
product
particular
physical/chemical
parts/monomers