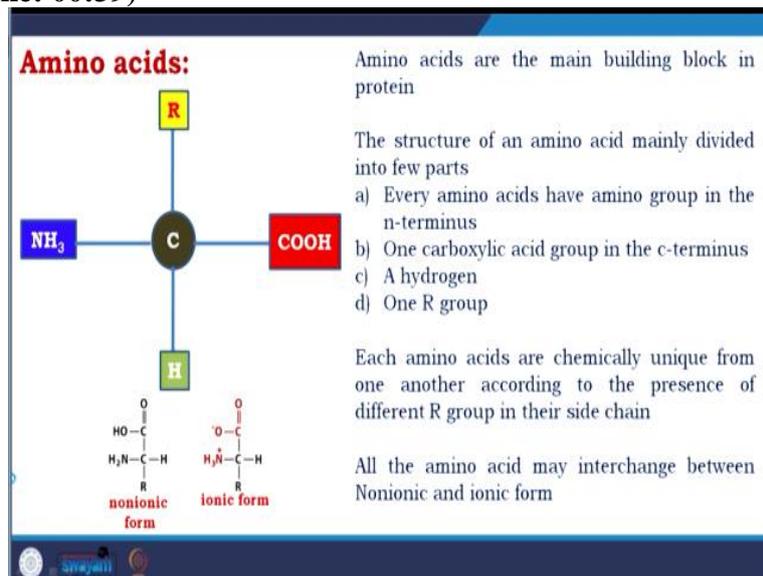


Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology - Roorkee

Lecture - 06
Protein: Amino Acids and Their Properties

Hello everyone, welcome again to the new class of structural biology. Today, we are starting a new module on protein and the first class we will discuss amino acids and their properties.

(Refer Slide Time: 00:39)

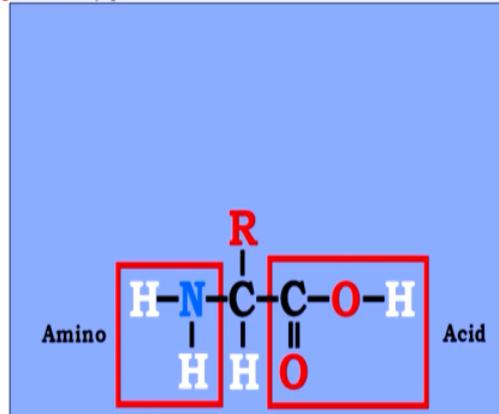


So, amino acids, as the name suggests, have one amino group, an acid group, hydrogen, and an R group. They are the main building block in the protein. As we have discussed earlier, the structure of amino acid, as I told now, is divided into a few main parts one, the N terminus, which is the amino group, the carboxylic acid group in the C terminus, a hydrogen and an R group each amino acids are chemically unique from one another and that happened because each of them has a different R group rest of the structure is same, all the amino acids may interchange between non-ionic and ionic form as you have seen in the picture.

(Refer Slide Time: 01:40)

An amino acid is a relatively small molecule with characteristic groups of atoms that determine its chemical behaviour.

The **R group** is the only part that differs between the 20 amino acids



So, there is an amino group; there is a carboxylic group. Then, the R you could have, as I showed earlier, hydrogen, which is glycine, CH_3 group, which is alanine, 2CH_3 group, which is valine, then a sulfur CH_2SH which is cysteine, a CH_2 aromatic ring which is phenylalanine, and they are called R as you guys know mathematics we call an unknown entity as x.

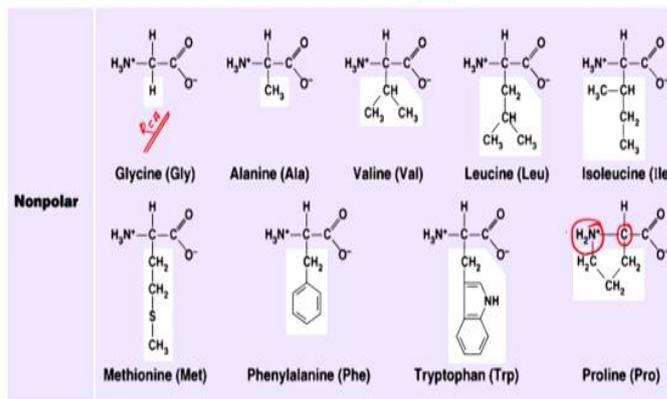
Similarly, in amino acids, the unknown part of the amino acid which designates a unique amino acid is considered or denoted as R. So, an amino acid is a relatively small molecule with a characteristic group of atoms that determine its chemical behavior. The R group is the only part that differs between the 20 amino acids.

(Refer Slide Time: 02:46)

Nonpolar amino acids:

nonpolar & hydrophobic

*20th
19th
Glycine*

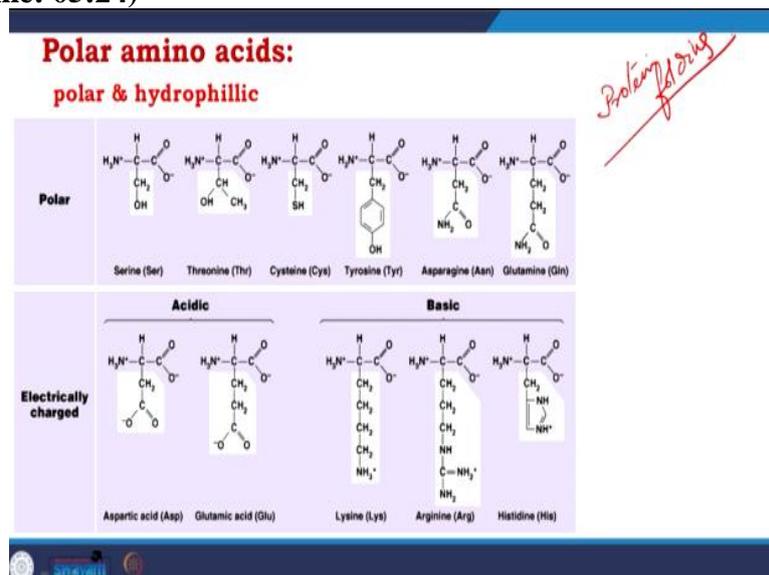


Let us look at what type of groups they are divided into. The first group is non-polar amino acids. In this group, you will get glycine, alanine, valine, leucine, isoleucine, methionine, phenylalanine, tryptophan, and proline. Among them, proline settles between the non-polar

and the polar groups. We will discuss that later. If glycine has two hydrogens, its R is hydrogen because it has two hydrogens. It is not chiral.

If you remember, we talked about chirality in our last module. We will discuss that in more detail. Alanine has a CH_3 . So it differs from glycine, and it is chiral so, among the 20 amino acids, 19 are chiral. Only glycine is achiral, valine has $\text{CH}(\text{CH}_3)_2$, leucine has $\text{CH}_2\text{CH}(\text{CH}_3)_2$, isoleucine have the same structure as leucine, but the CH_3 group is shifted CHCH_3 CH_2CH_3 , methionine has a sulfur group in it, but despite having that sulfur group, it is not polar. Because of the presence of a methyl group after the sulfur, phenylalanine has an aromatic ring it is CH_2Ph . Tryptophan also has an aromatic ring. There is one six-membered ring followed by a five-membered ring. Proline is very interesting if you look at its alpha carbon and the amino group fixed in a ring structure. We are going next, but I should tell here amino acid chemistry, and amino acid structures are extremely important for understanding the course to understand biological macromolecules, especially protein knowledge of those is very critical.

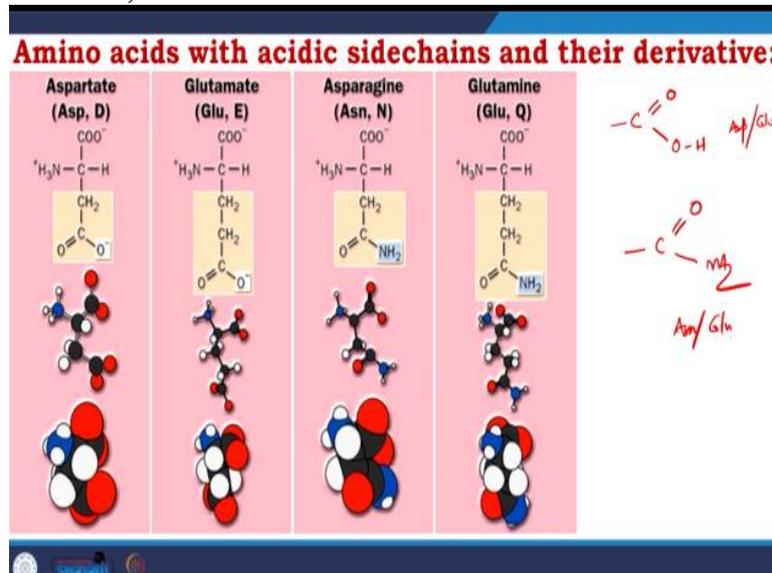
(Refer Slide Time: 05:24)



Polar and hydrophilic are serine, threonine, cysteine, tyrosine, asparagine, and glutamine. They are considered polar amino acids, whereas aspartic acid and glutamic acid are acidic whereas, lysine, arginine, histidine are basic amino acids. So, those polar have groups that could be ionized, but the ones that are acidic and basic or develop real charges. In our later section of this module, we will see how those polarities or real charges have helped those amino acids being in the protein being non-covalent interactions, non-covalent interactions are an extremely critical part of protein folding. From the beginning of this course, I told it is a journey from the protein sequence to protein function in between protein structures, the

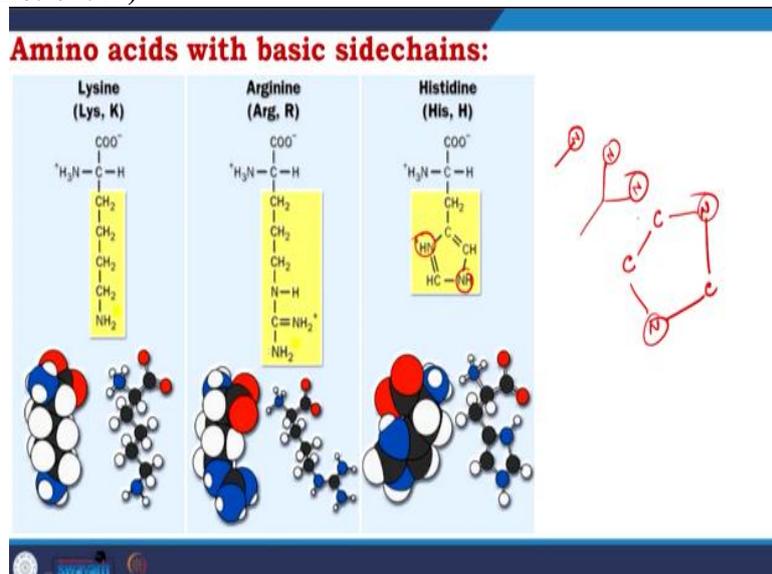
journey is about understanding these non-covalent bonds, which are actually forming and breaking continuously and giving unique protein architecture, the goal of this course and the goal of biochemists, biophysicist, structural biologists to understand or to register those non-covalent interactions.

(Refer Slide Time: 07:16)



Amino acids with acidic side chains and their derivatives aspartate and glutamate are acidic amino acids, whereas asparagine and glutamine have CONH whereas, acidic means COOH. So, this is aspartate glutamate, and this is asparagine, glutamine.

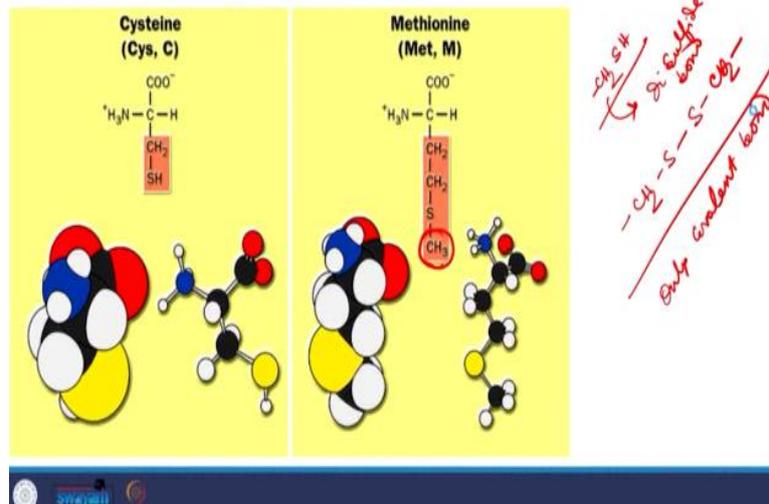
(Refer Slide Time: 07:47)



Again, an amino acid with basic side chains is lysine. Lysine creates one nitrogen, arginine has two nitrogen, histidine has its nitrogen inside the ring

(Refer Slide Time: 08:25)

Sulfur containing amino acids:



Sulfur-containing amino acids there are two sulfur-containing amino acids, cysteine, and methionine. Whereas cysteine, which is CH_2SH , can form a disulfide bond, methionine cannot form a disulfide bond because of the poor leaving group in the methyl group. So, when there is an oxidizing condition, two cysteines from the protein come together and form a disulfide bond. This is the only covalent bond that could form in physiological conditions. All other making and breaking come into the non-covalent bond category. This is the only covalent bond, and as I explained, methionine cannot form disulfide bonds.

(Refer Slide Time: 09:38)

Looking back to Chemistry:

Periodical Table

Life prefers lighter atoms

- (1) More abundant on Earth
- (2) Stronger bonding between small atoms

Chemical properties

The atomic composition of living organisms is more complex than others

Atomic sizes

1H 6C 7N 8O

Electrons on the outer shell

$1s$ $2s$ $2p$

sp^3

The tetrahedral structure of carbon orbital has rigid steric strain which makes the basic building unit of protein conformation

Components of the early atmosphere: H_2 CH_4 NH_3 H_2O

Handwritten notes:

- More/less
- H/C/N/P/S
- sp^3 hybridization
- rigid steric strain
- basic building unit of protein conformation

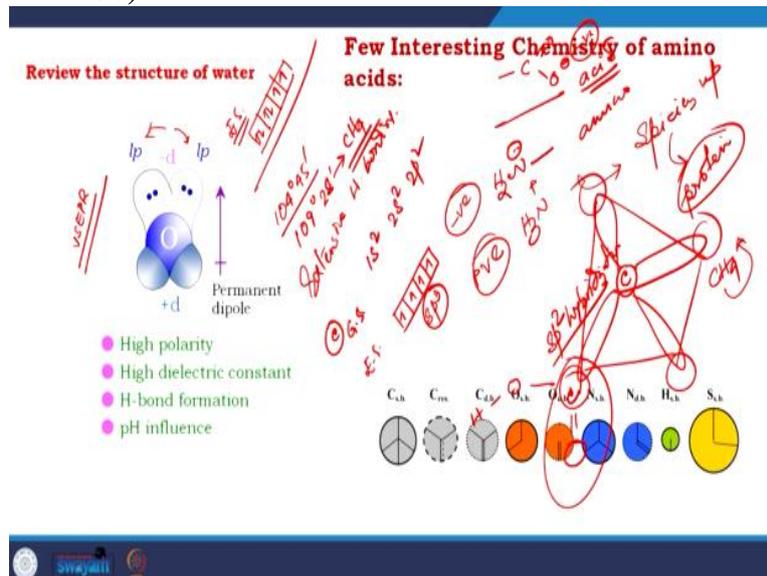
Jiang RH (2004) BCbasics

So, now we know about these 20 amino acids, but our goal is to go for structure, as we talked about. This knowledge is not good enough. We will look back to the chemistry. I have already discussed the position of those hydrogen, carbon, nitrogen, oxygen in the periodic table. We talked about that the atomic composition of a living organism is more complex than others, but, very interestingly, the biological macromolecules are formed with just a small

plastic is so popular because its inertness helps us develop materials. And now today, when we have already developed a huge amount of waste out of plastic and the inertness which was advantage ones, now coming as a disadvantage, because we are not being able to destroy them, because the inertness. So, you understand when there is a carbon polymer carbon backbone, they have the inertness now, in a functional molecule like protein, you cannot make it inert. So, as in my language, you have to spice up the inertness of the polymer. How could you do that? You could do that by introducing the positive and negative charges. So, let us look at the electronegativity table. You will see that oxygen and nitrogen are way more electronegative, oxygen being the highest 3.5 and nitrogen is 3.

Very interestingly, oxygen and nitrogen both could contribute as negative charge providers. You get lithium, beryllium, sodium, calcium as a positive charge. So, we could not use metals as positive charge contributors in biology. We have NH_3 that has $1s^2, 2s^2, 2p^3$. So, when it forms sp^3 hybridization, it forms bonding with hydrogen. In addition, it could form a coordinate bond. It could take another hydrogen, take the electron, share between two, and form NH_4^+ .

(Refer Slide Time: 19:38)



If you remember, we have talked about water's unique properties and how it is used as a universal solvent? Because if you remember, it had two lone pairs, so in the excited state, because of the lone pair-lone pair repulsion, which is called this VSEPR effect (valence electron pair repulsion theory), the angle between the two hydrogens forming a covalent bond with oxygen become 104 degrees 45 minutes instead of 109 degrees 28 minutes which is in within the classic data structure.

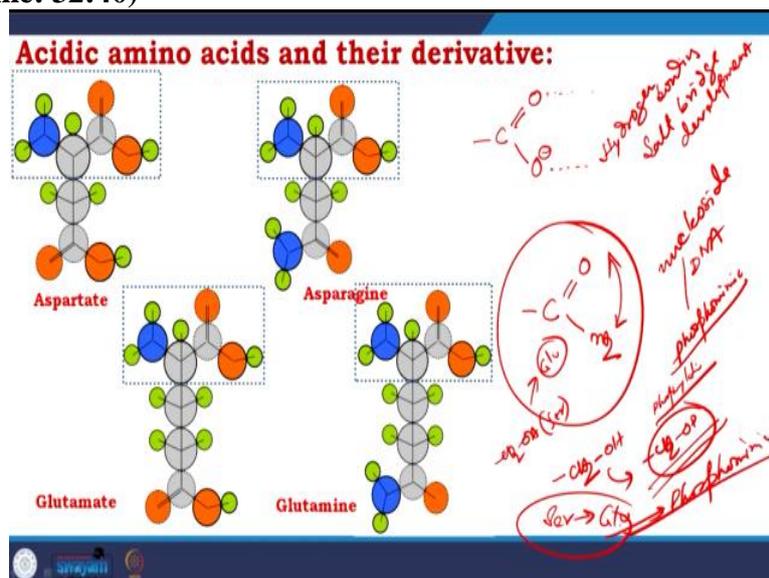
Because the angle is 104 degrees 45 minutes, there is extensive hydrogen bonding. In addition, we get high polarity, high dielectric constant hydrogen bond formation, which I told and pH influence that makes water a universal solvent. But now, as I told you, we have oxygen present in the acid part of the amino acid, and nitrogen in the amino part of the amino acid has the lone pair. It can be a negative charge provider as a lone pair contributor, but then it forms NH_3 plus, which is a positive charge provider. So, because nitrogen and oxygen provide positive and negative charges, they are spicing up the amino acid. Hence, they are spicing up the protein, the functional module of protein. The way it would be performing the function is taken care of by these groups being into different amino acids now, if you see, when this chemistry works, as I talked about sp^3 hybridization.

In the ground state, carbon has $1s_2, 2s_2, 2p_2$ in the excited state. Each has four unpaired electrons and forms four sp^3 orbitals that give a perfect tetrahedron forming methane. Similarly, if you modify from methane, you will see that it forms a bond with another carbon that forms a bond with oxygen and a double bond with oxygen. Here, there is sp^2 hybridization.

(Refer Slide Time: 25:45)

turn glycine because of its small size makes a flexible turn coming to serine, threonine, and tyrosine all of them have hydroxyl group, because of the presence of hydroxyl group, they could make reactions most popular of them is phosphorylation. So, we talked about post-translational modification, phosphorylation mainly happened to serine, threonine, and tyrosine, cysteine sulfur group is not crowded, and it could form disulfide bond and in reducing condition is reduced it makes SH-SH. So, because of the presence of the disulfide bond, there are many motifs developed. We will discuss those in a later section.

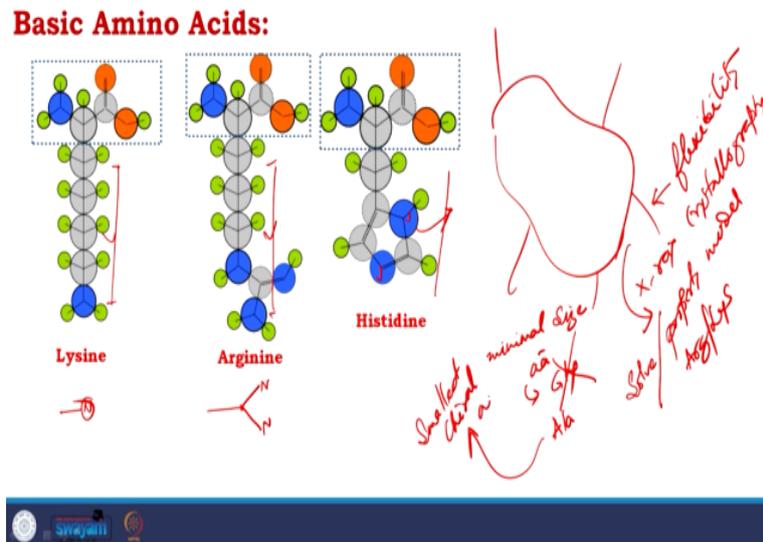
(Refer Slide Time: 32:40)



We have already talked about aspartate and glutamate. Aspartate has its derivatives asparagine, and glutamate has its derivative glutamine. When you consider acidic groups, you see that they could be too strong for hydrogen bonding or salt bridge development. So, both aspartate and glutamate have the potential to do that. But when it comes to asparagine or glutamine, they could use as both donor and acceptor, and that is why when there is a nucleoside or DNA base, you will see that there is asparagine or glutamine in the active site that is developing interaction with the adenosine, guanosine, thymine, cytosine. So, that is one of the very interesting properties they have. They could switch their residues by rotating their rotamers, and in that way, they could fit with many ligands, providing flexibility towards ligand binding. You could also see how the group looks when you take the common binders. Suppose you could mutate the serine to glutamate. In that case, you will get similar interaction if it is phosphorylated, so, if phosphorylated serine and a serine mutated glutamate behave similarly, used to use mutation of serine to glutamate as a strategy to phosphate mimic, that is why glutamate is called a good phospho mimic to phosphorylated serine, very

interestingly, where glutamate used to do that aspartate never did that. So, the size and charge both are important here.

(Refer Slide Time: 36:49)

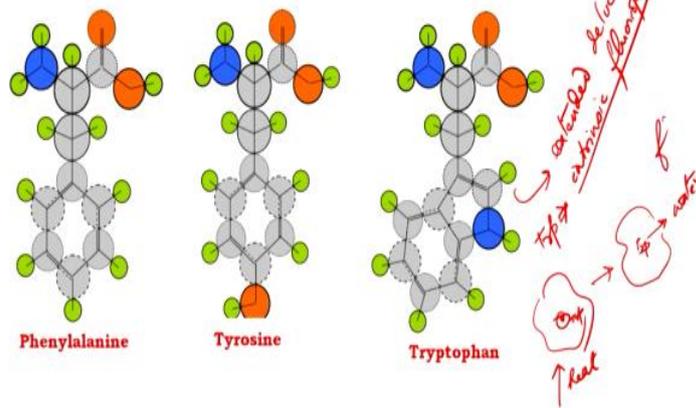


We talked about the basic amino acids: lysine, arginine, and histidine. Interestingly, histidine, as I told is in the ring, so the nitrogen which one will bind is very interesting to find. As I told because of the big length, when they are specifically at the protein's surface, they generally show extreme flexibility. When we perform X-ray crystallography most of the time, we cannot solve properly model arginine or lysine. We model it with the minimal size amino acid, which is glycine. We do it with alanine because alanine is the smallest chiral amino acid. So we want to model at least the chirality so, when we cannot model lysine or arginine, we model it as alanine.

Sometimes you see the structure, and you find alanine in the structure. And in the sequence, you find arginine or lysine. It is because there is a lot of disturbance. And that is why the crystallographer could not be able to model the arginine or lysine properly.

(Refer Slide Time: 39:51)

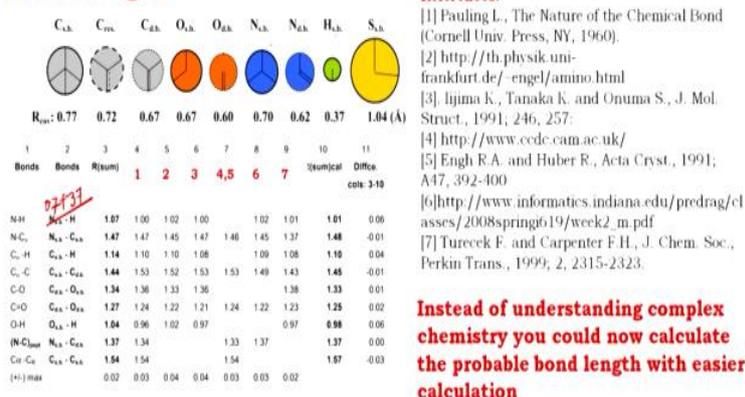
Aromatic Amino Acids:



We will discuss aromatic amino acids later in detail, but they all have an aromatic ring, whereas tryptophan having two rings has extended delocalization. And that is the reason tryptophan is used as an intrinsic fluorophore. So, when you have a pure protein, without labeling it, you could perform a fluorescence spectroscopy study of that protein. If you have tryptophan in the protein, that is an advantage. You cannot use phenylalanine or tyrosine as a fluorescent marker. Whereas you could use tryptophan as a marker, how could you use it? Let us say you have the protein and the tryptophan is inside. So, what could you do if you want to test how that protein is stable? You have the protein, tryptophan in the core, provide heat, the protein would open up, and tryptophan would be getting access to the water. So its fluorescence could be quenched. So, you could even measure the rate of folding if you are lucky enough to have one tryptophan inside the hydrophobic core of the protein. You could set up an assay of your ligand binding if your tryptophan is close to the active site or is affected by the ligand binding.

(Refer Slide Time: 42:44)

Understanding an alternative process of calculating bond length:



But one thing you know, if you have to calculate the bond lengths with the changing chemistry, then it would be not easy. Here is a method I will talk about now, which will give you a very easy calculation of the bond lengths. I talked about carbon single bond, carbon resonated, carbon double bond, single oxygen bond, double bond, single nitrogen bond, double bond, single hydrogen bond, and sulfur single bond. Here is the process, you see all of them have their covalent radius, single carbon bond has 0.77 angstroms, carbon resonated has 0.72, carbon double bond at 0.67, reduced because of the development of the pi bond, single oxygen bond is 0.67. Oxygen being more electronegative than carbon holding the electrons closer double bond is even lesser than single bonds 0.6, single nitrogen bonds 0.7, nitrogen double bond 0.62, hydrogen single bonds 0.37, sulfur single bond 1.04.

You could calculate the bonds like NH, N_{alpha}, C_{alpha} hydrogen, C_{alpha} carbon, C_{alpha} carbon oxygen single bond, carbon-oxygen double bond, oxygen-hydrogen NC, which is in the peptide bond CONH C_{alpha} C_R side chain all of them you could calculate. So, NH means N single bond and hydrogen single bond. So, it takes N single bond, which is 0.7 plus hydrogen, which is 0.37, which means 1.07. So, you understand similarly single nitrogen bond is 0.7 single carbon bond is 0.77 so, 1.47. So we fixed the curve covalent radius of the single bond and double bonded elements involved in protein and then added those to get the bond.

So, here you see the red mark 1, 2, 3, 4, 5, 6, and 7. These are references starting from Pauling's nature paper, where he posts published the nature of the chemical bond. We have taken relevant references that people have calculated, and you see that the difference in the validity ranges from 0.06 to 0.00. So, this is very accurate, and you could use it.

(Refer Slide Time: 47:02)

Physico-chemical properties of different Amino acids:

M.W.
A.A. → 110/20

Amino acids	Chemical properties	Molecular mass (Da)	Number of atoms	Volume (Å ³)	Hydrophathy index
Leucine	Nonpolar, aliphatic	131	22	166.7	3.8
Lysine	Positively charged basic	146	24	168.6	-3.9
Methionine	Nonpolar, aliphatic	149	20	162.9	1.9
Phenylalanine	Aromatic	165	23	189.9	2.8
Proline	Nonpolar, aliphatic	115	17	112.7	-1.6
Serine	Polar, uncharged	105	14	89.0	-0.8
Threonine	Polar, uncharged	119	17	116.1	-0.7
Tryptophan	Aromatic	204	27	227.8	-0.9
Tyrosine	Aromatic	181	24	193.6	-1.3
Valine	Nonpolar, aliphatic	117	19	140.0	4.2

● Amino acid with highest molecular mass and volume

Coming to Physico-chemical properties of different amino acids, you will see the chemical properties we have already talked about leucine, non-polar aliphatic, lysine positively charged basic, methionine non-polar aliphatic, phenylalanine aromatic, proline non-polar aliphatic, serine polar uncharged, threonine polar uncharged, tryptophan aromatic, tyrosine, aromatic and valine non-polar aliphatic.

(Refer Slide Time: 49:51)

Physico-chemical properties of different Amino acids:

Handwritten calculation for Glycine:
 $12 \times 2 + 16 \times 1 + 14 = 70 + 5 = 75$
 $12 \times 1 + 16 \times 1 + 14 = 42$
 $42 - 16 = 26$
 $26 - 14 = 12$
 $12 - 1 = 11$
 $11 - 1 = 10$

Amino acids	Chemical properties	Molecular mass (Da)	Number of atoms	Volume (Å ³)	Hydrophathy index
Alanine	Nonpolar, aliphatic	89	13	88.6	1.8
Arginine	Positively charged basic	174	26	173.4	-4.5
Asparagine	Polar, uncharged	132	17	114.1	-3.5
Aspartic acid	Negatively charged acidic	133	16	111.1	-3.5
Cysteine	Polar, uncharged	121	14	108.5	2.5
Glutamine	Polar, uncharged	146	20	143.8	-3.5
Glutamic Acid	Negatively charged acidic	147	19	138.4	-3.5
Glycine	Nonpolar, aliphatic	75	10	60.1	-0.4
Histidine	Positively charged	155	20	153.2	-3.2
Isoleucine	Nonpolar, aliphatic	131	22	166.7	4.5

● Amino acid with lowest molecular mass and volume

Similarly, alanine is a non-polar aliphatic, arginine positively charged basic, asparagine polar uncharged, aspartic acid negatively charged, acidic cysteine polar uncharged, glutamine polar uncharged, glutamic acid negatively charged, acidic, glycine non-polar aliphatic, histidine positively charged, isoleucine non-polar aliphatic.

(Refer Slide Time: 53:15)

Chirality in Amino acid:

Amino acids (except for **glycine**) have a **chiral** carbon atom adjacent to the carboxyl (COOH) and amine (NH₂) groups. This chiral center allows for stereoisomerism as there are four different chemical groups surrounding the carbon atom. The amino acids with two stereoisomers are mirror images of each other and are called L-amino acid and D-amino acid respectively. The structures are not superimposable on each other, much like the left and right hands.

Stereoisomers

Achiral carbon atom

Chiral carbon atom

L-amino acid D-amino acid

Glycine Glycine

Alanine Alanine

Remember three things I talked about: the three secrets, the 3 C's of covalent bond, chirality, configuration, and confirmation. So chirality in an amino acid is very important. Amino acids, except for glycine, have a chiral carbon atom adjacent to the carboxyl and amino groups. This chiral center allows stereoisomerism as four different chemical groups surround the carbon atom. It is tetrahedral. It has four groups with a 109 degree 28-minute angle with that.

(Refer Slide Time: 55:37)

Acid-Base Behavior of Amino Acids:

Amino acids exist as a zwitterion: a dipolar ion having both a formal positive and formal negative charge (overall charge neutral).

$$\begin{array}{ccc}
 \begin{array}{c} \text{R} \\ | \\ \text{H}_2\text{N}-\text{C}-\text{CO}_2\text{H} \\ | \\ \text{H} \\ \text{pK}_a \sim 5 \end{array} & \rightleftharpoons & \begin{array}{c} \text{R} \\ | \\ \text{H}_3\text{N}^+-\text{C}-\text{CO}_2^- \\ | \\ \text{H} \\ \text{pK}_a \sim 9 \end{array}
 \end{array}$$

Amino acids are *amphoteric*: they can react as either an acid or a base. Ammonium ion acts as an acid, the carboxylate as a base.

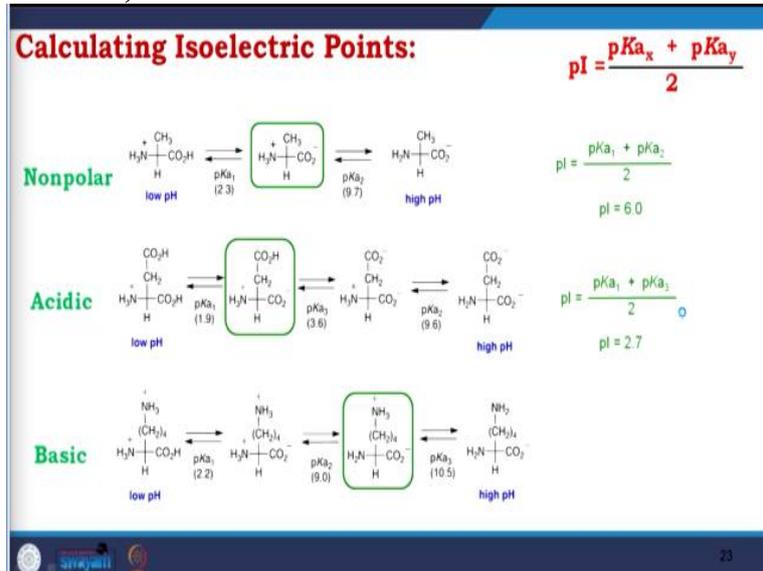
Isoelectric point (pI): The pH at which the amino acid exists largely in a neutral, zwitterionic form (influenced by the nature of the sidechain)

$$\begin{array}{ccccc}
 \begin{array}{c} \text{R} \\ | \\ \text{H}_3\text{N}^+-\text{C}-\text{CO}_2\text{H} \\ | \\ \text{H} \\ \text{low pH} \end{array} & \xrightleftharpoons[\text{pK}_{a1}]{} & \begin{array}{c} \text{R} \\ | \\ \text{H}_3\text{N}^+-\text{C}-\text{CO}_2^- \\ | \\ \text{H} \end{array} & \xrightleftharpoons[\text{pK}_{a2}]{} & \begin{array}{c} \text{R} \\ | \\ \text{H}_2\text{N}-\text{C}-\text{CO}_2^- \\ | \\ \text{H} \\ \text{high pH} \end{array}
 \end{array}$$

Amino acids exist as a zwitterion, a dipolar, and have both a formal positive and formal negative charge. Remember, I talked about amino acid means carbon, NH₂, CO₂H, R group, and hydrogen. Now, this could be COO⁻ and this could be NH₃⁺. When they are both plus and minus, it is charged. But interestingly, if you put it on an electrode, it is not going here or there. It is staying that stage is called the zwitterionic stage.

They are amphoteric. They can react as either an acid or a base ammonium ion act as an acid, carboxylate as a base. What is isoelectric point or pI? The pH at which the amino acid exists largely in a neutral zwitterionic form is influenced by the nature of the 20 side chains we have. So, when it is low pH high amount of proton makes NH_3^+ . Whereas when it is high pH, it is more COO^- minus.

(Refer Slide Time: 57:35)



So, you could calculate the isoelectric points so that non-polar you see that in low pH it is NH_3^+ then in high pH it is COO^- in between it is due to ionic form.

So, the pI is $pK_{a_x} + pK_{a_y} / 2$.

(Refer Slide Time: 59:02)

Summary:

- At low pH, proton concentration $[\text{H}^+]$ is high. Therefore, both amines and carboxylic acids are protonated. ($-\text{NH}_3^+$ & $-\text{COOH}$)
- At high pH, proton concentration is low. Therefore, both amines and carboxylic acids are deprotonated. ($-\text{NH}_2$ & $-\text{COO}^-$)
- At neutral pH, amines are protonated ($-\text{NH}_3^+$) and carboxylates are deprotonated ($-\text{COO}^-$)

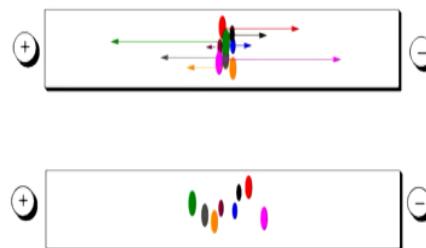
So summary, at low pH, proton concentration is high. Therefore both amines and carboxylic acids are protonated. At high pH, proton concentration is low they are both amines and

carboxylic acid are deprotonated. At neutral pH, amines are protonated and carboxylates are deprotonated, forming the zwitterion, which differs from the side chain.

(Refer Slide Time: 01:00:06)

Electrophoresis:

Separation of polar compounds based on their mobility through a solid support. The separation is based on charge (pI) or molecular mass



So, what I talked about, you could use this in separating the protein. The separation of a polar compound based on their mobility through solid support separation is based on charge or molecular mass.

(Refer Slide Time: 01:01:52)

Levels of Protein Structure:

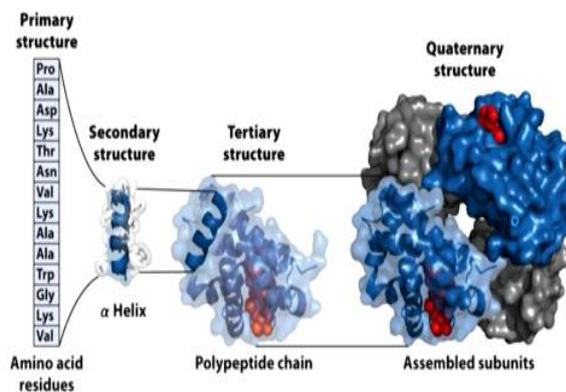


Figure 3-21
Lehninger Principles of Biochemistry, Seventh Edition
© 2017 W. H. Freeman and Company



So, we will talk about the levels of protein structures. We talked about these that our goal is to go to the quaternary structure, journey from the primary structure, which is the amino acid forming peptide bond to the secondary structure, then different secondary structure comes together to tertiary structure. And the tertiary structure is an independent chain. When multiple chains come together, they form assembled units or quaternary structures.

I will finish here today. In the next class, we will follow up with the properties of protein identification and techniques that will help us characterize protein, identify the protein, and sequence protein. Thank you very much.