Biochemistry - I

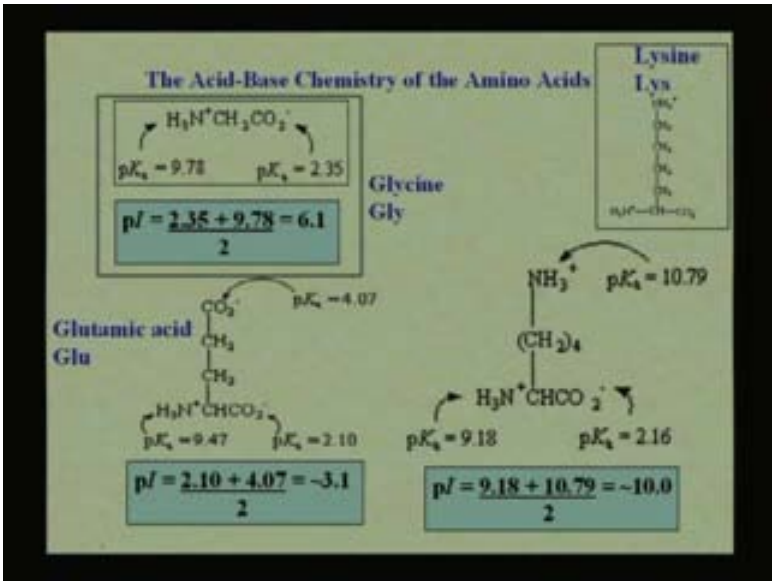
Prof. S. Dasgupta
Department of Chemistry
Indian Institute of Technology

Lecture –3
Protein Structure I

Welcome, we continue our discussion on the Acid Base chemistry of Amino acids which we were discussing in our last class. We understood that if we have structure such as Glycine we will have basically protonated NH_3^+ and we have a COO^-. Now it has zero charge because it has reached the point where it has lost carboxylic H^+ the proton belonging to COOH group and it is yet to loose a proton which attached to the amino group of the amino acid. When calculating the p*I* you have to go from +1 charge to a −1 charge through the zero. So for each structure that you are considering if I want to go from the +1 to the zero to the −1 these are the two steps that I am going to follow. So If I want to calculate the p*I* of Glycine I have to sum up the p^{K1} and the p^{K2} or p^{Ka} of the both acids then divide them by 2 to get the p*I* value corresponding to Glycine.

Now, if we go to the Glutamic acid initially if all of these are protonated so the charge is +1. When I loose one proton from the COOH group I am going to loose the acid proton belonging to the amino acid first but not the side chain. When I lose this proton then it has a plus charge here and a minus charge here and the side chain is still protonated.

(Refer Slide Time 3:00 min)

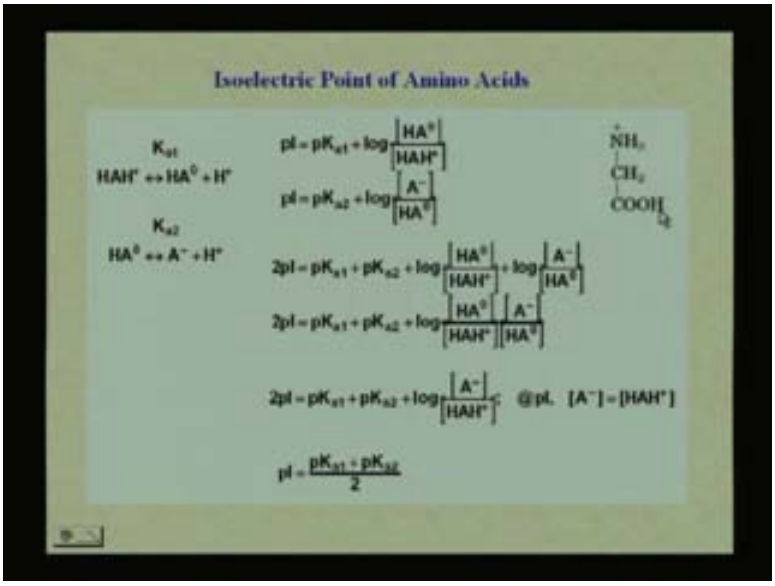


So now this is at the zwitterionic form. So the p*I* is going to be sum of 2.1 and 4.07 divided by 2. (Refer Slide Time 3:12 min)

In the next case, suppose we have a lysine. Then initially the charge is +2. When I loose the first proton here then this becomes COO^- but it still has a charge of +1. Now I have to loose the next proton. So the next proton that is going to be lost is the amino group proton of the amino acid. Then to calculate the p*I* in this case going to a +1 through a zero to a -1, this going to be (9.18 +10.79) / 2. So we just considering is in calculating the p*I*'s you have to go from a charge of +1 to -1 obviously through the zero. And the zero is the zwitterionic form of the amino acid. So basically if you want to calculate the isoelectric point of the amino acids, if we just consider the simplest one the Glycine then I have an NH_3^+ here, I have a COOH here. This carboxylic acid H^+ is going to be lost first because it can is easily removed. So when going from HAH^+ I am going to an $HA^0 + H^+$.

Here it is a +1 form and to go to zero form by losing this hydrogen first. After loosing this hydrogen based on the Henderson Hesselbach equation you can calculate the p*I* form for both K_{a1} and K_{a2}. When you are calculating it from p^{Ka2} you go to the A^- form because you are loosing this proton as well you have already lost this proton. Here the p*I* is actually going from the +1 to the -1 through the zero, the p*I* is the summation of the p^{Ka} values divided by 2.

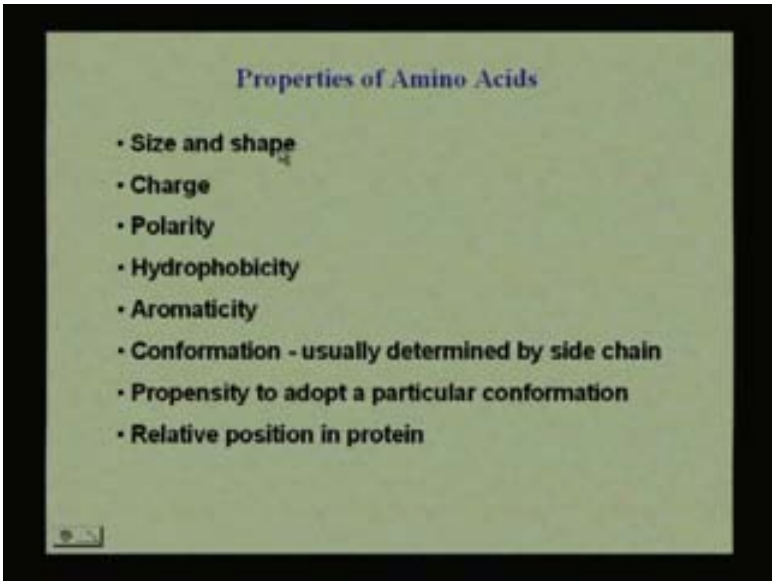
(Refer Slide Time 5:42 min)



We have to consider the +1 form to the -1 form going through the zero form for which we calculate the p*I* value of you specific amino acid.

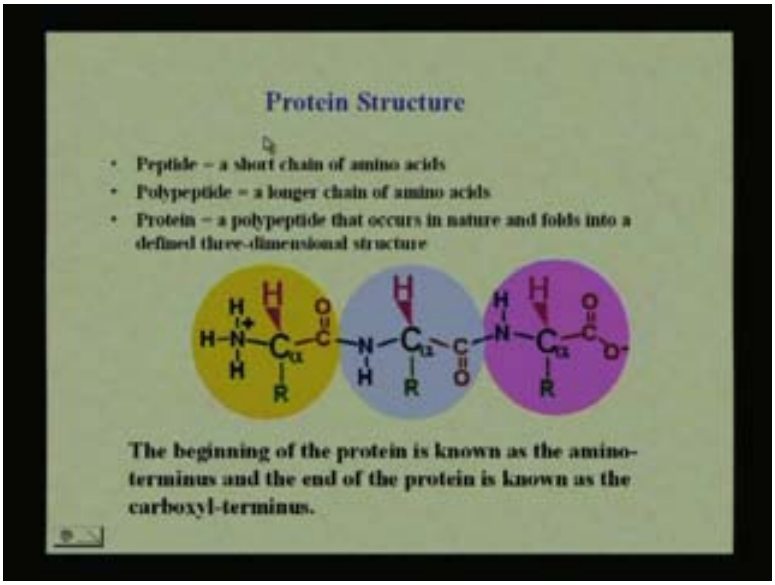
Now going back to the properties of amino acids where we have consider the Size and shape, the Charge, the Polarity, some of the Hydrophobicity, part of the Aromaticity and the Confirmation that is determined by the side chain due to the different R groups that are present and obviously the properties of these R groups.

(Refer Slide Time 6:18 min)



Next we are will consider protein structure in general how we have learnt, how these amino acids are linked together by peptide bonds. So, if we consider just the protein structure we have is a peptide bond.

(Refer Slide Time 6:44 min)



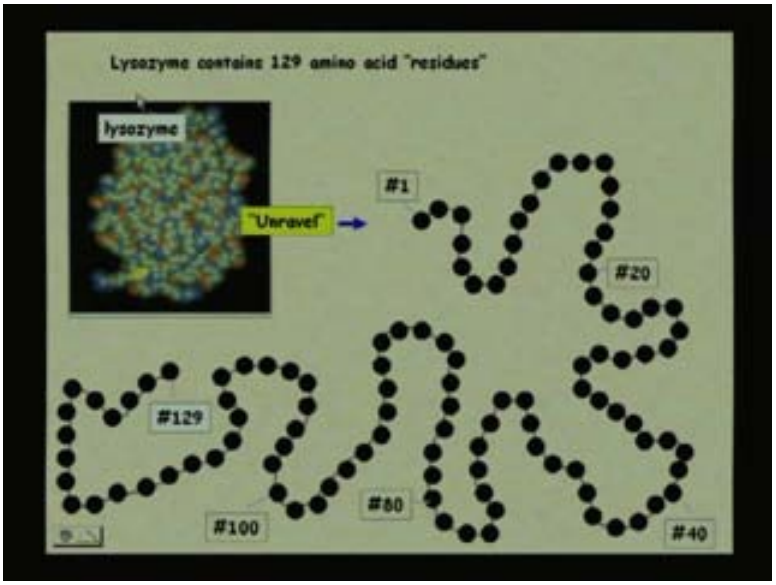
The difference between a peptide, a polypeptide and a Protein is a peptide is usually a short chain of amino acids, a polypeptide is a longer chain a set of amino acids. A protein is a polypeptide that occurs in nature and folds into a defined three dimensional structure. This means all Proteins are polypeptides but all polypeptides are not Proteins. The reason being is all polypeptides may not fold into a definite three dimensional structure which is

true for a Protein. So a protein is a polypeptide chain of set of amino acids that are linked together by peptide bonds and it has a definite three dimensional structure. So we have a peptide, we have a polypeptide, and we have a Protein. (Refer Slide Time 7:54 min)

Now if we look at these linkages you can say that at when we are at a $p^H = 7$ we have lost the carboxylic acid proton. So a protein is always represented in this form. This is a try peptide, the try peptide at physiological p^H which is $p^H = 7$ we have not lost the proton that belongs to the amino group but we have lost the proton that belongs to the carboxylic acid group because this is represented in this form. We always begin the protein with the amino terminus. So we have an amino terminus and we have a carboxylic acid terminus. We have the different R groups which are going to behave differently based on the different properties that it has.

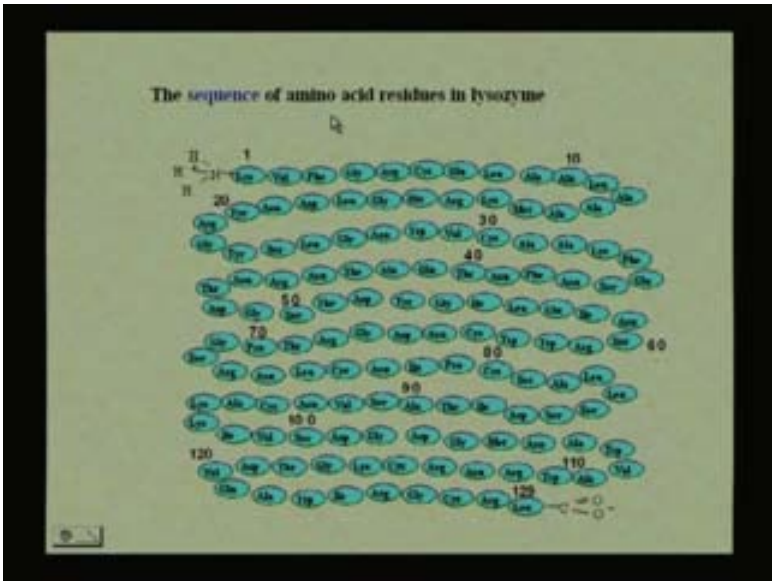
So, if we consider one of these proteins called Lysozyme. If I have a protein in its three dimensional structure and I just open it up into its polypeptide chain I will find that it opens up into a random polypeptide chain. Here these bonds are covalent peptide bonds linking the beads that are amino acids together.

(Refer Slide Time 9:22 min)



Here we have these amino acid residues which are linked by the peptide bonds to form the polypeptide chain. This polypeptide chain will form a Protein because it is going to fold into a definite three dimensional native structure.

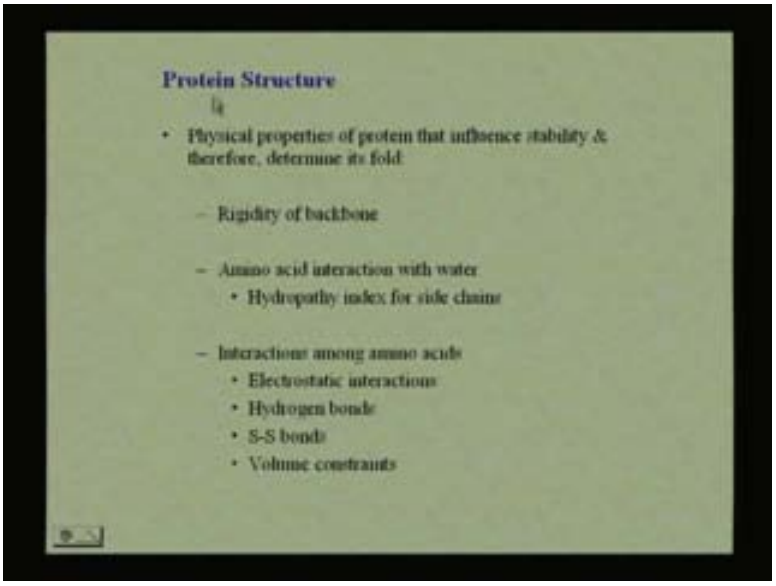
(Refer Slide Time 9:43 min)



So the sequence of amino acid residues of this protein is it start with a protonated amino terminus and end with a carboxylic acid terminus which does not have the proton attached to it and this is the sequence of my protein.

In this sequence we know that Lysine is followed by Valine followed by Phenylalanine followed by Glycine and so on and so fourth. Now you can recognize the 3-letter codes for the whole polypeptide chain that is comprises the sequence of Lysozyme. This is going to fold into a definite structure. Therefore when we consider the protein structure we have to consider the physical properties of the protein that influence the stability of the protein and determine its fold.

(Refer Slide Time 10:37 min)



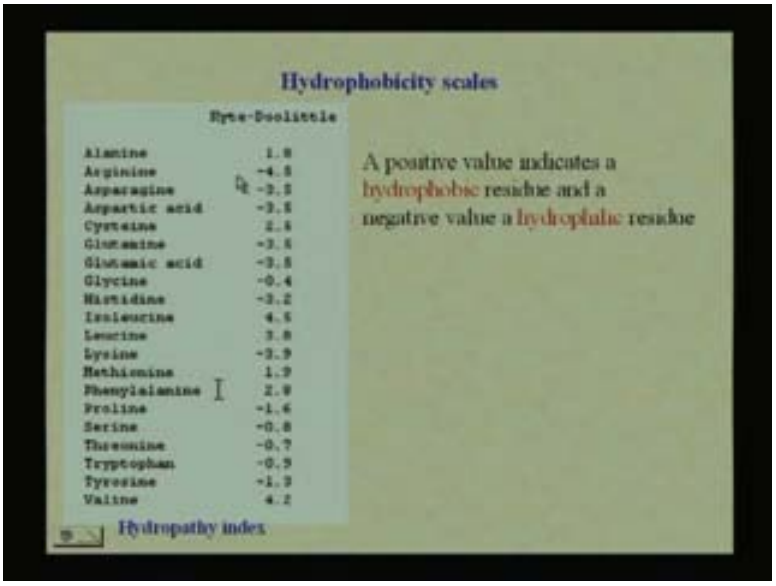
Here there is a specific Rigidity of the backbone that is essential for the structure, there are interactions of the amino acids with water depending on the Hydrophobicity of the amino acid. If the side chain has oxygen or a nitrogen atom so it is possible to form a hydrogen bond but if a side chains contains only carbon and hydrogen it is not possible. so you have a specific index called a hydropathy index which actually tells how the amino acids interact with water and how we can actually determine different regions of the protein that are going to be embedded in the protein because it has a stretch of amino acid. If you have a stretch of hydrophobic amino acid they will not be on the surface of the protein.

Then we have to consider the interactions among the amino acids. There are electro static interactions because we have charges on the amino acids. For example, Lysine which is a positively charged amino acid can interact with the Aspartic acid or Glutamic acid which is negatively charged amino acids. So we can have these acidic amino acids interact with the basic amino acids. We also have the hydrogen bonds between the polar amino acid residues and the solvent or between the polar amino acid residues themselves.

The next one that we already mentioned was the S-S bonds. The S-S bonds are the disulfide linkages. They arise from Cysteine residues coming together and forming Cysteines. So we have these S-S bonds and these are the only other covalent bonds apart from the peptide bonds that are present in proteins. And we also have volume constraints. Volume constraints means we have the sizes of the different amino acid side chains and obviously there is going to be a steric hindrance, a steric hindrance into the accommodation as to how they occurred in the folded protein.

If we look at a Hydrophobicity scale you can see the Kyte Doolittle Hydropathy index.

(Refer Slide Time 13:22 min)



It is nothing but the scale for Hydrophobicity values. Here a positive value indicates a hydrophobic residue which we have for Alanine and the Alanine is just a methyl group so it hydrophobic in nature. We know that Valine is also a hydrophobic residue. The hydropathy index for valine is 4.2. It is a high positive value indicating that it is hydrophobic in nature. Then looking at the Lysine, the Lysine is a positively charged amino acid residue but has a negative hydropathy index indicating that it is a hydrophilic residue. It would rather remain on the surface of the protein than embedded in the protein. Then Aspartic acid, Arginine, Glutamic acid, Glutamine all of these are hydrophilic residues. And the once that have positive values here are hydrophobic residues.

Later on we will see how this index actually helps us in determining which stretch of amino acids are on the surface of the protein and which stretch of amino acids are embedded in the center of the protein. So let us look at the structures of the two proteins. The right hand side one is a membrane protein. You can see this is the cell membrane which is a lipid bilayer.

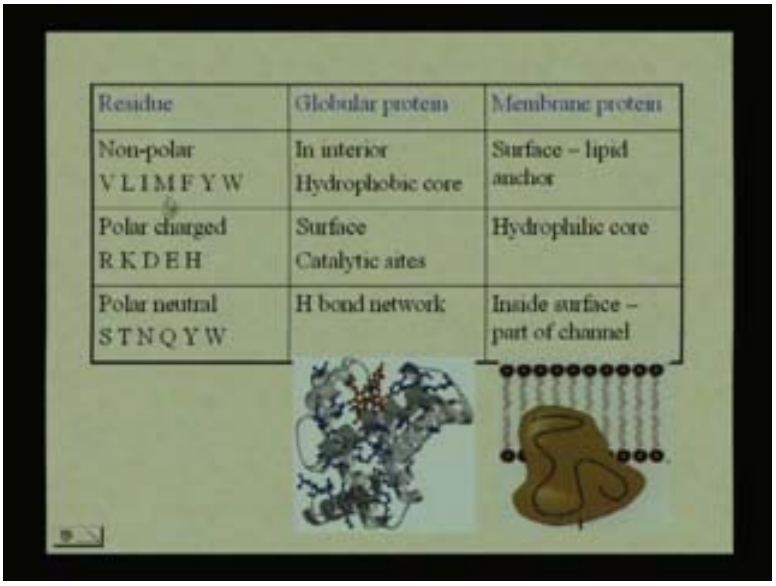
(Refer Slide Time 15:09 min)

You have polar head groups and hydrophobic tails in a lipid bilayer. We have a cell membrane that happens to be a lipid bilayer, a lipid is basically a polar head group with hydrophobic tails.

If we look at the structure of a solvent protein or protein in a polar solvent then there will be specific characteristics of the residues that are on the surface of the protein. For example, if we look at this chain is a hydrophobic chain. So we can expect the hydrophobic amino acid residues will be on the surface of the membrane protein. A polar residue on the surface will not interact favorably with a hydrophobic tail of the lipid. So you will have preferably hydrophobic interactions here.

But if we are looking at this protein we can expect hydrophobic amino acids residues to interact amongst themselves in the center of the protein. And I will expect polar residues on the surface which is just reverse of what I would expect for a membrane protein. So any protein that is embedded in the Membrane would rather have a hydrophobic surface to it because it can interact with the hydrophobic tails of the lipid bilayer. And a Globular protein that would be in a solvent would rather have its hydrophobic amino acid residues embedded in the center of the protein. So here we have a list. So we have Globular protein that is this, we have a Membrane protein that is this. I look at non polar residues. Here the single letter codes for the amino acid are given.

(Refer Slide Time 17:33 min)



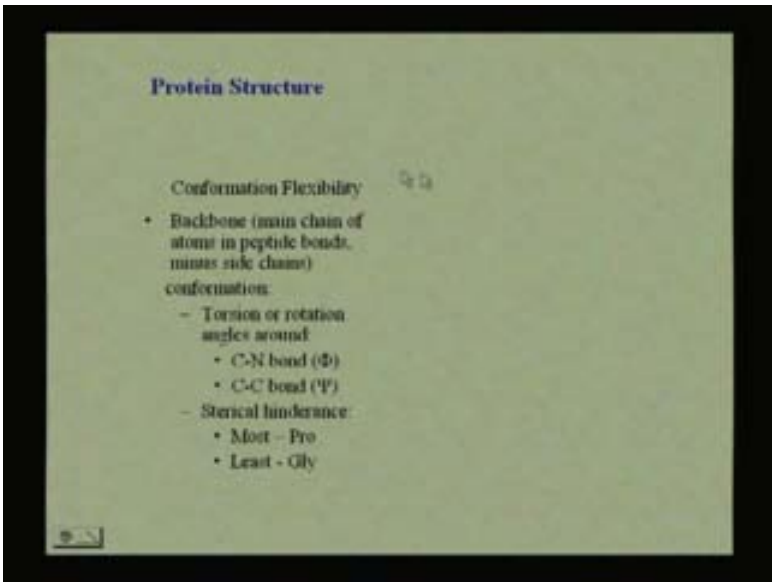
So we have Valine, Leucine, Isoleucine, Methionine, Phenylalanine, Tyrosine, Tryptophan. In a Globular protein they prefer to be in the interior to form the hydrophobic core. So they are some way here. The polar charged residues in this case are on the surface and they form the catalytic sites. They are on the surface of the protein. And the polar neutral ones would preferably form the hydrogen bonding network. So, if we look at a Globular protein and we look at the specific types of amino acids residues we know some extent where they might be located in a protein. On the other hand if we look at a Membrane protein, the non polar residues are not in the interior. They are on the surface because they interact with the lipid because they have to anchor themselves to the lipid and this hydrophobic tail would not want to interact with the polar residue. So the non polar residues would preferably be on the surface in the case of Membrane protein.

The polar charged groups would rather be in the core of the protein because they want to be as away from the surface is possible. So instead of hydrophobic core which you find in Globular proteins that are in solvents you have a hydrophilic core for Membrane proteins. And the polar neutral parts would form the inside surface are the part of the channel. Here a part of a channel means suppose on the bottom we have the inside of the cell and on the surface we have the outside of the cell. This is a cell membrane. So if I have a protein that actually integrates through the membrane there has to be a channel that goes in and out whether the salts inside the cell or the salts from outside the cell. And if you want to transport say sodium ions or potassium ions would prefer a hydrophilic environment. So it makes sense that for the lipid the inside surface of the lipid has the pattern of the hydrophilic amino acid residues that are eventually going to allow the passage or the transport of ions from the inside to the outside or the outside to the inside.

If you have a hydrophobic core in here it would be extremely difficult for you to allow this transfer to occur. So we have hydrophobic amino acid residues on the surface in the Membrane proteins that are going to link up or form interactions with the hydrophobic

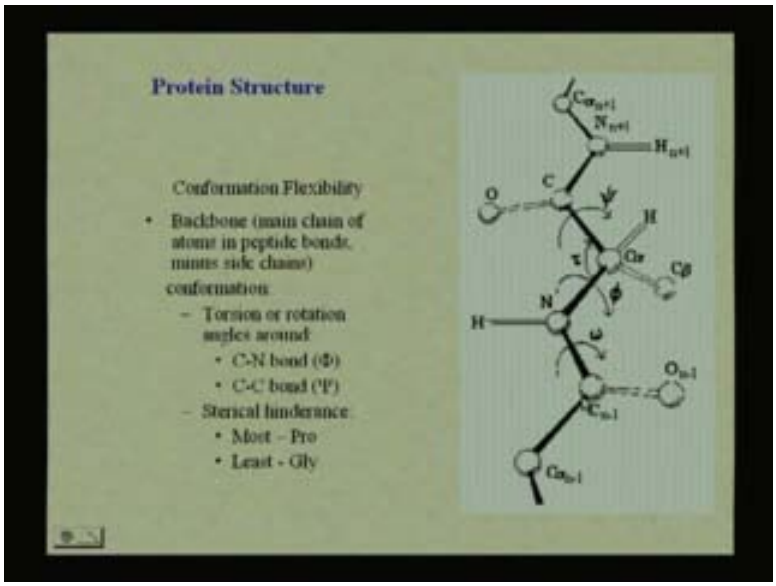
tails of the lipid bilayer and inside we have a hydrophilic core, we have a hydrophilic surface that allows the transport of ions in and out of the cell. So this is very clear to understand once you know the characteristics of the amino acid residues. And you also know the characteristics of the protein whether you are talking about the Globular protein or you are talking about a Membrane protein. So considering the protein structure there is conformational flexibility. The protein polypeptide chain is formed by the linking of the amino acids residues.

(Refer Slide Time 21:43 min)



Now we have a lot of geometrical considerations that we have to make once we link the amino acid residues together.

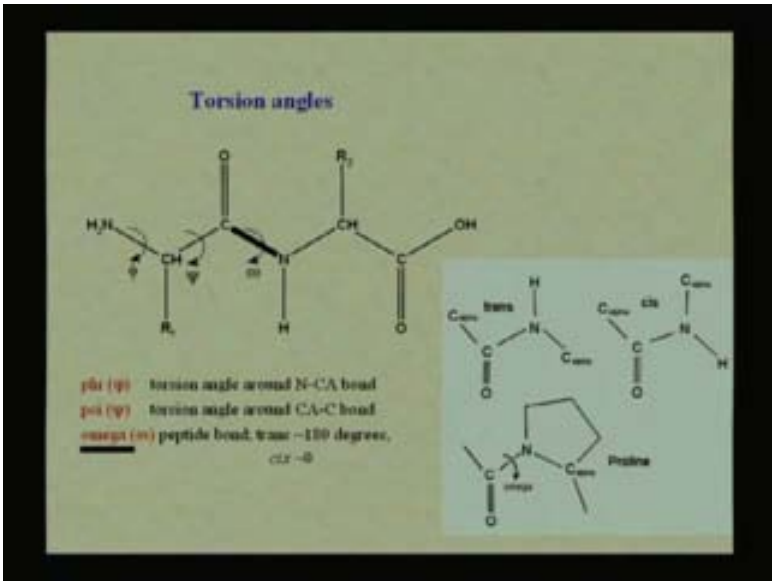
(Refer Slide Time 21:51 min)



Here we have the C_α carbon atom which is linked by the amino group and is linked by the carboxylic group. Following the chain the one in black is called the back bone that is where we have our linkages. Suppose this is just part of the chain, here we are looking at this particular amino acid I have this NH which was part of NH_2 originally, this is the C_α which is the one to which the amino group attached and the carboxylic group is attached. So this forms the part of one amino acid in the chain.

The R group has linked to the C_α. One side you have the hydrogen atoms. You have hydrogen, you have the R group, you have the NH_2 group and you have the COOH group so this makes the asymmetric. So this is part of the amino acid which is now present in the protein structure. it is part of the protein structure. So now we have torsion angles where we are looking at rotation about bonds. These torsion angles actually define the rotation about the single bonds that are present. This is the amino part of the chain, this is the C_α which is attached to the R group. This is the C double bond O. here it was attached originally an OH group which has lost with this amino group of H to form the peptide bond. Now this peptide bond actually has a partial double bond character to it.

(Refer Slide Time 24:35 min)



This is because there is a lone pair on nitrogen that actually can be donated to this bond where this is looses a COO−. So here it has a partial double bond character. Now in the partial double bond character will impart the rigidity to the polypeptide chain. The rigidity will not allow a rotation about the polypeptide chain about the peptide bond. It will restrict the rotation. Now it is partially double bonded in character. This bond is free to rotate, it is a single bond, it can rotate freely. Similarly this rotate freely, this can rotate freely but this has some restrictions due to the partial double bond character. Here the lone pair on nitrogen will results in a partial double bond character.

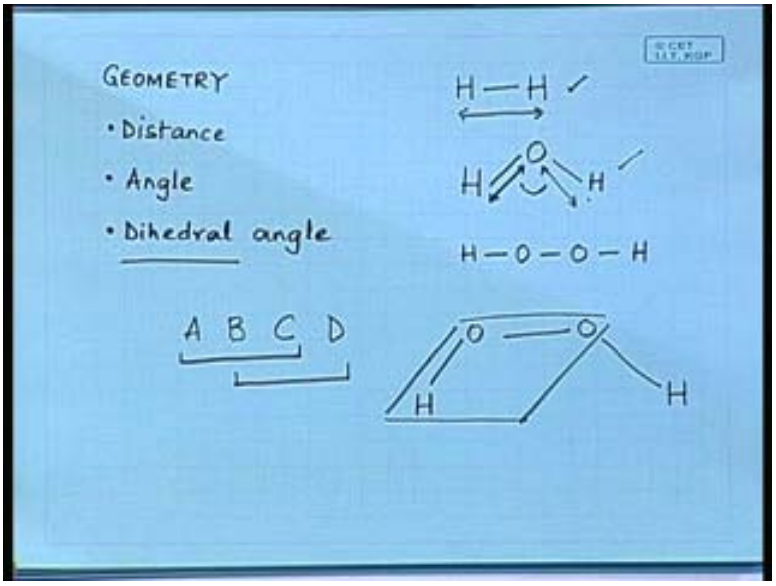
Now we can look at the disposition of the C_α residues. This is residue number one R_1, this is residue number two R_2. So this is the C_α of R_2, this is the C_α of R_1. So this is the C_α, this is C double bond O, this is NH, this is C_α of the next. Now this is R_1 because its tail is carboxylic acid. NH is the beginning of the next one, it has to be. So this is R_2. So here the orientation is Trans if this R_1 and this is R_2. If the rotate is about the peptide bond so this becomes Cis. So normally you can have a trans, in very rare cases you have cis. It is because you just have the C_α attached to the rest of the side chain. So here might steric hindrance or clashed of the side chains unless they are very small. So we would expect the peptide bond to be preferably in nature.

So now there are different definitions for the torsion angles. If we have the geometry of any compound for say to tell you about the geometry of hydrogen the only information you just need is the distance between the two hydrogen atoms. Because if you know where one is located you can find out where the other one is just by knowing the distance between the two atoms. So all you need in this case is this distance.

If you look at the geometry of water where you would look at H-O–H. it is not sufficient to explain by knowing only the distances. If I tell you this H is this distance away from the oxygen and this H is this distance away from the oxygen that does not suffice because

I have to tell you the angle between HOH if you have to get the correct geometry of the water molecule. So the distance is fine in the case of a diatomic molecule. If you have a tri atomic molecule you need two distances and an angle. Now this is the dihedral angle. Suppose you have four atoms A B C D. You know that a plane is defined by a minimum of three points. So I can define a plane using the atoms ABC, I can define another plane using the atoms BCD.

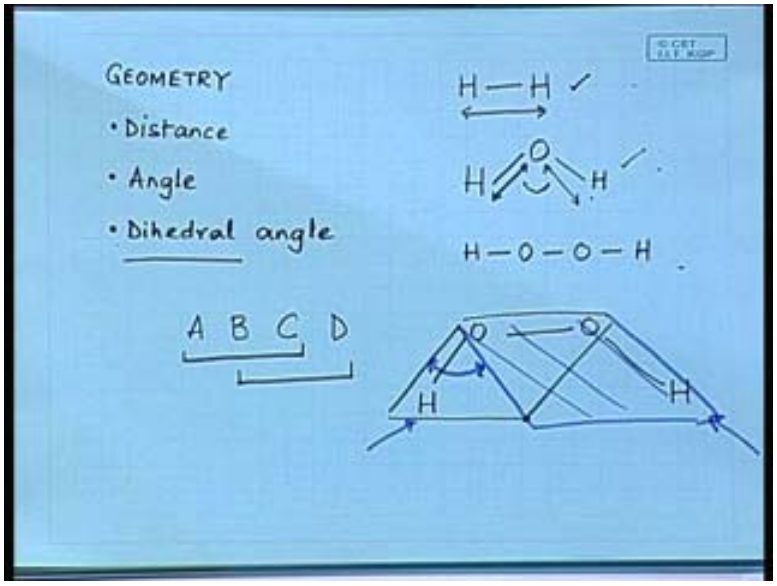
(Refer Slide Time 30:30 min)



So now I have two planes. The dihedral angle is nothing but the angle between these two planes. For example let us take hydrogen peroxide which is an example of ABCD. Here we will have one plane with HOO and the other is with BCD which is OOH. So we have two planes one is this plane that we have here and one is the plane behind that. We do have an angle between the two planes. This angle is the dihedral angle. (Refer Slide Time 31:43 min)

So when you are defining the geometry for these two atoms you just need the distance, when you defining it for three atoms you need two distances and an angle, here we do not need a dihedral angle because it does not occur. And we do not need an angle in the first case because it is not necessary. But when we come to the dihedral case we need everything we need the distance, we need the angle, we also need the dihedral angle.

(Refer Slide Time 32:24 min)



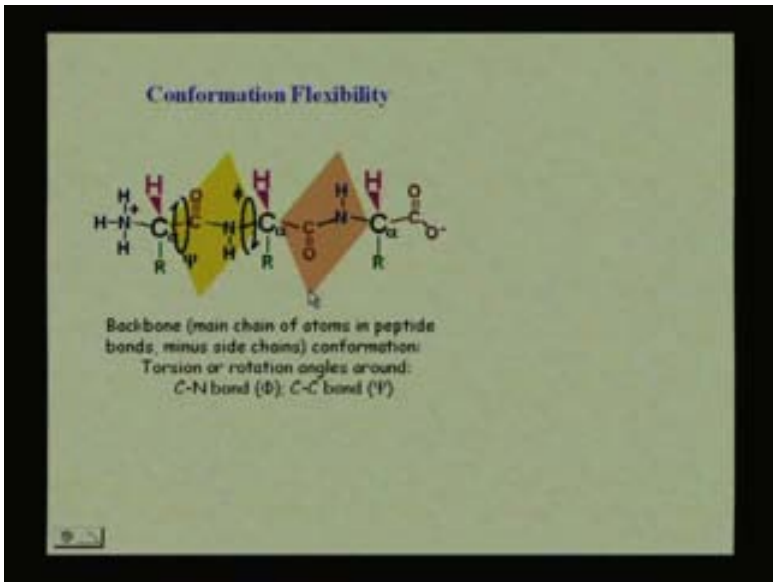
We need to know HOO angle on that plane, we need to know the OOH angle on this plane. We also need to know the dihedral angle to exactly know the disposition of this H and this H with respect to the oxygen and oxygen. So we have a torsion angle that is defined by rotations. For say we have the ø angle that is defined by rotation about the N-C_α bond. It mean this is N, this is C_α and it is similar to the hydrogen peroxide where these are actually the two oxygen atoms.

Now when you mention a polypeptide chain you can speak about the back bone. So the four atoms are the C, the N, the C_α and the C that belongs to the next or other this amino acid. So this C belongs to this amino acid because here this has an amino group, a C_α group and a carboxylic group. So this has to be one amino acid. This is the previous amino acid, it has to be because it begins with an amino terminal, has a C_α and the C of the carboxylic group.

So when I am talking about in similar cases that I mentioned about the hydrogen peroxide I have to have four atoms that are going to define the torsion angle or the angle between the planes. Here the points are A B C D. I have a plane that will define A B C and this plane defines C N C_α. the next plane is N C_α C. the ø angle is the angle between these two planes.

If I want to define rotation about the N-C_α I have to go one atom before it and one atom after it, it is just like doing HOOH. we have two planes, the plane that contain atoms number one, two, three and another plane that contains atoms number two, three, four. So here I have C N C_α C that is defined an ø angle. The C N C_α C means we have two planes in which one plane is C N C_α, the other is N C_α C plane. The angle between these two planes is the ø angle. Similarly you can define a ψ angle. The ψ angle is the angle between the C_α, C atoms. So the four atoms we have here is N C_α C N, the two planes we have here is NC_αC, C_αCN.

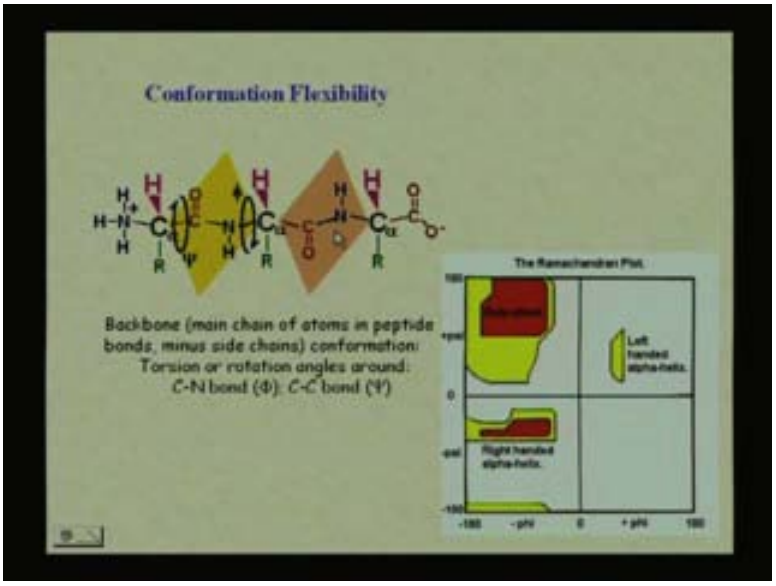
(Refer Slide Time 36:52 min)



Now here this is the back bone, and here the marked planes are the peptide bonds. This also forms the plane. I can also have a dihedral that defined this but I am not considered that importantly because it has a partial double bond character. There are restrictions to its rotation. So it is less flexible and it is more rigid. But does not mean it can not rotate you do get trans and the cis.

The angle between C_α C is ψ angle. So here N C_α C is one plane, C_α C N is the other plane. The ø is the rotation about the N C_α. So the four atoms in this case are C N C_α C and two planes are C N C_α and N C_α C. So defining specific ø, ψ angles that can defines the geometry of the polypeptide chain the amino acid disposition. If we rotate this, the R has also rotated along with it. So the whole polypeptide chain is going to have a large dependence of the ø, ψ angles. So in the whole polypeptide chain each amino acid will have an ø, ψ angles associated with it.

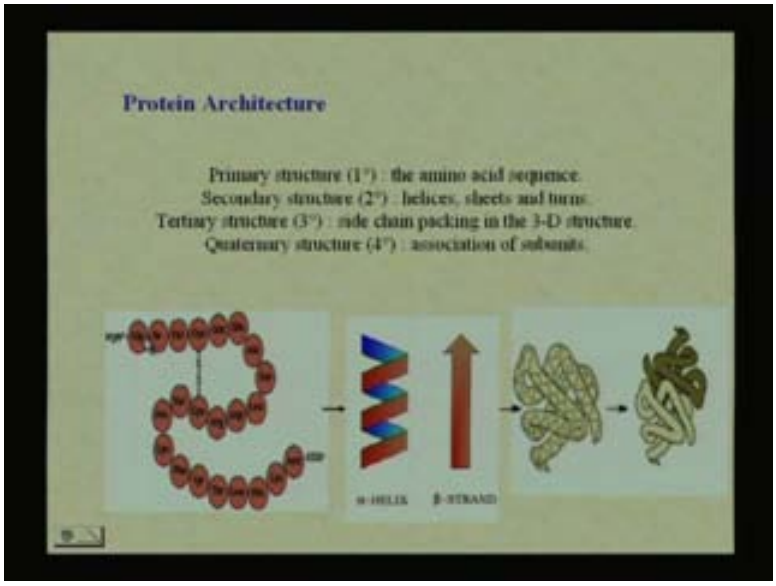
(Refer Slide Time 39:15 min)



Now if you plot the ø, ψ angles you will get the Ramchandran plot which extremely important plot in Protein structure. And we have to proud that it was GN Ramachandran who actually pointed this out. He has a land mark paper where he showed the definite geometry of the ø, ψ angles that are defined here belonged to the specific regions. Today this is called as Ramachandran plot. And this is must for any protein structure that is solved you will have a correct Ramachandran plot for that protein. (Refer Slide Time 40:00 min)

Here we have the primary structure, we have the secondary structure, we have the tertiary structure and we have the quaternary structure. The primary structure is the amino acid sequence. It is nothing but just the sequence of amino acids which give you no information about the structure. You only get the structure once you fold it.

(Refer Slide Time 40:38 min)

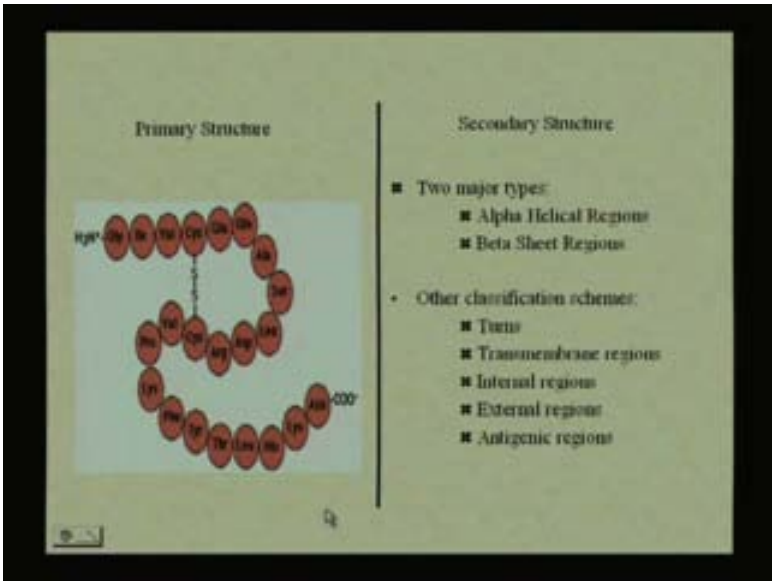


Next we have the secondary structure. Here this is the primary structure and these linkages are peptide linkages. This is a disulfide linkage and these are the only two covalent linkages that we have in protein structure. And this is amino terminal and this is carboxylic acid terminal. So this is the amino acid sequence of the polypeptide chain.

Now this sequence you found out with the ø, ψ angles can bend out the flexible in any direction. So you can have some thing that looks like a helix because you have flexibility to rotate the polypeptide chain. So the polypeptide chain can rotate and it can form α helix. It can also form a β strand. So when we want to form a total polypeptide or rather a protein structure it is just a linking of all these types of different secondary structures together. So we basically build up this chain from bits and pieces of helices and some turns or random coils at linked these two together. So if these were in the same protein they would be just linked by another part of the polypeptide chain. So we get the tertiary structure which is the side chain packing in the three dimensional structure.

Now this could be a specific unit called as a sub unit. It means that if you have a dimeric protein which means the two monomer units linked up together non covalently, so they just associate with one another. So you have a dimeric structure. For example Hemoglobin, the Hemoglobin is a tertiary structure. Hemoglobin has a tertiary structure, it has a quaternary structure and it has four sub units. It means that it has four such monomeric units that form the protein. So all protein will have a primary structure, a secondary structure and tertiary structure but all proteins will not have a quaternary structure. You only have a quaternary structure if it is not monomeric in nature. If the protein is monomeric there is just primary, secondary and the tertiary. But if there are a number of sub units that have to link together then you have a quaternary structure. So this is the primary structure, this is a secondary structure. The two major types of secondary structure are α helices and β sheets.

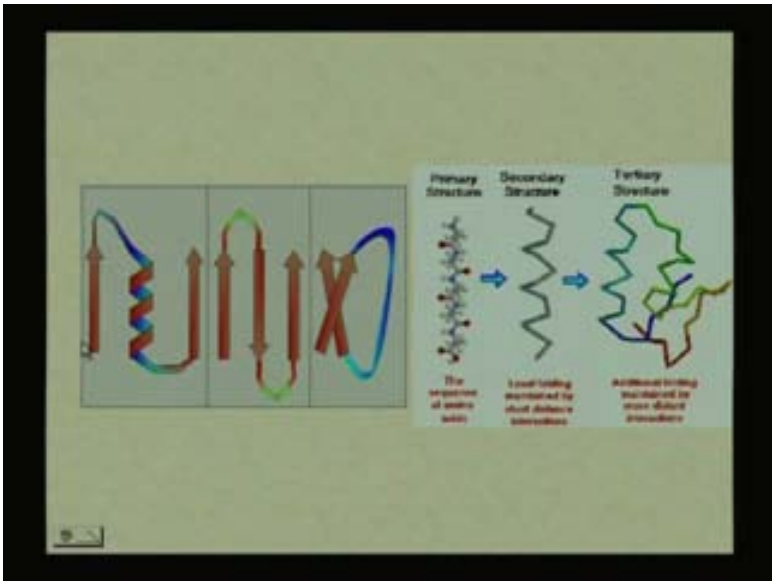
(Refer Slide Time 44:05 min)



The β sheets will form by linking the β strands. Apart from the primary structure all the other structures do not have any covalent linkages. So we will see α-helix or some thing that looks like α-helix but there is no covalent linkage between the polypeptide chains informing the α-helix which is extremely important.

When forming the secondary structure that is α-helixes or β sheets the only non covalent interaction that is important in this case are high hydrogen bonds. So we will consider this when we linking them up. Here we have a beta strand. This is the polypeptide chain. So here the protein will begin with an amino terminus. The arrow is pointing in upward direction so this has to be amino terminus of the protein, it is going up.

(Refer Slide Time 45:30 min)



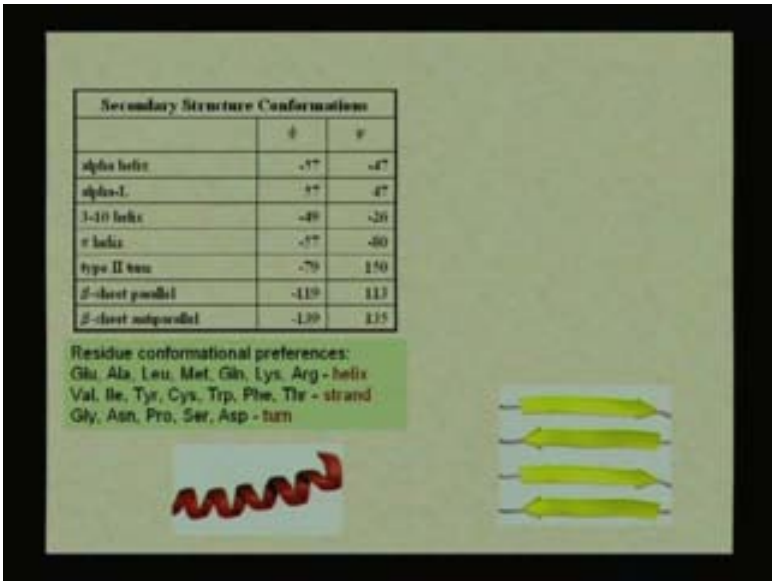
This is a β strand and it will form an α-helix then it forms another β strand. So together this is going to form a polypeptide chain structure or folded structure.

In the second case we have an amino part, it goes up, it comes down and it goes up again. So here we have three β strands that form a β sheet. Then in the third case we have again the protein begin here then we have just one huge loop that has no characteristic structure to it. But here we can have different sorts of linkages. We already knew that the polypeptide chain is just a linkage of the amino acid residues one of the other. For example, if I take a necklace or beads and I drop it on the table what is it going to form? It is going to have some structure. If I pick it up and drop it again it is going to have some other structure. But the protein will always fold into the same structure every time.

We have the primary structure which is the sequence of amino acid residues. The secondary structure is the local folding of the α-helix. It is maintained by short distance interactions and in this case hydrogen bonds only. When we consider the tertiary structure it is additional folding that is maintained by more distant interactions. It means suppose I have one polar amino acid here and I have another polar amino acid on this side then they are likely to form a hydrogen bond so they will come together. What about the disulfide linkages? We can have a disulfide linkage between residue number one and residue number seventy eight. It means that distant parts of the protein structure come together to form the overall structure. (Refer Slide Time 48:31 min)

Here we have different secondary structure conformations. The secondary structure conformation means you can have α-helix and this is the preferred ø and ψ of the α–helix.

(Refer Slide Time 48:58 min)



We have a right handed alpha helix and left handed alpha helix. What do we mean by a right handed alpha helix? We mean that the helix can go actually in two directions it could go up or down. Now if this is the direction of the propagation and this is the way your polypeptide chain goes so this is a right handed helix. If you consider your left hand and this is the polypeptide chain and this is the direction of the propagation the left hand. Then this is a left handed α–helix. Left handed α-helix is very rare in proteins. You do not generally see them. You just have right handed α helices. You can see the conformation of amino acids is always L conformation.

So, if you look at α- helix usually when we mention α- helix we do not say it is a right handed α-helix because by default it is. So this is the preferred region for the α-helix in proteins, this is the preferred region for the β- sheets in proteins. These are β strands that form a β sheet. Now in between these β strands in this blank space we could have any thing we could even have a helix. Because this is one part of the polypeptide chain linking with the some other part of the polypeptide chain. But the difference between the α-helix and the β helix is the α-helix has to be contiguous. We cannot have the part of the α-helix here and the rest from the other part of the protein that is not possible. (Refer Slide Time 51:00 min)

If an α-helix begins at residue number sixteen and continuous to residue number thirty means residues from sixteen to thirty will all be part of the an α–helix. But if I have a β sheet whether it can have residue three to seventeen or residue 40 to 46 it does not matter they just form β strands that are part of the same β sheet. Now when I consider beta sheet here and right handed an α-helix here then I have specific bonding characteristics that are linking the secondary structures.

We have different types of secondary structure conformations. We have considered a β sheet parallel or anti parallel. The strands are might be parallel to one another. here we

have an anti parallel β sheet because one of the strand is in this direction, the other is in this direction then the next one is in that direction and then I have one in this direction so this would be an anti parallel β sheet. Now some residue conformational preferences are listed here. It means these residues would like to be in α-helix, they preferred to be in an α helices. These residues preferred to be in strands and there are certain other residues are preferred to be in turns. Here turns are going to link the two secondary structural elements together. So we did learned today was how we can look at the geometry of the polypeptide chain, we have learned definite definitions of the polypeptide chains in terms of different torsion angles.

These torsion angles are one was we learned at the peptide plane would be planar in nature because of the double bond characteristic. Due to the partial double character of the peptide bond makes it rigid. But we can have rotation about the other single bonds that are present. These can form the ø and ψ angles. Rotation about these gives the definite conformational considerations that have to make and they give rise to specific secondary structural considerations. So I have my primary structure that is linked by peptide bonds, I have my secondary structure which we only learn the α-helix and the β sheet. And the α-helix and the β sheet will have definite ø, ψ angles that they correspond to and they would form a hydrogen bonding network. Thank you