Biochemistry - I

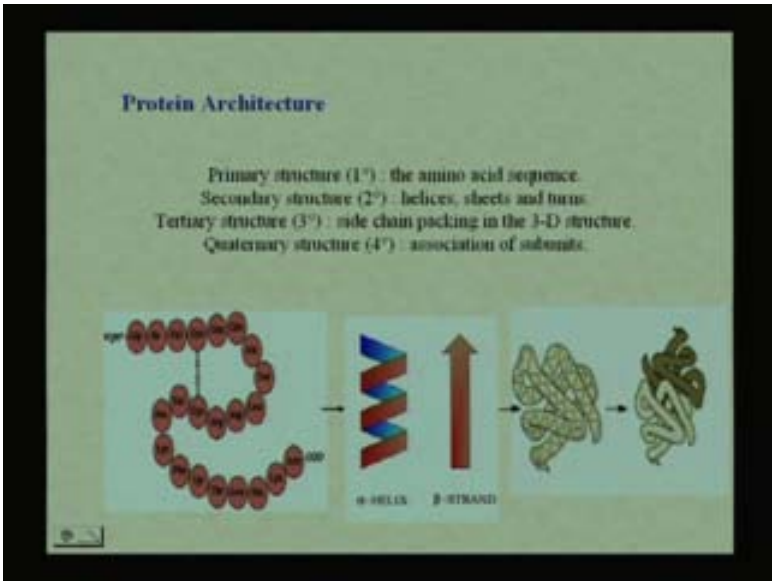
Prof. S. Dasgupta
Department of Chemistry
Indian Institute of Technology

Lecture – 4

Protein Structure II

Welcome, we continue our discussion on Protein structure and Protein architecture. Here we have the primary structure which is just the amino acid sequence of the protein, the secondary structure that comprises helices, sheets and turns.

(Refer Slide Time 1:00 min)

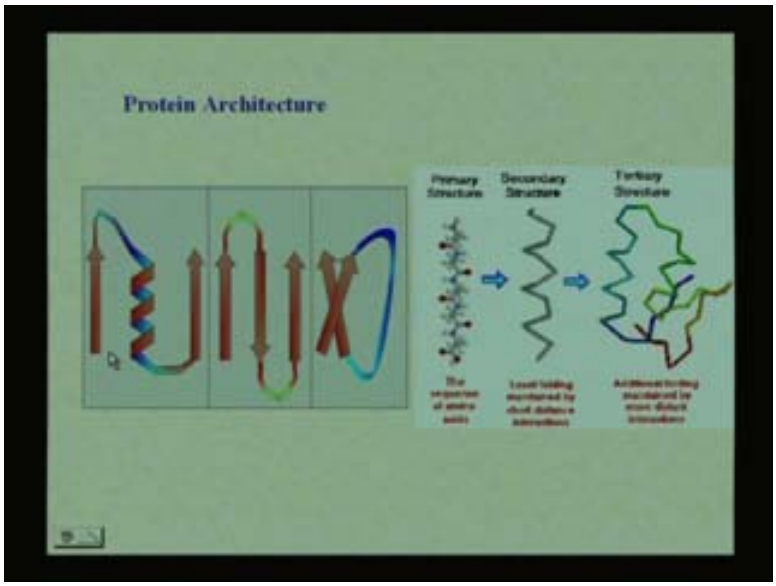


Then we have the tertiary structure of the protein that is the side chain packing in the three dimensional structure. Finally we have the quaternary structure of the protein that is the association of sub units.

Here we have the primary structure followed by the different elements of secondary structure. Then we have the tertiary structure of the protein. And this monomeric sub unit has associated with another sub unit where you have a connection that is not a covalent bond. It is just an agglomeration or aggregation of these two units together is forming the quaternary structure of the protein. There is usually no covalent formation between the quaternary sub units in the quaternary structure.

All proteins will have a primary structure, a secondary structure and a tertiary structure. But all proteins will not have a quaternary structure because all of them do not have a polymeric or an oligomeric structure. Now when we consider the specific elements of secondary structure associated to form the tertiary structure.

(Refer Slide Time 2:25 min)

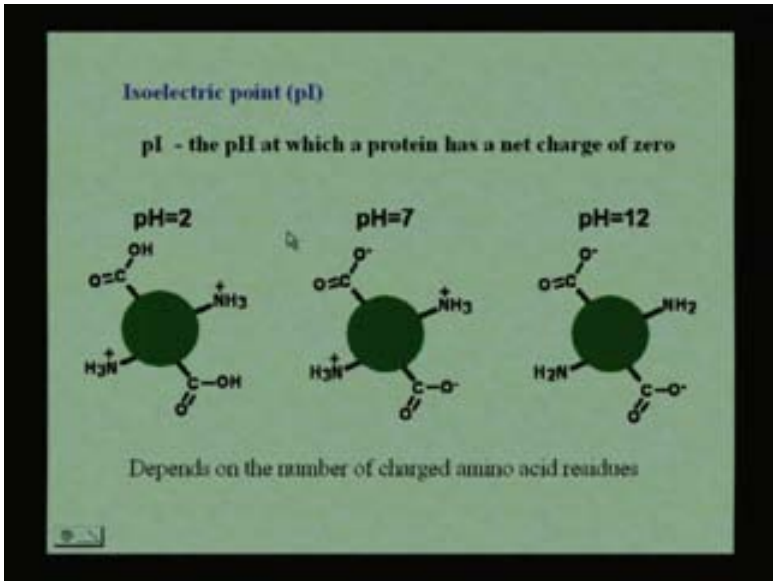


So from the primary amino acid sequence we go on to the helix and then we go on to the tertiary structure. We already know that each amino acid will have its own Isoelectric point. Each amino acid will have its own definite point where it will loose its protons or it would be zero in net charge.

The Isoelectric point also exists for the protein. So, the p^H at which net charge on the protein is zero. A pH at which a protein has a net charge of zero means the protein will have a large number of amino acid residues, some of them may be acidic and some of them may be basic so each of them will loose their protons at different times. So, for example if we have a $p^H = 2$ then we have all the side chains are protonated as well as the amino terminal and the carboxylic acid terminal are protonated. We cannot say which is which. We know that one of this has to be the terminal and we know that one of this has to be the terminal but this may belong to Lysine and this may either belongs to an Aspartic acid or a Glutamic acid.

Now as we increase the p^H we will loose the carboxylic acids protons first. In this case pH = 7 so the net charge on the protein is zero. So this would correspond to Isoelectric point of such a protein.

(Refer Slide Time 4:32 min)

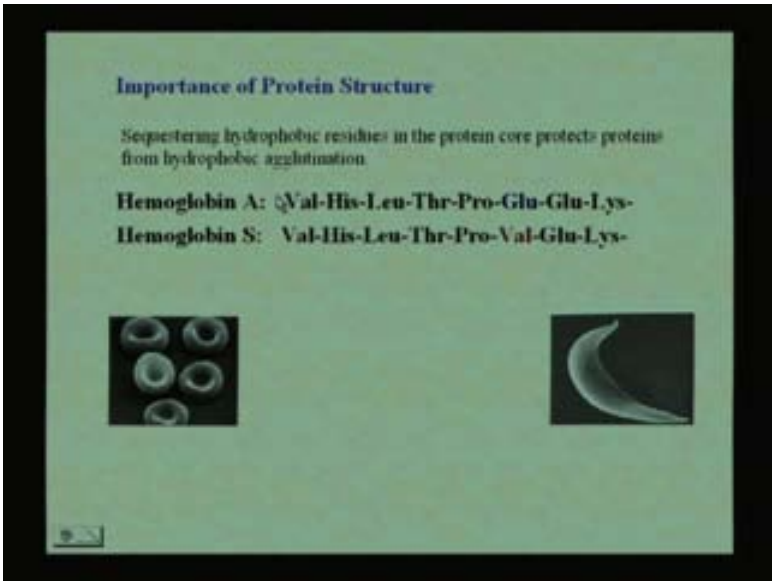


Now here you may have another protein that has an extra NH_3^+ to reach a net charge of zero or have to go even further high in the p^H to achieve a net charge of zero. So we will reach the p*I* of the protein by once achieving the net charge of zero. So as I keep on increasing or rather deprotonating all these protons that are available it will come to a very high p^H where it has lost even all the protons that were associated with the amine groups. (Refer Slide Time 5:04 min)

So, if you were to have a basic protein means a protein that has a large number of Lysines or Arginines associated to it then the p*I* of the protein will be high. If I have an acidic protein rather then p*I* will have a low value because it will reach a negative charge or a negative charge after the deprotonation of the carboxylic acid groups will achieve at a lower p^H. So the protein will have a net charge is zero at a lower p^H then would a basic protein. It means when we consider the p*I* of the protein it depends on the number of charged amino acids residues. So it will depend upon the number of Lysine, Arginine, Aspartic acid, Glutamic acid that you have in the protein. Because you are going to associate the charge that are going to be essential in understanding how you can actually determine the p*I* of the protein. So, if you have a polypeptide you can determine the p*I* from just knowledge of the content of the polypeptide.

You know the number of charges you know you have to come to a net charge of zero to attain the p*I*. Now we will consider the importance of the charge or the importance of the protein structures, a very slight change in the structure mutagen.

(Refer Slide Time 6:47 min)

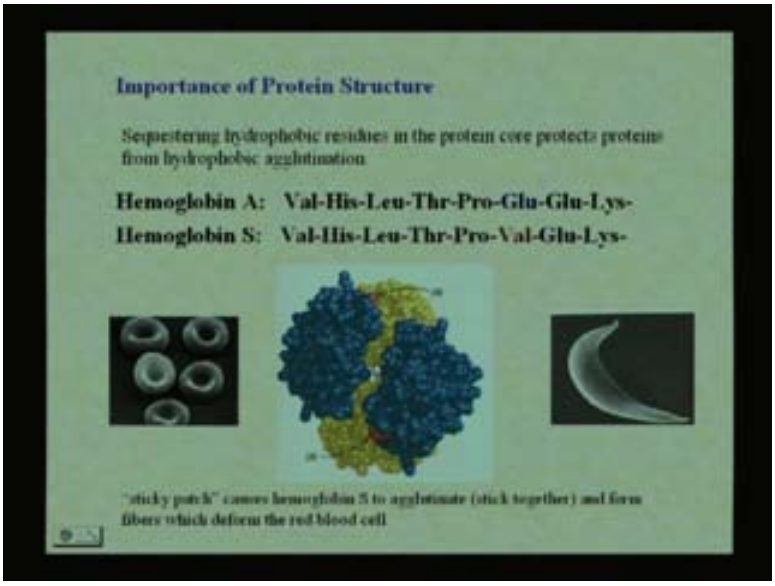


The red blood cells which are originally circular in shape they become sickle shape. This is the β chain of hemoglobin. Let us consider it as a polypeptide sequence of one of the chain that is presented in hemoglobin. Here you can see everything is the same except this Glutamic acid. It has become Valine, the rest of the protein chain is exactly the same. It has a hundred and seventy four amino acid residues and this is the sole change that makes this round red blood cell as this. Now here essentially you have a Glutamic acid, a Glutamic acid is a charged amino acid residue. (Refer Slide Time 8:00 min)

It would prefer to be on the surface of the protein. Because it has a negative charge associated to it. And it would like to associate with the solvent around it. Now when we change it to a Valine we can make it a hydrophobic amino acid residue that does not like to be on the surface at all. So here you can see this is hemoglobin, this is the beta chain. The blue ones are the beta, the yellow ones are the alpha. So here I have the quaternary structure of the hemoglobin comprised of four sub units in which two of them are α type, two of them are β type. The one that is marked in red here is the Valine.

Now this Valine is on the surface because the Glutamic acid was on the surface. So another hemoglobin molecule comes here and this bottom one of the next hemoglobin molecule sticks with this one means it forms a hydrophobic interaction because it ones to be away from the solvent.

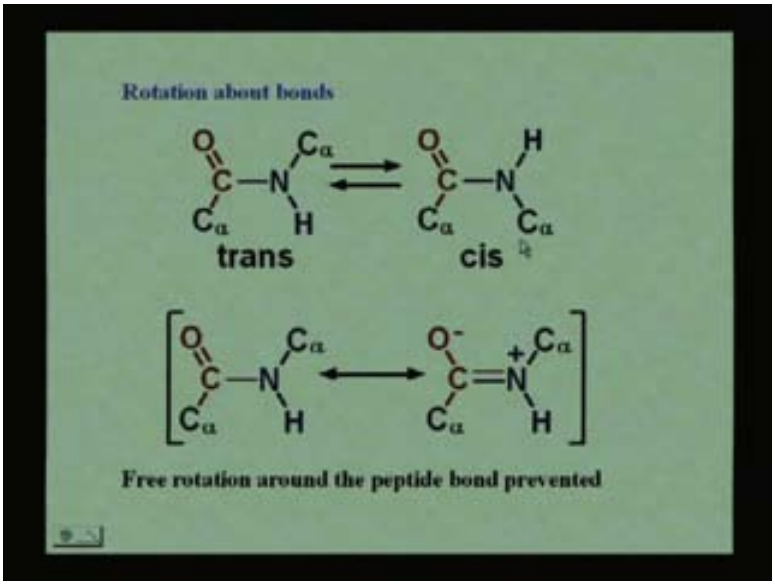
(Refer Slide Time 9:20 min)



So we will get a fiber because we can have other hemoglobin that will stick to the bottom of this one. So you are eventually get fibers and these fibers are in the form of the sickle cell. This was the first molecular so called disease that was determined to be solely on the amino acid sequence of the protein. You see the importance in understanding the whole protein structure. It is just a single property of that amino acid that results in such a detrimental effect to red blood cell forming the fibers to form your sickle cell anemia. And the reason being is that the Glutamic acid was on the surface which is no longer there. You have Valine which would prefer to be away from solvent. So we have the amino acids that are linked by these peptide bonds. And we have a partial double bond character to the peptide bond. In this case we have the C_α, this is the C_α of the next amino acid and both of them are attached side with chains. This is trans in nature because the C_α's are on two opposite directions.

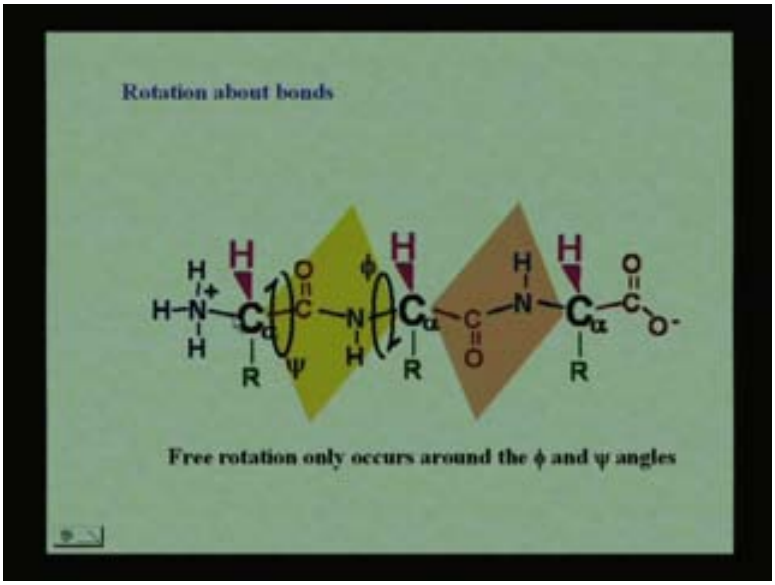
Here we can have a cis peptide also. But in the case of cis peptide the C_α's are on the same side. And if we have C_β's on top of this so there might be a steric clash. This is the reason why you have more trans peptide bonds then you have cis peptide bonds. Also we have prevention of rotation about this single peptide bond due to the formation of this partial double bond. You have the lone pair on nitrogen so it forms this partial double bond and because of its partial double bond character it imposes the rigidity, it is not as flexible as the other single bonds in the peptide chain.

(Refer Slide Time 11:50 min)



These are the other single bonds will be amount for the rotation around these single bonds. Here we have the peptides that form two rigid planes. We are calling these as rigid planes because of the partial double bond character restrict rotation about the CN peptide bond.

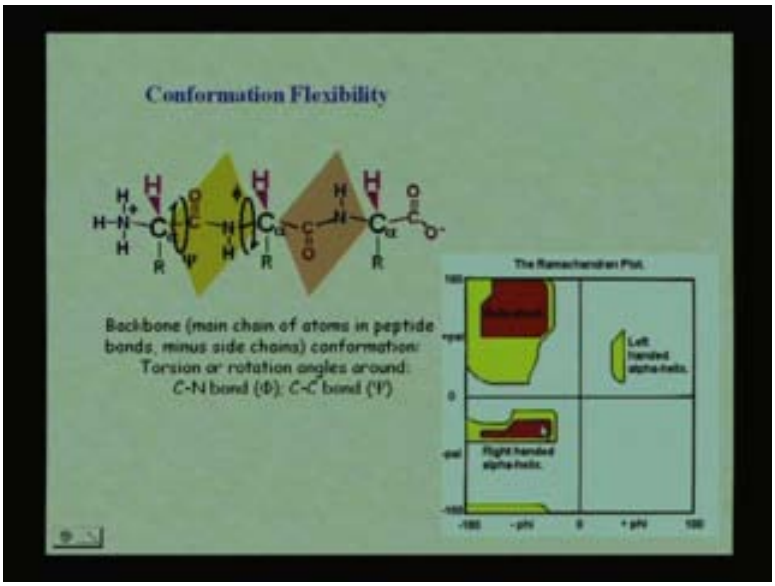
(Refer Slide Time 12:15 min)



Here we have free rotation about the ø and the ψ angles. We define it by four atoms. For example we can define ø angle by the C, the N, the C_α and the C. It means that if I construct a plane that has the points C, N and C_α on it and I construct another plane that has N, C_α and C on it then the angle between those two planes will give me the ø angle.

Similarly I can get the ψ angle that will be the angle between the planes that have N C_α C and C_α C N. So we have ø, ψ rotations and this gives rise to the Ramachandran plot.

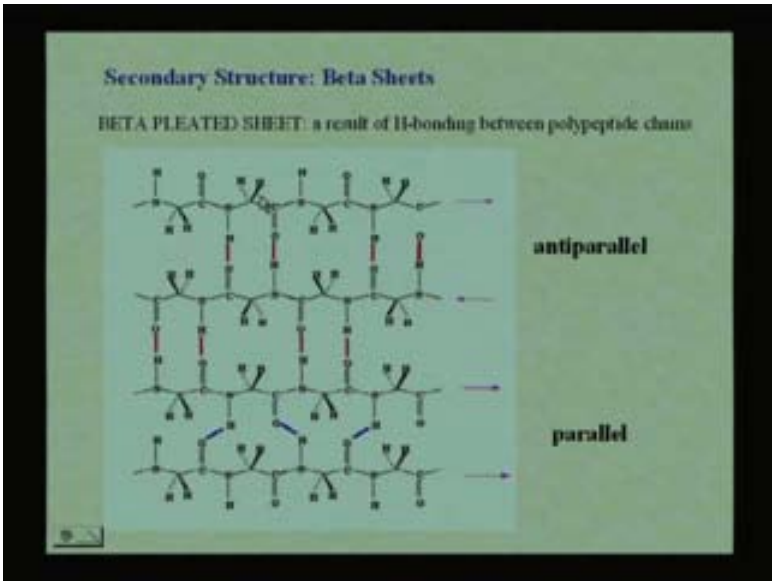
(Refer Slide Time 13:30 min)



We will understand how these locations are restricted in their ø, ψ values based on the geometry that they have. This is a beta sheet, when you look at a beta sheet you can straight away tell in which direction the polypeptide chain is moving because you know that you always begin with an N so you have N, a C_α followed by a C. So this amino acid is a Glycine. Here we have N C_α C and N C_α C so the protein chain is in this direction. (Refer Slide Time 14:28 min) Then we move to the next one, if I go from here I have N C C_α that is a wrong direction. The right direction is N C_α C and it the back bone, N C_α C. so the direction of the polypeptide chain is opposite to the previous one, it is anti parallel.

Then we have two strands at the bottom. We have N C_α C so it is going from left to right. We also have N C_α C in the bottom strand so it is also going from left to right. So we have two parallel strands here and we have two anti parallel strands here. So it is just the direction of the beta strand that is going to determine whether it is in anti parallel beta sheet or it is a parallel beta sheet.

(Refer Slide Time 15:44 min)



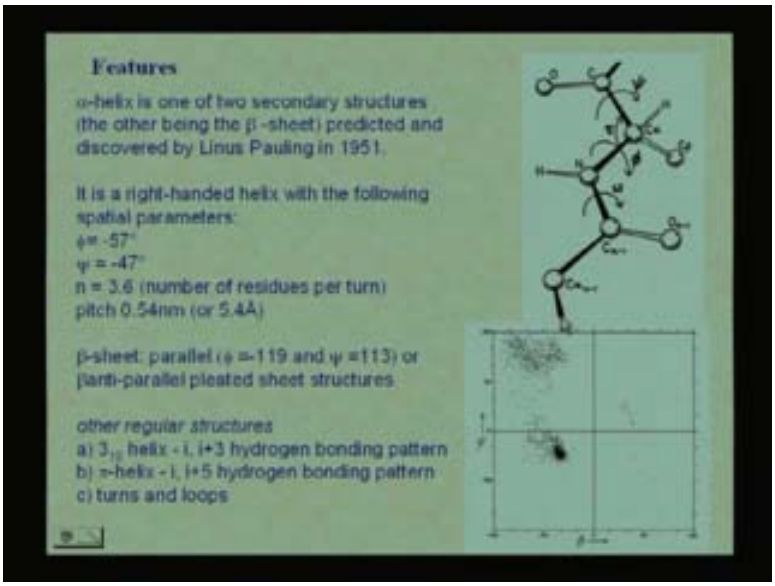
Now very carefully we can look at the hydrogen bonding between the sheets. There are two types of secondary structure are important. They are the alpha helix and the beta sheet. Both of these have hydrogen bonds that associate with them. Here the anti parallel beta sheet and the parallel beta sheet have different types of hydrogen bonds. If you look carefully you will be able to distinguish the difference. The difference this N C_α C is a single amino acid. If we look at this hydrogen bonding it is in the opposite direction from right to left. So I have N C_α C, it is the same amino acid. (Refer Slide Time 16:55 min)

So we have the N of this amino acid link with the O of another amino acid on another strand and the same amino acid will have its amino NH link with the carboxylic O of the other amino acid. So we have an amino acid N C_α C on top. On the second strand we also have N C_α C and the hydrogen bonds are between the same set of amino acids.

Looking at the difference in the parallel set, we have chains in the same direction and the strands are also in the same direction in the parallel set. So we have N C_α C on the top strand and we also have N C_α C in the bottom strand. This has a hydrogen bond with this amino acid and the CO of this amino acid here has a hydrogen bond with another amino acid.

In the first case the hydrogen bonding is the NH and CO bond with the same amino acid. In the second case the NH and CO bond with two different amino acids. So when you have the amino acid in a beta strand will form a hydrogen bond with two different amino acids then it is a parallel beta strand. When you have the amino acid forming hydrogen bonds with the same amino acid then it is an anti parallel beta strand. So this is the hydrogen bonding difference between the two irrespective of alpha helices.
When we speak about alpha helices that is another type of secondary structure that also have specific ø, ψ angles that we have to consider and they are ø = -57 and ψ = -47 which on the Ramachandran plot.

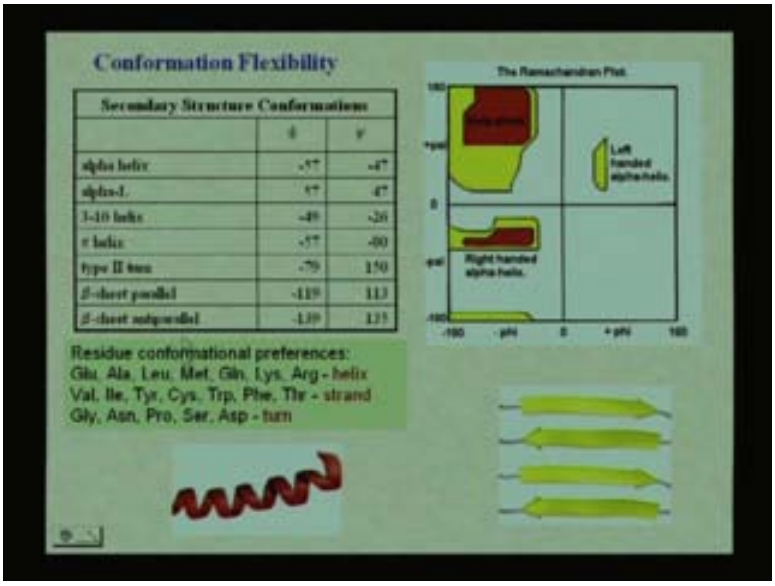
(Refer Slide Time 19:58 min)



So this would be a typical Ramachandran plot for a protein where you see some α–helices, some in the β-sheet region and some here which are usually Glycine. The ones on the right hand side with positive ø are usually Glycine.

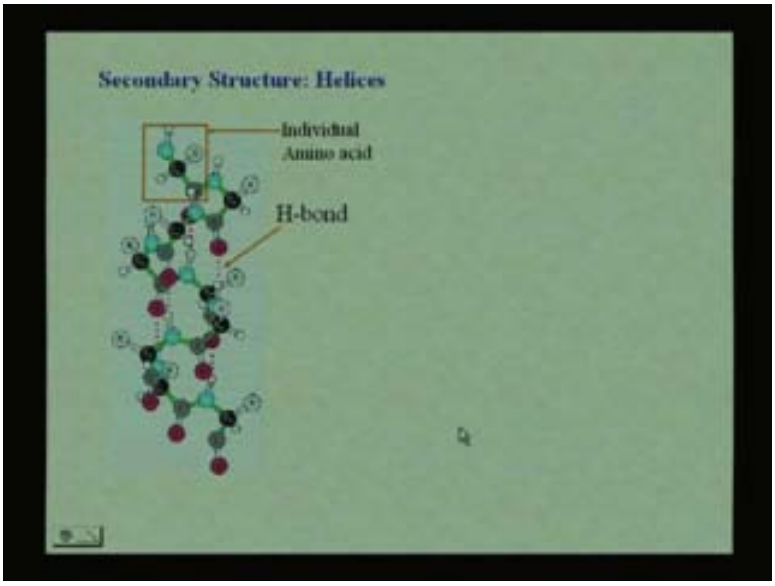
Here when we look at the chain you can say from the location of the nitrogen, C_α and the C. We know this is C N C_α C so I am going from the bottom to the top. And we know the type of angles we are considering. We have an ø angle, we have a ψ angle and the ω angle is the angle of the peptide bond. So this ω angle is usually trans so it is usually a hundred and eighty degrees. You can at times get a cis peptide bond at that point ω will be close to zero. Now let us look at the hydrogen bonding pattern of helices. When we have a helix we turn it round to form a spring. The n is 3.6 for a normal α-helix means that this is the number of residues per turn. You have understood that the helix looks like a spring so it will have a number of turns associated to it. These are features of secondary structure conformations. And you already know about the parallel beta sheet and anti parallel beta sheet. And these are the ø, ψ locations where you might see them.

(Refer Slide Time 22:15 min)



But usually we know that a right hand helix is here, a β sheet is here and this is a left hand α-helix or mostly as we mentioned it is Glycine because it is most flexible amino acid. And these are the residue conformational preferences. Now we will come to the secondary structure called the helices. Here we have a specific individual amino acid. If we were to follow a direction of the polypeptide chain we can say in which direction it is going. The nitrogen atoms are in blue, the oxygen atoms are in red and these are the R groups that are circled in gray. So we have the nitrogen here, this is the C_α because we have the R group connected to it. This is the CO so we are going from the top down because this is the amino acid N C_α C after that N C_α C and then N C_α C then next N C_α C then after that N C_α C.

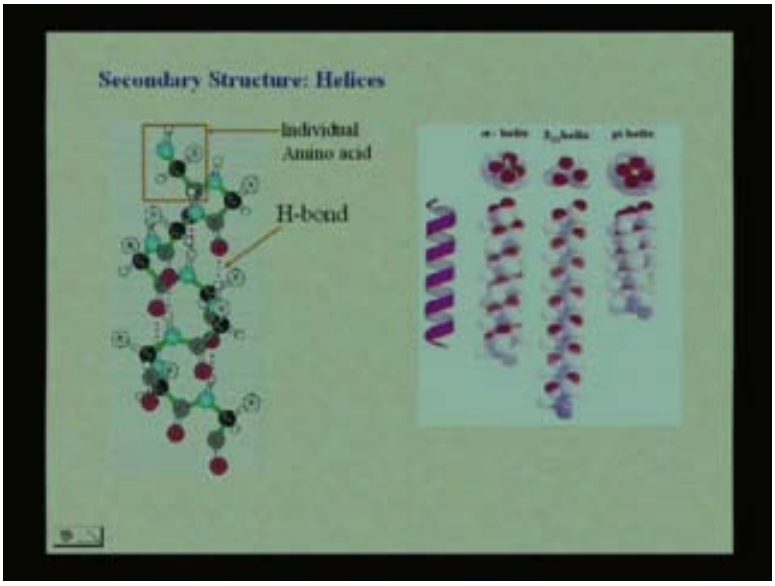
(Refer Slide Time 23:54 min)



So you are traversing a helix. Now you will go from one point to the next point as you traverse a helix. So as you go form one point to the next point you have a certain turn associated to it. Now for this particular one the direction of propagation is down. The helix is going in this direction and down so it is a right handed helix. If it were going up it would be a left handed helix. You will have to follow the polypeptide chain. In this case the nitrogen is here and there is C_α then you have C so you are going in that way. If you are going that way then the direction of propagation is down so the helix is a right handed helix.

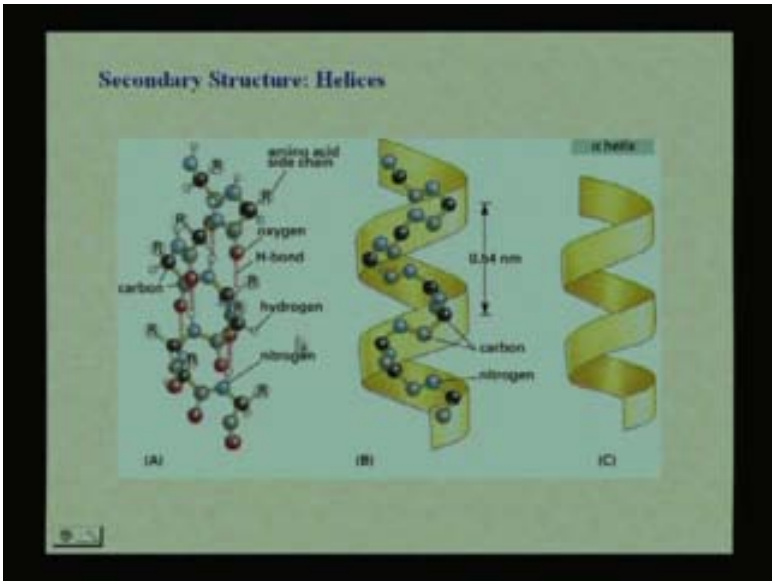
If the polypeptide chain is turning in this way and the direction of propagation is this so it is a left handed helix. So this is a right handed helix. You have to follow the polypeptide chain and have to follow its direction of propagation by looking at the ends of C_α C. the N C_α C and N C_α C and N C alpha C. Look at which direction it is going by following the whole chain. Once you follow the spring you can determine in which direction it is going and you can find out that this happens to be a right handed alpha helix and it has specific connectivity. We can see connectivity in the form of hydrogen bonds much likely saw in beta sheets but these hydrogen bonds sort of link the spring together. So we do not have an extended polypeptide chain we have a helix. These are the different types of helices you can have.

(Refer Slide Time 26:10 min)



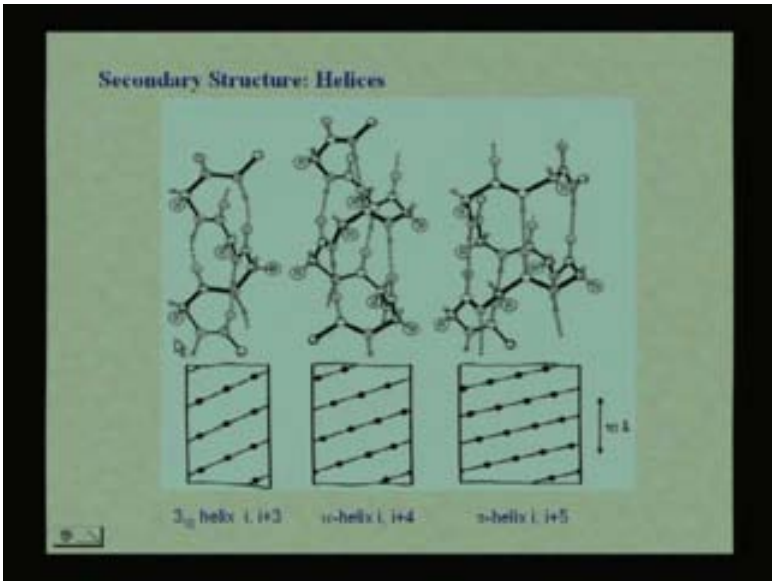
You can have α-helix, a 3_{10} helix and a pi helix. These just tell you how many turns there are because they look kind of different. You can see here there are about four atoms are in red, here there are three and here there is just about five one picking out there. Now again we are looking at the same thing we have N C_α C. So this is the traversing of the chain. So the polypeptide chain is now shaped like a helix. And we have the nitrogen of the amino acid, we have the C_α of the amino acid and we have C of the amino acid along the helix. So we have the two carbons and we have the nitrogen. Now this is the pitch. The pitch is for one complete rotation the height that the helix or the polypeptide chain has traversed. The helix should have a pitch because it is also a spring just like a screw. So it is basically a rotation of the polypeptide chain and you have the pitch.

(Refer Slide Time 28:01 min)



Now if we look at the pitch of these helices, the difference between a 3_{10} helix and α-helix and a π-helix is the hydrogen bonding.

(Refer Slide Time 28:23 min)

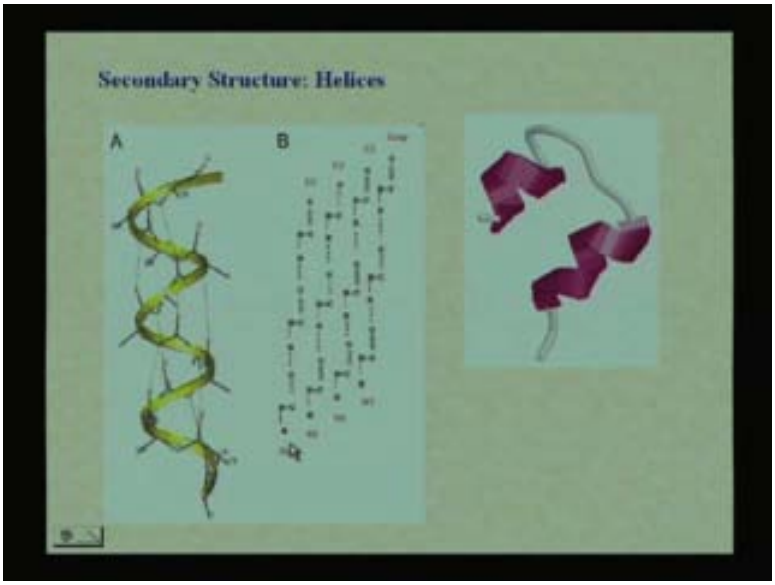


They all form hydrogen bonds and it is just with whom they are forming the hydrogen bonds. So we have a hydrogen bond formation from the i^{th} residue to the $(i + 3)^{rd}$ residue in the 3_{10} helix and the 3_{10} tells you that you have three residues for turn and there are ten atoms in those in that turn around. So in the 3_{10} the 3 stands for three residues in a turn and then 10 stand for ten atoms in the turn. And the hydrogen bonding of the helix is from the i^{th} residue to the $(i +3)^{rd}$ residue.

Now we will look at the normal α-helix. In this α-helix the hydrogen bonding pattern is slightly different. It has the hydrogen bonding from the i^{th} residue to the $(i+4)^{th}$ residue. So you have basically a polypeptide chain that is rotating like a spring. You have a hydrogen bonding between the CO and NH of the back bone. In the 3_{10} case it is from i to (i +3), in the normal α-helix it is from i to (i + 4) and in the π-helix which is extremely rare it has from i to (i +5).

Now the nomenclature for the 3_{10} helix can also be written for the α-helix. In this case it is 3.6 residues per turn and thirteen atoms. So if you were to write it in the same manner instead of writing 3_{10} we would write 3.6_{13}. You need to understand the difference in the different types of helices is solely due to the difference in the hydrogen bonding pattern.
So basically this makes the sort of a tighter helix, this is slightly loser than that and this is even loser because basically you are increasing the pore. You will have a tighter helix if the hydrogen bonding is from i to (i + 3). So the inside cylindrical part is smaller. If you have it from i to (i + 4) then it is slightly larger and if it is from i to (i + 5) is even larger. So you can see how this diameter gets larger. This shows the hydrogen bonding pattern even better.

(Refer Slide Time 31:45 min)

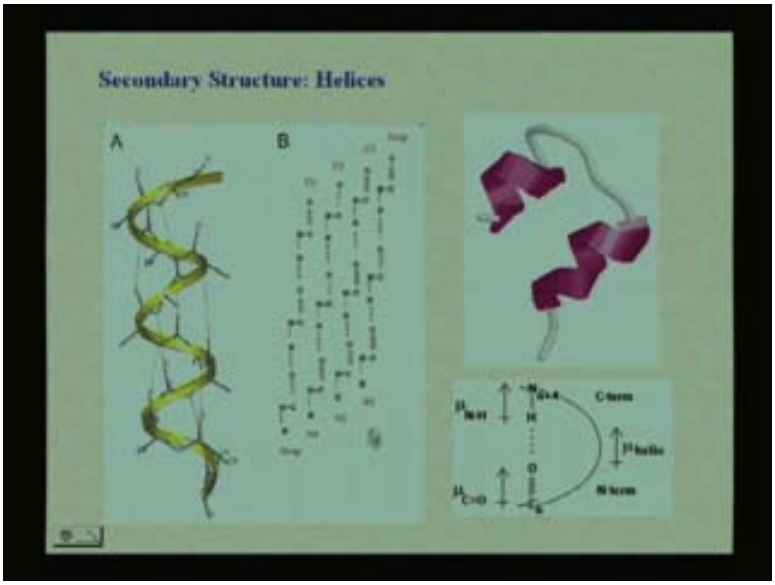


The normal alpha helix is from i to i + 4 which is the hydrogen bonding pattern. You can see the C double bond O here. So we have N C double bond O, N C double bond O. The C_α is in between these here.

Here the amino acid is N C_α C because there is the amino terminus on one side, the carboxylic acid terminus on the other side and in between is C_α. So the representation is in this fashion to show the hydrogen bonding pattern. So here if I have the first amino acid then fifth one is supposed to link to it. So if this amino acid is one then the remaining are 2, 3, 4, 5. Here we have a turn which is helix.

So the number 2 should have a link with number 6 then 2, 3, 4, 5 and then 6. We have the hydrogen bonding between the NH of this and the CO of 2. It is the second amino acid number 2 that is what is contributing to the hydrogen bonding here is the CO of residue number 2 and NH of residue number 6. If I go to one it is the CO of 1 and the NH of 5. If I go to number three it is the CO of 3 and the NH of 7. Here you notice something that all COs are pointing in the same direction. It means this alpha α-helix can have a dipole associated with it which imparts a polarity to this. Due to the fact that all the COs in this helix have to point in the same direction if they are to form this hydrogen bond which is specific for an amino α-helix.

(Refer Slide Time 35:05 min)



So we have basically the N terminus of the helixes to the C terminus of the helix. We have an over all dipole that is from the N term to the C term. Why do we have it? We have it because we have the hydrogen bonding pattern of the helix. This hydrogen bonding pattern of the helix is going to impart a polarity to the helix. And because of this polarity in the helix we have all the oxygen atoms pointing in the same direction, we have a dipole associated to that and we have the helix dipole in that direction.

Usually when you have two helices they are anti parallel. The reason being is the dipoles cancel one another out because you would not want the protein to be Polar in that sense. So usually when you have helices they are anti parallel in nature so that it can counteract the dipole that is associated with the α-helix. So whatever I have shown here in the pink are the alpha helices which are connected by just some part of the polypeptide chain. Now we will learn more about a helical wheel. You are going to construct a helical wheel.

Now, looking at this diagram, this is the picture of a protein here I have a strand this which is a strand of a particular β-sheet and these are just connectors. This is the α-helix. If I look at the α-helix I know that I have solvent surrounding here, I have particular

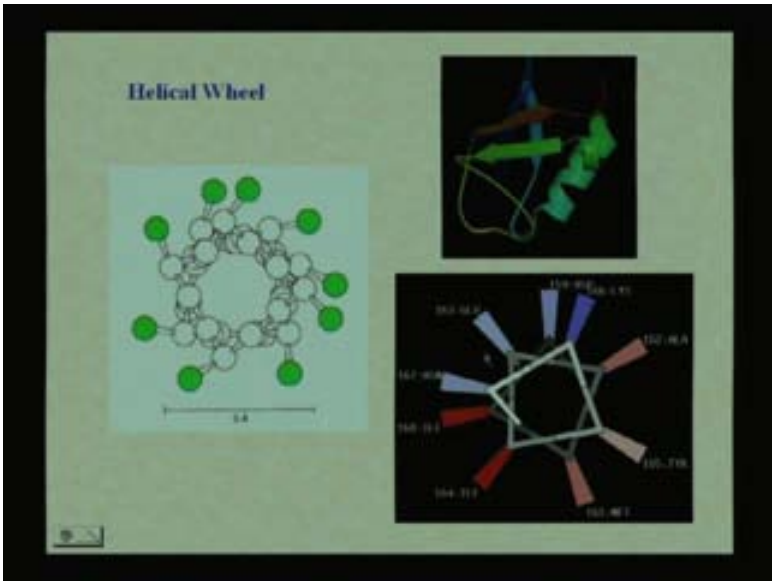
residues that are actually sticking out from the helix that all the R groups are sticking out from the helix. If I look at the helix I can say that this part is associated to the solvent and this part is associated to the core of the protein. These amino acids should be hydrophobic nature and one's that are sticking out on this side should be hydrophilic in nature.

So if I consider every turn say here I have one residue and the fifth residue here i and i + 4, so if I have a hydrophilic residue sticking out here then the fifth residue should also be hydrophilic in nature and so on and so forth because I my helix is a surface helix and here I expect hydrophobic types of residues because this is facing the central core of the protein. I can expect hydrophilic amino acid residues on the surface because it is facing the solvent.

Now I am looking down the helix and these are all back bone atoms. The ones that are in white are all the N's, C_α's and C's and so on, the green one's are the R groups. Now what happens to the R groups is they are surfaced around here. So what I should expect is this set should be one type, this set should be another type because I am looking down the helix. If I am looking down the helix, all the ones that are one side here are facing inside the protein and all of them should be hydrophobic in nature and all these ones should be hydrophilic in nature.

This is an example for a typical helical wheel that you could construct. The ones that are in blue in the helices are hydrophilic in nature. How can you say that? Here we have Asparagine, here we have Glutamic acid, here we have Aspartic acid here we have Lysine and these are all hydrophilic in nature. And here these are Isoleucine, Isoleucine, Methionine and Tyrosine to some extent and Alanine.

(Refer Slide Time 40:16 min)

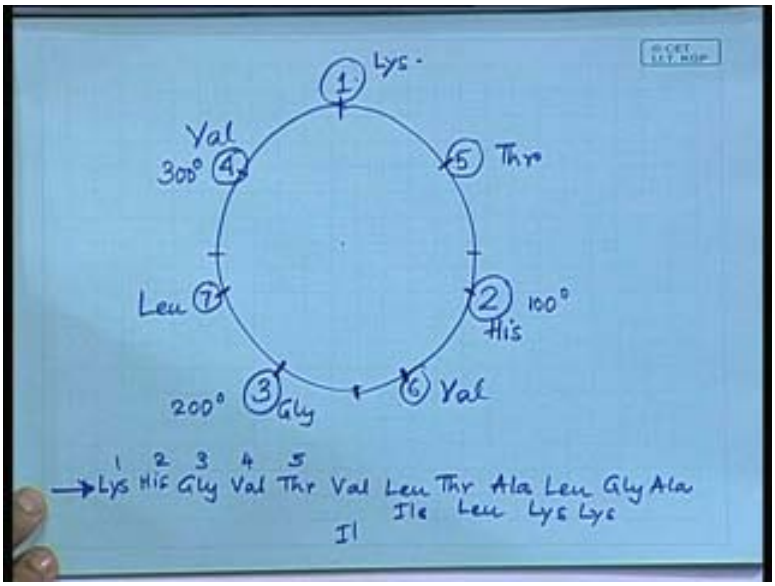


These are all mostly hydrophobic in nature. So having a picture of the helical wheel we can straight away say which part is going to face the solvent because we know the

characteristics of the amino acids. So all these blue ones will be outside because they are hydrophilic in nature and they will interact with the solvent. All the red ones will be inside because they are hydrophobic in nature. So how do we construct the helical wheel? This is how you construct the helical wheel. Now you have 3.6 residues per turn or usually we can say there are four residues per turn.

If you want to construct the helical wheel we would put residue number one say on the top, for every residue there is a rotation about 100º so the next residue is will come here that is residue number two. So if we want to construct a helical wheel we are going to have residue one here which is on the top, we have the center. So the residue number one is here, residue number two is here. Where is residue number three going to be? This is the 180º but residue number three is going to be around 200º that will be some where here. Where residue number four will be? I have this is at 100º, this is at 200º, so the fourth is going to be at 300º. This is approximately 300º so this is residue number four. Then you can go on.

(Refer Slide Time 43:06 min)



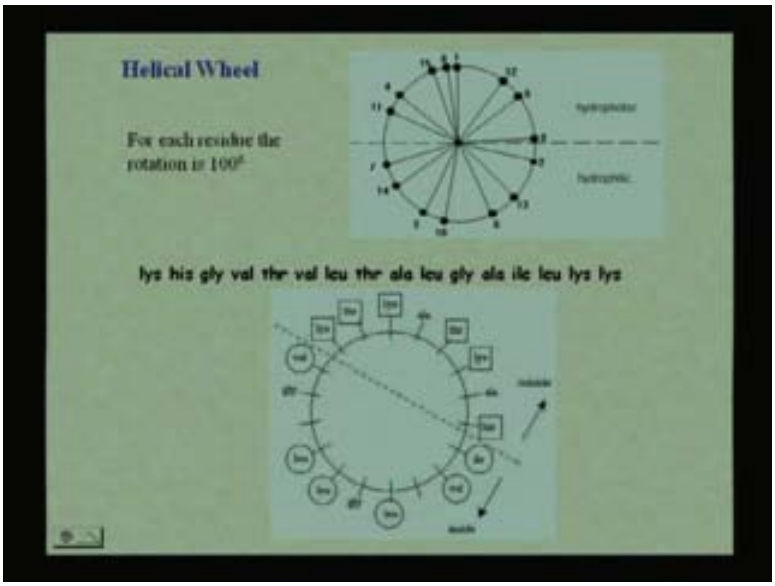
Now we want to construct a helical wheel for this so let us construct it. The first one I have is Lysine and the sequence we have here is Lys His Gly Val Thr Val Leu Thr Ala Leu Gly Ala Ilc Leu Lys Lys.Now, where is residue number 1 going to be? It is going to the top. The residue number two is Histidine. so this will be Histidine, residue number three is Glycine so this will be Glycine and residue number four is Valine. And then residue number five will be at 400º which is basically forty degrees after this so it could be some where here so this is going to be residue number five which Threonine.

So the residue number 6 is going to be somewhere here which is Valine, residue number seven is going to be some where here which will be Leucine. So what you can do is you can construct the whole helical wheel. What do you need to know? All you need to know is this sequence and you need to know for every residue the rotation is 100º that is

sufficient information for you to construct a helical wheel. Then once you construct a helical wheel, from the wheel itself you can predict which part is going to be outside, which part is going to be inside if it were to fold into a protein. So you can actually say which part will be hydrophilic in nature and which part will be hydrophobic in nature.

Thus if we do whole construction for a helical wheel then it will be look like this. Now if we look this picture it should have some of these at least Lysine, Histidine, Glycine, Valine, and Threonine and so on and so forth. So now you know how to construct a helical wheel. When you construct this helical wheel we have a dotted line here where it tells that these are mostly hydrophilic in nature so this is going to be the outside, these are mostly hydrophobic in nature so they are going to be on the inside. Even though it is a very simple construction it gives a lot of information.

(Refer Slide Time 47:02 min)



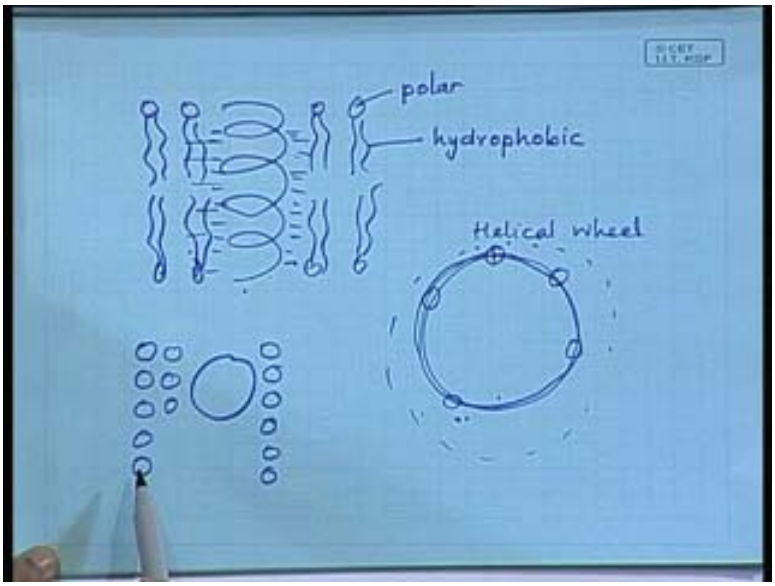
All you need to know is just the sequence of the protein. Once you know the sequence of the protein you can construct a helical wheel. For example, as I said this sequence forms a helix then you can say which part of the helix is inside and which part of the helix is outside. All you need to know is for every residue that you consider you have to go 100º. And then obviously you have to know types of amino acid residues to know which one can forms hydrogen bonds and which one can not forms hydrogen bond. Now what happens is this line does not have to be at the center as you can see it is not at the center. It may be just up here, just saying that most of it is bedded or most of it inside but a little part is outside.

So when we are looking at this helical wheel, we can straight away say that we have an outside region and an inside region based on the types of amino acid that we have. Now we are considering a helical wheel for say a membrane protein. What do we have in a membrane? We have some thing like this. If I have a helix here what should I get when I construct a helical wheel? If I construct a helical wheel I will have different residues for

say one, two, three, four and five and so on and so forth. What do I expect on the surface here? Here these are hydrophobic and these are Polar.

If I look at this side so I have interactions here and here in the helix, the interactions have to be with the hydrophobic tales of the lipids. So what do I expect the helical wheel to comprise of all hydrophobic residues? Does that make sense? It does because you should have hydrophobic interactions. If you do not have hydrophobic interactions then this helix will not stay here at all because the helix need to have a favorable interaction with a lipid tales which are hydrophobic in nature. The surface of this helix has to be hydrophobic in nature. So if I construct a helical wheel for this helix I would not have a distinction where I would get an outside and inside I would only get all hydrophobic. So if you do get all hydrophobic it means that your helix is part of membrane. Now if we have it such that I have four helices here.

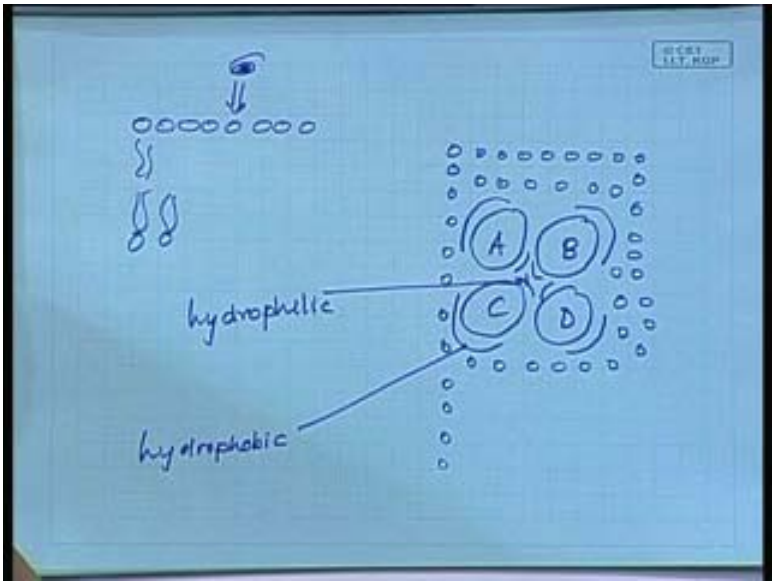
(Refer Slide Time 50:48 min)



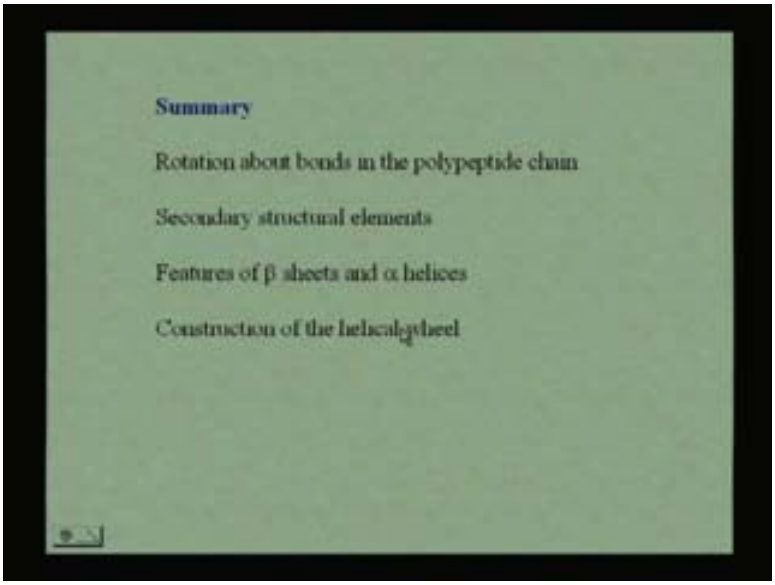
If I am looking from the top to down these are my polar heads and I have the helix here. That is also possible because you have to remember that your surface actually looks like this.

If I look at the membrane from the top it looks like this. This is a cross section. If I look from top then I can look at the surface which is actually all polar heads of lipids. I have four helices in the protein that is embedded in that. if in construct a helical wheel for A, B, C and D then helical wheel will be such that this part is going to be hydrophobic because it is interacting with the hydrophobic tales of the lipids that are down there. Then the inside region is hydrophilic which is exactly what you see in such types of proteins. And it is convenient to have this part as hydrophilic so that it can transport ions. And this part is hydrophobic. So what we learned today was we have rotations about the bonds in the polypeptide chain that give rise to ø, ψ angles and we know that the ω angle is present in proteins.

(Refer Slide Time 53:02)



(Refer Slide Time 53:30)



But we rather have a Trans peptide than a cis peptide due to the restricted rotation because of the partial double bond character of the peptide bond. Then we have talked about the secondary structural elements where we looked at the features of the β-sheets and α helices. What are the features of the β-sheets that we looked at? They are parallel and anti parallel in their differences in the hydrogen bonding patterns. Then we looked at the hydrogen bonding pattern for α helices. And we learned how to construct a helical wheel that is going to exactly tell us where a helix is located. So today we will stop here. Thank you.