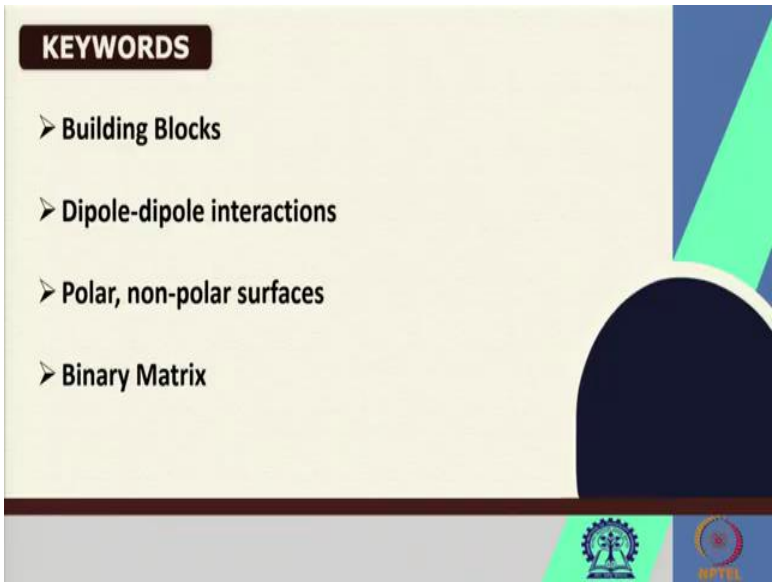
Fundamentals of Protein Chemistry
Prof. Swagata Dasgupta
Department of Chemistry
Indian Institute of Technology, Kharagpur

Module - 02
Protein Architecture
Lecture - 09
Protein Interactions

(Refer Slide Time: 00:19)

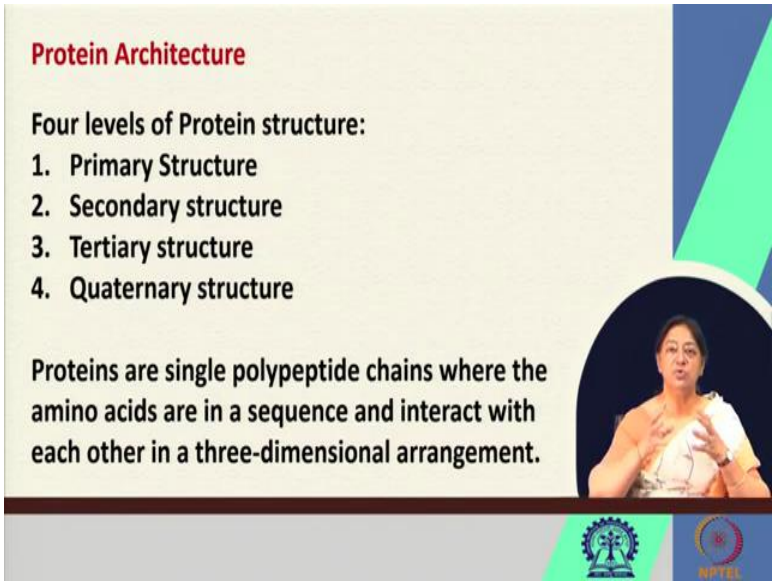
In this lecture of module 2, we will be talking about protein interactions. We had looked at the different types of structures in proteins. The aspects of protein folding, mean that we have to understand hydrogen bonding, salt bridges, van der Waals interactions, and hydrophobic interactions.

(Refer Slide Time: 00:38)



Now what we mean by all these, is going to give us an idea of what the protein structure in terms of how it folds into its compact three-dimensional structure, is important about.

(Refer Slide Time: 00:49)



So, when we look at the protein architecture, we see that the primary structure is just the string of amino acids that are connected together by the peptide bonds. When we look at the secondary structure we see the alpha helices and the beta sheets. The different types of alpha helices are the regular right handed alpha helix, then we have the 3_{10} helix and the π helix.

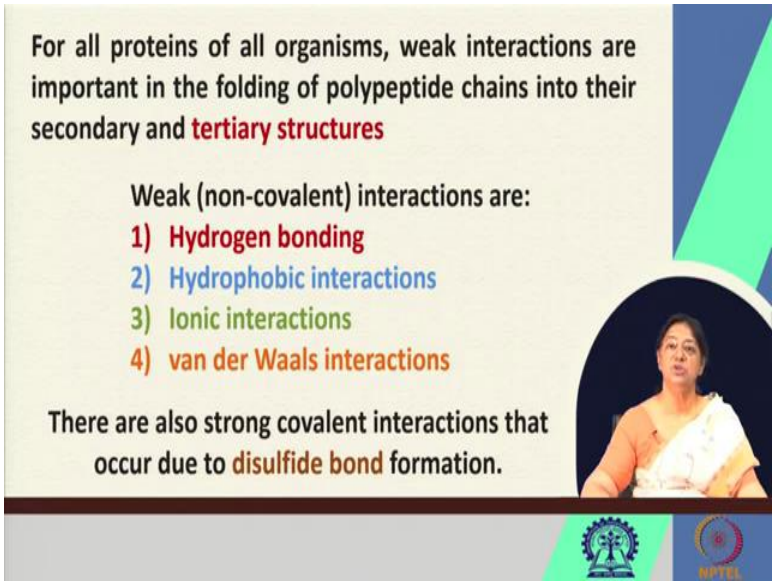
For the beta sheets, we have the anti parallel beta sheets and we found out the different levels of hydrogen bonding in terms of the secondary structural elements of the protein as they are called.

We then went on to look at the tertiary structure of the protein that formed the final native structure of the protein; in the sense, that we have the different interactions that we are going to look at in this lecture to study how they come together.

And in the quaternary structure, we learnt that there are different subunits of some proteins that come together to form a fully functional unit. Now what these proteins are. They are single polypeptide chains that have these amino acids in a specific sequence, that interact with each other in a three-dimensional arrangement.

So, the fact that this chain is coming together to form a compact three-dimensional structure is what is extremely important because the way it forms its three-dimensional structure is important. We are going to have amino acids at different parts of the protein come together to form specific geometry, which is important for the active site or for other interactions of the protein.

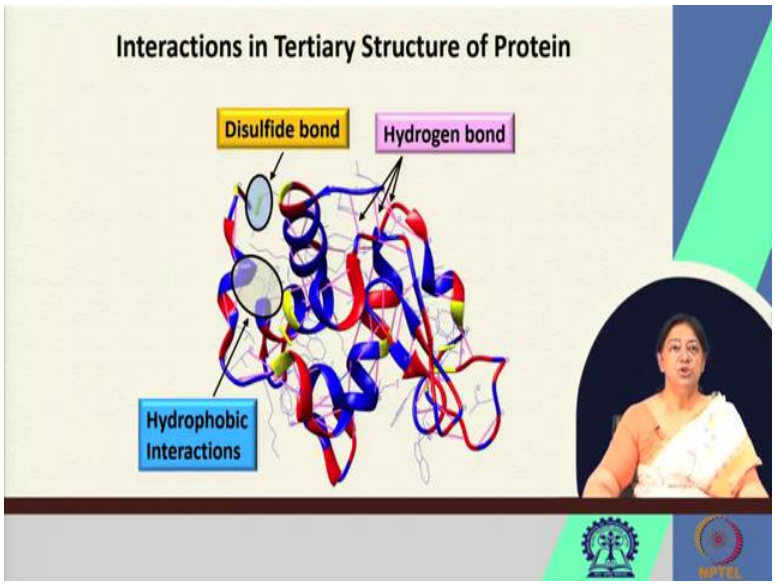
(Refer Slide Time: 02:34)



When we talk about the weak interactions, they are important in the folding of the polypeptide chains into their secondary and tertiary structures. So what are these? These are the weak non-covalent interactions. We know that the connection is through the peptide bond and the other covalent bond that we have is the disulfide linkage.

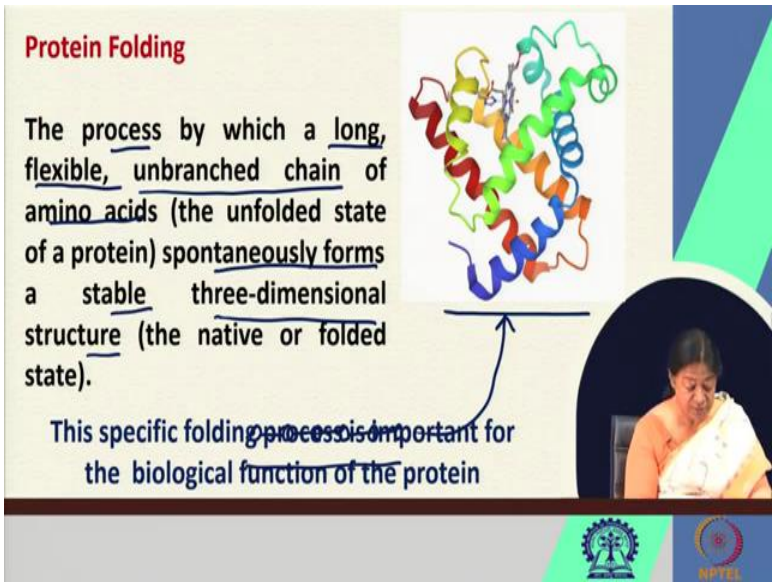
So we have hydrogen bonding, hydrophobic interactions, ionic interactions, van der Waals interactions. And of course, we have the disulfide linkages, that bring the different parts of the protein together in the formation of another type of covalent bond that is present.

(Refer Slide Time: 03:09)



So, when we look at these different types of the different structures that are possible, our idea is now to see how we can understand them. The disulfide bond we know is a formation of the cysteine-cysteine residues that form the cystine bond. We have hydrophobic interactions, we have hydrogen bonds. Hydrogen bonds we have seen between specific residues in the alpha helix, specific strands for the parallel and the antiparallel beta strands.

(Refer Slide Time: 03:52)

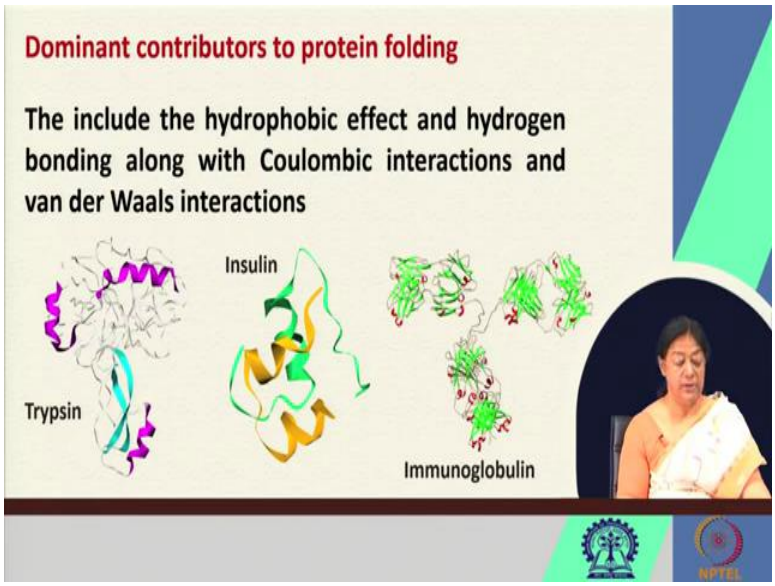


Now, we are going to look at the specific types that bring about what is called protein folding. What is protein folding? We have this long chain of amino acids that are linked to each other by the peptide bonds. This forms a three-dimensional structure. There is a process involved and this process is called protein folding. What is this process? It is the process by which the long, flexible, unbranched chain of amino acids, that is in the unfolded form, will spontaneously form

a stable three-dimensional structure, that we refer to as the folded form or the native form of the protein.

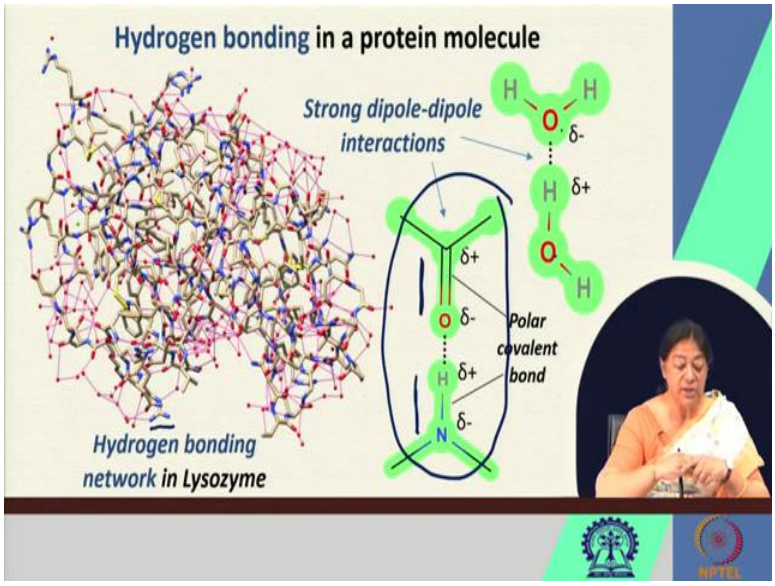
This is important because for this protein to function, it has to fold in this fashion. So, we are going to look into the process of protein folding; but before we do that we have to look at the specific interactions. The specific folding or the specific interactions that we observed are important because the folding will ultimately decide the biological function of the protein.

(Refer Slide Time: 05:12)



So the dominant contributors to protein folding are the hydrophobic effect, hydrogen bonding Coulombic interactions and van der Waals interactions.

(Refer Slide Time: 05:26)



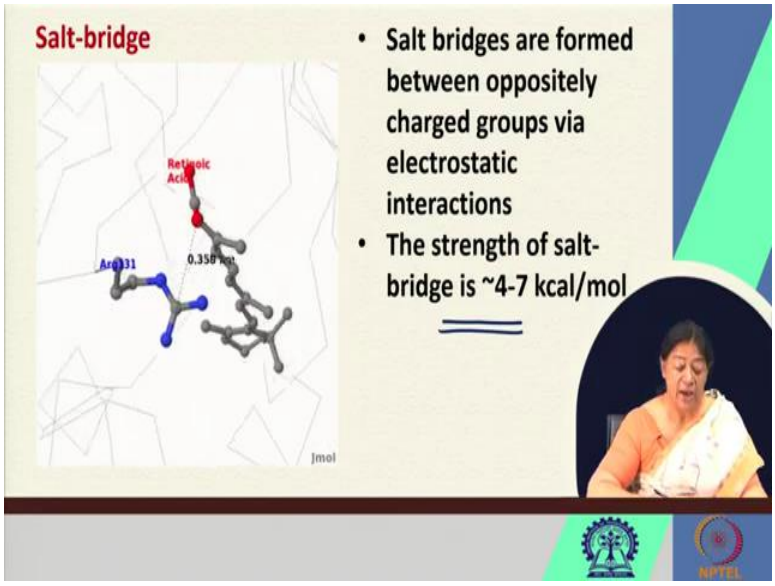
Now hydrogen bonding; we looked at hydrogen bonding when we considered the peptide bond. We know that the protein backbone contains the C = O and the NH. Every amino acid is interacting with the next amino acid through a peptide bond. And what is this peptide bond? It is the C = O NH. When we have specific interactions like this between the C = O and the NH, we have a specific hydrogen bond formation.

A hydrogen bond formation we know is between electro negative atoms, where we have a hydrogen connected to one of them. There is a strong interaction between the electro negative atom of one and the hydrogen of the other, giving rise to hydrogen bonds. Now these hydrogen bonds are important in proteins not only in the backbone, but also in the side chains.

There are specific polar amino acid side chains that are involved in hydrogen bonding. So a knowledge of the amino acids is important to understand which one of these can actually be involved in hydrogen bonding. So, when we look at the hydrogen bonding network say for a protein like lysozyme, we will see that this network involves not only the backbone atoms; but also side chain interactions.

This would be a side chain interaction, where we are looking at contributions not only within the protein itself, but also with the surrounding water molecules that can be involved in a hydrogen bonding network with the proteins. So, this important network is essential for the protein to remain stable in its three-dimensional framework.

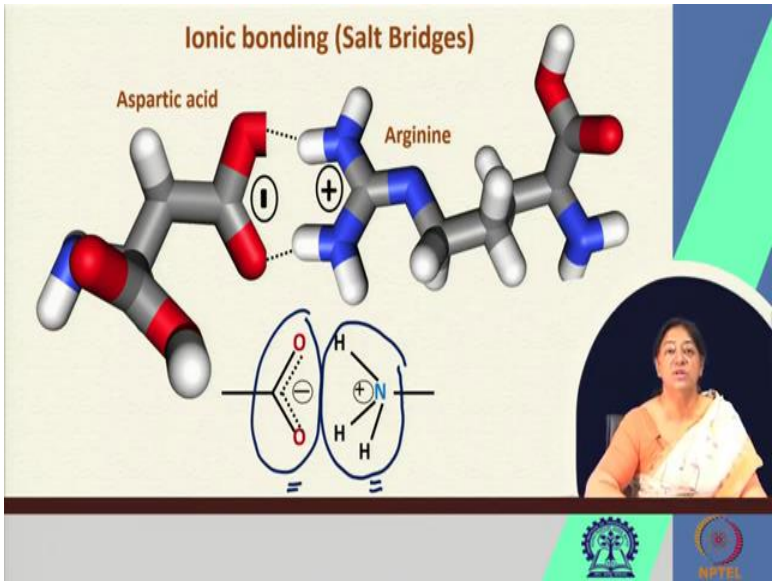
(Refer Slide Time: 07:28)



The salt bridge as the name implies, are formed between oppositely charged groups by electrostatic interactions and we will look at the relative strength of each of these types of interactions as we go on in the course, in looking at protein stability which will be part of our next module.

Now, when we look at salt bridges, we look at a positive charge and a negative charge. So, [refer to slide] in this case we have a protein that has an arginine residue and a retinoic acid which is a ligand and we see the possibility of a formation of a salt bridge that brings about a stability.

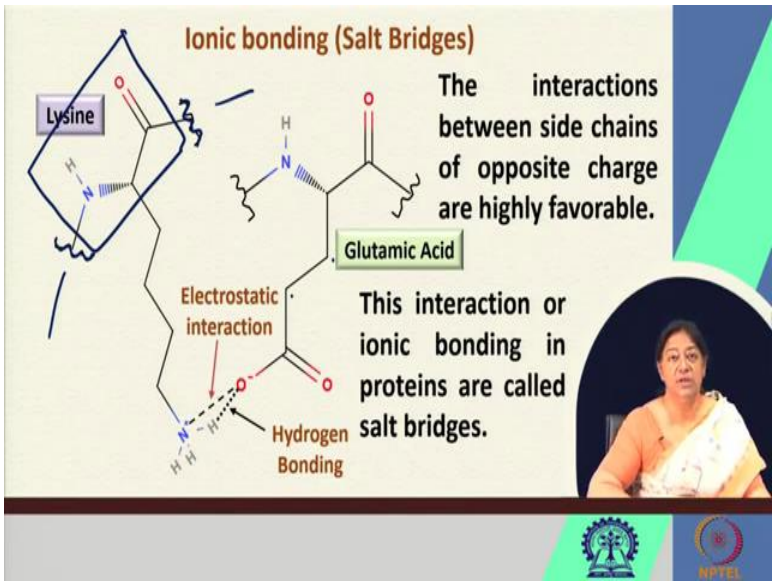
(Refer Slide Time: 08:11)



So, this is important in understanding how even amino acids of different charge can come together in an electrostatic interaction or electrostatic attraction that is going to bring different parts of the molecule together or also interact in a manner that is going to be important for the protein to maintain its three-dimensional structure.

When we look at an aspartic acid residue or for that matter for any such residue; the example of aspartic acid and arginine is given [refer to slide], we have an NH_3^+ and a COO^-. So, when we have an NH_3^+ and a COO^- again we have the possibility of ionic bond formation, because we have opposite charges that are going to interact with each other.

(Refer Slide Time: 09:09)

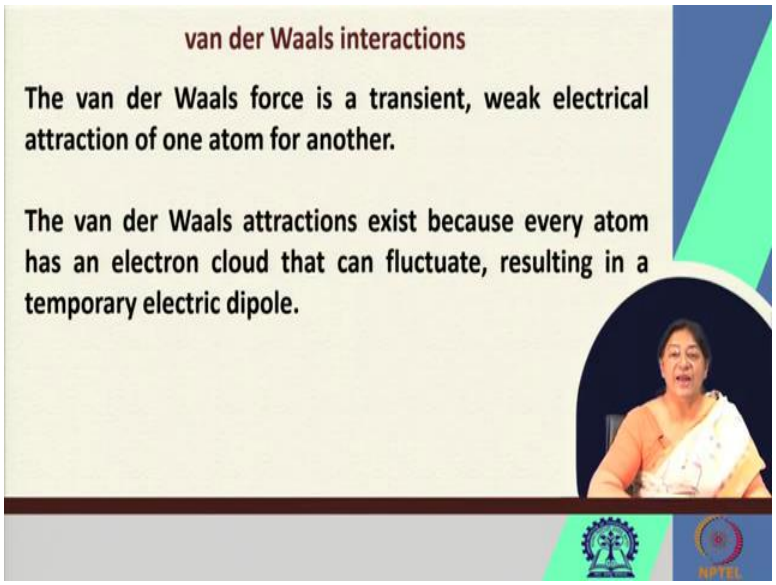


So we have favorable interactions between the opposite charges and another example would be lysine. When we have a lysine example, [refer to slide] we have these wiggly lines, which usually mean that the rest of the protein follows here and what we have is the CO and the NH.

This is where we have the connectivity, the peptide bond. So, this is a lysine residue and we can have it connect to a glutamate acid residue. How do we know this is glutamate? We know that there is a CH_2 here [refer to slide] and a CH_2 here. So, this is a glutamic acid.

So, we have to know how to identify the amino acids and also know the type of side chain that they are involved in and what kind of interactions are possible between the specific side chains. When we are look at the side chain of lysine and glutamic acid, we see that there is a possibility of an electrostatic interaction because of the charges, the negative charge and the positive charge. In addition to that we can also have hydrogen bonding possibilities. We have a mode by which we can have a strong interaction and this helps in the stability of the protein in its overall native's confirmation.

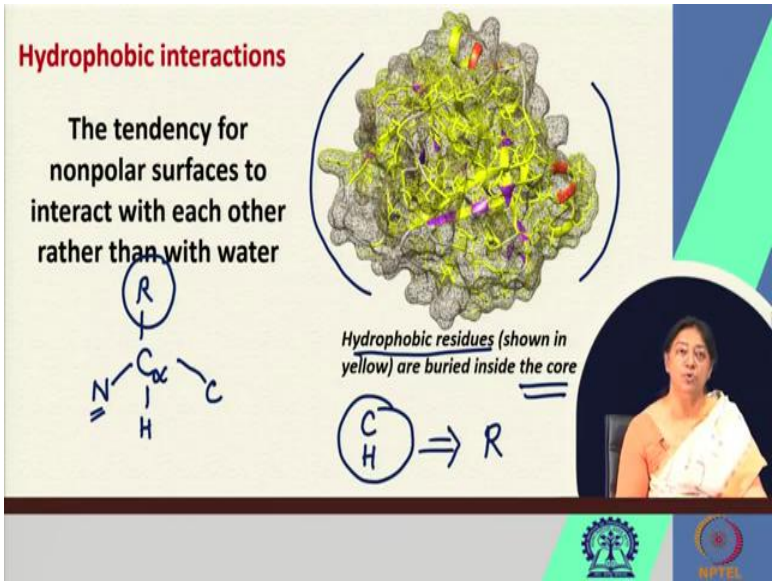
(Refer Slide Time: 10:49)



The other information that we need to know is van der Waals interactions. Now, in van der Waals interactions, we know that every atom has an electron cloud around it. This electron cloud is constantly fluctuating and because of this there is a temporary electric dipole that is formed.

Now because of this temporary dipole, there can be favored interactions between the atoms. They are weak transient electrical attractions from one atom to the other that give rise to what are called these van der Waals interactions, that even though weak in nature they can accumulate, because there can be a large number of such van der Waals interactions in the protein.

(Refer Slide Time: 11:35)



Now what do we see in the hydrophobic interactions? There is this tendency or preference for the non-polar surfaces to remain in contact with each other rather than the polar solvent water. So, in

this case what happens is if we have specific side chains, these specific side chains do not have an oxygen and a nitrogen associated with them.

We have to remember that when we are talking about the side chains in this case, we are talking about the additional attachments to the Cα because each amino acid is always going to have its N terminus, its C terminus and the hydrogen, and to this is going to be attached an R group. Now this R group is what we are interested in.

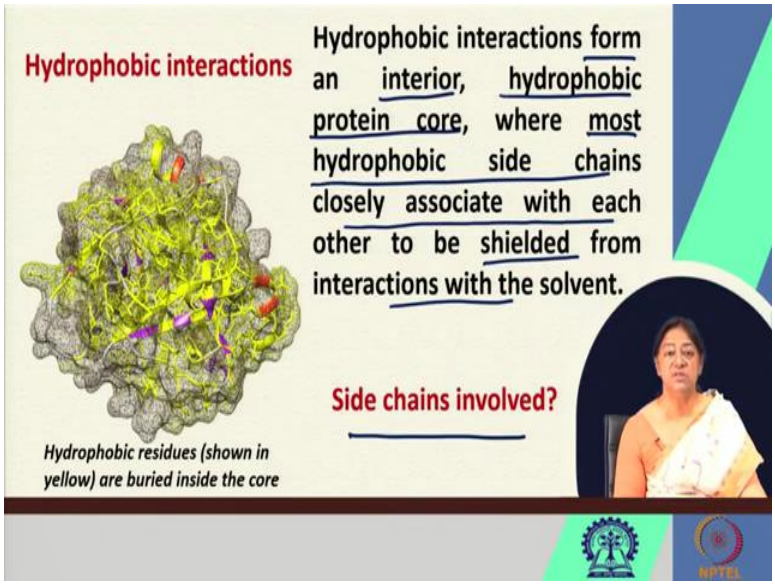
So if this R group happens to have carbon and hydrogen in it, then no electrostatic interactions are possible. No hydrogen bonding is possible because the side chain does not have an oxygen or nitrogen.

So what happens is, we have these hydrophobic residues that interestingly for the globular proteins, would tend to remain away from the surface of the protein. What happens when we look at a regular globular protein and when we look at all the hydrophobic residues.

We have to identify which are the hydrophobic residues and that is something we have to know; which ones are the side chains that have carbon and hydrogen present in their R group.

So these, [refer to slide] are the ones that are going to be hydrophobic in nature and they would preferably interact with each other rather than with the solvent. So then what would happen? They would tend to remain away from the surface and what we would see is we would see them bunching up together or forming what is known as the hydrophobic core of the protein, when it folds in this fashion in a regular globular protein.

(Refer Slide Time: 13:58)

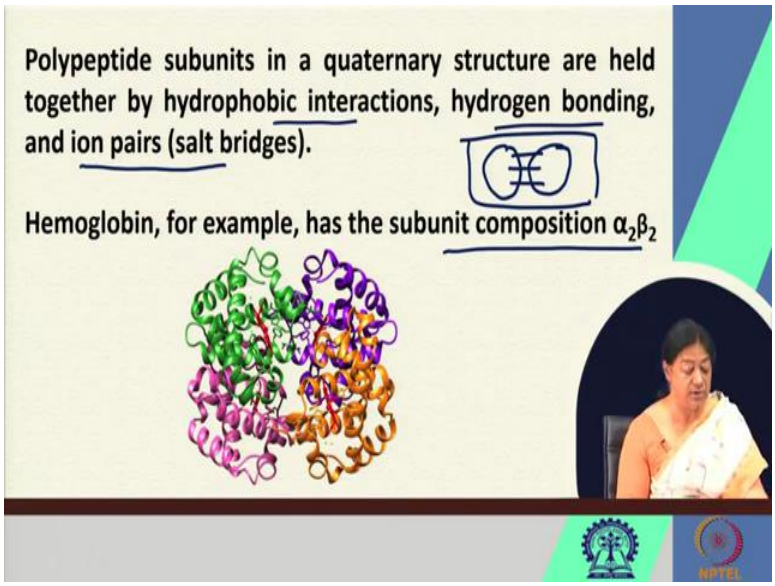


When we look at these hydrophobic interactions, we realize that because of the specific side chains that are involved, they are going to interact in a manner that is going to form the interior hydrophobic protein core. Here most of the hydrophobic side chains remain closely associated

with each other to be shielded from interactions with the solvent and with the knowledge of the different types of amino acid residues, looking at which side chains can be involved in what type of interactions, that is extremely important.

The basic amino acid residues and the acidic amino acid residues can be involved in ionic or electrostatic interactions. The polar type that have an oxygen and nitrogen to it, would preferably be involved in hydrogen bonding type interactions and the hydrophobic ones would be the ones that have the C and the H in them. So these would preferably remain in the hydrophobic core of the protein.

(Refer Slide Time: 15:09)

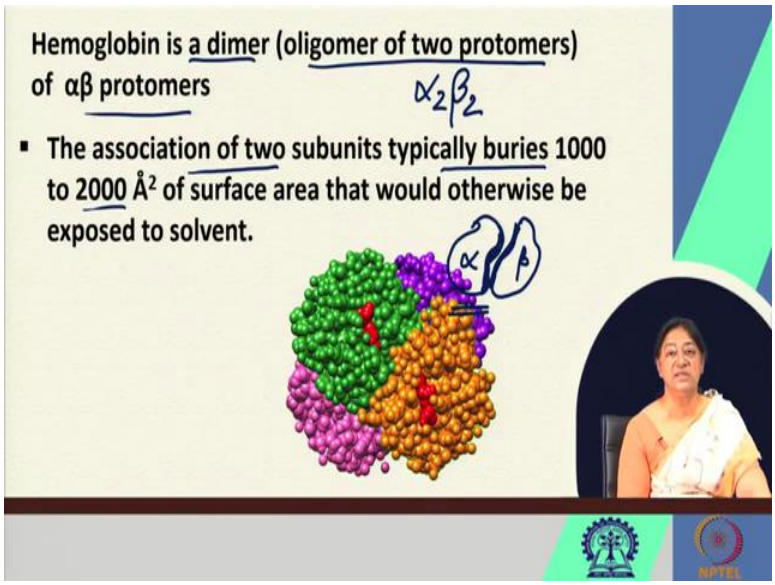


So, when we have these polypeptide subunits in the quaternary structures, they are held together by the hydrophobic interactions, hydrogen bonding and ion pairs. Now, that we have the folded compact structure and in a regular globular protein, we know that we have the hydrophobic core and the rest of the amino acids on the surface. Now what happens if we have a dimeric protein?

So we have two tertiary structural units; both in their subunit form, two monomeric units come to form a dimer. Our question is, what is connecting these two together and why do they remain as a unit like this? What are the interactions that are involved in the subunits coming together?

For example, when we look at hemoglobin, we have a subunit composition, as we saw in the last lecture, $\alpha_2\beta_2$. So, why is it not that the alpha and the beta are disintegrating? What is holding these together is also important.

(Refer Slide Time: 16:16)



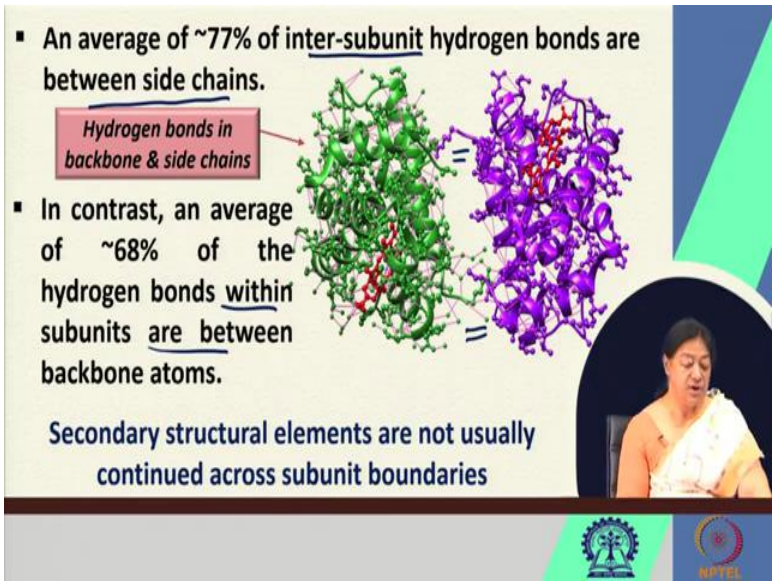
So we have a dimer, an oligomer of two protomers; these are αβ protomers. In the sense that we have the α_2 and the β_2. Then we have the association of two subunits.

There is a certain terminology that we call the surface area. So, the surface area of the protein or of that specific subunit is lost when it interacts with the surface of another subunit. In a normal case, this part [refer to slide] would be exposed to the solvent, so would the darkened part on the other chain.

Say we are looking at an alpha and a beta. The alpha chain part would have been exposed, the beta chain part would have also been exposed. But when they come together, the association can typically, what is called bury around 1000 to 2000 Angstroms of surface area that would have been otherwise exposed to the solvent.

Now, the fact that we have these two units coming together means there are specific interactions that are going to hold these together in a specific formation. And what is observed in the oligomeric unit's formation; we preferably see hydrophobic interactions to a greater extent in these types of interactions.

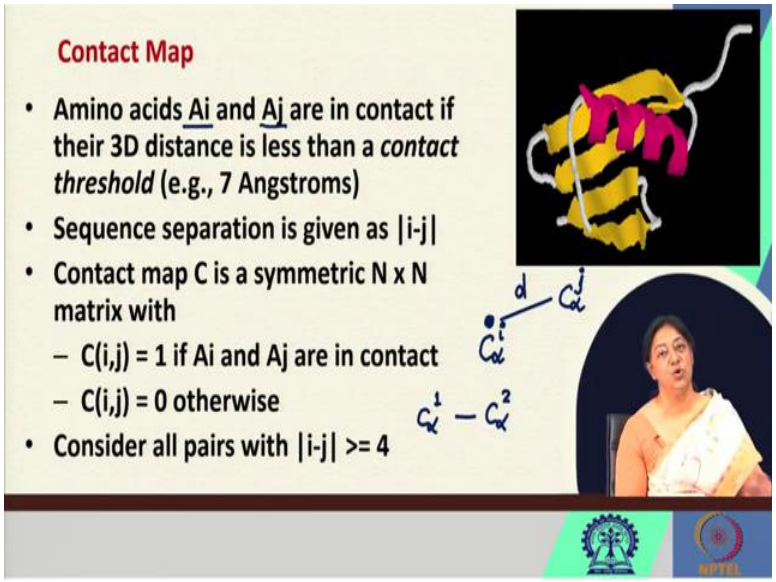
(Refer Slide Time: 17:52)



So an average of 77 percent of inter-subunit hydrogen bonds are between the side chains. We have to hold the subunits together and the hydrogen bonds generally within the subunits, are between the backbone atoms and the secondary structure elements that we observe, are not generally carried over the subunit boundaries.

These would be two subunits that would be held together in a fashion that could include hydrogen bonds, in addition to preferable hydrophobic interactions. Now that we have a knowledge of a three-dimensional structure of a protein, an idea of what the forces that are holding these together could imply, we now look at another aspect of a protein structural sense in understanding what a contact map is.

(Refer Slide Time: 18:47)



Now, to have a contact map of a protein, what we need to know here is, we need to know what are called the coordinates. What do we mean by coordinates? When we have a structure, for example, if we look at the specific structure here [refer to slide], we know that we have specific side chain atoms.

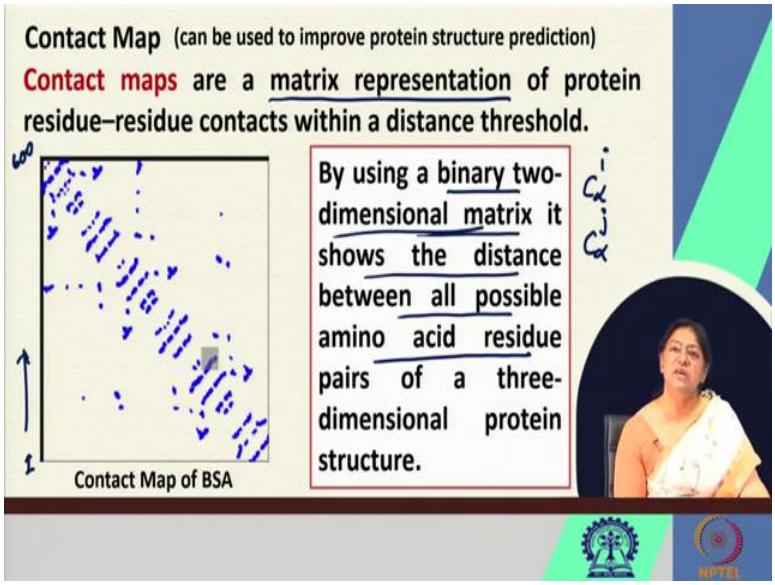
Each of these side chain atoms that are drawn in circles here, spheres here, they have a position in three-dimensional space. So, if we can look at the three-dimensional position of each of these atoms for the specific side chains or for the specific amino acids, we can construct what is called a contact map.

Now, the reason we try to look at a contact map is apparent later when we try and understand structural aspects. So, if we look at a specific amino acid and just look at the Cα of this specific amino acid and try to find out the distance between another Cα from somewhere far in the protein, there is a specific distance d.

So, the idea in the construction of the contact map is to look at all Cα of i connect all Cα of j and find the distances. We have say Ai and Aj, these are two different amino acids that are connected in contact if their 3D distance is less than a contact threshold.

So what would that mean? That means if we consider Cα of residue 1 to be close to Cα of residue 2, because they are in the sequence one after the other, we are looking at specific ways in which we could look at distances between the Cα carbon atoms in an identification of a structural aspect, in what we call a contact map.

(Refer Slide Time: 21:10)

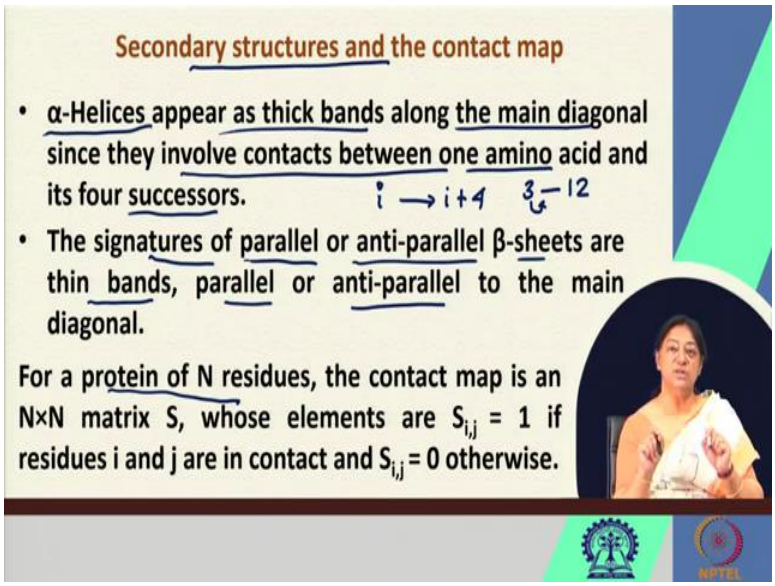


The contact map will be described in a later slide. So, what we have is we have a matrix representation. Where do we get this matrix representation from? We have a binary two dimensional matrix which shows the distance between all possible amino acid residue pairs of a three-dimensional protein structure.

So what this means is, that we have the Cα of residue i and we have the Cα of residue j. What do we have of the Cα of residue i? We have the x, y, z coordinates in three-dimensional space; for Cα j, we also have the three-dimensional coordinates in space and what we can calculate is we can calculate the distance from Cα i to Cα j.

Now, if we were to do this exercise from every atom to every other atom, this would be say from residue 1 to residue BSA is a very large protein, say close to around 600 residues. This would mean that we have a contact map between the Cα of 1 to Cα of 2, Cα of 1 to Cα of 3 and we create what we call this two dimensional binary matrix.

(Refer Slide Time: 22:36)



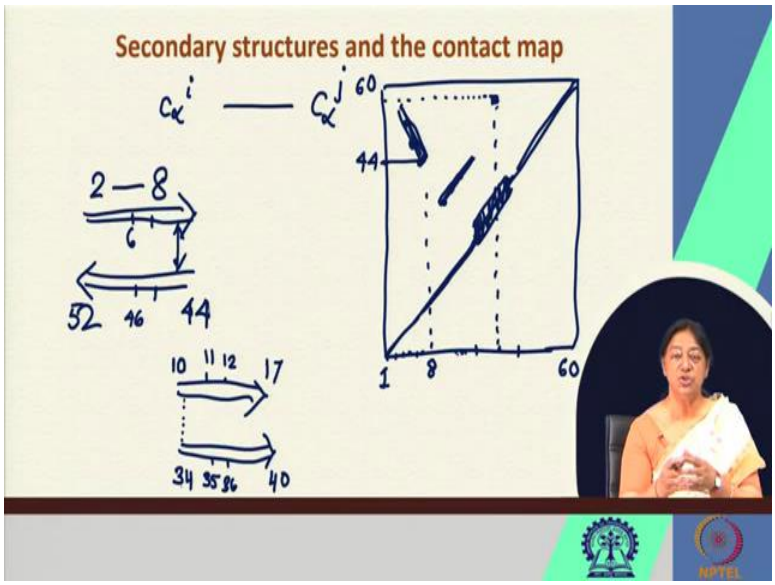
Now, what is the information that we can get from this matrix, this is important. Interestingly, we can look at secondary structures and the contact map and what we get from these secondary structures is we have alpha helices that appear as thick bands along the main diagonal because they involve contacts between 1 amino acid and its 4 successors.

What do we mean by that? We mean that we have a hydrogen bonding, we know from i to i + 4. Now because we have the contact structure in the alpha helix, it would mean that if residue number say 3 to 12 were to form an alpha helix, what would we expect?

This means that 3 is close to 4, it is close to 5, it is close to 6, it is close to 7 and there is a hydrogen bonding between 3 and 7; meaning, that 3 and 7 are going to be close to three-dimensional space. Also there are signatures of parallel or anti-parallel beta sheets that are thin bands, that are parallel or anti-parallel to the main diagonal.

So, essentially what we are doing is for a protein of N residues, it means we would have an NxN matrix and the elements would tell us whether the two parts of the protein or the two residues that we are looking at, are close enough together or not.

(Refer Slide Time: 24:12)



Let us look at what this means. Essentially what we are doing is we are taking the Cα of i and we are calculating the distance to Cα of j. What we do is we create a matrix, we create a matrix and we say that if we have say 60 residues. We look at 1 to 60 here [refer to slide] and we also look at 1 to 60 along the y axis.

What we have in this case is all along the diagonal, we see a thick line because 1 is close to 2 which is close to 3, 2 is close to 3, 3 is close to 4 and so on and so forth. So, we would expect a line and we also have to know that the upper and the lower triangle of this matrix is going to be the same because if residue number 3 is in contact with 7, then obviously, 7 is in contact with 3.

So now say if we have a helix around this region. Now, what do we know is, we know that helix are contiguous in nature. So, if we have a helix from around say 25 to 35, we would expect 25 to be close to 26 which is natural. But we also expect a hydrogen bonding between i and i + 4, which means there is going to be a thicker band along this helix region telling me that residue number say 25 is in contact with 29.

A thick region along the diagonal would imply that there is an alpha helix, because they are close together in three-dimensional space and we have a specific hydrogen bonding pattern that is bringing them close to each other. Now say we have a parallel beta sheet [refer to slide], where of the strands of which are say from 2 to 8 in this direction and again there are anti-parallel strands, say 44 to 52 in this direction, and we know we have hydrogen bonding patterns between this.

Now say 8 is close to 44. How would we plot this? We would have 8 say somewhere around here that is close to say 44, then we would have 7 that is close to 45. So then this would be 44 and this would be 8. Now we say 7 because we are one back here and 45 would be up here.

Then, if we look at number 6, it would be in connection with 46. So we go back a step here, further up here, what would we see? We would see a line that looks like this. So, this would be

an anti-parallel beta sheet. Now, say we have a parallel strand. We have a parallel strand that takes us say from 10 to 17 and we have a parallel strand that is from 34 to say 40.

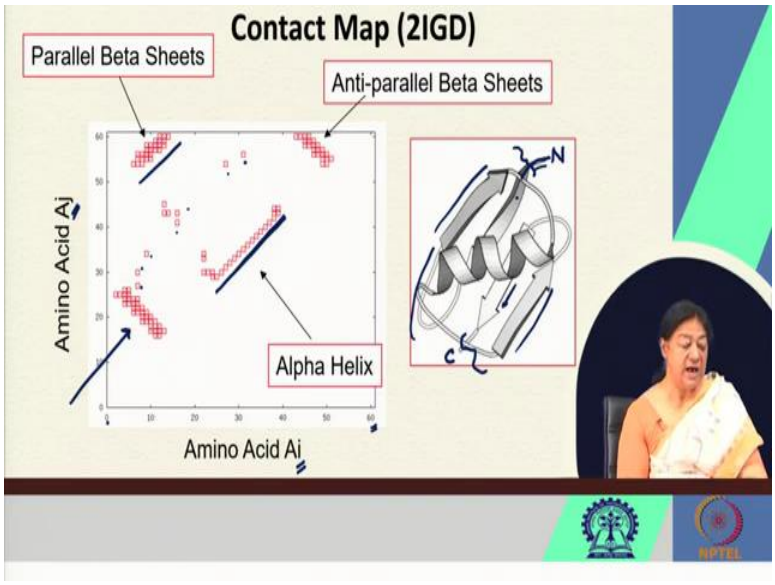
So now, when we are looking at hydrogen bonding patterns here, we would have 10 close to 34. 10 would be here and 34 say somewhere here; then we have 11 close to 35. So, we would look at 1 here and 35 which would be another point here, then we would look at 12 and say 36.

Then we would be again one on the right and see a parallel connectivity. So, when we look at specific regions of the protein in a contact map from the three-dimensional coordinates that is the x, y, z information that we have, we can identify specific secondary structures on the contact map.

For example, if we have now an extra dot here, this extra dot may give me some information. It would be 58 up here and 40 down here. Then we would say 40 and 58 are close together in three-dimensional space. Now the question may arise as to why are these two particular residues close to one another?

It could be that it forms a disulfide bond; it could be that there is an electrostatic interaction. So, from a knowledge of the contact map, it is easy to identify specific, not only secondary structures in the contact map, but also information about which parts of the protein are close to one another. This is particularly important when considering protein dynamics as to understand which regions are changing in three-dimensional space, when we are allowing a changing of the coordinates due to interactions or due to ligand binding.

(Refer Slide Time: 30:41)



While looking at the contact map [refer to slide] for a specific protein here, if we look at these two strands, we can see the N terminus. Why do we know this is the N terminus? Because this is where we cannot see the protein, but we can see that the arrow ends here, so this has to be the C terminus of the protein. We have the alpha helix and then let us try and trace this. So, here is

where it begins, then we go down this way around this to form this beta strand. Then we have the formation of the helix and then, we have this beta strand that then forms the other beta strand.

So, what we do have is we have two beta strands. This is an antiparallel beta strand, here also we have an antiparallel beta strand and we have an helix. So, now if we look, now we understand what amino acid Ai and what amino acid Aj means. Now what are these points on the graph or what are these red boxes that we have here?

The red boxes are indicating that these residues that are located here are close to each other in three-dimensional space. We can now identify the regions of the protein that are close to one another in the specific types of beta sheets or the beta strands.

So what we have is, if we look at this particular one, this one that starts and the one that ends, these strands are in the same direction giving us a parallel set and what we have in the other cases here, for this particular beta these two beta strands are anti parallel and so are these two beta strands, but the one in these two in the middle are parallel.

This is very clear from the structure of the protein in three-dimensional space and we also see the helix and what we did learn for the helix was along the diagonal, we are going to have an extended portion and this portion will correspond to the alpha helix. Why? Because we know that we are going to have this close in three-dimensional space because there is specific hydrogen bonding for the regular alpha helix say between i to i + 4.

So, what we have here is we have the parallel beta sheets, the anti-parallel beta sheets and the alpha helix. We have one antiparallel beta sheet here, one antiparallel beta sheet here, a parallel beta sheet shown here and the alpha helix.

If we did not have the structure on the right, from the information present here, we could go from residue number 1 to residue number 60 and identify where we have the alpha helix in the protein, where we have the antiparallel beta sheets in the protein and where we have the parallel beta sheets in the protein.

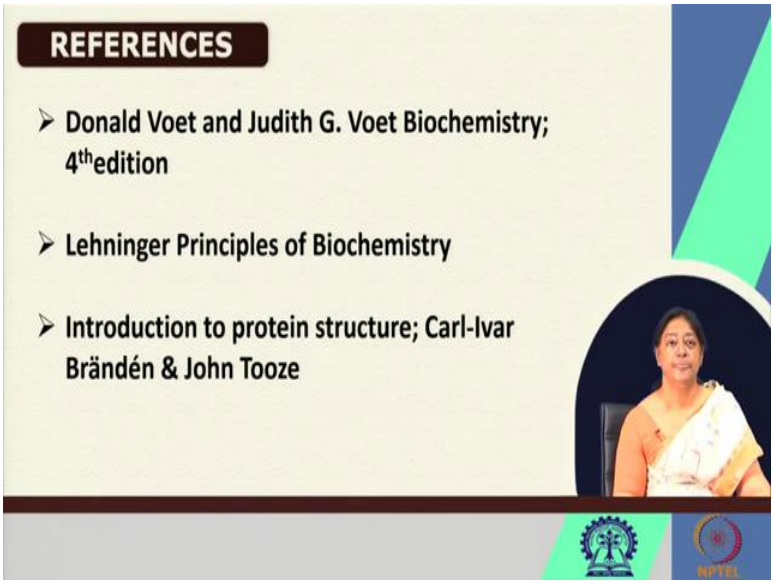
In addition to that, we have additional information about which parts of the protein are close to each other in three-dimensional space. So the information that we can get from the contact map is interesting, giving us an idea about how we can plot the information given that we know the structure. What is the structure? The three-dimensional coordinates of all the atoms present. But what we plot here for convenience is the Cα coordinates and we have the specific distances associated with it.

(Refer Slide Time: 35:11)

So, what we looked at is we looked at general protein folding, the process that brings the amino acids that are linked in the primary structure to form the secondary structure, then the tertiary structure forming the native three-dimensional unit of the protein. We have the hydrogen bonding that we looked at. And what we realized is we need to know which amino acid residues can actually be involved in the hydrogen bonding that is present.

The salt bridges, where we would look at electrostatic interactions and what kinds of residues again could be involved in these salt bridges; van der Waals interactions and hydrophobic interactions. So these are the concepts that we looked at. We will look at further interactions and a further understanding in our next lecture.

(Refer Slide Time: 36:05)



Thank you.