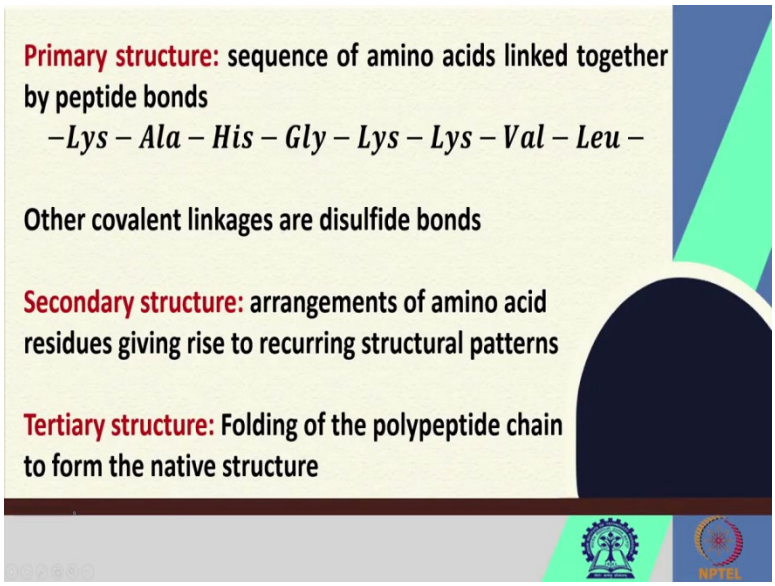
Fundamentals of Protein Chemistry
Prof. Swagata Dasgupta
Department of Chemistry
Indian Institute of Technology, Kharagpur

Module - 02
Protein Architecture
Lecture - 10
Discussion Class

In our discussion class for module 2, we will revisit some of the topics that were covered in this module and we will look at specific problems related to those topics.

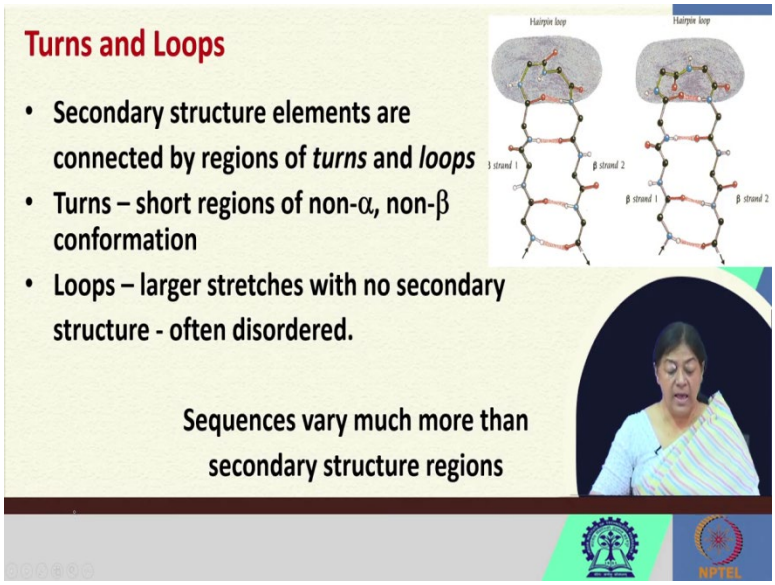
(Refer Slide Time: 00:23)



We begin by understanding or looking back at the primary structure, which we knew was the sequence of amino acids linked together by peptide bonds, followed by a knowledge that the other covalent linkages are the disulfide bonds. The secondary structure gave us the arrangements of amino acids, giving rise to recurring structural patterns with specific geometry associated with them.

The tertiary structure followed with the folding of the polypeptide chain to form the native structure, and we looked at the forces involved. In the next module we will be also looking at the specific thermodynamics involved in this folding process.

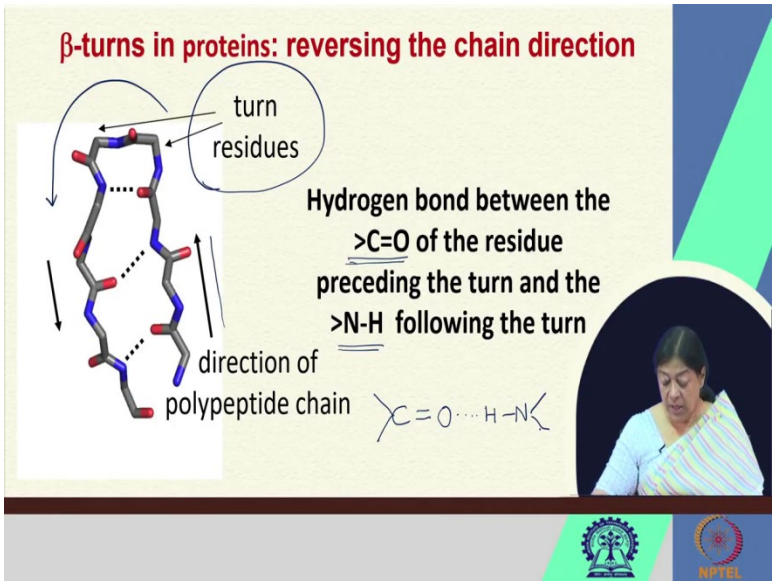
(Refer Slide Time: 01:04)



There are linkages in the tertiary structure, we know that there are secondary structures also, the alpha helices, and the beta strands that form the beta sheets, but these are linked together. These are linked by certain linkers that we refer to as turns and loops. Now these turns do not necessarily have any geometry associated with them, but nevertheless have specific hydrogen bonding patterns.

So, these secondary structural elements are connected by these regions called turns or loops. And they are short regions of non-α or non-β conformation, indicating that they would not fall in a specific region of the Ramchandran plot. The loops are larger stretches with no secondary structure and they are usually disordered, in the sense that they do not have any geometrical constraint to them. And the sequences here vary much more than the regular secondary structural regions.

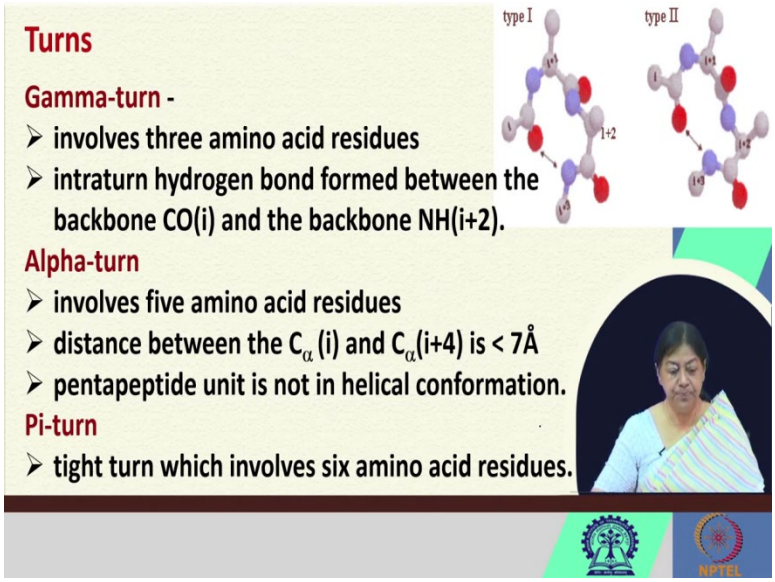
(Refer Slide Time: 02:22)



If we look at the different types that we know, there are turn residues as we would call them. And if we look at [refer to slide] say these two strands that are connected by these turn residues that have been shown here, what we see is these turn residues that are involved here have specific characteristics in terms of their geometry. What is happening to the direction? If we look at the direction of the polypeptide chain there is a reversal in three-dimensional space.

So, this reversal is caused due to this turn and there are oftentimes formation of hydrogen bonds between the C = O of the residue that is preceding the turn and the N-H following the turn. As we learnt in our discussion for the peptide bonds, this hydrogen bonding between the C = O and the N-H is very common in protein structures.

(Refer Slide Time: 03:29)



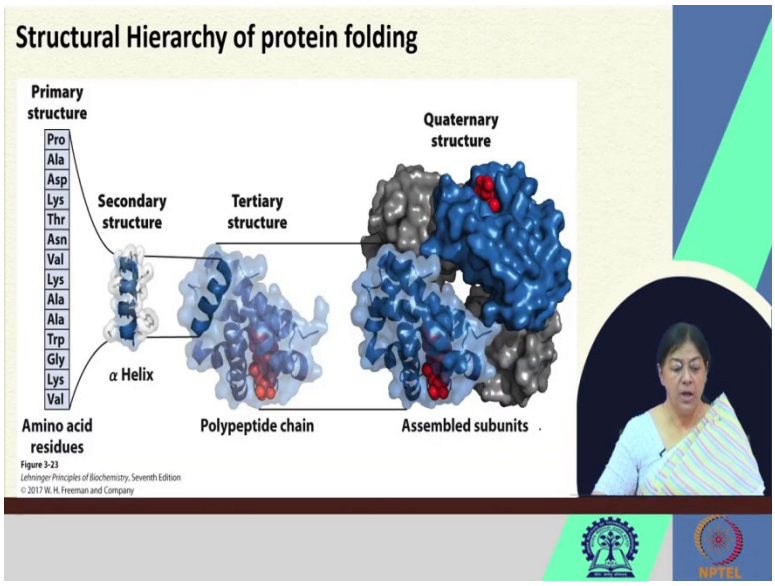
Now, the other types that we do see here are regular turns, different types of turns. For example, there is the gamma-turn that is characterized by an involvement of three amino acid residues and intraturn hydrogen bond and this bond again is formed between the backbone CO of the ith residue and the backbone NH of the i + 2.

Similarly, if we look at an alpha-turn this involves 5 amino acid residues and the characteristics of this specific alpha-turn is that the distance between the $C_\alpha(i)$ and the $C_\alpha(i+4)$ is < 7 Angstroms.

There are other examples also, sometimes we call them an omega loop because it's shaped like an omega. These loops and turns have different terminology. Another one is the pi-turn that is a tight turn which involves around 6 amino acid residues. These are linkers of the secondary structural elements and they do not have a very specific structure for themselves.

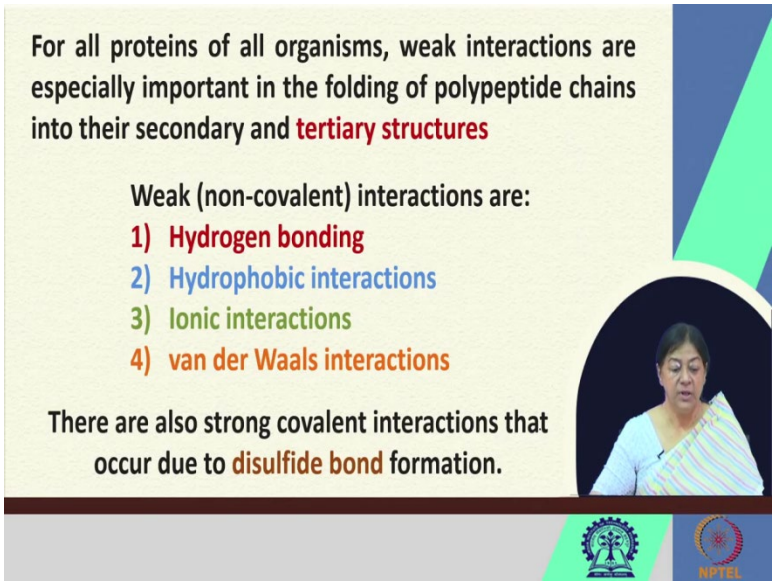
But nevertheless because of their reversal in three-dimensional space, they do have hydrogen bonding possibilities between the backbone C = O and the NH that bring the parts of the protein chain, the polypeptide chain close to each other.

(Refer Slide Time: 04:56)



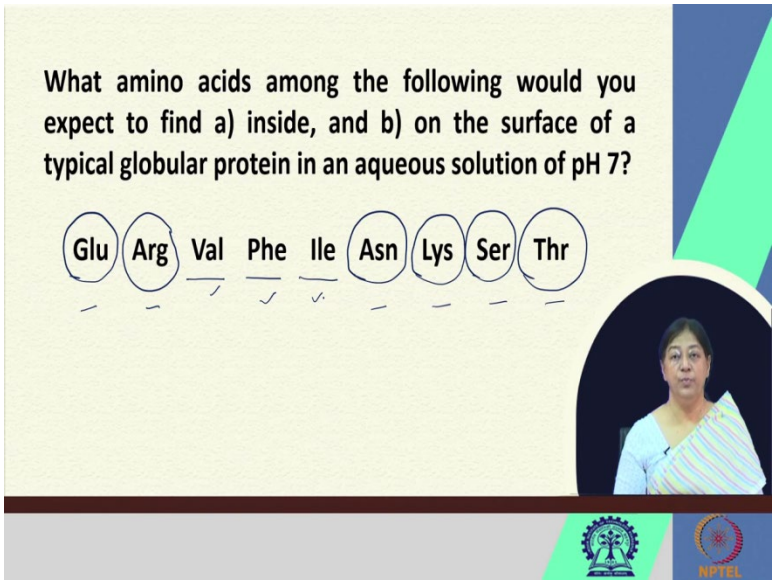
So, for the structural hierarchy we have the amino acid residues, the secondary structural elements going to the tertiary structure and the quaternary structure. Let us now revisit some of the associated issues that we can consider with a knowledge of what we have learnt in this module.

(Refer Slide Time: 05:16)



For all proteins of all organisms, we learnt that the weak interactions that are important for the formation of the secondary structure involved hydrogen bonding and then when we go on to form the tertiary structure, we have hydrogen bonding, hydrophobic interactions, ionic interactions, van der Waals interactions and the covalent disulfide linkage.

(Refer Slide Time: 05:39)



If we were to look at a specific topic and say what amino acids of the residues would you expect to find inside and on the surface of a typical globular protein in an aqueous solution of pH 7.

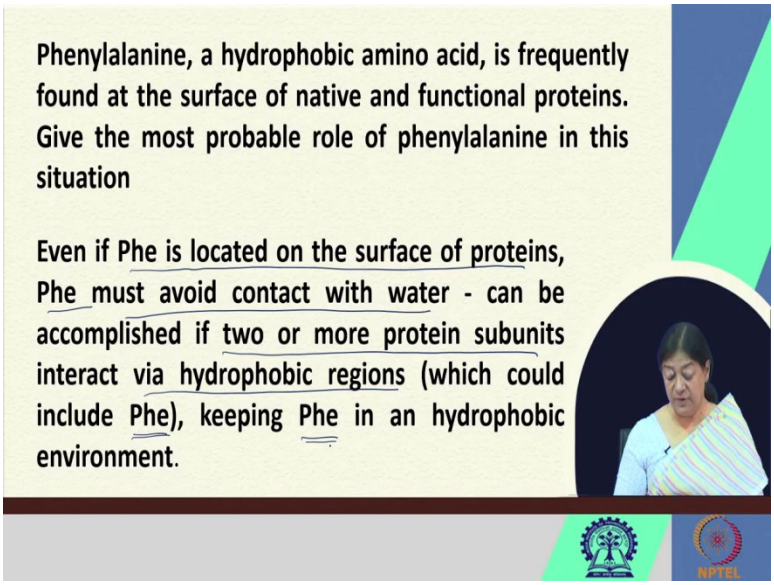
The fact that we have an aqueous solution or a polar solution indicates that for this globular protein the ones on the surface should be polar amino acid residues. And what do we mean by this? We mean that they should have a heteroatom in their side chain.

Now let us look at the specific amino acids. Glutamic acid we know is an acidic amino acid. It has CH_2CH_2COOH. Arginine is a polar basic amino acid that has the guanidine group with the nitrogens that can be involved in hydrogen bonding with water. For valine, phenylalanine and isoleucine, in all the side chains of these amino acid residues we have only carbon and hydrogen. They would be hydrophobic in nature.

Asparagine has an amide group. So it has a C=O NH_2 which is going to be involved again in hydrogen bonding. Lysine is the other basic amino acid that has an epsilon NH_2 group that can be involved in hydrogen bonding. And serine and threonine are small polar residues that also have an OH group attached. For serine we have CH_2OH and for threonine we have a beta branch at one with CH_3 and one with OH.

The ones that are circled here [refer to slide] are the ones that have in their side chains additional heteroatoms that could be involved with hydrogen bonding in water. So we would expect these to be on the surface of the protein. And valine, phenylalanine, isoleucine, having only carbon and hydrogen would preferably be inside the globular protein in the hydrophobic core as we discussed earlier.

(Refer Slide Time: 07:58)

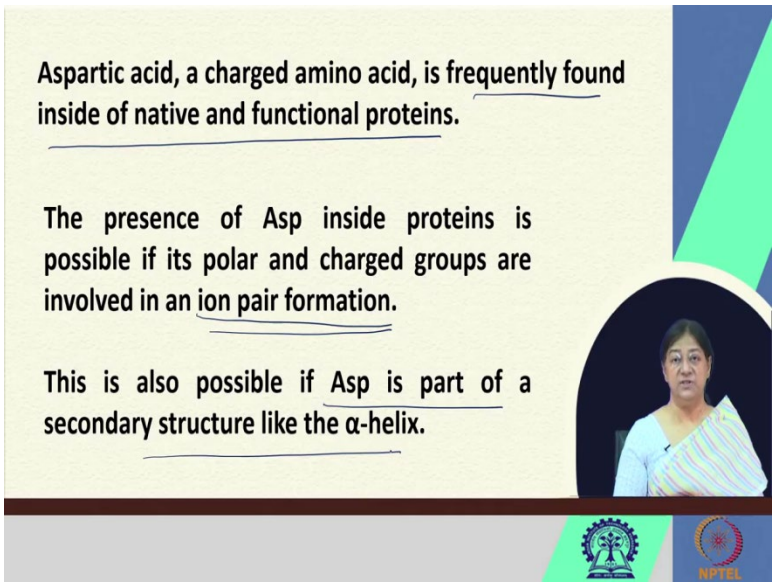


If we look at phenylalanine now, there is another interesting aspect of this. It is a hydrophobic amino acid residue. It is also frequently found at the surface of some native and functional proteins. What would be the most probable role in this situation? Now, when we looked at the quaternary structure we found out that there were different subunits that came together.

In a case where we do see a large predominance of phenylalanine on the surface it could so happen that even if phenylalanine is located on the surface of the protein, it should have to avoid contact with water because it is hydrophobic in nature. So how can this be achieved? This can be achieved if two or more protein subunits that have phenylalanine on the surface interact via

hydrophobic regions, (which could include this phenylalanine), thus keeping phenylalanine in a hydrophobic environment and allowing the association of the specific subunits.

(Refer Slide Time: 09:05)

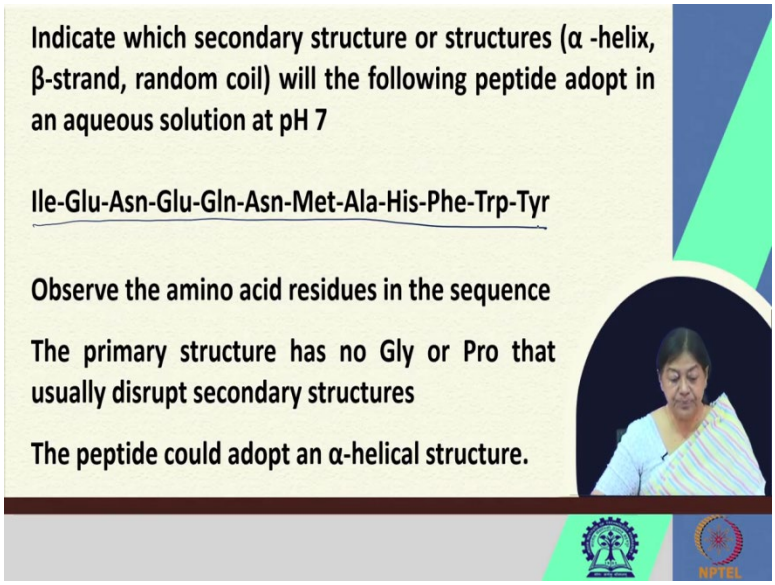


When we look at aspartic acid it is a charged amino acid, it is frequently found inside of native and functional proteins.

Now, if an acidic amino acid or a basic amino acid is found inside a native and a functional protein, then we expect this to have an ion pair formation involvement because the complete polar and charged amino acid will not remain safe inside unless it is balanced in its charge. So there has to be an ion pair because there will not be as many water molecules inside the native protein.

This is also possible if aspartic acid is part of a secondary structure like the α-helix, where it has to be involved in some specific interaction that is going to counter its charge.

(Refer Slide Time: 10:01)

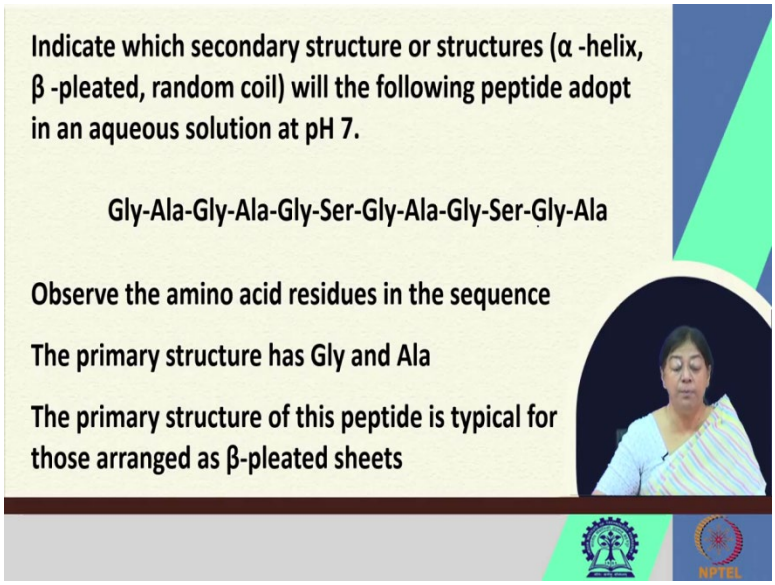


Now we would like to understand whether the secondary structure or structures are going to be adopted by this specific peptide that has been drawn here [refer to slide]. We have to look at the specific types of amino acids that are involved.

For example, we see it starts with isoleucine, then we have glutamic acid, asparagine, another glutamic acid, glutamine, asparagine and so on and so forth. So, there is a mix of hydrophilic and hydrophobic amino acid residues.

We have to observe the amino acid residues in the specific peptide sequence and then, we have to see whether the primary sequence has specific residues say for example, like glycine or proline, that actually disrupts secondary structures. But we see that there are no such glycine or proline present in this particular sequence. So, it is likely that this could adopt an alpha helical structure.

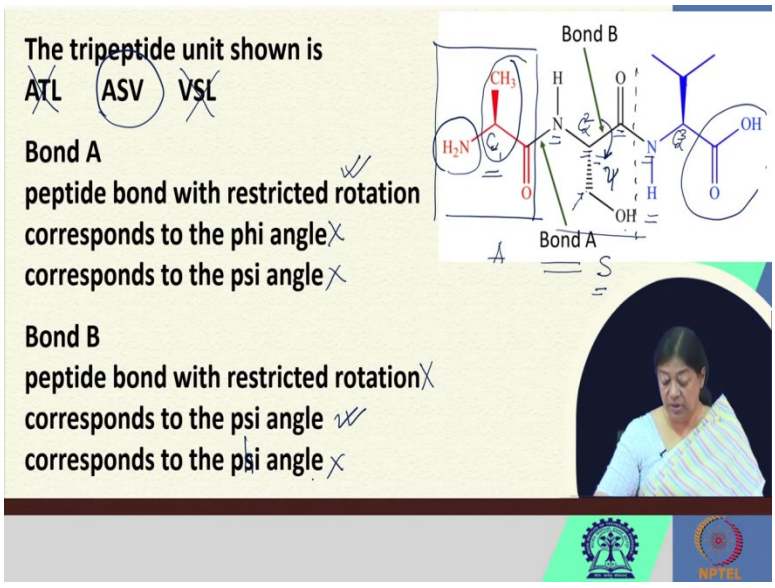
(Refer Slide Time: 11:12)



If we look at this specific secondary structure that this peptide could adopt, again in an aqueous solution at pH 7, and we try and observe the amino acid residues present in the sequence, we observe a large number of Gly and Ala. We know that Gly and Ala are the small amino acids that are present in the protein and the primary structure shown here has Gly and Ala present in the sequence. So this is likely one of those arranged in beta sheets, where this would probably belong to a specific strand.

So the identification of a specific amino acid sequence and then trying to determine what structure this could adopt, is a methodology that comes from the types that have been observed, where they are found and so on and so forth. This observation is important as to which amino acid residues are adopted in this sequence and what kind of a structure this may conform to in a aqueous solution at pH 7.

(Refer Slide Time: 12:25)



Now there is a tripeptide unit shown here [refer to slide]. We know that this is the amino terminus, we know that this is the carboxy terminus. What we observe here is this is the Cα of residue 1, this is the Cα of residue 2, and this is the Cα of residue 3. So the question is, whether the tripeptide unit shown is ATL, ASV or VSL? Now if we look at the first amino acid residue, we see that this is alanine.

If we now look at the second amino acid, we see that this is a CH_2OH, which would mean that this is serine. If this had another CH_3 attached to this carbon, it would have been threonine. So, this is serine. We have A, we have S and this is valine where we have CH_3CH_3 attached to the Cβ. So, ASV is the tripeptide that is shown here.

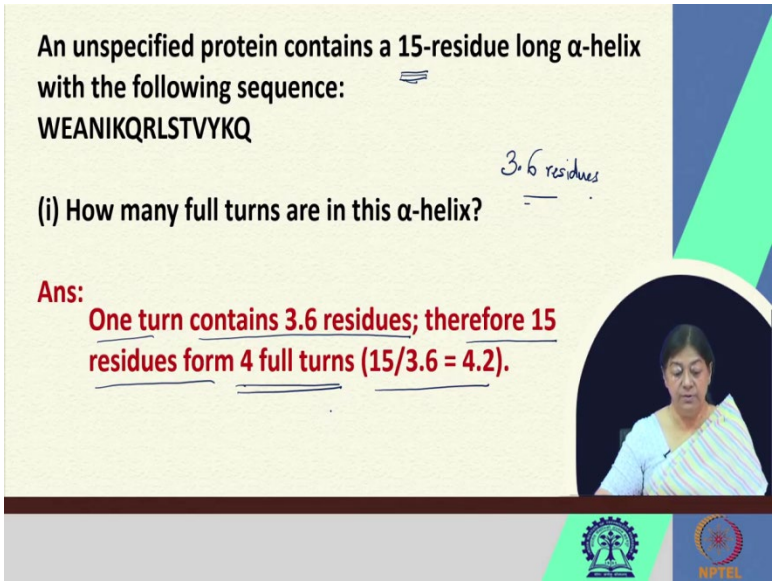
There is bond A shown here [refer to slide]. The next question is whether bond A is a peptide bond with restricted rotation or it corresponds to the phi angle or it corresponds to the psi angle. Let us look at the specific bond. We see that the amino acid marked in red is alanine.

This was the COOH of alanine and this is the NH_2 of the next amino acid serine. So the connectivity that we see for bond A is a peptide bond, because we see that it is connecting amino acid number 1 with amino acid number 2. So, bond A is a peptide bond with restricted rotation.

We know that this is the second unit because this is the C = O OH of the amino acid S and this is the NH of the third amino acid. So, what do we have here now? We are looking at bond B. It is not the peptide bond because it is within the amino acid side chain. Now the question is whether this is the psi angle or the phi angle.

Now when we look at rotation about the N, the Cα, C and this N we know that we need 4 of these atoms to define the dihedral angle. The N, Cα, C, N corresponds to the psi angle. So, bond B corresponds to the psi angle. These are types of identifications we might need to do for a specific peptide unit, identify the peptide bond, identify the dihedral angles.

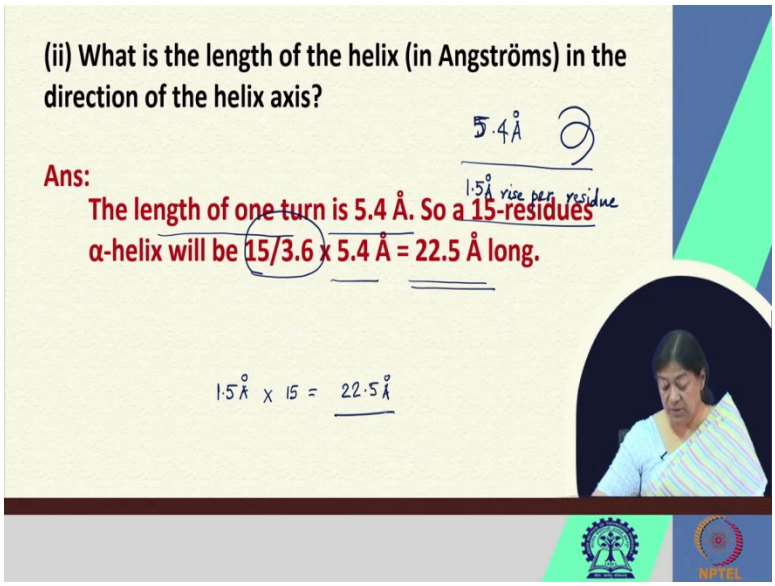
(Refer Slide Time: 16:13)



Now we have in this specific problem, a protein sequence that contains a 15 residue long α-helix with a sequence that has been shown here [refer to slide]. Here the question is how many full turns are in this α-helix? What we know is, we know that for the regular α-helix we have 3.6 residues per turn. So, if we look at this we have 3.6 residues per turn.

So how many residues do we have? We have 15 residues. What can we expect? We have one turn that contains 3.6 residues. Therefore, if we have 15 residues we would expect 15 divided by 3.6; that is 4 full turns of this particular helix. So the question is that how many full turns we would expect? Depending upon the length of the helix that we have and a knowledge that one turn is going to be 3.6 residues we can find out how many residues or how many turns there are for this particular helix.

(Refer Slide Time: 17:38)

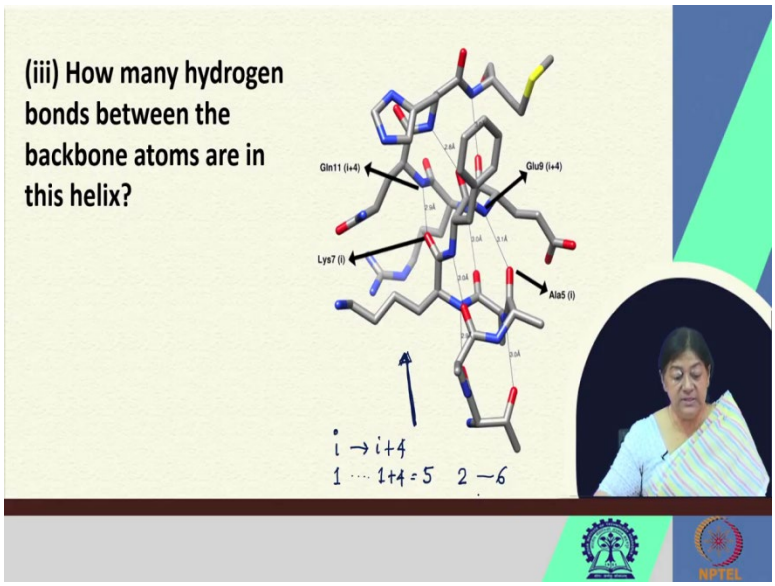


Our next question is, what is the length of the helix in Angstroms, in the direction of the helix axis? Now, we have to know what information we know about this axis or in general for all alpha helices. What we do know is that, there is 5.4 Angstroms as a pitch. What do we mean by the pitch? By a pitch we mean a complete turn that is going to give us a length of 5.4 Angstroms.

Another information we know is, there is going to be 1.5 Angstroms rise per residue. What do we mean by this? Depending upon the specific residues, there is going to be 1.5 Angstroms rise per residue. So, we can approach this in two ways with the knowledge of how many turns we have or a 1.5 Angstroms rise per residue with the knowledge that there are 15 residues in this particular helix.

So, when we look at this we say the length of one turn is 5.4 Angstroms. We know that there are going to be 15 divided by 3.6 turns as we found out in the previous slide, that multiplied by 5.4 is going to give us 22.5 Angstroms. Again, if we look at 1.5 Angstroms rise per residue and we know that there are 15 residues, again we get 22.5 Angstroms length of this particular helix.

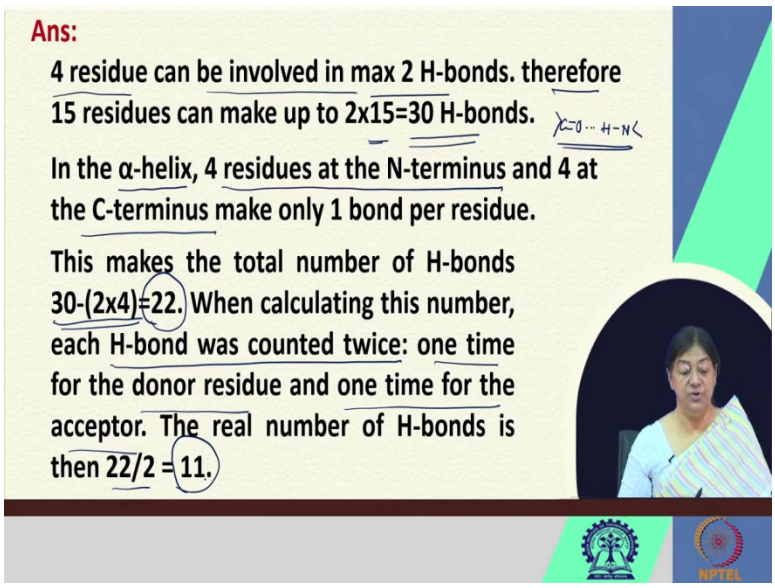
(Refer Slide Time: 19:23)



Our next question is, how many hydrogen bonds are there between the backbone atoms in the helix? We know that when we look at the hydrogen bonds as shown here [refer to slide], we have a specific direction, this is the direction of the helix. So, if this is the direction of the helix, it is going in an upward direction. What we have here is, we have now a hydrogen bond between i to i + 4.

So residue number 1 of the helix is going to have a hydrogen bond with residue number 1+ 4 that is 5. Again, residue number 2 is going to have a hydrogen bond with residue number 6 and so on and so forth. But how do we calculate this?

(Refer Slide Time: 20:19)



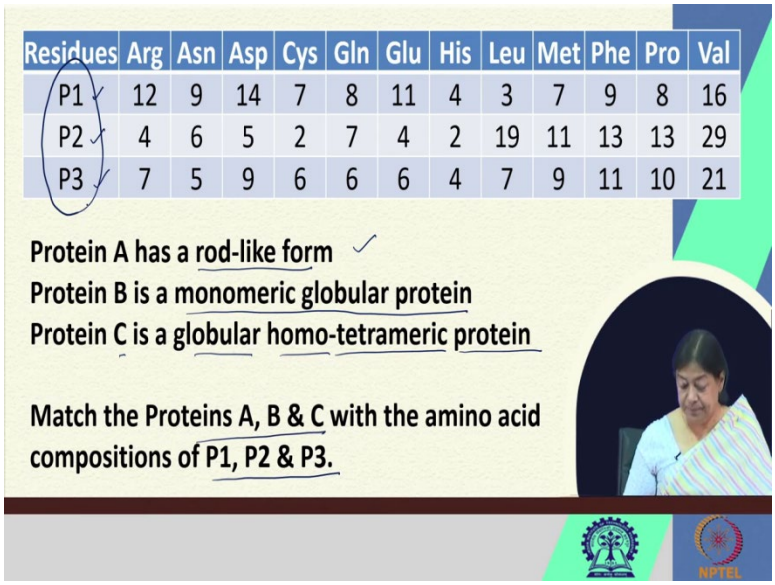
We have 4 residues that can be involved in a maximum of 2 hydrogen bonds. Why 2 hydrogen bonds? And where can hydrogen bonds form? They can form between the C = O and they can form between the NH. This is where a hydrogen bond will form. So we can have each residue involved in 2 specific hydrogen bonds. Why?

Because each residue has a C = O associated with it and also has an NH associated with it. So the 15 residues can have at maximum 30 hydrogen bonds there. But in the α-helix there are 4 residues at the N-terminus and 4 residues of the C-terminus that make only 1 bond. The i to the i + 4, when we go to the end of the α-helix is not possible because the helix has now ended, one i will not have an i + 4 to bond with.

Similarly, at the beginning also this will be the case. So, this means that the total number of hydrogen bonds can actually be leaving out the 4 residues at the beginning N-terminus, leaving out the 4 residues at the C-terminus. We can say that 30 minus 2 into 4, those are the 2 hydrogen bonds for the 4 residues leaving us with 22 possible hydrogen bonds.

Again, each hydrogen bond was counted twice, one time for the donor residue and one time for the acceptor. So the real number is 11, meaning that we have 11 hydrogen bonds possible for this particular case. You can actually draw the residues and count it for this particular α-helix and then confirm that we do see these many hydrogen bonds.

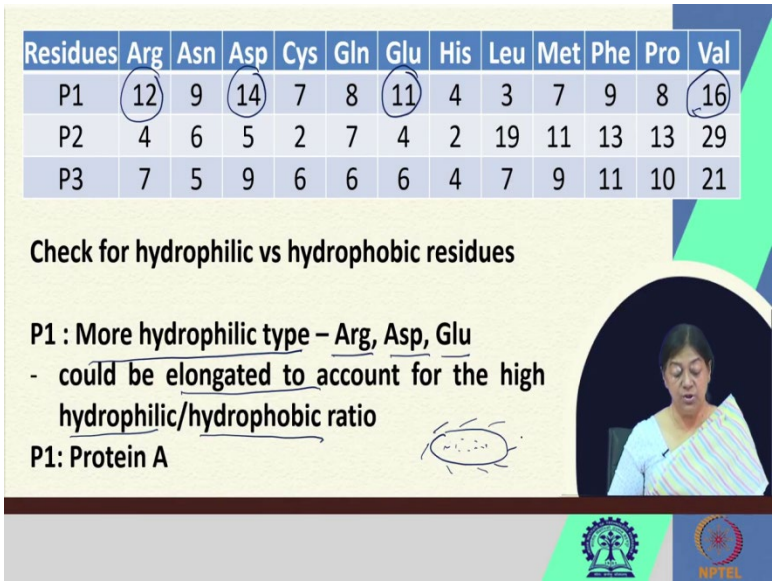
(Refer Slide Time: 22:31)



In our next problem there is a series of residues given [refer to slide], the number of each type and there is information about 3 different proteins. Protein A has a rod-like form, protein B is a monomeric globular protein, and protein C is a globular homo-tetrameric protein. And the problem here is, that we have to match the proteins A, B and C with the amino acid compositions given for P1, P2 and P3.

So, we now have to observe the types of amino acids and the specific numbers of each type. The information here is that protein A has a rod like form. So among P1, P2 and P3 which is protein A? Let us see how we can approach this problem?

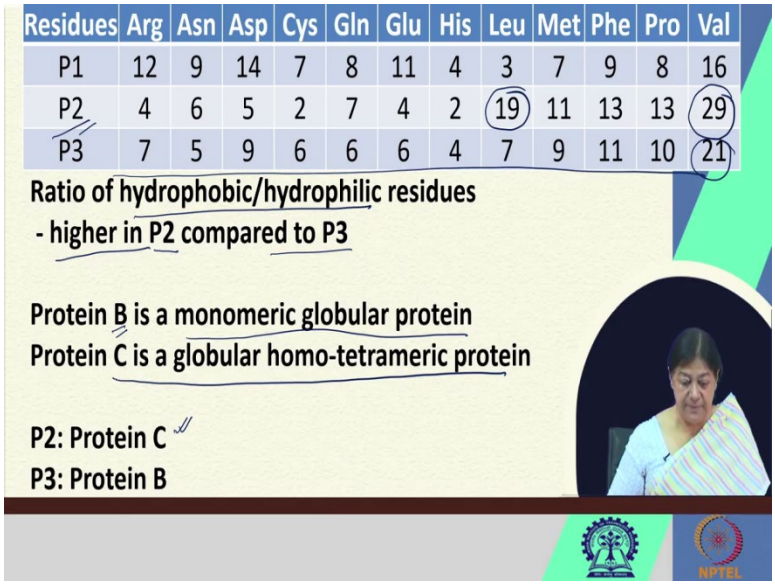
(Refer Slide Time: 23:40)



We check for hydrophilic versus hydrophobic residues and we look for a more hydrophilic type where we see arginine, aspartic acid and glutamic acid. So there is a large arginine here, aspartic acid here, glu here and also there is quite a number of valine here. There is what we can say an elongated type because there are more hydrophilic than hydrophobic ratio.

What is going to happen on the surface if we look at something like this. There are going to be the valines that are probably inside the protein and then hydrophobic ones that we see are going to be on the surface. So this could correspond to protein A.

(Refer Slide Time: 24:30)

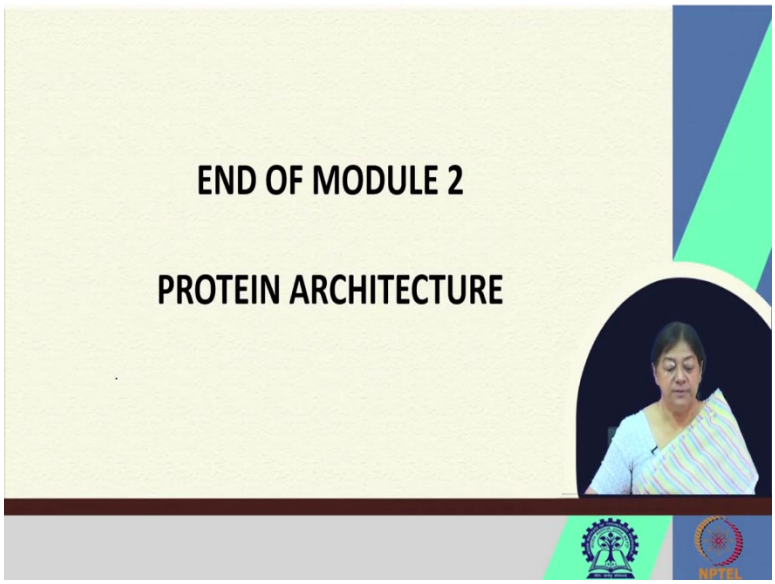


When we look at the residues again for a ratio of hydrophobic to hydrophilic residues, we see that there is a higher ratio in P2 compared to P3.

Why? Because, we see 29 here. We see 21 here also, but if we look at the overall setting here, we see 19 here for P2 again, which means that there is a higher hydrophobic to hydrophilic ratio for P2 compared to P3. Which means that, protein B is most likely the monomeric globular protein that we were talking about, and protein C is the globular homo-tetrameric protein.

Now we know for a tetrameric protein there are going to be more hydrophobic contacts. So, we want to look for a higher hydrophobic ratio. And what do we see? We see a higher hydrophobic ratio for P2, which means that P2 corresponds to protein C and P3 will correspond to protein B.

(Refer Slide Time: 25:44)



This will be the end of module 2, where we looked at protein architecture, primary, secondary, tertiary, structures and the interactions between them.

Thank you.