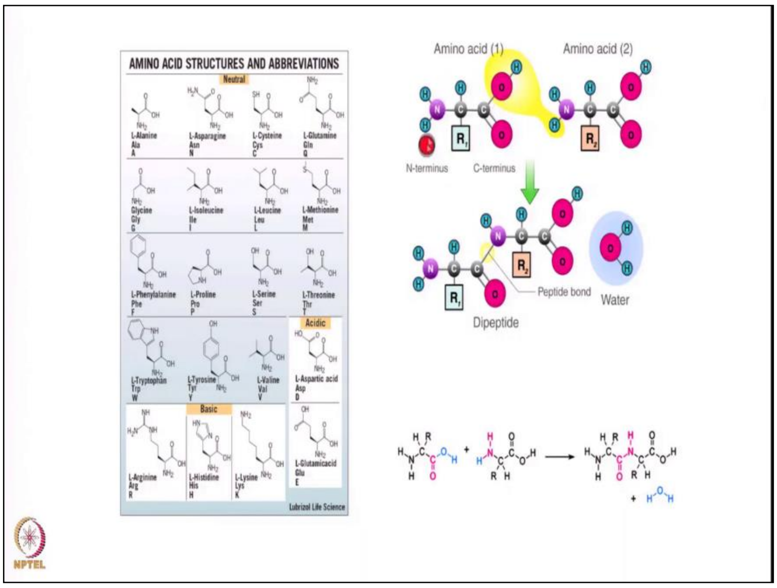
NMR spectroscopy for Structural Biology

Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 31
Determination of Structure and Dynamics of Proteins - 1

We will now start a new application of NMR in structural biology that is with regard to the determination of the structures, interactions, and dynamics of proteins. What are proteins? Proteins are polymers. They are built from amino acids. A typical structure of an amino acid is indicated here.

(Refer Slide Time: 00:42)



So, this is the NH_2 here. Then there is an atom called Cα carbon and to that is an attached proton and then it is R_1 which is called a side chain and then it is a carboxylic acid group here. So, this is the basic structure of an amino acid. You have the so-called N-terminal, which is an amino group and then you have the Cα, one carbon atom here. So, of course, you can have different kinds of groups attached which are called the side chains and then this is called a C-terminal which is the carboxylic acid.

Now, two amino acids can join together in this particular manner as it is indicated here. The OH group of the carboxylic acid and this one hydrogen of the amino group, get removed by the formation of a bond between this carbonyl carbon and the nitrogen here, and that is how you see this bond is formed and a water molecule is removed. So, this remains R_1 and another amino acid which is here, amino acid 2, which has a side chain which is called R_2.

You have formed a bond here with the C-N bond. The N-terminal of one residue is now bonded to the C-terminal of another residue and this is called the peptide bond. So, the chain continues like this. So, we have the OH group once more, then of course this carboxyl of this second residue and can form a peptide bond with another residue with the amino group of another residue.

So, it will continue like this. So, therefore this is indicated more schematically in this. So, you have this COOH group here and the amino group here. This hydrogen and this OH group are combined to form a water molecule and then there is a carbon-nitrogen bond formed here. So, therefore you have this C-R and then you have this carbonyl, carbonyl tends to the NH group. So, this is the peptide bond.

So, this group is called the peptide group CONH and then you have the CHR here. In these residues, R can be the same or it can be different. So they can have different kinds of linkages between the various amino acid. Now, there are 20 different types of amino acids. What is the difference between them? The difference is in the R Group here.

For example, this is called the L-alanine, this is called asparagine, and this is called cysteine and all these are L-groups. This depends on the configuration at the Cα carbon. So, which way the groups are oriented here, the CH is indicated on the top or one is indicated here. So, which way these are relatively oriented determines whether it is L-amino acid or the other thing is D-amino acid.

So, by and large in nature, we only have the L-amino acids, where the proton is on a particular side and the side chain is on the opposite side. So, all these are L-amino acids. So, for these residues which are indicated here, we group them into different categories. So, you have the L-alanine, L-asparagine, L-cysteine, and L-glutamine and then you have the glycine isoleucine, leucine, and methionine.

Then you have the phenylalanine, then you have the proline, and you have the serine and threonine. So, all these amino acids are so-called neutral amino acids. Neutral in the sense that after the peptide bond is formed they do not have any charge anywhere. So, there is no protonation site, after the peptide bond is formed in the polypeptide.

So, what is the nature of these residues? The alanine has the side chain which is a methyl group and then in the asparagine, you have the $CONH_2$ group. These are all neutral, although they are polar, there is some polarity there, it has a $CONH_2$ group, and then the cysteine has an SH group here. The side chain has an SH group. So, we call this as, after the Cα you have this Cβ and then the SH group. So, this is the Cβ carbon, from the Cβ carbon we see the nomenclature afterward, and the glutamine has this carbon here, another carbon then you have the $CONH_2$ group.

So, there is a common thing between this glutamine and asparagine. They both have the $CONH_2$. So, there is a CO and the NH_2, and in between there is another carbon in the case of glutamine whereas that is not there in the asparagine. And then you have the glycine, glycine does not have anything else, it only has 2 protons at the Cα positionα. So, this carbon here has 2 hydrogens, therefore there is the CH_2 group there.

And isoleucine, isoleucine has from here three groups, carbon branching out like this. There is a proton here, there is a carbon here and that carbon is a methyl group. All the protons are not drawn here only the carbons are drawn. So, there is a carbon here and a carbon there. There is the branching at this site at the Cα, there is another carbon here and one carbon here and one carbon here.

So, obviously from here, it is clear that if you have a stick just staying by itself, then it would mean that this carbon will be a CH_3 group because it is an SP^3 hybridized group. There are no other bonds indicated, it is only protons attached therefore this will be a CH_3 group. So, these are the CH_3 group here and then this will be a methylene group the CH_2 and then at the end, you have another CH_3.

So, at the end, if there is no other branching, it will be CH_3. And the leucine has from the Cα to the Cβ and then the Cγ and there are 2 methyls sticking around here, these are 2 methyl groups because their bonds are drawn only to the carbon-carbon bonds. So, this will be a methyl group there.

And then for the methionine, it goes in this manner from the Cβ. So, you have an S here, but now S is connected not to a proton but to a methyl. There is an SCH_3, unlike the cysteine had an SH here. Now, this does not have an SH but it has an SCH_3 and it also has methylene in between. Of course, there is a methylene group with the CH_2 and then the SCH_3.

Now, these are aromatic, this is phenylalanine, and you have the NH_2 from the Cα. You have to the Cβ, and to the Cβ is attached a phenyl ring. So, this is a phenyl ring. And now look at Proline, which is an interesting structure and this formed itself within, on the backbone itself formed the closed ring all these are CH_2, CH_2, CH_2, NH. One of the H of the proline is the one that is normally remaining, it participates in the closed ring formation.

Therefore in this hydrogen bond is formed and of course, this proton will also go and there will be no other proton remaining on the proline ring. So, therefore the proline ring in the polypeptide chain will not have any hydrogen at the nitrogen position. And the serine now has a CH_2OH group in the side chain. So, this is the CH_2 here and then there is the OH there. From the Cα, there is the CH_2, the CH_2 will be called the β-proton.

And the threonine is similar to the serine except that you have a methyl, and there is no proton there. So instead of the proton, there is the methyl group here. In the case of serine, there is a CH_2 and then the OH. Now here you have the C and then there is the proton there and then there is a CH_3 group attached to this. So, these are the different structures of amino acids.

Now all these are neutral amino acids. Once the polypeptide chain is formed, there is no charge anywhere on these. Now so, here we have the next set of amino acids these are called acidic amino acids and these are here L-tryptophan, tyrosine, and valine. So, these are also neutral because there is in a color-coded manner. You can see tryptophan has an indole ring here at the Cβ.

So, we have this tryptophan, and the tyrosine is very similar to phenylalanine except that it has an OH group here. So, all these amino acids, therefore 4+4=8 +4=12 +3 =15 amino acids are neutral in nature. And then you have these white ones this is acidic, acidic there are 2 there. So, these are aspartic acid and glutamic acid. Why? Because in the side chain, you have a COOH group and the COOH group is acidic.

So, you have the CH_2 then you have the COOH. The backbone COOH is different. It is the backbone COOH that participates in the polypeptide chain formation, but on the side chain you also have a COOH group, therefore, there is a chance of having a negative charge there because this will carboxylic acid group and that will OH group will dissociate. So, you will form a negative charge there.

So, you can have a negative charge in this. This is aspartic acid and it will be acidic. Similarly the glutamic acid, glutamic acid is similar to this except that it has one more carbon. There is a CH_2, CH_2, and then COOH. So, these 2 are acidic amino acids. Then you have the basic amino acids, there are 3 basic amino acids: Arginine, Histidine, and Lysine.

So, these have NH_2 groups or $NHNH_2$ groups in the side chain. So, you have the Cα, Cβ, then you have the γ, then you have this δ, then you have the ε. So, you see this goes to the ε, ψ. So, these side chains are very long. So, there are many carbons here in the side chain. Similarly, the histidine has a 5-member ring. There is a 5-member ring, there are 2 NH, there are 2 nitrogens and one of them has hydrogen there.

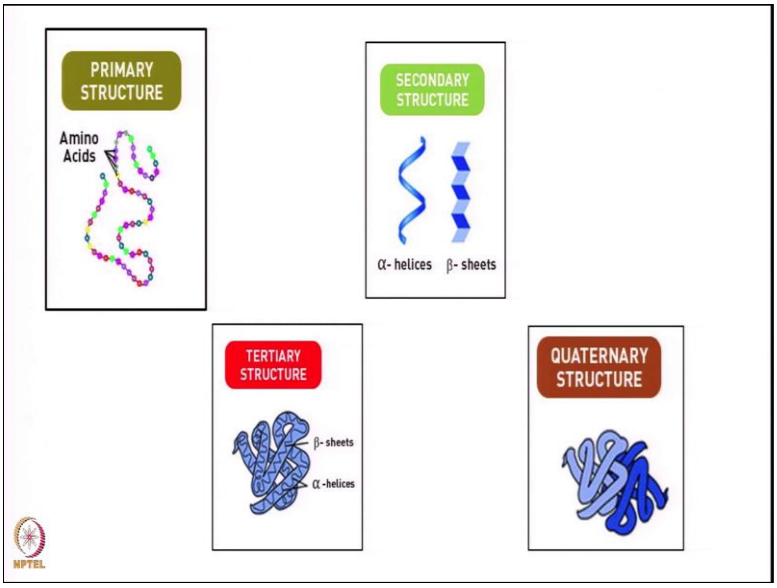
So because of the nitrogen, they can get protonated, then when it gets protonated it forms an NH_3 or NH wherever the charge is there it will produce a positive charge. The acidic one produces a negative charge, the basic one produces a positive charge. And this is the lysine, the lysine has so many carbons on the side chain, this is the Cβ, from the Cβ, you have the γ, δ, ε, ζ, then you have the NH_2 group there.

So, all of these are very distinctly different amino acids. Typically all these are denoted by one-letter codes, you can have the 3 letter codes as well as the one-letter codes. Alanine, asparagine, cysteine, and glutamine are represented as A N C and Q. Like that Gly I L M and then you have F P, F is phenylalanine, and you have the P is proline, S is serine, then T is threonine, then your W is tryptophan, and Y is tyrosine, V is valine.

Valine also has 2 methyl groups. There is a Cβ directly that is attached to the 2 methyl groups in this. And Arginine is represented as R, then histidine (H) and Lysine (K). So, this is the structure of the various 20 different amino acids. Therefore clearly the polymer chain can have a great variety of sequences, depending upon the combination, what is the nature, how long is the polypeptide chain, and which amino acid is appearing where.

How the structure is getting formed and what determines the first level of the structure of the proteins?

(Refer Slide Time: 13:08)



And that is what is indicated here as the primary structure of the protein and these are like beads on a necklace. So, each one of these beads is an amino acid. These are all amino acids. There can be many amino acids. This kind of protein has 100 amino acids, 200 amino acids, thousand amino acids. So, various kinds of amino acids are present.

Now each of these amino acids has a certain degree of degrees of freedom, how they are oriented with respect to each other. So, in the entire chain, there are many different possibilities of orientations of the individual amino acid side chains with respect to the previous one and the following one. So depending upon that, you get some regular structures and these regular structures will show up.

When you have a large chain like that going with certain preferences for the certain relative orientations of the individual side chains along the backbone then you get different kinds of structures. These are classified in a rough way like this. This is called the α-Helix. So, it goes in a helical manner and these are called the β-sheet. It looks like this. We will look into this in greater detail soon and these are called secondary structures.

This is called a primary structure and this is the secondary structure. You have 2 kinds of basic secondary structures α-helices and β-sheets. There are 2 types of β-sheets that we will see very soon and then the relative orientations of these secondary structures in a protein can vary. In some places, it can have this structure, some places it can have this structure.

So, therefore you can have a combination of all these secondary structures in a large protein and that is called a tertiary structure. The tertiary structure represents the 3-dimensional structure of the protein. So, you see you can have all the β-sheets or you can have a helix here, the Helix is indicated in this manner. You can have many helices here, then again β-sheets, and a combination of all of these will form the tertiary structure.

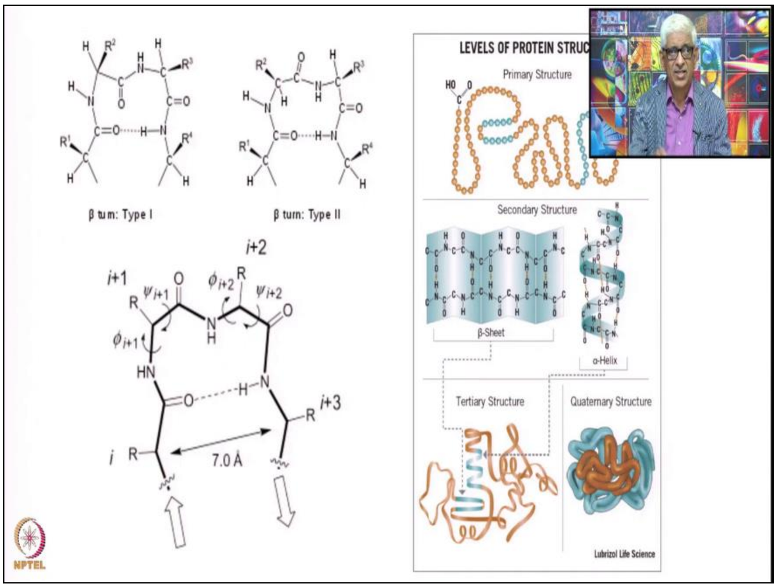
Of course, the chain can loop here and then those will be called the turns. In a particular molecule, there can be combinations of alphas and the β-sheets and there can be connections in between and those are called loops. There can be turns and they have specific geometry of the relative orientations of the individual amino acids. Then you have the quaternary structure have a well-folded structure like this, there can be multiple domains.

There can be more of such ones which can get associated because of various interactions and these are called quaternary structures. Therefore the protein structures can be categorized into 4 different types. So, you have the primary structure which is the basic, thing that determines the composition of the protein, what all amino acids are there, whether it is the basic protein or it is an acidic protein or it is a neutral protein or it is a hydrophobic protein.

Because certain amino acids have various methyl groups, and aromatic rings, they all constitute the structure and the physical characteristics of the amino acids. The physical characteristics of the amino acids pass on their nature to the entire structure of the protein, because of that they also define the energetics and then they can have different orientations and then you will have different kinds of folded structures.

And then these can get associated through various kinds of interactions, they may be charged interactions, hydrophobic associations, and different kinds of associations can happen and that will constitute the so-called quaternary structure.

(Refer Slide Time: 16:51)



So, let us look at this in somewhat more detail. So, we have here the primary structure as indicated there. So, you have different kinds of chains. This is the so-called carboxyl end, this is the N-terminal, and the chain is running like this. Of course, it has various loops going around and different kinds of interactions can happen there. So, then we show here what is the nature of the Helix in a little bit more detail. If you see the Helix, the N-terminal is on this side.

So, NH-Cα-CO-NH-Cα-CO-NH-Cα-CO is the N-terminal. So, it goes from the N-terminal to the C-terminal like this. So, now you see it taking this sort of a structure, because this is stabilized by what is called the hydrogen bonds. So the NH of this amino acid is hydrogen bonding to the carboxyl oxygen of this.

This is the NH group here and there is a hydrogen bond between the carboxyl group of this amino acid. This comes at a certain distance. After it has taken a certain turn, here in the Helix this forms a helical hydrogen bond. This provides stability to it likewise every amino acid, this carboxyl group will be hydrogen bonded to the next of somewhere there and similarly. So, this next amino acid here is hydrogen bonded to the carboxyl of here. So, this continues.

So, in every curve in this whole chain, every amino acid amide hydrogen is hydrogen bonded to the carboxyl of one of the residues in the polypeptide chain. So, this is called a turn of the Helix. So, this is the turn of the Helix and it continues like this and in the turn, the total rise of this is typically about 3.6 amino acids in a turn.

So, there can be different kinds of turns that are called 3 amino acid turns, they are called 3_{10}-helices and that determines how many; what is the size of this ring? If you consider this as the total ring what is the size of the ring? Then, how are the β-sheets formed? Here you see, within the same chain a 4th residue is forming a hydrogen bond with the amide group of this and that forms the hydrogen bonding scheme.

Now the β-sheet is slightly different. Now, look at what is this chain. How it is running here? So the chain is running NH-Cα-CO-NH-Cα-CO-NH-Cα-CO. So, this is the chain running like this, this is from the N-terminal to the C-terminal. Now the lower chain is going in the opposite direction here NH-Cα-CO-NH-Cα-CO-NH-Cα-CO and so on.

So, the chain is running in this manner here, therefore this is called an antiparallel β-sheet. There are 2 chains that are running in opposite orientations and then these are held together by these hydrogen bonds. This NH is hydrogen bonded to the carboxyl of a residue in the other strand. This is one strand, this is the other strand, and the 2 strands are held together by hydrogen bonds.

So they are perfectly positioned to form this sort of hydrogen bond NHCO. So, they are present at very regular intervals, and these many hydrogen bonds are possible. This forms one of the β-sheet, which is called the antiparallel β-sheet there can be others also. So, the tertiary structure here, the helical portion is shown in this.

The arrow comes here and points to the Helix here. It turns here; when it turns like this there is a change in the direction; one is going from here to here and then it goes from here to here. So, there is a change in the direction and this will form a β-sheet. The same thing is indicated here. This portion of the chain is supposed to be forming a kind of helix;

whereas this portion of the chain is turning here and then, of course, you will form hydrogen bonds in this and that is typically indicated in this sort of a structure. Now you see there are certain things which are hanging here; what are these? Because the chain has to fold and even here, if the chain has to turn here then something is here. What is this? What does this correspond to these are called turns.

Now there are 2 kinds of turns. So the chain goes in this manner, which way the chain is running we can look at this. So this is the Cα-CO-NH-Cα-CO-NH-Cα-CO-NH-Cα-CO. So, there are 4 amino acids here; 1, 2, 3, and 4. These 4 amino acids are forming a turn and the carboxyl of the first amino acid residue is hydrogen bonded to the NH of the 4th amino acid residue.

See it turns like this and comes back here, and in this situation, this NH is coming close to this carboxyl and it forms a turn. Notice here, the carboxyl of this second residue is pointing inside here and the NH is pointing outside. This is one configuration. This depends upon the relative orientations which are written by certain kinds of torsion angles which will describe a little bit later.

And in this case, this is called the type 1 β-turn. There can be another way this turn can happen and in this, it is the same hydrogen bond except that the configurations here are somewhat different. You see here the α-proton is down here whereas the α-proton was up there and the 2 things are in the same orientation.

And the carboxyl is oriented outwards, and here the carboxyl is oriented inwards, and the NH is outwards, and the NH is inwards here. Similarly, this R_3 and this portion is roughly the same as that, the difference is occurring in this. So this amino acid residue has a different relative orientation with respect to the others and that makes the configuration in the loop different or we also call it conformations is different and because of that the same hydrogen bond is found even so here.

So this is called the type 2 β-turn. So, now you see here, this is more explicitly indicated. What makes these different orientations? There are certain torsion angles indicated here. If this residue is i, this is the residue i + 1, this is the residue i + 2, and this is the residue i + 3. The chain is running like this NH-Cα-CO-NH-Cα-CO-NH-Cα-CO and continues like that.

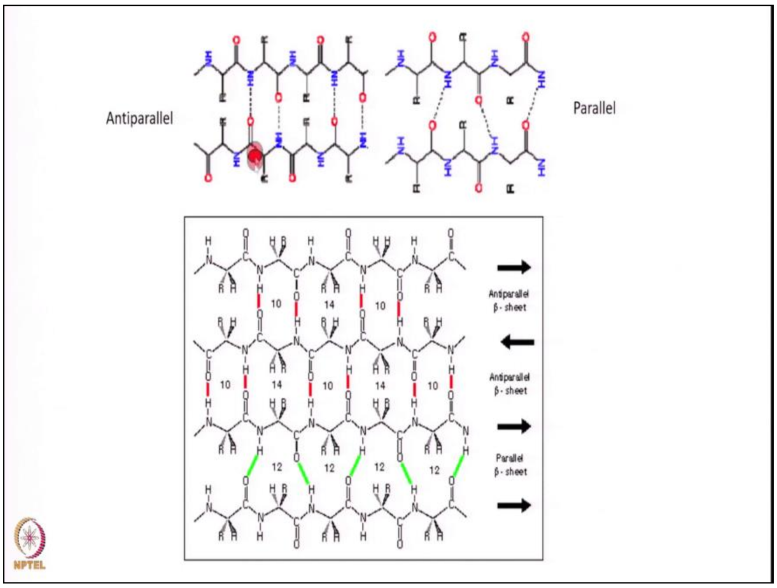
So, if I start this from this residue i + 1, NH-Cα-CO; this is the residue i, and this is NH-Cα-CO; and there are 2 torsion angles indicated here. This determines what are the relative orientations of these groups. These torsion angles are called Phi and Psi, which will describe more about it.

Similarly, this amino acid also has this Phi and Psi torsion angles and this defines the related orientations of these groups. These 4 atoms are nitrogen, this Cα, CO, and this nitrogen. What are the relative orientations of this? A rotation around this can change the orientation of this hydrogen with respect to this and that is what determines all of this. So, there are these 2 torsion angles which are called the Phi and Psi; these are there for every amino acid residue.

And that determines the variety of structures that can be formed by different combinations of these Phi and Psi torsion angles and when the chain turns around and comes back here you see this distance is roughly about 7 angstroms. So, one can calculate this. Depending upon what is the configurations of this, if the hydrogen bond has to be formed, this puts the limit as to how much should be the distance.

Because from N to O, this distance cannot be more than 2.8 to 3 angstroms, which defines the hydrogen bond, and once you have that the Cα- Cα distance is approximately 7 angstroms.

(Refer Slide Time: 26:09)



Now we talked about the anti-parallel β-sheet, there is one more possibility of a β-sheet and that is called the parallel β-sheet. How does this work? Now you consider the 2 strands NH-Cα-CO-NH-Cα-CO-NH-Cα-CO, the chain is running like this from left to right; and here also NH-Cα-CO-NH-Cα-CO, this chain is running in the same direction.

Whereas here, the 2 are running in opposite directions. Here the 2 chains are running in the same direction and yet of course the hydrogen bonds are possible. This NH now hydrogen bonds to this oxygen here, the hydrogen bond scheme is much cleaner, they are straight and of course, then it is more stable and here, of course, they are slightly tilted there.

So this NH is hydrogen bonded to this oxygen, this NH is hydrogen bonded to this oxygen. This will continue like that, alternately like here, this is NH to this CO, this NH to this CO, this NH to this CO, this NH to this CO; and similarly here. There is a shift in the resistor. So, therefore you have this CO to this NH, this CO to this NH, and this will of course go further down this. This is called the parallel β-sheet.

Now in a given protein structure, you can have combinations of these. The chains can be running in particular directions. So, here the chain is running like this NH-Cα-CO-NH-Cα-CO, the chain is running like this and it may have a turn somewhere and comes back, and then another portion of the chain runs like this. So, this is NH-Cα-CO-NH-Cα-CO. Therefore here it forms an anti-parallel β-sheet.

But again the chain may turn around somewhere and comes back and it goes once more here NH-Cα-CO-NH-Cα-CO. Now, this direction is antiparallel to the orientation. It can again form hydrogen bonds therefore this is once again an antiparallel β-sheet. So, these 2 antiparallel β-sheets are adjacent to one another but then, of course, this chain may go somewhere then again you have it you may have another chain that is running parallel to this NH-Cα-CO-NH-Cα-CO and so on.

Now this chain is running parallel to this. So, therefore this β-sheet which is formed here is the parallel β-sheet. So, such kinds of sheets can be formed in protein structures depending upon the relative orientations of the different helices or the chains. You can have anti-parallel β-sheets and parallel β-sheets. You can have combinations of these. It can form a huge sheet-like structure.

As a result of this, this will be extremely stable. Of course, this will be determined by the side chains which will dictate what sort of a structure can be formed, and that will depend upon the torsion angles Phi and Psi. And this of course is a very crucial parameter that generates a variety of protein structures. So, I think we can stop here.