Basics of Biology
Professor Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Module – V: Molecules of the Life (Part-II)
Lecture 21
Proteins (Part-II)

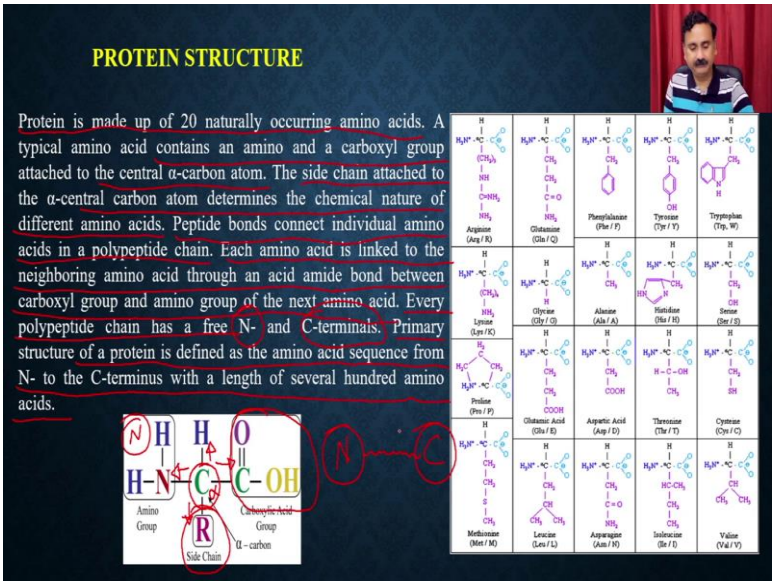
Hello everyone, this is Doctor Vishal Trivedi from department of biosciences and bioengineering IIT Guwahati. And what we were discussing? We were discussing about the bio molecules. And in the previous module, we have discussed about the three bio molecules, we have discussed about the nucleic acid, and we have discussed about the carbohydrates, and we also discuss about the lipids.

And in the previous lecture, we have discussed about the proteins. So, we have started discussing about the proteins, that the proteins are made up of, of the amino acids and these amino acids are having a general structure, where they have the central C alpha-carbon. And on this central C alpha-carbon, we have the four different types of functional groups which are attached to each.

On one side it has the amino group, on the other side it has the carboxyl group, and the third side it has the hydrogen, and the fourth side it has the functional side chains. And based on the functional side chain, amino acid can be classified into 4 different groups, 20 different types. It could be vary from the very simple glycine to a very, very complicated tryptophan. And depending on the side chain, it can be of the different molecular weights and different types of properties. So, it could be a hydrophobic amino acid, it could be a hydrophilic amino acid, it could be a polar amino acid, it could be negatively charged amino acid, or it could be a positively charged.

And in addition to that, we have also discussed in detail about the thin layer chromatography, and how the thin layer chromatography can be used to analyse the different types of amino acids. So, in today's lecture, we are going to discuss more about the protein structures. So, let us start discussing about the protein structures.

(Refer Slide Time: 2:48)



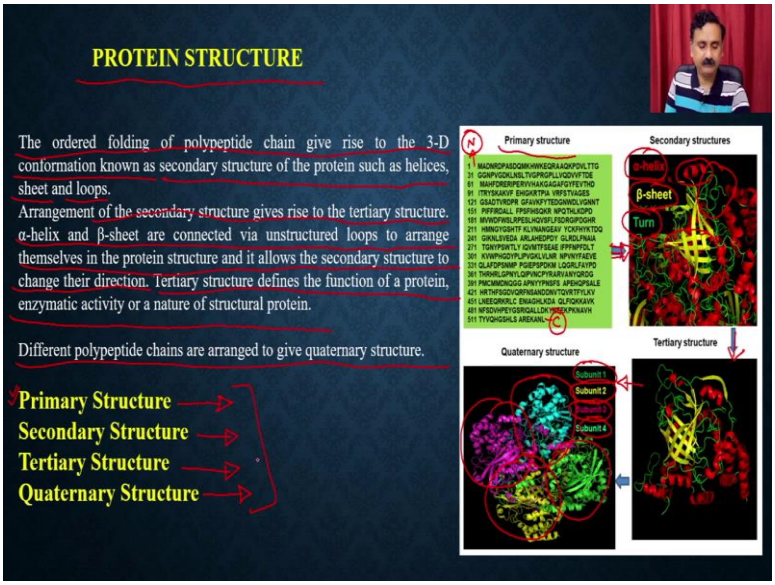
So, the as it said, protein is made up off, of the 20 naturally occurring amino acids, a typical amino acid contain amino group and a carboxyl group attached to the central alpha-carbon, the side chain attached to the central alpha-carbon determine the chemical nature of the different amino acids. So, what you see here is the you have the C alpha-carbon, and that C alpha-carbon is attached to the four different types of functional groups. And these four different type professional groups. And depending on the R side chains, you can have the 20 different types of amino acids.

Peptide bond is for connecting the two individual amino acids. And that is how it they are actually giving the polypeptide chain. Each amino acid is linked to the neighbouring amino acid through an acid amide bond between the carboxyl group and the amino group of the next amino acids. Every polypeptide chain has the free amino group and the free C-terminal groups. That is why the primary structure of a protein is defined as the amino acid sequence from the N- to the C-terminals with a length of the several hundred amino acids.

So, these are the 20 different amino acid, what we have already discussed in the previous lectures. So, the primary structure is defined as the amino acid sequence from the N-terminus to the C-terminus. So, do you see that it is it starts with the N-terminus. So, this is the N-terminus, and then it ends up with making a combination with the all of the amino acid to the peptide bond, but ultimately with the last amino acid, it is going to have the carboxyl group, which is going to be free. And that is why the protein is going to have the amino group and then it is going to have all the amino acids, and then it is going to have the carboxyl group.

So, that side, I mean protein has the two ends, one is N-terminus and the other one is called as the C-terminus.

(Refer Slide Time: 4:47)



The ordered folding of the polypeptide chain gave rise to the 3-D conformation, known as the secondary structure of the protein such as the helices, sheets and loops. So, this is what you see here is, the primary structures. So, all the protein, all the amino acids are present. And the first amino acid is actually going to have the N-terminus, whereas the last amino acid is going to have the C-terminus.

When this primary structure is getting folded, it is actually going to give you the secondary structure, where you have the alpha helices, these are the alpha helices, then we have the beta-sheets. So, you can see the beta-sheets, these are the beta-sheets, and then it also has the turn. So, these are the turn what you see here.

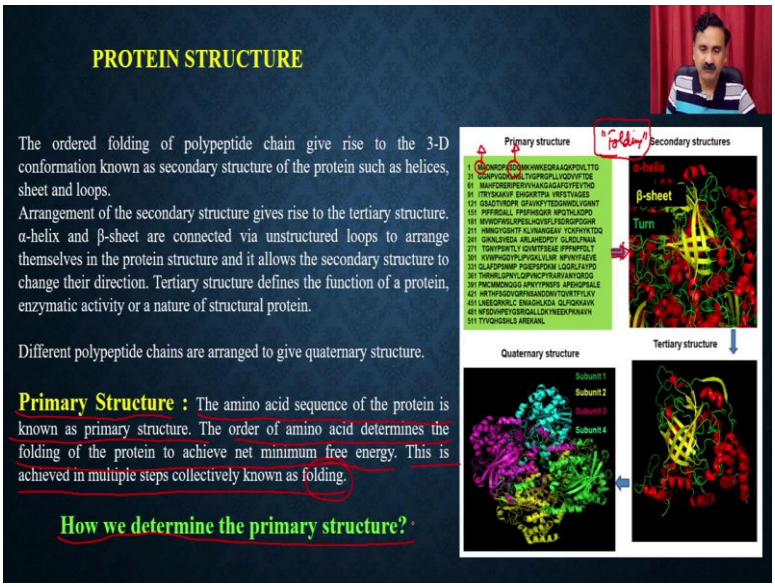
Arrangement of the secondary structure give rise to the tertiary structures, alpha helix and beta-sheets are connected by the unstructured loop to arrange themselves in the protein structure and it allows the secondary structure to change their directions. Tertiary structure define the structure of a protein and the enzymatic activity or the nature of the structural protein. So, once the secondary structures are joined together by the loops or the turn, they are actually going to give you the tertiary structure. So, this is what you see here is the tertiary structures.

And if the protein has the multiple subunits, then it is actually going to give you the quaternary structure. For example, in this case, we have the subunit 1, 2, 3, and 4. So, all

these three subunits, all four subunits are coming together. And that is how you are going to have the quaternary structures.

So, different polypeptide chain is arranged to give the quaternary structure. So, depend, that is why if you want to understand the protein structures, you have to understand all the three, four different types of structure. So, we have the primary structure, we have the secondary structure, we have the tertiary structure, and we have the quaternary structure. So, these are the different level of organizations, what if present into the protein structures.

(Refer Slide Time: 6:53)



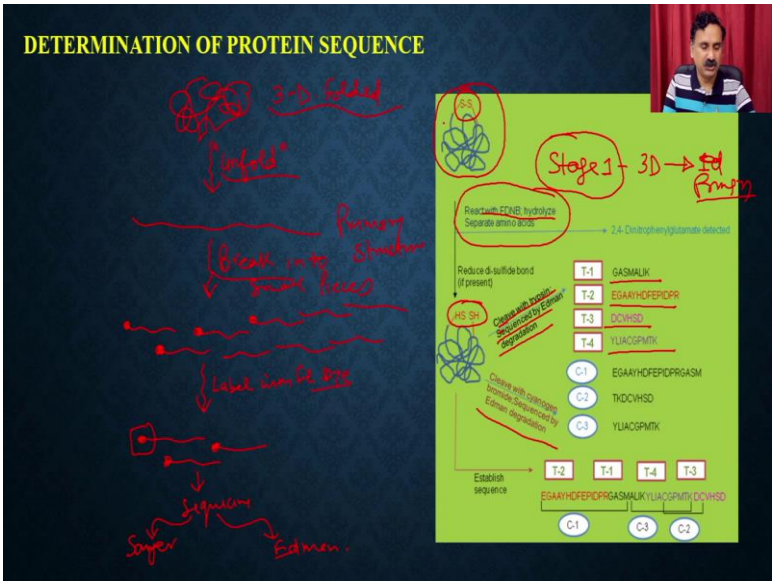
So, let us first start with the primary structures. So, the primary structure, the amino acid sequence of a protein is known as the primary structure, the order of the amino acid determines the folding of the protein to achieve the net minimum free energy and this is achieved in the multiple steps collectively known as the folding.

So, if the primary structure is going to fold, so this is what you see here is the primary structure, where you have the length of the amino acid. Now, see here, I am just showing you the single letter code of the different amino acid, because it is almost impossible to write the full length or the full name of that particular amino acid. because to save the space.

And these primary structures are actually going to fold to give you the secondary structure, and that event is called as the folding. So, when the primary structure is going to fold into a proper three-dimensional conformation, then it is going to give you the secondary structure. Now, the question comes how we can be able to determine the primary structures? So, if you want to determine the primary structures, you have to first achieve the primary structure, and

then you can be able to sequence the protein, and you can be able to know the amino acid sequence of that particular proteins.

(Refer Slide Time: 8:14)



So, these are the different steps what is being shown. So, what you can have is, you can have the starting with the protein structure. So, this is the three dimensionally folded protein. So, here you are going to have the tertiary and secondary structures. Now, what you have to first do is, you have to first convert that into a linear chain of amino acid, which means first you have to achieve the primary structure. So, which means you are actually going to unfold the protein by the chemical or the enzymatic method, then what you are going to do is because this length is going to be very large, then you are actually going to break the peptide or break the protein into in small pieces.

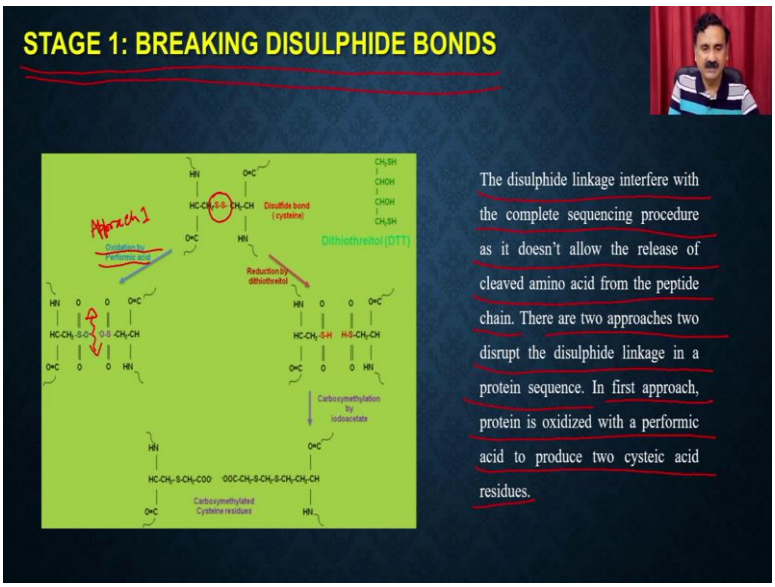
So, into small pieces. And the you are going to break the small pieces like this. There are going to be overlapping regions, so that you can be able to add put them these things separately. And then, what you are going to do is, you are actually going to label the terminse or the terminal amino acids.

So, you are going to have the labelling of the terminal amino acid in this way. So, you are going to have the labelling. So, labelling with the fluorescent dye, in different, and then you are going to do the sequencing, then you are going to identify that labelled amino acid. That sequencing, you can do by two methods. You can use the Sanger sequencing method, or you can use the Edman degradation method.

So, this is what it is showing here. First, you are going to start with the 3D fold structure, then you are actually going to make it unfold. So, you are going to use the different types of treatments, like you are going to use react with the FDNB, and all that, so that it is actually going to break the disulphide linkages. And then, once the it is going to adopt the primary structures, then that you are going to degrade the primary structure with the help of the different types of chemicals or the enzymes and that is how you are going to get the small fragments.

And once you got these small fragments, you can actually be able to do the sequencing, and then once the sequencing is over, then you can be able to put these blocks together and that sequencing you can do either by the Edman degradation method, or to the Sanger sequencing method. So, let us understand. So, in the stage 1, this is the stage 1. So, stage 1, you are going to convert the 3D conformations into the 1D or the primary structure. So, that you are going to achieve simply by in the stage 1, where you are first going to break the disulphide linkages.

(Refer Slide Time: 11:17)



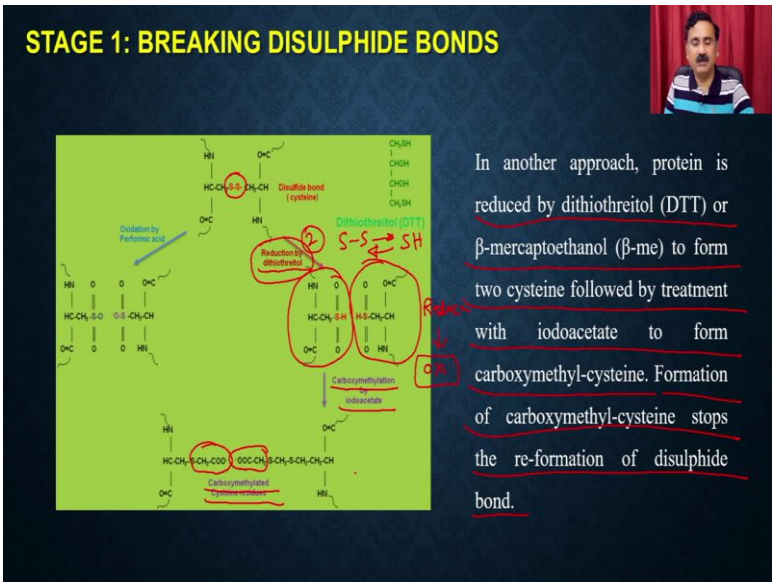
So, stage 1, stage 1 is the breaking of the disulphide bonds. So, we can imagine that this is the protein, which has a disulphide linkages. And the disulphide linkages interfere with the complete sequencing procedure as it does not allow the relief of the cleaved amino acid from the peptide chain.

There are two approaches to disrupt the disulphide linkages in a protein sequence. In the first approach, the protein is oxidized with a performic acid to produce the two cystic acid. So, why the, there is a need to break the disulphide linkages, because if you do not break the

disulphide linkages, even if this particular amino acid is actually going to be labelled, and it is going to be hydrolysed, it is not going to be released from the main chain, because it is still having a, it is bind to the main chain through a disulphide linkage.

And that is why it is important to break the disulphide linkages. So, disulphide linkages, we have the two approaches, in the approach one, you can actually use the oxidation with the performing acid. And that is, when you do the performing acid treatment, it is actually going to break the linkage between the disulphide linkages. And that is how it is actually going to give you the two fragments.

(Refer Slide Time: 12:37)

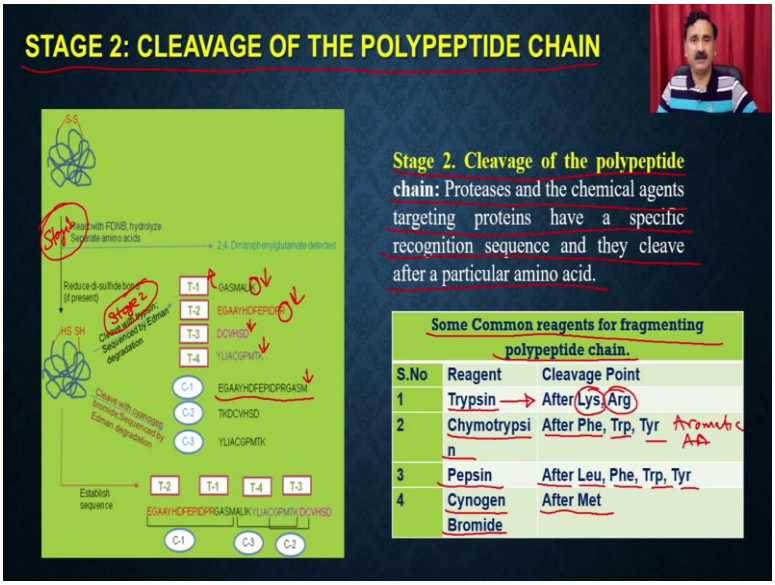


Whereas in the approach number 2, the protein is reduced by the DTT or beta-mercaptoethanol to form the two-cysteine followed by the treatment with the iodoacetate to form the carboxymethyl-cysteine. Formation of the carboxymethyl-cysteine stop the reformation of the disulphide bond. So, in the approach 2, what you going to do is, you are going to add the DTT.

So, DTT is a reducing agent. So, once you are going to reduce the disulphide linkages. The S is, S-S is actually going to get converted into S-H. And that is how it you are going to have the two peptide bonds where the disulphide bond is broken, but this this has a problem because as soon as you have the S-H and you have the reducing environment, it is going to be remade as S-H but once it is actually going to be acquired the oxidizing environment again the S-S is going to be oxidized and again the S-H is going to be get converted into the S-S double bonds.

So, to avoid that, you are again going to react this with the carboxymethylin by the iodoacetate. So, the in that case, then what will happen is that the S is actually going to be tagged with this particular functional group, this is going to like a. And that is how it is actually going to form the carboxymethylated cysteine residue. And once you have this, then they will not be able to come together even if the conditions are oxidizing in nature. Now, once this is done, you can actually go back to the stage 2. In stage 2, you are actually going to break the big polypeptide chain into the multiple fragments.

(Refer Slide Time: 14:21)



So, the stage 2, stage 2 is the cleavage of the polypeptide chain. So, in this stage 2 is the cleavage of the polypeptide chain, the protease and the chemical treatments are targeting protein have a specific recommendation sequence and they cleave after a particular amino acid. So, this is 1. So, stage 1 is this. So, stage 1 is over where you have actually destroyed the disulphide linkages by the two approaches, what we have just discussed. And now, in the stage 2, you are actually going to cleave the protein with the enzymatic or the chemical methods. So, some of the common reagents what you are going to use for fragmenting the polypeptide chain is, that you can use the enzyme.

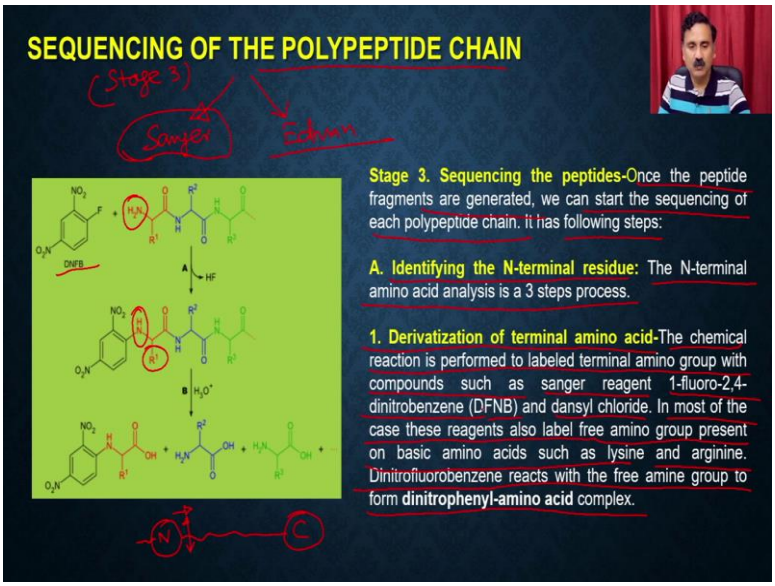
So, you can, if you have the trypsin enzyme, the trypsin has the cutting side which is actually after lysine or to the arginine which means wherever the lysine or the arginine is present. For example, see these are the peptide which are being generated by the trypsin. So, if you treat it with the trypsin, it is actually going to cut wherever you have the lysine or the arginine. So, for example, here it has cut, here it has cut. So, wherever you have the lysine and arginine it

is actually going to cut and that is why, it is actually going to generate the different types of fragments.

Similarly, you can use the chymotrypsin. So, chymotrypsin is actually going to cleave the peptide after the phenylalanine, tryptophan or tyrosine, which means after the aromatic amino acids. So, then we can also use the pepsin. So, pepsin is actually going to cleave the polypeptide chain after the leucine, phenylalanine, tryptophan or the tyrosine. And then you also have the different types of chemicals. For example, you can use the cyanogen bromide, and cyanogen bromide is actually going to cleave the polypeptide after methylene.

For example, these are the peptide sequences that have been generated by the cyanogen bromide. So, you see, this is the methylene and after the methylene it has been cut by the cyanogen bromide. Now, once you got the these small fragments, then what you can do is, you can take the individual fragments, and then you can sequence this small fragments, once you got the sequence of these small fragments, then you can have to put them together and that is how, you are going to get the sequence of the complete proteins.

(Refer Slide Time: 16:44)



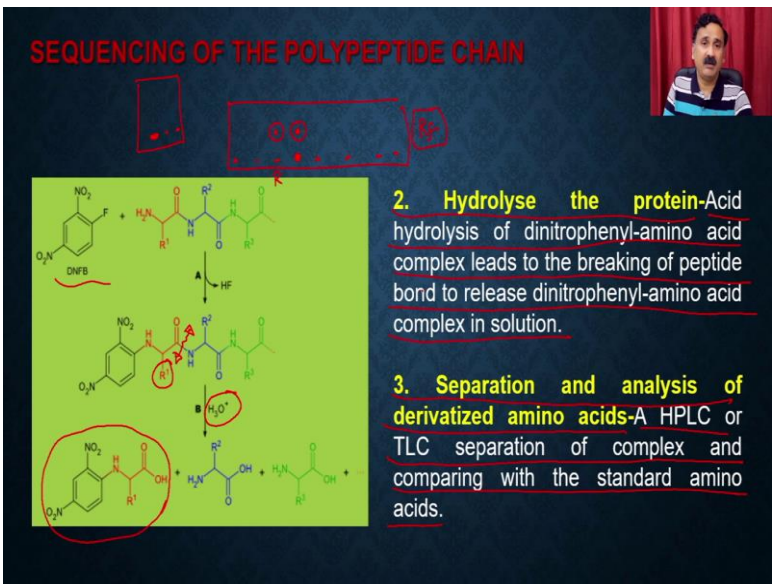
Now, in the stage 3, the stage 3, you are going to do the sequencing of the polypeptide chain. So, these are the you are going to have multiple polypeptide chains. So, for the sequencing of the polypeptide chain, you can have the two methods. One is you can use the Sanger's method, or you can use the Edman degradation method. So, let us first discuss about the Sanger's method.

So, once the polypeptide fragments are generated, we can start the sequencing of the each polypeptide chain, it has the following steps. So, first thing is, first thing is you have to identify the N-terminus residue, the N terminal amino acid analysis is being performed in the three step. Number 1, you are actually going to label the terminal amino acids. So, as I said, when we were talking about the primary structure, so, primary structure has the internal thing and then it has the C-terminal.

So, what we are doing is, we are first sequencing the protein from the N-terminus. So, first amino acid we have to first, do the sequencing from the N-terminus. So, for the first amino acid we are just using the labelling. So, we are labelling the terminal amino acids. So, the chemical reaction is performed to label the terminal amino acid with the compound such as the sanger reagents like 1-fluoro-2-4-dinitrogrnzene, or DFNB, and the dansyl chloride. In most of the cases these reagents also label the free amino acid which are present on the basic amino acids sidechain such as lysine and arginine.

Dinitrofluorobenzene reacts with the free amino group to form the dinitrophenyl-amino amino acid complex. So, what we are going to do is, you are going to take the DFNB and then you if you add DFNB to the first amino acid, because it has the free amino group, it is actually going to and in the presence of THF so there will be a release of this group. And then it is actually going to form a bond with the terminal amino groups. And that is how the first amino acid R1 is actually going to be labelled.

(Refer Slide Time: 19:00)



Now, what the step 2, what you are going to do is in the step 2, you are going to hydrolyse the protein to show that. When you add the acid hydrolysis of the dinitrophenyl-amino amino

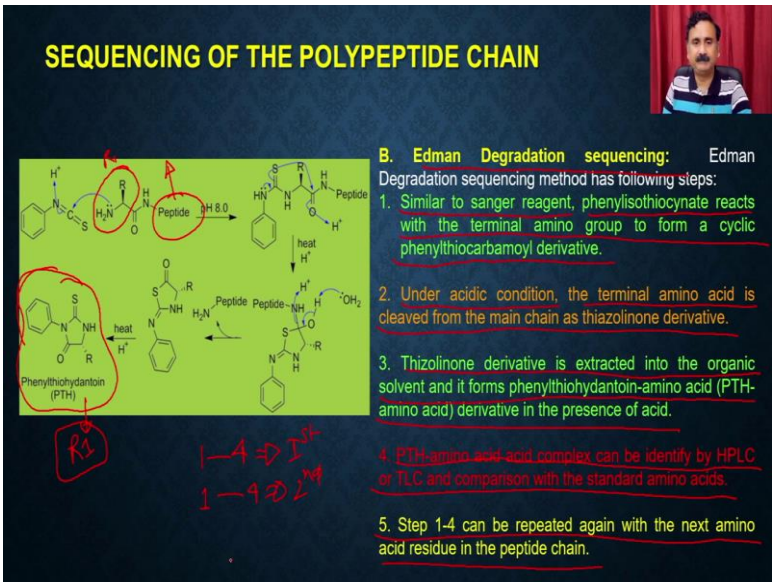
acid complex that lead to the breaking of the peptide bond to release the dinitrophenyl-amino amino acid complex in the solutions. So, after this, once the first amino acid is being labelled, then you are going to do the acid hydrolysis. And as when you do the acid hydrolysis, it is actually going to break the bond between the first amino acid and the second amino acid. And as a result, the first amino acid which is already been labelled with the DFMB. It is going to be released from the main chain.

Then you are going to do the separation and the analysis of the desired amino acid. So, A HPLC, or the TLC separation of the complex and comparing it with the standard amino acid is actually going to give you the name as well as the identify of this. So, what you are going to do is once you got this amino acid, then you can actually be able to run the TLC along with the standard TLC. So, what you can do is like for example, you can run a TLC like this.

So, you can run all the 20 amino acids. So, you can actually make the all the 20 amino acid in the DFMB complex and then you can run, and then you can also run the unknown sample. So, if you run the unknown sample, it is supposed it goes to the spot here. And suppose this is the arginine. So, if it goes to this, then you can say that, the this is the arginine or you can actually be able to calculate the Rf value of your unknown sample.

And since you know the Rf value of all the other amino acid complexes, you can be able to identify this. The other option approaches that you can do the HPLC and you can be able to calculate the retention values. So, this is about the how you can be able to use the Sanger method to sequence proteins.

(Refer Slide Time: 20:59)



Now, if you talk about the Edman degradation method. So, in the Edman degradation method, it also has a similar kind of steps. For example, the similar to the Sanger reagents, the reagents are different like where you are actually going to use the phenylisocyanate reacts with the terminal amino group to form a cyclized phenylthiocarbamoyl derivative.

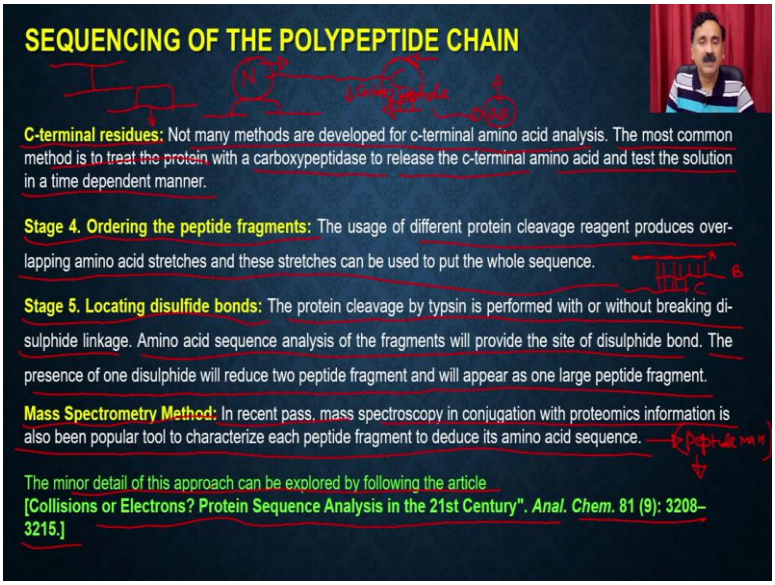
So, in this case, you are going to use the phenylisocyanate. And when it reacts with the terminal R1 group, onto the peptide, then it is actually going to form a cyclized product. Under the acidic conditions, the terminal amino group is actually going to be cleaved from the main chain as a thiazolinone derivatives.

And that is how you are actually going to have the first amino acid as the PTH. So, thiazolinone derivative is extracted into the organic solvent and it forms the phenylthiohyddantoin-amino acid PTH-amino acid complex into the presence of acid. So, ultimately, you are going to get the PTH complexes. So, this is for the PTH complex of the R1. And then what you are going to do is, you are going to run the PTH-amino acid complex, can be identified by the HPLC or TLC in comparison to the standard amino acid.

Now, once you have done this 1 to 4. For the first amino acid, so when you do the first step 1 to 4, for the first amino acid, you can do the 1 to 4 again for the second amino acid. Because the first amino acid is been released, so the remaining peptide chain is still there, you can use that and again do another round of this.

So, if you continue this like this, it is actually going to keep giving you the amino acid sequence from the N-terminus side. So, if you step 1 to 4, can be repeated for the next amino acids in the polypeptide chain. And that is how it is actually going to give you the whole sequence, and that you sequence if you have the different fragments, if you put them together it is actually going to give you the complete sequence of the all the amino protein peptides.

(Refer Slide Time: 23:06)



Now, as I said, the protein is having the two chain type. So, you have the N-terminus side and then you have the C-terminus side. So, we have just discussed the method like the Sanger method or the Edman degradation method to identify the N-terminus amino acids, but we can also do the sequencing for the C-terminus. And that is how you can be able to identify the C-terminal residues.

So, how we can do that, the C-terminal residues not many methods are developed for the C-terminal amino acid analysis. The most common method is to treat the protein with the carboxypeptidase, to release the C-terminal amino acid and test the solution in a timely dependent manner.

So, what you can do is, you can just treat this with a carboxypeptidase. So, carboxypeptidase is a specific enzyme, which actually releases amino acid from the C-terminal side, rather than the N-terminal side. So, if you use the carboxypeptidase, it is actually going to release the amino acid and, these amino acids are the C-terminal amino acid.

So, once the amino acid is released, you can identify that amino acid by the Sanger, or the Edman degradation methods. Then the stage 4, you are going to do the ordering of the peptide fragments. So, usage of the different peptide cleavage reagents produces the overlapping amino acid stretches, and these stitches can be used to put the whole sequence.

For example, when you generate that trypsin, it is going to generate the first fragment like this, the second fragment like this, the third fragment like this, so if you sequence this

fragment, if you sequence the A fragment, if you sequence the B fragment, if you sequence the C fragment.

So, what you see here is, that A is having this portion which is overlapping. C is having this portion which is overlapping with the B. So, by doing this overlapping sequencing, you can be able to deduce the final sequence of the final length, or the you can be able to put them these fragments, and that is how you can be able to do the sequencing,

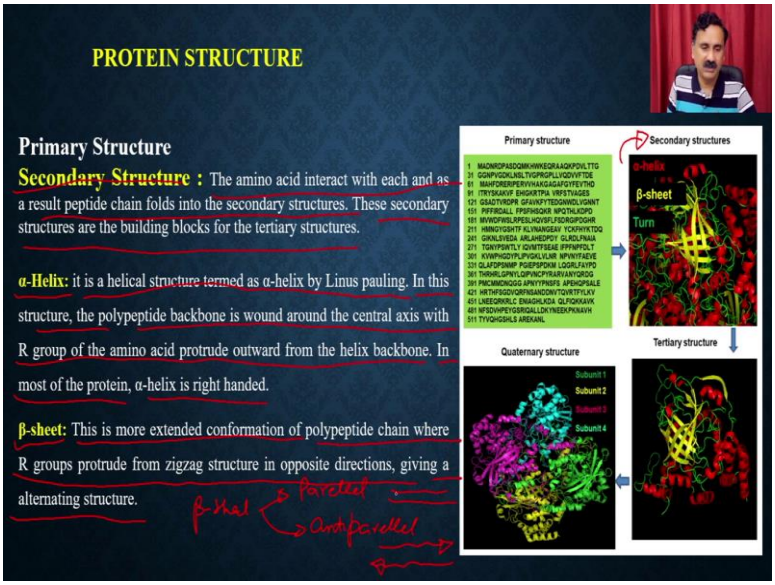
Then the stage 5, you can actually be able to locate the disulphide bonds. So, the peptide cleaved by, the protein cleaved by the trypsin is performed with or without breaking the disulphide linkages. Amino acid sequence analysis of fragment will provide the side of the disulphide bond. The presence of a disulphide bond will reduce the two fragments, will reduce two fragment and will appear as a single large fragment. As we said if there is a disulphide linkage present, it is not going to allow the release of the amino acid fragments.

So, because of that, if there is a suppose, this is the disulphide linkages and so, even if you cleave the this particular fragment. So, for example, if there is a disulphide linkage like this, if there is a disulphide linkage like this, if you cleave this with the protease it is it ideally should give you the two factor.

But if there is a disulphide linkage, it is still be going to bind like this. And that is how it is actually going to give you the single fragment. So, if you get the single fragment, then there is a disulphide linkage which is present. Apart from these kinds of methods, the way and with the advancement of the mass, mass data of different types of peptide fragments, and all this kind of thing, the mass spectrometry method is also been used.

So, in the recent past, the mass spectrometry method in conjugation with the proteomics information, it is also been a popular tool to characterize each fragment to deduce its amino acids. So, in this mass spectrometry data, what you are going to do is, instead of doing this sequencing, what you can do is you can just simply calculate the peptide mass, and that peptide mass actually, because there is a complete database can be used to deduce the sequence, I have given you the reference. So, if you are interested more about reading the protein sequencing, you can actually be able to read through this particular reference.

(Refer Slide Time: 27:17)

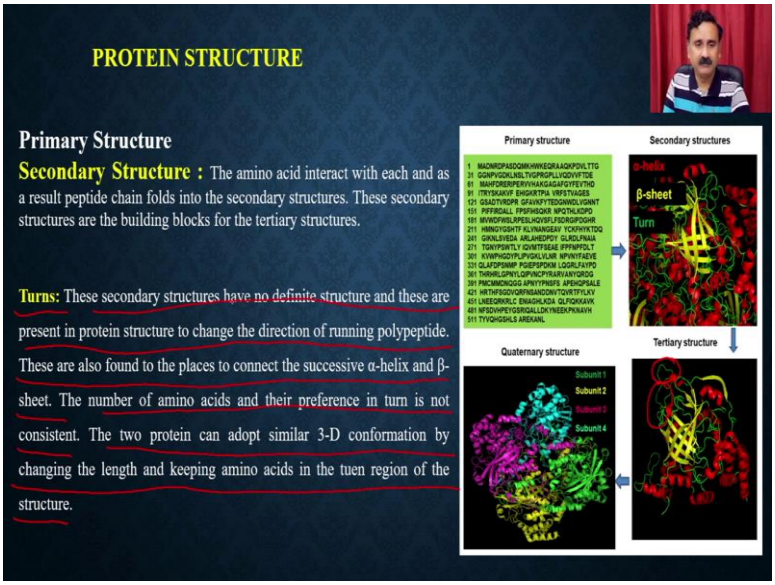


Now, let us talk about the secondary structures. So, secondary structures the amino acid interact with each other and as a result, the peptide chain folds into the secondary structures, these secondary structures are the building blocks for the tertiary structure. So, these are the primary structure, when they fold, they give you the secondary structures.

Secondary structures could be of two types, either it can be alpha-helix or to the beta-sheets. It is a helical structure termed as the alpha-helix by the Linus Pauling. In this structure, the polypeptide backbone is wound around a central axis with the R group of the amino acid protrude outward from the helix bond band, in most of the protein the helix is handed, which means you are actually going to see the helix, and in the helix the R groups are protruding outside.

So, then we have the beta-sheets, this is a more extended conformation of the polypeptide chain, where the R groups protrude from the zigzag structure in the opposite direction giving a alternate structure. So, beta-sheets could be the two types, it could be the parallel beta-sheets, or the anti-parallel beta-sheets. So, it can be parallel or the anti-parallel beta-sheets. Which means, either the beta-sheets are running in the same direction, then it is called as the parallel beta-sheets, if they are running in the opposite direction, then it is called as the anti-parallel beta-sheets.

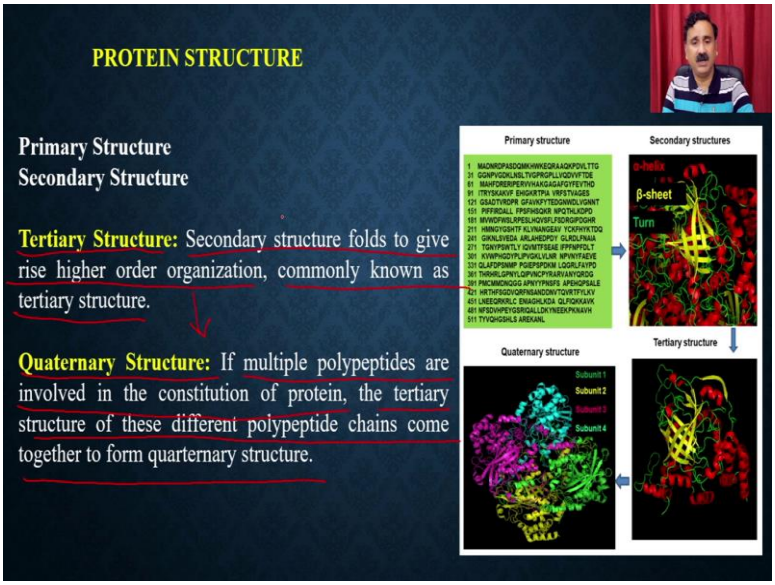
(Refer Slide Time: 28:47)



Then we have the turns. So, these secondary structures have no definite structure, and they are present in the protein structure to change the direction of the running polypeptide. These are also found two places to connect the successive alpha-helix and beta-sheets, the number of amino acid and their preference in turn is not consistent, the two proteins can adopt the similar 3-D conformation by changing the length and keeping the amino acid in the turn region of the structures. So, turn is actually a unstructured region, and it is actually not having a defined structure, but it has a very huge significance in terms of providing the flexibility of the different types of protein structures.

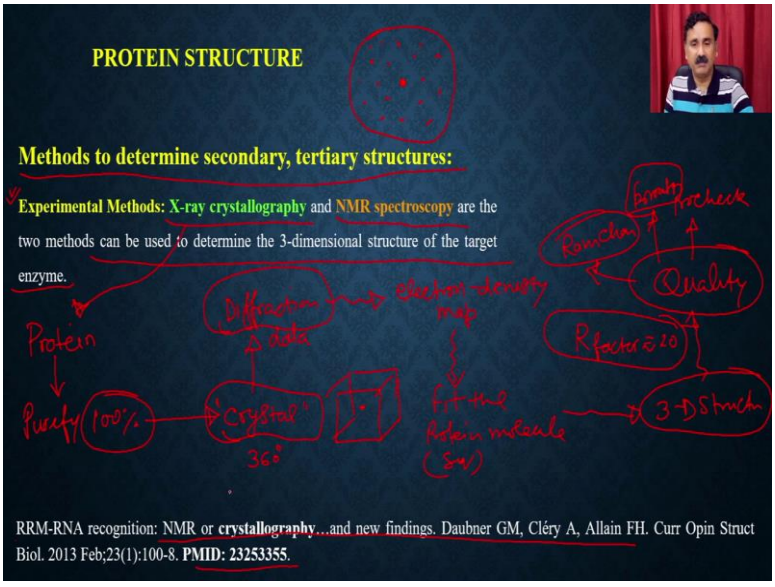
So, for example, you can have the two different types of protein structures their sequence, their amino acid sequence could be different, but they may adopt the identical structure simply by changing the length of the loops within the length. So, you can see that these are the loops. So, what you see here is this green coloured region is actually called as the loop, or the turn.

(Refer Slide Time: 29:56)



Then we have the tertiary structure So, tertiary structure, secondary structure folds to give the rise the higher order organizations commonly known as the tertiary structures, and then the tertiary structures can still be packed and that is how you can have the quaternary structure. So, if the multi polypeptide are involved in the constitution of the protein, the tertiary structure of these different polypeptide chain come together to form the quaternary structures. Now, as we discuss about the methods to determine the primary structure, we have discussed about the Sanger's method, and the Edman degradation method. We also would like to discuss about the methods to determine the secondary or the tertiary structures.

(Refer Slide Time: 30:41)



So, what are the methods to determine the secondary, or to the tertiary structures. So, so, there are two approaches, one is you can use the experimental methods. So, experimental methods, there are two methods, so, you can use the X-ray crystallography, or the NMR-spectroscopy.

And these are the two methods which you can use to determine the three-dimensional structure of the proteins. For the X-ray crystallography, you can actually be able to have the very different step. So, what you are going to do is, first you are actually going to isolate the protein, for which you are actually… you want to identify the secondary or the tertiary structures. Then what you going to do is, you are going to purify this protein at 100 percent purity or more than 90 percent purity.

And once you are done the purification, then what you are going to do is, you are actually going to crystallize this protein, so, you are going to produce the crystals, and once you produce the crystal, then you are actually going to put these crystals for the diffraction. And what you are going to get.

So, you are going to do the diffraction of these crystals and once you do the diffraction, it is going to give you the diffraction patterns. So, what if diffraction pattern? Diffraction pattern is actually going to give you the spots around the axis. So, it is going to give you the whereabout the diffracted X-ray beam has hit the film. And that is why, it is actually going to give you a diffraction pattern, and ideally when you are want to collect the complete diffraction pattern of a protein, it has to be rotated this this crystal has to be rotated for angle of 360 degree.

Because you can imagine, this is a crystal. You can have, and this crystal actually has to be rotated for 360 degrees, then only you can be able to collect the diffraction pattern of, or diffraction of all the electrons what is present. Once you have the diffraction pattern or the diffraction data from the X-ray, then you are actually going to put that and you would going to analyse the diffraction data, and that is how you are actually going to get the electron density map. Once you collect the electron density map, it is actually going to give you the position of the electrons within the three-dimensional.

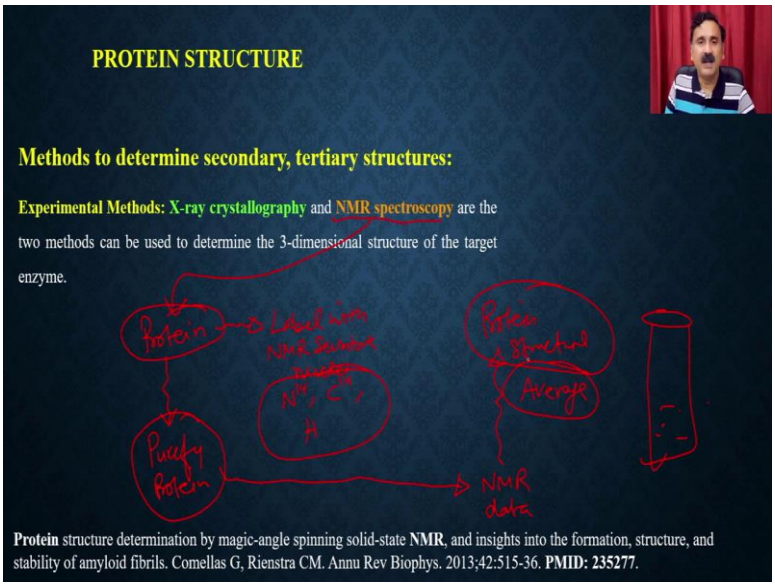
So, once you got the electron density map, then you are actually going to fit the protein molecules. So, you can do fit protein molecules, proteins molecule, protein, which means the protein sequence you are going to fit the protein sequence. And once you are done the fitting, then it is actually going to give you the 3-D structure of the protein. Once you got the 3-D

structure of the protein, then you can actually do the quality assessment, you can do the quality assessment with the help of the three programs. You can do the Ramachandran plot, you can do the Procheck, and you can also do the Errata plot.

So, if you do all these kinds of thing, it is actually going to tell you whether the 3-D structure what you have solved by fitting the protein sequence into the electron density map is correct or not, if you get the R factor, which is called as the error factor, which is approximately around 20. Then you are going to say that the protein what you have solved, or the protein structure, what you have solved using the X-ray crystallography is very good.

If you want to read more about, since this course is not about the X-ray crystallography, you will find a very good that kind of course, on in when in the books or SLS, you can actually get lot of good resources if you are interested to understand each and every detail about the X-ray crystallography. You can actually be able to even go through with this particular article and that actually will give you the very good idea about the X-ray crystallography as well as the NMR spectroscopy.

(Refer Slide Time: 34:56)

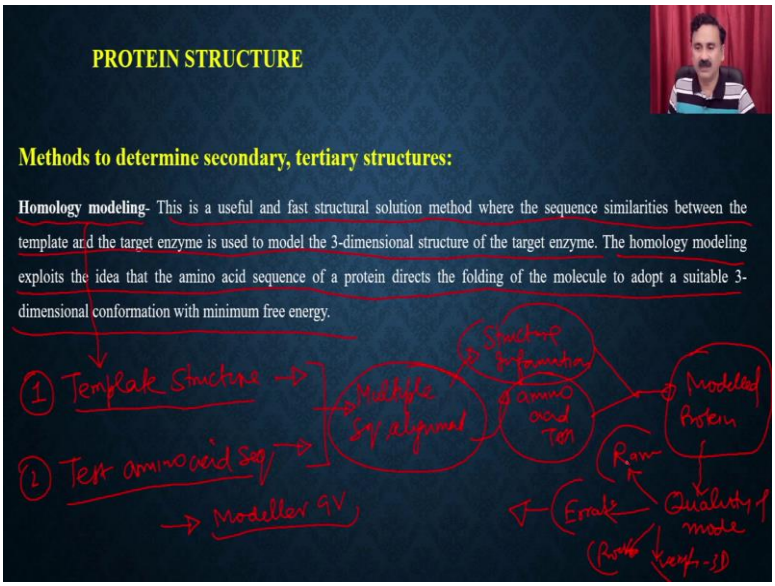


Same to this NMR spectroscopy also has the different type of steps, where you are first going to use the protein, you are going to first. The first step is you are going to purify the protein. And that the normal protein but we are going to get from the bacteria is not going to be labelled, not going to be NMR sensitive because it is so, so, first thing is you are actually going to label the protein with the NMR sensitive nuclei's.

For example, you can use the N14 Nitrogen's, you can use the Carbon 14. And you can also use the hydrogen like the deuterium and so on. And because of that the purified protein what you are going to produce is going to be NMR sensitive, and then you are actually going to collect the NMR data, and once you are going to collect the NMR data, and you are going to analyse that data it is actually going to give you the protein structures.

And that protein structure is actually going to be called as the average structure. So, this protein structure is called as average structure, because the NMR is actually going to be performed in the liquid. So, you can imagine that if I have a protein into the test tube, so this protein actually is freely moving, and its domain and all other kinds of structures are also moving, because of that it is actually going to give me the average structures.

(Refer Slide Time: 36:37)



Apart from that, you can also have the non-experimental matters, such as you can also do the homology modeling. So, this is a useful in a fast-structural solution method where the sequence similarity between the template and the target enzyme is used to model that 3-D structure of the target enzyme.

The homology modelling exploits the idea that the amino acid sequence of a protein directs the folding of a molecule to adopt the suitable three-dimensional conformation with the minimum energy. So, what you are going to do is, you are going to take, so in this homology modeling, homology modeling depends on the reliability of or depend on the phenomena that the two proteins when they are actually having the similar kind of amino acids, they are actually going to adopt the similar type of folds.

Because that the primary structure is actually going to direct the folding of these amino acids. And that is why, the even the primary structures or when the primary structure or the amino acid sequence is identical, it is actually going to fall into the same scene, which means if you want to use the homology model, you are actually going to have the two things, you are going to have a template structure.

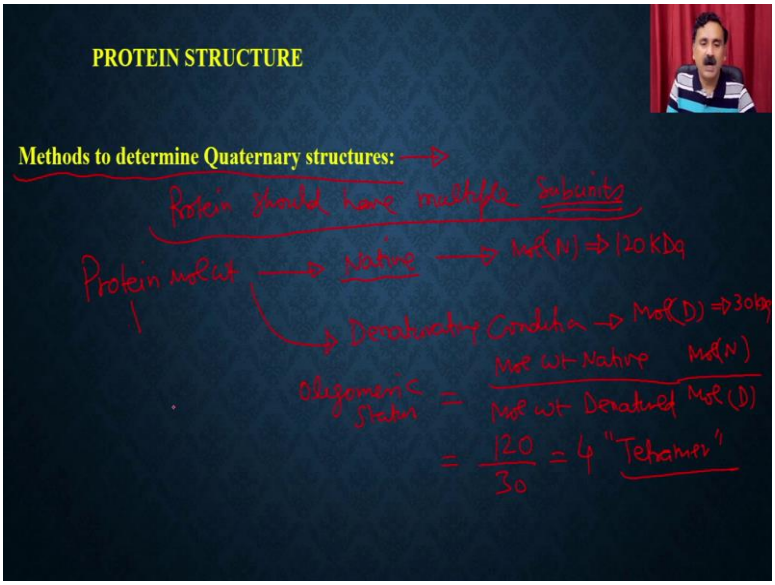
So, you going to have a template structure. And you are also going to have the test amino acid sequence. Now, first what you have to do is you are actually going to use this particular template, and you are going to use this sequence and you are going to do the multiple sequence alignment, so you are going to do the sequence alignment.

And this sequence alignment is actually going to tell you whether this particular template is good for modeling this particular amino acid sequence or not, once it is done, then you are actually going to do two things, you are going to take the structure information from the template. And on this structure information, you are actually going to put the amino acid sequence, what you have from the test. And that is how you are actually going to prepare the modelled protein. So, once you got the model protein, then what you are going to do is, you are going to test the quality of the model.

And we have already discussed the quality of the test of the model can be done by Ramachandran plot, or you can use the Errata plot, or you can do the Procheck, or you can do the Verify-3D. So, these are the three, four different types of groups or the different types of programs, what you can use and you can be able to do the error measurements. Once you have found that the structure is good, then what you can do is you can actually be able to utilize this model for the different types of applications. The programs what you can use very affluently is called as the Modeller.

So, you can alternatively use the Modeller ninth version, and that is actually going to allow you to do the modeling. So, all these steps, so, you can actually be used with the help of the modeler, and then you can use the different types of programs to do the quality of the model verification with Ramachandran plot, and all these kinds of things.

(Refer Slide Time: 40:30)



So, now, will us move on to the methods to determine the quaternary structure. So, method to determine the quaternary structures. You know that the condition for the quaternary structure is, that the protein should have the multiple subunits. So, if the protein has the multiple subunits, then only you can be able to perform the quaternary structures.

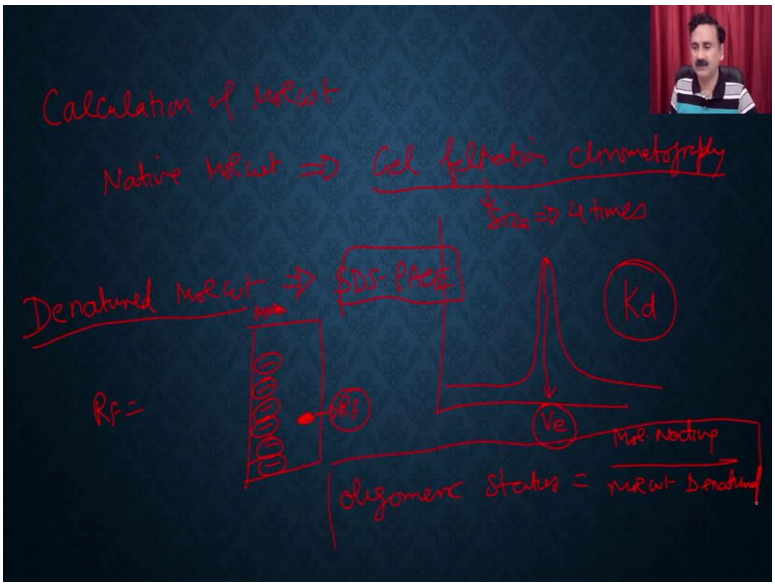
So, how we can actually be able to know that the protein has the multiple subunits? So, what you can do is, you can calculate the proteins molecular weight under the two condition. One, you can actually be able to calculate proteins molecular weight under the native conditions, or you can actually be able to calculate the proteins molecular weight under the denaturating conditions.

So, when you denaturating conditions, imagine that the molecular weight under the native condition is Mol N. So, it is N, and under the denaturating condition if the molecular weight is Mol D. Then, the you can actually be able to calculate the oligomeric status by the molecular weight native divided by molecular weight denature.

For example, so, like molecule N, versus molecule D. Let us take an example. For example, if I have calculated molecular weight and if I calculated the native molecular weight is 120 KDa, and if it I have calculated the denatured molecular weight which is the 30 KDa. Then the oligomeric status would be the 120 divided by 30 which means the 4, which means it is a tetramer.

So, once I calculated that it is actually a tetramer then, I can be able to show that there is a quaternary structure, what is present. Now, the question comes how you can be able to calculate the molecular weight of a protein under the native or the denaturating conditions?

(Refer Slide Time: 43:00)



So, calculation of molecular weight. So, for the native molecular weight, what you can do is you can run the protein under the gel filtration chromatography. So, if you do the gel filtration chromatography, although the scope does not allow you to explain the gel filtration chromatography, but gel filtration chromatography is a chromatography technique, which is actually going to filter the molecule based on the size.

So, if it is based on the size, so, if it is going to be four times because you are taking the one monomer, and you are making the four monomers, the size is going to be four times. So, that is why it is actually going to give you a pattern, it is going to give you a peak, which is actually going to tell you that okay, this is the size at which the protein is eluted.

So, this is called as the illusion volume, and utilizing this information, and as well as the distribution coefficient, you can be able to calculate the molecular weight of this particular protein. And that is going to be called as native molecular weight. And now, under what how to calculate the denatured molecular weight? you can be able to calculate the denatured molecular weight with the help of the SDS-PAGE. I am sure, we all know about the SDS-PAGE. So, when if you run the protein on the SDS-PAGE, it is actually going to give you the molecular weight. So, if you run first you run the molecular weight marker, so you run the marker, and then you are going to run your proteins.

So, depending on the and then you calculate the Rf values for the, these marker proteins. So, for the individual marker protein, when you calculate, you are actually going to get the Rf values for each and every spot, and then you can also be able to calculate the Rf value for the your protein, and depending on the Rf value of this particular spot, you can be able to calculate the denatured molecular weight.

I have already discussed this whole thing in a and so that is how you can be able to calculate the denatured molecular weight, and once you have the native molecular weight and you can have the denatured molecular weight. You can be able to utilize that for calculating the oligomeric status, which is the molecular weight native versus molecular weight denature.

So, with this, we have discussed about the protein structures, what we have discussed? We have discussed about the different types of organization, what is there in the protein structures, we have discussed about the primary structures, we have discussed about the secondary structures, we have discussed about the tertiary structures, and we have discussed about the quaternary structures.

We, what we have. So, with this we have discussed about the protein structures, and what we have discussed so far? We have discussed about the primary structures, secondary structure, tertiary structure, and quaternary structures, while we were discussing about the primary structures, we have also discussed about the method to determine the primary structure.

So, we have discussed about the Sanger's method, or the Edman's degradation method. So, both the methods are utilizing the similar approach, where you are actually going to label the terminal amino acid, and then you are going to use the acid hydrolysis. So, that the terminal amino acid is going to be released and that terminal amino acid, the labelled terminal amino acid going to be identified by running it on to the thin layer chromatography. And apart from that, we have also discussed about the secondary as well as the tertiary structures, we have discussed about the methods to determine the secondary as well as the tertiary structures.

And the lastly, we have also discuss about the homology modeling, we discuss about the X-ray as well as the NMR-spectroscopy. And lastly, we have also discussed about how you can be able to determine whether the protein is going to have the quaternary structures or not. So, with this, I would like to conclude my lecture here. In our subsequent lecture we are going to discuss some more aspects related to proteins.

So, with this I would like to conclude my lecture here. Thank you.