Molecular Biology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Module - 03
Basics of Biomolecules
Lecture-15 Protein

Hello everyone, this is Dr. Vishal Trivedi from department of biosciences and bioengineering IIT Guwahati. And what we were discussing we were discussing about the biomolecules 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 the amino acids and these amino acids are having a general structures 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 third side it has the hydrogen and the fourth side it has the functional side chains and based on the functional side chains the amino acid can be classified into four 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 acids, it could be a hydrophilic amino acids, it could be a polar amino acids, it could be negatively charged amino acid or it could be a positively charged amino acid.

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 analyze 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. So, as we said you know protein is made up 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 4 different types of functional groups and these 4 different types of functional groups and depending on the R side chains, you can have the 20 different types of amino acids. Peptide bond is connecting the 2 individual amino acids and that is how they are actually giving the polypeptide chain. Each amino acid is linked to the neighboring 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 end to the C terminals with a length of the several hundred amino acids. So, these are the 20 different amino acids, but we have already discussed in the previous lectures. So, the primary structure is defined as the amino acid sequence from the N terminals to the C terminals. So, do you see that it starts with the N terminals.

So, this is the N terminals 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 is why the protein has the two ends one is N terminals and the other one is called as the C terminals. The ordered folding of the polypeptide chain give rise to the 3D conformation known as the secondary structure of the protein such as the helix 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 terminals whereas the last amino acid is going to have the C terminals. When this primary structure is getting folded, it is actually going to give you the secondary structures 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 turns. So, these are the turns 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 defines 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 3 subunits, all 4 subunits are coming together and that is how you are going to have the quaternary structures.

So, different polypeptide chains are 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 3, 4 different types of structures. So, we have the primary structures, we have the secondary structures, we have the tertiary structure and we have the quaternary structure. So, these are the different level of organizations, what is present into the protein structures. So, let us first start with the primary structures.

So, primary structures, the amino acid sequence of a protein is known as the primary structure, the order of the amino acid determine 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 structures where you have the length of the amino acids. 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 structures and that event is called as the folding. So, when the primary structure is going to fold into a proper three dimensional conformations, then it is going to give you the secondary structures. 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 structures 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. 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 structures. So, it 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 small pieces. So, into small pieces and you are going to break the small pieces like this. So, they 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 terminides or the terminal amino acids.

So, you are going to have the labeling of the terminal amino acid in this way. So, you are going to have the labeling. So, labeling with the fluorescent dye in different and then you are going to do the sequencing, then you are going to identify that labeled 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 structures, then you are actually going to you know make it unfold.

So, you are going to use the different types of treatments like you are going to use react

with the FD and B and all that. So, that it is actually going to break the disulfide 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 state 1 this is the stage 1.

So, stage 1 you are going to convert the 3D conformations into the 1D or the primary structures. So, that you are going to achieve simply by the stage 1 where you are first going to break the disulfide linkages. So, stage 1 is the breaking of the disulfide bonds. So, you can imagine that this is the protein which has a disulfide linkages and the disulfide linkages interfere with the complete sequencing procedure as it does not allow the release of the cleaved amino acid from the peptide chain. There are 2 approaches to disrupt the disulfide linkages in a protein sequence.

In the first approach the protein is oxidized with a performic acid to produce the 2 cystic acid. So, why there is a need to break the disulfide linkages because if you do not break the disulfide linkages, even if this particular amino acid is actually going to be labeled and it is going to be hydrolyzed, 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 disulfide linkage and that is why it is important to break the disulfide linkages. So, disulfide linkages we have the 2 approaches in the approach 1 you can actually use the oxidation with the performic acid and that is when you do the performic acid treatment it is actually going to break the linkage between the disulfide linkages and that is how it is actually going to give you the 2 fragments. Whereas in the approach number 2, the protein is reduced by the DTT or beta mercaptoethanol to form the 2 cysteine followed by the treatment with the iodoacetate to form the carboxymethyl cysteine formation of the carboxymethyl cysteine is stopped the reformation of the disulfide bond. So, in the approach 2, what you are 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 disulfide linkages the S is S is actually going to get converted into SH and that is how it you are going to have the 2 peptide bonds where the disulfide bond is broken. But this has a problem because as soon as you have the SH and you have the reducing environment it is going to be remain as SH. But once it is actually going to be acquired the oxidizing environment again the S is going to be oxidized and again the SH is going to be get converted into the SS double bonds. So, to avoid that you are again going to react this with the carboxymethylene 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 when once this is done you can actually go back to the stage 2. In the stage 2 you are actually going to break the big polypeptide chain into the multiple fragments. So, in 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 recognition sequence and they cleave after a particular amino acid. So, this is one. So, stage 1 is this. So, stage 1 is over where you have actually destroyed the disulfide 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 use the trypsin enzyme the trypsin has the cutting side which is actually after the lysine or to the arginine which means wherever the lysine or the arginine is present for example, this is 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, the 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 can 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 methionine. For example, these are the peptide sequences what has been generated by the cyanogen bromide.

So, you see this is the methionine and after the methionine it has been cut by the cyanogen bromide. Now, once you got these small fragments, then what you can do is you can take the individual fragments and then you can sequence these small fragments. So, 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.

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 2 methods one is you can use the Sanger's method or you can use the Edmund-Diggard-Dahn 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 you have to identify the N-terminus residues, the N-terminal amino acid analysis is being performed in the 3 step. Number 1, you are actually going to label the terminal amino acids.

So, as I said, you know, when we were talking about the primary structure, so primary structure has the N-terminal thing and then it has a C-terminal thing. So, what we are doing is we are first sequencing the protein from the N-terminus. So first amino acid we have to first, you know, do the sequencing from the N-terminus. So, for the first amino acid we are just using the labeling. So, we are labeling the terminal amino acids.

So, the chemical reaction is performed to label the terminal amino acid with the compound such as the Sanger's reagents like 1-fluorot-2,4-dinitrobenzene or D-PhenB and the Denzyl chloride. In most of the cases these reagents also label the free amino acids which are present on the basic amino acid such as lysine and arginine. Dinitrofluorobenzene reacts with the free amino group to form the dinitropenyl amino acid complex. So, what we are going to do is you are going to take the D-FNB and then you if you add the D-PhenB to the first amino acid because it has the free amino group, it is actually going to and in the presence of the HF. So, there will be a release of the 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 labeled.

Now, what the step 2 what you are going to do is in the step 2 you are going to hydrolyze the peptoene. So, that when you add the acid hydrolysis of the dinitrophenyl amino acid complex that lead to the breaking of the peptide bond to release the dinitrophenyl amino acid complex in the solutions. So, after this once the first amino acid is been labeled 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 labeled with the D-PhenB it is going to be released from the main chain. Then you are going to do the separation and the analysis of the derived amino acids. 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 identity 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 D-PhenB complex and then you can run and then you can also run the unknown sample. So, if you run the unknown sample it is suppose it goes 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 approach is 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's method to sequence the proteins. 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 phenyl isothionate reacts with the terminal amino group to form a cyclized phenylthiocarbamoyl derivative. So, in this case, you are going to use the phenyl isothiocyanate and when it reacts with the terminal R1 group on to the peptide, then it is actually going to form a cyclized product. So, under the acidic conditions, the terminal amino group is actually going to be cleaved from the main chain as a thiazodiline derivatives and that is how you are actually going to have the first amino acid as the PTH. So, thiazoline derivative is extracted into the organic solvent and it forms the phenylthiodine 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, that is the remaining peptide chain is still there, you can use that and again do the 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 acid in the polypeptide chain and that is how it is actually going to give you the whole sequence and that the 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 protein peptides. Now as I said you know the protein is having the 2 chain types. 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's method or the Edmond degradation method to identify the N-terminus amino acids, but we can also do the sequencing from 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 the 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's or the Edmond 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 stretches can be used to put the whole sequence. For example, when you generate the 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 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 disulfide bonds. So, the peptide cleaved by the protein cleaved by the trypsin is performed with or without breaking the disulfide linkages. Amino acid sequence analysis of fragment will provide the side of the disulfide bond. Depressants of a disulfide bond will reduce the two fragments, will reduce two fragments and will appear as a single large fragments.

As we said, if there is a disulfide 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 disulfide linkages and so even if you cleave the this particular fragment. So, for example, if there is a disulfide linkage like this, if there is a disulfide linkage like this, if you cleave this with the protease, it is it ideally should give you the two fragments. But if there is a disulfide linkage, it is still be going to bind like this and that is how it is actually going to give you the single fragments. So, if you get the single fragment, then there is a disulfide

linkage which is present.

Apart from these kind of methods and with the advancement of the mass data of different types of peptide fragments and all those 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 is also been a popular tool to characterize the 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.

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 secondary structures. These secondary structures are the building blocks for the tertiary structure. So, these are the prime 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 linear spalling.

In this structure, the polypeptide backbone is wound around a central axis with the R group of the amino acid protrudes outward from the helix bond band. In most of the protein the helix is right 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 some more extended confirmation of the polypeptide chain where the R groups protrude from the zigzag structure in the opposite direction giving a alternate structure. 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. Then we have the turns. These secondary structures have no definite structures and they are present in the protein structure to change the direction of the running polypeptide. These are also found to 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 protein can adopt the similar 3D 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 definite 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 color region is actually called as the loop or the turn.

Then we have the tertiary structures. So, tertiary structures, secondary structures forced 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 multipolypeptides are involved in the concentration of the protein, the tertiary structure of these different polypeptide chain come together to form the quaternary structures. Now, as we discussed about the methods to determine the primary structures, we have discussed about the Sanger's method and the Edmond degradation method. We also would like to discuss about the methods to determine the secondary or the tertiary structures. So, what are the methods to determine the secondary or to the tertiary structures? 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 steps. So, what you are going to do is first you are actually going to isolate the protein what which you are actually identity you want to identify the secondary or the tertiary structures. Then what you are going to do is you are going to purify this protein at 100% purity or more than 90% purity and once you have done the purify 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 is diffraction pattern diffraction pattern is actually going to his spots around the axis so it is going to give you the wherever the diffracted x-ray beam has hit the film and that is how 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 rotate this crystal has to be rotated for an angle of 360 degrees because you can imagine this is the crystal you can have and this crystal actually has to be rotated for 360 degree 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 dust and you are going to analyze 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 the protein molecules proteins molecule 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 3D structure of the protein.

Once you got the 3D structure of the protein then you can actually do the quality assessment right you can do the quality assessment with the help of the three programs you can do the Ramchandran plot you can do the Pro-check and you can also do the Erato plot. So if you do all these kind of thing it is actually going to tell you whether the 3D 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 the books or as well as you can actually get a 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. Same to this NMR spectroscopy also has the different types 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 you know that the normal protein what we are going to get from the bacteria is not going to be labeled not going to be NMR sensitive because it is 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 you can use the carbon 14 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 analyze 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 kind of structures are also moving because of that it is actually going to give me the

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

The homology modeling exploits the idea that the amino acid sequence of a protein direct 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 a similar type of folds because you know that the primary structure is actually going to direct the folding of these amino acids and that is why the when the primary structures are when the primary structure or the amino acid sequence is identical, it is actually going to fold into the same shape, 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 are 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 this is done, then you are actually going to do two things, we 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 model 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 ratah plot or you can do the pro check or you can do the verify 3D. These are the three or 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 oftenly is called as the modeler. You can use the modeler 9th version and that is actually going to allow you to do the modeling. So all these steps 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 the Ramachandran plot and all these kind of things. So now let us move on to the methods to

determine the quaternary structures. 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 protein molecule weight under the 2 conditions. One you can actually be able to calculate the protein's molecule weight under the negative conditions or you can actually be able to calculate the protein molecule weight under the denaturating conditions. So, when you denaturate in conditions, imagine that the molecular weight under the native condition is mol n. So, it is n right and under the denaturate in condition if the molecular weight is mol d then you can actually be able to calculate the oligomeric status by the molecular weight native divided by molecular weight denature right.

For example, so like molecule n versus molecule d. Let us take an example. For example, if I have calculated the molecular weight and if I calculated the native molecular weight is 120 kDa and if 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 assure 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. So calculation of molecular weight right. 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 4 times because you are taking the 1 monomer and you are making the 4 monomers the size is going to be 4 times right. So that is why it is actually going to give you a pattern right 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 elution 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. 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 even if you run the protein on the SDS

page it is actually going to give you the molecular weight right. So you run first you run the molecular weight marker right 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 right. So if 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 right.

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 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 structures.

So, we have discussed about the Sanger's method or the Edman degradation methods. So, both the methods are utilizing the similar approach where you are actually going to label the terminal amino acids 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 labeled terminal acid going to be identified by running it onto 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 lastly, we have also discussed about the homology modeling we discussed 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 protein. So, with this, I would like to conclude my lecture here. Thank you.