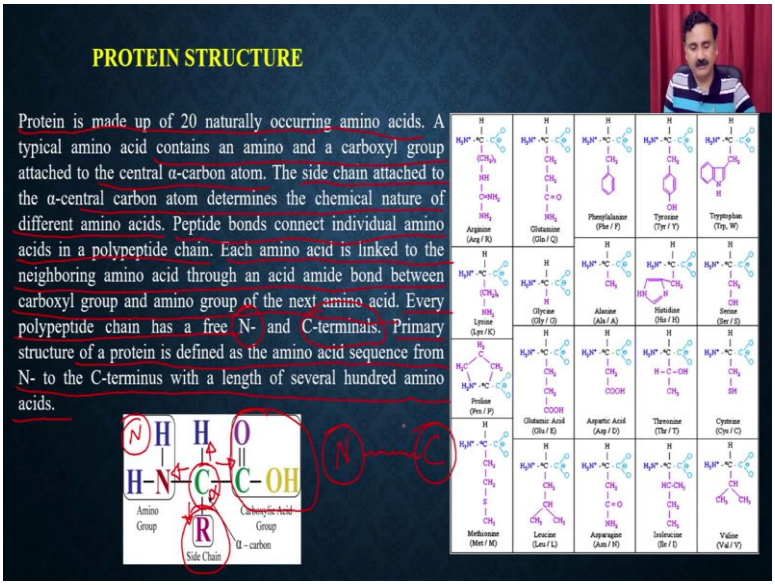
Enzyme Science and Technology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati

Module - II
Enzyme Structure
Lecture - 07
Determination of Primary Structure

Hello everyone. This is Dr. Vishal Trivedi from Department of Biosciences and Bioengineering, IIT, Guwahati. And what we were discussing in the course Enzyme Science and Technology, we are discussing about the Different Aspects of the Enzymes. In the previous module, we have discussed about the enzyme classification, nomenclature, and we have also discussed about how the field of enzymology developed.

And if you recall in the previous lecture, we have discuss about the enzyme structures. So, we have discuss about the primary structures, where we have discuss about the different types of amino acid, what are present in the enzyme structures and so on. So, in the today's lecture, we are going to discuss about the how you can be able to determine the protein structures.

(Refer Slide Time: 01:30)



So, as we said you know, protein is made up of the 20 naturally occurring amino acids. A typical amino acid contain a amino group and a carboxyl group attached to the central

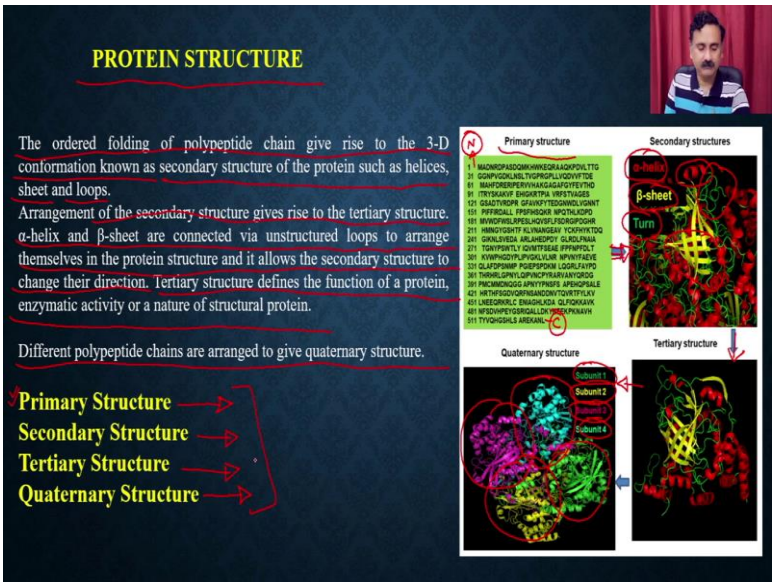
alpha carbon. The side chain attached to the central alpha carbon determine the chemical nature of the different amino acids.

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 type of functional 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 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, right. 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 100 amino acids. So, these are the 20 different amino acids 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, right.

So, did you see that its 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. Then, 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 I mean protein has the two ends, one is N-terminus and the other one is called as the C-terminus.

(Refer Slide Time: 03:27)



The ordered folding of the polypeptide chain give rise to the 3D 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, right. 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 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.

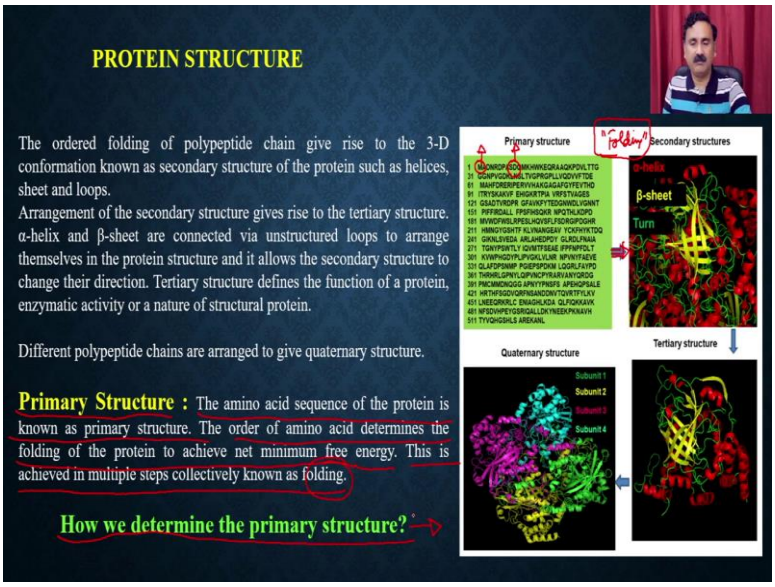
Arrangement of the secondary structure give rise to the tertiary structures. Alpha helix and beta sheets are connected by the unstructured loop to arrange themselves in the protein structure and it allows the secondary structure to change their directions. Tertiary structure defined the structure of 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 sub, so all sub 4 subunits are coming together and that is why you are going to have the quaternary structures.

So, different polypeptide chains are arranged to give the quaternary structure. So, different; 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.

(Refer Slide Time: 05:34)



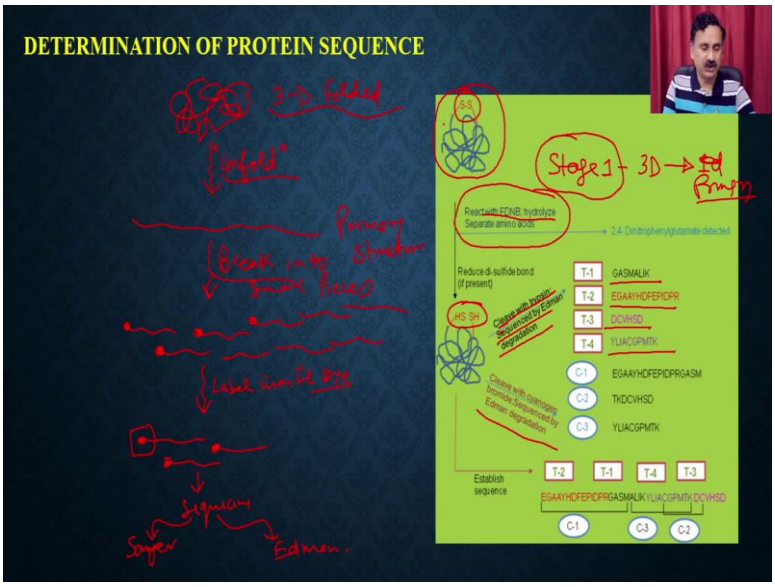
So, let us first start with the primary structure. So, primary structures, the amino acid sequence of a protein is known as the primary structure. The order of the amino acid determine the folding of the protein to achieve the net minimum free energy and this is achieved in the multiple steps collectively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

So, the primary structure is going to fold. So, this is what you see here is the primary structures, right, 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, right.

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, 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: 06:55)



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, right. So, this is the three-dimensionally folded protein, right. So, here you 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, right.

So, you it means you are actually going to unfold the protein by the chemical or the enzymatic method. Then, what you 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, right, so into small pieces. And the you are going to break the small pieces like this. So, they are going to be overlapping regions, right. So, that you can be able to put them these things separately.

And then what you are going to do is you are actually going to label the terminize or the terminal amino acids, right. So, you are going to have the labelling of the terminal amino acid in this way, right. 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, right, then you are going to identify that labelled amino acid.

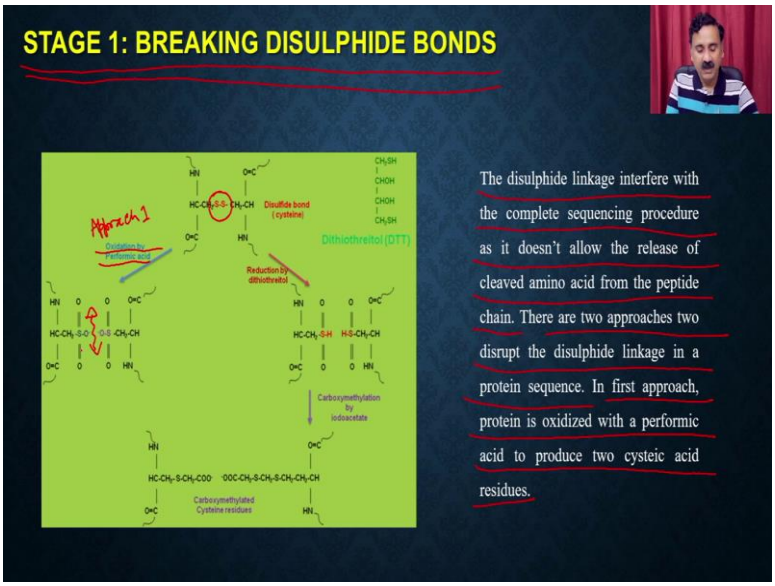
That sequencing you can do by two method, you can use the Sanger's Sequencing Method or you can use the Edman Degradation Method. So, this is what it is showing here, right. First you are going to start with the 3D fold structures, then you are actually going to you know make it unfolds.

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 state 1, this this is the stage 1, right. So, stage 1 you are going to convert the 3D conformations into the 1D or the primary structures, right. So, that you are going to achieve simply by the stage 1, where you are first going to break the disulphide linkages.

(Refer Slide Time: 09:58)

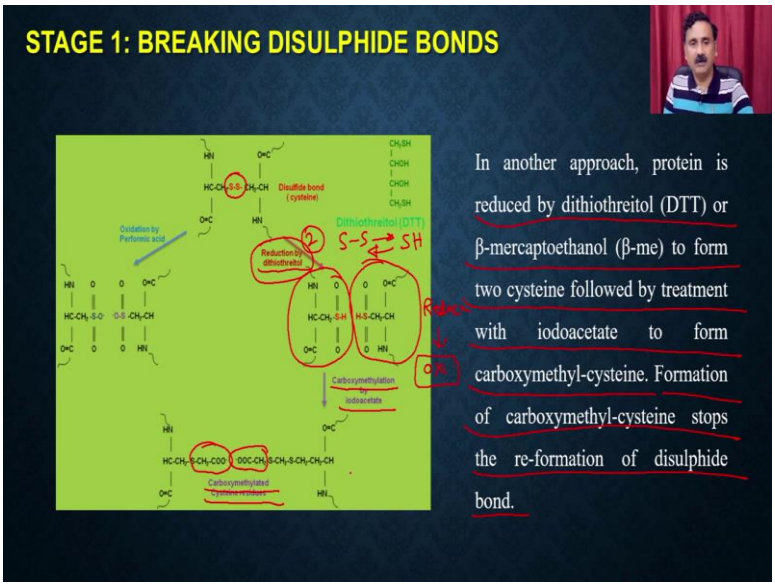


So, stage 1, stage 1 is the breaking of the disulphide bonds. So, the you can imagine that this is the protein, right which has a disulphide linkages and the disulphide linkages interfere with the complete sequencing procedure as it does not allow the release of the cleaved amino acid from the peptide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

There are 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 cysteic residue. 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 released from the main chain.

Because it is still having a it is bind to the main chain through a disulphide linkages. And that is why it is important to break the disulphide linkages. So, disulphide linkages we have the two 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 disulphide linkages and that is how it is actually going to give you the two fragments.

(Refer Slide Time: 11:20)



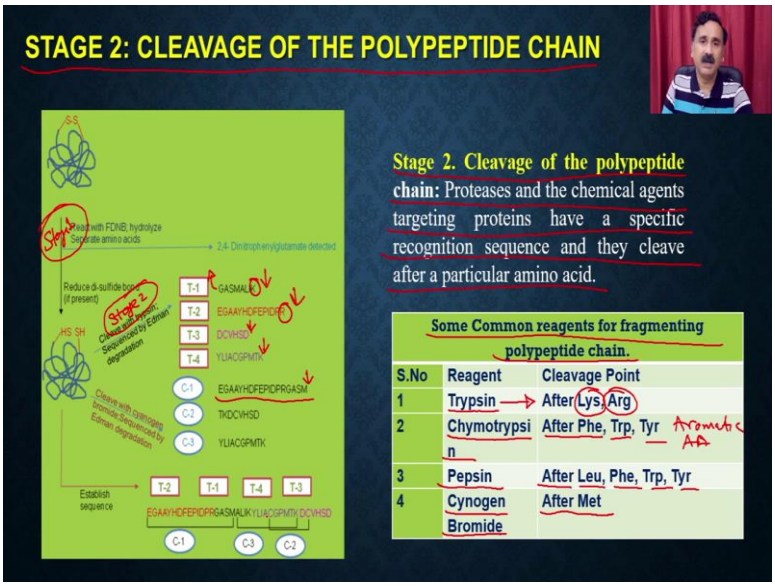
Whereas in the approach number 2, the protein is reduced by the DTT or beta mercaptoethanol to form the two cysteine followed by the treatment with the iodoacetate to form the carboxymethyl-cysteine. Formation of the carboxymethyl cysteine stop the reformation of the disulphide bond. So, in the approach two 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 disulphide linkages the S is S-S is actually going to get converted into SH, and that is how it you are going to have the two peptide bonds, where the disulphide 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 acquire the oxidizing environment, again the S-S is going to be oxidized and again the SH 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 carboxymethyl 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.

(Refer Slide Time: 13:02)



So, in the stage 2, stage 2 is the cleavage of the polypeptide chain. So, this stage 2 is the cleavage of the polypeptide chain, the protease and the chemical treatments targeting protein have a specific recognition sequence and they cleave after a particular amino acid. So, this is 1.

So, stage 1 is this, right. So, stage 1 is o, right, 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 a 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 see this these are the peptide which are being generated by the trypsin.

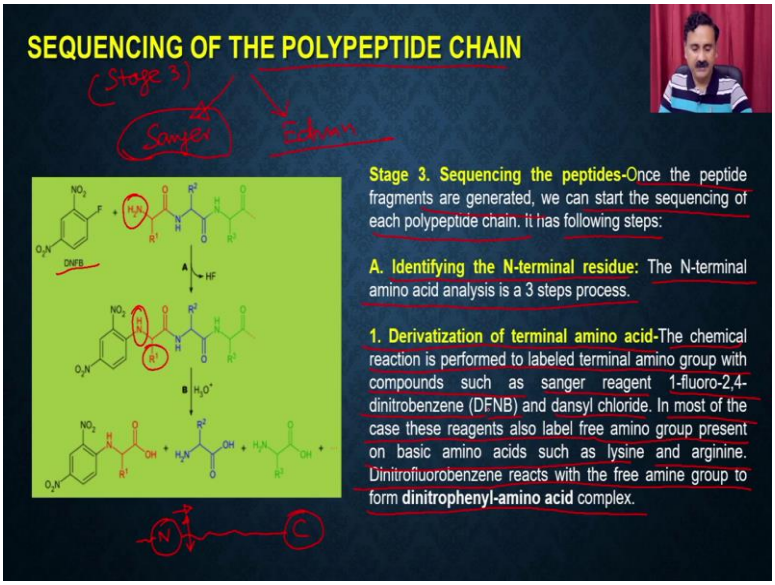
So, if you treat it with the trypsin, it is actually going to cut wherever you have the lysine or the arginine. So, for example, here it has cut, right here it has cut. So, wherever you have the lysine and arginine, it is actually going to cut. And that is why it is actually going to generate the different 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 poly peptide 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 the 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 why you are going to get the sequence of the complete proteins.

(Refer Slide Time: 15:25)



So, 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 a multiple polypeptide chains. So, for the sequencing of the polypeptide chain you can have the two method, one is you can use the Sanger's method or you can use the Edman degradation method, right. So, let us first discuss about the Sanger's method.

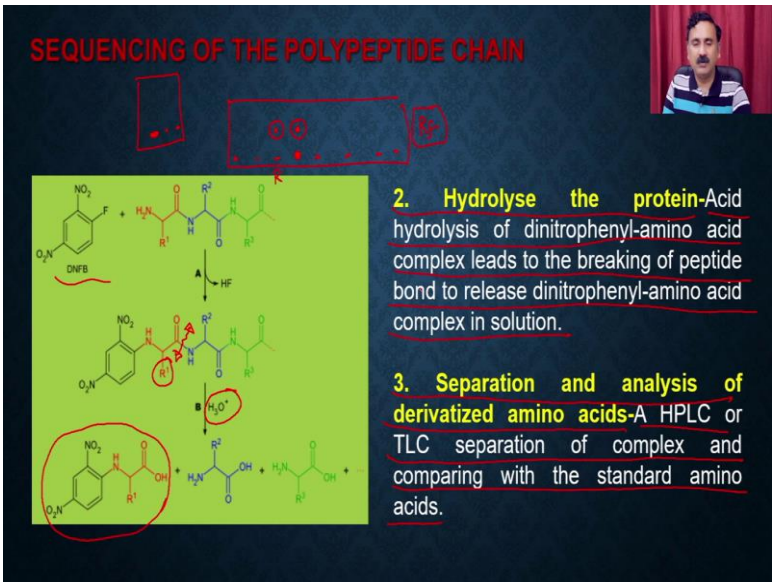
So, once the polypeptide fragments are generated, we can start the sequencing of the each polypeptide chain. It has the following step. So, first thing is first thing is you have to identify the N-terminus residues. The N terminal amino acid analysis is been performed in the 3 step. Number 1, you are actually going to label the terminal amino acids. So, as I said you know when the we were talking about the primary structure.

So, primary structure has the N-terminal thing and then it has a C-terminal thing, right. 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 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's reagents like 1-fluoro-2, 4 dinitrobenzene 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 acid side chain, such as lysine and arginine. Dinitrofluoro benzene reacts with the free amino group to form the dinitrophenyl-amino acid complex.

So, what we are going to do is you are going to take the DFNB, and then if you add the DFNB 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 R 1 is actually going to be labelled.

(Refer Slide Time: 17:42)



Now, the step 2, what you are going to do is in the step 2 you are going to hydrolyse the protein. So, so the 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.

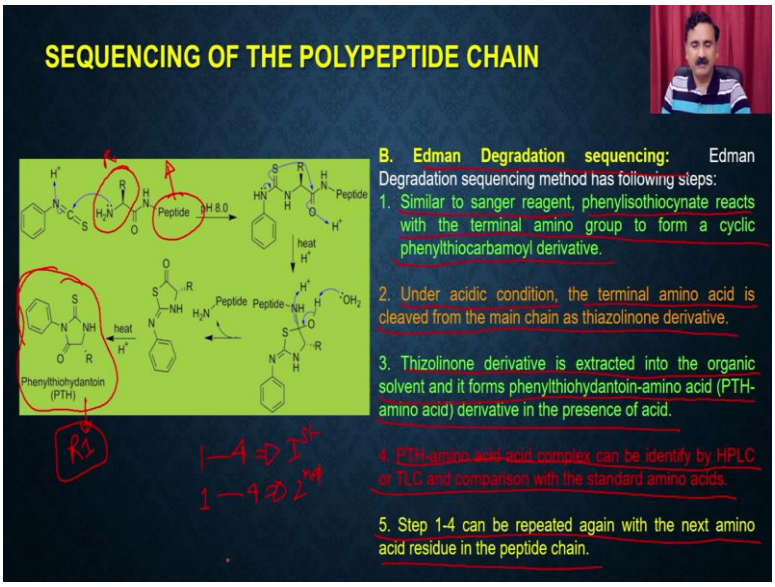
So, after this, once the first amino acid is been labelled, then you are going to do the acid hydrolysis and as when you do the acid hydrolysis, it is actually going to break the bond between the first amino acid and the second amino acid. And as a result, the first amino acid which is already been labelled with the DFNB is going to be released from the main chain.

Then, you are going to do the separation and the analysis of the derived domino 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 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 acid, right. So, you can actually make the all the 20 amino acid in the DFNB complex and then you can run and then you can also run the unknown sample, right. So, if you run the unknown sample, it is suppose it goes to the spot here and suppose this is the arginine, right.

So, if its 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.

(Refer Slide Time: 19:40)



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 isothiocyanate reacts with the terminal amino group to form a cyclized, phenyl thio carbamoyl derivative.

So, in this case, you are going to use the phenyl isothiocyanate. And when it reacts with the terminal R 1 group, right 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.

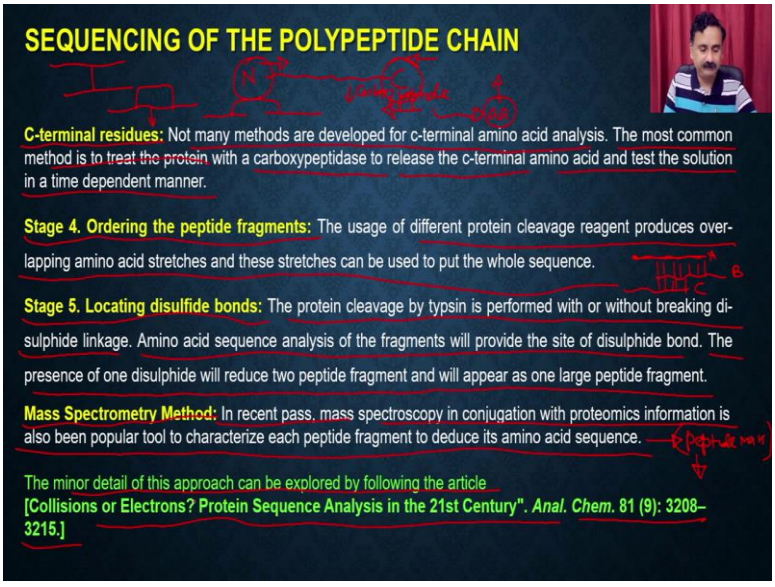
So, thiazolinone derivative is extracted into the organic solvent and it forms the phenyl thio hydantoin amino acid, PTH amino acid complex into the presence of acid. So, ultimately you are going to get the PTH complexes. So, this is for the PTH complex of

the R 1. 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, right. Because the first amino acid is been released, right, so 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 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.

(Refer Slide Time: 21:47)



Now, as I said you know the protein is having the two type. So, you have the N-terminal side and then you have the C-terminus side, right. So, we have just discussed the method like the Sanger's method or the Edman 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, right. 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 Edman degradation methods. Then the stage 4, you are going to do the ordering of the peptide fragments.

The, 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, right 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 sequent 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 put doing this overlapping sequencing, you can be able to reduce 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 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, right, will reduce two fragments and will appear as a single large fragments. As we said, right if there is a disulphide linkage present it is not going to allow the release of the amino acid fragment.

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

disulphide linkage it is still be going to bind, right 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 disulphide 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 the this particular reference.

So, what we have discussed so far? We have discussed about the primary structure, right. So, in the if you recall in the lecture 1, we have discussed about the properties of the amino acid which are present in the protein structure. And then we have also discussed how you can be able to generate the amino acid pool of a particular protein, and how you can be able to identify the different types of protein with the help of the thin layer chromatography.

So, in the thin layer chromatography, we have discussed that how you can be able to calculate the RF values, and how you can be able to use those values to determine a unknown amino acids which are present in the protein structure. Subsequent to that, in the current lecture we or in today's lecture, we have also discussed about the how you can be able to determine the amino acid sequence what is present in the protein structures.

So, we have discussed how you can be able to digest the proteins into the multiple fragments, so that the amino acids are going to be released. And then, how you can be able to do the different types of coupling reaction, so that you can be able to label that amino acids. And then, how you can be able to purify that labelled amino acid to determine the particular amino acids.

So, with this, I would like to conclude my lecture here. And in our subsequent lecture, we are going to discuss about the secondary structures. And in that, we are going to discuss about the alpha helix and beta sheets, and how you can be able to use the different types of methods to determine the secondary structures. So, with this, I would like to conclude my lecture here.

Thank you.