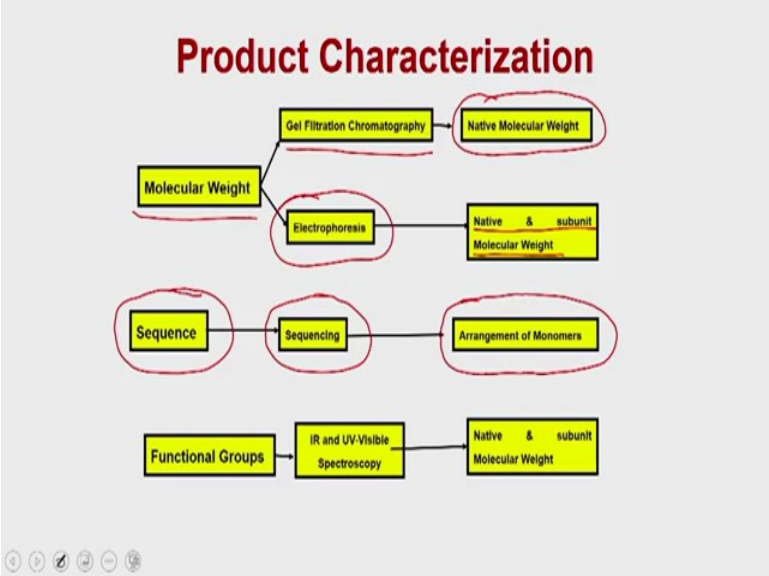
Genetic Engineering: Theory and Applications
Professor Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati, Assam, India
Lecture 32
Protein Sequencing

Hello everybody, this is Dr. Vishal Trivedi from the Department of Biosciences and Bioengineering IIT Guwahati. And in this module we were discussing, we were discussing about the characterization of the over express products which you are getting from the host cells and in this context, so far what we have discuss? We have discussed about the identification or the determination of the molecular weight.

(Refer Slide Time: 0:56)



What we have discussed so far? We have discussed about how to determine the molecular weight of the isolated product. It either it could be by Gel filtration chromatography which will give you the native molecular weight or it could be by the electrophoresis which will actually be going to give you native as well as the sub-unit molecular weight. In addition, while we were discussing about the electrophoresis we have also discussed about what are the different applications of electrophoresis.

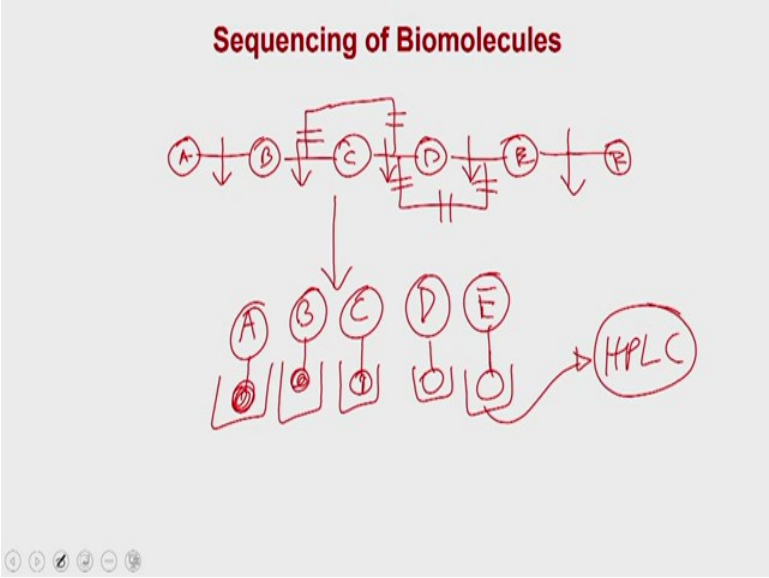
And in that context we have also discuss how the electrophoresis can be used in conjugation with the electro blotting techniques to determine or to check whether the isolated product has the any kind of amino reactivity with the antibodies or the developed antibodies. Now the

next thing what you can do in this series is, you can deduce the sequence of the biomolecules and that will be always been done by the sequencing techniques. That sequencing techniques will always going to tell you about the arrangement of the monomer.

Because in the case of biomolecule, whether it is a protein or the DNA it is not important that you have all the functional group present. But it is also important that in what sequence these functional groups are being arranged. So if you remember in the previous lectures while we were taking of the confirmation of the clones we have discuss about how to sequence the DNA or the recommitment DNA what you are going to produce so that you can be able to use that DNA sequencing techniques to verify the clone DNA.

Now in todays lecture we are going to discuss about how you can sequence the proteins. So whether it is protein or the DNA, the sequencing techniques for a biomolecule always follow a flowchart and when you talk about the biomolecules and the arrangement of the monomers you have to have a schematic way of doing it. So that you will not be able to miss out any residue and on the other hand you should be able to deduce the sequence in the right orientations.

(Refer Slide Time: 3:17)



So you can imagine that you have a biomolecule, so all these balls what you see is the monomer which are being associated with each other. So and in addition to that sometime the biomolecules are also being having the intermolecular bonding. Sot his intra-molecular bondings also are very important in especially when you are talking about in terms of the

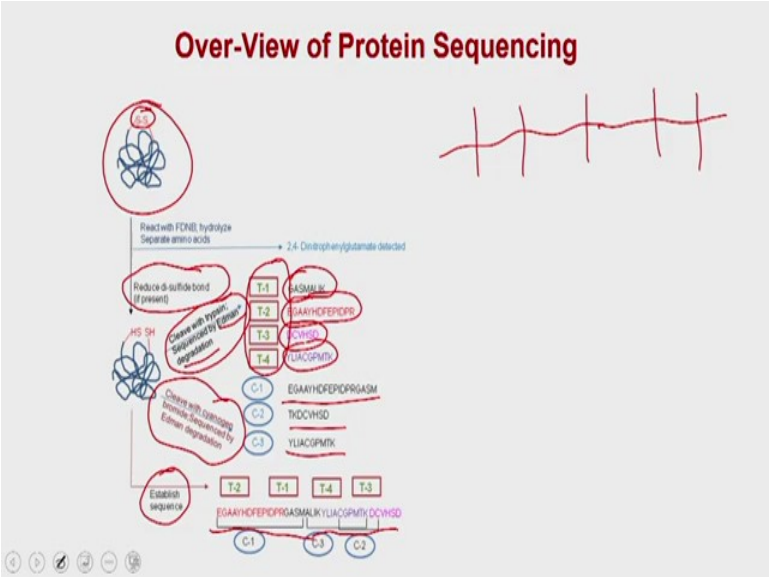
proteins. So if you would like to sequence this biomolecule which is actually made up of these many substituents or the monomers.

What you are going to do is, first you are have to destroy all these bondings which are actually intra-molecular binding and which you are going to destroy. The second thing is you also have to cleave this particular biomolecule at every amino acids or every monomer at the end of every monomer. So what will happen is, that actually suppose this is A, this is B, this is C, this is D, this is E, this is F, so if you cleave after everything, what you are going to get?

You are going to get a combination of all these alphabets into the solution. Now the next step is that you are going to capture these alphabets, so what you are going to do is, you are going to put some kind of reactive groups so that it will react with A and form a complex. It will also going to react with B, it will also going to react with C, it will going to react with D and it will going to react with E.

Now what you going to do is? You are going to isolate the individual A, B, C, D which are modified and then by using the HPLC techniques you will be able to deduce the sequence.

(Refer Slide Time: 5:15)



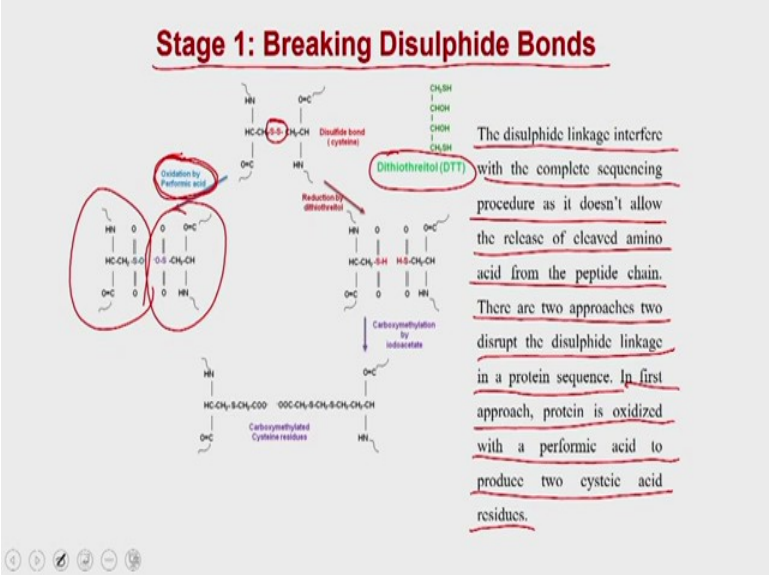
So now let us take us overview of what are the things you have to do when you would like to do the protein sequencing? So you know that protein is a very-very complex molecule it also has the peptide chains which are connecting the monomeric amino acids and on the other hand is the sulfide bridges which are actually forming the intra-molecular bondings. And as I said in the previous slide that the first thing is that you have destroy the disulfide linkages.

So you have to destroy disulfide linkages simply by the reduction. And once you destroy the disulfide linkage it is actually going to have a linear chain of polypeptide chain, then what you are going to do is, you are going to cleave this polypeptide chain with the help of the different types of the trypsin or different types of cleavage agents. It could be proteolytic enzymes or it could be chemical reagents.

Then you are going to have the option of doing the sequencing either by Edman degradation method or this, so then you will do the sequencing and that will actually going to give you the sequence of all these small fragments which you are going to generate after the cleavage.

Now all these and on the other hand if you would use the chemical molecules that such as cyanogen bromide that is also going to give you a different types of fragments and all these fragments then you are going to put them together with the help of the different types of software and that actually is going to give you the original sequence. So this is what we are going to discuss today. And in following this story, the first step is that you have to destroy the disulfide linkages.

(Refer Slide Time: 7:09)

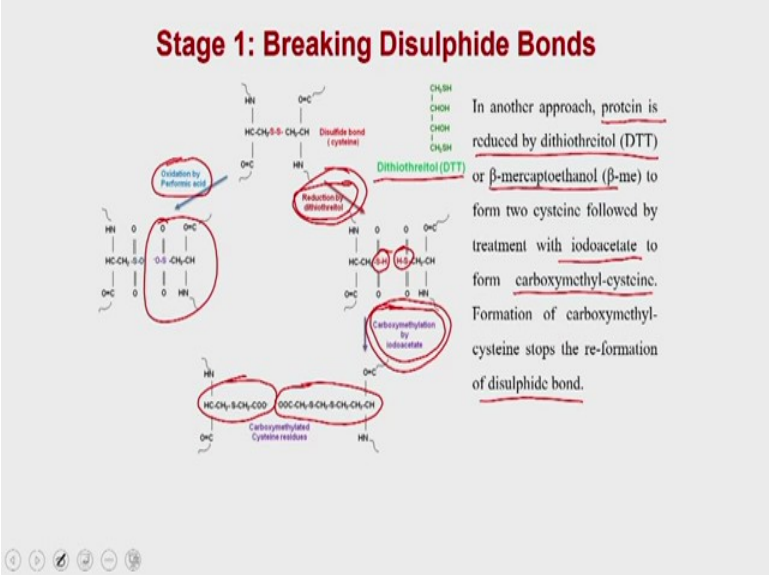


So in the stage one, when you start with the protein you have to break the disulfide bonds. So disulfide bonds are always being present between the two system residues and what you are going to do, you are have the two options or the two approaches and why there is the requirement of linking of disulfide linkages? Because disulfide linkages always interfere in terms of the final sequencing or setting or actually deducing the sequences because it does not allow your monomeric sequences to put together.

And on the other hand the disulfide linkages also interfere in the cleavage of this particular peptides. That is why in the first step itself you have to remove the disulfide bridges. So you have those disulfide linkages actually interfere with the complete sequencing protocol as it does not allow the release of the cleaved amino acid from the peptide chain. So there are two approaches to disrupt the disulfide linkages in a protein sequence.

So you have the two approaches in the one approach you are using DTT and in the other approach you are using the oxidation by the performic acid. In the first approach, the protein is oxidized with a performic acid to produce two cystic acid residues. So in the first, you are going to do oxidation with the performic acid that is going to break the disulfide linkages and that actually is going to give you the cystic acid.

(Refer Slide Time: 8:51)



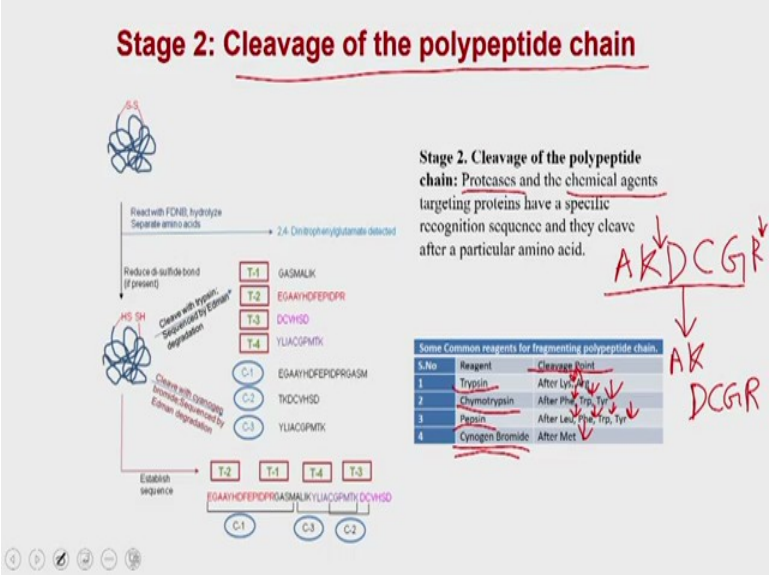
And on the other hand, in the alternate approach what you have is, that protein is reduced by the DTT and once you or beta mercaptoethanol both are actually the reducing agents. So you reduce the disulfide linkages because of that what will happen is, that S-S bond is going to be get converted into S-H, S-H and the linkage between them is going to be broken. But when you convert this by reduction the S H, S H is spontaneously going to come together and form the disulfide linkages.

To avoid that what you are going to do is, you are going to treat with this the iodoaectamide. So you treat this with the iodoacetamide and that actually is going to create or generate the carboxymethyl cysteine. And once it is generating the carboxymethyl cysteine derivatives of the two cysteine present on the protein then these two molecules will not come together and it

is not going to form a disulfide bonds. So these are the two approaches what you can use to break the disulfide linkages.

In approach 1, you are using the performic acid which is going to give you the cystic acid and in the approach 2, you are going to the reduction with the DTT or beta mercaptoethanol and then followed by you are going to have the reaction with the iodoacetamide.

(Refer Slide Time: 10:24)



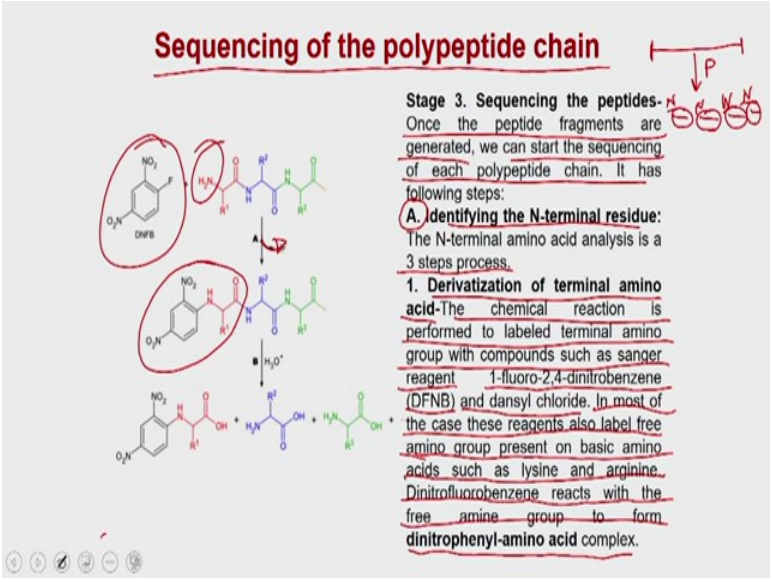
Now let us move on to the next, so the in the stage 2, you are going to do the cleavage of the polypeptide chain. So once your disulfide linkages are gone then the second step you are having to, you have to cleave the linear polypeptide chains. The cleavage can be done simply by the proteases or the chemical agents. These are the some classical proteases which a people are normally using for generating fragments such as Trpysin, Chymotrypsin, Pepsin or the chemical agents such as Cynogen bromide. All these molecules are having a specific cleavage reaction.

For example, if you take the trypsin, it is always going to cleave just after the lysine. So suppose you have a protein like this then if you treated with the trypsin and as I said like trypsin is going to cleave just after the lysine then it is going to cleave here okay. And if there is a arginine then it is going to cleave just after the arginine. So if you treat this particular peptide, what you are going to get, you are going to get two fragments, these are the two different fragments what you are going to get if it is a seven amino acid long polypeptide chain.

Similarly the chymotrypsin, it is going to cleave after the phenylalanine, tryptophan or the tyrosine. Similarly the pepsin, pepsin is a more non-specific proteases but it is still being going to cleave the leucine, phenylalanine, tryptophan or tyrosine whereas if we use the cynogen bromide as a cleaving agent it is going to cleave just after the methionine.

So whatever the enzyme you will use for cleavage reactions it is going to chop off the protease or it is going to chop off that particular peptides or protein sequence into the individual small fragments.

(Refer Slide Time: 12:32)



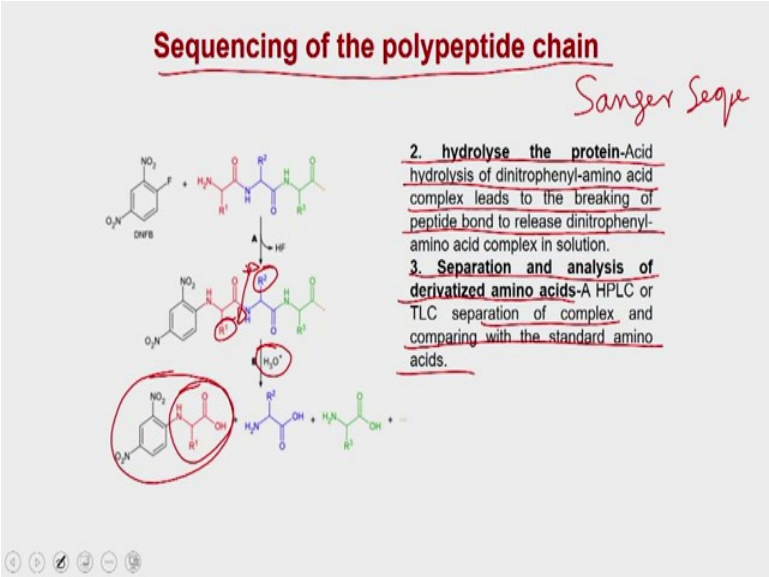
Now, we will move on to the stage 3, so then the stage 3 you are going to do the sequencing of these small polypeptide chains. And so once the peptide fragments are generated we can start the sequencing of each polypeptide chain. It has a following step, so the first step is that you have to identify the N-terminal residue okay. So every polypeptide, so for example, if you started with the one protein sequence like this and then you suppose you cleave this with the protease you are going to get multiple fragments.

Now all these multiple fragments are going to be the sequence individually and the first thing is that you have to identify the first amino acid of all these polypeptides okay. So the identification of the N-terminal residue is a 3 step process. In the step 1, you are going to derivatize the terminal amino acid. The chemical reaction is performed to level the terminal amino group with a compound such as the Sanger reagent or 1 fluoro 2-4 dinitrobenzene DFNB which is actually this.

So what you are going to do is, you take the DFNB and react it with the your polypeptide chain and the N-terminal amino group is going to react and the dansyl chloride in the most of the cases these reagents also label free amino acid present in the basic. So these reagents are going to label the amino group present on the basic amino acids such as lysine and arginine. So the dinitro fluoro benzene react with the free amino groups to form the dinitrophenyl amino acid complex.

So once you do that it there will be a release of hydrogen fluoride and then it is going to form the complex with the terminal amino group, so that is the first step the derivatization of terminal amino acid.

(Refer Slide Time: 14:35)



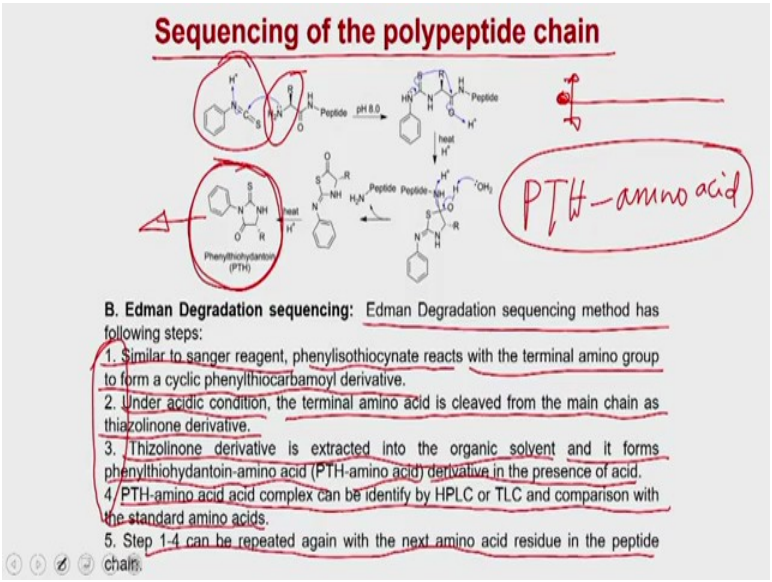
Then, you move on to move on the second step, in the second step you will hydrolyze, so you will hydrolyze this particular peptide complex. Then acid hydrolysis of dinitrophenyl amino acid complex leads to the breaking of the peptide bond to release. So once you do the acid hydrolysis, the peptide bond is going to be broken down between the R 1 and R 2. And as a result what will happen is, the the first amino acid is going to be released from the along with the complex.

So the dinitrophenyl amino acid complex is going to be released and the remaining peptide chain will remain intact. And then the third step is the separation and the analysis of derivatized amino acid. Then what you do is, you take this whole reaction and inject into the HPLC machine or the TLC plates and that is actually going to separate the complex. And the

will the help of the amino acid or the standard amino acid you could be able to identify this particular amino acid.

So what you are going to do is? You are going to run the standard amino acids as a complex and in addition to that you also going to run the reactions. So there, if you compare these two chromatograms in the HPLC or if you run, if you compare the TLC plates of these two reactions it will tell you that which amino acid is this.

(Refer Slide Time: 16:12)



In an alternate method of sequencing of the polypeptide chains, so this method what we discuss so far, this is the method what is being developed by the Sanger and that is why it is called as the Sangers sequencing technique. Whereas in the alternate method what you are going to do? You are going to do the similar steps as we have done in the Sangers sequencing method but these different sets of reagents.

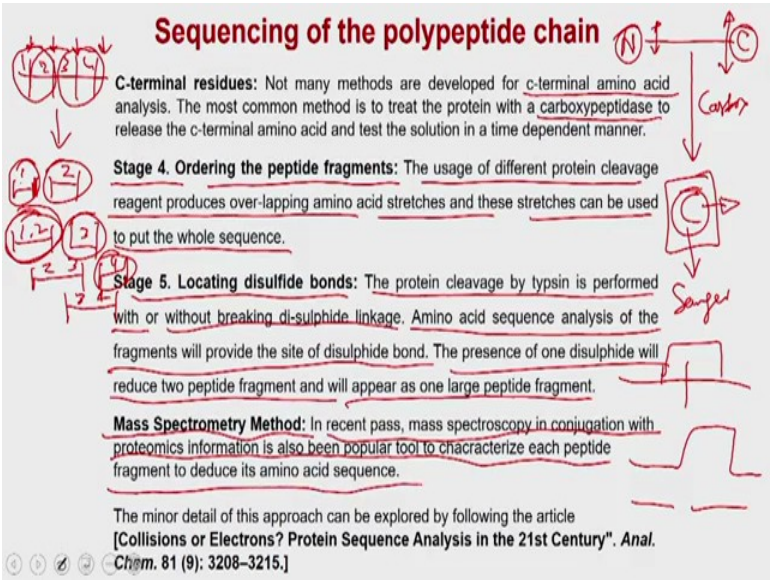
The Edman degradation sequencing method, you have the following steps. Similar to this Sanger reagent the phenylisothiocynate which is actually this reacts with the terminal peptide chain or terminal amino groups to form a cyclic phenylthiocarbamoyl derivatives. So in next step 2 under acidic conditions, the terminal amino acid is cleaved from the main chain as a thiazolinone derivatives and once the thiazolinone derivative is extracted into the organic solvent and it forms the phenylthiohydantoin amino acid PTH amino acid complex which is this one, in the presence of acid the PTH amino acid complex can be identified by HPLC or the TLC in comparison with the standard amino acid.

Step 1 to 4 can be repeated again and again. So you can imagine that you have a peptide chain, the terminal peptide is going to react with the phenyl isothiocynate and it is actually then going to label the terminal amino acid to form a cyclic phenyl thiocarbamoyl derivatives and the peptide chain is going to be cleaved from the main chain. And this particular derivative if you extract it into the organic solvent then it will be actually going to form the PTH amino acid complexes or and this is for terminal (amino group) terminal residues.

Now what you have is, you have the PTH amino acid complexes of this particular complex. So what you can do is, you can take this and run it into the HPLC add or you can load it onto the you can spot it onto the TLC plates. And along with that you can also run the standard PTH amino acid complexes and by comparing these two chromatograms or the TLC plates, you could be able to identify the terminal amino acid.

And then what you can do is, you can repeat the one to four with the remaining peptide chain and that is all if you continuing this particular cycle it will tell you the full sequence of that particular peptide chain.

(Refer Slide Time: 19:04)



Apart from that so, so far what we have discussed? We have discussed about the sequencing of the N-terminal amino acids. You can also do C terminal re-sequencing because the protein has the two groups is the amino group as well as the carboxyl group. So there is no method available for sequencing the carboxyl residues or the C terminal residues. Except that you can actually get the C terminal amino acid residues, simply if you treat the protein by the carboxypeptidase.

So if you know that the carboxypeptidase is a specific protease which actually choose of the protein from the carboxyl side which means it is chewing up from the reverse side. So in those cases what is going to do is, so you have the protein which is actually having the N-terminals on one side and the C terminals on the other side. So under ideal conditions when we are treating this particular protein, we are treating it with the normal protease and normal proteases are cleaning the protein from this side.

But if you treating this with the carboxypeptidase with the carboxypeptidase is going to start cleaving it from the C terminal side. So because of that it is going to release the C terminal amino acid and then once the this particular amino acid is going to be released then you have the option either you use this Sangers method or you can use the other method to identify this particular amino acid. And that is the only way you can be able to get the identity of the C terminal amino acids.

Now in the stage 4, you have to do the ordering of the peptide fragments. So, so far what we have discuss? We have discussed that you take the protease or you take the peptide chain, you first remove the disulfide linkages, now it is going to be a linear peptide chain. Now what you do is, you chop off these particular peptide chains with the different proteases. Now you are going to have the small-small-small peptide fragments.

Now this peptide fragments you have sequenced either by the Sanger method or the other method and then now you have the sequences. But the sequences are not going to give you the original sequence from where you the or the proteins sequence. Now the second step is you are going to put this peptide fragments into a particular order so that you will be able to get the original sequence.

The usage of different proteins cleavage reagents produces overlapping amino acid stretches and these stretches can be used to put whole sequence, let us understand this. Now suppose you take you have taken a protein okay you cleaved up into this like, so it is going to have the 4 fragments, fragment number 1, fragment number 2, fragment number 3, fragment number 4.

Now, let us see how you are going to get the fragment? This is the fragment number 1, this is the fragment number 2, this is the fragment number 3 and then you are going to have the fragment number 4 but when? So you can imagine that a protease is cleaving it at a four sides but what you are going to have is also this which is actually the fragment number 1, 2. You are also going to have a fragment which is this, which is 2 and 4. You also going to have a fragment which is 4 and 4 and so on.

So you can see that you are not going to get the 4 fragments but you are going to get 1st fragment, 2nd fragment, 3rd fragment, 4th fragment. You are also going to get 1 and 2, you are going to get 2 and 4, you are going to get 3 and 4 and in some cases you might get 2, 3 and 4

as well but 2, 3, 4 is going to be slightly bigger so that may not be very helpful because you are not going to use those particular larger fragments sequence.

But you can see that if I have 1 and 2 separately and I also have 1 and 2 together I will not going to take much time to know that whatever the amino acid is present on the end of the first peptide fragment and the beginning of the second fragment is going to be the same amino acid. And in those cases it will be very easy for me to put these two fragments together.

Similarly, if I have 3 and 4 separately and I have a overlapping stretch which is covering the 3 and 4 then it will not going to take much time to understand that these this is the junction point and that is how you are going to put all the 1, 2, 3, 4 and that is how it is going to be placed one after other and that is how to you are going to have the full sequence ready.

The stage 5, you also since we have broken down the disulfide linkages okay, so you also since protein sequencing means you are going to generate to the original protein sequence. So you also should have to know that where the disulfide bonds are present. The protein cleavage by the trypsin is performed with the width or without breaking the disulfide linkages. So to locating the disulfide bonds, what you are going to do is, you cleave the protein which trypsin in the without or with breaking the disulfide linkages.

Now when you do the protein sequence analysis the fragment will provide the site of the fragments. The presence of one disulfide link will reduce two peptide fragments and will appear as one large fragment. You can understand this, if you have a peptide sequence and you are having a disulfide linkages and suppose I am not going to break this disulfide linkages then what will happen is?

If I am cutting this from here, I am going to have this large fragment whereas if I do not have this, I am going to have the two fragments. So if I do not do the removal of disulfide linkages. And suppose I cut the trypsin then I am going to get a large fragment and that kind of analysis is going to tell me that the position of the disulfide linkages between that particular protein sequence.

Lastly because we have to deduce the sequences there is a method which is people are where oftenly using is mass spectrometry. So in the recent past, mass spectroscopy in conjugation with the proteomics information is also being popular tool to characterize the each peptide fragment to deduce it amino acid sequence. And you can if you would are more interested to

know this technique you can be able to read this particular articles such article such article and that will actually tell you the potential of mass spectrometry in conjugation with the proteomics information.

And there are a lot of exciting course works or course material is also available about the proteomics in NPTEL as well as the other online web sources and that also you can able to consult if you are interested to know about the potential of mass spectrometry in deducing the peptide sequence. So with this we would like to conclude our lecture here and in our subsequent lecture we are going to discuss about how to characterize the functional groups with the help of UV visible spectroscopy as well as the IR spectroscopy. Thank you.