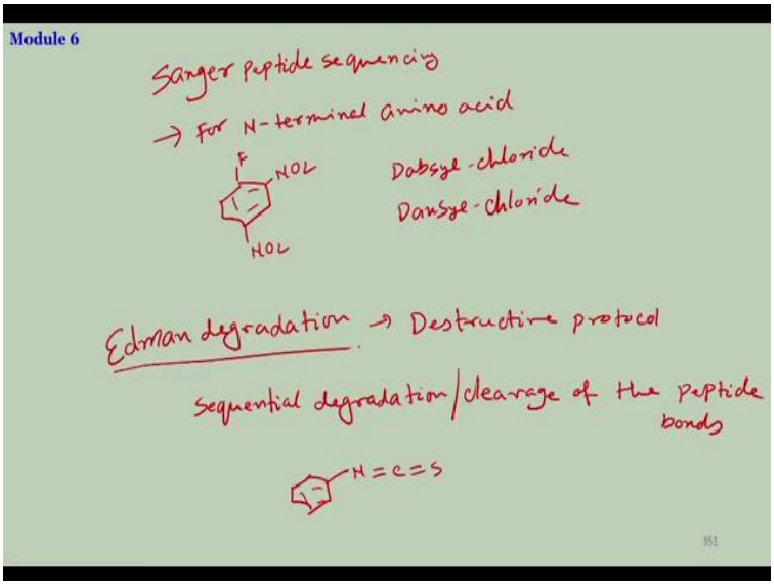
Essentials of Biomolecules: Nucleic Acids, Peptides and Carbohydrates

Prof. Dr. Lal Mohan Kundu
Department of Chemistry
Indian Institute of Technology-Guwahati

Lecture-25
Mass Spectroscopy and Other Sequencing Methods For Large Proteins

Hello everybody and welcome back to biomolecules. So, in the previous lecture we have started with the sequencing of proteins or peptides and we have seen 2 sequencing methods.

(Refer Slide Time: 00:47)



One is the Sanger peptide sequencing method and this method is usually used to determine the N-terminal amino acid of the peptide; or the protein sample for N-terminal amino acid. So, mostly what is the amino acid present in the N-terminal of the protein that to determine that Sanger method is used. Of course, we have seen that this is our chemical method and you usually use this compound aromatic compound with a fluoride here and 2 nitro group.

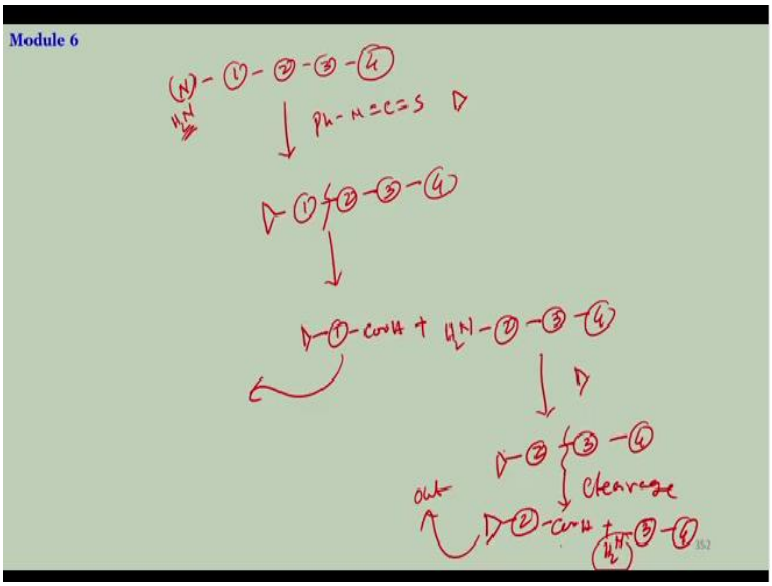
This is the reagent; this is the region that is used for Sanger's method initially. Nowadays, this can be replaced with the fluroforce or fluorescence containing molecules such as I have shown the structure in the previous class, dabsyl-chloride or dansyl-chloride. So, that was the one of the first method. The Sanger's method was one of the first methods to determine the sequence of proteins.

But ideally we use it for to determine only the N terminal amino acid. Next we have seen the in order to determine the full sequence of the peptide or the full sequence of the protein, we have used Edman degradation method and as we have seen, that this is also our destructive method or destructive protocol, because you do not get back your sample and this is the idea of Edman's degradation method is to cleave the peptide bonds in a sequential manner, one after the other.

So, sequential degradation of the peptide bonds or cleavage of the peptide bonds that are present in your protein sample and the reagent that we have used is also a chemical reagent that was isocyanide phenyl isothiocyanate. So, nowadays Edman's degradation method is used, it is pre programmed and there is just like PCR machine. There is also the peptide sequencer like you have DNA synthesizer or DNA sequencer.

Similarly, for sequence a peptide nowadays, the machines are there which are pre programmed and so it basically cleaves the peptide wants one after the other and cleaved part is eluted out every time and is getting determined by some characterization techniques usually with the mass spectrometry.

(Refer Slide Time: 05:09)



So, very briefly we have seen how the Edman's degradation method worked. So, if you have a peptide and the N-terminal end 1 2 3 4. So, these are the peptide bonds, here it is a peptide bond,

this is another peptide bond, this is another peptide bond and so on. So, first thing is that the free amine at N-terminal end that reacts with the phenol isothiocyanate. And we have given it a symbol as this. So, you have basically form this kind of complex N you can write N or I am not writing the N directly this.

So, you get a marker of this reacting with the amine group. So, once this is done then comes the degradation, it breaks this peptide bond once this is formed by sequential reaction. So, this is basically a carboxyl terminal plus to cleave this peptide bond and that releases the free amine at the second amino acid. So, this is the first amino acid eluted out and then the entire peptide sequences left over with an exposure of another amine.

So this amine further reacts with your phenyl isothiocyanate. So you get the next marker this is eluted out and it can be determined immediately. So, that will tell you the identity of the amino acid 1. So, you can know what is the first amino acid. Similarly, once this is done, then comes the cleavage of this one at the C terminal end, the cleavage of the bond which is present of the peptide of bond which is present after the amino acid which is that. So, after 2 between 2 and 3, this bond would be cleaved.

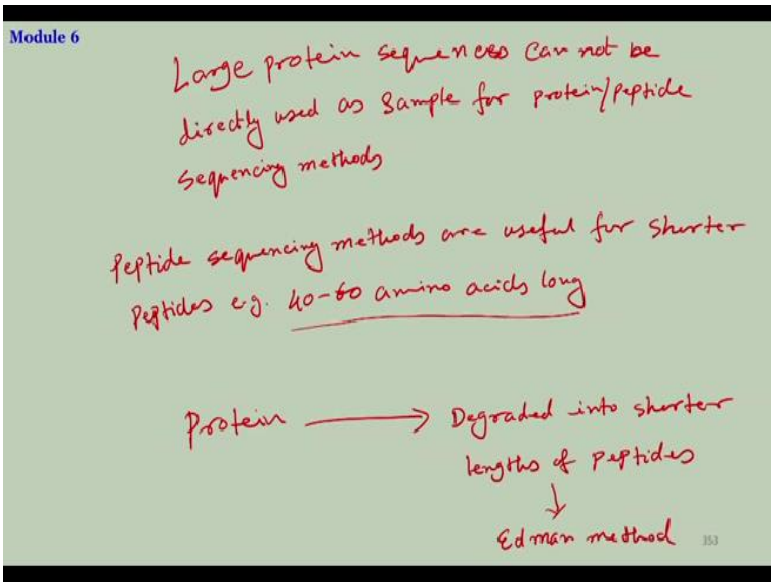
And it will leave you with 2 with the free carboxyl that would be eluted out. So, this is out plus the rest of the peptide or the protein starting with 3 4 and the other with again exposure of another free amine and this will go on and one. So, by determining or by characterizing each of these fragments that are coming out every time; you can know the entire sequence of the protein that was the Edman's degradation method.

So, ideally these 3 are the methods that are used for determining the sequence of a given protein sample. One is the Sanger's method mostly used only for the N-terminal nowadays. Second is and most widely used is the Edman's degradation method, which is now it is done automatically in an automated machine. The third one is the instrumental technique known as mass spectrometry.

So, using high end mass spectrometric analysis nowadays you can find out about the sequencing of the peptide. So, later on when we will be on the module of learning the modern techniques, I might talk about this mass spectrometry technique. So, these are the 3 methods that are 3 ways that are adopted to determine the protein sequencing. Now, all these techniques be it the Sanger method, be it the Edman's degradation method or be it the pure mass spectrometry.

All these methods are usually work better or are usually at applicable only for short peptides. If you use a very long protein sequence which has 500 amino acids for example, very long protein sequences that cannot be used as a sample for all these methods. So, that is the limitation of the protein sequencing methods.

(Refer Slide Time: 10:18)



So, large protein sequences cannot be directly used as sample for protein or peptide sequencing methods. So, that is the limitation of these techniques, large proteins or original proteins that you have isolated from some source cannot be used directly to determine it is amino acid sequence, what these methods are usually for Sanger's method, Edman's degradation method or the mass spectrometers method.

They usually use shorter peptides sequences. So, sequencing methods, peptide sequencing methods right now, peptide sequencing methods are useful for shorter peptides such as maybe 40 to 60 70. Let us 60 amino acids long, so 300, 400, 500 amino acids long protein sequences

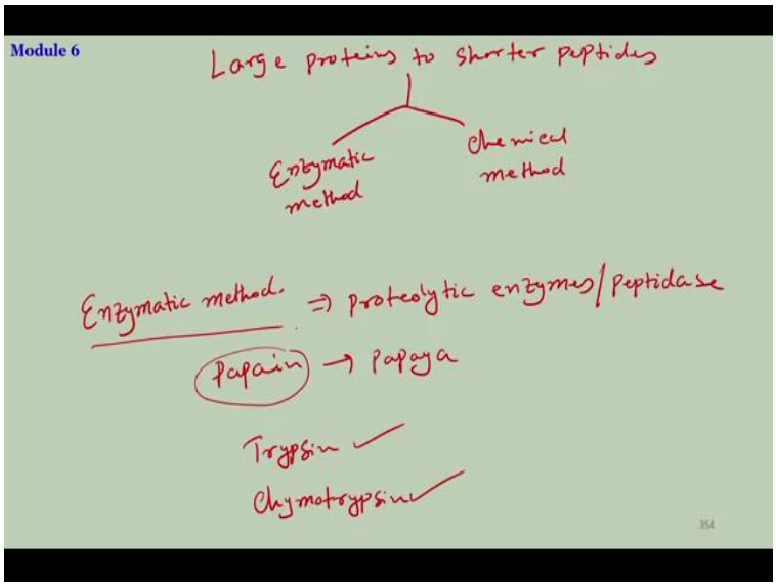
cannot be determined short sequences of peptides are the best as a sample in these methods that you can use these short sequences to find out they are exact amino acid chain length.

So, what is the way out. How can you then, if you cannot use the full protein as directly that are using these methods, if you are not able to determine the sequence of the full protein, then what is the way out to understand the sequence of this full protein. The idea is what you have to do now is if you have a protein, full length protein that you have to break down into short pieces of the peptides, which will have in this length maybe 40 to 60 amino acids length.

So, protein first degraded into shorter lengths of peptides. So you take your protein sample treated in some way, so that that produces a number of short peptides. Each of them maybe have the or each of them would have the length below 40 to 60 amino acids and then all those fragments, all those short lengths of peptides that you have made out of this protein can be used separately to find out the sequence.

After this you can use the advanced degradation method to determine the sequence of these shorter fragments.

(Refer Slide Time: 14:47)



So, in this case you have to break down first the large proteins to shorter peptides, that is very important. Now question is how to do that. So I will show you 2 methods. There are actually 2

ways of doing that. One is the enzyme meeting method. Second is the chemical method. So first let us take so there is enzymatic method, and the second is chemical way to break down a protein into pieces.

Enzymatic method, so you have to use an enzyme which can cleave a protein into pieces. Not only that, it has to be specific the enzyme has to be such that it will cleave the protein at a specific amino acids, then only you can know where the cleavage is occurring it cannot be arbitrary cleavage, it has to be pre determined cleavage. So, that you know where my bonds have been broken and then you can take the actions accordingly.

So, such kinds of enzymes which cleaves the proteins are known as proteolytic enzymes or they are also known as peptidase. Peptidase means that cleaves a peptide bond. There are many examples of such peptides or such proteolytic enzymes that are present in our body, that are present in other animals that are present in plants. And they have very important application actually, especially the proteolytic enzymes, or the peptides are very important in our metabolic system.

When we eat some food, protein rich food for example, then because the protein is so long, it is difficult to get digested. So, these peptidases or these proteolytic enzymes would cleave your food or cleave the proteins present in the food into short sequences into pieces, and then your digestion or the metabolism becomes easier. So these kinds of enzymes are present everywhere. I will give you some examples.

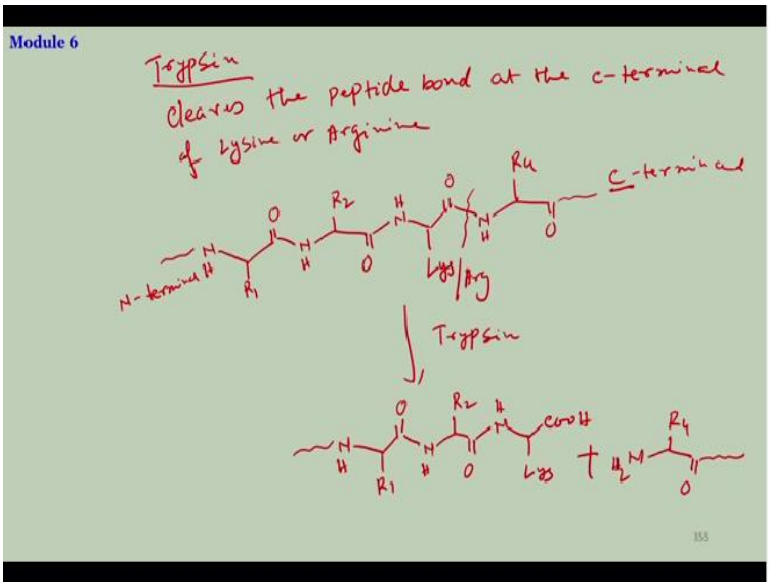
So papain, pipin or papain is one such enzyme that is present in papaya. And this enzyme cleaves the proteins into short sequences. So, I do not know it says school chemistry. If you, have noticed that when you cook food, some especially the protein rich food, for example meat. If you add papaya to it, then it takes less time to boil or it takes less time to cook the food. Your meat gets soften very easily if you add papaya there.

So the reason is this because you get the enzyme papain, which is a proteolytic enzyme and that cleaves the protein in the meat into the smaller fragments. And then your cooking becomes

easier. So other kinds of examples of the proteolytic enzymes are, which we will be using now is trypsin or chymotrypsin. So, this is one proteolytic enzyme, this is another proteolytic enzymes. Trypsin and chymotrypsin are usually that you find in the market are usually isolated from pigs.

So, how are they specific to sequences where do the cleave in the protein structure.

(Refer Slide Time: 19:44)



If you use trypsin it actually cleaves the peptide bond because it is the peptides, so it cleaves the peptide bond at the C terminal position, C terminal of the amino acid either lysine or arginine. So if you have I am starting somewhere in the middle, this is the N-terminal NH. This is your R 1 R 2, if you have a lysine here or an arginine here income CO NH R 1 R 2, so, this is R 3.

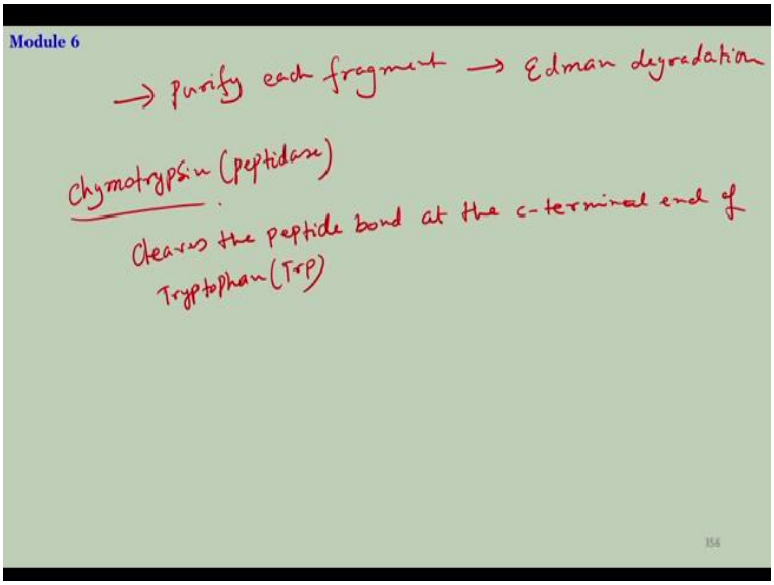
So, I will write R 4 here and so on it goes on. So, this is the C terminal. This is the N-terminal. If you treat these with trypsin then it cleaves at the C terminal of lysine or arginine. That means, here this peptide bond, this is the peptide bond after lysine**,** and this is the carbonyl group. So, this is the C terminal part of the lysine. So, after lysine the following peptide bond would be selectively cleaved by the enzyme trypsin.

It can be either lysine or it can be arginine or both. So, wherever there is lysine, wherever there is arginine the following peptide bond would be chopped off, by trypsin. So, what you will get, you will find amine R 1 CO NH R 2 CO NH lysine this is cleaved. So, you have a free carboxylic

acid here plus your protein is now fragmented into 2 pieces, at least one is this, this here. Other one is this amine R 4 CO and the other part, the rest of it.

So this would be one fragment, this would be another fragment. So that is how trypsin would cleave at a specific amino acid sequence in the protein. So, if you take a whole protein and if you treat it with trypsin, then you can know where your trypsin is cleaving, which amino acids or which peptide bonds this is cleaving. And the fragments that you get, after the reaction that you can isolate each one of them you can separate out. And each of them, you can put it or you can use further for Edman's degradation method.

(Refer Slide Time: 23:57)

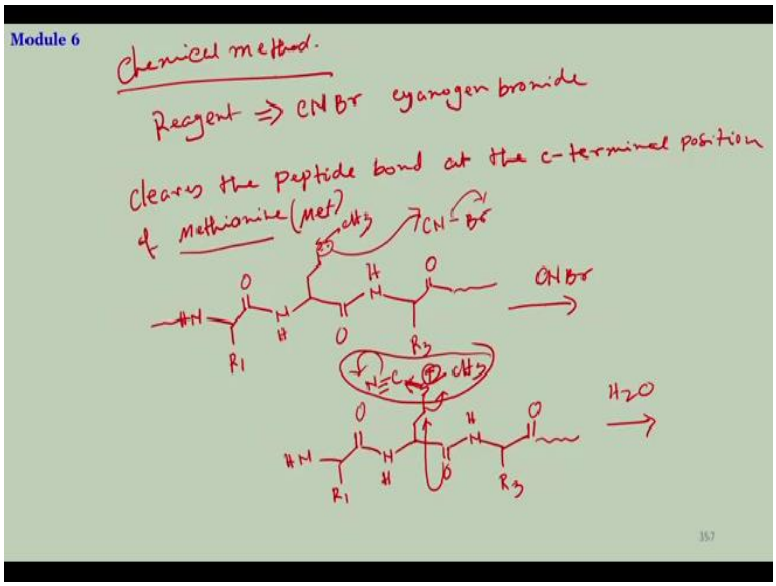


So, once you have the fragments, next step is you purify each fragment and then use it for Edman degradation. So, in this way you will know the sequence, exact sequence of each fragment and then you can basically tie them up to get the full sequence of the protein. So, that is the mechanism of action of trypsin. Now, if you have chymotrypsin, chymotrypsin is another proteolytic enzyme or another peptide is and it cleaves the peptide bond at the C terminal end of tryptophan Trp.

So, if you have a protein sequence, if you treat it with chymotrypsin wherever there is tryptophan, it will cleave the following peptide bonds and you will get the shorter fragments peptides, again you can purify each of these fragment and then use them for Edman's

degradation method. So, that is how you know whenever you will be using an enzyme you have to select an enzyme for which you know where it cleaves. That is very important actually predetermined sequence have to be cleaved.

(Refer Slide Time: 26:11)



Now comes a chemical method. In chemical method we use a reagent of course, and this reagent is CN Br cyanogen bromide. This is the region that is used to cleave a peptide bond at a pre determined amino acid. So this against cleaves the peptide bond again at the C terminal C terminal end or C terminal position of a specific amino acid. In this case is methionine met? So same here cleaves all the peptide bonds wherever there is a methionine.

How does it work. If you have so this is N-terminal there can be other peptide, other sequence here this is R 1 CO NH. Let us say here you have a methionine CH 2 CH 2 methionine is S and CH 3, this is methionine. This is the C terminal end. So this is the C terminal peptide sequence R 1 R 2, this can be R 3 and it goes like this. So, it should be CO and then the rest of it. Let us say that this is the sequence of the protein.

And if you treat it with CNBr, then what happens, bromide is a good living group. It is the electronegative group. So it will try to drag the electron density and move away, sulfur has one pair of electrons and sulfur has a large size. So it has the one pair of electrons is highly donatable

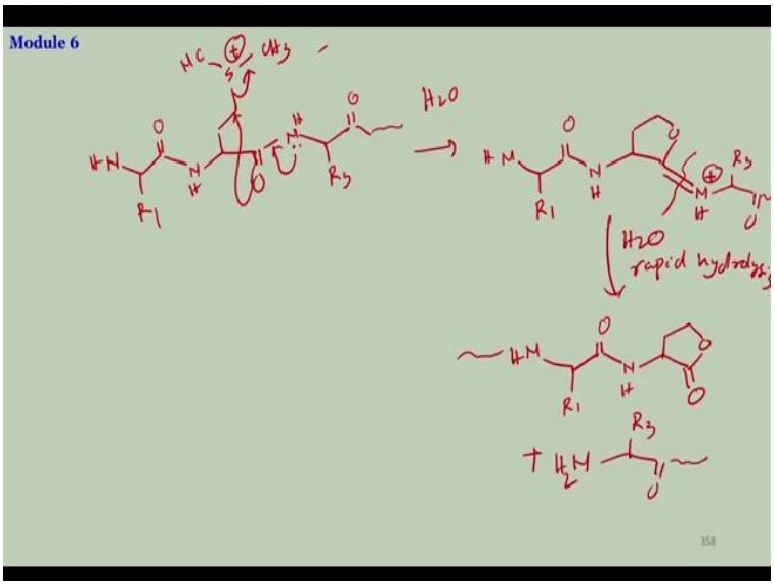
because it is to some extent electro positive. So the one pair of electrons will attack the carbon, attack the carbon of the cyanide eliminating the bromine or the bromide.

So, you will have CO NH here CH 2 CH 2 S CH 3 and cyanide and this would be now plus because it has donated the electron So, sulfur plus, in each R 3 run this, now this has formed a polypeptide here and it has this, but if you look at cyanide that is why cyanide has been used cyanide or CN minus has dual character. So, in this case in cyanide, cyanide also is pretty stable as cyanide minus. So, it has the tendency also to drag the electron density towards itself because nitrogen is electronegative.

So, the electron density would be dragged towards the nitrogen that forces the electron density of this bond to be dragged towards the cyanide, that makes the sulfur even more electro positive, even more electron deficient, it has already a plus charge and because of the cyanide attraction, electron withdrawing effect, this the positive charge would be or positive charges density will be higher here. So, this whole thing becomes a very good leaving group.

Since this is highly electron deficient, it will try to take away the bonded electron and move away. So it is very easy to cleave this bond. So, if you hydrolyze it in presence of water, the carbonyl here, that is where the chemistry of the peptide bond is coming. This although it is not very reactive, peptide bond is not very reactive, but because you have a pretty good living group, this reaction happens, it will cleave, it will attack this carbon eliminating the whole thing, the sulfur.

(Refer Slide Time: 31:54)



And you get a cyclic compound which will look like this is CH 3 CN. This is positive charge that makes it a good leaving group CO NH yeah this is R 1 R 2, this is R 3. So, that will attack this carbon and they should leave. So, what you have is of course, this is with treatment with water R 1 CO NH, these O here. So this bonded electron go and then this would be the resonance, this is gone, this is coming here, this is attacking.

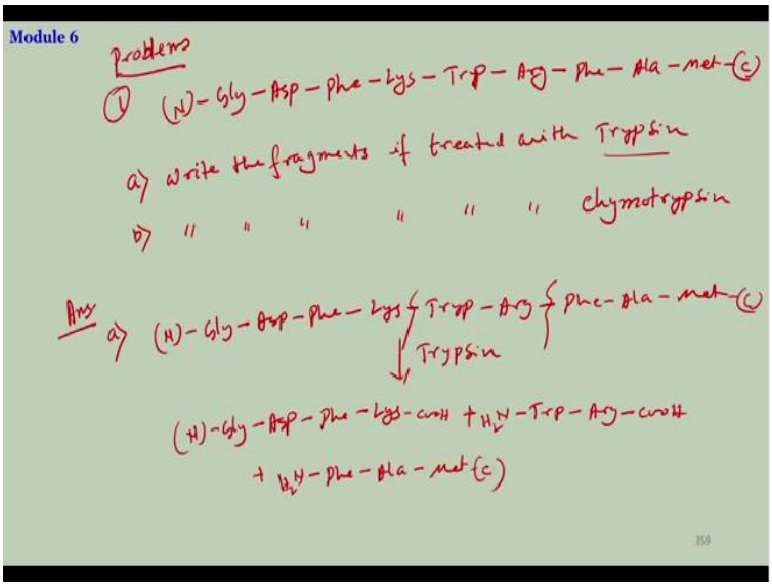
This is the carbon and this would be your NH. This is your R 3 CO and the rest of it. Here you have formed the double bond and would be NH plus. So, this is the cyclist compound that will be forming through the oxygen and to stabilize this, there will be this resonating structure. Now, this becomes a double bond with the positive charge on NA amine or with the positive charge on nitrogen which is also not very stable.

So, it will undergo a rapid hydrolysis and this bond would be cleaved. If you hydrolyze a carbon nitrogen bond what you will have water only attack here will move this. So, you will basically get a carbonyl compound here CO NH, this here will have a carbonyl plus you will have the free amine H 2 R 3. So, you have breaking or the cleavage of the peptide bond in the position where there was a methionine or the following peptide bond after the methionine.

So, this is the chemistry here. And now you can again separate the individual fragments and do the Edman's degradation. So, the number of fragments will tell you. So if you have 5 fragments,

for example, in the whole protein sequence, then you know, you have at least 5 methionine, present in this at the terminal of the, wherever there was the cleavage, okay. So that is how you can actually break a given protein into it shorter peptide sequences.

(Refer Slide Time: 35:43)



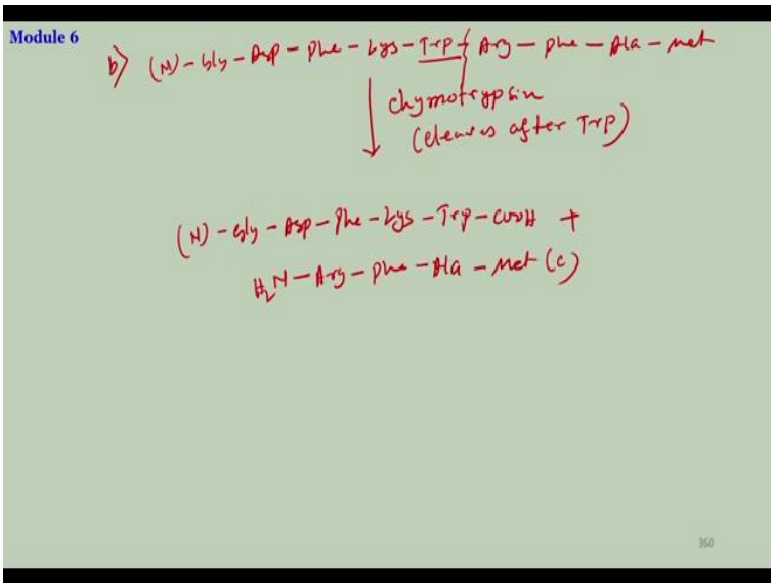
So, now I will give you 2 problems. Problem 1is let us consider a protein sequence N-terminal glysine, aspartic acid, phenylalanine, lysine, tryptophan, arginine, phenylalanine, alanine, methionine and this is your C terminal. Let us say this is a given protein sequence. Question a, is write the fragments if treated with trypsin. Question b; write fragments if treated chymotrypsin. So, can you write down the fragment of the proteins or what are the fragments that we will obtain if you treat this given sequence which is trypsin.

What will happen if you treat it with trypsin, it will cleave the peptide bond at the C terminal position of wherever there is a lysine or wherever there is an arginine. So glycine, aspartic acid phenylalanine, lysine, tryptophan, arginine, phenylalanine, alanine, methionine. So, if we treat this with trypsin then it will cleave here. There is a lysine here. So after the lysine you have another arginine here to cleave keep also here.

So, the fragments that will see is lysine, aspartic acid, phenylalanine, lysine, free acid plus N H 2 tryptophan, argentine free acid plus it will expose the amine at the phenylalanine methionine. So,

this is the C terminal end. So, this is the C terminal. So, these are the 3 fragments that you can expect if you treat in this specific protein sequence with trypsin.

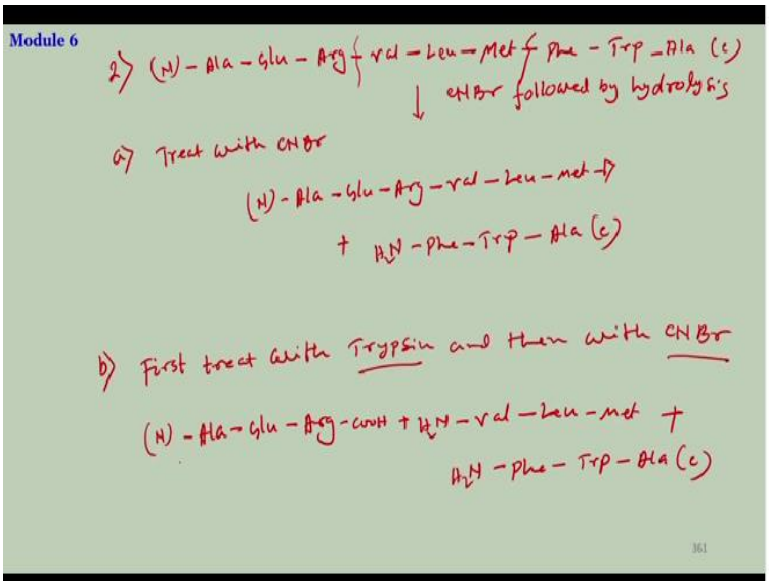
(Refer Slide Time: 39:34)



Now b is glycine, aspartic acid, phenyl alanine, lysine, tryptophan, arginine, phenylalanine, alanine, methionine and you are treating this with chymotrypsin, and we know that chymotrypsin cleaves the tryptophan, the C terminal end of the tryptophan, so it will basically cleave here, cleaves after tryptophan and there is only I think 1 tryptophan. So you will have a single cleavage here.

And your fragments would be glysine, aspartic acid, phenylalanine, lysine, tryptophan. This should be chopped off it will expose free carboxylic acid plus free amine of the next arginine and the rest of it, alanine, methionine. So these are the 2 fragments that you can expect from here.

(Refer Slide Time: 41:23)



So, I will give you another question. Let us take a different sequence alanine, glutamic acid, arginine, valine, leucine, methionine, phenylalanine, Trp tryptophan, alanine. So, this is the given sequence. Question a is if you treat this treat with CNBr then what are the fragments CNBr followed by of course, this is very important followed by hydrolysis, treatment of this sequence with CNBr followed by hydrolysis, what are the fragments that you will find.

CNBr will break after methionine right. So here this should be broken and you will end off with alanine, glutamic acid, arginine valine, Leucine, methionine, of course, is no free carboxylic acid, it will finish here with the cyclization. So I am writing this as cyclic. You do not have to write anything actually. So it is methionine plus it will expose the amine of the next amino acid which is phenylalanine, tryptophan alanine.

And that is your C terminal. So these are the 2 fragments that you will obtain. Second question is first treat with trypsin and then with CNBr, so it is a combination of treatment with enzyme as well as the chemical reagent, you are treating both with trypsin and with CNBr, then what are the fragments that you can expect. This of course with the expectation that they will not interfere into each other business.

So, if you treat it with trypsin what you will have, you will have a cleavage of arginine, you will have only arginine here. So, let us write that alanine, glutamine, arginine COOH plus NH 2

Valine, Leucine, then you have methionine, you have used CNBr. So, it will cleave after methionine to stop there plus amine, phenylalanine, tryptophan, alanine. This would be your products. These are the 3 fragments that you will obtain. I guess I was correct trypsin, arginine.

There is only arginine no lysine, so only 1 cleave is here for this and another for methionine 2% 2 cleavage that will give you 3 components. So 3 fragments that you can expect by a combinatorial treatment of trypsin and cyanobromide or cyanogen bromide, this plus, this plus this. So, likewise, you can actually find out you can play around with the different sequences. So, if you take a large protein sequence and treat it with the combination of trypsin and cyanogen bromide or even a combination of trypsin plus chymotrypsin plus cyanobromide, then what are the fragments.

You can expect huge number of fragments that will be present in your final products and you can separate each one of them, you can expect that your protein is now degraded enough into the short sequences, which you can use for Edman's degradation method to know its sequences. So, this is roughly about the sequencing of the proteins, if you have an unknown protein sequence, how would you find out it is the amino acid chain.

What are the amino acids present one after the other. We have talked about 3 methods for the sequencing itself. And then today, mostly we have talked about the methods through which you can actually cleave or you can actually make fragments from a large protein into its smaller pieces. That is for today. And then we will take the next lecture and we will talk about peptide synthesis, solid phase peptide synthesis that we will be using in the laboratory method to construct a peptide sequence. Thank you.