Organic Chemistry In Biology And Drug Development

Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture - 06
Selective Peptide Bond Cleavage: Enzymatic and Non- enzymatic Methods

Welcome back to this interdisciplinary course on Organic Chemistry in Biology and Drug Design. Now, last time when we ended, we were discussing the sequence determination of proteins. Sequence determination means the determination of the amino acids that are arranged one after another and the process to do that. We will have a recapitulation of that; first it was discovered by Sanger that how to know the N-terminus amino acid by using a reagent which is now popularly called Sanger's reagent (2,4-dinitrofluorobenzene).

And, then a very similar reagent that was also introduced just to enhance the sensitivity of the determination was the dansyl chloride. And then for the other end; (the C-terminus), carboxy peptidase was used which is an enzymatic method to know the C-terminus amino acids. While doing so, you can actually tell what are the last few amino acids because, they will come sequentially one after another. So, if one determines which amino acid is released at different time points; one is able to determine up to certain extent (the first 3-4 amino acids).

However, the determination of this complete sequence of amino acids in a protein or a peptide is called the Edman degradation where, the reagent is phenyl isothiocyanate. And, phenyl isothiocyante has a beautiful chemistry evolved around it and that is that after the attack by the N-terminus amino acid to phenyl isothiocyanate which is an electrophile. That ultimately induces a cleavage of the first peptide bond, resulting in a peptide which is one amino acid less than the starting one. And, you can isolate that peptide and also you can identify which amino acid has reacted with phenyl isothiocyanate and you can repeat the process.

However, there are some limitations not due to the fact that Edman degradation reactions are not very quantitative, infact the reactions are very high yielding with more than 90 percent yield. However, you have to remember that even if there are steps which are 90 percent yielding reactions, after the second step the reaction becomes 81 percent overall

yield. And so, it drastically reduces because it goes down for every step; if we consider that the yield is 90%then after 10 steps it will be ~35%. Similarly, the overall yield after 10 steps will be close to 9% if the yield of individual step is taken as 80%. The percentage is obtained if you just multiply ten 9's and that will be give you the yield and that will divided by remember divided by it is a percentage.

So, ultimately the percentage will be very low, just an example, suppose you do four steps with each step having 90% yield. So, the amount will be 9 x 9 x 9 x 9 so, that will make it 81, then 73 and 66. So finally, it will be about 63 % yield (65.6% to be precise) after 4 steps. So, the yield drops down and currently the limitation is that you can go possibly up to about 30 amino acids sequence per cycle per in one go, in one shot of Edman degradation. So, if you have a protein, we ended up there; but the problem lies if you have the protein which is a very large and proteins are usually very large.

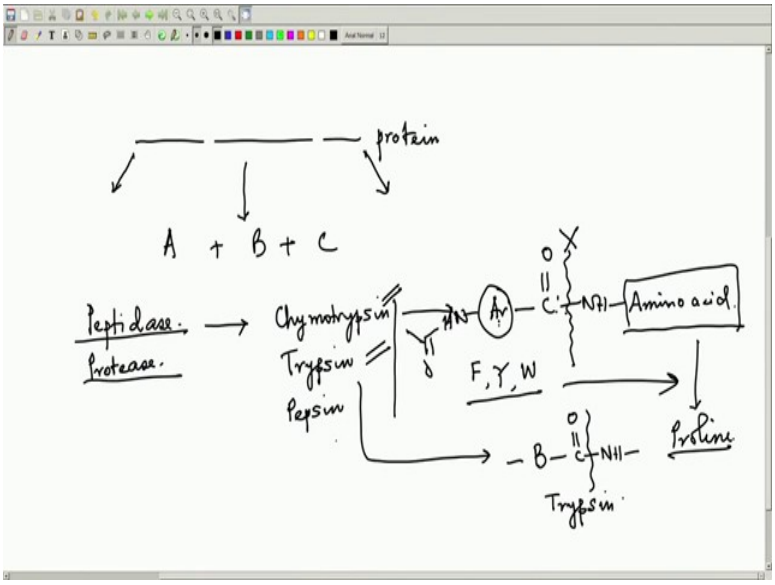
Now, the question is what is the difference between protein and peptide? Sometimes we call peptide, sometimes we are calling proteins; usually it is basically the molecular weight. There are no hard and fast rules, but if something contains less than 50 amino acids then that will be called as a peptide. If it has more than 50 amino acids, usually then that is classed as protein. But, the general term which will cover both is 'polypeptide'; polypeptides contain peptides less than 50 amino acids as well as proteins more than 50 amino acids. But, there is no a sharp demarcation between the both.

Now, again coming back to the question that if you have an amino acid which has got a large number of amino acids, then what are you going to do? Because, in one sort you can only determine up to about 30 amino acids and if it is more than 30, if it has more than say 100 amino acids then you have to cut down the protein into smaller pieces. I told you that, there are enzymes which can do that; some enzymes are very specific. You cannot do it by treating the protein with acid or base which hydrolyzes amide bonds; however, they are not specific. So, every amide bond will be hydrolyzed.

So, you have to have some specificity in your breakage of the protein into smaller pieces. It is basically what is called tailor made breakage; tailor made means the tailor cuts the piece of a cloth into such smaller pieces which ultimately they tie up together and to make the dress. This is a very similar process; you cut the protein into smaller pieces, determine the sequence of those smaller pieces, sequence of amino acids of those smaller

pieces and, again tie up, join them. in this joining means not physical joining, you again just write that what are the sequences and then finally, combine the sequences to find out the correct sequence.

(Refer Slide Time: 06:58)



Now, there are again problems with that because if you suppose you have a protein like this. I told about that if you cut it into say three pieces A B C and you determine the amino acid sequence of all these three. Then the next question comes, that what is the arrangement?

After you cut the protein into three pieces, then you really know that which piece is coming first followed by which one and what is the third one. So, what you need to do, if you have three pieces then you can add up that what is the N terminus (by Sander's method), what is the C terminus (by the carboxypeptidase method). And, if you do that then you know which one is actually in between A and C.

Now, let us talk about some of these enzymes which can selectively cut the protein molecule into three pieces like this. So, they are called artificial scissors. Let us now talk about what are these artificial scissors or the enzymes, that we are talking about which has very special selectivity in cleaving the peptide bonds at certain positions.

Enzymes are are also proteins. So, basically proteins which have catalytic power they are called enzymes, now here we are talking about enzymes. So, these enzymes break the

protein (the peptide bond) so, they are collectively called peptidase because, they are breaking a peptide bond.

Now, what are some of these enzymes? They are all naturally made and one particular one is called chymotrypsin, another one is called trypsin, then you have pepsin. So, these are some of the proteases or peptidase; you can also call protease because, they are breaking the peptide bond, they are breaking the protein into parts.

So, they are also called proteases, sometimes called proteases sometimes peptidase. And, these are the enzymes which are mostly used trypsin and chymotrypsin as they are the ones which are extremely specific. Specific means they only recognize certain amino acids and only when they recognize those amino acids, they will cleave the peptide bond associated with those amino acids.

Any enzyme which has got a broader specificity, broader specificity means identifying many amino acids out of this 20, then you have the possibility of cleaving the peptide or the protein into many fragments which we also do not want. Because, we want fragments which are sizable, which has got optimum size as well as the number should not be very large.

Unnecessarily if the protein is cut into 10 pieces of 10 amino acids, then that does not serve the purpose because, one is Edman cycle ultimately will lead to the determination of our 30 amino acids; so no point in selecting an enzyme which has got a very broad specificity.

So, that is why people have selected this chymotrypsin and trypsin because, they are extremely selective, their selectivity is pretty narrow. Pretty narrow means they only recognize very specific (certain minimum number) amino acids. Chymotrypsin, what it does is that if you have a peptide bond CONH and if it happens to be have an aromatic amino acid on this side and another amino acid on this side.

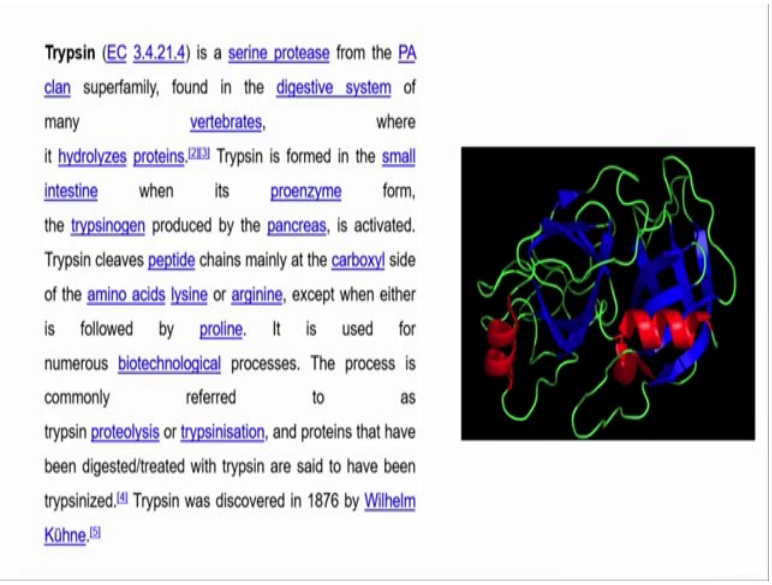
Then chymotrypsin cleaves the peptide bond between this aromatic amino acid and the other amino acid from the C terminus end. Chymotrypsin will cleave only when there is phenylalanine that is one aromatic amino acid, remember those aromatic amino acids, then tyrosine that is Y and W that is tryptophan. So, these are the three aromatic amino acids (F, Y and W). So, if they are present then the peptide bond will be cleaved, but

remember that this aromatic amino acid is attached by 2 peptide bonds. One is on the N-terminus side and another is in the direction of the C terminus.

So, it actually it is breaking the peptide bond involving the aromatic amino acid at the carboxylic end. So, what I mean is basically this is also again connected to another peptide bond on both left and right side. Here you have to say for aromatic amino acid, it breaks the peptide bond involving amino acid at the carboxy end. So, the C double bond O at this end because, that is the carboxy end on this side. So, that is about chymotrypsin and, but one thing you have to remember is this adjacent amino acid there is one restriction, that if this amino acid happens to be proline, then the hydrolysis cannot take place. So, in the next amino acid which is connected to the aromatic amino acid happens to be proline, then this hydrolysis cannot take place. What about trypsin? Trypsin hydrolyzes very similarly, but it only recognizes highly basic amino acids like lysine and arginine. So, in this case you have the basic amino acid, a basic amino acid means amino acids having a very highly basic side chain; we are not talking about histidine here. I told that histidine although it is grouped into the basic amino acid but its basicity is much less as compared to arginine and your lysine ok.

So, if there is a basic amino acid then also the hydrolysis will take place, but this time the enzyme that is required is trypsin ok. So, trypsin hydrolyzes again the amide bond at the carboxy end of a basic amino acid, on the other hand pepsin has a much broader specificity, it actually can hydrolyze quite a few of the amino acids. So, pepsin is generally not used to cut the protein into smaller pieces. Now, let us go to the next page. So, the next point before I go into any other topic let me first solve one problem that based on whatever we have learned so far.

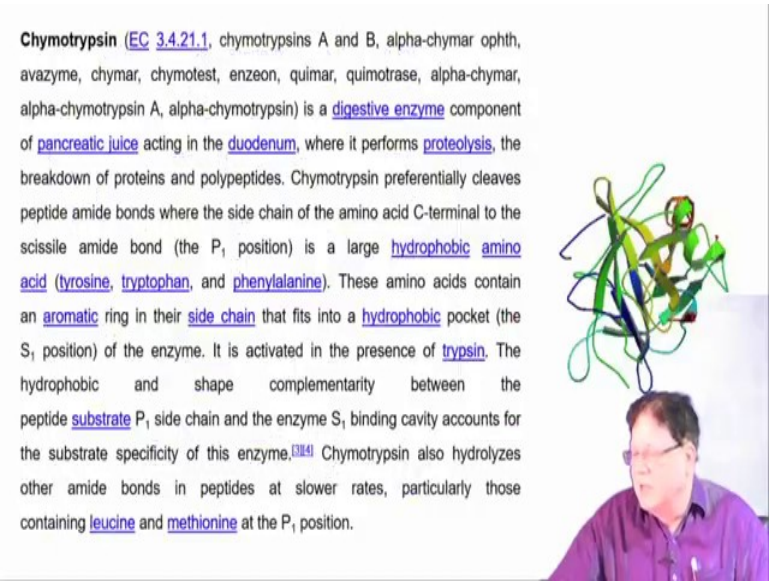
(Refer Slide Time: 15:01)



I have some pictures here which will show you that what is trypsin, and how it looks like. So, this is the crystal structure of trypsin and this trypsin, chymotrypsin and pepsin are essential enzymes for food digestion. So, when we take a food which contains lot of proteins; these proteins have to be broken down into smaller fragments and ultimately into amino acids. So, that it is absorbed through the intestine and then they are again used to make the different proteins which we need in our body.

So, first there is a catabolism, catabolism is breaking down the large molecules. So, we take food containing proteins and these proteins are hydrolyzed with the help of trypsin, chymotrypsin and pepsin to ultimately the individual amino acids which are then absorbed and then used by the system. Now, trypsin it looks like this and it has specificity as it recognizes basic amino acids, lysine and arginine strongly.

(Refer Slide Time: 16:18)

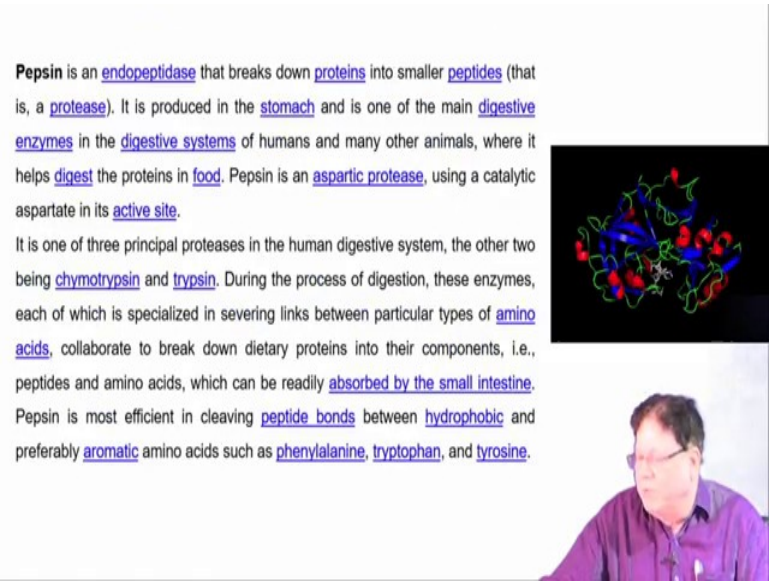


Then you have chymotrypsins and this is the crystal structure of chymotrypsin. What it does? It recognizes only the aromatic amino acids. The aromatic amino acids have a big hydrophobic side chain. So, the question is why they are recognizing certain amino acids and not the others?

When you will learn enzyme chemistry, we will see that what happens is that the molecule (prospective substrate) goes inside and there is a big hydrophobic pocket where, the aromatic groups first binds through weak interactions, hydrophobic interactions and then the hydrolysis or the reaction goes on.

In case of trypsin there must be a pocket which is negatively charged and it recognizes basic amino acids which are positively charged at the normal pH of a 7.2. So, that will go and bind there and that is how the specificity comes.

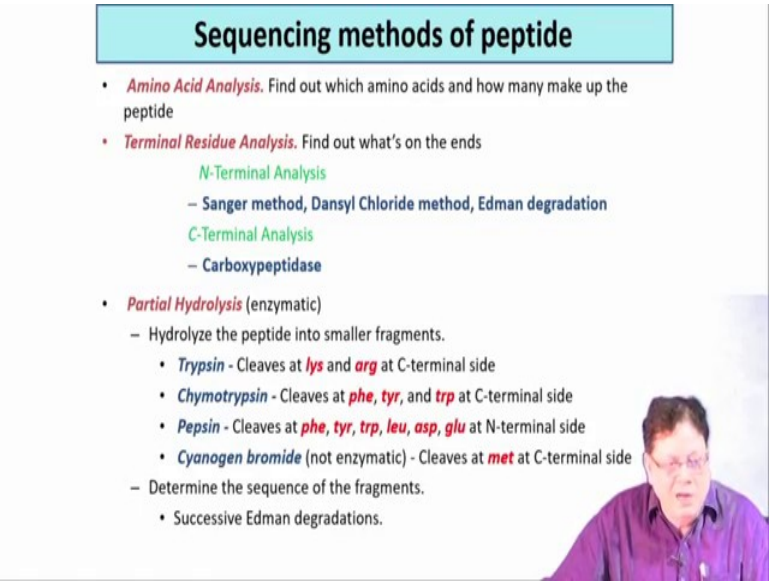
(Refer Slide Time: 17:20)



And then you have pepsin, pepsin is the other one whose selectivity is little bit broader, it recognizes other peptide bonds. Although here it is written only phenylalanine, tryptophan and tyrosine; that means, the aromatic amino acid. But I think later on it was found that it can hydrolyze even other, as it can recognize even other amino acids which are also hydrophobic not only aromatic.

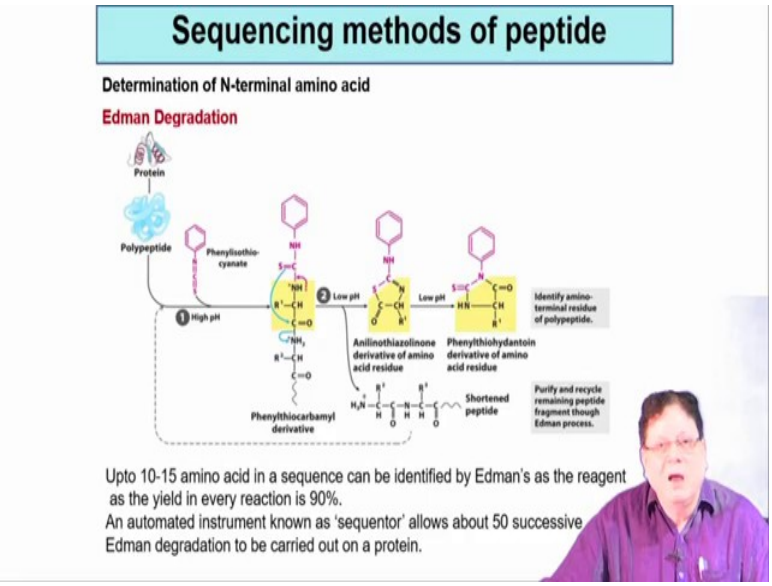
Because, hydrophobic groups belong not only to the aromatic groups, it is also present in leucine, isoleucine, valine, which have aliphatic hydrophobic groups. So, for pepsin, its specificity is little less that is why people do not rely on pepsin so much.

(Refer Slide Time: 18:11)

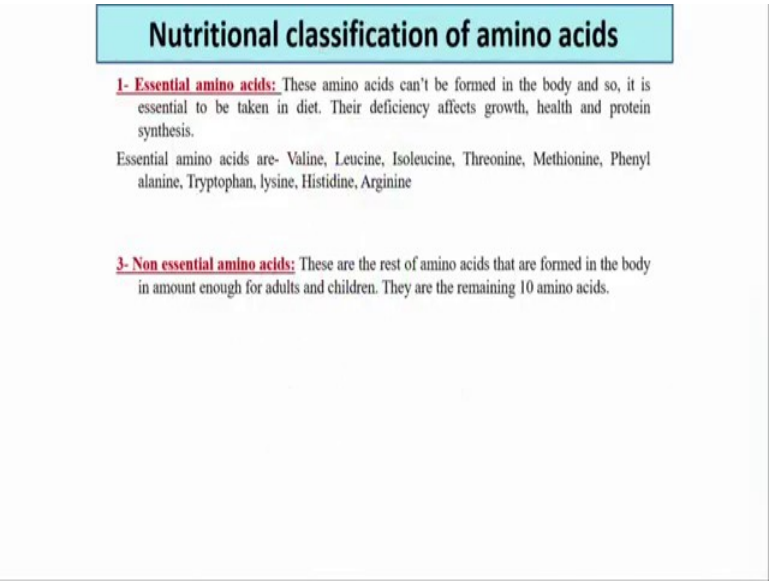


Now, I told you that when you take the food it has to be digested and fragmented into smaller amino acids and then these amino acids are absorbed and are used to make the proteins that we require.

(Refer Slide Time: 18:21)



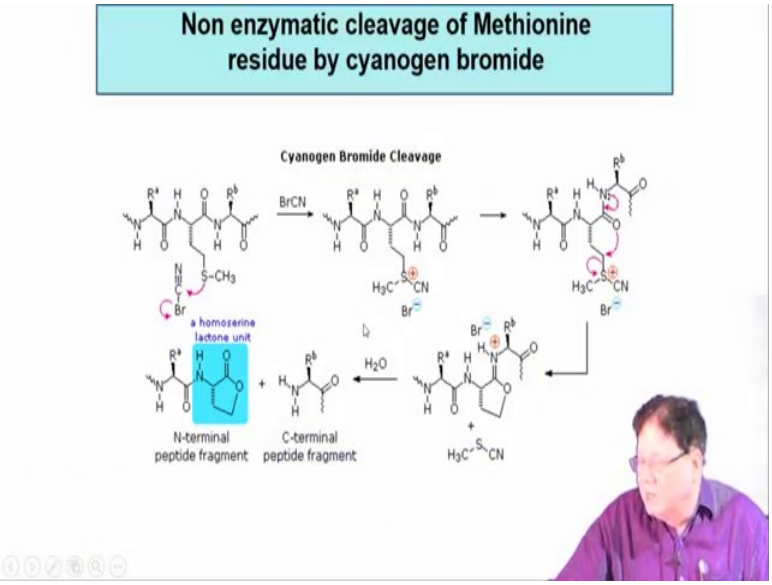
(Refer Slide Time: 18:24)



I think with that classification; that means, the question is that whether can we make our amino acids inside the body or not.. And, you know that there is a classification of amino acids based on that whether we can prepare the amino acids in our body or not. And, when the amino acids cannot be bio synthesized in the body, those amino acids are called essential amino acids.

There are 10 essential amino acids which have to be supplied from the diet and the rest 10 which we can prepare inside the body. Possibly you know this, this is the essential amino acids cannot be formed in the body, it is essential to be taken through the diet. And, the others are non-essential amino acids, they can be made in normal enough amount in adults and children both. So, that is another type of classification of amino acids.

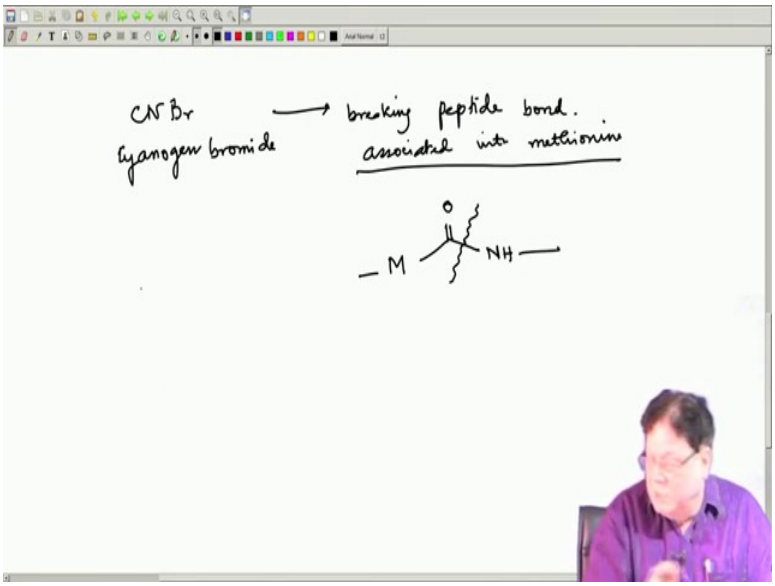
(Refer Slide Time: 19:27)



Now, let us do one problem. Now, let me again remind that this trypsin or chymotrypsin and even other proteases they failed to break the bond (even if there is amino acid of their choice which they recognize) when there is a proline adjacent to it. And, the reason is that proline has a secondary amino group which distorts the structure at that point.

Hence enzymes cannot accommodate the side chain of the amino acid which they prefer, that is how proline stops the hydrolysis at that point. Is there any chemical method of selectively cleaving a peptide bond? The answer is yes, there is one reagent has been developed and that is called cyanogen bromide.

(Refer Slide Time: 20:36)



Cyanogen bromide is a very simple reagent having the formula CNBr. Cyanogen bromide is very specific in breaking the bond peptide bond associated with only methionine. So, it only recognizes one amino acid. And what is the reaction that takes place here? The chemistry that takes place is first of all associated which means you have to be very clear that if it is methionine, the code is M and if it is NH, then this is the bond which is being cleaved by cyanogen bromide.

Now, let us try to draw the mechanism; the mechanism might be given here in the slides, I can show it here. See this is a peptide in the extended form that has been shown. And where is the methionine? This is the methionine with the side chain is CH_2CH_2 and then SMe. It is a very important amino acid you will later learn that this methionine has a distinct role. If we cannot make methionine in our body, our protein synthesis is going to collapse. But that is not the issue right now; we are trying to see that how cyanogen bromide is able to cleave the peptide bond associated with methionine.

So, this is your methionine part and this is the peptide bond. And, now as you treat cyanogen bromide, you know in cyanogen bromide, this carbon is highly electrophilic being flanked by two electronegative atoms. So, it attacks this cyanogen bromide, specifically the carbon and the bromine leaves as that is a good leaving group here. So, the bromine leaves and as soon as bromine leaves so, you get a species which is like this. S methyl was already there and now S that will be attached to a cyanide and the bromide

ion leaves from the cyanogen bromide. But, in the process what you have created is a sulfonium ion, where sulfur is having a plus charge.

And, now what will happen as you have done this; this sulfonium moiety has become a very good leaving group now. What the sulfur will want to do now? Sulfur will want to break this bond. So, how it will happen? It will happen by any of the adjacent nucleophilic center.

So, if there is any nucleophilic center nearby which can attack this carbon and if the size of the ring is quite appropriate that is also important; if there is an appropriate nucleophile which attacks this carbon, but that can form only a 4 membered ring or a 3 membered ring (which are strained), then that may not happen. In this case what happens is that there is a nucleophilic center and that is basically this nitrogen and this amide carbonyl.

Now, between these two, nitrogen lone pair of electrons is in conjugation with the carbonyl. So that means, the oxygen is having the greater share of the negative charge between the nitrogen and the oxygen. So, you just write it in a different fashion to show another confirmation; that means, the carbonyl on this side and this is the same peptide bond that is written in another confirmation; just to show you what is happening here. If you write in this fashion then you see this oxygen is quite near to this carbon which is now suffering the electron deficiency imparted by the positively charged sulphur; sulphur wants to go away taking the bonded electrons along with it.

So, what will happen now? This whole thing will be assisted by the nitrogen, hydrogen lone pair flows in here and then the oxygen becomes negative that attacks the carbon and this goes off. So finally, the molecule that comes out is this CH_3SCN as the by-product. Now, question is what happens to our protein? The protein is now in this iminium form; the nitrogen with a plus charge and then this will not be very stable.

Now, water comes and attacks this carbon and then ultimately hydrolysis occurs. So, this hydrolysis ultimately leads to a cyclic 5 membered lactone and this lactone comes from the methionine.

And what was the result of this? The result is that this peptide bond with which we have started, that peptide is now gone; that peptide bond is not there. And, you have a new

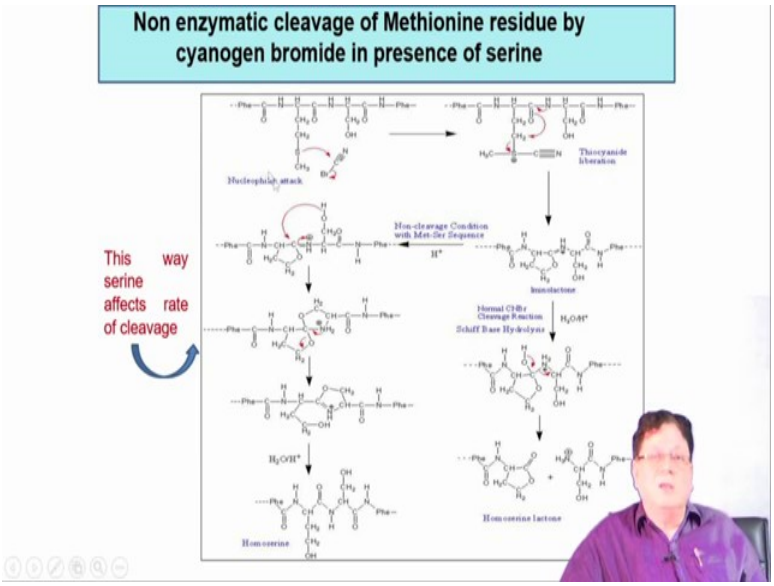
peptide which is devoid of this, from this left side, whatever is residing here, whatever the be the number of amino acids here, that is gone. So, the methionine part is converted to the N-terminal peptide fragment. So, this is a very important reaction. Cyanogen bromide is very specific only for methionine. However, this lactone now is called homoserine lactone. Why is it called homoserine lactone? What is serine?

Serine is a an α amino acid and then there is a CH_2 and then a OH. In this case, what is happening? There is a CH_2 that is serine, then you have another CH_2 so, that gives you the higher homologue of serine and then you have the oxygen.

So, it is basically homoserine which is converted into the lactone. So, ultimately it becomes a homoserine lactone

. So, if that is formed, then you know that there is a methionine at this position and the rest of the fragment, that is the other peptide portion from the C terminal end that remains.

(Refer Slide Time: 27:23)

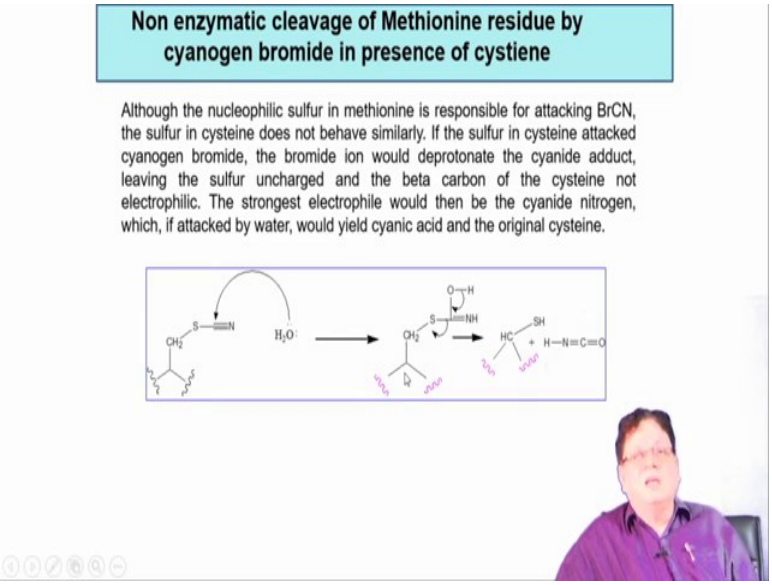


However, the cyanogen bromide reaction does not work if the second amino acid happens to be a serine. You can see here that this is serine and this was your methionine. The first reaction takes place without any problem, the bromine goes out so, you get this sulfonium ion. Now, in the next step, a peptide bond oxygen comes and attacks this carbon, the sulfur leaves; the chemistry is very similar to what was happening earlier.

But, now if you just go along this whole route, what you will see is that ultimately you end up not with the homoserine lactone, rather you end up with the homoserine itself. That means, you are not breaking the peptide bond here. So, although it has reacted with cyanogen bromide, the ultimate effect is that your methionine is converted into a homoserine which is the end product of each. This mechanism you can try by yourself, but it is very similar to the earlier one.

But I leave it to you as a problem that what is the mechanism, I am not going to share the slide right now; this slide shows that how does it ultimately go into the homoserine. So, again I repeat if there is a serine after methionine, then what happens is that methionine becomes homoserine and that is the end product and there is no breakage of any peptide bond. So, if there is no breakage it will be really difficult to now tell that whether there was a methionine present or not.

(Refer Slide Time: 29:25)



The reaction also fails if there is a cystein after methionine; what happens to cystein? You can expect a very similar type of reaction like methionine. If there is a cysteine, what is going to happen? Will cystein also react with cyanogen bromide and finally, it will also break the peptide bond? No, it does not. What happens is that the cystein will definitely react, it has a SH, it reacts with the cyanogen bromide. And so, it forms SCN and the bromide ion leaves.

The problem is now that the sulfur does not have any positive charge because, earlier it was having the hydrogen; in methionine it has a Me, but in your cystein it is SH. So, the hydrogen is lost here and thus sulfur does not have any positive charge. So, it is not a good leaving group in this case and hence the ultimate result is that the cyanide gets hydrolyzed. So, if the cyanide gets hydrolyzed you get isocyanic acid and you get back to your serine.

So, the serine reacts with cyanogen bromide and the cyanogen bromide ultimately is converted into this isocyanic acid. Isocyanic acid is the end product, but there is no cleavage of any peptide bond. So, that is a chemical reagent that we are talking about.

So, there is a chemical method of breakage. So now, we know how to cleave the protein into smaller pieces by using enzymes like trypsin or chymotrypsin. Or you can use a chemical reagent, if you want to know that whether methionine is present or not; or how many methionines are present; all these can be done by treating with cyanogen bromide.

So, that is the how the whole machinery works. Today the protein sequencing has become an automated, there is a machine; you program it and everything will be done. First it will be a digestion with trypsin or chymotrypsin and we will call it tryptic digestion or chymotryptic digestion. And, then it will be separated, the pieces will be separated and it will directly go to the amino acid analyzer. And, then do the Edman degradation and ultimately the data will come out in the computer screen that these are the sequences.

And finally, by the use of bioinformatics tool you can ultimately determine what are the fragments, what is the attachment of the individual fragments; with that you determine what is called the primary structure of the protein.

Thank you.