Experimental Biotechnology
Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology - Guwahati

Lecture - 42
Sequencing Techniques

Hello everybody, this is Dr. Vishal Trivedi from department of biosciences and bioengineering, IIT Guwahati and what we were discussing? We were discussing about the PCR in the last lectures. And now, what we are going to start is? We are going to start the new topic and that new topic is the sequencing reactions or the sequencing of biomolecules. So, the first question comes, what is the requirement of sequencing a biomolecule?

When you are generating a recombinant DNA for example, if you are you know PCR amplifying a particular fragment and then you are putting the restriction enzymes and then you are integrating this into a vector, the first question comes whether the PCR product what you have synthesized from the genome or from the target DNA, whether it is the same identical copy.

Because, we assumed that the DNA polymerase what we are using for example, taq DNA polymerase or the pfu DNA polymerase is actually giving you the identical sequence. But, whether that is really been the case, because, if there are mutations, if there are replacement of some of the nucleotides from the template DNA that actually may lead to a wrong amplification or the wrong synthesis of the protein and subsequently it may affect the downstream work.

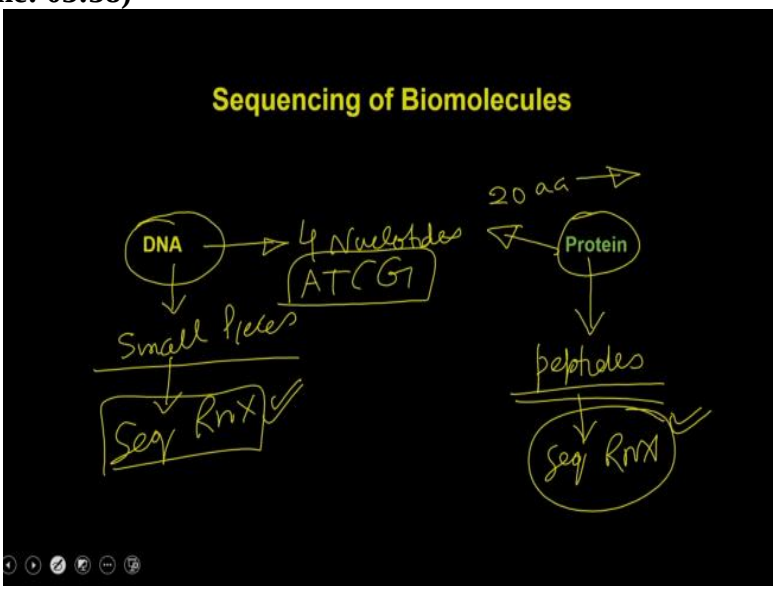
For example, if you are interested to study the activity of an enzyme, you have cloned that enzyme from the genome and then if you have some of the mutations like then some of the nucleotides have been replaced from the template DNA, then what will happen is and if these mutations are present within the active site or some of the crucial sites, then you may not see the desired results, you may see some changes in the activity and other kinds of shows. So, this can be verified simply by sequencing of your recombinant DNA.

Similarly, if you are synthesizing a protein or suppose you are isolating a protein from an unknown sources then the first question comes what is this protein so, if you want to identify

your protein, you also have to deduce it is you know amino acid sequence. So, that actually is going to give you the idea about the identity of this particular protein and then eventually you can be able to identify the gene and then using that information, you can be able to clone the protein.

So, when we are talking about the sequencing reactions, the sequencing reactions can be done for the 2 macromolecules one is DNA and the other is protein.

(Refer Slide Time: 03:38)



So, when you are doing a sequencing a biomolecule, it could fall under the 2 category either it could be a DNA or it could be a protein, irrespective of whether you are doing the sequencing of the DNA or whether you want to do a sequencing of the protein, the discrete steps are always been remained identical. For example, the first step itself that you have to break the DNA into the small pieces, if in case the DNA is of a very large number.

So, you have to, you know, first break it into the small pieces, and then you are actually going to perform the sequencing reactions. So, sequencing reactions could be different for the DNA and could be different for the protein. Similarly, for the protein also, you know that the protein is made up of the 20 different types of amino acids, whereas, the DNA is made up of the 4 nucleotides.

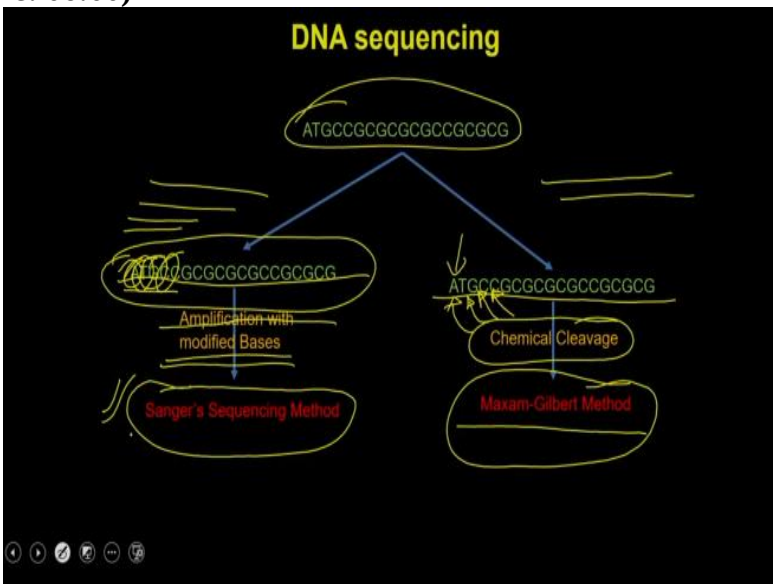
That is why the complexity for the DNA sequencing is less because, what are these nucleotides you have the 4 different nucleotides like ATGC, in whereas, in the case of protein, you have the 20 different types of amino acids, which are actually going to make the

system or make the things more and more complicated because now, you have to set these all the 20 amino acids in such a way that what is the this been present.

But with the complexity it also gives the flexibility that you cannot, you know because, once you in the first step once you are going to generate the peptides from the protein sequence, these peptides are then you have to go for the sequencing reactions. So, in the sequencing reactions, you might have to do sequencing of the each and every amino acid. So, the sequencing reaction what you do for the DNA or the sequencing reaction, what you do for the proteins are different in terms of the chemical reactions.

Because, the DNA is different from the protein and that is how, the sequencing reaction has to be different. So, first we are going to start with DNA and then we are going to discuss about the different sequencing protocol what you can follow to sequence the DNA.

(Refer Slide Time: 06:00)



So, as far as the DNA is concerned, for example, if this is the nucleotide sequence, what you would like to sequence what you have to do is you have to take this DNA and then you have to do a PCR amplification with a primer. So, you have to set a primer of this particular sequence. So, in most of the cases when we are doing a DNA sequencing you know, that where the my insert is present.

So, in some cases for example, if you are talking about recombinant DNA, you are always using the primer which is for the some of the tags or like promoter sequences. So, you can just use the for example, you if you have any pet series vector and you are cloning the DNA

into the pet series vector, you can easily take the T7 promoter primers, so, that actually is going to amplify this particular sequence and while it is amplifying you can be able to put some modified bases.

So, what will happen is, when you add the modified bases, it is actually going to terminate the synthesis at that particular sequence. So, you can have the termination at A, you can have the termination at T, you can have the termination at G and you can have the termination at C. So, that is how you can actually have the fragments of the different lengths and all these fragments can be analyzed in a gel or as well as in the capillary electrophoresis.

And that is the major method which has been developed by the Sanger and that is it this method is called as the Sanger sequencing method. Whereas, in the other case, what you have to do is you have to have this sequence which you are interested to synthesize. So, now, instead of doing the amplification, what you can do is you can simply use some chemical reagents and you can do a chemical cleavage for example, you can do a cleavage for A you can do cleavage for T, you can do a cleavage for G and you can do a cleavage for C.

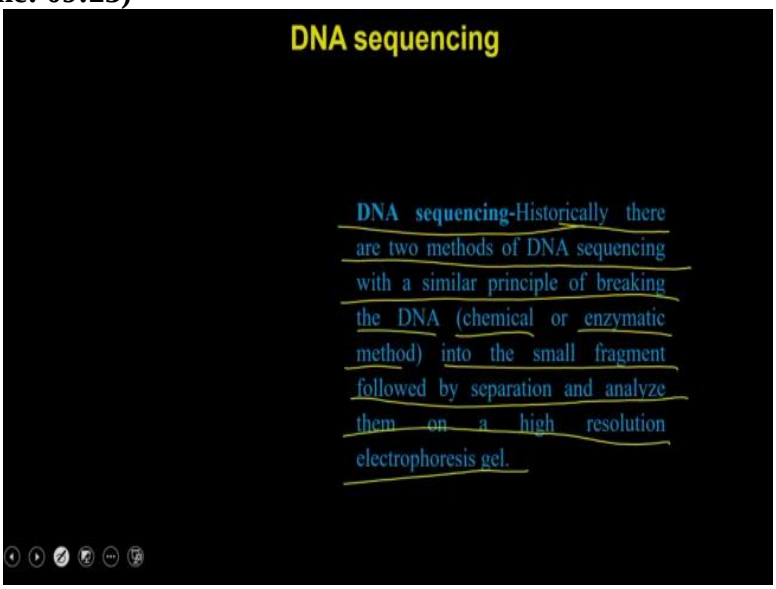
So, that case what will happen is that you are going to get the fragments and all these cleavages are actually going to break the DNA at that particular site. So, because of that, you are going to get the DNA of different lengths. And now all these lengths can be analyzed. And depending on the pattern, you can be able to deduce the sequence and this method is being discovered by the Maxam Gilbert and that is why this method is called as Maxam Gilbert method.

So, as for the DNA is concerned you have the 2 methods one is called Sanger method where you are using the PCR as a tool and you are adding the modified bases and that modified base is actually terminating the PCR synthesis, whereas, the other method is called as the Maxam Gilbert method where you are actually doing a chemical reactions with the DNA. And that chemical reaction is breaking the DNA at the place where the reaction is actually modifying the bases.

For example, if it is against the A T or G or C, it is actually going to broken the DNA at every A, whatever the A is present in this particular sequence. And that is how you are actually going to get the nucleotide or the stretch of DNA. And that can be, you know, separated and

the pattern of that fragments is actually going to help you to deduce the sequence. Let us start with the Sanger sequencing method.

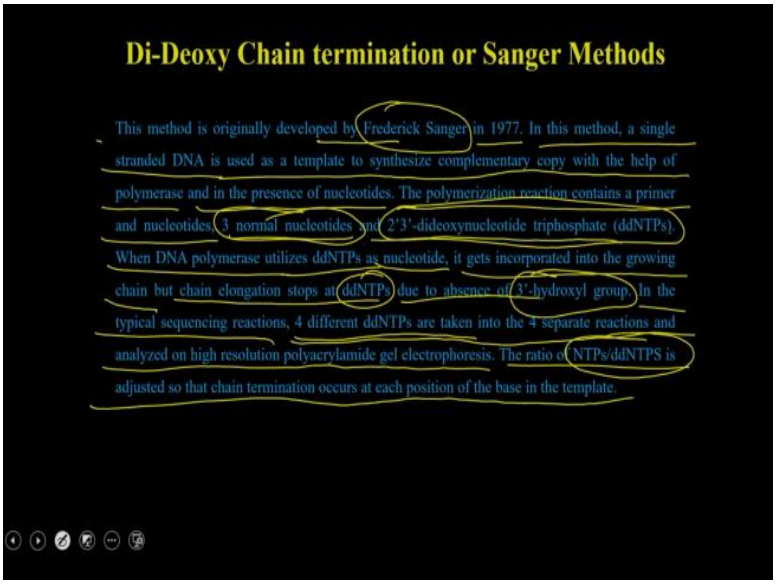
(Refer Slide Time: 09:23)



So, in the Sanger sequencing method, you have the, there are 2 methods of DNA sequencing with a similar principle of breaking the DNA either the chemical or the enzymatic method into the small fragment followed by the separation and analyze them on a high resolution electrophoresis gel. So, this is the basic principle of the DNA sequencing where you are actually taking a DNA sequence you are breaking it either with the help of the PCR with the help of the modified bases or with the help of the chemical or the enzymatic method.

And then it is actually going to give are different types of fragments then you separate these fragments and it is actually going to give you the, if you analyze that sequence pattern, you can be able to deduce the sequence.

(Refer Slide Time: 10:12)

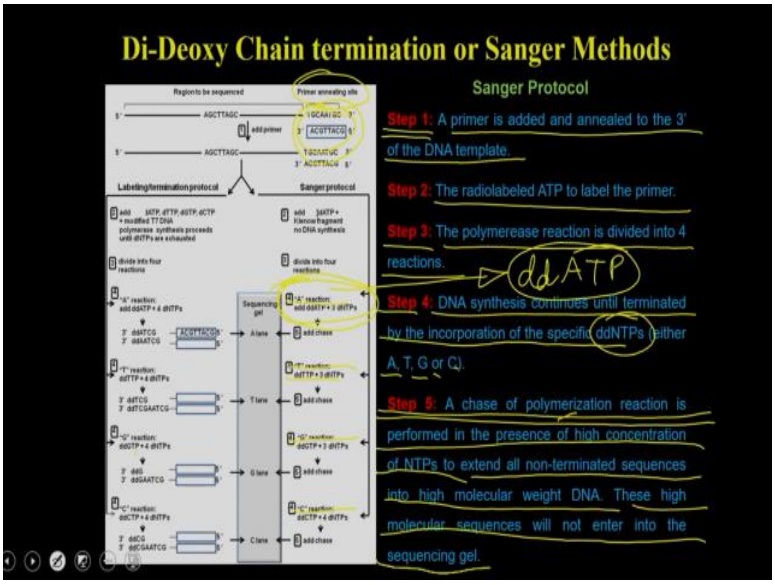


So in the Sanger's method this method is originally been developed by the scientist called Frederick Sanger in the year of 1977. In this method, a single stranded DNA is used as a template to synthesize the complementary copy with the help of polymerase and in the presence of nucleotide. The polymerization reaction contains a primer and a nucleotide. There are 3 normal nucleotides, which means the standard nucleotide like dATP, dTTP and all that and then are modified base like the ddNTP like 2, 3 dideoxy nucleotides triphosphate.

So, when the DNA polymerase utilizes the dideoxy nucleotides as a nucleotide, it gets incorporated into the growing chain, but the chain elongation stops at the ddNTP because due to the absence of the 3 prime hydroxyl group, so, actually in a dideoxy nucleotides, what you have is a cyclized product. So, what will happen is that, because you have known 3 prime free end the incoming nucleotide will not be able to connect in this particular nucleotide and because of that, the chain is going to be terminate at that particular site.

So, in the typical sequencing reaction, what you are going to do, you are going to do a 4 reactions. So, you are going to have the 4 different types of dNTP's taken into the 4 separate reaction and analyze them on a high resolution polyacrylamide gel electrophoresis. The ratio of the NTP's to dNTP is adjusted so that the chain terminations occur at each position of the base in the template.

(Refer Slide Time: 11:53)



So, in the first step, what you are going to do is you are going to do a primer. So, in step 1, what you have to do is you have to synthesize a suitable primer so, that it will be able to anneal to your target sequence. So, in the step 1 a primer is added and annealed to the 3 prime end of the DNA template. So, that is what you are going to do to first you have to design a primer that primer is going to be directed against a particular gene or particular promoter which you are interested to synthesize means your template DNA.

Then you are going to add the primer to radio labeled ATP to label the primers then step 3 polymerase the action is divided into the 4 reactions. So, then you are going to do a radio labeling of your primers so that wherever the primer is present, it is actually going to give you the signal. And after that you are actually going to divide these reactions into the 4 reactions in the step 4 the DNA synthesis continue until the terminated by the incorporation of the specific dinucleotide either A, T, G or C.

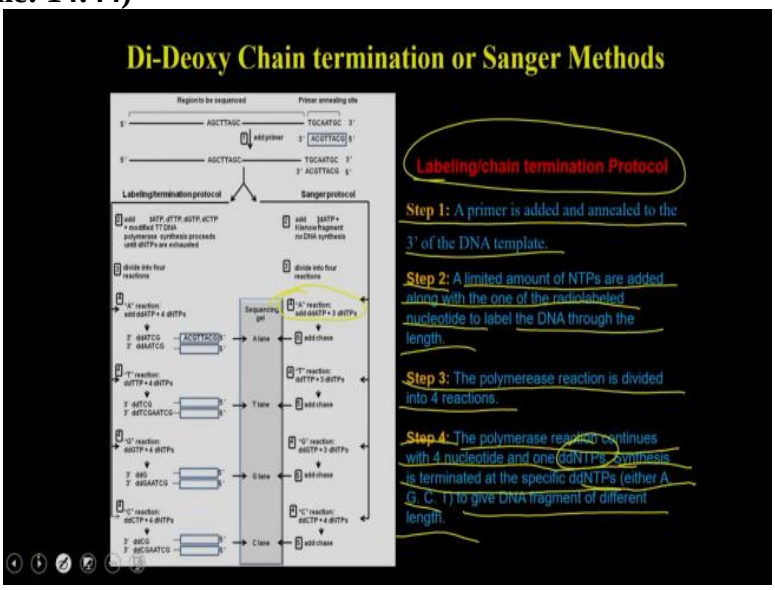
So, what we are going to happen is in the step 4, you are going to divide that into the 4 reactions. So, first reaction is for the modified ddATP. The second is for the ddGTP, the third is for ddTTP. And so that is how you are going to do is the 4th reactions you are actually going to divide into the 4 parts like A reactions, D reactions, G reactions and the C reaction which means in the A reactions.

You are going to have all the nucleotides normal except that A nucleotide is going to be replaced by the ddATP which means the dideoxy adenine dinucleotide, which means the A is going to be modified if it is in the A reactions similarly, for the T reactions, the T is going to

be modified this means for the A reactions, all the termination is going to occur just after the A where as in the T reactions, the termination is going to occur just after the T similarly, the same is true for the G and the C reactions.

Now, you have to do a chase of the polymerizations which means you have to allow the enzyme to go for the elongation step so that it can be able to synthesize the DNA, but it will terminate wherever it will find their corresponding nucleotide for example, in the case of A it will terminate at the A in the case of T it is going to terminate at T so when in the chase in the step 5, a chase reaction is performed in the presence of high concentration of NTP to extend all non terminated sequence into the high molecular weight DNA, these high molecular sequences will not enter into the sequencing gel.

(Refer Slide Time: 14:44)



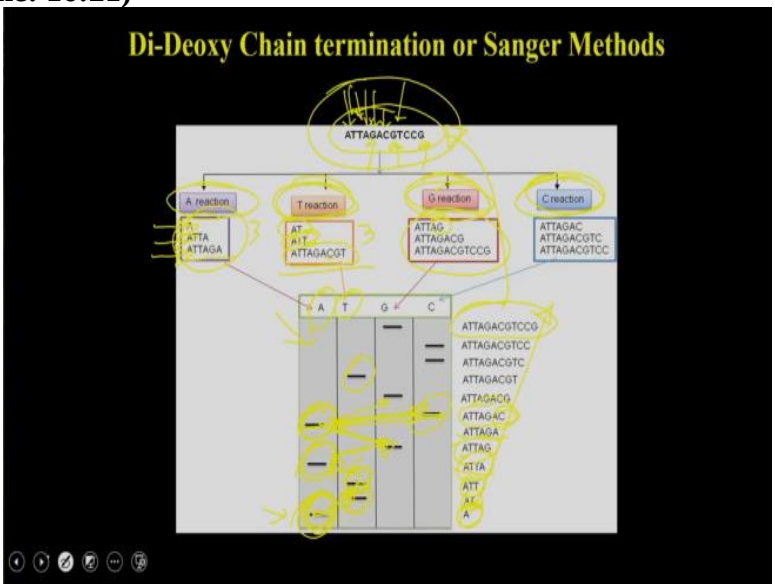
Apart from that, you can also do another protocol which is called as the labeling or the termination protocol. So what you have seen just now is actually the original protocol what is being developed by the Sanger but in this modified protocol if you are doing a labeling as well as the termination protocol. So, in this one the step 1 remains the same that you are actually going to anneal a primer to the 3 prime end of the DNA.

Then in step 2 limited amount of NTP's are added along with the one of the radio labeled nucleotide to label the DNA throughout the length, which means if you are going to radio labeled the DNA then the in the step 3 the polymerase reaction is divided into 4 reactions, which means the A reaction, T reaction, G reaction and C reactions. And then the step 4, the

polymerase reaction continuous with the 4 nucleotide and one of the ddNTP's which means the modified bases.

So, in the case of A it is going to be a modified base for the A in the G it is going to be modified phase for G and so on the synthesis is terminated at the specific dideoxy nucleotide either A, G, C or T to give the DNA fragments of the different length. So, irrespective of whether you are using the Sanger's protocol or whether you are using this modified labeling and termination protocol, ultimately you are going to get the DNA fragments of the different length and now, these DNA fragment has to be resolved. So, that you will be able to identify or you will be able to understand the size of these DNA.

(Refer Slide Time: 16:21)



So, in the third step, what you are going to do is you are going to take this DNA sequence and then you are already divided this into the 4 reactions like A reaction, T reaction, G reaction and C reactions. So, what will happen is, when you are dividing this into A reaction, it is actually going to terminate at every place where you have the A which means it is going to terminate at this point it is going to terminate at this point it is going to terminate at this point, and so on.

So, what you are going to get the fragment is you are going to get A you are going to get ATTA and you are going to get ATTAGA which means all the 3 fragments you are going to get when you are running the reactions in the A, similarly, you when you are doing the T reaction, it is going to terminate at every T which means this all these T's it is going to terminate.

So that is how it is going to give you the AT, ATT and the ATTAGACGT so you are going to get the 3 fragments here as well, you are going to get the 3 fragment here as well. And after in the G reaction, it is going to terminate at every G so it is going to terminate at this point it is going to terminate at this point, and it is going to terminate at the last point. So that is how you are going to get all these fragment.

What you see is that the last nucleotide in any reaction is going to be the reaction to be the that nucleotide which has been modified which means the G is actually the modified nucleotide and the same is true for the C reaction. So now what you have to do is you take all these 4 reaction which is corresponding to this particular sequence, and then you analyze them on to a high resolution polyacrylamide gel electrophoresis.

So what will happen is the A is going to give you the 3 fragments, like this, T is going to give you the 2 fragments like this, at the end what you are going to get you are going to get the fragments and you are going to now start analyzing these sequences so that you will be able to deduce the final sequence. So, what you have to do is you have to first start with the, whatever the nucleotide is giving you the fragment at the end because that is the smallest fragment present.

So, you have to take that into the first then you have to go like this you have to go like in a zigzag manner like this like this. So, in case the 2 fragments are nearby, then you can actually take up like so, you have to start reading from the wrong orientation like so far example A then AT then ATT and ATTAATTAG like that. So, if you go into the reverse orientations, you will be able to deduce the sequence.

So you start from the reading from the bottom, and that is how you have written A because this is a single nucleotide what you are getting and then you entered into the T because T is the second base pair. So, what you have done, you have added a sequence like A and T because the terminal nucleotide is going to be T and then you went a little ahead so it is going to be give you the ATT because the next is also T and then you went to this one.

And it is actually going to give you a sequence of A that is how you said the sequence is going to be ATTA and then you the second nucleotide is this so you said ATTAG and then you went to this one because this one is going to give you the this is on the lower side this is

on the higher side. So that is how it is going to give you the ATTAGA and then you enter into the C and it is going to give you the ATTAGC. So, you see, wherever I am taking up this nucleotide, I am just adding the C at the end because we are taking the reaction from the C reactions.

So, if you read like a in a reverse pattern, you will be able to deduce the sequence and ultimately what sequence I am getting, I am getting the same sequence what I started with. So, that is how you can be able to analyze the sequencing reactions what you are going to get after the you know Sanger's protocol or the chain termination protocols. And once you got this pattern, you will be able to deduce the DNA sequence. So, in a dideoxy chain termination method what we have discussed so far.

(Video Starts: 20:46)

So, in a Sanger sequencing method what you have is you have a you know DNA what you have a target DNA and that will have to sequence. So, what we have if you want to do a Denis sequencing using the Sanger sequencing method, you have the 2 way either you can go with the gel filtration chromatography, gel electrophoresis or you can do the capillary electrophoresis.

So, the first step is that way, you have to take the your DNA into the eppendorf and then you have to add the primers, these primers, you have to add into the 4 reactions, if you remember, we have said that we have to divide the reactions into the 4 reactions, and then you have to add the DNA polymerase into the each reactions. So, you have to add the reaction from number 1, 1, 2, 3 and 4 and once you added the DNA polymerase into the 4 reactions.

Then you are going to add the nucleotides you have to add the all the 4 nucleotides like C, T, A and G in all the 4 reactions. And in the subsequent step, you are going to add the dideoxy nucleotides. So, as if you recall, you can have the 4 different reactions A reactions, T reactions, C reaction and the G reaction and in all of these you have added the dideoxy nucleotide.

And what is the difference between a normal versus dideoxy nucleotides the difference between a normal dNTP is that it has the 5 prime phosphate and you have the hydroxyl group at the 3 prime whereas, in the case of dideoxy nucleotides, you have this OH is missing and because of this OH is missing, it is actually going to do the nucleotide is going to stop the

synthesis let us see how it is actually going to stop the synthesis. So, you can imagine that if there is a dNTP's it actually will going to form a bond by the phosphodiester linkage.

And that OH is still there so, that will continue the synthesis whereas, in the case of the dideoxy nucleotides once the dideoxy nucleotide is going to use it is phosphate and going to form the phosphodiester linkage. Since that OH is missing on this site, it will not going to allow the incoming nucleotide to bind. So, that is how it is actually going to stop the synthesis of the DNA synthesis by the DNA polymerase.

Now, what you have to do is you have to take the 4 reactions and put it into the thermal cyclers where you have the all the reagents and the thermal cycler you have the different steps like in the first step you are going to do the denaturations. So, in the denaturation step, you are going to increase the temperature of the thermal cycler. And once you increase the temperature of the reactions.

The 2 strands of the DNA are going to remove are going to be attached and then you are going to add the primers and the primer will anneal and then there will be an extension, but what will happen is if there will be a dideoxy dNTP's then it is actually going to terminate the reaction wherever the dideoxy the enzyme will find that dideoxy what you see here is that the termination is happening at every A.

So if it is going to find the A so it will going to give you the DNA of different reactions same is going to happen for the dideoxy T reactions. So in all the T wherever you have the T it is actually going to terminate the reactions, same is true for the C reaction. So wherever you could find the C it is actually going to terminate like here it is going to terminate and so on. So, same is going to happen even for the G reactions that wherever it is finds the G it is actually going to terminate.

For example, in this case it has finds the G at the end, so it is going to terminate at this point, then it is going to terminate at this point. And that is how you will see they are actually going to give you the different reactions. Now, what you have to do is you have to take out these reactions from the thermal cyclers and then you have to resolve these samples on to the gel electrophoresis. So, you have to take the all the 4 reactions and load it into the 4 different wells.

And you know that that DNA is negatively charged. So it is actually going to dissolve onto the gel electrophoresis. So, you load the 4th reactions, and then you connect it to the power pack and you turn on. So when you turn on the DNA is going to run from the negative to positive because the DNA is negatively charged, because the DNA is having a phosphate backbone and that actually gives the negative charge to the DNA for because of the negative charge it goes towards the positive electrodes in the gel electrophoresis.

And you know that this migration is inversely proportional to the size of the DNA for the larger, smaller DNA will run faster, and the larger DNA will learn slower. Now, what you have to do is, once you have resolved the DNA, you have the 2 ways in which you can be able to visualize this DNA, you have the either utilization of the radio labeled primers, or you can use the labeled dNTP's that labeling you can do with the radioactivity.

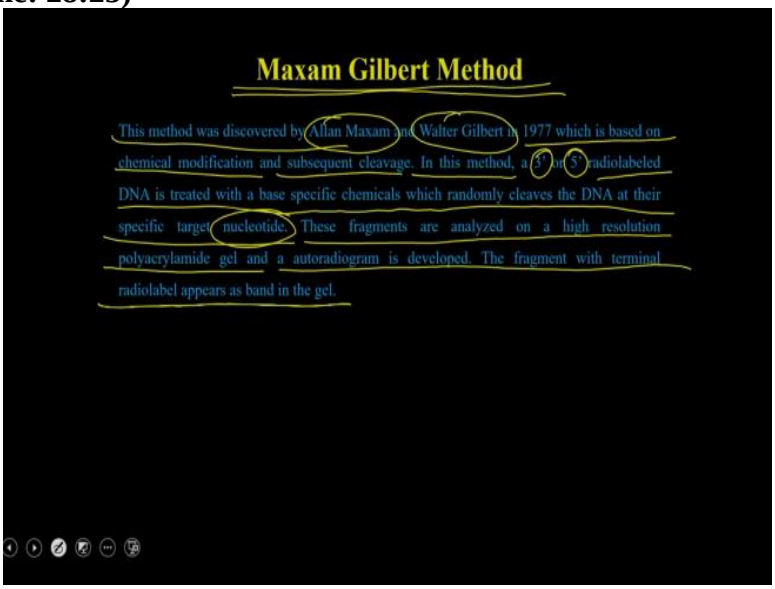
So what you can do is you can use the P 32 labeled DNA, a label basis and that is actually going to label the DNA when it is actually going through with the synthesis. So irrespective of whether you use the label primer or the label ATP's, once you have got the DNA being resolved onto the agarose, then you have to do is you have to visualize that DNA band with the help of the autoradiography. So what you have to do is you have to take this gel, you have to take the agarose gel and put it into the gel cassette.

Put the X ray film and then you close it and let it be exposed for overnight or 72 hours. During that period, the radioactivity what is present on the gel is going to exposed to the X ray film. And that is how you are going to get the bands of the DNA. Now what as we discussed before, you have to read it in the reverse orientation, which means you have to read it first this sequence then this sequence then this sequence that this sequence, then this sequence.

And this sequence, and that is how you are going to get the sequences from each band. And what you have is this is yours DNA sequence. So what you have to do is you have to take this sequence and then these sequences and that is actually is going to be the DNA sequence what you are going to get from the Sanger's method. So, this is all about the Sanger's sequencing method, what we have discussed.

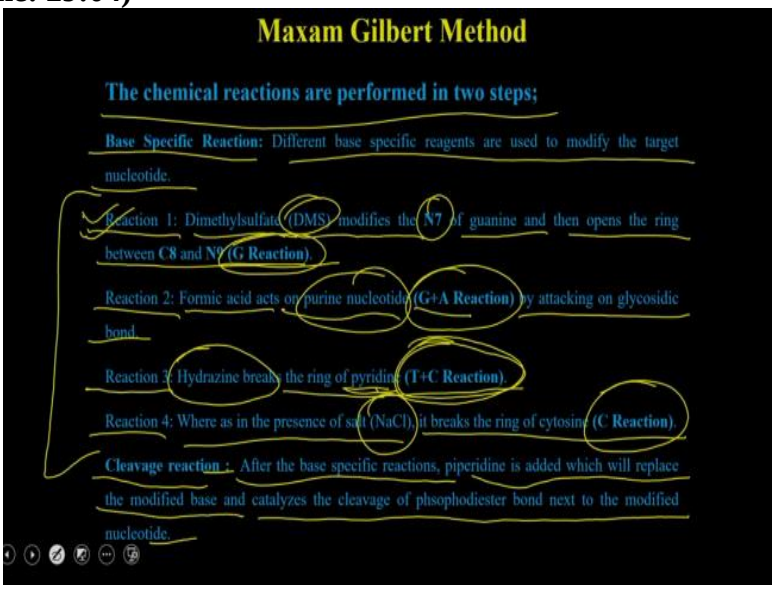
(Video Ends: 28:24)

(Refer Slide Time: 28:25)



Now let us move on to the next method. And the next method is called as the Maxam Gilbert method. So, this method was developed by the Allan Maxam as well as the Walter Gilbert in 1977, which is based on the chemical modification and the subsequent cleavage. In this method, a 3 prime or the 5 prime nucleotide radiolabeled DNA is treated with a base a specific chemicals which randomly cleaves the DNA at a specific target nucleotide. These fragments are analyzed on a high resolution polyacrylamide gel an autoradiogram is developed the fragments with the terminal radiolabeled appear as a band in the gel.

(Refer Slide Time: 29:04)



So, chemical reactions are performed in 2 steps. So, the you have the 2 reactions when you have a base specific reaction and so, during the base specific reactions, different bases reagents are used to modify the target nucleotide. For example, in a reaction one you are

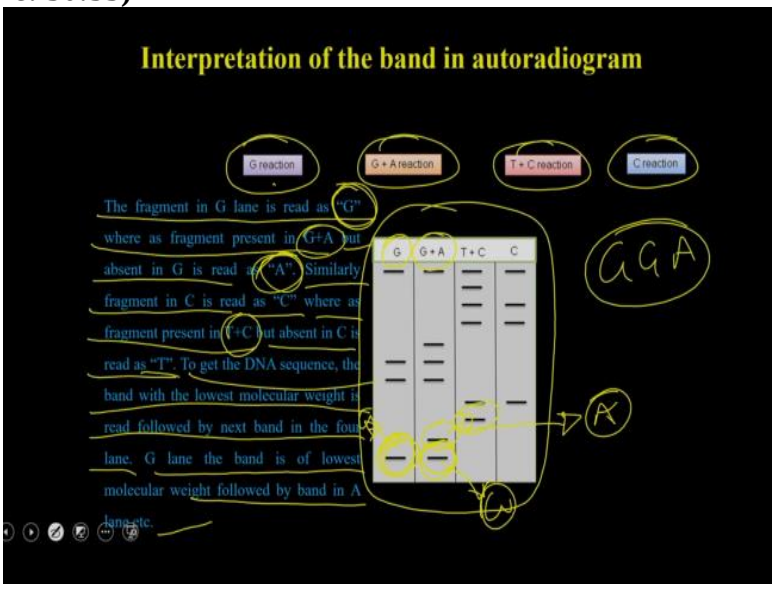
going to use the dimethyl sulphate like DMS modified the N7 of the guanine and then opens a ring between C8 and C9 so that is called as G reactions.

So, in the reaction 1 you are modifying the A nucleotide, G nucleotide, then the reaction 2 you are using the formic acid which is act on purine nucleotides. So that is actually going to modify the G plus A reaction by attacking on the glycosidic bonds. Then the reaction 3 you are going to use the hydrazine which is actually going to break ring the open so that is actually good.

So that is going to be attack on the pyridine which means it is these are the reactions which are going to cleave just after the T or the C. So, these are called as T plus C reactions. And then the reaction for where in the presence of salt it breaks the ring of the cytosine and that is called as the C reactions and the cleavage reaction after the base was fixed reaction the piperidine is added, which will replace the modified base and catalyze the cleavage of the phosphodiester bond linked next to the modified nucleotides.

So, first you are going to do a bases specific reaction and then you are going to do a cleavage reactions with the help of the piperidine and that actually is going to cleave the DNA just after that particular modifications. So, what you have you have the 4 reactions, you had the G reactions, you have G plus A reactions, you have T plus C reactions and you had the C reactions.

(Refer Slide Time: 30:53)



Now, that you have the G reactions G plus A reaction T plus C reactions and the C reactions, the fragment in G lane is read as the G whereas, the fragment present in G plus A, but absent in G is read as A similarly, so, what you can see is now you have 4 reactions, all these 4 reactions I have analyzed onto a polyacrylamide gel, the rule is that the G reactions are considered as G whereas, the G plus A reactions in case suppose you have the band for example, here you have a band in G you have also a band in A.

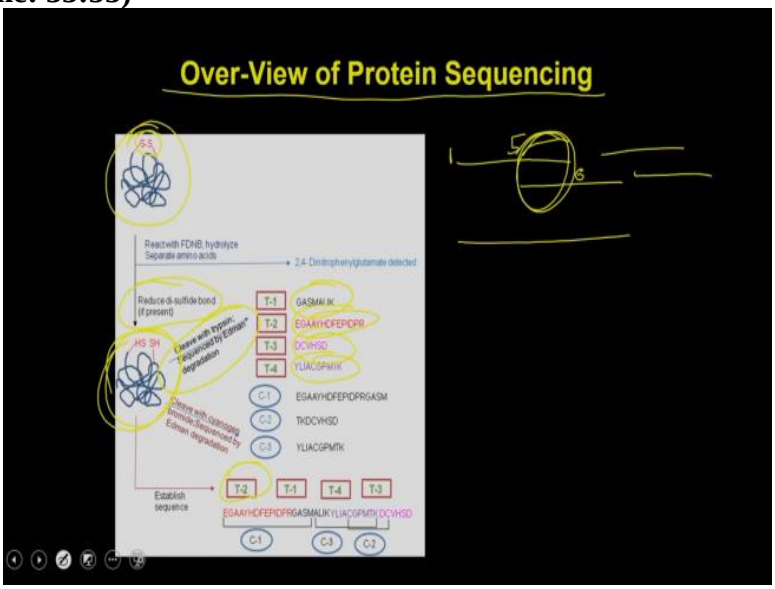
So, in that case the it is going to be read as A which means in this particular reaction you are going to have the A similarly, the fragment in C is read as C whereas, the fragment present in T plus C, but absent in C is known as the T to get the DNA sequence the band with the lowest molecular weight is read followed by the next band in the 4 lane G lane the is of lowest molecular weight followed by the band in the A lane which means, if you have the 2 A's like that.

And so, you have to read the bases exactly the same way as we just discussed in the Sanger method, the conventions are different, you are going to have the G reaction, you are going to have G plus A the actions for if there are is same height, then it is actually the you are going to add you are going to read the G reaction, but if it is a G reaction and you have the slightly higher length, then this is going to be considered as A reaction.

Whereas in this case it is going to be considered as G reaction whereas this is also going to be considered as G reaction which means the sequence is going to be GGA like that. Similarly, if you go to here it is going to be considered as T plus C reaction, which means it is going to be considered as T reaction because you do not have any DNA next to the in the C reaction. But for example here this is going to be called a C reaction here this is also going to be considered a C reaction, because it is actually going to be of a same molecular weight.

So, that is what you have to do when you are would like to utilize the Maxam Gilbert method to through the sequencing. So, this is all about the sequencing of the DNA molecules. And with the help of the both method like Sanger method as well as the Maxam Gilbert method, you can be able to deduce a particular DNA sequence and it has an advantage that you can be able to verify the clone or whatever you have synthesized or you whatever you have amplified and verified that I am whatever I am going to use the DNA is of the same sequence what I have started with.

(Refer Slide Time: 33:53)



So, let us move on to the next molecule and the next molecule is called as the protein molecule. So in the protein molecule, the steps remain the same except that the reagents are different. For example, in the case of protein, the first event is that you are going to use and denature the protein structures. So first, you are going to break the disulfide linkages. So you are going to reduce the disulfide linkages.

So that is how you are going to have the protein which note as disulfide presents and then you are going to use the trypsin or some other degrading enzymes. And that is actually going to give you the different peptide fragments. And all these peptide fragments are then going to be analyzed with the help of the sequencing reactions and that actually is going to give you the different sequence then you are going to set the sequence and you can be able to deduce the complete length or complete sequence of this particular protein.

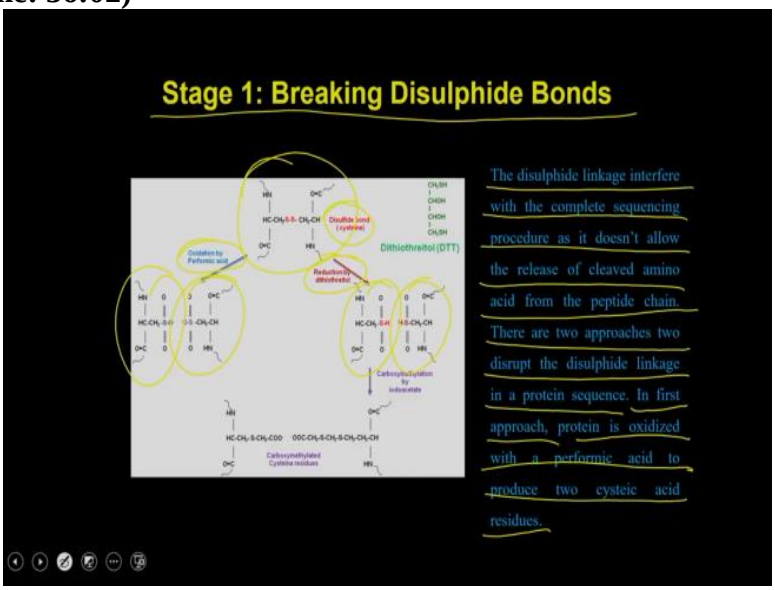
As if you remember I have said in the beginning itself because you have the 20 different types of amino acids. It brings a complexity because you have to have the more and more diversity in the sequence. But at the same time, it also brings you the convenience, because when you got the sequence of the small peptide sequences, the matching these sequences are easy.

Because you are going to have 1 sequence like this, you're going to have the second sequences like that. So you are going to have some of the flanking sequence, which is actually going to match with the next fragment. And because of that, you can be able to use

that to say, this is the number 1, this is going to be number 5, and this is going to be number 6, like that.

So you can be able to deduce the whole sequence of these, because you have the multiple overlapping sequences, you can be able to put these fragments back, because once you are digesting with the trypsin, you are going to have the different fragments. Now once you are done the sequencing, you know the sequence of each and every fragment, and but these fragments are going to have the overlapping regions, and all these overlapping regions can be used to deduce the sequence of your particular protein.

(Refer Slide Time: 36:02)



So in the stage 1, you are going to have the breaking of the disulfide linkages. So this is you can imagine that this is a protein, you can have the breaking of disulfide linkages either by the 2 method either you can use the oxidation by the performic acid and that is actually going to break the disulfide linkages, which is present between the 2 cysteine residues or you can do a reduction by the DTT.

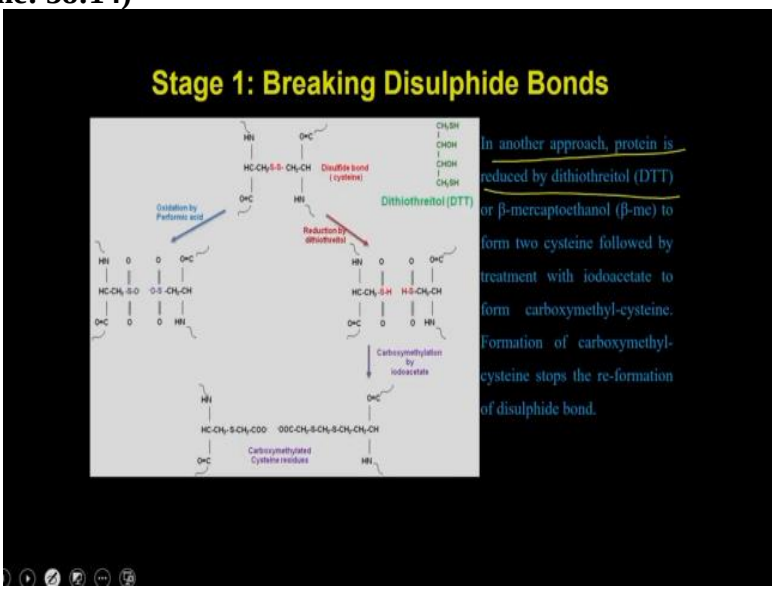
And that is also going to break the disulfide linkages. So, the disulfide linkages interfere with the complete sequencing protocol as it does not allow the release of the cleaved amino acid from the peptide chain, there are 2 approaches one disrupt the disulfide linkages in a peptide sequence in first approach, the protein is oxidize with the performic acid to produce the 2 cysteic acid residues, whereas in the second approach, you can use DDT.

And that is actually going to break the bond and once you got the disulfide linkages broken, then you what you can do is you can add some reagent so that it will remain like that. So, what you have to do is you have to block these SH groups in with some enzyme or some blockers, so that it should not get back because as soon as you are going to make the conditions reducing again these disulphide linkages are going to be formed.

And the way these disulphide linkages are interfering is that for example, you have a disulphide linkage like this. So when you are going to do trypsin digestion and trypsin is going to cut this fragment like this, even if it has cut the there will be a disulfide linkages. So that is how it is actually going to give you a bigger peptide, which means it is going to give you a peptide like this.

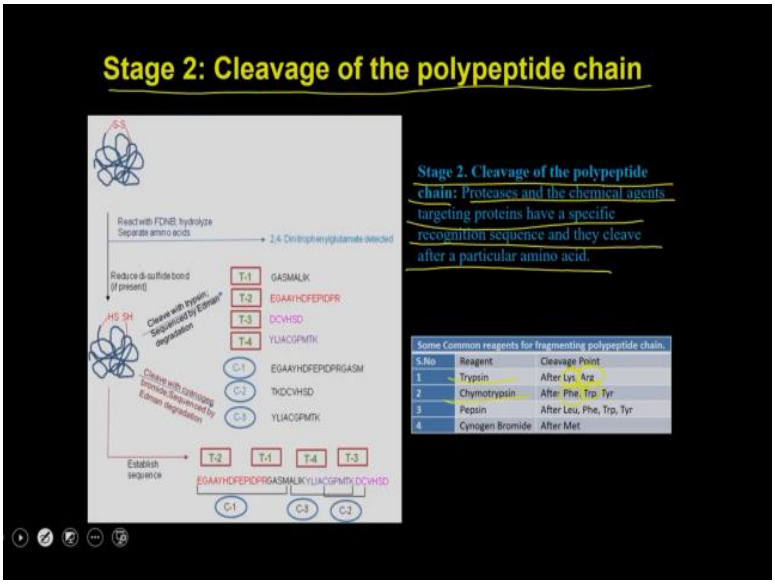
And it is not going to allow these peptide fragments to be separated. And that is actually going to create a lot of interference, because you will not be able to get the flanking sequences, you will not be able to deduce the sequence because now the sequence is going to be go through this disulfide linkages and actually, you do not have the sequence connected like this, you have a connection like this. So, because of that, you will not be able you it is actually going to interfere in the complete sequencing reactions.

(Refer Slide Time: 38:14)



In another approach the protein is reduced by the DTT that is the only way we have discussed so far.

(Refer Slide Time: 38:20)

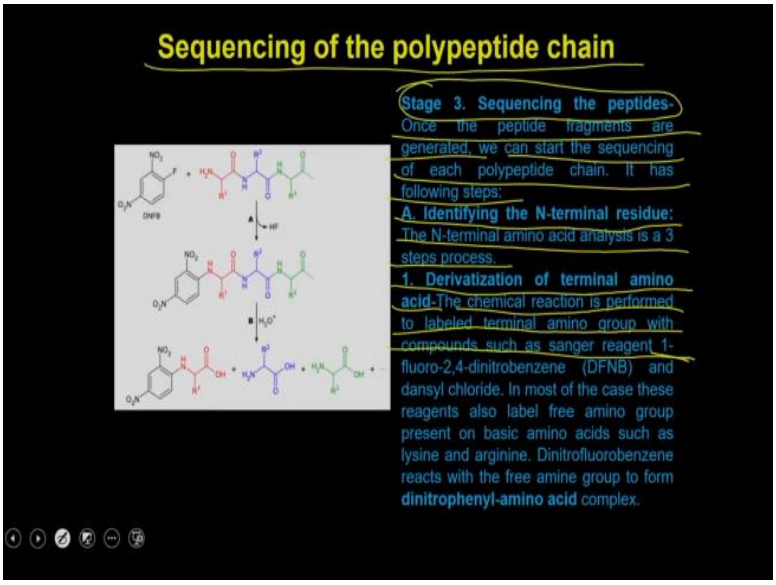


So in the step 2, you are going to do the cleavage of the peptide chain. So in the stage 2, the cleavage of the polypeptide chain protease and the chemical reagent targeting proteins have a specification sequence and they cleave after the particular amino acid, one of the classical example is that you can use the trypsin and that is actually going to cleave just after the lysine or the arginine. Similarly, you can use the chymotrypsin that is going to cleave after the aromatic amino acid like phenyl alanine, tryptophan or the tyrosine.

Similarly, you have the pepsin, which is actually going to cleave after the leucine, phenylalanine, tryptophan or tyrosine. Similarly, you can have the reagent like cyanogen bromide, which is actually going to cleave the methylene. So you actually know where these chemicals or the enzymes are claiming and that's all it is actually going to give you an indication that the last amino acid is going to be lysine or arginine or whatever.

So that is actually going to help you in you know, in making the sequences back and you can be able to use that information to put the sequence back and that is how you can be able to deduce the original sequence.

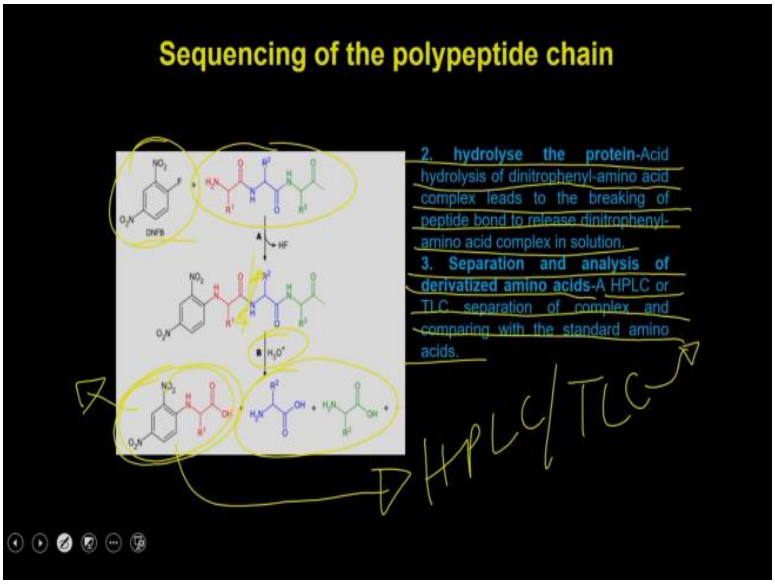
(Refer Slide Time: 39:31)



Then you have to do a sequencing of the polypeptide chain. So in the stage 3, you are going to do a sequencing of the polypeptide. So once the peptide fragments are generated, we can start the sequencing of the each polypeptide chain it has the following steps number A you identify the N terminal residue so the N terminal residue analysis is a 3 step process. Number 1 you do a derivatization of the N terminal amino acid.

The chemical reaction is performed to label the terminal amino acid with a compound which is called as the Sanger's reagent or the 1 fluoro 2, 4 dinitrobenzene DFNB and dansyl chloride. In most of the case, these reagents also label free amino acid present in the basic amino acid such as the lysine and arginine, then di tri fluoro benzene reacts with the free amino acid group to form the di tri phenyl amino acid complex. So, in the first event itself do are going to label the N terminal amino acid so that you can be able to tag that this is the N terminus of that particular peptide sequence.

(Refer Slide Time: 40:38)



Then in you the step 2, you are going to do the hydrolysis. So acid hydrolysis of the phenyl amino acid complex leads to the breaking of the peptide bond to release the complex which means first you are going to label the peptide with the help of the Sanger's agents, which is called as DFnB. And then you are actually going to do an acid hydrolysis. Once you do the acid hydrolysis is going to break the bond between the first residue as well as the rest of the peptide.

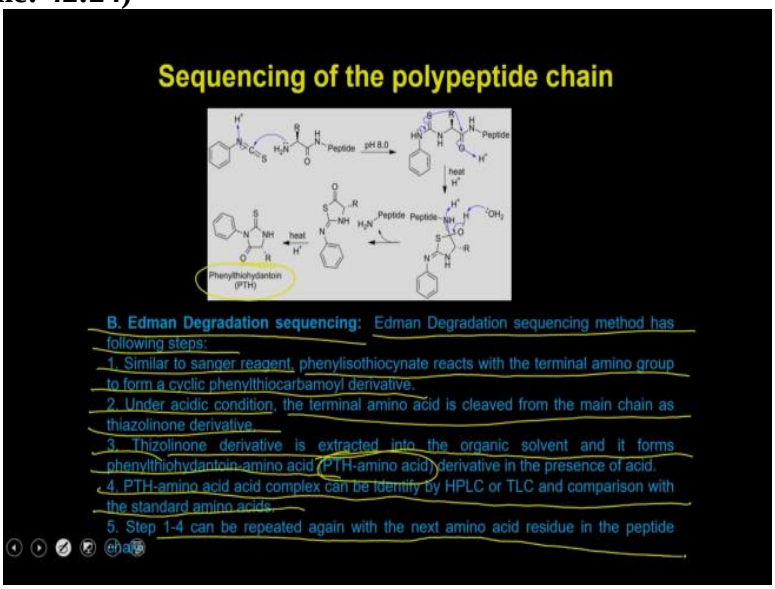
And it is actually going to generate the this complex, then what you are going to do, you are going to take this whole mixture and you are going to do a separation and analysis of the derivatized amino acid HPLC, or the TLC separation of the complex and comparison with the standard amino acid which means now what you're going to do is you are going to analyze this on to HPLC or the TLC and that is actually going to give you the spot for this particular amino acid.

What you can do also is you can also run the standard amino acids like you can do a standard amino acid which is coupled with the DFNB and which will going to tell you that at what spot is corresponding to that particular amino acid and that is how you can be able to identify the complex which is whether it is a making a complex with arginine or whether it is making a complex a lysine and so on.

So, once this is done, you can again do another reactions for the R 2 and R 3 and R 4 and all that and you can complete these reactions every time and every time it is going to react with

the new N terminus amino acid and that so you can be able to deduce the complete sequence of that particular peptide.

(Refer Slide Time: 42:24)



Now, the second method is called as the Edman degradation method. So, in the Edman degradation sequencing method has the following steps. Similar to the Sanger reagent, the phenylisothiocyanate reacts with the terminal amino group to form a cyclic phenolthiocarbamoyl derivative. So, and then under the acidic condition, the terminal amino acid is cleaved from the main chain as the thiazolinone derivatives number 3.

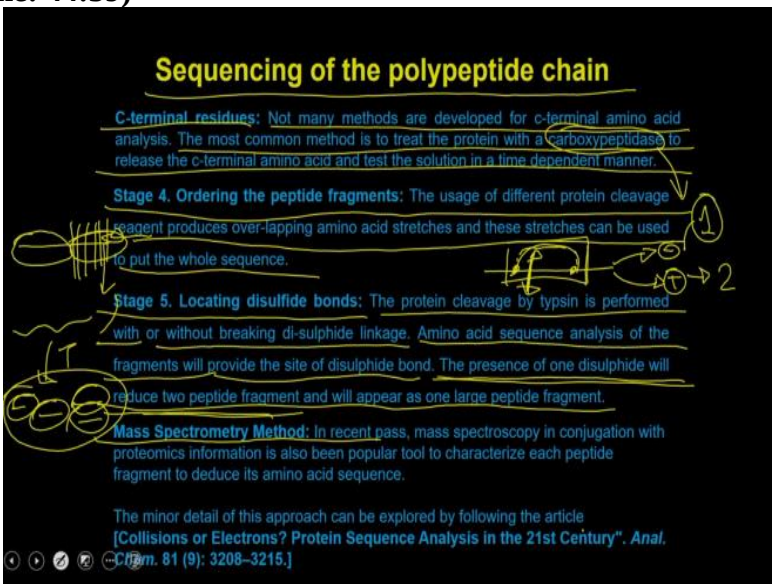
Then the thiazolinone derivative is extracted into the organic solvent and it forms the phenylthionate amino acid or the PTH amino acid complex. And the PTH amino acid complex is actually going to be separated and analyzed on to the HPLC or TLC, just like as we discussed before also. And that can be compared with the standard amino acid and that is how it is actually going to give you the presence of that terminal amino acids.

So the PTH amino acid complex can be identified by the HPLC or TLC and it is compared with the standard amino acid, the step 1 to 4 can be treated again with the next and terminal amino acid. So the method remains the same whether it is a Sanger's method or the Edman degradation method. The reagents are different, but the process remains the same, that you have a peptide sequence where you have 1, 2, 3, 4, 5 like that. So what you're going to do is first you are going to react with this amino acid it is actually going to form the complex.

And then you are going to cleave this off and that is actually going to give you the second third, 4th and 5th and then it is actually going to give you this modified amino acid plus this all these are again going to be present and then you can do a reaction with this one and it is continued like that. It is going to give you the sequence first the presence of this amino acid then it is going to give you the information about this amino acid and it is going to give you the information about the next.

So that is how you if you do a cyclic event of 1 to 4. It is actually keep giving you the presence of that particular amino acid starting from the N terminus, it is going to give you the full sequence of that particular protein.

(Refer Slide Time: 44:39)



Now, this is all about the N terminal sequencing for the C terminal sequencing, which is very, very difficult because, from the N terminus you have the reagents either the Sanger's reagent or the in the Edman degradation method, you have the reagents which are actually attacking onto the N terminus amino acid and that is how it is actually going to give you the complex and then you can be able to cleave but for the C terminal amino acids, because sometimes what happened is you can even do the you know analysis from the C terminus.

Because suppose, you have a very long protein like this. So, first some analysis you can do from the N terminus some analysis you can do from the C terminal and that is how, you can be able to do the complete sequencing of that particular protein. So, for the C terminal residues, not many methods are developed for the C terminal amino acid, what you have to

do is for the; you what you can do is you can just treat the protein with the enzyme which is called as the carboxy peptidase.

So, carboxy peptidase is a very, very specific protease, which actually cleaves the protein from the C terminus, which means they are going to chew from this side. So, if you treat the protein with the carboxy peptidase, it is going to start chewing from the from the backside of the protein which means it is actually going to start giving you the C terminal amino acid available and then what you can do is you can simply do the Sanger's method or the Edman degradation method.

And that is how it is actually going to give you the sequence of this C terminal fragment and that is how you can be able to deduce the sequence from the C terminal residue as well. Now, once you have done you know, appearance of these methods, then the stage 4 is that you have to order the peptide fragment, the usage of the different protein cleavage reagent produces overlapping amino acid stretches, and these stretches can be used to put the whole sequence back, which means, once you got the peptide sequence of a particular protein.

Then what you can do is you take these all the peptide fragments and you can just put them so, what will happen is you are going to have some of the portion which is overlaying you know, so, that overlapping portion can be used to deduce the final sequence then the stage 5, you also have to locate the disulfide bonds protein cleaved by the trypsin is performed with or without breaking the disulfide linkages.

The amino acid sequence analysis of the fragments will provide the site of the disulphide bond the presence of 1 disulphide will reduce to peptide fragment and appear as a 1 large peptide which means, what is mean by is that suppose you are expecting that these are the 2 places where you are going to have the disulphide linkages which means it may be connected like a disulphide linkage what you can do is you can just do this analysis in the absence or the presence of trypsin.

So, what will happen is that if there will be a disulphide linkages will present. If disulphide linkages are present, then in the presence of trypsin you are going to get the 2 fragment because it is going to be cleaved like this. So, you are going to get the 2 fragment if it is the disulphide linkages are not present, but if the disulphide linkages are present, then the 2

peptide fragments are going to be joined by these disulphiide linkages, which means it is going to be joined like this.

So, even if it cutting here, it is only going to give you the 1 peptide fragment because still the trypsin has cut this peptide fragment, but the fragments are not going to be separated because they are joined by the disulfide linkages and that is how it is going to give you the one peptide and that is going to give you the information that the disulphide linkage is present between these 2 peptide sequences.

Apart from that, you can also use the mass spectrometry method. So, in the recent past, people are not going with the extensive reactions and then you are isolating the amino acid complexes and then you are doing a separation by the HPLC or TLC and all that instead of that what people are doing is they are just taking a protein, you they are treating it with the trypsin.

So, they are getting the small, small, small fragments and then what they are doing is they are simply taking these fragments and do a mass spectrometry because once you done the mass spectrometry it is going to give you the masses of these peptide sequences and then utilizing these masses you can be able to explore the probable sequences which can fit to this particular mass and that is a proteomics approach people are more often using to deduce the sequence of this original protein sequence.

And that is a more and more complicated it is having more and more details. So, I think I have done I have said in the past also if you are interested to study these processes, there are classical and the one of the several you know the courses are available on the IIT Bombay smocks courses as well that you can easily take to understand how this process, but this is a more and more advanced technique where you are not going through with the extensive chemical reactions and analyzing the each and every amino acid.

Instead, you are just simply doing a mass spectrometry, you are getting the masses of these peptides. And then since the database is increasing day by day, you will get a particular amino acid sequence which is definitely going to match with this particular mass. And that is how you can be able to deduce that particular sequence some of the details you can easily get from this particular articles.

So, I will strongly suggest that you should read all these articles. So, that actually is going to tell you the whole the possible methods what has been available to sequence the protein sequences. So, this is all about the sequencing techniques for sequencing the DNA or the proteins what we have discussed we have discussed about the DNA sequencing method whether it is the Sanger's DNA sequencing method and as well as the protein sequencing method. So, with this I would like to conclude my lecture here. Thank you.