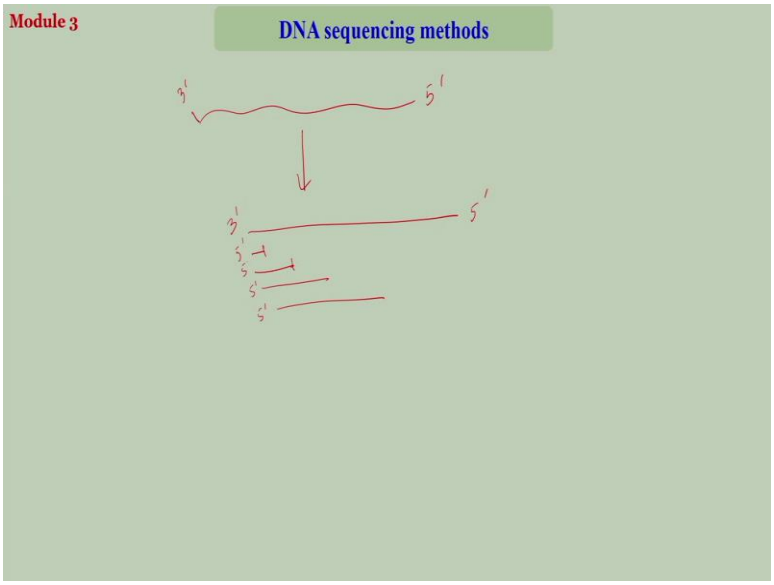
Essentials of Biomolecules:
Nuclic Acids, Peptides and Carbohydrates
Prof. Dr. Lal Mohan Kundu
Department of Chemistry
Indian Institute of Technology-Guwahati

Lecture No. 14
DNA Sequencing Maxam - Gilbert method

Hello everybody and welcome back to the lecture. So, we have been discussing about the DNA sequencing methods and we have talked about Sanger's method, Sanger's di-deoxy method or also known as chain termination method to sequence or whole length of DNA and we have seen how does it work? It basically uses DNA polymerase and it basically uses the same principle of DNA replication to construct the complimentary strand of the target single stranded DNA. And then we have seen that how the fragments can be obtained of different chain length.

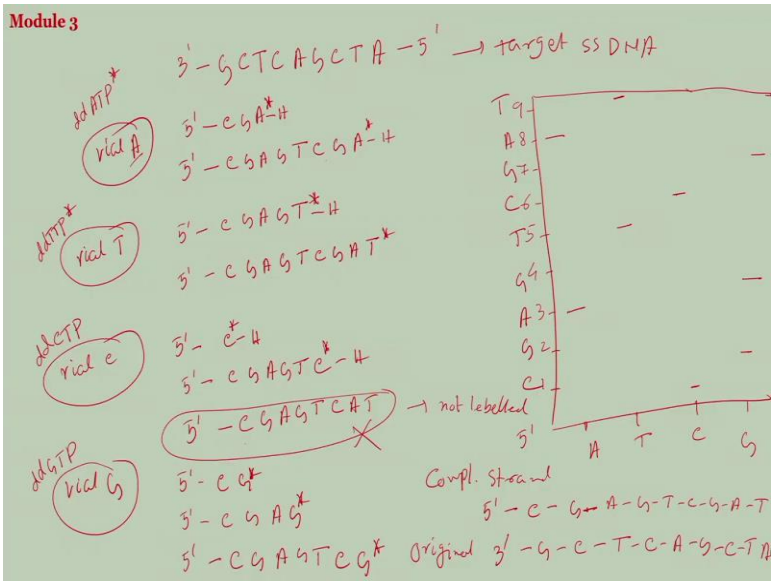
(Refer Slide Time: 01:25)



So, that basically if you have the DNA, single stranded DNA, 5 prime to 3 prime and this is phosphate and we have used certain ingredient which we call the di-deoxy DTNTP. And then you get the idea is to stop the replication to stop the constructions of the parents strand at every base period. So, your chain will be terminated the synthesis of the complimentary strand, this should be 3 prime this should be 5 prime.

So your synthesis of the complimentary strand would be stopped at every base and that is how you can get the fragments of all length of DNA and you can analyze the fragments using gel electrophoresis that is what we have seen in the last class. So today, we will take another example, shorter one, and we will try to see, we will see what the fragments are.

(Refer Slide Time: 02:33)



So I am taking the DNA from the 3 prime to the 5 prime let us make an arbitrary the GC TCAGCTA that 123456789 the very shorter 1, 2, just to show another example, how the fragments are coming up. So, obviously, so, this is your target single stranded DNA for which you want to know the sequence. Now, obviously, such short oligonucleotides or such short DNA cannot be sequenced using the Sanger method.

They are usually longer target DNA that we use for the sequencing. It was in the Sangers method, minimum they prefer that it has to be longer than 50 nucleobase long or 50 base pairs long should be the target DNA then only you can proceed for the sequencing. The reason is you have to use a primer and the primer itself is 15 to 20 miles long. So, your target sequence has to be a little longer than that.

So, I am not drawing the primer part, I will start straight from here. So first, what you have, vial let us take vial A contents dd ATP along with all the 4 other dn TPS plus the polymerase. So now and the dd ATP is radioactive labeled. So what are the fragments that you are going to get in

a vial A 5 prime? So it will stop wherever there is a possibility of picking up it. So this would be c, this would be G. Here is the possibility of picking up A and it will stop there because it does not have the 3 prime hydroxyl group.

And it will show you the radioactive level. That is the first fragment second is CGAGTCGA. It will stop here to fragments. Second is vial let say T and that has dd TTP with the label one. What are the fragments that you are going to get here? 5 prime. See CGAGT so, here you have the possibility of picking up the timing and that timing can be the dd TTP. And here it will stop the chain. If there is anything else L, the last one CGAGTCGA and the last one is T.

So, this will also show you the radioactive level and third one is vial C, where you have added dd CTP what are the fragments here the C would be picked up and therefore it will be stopped 5 prime CGA, this is your G then comes T and here is a C it will stop there equal right down along with me also CGAGTCAT but this will not have leveled. So, this will be the full length DNA not labeled. So, this will not be seen in your gel.

And vial G that has dd GTP fragment are CG this CGAG another here, still nothing else no other possibility wanted to see 1C, 2C 3C shows there is 3g, AG show CG, 123c show there should be 3g. First one is CG 1234 position, and then their way another CGAGTC sorry by mistake. This is C, this is G, so suppose this is A, then this is G then it is T, G opposite to C then should be G here so, I need to recap a little bit CG fine that will stop here.

Wherever there is a so, opposite T 12345678 here vial T would be 1 and 2, 2 T is 1 and their 2 T see, now, see there are how many G is, 1 and 2 G. So, there will be 1 the first 123456 position that is fine and here there would be this one is in the second position and other is the fourth position and another would be 1234567 position now you have the fragments. Now if you write one to draw the gel let us say this is your vial A this is your vial T.

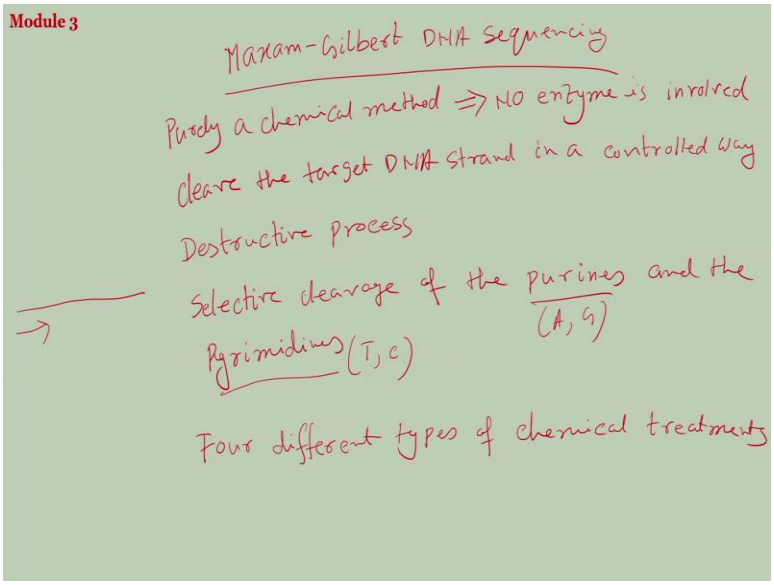
This is your vial C and here this is your G. Now let us count the fragment we have got 123456789, 12345678 and 9. So first length is here in the vial C, vial A 312345678, 2 spots there, vial T 12345, and then 123456789 that last one. Then we will see the first 123456 should

have won at 6, and then the vial G 12, 1234, 1234567. Now at the length of 7 now do we have all the length this is first, second, third, fourth, fifth, sixth, seven, eight and nine.

Now your write the compliment strand of which you are synthesizing from the 5 prime and so, this would be your C, next here will be your G, third one is A, fourth one is G, fifth is T, sixth is C, seventh is G, eighth is A and ninth is T. So your complimentary strand sequence is 5 prime CGAGTCGAT. Therefore, the originals strands the target should be your 3 prime to GCTCAGCTA, so it matches with the original strength.

So, that is how obviously, you can write the fragments and make the gel and then read the gel accordingly. So, now we have learned how to analyze the gel electrophoresis or the auto radiogram and get the sequence of your target DNA using the Sanger's method. If we just gave you the diagram of the gel, you can find out what is the originality DNA and of course the vice versa if I give you the original DNA, then you can find the gel out also so now, let us move on to the other sequencing method that is that Maxam Gilbert DNA sequencing.

(Refer Slide Time: 14:16)



Maxam Gilbert DNA sequencing method this is purely a chemical method which means we do not use any enzyme, no enzyme is involved. It is only chemistry. And so how is it done? It is done. The idea is to cleave the DNA strand in its backbone. So, you cleave the target DNA

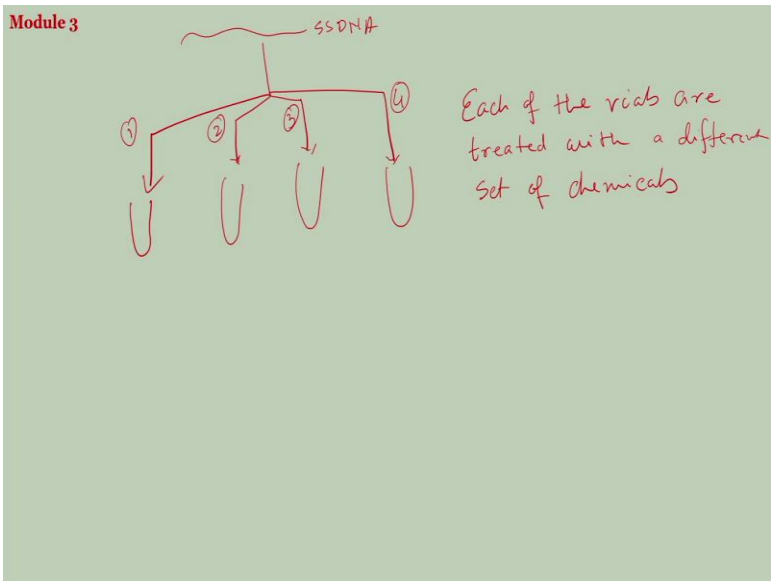
strand in a controlled way. So, if you have the target DNA using chemical treatment, you actually break the strand, obviously in a controlled fashion.

So that you can have restrictions to certain quantities that some groups should react under that condition some other nucleobases will not react under that condition will see shortly. So, since you believe that DNA target obviously is a destructive process during Sanger's method, you have seen that you had your target DNA. What you have been doing you have been trying to synthesize the complementary strand, which means your target DNA was always untouched and it could be reused again and again in the cycles.

So, it is a non-destructive process. On the other hand, Maxam Gilbert method is a destructive process. You cannot recover your sample because your sample is getting clipped or get destroyed under the treatment of the, by the chemicals. It is a destructive process and selective cleavage of the purines and the pyrimidines we will see, there will be some chemical treatment done, which will only react with the purines nuclear basis, like adenine and guanine.

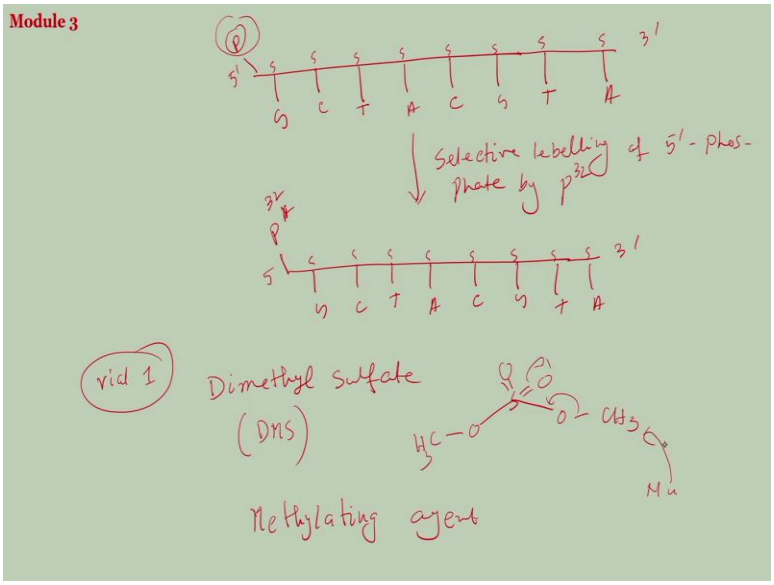
And other chemical treatment will act only on the pyrimidines which means thymine and cytosine so there will be total 4 different kinds of reactions, first one is 2 set for the purines and other 2 set of reactions for the pyrimidines. 4 different types of chemical treatments that is what would be done in this method.

(Refer Slide Time: 18:39)



So let us take a target DNA I will write the sequence later. So this is your again we use the single stranded DNA and this is again divided into 4 vials, vial 123 and 4. So, basically all the vial content, a little bit of your targeted DNA. Now each of the vials are subjected to different chemical treatments are treated with different set of chemicals. So let us take 1 vial at a time. And we will see, what are the chemical treatments that are done and what is the outcome of it?

(Refer Slide Time: 20:16)



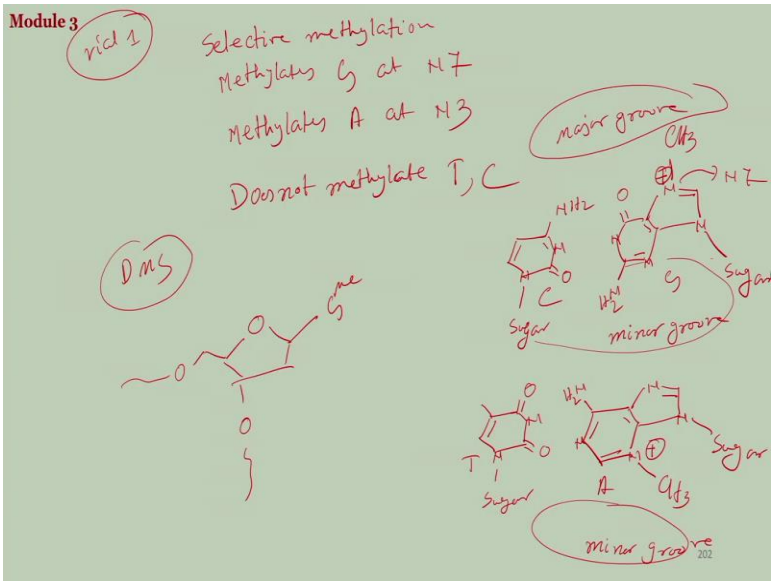
How does it cleave the DNA in a specific manner so vial 1 so before that there is another step that we have to do to start the reactions? So let us take a sequence. Now the P and this is the 5 prime end I am writing now the sugar S for sugar. Your skeleton has a sugar CGCTACGTA, 3

prime it so let us take that this is your target DNA, the sugar, that base sugar base. And again now the same thing as hangers that you need to track your molecule.

Whatever the fragmentations you are doing whatever the chemical treatments are doing, It is going to ale some products and you need to identify the products or you need to track the product, where are they located. So we also will need a label in here as well. And that is the first step that is being done. That selective labelling of these 5 prime phosphate by P 32. This is usually done by enzyme. So, what do you get your actual target P 32, 5 prime GCTACGTA?

So now you can track wherever it breaks, no vial 1. So first treatment is done with our the chemical reagent that is known as dimethyl sulfate I will show you the structure this is Sulphur, O CH3 this is called dimethyl sulfate I am giving an abbreviated name DMS. So, this compound is usually used as a methylating agent it means if you want to methylate something then this is used so that 1 methyl from here would be transferred from here to your compound. The reason is if you have a suitable nucleophile, so, because of this internal regiments, this is very label. So you can easily pick it up so methylation can be done.

(Refer Slide Time: 24:21)



And this methylation is required selective. It does not methylate all the nuclear bases. So primarily the nucleus bases would be methylated by the DMS but it is a selective methylation. It does not matter it all an equal bases it methylated guanine at N7 position will see the structure

and it methylated adenine at N3 position. So at the centers are different. Do not methylate or it does not methylate type thymine and cytosine.

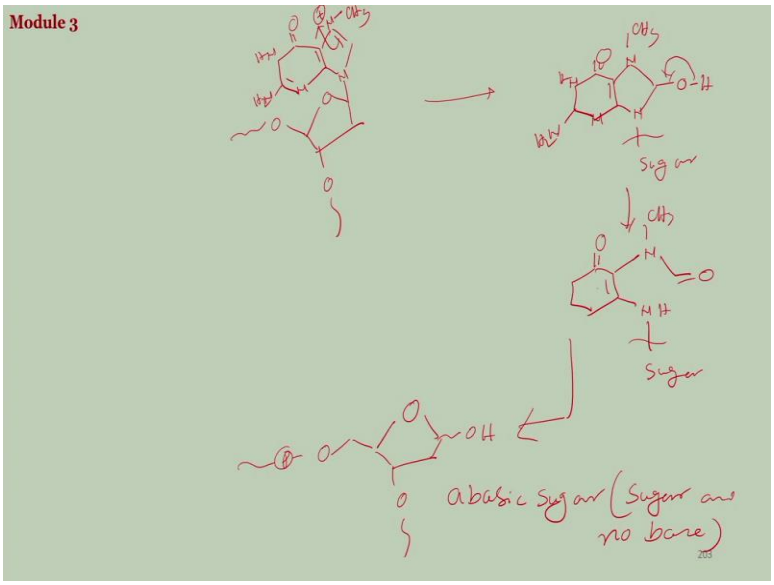
So it is a selective methylation only for the pure in nuclear basis. Now if methylate what happens? So if you look at the DNA helix, there is major group and minor group. And if you look at the base pyrimidines look at the CG base pyrimidines first. Here is your sugar and carbonyl this is the fiber memory here is your sugar NH2 correct here is a sugar and this is your in N7. So this nitrogen would be methylated by the methyl and this will be plus.

Now if you look at the helical form, so the interior part of the sugar this is the sugar moiety. This is the interior part of the helix. And here this part actually constitutes a minor group of the DNA. And this part the upper part is exposed to the major group of the DNA. So, which essentially means that you are methylated version methylated warning this is OG and this is your C is exposed towards the major group of the DNA.

Now let us look at the AT sugar here would be in this is ray and this is a T is this is your sugar here is the double bond. So, this is your A, this is your T and for adenine, this is the N3 position so here it is methylated. And now again this is the interior part this forms the minor group the upper one is the major group. So you see here the methylated version of the timing is exposed to the minor group of the DNA that actually makes quite a difference.

Because, so once you treat this with dimethyl sulfate with DMS then you get these things with the sugar now followed by treatment with hot piperidine. So if you write down this is the 3 prime app to go to the next page I think and here, your G with the methylation portion will be not start I will see that actually what happens is when you treat it with dimethyl sulfoxide and then followed by treatment with piperidine, then only the G which is the that would be clipped off the sugar. Adenine will remain intact because the methylene happens in the minor group kind of.

(Refer Slide Time: 30:52)

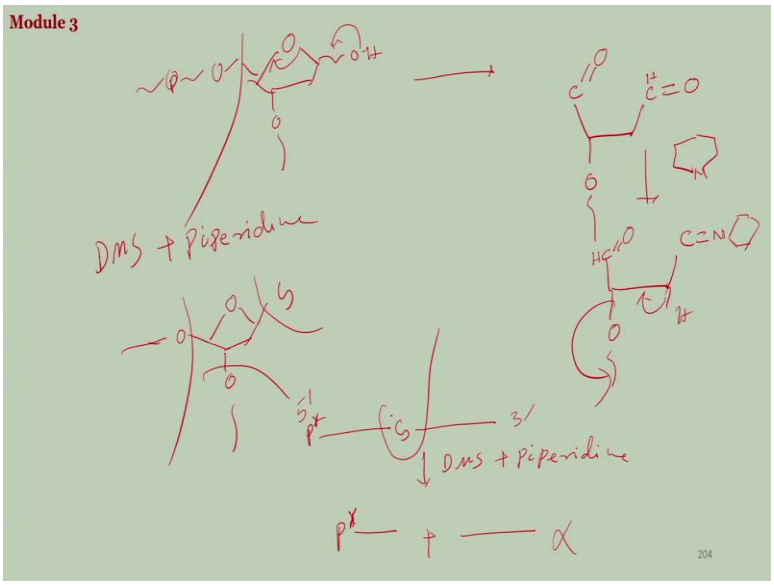


So if write down the actual structure, here is super I have drawn in opposite direction that is not correct. It has to be the other way around. So this is DNA. And here is your N here is the methylated version. So it is plus and in here you will have I mean in each CO, this is a warning. Now what happens is I will shortly show you briefly, I will not go into the details. So once these have the methylation, this will be picked up.

So that creates a positive charge here and it will be attacked by a water molecule. And what you are going to get there will be water in CH3 here is your sugar. And here you have this now, this form ketone and it actually opens the ring up, this breaks this up and you gate then ultimately chops it up, it will open up as let us write in CH3 which would be NH and it would be sugar and here it would be CO. So your nucleobase is clipped and now in another condition of piperidine in treatment, this will be clipped up water will come here.

So, basically what you get is these temporarily your phosphate and the DNA. So you get the sugar without the nucleobase and this is called a basic sugar a basic means sugar and no base. Sugar without the base is called a basic site.

(Refer Slide Time: 33:45)

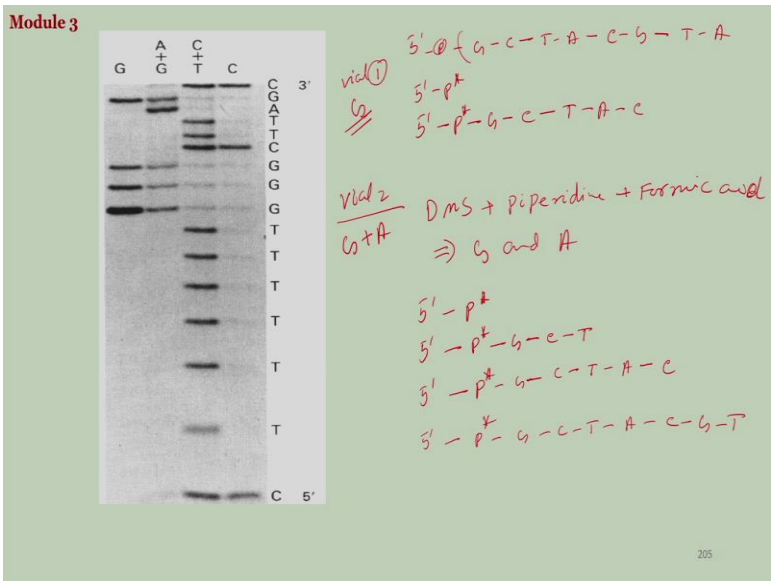


Once you get a basic site then in piperidine what happens it gets oxidized and that will basically open up it will make the cleavage of the 5 prime and so different people show different kinds of mechanism. So it is something like this so you have OH here that has clipped out and it will have the ketone the aldehyde here the aldehyde reacts with piperidine which is the structure of piperidine and you get this will be ultimately be chopped off.

So effectively, you get these it is not a wage this is and there is see it may be in terms of O, or it may be in terms of that other different thing. And this prompts cleavage of this also so, ideally what you get is by the treatment of DMS + Piperidine you get a cleavage. If you have the G, then you get the cleavage here that prompts you to get cleavage here and at cleavage there. So both the 5 prime ends of the DNA as well as the 3 prime end of the DNA is clipped off.

What does that mean? So if you have a DNA, G and if you treat it with the DMS + Piperidine what essentially does it will cleave these off you will get a fragment this plus our fragment this now, we have started with the P 32 leveling here so, therefore, this fragment will have the level 1 this fragment does not have the level and we will not be able to visualize it. So, we do not consider this trend, we will only consider the fragments which will have the label phosphate. And we will see in the all the subsequent steps it basically does the same thing. It chops of the nucleotide from the 5 primate as well as the 3 primate.

(Refer Slide Time: 36:54)



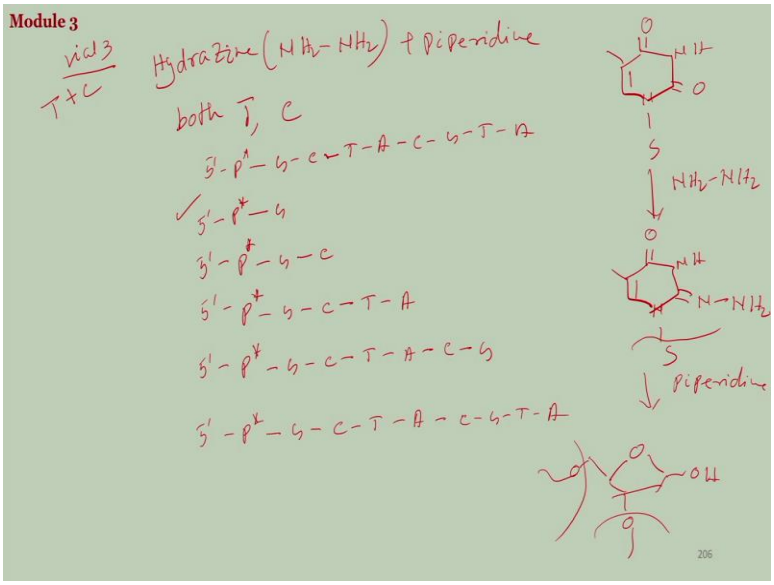
So, that is the first treatment in forget about this so, we had the sequence please do not consider this one for the timing which was 5 prime phosphate. And we had the sequence which was GI now I do not have to write this because we have already seen how this is clipped off phosphate then you have GCTACGTA now, so vial 1 what kind of vial 1 we now call it a G. Because vial 1 will cleave only the G, not the A. So, fragments first one would be G would be chopped off here, it is gone so, you are left to it to only the phosphate.

Next P I have to write the phosphate, but all the time because this is one band that will that you are also getting phosphate, GCTAC because G would be chopped off here, and that is all. So, this is would be the fragments that will get in vial 1 now vial 2 the treatment is same dimethyl sulfate plus piperidine so it will basically methylate going in and adding in. Now if you do this treatment in presence of formic acid so, that brings the pH low.

And in formic acid, your Glycosidic bond like cosidic bond means this bond basically this bond is called like a Glycosidic bond. So, this bond after the methylation gets very weaker and therefore, this cleaves both G and A. So, wherever you have the methylation wherever you have a G wherever you have A both will be cleaved off under this condition. So, vial 2 we call it now G + A, it will clip G it will clip A both so, what are the fragments you will get? You will get this all fragments that you get from G plus the A.

Second is PGCT here it will stop it t said anything else PGCTAC, P would be chopped off GCTACGT then A would be cleaved. So we will get these fragments from the vial 2, which basically destroys both G and A. So this is also a kind of double check method, where you can get 2 bands and can verify with the other. So that is vial 1 and 2 now purines are done.

(Refer Slide Time: 41:09)

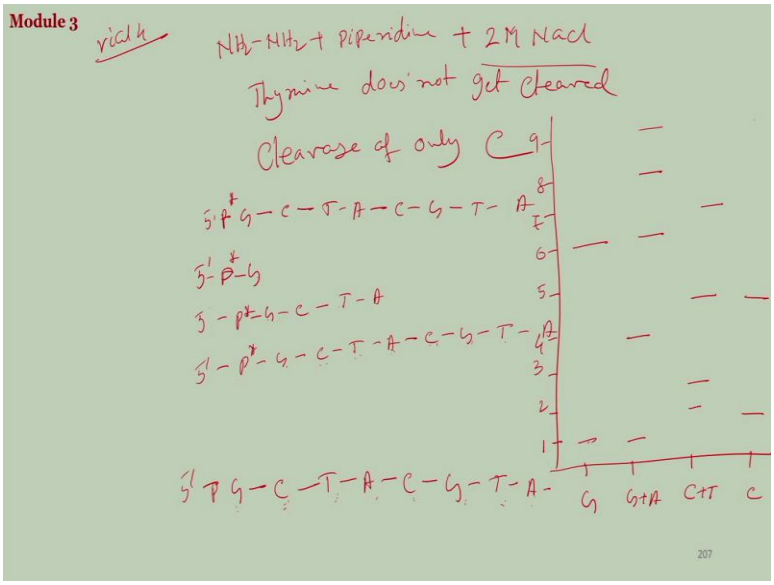


Now vial 3 will treat it not with the DMS or not with people who did not will be your target DNA with hydrazine. Hydrazine is NH2 + piperidine. So what this will do is hydrazine will react with the pyrimidine nucleobases. So it reacts with both thymine and cytosine. I will just with 1 example I will show if it is thymine and this is sugar this NH this is your thymine it is treated with hydrazine. So if you treat hydrazine with a carbonyl compound it will give you hydrayl zones, there will be condensation reactions here, I will take the example of this. So here it will form this kind of condensed product.

Now once this is formed, this undergoes intramolecular rearrangement and basically breaks this up. So under piperidine in treatment, you basically get again, the same a basic sugar followed by again the breakage of the T primate, 5 primate so, this treatment hydrazine with piperidine cleaves both thymine and cytosine. So, if your sequence I have to read again GCTACGTA that is the prime sequence. So, the fragments in vial 3 would be here early be wherever there is T and C G to be clipped here and it will stop,

It will have that this is one GC it will cleave the T also done GCTA another series coming it will stop here GCTACG, T is coming next. So it will stop there and that basically it will not have any other the last one the full length GCTACGTA and the full length. So, these are the fragments that you are going to get from vial 3 so vial 3 now call it T + C, or C + D.

(Refer Slide Time: 44:54)



And next, now, the third last one vial 4 is treated with the same hydrazine + piperidine in presence of 2 molar sodium chloride. Now, if you do these in presence of sodium chloride then what happens? This does not your thymine does not react here under the presence of sodium chloride I mean does not get cleaved so all you see the cleavage of only cytosine. So, wherever there is hydrazine it will be cleaved not the thymine.

So, again I have to write the target GATACGTA, what are the fragments, if it clips only the CG P it starts with P. Here is it phosphate is important actually. Because this we have to count GCTA then GCTACGTA the full length. So again, if you compare with vial 3, you will get all the fragments that are there in vial 3. And vial 3 will content some more fragments for thymine, now you are basically done. So, you again have like the Sanger's method, you have the fragmented DNA with all the lens.

Now if you try to construct the, do I have space, I have to do it here. The gel your gel will be so G next one is G + A next one would be your C + T and the last one would be, your C, and here at

the length, starting with the first phosphate, so we will count the phosphate also 123456789 now you can actually do it the first length with a single phosphate that we have got single phosphate here that was obtained from vial G.

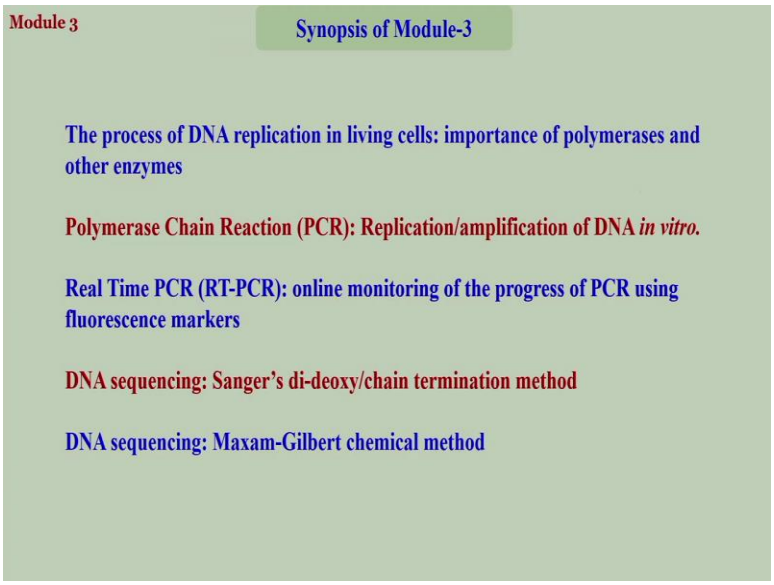
So G vial first vial G well contents the first one, and 12345, 123456, 1 and 6; 1 and 6, 2 bands here. Now G + A. 11234, 14123456, 14612345678, 1468. So you see, you can match G and G + A or whether your gel has or the fragmentations has occurred or not. It is a double check up actually. So vial 3 is C + T, what you have is 12, 123, 23, 2 you have 3312345, 1234567 and 9. So 5, 7 and 9, I guess 5, 7, and 9.

And lastly, the C has 123 is here 2, 12345, 5123456789 and the last one is 9 and you can read your sequence now so, this would be 5 prime to write, this is G, then G means G+ A also you can match so, this would be your G, next is second length is C + T. So, and it is C is there, it is C, third length is C + T, but it does not show up in C. So, this has to be T fourth is G + A, but there is nothing in G. So, this is A fifth is C + T and it is presenting C so, this has to be C 12345.

So, with the phosphate, so, this is 6, next is this is G, G and G + A. So, this has to be a G. Next is this C + T, but there is nothing in C. So, it has to be a T number 8 is G + A, where there is nothing in G. So, this has to be an A and the last one is C + T and C. So, it is CG, G + A fine then C + T then G + A so far is good CG is also correct. CG is 1234567 is this 7 when is C + T so, this is A this is T and the last one this has this has been a mistake, I think this has been a mistake or it has to come here.

So, that would finish it 1234 GCTACGTA that will be okay. So from the gel, you can read the strand and that will be your original strand. So, these are the fragments. It is a little bit complicated process, but very fine chemistry actually, and very controlled chemistry that are quite selective to shorten nucleobases and will not disrupt the other nucleobases. So primarily, this is a destruction of the DNA and destruction of the main DNA strand.

(Refer Slide Time: 53:10)



So with this, I will complete the module 3. So, what we have covered is that we have seen the process of DNA replication in the living cells, how the polymerases and how other enzymes functions also or take part in the process of DNA replication. Second, we have seen the polymerase chain reaction PCR, how you can do the replication same replication process or you can amplify the quantity of your DNA in a laboratory in the chemical technique that is called the PCR in vitro.

So, third and then real time PCR, how can you monitor the PCR that is going on? How can you monitor whether you are getting your correct DNA amplified or not using the fluorescence tax in the fluorescence markers and then DNA sequencing we have covered the Sanger's method which is di-deoxy method or also known as chain termination method. And finally today we have covered the Maxam Gilbert method for the DNA sequencing. Thank you.