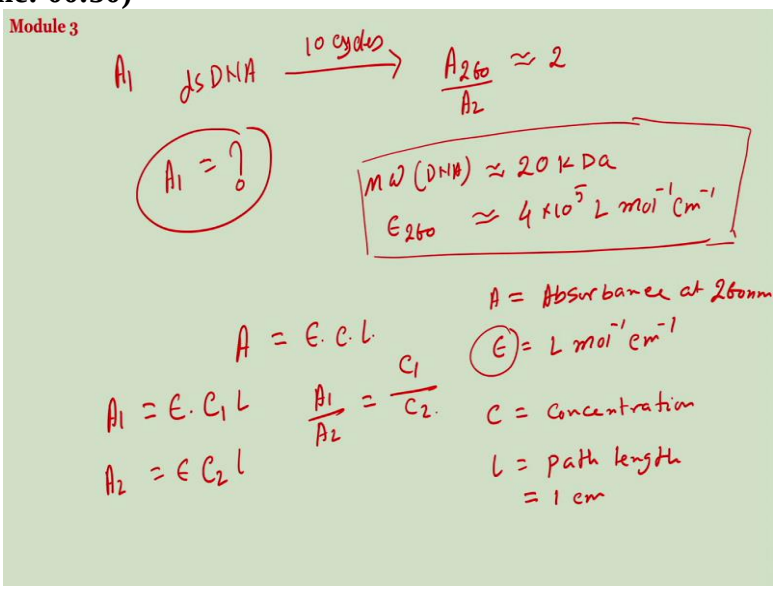
Essentials of Biomolecules : Nucleic Acids, Peptides, and Carbohydrates

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

Lecture - 13
DNA Sequencing: Sanger's di-deoxy method

Hello, everybody, and welcome back. So in the last lecture, I had given a problem, numerical problem. And today I will quickly show you the answer. It is very easy actually. The problem was that if you do a PCR, let us say you have a double stranded DNA.

(Refer Slide Time: 00:50)



For which you are doing the PCR and after about 10 cycles then you get a certain quantity of the DNA and the measured optical density or the measured UV absorption at 260 nanometer was coming as approximately 2. That was the experimental quantity that you have opted. After the cycle, the question is if at the beginning if you have the observance we call A 1. And if this is now A 2 in the question is what was A 1.

That was the problem and I had given a I think few data also that the molecular weight or the molar mass of the DNA given DNA was around, let us say 20 kilo Dalton. And the calculated epsilon value at 260 nanometre was approximately this is the molar extinction coefficient 4 into 10 to the power 5 litre mole inverse centimetre inverse. So, what is A1 actually?

So, you must have already done this this is very easy and this data I guess, you probably would not require them at all. They were meant to confuse you. So, if you want to calculate what was the basic equation, basic equation is the Beer Lambert law? Absorption is equal to

epsilon into C into L. This A is as you know A is the absorbance at a certain wavelength in this case, this is 260 nanometre.

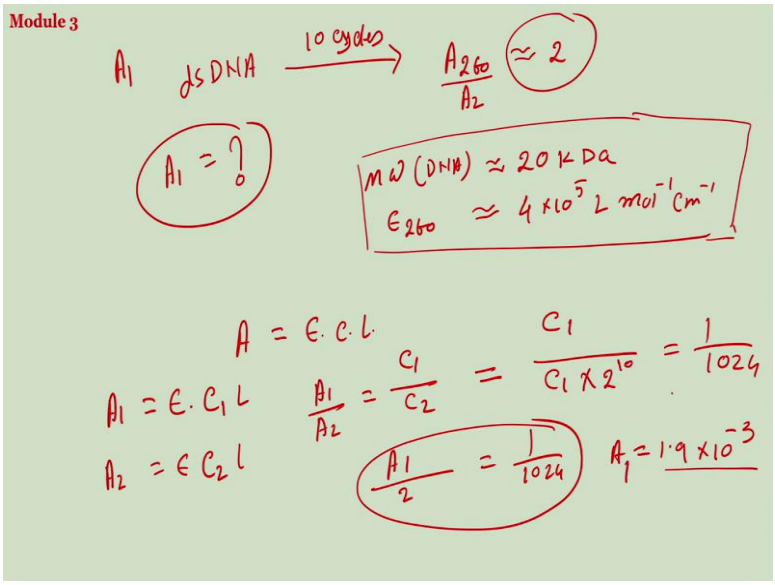
Because we are dealing with DNA and DNA absorbs maximum at 260 nanometre and this is unit less there is no dimension for absorbance. Epsilon is the molar extinction coefficient. And it has the stranded dimension of litre mole inverse centimetre inverse. And this is a characteristic of a molecule. So you can calculate epsilon for depending upon the DNA sequence.

But it is kind of characteristics are so it does not change over dilution, or it does not change over. If you change the concentration of your sample in epsilon the deviants intact, so it is a pretty much constant quantity and kind of a characteristic of the sample. C is the concentration of your DNA in this case. L is the path length. Path length means the length through which the light travels.

What is the distance that the incoming light travels through your sample? So ideally, the path length is actually the length of your qubit. If you are using we of course has to we have to use the qubit to measure the UV. So it is basically the dimension of the qubit. And, the most stranded dimension of the qubit that we use is diameter of the qubit that we use is 1 centimetre.

So we have to measure basically, the concentration first, or our target is to measure the absorbance at the beginning. So if you start A1 is equal to epsilon, at the beginning, let us say the concentration was C1 and L, you are using the same qubit. So essentially L is same for both cases. A2 is epsilon does not change because the same DNA. Let us say the concentration has been changed to C2 and L. Now, A1 by A2 basically comes down to C1 by C2.

(Refer Slide Time: 05:44)



Now so, other parameters are not required in this case. So, what is A 2 so, C 1 and C 2 is there any relation. So C 1 is the concentration at the beginning. What is C 2? C 2 has to be C 1 into 2 to the power 10. This would be your C 2 or 1 by 1024. You have to calculate your A 1, your A 2 is given as 2. This is your A 2 1024. So, your calculated A will be around 1.9 into 10 to the power - 3.

So, A1 is 2 by 1024. I hope I have done the calculation correct, but this is it from here you can get your A1. So now, let us move our attention to another important discovery and then important aspect that is called the DNA sequencing highly important and highly fascinating subject is how to find out the sequence of a given DNA. So, so far we have seen the structure of a DNA.

We know that it is forms a double helix, we have seen what is the composition of the DNA that it has 4 nuclear bases, and it has a phosphate unit, it has a sugar unit. And we have also come to know that the DNA is the most important material in the cell because it dictates most of the other cellular functions. For example, it dictates the synthesis of protein and it passes on the genetic information to the offspring.

So, now, the question is, what is the sequence of a given DNA if you have isolated a DNA from a source, then what nuclear bases are coming after one another. That is of utmost important. Because, if you consider a molecule, if I give you a molecule, maybe an organic molecule or an inorganic molecule, and I ask you to find out more about the molecule. So, the first thing you would like to know is what the molecule looks like.

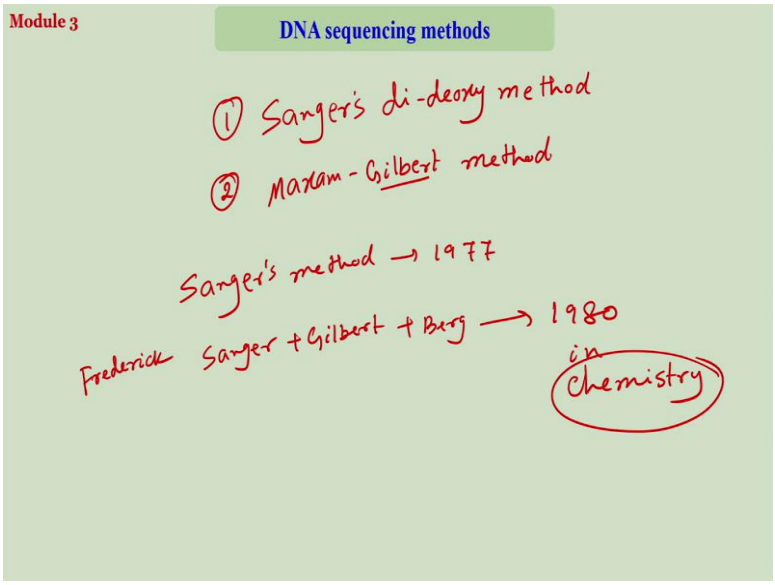
What is the structure of the molecule? So, once you find out the structure of the molecule, once you find out how the molecule looks like, what are the atoms involved how the atoms are attached together? What are the functional groups that are present in the molecule? Then only, you can think of the property of the molecule you can get some idea what this molecule is going to do, what can be the characteristics of the molecule same for DNA.

So, once you understand the whole sequence of a gene, how the nucleus are attached one by one, then you can figure out or you can have some certain ideas about what should be the characteristics or what should be the property of the given DNA. So, and it will remember that I have talked about hair samples or the samples DNA is routed from the diseased cells that are of low quantity.

So, that one main purpose of studying those genomes is of course, to see they are sequence once you know the sequence of a DNA, then you can offer diseased DNA then you can understand whether there is any mutation or not. Or even you can use the DNA to match one identity to measure the; or to find out the identity of the persons or identity of the species also. So Human Genome Project was a very big project that actually has done this entirely.

It was entirely about studying the whole sequences that are present in human genome. What are the variations that are present from one community to the other. And it was one of the biggest project that actually was counted out all almost all over the world. So it is of that important. So, today, we will see the methods through which you can sequence a DNA. Primarily, there are 2 most popular methods that used for sequencing a DNA.

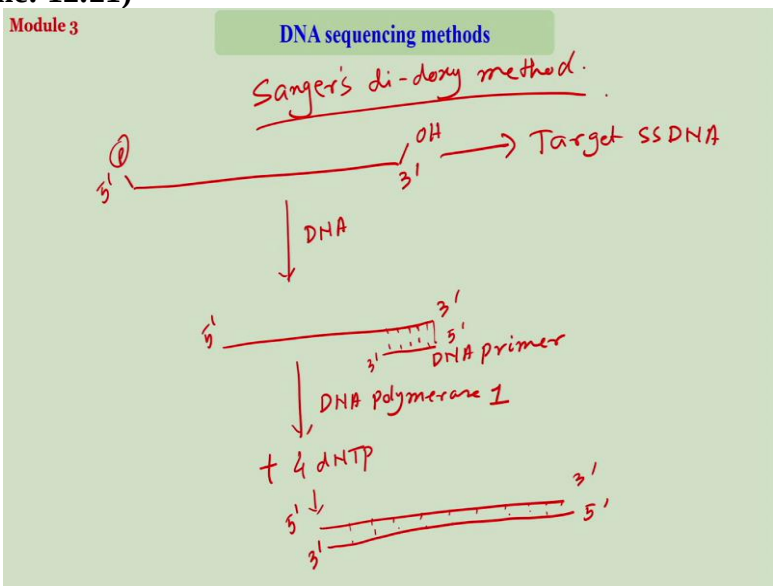
(Refer Slide Time: 10:21)



One is the Sangers method. Sangers, this is called di-deoxy method. Second one is Maxam - Gilbert method. So, these 2 methods will discuss. The first is Sangers di-deoxy method. So, by the way both of these methods have been developed much before the PCR technique was evolved or developed. For example, Sangers method was developed in 1977 and both Sanger is Frederick Sanger actually.

Frederick Sanger received the Nobel Prize. Sanger actually has received his second Nobel Prize along with Gilbert. And I think Berg also in 1980. And, again, most importantly, he received the Nobel Prize in Chemistry. So you can see the enormous contribution the chemistry has in biology or in understanding the biological molecules. So, we will start with the Sangers method or Sangers di-deoxy method.

(Refer Slide Time: 12:21)



Sangers method uses the same idea as DNA replication. Only difference is that it uses a single stranded DNA instead of our double stranded DNA. So your target DNA is a single stranded DNA. So this is your phosphate 5 prime end. And this is your hydroxyl. This is the 3 prime end. Let us say this is your target DNA. And this is a single stranded DNA. So it uses the single stranded DNA as the template.

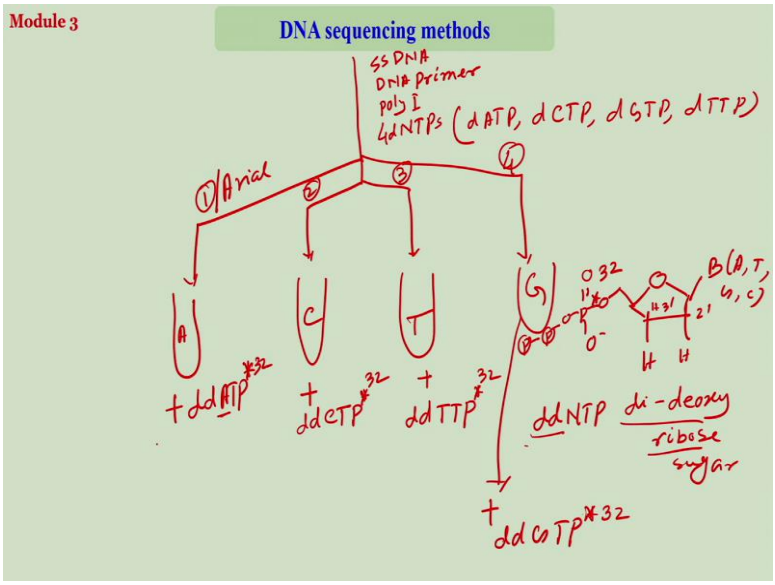
And most often you can use the prime strand or the leading strand 5 prime to 3 prime direction. It also uses a DNA primer, a primer, unlike DNA replication in normal cell that uses the RNA primer. The Sangers method uses a DNA primer. I am not writing the phosphate again and again. So 5 prime, 3 prime and this is your DNA primer, which is complimentary to a part of the target. So that has to be 5 prime to 3 prime end.

And it also uses the polymerase and a single polymerase. But it does not use tack polymerase. They have used the normal polymer is normal DNA polymerase, or DNA polymerase 1. So, that polymerase was actually not discovered in Sangers style, so, they have used the normal DNA polymerase. And since they are not using heat since they are using only the single stranded DNA, the normal DNA polymerase works fine.

And of course, you have to remember that all of these things actually happened before the PC era was developed. And plus, if you add now, the DNTPs so, Sanger has also used the all the 4 dNTPs if you add all of this what will happen? You will basically get the synthesis of the complimentary strand. So, using the DNA primer as the template here, you can synthesize the whole of the complimentary strand.

So, if you use all these ingredients that are what you are going to get. Now, it does not tell you the sequence of your DNA. So, nothing special here, what Sanger has done is that so, he has divided the sample into 4 different vials.

(Refer Slide Time: 15:49)



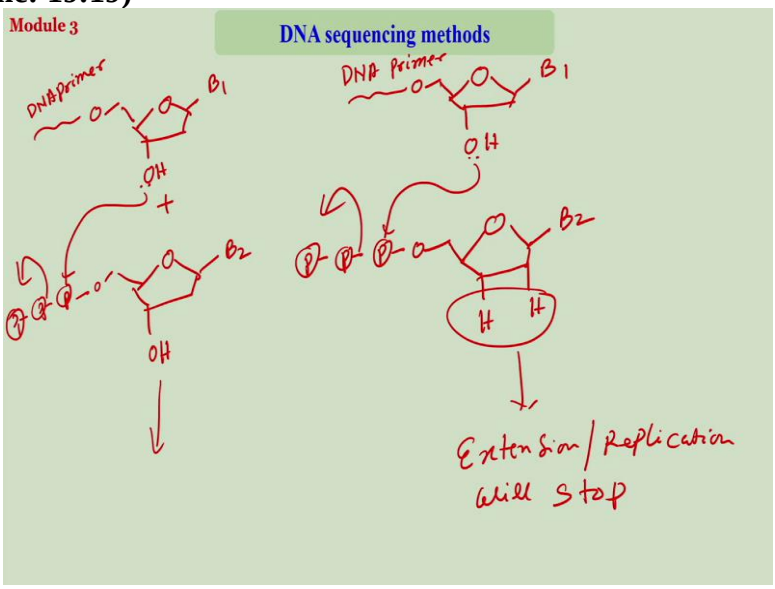
So, if you have these, what is the competition now, you have your single stranded DNA that is your target. You have your DNA primer. Only 1 primer you had to use. You have the polymerase 1 and you have the 4 dNTPs dATP is important. That is why I am writing. dCTP, dGTP and dTTP. So, that is what you have in your mixture and then Sanger has divided into 4 different vial.

So, vial 1, vial 2, vial 3 and vial 4. So vial, So, all of these vials, essentially content, the same composition. The same thing and if you allow it to carry on the replication, all of them will give you the same double stranded DNA. So, again nothing special here. Now, what Sanger has done he has done something extra. And that is where it is a brilliant combination of organic chemistry and biology. So, along with all the 4 dNTPs Sanger has used this molecule.

This is nuclear base A, T G, C H H, O, phosphate 5 prime and I am not writing the whole of it P and P triphosphate. So in each vial he has added a beat of this molecule. This is called you can see now do you see the trick. See, this is there is no hydroxyl group. So it is a deoxyribose we have called when there was a DNA sugar. And there was a hydroxyl group here. So this is 2 prime, this is 3 prime. In this case, you do not have any hydroxyl group.

Both of these hydroxyl groups have been replaced by just simple hydrogen. That is why this is called di-deoxy. So, it is a basically di-deoxy ribose sugar. This unit and it is a triphosphate. So just like dNTPs, it is a triphosphate. Now, why this molecule? What this will do, if you see if you look at the reaction by now I think you know it by heart.

(Refer Slide Time: 19:19)



That when the polymerase was reacting, what was happening? You have the base B1, O and here, let us say that DNA primer. We have started with the primer. This is the last base of the primer with the; 3 prime hydroxyl free. What was the reaction that it was reacting with the dNTP O, phosphate, phosphate, phosphate. So polymerase was catalysing this reaction. This was eliminating out and that is how you get other new phosphodiester bond.

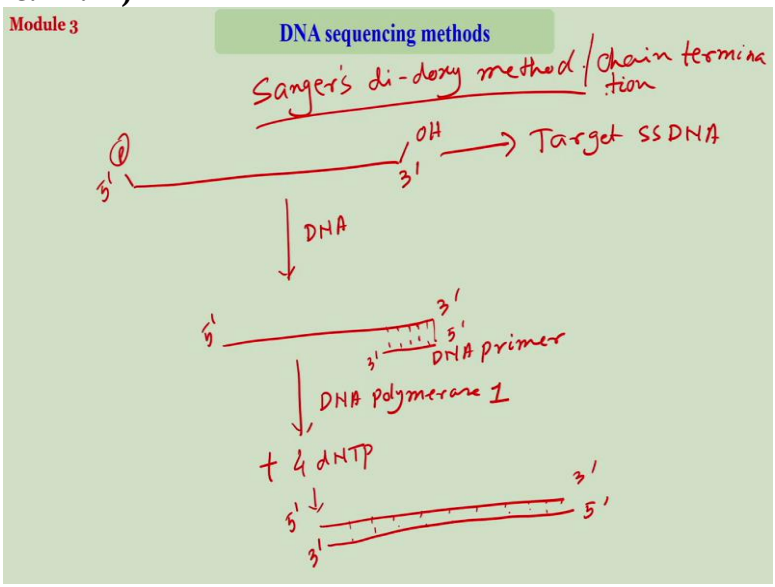
O, this and then again the 3 prime hydroxyl of this B2 will be free and it will continue the synthesis. That was the normal reaction that happens? That is catalysed by the enzyme polymerase during the replication process? Now once you use the di-deoxy what could happen? DNA primer OH. And here instead of the dNTP you have the ddNTP. So I will forget otherwise going back di-deoxy ribose or ribonucleic acids is known is called abbreviated as ddNTP.

So dNTP is the deoxyribose DD is di-deoxy ribose nucleotide triphosphate. This is the abbreviation that will be using from now. So back here. B2 here as usual you have the phosphate and here you have 2 hydrogen catalyst, you use the polymer polymerase the same reaction as it goes. And there will be this formation of the phosphodiester want. Here, one base would be attached. What next? This does not have a free hydroxyl group at the 3 prime end. Since this does not have a hydroxyl group, it cannot do the further process.

So it will stop the replication in your solution once this molecule is picked up by the polymerase, and this reaction happens after that your replication would be stopped or

extension whether you call extension or replication will stopped. It cannot move any further. That is where the chemistry is involved and that is where the beauty of it.

(Refer Slide Time: 22:47)



So, Sangers method is known as di-deoxy method. It also is known as a chain termination method, chain termination method. Because, as you have seen it stops the replication in between. So, the complimentary sequence or complimentary chain DNA chain would be terminated would be stopped. So, that is what the entire idea is. Now, if I come back here. What Sanger has done, so here you have, I am not writing the all of this here. Additionally along with all the 4 dNTPs Sanger has used a little bit quantity of this dd thing.

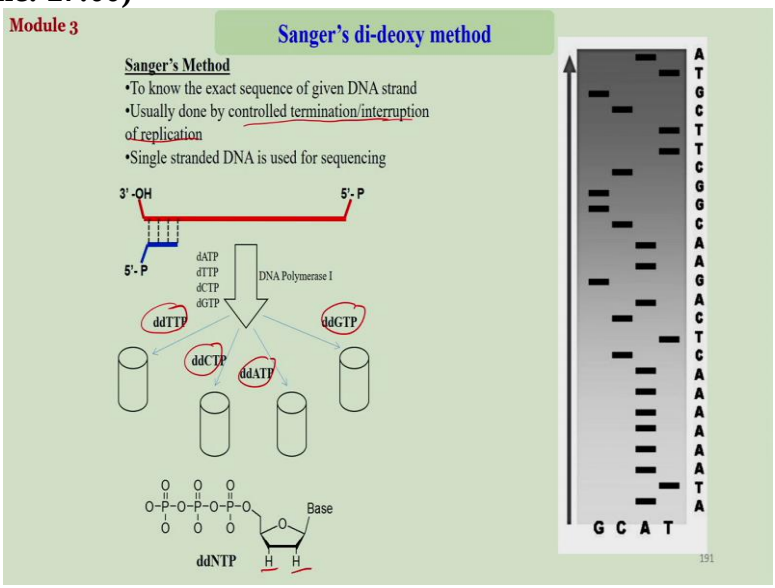
For this particular vial and here comes that differentiation. ddATP was added in this case, so while one had all the 4 dNTPs + the ddATP and now I call this vial instead of 1 I call this vial A, A vial, because it has ATP added ddATP that was added. Similarly vial 2 along with everything that it had another ddCTP was added. So, this is A vial. I call this now a C vial. In this case dd maybe TTP was added. So this is a T vial, because it will be easier to figure out and for this plus ddGTP. We call it G vial.

So, here comes the difference between the vials, each vial is now different. And now let us see how the reactions proceed. What good this is going to do for this addition of the new ddNTPs are going to do. So there was another trick that is the Sanger has levelled this phosphate with a 32 radioactive element of that ddNTP. So, nowadays we do not use radioactive levelling, most often we use the fluorescence levelling and that is done on the nucleo base not here.

So, anyway, be it the radioactive level, be it the fluorescence level, the labelling was done on the ddNTP. All of them what is the use the levelling are done to locate a thing to track where the molecule is going. For example, if you keep your GPS on in your mobile, it is to know where exactly you are. People can track you or you can know your location where exactly you are. So it basically to track somebody to track something that we do levelling. Here also so the radioactive labelling was done on P32 to track the molecule.

So all of these now has a star with P 32. And here P 32. And now let us take 1 vial at a time, and let us see, what are the reactions that are happening. So this is basically here.

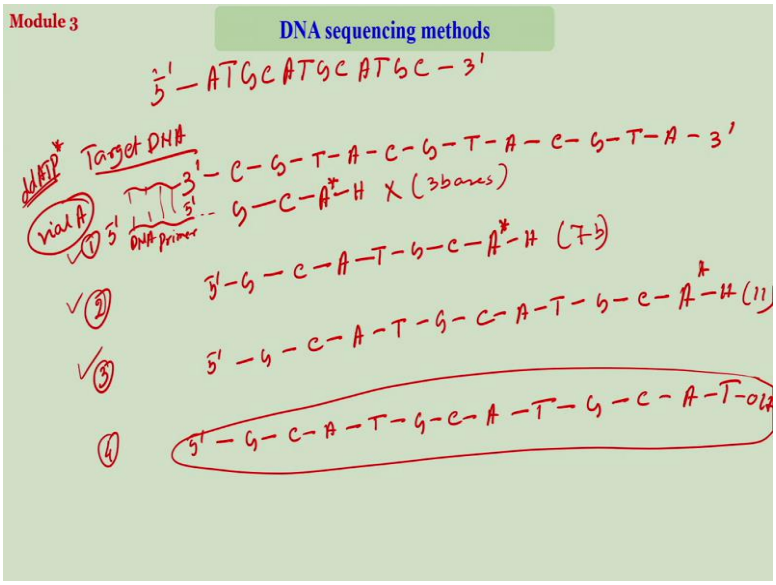
(Refer Slide Time: 27:00)



This. So you have the DNA, you have the primer, you have added initially all 4 dNTPs, ATP, TTP, CTP, GTP. You have added DNA polymerase 1 then it was divided into 4 individual vials and in each vial are separate a ddNTP was added. This is the structure of the ddNTP is the nucleobase 2 prime deoxy 3 prime also deoxy. So, it cannot go further extension for the when the synthesis of the complementary strand would be happening.

So, use only done by controlled termination or interruption of the replication. That is the whole idea behind the Sangers method. So, let us take a target DNA.

(Refer Side Time: 27:57)



5 prime and maybe I am taking an arbitrary sequence ATGC ATGCATGC, 3 prime for the sake of simplicity of for me it would be easier if I write it on the other way around. So my target DNA single strand DNA, I am writing now, from the 3 prime to 5 prime it, it would be easy for me C G T A C G T A C G T A. So this is the target DNA. Let us consider vial A. Where you had added the dd ATP and that had a radioactive level.

So what will happen? Let us say there was something else here. The other part of the DNA that we are not considering, and here, the DNA primer was bound, because you had to start with a DNA primer. This one was the DNA primer. So we will not consider that part now. We will see what happens after this after the primer after you start your reaction basically. So this is your 5 prime. Here, it is basically starting with the 5 prime.

Now, what will happen you have all the 4 dATPs you have polymers and you have that primer ready. So it will start the extension. So it will have the first G that will be picked up again C, then it will be C here then there has to be 1 A here. Now here comes the trick, or here comes the difference of reactivities that you have added the normal d ATPs there in the vial A that was there.

So if it takes the normal dATP that has a 3 prime hydroxyl group free and it can move on. So, if it takes up if the polymer is speaks up the dATP it will move on forward direction, no problem, but instead of the normal dATP, if it picks up the ddATP, because you have also added this compound into our mixture. So, there will be competition between the normal ATP and the ddATP and it may happen that is a high probability of course.

That certain molecules are in certain molecules, the ddATP will be picked up and once this is picked up, you have A 3 prime hydrogen here. So, your chain is terminated. There will be no further extension and that is what the labelling is also coming. So, since you have the labelling, it will amplify the signal and you can track where this molecule is now. So, what Sanger has done in each of these vials that you have added the ddATPs the quantity of the ddATPs.

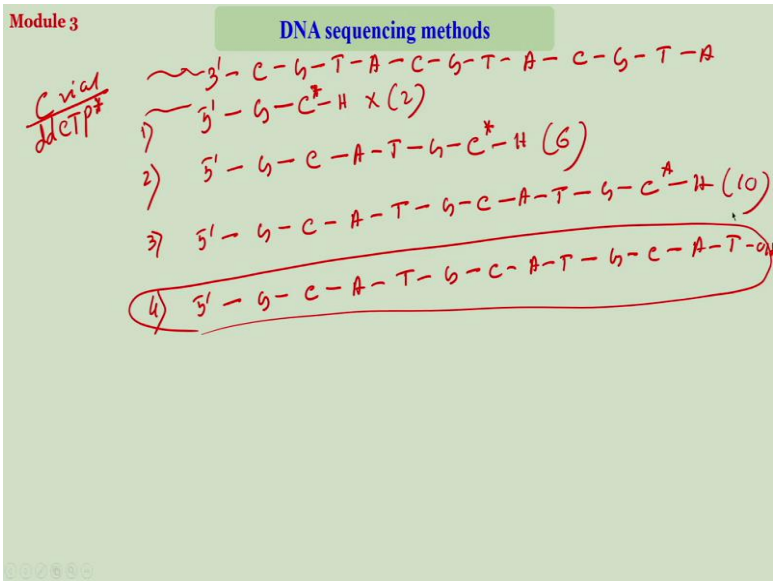
Or ddCTPs ddTTP were small compared to the normal dATPs or the normal dNTPs. The idea is, you allow the extension to happen, but at the same time, it can stop the certain number of molecules are a fraction of the reactions can be stopped due to the presence of these molecules. So, if you have, for example, 100 copies of DNA, or 200 copies of DNA at the beginning, then what will happen? Maybe 10 copies of DNA will be stopped here and rest 90 or rest 190 will move along.

That is where you get the more fragments. So, in this case the case number 1. So, first case is that your reaction may stop here. Case number 2, I am not writing the primer again anymore move directly here. G C that it takes up the normal dATP. So, it will move along here is a T here would be G here would be C here would be A. Again the same issue that this may be the normal dATP so it will move on and the probability is also there that it may be the ddATP. So in this case, it will stop here.

So you will get the 2 fragments and the radioactive levelled. Case 3 5 prime G C A T G C A T G C again it has to come and it may stop right here. And case for that it does not stop at all A, T, C, D C A T, I think it should be G C GC AT GC AT GC AT. So, this would be G this would be G C A T and the G A T, the C A T G, G, C, A. and then finally T hydroxyl. But this molecule does not have any level here. So you would not see this and it will not interfere into your data.

Because there is no tracking here so it does not give you any signal, even if it is present. So ideally, with radioactive levels, you get 3 different fragments. One this, one this, one this, we will use this just in a week time. So A T G C C G T A.

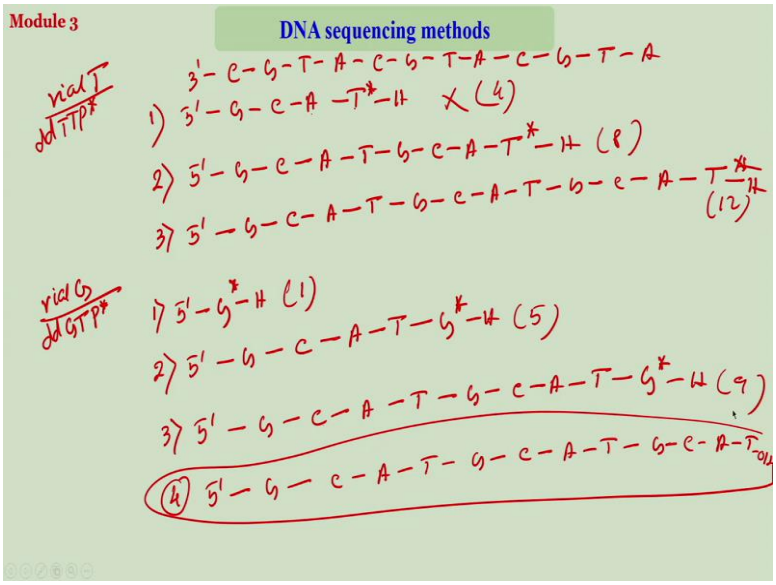
(Refer Side Time: 35:13)



3 prime C G T A C G T A. Now, come to the second vial second vial is your CTPIS here CTP. So, you call it C vial. C vial you have the ddCTP with regularity level. Start with the 5 prime, I am not showing that this is a primer, this is the other part of the DNA this is the DNA primer. So, it has to take a G which will be a normal G now, it has to take a C. So, again this C can be the dd CTP or it can be the D G CTP if it is ddCTP then this will stop it has the hydrogen at the 5 prime.

So, far the elongation is stopped and you have that radioactivity level. So, this is 1 it is case 1 what can be 2? 5 prime G C that it goes on A T G C again here. So, wherever there is a probability of picking up a C here, maybe a ddCTP can be picked up and then it will be stopped. 5 prime G C A T G C A T G C stopped. And then the fourth one is the full length. I guess T, G, C, T, G C A T and the hydroxyl group here. But again, this guy you cannot see because it does not have a level. So 3 fragments basically also from this vial.

(Refer Side Time: 37:39)



C G T A C G T A C G T C G T A so vial, what was the original one? I am following this. So TTP. T vial and then we will come the G. Vial T so we call it T vial, where you have added the ddTTP levelled. So case 1, 5, this is G C A T . So wherever there is a probability taking up a T, it can be either the normal T or it can be the levelled the radio activity. Once it is there, it is hydrogen. The chain is terminated.

Case number 2 G C A T G C A T star 5 prime, G C A T G C A T G C A T the last one and it would be H and of course, that is the full length. So, again 3 cases. So, I think here itself we can do the vial G where we had added ddGTP. So, here first case will be 5 prime G the very first case and that would be stopped no further. So, please remember that there are other sequence before this that I have not considered and I am assuming that the DNA primer that we have used would be complementary to that sequence for all cases.

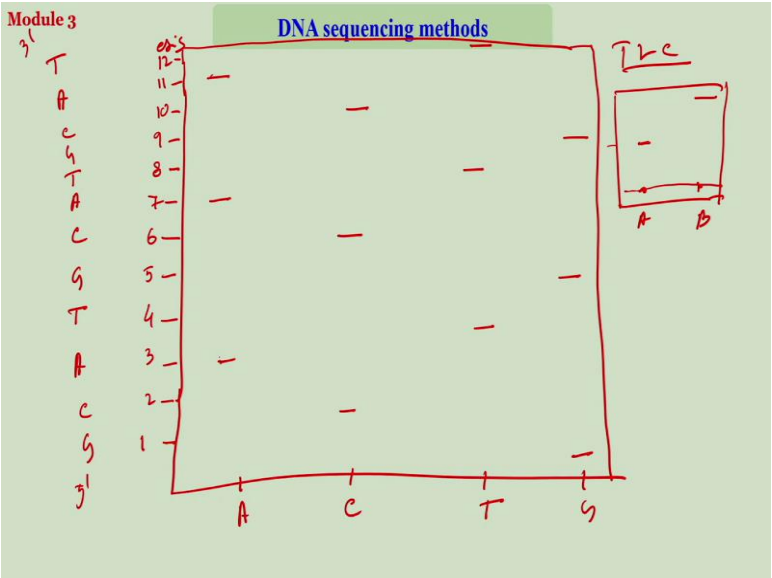
G C A T here there has to be another G stopped 5 prime G C A T G C A T here There has to be G and the full length. I think, 5 prime, G C A T G C A T G C A T OH. And again, this is not levelled. So now you can see so many fragments you have. Now here comes the chemistry. Since you have fragments, now you can study them the whole long DNA is hard to study at a time. So now if you have the smaller sequences, you can study them well.

Another thing did you notice that all of these sequences have different length of the DNA For example, this is 1 Nicholas long, this is 1 2 3 4 5 Nicholas long. If you see if you look at all the fragments that we have done, all of them are of different length of DNA. So, since DNA

has negatively charged phosphates, so overall charge of the DNA is negative. So number of phosphate means the number of negative charges.

So as the length of the DNA will increase, the number of charge will also increase negative charge will also increase. And of course, by the charge density, now you can separate the molecules. If we call the fragments as individual molecules now, all these molecules now have different charges. Therefore we can easily separate them it is just like doing your organic chemistry.

(Refer Side Time: 42:32)



If you remember your basic organic chemistry that we do thin layer chromatographic TLC. To see the mobility of compounds and if you have a plate for example, a silica plate or an aluminium plate and if you have organic molecules or a sample which has mixture of compounds and you want to know how many compounds there are, what is their polarity property.

So, if this is compound A and if this is compound B and if they have difference of polarity, then if you put a drop of this molecule here, if you put a drop of this molecule there and allow them to move, then they will move along the plate and you can see depending upon their polarity, there are degrees of movement or they are the speed or the velocity of their movement would be different.

So, over a certain time, for example, this compound will move until here. On the other hand, this component will move in a more length will move faster. And that is because they have

different polarity. So, more polar molecule. In this case the polar molecule moves slower, nonpolar molecule moves faster. It is also a kind of the charge difference. Same for DNA that the more number of negative charge they have in this case is a little bit other way around. More number of negative charges it has it will move faster.

And that is called it is also a plate it is also a chromatic kind of chromatography. But this is called electrophoresis gel electrophoresis. So this is called gel electrophoresis is also known as auto radiogram. So, you say electrophoresis technique, and any electrophoretic technique means, they will separate molecule in terms of charge or in principle, charge density basically. So, for DNA, the longer the DNA is, movement will be higher.

So, now if you have the mixture from the A vial from C vial, from T vial and from the G vial if you put that certain quantity here, here, here from the individual 4 vials and you allow them to move, then according to the length, they will move and you have the tracker, the V 32 levelled or the fluorescence level where you can see how far they have moved. So that is where is done for the analysis.

And let us say we divide the y axis in terms of the length of the DNA, this is one base long, this is 2. This is 3 4 5 6 7 8 9 10 11 and 12 here. So you have scaled your y axis to see the number of. Now look at the A vial and let us see which length you get the spots. It is very simple. A vial you have this is 3. So 3 basis I am writing 3 B. So, this is for 3 this is 1 2 3 4 5 6 7 3 7 is 1 2 3 4 5 6 7 8 9 10 11.

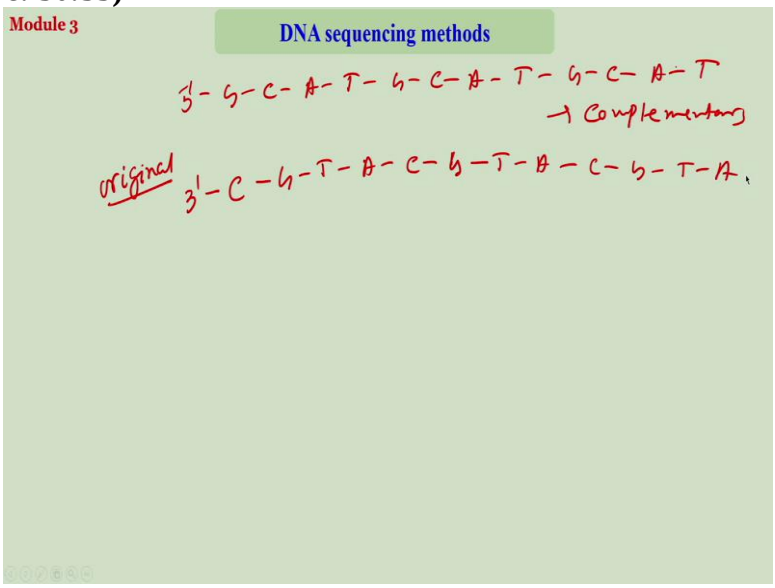
This is your 11th. So, 3 7 and 11 you should see the spots 3 here, 7 here and 11 here. So, you should see since you have 3 fragments, they should have different movement and it should come in these positions. And of course the full length does not have the tracker so you cannot see DNA is invisible. The quantity is very low or also normal the end does not have any color.

So, you cannot see that. All you will see where there are the level. So 3, 7 and 11. Now, C vial. What you have 2, 1 2 3 4 5 6 1 2 3 4 5 6 7 8 9 10. 2, 6, 10. 2, 6 and 10. 3 fragments. T vial you have 1 2 3 4. 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 9 10 11 12. So that is the full length 4, 8 and 12. 4, 8 and the last one 12. Quickly for the G 1, 1 2 3 4 5, 1 2 3 4 5 6 7 8 9, 1, 5, 9 1,5, 9 this is your data, this is your result.

Now, if you see here every length is here, you have spot for every length, what you have to simply do is just read the gel and put the corresponding base here. So, we have started basically we are would actually synthesizing the complimentary strand from the 5 prime end. So, this is your 5 prime end where is the first base here G, so, this has to be a G second base is C. So, there is a C here third base is A. So A. Fourth is your T.

Fifth is your G. sixth here C seventh A 8 T 9 is G 10 C 11 A and 12 is your T. This is the 3 prime. Now this is so simple to analyse the data. By the way, this is your sequence of the complimentary strand that you have been synthesizing. So, if you can figure out now 5 prime G C A T G C A T G C A T. So your original strands.

(Refer Side Time: 50:35)



So, 5 prime G C A T G C AT G C A T. So this was the complimentary strand. So your original strand or the your target the sequence of your target DNA would be, of course, C G T A C G T A C G T A. Now, let us match what was your target. C G T A C G T A C G T A. So just by reading your data just by reading the gel, you can find out the sequence of your desired DNA. This is so simple analysis.

So, if you think of small organic molecules, and if you want to characterize, if you want to know the structure of the compound, what we do have to do maybe elemental analysis, you have to do NMR, you have to do FTIR and so on. And all these techniques require a lot of knowledge or require hard work to analyze the data itself then you come you figure out the structures. So, this is something which is so simple to interpret the data does not need much

hard work or much knowledge to analyse. It is so simple it would just see the gel and put up your DNA strength. This is what the beauty of the Sanger's method. Thank you.