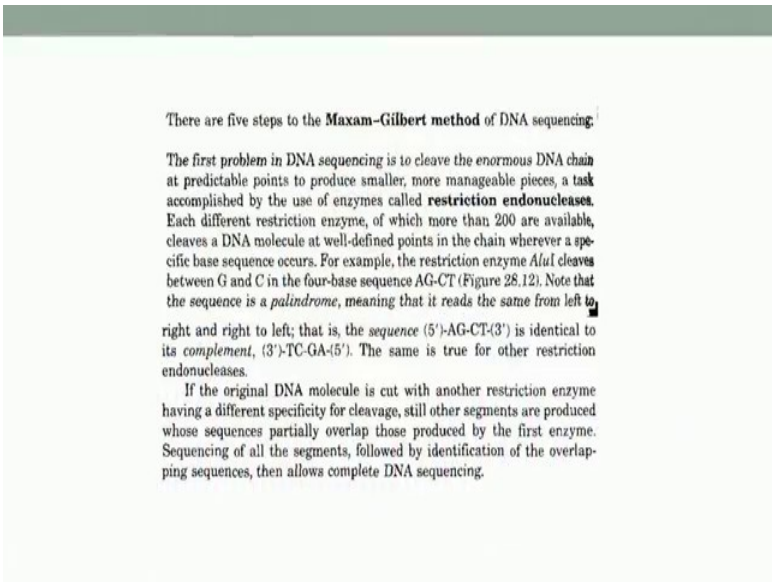
Organic Chemistry In Biology And Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture - 25
DNA Sequencing Method (Contd.)

Welcome back to this course on Organic Chemistry In Biology and Drug Development. Now last time we were discussing the determination of sequence in a piece of single stranded DNA. Sequence in DNA means the sequence of bases that are arranged one after another; this is extremely important because ultimately the sequence of bases will dictate the type of protein that will be ultimately made in the living system.

So, determination of sequence is the primary requirement in order to predict the flow of information towards the protein.

(Refer Slide Time: 01:19)



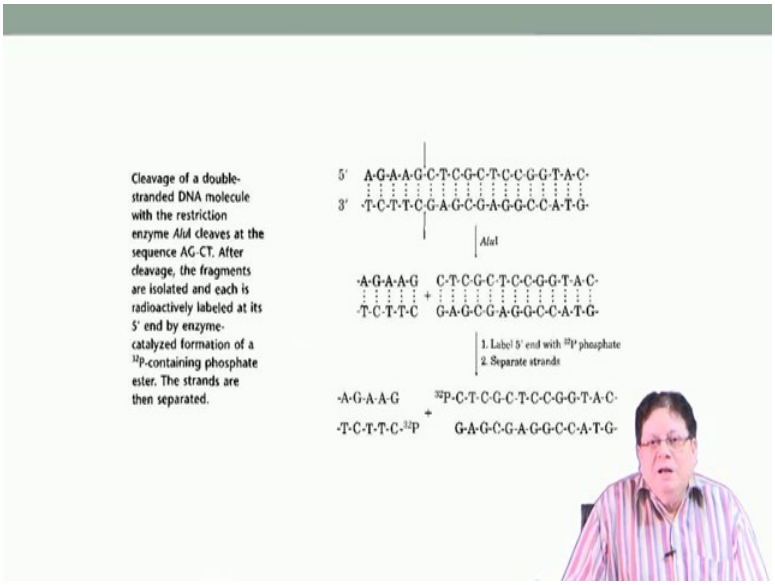
There are five steps to the **Maxam–Gilbert method** of DNA sequencing:

The first problem in DNA sequencing is to cleave the enormous DNA chain at predictable points to produce smaller, more manageable pieces, a task accomplished by the use of enzymes called **restriction endonucleases**. Each different restriction enzyme, of which more than 200 are available, cleaves a DNA molecule at well-defined points in the chain wherever a specific base sequence occurs. For example, the restriction enzyme *AluI* cleaves between G and C in the four-base sequence AG-CT (Figure 28.12). Note that the sequence is a *palindrome*, meaning that it reads the same from left to right and right to left; that is, the *sequence* (5′)-AG-CT-(3′) is identical to its *complement*, (3′)-TC-GA-(5′). The same is true for other restriction endonucleases.

If the original DNA molecule is cut with another restriction enzyme having a different specificity for cleavage, still other segments are produced whose sequences partially overlap those produced by the first enzyme. Sequencing of all the segments, followed by identification of the overlapping sequences, then allows complete DNA sequencing.

Now, I said that there are primarily two methods; one is the chemical method and the other is an enzymatic method. The chemical method was discovered by Maxam and Gilbert. In our last lecture, we had a detailed analysis of the Maxam-Gilbert sequencing technique.

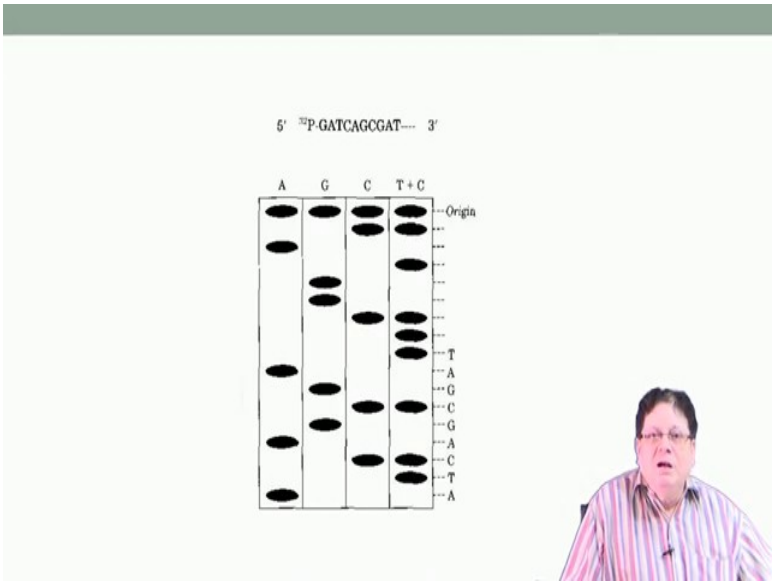
(Refer Slide Time: 01:39)



Now, like the proteins, DNAs are very large and it is very difficult to handle such large pieces of DNA and also it is not possible to determine the sequence of the entire DNA in one attempt. So, you have to first cleave the DNA into smaller pieces and these smaller pieces can be obtained by the use of restriction enzymes, a particular class of enzymes which are called endonucleases. They recognize typical palindromic sequences within the DNA and accordingly they cut the DNA. There are two types of cuts; there are blunt cuts and there are sticky cuts.

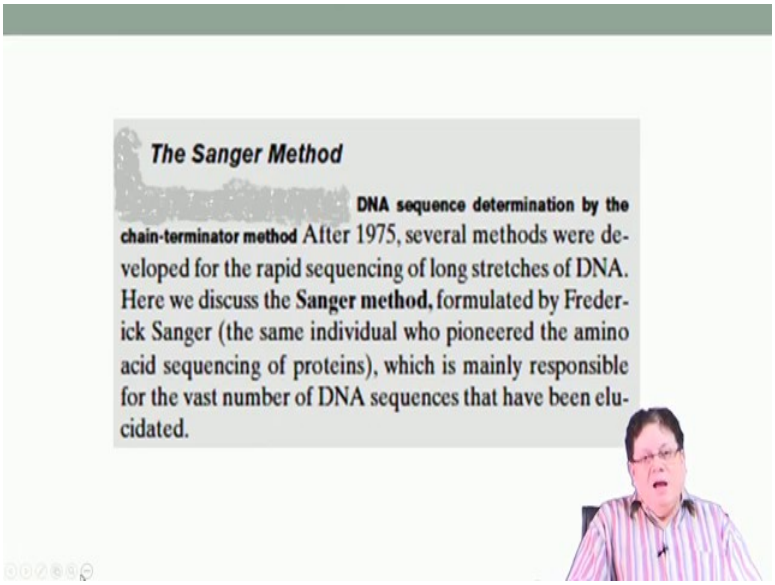
So, basically you have to take the DNA first; that is the first requirement for any sequencing technique; then you have to cleave the DNA into smaller manageable pieces using restriction enzymes.

(Refer Slide Time: 02:40)



And this was the gel pattern that you got in case of Maxam-Gilbert sequencing; I just remind you that some books have written that there are 4 lanes; one is A-specific lane, another is G, another is C, another is C plus T. But in practice what happens is A and G; some bands which correspond to the cleavage of the G can also be there. That is why I said you a G specific reaction and A plus G specific reaction, C specific reaction, and C plus T specific reactions; that serve the purpose.

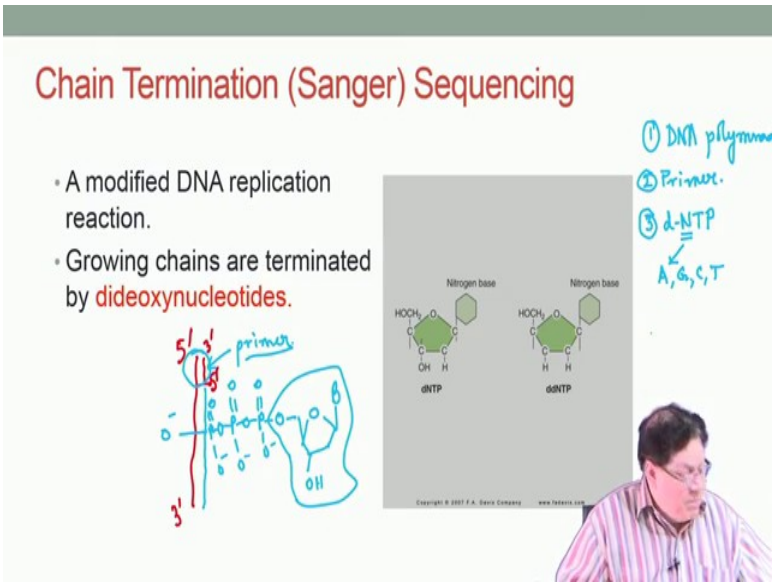
(Refer Slide Time: 03:33)



Now, we will go to the other technique which is the Sanger's method. It was discovered by Frederick Sanger at Cambridge University, England and that was done in 1975; at the same time, Maxam-Gilbert sequencing technique was also discovered.

Now, this Sanger's method is an enzymatic method, but it is also known as a chain termination method or a dideoxy method. So, these are the other names given to Sanger's method of determining the sequence of bases in a DNA.

(Refer Slide Time: 04:10)



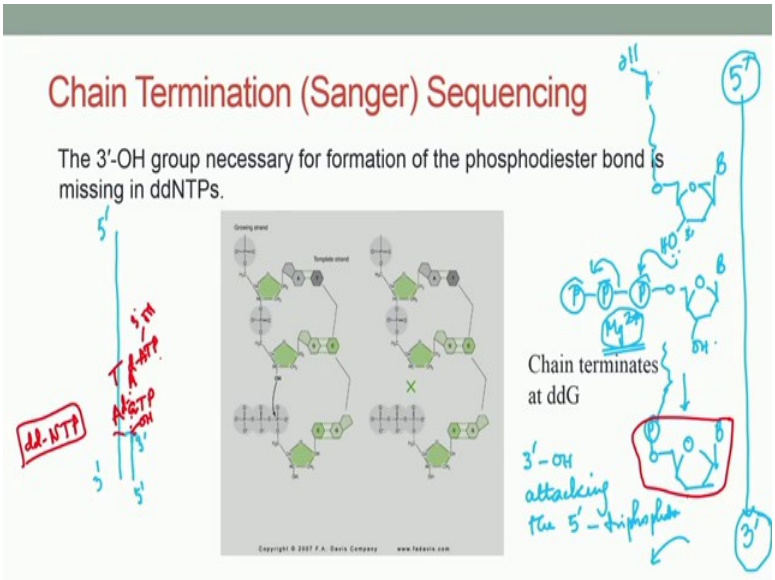
Now, what is the principle behind Sanger's method? There is an enzyme called DNA polymerase. The function of DNA polymerase is to make that DNA polymer. If you have a single strand of DNA running from 5' to 3', then the DNA polymerase can make the complementary strand of this primary DNA strand. Suppose we have this DNA strand and we want to know the base sequence in this strand.

So, we have to use the DNA polymerase in order to make the complementary DNA strand. However, there are certain important points that one needs to consider. One is that DNA polymerase works only when there is a small piece of DNA which is attached to this primary piece of DNA; primary means for which you are trying to determine the sequence. So, a small piece of DNA needs to be attached and then only DNA polymerase can start from there and make the remaining part of the chain; this piece is called the primer. Now this is made up of deoxynucleotides. So, this is a DNA primer. Unless this primer is there, DNA polymerase cannot work.

So, basically there must be a small segment of double stranded DNA and then the remaining part can be joined together one after another by the DNA polymerase. So, in Sanger's method, first thing that you need is DNA polymerase and along with DNA polymerase, you also need a primer. And then the third thing is that in order to make this DNA strand, what you also need the building blocks for the DNA synthesis; like if you want to make a peptide, you need the building blocks (amino acids). So, here also the DNA will need the building blocks; now what are those building blocks? They are the deoxynucleoside triphosphates.

So, basically I can write the structure. So, what you need is this thing: A base, the OH and triphosphate; that means, P double bond O, O minus then O another P, double bond O, O minus and then again O, P double bond O, O minusand on the top of it that is O minus. So, this is what is nucleoside triphosphate NTP, deoxynucleoside triphosphate because up to this point is what is known as nucleoside; so it is called the nucleoside triphosphate. N stands for all the four bases A, G, C and T. So, basically what you need is all the four deoxynucleoside triphosphates.

(Refer Slide Time: 08:18)



Why do you need this? Suppose you have made a part of the DNA (primer) and the 3' OH is free, now if you add a nucleoside triphosphate, then DNA polymerase comes. First let me write phosphate, you can always write like this (circle), it is a triphosphate. So, DNA polymerase attacks this phosphate in presence of magnesium; a magnesium ion as

a cofactor is needed to neutralize the charges of the of the phosphate ion. Because the phosphate is negatively charged and you are doing a nucleophilic attack; so someone has to neutralize these charges and that is taken care by the magnesium ions. So, the OH attacks the phosphate and inorganic diphosphate or a pyrophosphate goes out. So, in the process you have extended the DNA chain by one nucleotide.

So, this was the earlier DNA part. So, then this again attacks another triphosphate and that is how the DNA chain is extended. This is the principle of functioning for DNA polymerase. Some important points that we need to consider is that it is the 3' OH of the oligonucleotide chain that works as the nucleophile to attack the phosphate. Thus when the DNA chain is made, it is the 3' OH attacking the 5' triphosphate. Now on the top side ultimately you have a oligonucleotide here. We have a nucleotide here where the 5' OH will be free and as you go down, finally it is truncated when it achieves the desired length, so, your 3' OH will be free.

So, the direction of synthesis is from 5' to 3' end. Thus it is the 3' OH that is attacking the 5' triphosphate. So, this is the job of DNA polymerase. The important thing is that the direction of the synthesis takes place from 5' to 3' and while doing so, it is a 3' OH attacking the 5' triphosphate. Now what Sanger did? Sanger thought that if suppose if for one of these deoxynucleoside triphosphate, if this 3'-OH is not there, then what will happen? Then if this is incorporated in the chain, then the chain will stop here; it will truncate, because that nucleophilic OH is absent here.

Suppose you have the primary DNA strand and you have the primer up to thispoint, that is your 3'-position. So, the synthesis will not take place from here because already I have said that the synthesis takes place from the 5' to the 3' direction. So, if this is your 5', going from 5' to 3', then your primer 5' should be here and 3' here. As you add all these deoxynucleoside triphosphates, depending on the sequence here, whether A is there or G or all these sequences, the complementary base containing nucleoside triphosphate will be brought here. And these 3' OH is free; so that will now react with the nucleoside triphosphate and then extend the chain according to the base, whatever base sequence here.

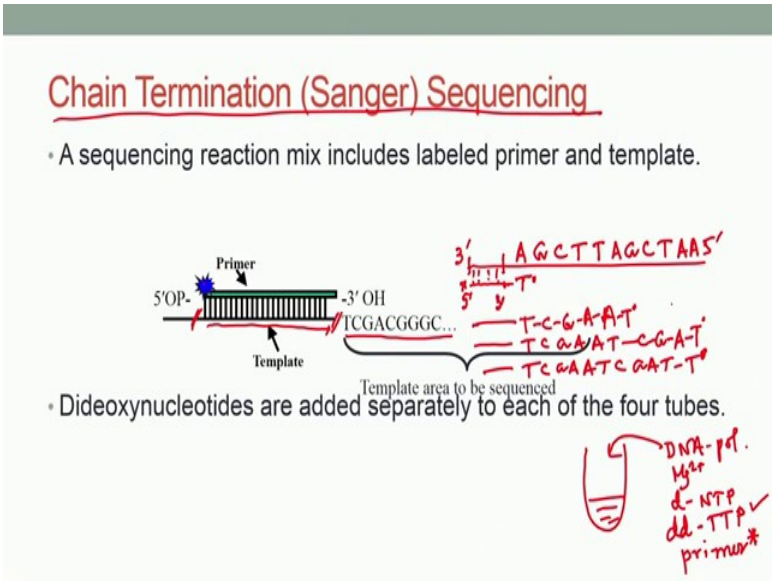
Suppose the primer the first base is A; so definitely the it will require a deoxy TTP, dTTP that will be brought; the 3' OH will attack the 5' triphosphate, and in the process a

phosphodiester bond will be formed between the primer and the T. Suppose the next one is T; so now, there will be a dATP that will be brought. Now the 3' OH of the T will attack this triphosphate of ATP. It was only the thymidine mono phosphate because the diphosphate is already out. So, that that is how A will be incorporated here. I hope this is clear.

So; that means, it goes one after another. The important thing is that always the 3' OH is attacking the 5' triphosphate. Now Sanger cleverly did one thing; that along with deoxy nucleoside triphosphates, he also added the dideoxy nucleoside triphosphate (ddNTP). So when there is requirement of suppose A, if instead of dATP, ddATP is taken up by the enzyme by mistake, then the chain will truncate at that position, the chain will not proceed any further chain.

If the correct nucleotide (deoxy-ATP) is taken up then the chain will progress but the chain growth will terminate if a dideoxynucleoside triphosphate is taken instead of the normal deoxy ones, that is why this method is also is called the chain termination method by Sanger.

(Refer Slide Time: 15:55)



Now, let us take an example which will make it clearer. Let us see, so suppose this is the DNA that I want to sequence and suppose this is the one which I need to know the base sequence, this part is what is called the primer.

So, when I have the DNA, it goes up to that point; that means I know the sequence up to this part. There are ways of synthesis of oligonucleotides. So, I will synthesize the oligonucleotide which is complementary to this part of the DNA. So, what will happen, the primer will now hybridize, will form a duplex here by the hydrogen bonds; and in addition, this primer is such that the 5' OH is having radioactive phosphorus (like Maxam Gilbert method). The 5'-OH is tagged with a radioactive phosphate; the phosphorus is radioactive. That is needed because ultimately you have to find the pieces where that label is present; otherwise it will be all mixed up with the other nucleotides that are present there.

So, you need a tag which will ultimately act to determine the nucleotides that are truncated. You may not understand right now, once I give the example it will be very clear. This radioactivity is needed to know which nucleotides are truncated by taking the dideoxynucleoside triphosphates. So this is the primer. One important issue of Sanger sequencing is that, if I have the DNA, I need to know the sequence of few bases at the end; then I can design the primer and the primer goes here and binds.

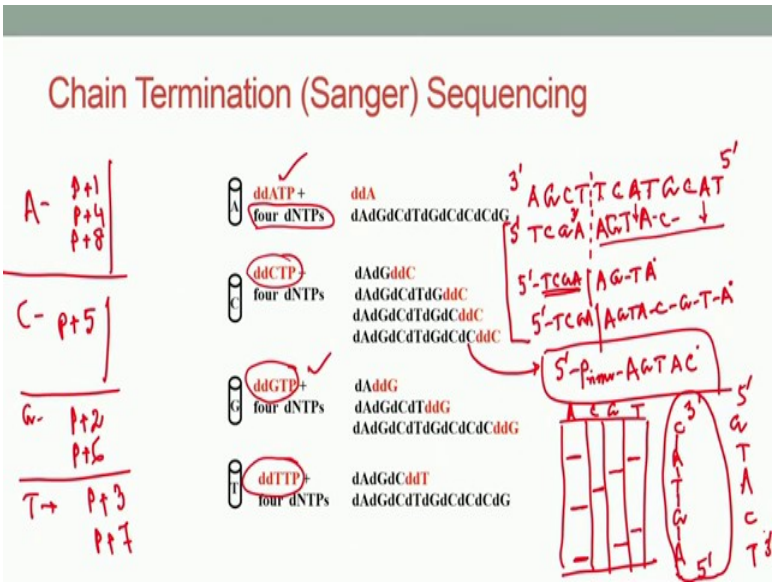
Let us take a new example. Suppose this is the sequence that this DNA has: T T A G random just I am writing just random some sequences, ok. Remember this primer starts from 5' to 3' and that has to be so, because the DNA sequence goes from the 5' to 3' direction. So, now, what will happen? So, whenever there is A, so it should take a T. Now suppose instead of T the enzyme picks up a dideoxy T, then this piece will not grow any further.

So, let us make a; some kind of notation suppose this is dot; dot means didoexy T. So, if didoexy T is taken, then the whole thing will stop here. Now if the correct T is taken; that means, the deoxy T, then it will continue up to the point when there is requirement of another T. So, C then G then A then A and now you again come to this point where there is requirement of a T. So if the enzyme now takes the dideoxy T; again you will have a truncated chain here. So, you can have 2 truncated chains; you can have more also if there are more As because it depends on the number of As in the chain. So, you will have others also, so T C G A A T because when this is taken when the correct T is taken, this is going to proceed forward. So, then C because it is G then a G then A, no problem; now then there is a T. So, there is a chance of truncation.

So, it takes a dideoxy thymidine and that will truncate here. And but if it takes the right one then you will get another one T C G A A T C G A T; suppose this is the correct T that the enzyme has taken, but the next A again there is a possibility that it may take a wrong T; wrong T means the dideoxy T. So, in a test tube, suppose what I do, I take the DNA, single piece, then I add it the DNA polymerase, I add magnesium, I add all the deoxy NTPs (the normal ones) then I add ddTTP; that means, the dideoxy TTP, other dideoxy I do not add to this test tube or the eppendorf and what else I need to add radioactive primer. So, the primer goes and binds and because I have only ddTTP. So, the chain will terminate only when there is requirement of T, others will not terminate because I have not added the ddCTP or ddGTP or ddATP. So, that has to be made clear.

So, if you add ddTTP in one of the test tubes or one of the eppendorfs, then you will get four pieces which are all truncated at T.

(Refer Slide Time: 22:35)



So, if that is clear, now the other part will be very clear. This is what is shown here, these are the eppendorfs. In one eppendorf, you add all the four dNTPs, DNA polymerase, magnesium and the primer has been added. In addition, you add is ddATP. So, if you add ddATP then the chain will terminate whenever there is a requirement of A. I am not saying that it will always terminate, some will take the correct A and that will proceed forward, but whereever the enzyme picks up the wrong A; that means, the dideoxy nucleotide, there the chain will terminate.

So, there are four eppendorfs; in one I have added ddATP; in the second eppendorf, I added ddCTP; in the third one I added ddGTP and in the fourth one I added ddTTP. I take another example, suppose the DNA has this type of sequence A G C T T C A T; just to make it little bit shorter, because the principal is most important, so take a short one. Now suppose my primer is this: T C G A; this is the length of my primer; usually primer lengths are longer than this; but for the sake of the problem, I am just taking this part, say up to this point. The primer has to start from 5' and then it should go to the 3'. The main strand runs from 3' to 5' direction. I have written that sequence and so the primer will be 5' to 3'. So, up to this point is my primer. Now what will happen? In the first test tube what will I have? Since I have added ddATP, so, I will get a truncated chain like this: TCGAA°; where A° denotes dideoxy adenosine derivative. Next I will get 5'T C G A (this is the primer) and then the important thing. So, if the proper A (deoxy ones) is picked up, then it will proceed further to G. Now this test tube does not have any ddGTP. So, G will always be the right one that will be picked up; then T, then there is another requirement of A, so it may terminate here and finally and we will get 5' T C G A that is the primer then A G T A°. So, you just see whether there is requirement of A anywhere else. So the third truncated fragment that we may obtain is 5' T C G A A G T A C G T A°.

Thus, I will get 3 pieces in the first test tube where I have added ddATP. So, in this way I can predict without writing that if you added ddCTP, so how many pieces you will get. So, you will see that where are the Gs, at that point C will be taken up. So, if you add the ddCTP, so you will get truncated chain only there. I could see only one G after the primer. So, only there it will be taken up. So, for this one, I will get 5' T G C A A G T A C°. So, it will truncate here that is the only piece that you will get from adding ddCTP.

So, from ddGTP test tube how many pieces you will get? There are 2 Cs in the template (primary) chain after the primer, so we will get two truncated fragments. And for TTP you will get truncated fragments wherever there are As in the template (primary) chain. So, you will get two such truncated fragments. Now if you count these number of truncated fragments obtained from each test tube, for the first test tube, where you have added ddATP, you can get oligonucleotides which are primer(P) plus 1, also you will get P plus 4; and you will get primer plus 8. In the case of ddCTP, you will get primer plus 5. For your G specific reaction, you will get primer plus 2 and primer plus 6 fragments.

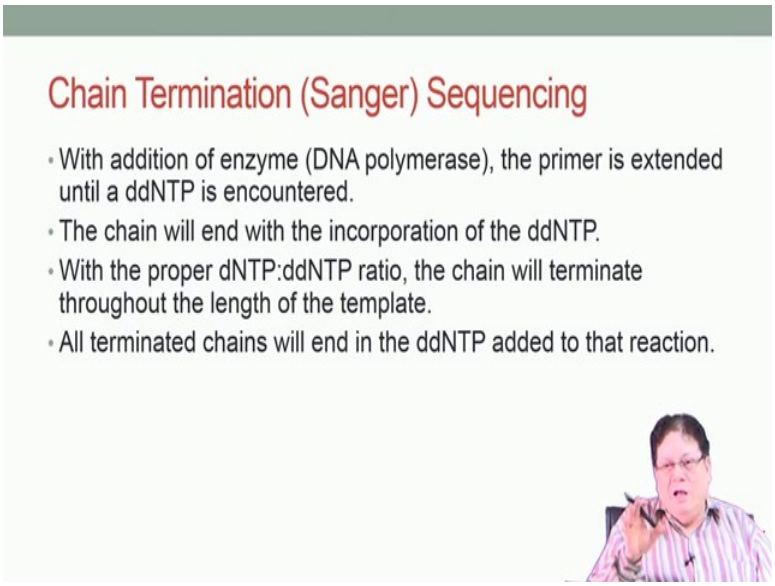
For the T specific reaction, you will get primer plus 3 and you will get primer plus 7. That means, all these test tubes are giving oligonucleotides which are varying in length. If we go in ascending order of molecular weight, then P+1 (in the A lane) is first then P+2 (in the G lane) is second; the third one is here in the T lane; similarly we can identify the the fourth, fifth, sixth, seventh and eighth in their respective lanes.

So we do a gel electrophoresis. We take a gel electrophoresis and we run four lanes. So, one is A, another is C, another is G and the third is T. So, what will happen in the A lane, you will have three pieces. So, one will be the minimum molecular weight, then you have somewhere here that is the P plus 4 and somewhere here that is the P plus 8. So, and then for the C lane, you will get only one piece and that will be here because that will be little bit below than the P plus 4, then for the G one, you will get somewhere here that is P plus 2 and then P plus 6 which is somewhere here; and for the T lane you will get P plus 3, so that will be here and then P plus 7; so P plus 7 will be little bit here, because this is your P plus 8. So, now, you read it from the bottom to the top.

So, what will happen first is your A then comes the G, then comes T, then comes A, then comes your C. So, that way you just read it from this to that side. So, what is this A G T A C; now you write the sequence of the complementary strand of the DNA. So, that will be your 5’ end and this will be your 3’ end. So, basically you read the gel which is very simple; reading the gel is not a problem; you just see what are the bands; which is the fastest moving band;that is the first base then the second then the third then the fourth then the fifth and sixth seventh and so on.

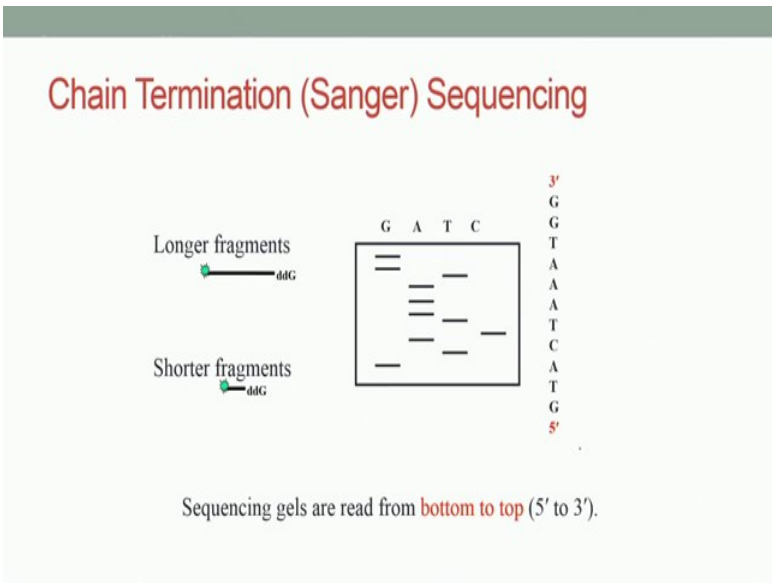
One important issue is here that whatever sequence you get, is the sequence of the complementary strand to your primary sequence. So, you just write the actual sequence. So, that will be 5’ G T A C T 3’. So, that is the basic principle of Sanger’s method. It is called dideoxy because the whole thing is based on the dideoxy approach and the important issue is set finally; at the end of the day, you get the sequence of the complementary strand; from that you have to write the sequence of the primary strand.

(Refer Slide Time: 33:22)



I hope this is clear that this is a very intelligent way of doing the sequencing; because this Sanger's method of using dideoxynucleoside ultimately showed the way to make antiviral compounds. We will discuss when we go to the medicinal chemistry part, but many of the antiviral drugs that we have today are literally based on this principle of using the dideoxy nucleoside. Instead of OH, you can have other group which is not a nucleophile. In Sanger's case, he used dideoxy nucleotides, he just did not bother to put any other group; he removed the 3'-OH. So, that is the Sanger's method.

(Refer Slide Time: 34:11)



So, ultimately you get the this type of gel; and you read from here, but as I told you this is the complementary strand and finally, you can write the actual sequence of your primary strand. So, that is all about Sanger sequencing. Now in the next session, we will first start with what are the problems of Sanger sequencing because that is how science is developed; that we consider the problems. As such, there is no problem; but there are constrictions which are present. So, we will discuss that in the next session.

Thank you.