# Introduction to Complex Biological Systems

**Professor Dibyendu Samanta and Professor Soumya De**
**Department of Bioscience and Biotechnology**
**Indian Institute of Technology, Kharagpur**

## Lecture 52
## Visualizing and analyzing nucleic acids: DNA Sequencing

Welcome back to Lecture 52; I will be discussing DNA sequencing. During the last lecture, I was discussing agarose gel electrophoresis and polymerase chain reaction, but Through polymerase chain reaction we can increase the amount of DNA, but through DNA sequencing we can establish the exact sequence of nucleotides present in a particular DNA segment. So, particularly I will be discussing the principle of DNA sequencing, then the sequencing reaction followed by new development, I would say, automated DNA sequencing method. Now, here, before starting the DNA sequencing method, I would like to mention that sequencing means the process by which you determine the exact order of the nucleotides in a given region of DNA. So, we do DNA sequencing to decode the genetic information present in DNA and how we can carry out this reaction, how we will determine the sequence that I am going to discuss in more detail in the next slide. You can see the scientist is very famous, Frederick Sanger. He received the Nobel Prize twice.

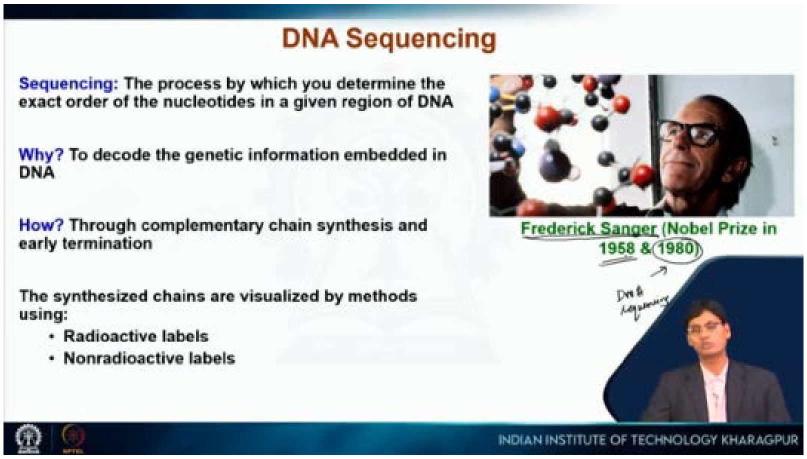
CONCEPTS COVERED

1. Principle of DNA sequencing
2. Sequencing reaction
3. Automated sequencing

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Only out of five people across the world who received the Nobel Prize twice, he is one of them. Again, he received the Nobel Prize in the same category. Both the time he received the Nobel Prize in chemistry and for DNA sequencing, he received the Nobel Prize in 1980. So, this is for DNA sequencing. Then what about 1958? So, he also developed some technique to sequence protein. He determined the sequence of one protein called insulin.

So, for that he also received the Nobel Prize in 1958. This is just interesting, that is why I mentioned. Now let's discuss DNA sequencing methods. This is a DNA sequence for example if I write here 3′ ATGCGTAGATGCTATGCAG, just some sequence here.



DNA Sequencing

**Sequencing:** The process by which you determine the exact order of the nucleotides in a given region of DNA

**Why?** To decode the genetic information embedded in DNA

**How?** Through complementary chain synthesis and early termination

The synthesized chains are visualized by methods using:

- Radioactive labels
- Nonradioactive labels

Frederick Sanger (Nobel Prize in 1958 & 1980)

Now, I am writing this sequence so that I will be able to explain it, but our goal will be to determine the sequence of this stretch of DNA. Without writing the sequence, I cannot explain it; that is the reason I am writing here. So, now, this is just one strand of the DNA. For the DNA sequencing, we have to just sequence one strand of DNA. As you know that DNA strands are complementary, so that does not matter if you just decode the sequence of one strand; that is good enough for us. So now if this is the DNA sequence, then we can use some primer, say for example 5′ TACGCATC. So this is a primer, 5′ to 3′ so this primer can grow so this is primer. May be we are not aware about this sequence as I already mentioned.

So, we have to determine the sequence of that segment. So, how can we do this? So, this primer will bind to this template sequence and then if we use DNA polymerase then this primer will be extended. So, whatever just like PCR, just like replication, we have to carry out this reaction and then we will add some additional trick and strategy so that we can decode the sequence here.

So, that is why I initially explain about replication and then I explain about polymerase chain reaction so that it will be easier to understand about DNA sequencing here also. So, as a result of that in this tube if we are carrying out the reaction, then in this tube you should have DNA polymerase. You should have this DNA template, whatever sequence I mentioned. You should have a primer also, but the only thing here is that you just need one primer or one primer sequence. You do not need two primers like for polymerase and the reaction. Here we are trying to determine the sequence of one strand; that is why. DNA polymerase will start adding nucleotides so then this primer will be extended. So now what will be the next nucleotide here? It will be T. T should come here, then A, then C, G; it will go like this, but if it is going like this, just like in polymerase chain reaction, we are synthesizing the whole DNA. But that is not our goal.

We need to find out the sequences of this region, we have to determine the sequence. So, as a result of that I am just erasing this part. So, this is not our goal. What we are going to do is we will add some additional reagent in this tube, some kind of I would say defective nucleotides, some kind of problematic nucleotides there.

So, whenever those nucleotides are incorporated during chain elongation, the chain will not be extended anymore; it will stop. So, that is the strategy. So, as a result of that, I would say, if in this tube, for example, this is my tube number 1. In this tube, in addition to all the reagents required for this chain synthesis, the DNA polymerase, magnesium ion, all dNTPs, everything is there. So, I have dNTPs, polymerase, then template, primer, and everything is there. In addition to that, I have some defective T, so some additional nucleotides, I am saying T*. That is the defective T, so I have all these four nucleotides: ATGC, but in addition to that, I added another nucleotide in this tube, that is, some defective T, some problematic T. Whatever is happening here, I am trying to explain. This is happening inside this tube. So then, the next nucleotide will be added here, the T. Now we have two options: one option is that the good T, the normal T, which is present in this dNTP mixture, can be incorporated here, or this defective T can also be incorporated here. If the normal T has been added here, then the chain will grow. But in case this defective T* gets incorporated, then you will have the T*, so it will not grow anymore; it will stop there because of this defect. Then, in this tube, I have many primers. So, as a result, if I take another primer here TACGCATC and then this good T, or the normal T, is incorporated. So this end will be extended. So, then the next nucleotide should be ACG, then again A, then again another T. So, maybe this T is the defective T. So, as a result of that, again, it will stop here. It will not grow anymore.

So, another primer, $5'$. I am not writing the primer sequence anymore. So, here this is the primer and then again TACGA normal T incorporated here. So again it will be extended ACG then again the bad T or the defective T. So, it will stop here. So, then the idea here is in this reaction in tube 1, I have every component in addition to that I have this defective T. So, as a result of that, whenever this defective T is getting incorporated, then the reaction stops. That chain cannot be extended anymore. Then as a result of that I am going to get it. In this tube after the reaction we will get different length products and all those products will end with the defective T at their $3'$ end. So, this is the result from tube 1, but here we have to be careful that the concentration of defective T should be very less. Generally, the ratio of this defective T should be almost 100 is to 1, 100 times it should be the normal nucleotide normal T and only 100 is to 1, only 1 percent should be this defective T so that at the beginning only defective T will be incorporated, then the reaction will stop, it will not proceed further, but that is not our goal. Our goal is to try to carry out the reaction as long as possible, so that we will get different length products and all of those products are done with T at the $3'$ end. Now I am just mentioning this tube 1 here, what will happen then if I have 4 tubes, like tube 1 already I mentioned and now tube 2, tube 3, tube 4, and in this tube 1 we have defective T and in tube 2, for example, defective A, tube 3 defective G and tube 4 defective C . So, as a result of that in this tube 2 we will have different in lane product. This is $5'$ to $3'$ direction and everywhere at the end what nucleotide should be there?

It should be defective A at the end because I have defective A in tube 2. So, as a result of that, now if we try to understand what is happening in all these 4 tubes, we will get different length products, all the possible lane products in these 4 tubes. Now, what will happen if we run a gel here? If we run a gel, this gel is a little bit different from agarose gel electrophoresis because our goal here is to separate DNA molecules if they differ by one nucleotide. As you can see, this DNA molecule here and the next one differ by just four nucleotides. Similarly, here, this one has just one extra residue of defective A compared to this DNA strand because after this, in this template, you have T, so that is why this is the smallest product in tube number 2. So, as a result of that, our goal is to separate these DNAs, these DNAs of different lengths. They differ just by one nucleotide. So, now we will be loading our DNA sample here in this gel. This gel is really made of polyacrylamide, not agarose gel, and some additional parameters have to be incorporated in order to separate those DNAs properly so that we can even see just one nucleotide difference. Now, the thing is, if we load the sample from tube number 1, tube number 2, so, the sample from all these 4 tubes has been loaded in this gel. Similarly, it will be electrophoresed from the negative to the positive direction. In which tube do we have the smallest product? Based on the template, if you see, actually, we have the primer same in every tube. So, we have the smallest product in tube number 1, as you can see. It was A here, and that is why this is the defective T. So, in tube 1, we have the smallest product.

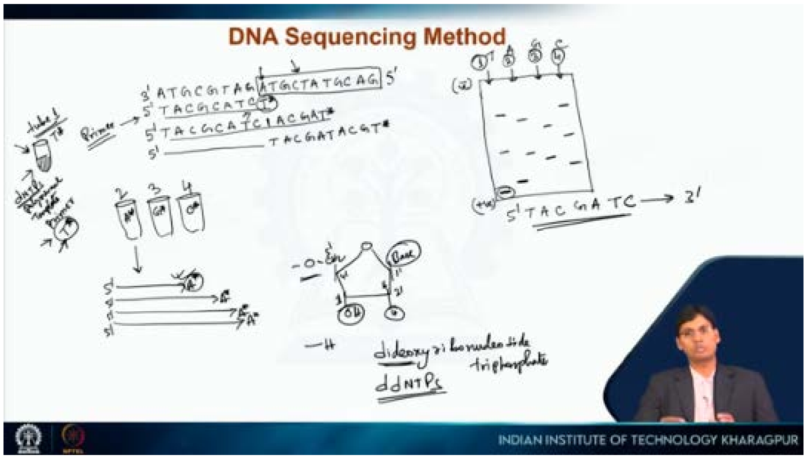
So, as a result of this gel run we will be seeing the smallest product here and then the next one you have T. So, that it should be defective A. So, you will be seeing something like this here and then C. So, defective C in is lane 4 so here then here. So, I am not following whatever is mentioned in the sequence in the template, but if I just put it randomly then we will be getting something like this. So, as a result of that, if this is the gel, then you can determine the sequence. From here, you will get the smallest product; it migrated maximum.

So, this is the first nucleotide. So, what is the first nucleotide? it should be the T because defective T was there in tube 1. So, the sequence will be $5'$ T then so you have here defective TAG and C. So, the sequence will be TA here this one CGATC is going on like this. So, that way from the gel itself you will determine the sequence. Now, here the question is what was the defect? I just mentioned that it is defective T defective A defective G what was the defect? The defect was here nothing, but a small change in the nucleotide.

So, as you know, this is the pentose sugar present in DNA. So, this is the $1'$, $2'$, $3'$, $4'$, and $5'$ carbon. Now, in DNA, you have OH here. That means in deoxyribonucleotides, you have OH at the $3'$ ends. This defect was a very small change; instead of OH, you have just H. So, if you have H, then what will happen? Whenever this nucleotide is incorporated, it will not be extended because, as I mentioned before during replication, this OH helps in the nucleophilic attack.

So, the next nucleotide can be added. So, as a result, if you replace this OH with H, then we would say this is dideoxyribonucleotide triphosphates. I am not showing the phosphate part here so dideoxy because you do not have oxygen here at the 2′ position. This is the dNTP analog, but here also, you do not have the OH, that is why dideoxy. Both oxygen atoms are missing here. So, that is why we mentioned this as ddNTPs, or dideoxyribonucleotide triphosphates.

So, as a result, you can understand that defective T means you have the dideoxythymidine. So whenever it is incorporated, it will not be extended anymore. This is the trick, and this was developed by Frederick Sanger, which is why he received the Nobel Prize. So, he discovered this principle and this method. So, now if you see here, whatever I tried to explain, you will also get almost the same thing. So, as you can see, this is the DNA that needs to be sequenced, and this is the sequence we are mentioning here. But after the reaction, you will get the sequence. Just for explanation, it has been shown.



DNA Sequencing Method

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

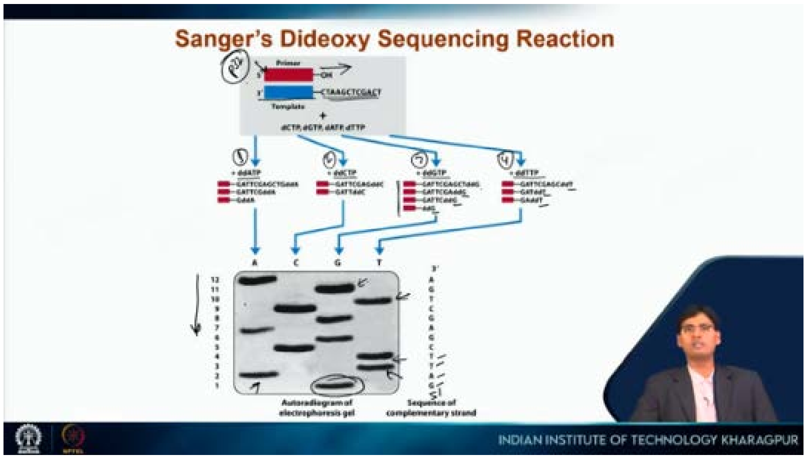
Now, you have this primer, and this primer will be extended in this direction. So, now, here you have these tubes: number 1, 2, 3, and 4. In tube 1, you have dideoxy ATP and dideoxy CTP; here in tube 3, dideoxy GTP; and here, dideoxy TTP. So whenever this reaction happens, you will get this sort of product here. So, you can see all these products are done with the G residue at the end. Similarly, here the T residue is at the end. Now, if you analyze this through a gel, as I explained, you will see something like this sequence.

Here, the migration is from this to this direction. So, as a result of that, this is the smallest product. So, this is G, as you can see. So, as a result, you will get this sequence: from 5' GA, because this is present here A, and then TT, so the same sequence is present here. So, that way, you will get the sequence.

So, this is how we determine the sequence of a particular segment of DNA. Now, in this method, some important features I should mention. At least a few of them I should mention. The amount of DNA is very low here, but how will I observe that? So, because of that, we use some kind of radioactive isotope. This primer here, this 5′ end is labeled with P32.

This is the radioactive phosphorus. It is leveled with that. So, as a result of that, after this gel run, although we have very small amounts of product in that gel, you can do some additional experiments so that we can put in some extra film and then we can develop that film. So, all this DNA molecule is like whatever you can see; this band is getting extended from that primer, and each primer contains this radioactive atom, radioactive phosphorus.

So, as a result of that they will because of this radioactivity put some kind of impression on that extra film, and that is what you are getting here. So, just to make this thing more sensitive, we add this radioactive phosphorus with the primer, and you can analyze this gel in this way. So, this is all about the idea of the sequencing method. This is Sanger sequencing, commonly known as. Now a little bit of modification of this technique is very robust, and generally we say automated DNA sequencing.



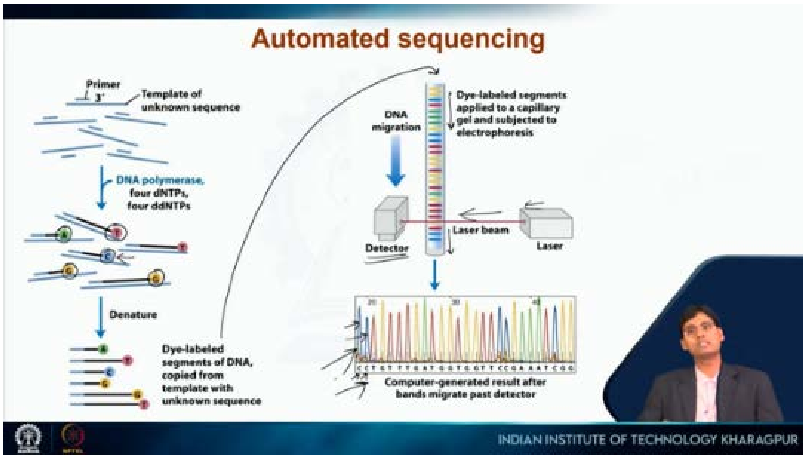
Sanger's Dideoxy Sequencing Reaction

This is nothing; but the same, whatever I explain, the basic principle is the same. The only thing is a little modification instead of the radioactive thing here we use some fluorophore. So, we do not add radioactive phosphorus with the primer here; rather, here this dideoxyribonucleotide triphosphate, that ddNTP. So, they are actually labeled with four different colors. For example, as you can see here, you have the dideoxy T in a color mentioned here, and dideoxy A here is also green; dideoxy G and here dideoxy C are in blue color.

Just four different fluorophores should be added with this defective ATGC that is all. So, if we do that, actually we can carry out this reaction in just one tube; there we are doing it in four different tubes to analyze that in the gel, but here since all these four nucleotides. These dideoxy nucleotides are labeled with different colors. We can carry out the reaction in the same tube. The reaction will be happening in the same tube. So after the reaction is done, you will get all possible different-length products in a single tube itself and then whatever product you get after the reaction, you have to load it on a small or narrow capillary tube. This is called capillary gel electrophoresis or capillary electrophoresis. Some matrix is present inside this capillary, and now again, this DNA will migrate from top to bottom in this direction. This is just capillary electrophoresis, the same principle as electrophoresis. Now again, the smallest product will migrate faster just like gel electrophoresis.

So, as a result, when these products are coming out from the capillary tube, here you have one detector and here you have the laser source. So, the laser will go in this way, and the detector will detect which color is passing. So, now during this separation in the capillary tube, as you can see, even the single nucleotide difference, those DNA can be separated, and we can plot that data in this way, this kind of curve. So whatever this color is getting detected so it will plot here. For example, this blue color peak here, consecutively two blue color peaks. What does that mean?

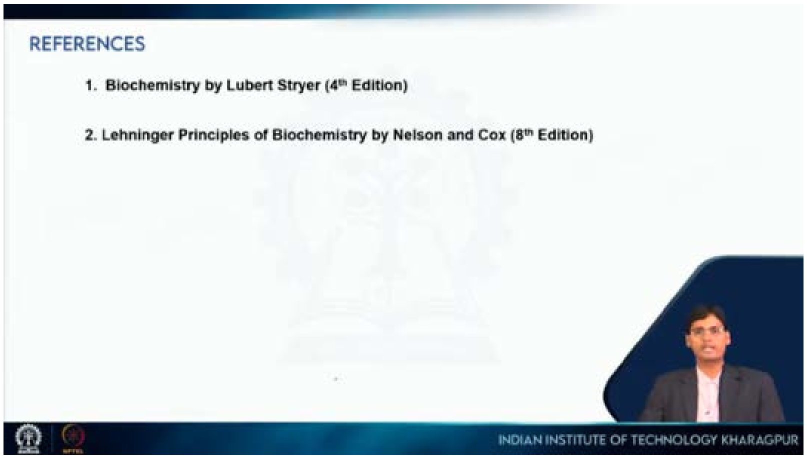
That means, as I already mentioned here, the CC that dideoxy C here is colored with this blue color fluorophore, for example. So, that means here you have the C residues coming. So, that way continuously these DNA molecules come out from the capillary tube, the detector detects them, and you get this kind of chromatogram and this kind of peak of different color, and from there you will get the sequence of the DNA present in the templated cell. So, but this is very powerful because, as you can understand, this is a very efficient process. We can carry out this reaction in just one tube, and you do not need to use any radioactive thing and that kind of complicated gel analysis and all those things, but this is a very efficient method. So, we can sequence a larger fragment of DNA at a stretch and that is all.



## Automated sequencing

Primer
3'
Template of unknown sequence
DNA polymerase, four dNTPs, four ddNTPs
A
T
C
G
Denature
Dye-labeled segments of DNA, copied from template with unknown sequence
DNA migration
Dye-labeled segments applied to a capillary gel and subjected to electrophoresis
Detector
Laser beam
Laser
20
30
40
Computer-generated result after bands migrate past detector
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

You can follow any of these textbooks for this sequencing part. That is all. Thank you very much.



REFERENCES

1. Biochemistry by Lubert Stryer (4th Edition)

2. Lehninger Principles of Biochemistry by Nelson and Cox (8th Edition)