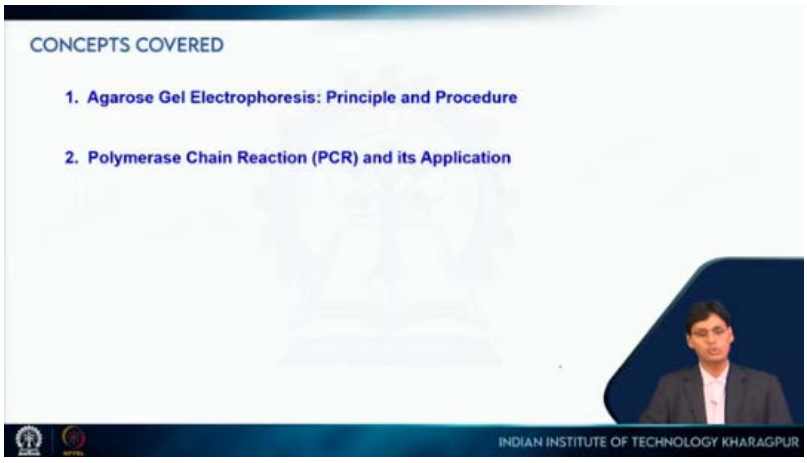
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur

Lecture 51
Visualizing and analyzing nucleic acids: Gel electrophoresis and PCR

Welcome back to the NPTEL online course on Introduction to Complex Biological Systems. Today we are on Module 11, and here I am going to discuss genetic engineering and its impact on human well-being. Particularly in this lecture, Lecture 51, I am going to discuss two major techniques: agarose gel electrophoresis and polymerase chain reaction, which we generally use to visualize and analyze DNA. So during the first week and second week, we discussed nucleic acids, particularly DNA and RNA, and based on their dimensions, you can understand that it is really difficult to visualize DNA.

So, here, since we are going to discuss genetic engineering, we have to engineer genes or engineer DNA. So, somehow we have to cut and paste DNA, but even before doing that, we have to understand how we can visualize DNA molecules. How will I understand which DNA is longer, which DNA is shorter, and many more details about DNA, particularly what different sequences are present in a particular segment of DNA? All those things we need to understand before doing genetic engineering. So, here, mostly I will concentrate on agarose gel electrophoresis, its principle and procedure, followed by polymerase chain reaction.



So, what is gel electrophoresis? As I told you, our goal here is to visualize DNA. Now, agarose is a polymer, particularly a linear polymer made up of a disaccharide called agarobiose. So, if I say what is agarose. It is a linear polysaccharide extracted from algae, and particularly it is made up of agarobiose, which is a disaccharide. So, now this component or this agarose is insoluble in cold water. So, what we do is, we put it in water, not really water, but some kind of buffer, and then we warm that. At higher temperatures, this agarose will be solubilized, and if we cool it down to around 40 to 45 degrees centigrade, it will solidify again. So, then we will get some kind of jelly-like structure, which is the agarose gel. So, before cooling down, we have to pour this solution into a mold so that we will get some kind of gel, and whenever it cools down, it will form a hydrogel. Depending on how much agarose you add, you will get different densities of this gel. So, the thing is, when it is getting solidified, we will get something like this. So, this is the agarose gel here.

The blue color you can see here is not really blue; it is just shown like this for representation. So, agarose gel, and now this is some kind of matrix. So, it has some pores, and our goal is to load our nucleic acid sample, particularly we analyze DNA samples through agarose gel mostly. So while we are making this gel, we put some kind of comb, and when it gets solidified, then we take out the comb; so as a result, we have this sample well. Now you can see this sample well for example, if I put some numbers: one, two, three, and four. These four are the wells, so some kind of depression on the gel so that we can load our DNA sample into these wells, and then let us see what happens. I will discuss that. So, if we put our DNA sample in this well, for example, lanes 1, 2, 3, and 4, and then we are actually connecting this system through some power supply in such a way that this side here is the negative terminal and this is the positive terminal. So, as you already know, nucleic acids, both DNA and RNA, are negatively charged due to the phosphodiester bonds between two nucleotides. So, because DNA is negatively charged, it will migrate from the negative to the positive terminal. Now, based on their size, the DNA molecules will separate accordingly.

The reason is that all DNA is negatively charged, but now in this gel it has some kind of pore size. So, smaller DNA fragments will migrate faster. They will come faster towards

this positive end. So, on the basis of that we can separate different size fragments of DNA. So, this is the major goal of agarose gel electrophoresis. So, we can really analyze different length product of DNA molecule. So, if we do this experiment, but still I understand that whatever I explain, we will be able to separate DNA molecules based on their size. But then how can we visualize that? We have to bring some method so that we can understand where this DNA is present, so that we will be able to understand which one is the smaller and which one is the bigger fragment. In order to do that, we use some dye, so the dye is called ethidium bromide. So, if we add this dye to the DNA sample, then ethidium bromide will get intercalated between the DNA bases. If you consider the DNA double helix and now those bases are forming some base stacking.

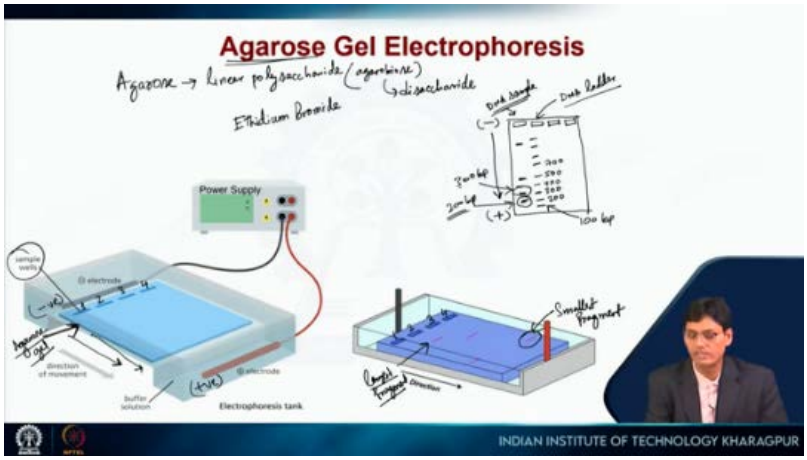
So, one base stack with another one is going like that. So, in between two bases it will go inside. So, it will be intercalated in between the bases, this ethidium bromide, and the property of this ethidium bromide is something like this when it is getting intercalated within the DNA. Then, if we expose this gel on some UV Transilluminator, that means, some small machine where you have the UV light source, if you put the gel on top of that UV light source, then because of the ethidium bromide present in DNA, it will fluoresce. So, generally we use around 300 to 350 nanometer wavelengths of UV light, and then some energy will be utilized, and then it will emit a higher wavelength.

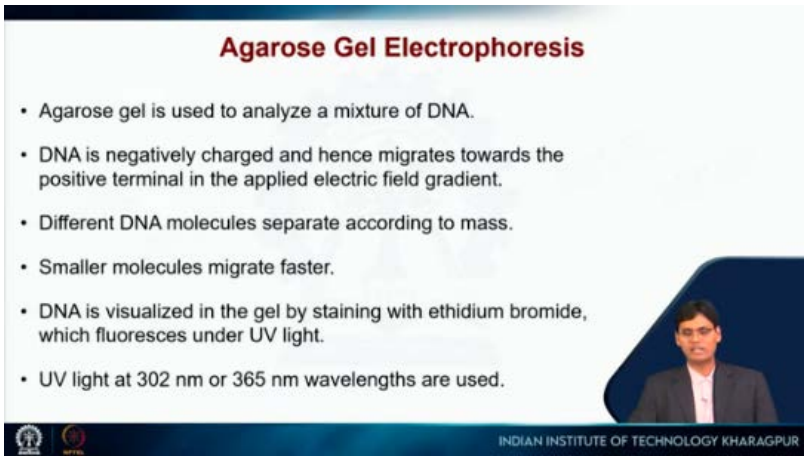
So, we will see something like pink color of band. So there, our DNA is present because this is coming because ethidium bromide gets intercalated in DNA. So, wherever this DNA is present we will see some kind of band, something like that. So, as you can see after running this gel. I am talking about T this one. So, after running this gel if we put this gel on some UV light source or UV Transilluminator, then you can see 4 bands here in 4 different lanes, whatever DNA sample is loaded. Based on this result, we can interpret that those DNA are different in terms of their size. So, for example, in lane number 4 you can see this DNA it migrated maximum towards the positive end. So, this is the smallest fragment in comparison to the other 4 DNA molecules we are analyzing here, and this is in lane 2, this is the largest fragment of DNA as it migrated minimally compared to the other 3 DNA molecules presented here. So, on the basis of that we can analyze DNA molecule. So, the thing is here in these four lanes we have loaded only

DNA sample. We have just one lane product, but it is also possible like this is an agarose gel and here, for example, I have one lane here. So we can make many more lanes also on the basis of the comb, what kind of comb we are using. I am loading here some DNA sample, not just one DNA present here; it is a mixture of DNA.

So after loading it, if we carry out agarose gel electrophoresis, then again, DNA molecules will be separated on the basis of their size. So, they will move from the negative to the positive electrode. So, then, for example, if I am getting these four bands in this sample, that means four different types of DNA molecules are present in my DNA sample. Then, how will I determine their length? Very frequently, what we do in the lab is run another sample here at the same time. This is a DNA ladder, which consists of known DNA fragments. For example, I have 10 different DNA molecules of different lengths, and I know their lengths. How many nucleotides are present; I already know that.

So, now, if I get something like this, for example, these are the different fragments present in this DNA ladder and I know that this one is 100 base pairs, and then this is 200, 300, 400 500, and 700, something like that. It is going on like that. So, as a result of that, if I know the length of these fragments because this is in our control, we are loading this DNA ladder. So now I will be able to tell that this DNA present in my sample is 200 base pairs, and the next one here is 300 base pairs. So, this is the technique, the agarose gel electrophoresis, by which you can really separate different DNA fragments, analyze them, and interpret their lengths and all the information you will get from gel electrophoresis. This is absolutely required. It is a very important technique we frequently use during recombinant DNA technology or genetic engineering. So, I will discuss it in more detail whenever I discuss the genetic engineering part. Now, whatever I mentioned, some of the information is just listed here. It will be helpful for you, and then, the next topic: polymerase chain reaction.





In the case of agarose gel electrophoresis, whatever DNA sample you are going to analyze, you have to load it on that gel, and then we have to conduct this electrophoresis experiment, and after the experiment, we have to see it on some UV Transilluminator to understand what kind of DNA fragment is present in our DNA sample. But now the problem is sometimes the DNA sample itself is very low in amount. The amount is very less. So neither we can see on gel nor we can proceed for the DNA sequencing part. So then what will we do? So, this is the technique where we can actually amplify the DNA amount. Whatever DNA is present with us, we can increase the amount of that particular DNA by this method, and this method is called polymerase chain reaction. So, polymerase chain reaction.

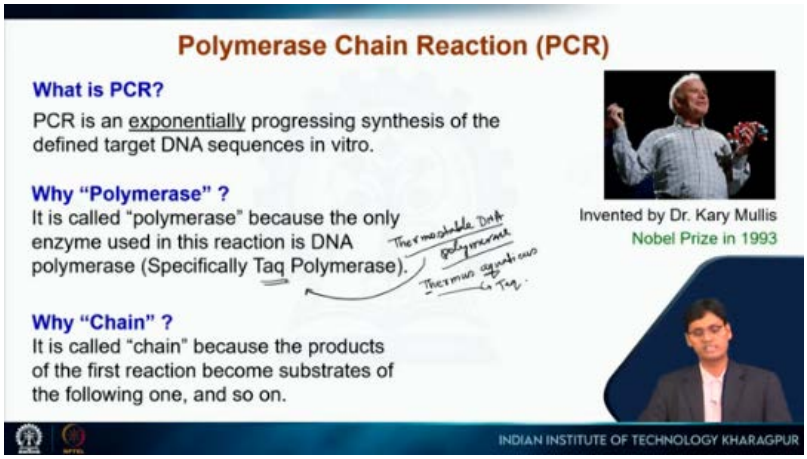
So, this is an exponentially progressing synthesis of the defined target DNA sequence in vitro. So, basically I would say this is very similar to replication, but this is happening inside a tube in vitro. So, we are just synthesizing DNA in vitro, but that is also a defined

target. It is in our control which segment of the DNA we are going to amplify. Now why the name is polymerase? It is called polymerase because the only enzyme used in this reaction is DNA polymerase. As you can remember, I just mentioned here that this is almost like replication but this is happening inside the tube. But during replication, there are many enzymes involved, as you can remember, like different types of DNA polymerase, topoisomerase, helicase, and many enzymes are involved there, but in this system, in this reaction, we just need one enzyme. That is DNA polymerase, but this DNA polymerase is some kind of special DNA polymerase. What is the specialty? This DNA polymerase is thermo stable. So, thermo stable DNA polymerase. So, one example is Taq polymerase. Here you can see Taq polymerase. So, this Taq polymerase is nothing but a DNA polymerase, but it was isolated from some bacteria that are naturally present in high temperature, maybe in a hot spring. So, now, in their system, they are used to carry out all reactions, all metabolism, at very high temperatures. So this enzyme, the Taq polymerase, why this name Taq polymerase? Because the bacteria's name is Thermus aquaticus. So from Thermus this 'T' came and from aquaticus this 'aq'.

So, together, this is Taq polymerase. So, this is just the DNA polymerase, but it can work at very high temperatures because it is extracted from some thermophilic bacteria. The idea is, if we can use this enzyme, we do not need helicase or topoisomerase. At very high temperatures, the DNA double strand will be separated, and this enzyme can work. I am going to discuss this in more detail in the next slide. Now, why 'chain'? It is called 'chain' because the products of the first reaction become the substrate of the following one, and so on. So, as a result, you can exponentially synthesize DNA from a single DNA molecule, starting at the beginning.

So, that is why this is called a chain reaction or polymerase chain reaction. Now, the polymerase chain reaction and its application. As I already mentioned, the major use is amplification of a desired DNA segment. Whatever our interest, on the basis of that, we can amplify that segment through this reaction, this polymerase chain reaction, but what are the goals? There are huge utilities of this reaction in day-to-day life also. Because, as you can understand, for example, forensic teams in forensic science use genetic fingerprinting to identify criminals, for example, from a crime scene, the forensic team

may have isolated a little bit of sample, maybe some body fluid or some small drop of blood or something like that from a crime scene, then, if they have extracted DNA from that sample, but that DNA might be very low in amount, neither can you do sequencing to find out the identity of the criminal.



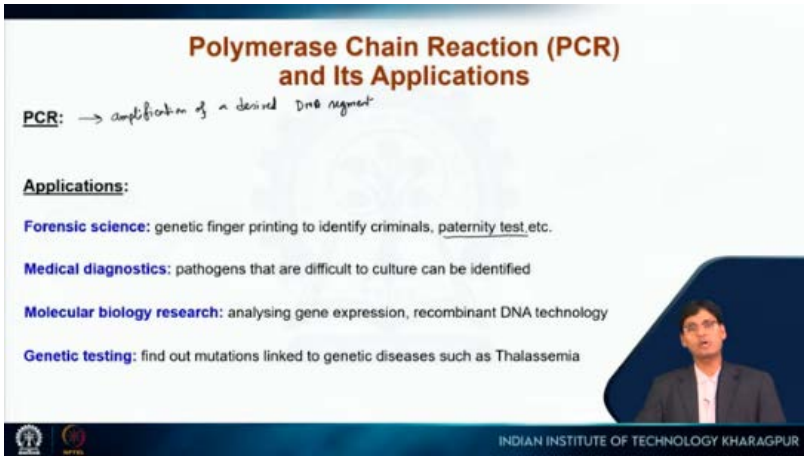
So, as a result of that, what we can do is use that small amount of DNA as a template and amplify it through this PCR technique. So, we will have a huge amount of DNA, and that DNA we can further analyze to get all sorts of information about whoever was involved in this crime scene, and so on. Similarly, PCR can be used for paternity tests also in forensic science. Many times, forensic teams do this, and this is a very powerful technique. We can also use it in medical diagnostics. For example, some pathogens, like certain bacteria, if they are very difficult to culture in the lab, then what can we do?

We can take a sample from the patient and use that sample again through polymerase chain reaction to amplify the DNA. Finally, based on sequencing and different downstream techniques, we can find out what kind of pathogen is involved in this condition. So, that is why it is a very powerful technique in medical diagnostics also. Similarly, polymerase chain reaction is used routinely in molecular biology research. For example, in our lab, we use it on a routine basis almost every week. Many times we do this polymerase chain reaction to analyze gene expression.

Sometimes to make different types of DNA constructs in recombinant DNA technology for protein expression, purification, and many other things we do, PCR is the backbone of many of these techniques. So, we rely on PCR on a routine basis. Also, genetic testing is

very important. Like PCR technique, we use it to find out mutations. Mutation means some changes in nucleotides that should not be present in normal conditions, but because of some disease conditions, some changes happen and we see the effect of that mutation, like some changes. We can find out those changes through PCR. We can amplify the DNA, then go for sequencing, and we can do that.

So for example, to identify thalassemia, thalassemia is some kind of disorder with our hemoglobin. It can be beta thalassemia or alpha thalassemia where the beta globin chain or the alpha globin chain of hemoglobin is involved. So, those mutations can be tracked through this PCR technique. So, as a result of that, what I want to say is that this polymerase chain reaction is a very powerful technique, and we use it for many different purposes. As you already know, in the previous slide I was showing that Kary Mullis received the Nobel Prize for inventing this polymerase chain reaction because it has huge applications in different fields of science and biomedical science.



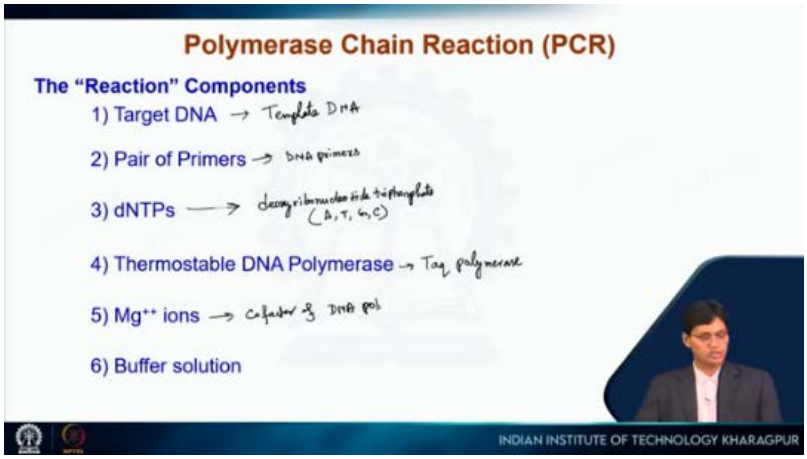
Now, the reaction components so, as I already mentioned this are very similar to DNA replication, but it is happening inside some tubes. So, it is an in-vitro reaction. So, as a result of that, in order to replicate something or to make something, we need to have the template DNA that is the target DNA. So, this is what we used as template DNA and now pair of primers because as we know that we need primers to start the DNA replication because DNA polymerase cannot start it from the beginning. So, as a result of that we need a primer. So, in this case, I will explain why we need two primers, just two primers, but also remember since we are doing this in vitro.

So, we are adding DNA primers here. So, this is oligonucleotide made up of dNTPs. So these are DNA primers not like RNA primers used in replication steps. So, their primase was involved, but here we are actually synthesizing primer or commercially you can buy primer from some company based on a particular PCR reaction.

These are DNA primers. So, you do not need to remove this primer again from what we have seen during the replication. Now, the raw material that is our dNTPs, this is deoxyribonucleotide triphosphate. So, dNTPs, but here all dNTPs should be there, all four like A, T, G and C and then thermostable DNA polymerase that I have already mentioned, this polymerase's special feature is that it will not be destroyed in high temperature.

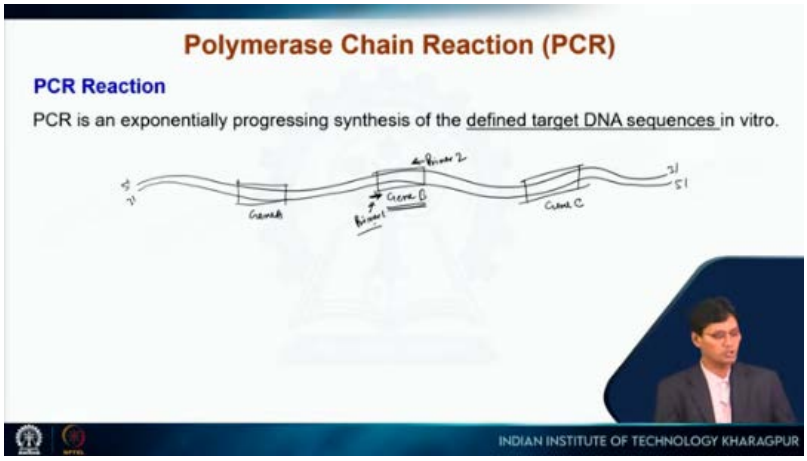
So, it will be fine even at 90^0C. So, thermostable DNA polymerase one example is Taq polymerase, which I have already explained. Then magnesium ion is the cofactor of DNA polymerase.

So, magnesium ion is also required for this reaction, and finally, a buffer solution is needed to maintain the ionic strength and the pH of that reaction. So, that reaction can go smoothly, that is all. So, these are the components we need to mix together, and then we have to carry out the reaction. So, let us see what reaction steps are involved here. So, as I already mentioned, PCR is an exponentially progressing synthesis of the defined target DNA sequence.



So, a defined target DNA sequence in vitro. So, here, I would especially like to mention or explain why I am mentioning the defined target. So, for example, this is one long DNA molecule. So, now this is 5′, 3′, 3′, 5′, a long DNA molecule. Now, you have one gene here: gene A, gene B, gene C. So, now maybe I am interested in this gene B so I do not need to amplify the whole DNA segment here. So, what I can do is design primers in such a way that they will amplify only this gene B. As a result, I will be using primers here so that it will bind here. This is one primer, and this is another primer so primer 2 and here, primer 1.

So, it will bind in this segment then B will be amplified. So, as a result of that, I have the control over which segment I would like to amplify. So, that is why we can actually increase the amount of a defined target or defined portion of the DNA. That is why this is very powerful. I do not need to synthesize the whole DNA molecule. Now I will mention particularly what are the steps involved here.



So, as you can see here at the beginning, I have this DNA molecule here; this is double-stranded DNA. You already know that double-stranded DNA means it is hydrogen-bonded, and two strands are hydrogen-bonded together. Now, in the first step what we have to do is, in this reaction tube, just imagine that in the reaction tube, whatever I mentioned, the components are all there inside the tube. Now, in the first step, we have to increase the temperature of the tube. So I would say we can increase it up to 92^0C. As a result of that, what will happen is this DNA.

The template, which is double-stranded, will be separated into single strands. The hydrogen bonds will be broken because of this high temperature. So, as you can see, this is one DNA strand here, and another DNA strand is getting separated so now again, in the next step, we will cool down the temperature of the tube. So, in the next step, if we cool down, then I would say the temperature is around 55^0C, depending on the sequence of the primer, but I am just mentioning that for your explanation. So, what will happen? In the reaction tube, I have an excess amount of primers.

So, now, at lower temperature those primers will bind here as you can see in these two colors I am showing here. So, for example, this is 5′ 3′ DNA. So, this is 5′ 3′ DNA. So, this primer now will bind here and it will go in this reaction because of the DNA polymerase or Taq polymerase in this case and this primer 5′ to 3′ in this direction, it will grow.

Now this DNA polymerase or Taq polymerase, specifically its optimal temperature is around 72^0C, at that temperature it will work best. So, as a result of that again we will increase the temperature. So I would say here that the temperature in this step is 72^0C. So, in this temperature the DNA polymerase will start adding specific nucleotides based on the template sequence. So you can see that this stand has been synthesized from 5′ to 3′ directions.

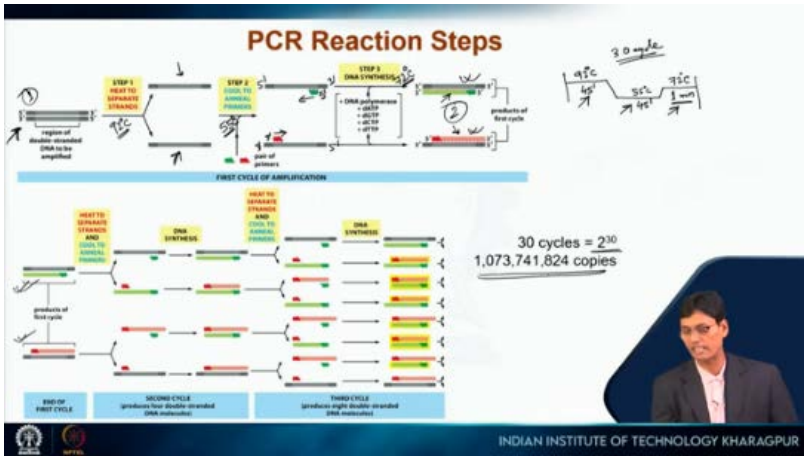
We started with one DNA molecule here, now we have two DNA through this method. So now if I summarize this PCR cycle, that means nothing but just different types of temperature, three different temperatures we are using, and on the basis of that we are carrying out our reaction if all the components are present inside the tube. So the first thing was the high temperature. So here, I would say the temperature was 92^0C; it can be 92^0C, 93^0C, and 94^0C that does not matter. So, now for example, I am putting here 45 second time and then again we are lowering the temperature a little bit, say 55^0C, and here again 45 second. Then we are increasing the temperature a little bit, as I mentioned that 72^0C temperature is the optimal temperature for the Taq polymerase.

So, here again, maybe 1 minute so here, the DNA will be synthesized. This time, this 1 minute can vary depending on how long the flanking sequence is? That means, what your

target sequence is and how big it is. On the basis of that, you have to decide how much time you will provide here, whether it is 1 minute, 2 minutes, or 3 minutes. You have to decide that. So, now, these are the three different types of temperatures we use for the PCR reaction. So, now, in the first step, the two strands will be separated. Here, in this step, the annealing temperature will allow the primer to bind to your DNA molecule, which is already separated. Then, in the third step, the polymerase will add nucleotides, and the DNA will be synthesized. So, this is just one cycle of the reaction.

Now, if we continue this cycle for, say, 30 cycles, then what will happen? Whatever these two DNA molecules we have after one cycle, they will again act as the substrate for the next cycle, and then we will get 4 DNA molecules in the next step. So, exponentially, we can increase the number of DNA molecules. So, as a result, just to explain here: if you start there, then after the first cycle, you have 2 DNA molecules, and then after the second cycle, you have 4 DNA molecules, and it continues like this. So, now the thing is, after 30 cycles, this will be the number, 2^{30}.

So, this is the theoretical number of DNA molecules. So, as you can understand, that's how many DNA molecules are present within a short period of time. This is why this technique is very powerful. We use this technique to amplify DNA molecules or increase the amount of DNA, which can be used for further analysis. As I already mentioned, this is a very important technique in forensic science.



That is all. You can refer to any textbook.

REFERENCES

1. Biochemistry by Lubert Stryer (4th Edition)

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

3. Essential Cell Biology by Alberts et al., (4th Edition)