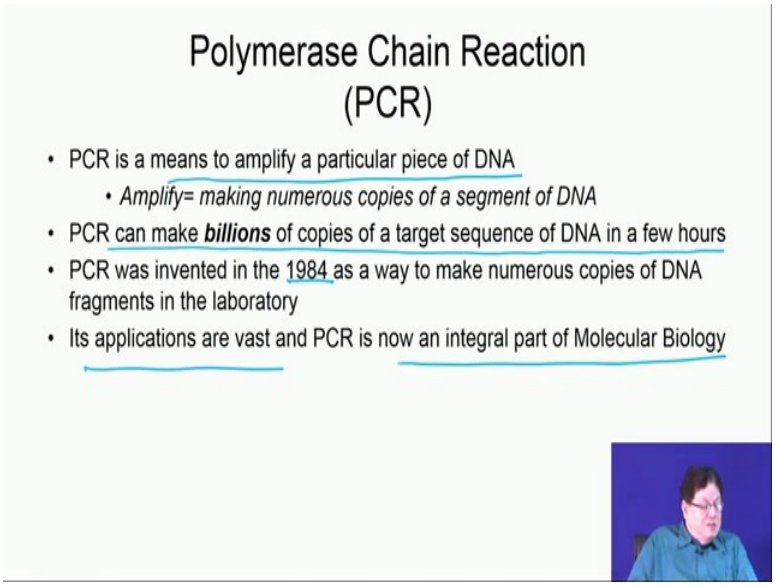
Organic Chemistry In Biology And Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture – 34
Molecular Biology (Contd.)

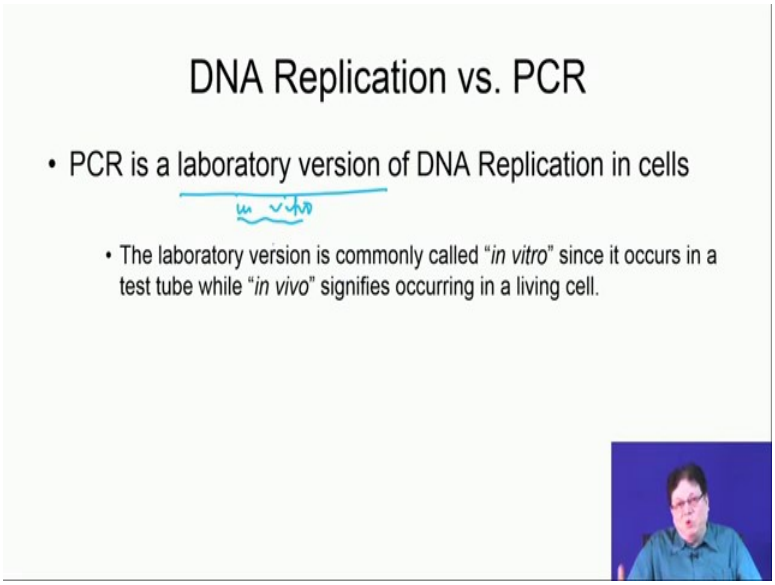
(Refer Slide Time: 00:39)



In the last session, we have discussed the gene cloning technique utilizing recombinant DNA technology *i.e.* the r DNA technology. Now, we will discuss a completely in vitro method of amplifying a piece of DNA by PCR. Amplify means making numerous copies or clones.

PCR can make billions of copies of a target sequence of DNA in a few hours. So, it is very useful. It was invented in 1984 by Dr. Mullis. He received the Nobel Prize within after 5 years of this discovery. Every molecular biology lab will have one or many more PCR machines because this is fully automated now.

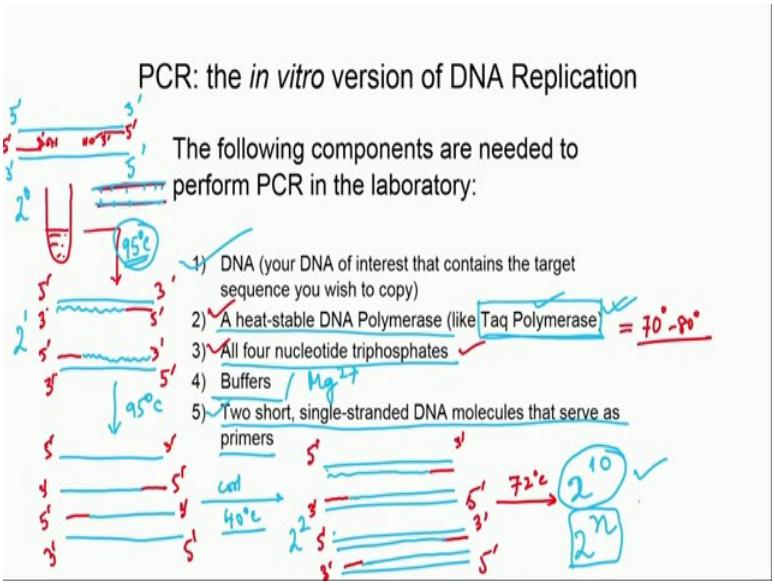
(Refer Slide Time: 02:12)



How do you do this amplification? As I told you it is a laboratory version. It is an *in vitro* version of DNA replication in cells. In last lecture, I have discussed how recombinant DNA technology can be utilized in the cells to make copies of the DNA and to make enough of the substantial amount of proteins that you require.

In PCR, you make copies of the DNA but you do not get the proteins out of it. Here the DNA is in the test tube. It is not inside the living system. So, you do not get the proteins. It is just a way to make the DNA. Remember in r DNA technology, you can make the copies of the DNA as well as proteins also because it is being amplified inside the bacterial cells.

(Refer Slide Time: 03:20)



In PCR, you need the DNA that you want to amplify. Then it requires a heat stable DNA polymerase like Taq polymerase. DNA polymerase makes poly oligonucleotide in 5'-3' direction. This enzyme works at a higher temperature.

Usually our proteins work in the biological system at 37 degree. It is the optimum temperature. If the temperature goes above 50^0 C the enzyme slowly loses the activity. Here it is a heat stable DNA polymerase like the Taq polymerase.

Now, these enzymes are isolated from volcanic regions or hot spring where the temperature is very high. So, if some bacteria grows there those bacteria must be having and producing enzymes which work at a higher temperature. So, somebody must have thought that in the extreme conditions you can get enzymes which work under extreme conditions. Like heat is one example, similarly at very cold temperatures you get the cryo enzymes which work under cryo conditions. So, this is we are talking about a heat stable DNA polymerase.

All the four oligo nucleotide triphosphates are mandatory requirements for synthesis of DNA. As we are doing it in vitro all the four nucleotide triphosphates are required. You have to add buffer and magnesium. When polymerase works the 3 prime OH attacks the 5 prime triphosphate and then the pyrophosphate is released. Magnesium takes care of

the negative charges through chelation. These two short single stranded DNA molecules serves as primers.

These short double strands are called primers. We have seen the use of primers in Sanger's dideoxy sequencing method. Remember RNA polymerase does not require any primer. RNA polymerase can work only on a single strand but DNA polymerase requires a small segment of double strand in order to make the complementary strand.

I need to take this in a Eppendorf and then add primers.

Now, what are these primers? The primers bind to the 3 prime end. This is binding at the right most corner and this is binding at the leftmost corner. It has to bind from the 5 prime to 3 prime and then the DNA synthesis takes place by attack of the 3 prime OH because here 3 prime OH is free.

Here also 3 prime OH is free. If you add nucleotide triphosphates they will be taken up one by one depending on the sequence here. The reaction will proceed and you get the DNA piece. So, you add the primer and then you also add a heat stable DNA polymerase whose optimum activity is usually between 70 to 80 degree centigrade.

Taq polymerase works very well at this temperature. But there is a particular temperature usually 72 degrees maintained by the machine. You have this piece of test tube or Eppendorf and then you take the buffer, the magnesium, the DNA, 4 oligonucleotide triphosphates, a heat stable DNA polymerase and the primers. So, initially the primers will not join because it was basically the DNA was present as the double strand.

So, in order for the primer to bind you have to heat it. The heating is done till the temperature reaches 95 degree centigrade. 95 degree centigrade means all the DNA that are present in the universe will melt at this temperature.

So, these DNA will melt now. This is my original DNA that will form single strands. Now you have the primers also and primers are given in excess large excess. If you do not add the primers then they are going to self-anneal with each other. The primers will hybridize with these single strands that are separated by heating at 95.

If you cool down to about 40 degree the primers will join at the two ends and then you heat it again to keep the temperature between 70 to 80 degree. The Taq polymerase t

works very well at 70 to 80 degrees. Now you get two copies of the double stranded DNA. This is called one cycle. What is the cycle? Cycle means you heat everything to around 95. It becomes single stranded, then cool it to around 40. So, the primers bind to the single strands and then you heat it back to about 72 degree centigrade. Then the polymerase will complete the extension of the chain. So, this is the cycle.

Then what you do? Then again you heat it to 95 degree centigrade. If you heat it to 95 degree centigrade these 4 strands will separate. Now you cool to about 40 degree. The temperature may not be very accurate, but the science is very clear. You heat it, make it a single strand, you cool it so, that it hybridizes with the primer and then you heat it to about 70 degree so, that polymerase can complete the elongation of the chain.

All the DNAs will have the primers. This is your 5 prime end and this is your 3 prime end. The first one is 5 prime to 3 prime, the second one is 3 prime to 5 prime, the third one is 5 prime to 3 prime and this last one is 3 prime to 5 prime. That is your original piece of DNA. When it is cooled the primer will bind to the 3 prime end. This is 3 prime to 5 prime so, now the primer will bind here. And this is 5 prime to 3 prime so, the primer will bind here and this is 3 prime to 5 prime so, the primer will bind here.

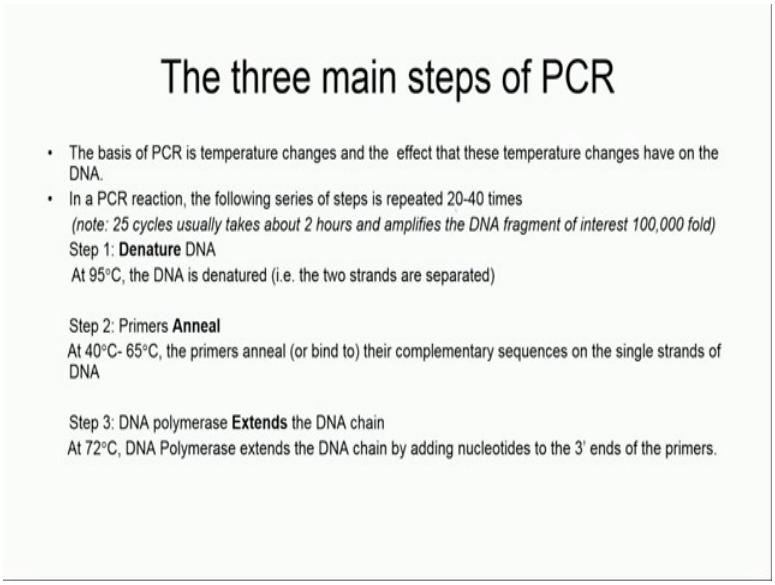
Now, what you will do? You again heat it to 72. Now this will be your piece new piece of DNA and this will be your new piece of DNA. After two cycles you have $2^2 = 4$ strands of double stranded DNA.

So, now if you do n number of times you will get 2^n number of strands. If you do it 10 times it will be 2^{10} number of strands. You copies of double stranded DNA within a couple of hours. Most interesting the breakthrough came in the polymerase chain reaction process is the discovery of this Taq polymerase. If the temperature is kept at 95 the enzyme loses its activity and it becomes denatured. You cannot really get back the original activity that was a problem.

After the discovery of Taq polymerase this whole thing can be automated. Earlier before the discovery of Taq polymerase you have to heat, then after cooling you have to add the DNA polymerase in every cycle because the polymerase will lose its activity and denatured at 95.

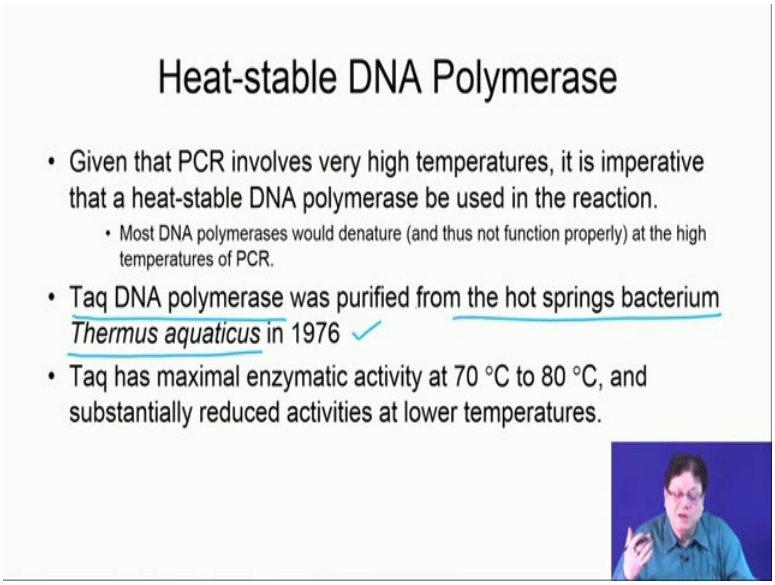
But once this Taq polymerase was discovered it survives at 95 degree centigrade. It survives definitely at 45, 40 degree centigrade but its optimum reactivity is at between 70 to 80. Depending on the number of copies, you just do this number of cycles. So, that becomes $2^{n.}$

(Refer Slide Time: 17:40)



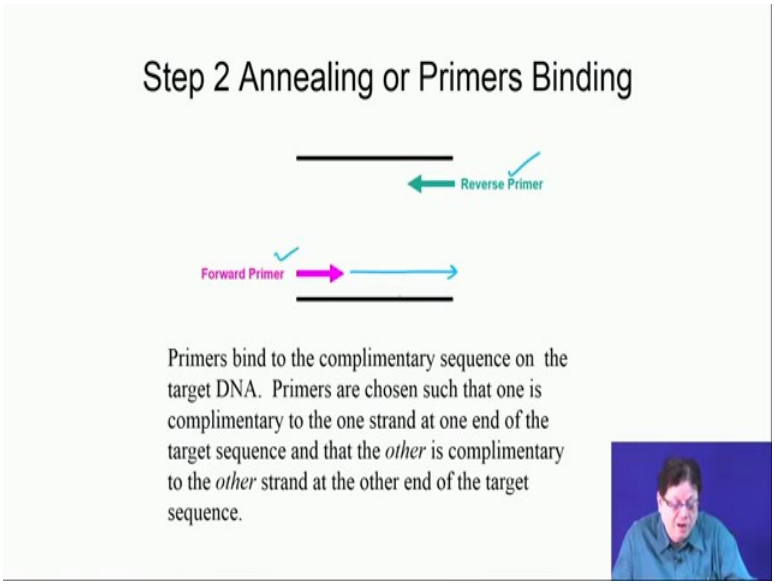
So PCR reaction is repeated usually 20 to 40 times. 25 cycles usually takes about 2 hours. So, 2 to the power 25, so, 100000 fold you increase. Now these are done by a machine called thermocycler.

(Refer Slide Time: 18:38)

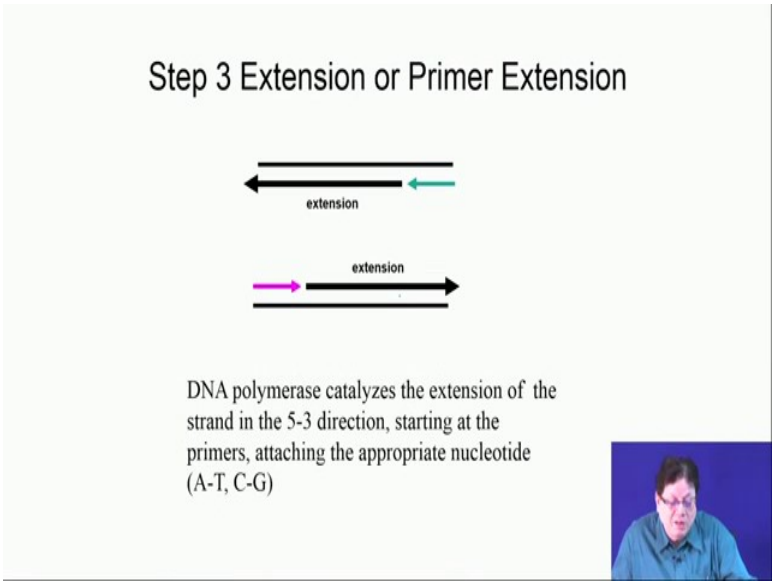


So, you can change the temperature and can fix it at particular temperature. These thermo cyclers are extremely good in maintaining the temperature. You have to go 95 then quickly you have to drop to 40 and then you have to take it to 72. So, that can be done in a machine called thermo cycler. Actually the Taq DNA polymerase was purified from the hot spring bacterium thermos aquaticus in 1976 and that actually led the foundation of this automated DNA polymerase chain reaction.

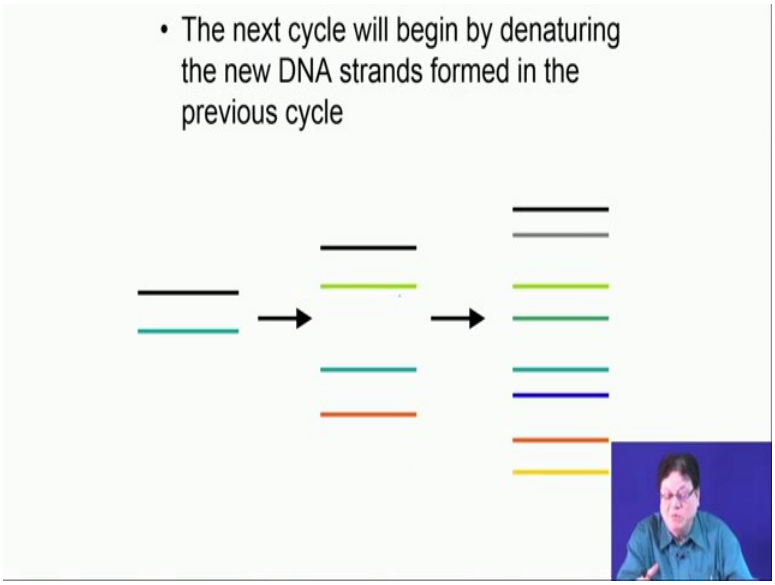
(Refer Slide Time: 19:04)



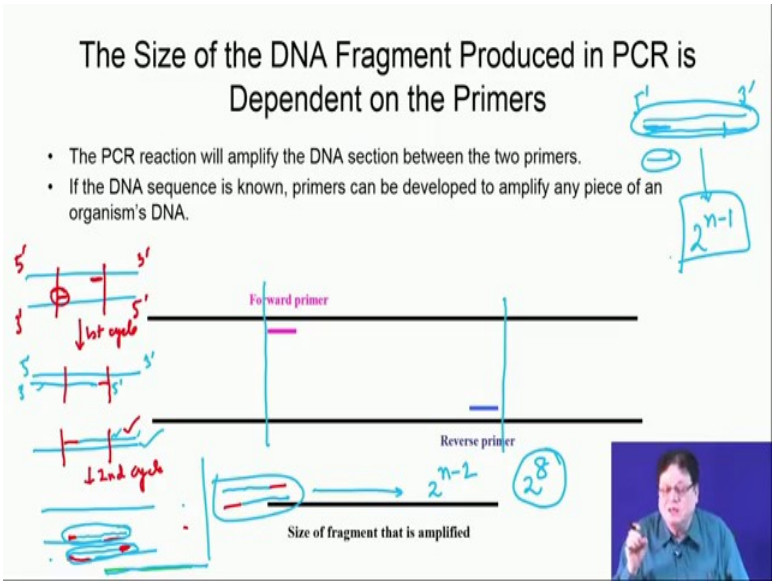
(Refer Slide Time: 19:24)



(Refer Slide Time: 19:26)



(Refer Slide Time: 19:29)



By the way, this is called the reverse primer because the extension goes in the reverse direction of the DNA. This is called the forward primer because this goes in the forward direction. Now, suppose you have a DNA and you are interested only to copy from this region. So, suppose this is your 5 prime end this is your 3 prime end. Now if I want to copy only this region I have to use a primer which recognizes this part. Because then only copying can be done at this zone. I do not need a primer from this side and from this side then, I will get the entire piece of DNA as the copy. The original piece of DNA is this one and then you have the red portion which started from this one and then that goes

up to the end here. So, this is the piece of DNA that you will get and from the other strand you will get copy. You really do not get your copy that you wanted. You have extra one. This is your region of interest but you are getting extra after the 1st cycle.

Now, come to the second cycle. So, what will happen here? When that will extend up to this point and similarly for this part, the primer will be attached somewhere here and when this is extended that will form up to this part. So, then from the 2nd cycle onwards you are getting one double strand. So, this is the other strand again contains only the original strand. The other strand will be the normal strand containing the entire piece.

So, again I repeat, the first cycle you have this is the primers where that will be attached. You have to design the primers according to the sequence here and according to the sequence there. In the 1st cycle we are seeing that you do not get the DNA only containing your zone of interest. You get extra. This is extra on this side and on the other side this is extra. In 2nd cycle, when you melt it from this strand you will get the actual the zone of interest. From the other strand, you get the piece of interest.

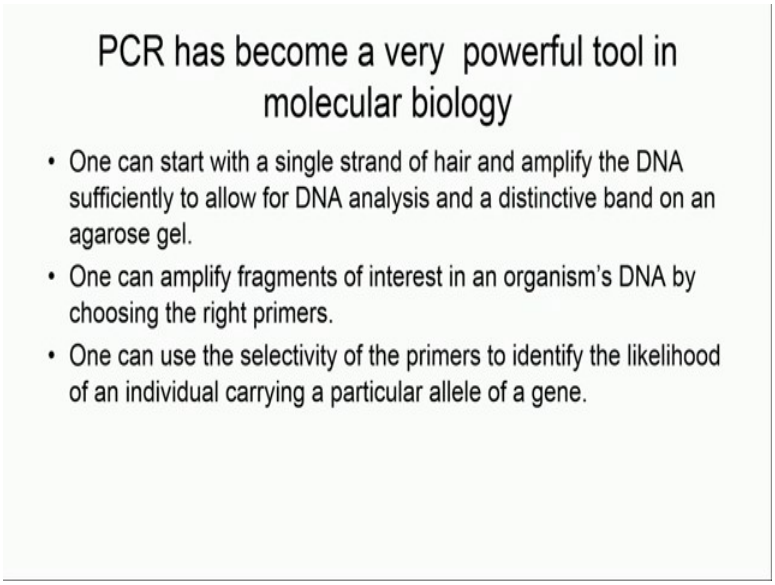
So, this is your piece of interest and that is your piece of interest. After the second cycle, this is your starting point. So, that is your starting point. You have to complete two cycles in order to get the first copy of the double strand DNA that containing the region of interest. So, from then onwards you will get only this part. That will be copied because all the primers will either bind here or bind there. That will be extended up to the zone of interest.

The number of copies of the zone of interest after 10 cycles will be 2^{n-2} because you have to first complete 2 cycles in order to come to a first strand of the double strand containing the zone of interest. So, after 10 cycles you will get 2 to the power 8 copies. If you have only one strand of the DNA which you want to copy then how many copies you will get after 10 cycles.

First of all, can we apply PCR to this or not? Yes, you can apply you just make a primer that is the reverse primer. You also can write forward primer for the complementary strand. So, you add both these primers, the single strand and you heat it and then cool it. This primer will remain free because it did not have the complementary strand to start.

Now, you have a double strand. One cycle is required to make the double strand and from then onwards you will get the double strand because now this primer has the complementary strand so that it can bind here. Now, you will have basically 2^{n-1}. It is possible to do PCR reaction on a single strand. Here first cycle is needed to make the double strand and from then onwards you can copy that DNA.

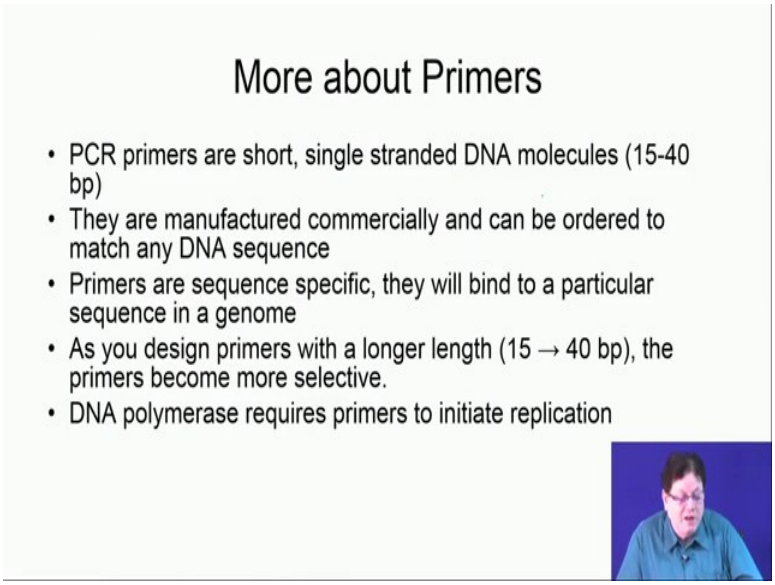
(Refer Slide Time: 27:33)



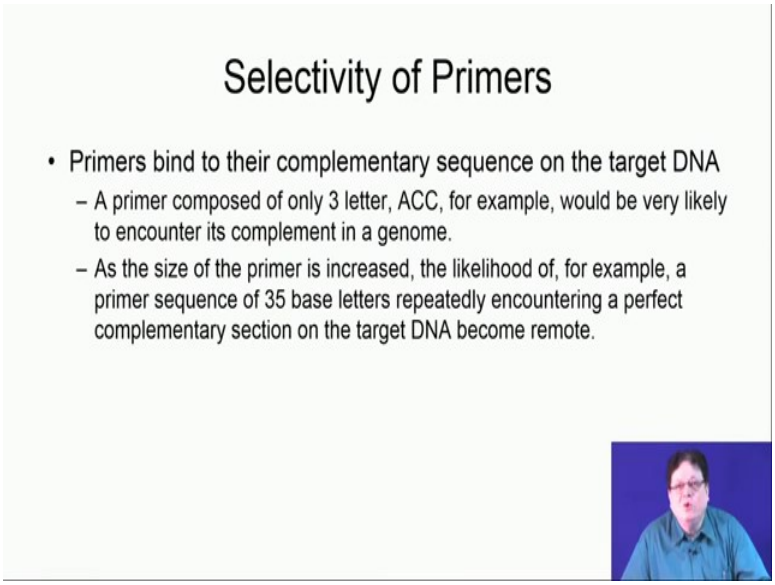
What are the usefulness of PCR? PCR has become a powerful tool in molecular biology. One can start with a single strand. Suppose there is a homicide somewhere and you want to identify who is the murderer. Now usually a hair of the murderer can be isolated at the crime scene, but from the single piece of hair is very difficult to do the DNA analysis. So, for that you have to amplify the DNA.

So, you isolate the DNA and do the PCR and then from the PCR you do the DNA sequence analysis. Then you compare with your suspect DNA. You have a group of suspect and from there you can take the blood and then you isolate the DNA. Then you can again do the analysis of the suspect and from that you can make out who is the likely murderer. That is usually not much uncertainty. However, it has been found that it is 1 in a billion that there could be a chance of misidentifying the murderer.

(Refer Slide Time: 28:41)

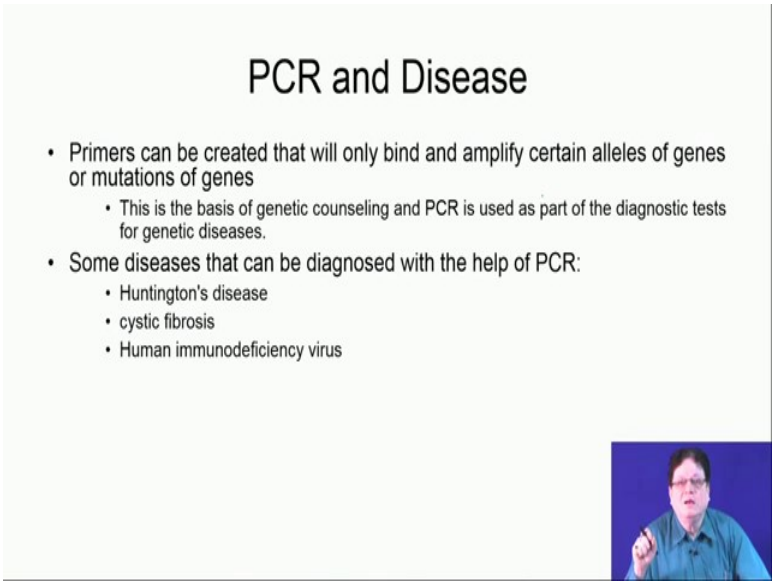


(Refer Slide Time: 29:21)



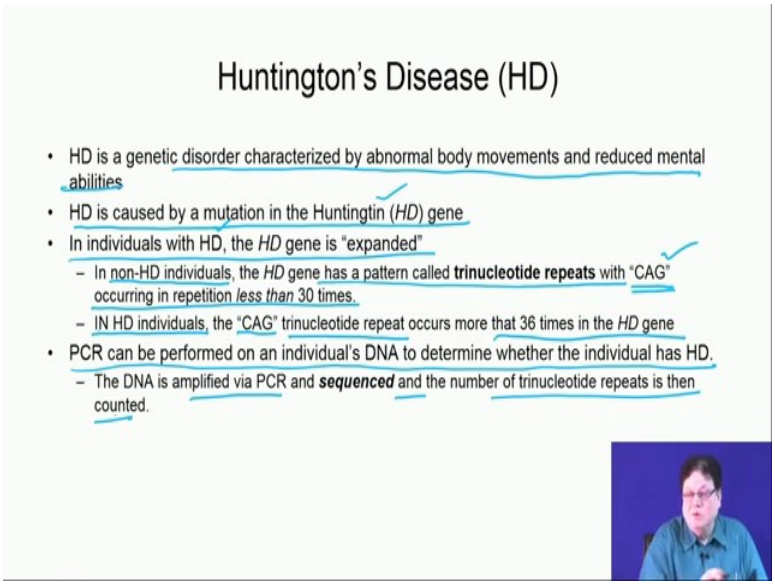
So, In forensic where the very small amount of DNA is available and only the sequence of DNA may not be enough to identify the murderer. There are certain repeat of base sequences which is present in every individual. So, if you can identify those repeats and then compare, that gives a better way of identifying the person doing the homicide. So, not only the DNA sequence, you have to see the short repeats that are present in every individual which vary from individual to another individual. So, that gives you more rigorous way of identifying the culprit.

(Refer Slide Time: 30:39)



There are many genetic diseases like Huntington's disease or cystic fibrosis and there are some viral diseases which remains dormant for a long time. It is like the HIV, the Human Immunodeficiency Virus where the viral load means the amount of viral DNA will be very tiny in the biological fluid of the body system. So, some of these genetic disorders can be detected with the help of DNA like this Huntington's disease.

(Refer Slide Time: 31:27)



Huntington's disease is a genetic disorder characterized by abnormal body movements and reduced mental abilities. In this case, mental function is defective and also there are

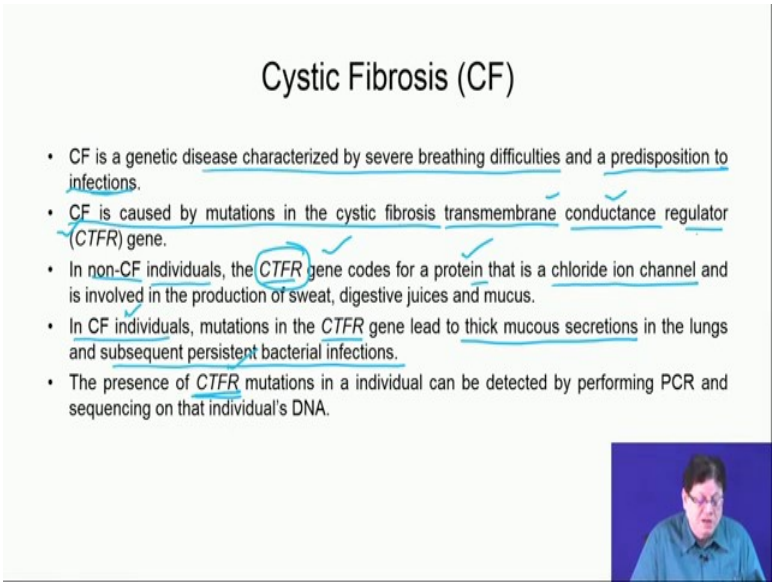
abnormal body movements. We have seen babies unfortunately born with Huntington's disease, HD. Huntington gene is expanded.

Huntington gene is having a repeat of CAG at regular intervals. The people who are not suffering from this disease or this abnormality are called non HD individuals. For non HD individuals repetition number is less than 30.

In HD individuals the CAG trinucleotide repeat occurs more than 36 times. If the repeat of number is more than 36 you get these Huntington's disease. If it is less than 30 you are perfectly normal.

So, it has now incorporated originally when the baby is born. It is a genetic defect. You isolate the DNA and then amplify by a via PCR and see the number of repeats of sequence. If the number of repeats is greater than 36 then that individual will have Huntington's disease.

(Refer Slide Time: 34:15)

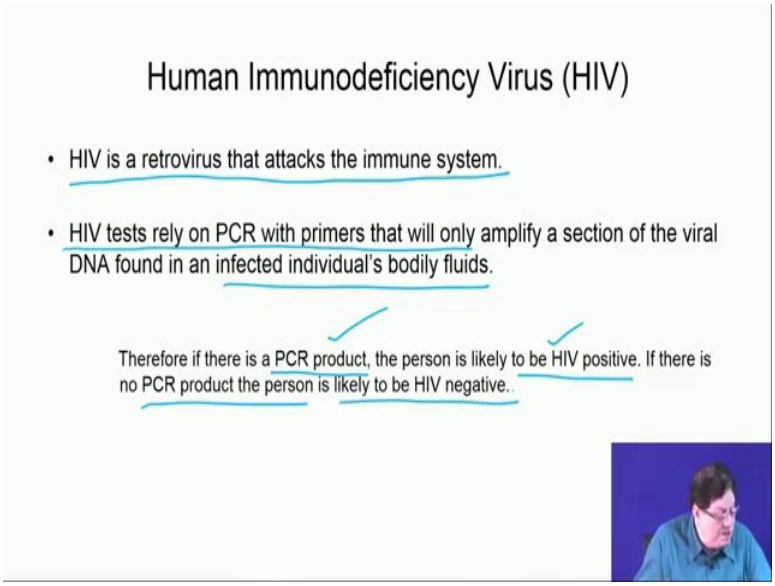


Another cystic fibrosis is a genetic disease characterized by severe breathing difficulties and a predisposition to infections. So, they suffer from infections very frequently. CF, the cystic fibrosis is caused by mutation in transmembrane conductance regulator gene called CTFR. It is a specific gene where there is a mutation. In non CF individuals the CTFR gene codes for a protein that is a chloride ion channel. That means it makes a channel through which the chloride ions move into the cell or out of the cell. That is very

important for transmembrane conductance. That is extremely important migration of the chloride because that is a charged anion.

The CTFR gene that is called the cystic fibrosis transmembrane regulated gene conductance regulator gene. If it is normal then it expresses a protein which creates a chloride ion channel. Mutation in CTFR leads to thick mucus secretions in the lung and subsequent persistent bacterial infection. Again, you can check whether there is any mutation by comparing with the healthy individual versus a CF individual and then see whether there is any mutation in the CTFR gene.

(Refer Slide Time: 36:24)

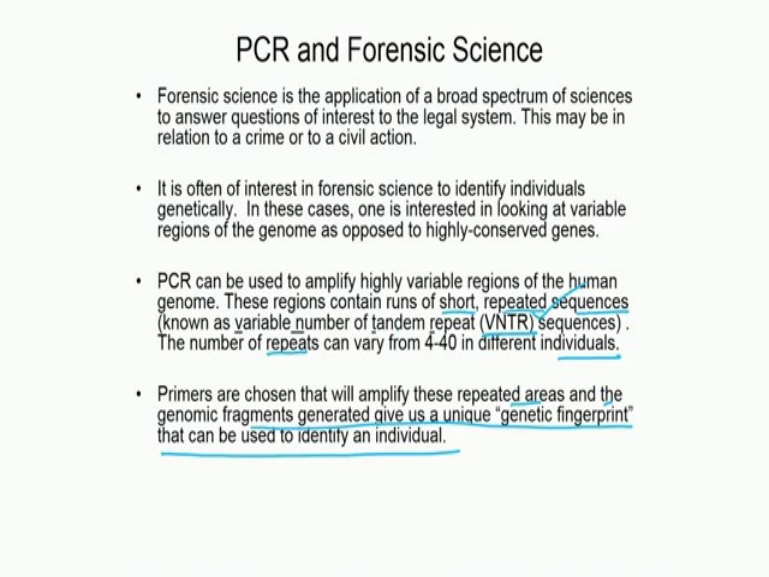


The HIV human immunodeficiency virus does complete destruction or reduction of immune response. So, it basically destroys the immune response of a person but it stays in a very dormant state for a long period of time. HIV is a retrovirus. We will talk about that how it attacks the immune system in medicinal Chemistry part. If the viral load is very little then tiny amount of viral DNA will be found in the infected individual's body fluid.

In spite of having tiny amount of DNA, it can be multiplied using PCR with addition of the right primer. You know what the HIV gene is. So, you know what the primer is. So, add those primers to body fluid to do the PCR.

PCR product is generating by the primer that you have added. There must be the viral DNA. If this viral DNA is coming the person is likely to be HIV positive. If there is no PCR product the person is likely to be HIV negative.

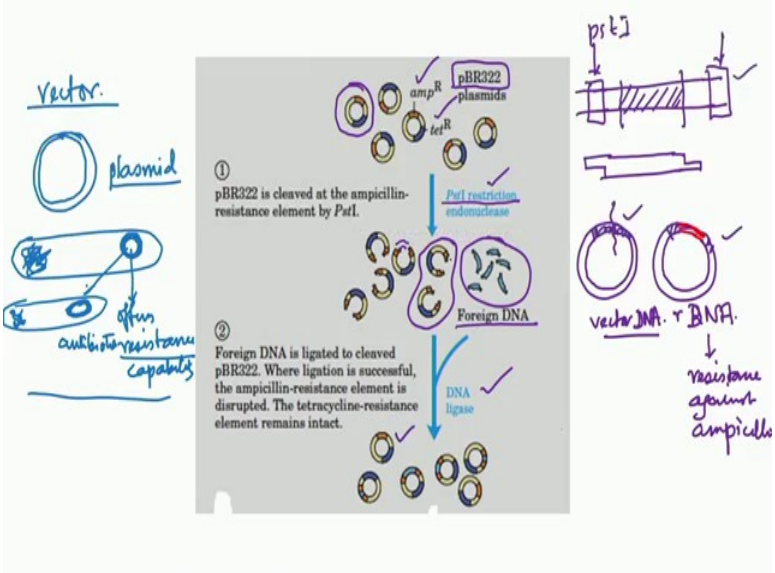
(Refer Slide Time: 38:25)



These are the application of PCR in forensic science. There are short repeated sequences known as variable number of tandem repeats, VNTR. This VNTR can vary from 44 to 40 in different individuals. Primers will amplify these repeated areas and the genomic fragment generated gives us an unique genetic fingerprint that can be used to identify an individual. You have to use this VNTR. The variable number of tandem repeats sequences will give a genetic fingerprint.

This is a fingerprint of the gene. So, that is why it is called a genetic fingerprint. Fingerprint means a method of identification of an individual. So, through the genetic fingerprint we can now match the genetic fingerprint of the blood or the piece of hair that we obtained at the crime scene and compare with the genetic fingerprint of the suspected persons and by that you can tell who the culprit is.

(Refer Slide Time: 39:48)



So, that is I think we have now discussed the r DNA technology and polymerase chain reaction. Polymerase chain reaction is an in vitro process. It is a very rapid one by which you can multiply DNAs and then it has got many utilities. On the other hand, r DNA technology gives rise to the protein. It is very important. If your target is protein you apply the r DNA technology and then you amplify the DNA and get the protein out of it.

So, that completes our nucleic acid. We covered structure of DNA RNA, the processes, the melting temperature, the flow of information. Then we have studied all these in vivo and the in vitro process of the multiplication.