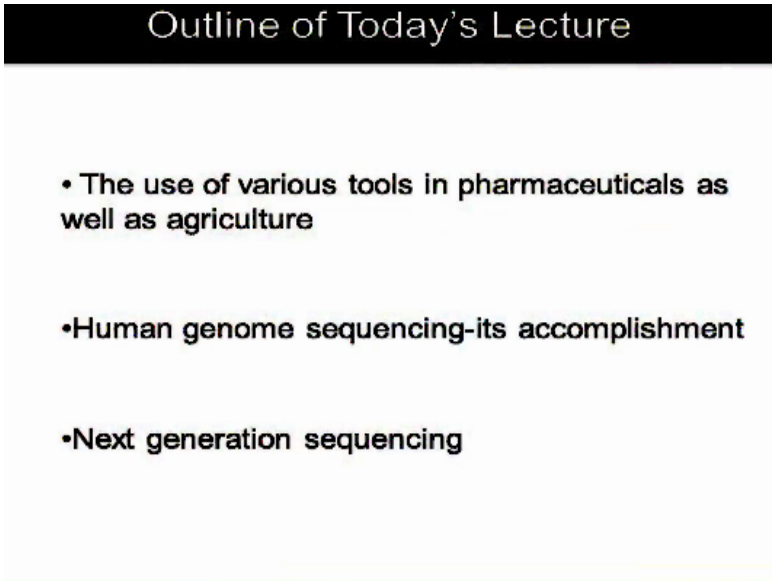
Bioengineering: An Interface with Biology and Medicine
Prof. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology – Bombay

Lecture - 11
DNA Tools & Biotechnology-III

Today, we are going to continue on some of the biotechnology tools which we discussed last class.

(Refer Slide Time: 00:20)



So outline for today's lecture is we talked about polymerase chain reaction in last class. We discussed about the gene cloning. We looked at different vector maths. Continuing to that we will talk some of the sequencing technologies as well as we will talk about how to do the protein production and which way many type of gene expression analysis can be performed. So those things we are going to talk in biotechnology tools.

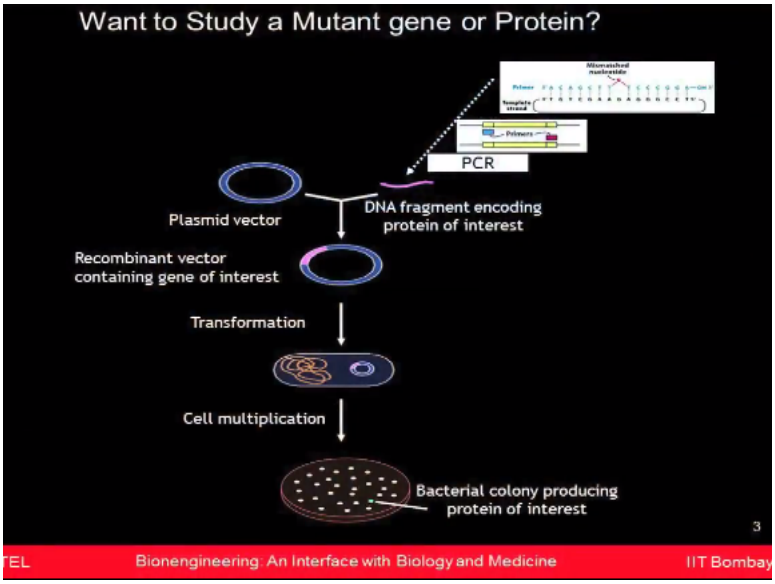
Latest field in the overall biology area is moving towards looking at all the biomolecules together because now we realize that if you are studying one gene at a time or one protein at a time probably you are not able to get the right picture of the whole system. So to know a system for example human system or any of the model organism if you want to study them well probably you need to study that whole system as a whole.

And therefore you need to know all the biomolecules and their complete properties and that is the kind of field which is known as omics, where you are looking at all the gene of an

organism like genomics or all the proteins of your organism like proteomics. Something of that kind you have seen in the beginning where scientists are now aiming to look for all the proteins present in the given system which is really much more complex, much more tedious even as compared to what you already know from genomics.

So we will talk very briefly about some of those latest developments in the field. So if I go back in the last lecture, if you remember we were talking about gene cloning, you remember this workflow that we have a plasmid vector.

(Refer Slide Time: 01:59)



And then along with them if we want to fuse a gene of interest, you are making a recombinant DNA molecule and that recombinant DNA molecule because it also contains and vector contains some of the marker genes like antibiotic resistant genes. So you can select the right clone.

So after doing the transformation and moving this plasmid in the host in the bacteria, now you can select the right bacterial colony which is containing the gene of your interest in the plasmid and then that can be used to further multiply grow to make more copies of that gene as well as if you want to do the protein production you can use some of the expression systems and do the protein expression as well.

Let us think one step back. You have studied even PCR right so let us say gene of interest which you want to study that is present in very low amount. So first of all you can do a polymerase chain reaction to increase the overall DNA component of that. I am sure you

remember the primers which we talked in context of polymerase chain reaction right. The short nucleotide sequence which is going to have the complementary sequences to bind to the region of DNA which you want amplify.

Now let us imagine that think about Morgan's experiment right. So we were talking about wild type and some of the mutant phenotypes. So if let say you want to study a gene in its wild type and also you want to see that what will be the change in an organism if I make just one or two small changes in that gene sequence which is the mutant form of that gene. So to do that part then ideally you are making the changes in the gene sequence itself right.

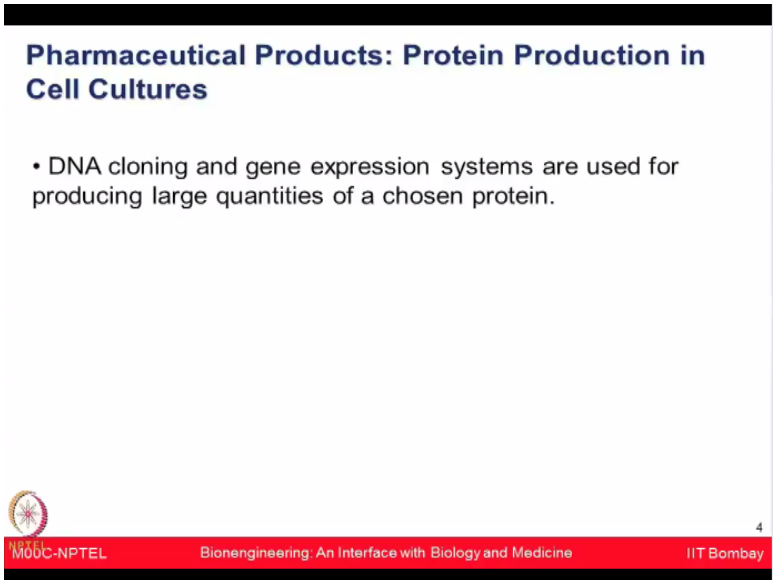
And to introduce that gene sequence change you can actually introduce that at the primer designing level itself. When you are designing the primer, you are selecting the short nucleotide stretch. In that short nucleotide stretch, if you make some changes in the primer sequence itself probably that can actually make the changes subsequently in the gene and then followed by in your cloning procedure.

So this is a place when people can introduce the changes and you can study the mutant forms of the gene or if you want to study the protein production from that, so you can study both gene or the protein, you can make the changes at the sequence level itself and then rest of the procedure remains the same. So this is I think important slide from many experimental and schematic point of view.

Because it conveys you the process of cloning, it conveys you the process of polymerase chain reaction and also conveys you how can you introduce some changes at the gene sequence level by designing the primers and then use those information to then further amplify the gene segment and then do the further rest of the cloning step in the same manner. This is what we talked in the last class.

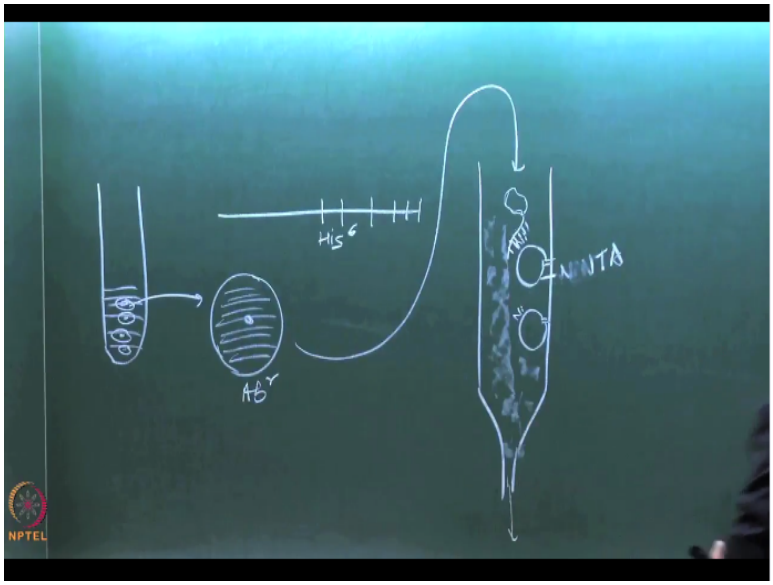
Let us now move on to see that you know there are many applications of doing a cloning, not only that you can make large amount of DNA but also you can do the protein production and that has huge value for lot of pharmaceutical products.

(Refer Slide Time: 05:23)



So DNA cloning and gene expression systems can also be used for producing large quantity of a chosen protein, just kind of you know for your curiosity let us imagine that you know how to use this expression systems for the protein production. So for example we talked about you know you are doing a cloning experiment and you can make multiple copies of that gene right.

(Refer Slide Time: 05:45)



And then you are growing that bacteria and you are multiplying that bacteria and you have now multiple copies of that bacteria which is having gene of your interest. So imagine this particular vector which you had used that was an expression vector which has the ability to induce protein production. So now once you have selected this bacteria on the plate here, you have the antibiotic resistant plates you know and let us say you can use any of the antibiotic resistance gene for selection.

So now you know this is for sure this is the colony of your interest. This colony you are further you want to use it for the protein production. Now at the time of a primer designing level if let us imagine that you had the short nucleotide sequence and in that sequence you had introduced certain nucleotide which corresponds to an amino acid histidine. So what people do?

People tried to use certain tags, introduced some tags for the protein purification process. So let us say if my gene sequence also has certain histidine tags here, which I have added at the primer designing level and people use sometimes 6-histidine tag continuously. So now whatever proteins are going to produce in addition to the protein which is going to come from the native protein structure you are also going to have certain histidine in addition to that.

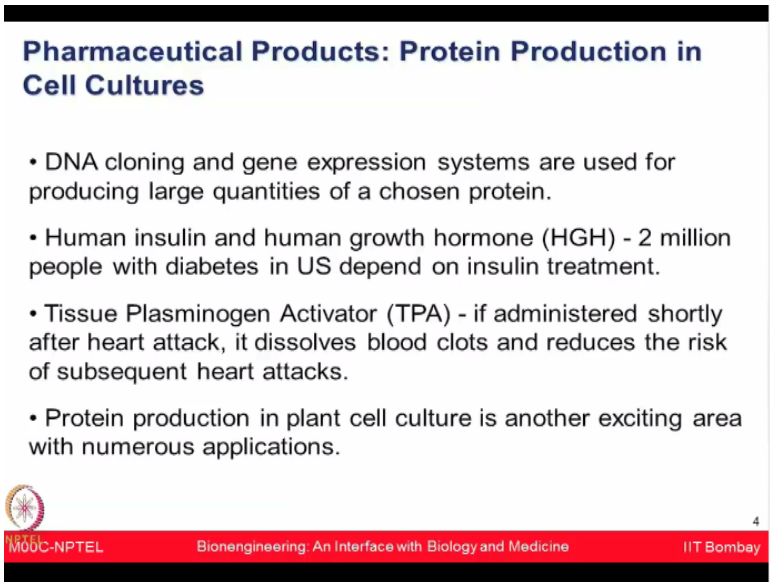
So now imagine a chromatography technique where you want to purify the proteins and which is another affinity chromatography technique. So let us say you have these columns and in these columns you have certain resins which are having Nickel NTA. Now you are having this particular bacterial lysate and that lysate you are passing it in this particular column.

This particular column has these resins which are having Nickel on their surface. Now Nickel binds with histidine with the coordination chemistry. So now all the proteins which are coming here, those protein which are having so let us say this is the protein, this protein which is having this 6-histidine sequence is going to come and bind to the Nickel beads. Rest everything else is going to just pass and get into the flow through.

So all the nonspecific things can be washed off and then those which are specific your protein of interest you can actually enrich those using affinity chromatography and then you can be purify those proteins further for your further analysis. This is one of the ways of doing the protein production. So concept is now you get it that you know once you have made a clone, once you have made multiple copies of the gene of your interest now you can also use that information to do the protein production.

And there are some very specific chromatography techniques; by using them you can actually have the large scale protein production.

(Refer Slide Time: 08:44)



So just imagine that you know people who are diabetic who need insulin every day, they need every day pure protein right. There you cannot you know compromise on the quality of that protein then because somebody needs every day that particular protein and the large number of diabetic patient whole world who require human insulin every day.

So as many people need different type of human growth hormones, millions of people are depending on these kind of treatments, so you need to produce these proteins in really, really large batches in large culture conditions and that is where chemical engineering a lot of reactors became very handy where the same concept which I talk to you can be done in a small chromatography level, can also be done at the very large you know reactor level when you can produce in the several litres or in gallons those kind of cultures you can grow.

And then from there you can do the protein production, so you need really, really large amount of protein and very high quality and purified protein for various therapeutic requirements. Another protein which is tissue plasminogen activator TPA, this protein is required for people who are having let us say heart attack, immediately after you have that sensation if this kind of proteins can be injected it actually helps to dissolve the blood clots.

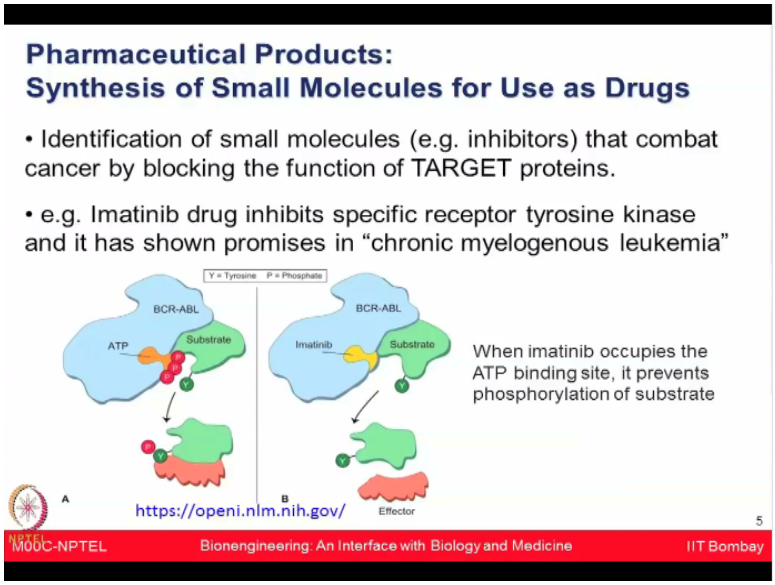
So then patient could be actually saved if this kind of proteins are available. So therefore you can just imagine that how useful these proteins are and you need to produce these proteins in really, really large amount and that is why many big pharmaceutical companies they are

producing lot of these kind of proteins which are having large requirements and they need large you know the quality control checks as well.

So protein production in the plant cell culture system is one of the area in the plant biotechnology and the biotechnology in general, which is kind of very much progressing and a lot of examples one could see in that area. Also imagine that you know people who are studying different type of drug targets, so let us say you have identified a protein now and you are developing some drug targets or the inhibitors.

So you have realised that you know for a given cancer patient for example there is some change happening in that one given protein and then if you can block that from some inhibitor probably you can control that activity.

(Refer Slide Time: 11:07)



So lot of companies they are developing different drugs and in this case if I show you this is the protein BCR-ABL. Now people have studied that you know how it binds with ATP and if it binds with ATP and the substrate binding happens then that results into the phosphorylation activity which kind of becomes very detrimental for these individuals. Now people have exactly identified the same pocket for the binding.

And an inhibitor imatinib has been designed which can go quickly and block that site and now once you block that site therefore the phosphorylation activity will not happen and then the whole of the cascade of the events can be stopped. So just by knowing these molecules properly and knowing that you know where is the site where you can do the binding or the

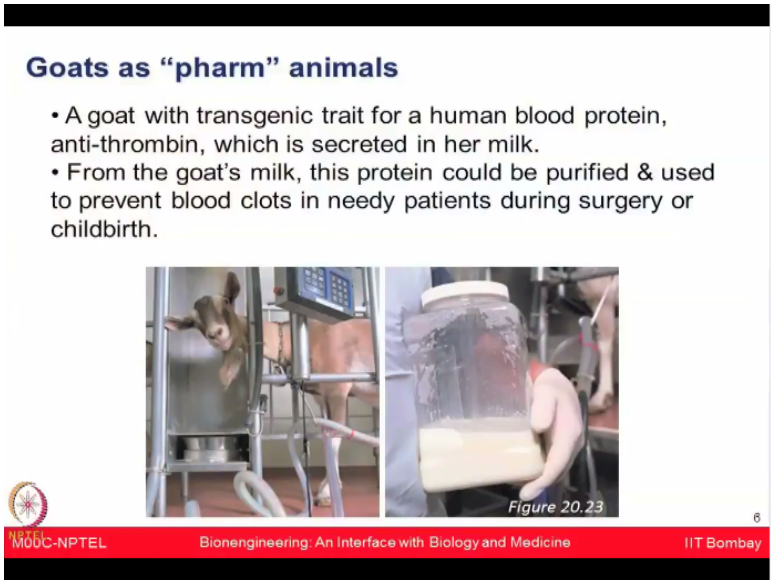
docking and then you can develop some of the drugs which can mimic that exact binding site that could be very useful.

And that is an area again lot of pharmaceutical sector is doing lot of work yes. "**Professor - student conversation starts**" What is BCR-ABL? Can you elaborate? So BCR-ABL fusion gene is found in CML leukaemia patients. It is produced when a segment of Abelson proto-oncogene ABL from chromosome 9 translocates to the major break point cluster region or BCR on chromosome 22.

So I want to really listed not too much about knowing telling you about what BCR-ABL protein does because that is much more difficult and not so much of the right relevance at this context here but rather that you know how people are using protein production as well as different type of drug molecule production and inhibitors designing which can be so useful for the overall field, in general for the medical biotechnology field. "**Professor - student conversation ends**."

Also people are using the concepts of transgenic if you remember we said that you know we can make transgenic animals, transgenic organism as well right.

(Refer Slide Time: 13:08)



So imagine there is a goat, in that goat you have just changed one gene and that gene you are using to synthesise the human blood protein which could be used for the anti-thrombin or the clotting factors. So now goat is doing all the other protein production the way it normally

does but it is also synthesizing one of the protein of human interest and that particular protein from the goat milk people are trying to purify that which can be very helpful.

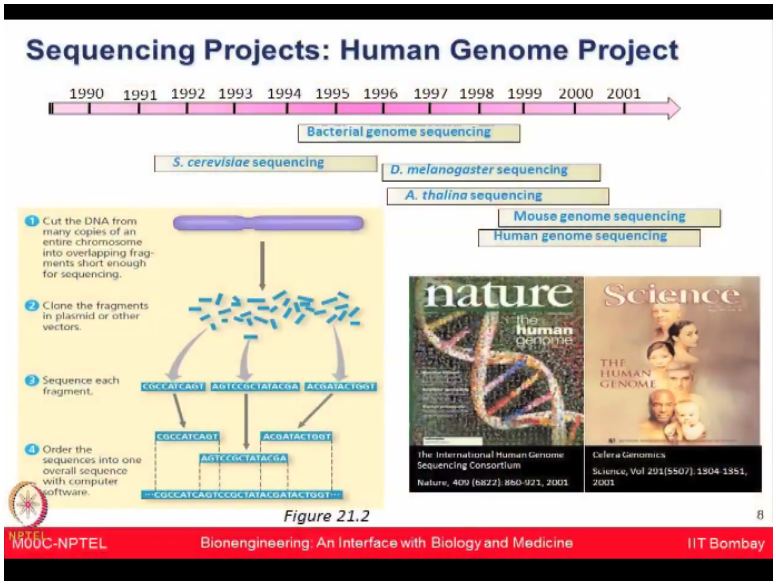
And that can if you can you know have enough of the milk produced from this goat and then you can synthesise or you can simply purify this particular protein that could be used for preventing the blood clots. So these are just various examples to illustrate you in which way our understanding at the genome level at the modern biotechnology tools which we are studying can be so helpful at different levels.

So we have been talking about the techniques in isolation. Now I am just trying to illustrating you how some of these techniques have resulted into some of the practical applications which we all need on day-to-day basis.

(Refer Slide Time: 14:14)

Well, let us move on to another exciting area, another exciting technology which is doing the gene sequencing. So I have been discussing to you that now we have some idea for all the human genes, how many genes we have in human body, which is around 20,300 genes.

(Refer Slide Time: 14:34)



And there has been some very large sequencing projects in whole world where people were trying to study that you know how many genes we have, so first of all people had curiosity to know that how many human genes we have or for the same matter different model organisms how many genes they possess.

So you know we felt that probably as a human we have of course large number of genes which is much more as compared to what one would assume the fruit fly or Drosophila or in the plants like Arabidopsis. So because we are human we are so superior, we are so intelligent, so probably we have you know much larger number of genes which are coding for all of those information.

So before starting this projects of sequencing projects people had kind of you know different apprehensions and lot of you know interesting ideas in mind but in 1990 lot of major projects started which aim towards doing the sequencing of the whole organisms. So many projects for the bacterial genome sequencing or the yeast saccharomyces cerevisiae, fruit fly Drosophila melanogaster, plant Arabidopsis thaliana, mouse genome and human genome sequencing.

All of these projects were actually started and happening simultaneously in that time. So 1990s and 2002 to 2003 that much time was pretty exciting time for all kind of you know sequencing projects. Intention was to see can we know which are all genes present in a given system. To do that people have taken the DNA which we have been talking, cut those DNA into the you know small fragments.

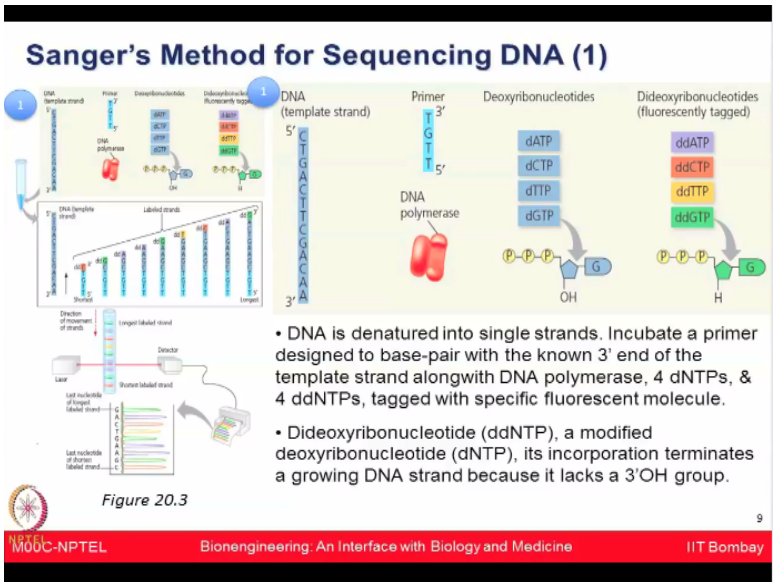
And now you can clone those small fragments each one of those small pieces into different vectors and now each vector you have amplified and now you can sequence each one of those vectors, what is the you know ATGC contents of those and then finally you can put all of them together in the bioinformatics manner which will help you to put together the whole sequence of a given organism.

Just to convey you the complexity of the project aim was to do one human genome sequence, only one human and that took almost 15 years' time of 15+ countries involved and both government fund and private funds were used and this was kind of one of the you know the mega projects in biology which really aim first trying to see inside what is happening and that published the draft human genome published in 2001.

Subsequently, 2002 and 2003 they made more progress in that and then we got some idea that how many genes we have but you know what was interesting here that we knew that which are all genes we have but the number of genes were much lower as expected and the number of genes was pretty much uniform across different organism. So numbers are not very heavily different in different model organism.

So that were kind of major findings from the Human Genome Project. How this was being done?

(Refer Slide Time: 17:31)



So one of the scientists Frederick Sanger, he developed a method for sequencing DNA and we are just showing you briefly about it like how this particular thing was done. So just imagine that you know you have double stranded DNA and now you have separated those, you have denatured the DNA into the single strand. Now imagine from your previous PCR experiments, what are the reagents you are adding? You are adding some of the primers.

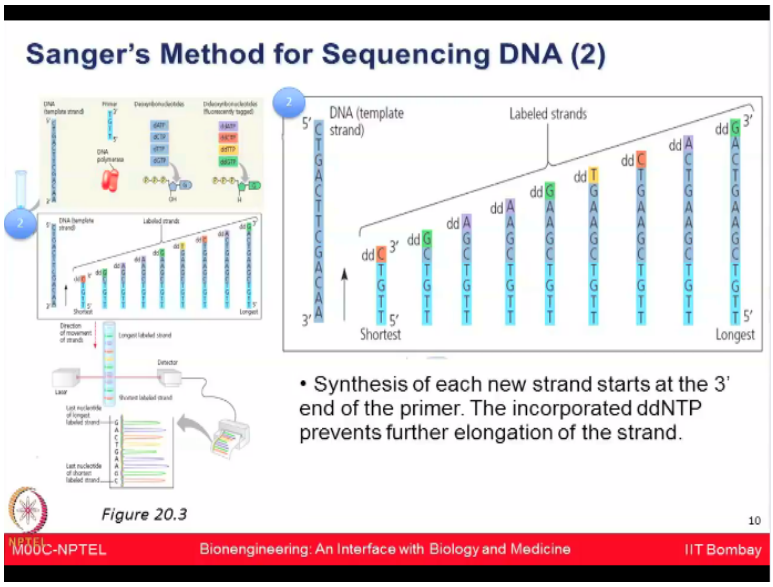
You are also adding DNA polymerase, you are adding different dNTP deoxyribonucleotides. So these are the things which you were added as a part of the reaction for sequencing. Now in addition to this deoxyribonucleotides for sequencing purpose they added one more reagent which is dideoxyribonucleotide, which is different. I will tell you why they did that but before that let us look at the structure of these two.

So in the deoxyribonucleotides whether it is dATP, dCTP, dGTP or dTTP, in all of them you will see that you know 3-5 hydroxyl and OH is present here whereas in all the dideoxyribonucleotides this O is missing so it is only hydrogen is there, OH hydroxyl group is not there. That was one change introduced. Second change here is all these dNTP's are fluorescently labelled.

Each one of the ATGC these are all fluorescently labelled as compared to these which are non-florescent ones. So by doing this reaction now they have added DNA polymerase, 4 dNTP's, 4 ddNTP's which are having the fluorescent molecules. So whenever a ddATP or GTP or CTP or TTP will go, it will stop the reaction because it does not have the OH group on the side.

So it will terminate that particular reaction and because it is fluorescent labelled, so you can actually try to read that from the fluorescent scanners that what this base pair is, what that is A or T or G or C. So the first step was to just add the reaction components right.

(Refer Slide Time: 19:36)

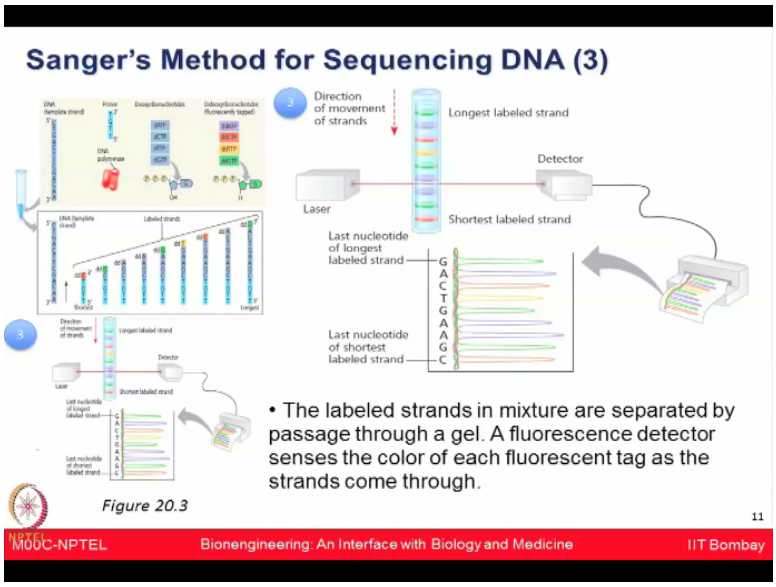


Then, second part was to do synthesis of the new strand. So if this was 5 prime to 3 prime of the original DNA template now because of the primer which you have already provided, a short stretch started synthesizing from it and then from here you can see whenever a C is coming which is dideoxynucleotide then will stop that reaction, whenever dideoxy G is coming here so it is actually complementary here right.

If you look at here this particular stretch of the sequence when you have A it is T, A is T, C is G, A is T and G is C. So whenever this C is coming which is dideoxynucleotide it is stop that reaction. Now whenever you see another G is coming that is stop that reaction. So you have many incomplete reactions, you have many incomplete sequences but what is interesting here whenever they terminate there will be some florescence colour being produced from those base pairs.

So this was the second part to incorporate these dideoxy labels and then elongate the strands.

(Refer Slide Time: 20:42)



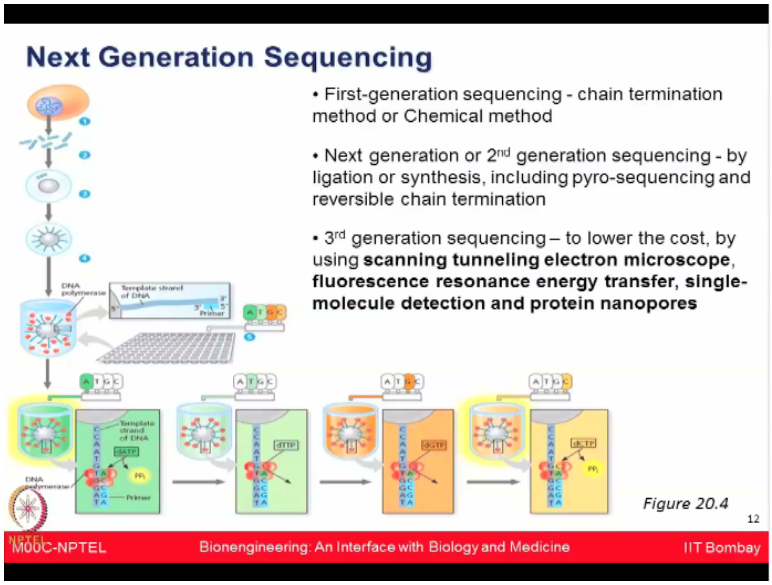
And then you are passing each one of these bases which are labelled now. The sequences from these fluorescent scanners and then you are reading what are the base pairs GACT, GAAGC and then you can distinguish them from different colours because you have already labelled each one of these bases from different colours. So by using this method they were able to do the know the sequence of that particular small DNA piece.

And then like that you have done sequencing for many DNA pieces and then everything you have to put it together to see which are the complementary regions which are the overall sequence for that particular genome. This was kind of you know much more tedious and all that was very accurate but it takes lot of time to do this kind of sequencing but then eventually this is very you know kind of gist of the lot of information here sequencing technologies have really matured quite well.

So from 2001 onwards if you think now in 2017-18 a huge progress has been made in the sequencing technologies. So what was kind of you know billion dollar project which was done in so many years time, now you can do one full genome sequencing probably in 2 or 3 days and maybe in 50,000 to 60,000 rupees. Just imagine how much progress has been made and that is only because of people have from the physics background they brought a lot of principles of how to read these bases much faster.

How to increase the cycle of our scanning and in which way different types of physical principle could be used to separate these bases and one could read them much faster.

(Refer Slide Time: 22:16)



So there are different generations of sequencing technology which has come. The original one was first generation sequencing either the chain termination or the chemical method. Then, we have the second generation sequencing methods which includes the pyro-sequencing and the reversible change termination.

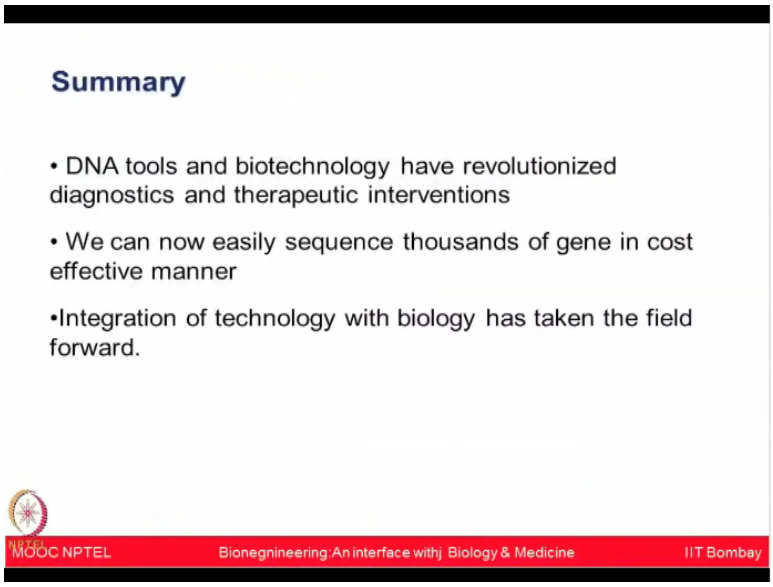
The third generation sequencing is where we now started seeing lot of engineering and physical principles being involved in sequencing and that has really accelerated the pace of sequencing technology. One of the latest example is protein nanopore technology, some of the single molecule detection, FRET technology.

And many of these are the you know latest third generation sequencing methods which are just aiming to look at the some of the physical behaviour of these bases and how fast and accurately one could read and detect them. So as a result now our capability to sequence any organism is much faster. So what people used to think about only sequence 1 Drosophila or 1 human.

Now people are thinking can we sequence you know maybe 100,000 individuals who are all diabetic and then look at you know what is happening at their entire genome level in all of that you know 1 lakh or 1 million population. So now you will have much robust data because you are looking at much larger population and you can do it in the still much shorter time.

So this is one thing that you know you can now sequence the entire organism, entire genome much accurately, much faster way and much more cost reduction has happened as well.

(Refer Slide Time: 23:46)



(Refer Slide Time: 23:51)