Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Arindam Ghosh

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Lecture 22 : DNA Sequencing (Part 2) - Sanger Sequencing

Hello dear students. Welcome back to your lecture series on comprehensive molecular diagnostics and advanced gene expression analysis. In the last class we discussed about Maxim Gilbert sequencing, we discussed about the overview of DNA sequencing, how is the procedure done, what are its applications and as I had mentioned Maxim Gilbert's method and today is the second part of the general DNA sequencing video, where we will be studying the Sanger sequencing method. We will be discussing what is the principle of this method, we will be discussing what is the methodology, what are the advantages as well as the disadvantages and lastly how in modern times the methods have been modified alright. So there is a lot to discuss, so let us start right away this is the overview of the Maxim Gilbert sequencing which is the chain termination method I am not going into detail this is the single slide overview we already discussed in detail how the fragments of DNA separated, how the radio level, how they are cut at various places by the specific chemicals and how the gel electrophoresis is done and how the reads are interpreted using autoradiography using an x-ray plate. Now the next method that is the Sanger sequencing method is actually the second method that Frederick Sanger discovered, mind it while studying Maxim Gilbert sequencing we already discussed that Maxim Gilbert sequencing followed Sanger and Coulson's plus minus method of sequencing which required cloning.

However, this is a totally novel method and which was developed by Frederick Sanger and his colleagues in 1977 and it gained widespread popularity. Maxim Gilbert was the first method that was widely adopted, Sanger sequencing is the next method that actually outdated Maxim Gilbert's method and hence it became widely popular and it remained so and it still remains for the next 4 decades, 40 years, for 50 odd years right. The next method was also automated subsequently almost a decade later by applied bio systems. Nowadays it has been replaced largely for some case to case basis using next generation sequencing which will be our topic of discussion in subsequent lectures.

Now what are the reagents that are requirements of Sanger sequencing? Now you know Sanger actually the his research work was basically an extension of the work that was

already done by Arthur Kornberg, you know Arthur Kornberg who discovered DNA polymerase how replication of DNA happens so his work was further extended and by Dr. Frederick Sanger and his process of DNA amplification actually mimics is almost the same like that of a DNA replication alright. How DNA replication occurs you already know with the help of an existing strand you need a primer you need DNA polymerase like that alright. So similarly all of these are required in Sanger sequencing as well right. So you know a DNA polymerase enzyme, a primer, dNTPs, deoxyadenine, deoxythymidine, deoxycytidine and deoxygonine triphosphate and the template DNA source DNA that we need to sequence.

Deremicing means determining the nucleobases one after another. However, there is one very distinctive component that was the innovation ingenuity of Sanger and his team is use of a special type of nucleotide that is dideoxynucleotide which is chain terminating in nature. So he used all four types of dideoxynucleotide dDATP, dDTP, dDCTP and dDGTP. What are those? Let us understand what we mean by dideoxy. Now this schematic diagram you should be familiar with by now.

This is nothing but a structure of a nucleotide. There is the nitrogenous base which might be ATG or C in case of DNA. The pentose sugar ribose or deoxyribose and this is the phosphate group it makes the nucleotide molecule right. In case of ribose it has got 5 carbons and presence of H and OH group right all over. In places 2 and 3 there are 2 OH groups.

However, if we remove the OH group from the second position that is O is gone if there is nothing it means there is only hydrogen right. So in this case there is hydrogen. It is known as deoxy this is a deoxyribose is the sugar that is present in DNA. What Frederickshanger did so this is the basic structure of dNTP right. What Frederickshanger did he used a special type of nucleotide we will get to it.

Now what happens? Now this dNTP this 3 prime hydroxyl group this is 1, 2, 3 sorry 1, 2, 3, 4 and 5 ok. See the carbons here are actually denoted using 1, 2, 3, 4 numbers right and the bases of the I mean the carbons of the pentose sugar are denoted by prime numbers all right. Anyways so let us I will erase these numbers so it is easy for you to visualize without the cluttering. Now these 3 prime hydroxyl group actually undergoes a reaction with the 5 prime incoming 5 prime phosphate group. So what happens this OH group attacks the 5 prime phosphate group and results in formation of a bond ok phosphodiester linkage that is the basic linkage of a DNA.

So for this amplification to occur you see this is a you can visualize the template DNA and DNA polymer this is a primer and DNA polymer is adding a nucleotide one by one. It is very important that this 3 prime hydroxyl group is present so that 1 dNTP can be

picked up to form another phosphodiester bond just to extend this chain ok. So this is very vital. However what Frederickshanger did he constructed he used a special type of nucleotide where both the oxygen from position 2 and 3 are gone so it is replaced by hydrogen right. So 2 deoxy hence it is known as dideoxy all right so this is dideoxy nucleotide diphosphate d dNTP.

So what is the issue or what is the change when 3 prime hydroxyl group is absent this reaction cannot happen. What is the reaction? The 3 prime hydroxyl group which actually acts as a hook to fetch incoming nucleotide is absent. Hence the moment 1 d dATP is added there will be no further extension because there is no hydroxyl group to form another phosphodiester bond ok. So this was used this principle was used all right. Now what extra did Frederickshanger and his team use for the experiment? They used one small a very small amount of radioactive NTPs as well radioactive dNTP.

So they used DNA polymerase they used template DNA they used d dNTP all 4 dNTPs as well as a very small amount of radiolabeled dNTPs. There are the amount of radiolabeled dNTPs are very small compared to the normal amount of dNTPs in the reaction ok. Now the reaction occurred in a special way all the 4 just like Maxam Gilbert sequencing where they used 4 different types of chemicals to treat 4 different tubes. Here also the common reagent were added to all the 4 tubes, but in tube 1 only one type of radiolabeled dNTP and only one special type of d dNTPs were added ok. So now we have tube 1 which has got radiolabeled dNTPs in very small amount and all the 4 dNTP it has got a template DNA it has got a DNA polymerase and special in case of tube 1 there is d dATP.

All the same reagents are same for tube 2 except tube 2 has got d dTTP in addition to all the normal dNTPs tube 3 has got d dGTP tube 4 has got d dCTP all right. Now we will proceed towards the next step let us visualize what is happening in case of tube 1 all right. Now we should keep in mind that all the dNTPs has got an equal chance of being added see this is a template strand means we need to sequence this strand and we are forming a complementary strand from these dNTPs that are in the mixture all right. And at last we will decode what is the complementary strand that has been formed in the mixture and we will calculate the original strand by using the formula A opposite to T and G opposite to C right. Now and one more thing when Frederik Sanger devised this method mind it the original template DNA had to be denatured ok and then cooled so that the primer can bind why and subsequently when it was cooled only then the DNA polymerase could be added why can you answer this question having read PCR the earlier week it is because till now in this time lines 1977 polymerase chain reaction has not been discovered yet.

So after PCR was discovered the more the better automation of Sanger sequencing was

done, but till now at this point of time when Frederik Sanger and his group was working on that they did not have access to Taq polymerase ok the one thing very important fact you should keep in mind. Anyways so now that primer is there and DNA polymerase will add all the dNTPs based on the complementary sequence that is present in the template strand. So if there is T over here suppose there is T first A will be added. So like that next if there is suppose it was T and then it was A like that next definitely it T will be added so like this. So depending on the complementary strand that is on the depending on template strand the complementary bases will be added fine.

So it will be continued to add continued to be added this way till by chance one dDATP is added. Mind it when they are fetching the nucleotide there is no difference at the 5 prime end of either dATP or dDATP both will look the same therefore DNA polymerase has got no problem in picking up either dATP or dDATP. Now you may ask or you may think that this dDATP could have been added in the first step yes definitely it could have been added but just because there are so many millions and millions of bases over here is a high probability that it will be added or it may not be added. So wherever there are complementary wherever there are Ts and wherever there are complementary As to be added the chain can be terminated at that point of time. So what is the problem that is happening over here when this dDATP is added in the next step dDCTP which should have been added in the next step now cannot be added because this special type of NTP that is dDNTP does not have a 3 prime hydroxyl group ok.

So reaction is terminated over here. So imagine there are billions of DNA molecules that are present and they can be terminated at any place where supposedly one A has to be added at some point of time suppose in billions and billions of copies there will be few thousands of such copies where the dDATP will be added and it has to be terminated at that spot right. So these are all probability based additions where the dDNTP has been incorporated and chain has been terminated in the tube in the reaction mixture right. This is same for all three other terminator reaction. So it may be so that considering tube 1 the tube 1 reaction has been terminated in three areas and if we calculate the if we in fact if the number of bases let us presume are one has been terminated after addition of 5 bases, one has been terminated after addition of 13 bases and one has been terminated after addition of 10 bases right.

So next part now this tube means it has got different length of fragments already present in it. So how and likewise this will be same in all other three tubes as well just as discussed. So dDTTP will also be terminated in two or three places dDGTP, dDCTP depending on what were the position. So number of fragments can vary depending on the number of nucleotides that could have been incorporated depending on the templates and complementarity.

I hope I am clear. So now what is done all the material of all the tubes are loaded in all the four lane. This step is almost similar to that of Maxam Gilbert reaction, Maxam Gilbert sequencing step where you I mean the material of tube 1 is loaded in one lane, material of tube 2 is loaded in another lane so on and so forth all right. So then they are subjected to electrophoresis all right and this is the see DNA is a negative charge right. So from negative to positive this will migrate like this and this just like Maxam Gilbert sequencing is polyacrylamide gel electrophoresis. Why polyacrylamide gel is used? Because it can give a very good resolution up to a single difference of a single nucleotide can be determined all right.

So they will move through the electric field and over the course of electrophoresis the shorter fragments will move the farthest and the heavier fragments will be lagging behind all right and we can get a similar situation where there are multiple bands in a gel. Now mind it there were also radiolabelled dNTPs in the gel all right. So if we subject the gel I mean if we expose if we now treat the gel electrophore polyacrylamide gel with x-ray film it can record the radioactivity that is emitted by the radioactive bands all right the radioactive dNTPs and ultimately just like Maxam Gilbert sequencing we will get these type of information. Now it is very easy if we know what tubes were loaded in what lane to back calculate the sequence. So I hope you can do with me this time.

So A1 will be definitely A, 2 will be A, 3 will be G, 4 will be C, 5 will be T, 6 will be T, 7 will be A, 8 will be G, 9 will be C and 10 will be C. So again read direction is from lower to higher all right. Mind it this is the strand that has been synthesized in the tube. So the original strand will be complementary to this strand. So this was the method of Sanger sequencing contrast to Maxam Gilbert sequencing since there was A and A plus T, G and C plus C.

So there was a rule of exclusion, but we already read how to interpret autoradiograph of Maxam Gilbert sequencing as well. Mind it where whenever an x-ray film is detected I mean whenever a gel radiation is detected this type of band using an x-ray film the process is known as autoradiography ok. So what is the advantages of Sanger sequencing? It is very accurate, it is very accurate, it is highly reliable, the polyacrylamide gel can give us a resolution of single nucleotide ok very very very sharp resolution and it is very cost effective and for targeted sequencing ok. If you already know sequence of DNA and we can target segment where we need to determine where we need to determine the mutation it is very effective. And hence till date it is considered as the gold standard technique and this has been actually utilized and modified for the all the future varieties of sequencing.

Now you may ask ok does it have any disadvantages? Definitely it has got disadvantages that traditional Sanger sequencing the original one that we just discussed

over here it is actually very slow. Slow in the sense from the perspective of reading 200 bases ok it took 4 days mainly because of the fact that one day it took for all amplification reaction, chemical treatment, electrophoresis, radio labeling everything and the development that is autoradiography and analysis it took 3 long days. So it was very slow and it was manual and labor intensive and hence just because it is very slow even after it used radioactive chemicals maximum up to 400 bases was tried by the group of scientists and hence it is actually ineffective for very long sequences ok. So, hence it has been modified the first applied biosystems actually adopted I mean devised a veryclever method and it in they introduced their first commercial sequencer automated sequencer ABI 370 A which actually adopts the similar principle, but instead ofdetecting with the help of x-ray film which took long 3 days they used fluorescent chain terminators means they tagged each type of DDATP, DDTTP, DDGTP or DDCTP with a different color means whenever the reaction is terminated the if we try todetect using a special type of emission we already discussed how fluorescence works when we are discussing RTPC real time PCR right. So, using that fluorescence emission spectators very easy to get colored bands.

So, when x-ray was out of the picture when we are detecting fluorescence now it becomes very easier we can see colored bands. So, yellow corresponds to A, blue corresponds to T, green corresponds to G and red corresponds to C not only that when we have got all the 4 colors there is no need of 4 different lanes and 4 different tubes we can actually perform or combine all of the reaction in a single tube in a single lane all right. This method so, I hope you all agree with me because since we can now do the whole sequencing in one single reaction this was actually the ingenuity of the discovery that was done by applied bio systems where they took one step further and they introduced their ABI 377 DNA sequencer where they used the capillary electrophoresis is the same principle they are they are using same fluorescent chain terminators, but they are sending the reaction the products through a small capillary tube and there is a camera at the end of the capillary tube which can sense the colors. So, we call it a finish line. So, as one base actually flows through the capillary tube as one gets exited they can easily detect the signal and the signal is given by a peak of a colored wave that is detected by a camera here a CCD camera sensor is there and the mixture is actually illuminated by laser which helps in detection of the emitted fragment ok the emitted fluorescent chain terminator corresponding to the reaction.

Hence this was very easily automated. So, all the four colored dye registered one after another and depending on where they were terminated this is sent through a capillary it is illuminated by laser and it is detected by detector and we get this beautiful waves of color which is then analyzed which is called a chromatogram because this capillary electrophoresis and it is detected using a computer software alright. So, this was how Sanger sequencing was automated and ABI prism 377 was an automated capillary

electrophoresis based DNA sequencer means we do not need to add reagents separately reagents are already there we just add the sample in the capillary and everything using robotic arms was done internally and finally, just a matter of time we got all the sequence waves which was ready to be analyzed using the computer software. So, with further development you know the first commercially available sequencer was introduced in 84 then it was the successor was introduced in 87 early 90s multiple models followed. So, applied biopsy which is now a part of thermofisher they introduced the 310 in the mid 90s 370s in the late 90s and prism 3730 in the early 2000s which actually continues to be the machine of traditional Sanger sequencing that we use the latest model is 3730 excel alright. So, all of these machines are improvement over another and they all improve over the ease of use speed of automation number of samples that can be done the number of samples that can be done simultaneously so on and so forth the number of capillaries that are present in the machine.

So, this is ultimately the latest variety of automated DNA sequencer that is that has been developed by thermofisher scientific designer SEEK Studio 24 and it has got all the belts and will cells and nuances of the latest analyzer as it happens it use a touch screen there are multiple user friendly software it has got remote monitoring integrated table shooting solution. But the it all started with the geniusness or ingenuity of Dr. Frederick Sanger and his group which capitalize on the fact that using a dideoxynucleotide we can terminate the reaction. So, what was the impact of Sanger sequencing in genetics? You know after Sanger sequencing was developed we already we all knew that it was easily adaptable and it was adopted and afterwards just a few days later with the advent of PCR the whole ball game changed and we were ultimately able to sequence the entire genome. So, the human genome project we will be discussing in brief and in detail actually was very much possible due to the Sanger sequencing the discovery of the Sanger sequencing.

One is Sanger sequencing one is PCR made sequencing of the entire human genome possible. Now, regarding the further development as we have read modern machine we can read up to 900 pairs that are routinely done using this method and fragments are actually aligned based on overlapping portions. So, I will be discussing that in short gun sequencing. So, thus the entire chromosomes can be done. So, if you are not clear with this point you can choose to just read it now and understand it well fully and recently there was a COVID-19 pandemic and sequencing the SARS-CoV-2 spike protein none other than Sanger sequencing method was used.

So, you understand now what and how much Sanger sequencing is important in today's molecular diagnostics as well as gene expression analysis. So, to summarize in the last two classes we discovered I mean we discussed the overview of DNA sequencing the applications of DNA sequencing we covered Maxam Gilbert and today we covered in

detail what are the methods of Sanger sequencing and what are the methods and modern methods and modifications of Sanger sequencing. So, these are my references and I thank you for your patient hearing.