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 21 : DNA Sequencing (Part1) - Maxam Gilbert Sequencing

Namaskar. Hello students, welcome back to our lecture series on Comprehensive Molecular Diagnostics and Gene Expression Analysis. In today's module henceforth we will be discussing DNA sequencing. We will be covering all types of genetic DNA sequencing and in today's class we will be starting with the very basics fundamental and historical aspects of sequencing. We will be discussing a brief overview of DNA sequencing, we will be discussing what are the applications of sequencing and we will be discussing one of the most important discoveries that is Maxam Gilbert method of DNA sequencing. We will be discussing the history of it, what is the method, what are the advantages and disadvantages of this specific method of DNA sequencing right.

So what is DNA sequencing? DNA sequencing is nothing but a laboratory technique, it is an experimental method to determine what are the exact sequences of nucleobases in the DNA. You know DNA are made up of various four nucleotide bases adenine, cytosine, guanine and thymine, but not every DNA is the same and to determine what are the sequences that appear one after another in a DNA is basically DNA sequencing. Now this method is very important, why because it helps us to understand many things, the entire molecular biology is dependent on it right. So any variety of genomic research we need to know what we are working with to read the very basic element of DNA is the fundamental of all genomic research and it is got immense application.

Basically you can look into the whole module and you can fit that into the head of genomic research. Not only research in diagnostics, in medical diagnostics, diagnostics of various rare diseases what mutation is happening, what mutation leads to cancer, many forensic analysis for detection of you have seen paternity test or detection of any criminal anything we need to identify them with DNA and what do we tell you by identifying the DNA basically the DNA sequence all right. Next microbial genomics definitely we by understanding the DNA sequence we can detect whether there is any change in the DNA sequence, whether there is any mutation, whether there is any property of the bacteria or virus or fungi, what is the drug sensitivity, what is the drug resistance will be easily able to identify everything. Again phylogenetics evolutionary

biology so what organism is related to what ancestor, everything again the information lies in sequence of the DNA. Agriculture again all types of crops what are the genetic engineering is I mean producing any alteration of crops or we are achieving any transgenic variety of crops everything again answer we need to start with the very basic DNA sequence in order to alter it.

And finally, environmental monitoring every living organism in and around us have got some sort of DNA life form to start with and if we know the sequence we will know all their properties by further downstream experimentation. So, utility of sequencing is manifold and you can think of any molecular biology technique definitely it will start with the very fundamental that is the understanding the sequence or knowing the sequence or the serial of the bases in the DNA. Now it all started this Maxam Gilbert method of DNA sequencing started when they published the Alan Maxim and Walter Gilbert they worked in the department of biochemistry molecular biology in Harvard University Manchester right Massachusetts Cambridge. And they published the paper in 1977 and it widely became the very first adopted method right why why did it happen it was preceded by Sanger and Coulson's original method of plus minus sequencing mind it. However, this method the Sanger and Coulson's method of original Sanger and Coulson's method of DNA sequencing needed to purify the DNA with the help of cloning right and that added another layer of complexity, but this Maxam Gilbert method provided a mean of directly reading the purified DNA.

So, it was widely adopted and it became very popular. So, we should understand the method right. So, what exactly happens in this method this is this method also known as chemical degradation method of DNA sequencing why chemical degradation because this method involves cleaving or cutting of a DNA fragment at various known location and then by looking at the pattern of those cuts we need we the discoverers of this method back calculated in order to learn the sequence. We will also learn it in step by step method and it is very easy and intuitive once you understand the concept alright. So, what is done first? At first the double stranded DNA is treated with alkaline phosphatase to remove the 5 prime phosphate group and it is actually radiolabeled we need the radiolabeled DNA to start with.

So, with a p32 gamma labeled ATP using polynucleotide kinase the phosphate group is again reattached and the DNA is now radiolabeled. So, this radiolabeled double stranded DNA at the 5 prime end of both ends right it is treated with DMS or dimethyl sulfoxide and heated that results in separation of the two strands that results in denaturation and those two strands are actually separated by electrophoresis. We will be discussing electrophoresis is a method of proteomic separation as well. So, we will be discussing that again during proteomics, but the basic concept of electrophoresis is it is loaded in a gel and it is subjected to an electrical charge. Now this will separate the two bands, but

how definitely one band is heavy and one band is light.

You see not all the two strands of the DNA are alike one thing is complementary to each other right, but the strand with more number of purines you know purines have got double rings like that and pyrimidines have got a single hexagonal ring. So, pyrimidines are lighter less number of molecules purines are heavier. So, the heavier chain will move slower. So, the fragment that is lighter will contain more number of purine will lag behind and the fragment of DNA which has got less number of purine bases that is the complementary segment the single strand will go forward. Then so, this is the way how two strands they separated the Maxim and Gilbert and the strands were now taken up right.

So, we now have single stranded radiolabeled DNA at 5 prime end and ultimately the sequence will be obtained from the 5 prime end only sequence is read from the 5 prime end ok. So, I hope you understand the method how the DNA was radiolabeled first and then it was denatured into two single strands and then they are separate with the help of electrophoresis right. So, now, we have single stranded separated DNA to work with definitely we can work on both of them right, but for our experiments sake let us take up one strand to visualize what is happening. Now, what they did they poured all the radioactively labeled strand in four different tubes. So, all the four tubes have got same parent strand which is radiolabeled at 5 prime end.

Next what they did they subjected these tubes they exposed these tubes to special chemical treatments. So, all the four tubes underwent different chemical treatment. So, that they were cut specifically at different regions you see. So, in case of tube number 1 what happened tube number 1 they added dimethyl sulfate all right what dimethyl sulfate does dimethyl sulfate methylates the guanine residues. Thereafter in heating with hot piperidine the DNA will only specifically be cut at the places where there is guanine not any other region.

So, this is tube 1. Next again treatment with dimethyl sulfate, but now heating with alkali. So, when we are treating with dimethyl sulfate adenine and guanine residues will both be methylated, but it is the second treatment which is varying the cut in this case. So, in this case definitely adenine will also be methylated, but it will not be cut, but in this case since we are treating with 0.1 molar NaOH or caustic soda sodium hydroxide it is cut both in adenine and guanine residues.

So, if the strand has got adenine in two regions as well as guanine not only it will be cut at the guanine it will also be cut at the adenines as well where adenine. So, anywhere where adenine is present it will be cleaved where guanine is present it will be cleaved. So, both adenine and guanine zones will be chopped all right. So, we will get more

fragments when treating with dimethyl sulfate and NaOH. Likewise in the third tube what happens in the third tube they treated with hydrazine, piperidine and sodium chloride in presence of sodium chloride 1.

5 molar. This specific chemical reaction cleaves the DNA at places where there are cytosine residues fine. So, you see here there are cytosine residues and the DNA have been cut like this. So, one segment is of this size and another segment is of this size all right. This will come very handy. Next in the last tube same hydrazine, but without sodium chloride.

So, sodium chloride protects thymine residues. Here since there is no sodium chloride hydrazine and piperidine exposure cleaves the DNA segment where either cytosine or thymine residues are present. So, you see here there were three sites where either cytosine or thymine were present. Here there are two sites of cytosine hence these were cut like that. So, either it will be cut here or it will be cut here fine.

In this case either it will be cut here or it will be cut here or it will be cut here. So, we have probability of getting three fragments clear. So, this is the first step or the first phase of Maxam Gilbert reaction where their ingenuity their I mean invention is that they specifically devised methods to cut radiolable strands of DNA in specific areas of interest ok. So, what next? So, all the fragments from all the tubes are now poured in gels in a specific plane. So, this is a polyacrylamide gel also polyacrylamide gel alright and when an electrophoresis is done on a polyacrylamide gel it is referred to as polyacrylamide gel electrophoresis or PAGE alright.

So, the so when we are casting a gel we can design a gel in such a way using comb. So, that there are lanes or holes in a gel right. So, we need to load the DNA in these holes alright and then we expose the DNA or expose the whole gel to electric current and then depending on the weight of the fragment they will move or they will stay depending on the whether they are heavy or light not only weight also on the charge charge and mass charge by mass ratio M by Z ratio is a determinant factor how it will move alright anyways. So, tube or the material from tube 1 is poured in lane 1 material from tube 2 is poured in lane 2 lane 3 lane 4 so on and so forth and then the gel electrophoresis is run. So, definitely all the materials since they are very different in all the tubes depending on where they have been cut the on the exposure of electric current they will be separated based on their size alright.

So, smaller fragments will move further and larger fragments will remain more towards the zone alright I mean this loading zone. Next since these are radiolabelled alright radiolabelled p 32 will give a radiation alright they will emit rays and they can be detected with the help of x-ray film and this technique is known as autoradiography

alright. So, only the segments whose ends are labeled with radioactive probe will give a band right this is known as autoradiography. So, when exposing a next ray film to the gel we will only get information about those brands via autoradiography which has got a radiolabelled phi prime end right. So, now, this will be the final picture.

Now, this is the final step the beauty of Maxam Gilbert sequencing is to interpret what are the sequences from this picture of x-ray film alright. So, let us try to understand. So, if you understand this step it is very easy for you to determine the sequence alright. So, mind it the sequence will be read from phi prime to phi prime end. Now once again this is very important for students who are in their first year undergraduates ok because Maxam Gilbert method of sequencing is very important for any type of image based questions as well as short questions and even long question in first year undergraduate medical curriculum.

So, we start with the first lane or the I mean the zone which has travelled the farthest ok. So, when we are considering this so, this is the sequence where we need to plot what lies here so on and so forth right. So, let us understand in the third lane and the second lane we see two bands right. We need to keep in mind what are the lanes to start with the first lane comprises of bases which were cut in G, the second lane comprises of bases which are cut in C, the third lane comprises of base which are cut in both C and T and the fourth lane consists of bases which were cut in both G and A. So, by collective information we need to now back calculate what is the sequence alright.

So, for considering the first position when there are bands at both the tubes which were cut at either C or even if they were cut on C and T. So, can you guess what will be the nucleotide in the first position? Yes, the nucleotide that is common towards so, irrespective of whether it is cut at C or whether it is cut at C and T they are giving a same band. So, definitely the first position contains C alright absolutely right. So, considering the next lane the next lane I mean the next band ok the next band it has got only band where C and T were cut ok, but it does not have any band where C was cut right. What does it mean? It means the second has got nucleotide in the T position yes absolutely right.

Next we are considering the third band. In the third band we can see there is a band where both G and A position was cut, but there is no band where there was supposed to be cut if there was a G. Hence in the third position the probable nucleotide is A not probable it has to be the nucleotide is A. The fourth one can you guess? In the fourth one there is a nucleotide where there is a band where there is a cut in both G as well as there is a band where there is cut in both G and A. So, what what lies common between them yes the fourth one will have to be G.

Next band at both C and T and band at C. So, what fifth one yes if you have guessed it right it is C. Again the next one band at C plus T, but no band at C definitely it has got T. So, likewise what will be this one G plus A and G A. So, band at only G as well as G and A definitely it will be G.

Next band only at G plus A, but no band at G definitely it is A. Again the next one both band at C plus T and C what is it? It is right C and the next one band at both G as well as G plus A. So, it will be G right. So, this way the scientist Maxima Gilbert showed both the strands here we just detected one strand to start with. So, one strand if it is cut specifically using specific nucleotides at whatever places of interest by loading the fragments and reading their 5 prime end we will have definite idea what is the length and how much they have travelled.

So, Maxima Gilbert actually in their original paper they analyzed both the strands and they showed with their experiment that both were complementary to each other alright and therefore, it was proven beyond doubt that this method of chemical degradation get comprehensive idea about sequencing the entire DNA fragment and hence it was very widely adopted alright. So, recently I mean in 2017 the paper of maximum Maxima Gilbert received a citation for chemical breakthrough award in the class of 2017 and the award was presented by the division of history of chemistry is a division known as history of chemistry in American chemical society and the award was present to the very department where maximum and Gilbert used to work and from where they devised this whole method that is the department of molecular and cellular biology in Harvard university alright. However, as you know this method has not aged well it was it has got nowadays it has got historical importance mainly because of the fact that it used many toxic chemicals alright and extensive use of radiolabel isotopes which are hazardous to the experimenters very poisonous and them are unstable. And moreover it can only read a very small fragment of DNA up to maximum 500 base pair because the reaction is actually quite slow setup is very complex reaction speed is slow. So, it will take days to weeks in order to digest the whole product slowly and then depending on their extensive length the read length the product will be very slow to start with and moreover designing a kit which everyone else can use using the maximum Gilbert method was also not feasible to feasible right.

And hence it was replaced by Sanger's chain termination method this is an award winning method which will be discussing in the next class and you know this is what happens we in science there is an wonderful method which is adopted by everyone and then another wonderful method. So, it that original method sets the bar for discoveries and then another wonderful method comes out and then it is replaced and hence the science advances right. So, we will be learning the Sanger's chain termination method which is actually known as Sanger sequencing method and this method has been actually

replaced now by next generation sequencing. So, we will be learning all of them in upcoming classes. So, in today's summary these are the topics that we covered we covered overview of sequencing what are the applications of DNA sequencing and we studied in detail the historical sequencing method that is maximum Gilbert's method of sequencing we discuss the methods how the whole thing is done what are the chemicals that are used to cleave the whole process whole DNA how the gel electrophoresis done how it is auto reducible and then by back calculation and back calculating and looking at the fragment of DNA how we can re purpose or I mean backtrack in detecting the sequences we also learnt about its advantages as well as disadvantages.

So, these are the um references the I have included a review article as well as maximum Gilbert's original paper which you can read at your own interest and time and it will help you to further look into what problems they face what other details they were doing in their own lab in order to when they discovered the whole thing. So, thank you for your patient hearing and I will see you soon in the next class with a lecture on Sanger Sequencing Method. Thank you for your attention.