

Maxam-gilbert and sanger's method of sequencing



**ASSIGN
BUSTER**

This chapter provides an understanding of the history of sequencing. The Maxam-Gilbert and Sangers method of sequencing are explained in detail. A short note on pyro sequencing is also added.

Sequencing is a process by which the sequence of nucleotides is deciphered in a particular portion of DNA or RNA. This method offers several advantages in daignosis.. Firstly, in a PCR product, it helps him to determine if there is a mutation in the sequence. A classical example of this is in evaluating RAS gene mutations. The RAS gene usually shows mutations in codon 12 and 13. It may also rarely show mutations in codon 61. Without sequencing, the determination of a mutation is impossible. Sequencing is also useful to confirm the presence of a Single Nucleotide Polymorphism (SNP) or a point mutation in cases where Restriction Length Fragmentation Polymorphism (RFLP) is equivocal. Scientists use it to characterize newly cloned cDNAs and to check the fidelity of a newly created mutation.

HISTORY

Prior to 1970's, there was no system to check a DNA sequence. The only way to hypothesize a sequence was to determine the amino acid sequence and retrospectively determine the nucleotide sequence based on the appropriate codons. Given the degeneracy of the genetic code, this system was essentially intelligent guesswork at its best (see the chapter on transcriptions and translation Chapter 5).

In mid 70's, Maxam-Gilbert and Sanger developed methods to accurately determine a DNA sequence. These methods were cumbersome and time consuming. The automated sequencing method was a considerable

improvement over the previous methods. An analogy between them is best illustrated by comparing the joy in driving a Ford Model T and an S class Mercedes Benz.

MAXAM-GILBERT'S METHOD

What Maxam and Gilbert proposed to determine a nucleotide sequence was quite simple. They took a terminally labeled DNA molecule and with the help of chemical agents, broke it at the points of attachment with adenine, guanine, cytosine and thymine. They then produced radioactive fragments extending from the labeled end to the position of that base. They ran the entire product on Polyacrylamide Gel electrophoresis (PAGE) which resolved the points of breakage. They then took an autoradiograph which produced four different cleavages specific for each base.

The details of the chemicals used for cutting need not concern us here. Suffice to say that the method developed was both sensitive and specific. It provided a good chemical distinction between the bases. Analysis of sequence on both the strands provided adequate check. However, the problems associated with this method were enormous. It took few days to sequence 200 - 300 bases. Moreover, there were several 'ifs' related to the procedure such as 'if the radioactive labeling process did not work', 'if the cleavage reactions did not perform as expected', 'if the gel did not set up properly', 'if the electrophoresis did not work, if the gel were torn or otherwise destroyed during transfer', and 'if the X-ray film developer broke down during the development'. Even if everything worked perfectly, one would expect to get 200-300 bases of a confirmed DNA sequence every few days. The other associated problems were that a lot of radioactive material

<https://assignbuster.com/maxam-gilbert-and-sangers-method-of-sequencing/>

was used and hydrazine which was a chemical used for cutting happened to be a neurotoxin.

SANGER'S METHOD

At about the same time as Maxam-Gilbert DNA sequencing, was being developed, Fred Sanger developed an alternative method of DNA sequencing. Rather than using chemical cleavage reactions, Sanger opted for a method involving a form of ribose sugars.

The principle that Sanger used was based on a paper by Atkinson et al. Atkinson showed that when 2', 3'-dideoxythymidine triphosphate (ddTTP) was incorporated into the growing oligonucleotide chain in place of thymidylic acid (dT), the chain extension stopped and termination occurred specifically at positions where dT should have been incorporated. Sanger extended this technique to other dideoxy nucleotides (ddCTP, ddATP, ddGTP) and thus using four different tubes with four different ddNTP's, he managed to terminate DNA sequence at places where nucleotides were supposed to be incorporated.

Fig 11. 1 - In the upper panel, there is an OH group at position 4. This allows the chain to elongate. In the lower panel, there is an H atom which has replaced the OH group. This does not allow the chain to elongate and thus the chain terminates.

Fig 11. 1 illustrates this principle. The presence of an ' H' group instead of the ' OH' group does not allow the chain to elongate and thus, the chain terminates. To put this into practice, one can use four separate reactions. Each reaction has all the components for a PCR but in addition to dNTPs, a <https://assignbuster.com/maxam-gilbert-and-sangers-method-of-sequencing/>

small proportion of ddNTPs is also added. The four reactions have four different ddNTPs. The ddNTP concentrations are carefully adjusted so that they get incorporated into the growing DNA strand randomly and infrequently. Due to this, the elongating chain terminates randomly. When the entire product is run on a gel using separate lanes for each nucleotide, it is obvious that the position of the bands corresponds to the position of the nucleotides. Thus, the gel can be read off and one can easily decode the sequence. For further understanding of what the gel picture would look like, please refer to fig 11. 2.

Fig 11. 2 - An example of a sequence obtained using the Sanger's method. Note that the four lanes have been labeled as GA, CT, A G and T C respectively. This means that in the PCR reaction, in addition to the normal dNTP's, there were also ddNTP's of Adenine, Thymine, Guanine and Cytosine Guanine, Cytosine, Adenine and Thymine in the respective lanes. As soon as the ddNTP is incorporated, there is a termination of the extension. When the products are run on a gel, the termination of the sequence is seen as a band when labeled by autoradiography. As noted in the text, it can be seen that the sequence can easily be read off the gel

Box 8. 1

Sequencing is a process by which the sequence of nucleotides in a particular portion of DNA or RNA is obtained.

The Maxam and Gilbert method used the concept of taking a terminally labeled DNA molecule and breaking it at the adenine, guanine, cytosine and

thymine residues with chemical agents. This was run on a PAGE and the points of breakage were resolved.

Sanger used a 2' 3' dideoxythymidine triphosphate. This was incorporated in the growing chain and it prevented further extension. Four different labeled nucleotides are used and they terminate the DNA sequences at those places where the nucleotides were supposed to be incorporated.

AUTOMATED METHODS

The development of manual sequencing methods by Maxam Gilbert and Sanger et al was a dramatic improvement over the previous methods which were mainly based on guesswork and luck. Though the chemistry of both the methods was path breaking, it was difficult to sequence large portions of the genome.

Development of automated sequencing methods by Hood ensured that sequencing was faster and far simpler to perform as compared to the manual sequencing methods. The basis of automated sequencing is labeling the product with some form of a fluorescent dye that can be detected using a detector system.

Logically speaking, only two components in the Polymerase Chain Reaction can be labeled: the primers and the dideoxy sequences. In the method described by Hood, the primer was labeled with one of four different fluorescent dyes. Each labeled primer was placed in a separate sequencing reaction with one of the four dideoxynucleotides (to terminate the reaction) and of course all four deoxynucleotides. After completion, all the four reactions were pooled and run together in single lane of a polyacrylamide

<https://assignbuster.com/maxam-gilbert-and-sangers-method-of-sequencing/>

sequencing gel. A four-color laser-induced-fluorescence detector was used to scan the gel as the reaction fragments migrated past. The fluorescence signature of each fragment was then sent to a computer where the software was trained to perform 'base calling' (a computer program for identifying a base (nucleobase) sequence from a fluorescence "trace" data generated by an automated DNA sequencer). This method was commercialized in 1987 by the Applied Biosystems.

James M. Prober and colleagues at DuPont took the fluorescent sequencing method to its next level by developing "a more elegant method". Instead of fluorescence-labeled primers, they labeled the terminators themselves. The first 'dye set' was based on succinylfluorescein. Each ddNTP was labeled with a different chemically tuned succinylfluorescein dye which could be distinguished by its fluorescent emission. All four dye-labeled terminators were excited by an argon ion laser at 488nm to produce peak emission that could be distinguished by a detector. This detection system meant that the sequencing reaction could now be carried out in a single tube with all four terminators present and fragment resolution would require only one gel lane. For record, it must be mentioned that initially investigators used to run PCR products on a gel and then 'read' the fluorescence generated. The introduction of capillaries was a breakthrough in the development of automated sequencing methods. Small capillaries with 50 μ m inner diameter dissipate heat very efficiently due to their high surface area to volume ratio. A capillary based system can, therefore, be run with much higher voltages. This lowers their running time dramatically. Fluorescence can be detected through the capillary tubes. Thus, the capillary systems

could be automated as opposed to gel based systems. A schematic diagram of sequencing is shown in Fig 11. 3.

THE METHODOLOGY OF AUTOMATED SEQUENCING

Sequencing is performed on a short chain of nucleotides, which can be either a PCR product or a cloned DNA sequence. Only about 1000 bases can be sequenced accurately, a far cry from approximately 50 to 250 million bases that comprise a human chromosome. If one takes a PCR product, a primer of known sequence is required for each sequencing reaction. Thus, one cannot take any piece of DNA and “ just sequence it.” A known starting point, and thus some knowledge of the sequence, is required to begin the reaction.

There are two ways of making DNA manageable and thus beginning the cloning process. The simpler way would be to perform a PCR and sequence the products. The second method would be to clone the DNA. In cloning, a DNA sequence is introduced into a vector and several thousand copies are generated when the vector replicates. In this section, we will not elaborate on cloning as a preferred method because sequencing of PCR products is simpler and more commonly used on clinical specimens.

Sequencing the PCR product – Following an initial PCR reaction, it is necessary to confirm that the reaction has worked and a product has actually formed. This can be done by running the product on a gel and confirming that the product is of right size. Then, a second PCR reaction is performed using either fluorescent primers or fluorescent nucleotides as outlined earlier. Protocol needs to be adjusted based on the machine used

and will not be elaborated further. After the confirmation of a successful PCR, the products are purified.

There are several methods for purifying PCR products. These are ultrafiltration, ethanol precipitation, gel purification and enzymatic purification. In a functioning laboratory, it is however, advisable to use commercial kits for purification. Several manufacturers such as Sigma and Genetix manufacture such kits and it is advisable to follow their set procedures. One should remember that the basic aim of DNA purification system is to remove chromosomal DNA, proteins, enzymes, residual organic chemicals, detergents, residual agarose if DNA was extracted from a gel, primers, unincorporated nucleotides, and salts from enzymatic reactions; The commercial kit should be chosen keeping all this in mind.

SOURCES OF ERROR IN DIRECT DNA SEQUENCE ANALYSIS:

Errors introduced during production of the DNA template: Majority of errors are introduced during DNA template production by PCR based protocols. One major cause is the intrinsic error rate in incorporation of nucleotides by the thermostable DNA polymerases. Even polymerases which have an inherent proof reading function can end up with PCR products containing a mixture of different sequences.

SEQUENCING IN ANATOMICAL PATHOLOGY

INVASIVE CARCINOMA BREAST - Both the BRCA 1 and BRCA 2 genes are known to be mutated in families with high risk of breast cancer. These mutations are extremely rare in sporadic cases of breast cancer. The

problem with BRCA gene mutations is that there are a large number of mutations. Over 1500 mutations have been characterized till date. It is quite impossible to develop a standardized test for the evaluation of all these mutations. Therefore, PCR followed by sequencing of specific regions of the gene remains the main method of testing.

RAS GENE MUTATIONS - The RAS gene is commonly mutated in cancers like colonic, lung, pancreatic, and thyroid cancers. It is also commonly mutated in meningiomas and several other tumours. RAS gene mutations were first reported in the 1980's. There are three cellular homologues of viral oncogenes. These are HRAS, KRAS and NRAS. The most common mutations that occur in the KRAS gene are the mutations at codon 12 and codon 13. Less commonly, mutations at codon 61 occur. KRAS mutations are usually tested by sequencing.

P53 MUTATIONS - Inactivating mutations in TP53 tumor suppressor genes are the most common genetic events in human cancers. Majority of these arise from a single point mutation in the segment encoding the DNA-binding domain of TP53. These mutations render the mutant TP53 protein unable to carry out its normal functions, i. e., transcriptional transactivation of downstream target genes that regulate cell cycle and apoptosis. Most mutations cluster in the TP53 DNA binding domain, which encompasses exons five through eight and spans approximately 180 codons or 540 nucleotides. Analysis of the p53 mutations is usually carried out by PCR of exons 5 to 8 followed by sequencing.

OTHER USES IN ANATOMICAL PATHOLOGY

Large-scale re-sequencing of human genes has identified generally between 10 and 100 mutations in each tumor. The percentage of silent mutations is often quite high. However, careful analysis has led to the prediction that a limited number of the newly identified mutations other than TP53, KRAS, etc., are biologically significant. In future, it appears that the PCR followed by sequencing is likely to play an increasingly important role in pathology.

SEQUENCING IN GENETIC DISORDERS - Two disorders will be dealt with in this section, Von Hippel Lindau disease and Connexin gene mutations in sensorineural deafness.

Von Hippel-Lindau (VHL) disease - It is a hereditary cancer syndrome caused by germline mutations in the VHL tumor suppressor gene. The VHL gene contains three exons and encodes a mRNA of 4.5 kb. Germline mutations were identified in the latter half of exon 1, in the first half of exon 3, and in some part of exon 2. Missense, frameshift and nonsense mutations are known to occur along with deletions. Given the wide spectrum of mutations, the only standardized method for screening mutations is by sequencing.

Sensorineural deafness - Most hereditary hearing loss is inherited in a recessive manner, accounting for approximately 85% of non-syndromic hearing loss (NSHL). Deafness associated with DFNB1 locus on chromosome 13q11 is prevalent in many parts of the world. Two genes localised in this chromosomal region have been implicated in deafness. These include connexin26 (Cx26, gene symbol GJB2) and connexin 30 (Cx30, GJB6). The mutations in these regions are multiple and include missense, frameshift and

nonsense mutations. Given the large number of mutations, sequencing has been adopted as the standard method for mutation analysis.

SEQUENCING IN HEMATOLOGY - There are few instances of use of sequencing in hematology. Studies have mentioned sequencing as an adjunct investigation for clonality assessment in lymphomas. However, by and large, hematology does not use sequencing as an investigative modality; DNA and RNA based PCRs are preferred. Although not being done, a possible use of sequencing in the analysis of Factor VIII mutations is being outlined.

The factor VIII gene is extremely large (~ 180 kb) and structurally complex (comprising of 26 exons). Direct nucleotide sequence analysis using automatic DNA sequencers is becoming more mainstream and confident results can be expected for male DNA (hemizygous). Sequencing should be interpreted cautiously for female DNA because heterozygosity may fail to show in the sequencing data. Multiplex amplification of all of the essential regions of factor IX gene in a single PCR, followed by sequencing, represents a step forward and could be applied to factor VIII gene as well.

TROUBLESHOOTING DNA SEQUENCING: Although DNA sequencing usually works, there are times when it doesn't. This can be extremely irritating because a lot of painstaking work has gone into preparing the reaction. However, it must be mentioned that the causes of a failed sequencing are not many and usually the mistakes are amenable to correction. Some of these are given in table 11. 1.

Table 11. 1 : DNA sequencing reaction failures

Causes

Solution

Degraded/ poor quality/ absent PCR product. The reasons for this are numerous and have been explained in detail in chapter 5.

Under these circumstances, it is best to repeat the PCR reaction.

Poor quality DNA. Very common when sequencing plasmid miniprep templates.

The best way of avoiding this problem is to not sequence plasmid DNA and sequence a PCR amplified fragment of the plasmid insert. If this is not possible, it is recommended that a plasmid miniprep kit is used. One tip is to perform a final ethanol precipitation on the kit purified plasmid DNA. This often solves problems with the quality of the template.

Loss of reaction during clean-up. This can be a particular problem when using ethanol precipitation clean-up protocols.

This can be avoided by not using an ethanol precipitation protocol to clean up the sequencing reaction. Commercial kits are available for cleaning up PCR products. These kits work very well but maybe expensive. However, when one considers the other expenses involved in the sequencing process, use of a commercial kit adds only slightly to the entire process.

Bad water. The water used contains sequencing inhibitors.

Inhibitors can end up in lab water stocks that can kill DNA sequencing reactions. If there is a problem with water, it is best to throw out the water and use a fresh stock - remember water is cheap.

Degradation of Taq DNA polymerase or dye labeled nucleotides.

If this is suspected, then it is advisable to perform a control sequencing reaction before undertaking a large number of experimental reactions. Many problems can be prevented by storing the chemicals in small aliquots and avoiding repeated freeze/thaw cycles.

Blocked capillary. The capillaries need to be maintained as per protocols.

Can be identified by tracking trace quality on a trace by trace basis.

Box 8. 2

In automated sequencing, the product is labeled with a fluorescent dye that can be detected using a detector system.

Either the primers or the dideoxy sequences can be labeled.

A four colour laser induced fluorescence detector detects the reactions fragments as they migrate past.

Capillary based systems considerably reduce the run rate.

In automated sequencing, a PCR is initially run and the PCR product is either cloned and sequenced or sequenced without cloning.

The major source of error in direct DNA sequence analysis is because of error introduced during the production of DNA template. A mixture of different sequences maybe produced.

Causes of failed DNA sequencing reactions are because of degraded/ poor quality/ absent PCR product, poor quality DNA, degradation of Taq or dye labeled nucleotides and blocked capillaries.

PYROSEQUENCING

Ever since Sanger brought out his method to make sequencing the simple procedure it is today, workers have been looking for methods to improve sequencing. The main methods which are likely to be useful are sequencing by hybridization, parallel signature sequencing based on ligation and cleavage and pyrosequencing.

Pyrosequencing is a DNA sequencing technique that is based on the detection of the released pyrophosphate (PPi) during DNA synthesis. In a cascade of enzymatic reactions, visible light is generated that is proportional to the number of incorporated nucleotides.

Initially, there is a nucleic acid polymerization reaction in which an inorganic PPi is released as a result of nucleotide incorporation by polymerase. The released PPi is subsequently converted to ATP by ATP sulfurylase, which provides the energy to luciferase to oxidize luciferin and generate light. Because the added nucleotide is known, the sequence of the template can be determined.

The reaction of the pyrosequencing reaction is as follows:

<https://assignbuster.com/maxam-gilbert-and-sangers-method-of-sequencing/>

$(NA)_n + \text{Nucleotide Polymerase} (NA)_{n+1} + \text{Pyrophosphate (PPi)}$

Pyrophosphate ATP Sulfurylase ATP

ATP + Luciferin + Oxygen Luciferase AMP + Pyrophosphate + Oxyluciferin +
CO₂ + Light

It is to be remembered that dATP is a substrate for Luciferase. The addition of dATP $\hat{I}\pm S$ was a considerable improvement since dATP $\hat{I}\pm S$ was found to be inert for luciferase, yet could be incorporated efficiently by all DNA polymerases tested. The last step includes the addition of Apyrase. Apyrase, in the pyrosequencing reaction system, efficiently degrades the unincorporated nucleoside triphosphates to nucleoside diphosphates and subsequently to nucleoside monophosphate.

The sequence of nucleotides in the reaction is read as a pyrogram shown in fig 11. 4.

The problem in a pyrosequencing reaction is that the length of the sequences that can be analysed is usually quite small. Therefore, it is used mainly to confirm the sequences that have already been established. It may also be used in the analysis of hair pin structures which may not be amenable to sequencing by standard methods.