DNA SEQUENCING

Introduction

Sequencing in terms of molecular genetics and biochemistry means the determination of the sequence of the components of an unbranched biopolymer like DNA and certain proteins. It is the elucidation of the primary structure or the primary sequence. The result of sequencing is the linear depiction, which is called as sequence that gives summary of the atomic level structure of the sequenced molecule.

We know that the sequence of DNA contains coded information for the survival and reproduction by the organism. Therefore, sequencing holds equal importance in pure or fundamental research as well as in applied sciences. Knowledge of DNA sequence is now basically a useful information for any biological research work. In the field of medicine, it is useful to identify, diagnose and also to treat the genetic diseases and in case of pathogens it helps treatments of contagious diseases. For biotechnology, a growing discipline, it provides techniques of novel utilities.

The term DNA sequencing refers to sequencing methods for determining the order of the nucleotide bases viz., adenine, guanine, cytosine and thymine in the DNA molecule.

The sequence of DNA constitutes the heritable genetic information in nuclei, plasmids, mitochondria and chloroplasts that forms the basis for the developmental programs of all living organisms. Determining the DNA sequence has, therefore, become indispensable to conduct biological research in numerous applied fields such as diagnostic, biotechnology, forensic biology and biological systematics. Modern DNA sequencing technology has been instrumental in the sequencing of the human genome, in the Human Genome Project.

With reference to nucleic acid, DNA sequencing determines the order of nucleotides in a given DNA fragment. DNA sequencing includes several methods and technologies that are used for determining the order of the nucleotide bases adenine, guanine, cytosine, and thymine in a molecule of DNA. From past several years till now the DNA sequencing is done by chemical cleavage method developed by Allan Maxam and Walter Gilbert and also by chain termination method developed by Frederick Sanger. In 1977 two separate methods for sequencing DNA were developed: The chain termination method or cycle sequencing (Sanger and Coulson) and the chemical degradation method or Maxam-Gilbert sequencing (Maxam and Gilbert).

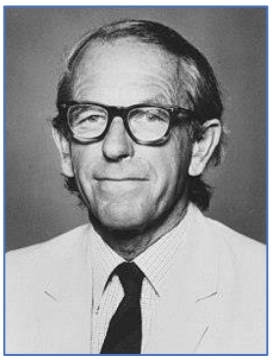
The former procedure was published a little earlier than the latter one but both of them appeared in the journal of National Academy of Sciences.

[**Interesting Fact**: Sanger was born at Rendcombe, Gloucestershire, England. He won his first Nobel Prize in 1958 for sequencing a protein structure, insulin and for the second time he was awarded Nobel Prize with Gilbert in 1980 and the third of awardees was Paul Berg. Sanger lived and worked both in England and USA. Walter Gilbert was from Boston, Massachusetts, USA. These two methods appeared respectively in following publications:

Maxam, A., and Gilbert, W. (1977). A new method of sequencing DNA. Proceedings of National Academy of Sciences, USA, 74, 560-64.

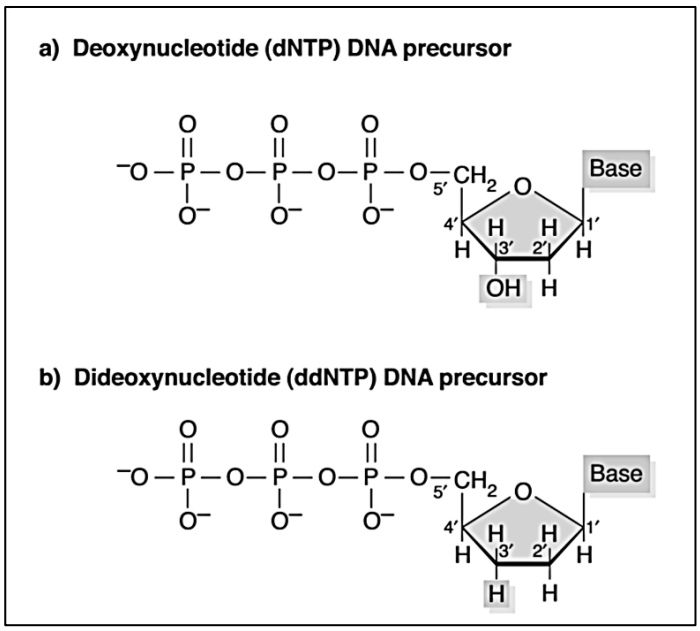
Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proceedings of National Academy of Sciences, USA, 74, 5463-67.]

SANGER'S SEQUENCING (DIDEOXY CHAIN TERMINATION METHOD OF DNA SEQUENCING)



Frederick Sanger

This chain termination method uses single stranded (ss) DNA. This method is prevalently popular and it has also been automated for very large sequencing such as "Human Genome Project". **The method requires DNA polymerase I enzyme from *E. coli* for making copies of single stranded DNA that has to be sequenced**. DNA polymerase I assembles the four deoxy-nucleoside triphosphates viz. dATP, dCTP, dGTP, and dTTP, into a polynucleotide chain that is complementary to the strand which is being replicated for being sequenced and hence is a template strand.



This Sanger's sequencing or chain termination or dideoxy method is based on two properties of DNA polymerases: the polymerases synthesize precisely correct complimentary copy of a single stranded DNA template and they can use **2', 3'-dideoxynucleotides tri-phosphate (ddNTP)** as substrates. **The 3' end of a dideoxynucleotide lacks a hydroxyl group and has only hydrogen so once this nucleotide is incorporated into the chain at the growing point**, the elongation of the chain stops at the ddNTP and no longer acts as a substrate for chain elongation. In practice, Klenow fragment of DNA polymerase I is used since it does not have 5 '- 3' exonuclease activity which is a property of intact enzyme. All DNA synthesis requires a primer, and hence a chemically synthesized oligonucleotide is used as a primer that gets annealed close to the sequence that has to be sequenced. (*The Klenow fragment is a large protein fragment produced when DNA polymerase I from E. coli is enzymatically cleaved by the protease subtilisin. It retains the 5' → 3' polymerase activity and the 3' → 5' exonuclease activity for removal of precoding nucleotides and proofreading, but loses its 5' → 3' exonuclease activity, i.e., the primer removal activity*).

The major steps are as follows:

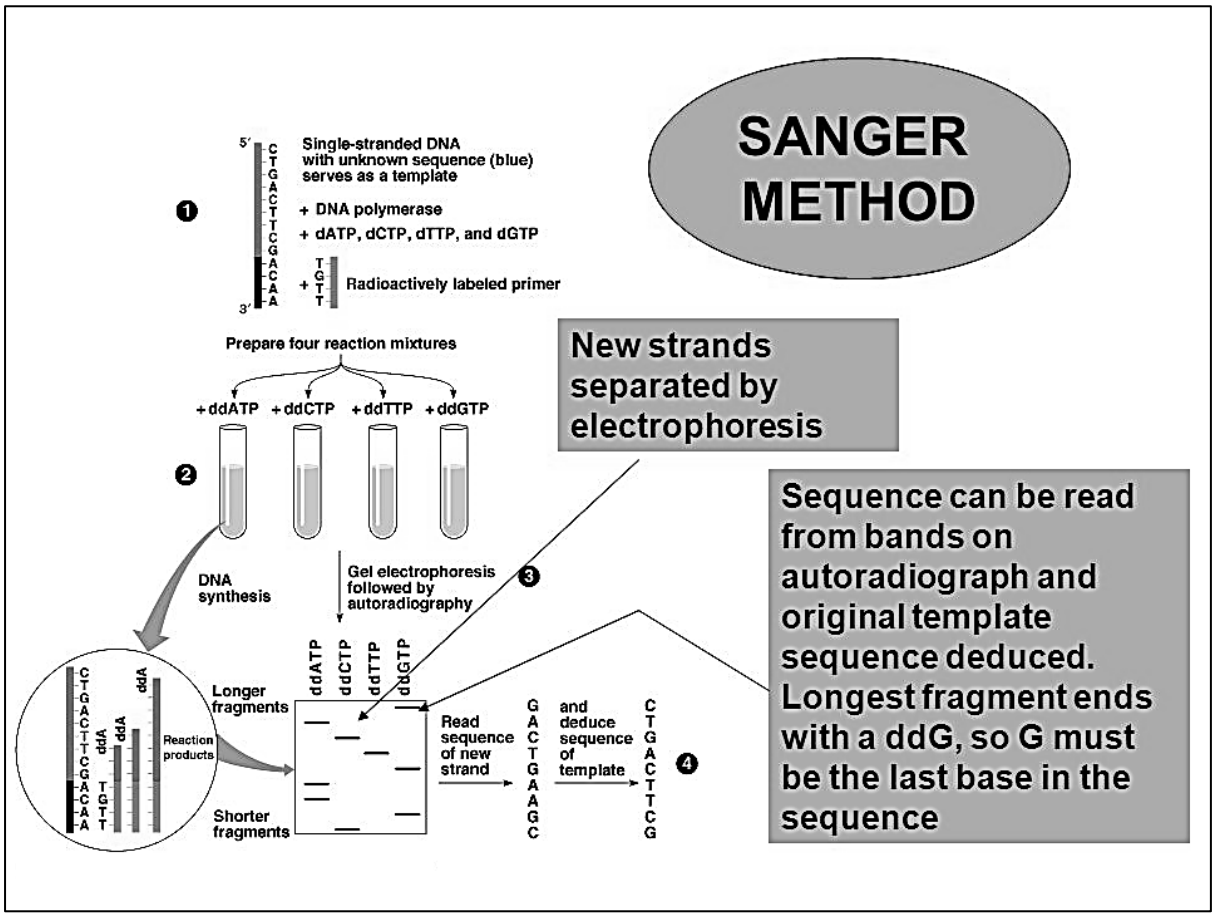
In brief, this chain termination method uses single stranded DNA and the fragment to be sequenced is first to cloned in M13 phage vector. This method is based on interruption by dideoxy analogues of nucleotides, of enzyme mediated synthesis of a second strand of DNA, which will be complementary to the fragment being sequenced. The consequence is the formation of a mixture of fragments of different lengths and the length will depend on the occurrence of the point of interruption. Finally, the mixture is run on a gel and the sequences are read.

1. The sample sequence to be sequenced is spliced in to M13 vector DNA and following this the infected *E. coli* host cell releases phage particles containing single stranded recombinant DNA that includes the sample sequence as well. DNA is extracted from phage and is put for sequencing.
2. A short oligonucleotide primer, a chemically synthesized one, is added to SS recombinant DNA. This primer anneals at a position that will be the starting point of the synthesis for the new i.e., complementary strand.
   [*DNA polymerase I has a 5' to 3' exonuclease activity that is to say it degrades the DNA by one nucleotide at a time from its 5' end. This degradation is catalysed by an active site, which is separate from polymerising site. The proteolytic cleavage of the enzyme gives two fragments one larger segment, called Klenow segment, has full polymerase activity, while smaller fragment has the 5' - 3' exonuclease activity.*]
   Therefore, only Klenow fragment is useful for DNA sequencing as it ensures that all replicated chains have the same 5' terminus.
3. The Klenow fragment is now added in the presence of: Four normal deoxynucleotides viz., dATP, dCTP, dGTP, and dTTP. One or more of the nucleotides may be radio labelled with ^{32}P. Four separate sets of the reaction mixtures are to be incubated with Klenow fragment. **A small amount of the 2',3'-dideoxynucleotide triphosphate (ddNTP) of any one of the bases is also added to one reaction mixture**. Thus, each reaction mixture has a different ddNTP, which is identical to normal deoxynucleotide (dNTP) except for the substitution of hydroxyl groups (OH) in the sugar ring by the hydrogen (H). DNA polymerase as normal substrate uses these analogues, as this enzyme does not distinguish the two. But as the above analogue is added the chain growth stops because the absence of 3 '-OH does not allow further bonding of incoming nucleotides.
4. Now on incubation the complementary synthesis begins away from primer end and also away from 5' end; but when an analogue becomes incorporated in to the new strand the chain is terminated and further elongation is stopped for the lack of OH group at 3' end [the presence of OH group allows further bond formation in a normal chain elongation reaction]. Consequently, in each of the four separately incubated reaction mixes there shall be double stranded DNA molecules that shall be partially synthesized and radioactively labelled. These DNA fragments will be of varying sizes depending on the point of incorporation of dd (dideoxy) analogue. As the incorporation is purely random, presumably the total population of synthesized DNA molecules will be a mixture of smallest to largest possible

fragments that will represent every position of a particular base in the total fragment that is being sequenced.

5. The above said four reaction mixtures are simultaneously run on a sequencing gel to separate fragments depending on their sizes. This gel contains urea which denatures the double stranded DNA and minimizes DNA secondary structures. The process is run at high voltage so that high temperature is generated that prevents annealing. The DNA strand in the gel are read or observed by autoradiography as ^{32}P used for labelling of dNTPs will form a fog on the photographic film.

Procedure for reading the DNA sequence from the autoradiogram: The DNA bands are then visualized by autoradiography or UV light and the DNA sequence can be directly read off the X-ray film or gel image. The X-ray film is exposed to the gel and the dark bands corresponding to DNA fragments of different lengths can be read. A dark band in a lane indicates a DNA fragment that is the result of chain termination after incorporation of a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP). The relative positions of the different bands among the four lanes are then used to read (from bottom to top) the DNA sequence. The autoradiogram is always read from the anode to the cathode. The sequence is that of the nascent (new) strand, which is in 5'→ 3' direction. The complementary strand can be deduced from this in 3'→ 5' direction.



MAXAM–GILBERT SEQUENCING

Walter Gilbert

Maxam–Gilbert sequencing is a method of DNA sequencing developed by Allan Maxam and Walter Gilbert in 1977–1980. This method is based on nucleobase-specific partial chemical modification of DNA and subsequent cleavage of the DNA backbone at sites adjacent to the modified nucleotides.

Maxam–Gilbert sequencing was the first widely adopted method for DNA sequencing, and, along with the Sanger dideoxy method, represents the first generation of DNA sequencing methods. Maxam–Gilbert sequencing is no longer in widespread use, having been supplanted by next-generation sequencing methods.

The fragments in the four reactions are electrophoresed side by side in denaturing acrylamide gels for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each showing the location of identical radiolabelled DNA molecules. From presence and absence of certain fragments the sequence may be inferred.

The sequencing requires DNA molecules, either double-stranded or single-stranded, that are labelled at one end of one strand with ^{32}P. DNA can be made single stranded by treating it with NaOH. The whole procedure requires modification of bases of DNA and then its base specific chemical cleavage.

1. The first step is the radioactive labelling of one end of DNA, preferably 5' end, with a radioactive ^{32}P.
2. The two DNA strands are separated by using dimethyl sulphate and heating to 90°C and are then separated and purified by gel electrophoresis, for example. The separation is achieved as one of the strands is supposed to be heavier than the other for the presence of more purine nucleotides i.e., adenine and guanine than pyrimidines i.e. cytosine and thymine.
3. The single stranded sample is divided in to separate samples and then each sample is treated with one of the cleaving reagents. This step involves alteration of bases followed by removal of altered bases.
4. Finally, piperidine is applied to cleave the strands at the points where bases are missing.
5. If reactions have been arranged to give only one, or a few, cleavages per DNA molecule, the consequence is the production of a nested set of end-labelled DNA fragments of different lengths.
6. The samples are run together on a sequencing gel, which separates the fragments, by electrophoresis depending on their size. DNA bands in the gel are visualized by autoradiography (^{32}P-labelled 5' end fogs photographic film).
7. The DNA sequence is read directly from the gel.

G base: Reaction of DNA with dimethyl sulphate (DMS) methylates, G bases at N(7), and so the glycosidic bond of methylated residue becomes susceptible to hydrolysis and so the following treatment with piperidine will cleave the polynucleotide chain at that point before the depurinated residue.

A+G Bases: DMS is not suitable for this purpose as it methylates A residue at N(3) rather than N(7) and so cleavage at A residue would be only al 1/5th of the rate as compared with that for G residue. But the treatment with **formic acid** releases both A and G at comparable rates and yields the depurinated product. The following treatment with piperidine cleaves the strand before both A and G residues and A residues arc identified by comparing the position of G and A + G cleavages.

C+T Bases: As indicated in the table DNA is cleaved before its C and T residues if it is treated with hydrazine (NH_2 - NH_2) followed by a treatment with piperidine.

C Base: If DNA is reacted with solution of hydrazine in 1.5 M NaCl, only C residues will react, notably, and then the comparison of C and C+T cleavage positions helps identification of T residues.

In brief:
Maxam–Gilbert sequencing requires ^{32}P radioactive labelling at one 5′ end of the DNA fragment to be sequenced and purification of the DNA. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T). For example, the purines (A+G) are depurinated using formic acid, the guanines (and to some extent the adenines) are methylated by dimethyl sulphate, and the pyrimidines (C+T) are hydrolysed using hydrazine. The addition of salt (sodium chloride) to the hydrazine reaction inhibits the reaction of thymine for the C-only reaction. The modified DNAs may then be cleaved by hot piperidine; $(CH_2)_5NH$ at the position of the modified base. The concentration of the modifying chemicals is controlled to introduce on average one modification per DNA molecule. Thus, a series of labelled fragments is generated, from the radiolabelled end to the first "cut" site in each molecule.

Separation of Cleavage Fragments: The samples are run together on a sequencing gel so that the fragments are eletrophoretically separated depending on their size. DNA bands are examined by autoradiography. Usually polyacrylamide gel electrophoresis is applied, to separate the fragments according to their sizes. Hence, the position of the chosen residue may be identified in the DNA fragment with reference to the relative position of the ^{32}P labelled fragments on the gel with the help of chromatography. However, unlabelled fragments will escape the observation. The gel must be efficient enough to separate fragments that differ for even one nucleotide.

Figure 4A.4 Sequencing an oligonucleotide by the Maxam-Gilbert method