**DNA SEQUENCING**

 The determination of the order or sequence of bases along a length of DNA is one of the central techniques in molecular biology. Although it is now possible to derive amino acid sequence information with a degree of reliability it is frequently more convenient and rapid to analyse the DNA coding information. The precise usage of codons, information regarding mutations and polymorphisms and the identification of gene regulatory control sequences are also only possible by analysing DNA sequences. Two techniques have been developed for this, one based on an enzymatic method frequently termed Sanger sequencing after its developer, and a chemical method called Maxam and Gilbert, named for the same reason. At present Sanger sequencing is by far the most popular method and many commercial kits are available for its use. However, there are certain occasions such as the sequencing of short oligonucleotides where the Maxam and Gilbert method is more appropriate.

 One absolute requirement for Sanger sequencing is that the DNA to be sequenced is in a single-stranded form. Traditionally this demanded that the DNA fragment of interest be inserted and cloned into a specialised bacteriophage vector termed M13 which is naturally single-stranded. Although M13 is still universally used the advent of the PCR has provided the means not only to amplify a region of any genome or cDNA but also very quickly generate the corresponding nucleotide sequence. This has led to an explosion in the accumulation of DNA sequence information and has provided much impetus for gene discovery and genome mapping.

**Sanger’s method for DNA sequencing**

 **Sanger sequencing**, also known as the “chain termination method”, is a method for determining the nucleotide sequence of DNA. The method was developed by two time Nobel Laureate Frederick Sanger and his colleagues in 1977, hence the name the **Sanger Sequence**. Sanger sequencing can be performed manually or, more commonly, in an automated fashion via sequencing machine. Each method follows three basic steps, as described below.



There are three main steps to Sanger sequencing.

**1. Chain Termination PCR**

 The DNA sequence of interest is used as a template for a special type of [**PCR**](https://www.sigmaaldrich.com/life-science/molecular-biology/pcr/routine-amplification.html) called chain-termination PCR. Chain-termination PCR works just like standard PCR, but with one major difference: the addition of modified nucleotides (dNTPs) called dideoxyribonucleotides (ddNTPs). In the extension step of standard PCR, DNA polymerase adds dNTPs to a growing DNA strand by catalyzing the formation of a phosphodiester bond between the free 3’-OH group of the last nucleotide and the 5’-phosphate of the next. In chain-termination PCR, the user mixes a low ratio of chain-terminating ddNTPs in with the normal dNTPs in the PCR reaction. ddNTPs lack the 3'-OH group required for phosphodiester bond formation; therefore, when DNA polymerase incorporates a ddNTP at random, extension ceases. The result of chain-termination PCR is millions to billions of oligonucleotide copies of the DNA sequence of interest, terminated at a random lengths (n) by 5’-ddNTPs.

* In **manual** Sanger sequencing, four PCR reactions are set up, each with only a single type of ddNTP (ddATP, ddTTP, ddGTP, and ddCTP) mixed in.
* In **automated** Sanger sequencing, all ddNTPs are mixed in a single reaction, and each of the four dNTPs has a unique fluorescent label.

**2. Size Separation by Gel Electrophoresis**

 In the second step, the chain-terminated oligonucleotides are separated by size via gel electrophoresis. In gel electrophoresis, DNA samples are loaded into one end of a gel matrix, and an electric current is applied; DNA is negatively charged, so the oligonucleotides will be pulled toward the positive electrode on the opposite side of the gel. Because all DNA fragments have the same charge per unit of mass, the speed at which the oligonucleotides move will be determined only by size. The smaller a fragment is, the less friction it will experience as it moves through the gel, and the faster it will move. In result, the oligonucleotides will be arranged from smallest to largest, reading the gel from bottom to top.

* In **manual** Sanger sequencing, the oligonucleotides from each of the four PCR reactions are run in four separate lanes of a gel. This allows the user to know which oligonucleotides correspond to each ddNTP.
* In **automated** Sanger sequencing, all oligonucleotides are run in a single capillary gel electrophoresis within the sequencing machine.

**3. Gel Analysis & Determination of DNA Sequence**

 The last step simply involves reading the gel to determine the sequence of the input DNA. Because DNA polymerase only synthesizes DNA in the 5’ to 3’ direction starting at a provided primer, each terminal ddNTP will correspond to a specific nucleotide in the original sequence (e.g., the shortest fragment must terminate at the first nucleotide from the 5’ end, the second-shortest fragment must terminate at the second nucleotide from the 5’ end, etc.) Therefore, by reading the gel bands from smallest to largest, we can determine the 5’ to 3’ sequence of the original DNA strand.

* In **manual** Sanger sequencing, the user reads all four lanes of the gel at once, moving bottom to top, using the lane to determine the identity of the terminal ddNTP for each band. For example, if the bottom band is found in the column corresponding to ddGTP, then the smallest PCR fragment terminates with ddGTP, and the first nucleotide from the 5’ end of the original sequence has a guanine (G) base.
* In **automated** Sanger sequencing, a computer reads each band of the capillary gel, in order, using fluorescence to call the identity of each terminal ddNTP. In short, a laser excites the fluorescent tags in each band, and a computer detects the resulting light emitted. Because each of the four ddNTPs is tagged with a different fluorescent label, the light emitted can be directly tied to the identity of the terminal ddNTP. The output is called a chromatogram, which shows the fluorescent peak of each nucleotide along the length of the template DNA.



Figure: Basic protocol for Sanger’s sequencing method.

**Direct PCR pyrosequencing**

 Rapid PCR sequencing has also been made possible by the use of pyrosequencing. This is a sequencing by synthesis whereby a PCR template is hybridised to an oligonucleotide and incubated with DNA polymerase, ATP sulphurylase, luciferase and apyrase. During the reaction the first of the four dNTPs are added and if incorporated release pyrophosphate (PPi). The ATP sulphurylase converts the PPi to ATP which drives the luciferase-mediated conversion of luciferin to oxyluciferin to generate light. Apyrase degrades the resulting component dNTPs and ATP. This is followed by another round of dNTP addition. A resulting pyrogram provides an output of the sequence. The method provides short reads very quickly and is especially useful for the determination of mutations or SNPs.

**PCR cycle sequencing**

 One of the most useful methods of sequencing PCR amplicons is termed PCR cycle sequencing. This is not strictly a PCR since it involves linear amplification with a single primer. Approximately 20 cycles of denaturation, annealing and extension take place. Radiolabelled or fluorescent-labelled dideoxynucleotides are then introduced in the final stages of the reaction to generate the chain-terminated extension products. Automated direct PCR sequencing is increasingly being refined allowing greater lengths of DNA to be analysed in one sequencing run and provides a very rapid means of analysing DNA sequences.

**Automated fluorescent DNA sequencing**

 Advances in fluorescent dye terminator and labelling chemistry have led to the development of high-throughput automated sequencing techniques. Essentially most systems involve the use of dideoxynucleotides labelled with different fluorochromes. Thus the label is incorporated into the ddNTP and this is used to carry out chain termination as in the standard reaction. The advantage of this modification is that since a different label is incorporated with each ddNTP it is unnecessary to perform four separate reactions. Therefore the four chain-terminated products are run on the same track of a denaturing electrophoresis gel. Each product with its base-specific dye is excited by a laser and the dye then emits light at its characteristic wavelength. A diffraction grating separates the emissions which are detected by a charge-coupled device (CCD) and the sequence is interpreted by a computer. The advantages of the technique include real-time detection of the sequence. In addition the lengths of sequence that may be analysed are in excess of 500 bp.



Figure: Automated fluorescent sequencing detection using single-lane gel and charge-coupled device.

**Maxam and Gilbert sequencing**

 Sanger sequencing is by far the most popular technique for DNA sequencing; however, an alternative technique developed at the same time may also be used. The chemical cleavage method of DNA sequencing developed by Maxam and Gilbert is often used for sequencing small fragments of DNA such as oligonucleotides, where Sanger sequencing is problematic. A radioactive label is added to either the 3’ or the 5’ ends of a double-stranded DNA sample. The strands are then separated by electrophoresis under denaturing conditions, and analysed separately. DNA labelled at one end is divided into four aliquots and each is treated with chemicals which act on specific bases by methylation or removal of the base. Conditions are chosen so that, on average, each molecule is modified at only one position along its length; every base in the DNA strand has an equal chance of being modified. Following the modification reactions, the separate samples are cleaved by piperidine, which breaks phosphodiester bonds exclusively at the 5’ side of nucleotides whose base has been modified. The result is similar to that produced by the Sanger method, since each sample now contains radioactively labelled molecules of various lengths, all with one end in common (the labelled end), and with the other end cut at the same type of base.



Figure: Maxam and Gilbert sequencing of DNA. Only modification and cleavage of deoxycytidine is shown, but three more portions of the end-labelled DNA would be modified and cleaved at G, G+A, and T+C, and the products would be separated on the sequencing gel alongside those from the C reactions.