ASMITA ROY

What is sequencing?

DNA sequencing is the process of determining the sequence of nucleotide bases (As, Ts, Cs, and Gs) in a piece of DNA. Today, with the right equipment and materials, sequencing a short piece of DNA is relatively straightforward.

Sequencing an entire genome (all of an organism's DNA) remains a complex task. It requires breaking the DNA of the genome into many smaller pieces, sequencing the pieces, and assembling the sequences into a single long "consensus." However, thanks to new methods that have been developed over the past two decades, genome sequencing is now much faster and less expensive than it was during the Human Genome Project.

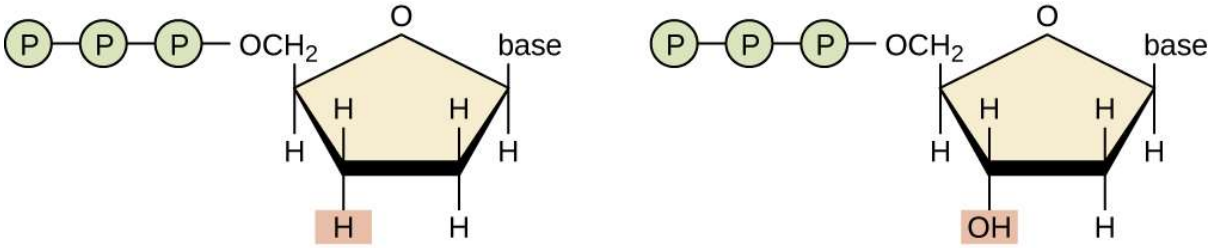
Sanger sequencing: The chain termination method

Regions of DNA up to about 900900900 base pairs in length are routinely sequenced using a method called **Sanger sequencing** or the **chain termination method**. Sanger sequencing was developed by the British biochemist Fred Sanger and his colleagues in 1977.

Its also known as the "chain termination method," was developed by the English biochemist Frederick Sanger and his colleagues in 1977. This method is designed for determining the sequence of nucleotide bases in a piece of DNA (commonly less than 1,000 bp in length). Sanger sequencing with 99.99% base accuracy is considered the "gold standard" for validating DNA sequences, including those already sequenced through next-generation sequencing (NGS). Sanger sequencing was used in the Human Genome Project to determine the sequences of relatively small fragments of human DNA (900 bp or less). These fragments were used to assemble larger DNA fragments and, eventually, entire chromosomes.

Principle

A DNA primer is attached by hybridization to the template strand and deoxynucleosides triphosphates (dNTPPs) are sequentially added to the primer strand by DNA polymerase.
The primer is designed for the known sequences at 3′ end of the template strand.
M13 sequences is generally attached to 3′ end and the primer of this M13 is made.
The reaction mixture also contains dideoxynucleoside triphosphate (ddNTPs) along with usual dNTPs.
If during replication ddNTPs is incorporated instead of usual dNTPs in thegrowing DNA strand then the replication stops at that nucleotide.



dideoxynucleotide (ddNTP) deoxynucleotide (dNTP)

The ddNTPs are analogue of dNTPs.
ddNTPs lacks hydroxyl group (-OH) at c3 of ribose sugar, so it cannot make phosphodiester bond with nest nucleotide, thus terminates the nucleotide chain.
Respective ddNTPs of dNTPs terminates chain at their respective site. For example ddATP terminates at A site. Similarly ddCTP, ddGTP and ddTTP terminates at C, G and T site respectively.

Ingredients for Sanger sequencing

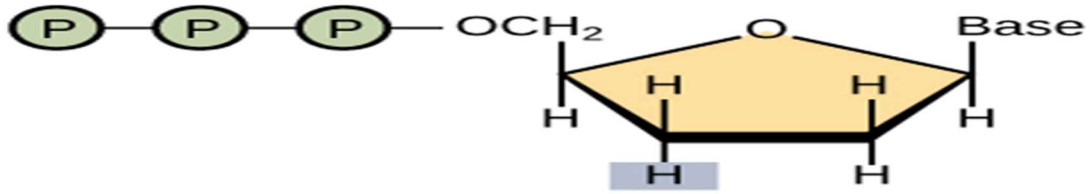
Sanger sequencing involves making many copies of a target DNA region. Its ingredients are similar to those needed for DNA replication in an organism, or for polymerase chain reaction (PCR), which copies DNA *in vitro*. They include:

- A DNA polymerase enzyme

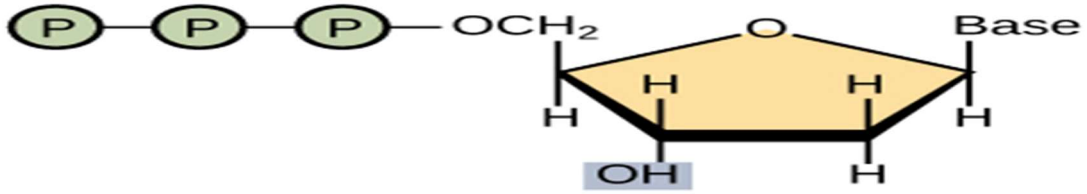
- A **primer**, which is a short piece of single-stranded DNA that binds to the template DNA and acts as a "starter" for the polymerase
- The four DNA nucleotides (dATP, dTTP, dCTP, dGTP)
- The template DNA to be sequenced

However, a Sanger sequencing reaction also contains a unique ingredient:

- Dideoxy, or **chain-terminating**, versions of all four nucleotides (ddATP, ddTTP, ddCTP, ddGTP), each labeled with a different color of dye



Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

Dideoxy nucleotides are similar to regular, or deoxy, nucleotides, but with one key difference: they lack a hydroxyl group on the 3′ carbon of the sugar ring. In a regular nucleotide, the 3′ hydroxyl group acts as a "hook," allowing a new nucleotide to be added to an existing chain. Once a dideoxy nucleotide has been added to the chain, there is no hydroxyl available and no further nucleotides can be added. The chain ends with the dideoxy nucleotide, which is marked with a particular color of dye depending on the base (A, T, C or G) that it carries.

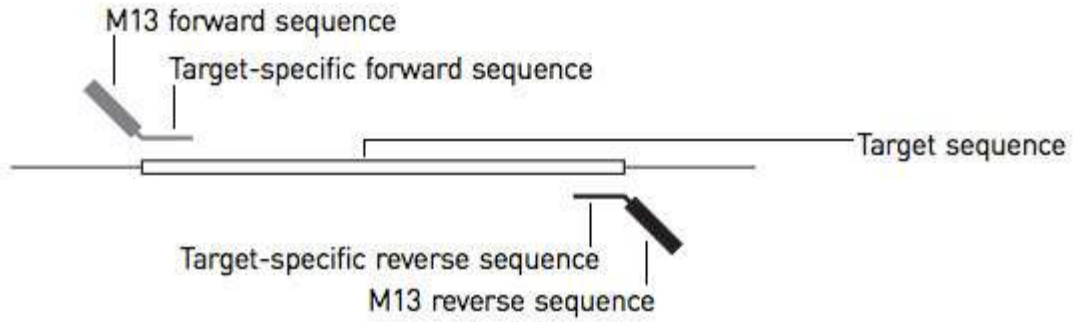
Sanger Sequencing Steps

ASMITA ROY

The Sanger sequencing method consists of 6 steps:
(1) The double-stranded DNA (dsDNA) is denatured into two single-stranded DNA (ssDNA).
(2) A primer that corresponds to one end of the sequence is attached.
(3) Four polymerase solutions with four types of dNTPs but only one type of ddNTP are added.
(4) The DNA synthesis reaction initiates and the chain extends until a termination nucleotide is randomly incorporated.
(5) The resulting DNA fragments are denatured into ssDNA.
(6) The denatured fragments are separated by gel electrophoresis and the sequence is determine

Procedure

1. Template preparation:



i)m13-forward-sequence Copies of template strand to be sequenced must be prepared with short known sequences at 3′ end of the template strand.
ii)A DNA primere is essential to initiate replication of template , so primer preparation of known sequences at 3′end is always required.
iii)For this purpose a single stranded cloning vector M13 is flanked with template strand at 3′end which serves as binding site for primer.

2. Generation of nested set of labelled fragments:

i)Copies of each template is divided into four batches and each batch is used for different replication reaction.
ii)Copies of standard primer and DNA polymerase I are used in all four batches.

iii)To synthesize fragments that terminates at A, ddATP is added to the reaction mixture on batch I along with dATP, dTTP,dCTP and dGTP, standard primer and DNA polymerase I.

iv)Similarly, to generate, all fragments that terminates at C, G and T, the respective ddNTPs ie ddCTP,ddGTP and ddTTP are added respectively to different reaction mixture on different batch along with usual dNTPs.

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.

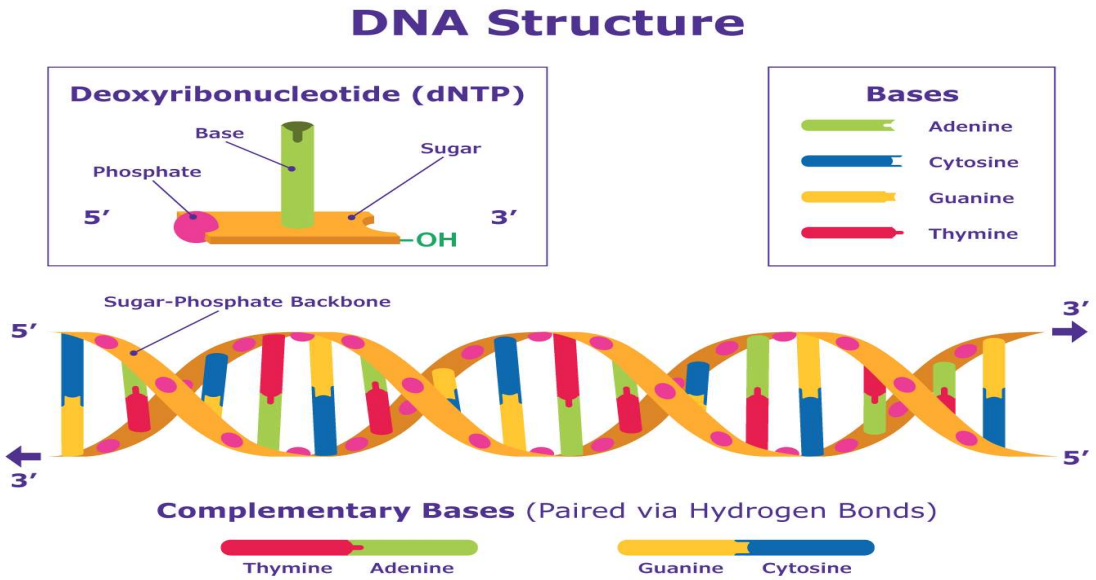


Figure - DNA Structure Schematic. DNA is a molecule composed of two strands that coil around each other to form a double helix. Each strand is made up of a string of molecules called deoxyribonucleotides (dNTPs).

Each dNTP contains a phosphate group, a sugar group, and one of four nitrogenous bases [adenine (A),thymine (T), guanine (G), or cytosine (C)]. The dNTPs are strung together in a linear fashion by phosphodiester covalent bonds between the sugar of one dNTP and the phosphate group of the next; this repeated sugar-phosphate pattern makes up the sugar-phosphate backbone.

The nitrogenous bases of the two separate strands are bound together by hydrogen bonds between complementary bases to form the double-stranded DNA helix.

HOW TO READ SANGER SEQUENCING RESULTS

Reading the Sanger sequencing results properly will depend on which of the two complementary DNA strands is of interest and what primer is available. If the two strands of DNA are A and B and strand A is of interest, but the primer is better for strand B, the output fragments will be identical to strand A. On the other hand, if strand A is of

interest and the primer is better for strand A, then the output will be identical to strand B. Accordingly, the output must be converted back to strand A.

So, if the sequence of interest reads "TACG" and the primer is best for that strand, the output will be "ATGC" and, therefore, must be converted back to "TACG". However, if the primer is better for the complementary strand ("ATGC"), then the output will be "TACG", which is the correct sequence.

In short, before starting, you need to know what you're targeting and how you're going to get there! So keeping this in mind, here is an example of the former example (TACG -> ATGC -> TACG). If the dideoxynucleotides labels are T = yellow, A = pink, C = dark blue, and G = light blue, you will end up with the short sequences primer-A, primer-AT, primer-ATG, and primer-ATGC. Once the fragments have been separated by electrophoresis, the laser will read the fragments in order of length (pink, yellow, light blue, and dark blue) and produce a chromatogram. The computer will convert the letters, so the final sequence is the correct TACG.

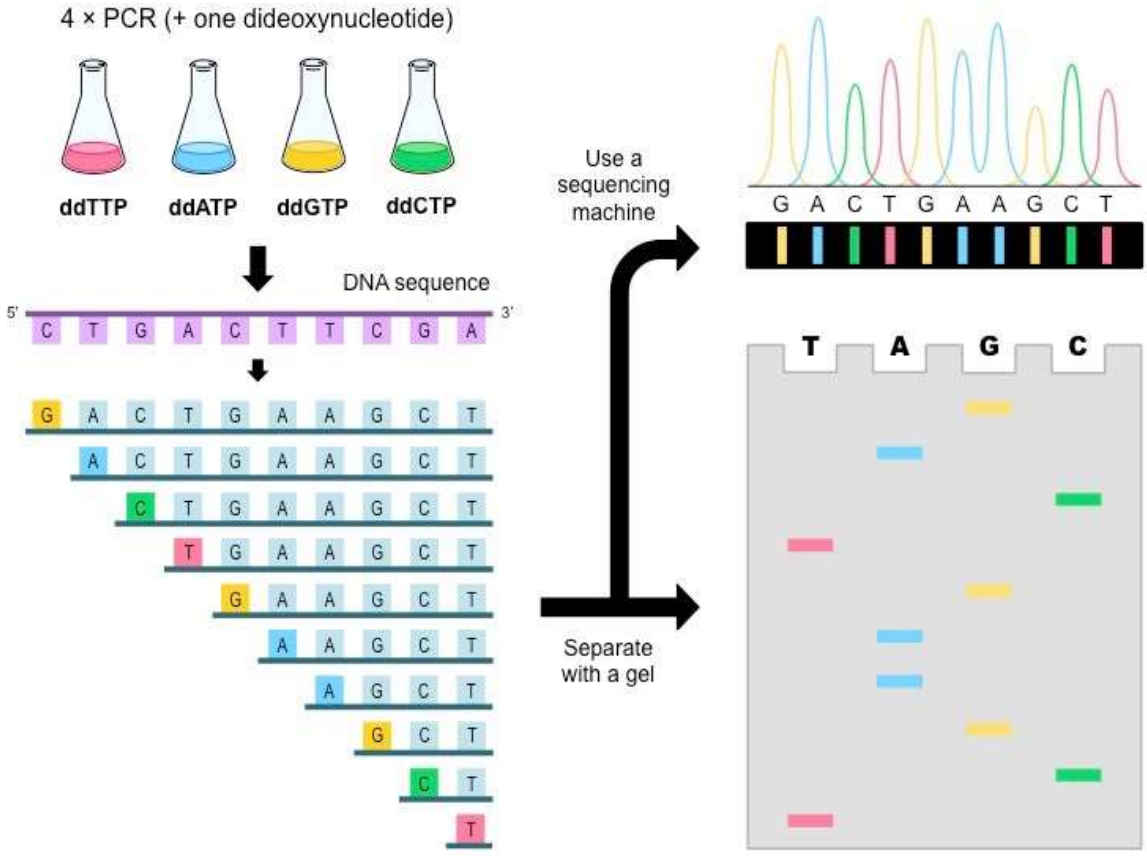
SANGER SEQUENCING VS. PCR

Sanger sequencing and PCR use similar starting materials and can be used in conjunction with each other, but neither can replace the other. PCR is used to amplify DNA in its entirety. While fragments of varying lengths may be produced by accident (e.g., the DNA polymerase might fall off), the goal is to duplicate the entire DNA sequence. To that end, the "ingredients" are the target DNA, nucleotides, DNA primer, and DNA polymerase (specifically Taq polymerase, which can survive the high temperatures required in PCR).

In contrast, the goal of Sanger sequencing is to generate every possible length of DNA up to the full length of the target DNA. That is why, in addition to the PCR starting materials, the dideoxynucleotides are necessary.

Sanger sequencing and PCR can be brought together when generating the starting material for a Sanger sequencing protocol. PCR can be used to create many copies of the DNA that is to be sequenced.

Having more than one template to work from makes the Sanger protocol more efficient. If the target sequence is 1,000 nucleotides long and there is only one copy of the template, it is going to take longer to generate the 1,000 tagged fragments. However, if there are several copies of the template, in theory it will take less time to generate all 1,000 of the tagged fragments.



Uses and limitations

Sanger sequencing gives high-quality sequence for relatively long stretches of DNA (up to about 900900900 base pairs). It's typically used to sequence individual pieces of DNA, such as bacterial plasmids or DNA copied in PCR.