

Sanger Sequencing as the Holy Grail of sequencing technology

*¹Ayobami A. L., ¹Kade E. A., ²Oladimeji K. A., ³Kehinde S.

¹ Department of Microbiology, University of Lagos (UNILAG), Akoka, Lagos, Nigeria.

² National University of Ireland Galway, Ireland and University of Limerick

³ Department of Cell biology and Genetics, University of Lagos (UNILAG), Akoka, Lagos, Nigeria;

* **Correspondence:** ayobamiadesiyan123@gmail.com; ORCID: 0000-0001-8240-5754

Abstract

DNA sequencing methods were first developed more than 20 years ago with the publication of two approaches to sequencing methodology that became known as Sanger sequencing (1), based on enzymatic synthesis from a single-stranded DNA template with chain termination using dideoxynucleotides (ddNTPs) and Maxim-Gilbert sequencing (2), which involved chemical degradation of end-radio-labeled DNA fragments. Both methods relied on four-lane, high resolution polyacrylamide gel electrophoresis to separate the labeled fragment and allow the base sequence to be read in a staggered ladder-like fashion. Sanger sequencing was technically easier and faster, and thus became the main DNA sequencing method for the vast majority of applications.

Keywords: Sanger Sequencing, DNA Sequencing, Chain Termination Sequencing

INTRODUCTION

The DNA stands for deoxyribose nucleic acid. It is the blueprint of life and it is the prime genetic molecule carrying all the hereditary information within chromosomes. Frederick Miescher first observed DNA in the late 1800s. But nearly a century passed from that discovery until researchers unraveled the structure of the DNA molecule and realized its central importance to biology. Most thought that DNA was too simple a molecule to play such a critical role. The importance of DNA became clear in 1953 with the work of James Watson, Francis Crick, Maurice Wilkins and Rosalind Franklin. By studying X-ray diffraction patterns and building models, the scientists figured out the double helix structure of DNA - a structure that enables it to carry biological information from one generation to the next [1, 2].

Sanger sequencing was a critical transition technique leading to the modern generation of methods that have completely dominated sequencing over the past 30 years. The key to this advance was the use of polyacrylamide gels to separate the products of primed synthesis by DNA polymerase in order of increasing chain length. DNA sequence studies had not developed because of the immense size of DNA molecules and the lack of suitable enzymes to cleave DNA into smaller pieces. Building on the enzyme copying

approach used by the Charles Weissmann in his studies on bacteriophage RNA, Sanger began using the enzyme DNA polymerase to make new strands of DNA from single-strand templates, introducing radioactive nucleotides into the new DNA. DNA polymerase requires a primer that can bind to a known region of the template strand [2].

A few problems remained with the plus and minus system. Sanger, Coulson, and British colleague Steve Nicklen developed a similar procedure using dideoxynucleotide chain-terminating inhibitors. DNA was synthesized until an inhibitor molecule was incorporated into the growing DNA chain. Using four reactions, each with a different inhibitor, sets of DNA fragments were generated ending in every nucleotide. For example, in the A reaction, a series of DNA fragments ending in A (adenine) was generated. In the C reaction, a series of DNA fragments ending in C (cytosine) was generated, and so on for G (guanine) and T (thymine). When the four reactions were separated side by side on a gel and an autoradiograph developed, the sequence was read from the film [3].

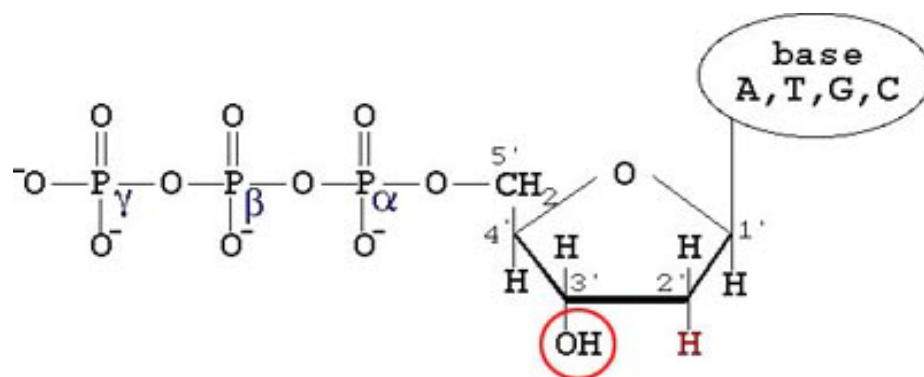
Maxam and Gilbert developed a DNA sequencing method that was similar to the Sanger and Coulson method in using polyacrylamide gels to resolve bands that terminated at each base throughout the target sequence, but very different in the way that products ending in a specific base were generated [3,4].

SANGER SEQUENCING METHOD

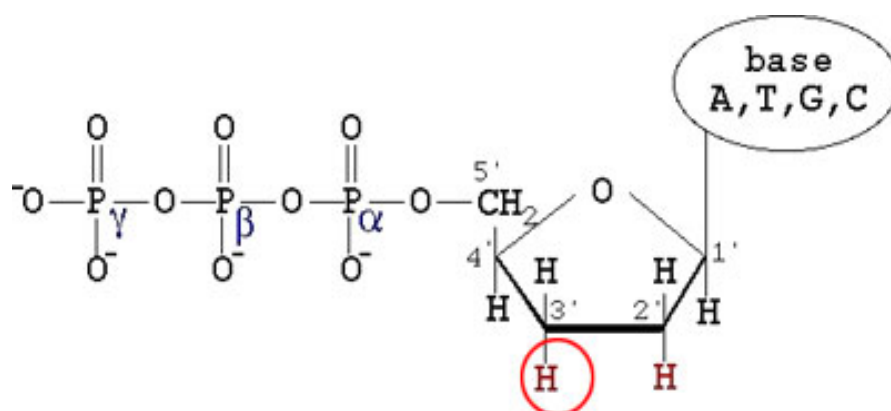
Sanger DNA sequencing relies on the use of a special molecule termed 2' 3' dideoxynucleotide triphosphate(ddNTP) which are ddATP, ddGTP, ddCTA and ddTTA. These molecules are structurally different from the normal deoxyribonucleotide triphosphate (dNTP) which are required for the elongation of the growing DNA strand by a DNA polymerase through the formation of phosphodiester bond. These special molecules sometimes called unusual sugar are synthesized through a well programmed targeted deprotonation of 3' carbon atom of the deoxyribose ring [2].

2'3' Dideoxynucleotide Triphosphate

2'3' Dideoxynucleotide Triphosphate are deoxyribonucleotid (dNTP) analogue with a triphosphate and a base group bound to 5' and 1' of a pentose sugar respectively and with a 3' hydroxyl group of a dNTP replaced with a hydrogen molecule. The 3' hydroxyl group is essential for chain elongation and without it, DNA polymerase cannot add further dNTP. The selective incorporation of this chain terminating molecule (ddNTP) by a DNA polymerase to a growing strand of a nucleotide prevents further elongation and when the complete set of the reaction mixture is run through a gel electrophoretic medium, they are separated according to their sizes which will help in the sequencing of the DNA [3, 4, 5].



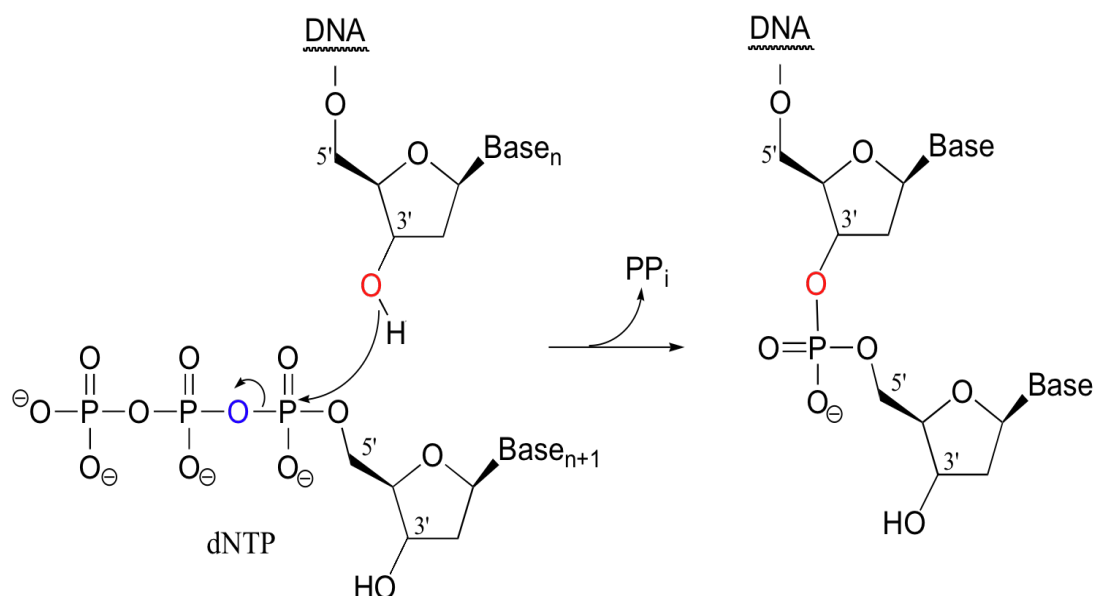
dNTP
deoxyribonucleotide triphosphate



ddNTP
dideoxynucleotide triphosphate

Mechanism of Formation and Prevention of Phosphodiester Bond

A phosphodiester bond is bond that that connects two nucleotide together in the backbone of a DNA molecule. It's composed of two ester bond (di-ester) joined together by a phosphate group. It's formed through a nucleophilic attack of a nucleophile (oxygen molecule of -OH) on an electrophile (phosphate molecule of the alpha phosphate group). A nucleophile is a chemical specie that donates a pair of an electron to an electrophile for the formation of a chemical bond in relation to the reaction. An electrophile is that chemical specie with the tendency of accepting an electron pair from a nucleophile [4].



When a dNTP is incorporated into a growing DNA strand

In the formation of a phosphodiester bond through the incorporation of a dNTP as in the above diagram, the oxygen molecule of $-OH$ group attached to the 3' carbon atom of the pentose ring of the incoming dNTP, initiates a nucleophilic attack on the alpha phosphate group of the already incorporated nucleotide. This nucleophilic attack breaks the bond connecting the alpha phosphate group to the beta phosphate group leading to the formation of a phosphodiester bond and the release of a molecule of an unstable pyrophosphate. Sanger DNA polymerase is a modified Taq DNA polymerase. Part of the modification involves the addition of divalent ion and a pyrophosphatase which function as a cofactor and for the breaking down of the pyrophosphate respectively. The divalent ion of the polymerase normally removes the hydrogen molecule of the hydroxyl group making the oxygen more nucleophilic and therefore initiating a potent nucleophilic attack which otherwise would have taken a long time to take place. The unstable pyrophosphate released due to nucleophilic attack can through the process of pyrophosphorolysis knock out the already incorporated monophosphate to form back triphosphate thereby countering the action of the DNA polymerase. To prevent this from happening, part of the divalent ion stabilize the unstable pyrophosphate and then the pyrophosphatase breaks it down to an inorganic phosphate releasing an energy required by DNA polymerase activity. In this, there will be smooth elongation of the DNA strand until the DNA polymerase breaks down usually after addition of about 900-1000 bases [6, 7].

ACTION OF DNA FOR SANGERS SEQUENCING

DNA isolation is a process of purification of DNA from sample using a combination of physical and chemical methods. The first isolation of DNA was done in 1869 by Friedrich Miescher. DNA isolation for Sangers sequencing is a critical first step in the experimental flow of DNA sequencing and fragment analysis. The overall quality, accuracy and length of the DNA sequence read can be significantly affected by characteristics of the sample

itself, and the method chosen for the nucleic acid extraction. The isolation purification methods used in earlier times were lengthy and tiresome with the use of ultracentrifugation. But now with the advancement of separation techniques, the procedure is very simple and short. Ideal methods will vary depending on the source, and how the sample was handled or stored prior to extraction [5, 6].

MICROFLUIDIC SANGER'S SEQUENCING METHOD

Microfluidic Sanger sequencing is a lab-on-a-chip application for DNA sequencing, in which the Sanger sequencing steps (thermal cycling, sample purification, and capillary electrophoresis) are integrated on a wafer-scale chip using nanoliter-scale sample volumes. This technology generates long and accurate sequence reads, while obviating many of the significant shortcomings of the conventional Sanger method (e.g. high consumption of expensive reagents, reliance on expensive equipment, personnel-intensive manipulations, etc.) by integrating and automating the Sanger sequencing steps [8].

In its modern inception, high-throughput genome sequencing involves fragmenting the genome into small single-stranded pieces, followed by amplification of the fragments by Polymerase Chain Reaction (PCR). Adopting the Sanger method, each DNA fragment is irreversibly terminated with the incorporation of a fluorescently labeled dideoxy chain-terminating nucleotide, thereby producing a DNA "ladder" of fragments that each differ in length by one base and bear a base-specific fluorescent label at the terminal base. Amplified base ladders are then separated by Capillary Array Electrophoresis (CAE) with automated, in situ "finish-line" detection of the fluorescently labeled ssDNA fragments, which provides an ordered sequence of the fragments. These sequence reads are then computer assembled into overlapping or contiguous sequences (termed "contigs") which resemble the full genomic sequence once fully assembled [6, 7].

Sanger methods achieve read lengths of approximately 800bp (typically 500-600bp with non-enriched DNA). The longer read lengths in Sanger methods display significant advantages over other sequencing methods especially in terms of sequencing repetitive regions of the genome. A challenge of short-read sequence data is particularly an issue in sequencing new genomes (de novo) and in sequencing highly rearranged genome segments, typically those seen of cancer genomes or in regions of chromosomes that exhibit structural variation [1, 2].

Applications of microfluidic sequencing technologies

Other useful applications of DNA sequencing include single nucleotide polymorphism (SNP) detection, single-strand conformation polymorphism (SSCP) heteroduplex analysis, and short tandem repeat (STR) analysis. Resolving DNA fragments according to differences in size and/or conformation is the most critical step in studying these features of the genome. The sequencing chip has a four-layer construction, consisting of three 100-mm-diameter glass wafers (on which device elements are microfabricated)

and a polydimethylsiloxane (PDMS) membrane. Reaction chambers and capillary electrophoresis channels are etched between the top two glass wafers, which are thermally bonded. Three-dimensional channel interconnections and microvalves are formed by the PDMS and bottom manifold glass wafer. The device consists of three functional units, each corresponding to the Sanger sequencing steps. The Thermal Cycling (TC) unit is a 250-nanoliter reaction chamber with integrated resistive temperature detector, microvalves, and a surface heater. Movement of reagent between the top all-glass layer and the lower glass-PDMS layer occurs through 500- μ m-diameter via-holes. After thermal-cycling, the reaction mixture undergoes purification in the capture/purification chamber, and then is injected into the capillary electrophoresis (CE) chamber. The CE unit consists of a 30-cm capillary which is folded into a compact switchback pattern via 65- μ m-wide turns [4, 7, 8].

Performance values for genome sequencing technologies including Sanger methods and next-generation methods

Technol ogy	Num ber of lanes	Inject ion volu me (nL)	Analy sis time	Avera ge read lent h	Throug hput (includi ng analysi s; Mb/h)	Gel pouri ng	Lane tracki ng
454/Roche FLX		< 0.001	4 hours	200–30 0 bp	20–30	No	
ABI/SOLiD			8 days	35 bp	5–15	No	
Capillary array electrophor esis	96	1–5	1–3 hours	700 bp	0.166	No	No
Illumina/So lexa			2–3 days	30–100 bp	20	No	

Performance values for genome sequencing technologies including Sanger methods and next-generation methods

Microchip	96	0.1–0.5	6–30 minutes	430 bp	0.660	No	No
Slab gel	96	500–1000	6–8 hours	700 bp	0.0672	Yes	Yes

SANGER'S DNA SEQUENCING APPLICATION

The method has been extensively used to advance the field of functional and comparative genomics, evolutionary genetics and complex disease research. Because of this suitability for routine validation of cloning experiment and PCR fragments, Sanger sequencing remains a popular technique in many laboratories across the world and due to its flexibility to support diverse ranges of application in many research areas Sanger's DNA is applied in

De novo Sequencing

De novo sequence refers to sequencing or construction of a new genome. Small read is not sufficient to sequence De novo sequence therefore there is need for a long block to encapsulate repetitive sequencing. To ease the bioinformatics process of alignment in de novo sequencing, overlapping sequences are built up into a few large contigs as possible. This is aided by using Sanger's DNA sequencing. Sanger DNA is used because of its high degree of accuracy, increases the amount of overlapping sequencing and long read capacity, therefore increases confidence in sequence assembly [9].

Targeted DNA Sequencing

Sanger's DNA identifies heterozygous base positions or small insertions or deletions in genomic DNA. It is often employed to locate mutations or polymorphisms in diploid organisms, detect genetic rearrangements and uncover rare variants [7, 8].

Next-generation Sequencing validation

One technique or method is not sufficient to tell the whole story therefore researcher use combinations of approaches that will deliver the insight needed to move research therefore there is need for confirmation of results by other technique. Sanger's DNA sequencing helps researcher using new generation sequencing based data to confirm their results by an orthogonal method [9].

Microbial Sequencing

This is perform for diverse application such as microbial identification, environmental monitoring, pathogen detection and routine testing of material for bacterial contamination [9].

Mitochondrial Sequencing

Mitochondrial is a double membrane bounded organelle which is find in almost eukaryotes such as mammals. Examples of where mitochondrial can be is liver cell, muscles etc. mitochondrial diseases is an heterogeneous inherited diseases can arise at any stage in life.

Sanger's DNA sequencing a useful tool for studying human diseases such as

- diabetes, certain cancers and mechanism of ageing
- Population genetics and biodiversity assessments
- To detect mutations in some copies of mitochondria genome
- Human identification of forensics application

Other application of Sanger's DNA

Agricultural biology research: in cooperated edible vaccine into food product, biopesticides, disease identification etc

Human identification such as establishment of paternity, identification potential suspect

Pathogen detection and analysis

Food testing

Acknowledgements

This work was supported by students and provided by Department of microbiology and cell biology and Genetics. The authors are also grateful to University of Lagos, Nigeria.

Conflict of interest declaration

All authors declare: No conflict of interest in this work

REFERENCES

1. Sanger F; Coulson AR (May 1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase". *J. Mol. Biol.* **94** (3): 441–8. doi:10.1016/0022-2836(75)90213-2. PMID 1100841.
2. Sanger F; Nicklen S; Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors". *Proc. Natl. Acad. Sci. U.S.A.* **74** (12): 5463–7. Bibcode:1977PNAS...74.5463S. doi:10.1073/pnas.74.12.5463. PMC 431765 . PMID 271968.
3. Smith LM, Sanders JZ, Kaiser RJ, et al. (1986). "Fluorescence detection in automated DNA sequence analysis". *Nature*. **321** (6071):6749. Bibcode:1986Natur.321..674S. doi:10.1038/321674a0. PMID 3713851.
4. Smith LM; Fung S; Hunkapiller MW; Hunkapiller TJ; Hood LE (April 1985). "The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis". *Nucleic Acids Res.* **13** (7): 2399–412. doi:10.1093/nar/13.7.2399. PMC 341163 . PMID 4000959. . Retrieved 2011-02-24.
5. Ledergerber, C; Dessimoz, C (2011). "Base-calling for next-generation sequencing platforms". *Briefings in Bioinformatics*. **12** (5): 489–97. doi:10.1093/bib/bbq077. PMC 3178052 . PMID 21245079.
6. Murphy, K.; Berg, K.; Eshleman, J. (2005). "Sequencing of genomic DNA by combined amplification and cycle sequencing reaction". *Clinical Chemistry*. **51** (1): 35–39. doi:10.1373/clinchem.2004.039164. PMID 15514094.
7. Sengupta, D. ; Cookson, B. . (2010). "SeqSharp: A general approach for improving cycle-sequencing that facilitates a robust one-step combined amplification and sequencing method". *The Journal of molecular diagnostics*. **12** (3): 272–277. doi:10.2353/jmoldx.2010.090134. PMC 2860461 PMID 20203000.
8. Kan, Cheuk-Wai; Fredlake, Christopher P.; Doherty, Erin A. S.; Barron, Annelise E. (1 November 2004). "DNA sequencing and genotyping in miniaturized electrophoresis

- systems". Electrophoresis. **25** (21–22): 3564–3588. doi:10.1002/elps.200406161. PMID 15565709.
9. Morozova, Olena; Marra, Marco A (2008). "Applications of next-generation sequencing technologies in functional genomics". Genomics. **92**(5):255. doi:10.1016/j.ygeno.2008.07.001. PMID 18703132.
10. Microchip Biologies Inc. Apollo 100
11. Sinville, Rondedrick; Soper, Steven A (2007). "High resolution DNA separations using microchip electrophoresis". Journal of Separation Science. **30** (11): 1714–28. doi:10.1002/jssc.200700150. PMID 17623451.