# Gene sequencing by Capillary Electrophoresis

**M. Prathibha Bharathi*[1], D . Vijaya Durga[2] , A . Durga Pani Kumar, Rishitha Choudary[3], Ceema Mathew[4]**
Department of Pharmaceutical Analysis
Gokaraju Rangaraju College of Pharmacy
Osmania University, Hyderabad, Telangana - 500090, India

**Corresponding Author**
**1***Dr. M. Prathibha Bharathi
Associate Professor
Gokaraju Rangaraju College of Pharmacy
Bachupally, Hyderabad-500090

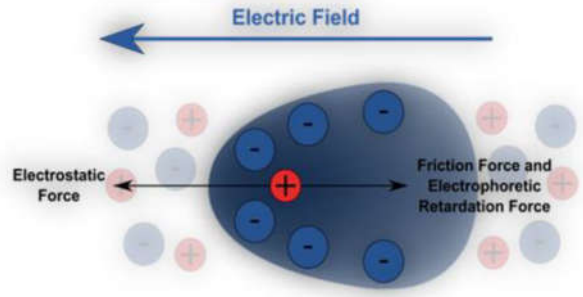
## ABSTRACT

Electrophoresis is an effective analytical tool for separation, as it does not affect the molecular structure, and it is also highly sensitive to small differences in molecular charge and size of the analytes. Capillary electrophoresis (CE) is a separation technique where the ions are separated based on their electrophoretic mobility with an applied voltage. Capillary electrophoresis took its place among established analytical techniques and, for instance, became the method of choice for fast high resolution Although with a considerably smaller footprint than liquid and gas chromatography, CE remains to play an essential role in contemporary analytics., with the strong advent of biopharmaceuticals, CE has shown to be particularly useful for routine quality control of therapeutic proteins, such as monoclonal antibodies. Current CE applications range from determination of small inorganic ions to characterization of high molecular weight biomolecules, and even particles and intact cells. It is one of the useful techniques because there is a large range of detection methods available. Hyphenation techniques based on capillary electrophoresis are (CE-MS, CE-ICP-MS, CE-ESI-MS, CE-MALDI-MS, and CE-NMR)

**Keywords:** Capillary electrophoresis, Gene sequencing CE-MS, CE-ICP-MS, CE-ESI-MS, CE-MALDI-MS, and CE-NMR

## II. INTRODUCTION

Electrophoresis is the movement of scattered particles comparative with a fluid under the influence of a uniform electric field Fig. 1.0. The electrophoresis of positively charged particles (cations) is sometimes called as cata-phoresis, whereas the electrophoresis of negatively charged particles (anions) is called as anaphoresis. The electrokinetic phenomenon of electrophoresis was discovered for the first time by Russian professors Peter Ivanovich Strakhov and Ferdinand Frederic Reuss at Moscow in 1807, who noticed that the application of a constant electric field causes the clay particles dispersed in water to migrate.




**History of Electrophoresis:** The history of electrophoresis had begun with the work of Arne Tiselius in the year of 1930s, and new separation processes and chemical analysis techniques based on electrophoresis continued into the 21st century. With help from the Rockefeller Foundation, the "Tiselius apparatus" for moving boundary electrophoresis was developed by Tiselius, which was described in 1937 in the well-known paper "then after a new apparatus for electrophoretic analysis named as "colloidal mixtures" is formed. This method spread slowly until the advent of effective zone electrophoresis methods in the 1940s and 1950s, where filter paper or gels are used as supporting media. By the 1960s, sophisticated gel electrophoresis methods increased widely. The method made it possible to separate biological molecules based on minute physical and chemical differences, helping to drive the rise of molecular biology. For a wide range of biochemical methods, such as protein fingerprinting, Southern blot and similar blotting procedures, DNA sequencing, and many more, gel electrophoresis and other related techniques have become an important basis. This kind of apparatus designed by Arne Tiselius enabled wide ranges of new applications of electrophoresis in analyzing chemical mixtures 1. Types of Electrophoresis: The main types of electrophoresis in practice are: 1. Zone Electrophoresis: includes; (a) Paper electrophoresis; (b) Cellulose acetate membrane electrophoresis; (c) Gel electrophoresis 2. Moving Boundary Electrophoresis: includes; (a) Capillary electrophoresis; (b) Isoelectric Focusing; (c) Isotachophoresis 1. Zone Electrophoresis: Zone electrophoresis includes methods that produce more or less differentiated zones completely of individual components that are being separated. It involves the migration of the charged particles on the supporting media like (paper, cellulose acetate membrane, starch gel, polyacrylamide) 2. Components that are separated are distributed into discrete zones on the support media. This supporting media is saturated with buffer solution in a small volume of samples and is being applied as a narrow band. Advantages**:** It is useful in investigations of biochemical. The small quantity of sample can be analyzed. The cost is low and easy maintenance. Disadvantages: It is unsuitable for accurate mobility and isoelectric point determination. Due to the presence of supporting medium, technical complications such as capillary flow, electro-osmosis can lead to adsorption and molecular sieving. Capillary electrophoresis (CE), even since its first introduction in 1981 ( Jorgenson JW, Lukacs KD (1981) , has been developed well into a mature and robust separation technique ( Harstad RK, Johnson AC, Weisenberger MM, Bowser MT. (2016). Compared to the traditional separation techniques like gel electrophoresis and liquid chromatography, CE carries the advantages of high simplicity in setup and miniaturization, fast separation with high resolution and efficiency, and low sample and solvent consumption. Thus, CE has been widely employed in biomedical research, forensic investigation, environmental monitoring, food quality organisms. Continuous advancement of CE has been focused on overcoming the key (Piñero M-Y, Bauza R,

Arce L. (2011). obstacles in its operation, such as low concentration sensitivity and strong wall adsorption of compounds with large molecular mass or unfavorable charges, (Timerbaev AR. (2013),which could be particularly problematic when employed to analyze complex samples with high amounts of interfering compounds in the matrix. Modification of the separation condition with new column coating or buffer additives, fabrication of new separation setup to facilitate multidimensional separation, and development of preconcentration methods and detection systems have been actively explored (Ramos-Payán Bello-Lopez MA. (2018), Voeten RLC, Ventouri IK, Haselberg R, Somsen GW. (2018). Additionally, CE can be coupled with bioassays, such as enzyme assays (Cheng M, Chen Z. (2018), and systematic evolution of ligands by exponential enrichment (SELEX) , to expand its application scope. Notably, exciting CE applications in analysis of proteins, pharmaceuticals, nucleic acids, metabolites, and carbohydrates, (Kristoff CJ, Bwanali L. ( 2020) which also provides an outlook for future directions towards improving the analysis of single cells and bio-particles.

Since the introduction of modern capillary electrophoresis (CE) by (Jorgenson and Lukacs in (1981), CE has evolved into a highly mature and versatile separation technique. After a first decade of development studies and instrument commercialization, CE took its place among established analytical techniques and, for instance, became the method of choice for fast high resolution DNA sequencing in the nineties of the last century. Although with a considerably smaller footprint than liquid and gas chromatography, CE remains to play an essential role in contemporary analytics. For example, with the strong advent of biopharmaceuticals, CE has shown to be particularly useful for routine quality control of therapeutic proteins, such as monoclonal antibodies. Current CE applications range from determination of small inorganic ions to characterization of high molecular- weight biomolecules, and even particles and intact cells. The research field of CE remains very active, as exhibited by a steady and significant flow of scientific reports on theory, separation modes, new instrumentation, and applications of An initial search on Web of Science, considering keywords related to all modes of CE including their acronyms, yielded about 7000 articles.

DNA stands for deoxyribonucleic acid. It is a type of nucleic acid that, together with RNA (ribonucleic acid), is both a repository and a functional expression of biological information. DNA is said to be a polymer of deoxyribonucleotides Wells SS, Dawod M, Kennedy RT. (2019). . Nucleotides are the building blocks of nucleic acids and are the monomers of DNA and RNA. Each nucleotide is a 5`-monophosphate ester of a nucleoside, and thus, in the case of deoxyribonucleotide found in DNA, it's a nucleotide containing 2-deoxy-D-ribose (monophosphate). Examples are deoxyadenosine 5`-monophospate (dAMP), deoxyguanosine 5`-monophospate (dGMP), deoxycytidine 5`-monophospate (dCMP), and deoxythymidine 5`-monophospate (dTMP) (David, L. N., Cox, M. M., & Hoskins, A. A. (2021). A DNA sequence refers to the specific order of the building blocks or bases in a strand of DNA. The common nitrogenous bases in DNA are adenine, guanine, cytosine, and thymine, each with a unique letter code given as A, G, C, and T, respectively.

**Sanger`s sequencing: the chain termination method**

The most important property of a DNA molecule is its nucleotide sequence.

Until the late 1970s, determining the sequence of a nucleic acid containing as few as 5 or 10 nucleotides was very laborious; that is to say, it was characterized by effort to the point of exhaustion, especially physical effort. However, sequencing technology has been improving steadily for the past two decades ( McMurry, J., Ballantine, D. (2020). Sanger`s sequencing, also known as the dideoxy chain termination method**,** was designed to determine the sequence of

nucleotide bases in a piece of DNA molecule. Even though the strategies of the two methods are similar, Sanger`s sequencing remains technically easier and more accurate, with a 99.99% base accuracy; thus, it`s considered the "gold standard" for validating DNA sequences. This development played an important role in molecular biology and revolutionized the Human Genome Project, where it was used to determine the sequencing of relatively small fragments of human DNA (Men, A. E., Wilson, P., Siemering, K., & Forrest, S. (2008).

**Principle: Sanger`s sequencing method**

The principle behind Sanger`s sequencing method for a DNA molecule lies in the random incorporation of modified nucleosides, in this case, dideoxynucleoside triphosphates (ddNTPs) by DNA polymerase to interrupt in the vitro DNA replication process at one of the four bases. Four different replication reactions produce fragments that terminate at A, G, C, and T, respectively.
The ddNTPs are analogs of dNTPs, only that they lack the 3` OH group on the sugar moiety. The absence of this OH group at the 3` carbon of the sugar prevents the formation of the phosphodiester bond (O-P-O) with the incoming nucleotide, effectively terminating the DNA synthesis process when added to a growing DNA strand since it hinders enzyme polymerase from adding additional newly synthesized strand terminates with the base N.

**Constituents for Sanger`s sequencing**

Classical Sanger`s sequencing method is similar to that of DNA replication in an organism, only that this method involves making many copies of the target DNA region. These constituents include . Corresponding ddNTPs of dNTPs terminate the chain at their respective sites. For instance, ddATP terminates at the A site. In like manner, ddCTP, ddGTP, and ddTTP terminate at C, G, and T sites, respectively. Hence, ddNTPs are used as chain terminators in Sanger`s sequencing method. In a regular nucleotide, like the dNTPs, the hydroxyl group (OH-) in the 3` carbon atom acts as a hook, allowing a new nucleotide to be added to an already existing chain (Roy, A. (1977).

**Constituents:**

1) A DNA polymerase enzyme: This enzyme is sometimes called the `builder enzyme`. It catalyzes the synthesis of the new DNA strand by adding nucleotides to the 3` end of the primer following the template strand.
2) A primer: This is used to initiate the DNA synthesis process at specific regions of the template strand of the DNA to be sequenced. The enzyme primase, also called the `initializer enzyme`, creates the primer.
3) The template of the DNA to be sequenced
4) Four DNA nucleotides (dATP, dGTP, dCTP, and dTTP) are the building blocks for DNA synthesis. Their presence allows for DNA strand extension during sequencing.
5) Dideoxy, or chain terminator version of all four nucleotides (ddATP, ddGTP, ddCTP, and ddTTP), each labeled with a different color of fluorescent dye for identification.

**Sequencing procedure**:

Sanger`s sequencing method takes place in four separate glass test tubes in the following procedures

**1. DNA Template Preparation:** The double-stranded DNA (dsDNA) is denatured into two single-stranded DNAs (ssDNA), and samples are put into each test tube. Denaturing of the dsDNA into ssDNA is achieved by heating the DNA at 96$^{o}$C to disrupt and break the hydrogen bonds between the complementary base pairs, causing the double helix to separate and unwind. At all times, in the DNA double helix, A forms double hydrogen bonds with T, and C forms hydrogen triple bonds

with G. After denaturation, the sample is rapidly cooled to prevent the single strands from binding together again. Denaturation of the DNA allows a sequencing primer to attach to the template DNA and the DNA polymerase to synthesize a new complementary strand.

**2. Primer Binding**: After the DNA has been denatured to be single-stranded, a sequencing primer that complements and corresponds to one end of the template ssDNA is added.

DNA polymerase enzyme solutions are added to each of the four test tubes, followed by four types of dNTPs and one type of ddNTP.

**3. Synthesis reaction:** The DNA synthesis reaction initiates, and the chain extends until a termination nucleotide is randomly incorporated, either ddATP, ddGTP, ddCTP, or ddTTP.

The resulting DNA fragments are denatured into ssDNA.

The denatured fragments are separated by gel electrophoresis, and their sequence is determined.

**Sanger`s sequencing steps**

Sanger`s sequencing method involves three basic steps: template preparation, generation of a nested set of labeled fragments, and electrophoresis and gel analysis reading. These steps are discussed in detail below

**Step 1: Template Preparation**

- Copies of the DNA strand to be sequenced must be prepared with a short sequence at the 3`end of the strand. The short sequence will be hybridized with the DNA primer, whose sequence is exactly complementary to that of the known sequence.
- The primer is essential to initiate the replication of the templates by the DNA polymerase enzyme.
- Thus, in Sanger`s sequencing method, the most convenient way for adding a known sequence to the 3`end of the DNA template strand is to clone the strand in a single-stranded cloning vector, M13, so that a known M13 sequence will always flank the unknown DNA insert and serve as a binding site for the standard primer. This helps to prepare the templates.
- The figure below shows a cloned template DNA strand of the sequence:
  3` – CTAAGCTCGACT – 5`

**Step 2: Generation of a Nested Set of Labeled Fragments:**

- Copies of each template are divided into four batches, and each batch is used for a different replication reaction in four test tubes. Copies of the standard primer and DNA polymerase I are used for all four batches.
- To synthesize fragments, all of which terminate at A, the dideoxy analog ddATP is added to the reaction mixture along with dATP, dGTP, dCTP, dTTP, the standard primer, and the DNA polymerase I. Similarly, to generate all fragments that terminate at G, C, and T, the respective ddNTPs, i.e., ddGTP, ddCTP, and ddTTP, are added, respectively, to the reaction mixture in different batches along with the usual dNTPs, which are dGTP, dCTP, and dTTP.

The ddNTPs and one of the dNTPs are labeled with a radioactive isotope in each batch, say radioisotope 32P, to produce radiolabeled strands.

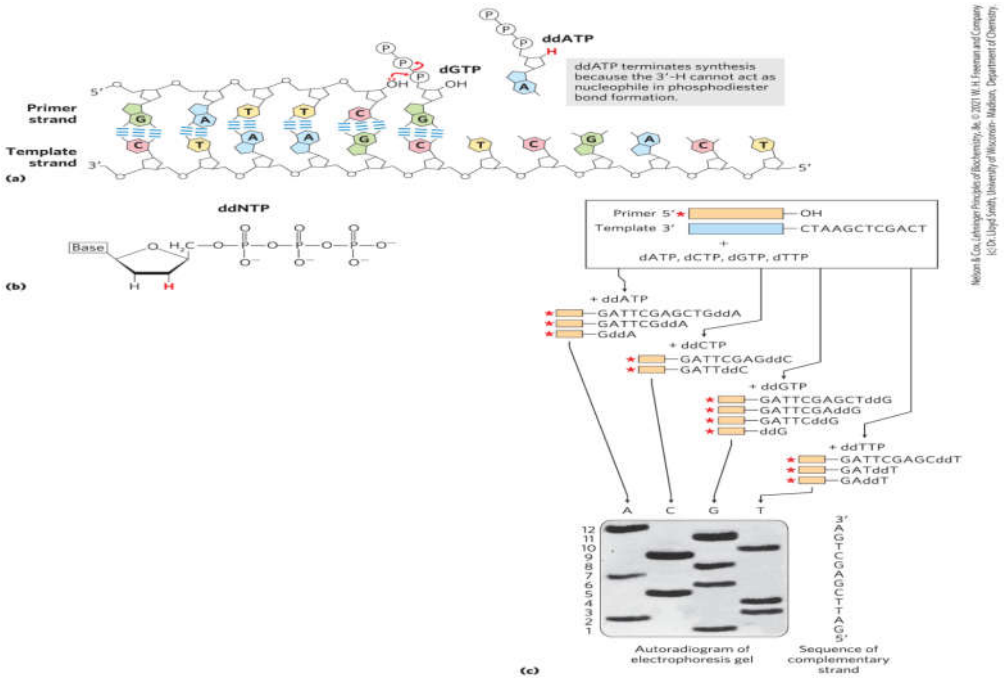
These synthesized fragments from the four reaction mixtures compose the set of nested fragments needed to determine the order of the bases in the strand complementary to the template strand.

**Step 3: Electrophoresis and Gel Analysis Reading**

- The fragments from the four reaction mixtures are loaded into four parallel lanes of a polyacrylamide gel and separated by length using gel electrophoresis. Thus, each lane

contains labeled fragments that end in only one of the four bases. During electrophoresis, shorter fragments travel farther than longer fragments.

- Therefore, copies of the shortest fragment form a band farthest from the end at which the fragment batches were loaded into the gel.
- Successively longer fragments form bands at positions closer and closer to the loading end. Following electrophoresis, the radio-labeled fragments are visualized by exposing the gel to an X-ray filter to make an autoradiogram. When reading the autoradiogram, one starts at the bottom and looks across the four lanes, each time identifying the lane containing the band corresponding to the next longer fragment, and then noting the end base of those fragments since each fragment ends at the base marked at the top of the lane. Figure 1DNA Sequencing by Sanger`s Method The sequence of the complementary strand is hence read from its 5' end, the common starting point, to its 3' end.
- Knowing this, one can determine the unknown sequence of the DNA template.
- From the figure 1 above, the nested fragments obtained are as follows:
- The sequence of the complementary strand is hence read from its 5' end, the common starting point.




**Applications of Sanger`s dideoxy chain termination method**

Apart from primarily being used in determining the nucleotide sequence of DNA molecules ranging from small DNA fragments to an entire genome, Sanger`s dideoxy chain termination method has been pivotal in numerous other scientific disciplines since its development.

Its applications include:

1. Forensic Science: Sanger`s sequencing is used in forensic science for identifying individuals based on their genetic profiles and DNA profiling. Also, paternity tests and criminal investigations make use of this method.

2. Viral Genomics: Sanger`s sequencing is employed in studying viral genomes and, as a result, aiding in virus classification. It is also used for tracking viral evolution and detecting emerging pathogens.
3. Validation of NGS Data: Due to its high accuracy, Sanger`s method is used in validating and confirming data from the new sequencing technologies, i.e., New-Generation Sequencing (NGS).
4. Drug Discovery and Development: Sanger`s sequencing helps in understanding drug resistance mechanisms and optimizing drug efficacy. It is thus employed in pharmaceutical research aimed at understanding drug responses.
5. Agricultural Biotechnology: For marker-assisted breeding, genetic engineering, and studying plant genomes, Sanger`s sequencing method is employed, hence aiding in crop improvement.
6. Cancer Genomics: This method plays a crucial role in cancer research and diagnosis by identifying somatic mutations in cancer-related genes to accurately predict patients` cancer-targeted therapies.

**Advantages of sanger`s dideoxy chain termination method**

1. The method is more accurate, especially for short- to medium-sized DNA fragments.
2. For small-scale projects or the sequencing of individual genes, Sanger`s sequencing can be cost-effective.
3. Since Sanger`s sequencing has been around for several decades, it is a well-established and understood method for standardized protocols. It's thus reliable.
4. Sanger`s sequencing can generate longer read lengths compared to some new sequencing technologies like NGS. It can still generate reads of up to 1000 bases long.

**Disadvantages of sanger`s dideoxy chain termination method**

1. This method becomes inefficient and expensive for large-scale sequencing due to its limited throughput.
2. Sanger`s sequencing is less suitable for sequencing very long DNA fragments due to the limitations in read length compared with some newer sequencing technologies, such as Next-Generation Sequencing (NGS).

**Common software used in the reading of DNA chromatograms**

Proficiency in interpreting DNA chromatograms stands as an indispensable competence within the realm of molecular biology and genetics research. The chromatograms from both the forward and reverse strands offer insightful revelations about the nucleotide sequence embedded within the scrutinized DNA fragment. This practice of deciphering chromatograms from both strands holds heightened significance, particularly in Sanger dideoxy sequencing, as it serves as a linchpin for the precision and dependability of sequencing data. Instances of misinterpretations, errors in base-calling, and artifacts might arise during sequencing endeavors, and their identification can prove especially formidable if solely reliant on the scrutiny of either the forward or reverse chromatogram. Numerous software tools are at one’s disposal to facilitate the scrutiny of Sanger sequencing chromatograms. For instance, the software Sequencher by Gene Codes Corporation enables concurrent visualization and analysis of both forward and reverse chromatograms. Correspondingly, Chromas, developed by Technelysium Pty Ltd., empowers users to peruse and manipulate chromatograms, discerning disparities and inaccuracies. A plethora of specialized programs designed to facilitate DNA chromatogram interpretation are accessible, including

4Peaks (exclusive to Mac), Snap Gene Viewer, FinchTV, Sequence Scanner (compatible with both PC and MAC), Chromas (exclusive to PC).

## CONCLUSION

In conclusion, using modified dideoxynucleosides, Sanger was able to precisely determine DNA sequences. Dideoxy chain termination method inspired DNA sequencing, providing an accurate and more reliable technique for determining the genetic code. Being an accurate, cost-effective, and well-understood method, Sanger`s sequencing technique has found application in various fields of science over the past decades. Regardless of the software employed for interpretation, the comprehension of DNA chromatograms demands meticulous attention to detail and an in-depth familiarity with the sequencing process.

**Acknowledgment**

The authors are thankful to the management and to the Principal, Prof. Dr. M. Ganga Raju, Gokaraju Rangaraju College of Pharmacy for providing facilities in doing this this research work.

## REFERENCES

[1] Jorgenson JW, Lukacs KD (1981). Zone electrophoresis in open-tubular glass capillaries. Anal Chem, 53(8):1298–302.
[2] Harstad RK, Johnson AC, Weisenberger MM, Bowser MT. (2016). Capillary electrophoresis. Anal Chem.;88(1):299–319.
[3] Pinero M-Y, Bauza R, Arce L. (2011). Thirty years of capillary electrophoresis in food analysis laboratories: potential applications.
Electrophoresis;32(11):1379–93.
[4] Timerbaev AR. (2013). Element speciation analysis using capillary electrophoresis: twenty years of development and applications. Chem Rev, 113(1):778–812.
[5] Ramos-Payan M, Ocana-Gonzalez JA, Fernandez-Torres RM, Lobera A, Bello-Lopez MA. (2018). Recent trends in capillary electrophoresis for complex samples analysis: a review. Electrophoresis, 39(1):111–25.
[6] Voeten RLC, Ventouri IK, Haselberg R, Somsen GW. (2018) .Capillary electrophoresis: trends and recent advances. Anal Chem, 90(3):1464–81.
[7] Cheng M, Chen Z. (2018). Recent advances in screening of enzymes inhibitors based on capillary electrophoresis. J Pharm Anal, 8(4):226–33.
[8] Kristoff CJ, Bwanali L, Veltri LM, Gautam GP, Rutto PK, Newton EO. ( 2020) .Challenging bioanalyses with capillary electrophoresis. Anal Chem. 92(1):49–66.
[9] Wells SS, Dawod M, Kennedy RT. (2019). CE-MS with electrokinetic supercharging and application to determination of neurotransmitters.Electrophoresis, 40(22):2946–53.
[10] David, L. N., Cox, M. M., & Hoskins, A. A. (2021). Lehninger Principles of Biochemistry (N. D. L., M. M. Cox, & A. A. Hoskins (eds.); 8th Edition). W.H. Freeman, Macmillan Learning.
[11] McMurry, J., Ballantine, D. S., Hoeger, C. A., Peterson, V. E., & Madsen, S. (2020). Fundamentals of General, Organic, and Biological Chemistry (J. Zalesky, C. Hess, & C. Morrison (eds.); 8th Edition). Pearson.

[12] Men, A. E., Wilson, P., Siemering, K., & Forrest, S. (2008). Sanger DNA Sequencing. Next Generation Genome Sequencing: Towards Personalized Medicine, 1–11. https://doi.org/10.1002/9783527625130.ch1
[13] Roy, A. (1977). What is sequencing ? Sanger sequencing : The chain termination method. DNA Sequencing. 20.