**DNA Sequencing Methods**

**1. Introduction**

The term DNA sequencing refers to methods for determining the order of the nucleotides bases adenine, guanine, cytosine and thymine in a molecule of DNA. The first DNA sequences were obtained by academic researchers, using laboratories methods based on 2-dimensional chromatography in the early 1970s. By the development of dye based sequencing method with automated analysis, DNA sequencing has become easier and faster. The knowledge of DNA sequences of genes and other parts of the genome of organisms has become indispensable for basic research studying biological processes, as well as in applied fields such as diagnostic or forensic research.DNA is the information store that ultimately dictates the structure of every gene product, delineates every part of the organisms. The order of the bases along DNA contains the complete set of instructions that make up the genetic inheritance. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of the human genome, in the human genome project.

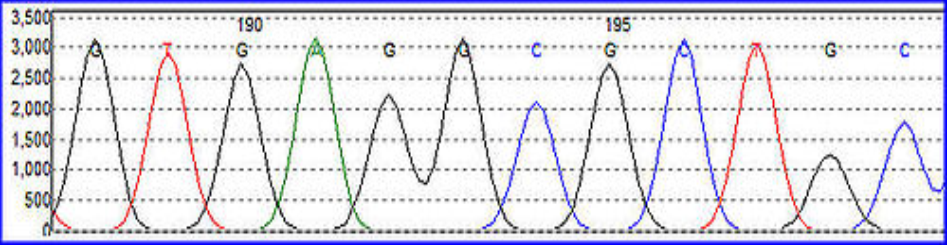


Fig. 1. DNA Sequence Trace

DNA can be sequenced by a chemical procedure that breaks a terminally labelled DNA molecule partially at each repetition of a base. The length of the labelled fragments then identifies the position of that base. We describe reactions that cleave DNA preferentially at guanines, at adenines, at cytosine and thymines equally, and at cytosine alone. When the product of these four reactions are resolved by size, by electrophoresis on a polyacrylamide gel, the DNA sequences can be read from the pattern of radioactive bands. The technique will permit sequencing of atleast 100 bases from the point of labelling. The purine specific reagent is dimethyl sulphate; and the pyrimidine specific reagent is hydrazine.

In 1973 , Gilbert and Maxam reported the sequence of 24 base pairs using a method known as wandering- spot analysis.

The chain termination method developed by Sanger and co-workers in 1975 owing to its relative easy and reliability.

In 1975 the first complete DNA genome to be sequenced is that of bacteriophage ΦX174.

By knowing the DNA sequence, the cause of the various diseases can be known. We can determine the sequence responsible for various diseases and can be treated with the help of Gene therapy.

DNA sequencing is very significant in research and forensic science. The main objective of DNA sequence generation method is to evaluate the sequencing with very high accuracy and reliability.

There are some common automated DNA sequencing problems:-

1. Failure of the DNA sequence reaction.

2. Mixed signal in the trace (multiple peaks).

3. Short read lengths and poor quality data.

4. Excessive free dye peaks “dye blobs” in the trace.

5. Primer dimer formation in sequence reaction

6. DNA polymerase slippage on the template mononucleotide regions.

So, we should have to do the sequencing in such a manner to avoid or minimize these problems.

DNA sequencing can solve a lot of problems and perform a lot of work for human well fare.

DNA sequencing can be done by different methods:

1. Maxam – Gilbert sequencing

2. Chain-termination methods

3. Dye-terminator sequencing

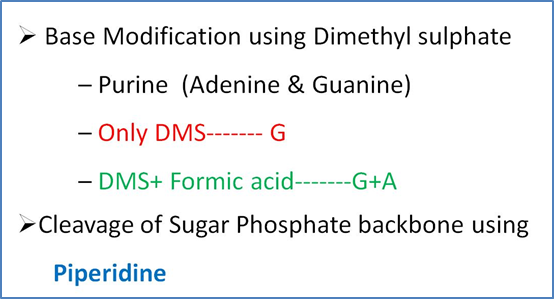
4. Automation and sample preparation

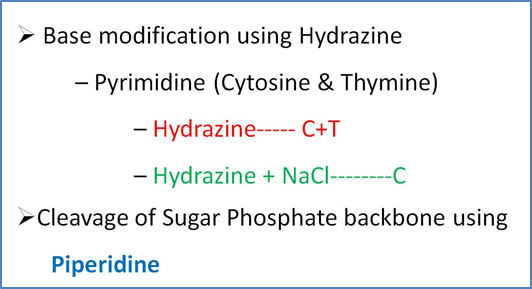
5. Large scale sequencing strategies

6. New sequencing methods.

**2. Maxam-Gilbert sequencing**

In 1976-1977, Allan Maxam and Walter Gilbert developed a DNA sequencing method based on chemical modification of DNA and subsequent cleavage at specific bases. The method requires radioactive labelling at one end and purification of the DNA fragment to be sequenced. Chemical treatment generates breaks at small proportions of one or two of the four nucleotide based in each of four reactions (G, A+G, C, C+T). Thus a series of labelled fragments is generated, from the radiolabelled end to the first ‘cut’ site in each molecule. The fragments in the four reactions are arranged side by side in gel electrophoresis for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabelled DNA fragment, from which the sequence may be inferred.

****

****

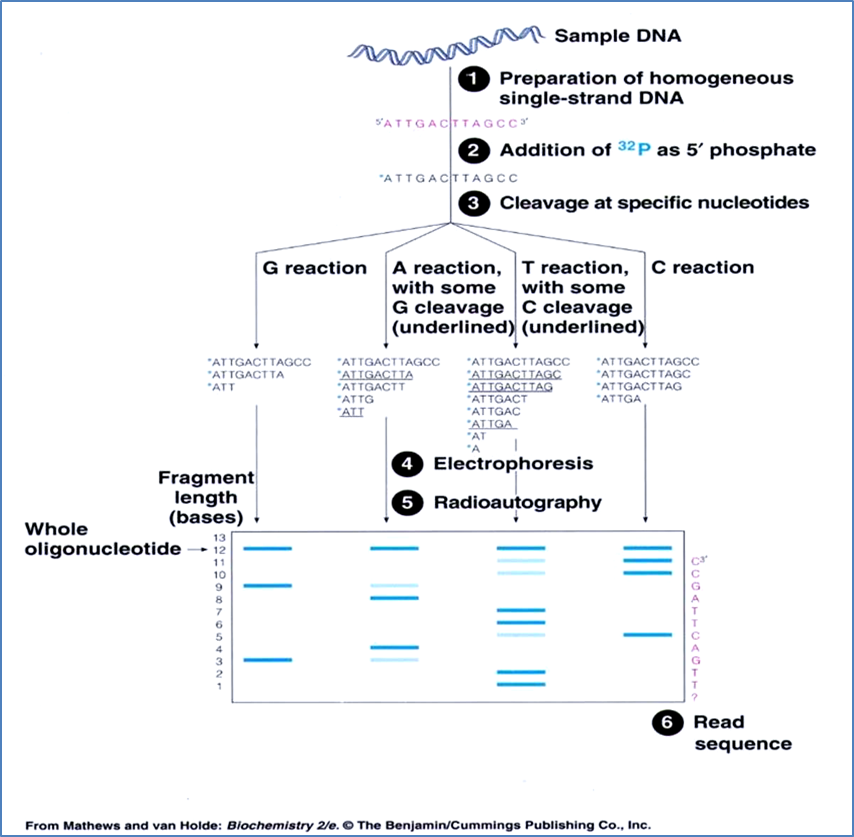
****

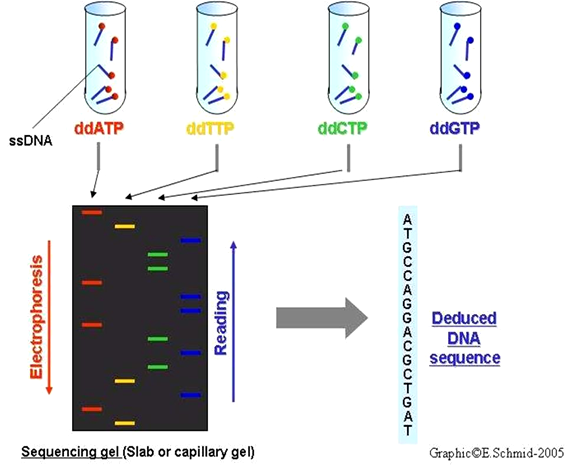
Fig.2. Maxam-Gilbert Method of DNA sequencing (Chemical modification method)

**3. Chain-termination method**

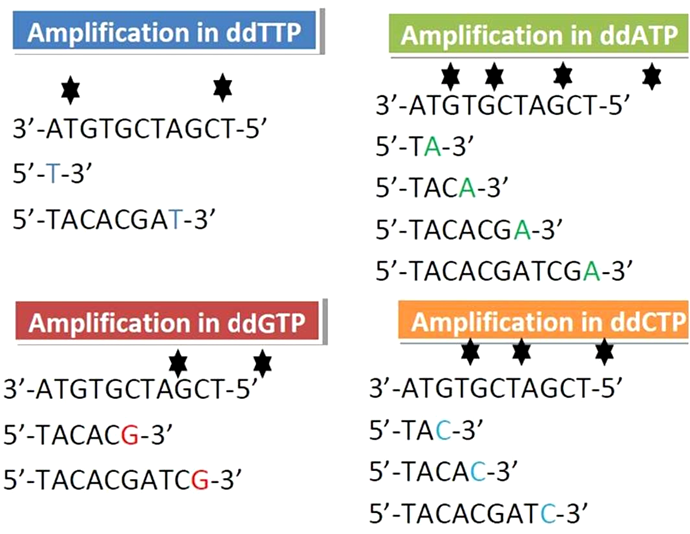
The chain terminator method is more efficient and uses fewer toxic chemicals and lower amount of radioactivity than the method of Maxam and Gilbert. The key principle of the Sanger method was the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators. The chain termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, radioactively or fluorescently labelled nucleotides, and modified nucleotides that terminate DNA strand elongation. The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP, dTTP) and the DNA polymerase. To each reaction is added only one of the four dideoxynucleotide (ddATP, ddGTP, ddCTP, ddTTP) which are the chain terminating nucleotides, lacking a 3’-OH group required for the formation of a phosphodiester bond between two nucleotides, thus terminating DNA strand extension and resulting in DNA fragments of varying length.

The newly synthesized and labelled DNA fragments are heat denatured, and separated by size by gel electrophoresis on a denaturing polyacrylamide-urea gel with each of the four reactions run in one of the four individual lanes (lanes A, T, G, C), 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. A dark band in a lane indicates a DNA fragment that is result of chain termination after incorporation of a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP). The relative position of the different bands among the four lanes is then used to read (from bottom to top) the DNA sequence.

The technical variations of chain termination sequencing include tagging with nucleotides containing radioactive phosphorus for labelling, or using a primer labelled at the 5’ end with a fluorescent dye. Dye- primer sequencing facilitates reading in an optical system for faster and more economical analysis and automation.



**Sanger method of DNA sequencing**



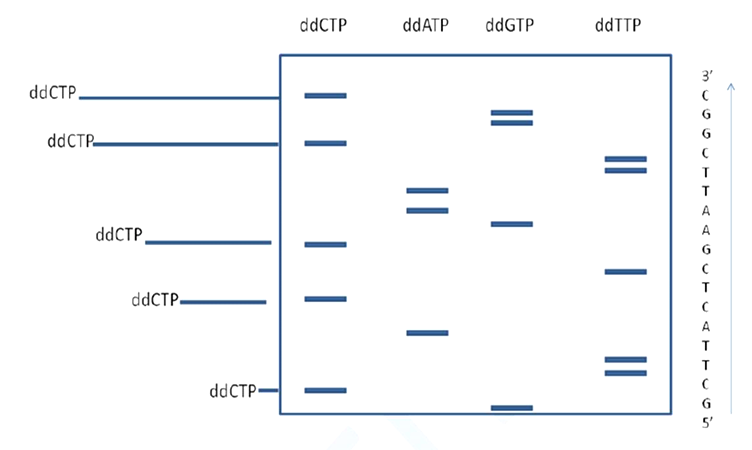


Fig.3 Schematic diagram of Sanger DNA sequencing method with an autoradiograph. The sequence is in 5’ to 3’ direction and is the complementary of query

Chain termination methods have greatly simplified DNA sequencing. Limitations include non-specific binding of the primer to the DNA, affecting accurate read-out of the DNA sequence and DNA secondary structures affecting the fidelity of the sequence.

**3.1 Dye-terminator sequencing** (Fluorescence DNA Sequencing)

Dye-terminator sequencing utilizes labelling of the chain terminator ddNTPs, which permits sequencing in a single reaction, rather than four reactions as in the labelled- primer method. In dye- terminator sequencing, each of the four dideoxynucleotide chain terminators is labelled with **fluorescent dyes**, each of which with different wavelengths of fluorescence and emission. Owing to its greater expediency and speed, dye terminator sequencing is now the mainstay in automated sequencing. Its limitation include dye effects due to differences in the incorporation of the dye-labelled chain terminators into the DNA fragment, resulting in unequal peak heights and shapes in the electronic DNA sequence trace chromatogram. The common challenges of DNA sequencing include poor quality in the first 15-40 bases of the sequence and deteriorating quality of sequencing traces after 700-900 bases.

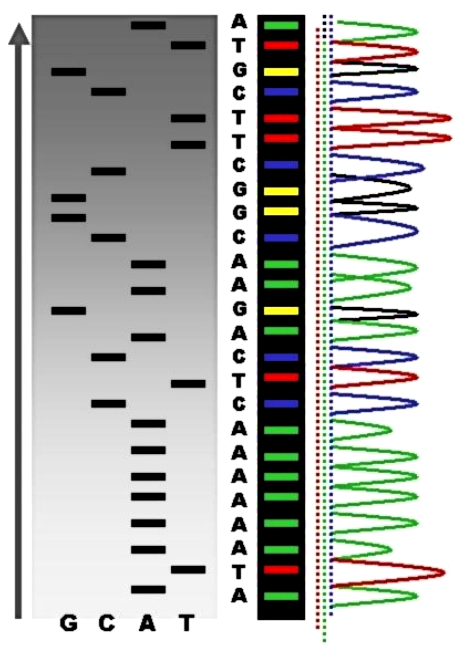
****

Fig.3. Sequence ladder by radioactive sequencing compared to fluorescent peaks

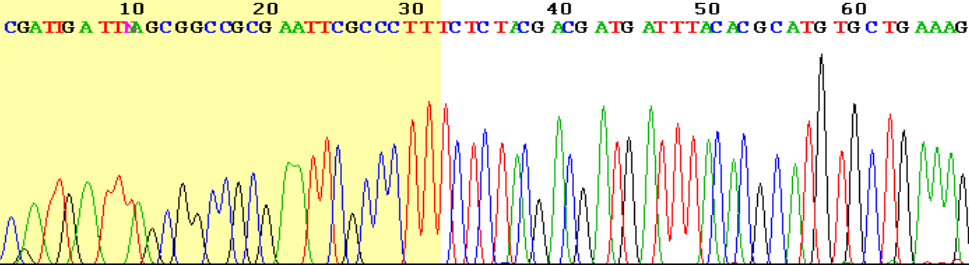
****

Fig. 4. View of the start of an example dye-terminator read

**3.2 Automation and sample preparation**

Automated DNA sequencing instruments (DNA sequencers) can sequence up to 384 DNA samples in a single batch (run) in up to 24 runs a day. DNA sequencers carry out capillary electrophoresis for size seperation, detection and recording of dye fluorescence and data output as fluorescent peak trace chromatograms. A number of commercial and non-commercial software packages can trim low-quality DNA traces automatically. These programmes score the quality of each peak and remove low quality base peaks (generally located at the ends of the sequence).

**4. Large-scale sequencing strategies**

Current methods can directly sequence only relative short (300-1000 nucleotides long) DNA fragments in a single reaction. The main obstacle to sequencing DNA fragments above this size limit is insufficient power of separation for resolving large DNA fragments that differ in length by only one nucleotide. Large scale sequencing aims at sequencing very long DNA pieces, such as whole chromosomes. It consists of cutting (with restriction enzymes) or shearing (with mechanical forces) large DNA fragments into shorter DNA fragments. The fragmented DNA is cloned into a DNA vector, and amplified in *E.coli*. Short DNA fragments purified from individual bacterial colonies are individually sequenced and assembled electronically into one long, contiguous sequence. This method does not require any pre- existing information about the sequence of the DNA and is referred to as de novo sequencing. Gaps in the assembled sequence may be filled by primer walking. The different strategies have different tradeoffs in speed and accuracy.

**5. New sequencing methods**

The high demand for low-cost sequencing has driven the development of high- throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing. Molecular detection method are not sensitive enough for single molecule sequencing, so most approaches use an in vitro cloning step to amplify individual DNA molecules. In microfluidic Sanger sequencing the entire thermocycling amplification of DNA fragments as well as their separation by electrophoresis is done on a single chip (appoximately 100cm in diameter) thus reducing the reagent usage as well as cost. In some instances researchers have shown that they can increase the throughput of conventional sequencing through the use of microchips.

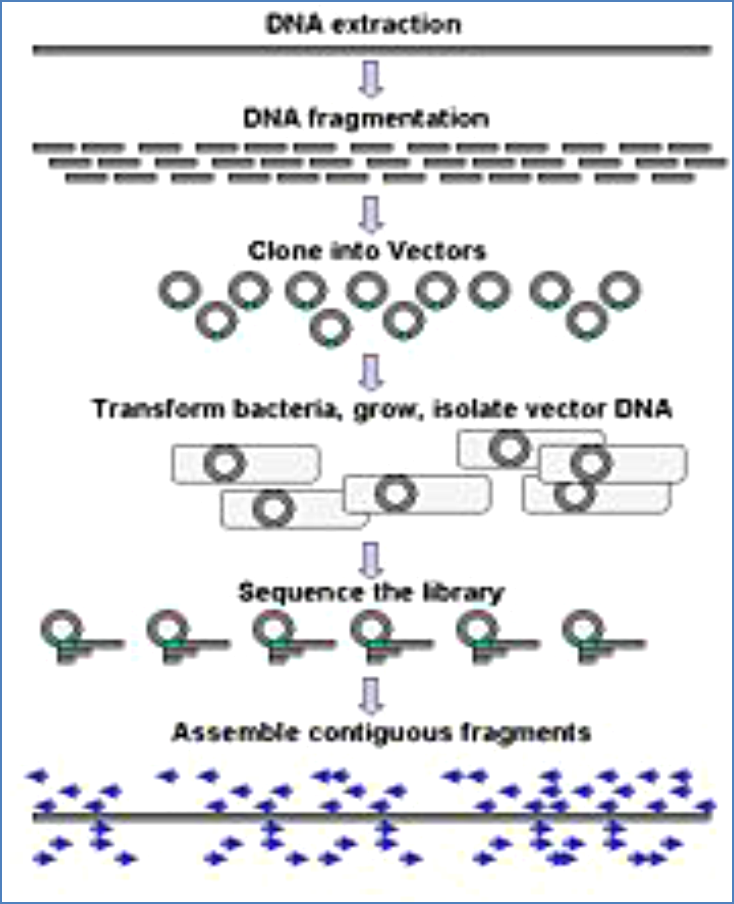
****

Fig. 5. Genomic DNA is fragmented into random pieces and cloned as a bacterial library. DNA from individual bacterial clones is sequenced and the sequence is assembled by using overlapping regions.

**6. High throughput sequencing**

The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods.

**6.1 Pyrosequencing**

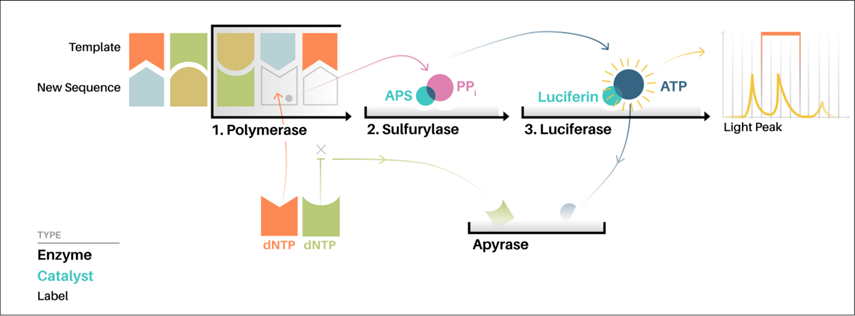
A parallelized version of pyrosequencing was developed by 454 Life Sciences, which has since been acquired by Roche Diagnostics. The method amplifies DNA inside water droplets in an oil solution (emulsion PCR), with each droplet containing a single DNA template attached to a single primer-coated bead that then forms a clonal colony. The sequencing machine contains many picolitre-volume wells each containing a single bead and sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA, and the combined data are used to generate sequence read-outs. This technology provides intermediate read length and price per base compared to Sanger sequencing on one end and Solexa and SOLiD on the other.

*Procedure:*

"Sequencing by synthesis" involves taking a single strand of the DNA to be sequenced and then synthesizing its complementary strand enzymatically. The pyrosequencing method is based on detecting the activity of [DNA polymerase](https://en.wikipedia.org/wiki/DNA_polymerase) (a DNA synthesizing enzyme) with another [chemoluminescent](https://en.wikipedia.org/wiki/Chemiluminescence" \o "Chemiluminescence) [enzyme](https://en.wikipedia.org/wiki/Enzyme). Essentially, the method allows sequencing a single strand of [DNA](https://en.wikipedia.org/wiki/DNA) by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step. The template DNA is immobile, and solutions of A, C, G, and T [nucleotides](https://en.wikipedia.org/wiki/Nucleotides) are sequentially added and removed from the reaction. Light is produced only when the nucleotide solution complements the first unpaired base of the template. The sequence of solutions which produce chemiluminescent signals allows the determination of the sequence of the template.

For the solution-based version of Pyrosequencing, the single-strand DNA ([ssDNA](https://en.wikipedia.org/wiki/SsDNA" \l "ssDNA" \o "SsDNA)) template is hybridized to a sequencing [primer](https://en.wikipedia.org/wiki/Primer_(molecular_biology)) and incubated with the enzymes [DNA polymerase](https://en.wikipedia.org/wiki/DNA_polymerase), [ATP sulfurylase](https://en.wikipedia.org/wiki/ATP_sulfurylase), [luciferase](https://en.wikipedia.org/wiki/Luciferase) and [apyrase](https://en.wikipedia.org/wiki/Apyrase" \o "Apyrase), and with the substrates [adenosine 5´ phosphosulfate](https://en.wikipedia.org/wiki/Adenosine_5%C2%B4_phosphosulfate) (APS) and [luciferin](https://en.wikipedia.org/wiki/Luciferin" \o "Luciferin).

1. The addition of one of the four [deoxynucleotide triphosphates](https://en.wikipedia.org/wiki/Deoxynucleotide_triphosphate" \o "Deoxynucleotide triphosphate) ([dNTPs](https://en.wikipedia.org/wiki/DNTP" \o "DNTP)) (dATPαS, which is not a substrate for a luciferase, is added instead of dATP to avoid noise) initiates the second step. DNA polymerase incorporates the correct, complementary dNTPs onto the template. This incorporation releases [pyrophosphate](https://en.wikipedia.org/wiki/Pyrophosphate) (PPi).
2. ATP sulfurylase converts PPi to [ATP](https://en.wikipedia.org/wiki/Adenosine_triphosphate) in the presence of adenosine 5´ phosphosulfate. This ATP acts as a substrate for the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a camera and analyzed in a program.
3. Unincorporated nucleotides and ATP are degraded by the [apyrase](https://en.wikipedia.org/wiki/Apyrase" \o "Apyrase), and the reaction can restart with another nucleotide.



Pyrosequencing Method of DNA sequencing

**6.2 Illumina (Solexa) sequencing**

Solexa developed a sequencing technology based on dye terminators. In this, DNA molecule are first attached to primers on a slide and amplified, this is known as bridge amplification. Unlike pyrosequencing, the DNA can only be extended one neucleotode at a time. A camera takes images of the fluorescently labeled nucleotides, then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing the next cycle.

**6.3 SOLiD sequencing** *(Applied Biosystem)*

The technology for sequencing used in ABI Solid sequencing is oligonucleotide ligation and detection. In this, a pool of all possible oligonucleotides of fixed length is labelled according to the sequenced position. This sequencing results to the sequences of quantities and lengths comparable to illumine sequencing.

**Some important applications of DNA sequencing are:**

1. To analyse any protein structure and function we must have the knowledge of its primary structure i.e its DNA sequence.

2. With its study we can understand the function of a specific sequence and the sequence responsible for any disease.

3. With the help of comparative DNA sequence study we can detect any mutation.

4. Kinship study (Analysis of blood relationship of child with parents)

5. DNA fingerprinting.

6. By knowing the whole genome sequence, Human genome project get completed.

The main problem with sequencing is its intactness. If we perform the sequencing of same sample with different methods the result may be different so we should have to do it in such a manner that atleast 40-50% sequence must be same of similar sample.