TEJASVI NAVADHITAMASTU

"Let our (the teacher and the taught) learning be radiant"

Let our efforts at learning be luminous and filled with joy, and endowed with the force of purpose

Dr. Prabhakar Singh
Department of Biochemistry
Faculty of Science, V.B.S. Purvanchal University, Jaunpur
PIN-222 003, (U.P.), INDIA.
Mobile No.: +91-9454695363
E-Mail: pruebiochem@gmail.com

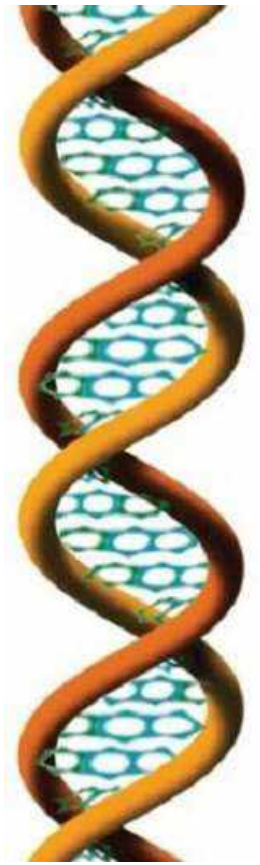
VEER BAHADUR SINGH PURVANCHAL UNIVERSITY JAUNPUR-222003

E –content

Course: M.Sc.
Subject: Biochemistry; Biotechnology, Microbiology, Environmental Science
Topic: Instrumentation and Analytical Techniques
S**ubtopic: DNA Sequencing**
Prepared by: Dr. Prabhakar Singh
Department : Biochemistry
Faculty : Science
Email: pruebiochem@gmail.com
Contact: +91-9454695363

Note-This E-content has been prepared as a reading material for students without any commercial interest. Original contributors are acknowledged.

DNA Sequencing



- Sanger sequencing (1977 – present)
- Next generation sequencing (2005 – present)
  - 454
  - Ion Torrent
  - Illumina
- 3rd generation sequencing (2012 – present)
  - PacBio or SMRT sequencing
  - Oxford Nanopore

Maxam-Gilbert sequencing

Allan Maxam and Walter Gilbert published a DNA sequencing method in 1977 based on chemical modification of DNA and subsequent cleavage at specific bases.

Also known as chemical sequencing, this method allowed purified samples of double-stranded DNA to be used without further cloning. This method's use of radioactive labeling and its technical complexity discouraged extensive use after refinements in the Sanger methods had been made.

Maxam-Gilbert sequencing requires radioactive labeling at one 5' end of the DNA and purification of the DNA fragment to be sequenced.

Chemical treatment then generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T).

The concentration of the modifying chemicals is controlled to introduce on average one modification per DNA molecule. Thus a series of labeled fragments is generated, from the radiolabeled end to the first "cut" site in each molecule.

The fragments in the four reactions are electrophoresed side by side in denaturing acrylamide gels 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 radiolabeled DNA fragment, from which the sequence may be inferred

Principle

The first DNA sequencing technique, using chemical reagents, was developed by Maxam and Gilbert (1977). This method is briefly described as-

A strand of source DNA is labeled at one end with ^{32}P. The two strands of DNA are then separated. The labeled DNA is distributed into four samples (in separate tubes).

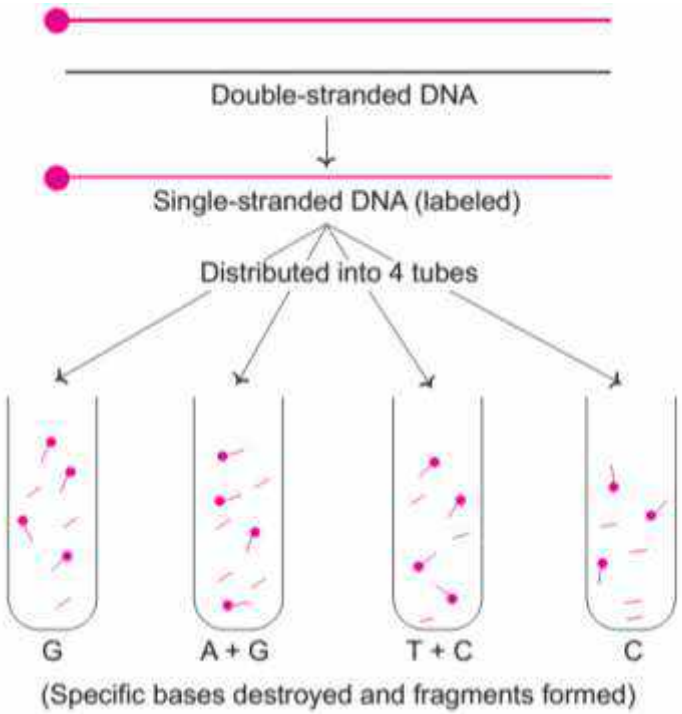
Each sample is subjected to treatment with a chemical that specifically destroys one (G, C) or two bases (A + G, T + C) in the DNA. Thus, the DNA strands are partially digested in four samples at sites G, A + G, T + C and C. This results in the formation of a series of labeled fragments of varying lengths.

The actual length of the fragment depends on the site at which the base is destroyed from the labeled end. Thus for instance, if there are C residues at positions 4, 7, and 10 away from the labeled end, then the treatment of DNA that specifically destroys C will give labeled pieces of length 3, 6 and 9 bases.

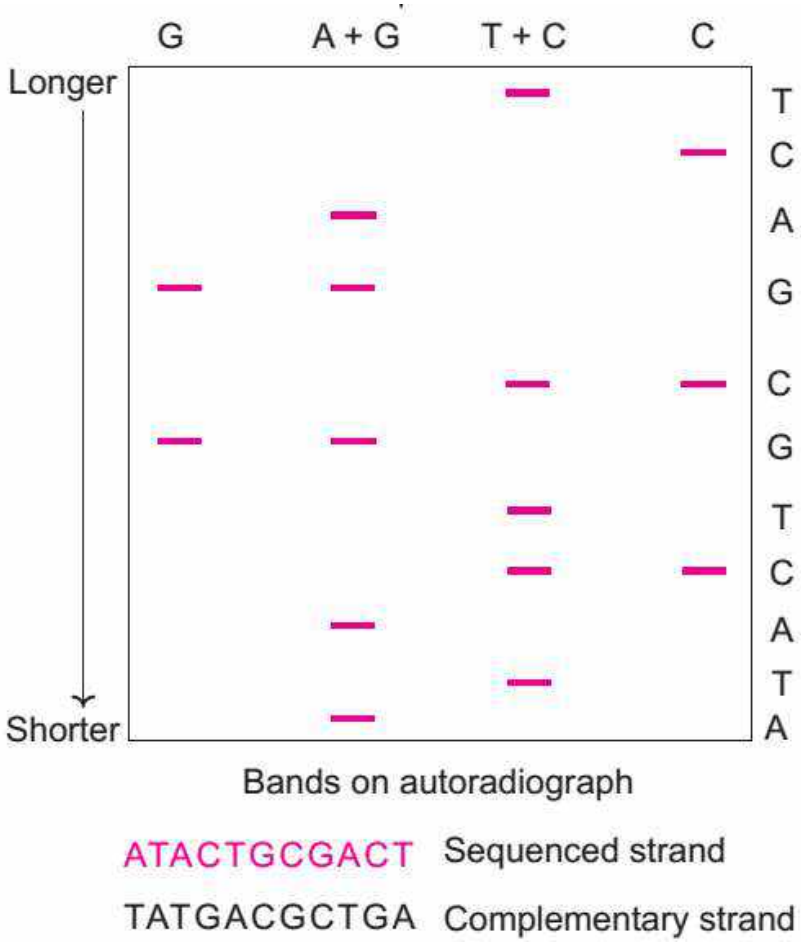
The labeled DNA fragments obtained in the four tubes are subjected to electrophoresis side by side and they are detected by autoradiograph.

The sequence of the bases in the DNA can be constructed from the bands on the electrophoresis.

Maxam and Gilbert method for DNA sequencing



Fragments separated by electrophoresis



SANGER SEQUENCING/ CHAIN-TERMINATOR/ DIDEOXY SEQUENCING

Sanger sequencing is a method of DNAsequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication.

Developed by Frederick Sanger and colleagues in 1977, it was the most widely used sequencing method for approximately 25 years.

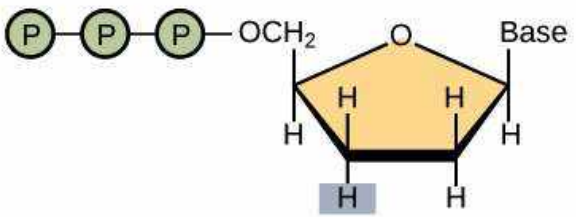
This method was superseded in 1977 by two different methods, that of Maxam and Gilbert (1977) and the chain-termination or dideoxy method (Sanger et al.1977b). For a while the Maxam and Gilbert method, which makes use of chemical reagents to bring about base-specific cleavage of DNA, was the favored procedure.

Principle

The chain-terminator or dideoxy procedure for DNA sequencing capitalizes on two properties of DNA polymerases: (i) their ability to synthesize faithfully a complementary copy of a single-stranded DNA template; and (ii) their ability to use 2', 3'-dideoxynucleotides as substrates.

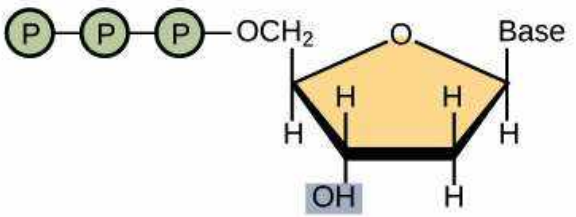
Once the analog is incorporated at the growing point of the DNA chain, the 3'end lacks a hydroxyl group and no longer is a substrate for chain elongation.

Thus, the growing DNA chain is terminated, i.e. dideoxynucleotides act as chain terminators. In practice, the Klenow fragment of DNA polymerase is used because this lacks the 5'→3'exonuclease activity associated with the intact enzyme.



Dideoxynucleotide (ddNTP)

Initiation of DNA synthesis requires a primer and usually this is a chemically synthesized oligonucleotide which is annealed close to the sequence being analyzed.

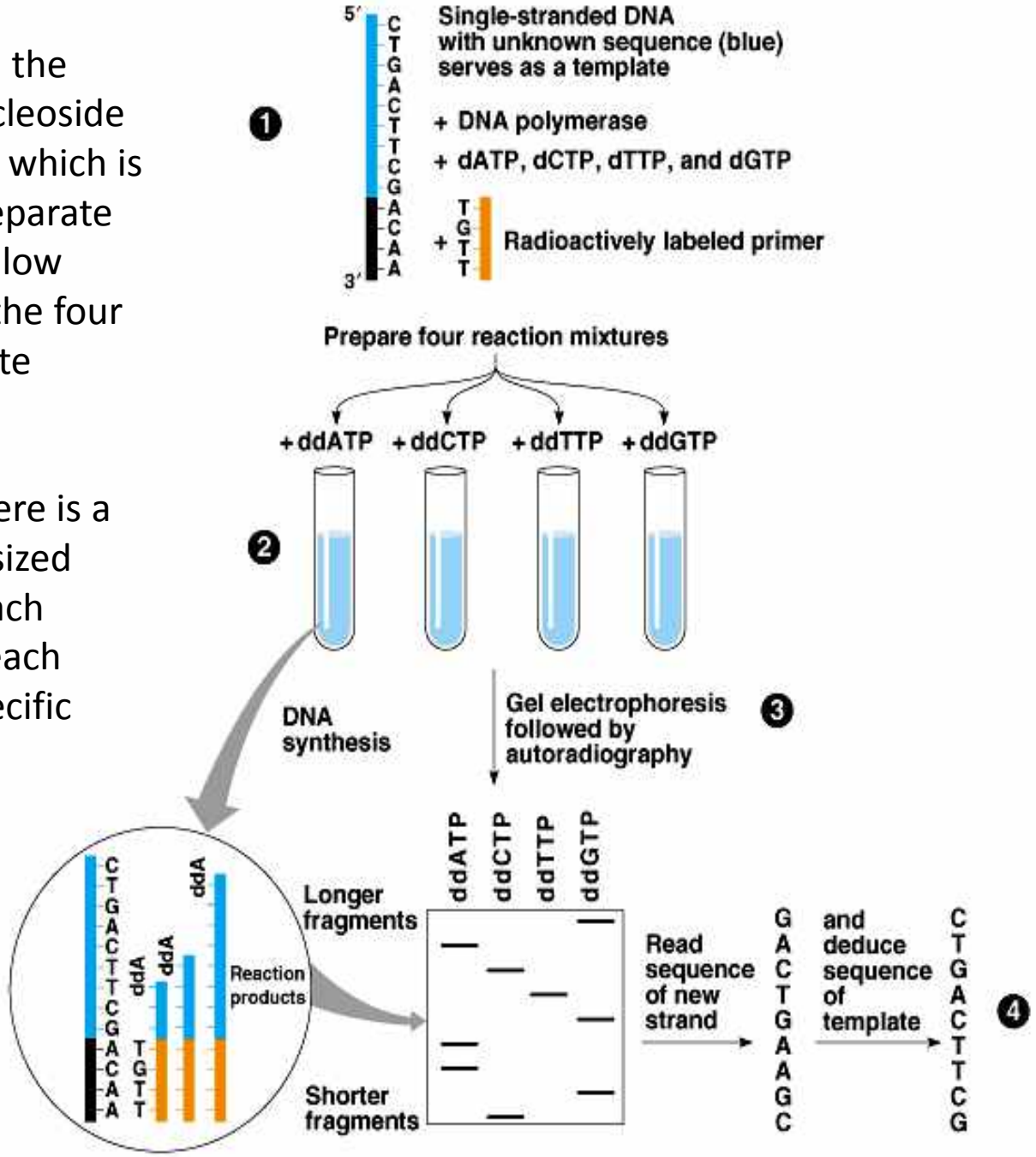


Deoxynucleotide (dNTP)

DNA synthesis is carried out in the presence of the four deoxynucleoside triphosphates, one or more of which is labeled with ^{32}P, and in four separate incubation mixes containing a low concentration of one each of the four dideoxynucleoside triphosphate analogs.

Therefore, in each reaction there is a population of partially synthesized radioactive DNA molecules, each having a common 5'end, but each varying in length to a base-specific 3'end.

After a suitable incubation period, the DNA in each mixture is denatured and electrophoresed in a sequencing gel

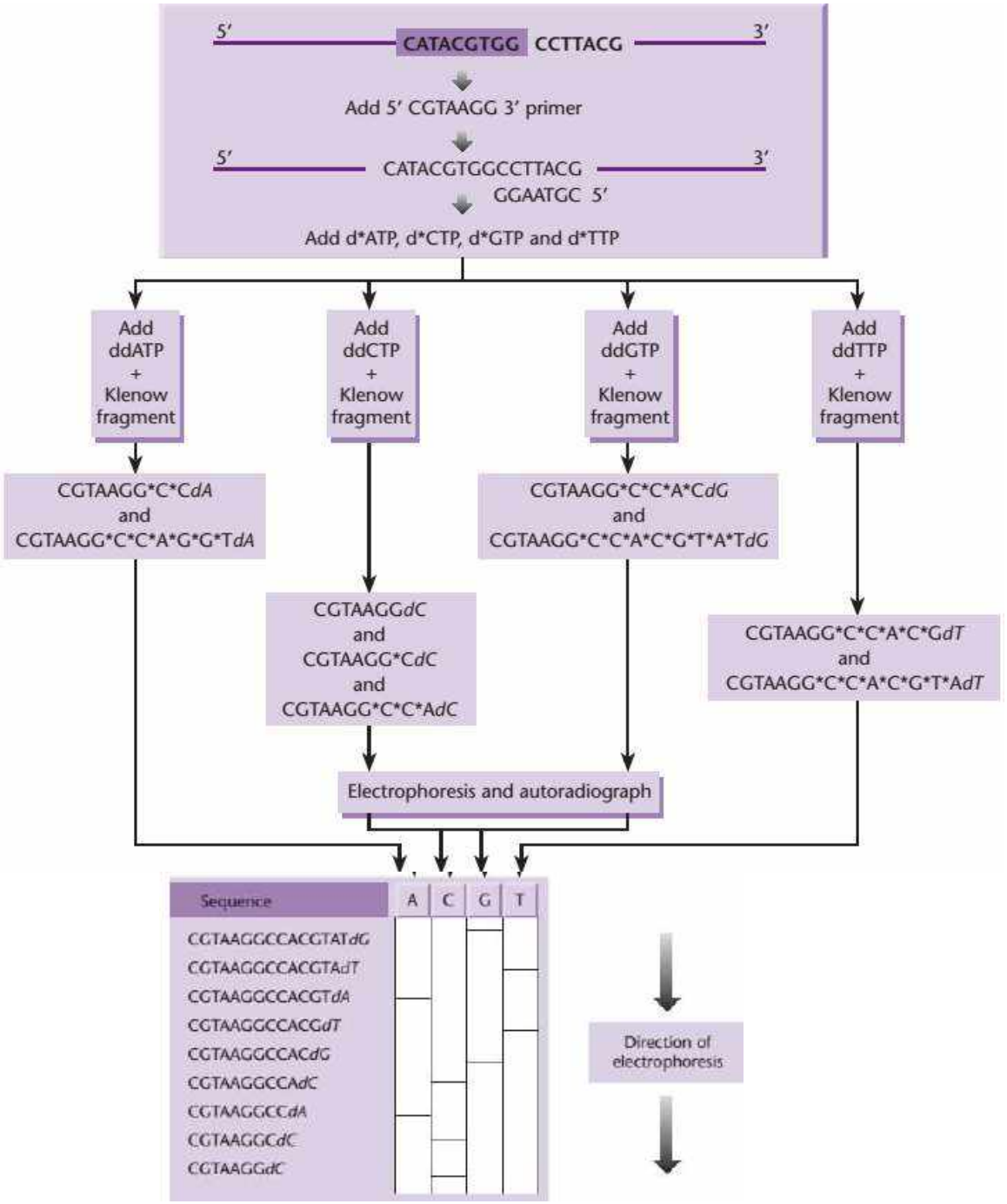


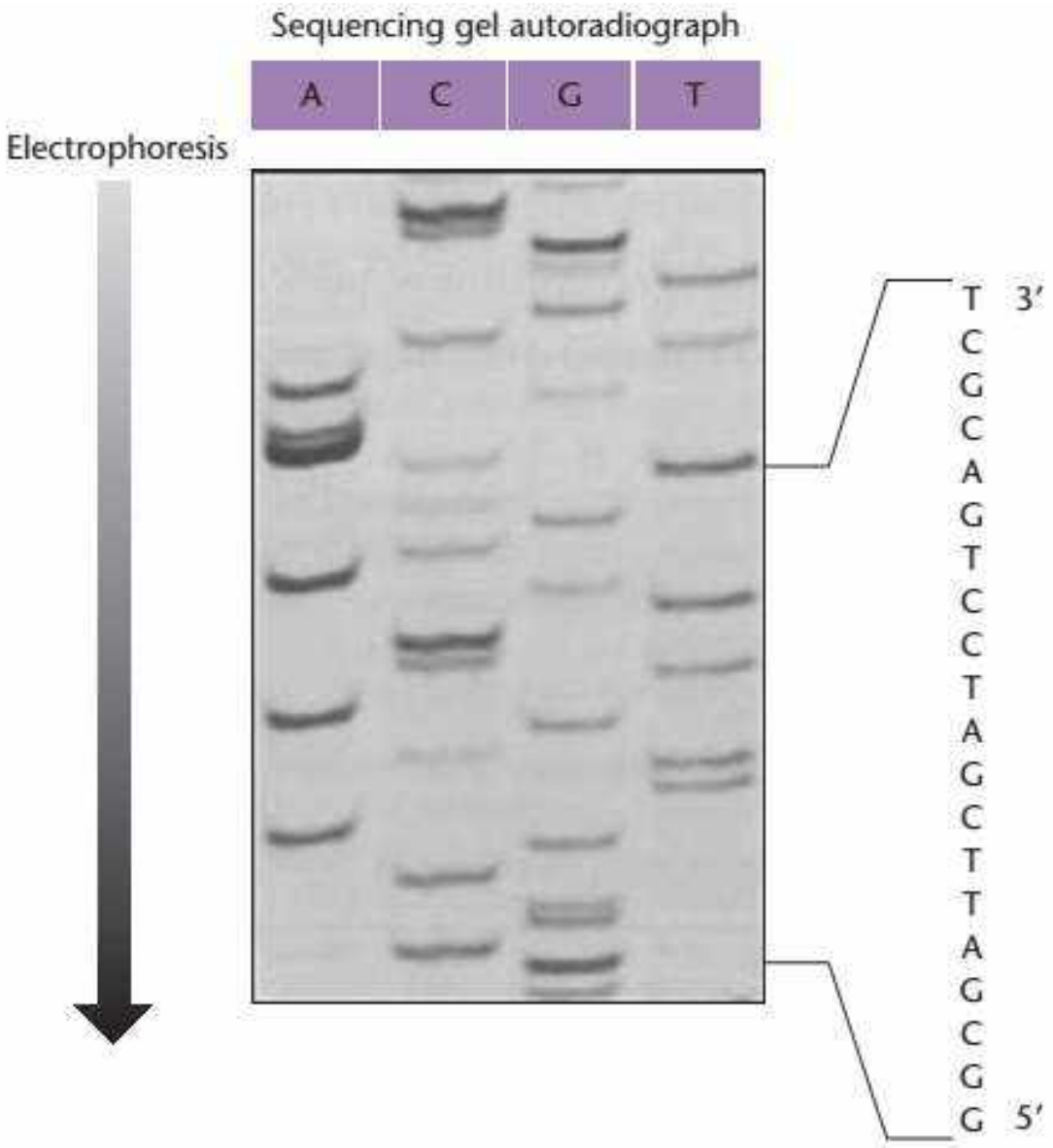
DNA sequencing with dideoxynucleoside triphosphates as chain terminators.

In this figure asterisks indicate the presence of 32P and the prefix “d” indicates the presence of a dideoxynucleotide.

At the top of the figure the DNA to be sequenced is enclosed within the box.

Note that unless the primer is also labeled with a radioisotope the smallest band with the sequence CGTAAGGdC will not be detected by autoradiography as no labeled bases were incorporated.





Enlarged autoradiograph of a sequencing gel obtained with the chain-terminator DNA sequencing method.

PYROSEQUENCING

Pyrosequencing is a DNA sequencing method that involves determining which one in the four bases is going to incorporate at each step in the DNA Template during DNA synthesis by DNA polymerase.

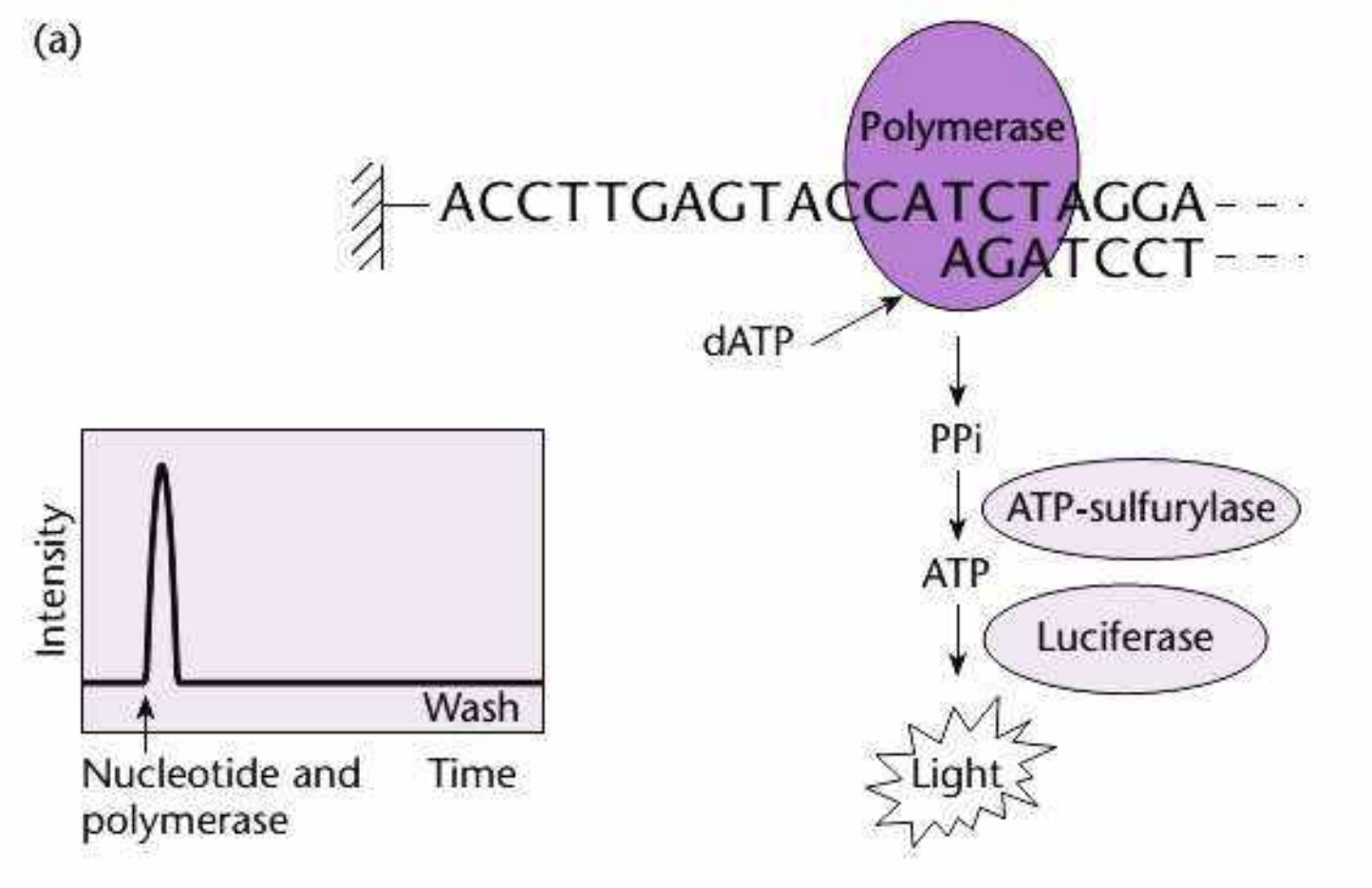
DNA polymerase moves along a single stranded template, each of the four nucleoside triphosphates is fed sequentially and then removed.

If one of the four bases is incorporated then pyrophosphate is released and this is detected in an enzyme cascade that emits light.

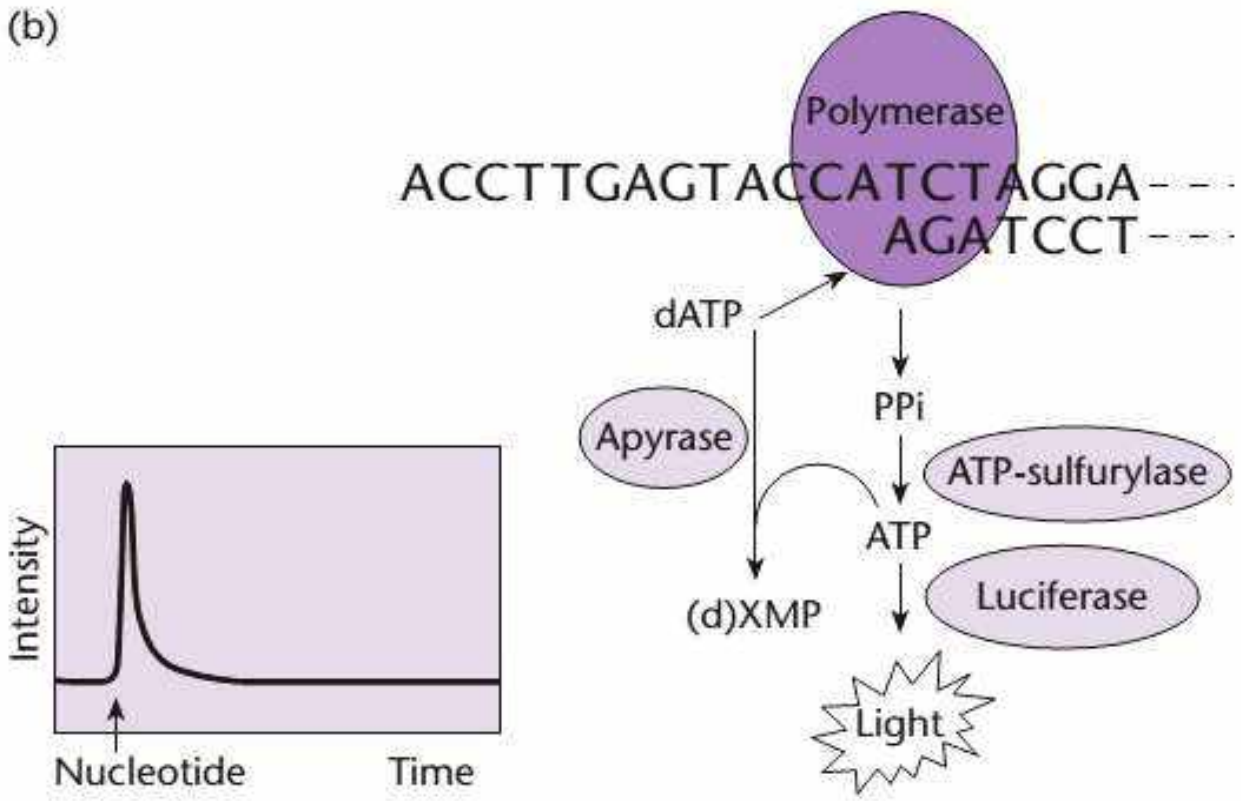
There are two variants of the pyrosequencing technique

1- In solid-phase pyrosequencing (Ronaghi et al.1996): The DNA to be sequenced is immobilized and a washing step is used to remove the excess substrate after each nucleotide addition.

2- In liquid-phase sequencing (Ronaghi et al. 1998b): A nucleotide degrading enzyme (apyrase) is introduced to make a four-enzyme system. Addition of this enzyme has eliminated the need for a solid support and intermediate washing thereby enabling the pyrosequencing reaction to be performed in a single tube.



(a) SOLID-PHASE PYROSEQUENCING: The four different nucleotides are added stepwise to the immobilized primed DNA template and the incorporation event is followed using the enzymes ATP sulfurylase and luciferase. After each nucleotide addition, a washing step is performed to allow iterative addition.

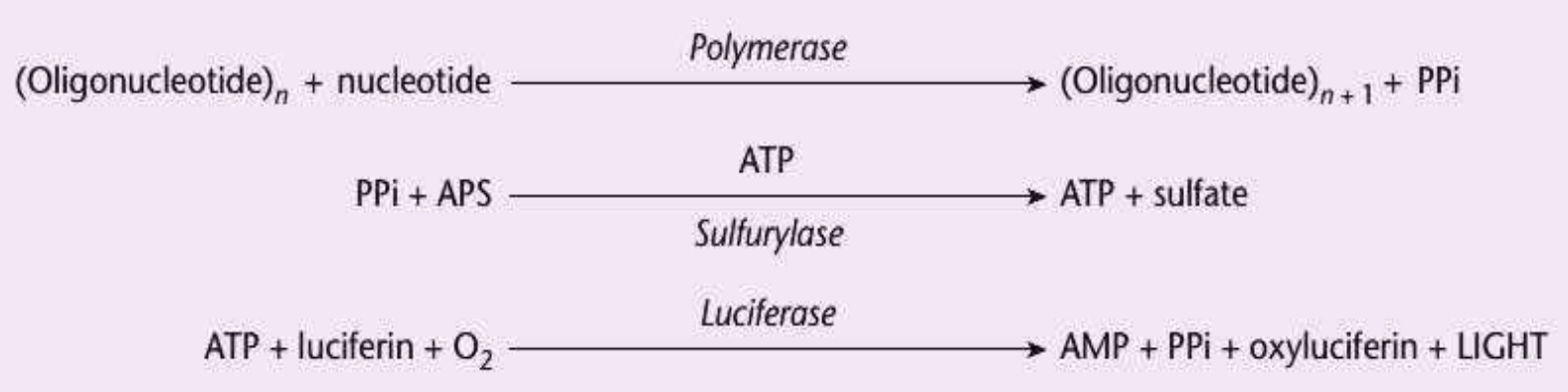


(b) LIQUID-PHASE PYROSEQUENCING:

The primed DNA template and four enzymes involved in liquid-phase pyrosequencing are placed in the well of a microtiter plate. The four different nucleotides are added stepwise and incorporation is followed using the enzymes ATP sulfurylase and luciferase. The nucleotides are continuously degraded by an enzyme allowing the addition of the subsequent nucleotide. dXTP indicates one of the four nucleotides. (Redrawn with permission from Ronaghi 2001.)

PRINCIPLE OF PYROSEQUENCING

1. **Pyrosequencing** is a method of DNA sequencing (determining the order of nucleotides in DNA) based on the "sequencing by synthesis" principle. it relies on the detection of pyrophosphate release on nucleotide incorporation, rather than chain termination with dideoxynucleotides.
2. The desired DNA sequence is able to be determined by light emitted upon incorporation of the next complementary nucleotide by the fact that only one out of four of the possible A/T/C/G nucleotides are added and available at a time so that only one letter can be incorporated on the single stranded template.
3. Light emitted by the enzyme cascade is directly proportional to the amount of pyrophosphate released, The intensity of the light determines if there are continuous more than one of these “Nucleotide" in a row.



The general principle of pyrosequencing. A polymerase catalyzes incorporation of nucleotides into a nucleic acid chain. As each nucleotide is incorporated a pyrophosphate (PPi) molecule is released and incorporated into ATP by ATP sulfurylase. On addition of luciferin and the enzyme luciferase, this ATP is degraded to AMP with the production of light.

TEMPLATE PREPARATION

Template preparation for pyrosequencing is very easy. After generation of the template by the PCR, unincorporated nucleotides and PCR primers are removed.

Two methods have been developed for this purification step-

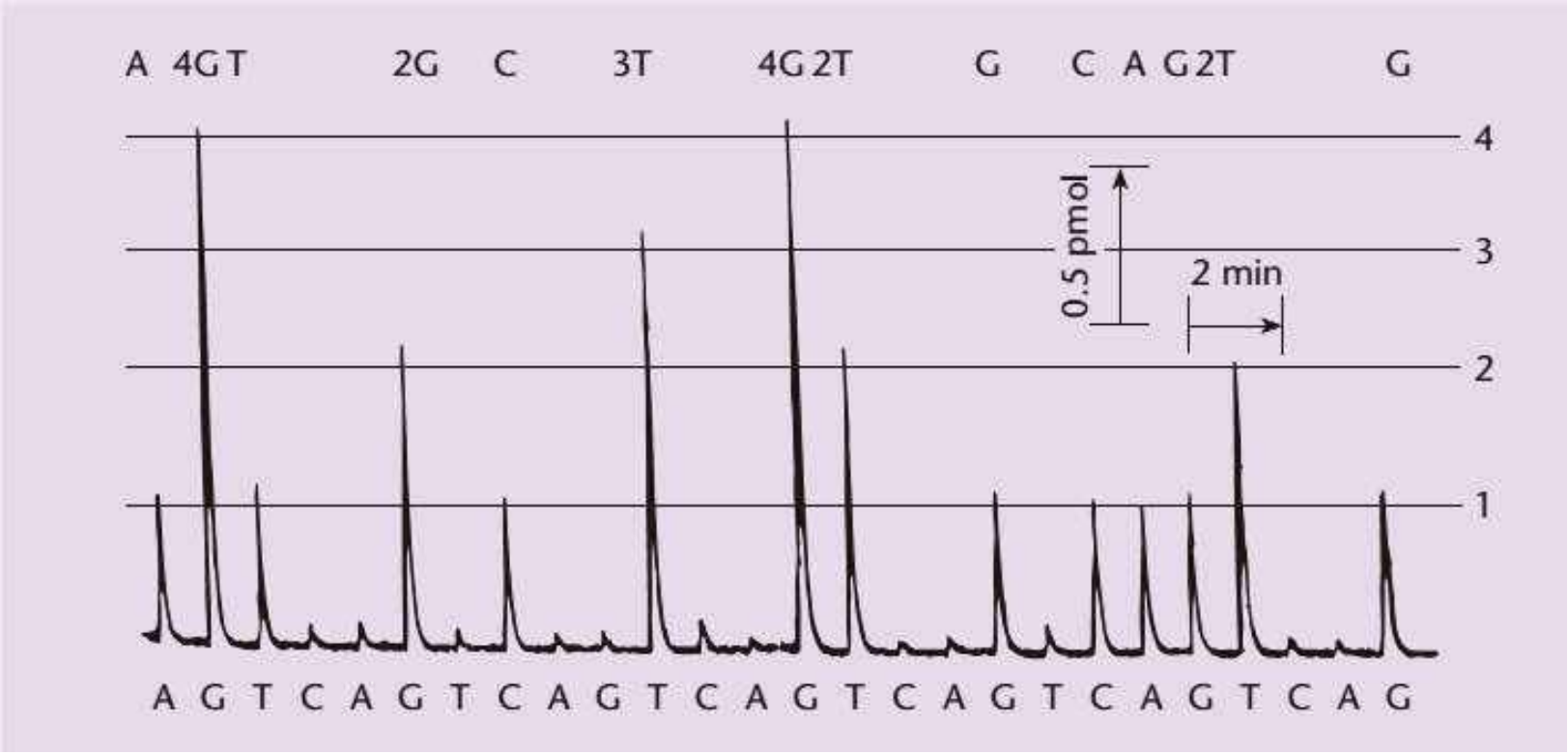
1. In the first, biotinylated PCR product is captured on magnetic beads, washed, and denatured with alkali (Ronaghi et al.1998a).

2. In the second method, akaline phosphatase or apyrase and exonuclease I are added to the PCR product to destroy the nucleotides and primers, respectively. The sequencing primer is then added and the mixture rapidly heated and cooled. This inactivates the enzymes, denatures the DNA and enables the primers to anneal to the templates (Nordstrom et al. 2000)

The acceptable read length of pyrosequencing currently is about 200 nucleotides, i.e. much lessthan is achieved with Sanger sequencing. However, many modifications are being made to the reaction conditions to extend the read length. For example, the addition of ssDNA-binding protein to the reaction mixture increases read length, facilitates sequencing of difficult templates, and provides flexibility in primer design.

PYROGRAM

Pyrogram of the raw data obtained from liquid-phase pyrosequencing. Proportional signals are obtained for one, two, three, and four base incorporations. Nucleotide addition, according to the order of nucleotides, is indicated below the pyrogram and the obtained sequence is indicated above the pyrogram. (Redrawn with permission from Ronaghi 2001.)



Sanger Sequencing Vs Pyrosequencing

Currently, a limitation of the method is that the lengths of individual reads of DNA sequence are in the neighborhood of 300-500 nucleotides, shorter than the 800-1000 obtainable with chain termination methods (e.g. Sanger sequencing).

This can make the process of genome assembly more difficult, particularly for sequences containing a large amount of repetitive DNA.

As of 2007, pyrosequencing is most commonly used for resequencing or sequencing of genomes for which the sequence of a close relative is already available.

The templates for pyrosequencing can be made both by solid phase template preparation (streptavidin-coated magnetic beads) and enzymatic template preparation (apyrase+exonuclease).

So Pyrosequencing can be differentiated into two types, namely Solid Phase Pyrosequencing and Liquid Phase Pyrosequencing.

Next Generation Sequencing (NGS)

Next generation sequencing (2005 – present)

- 454
- Ion Torrent
- Illumina

454 Pyrosequencing or Roche454

A parallelized version of pyrosequencing was developed by 454 Life Sciences, which has since been acquired by Roche Diagnostics in 2007. The 454 DNA sequencer was the first next-generation sequencer to become commercially successful.

454 utilizes the detection of pyrophosphate released by the DNA polymerase reaction when adding a nucleotide to the template strain.

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 picoliter-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 reads.

Roche currently manufactures two systems based on their pyrosequencing technology: the GS FLX+ and the GS Junior System.[11] The GS FLX+ System promises read lengths of approximately 1000 base pairs while the GS Junior System promises 400 base pair reads.[12][13]

In 2009, Roche launched the GS Junior, a bench top version of the 454 sequencer with read length up to 400bp, and simplified library preparation and data processing.

Advantages of 454 systems is their running speed, Manpower can be reduced with automation of library preparation and semi-automation of emulsion PCR.

Disadvantage: Price of reagents is relatively more expensive compared with other next-generation sequencers.

Homopolymer Error: 454 system is prone to errors when estimating the number of bases in a long string of identical nucleotides. This is referred to as a homopolymer error and occurs when there are 6 or more identical bases in row.

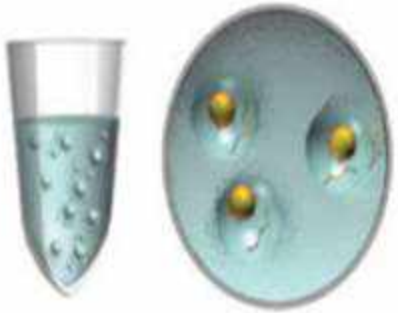
NOTE: In 2013 Roche announced that they would be shutting down development of 454 technology and phasing out 454 machines completely in 2016 but Roche produces software tools which optimised the analysis of 454 sequencing data.

GS Run Processor converts raw images generated by a sequencing run into intensity values. The process consists of two main steps: image processing and signal processing. The software also applies normalization, signal correction, base-calling and quality scores for individual reads.

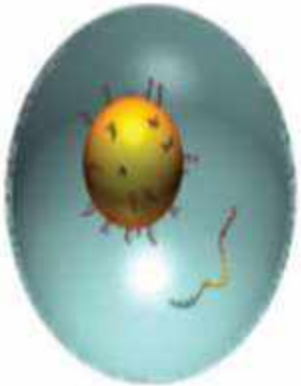
Roche

454 SEQUENCING

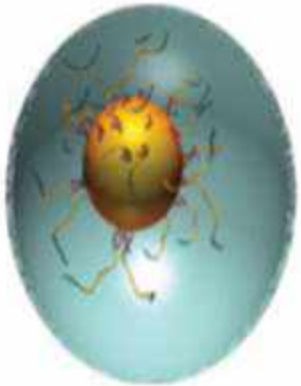
- DNA is fragmented, adapter ligated and denatured
- Emulsion PCR



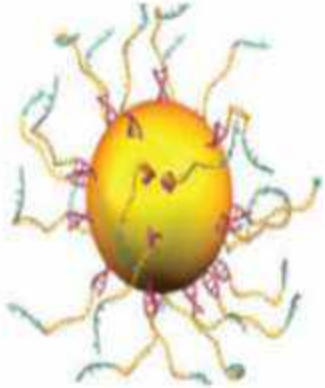
Anneal sstDNA to an excess of DNA capture beads



Emulsify beads and PCR reagents in water-in-oil microreactors

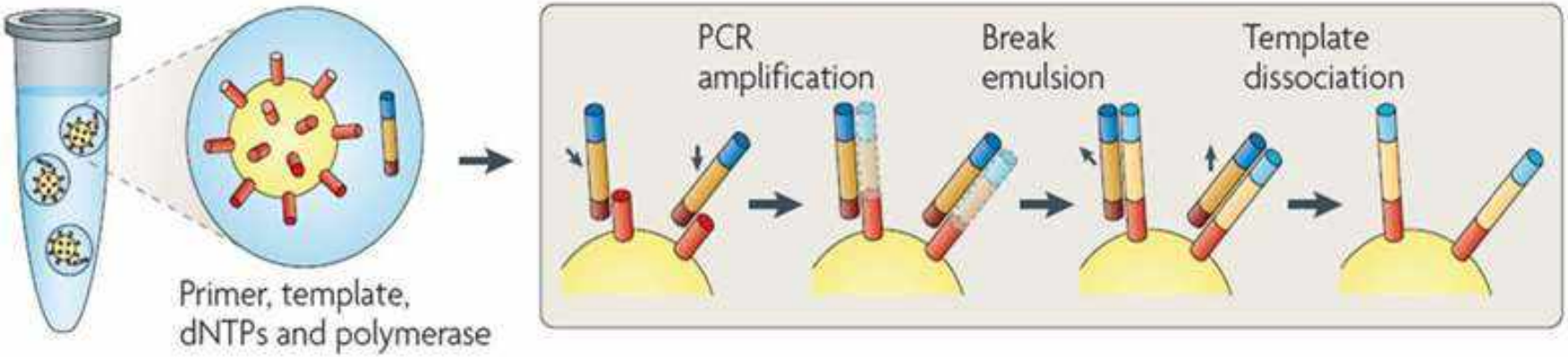


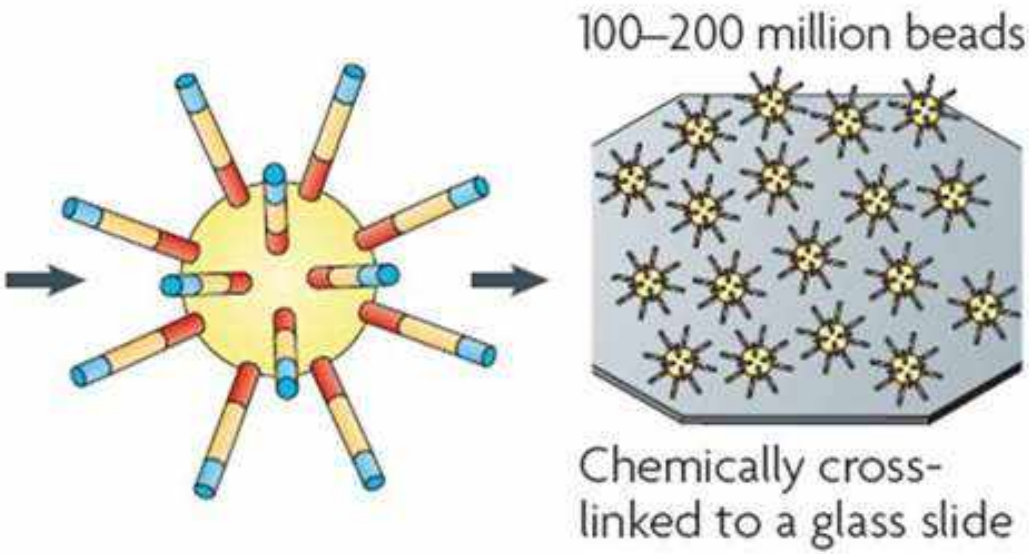
Clonal amplification occurs inside microreactors



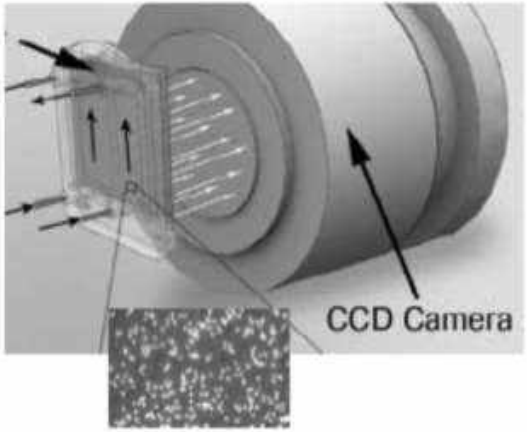
Break microreactors and enrich for DNA-positive beads

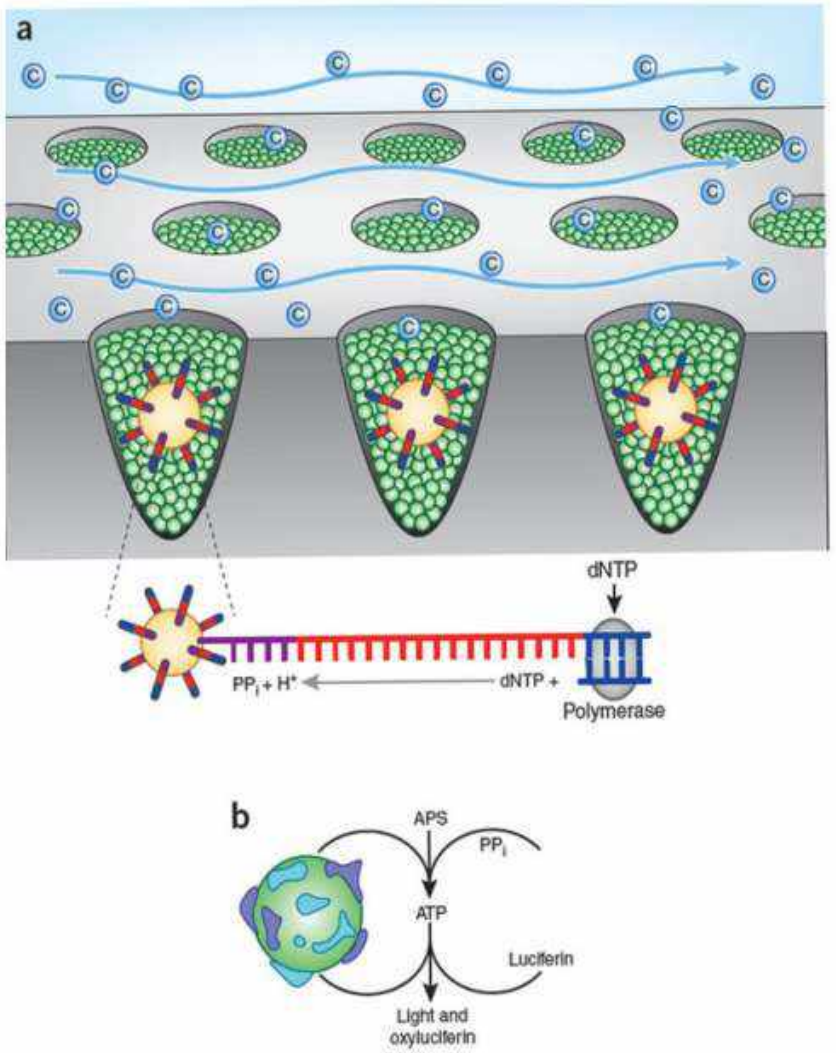
One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion





- Sequencing
  - One bead = one read
  - Reads up to 700 bp
  - 1 million reads / run



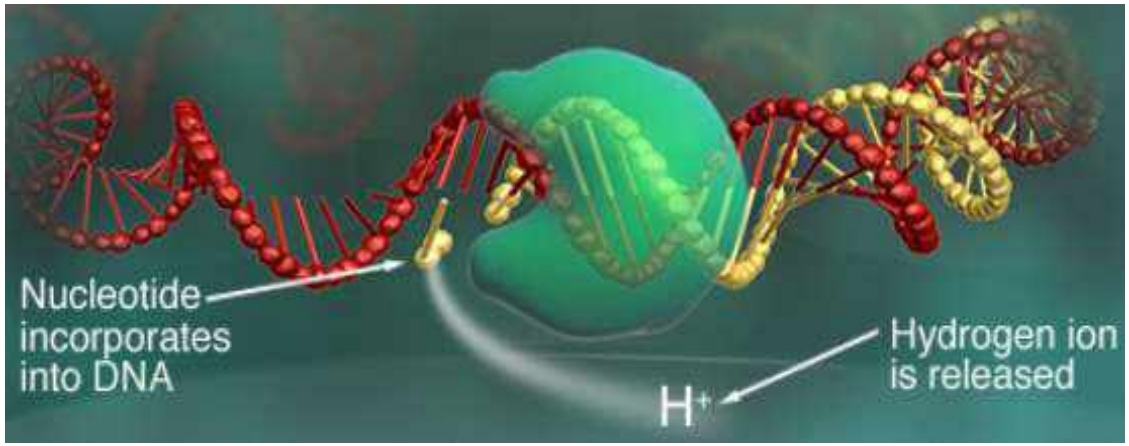


ION TORRENT / ION SEMICONDUCTOR SEQUENCING

The technology was licensed from DNA Electronics Ltd,-developed by Ion Torrent Systems Inc. and was released in February 2010. Ion Torrent have marketed their machine as a rapid, compact and economical sequencer that can be utilized in a large number of laboratories as a bench top machine.

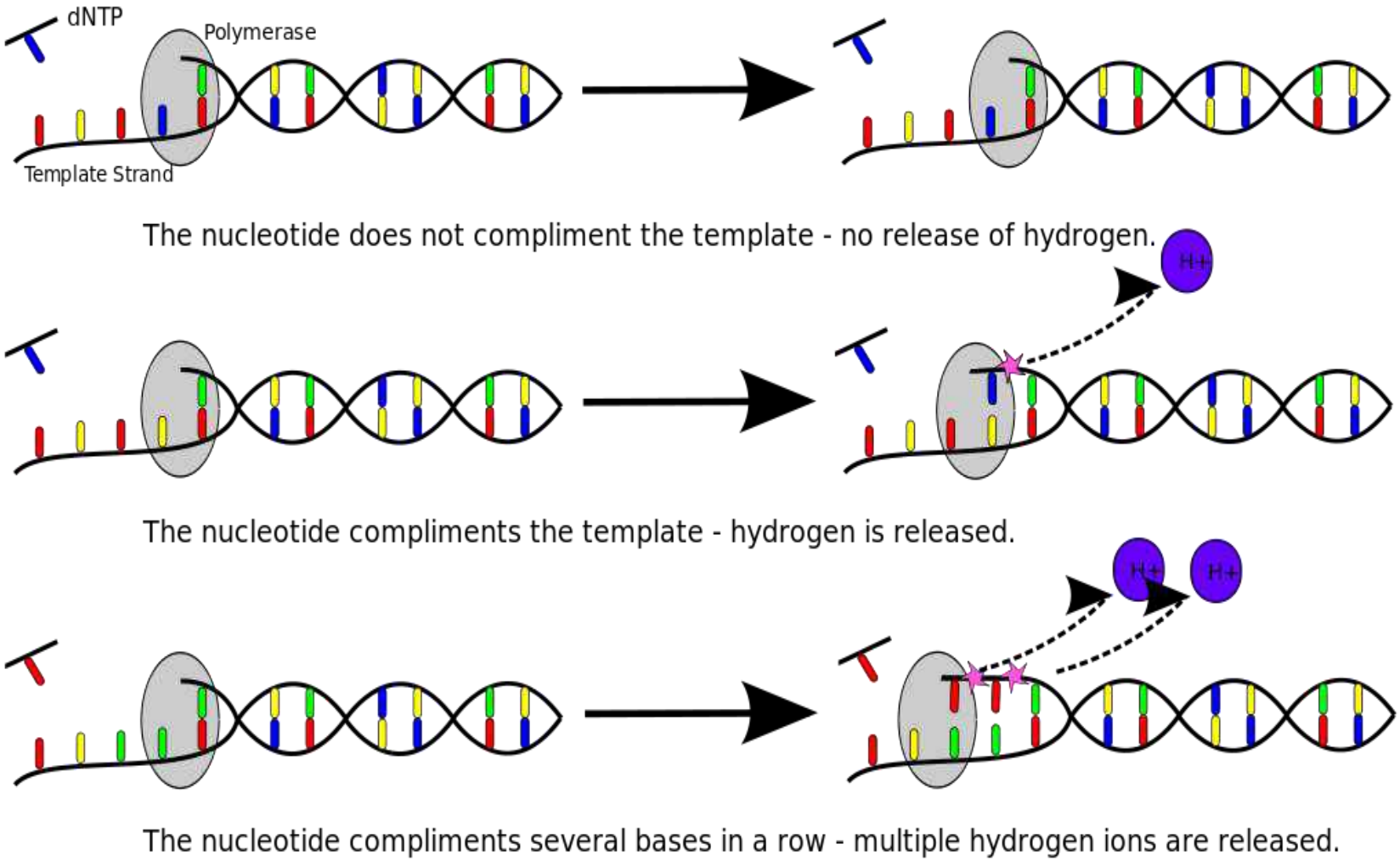
Roche's 454 Life Sciences is partnering with DNA Electronics on the development of a long-read, high-density semiconductor sequencing platform using this technology.

This technology differs from other sequencing-by-synthesis technologies in that no modified nucleotides or optics are used. Ion semiconductor sequencing may also be referred to as Ion Torrent sequencing, pH-mediated sequencing, silicon sequencing, or semiconductor sequencing.

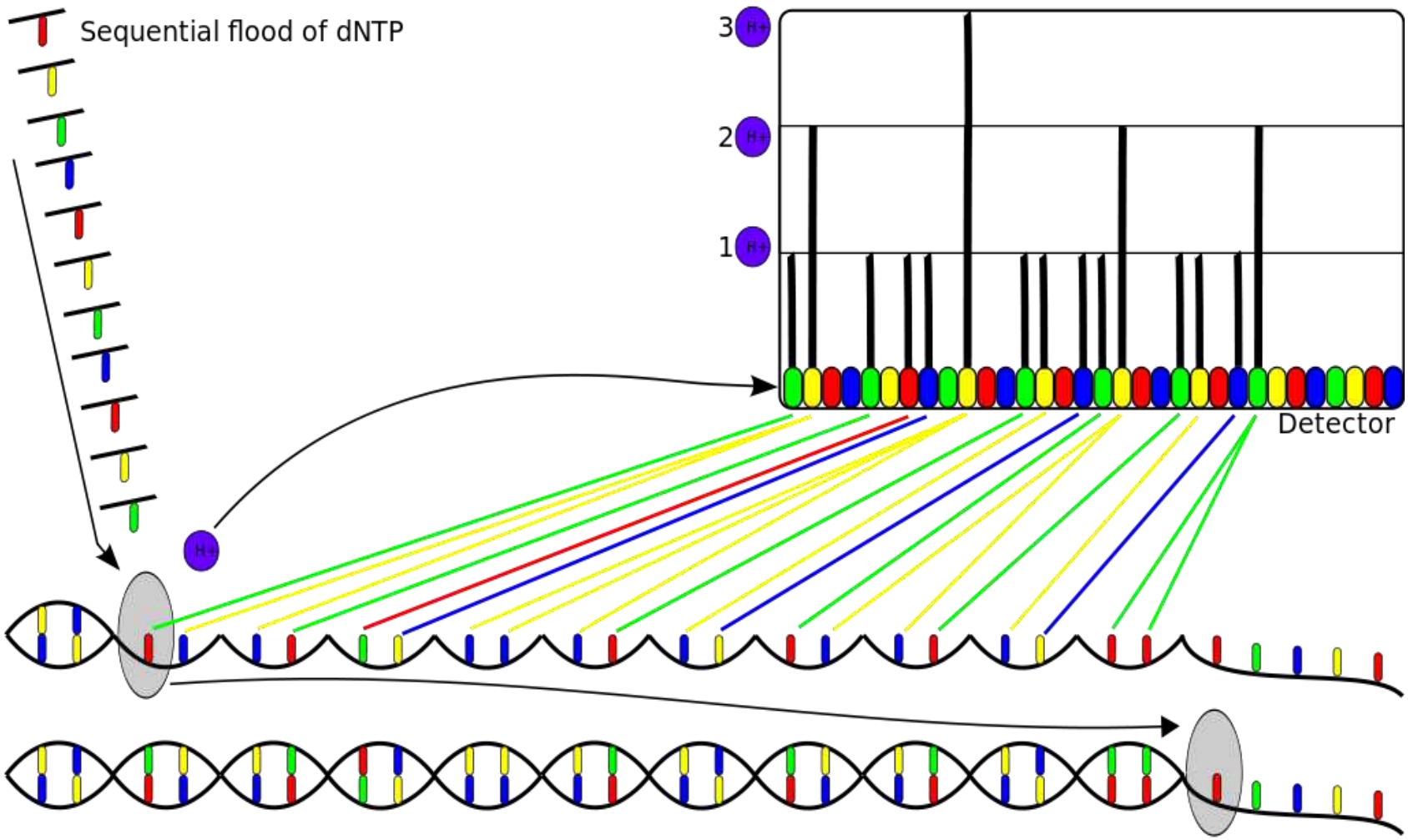


Principle

- **Ion semiconductor sequencing** is a method of DNA sequencing based on the detection of hydrogen ions that are released during the polymerization of DNA. This is a method of "sequencing by synthesis", during which a complementary strand is built based on the sequence of a template strand.
- the incorporation of a deoxyribonucleoside triphosphate (dNTP) into a growing DNA strand involves the formation of a covalent bond and the release of pyrophosphate and a positively charged hydrogen ion A dNTP will only be incorporated if it is complementary to the leading unpaired template nucleotide. Ion semiconductor sequencing exploits these facts by determining if a hydrogen ion is released upon providing a single species of dNTP to the reaction.
- A microwell containing a template DNA strand to be sequenced is flooded with a single species of deoxyribonucleotide triphosphate (dNTP). If the introduced dNTP is complementary to the leading template nucleotide, it is incorporated into the growing complementary strand. This causes the release of a hydrogen ion that triggers an ISFET ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence, multiple dNTP molecules will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher electronic signal.



The release of hydrogen ions indicate if zero, one or more nucleotides were incorporated.



Released hydrogens ions are detected by an ion sensor. Multiple incorporations lead to a corresponding number of released hydrogens and intensity of signal.

Microwells on a semiconductor chip that each contain many copies of one single-stranded template DNA molecule to be sequenced and DNA polymerase are sequentially flooded with unmodified A, C, G or T dNTP.

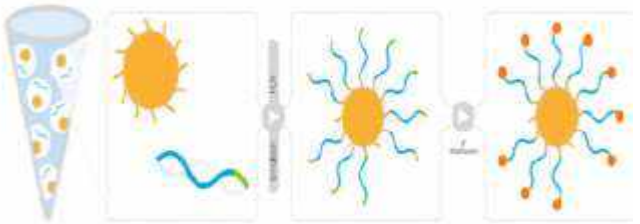
If an introduced dNTP is complementary to the next unpaired nucleotide on the template strand it is incorporated into the growing complementary strand by the DNA polymerase

If the introduced dNTP is not complementary there is no incorporation and no biochemical reaction.

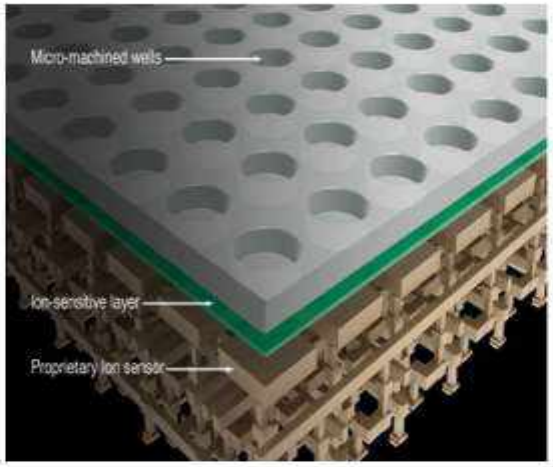
The hydrogen ion that is released in the reaction changes the pH of the solution, which is detected by an ISFET.

The unattached dNTP molecules are washed out before the next cycle when a different dNTP species is introduced.

- The first steps are similar to 454 sequencing
  - Bind DNA to beads
  - Emulsion PCR



- Chip with ion sensor



Signal detection

- Beneath the layer of microwells is an ion sensitive layer, below which is an ISFET ion sensor. All layers are contained within a CMOS semiconductor chip, similar to that used in the electronics industry.

- Each chip contains an array of microwells with corresponding ISFET detectors.Each released hydrogen ion then triggers the ISFET ion sensor. The series of electrical pulses transmitted from the chip to a computer is translated into a DNA sequence, with no intermediate signal conversion required.

- Because nucleotide incorporation events are measured directly by electronics, the use of labeled nucleotides and optical measurements are avoided. Signal processing and DNA assembly can then be carried out in software.

Sequencing characteristics

The per base accuracy achieved in house by Ion Torrent on the Ion Torrent Ion semiconductor sequencer as of February 2011 was 99.6% based on 50 base reads, with 100 Mb per run. The read-length as of February 2011 was 100 base pairs. The accuracy for homopolymer repeats of 5 repeats in length was 98%. Later releases show a read length of 400 base pairs.

Strengths

- The major benefits of ion semiconductor sequencing are rapid sequencing speed and low upfront and operating costs. This has been enabled by the avoidance of modified nucleotides and optical measurements.
- Because the system records natural polymerase-mediated nucleotide incorporation events, sequencing can occur in real-time. In reality, the sequencing rate is limited by the cycling of substrate nucleotides through the system Ion Torrent Systems Inc., the developer of the technology, claims that each incorporation measurement takes 4 seconds and each run takes about one hour, during which 100-200 nucleotides are sequenced. If the semiconductor chips are improved (as predicted by Moore's law), the number of reads per chip (and therefore per run) should increase.
- The cost of acquiring a pH-mediated sequencer from Ion Torrent Systems Inc. at time of launch was priced at around $50,000 USD, excluding sample preparation equipment and a server for data analysis. The cost per run is also significantly lower than that of alternative automated sequencing methods, at roughly $1,000

Limitations

- If homopolymer repeats of the same nucleotide (e.g. TTTTT) are present on the template strand (strand to be sequenced) then multiple introduced nucleotides are incorporated and more hydrogen ions are released in a single cycle. This results in a greater pH change and a proportionally greater electronic signal.This is a limitation of the system in that it is difficult to enumerate long repeats. This limitation is shared by other techniques that detect single nucleotide additions such as pyrosequencing.[v]Signals generated from a high repeat number are difficult to differentiate from repeats of a similar but different number; *e.g.*, homorepeats of length 7 are difficult to differentiate from those of length 8.

- Another limitation of this system is the short read length compared to other sequencing methods such as Sanger sequencing or pyrosequencing. Longer read lengths are beneficial for *de novo* genome assembly. Ion Torrent semiconductor sequencers produce an average read length of approximately 400 nucleotides per read.

- The throughput is currently lower than that of other high-throughput sequencing technologies, although the developers hope to change this by increasing the density of the chip

Application

- The developers of Ion Torrent semiconductor sequencing have marketed it as a rapid, compact and economical sequencer that can be utilized in a large number of laboratories as a bench top machine.The company hopes that their system will take sequencing outside of specialized centers and into the reach of hospitals and smaller laboratories.[17] A January 2011 New York Times article, "Taking DNA Sequencing to the Masses", underlines these ambitions.[17]

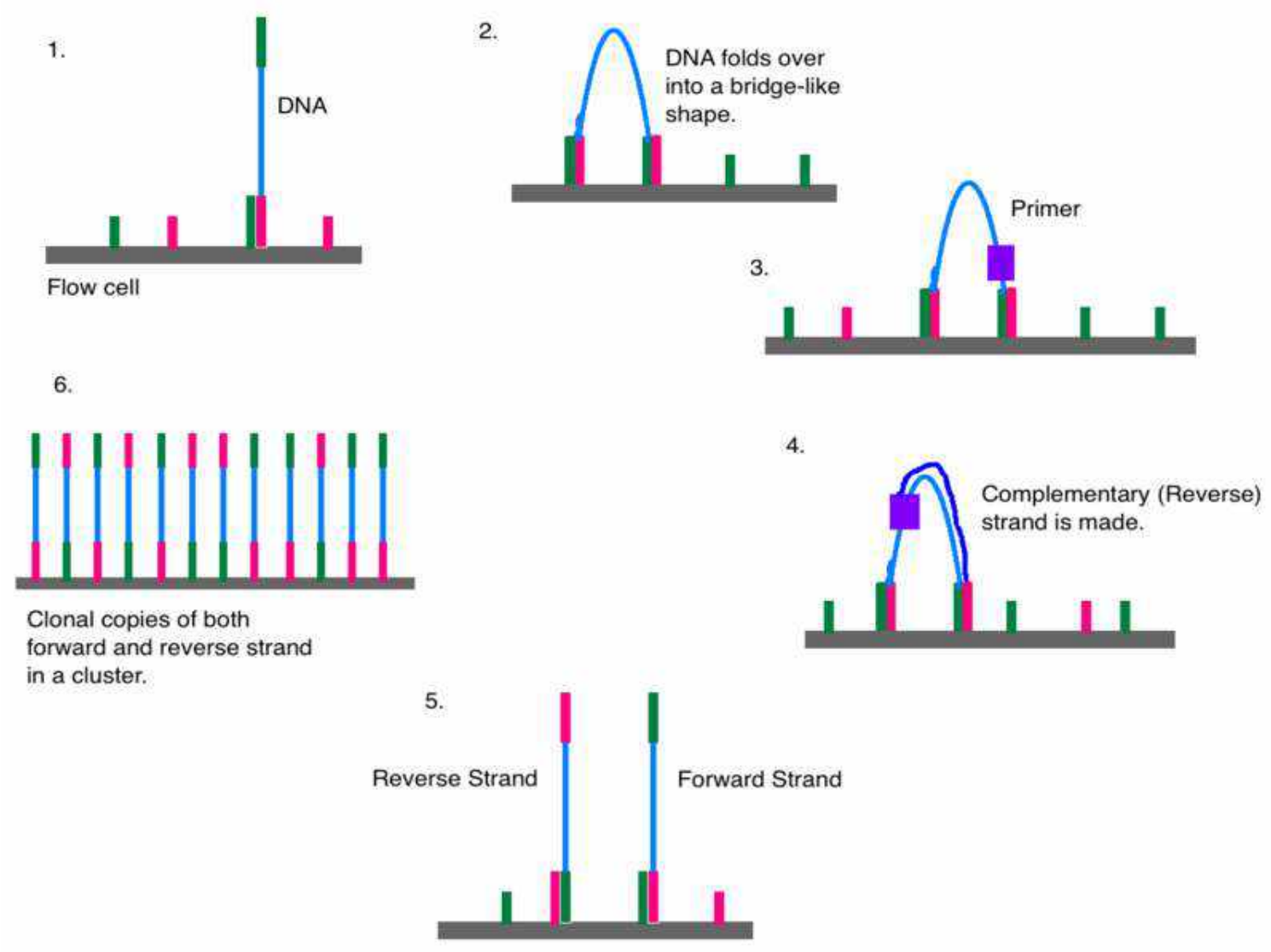
- Due to the ability of alternative sequencing methods to achieve a greater read length (and therefore being more suited to whole genome analysis) this technology may be best suited to small scale applications such as microbial genome sequencing, microbial transcriptome sequencing, targeted sequencing, amplicon sequencing, or for quality testing of sequencing libraries

Illumina- Next-generation Sequencing

- **Ilumina dye sequencing** is a technique used to determine the series of base pairs in DNA, also known as DNA sequencing.
- The reversible terminated chemistry concept was invented by Bruno Canard and Simon Sarfati at the Pasteur Institute in Paris.
- It was developed by Shankar Balasubramanian and David Klenerman of Cambridge University, who subsequently founded Solexa, a company later acquired by Illumina.
- This sequencing method is based on reversible dye-terminators that enable the identification of single bases as they are introduced into DNA strands.
- It can also be used for whole-genome and region sequencing, transcriptome analysis, metagenomics, small RNA discovery, methylation profiling, and genome-wide protein-nucleic acid interaction analysis

Mechanism

1. Illumina sequencing technology works in three basic steps: amplify, sequence, and analyze. The process begins with purified DNA. The DNA gets chopped up into smaller pieces and given adapters, indices, and other kinds of molecular modifications that act as reference points during amplification, sequencing, and analysis.
2. The modified DNA is loaded onto a specialized chip where amplification and sequencing will take place. Along the bottom of the chip are hundreds of thousands of oligonucleotides (short, synthetic pieces of DNA). They are anchored to the chip and able to grab DNA fragments that have complementary sequences. Once the fragments have attached, a phase called cluster generation begins. This step makes about a thousand copies of each fragment of DNA.
3. Next, primers and modified nucleotides enter the chip. These nucleotides have reversible 3' blockers that force the polymerase to add on only one nucleotide at a time as well as fluorescent tags. After each round of synthesis, a camera takes a picture of the chip.
4. A computer determines what base was added by the wavelength of the fluorescent tag and records it for every spot on the chip. After each round, non-incorporated molecules are washed away.
5. A chemical deblocking step is then used in the removal of the 3’ terminal blocking group and the dye in a single step. The process continues until the full DNA molecule is sequenced.With this technology, thousands of places throughout the genome are sequenced at once via massive parallel sequencing.

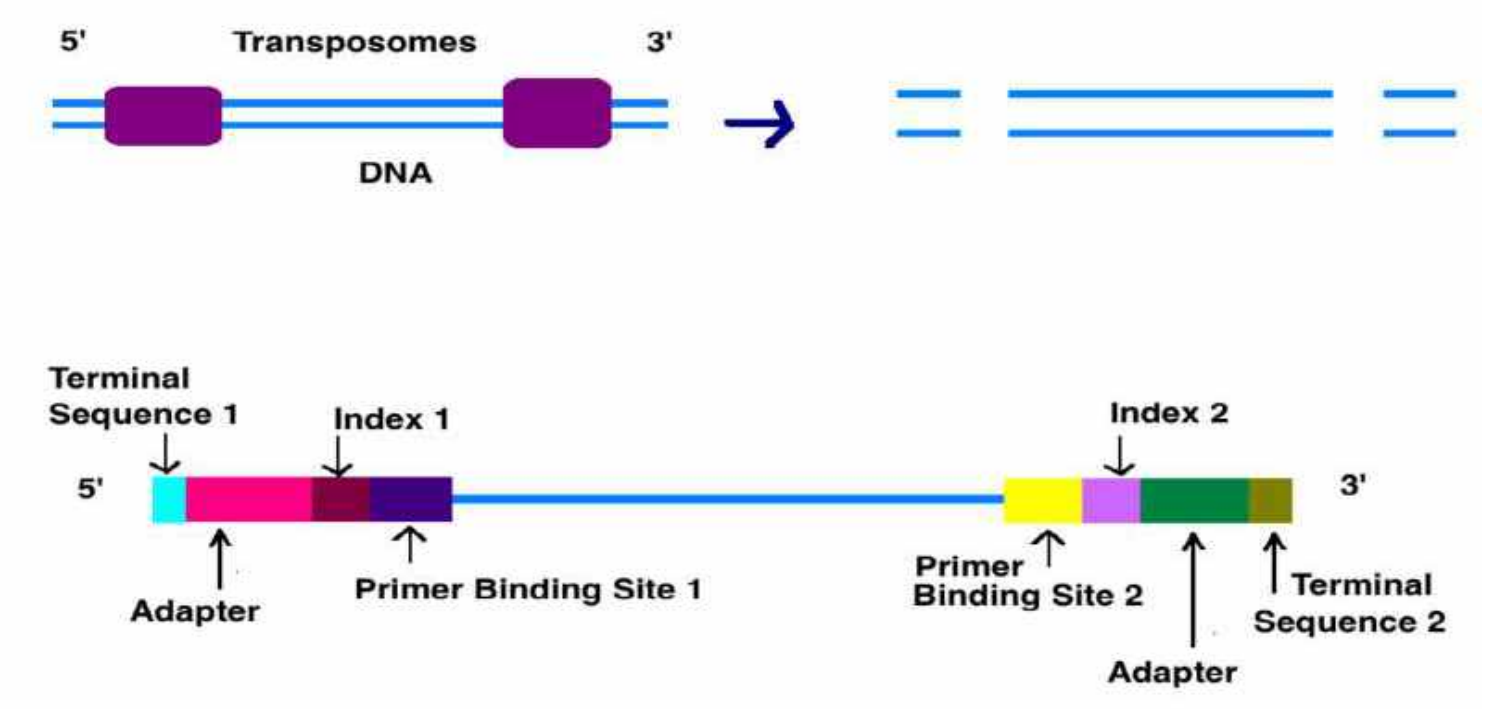


The DNA attaches to the flow cell via complementary sequences. The strand bends over and attaches to a second oligo forming a bridge. A polymerase synthesizes the reverse strand. The two strands release and straighten. Each forms a new bridge (bridge amplification). The result is a cluster of DNA forward and reverse strands clones.

PROCEDURE

1. Tagmentation

The first step after DNA purification is tagmentation. Transposases randomly cut the DNA into short segments ("tags"). Adapters are added on either side of the cut points (ligation). Strands that fail to have adapters ligated are washed away



Double stranded DNA is cleaved by transposomes. The cut ends are repaired and adapters, indices, primer binding sites, and terminal sites are added to each strand of the DNA.

2. Reduced cycle amplification

The next step is called reduced cycle amplification. During this step, sequences for primer binding, indices, and terminal sequences are added.

Indices are usually six base pairs long and are used during DNA sequence analysis to identify samples. Indices allow for up to 96 different samples to be run together. During analysis, the computer will group all reads with the same index together.

The terminal sequences are used for attaching the DNA strand to the flow cell. Illumina uses a "sequence by synthesis" approach.

This process takes place inside of an acrylamide-coated glass flow cell. The flow cell has oligonucleotides (short nucleotide sequences) coating the bottom of the cell, and they serve to hold the DNA strands in place during sequencing.

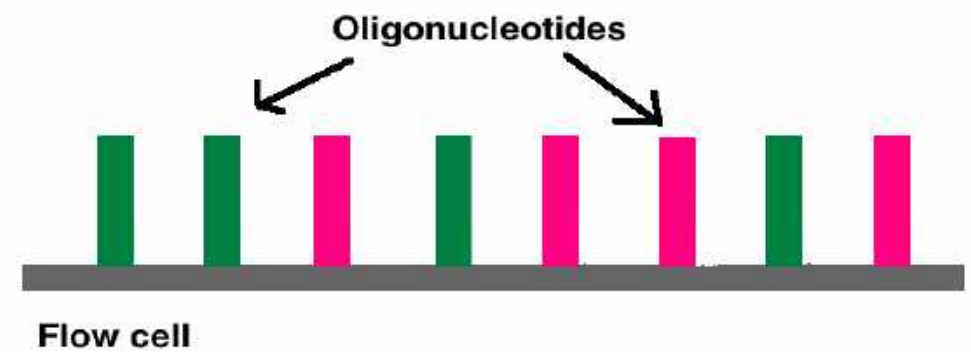
The oligos match the two kinds of terminal sequences added to the DNA during reduced cycle amplification. As the DNA enters the flow cell, one of the adapters attaches to a complementary oligo.

2. Bridge Amplification

Once attached, cluster generation can begin. The goal is to create hundreds of identical strands of DNA. Some will be the forward strand; the rest, the reverse. Clusters are generated through bridge amplification.

Polymerases move along a strand of DNA, creating its complementary strand. The original strand is washed away, leaving only the reverse strand. At the top of the reverse strand there is an adapter sequence. The DNA strand bends and attaches to the oligo that is complementary to the top adapter sequence. Polymerases attach to the reverse strand, and its complementary strand (which is identical to the original) is made.

The now double stranded DNA is denatured so that each strand can separately attach to an oligonucleotide sequence anchored to the flow cell. One will be the reverse strand; the other, the forward. This process is called bridge amplification, and it happens for thousands of clusters all over the flow cell at once



Millions of oligos line the bottom of each flow cell lane.

3. Clonal amplification

Over and over again, DNA strands will bend and attach to oligos. Polymerases will synthesize a new strand to create a double stranded segment, and that will be denatured so that all of the DNA strands in one area are from a single source (clonal amplification).

Clonal amplification is important for quality control purposes. If a strand is found to have an odd sequence, then scientists can check the reverse strand to make sure that it has the complement of the same oddity. The forward and reverse strands act as checks to guard against artifacts. Because Illumina sequencing uses polymerases, base substitution errors have been observed, especially at the 3' end. Paired end reads combined with cluster generation can confirm an error took place. T

he reverse and forward strands should be complementary to each other, all reverse reads should match each other, and all forward reads should match each other. If a read is not similar enough to its counterparts (with which it should be a clone), an error may have occurred. A minimum threshold of 97% similarity has been used in some labs' analyses

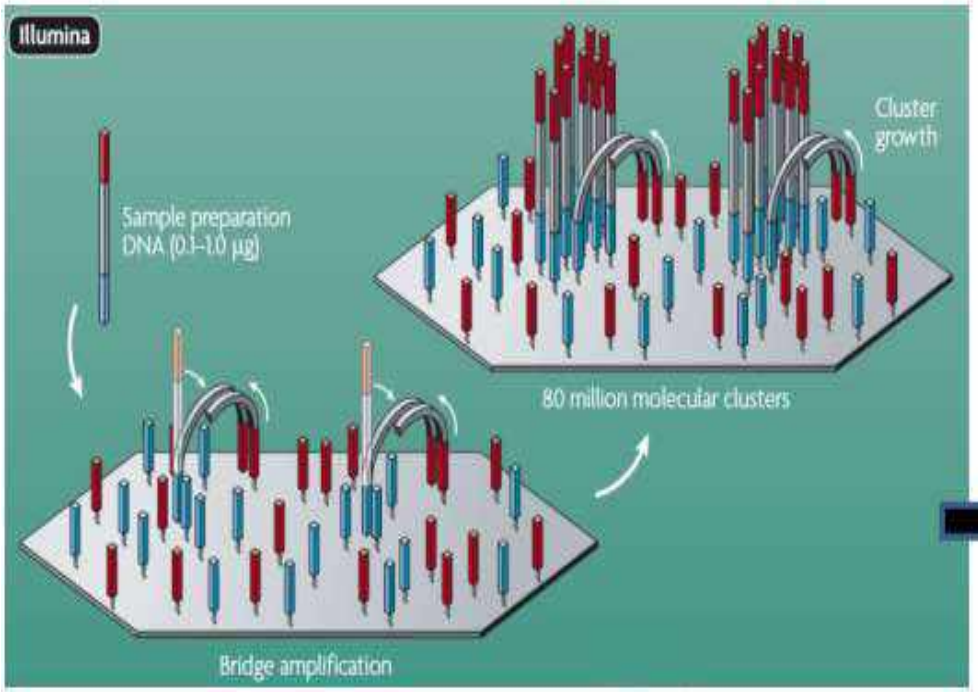
4. Sequence by synthesis

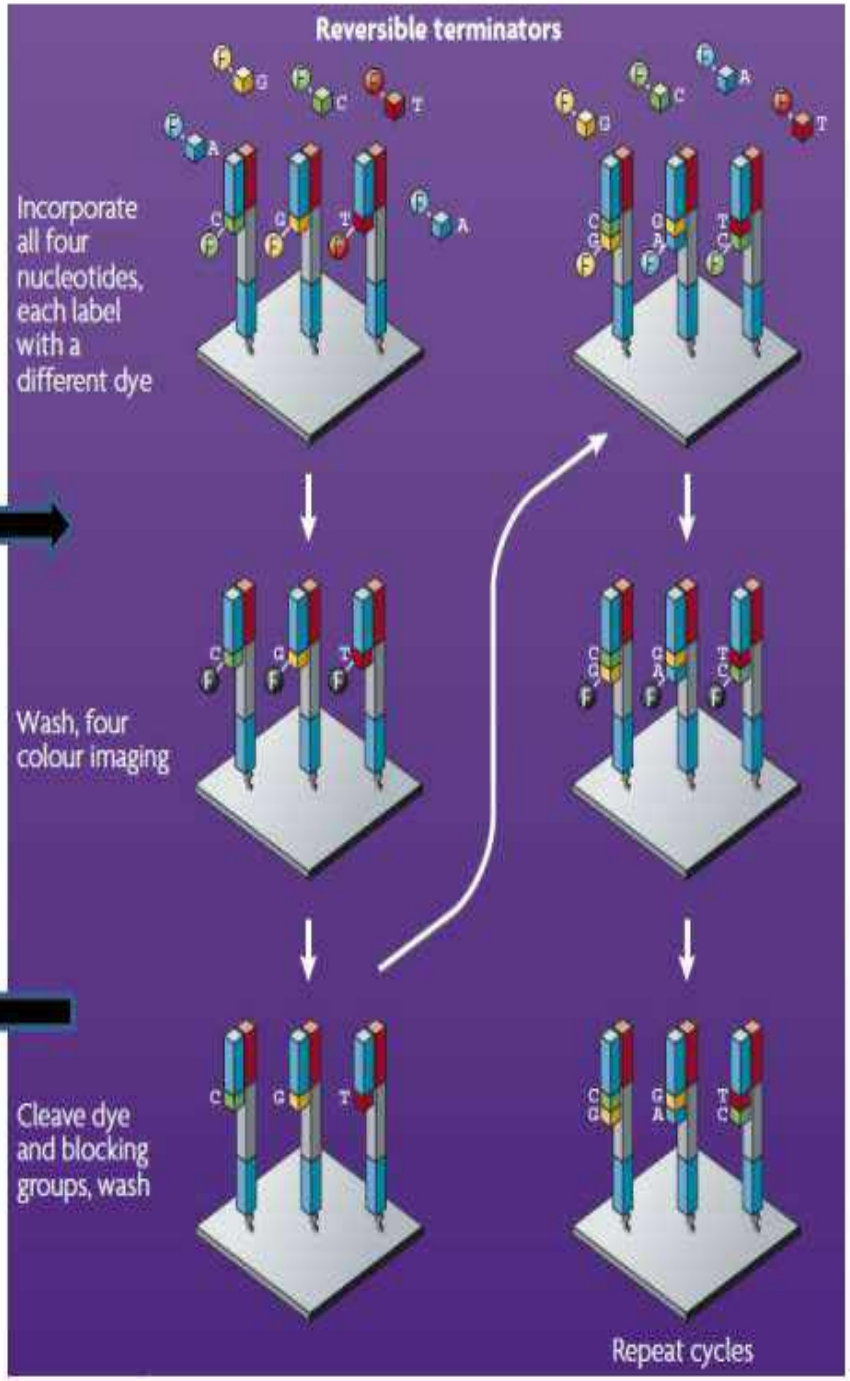
At the end of clonal amplification, all of the reverse strands are washed off the flow cell, leaving only forward strands. Primers attach to the forward strands and a polymerase adds fluorescently tagged nucleotides to the DNA strand. Only one base is added per round. A reversible terminator is on every nucleotide to prevent multiple additions in one round. Using the four-colour chemistry, each of the four bases has a unique emission, and after each round, the machine records which base was added.

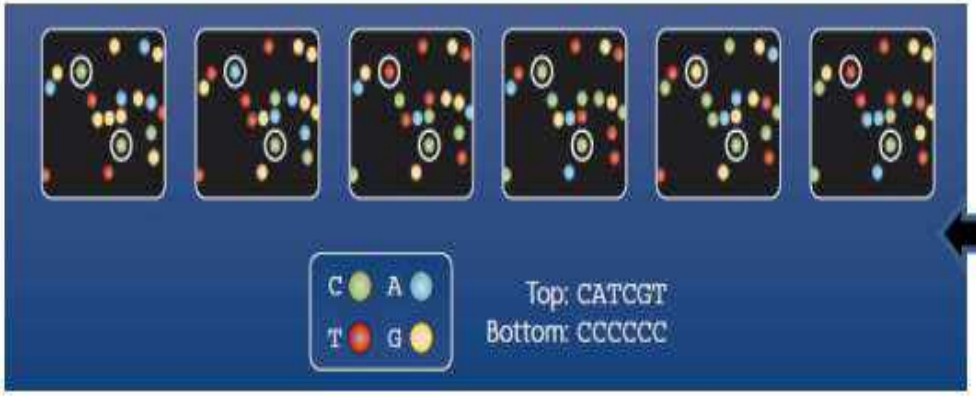
Starting with the launch of the NextSeq and later the MiniSeq, Illumina introduced a new two-colour sequencing chemistry. Nucleotides are distinguished by either one of two colours (red or green), no colour ("black") or binding both colours (appearing orange as a mixture between red and green).

Once the DNA strand has been read, the strand that was just added is washed away. Then, the index 1 primer attaches, polymerizes the index 1 sequence, and is washed away. The strand forms a bridge again, and the 3' end of the DNA strand attaches to an oligo on the flow cell. The index 2 primer attaches, polymerizes the sequence, and is washed away.

A polymerase sequences the complementary strand on top of the arched strand. They separate, and the 3' end of each strand is blocked. The forward strand is washed away, and the process of sequence by synthesis repeats for the reverse strand.







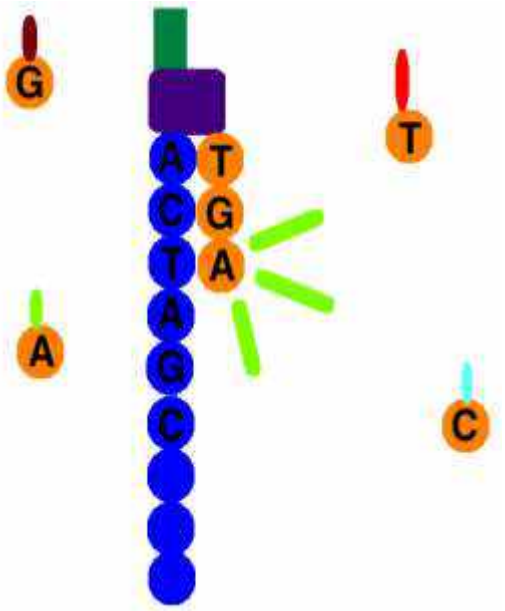
Metzker, Nature Reviews Genetics, 2010, 11, 31-46

Data analysis

The sequencing occurs for millions of clusters at once, and each cluster has ~1,000 identical copies of a DNA insert. The sequence data is analyzed by finding fragments with overlapping areas, called contigs, and lining them up. If a reference sequence is known, the contigs are then compared to it for variant identification.

This piecemeal process allows scientists to see the complete sequence even though an unfragmented sequence was never run; however, because Illumina read lengths are not very long (HiSeq sequencing can produce read lengths around 90 bp long, it can be a struggle to resolve short tandem repeat areas.

Also, if the sequence is de novo and so a reference doesn't exist, repeated areas can cause a lot of difficulty in sequence assembly. Additional difficulties include base substitutions (especially at the 3' end of reads) by inaccurate polymerases, chimeric sequences, and PCR-bias, all of which can contribute to generating an incorrect sequence.



Tagged nucleotides are added in order to the DNA strand. Each of the four nucleotides have an identifying label that can be excited to emit a characteristic wavelength. A computer records all of the emissions, and from this data, base calls are made.

3rd Generation Sequencing

3rd generation sequencing (2012 – present)

- PacBio or SMRT sequencing
- Oxford Nanopore

Single-molecule real-time(SMRT) / PacBio sequencing

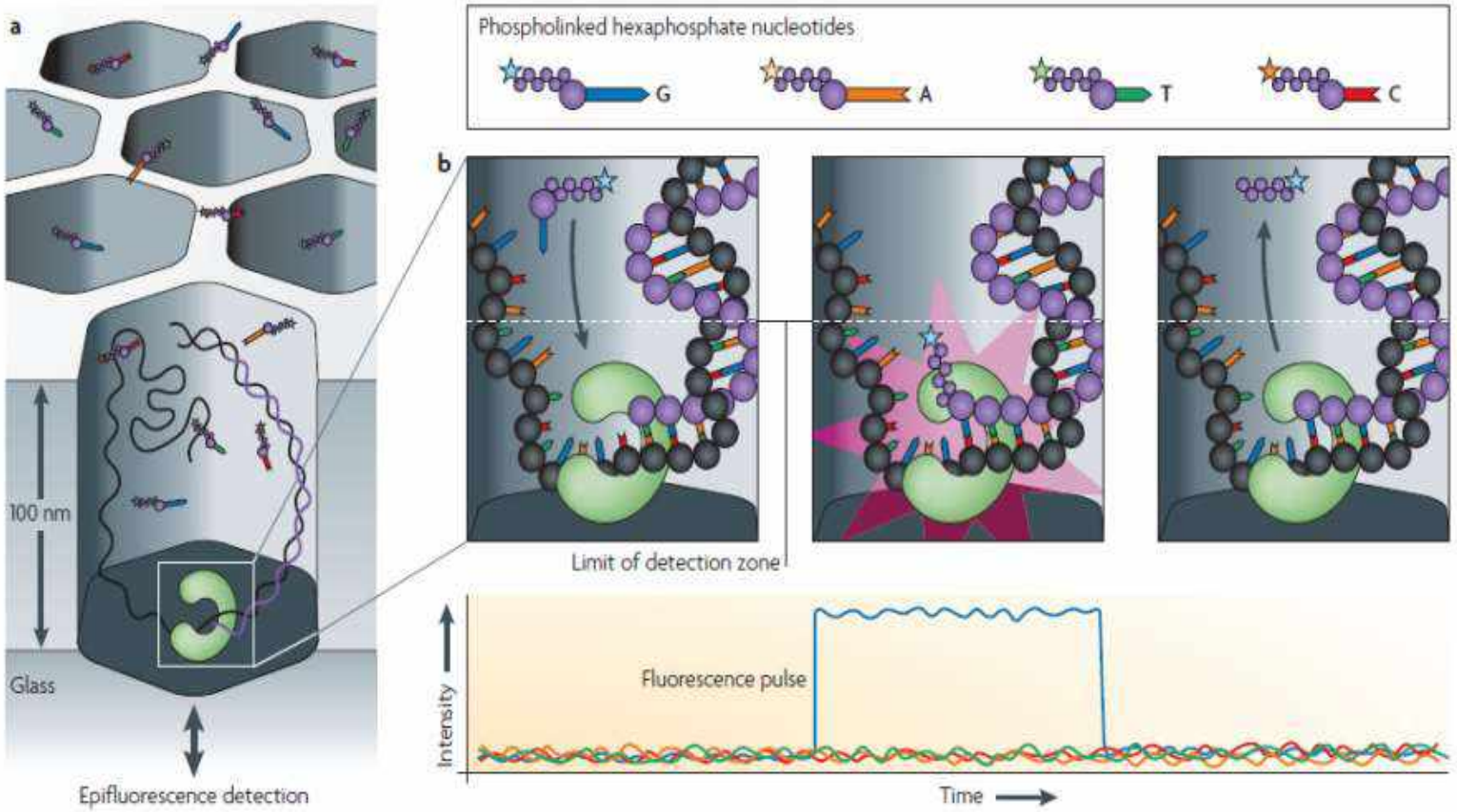
Single-molecule real-time sequencing (**SMRT**) is a parallelized single molecule DNA sequencing method. Single-molecule real-time sequencing utilizes a zero-mode waveguide (ZMW).

A single DNA polymerase enzyme is affixed at the bottom of a ZMW with a single molecule of DNA as a template. The ZMW is a structure that creates an illuminated observation volume that is small enough to observe only a single nucleotide of DNA being incorporated by DNA polymerase.

Each of the four DNA bases is attached to one of four different fluorescent dyes. When a nucleotide is incorporated by the DNA polymerase, the fluorescent tag is cleaved off and diffuses out of the observation area of the ZMW where its fluorescence is no longer observable.

A detector detects the fluorescent signal of the nucleotide incorporation, and the base call is made according to the corresponding fluorescence of the dye.

- 2011, Pacific Biosciences
- Single Molecule Real Time Sequencing (SMRT)



The DNA sequencing is done on a chip that contains many ZMWs. Inside each ZMW, a single active DNA polymerase with a single molecule of single stranded DNA template is immobilized to the bottom through which light can penetrate and create a visualization chamber that allows monitoring of the activity of the DNA polymerase at a single molecule level. The signal from a phospho-linked nucleotide incorporated by the DNA polymerase is detected as the DNA synthesis proceeds which results in the DNA sequencing in real time.

1. ZERO-MODE WAVEGUIDE

The zero-mode waveguide (ZMW) is a nano photonic confinement structure that consists of a circular hole in an aluminum cladding film deposited on a clear silica substrate.

The ZMW holes are ~70 nm in diameter and ~100 nm in depth. Due to the behavior of light when it travels through a small aperture, the optical field decays exponentially inside the chamber.

The observation volume within an illuminated ZMW is ~20 zeptoliters (20 X 10^{-21} liters). Within this volume, the activity of DNA polymerase incorporating a single nucleotide can be readily detected

2. PHOSPHOLINKED NUCLEOTIDE

For each of the nucleotide bases, there is a corresponding fluorescent dye molecule that enables the detector to identify the base being incorporated by the DNA polymerase as it performs the DNA synthesis.

The fluorescent dye molecule is attached to the phosphate chain of the nucleotide. When the nucleotide is incorporated by the DNA polymerase, the fluorescent dye is cleaved off with the phosphate chain as a part of a natural DNA synthesis process during which a phosphodiester bond is created to elongate the DNA chain. The cleaved fluorescent dye molecule then diffuses out of the detection volume so that the fluorescent signal is no longer detected

History

Pacific Biosciences [PacBio] commercialized SMRT sequencing in 2011,after releasing a beta version of its RS instrument in late 2010.

On 19 Sep 2018, Pacific Biosciences [PacBio] released the Sequel 6.0 chemistry, synchronizing the chemistry version with the software version. Performance is contrasted for large-insert libraries with high molecular weight DNA versus shorter-insert libraries below ~15,000 bases in length. For larger templates average read lengths are up to 30,000 bases. For shorter-insert libraries, average read length are up to 100,000 bases while reading the same molecule in a circle. The latter shorter-insert libraries then yield up to 50 billion bases from a single SMRT Cell

8M Chip

In April 2019 the company released a new SMRT Cell with eight million ZMW's, increasing the expected throughput per SMRT Cell by a factor of eight. Early access customers in March 2019 reported throughput over 58 customer run cells of 250 GB of raw yield per cell with templates about 15 kb in length, and 67.4 GB yield per cell with templates in higher weight molecules. System performance is now reported in either high-molecular-weight continuous long reads or in pre-corrected HiFi (aka CCS) reads. For high-molecular-weight reads roughly half of all reads are longer than 50 kb in length

Application

1. Single-molecule real-time sequencing may be applicable for a broad range of genomics research.
2. For *de novo* genome sequencing, read lengths from the single-molecule real-time sequencing are comparable to or greater than that from the Sanger sequencing method based on dideoxynucleotide chain termination.
3. The same DNA molecule can be resequenced independently by creating the circular DNA template and utilizing a strand displacing enzyme that separates the newly synthesized DNA strand from the template. In August 2012, scientists from the Broad Institute published an evaluation of SMRT sequencing for SNP calling.
4. The dynamics of polymerase can indicate whether a base is methylated. Scientists demonstrated the use of single-molecule real-time sequencing for detecting methylation and other base modifications. In 2012 a team of scientists used SMRT sequencing to generate the full methylomes of six bacteria. In November 2012, scientists published a report on genome-wide methylation of an outbreak strain of E. coli
5. Long reads make it possible to sequence full gene isoforms, including the 5' and 3' ends. This type of sequencing is useful to capture isoforms and splice variants.
6. SMRT sequencing has several applications in reproductive medical genetics research when investigating families with suspected parental gonadal mosaicism. Long reads enable haplotype phasing in patients to investigate parent-of-origin of mutations. Deep sequencing enables determination of allele frequencies in sperm cells, of relevance for estimation of recurrence risk for future affected offspring.

OXFORD NANOPORE

Oxford Nanopore Technologies has begun shipping early versions of its nanopore sequencing MinION sequencer to selected labs.

The device is four inches long and gets power from a USB port. MinION decodes DNA directly as the molecule is drawn at the rate of 450 bases/second through a nanopore suspended in a membrane.

Changes in electric current indicate which base is present. It is 60 to 85 percent accurate, compared with 99.9 percent in conventional machines. Even inaccurate results may prove useful because it produces long read lengths.

GridION is a slightly larger sequencer that processes up to five MinION flow cells at once.

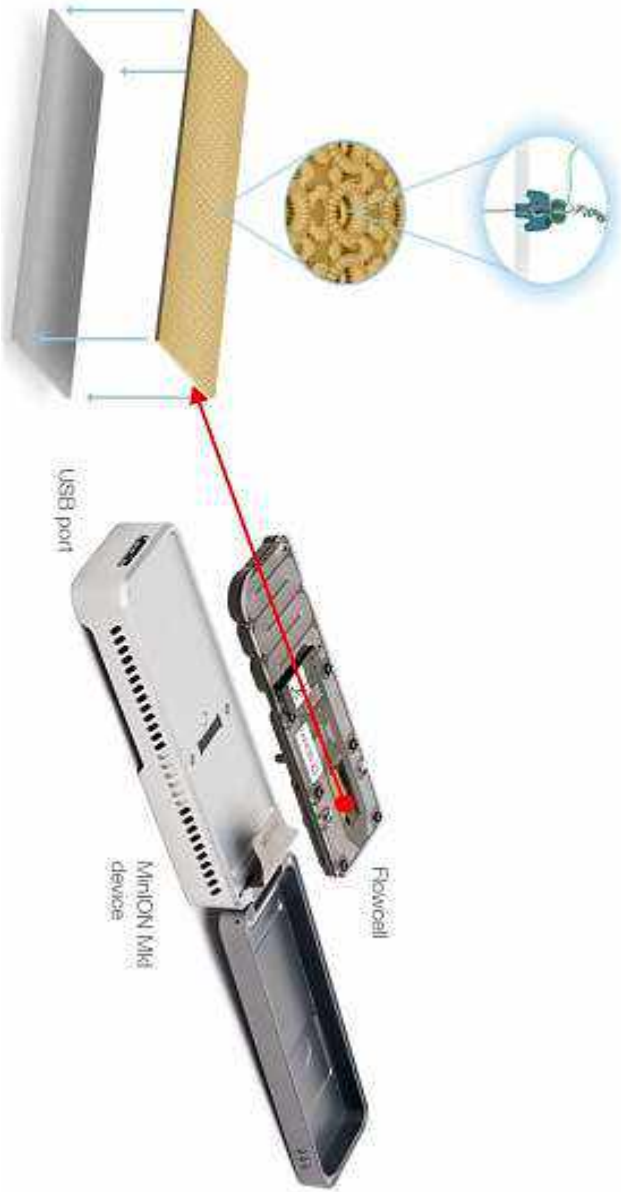
PromethION is another (unreleased) product that will use as many as 100,000 pores in parallel, more suitable for high volume sequencing

In July 2015, a group published on nanopore sequencing of an influenza genome, noting “A complete influenza virus genome was obtained that shared greater than 99% identity with sequence data obtained from the Illumina Miseq and traditional Sanger-sequencing.

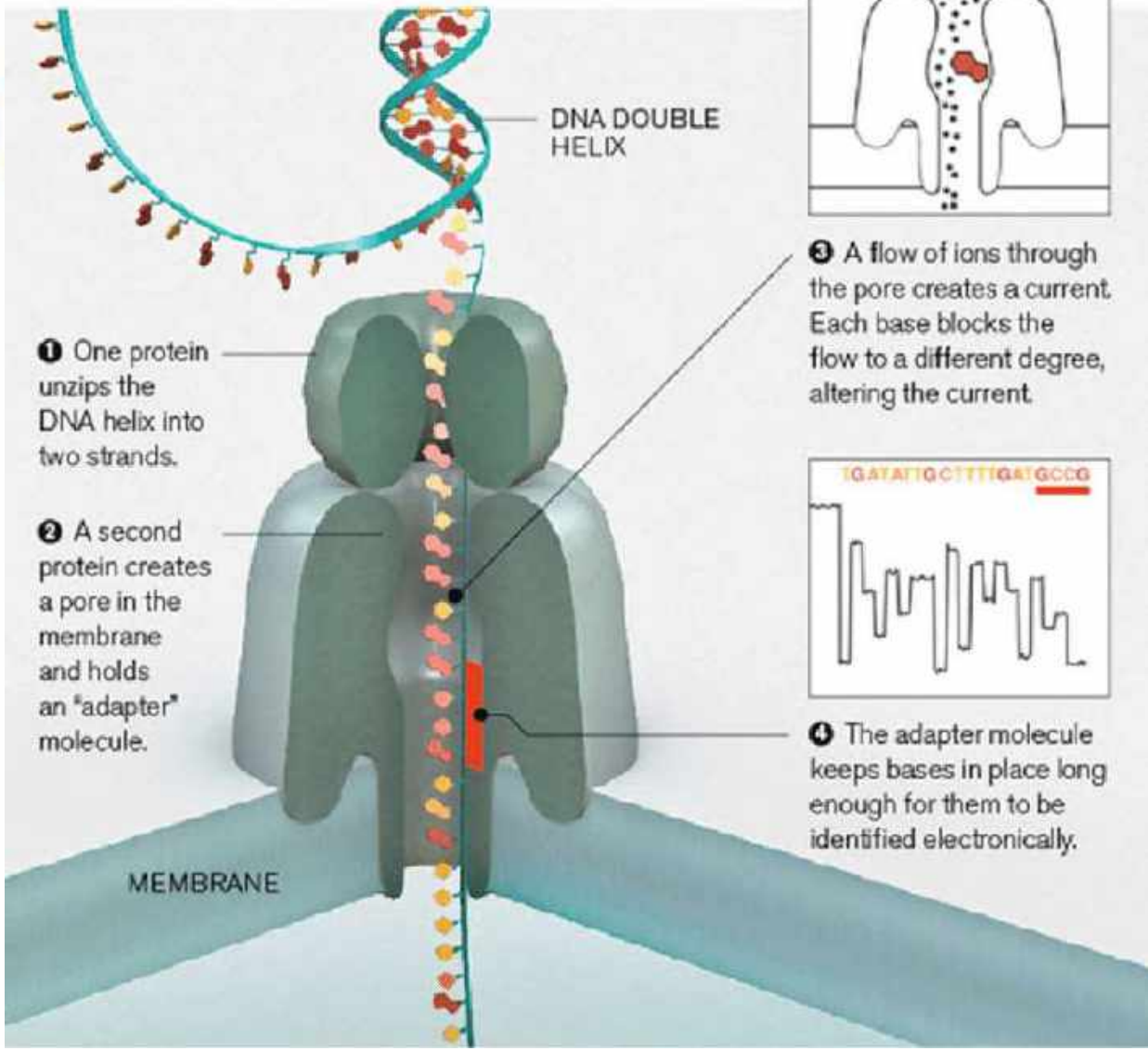
The laboratory infrastructure and computing resources used to perform this experiment on the MinION nanopore sequencer would be available in most molecular laboratories around the world.

Using this system, the concept of portability, and thus sequencing influenza viruses in the clinic or field is now tenable.“ In a paper and accompanying editorial published in October 2015, a group of MinION users wrote, “At the time of this writing, around a dozen reports have emerged recounting utility of the MinION for de novo sequencing of viral, bacterial, and eukaryotic genomes.”.

Oxford NANOPORE Technologies



DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



Other Products of Oxford Nanopore

- MinION: this portable protein nanopore sequencing USB device has been commercially available since May 2015
- GridION X5: this desktop device has been commercially available since March 2017. The device processes up to five MinION Flow Cells and enables generation of up to 100 Gb of data per run.
- PromethION: this desktop, high throughput device will be available through an access programme that opened for registration in July 2015. The device contains channels for 144,000 nanopores (in comparison to MinION's 512).
- VolTRAX: this device, currently in development, is designed for automated sample preparation so that users do not need a laboratory or lab skills to run the device. Registration for the early access programme was opened in October 2016.
- Metrichor: this spinout company from Oxford Nanopore was set up to provide end to end solutions for biological analyses, using nanopore sensing technologies.
- SmidgION: a mobile phone sequencer announced in May 2016, currently in development

PICODROP OR NANOQUANTA ANALYZER

The Picodrop Microliter UV/Vis Spectrophotometer is the only system that allows you to use the best format for your assay without compromise. Use our patented disposable UVpette Tips or the unique drop-and-read TrayCell for tip-free applications

Use the In-tip detection method for a range between 5-1200ng/ul of dsDNA or for an extended range switch the holder to use the TrayCell and benefit from the extended concentration range up to 8500ng/ul dsDNA.

Range of interchangeable cuvette accessories including compatibility with TrayCell, standard cuvettes and dip probes

No cleaning required when using tips and just a simple wipe when using the Cell

Sample volumes down to 1uL

No dilutions necessary

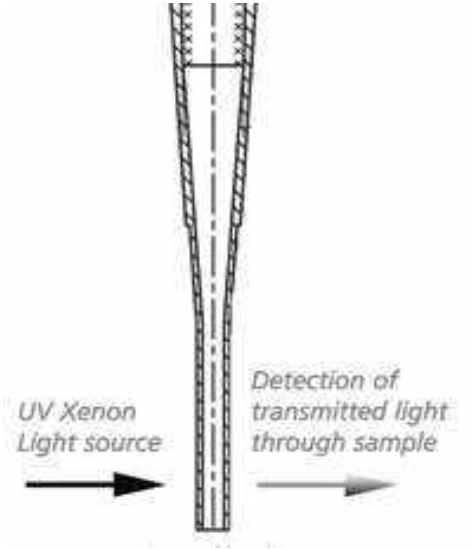
Kinetic measurements in both tip or cuvette

100% Sample recovery

Zero cross-contamination

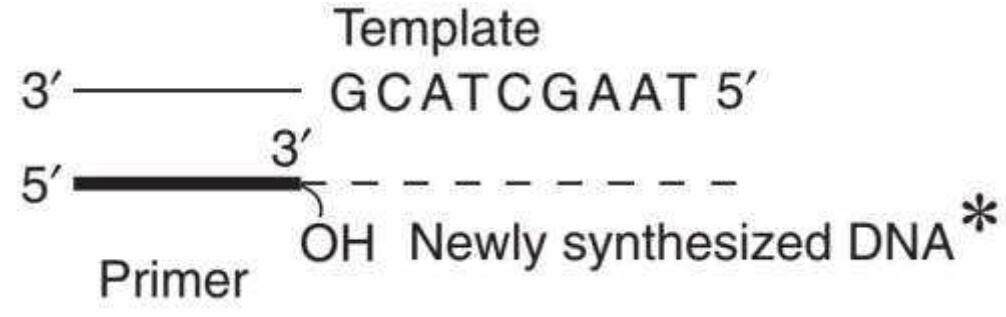
3 Second detection time

Scanning wavelength range from 220nm to 950nm



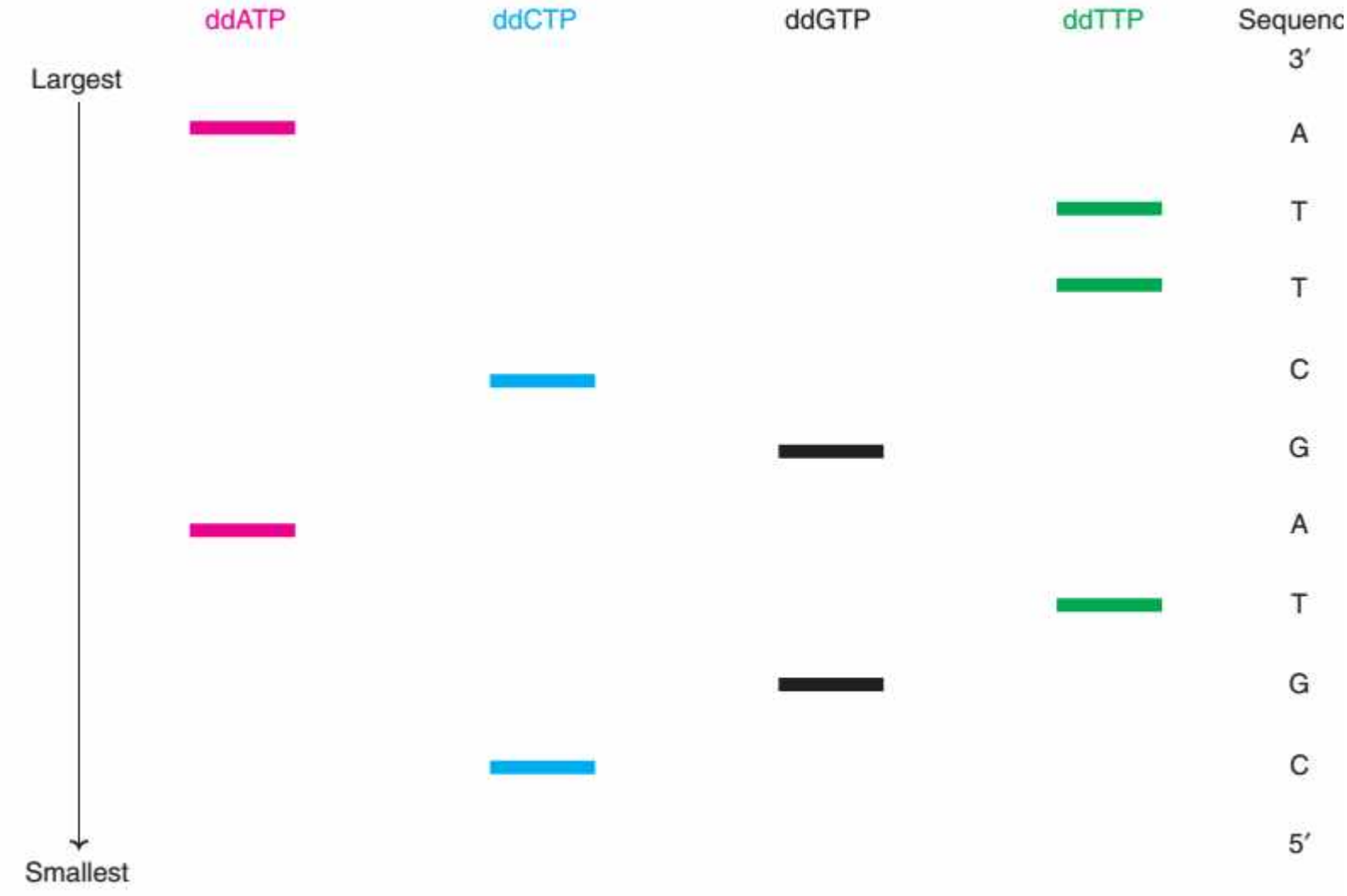
AUTOMATED DNA SEQUENCING

- DNA sequencing in the recent years is carried out by an automated DNA sequencer. In this technique, flourescent tags are attached to chainterminating nucleotides (dideoxynucleotides).
- This tag gets incorporated into the DNA molecules, while terminating new strand synthesis. Four different fluorescent dyes are used to identify chain-terminating reactions in a sequencing gel.
- The DNA bands are separated by electrophoresis and detected by their fluorescence. Recently, four dyes that exhibit strong absorption in laser are in use for automated sequencing.



Reaction tube with dideoxynucleotide	Primer with nucleotide extended	Primer with sequence of nucleotides extended
ddATP	Primer + 4	Primer–CGTddA
	Primer + 9	Primer–CGTAGCTTddA
ddCTP	Primer + 1	Primer–ddC
	Primer + 6	Primer–CGTAGddC
ddGTP	Primer + 2	Primer–C**ddG**
	Primer + 5	Primer–CGTA**ddG**
ddTTP	Primer + 3	Primer–CGddT
	Primer + 7	Primer–CGTAGCddT
	Primer + 8	Primer–CGTAGCTddT

Synthesis of new DNA fragments in the presence of dideoxynucleotides (*the size of the new DNA is variable, depending on the chain termination).



Sequence of the newly synthesized DNA fragment (complementary to original strand).

ADVANTAGES OF AUTOMATED SEQUENCING :

- It is a rapid and accurate technique.
- Automated DNA sequencer can accurately sequence up to 100,000 nucleotides per day.
- The cost works out to be not more than $0.2 per nucleotide.
- Automated DNA sequencing has been successfully used in the human genome project

METHOD	READ LENGTH	ADVANTAGES	DISADVANTAGES
Single-molecule real-time sequencing (Pacific Biosciences)	30,000 bp (N50); maximum read length >100,000 bases	Fast. Detects 4mC, 5mC, 6mA.	Moderate throughput. Equipment can be very expensive.
Ion semiconductor (Ion Torrent sequencing)	up to 600 bp	Less expensive equipment. Fast.	Homopolymer errors.
Pyrosequencing (454)	700 bp	Long read size. Fast.	Runs are expensive. Homopolymer errors.
Sequencing by synthesis (Illumina)	MiniSeq, NextSeq: 75–300 bp; MiSeq: 50–600 bp; HiSeq 2500: 50–500 bp; HiSeq 3/4000: 50–300 bp; HiSeq X: 300 bp	Potential for high sequence yield, depending upon sequencer model and desired application.	Equipment can be very expensive. Requires high concentrations of DNA.
Combinatorial probe anchor synthesis (cPAS-BGI/MGI)	BGISEQ-50: 35-50bp; MGISEQ 200: 50-200bp; BGISEQ-500, MGISEQ-2000: 50-300bp		
Sequencing by ligation (SOLiD sequencing)	50+35 or 50+50 bp	Low cost per base.	Slower than other methods. Has issues sequencing palindromic sequences.
Nanopore Sequencing	Read length (up to 2,272,580 bp reported	Longest individual reads. Accessible user community. Portable (Palm sized).	Lower throughput than other machines, Single read accuracy in 90s.
GenapSys Sequencing	Around 150 bp single-end	Low-cost of instrument ($10,000)	
Chain termination (Sanger sequencing)	400 to 900 bp	Useful for many applications.	More expensive and impractical for larger sequencing projects. This method also requires the time-consuming step of plasmid cloning or PCR.

References

1. Voet D, Voet J. .G “Biochemistry”, 4th Edition, John Wiley & Sons, Inc.,2010. ISBN: 978-0-470-57095-1
2. David L. Nelson; Michael M. Cox “Lehninger Principles of Biochemistry” Seventh Edition, Macmillan 2017, ISBN:9781464187957
3. Wilson, K., Walker, J. (eds.); Cambridge University Press, Cambridge, 2000, 784 pp., ISBN 0-521-65873-X (paperback)
4. Primrose S B, Twyman R. Principles of Gene Manipulation and Genomics, 7th Edition. John Wiley & Sons, Inc, 2006. ISBN: 978-1-405-13544-3

All the original contributors of the concept and findings published are gratefully acknowledged while preparing the e-content for the students of Biochemistry and allied sciences