Bioinformatics Lecture 2: DNA sequencing

Course name: Bioinformatics and Computer Application
Course Code: MSCCONBC401

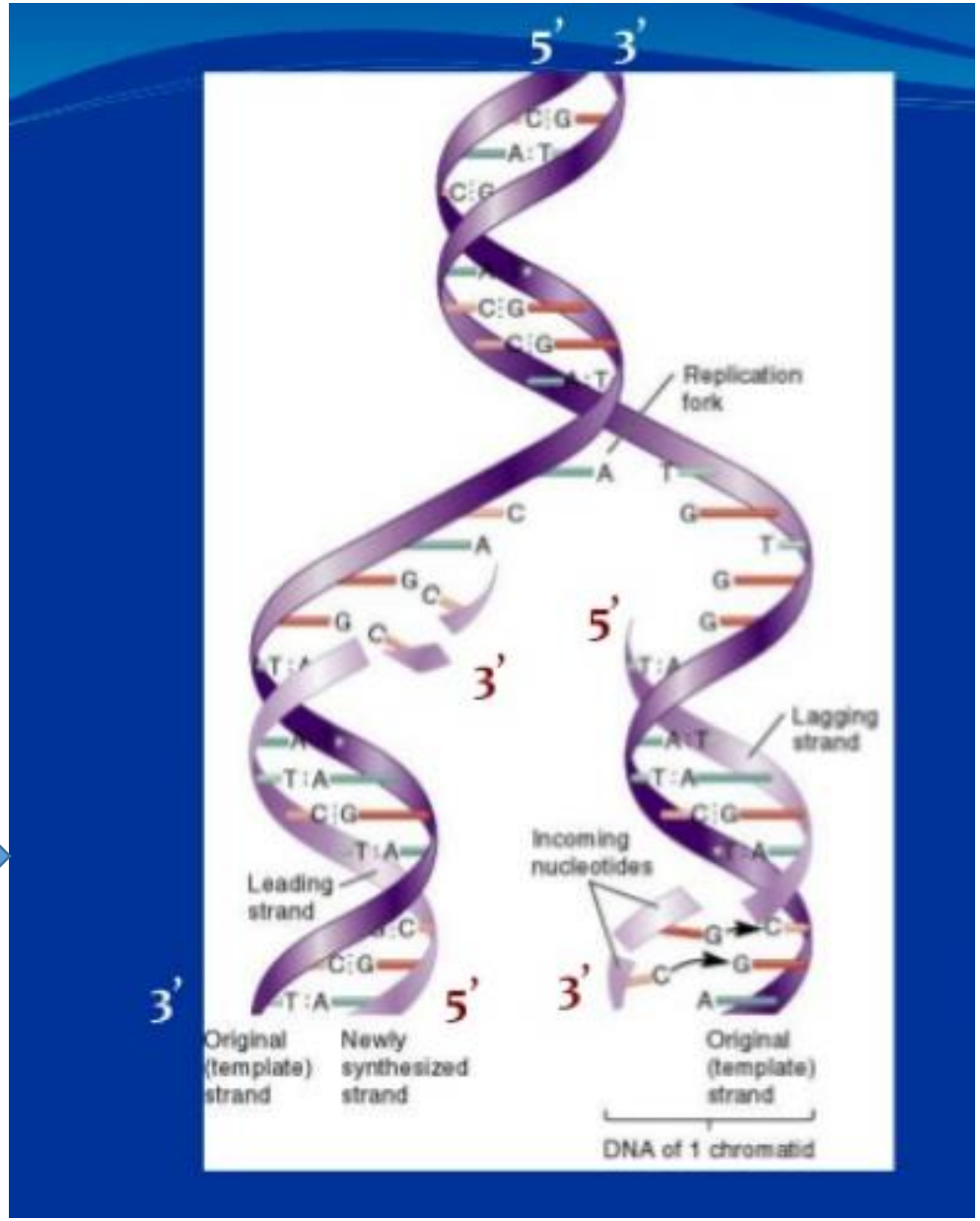
Dr. Moitreyee Chakrabarty
Assistant Professor
PG Department of Conservation Biology
Durgapur Government College

Sequencing of DNA is a method of DNA replication done in vitro to know the sequence of bases

What is DNA replication?

DNA Replication

- A reaction in which daughter DNAs are synthesized using the parental DNAs as the template.
- Transferring the genetic information to the descendant generation with a high fidelity.



Bases are added always in 5’ to 3’ in the daughter strand

- 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 sequence were obtained by academic researchers, using laboratories methods based on 2- dimensional chromatography in the early 1970s.
- 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 coworkers in 1975 owing to its relative easy and reliability.
- In 1975 the first complete DNA genome to be sequenced is that of bacteriophage ϕX174.

DIFFERENT METHODS FOR DNA SEQUENCING

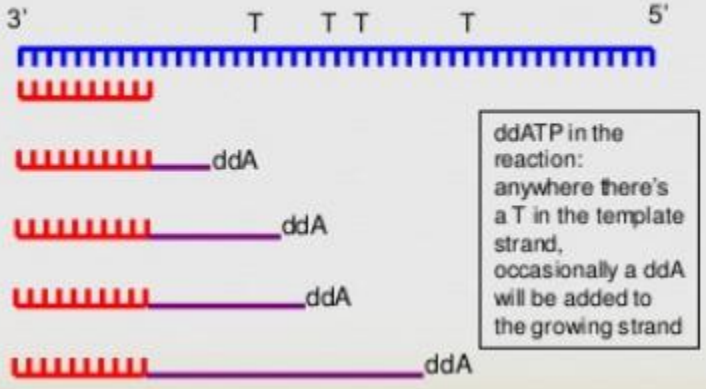
- **Basic Methods:**
  - Maxam-Gilbert sequencing
  - Chain-termination methods
- **Next-generation methods:**
  - Massively parallel signature sequencing (MPSS)
  - Polony sequencing
  - 454 pyrosequencing
  - Illumina (Solexa) sequencing
  - SOLiD sequencing
  - Ion Torrent semiconductor sequencing
  - DNA nanoball sequencing
  - Single molecule real time (SMRT) sequencing

Sanger Method or Chain Termination Method (Frederick Sanger, 1975)

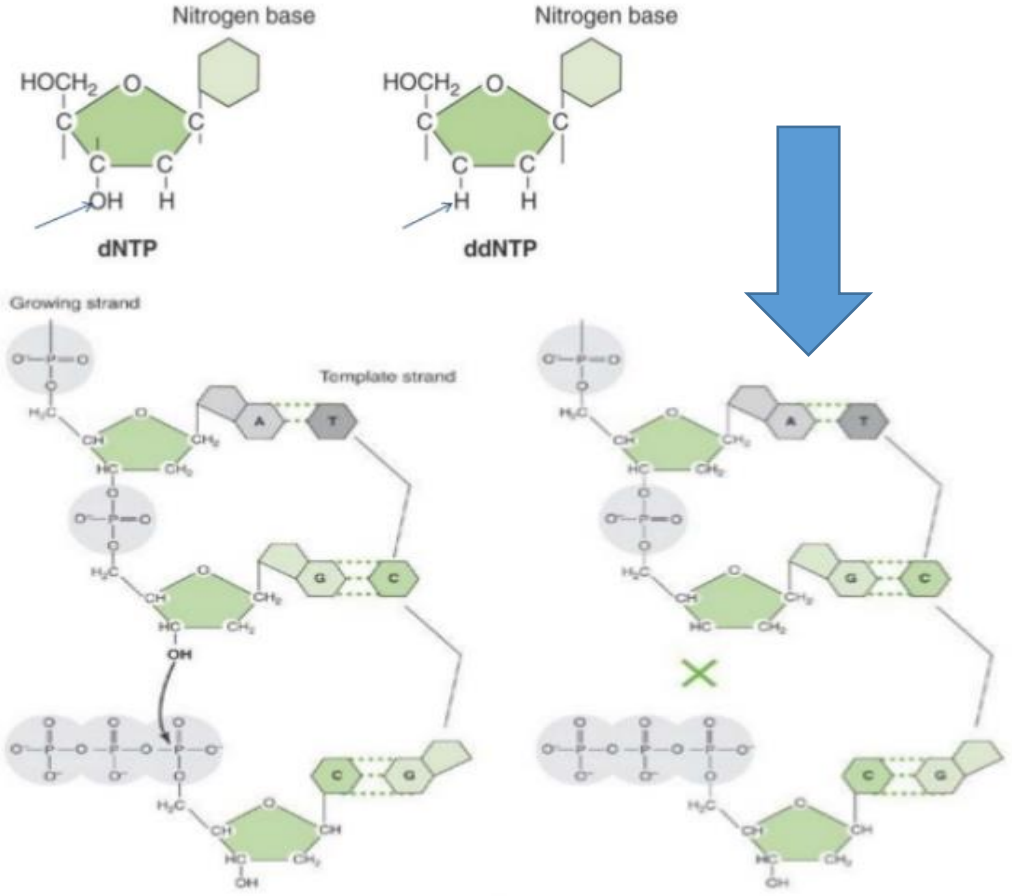
- It is method to find out the nucleotides Sequence of unknown DNA strand.
- More recently, Sanger sequencing has been upgraded as "Next-Generation" sequencing methods, especially for large scale genome analyses and for obtaining especially long DNA sequence reads (>500 nucleotides).

BASIC PRINCIPLE

- This method generally is an In-Vitro synthesis of DNA strand and by using terminators (di-deoxynucleotide) the growing strand terminates at specific site.
- Upon termination the strands are overlap to got original sequence of unknown DNA Strand.



- It is PCR based method
- A modified DNA replication reaction
- Growing chains are terminated by dideoxynucleotides



The 3'-OH group necessary for formation of the phosphodiester bond is missing in ddNTPs.

Requirements

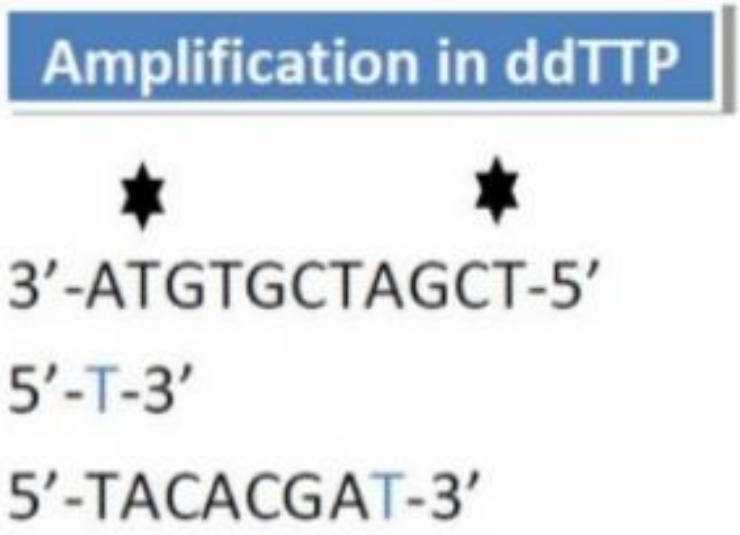
- Single Stranded template
- Primer
- DNA polymerase
- Di-Deoxynucleotide
- ✓ The 3′-OH group necessary for formation of the phosphodiester bond is missing in ddNTPs)
- ✓ Every nucleotide have its specific ddNTP form i.e., ddATP, ddGTP etc

- Steps:
  1. Denaturation
  2. Primer attachment and extension of bases
  3. Termination
  4. Gel electrophoresis

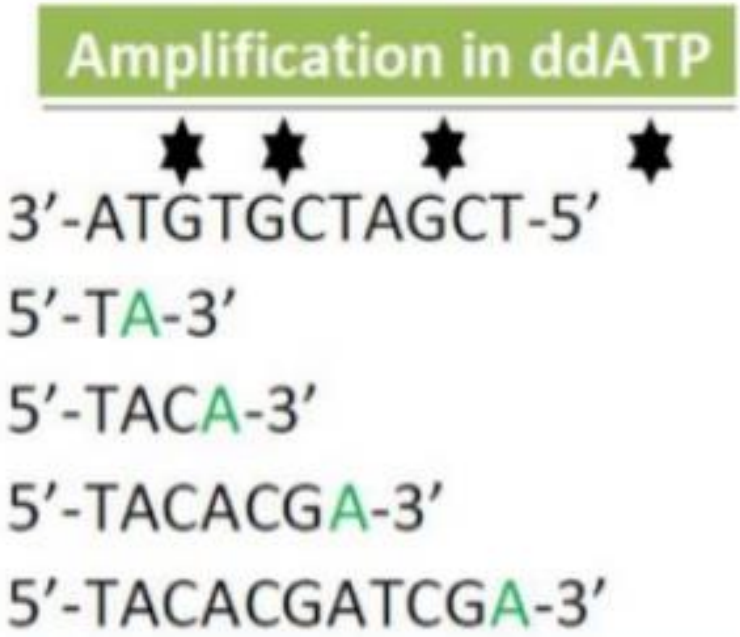
Detail process

- The DNA template is treated with heat so that it becomes single stranded
- A short, single-stranded primer which is radioactively labelled is added to the end of the DNA template
- Add template DNA and primer in 4 Tubes.
- Now add ddNTPs In tubes in the way that single tube contain one type of ddNTP.
- Extension is start and band formed of various sizes.
- The fragments of DNA are separated by electrophoresis
- Overlap these sequences to find out sequence of Target DNA.

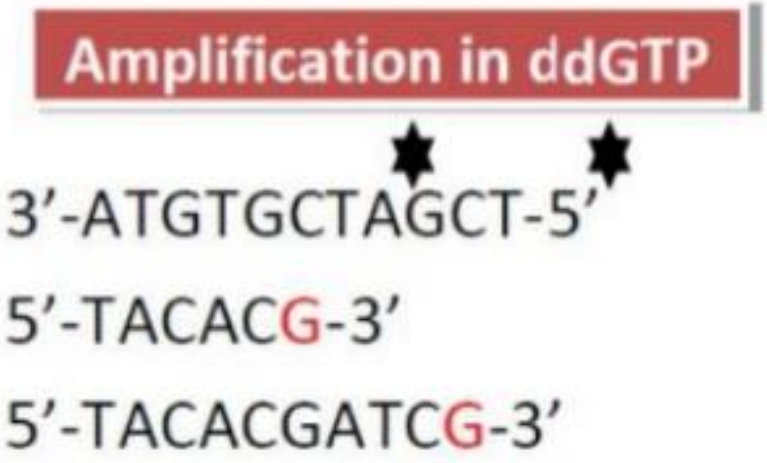
Tube 1



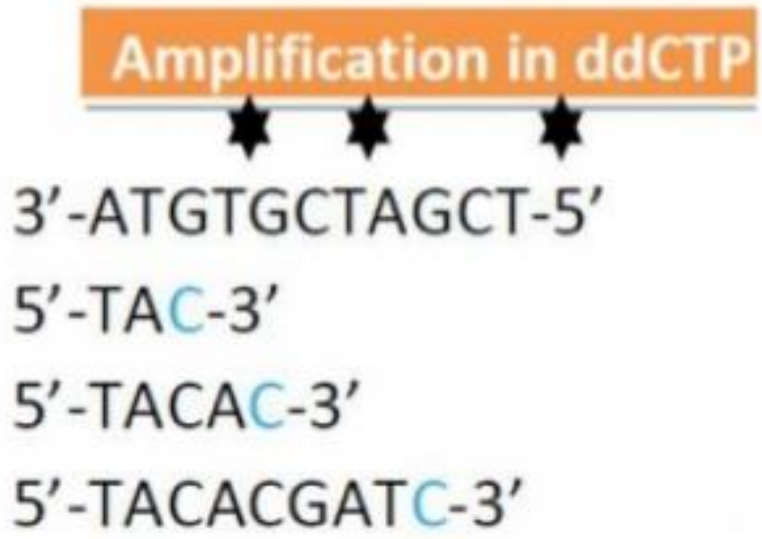
Tube 3

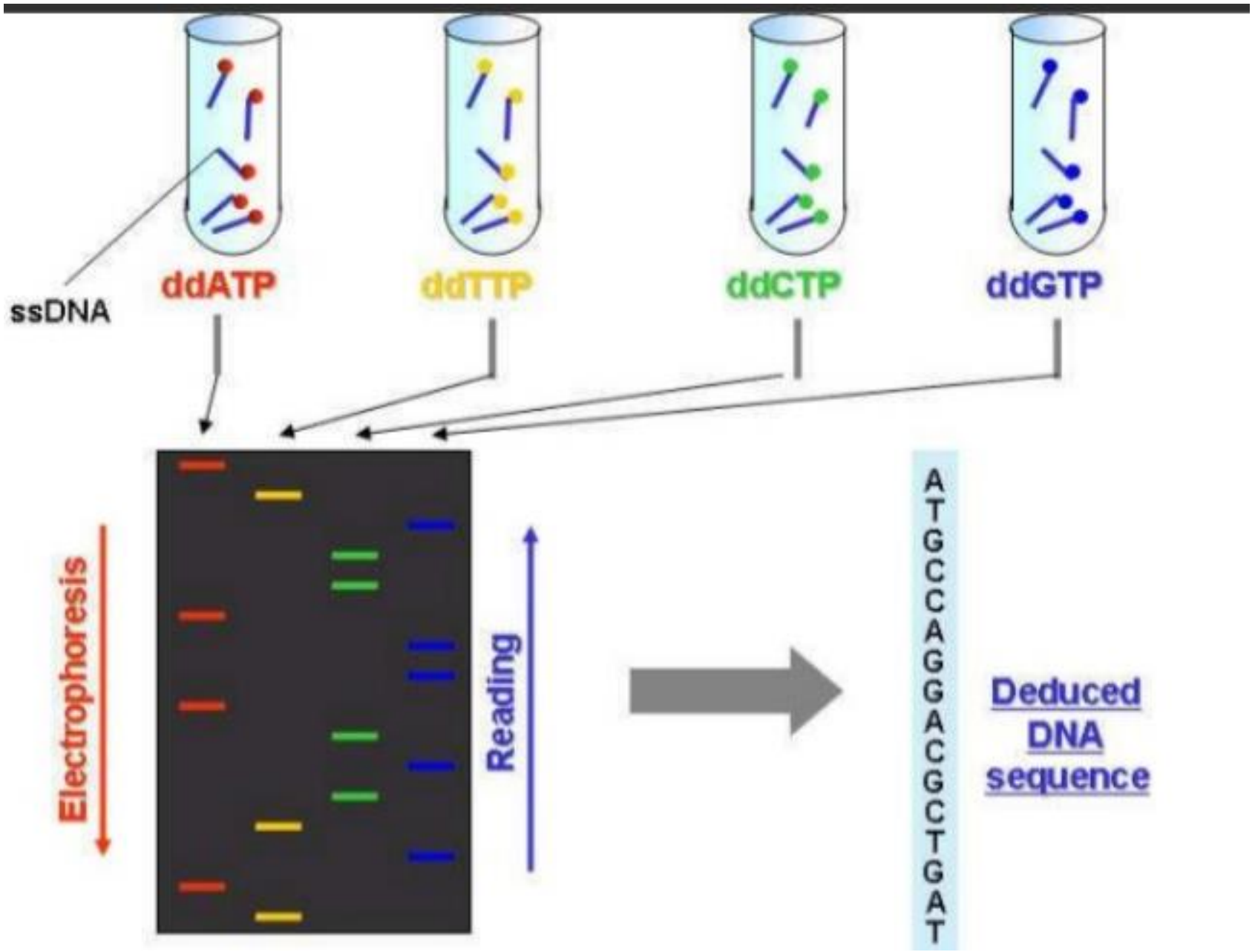


Tube 2



Tube 4





Hypothetical example of DNA sequencing:

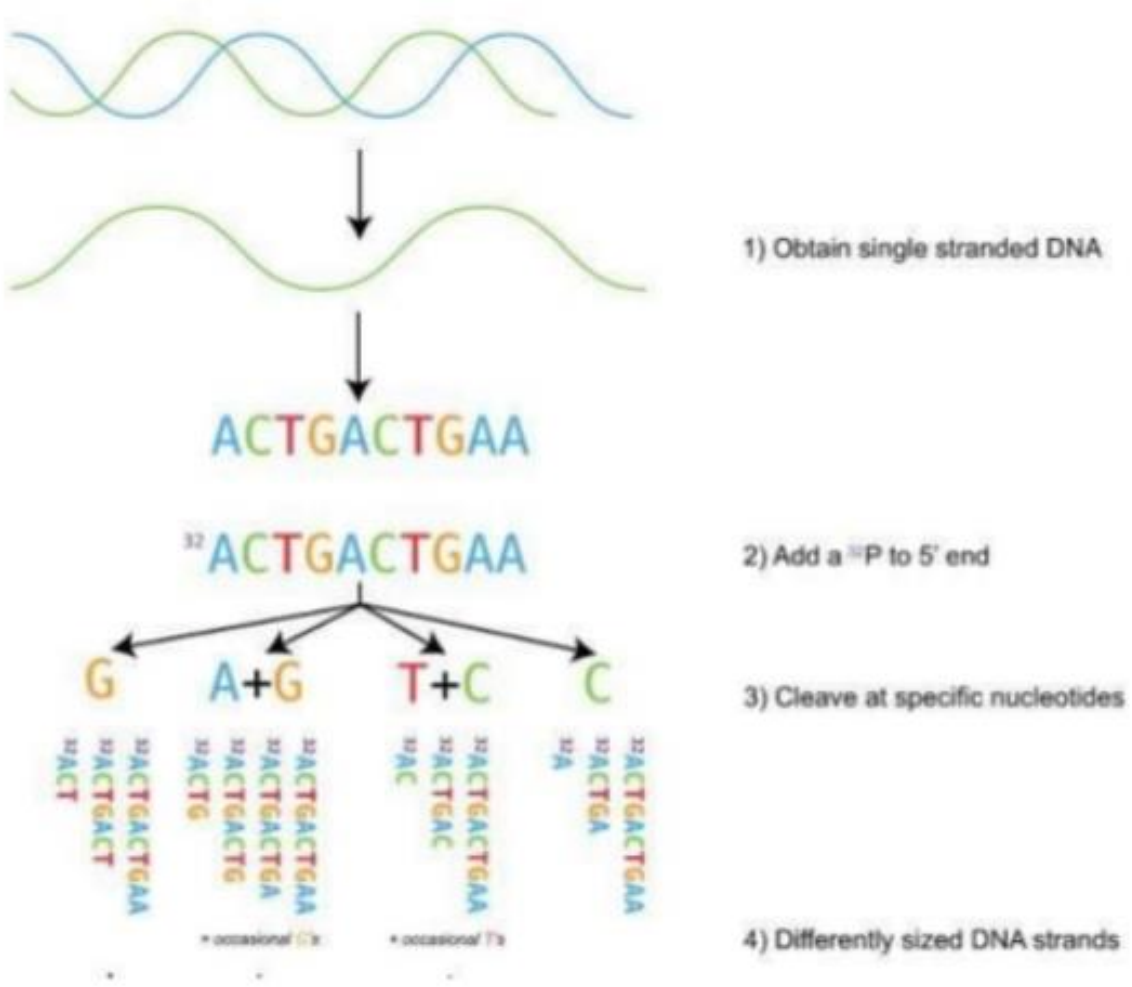
- The template used is the 3’ to 5’ strand or the antisense strand of the DNA.
- The fragments formed in Gel electrophorese are formed from 5’ to 3’ end and has to be read from behind to get the actual 5’ to 3’ DNA sequence as smaller fragments settle at the bottom.

Maxam-Gilbert Method or Chemical termination method

- Maxam–Gilbert sequencing is a method of DNA sequencing developed by Allan Maxam and Walter Gilbert in 1976–1977.
- Maxam–Gilbert sequencing was the first widely adopted method for DNA sequencing, and, along with the Sanger dideoxy method.
- method based on chemical modification of DNA and subsequent cleavage at specific nitrogenous bases.

PRINCIPLE

- purification of the DNA fragment that to be sequenced and labeled with radioactive material.
- Chemical treatment generates breaks at a specific nitrogenous bases and thus a series of labelled fragments is generated. The fragments in the four reactions are arranged side by side in gel electrophoresis for size separation.
- The fragments visualize in X-ray for autoradiography.
- 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.



Procedure

- Maxam–Gilbert sequencing requires radioactive labeling at one 5' end of the DNA fragment to be sequenced (gamma-32P).
- Chemical treatment 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). For example,

1. the purines (A+G) by using formic acid,
2. the guanines (and to some extent the adenines) by dimethyl sulfate,
3. the pyrimidines (C+T) by using hydrazine.
4. NaCl add to hydrazine for Cytosine.

- Add each chemical in separate tube.
- Thus a series of labeled fragments is generated.
- The fragments in the four reactions are electrophoresed side by side for size separation.
- To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each showing the location of identical radiolabeled DNA molecules.

Chemical Modification and Cleavage

➢ Base Modification using Dimethyl sulphate

– Purine

- Adenine
- Guanine

– Only DMS------- G

– DMS+ Formic acid-------G+A

➢ Cleavage of Sugar Phosphate backbone using Piperidine

➢ Base modification using Hydrazine

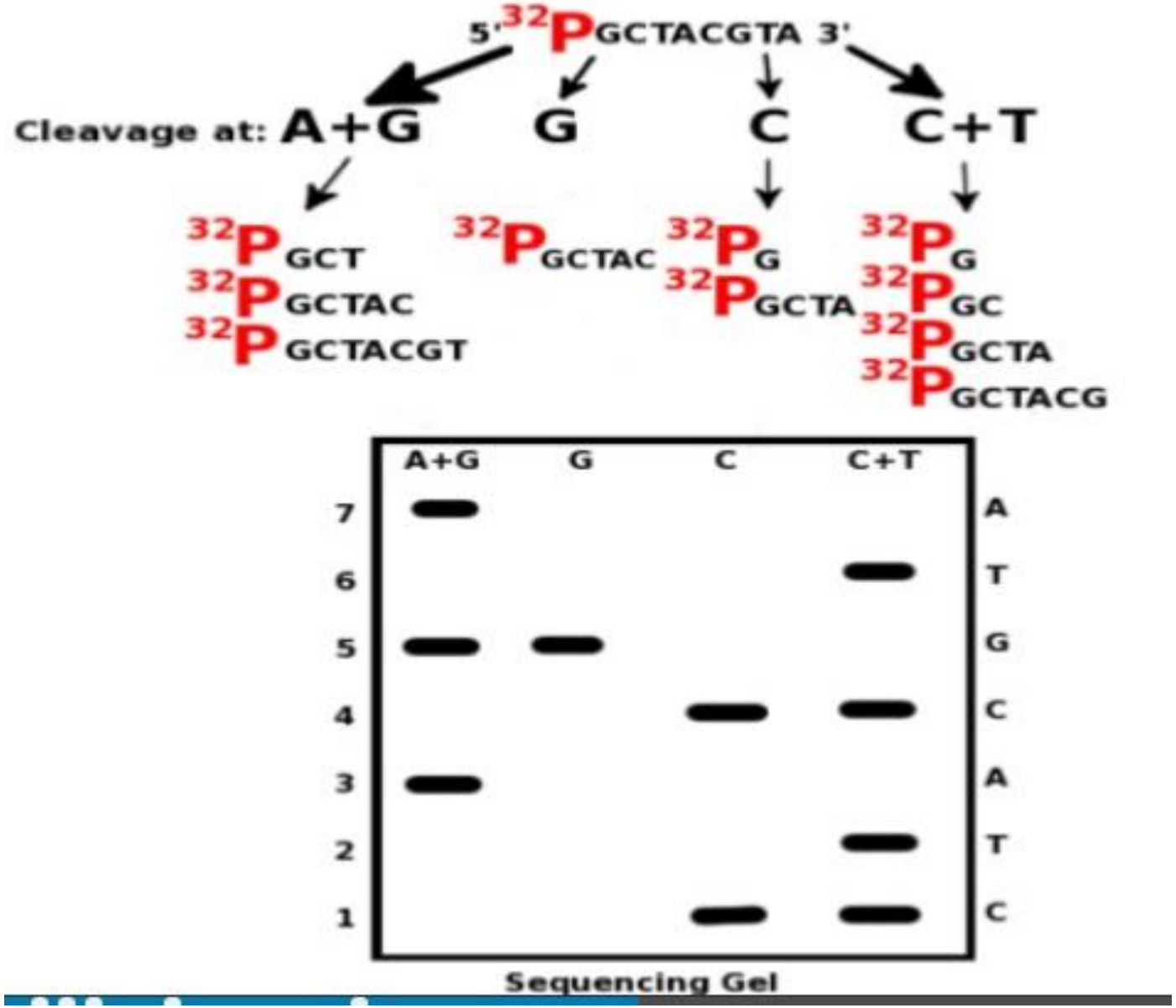
– Pyrimidine

- Cytosine
- Thymine

– Hydrazine----- C+T

– Hydrazine + NaCl--------C

An hypothetical example



Next Generation Sequencing (NGS)

➢The principle behind Next Generation Sequencing (NGS) is similar to that of Sanger sequencing, which relies on capillary electrophoresis.

➢The genomic strand is fragmented, and the bases in each fragment are identified by emitted signals when the fragments are ligated against a template strand.

➢The NGS method uses array-based sequencing which combines the techniques developed in Sanger sequencing to process millions of reactions in parallel, resulting in very high speed and throughput at a reduced cost.

Three general steps in NGS

1. **Library preparation**: libraries are created using random fragmentation of DNA, followed by ligation with custom linkers
2. **Amplification**: the library is amplified using clonal amplification methods and PCR
3. **Sequencing**: DNA is sequenced using one of several different approaches

Step 1

LIBRARY PREPARATION

➢Firstly, DNA is fragmented either enzymatically or by sonication (excitation using ultrasound) to create smaller strands.

➢Adaptors (short, double-stranded pieces of synthetic DNA) are then ligated to these fragments with the help of DNA ligase, an enzyme that joins DNA strands.

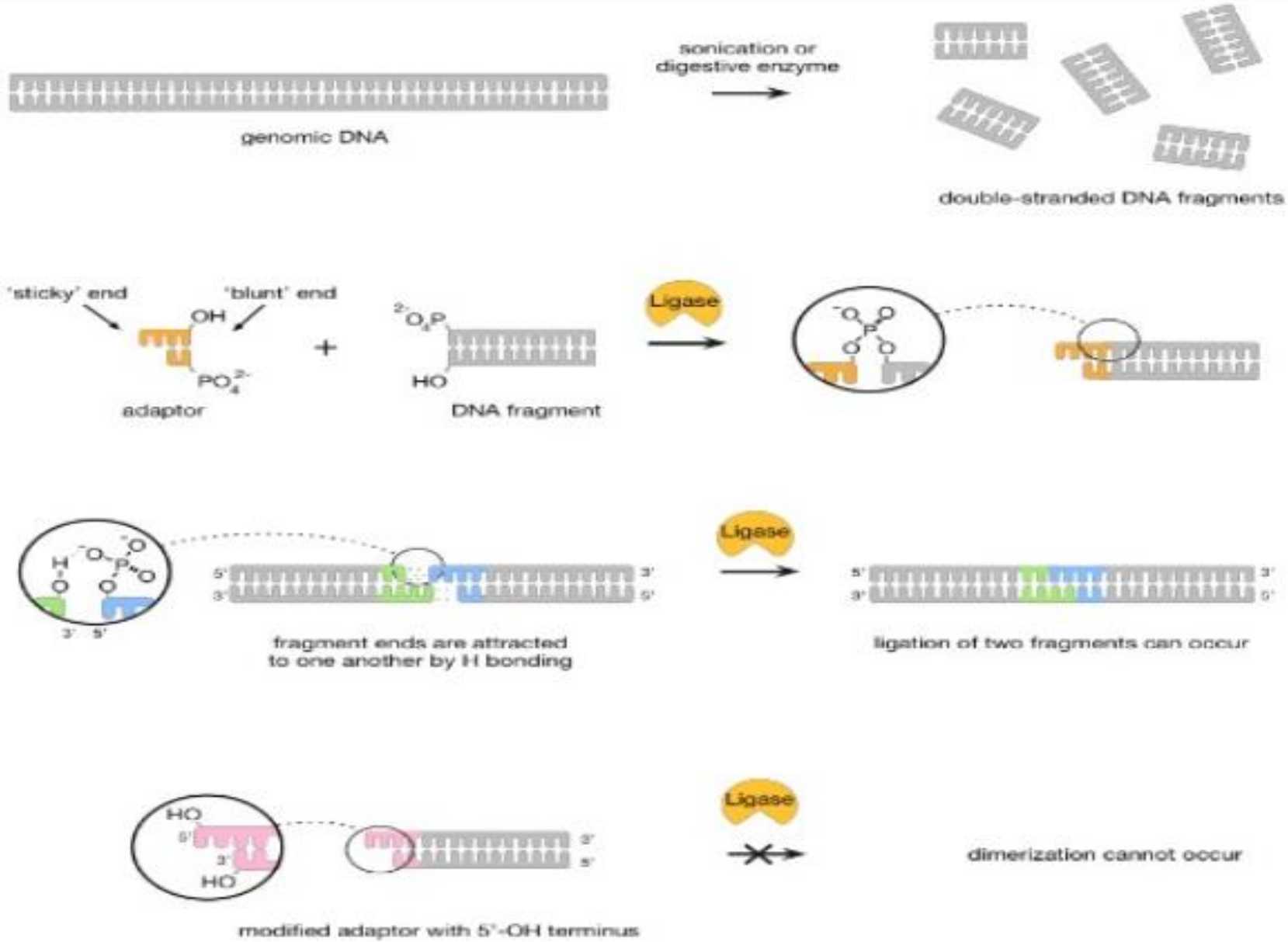
➢The adaptors enable the sequence to become bound to a complementary counterpart.

➢Adaptors are synthesized so that one end is 'sticky' whilst the other is 'blunt' (non-cohesive) with the view to joining the blunt end to the blunt ended DNA.

➢This could lead to the potential problem of base pairing between molecules and therefore dimer formation.

➢To prevent this, the chemical structure of DNA is utilised, since ligation takes place between the 3'-OH and 5'-P ends.

➢By removing the phosphate from the sticky end of the adaptor and therefore creating a 5'-OH end instead, the DNA ligase is unable to form a bridge between the two termini.


genomic DNA
sonication or
digestive enzyme
double-stranded DNA fragments
'sticky' end
'blunt' end
OH
PO_4^{2-}
adaptor
+
$^{2-}O_4P$
HO
DNA fragment
Ligase
fragment ends are attracted
to one another by H bonding
Ligase
ligation of two fragments can occur
HO
HO
modified adaptor with 5'-OH terminus
Ligase
dimerization cannot occur

Step 2

AMPLIFICATION

➢Library amplification is required so that the received signal from the sequencer is strong enough to be detected accurately.

➢With enzymatic amplification, phenomena such as 'biasing' and 'duplication' can occur leading to preferential amplification of certain library fragments.

➢Instead, there are several types of amplification process which use PCR to create large numbers of DNA clusters.

A

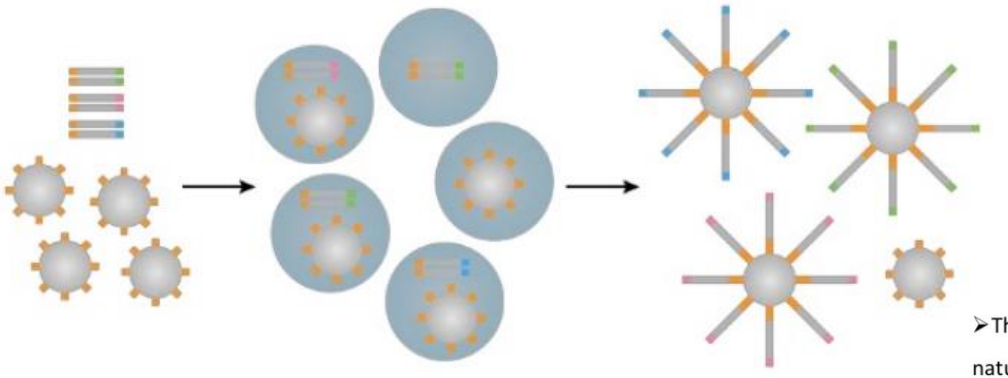
➢In order for the sequencing process to be successful, each micro well should contain one bead with one strand of DNA (approximately 15% of micro wells are of this composition).

➢The PCR then denatures the library fragment leading two separate strands, one of which (the reverse strand) anneals to the bead.

➢The annealed DNA is amplified by polymerase starting from the bead towards the primer site.

Emulsion PCR

Emulsion oil, beads, PCR mix and the library DNA are mixed to form an emulsion which leads to the formation of micro wells



B

➢The original reverse strand then denatures and is released from the bead only to re-anneal to the bead to give two separate strands.

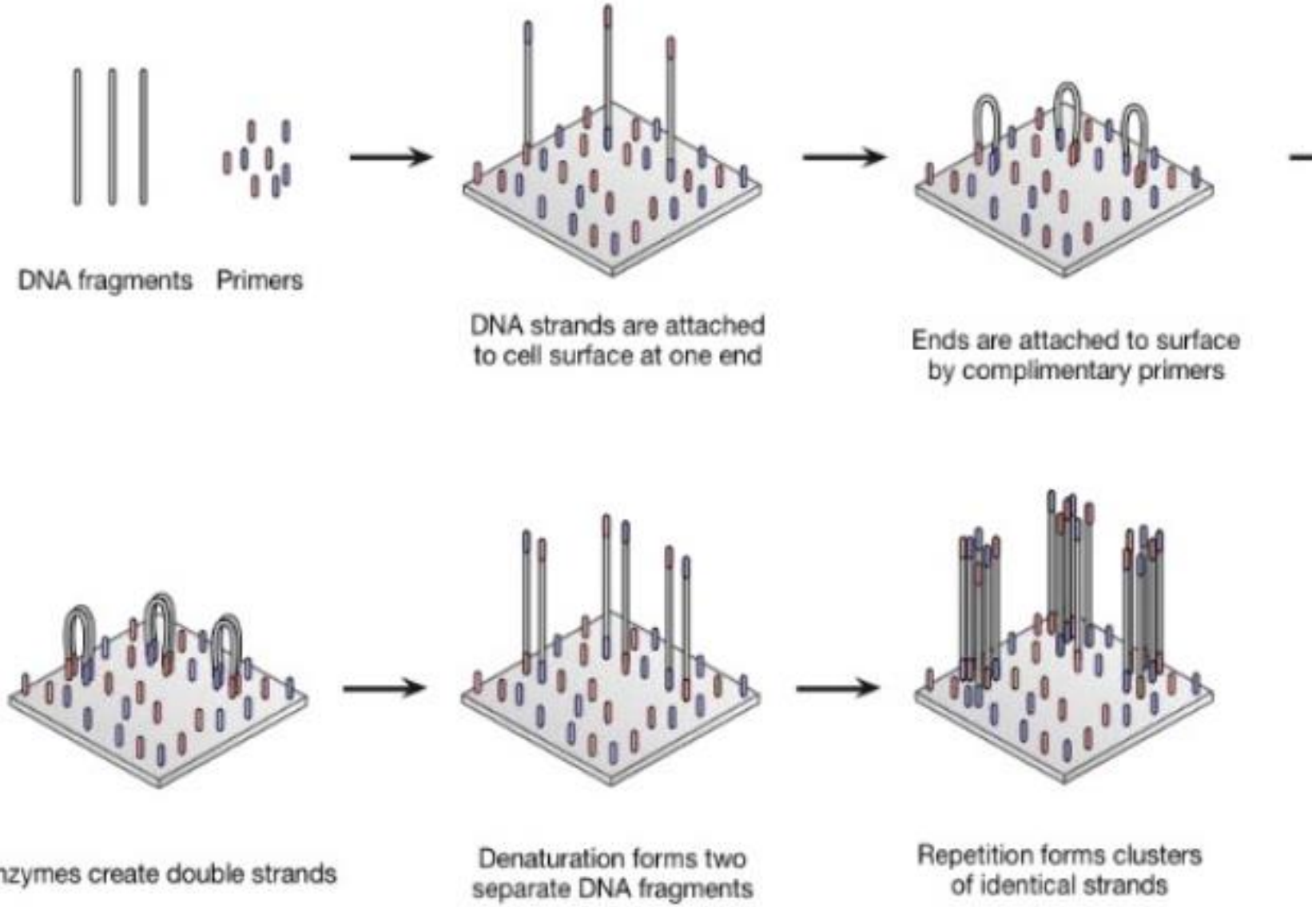
➢These are both amplified to give two DNA strands attached to the bead.

➢The process is then repeated over 30-60 cycles leading to clusters of DNA.

➢This technique has been criticized for its time consuming nature, since it requires many steps (forming and breaking the emulsion, PCR amplification, enrichment etc) despite its extensive use in many of the NGS platforms.

➢It is also relatively inefficient since only around two thirds of the emulsion micro reactors will actually contain one bead.

➢Therefore an extra step is required to separate empty systems leading to more potential inaccuracies.



This PCR is later modified by Bridge PCR

Bridge PCR

- The surface of the flow cell is densely coated with primers that are complementary to the primers attached to the DNA library fragments.
- The DNA is then attached to the surface of the cell at random where it is exposed to reagents for polymerase based extension.
- On addition of nucleotides and enzymes, the free ends of the single strands of DNA attach themselves to the surface of the cell via complementary primers, creating bridged structures.
- Enzymes then interact with the bridges to make them double stranded, so that when the denaturation occurs, two single stranded DNA fragments are attached to the surface in close proximity.
- Repetition of this process leads to clonal clusters of localized identical strands.
- In order to optimize cluster density, concentrations of reagents must be monitored very closely to avoid overcrowding.

Step 3

SEQUENCING

Several competing methods of Next Generation Sequencing have been developed by different companies.

1. 454 Pyrosequencing
2. Ion torrent semiconductor sequencing
3. Sequencing by ligation (SOLiD)
4. Reversible terminator sequencing (Illumina)
    - 3'-O-blocked reversible terminators
    - 3'-unblocked reversible terminators

*We will study the first two

1. 454 Pyrosequencing

➢Pyrosequencing is based on the 'sequencing by synthesis' principle, where a complementary strand is synthesized in the presence of polymerase enzyme.

➢In contrast to using dideoxynucleotides to terminate chain amplification (as in Sanger sequencing), pyrosequencing instead detects the release of pyrophosphate when nucleotides are added to the DNA chain.

➢It initially uses the emulsion PCR technique to construct the colonies required for sequencing and removes the complementary strand.

Clip slide

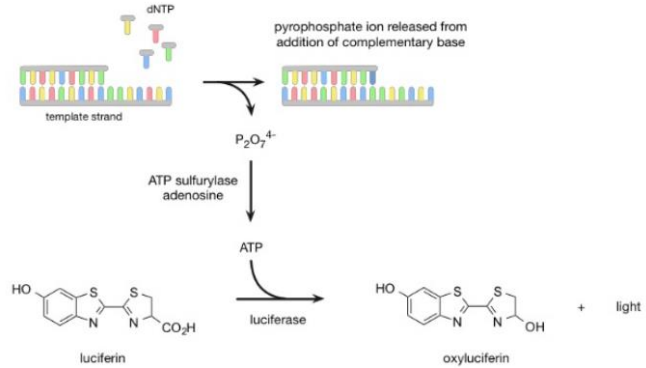
➢Next, a ssDNA sequencing primer hybridizes to the end of the strand (primer-binding region), then the four different dNTPs are then sequentially made to flow in and out of the wells over the colonies.

➢When the correct dNTP is enzymatically incorporated into the strand, it causes release of pyrophosphate

➢In the presence of ATP sulfurylase and adenosine, the pyrophosphate is converted into ATP.

➢This ATP molecule is used for luciferase-catalysed conversion of luciferin to oxyluciferin, which produces light that can be detected with a camera.

➢The relative intensity of light is proportional to the amount of base added (i.e. a peak of twice the intensity indicates two identical bases have been added in succession).



➢Pyrosequencing, developed by 454 Life Sciences, was one of the early successes of Next-generation sequencing; indeed, 454 Life Sciences produced the first commercially available Next-generation sequencer.

➢However, the method was eclipsed by other technologies and, in 2013, new owners Roche announced the closure of 454 Life Sciences and the discontinuation of the 454 pyrosequencing platform.

2. Ion torrent semiconductor sequencing

➢Ion torrent sequencing uses a "sequencing by synthesis" approach, in which a new DNA strand, complementary to the target strand, is synthesized one base at a time.

➢A semiconductor chip detects the hydrogen ions produced during DNA polymerization

➢Following colony formation using emulsion PCR, the DNA library fragment is flooded sequentially with each nucleoside triphosphate (dNTP), as in pyrosequencing.

➢The dNTP is then incorporated into the new strand if complementary to the nucleotide on the target strand.

➢Each time a nucleotide is successfully added, a hydrogen ion is released, and it detected by the sequencer's pH sensor.

➢As in the pyrosequencing method, if more than one of the same nucleotide is added, the change in pH/signal intensity is correspondingly larger.

➢Ion torrent sequencing is the first commercial technique not to use fluorescence and camera scanning.

➢It is therefore faster and cheaper than many of the other methods.

➢Unfortunately, it can be difficult to enumerate the number of identical bases added consecutively.

➢For example, it may be difficult to differentiate the pH change for a homo repeat of length 9 to one of length 10, making it difficult to decode repetitive sequences.