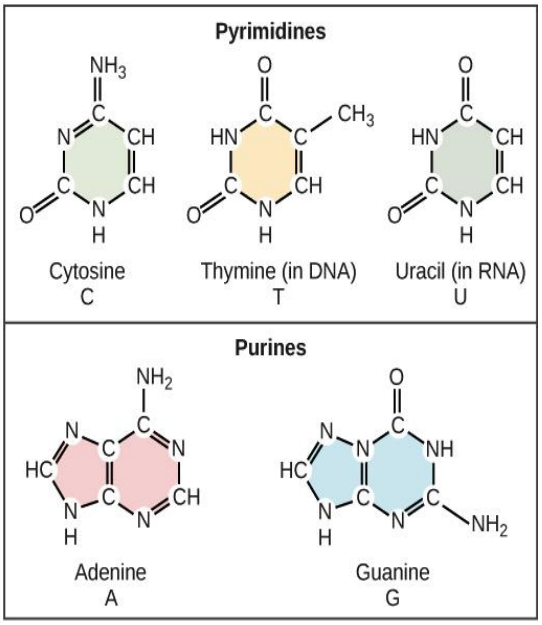
DNA Sequencing

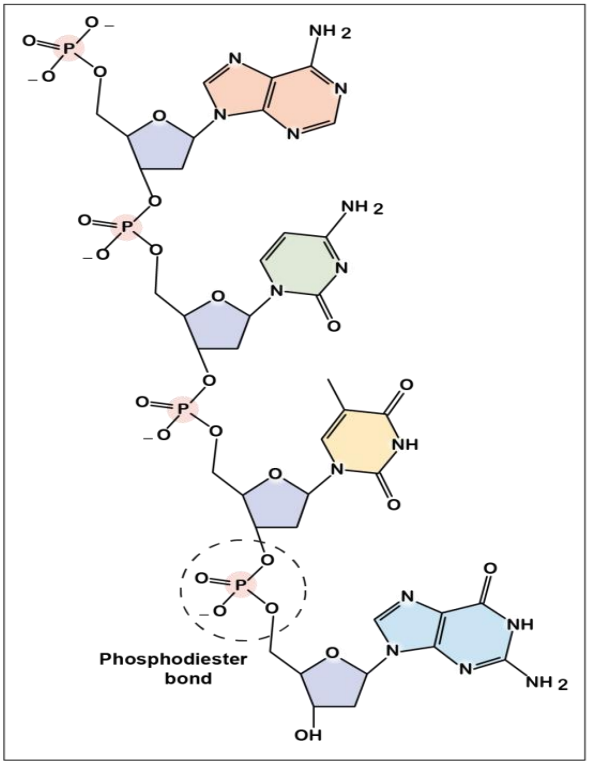
(NOTES)

DNA sequencing is the process of determining the exact order of **nucleotides** within a **DNA** molecule. This method is used to determine the order of the four bases—*adenine* (**A**), *guanine* (**G**), *cytosine* (**CY**), and *thymine* (**T**) in a strand of **DNA**. The advent of rapid DNA sequencing methods has greatly accelerated the biological and medical research.

Four nitrogenous bases of DNA



DNA structure

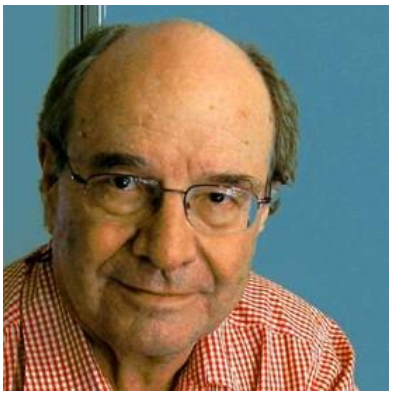
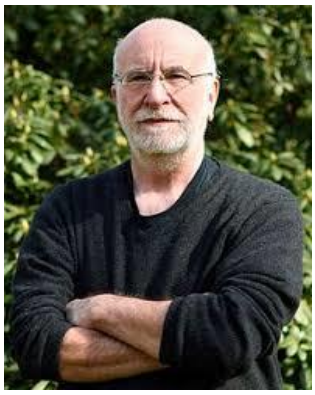


DNA sequencing has been feasible because of the following developments

1. Availability of restriction endonucleases
2. Development of highly sensitive gel electrophoresis technique, which can separate DNA fragments, differing by only one nucleotide.
3. Availability of large quantities of individual DNA fragments due to development of gene cloning and PCR techniques.
4. Development of reliable, easy and rapid DNA techniques

There are two basic techniques of DNA sequencing

- **Chemical sequencing (** explored by **Maxam** and **Gilbert**)
- **Enzymatic sequencing** (explored by **Sanger**)

		
Allan Maxam	**Walter Gilbert**	**Frederick Sanger**

MAXAM & GILBERT PROCEDURE (*Chemical Method*)

Allan Maxam and **Walter Gilbert** published a **DNA** sequencing method in 1977 based on chemical modification of DNA and subsequent cleavage at specific bases. This method allows purified samples of double-stranded DNA to be used without further cloning.

Maxam-Gilbert sequencing requires radioactive labelling at **5' end** or **3' end** of the **DNA** followed by purification of the DNA fragment to be sequenced.

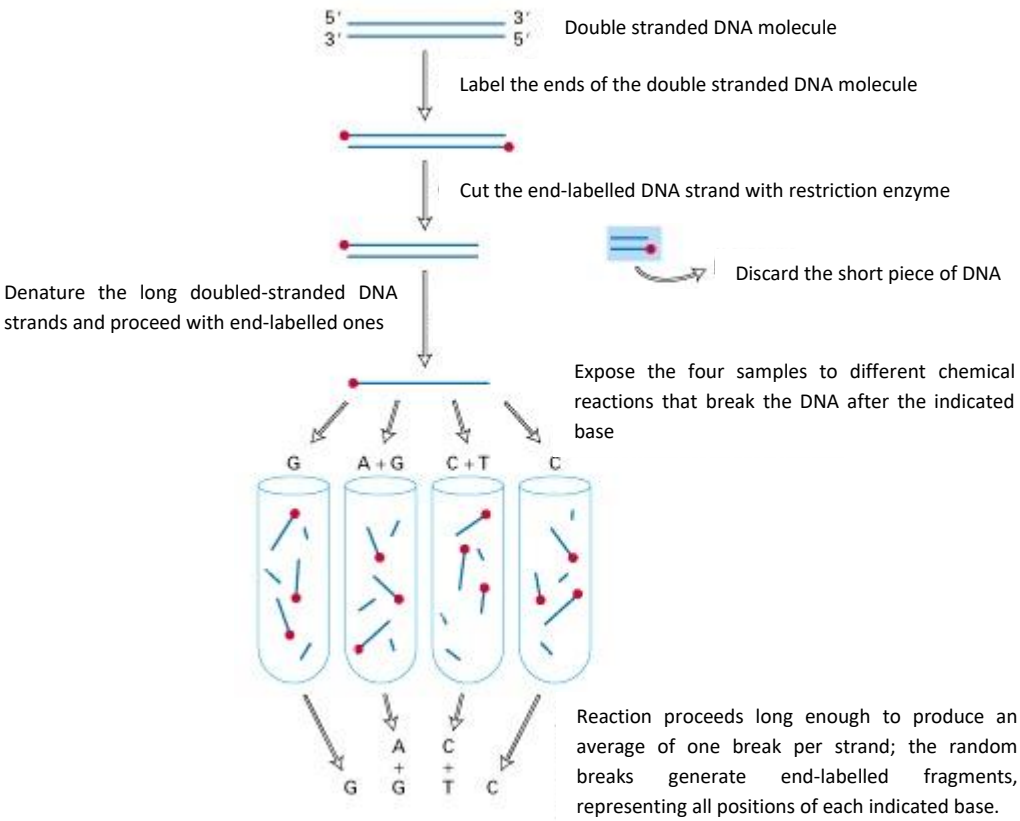
Procedure (STEPS**)**

1. Radioactive labelling of one end (**5' end** or **3' end**) of the DNA fragment to be sequenced by a kinase reaction using ^{32}**P**.
2. Cut the DNA fragment with specific restriction enzyme, resulting in two unequal DNA fragments
3. Denature the double-stranded DNA to single-stranded DNA by increasing temperature.
4. Cleave the DNA strand at specific positions using chemical reactions. For example, we can use one of the two chemicals followed by addition of piperdine. Dimethyl sulphate (DMS) selectively attacks purine (**A** and

G), while hydrazine selectively attacks pyrimidines (**C** and **T**). This is called **modification step**.

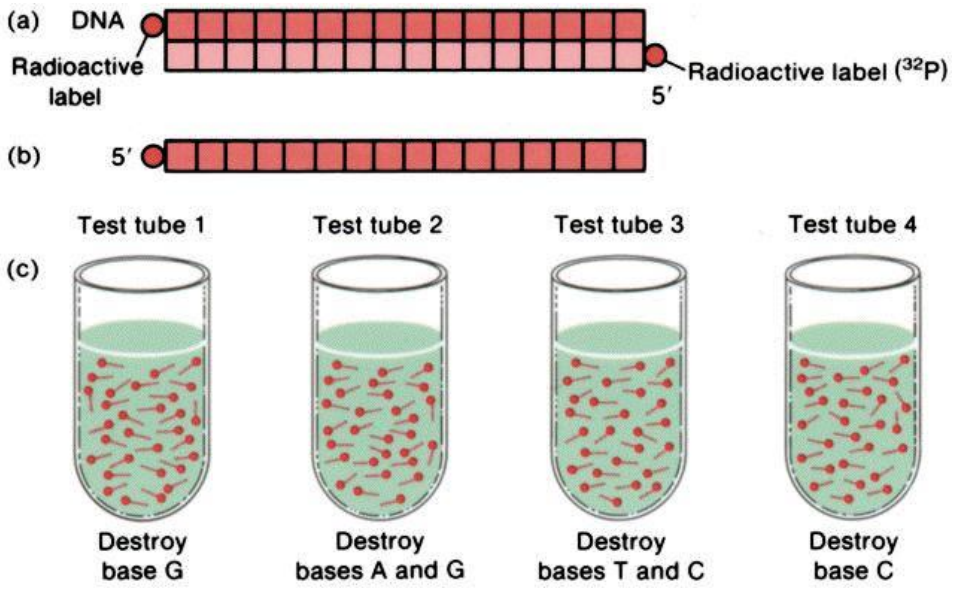
5. Chemical treatment generates breaks at the four **nucleotide bases** in the four reaction mixtures (**G**, **A+G**, **C**, and **C+ T**).

Maxam & Gilbert technique of DNA sequencing



Reagent mixtures

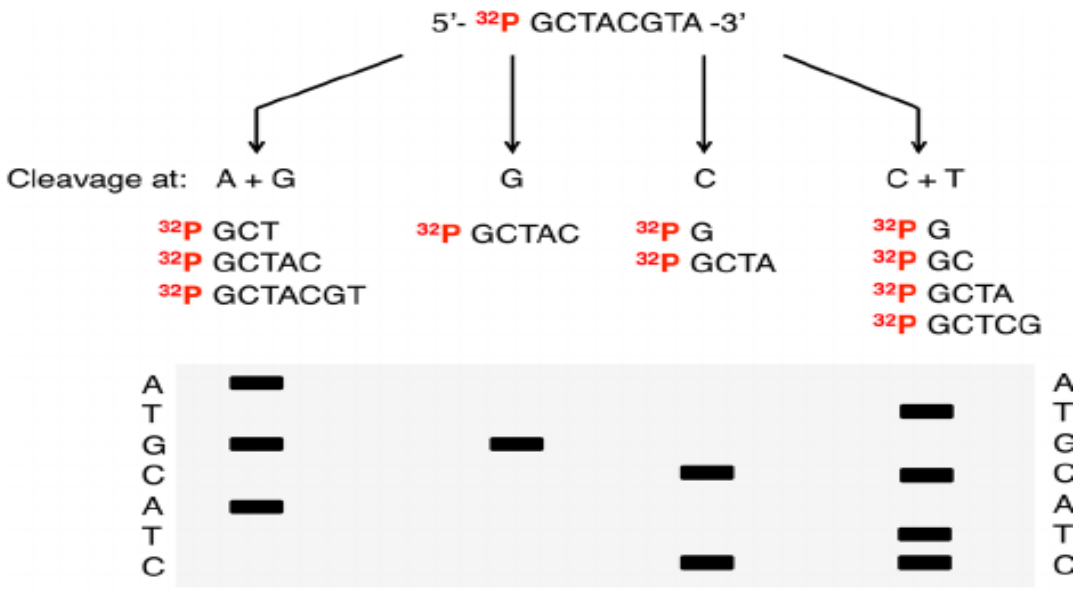
1. **Reagent G:** It breaks the DNA chain after guanine (**G**) base
2. **Reagent A+G:** It breaks the DNA chain after adenine (**A**) and guanine (**G**) bases
3. **Reagent C:** It breaks the DNA chain after cytosine (**C**) base
4. **Reagent C+T:** It breaks the DNA chain after cytosine (**C**) and guanine (**G**) bases



Four reaction mixture

(DNA modification steps**)**

The concentration of the modifying chemicals is controlled to introduce, on an average, one modification per DNA molecule. Thus a series of labelled fragments is generated, starting from the radiolabeled end to the first "cut" site in each molecule. As a result, we have several differently sized DNA strands in four reaction tubes.

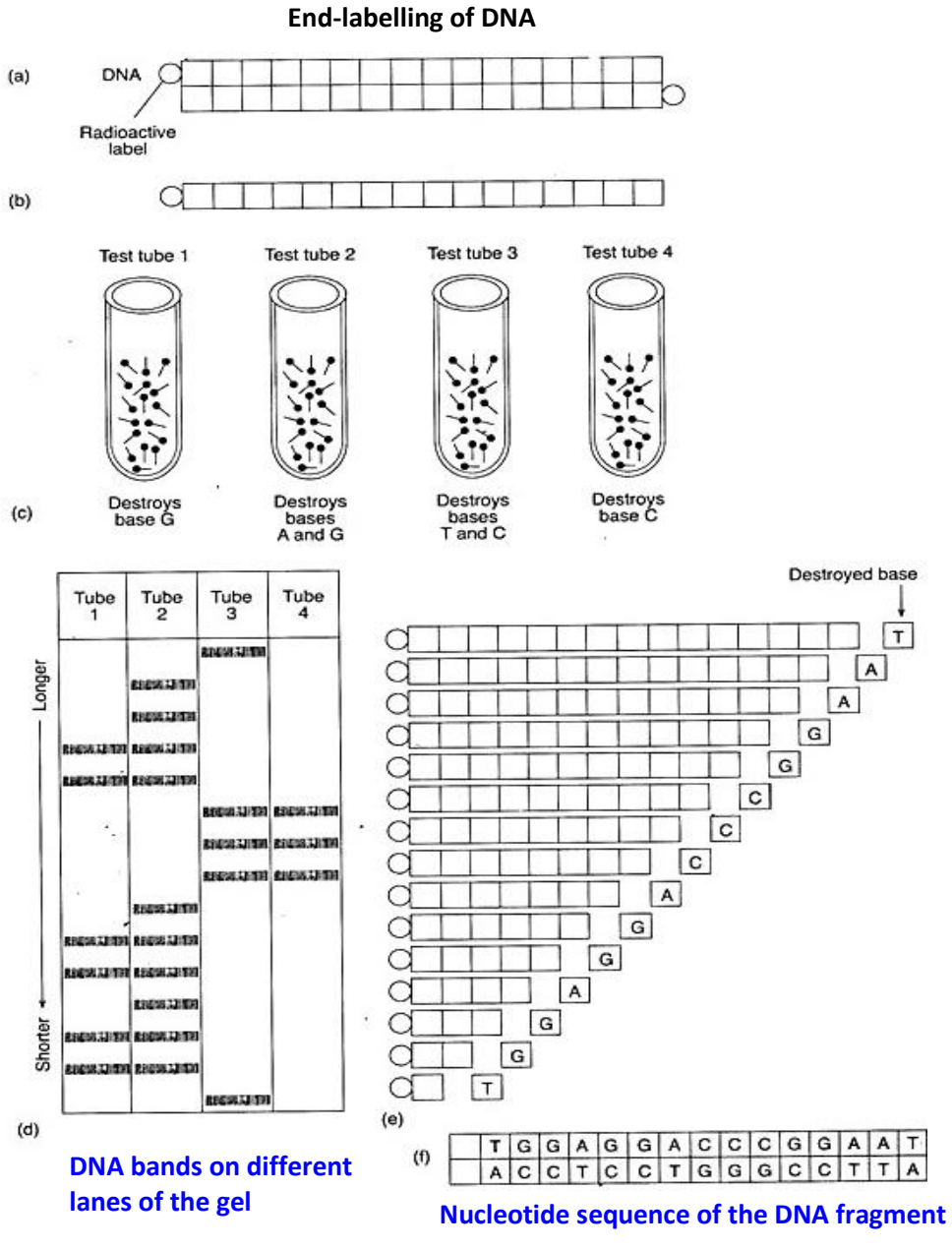


Nucleotide sequences as per the DNA bands depicted on the gel

Fragments are subjected to electrophoresis in high-resolution acrylamide gels for size-based separation.

To visualize the fragments, the gel is exposed to X-ray film for autoradiography, which yields a series of dark bands, each corresponding to a radiolabeled DNA fragment, from which the nucleotide sequence may be inferred. In the gel, the fragments are ordered by size and, thus, we can deduce the sequence of the DNA molecule.

DNA sequencing by Maxam & Gilbert Technique



DNA bands on different lanes of the gel

Nucleotide sequence of the DNA fragment

DNA sequencing evaluation: *Reading the gel*

1. The gel is read from bottom to top
2. The gel has nucleotide sequence differing by only one nucleotide; i.e. each subsequent base will be one nucleotide longer than the previous one.
3. The larger the fragment, the more it is slowed down by the gel; i.e. the largest fragment will be at the bottom, while the smallest one will be at the top of the gel.
4. Each band on the gel identifies the specific nucleotide; thus, the nucleotide sequence of the DNA fragment can be read off the gel '**end to end**'.

2. ENZYMATIC PROCEDURE (Chain Termination Method)

The enzymatic method is called as Sanger Method. **Sanger sequencing method** was developed by **Frederick Sanger** and his colleagues in **1977**. The development of this technique won Sanger the **Nobel Prize** in Chemistry in 1980.

From the 1980's to the mid - 2000's, Sanger sequencing dominated the DNA sequencing platform, bringing successful completion of the Human Genome Project (HGP) in 2003. Although this technique has been replaced by **next generation sequencing methods**, it is still used today for smaller-scale projects.

In order to perform the sequencing, one must first convert double stranded DNA into single stranded DNA. This can be done by denaturing the double stranded DNA with **NaOH**.

A Sanger reaction consists of the following components:

Single stranded DNA fragment (ssDNA template): a DNA strand to be sequenced (one of the single strands, which was denatured using NaOH).

All four deoxyribonucleotide triphosphates: i.e. d**ATP**, d**GTP**, d**TTP** and d**CTP**

NOTE: *At least one of the four **deoxy**ribonucleotide triphosphates should be radioactive in each of the four reaction mixture tubes in order to permit the autoradioautographic development of **DNA bands** after the gel electrophoresis.*

DNA polymerase: Each incubation tube will also carry DNA polymerase enzyme (**Sequenase**) in order to copy the **DNA template** by adding nucleotides to the **primer** as the synthesis proceeds.

NOTE: Sequenase is an engineered *E. Coli* **DNA polymerase I** (known as '**Klenow Fragment**'), which is used to copy the DNA template. Sequenase is obtained by removing the first 323 amino acids of the polypeptide (5'-------3'), using **exonuclease** enzyme.

DNA primers: The enzyme Sequenase needs a **primer** to start end to end nucleotide synthesis

NOTE: *A primer should have the following characteristics*

1. The primer should be either a DNA restriction fragment or short DNA sequence complementary to the single stranded DNA template.
2. The primer should have a free **3' – OH** group required to make **3' – 5'** phosphodiester linkage between the nucleotides to be added to the primer.
3. The primer should be radioactively labelled at the **5' end**.

Four reaction mixtures

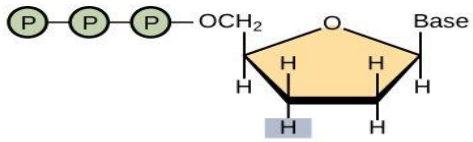
In addition to **radio-labelled primers** and **DNA polymerase** (*Sequenase*), each incubation tube (carrying a reaction mixture) should have all the four **deoxy**ribonucleotides (d**ATP**, d**CTP**, d**GTP** and d**TTP**) and a particular **dideoxyribonucleotide phosphate**.

The four reaction mixtures should differ from each other in having a different dideoxyribonucleotide phosphate analogue (dd**NTP** analogues: dd**GTP**, dd**ATP**, dd**TTP** and dd**CTP**).

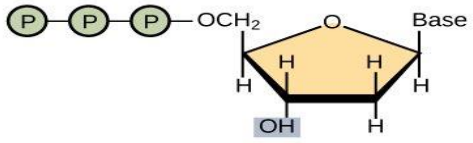
Example: Tube A will carry Sequenase, radio-labelled primers, all the four deoxyribonucleotides (d**ATP**, d**CTP**, d**GTP** and d**TTP**) and a particular dideoxyribonucleotide phosphate (dd**ATP** in this case).

Dideoxynucleotides (dd**NTP**s) are **chain-elongation inhibitors** of DNA polymerase, used in the Sanger method for DNA sequencing.

A **dideoxynucleotide** (dd**NTP**) is an artificial molecule that lacks a hydroxyl (**OH**) group at both the **2nd** and **3rd** carbons of the **sugar moiety** of DNA molecule. In contrast, a regular deoxynucleotide triphosphate (d**NTP**) has the hydroxyl group on the **3rd** carbon of the sugar.

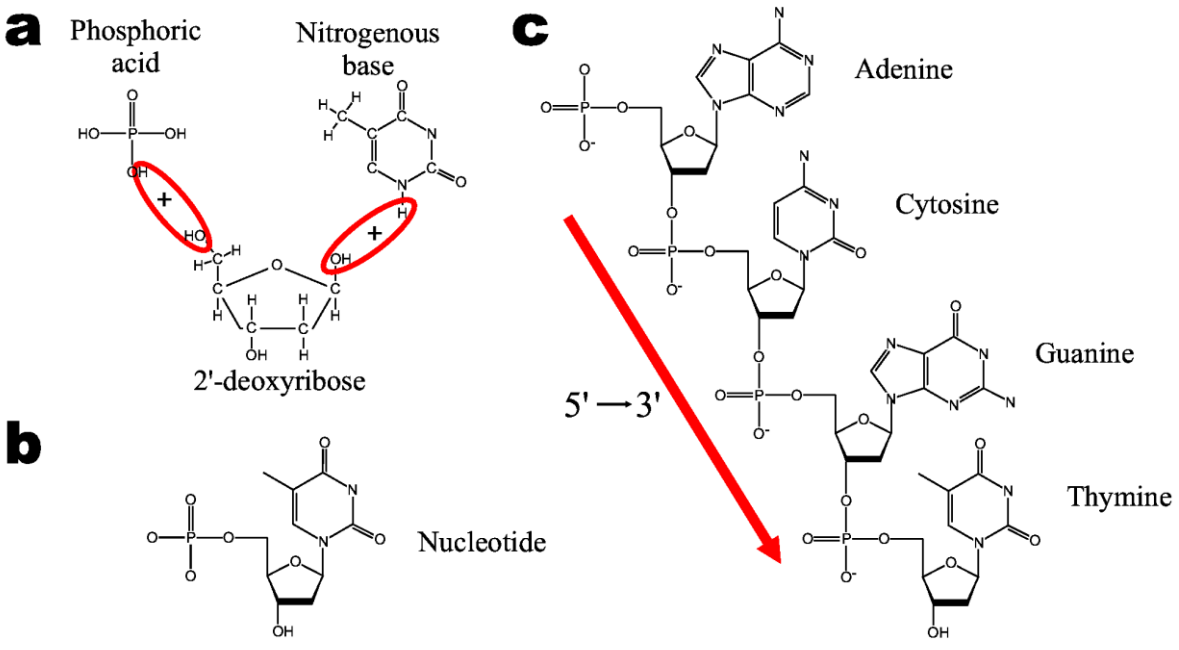


Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

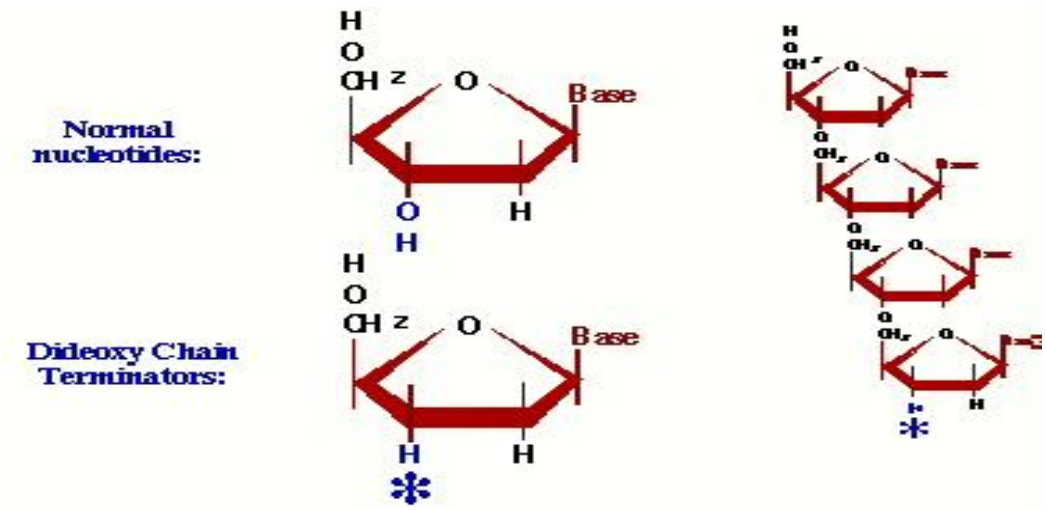
DNA-Chain Components



The main purpose of the **3'-OH** group is that it is used to form a phosphodiester bond between two nucleotides - this allows a DNA strand to elongate.

During **DNA replication**, an incoming nucleoside triphosphate is linked by its **5' α-phosphate group** to the **3' hydroxyl group** of the last nucleotide of the growing chain. With dd**NTP**, where there is no **3' - OH** group, this reaction cannot take place, so elongation is terminated. Thus, each new strand will stop randomly at positions where d**NTP** is replaced by dd**NTP**

DNA elongation cannot occur if dNTP is replaced with its analogue (ddNTP)



The concentration of dd**NTP** should be **1%** of the concentration of d**NTP**. The logic behind this ratio is that after **DNA polymerase** is added, the polymerization will take place and will terminate whenever a dd**ATP** is incorporated into the growing strand.

Thus, **four sets** of chain-termination fragments, corresponding to **A**, **C**, **T** and **G**, are produced in four reaction mixtures.

Procedure

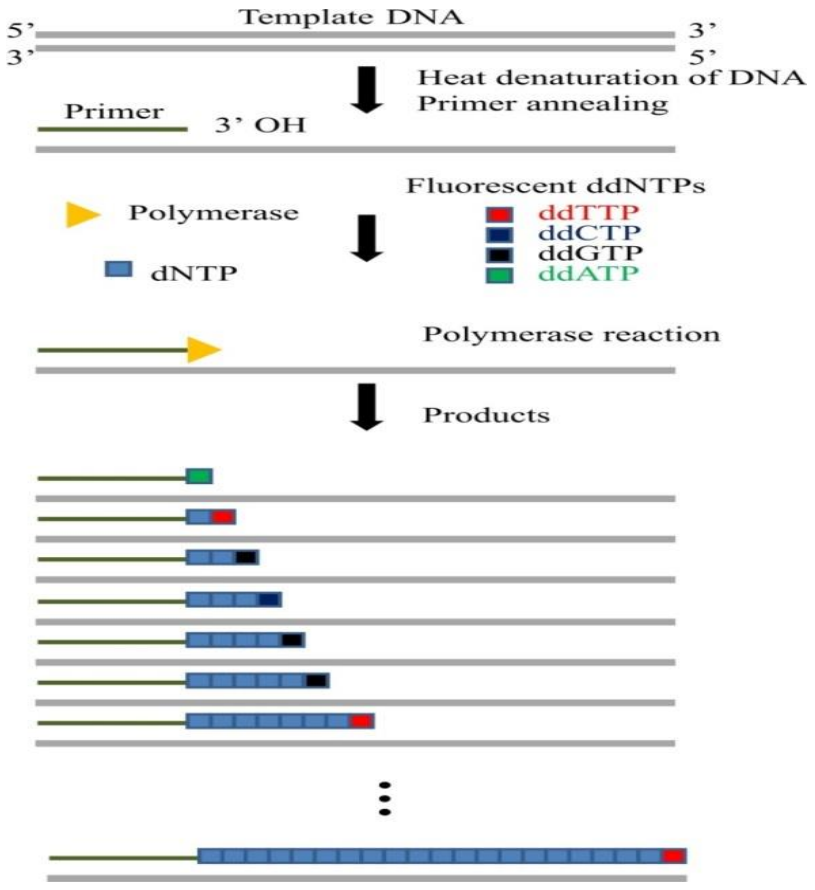
The single stranded DNA is mixed with primer and split into four reaction mixtures. Each **reaction mixture** contains DNA polymerase, four deoxyribonucleotide phosphates (d**NTP**s) and a replication terminator (dd**NTP**). Each reaction proceeds until a replication terminating nucleotide

(dd**NTP**) is added. The mixtures are loaded into **four separate lanes** of gel and the electrophoresis is used to separate the DNA fragments.

STEPS:

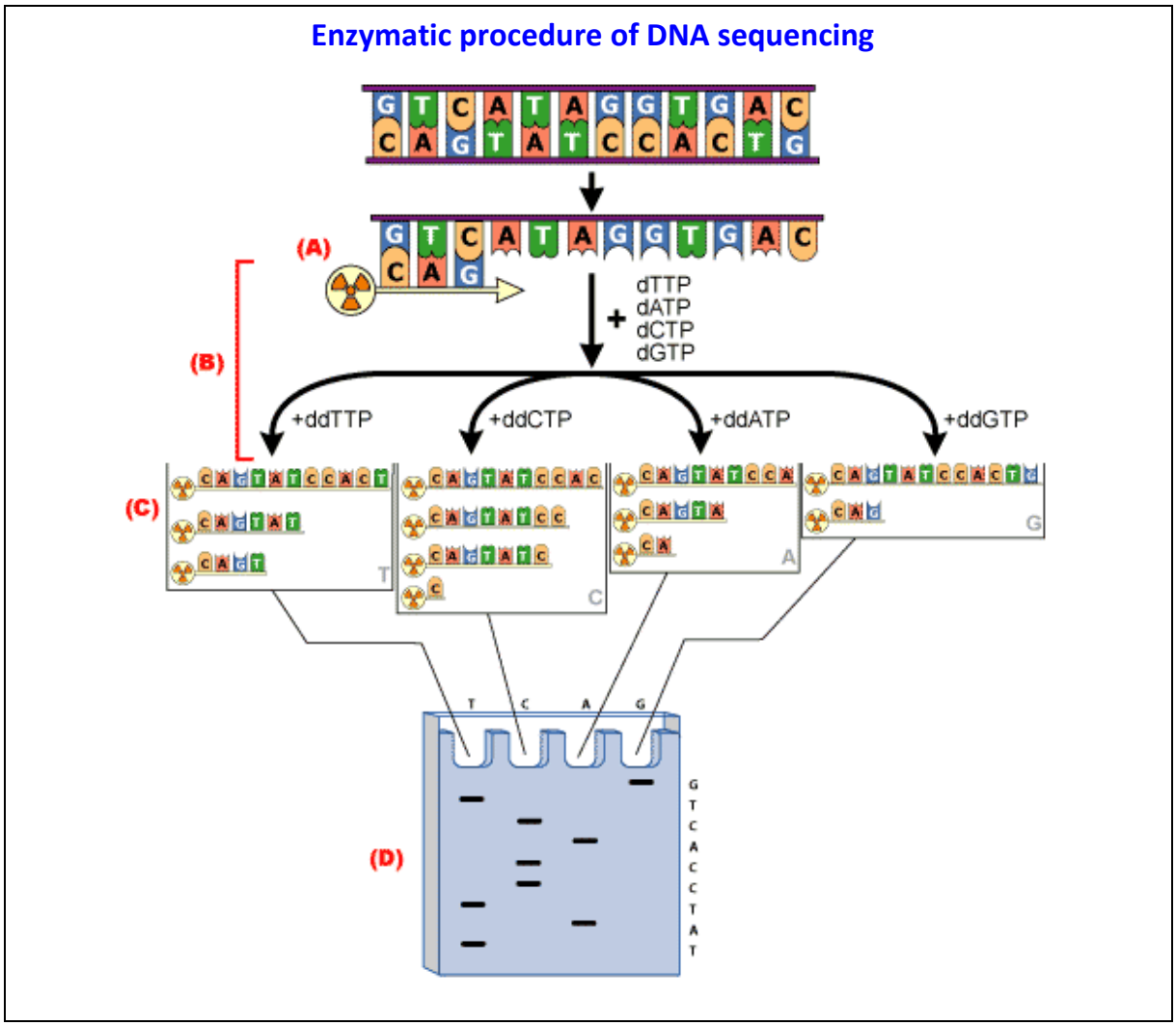
1. **Fragmentation and amplification:** Fragment the DNA and clone the fragments into vectors.
2. **Denaturation:** Denature the double stranded DNA (by heat or NaOH) into single stranded DNA fragments.
3. **Attach the primer:** A primer is a synthetic oligonucleotide, containing 17 to 24 nucleotides. The primer binds to the DNA molecule and provides a **3' OH** group, which is necessary to initiate DNA synthesis. The **3'-OH** group allows for **DNA chain elongation**.
4. **Add 4** d**NTPs + 1** dd**NTP**

Enzymatic procedure of DNA sequencing



5. Four different reaction vials are taken, each with the four standard d**NTP**s (d**ATP**, d**GTP**, d**CTP** and d**TTP**) and DNA polymerases. Difference among the vials is because of different type of dd**NTP**. Each vial will have **1** dd**NTP** per 100 d**NTP**s.
6. After the occurrence of DNA synthesis, each reaction vial will have a unique set of single-stranded DNA molecules of varying lengths. However, all DNA molecules will have the same primer sequence at its **5' end**.

7. **Find the nucleotide sequence using gel electrophoresis:** In the gel, we have varying sequences, lined up according to size.



Autoradiography

In order to identify various terminated chains of DNA (the DNA strands) the dd**NTP**s would have to be radioactively or fluorescently labelled beforehand.

The DNA strands are then separated using gel electrophoresis; then, the gel is read from top to bottom (3' to 5') to obtain the sequence.

Smaller strands migrate to the bottom, while larger strands stay up top (near the well). We can read each molecule in order to find the DNA sequence

Merits of chain termination method

1. Not as toxic and less radioactivity than Maxam and Gilbert method.
2. Easier to automate - Leroy Hood and co-workers used fluorescently labelled ddNTPs and primers for the first high-throughput DNA sequencing machine. This lowered the cost from $100 million to $10,000 USD in 2011.
3. Highly accurate long sequence reads of about 700 base pairs.
4. Easier to get started. The kits that are commercially available contain reagents necessary for sequencing - pre-aliquoted and ready to use.

AUTOMATED DNA-SEQUENCING (Modification of Chain Termination Method**)**

Automatic DNA sequencing is based on Sanger method of DNA sequencing. It makes use of Automatic Sequencers.

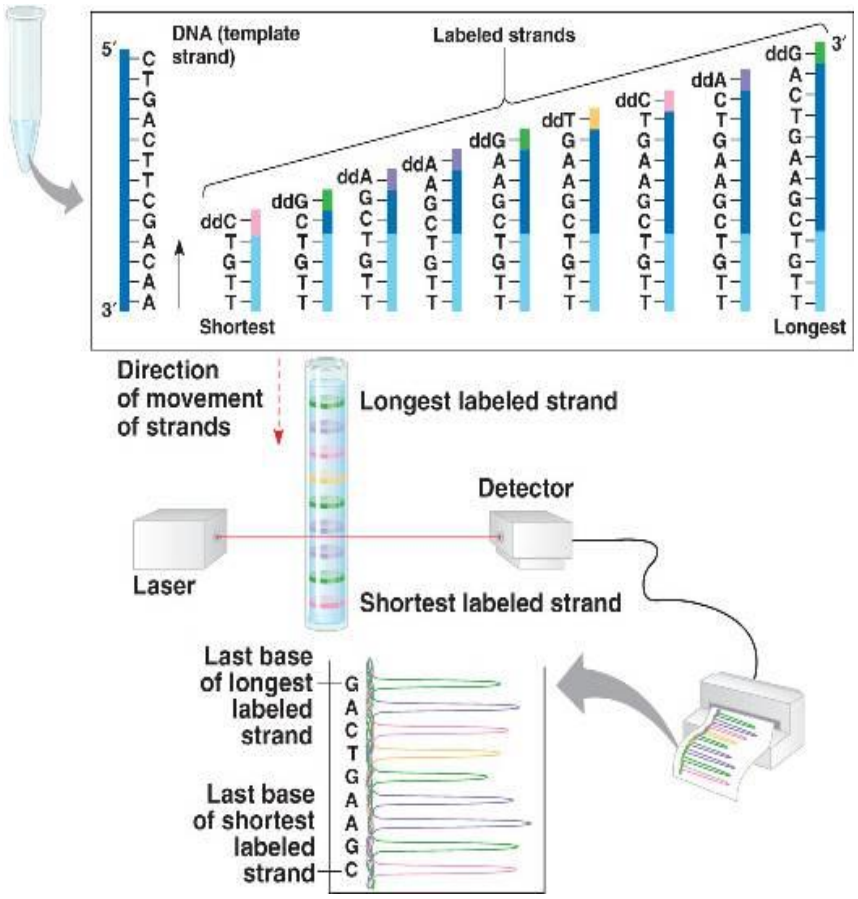
STEPS:

1. Labelling the products of polymerase chain reaction(PCR)
2. In place of radioactive labelling, fluorescent lables are used. The fluorescent lables are attached to the four dideoxyribonucleotides or ddNTPs (ddATP, ddGTP, ddCTP and ddTTP) used for chain termination.
3. If a different fluorochrome is attached to each of the four ddNTPs, all of them may be used in the same reaction (the **Single Track System**).
4. In a single track system, the reaction products are run in a single lane of gel or capillary.

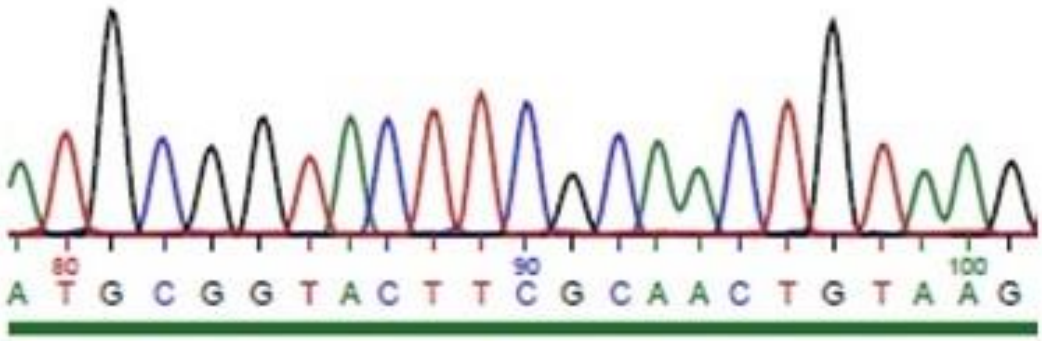
5. The reaction products are subjected to polyacrylamide gel electrophoresis (PAGE) under denaturing conditions.

6. The bands produced in polyacrylamide gel are identified with the help of fluorescent detector, which identifies the fluorescent signal emitted by each band.

Automated DNA Sequencing



Automated DNA Sequences



7. The fluorochromes are excited by laser beam and the resulting fluorescent signal is sensed by a photovoltaic cell.

8. The resulting data are fed into a computer, which converts these signals into the base sequence of the DNA molecule

9. The sequence information (data) can be printed out or stored in data storage device for future use

Automated DNA sequencing has the following advantages

1. Radioactivity is not used
2. Gel processing after electrophoresis and autoradiography is not needed.
3. The tedious manual reading of the gel is not required as the data are processed by computer.
4. The sequenced data is directly fed and stored in computer
5. The separation of the same reaction products can be repeated to check the results in case of doubts
6. The method is extremely fast.

<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<>>>>>>>>>>>>>>>>>>>>>>>>>>>