Core course
BMS361N
Genetic Engineering

DNA SEQUENCING

Prof. Narkunaraja Shanmugam
Dept. Of Biomedical Science
School of Basic Medical Sciences
Bharathidasan University

DNA SEQUENCING

- In mid 1970
  - Cloning was rapidly improving
  - Prof. S.Cohen in Stanford was one of the pioneer in improving this method.
  - He was using Bac. Genetic to clone genes from bacteria but same time they wanted to see the nucleotide sequence of DNA.
  - First nucleotide sequencing technique was developed by Robert Holley's lab group.
  - They were the one first sequenced $tRNA^{Ala}$

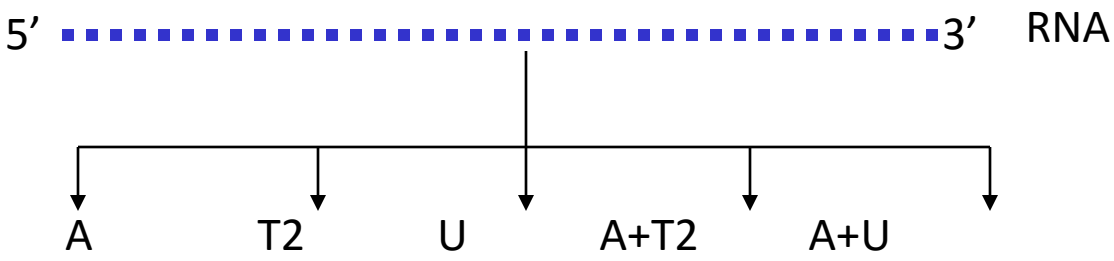
DNA SEQUENCING

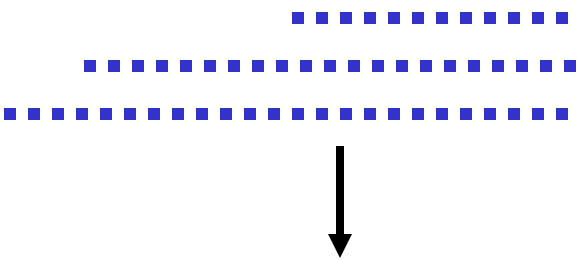
Holley's group

- Purified $tRNA^{Ala}$ was digested with various RNases like RNase A, T2, U etc.
- Digested $tRNA^{Ala}$ fragments were 5' end labeled and separated either 2-D homochromatography or ion exchange (DEAE cellulose) column chromatography.
- Fragments were then digested with nuclease to make mononucleotide. 5'end labelled mononucleotide was analyzed by TLC.

DNA SEQUENCING

Holley's methods





DEAE column (ion exchange)

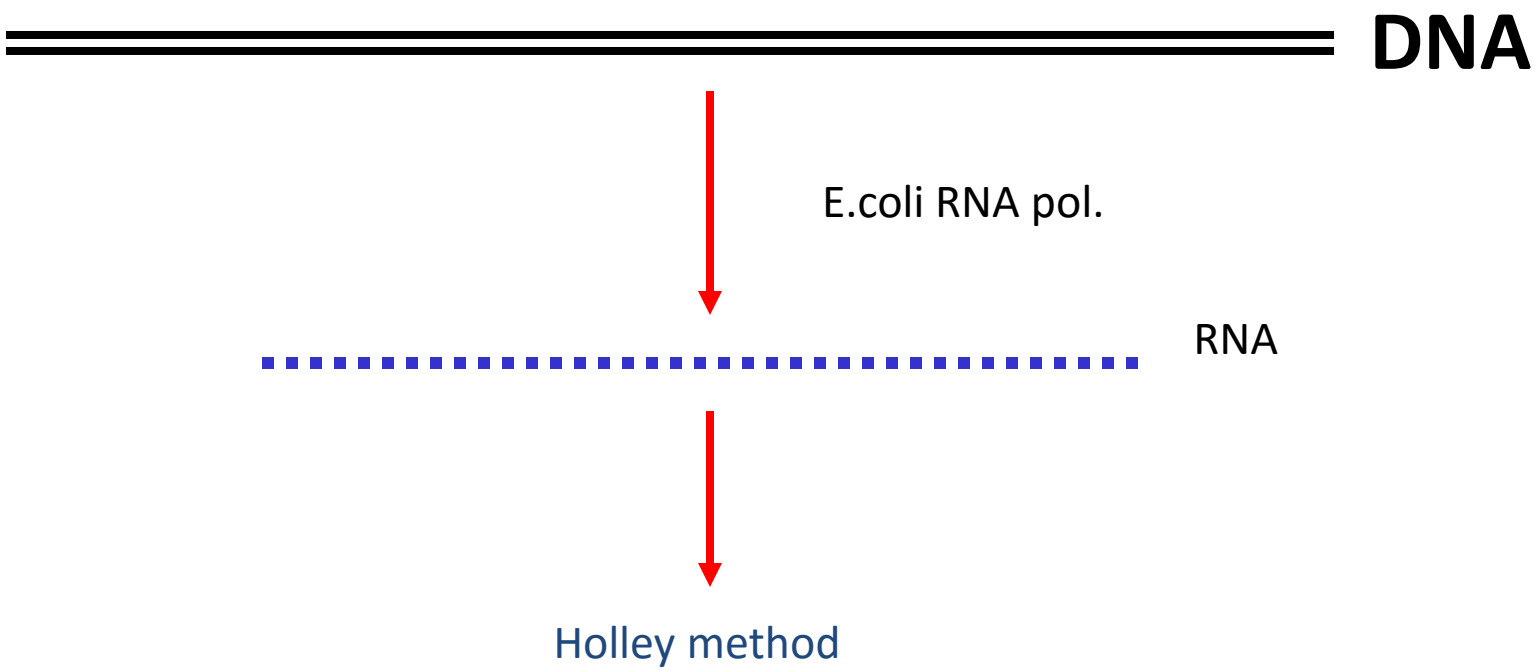
Separated fragments were 5' end labeled with ^{32}P then digested with nuclease and analyzed by TLC.

DNA SEQUENCING

Same type of method was tried to attempt for DNA sequencing by cleaving DNA into small fragment with “endonuclease IV or chemical like pyrimidine.

DNA SEQUENCING

Second type of method was transcribe “the DNA fragment” into RNA by E.coli RNA pol then go for Holley method.



DNA SEQUENCING

Sanger & Gilbert

- In late 1970 there were two groups developed 2 different methods of DNA sequencing

✓ F.Sanger & AR Coulson in Cambridge England.

✓ W.Gilbert in Harvard university in USA

DNA SEQUENCING

- Sanger's lab used enzymes for synthesis
- Gilbert's lab used chemical for degradation.

both method has its own cons and pros…

but generally Sanger's method is widely used b'coz it is relatively easy compare to Gilbert's methods.

DNA SEQUENCING

SANGER-COULSON SEQUENCING Method

- The method is based on the **interruption** by nucleotide analogues (di-deoxy nucleotide) of enzymatic synthesis of a second strand of DNA complementary to the template DNA.

- A mixture of different length fragments is produced depending where the interruptions occurred.

- The mixture of fragments is run on a sequencing gel (8%-7M urea PAGE) and the sequence read off.

DNA SEQUENCING

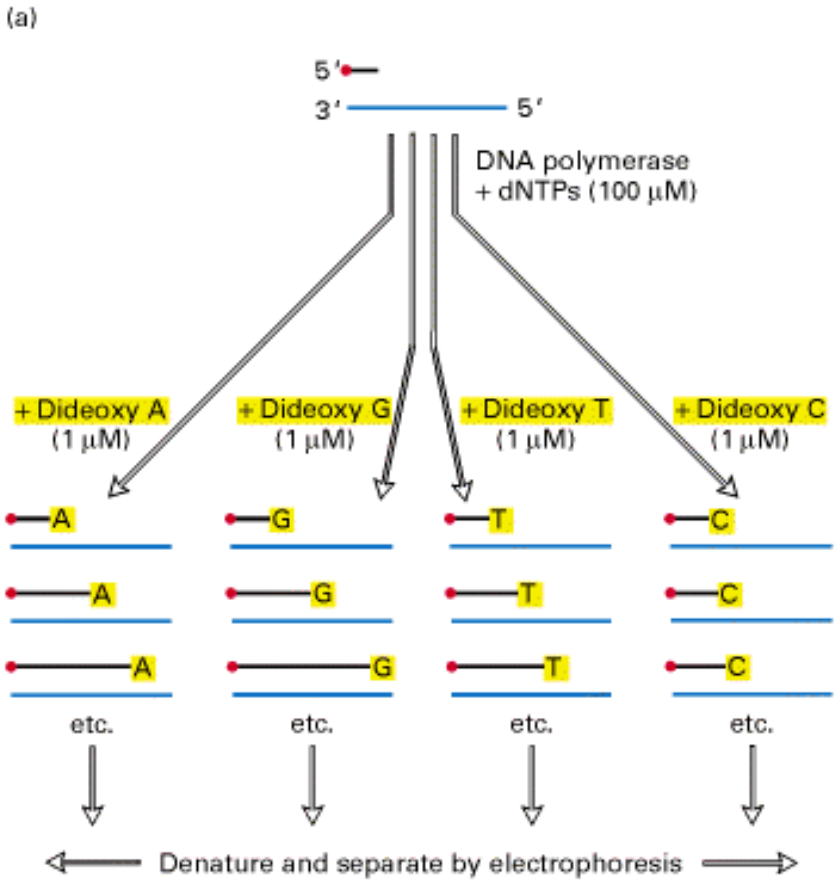
Chain-terminating event

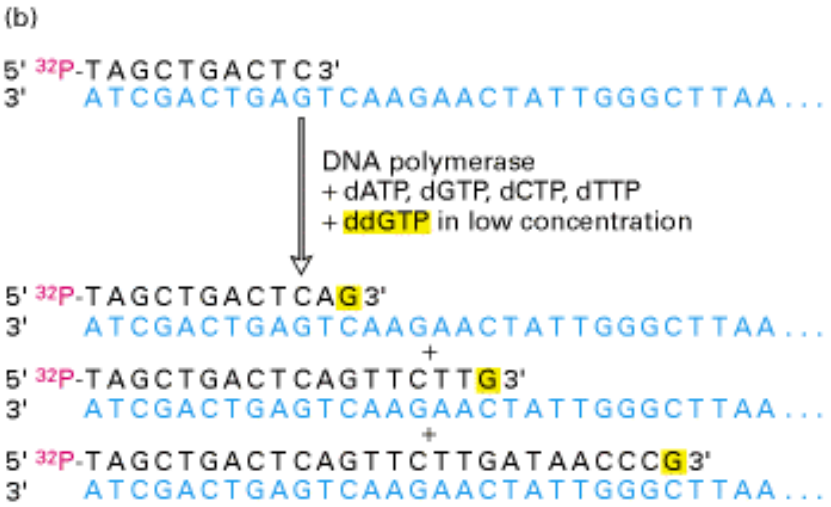
1. Before the DNA can be sequenced, it has to be denatured into single strands using heat or NaOH or cloned into M13 plasmid to make single strand DNA.
2. Next a primer (complementary to portion of template) is annealed to one of the template strands. This primer is specifically constructed so that its 3' end is located next to the DNA sequence of interest.
3. Either this primer or one of the nucleotides should be radioactively or fluorescently labeled so that the final product can be detected on a gel.
4. Once the primer is attached to the DNA, the solution is divided into four tubes labeled "G", "A", "T" and "C". Then reagents are added to these samples as follows:

    "G" tube: all four dNTP's, ddGTP (1/100th) and DNA polymerase
    "A" tube: all four dNTP's, ddATP (1/100th) and DNA polymerase
    "T" tube: all four dNTP's, ddTTP (1/100th) and DNA polymerase
    "C" tube: all four dNTP's, ddCTP (1/100th) and DNA polymerase

5. As the DNA is synthesized, nucleotides are added on to the growing chain by the DNA polymerase. However, on occasion a dideoxynucleotide is incorporated into the chain in place of a normal nucleotide, which results in a chain-terminating event.

DNA SEQUENCING





DNA SEQUENCING

SANGER's Dideoxy chain termination method

- Sangers technique is

"controlled DNA polymerase activity to generate different size of DNA length as little as one nucleotide in a set of independent, nucleotide specific reactions"

Primer– 10nt

2nd fragment is primer+1nt

3rd fragment is primer +2nt

DNA pol are the enzymes that make DNA

An enzyme that can synthesize a new DNA strand on a template strand (replication). Any enzyme involved in DNA repair or both is called DNA pol.

Both prokary- Eukaryotic DNA pol share the same fundamental sythetic activity.

DNA pol enzymes involved in two type of functions

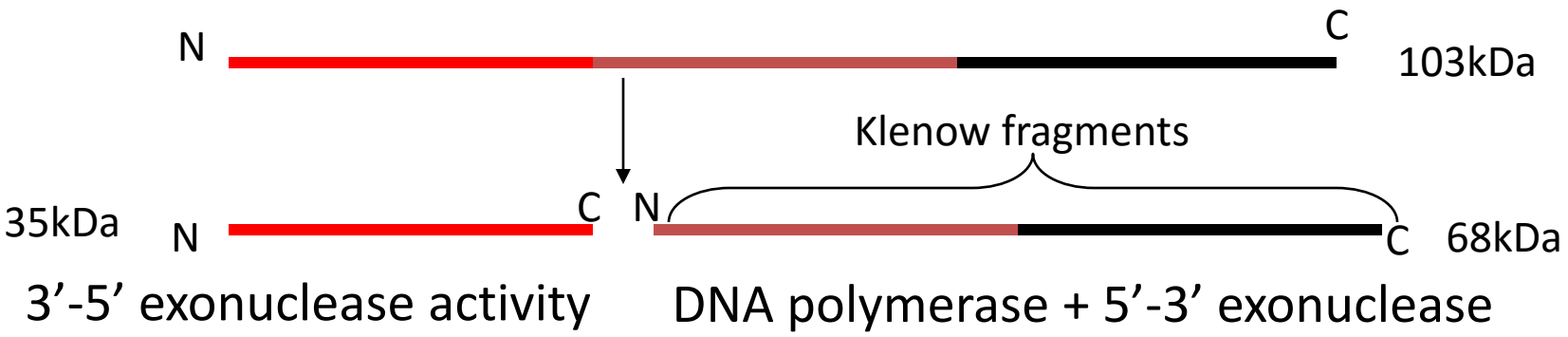
– Replication call DNA replicases

– DNA repair

DNA Pol I

- Mole. Wt of DNA pol I – 103kDa
- Proteolytic cleave yields two fragments
  - N-terminal 35kDa (1/3) fragments which contains 3’-5’ exonuclease activity.
  - C-terminal 68kDa (2/3) fragments called Klenow fragments

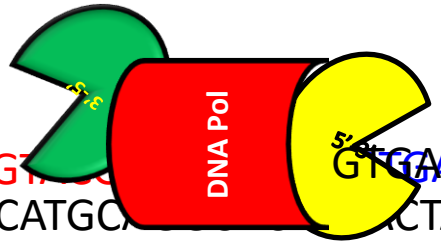
DNA pol are the enzymes that make DNA



DNA Pol I Polymerase activity

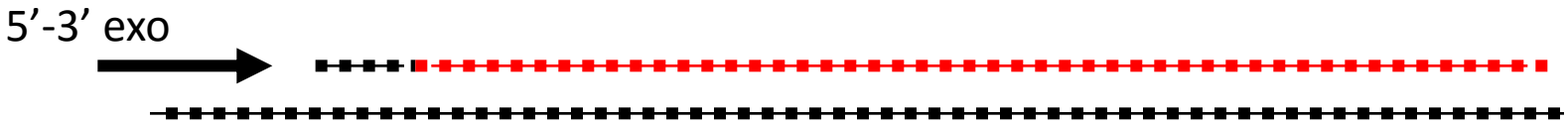
ACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGCT
TGACGCATGCAGGCACTGCACTAGCTGCACGTTGGCCAATCGA

DNA Pol I 5'→ 3' exonuclease activity



ACTGCGT
GTGACGTGATCGACGTGCAACCGGTT
TGACGCATGC
CTAGCTGCACGTTGGCCAATCGA

DNA Pol I 5'→ 3' exonuclease activity



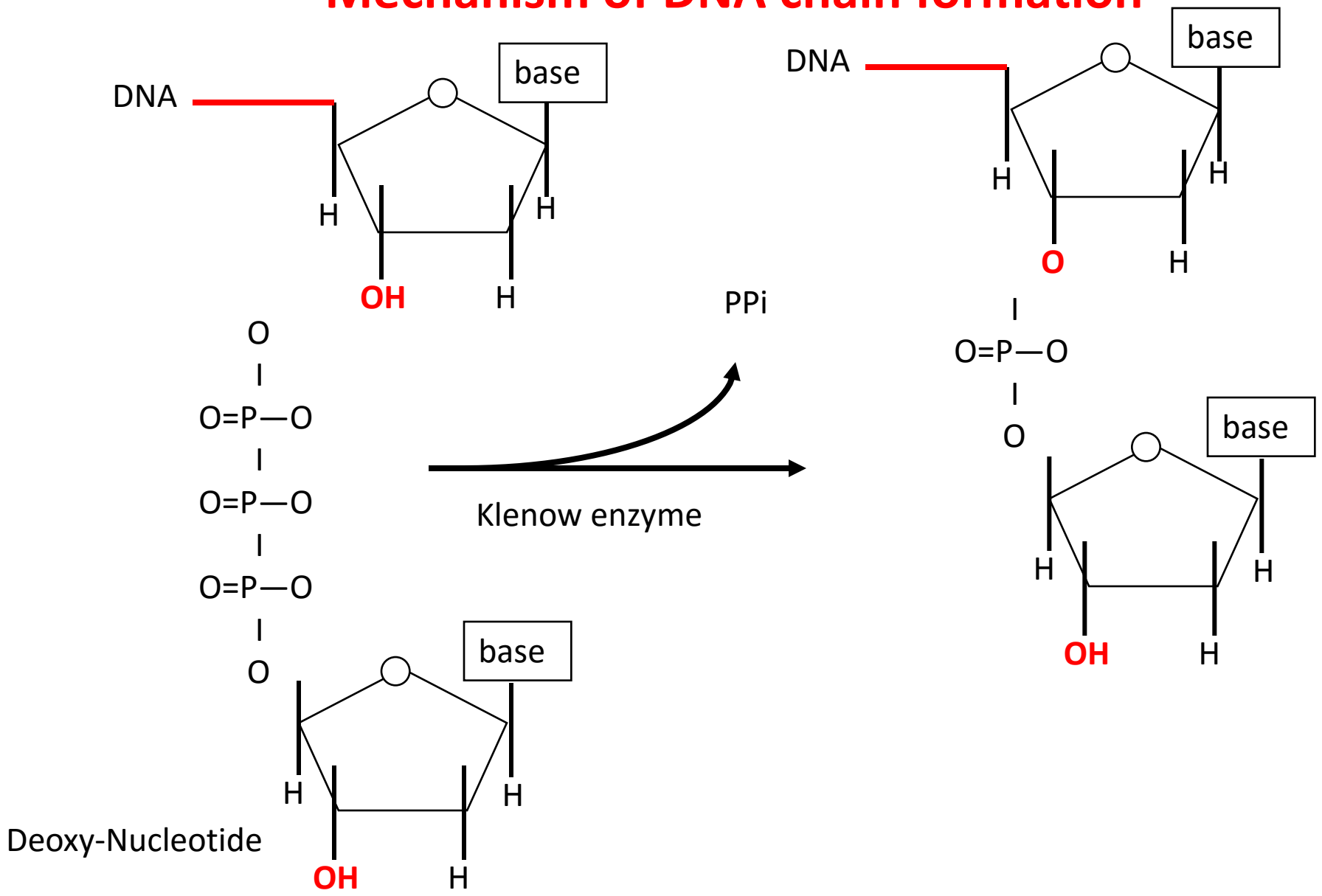
Because of 5'-3' exonuclease activity daughter strand is going to different size this will not give DNA fragments.

DNA Pol I 3'→ 5' exonuclease activity

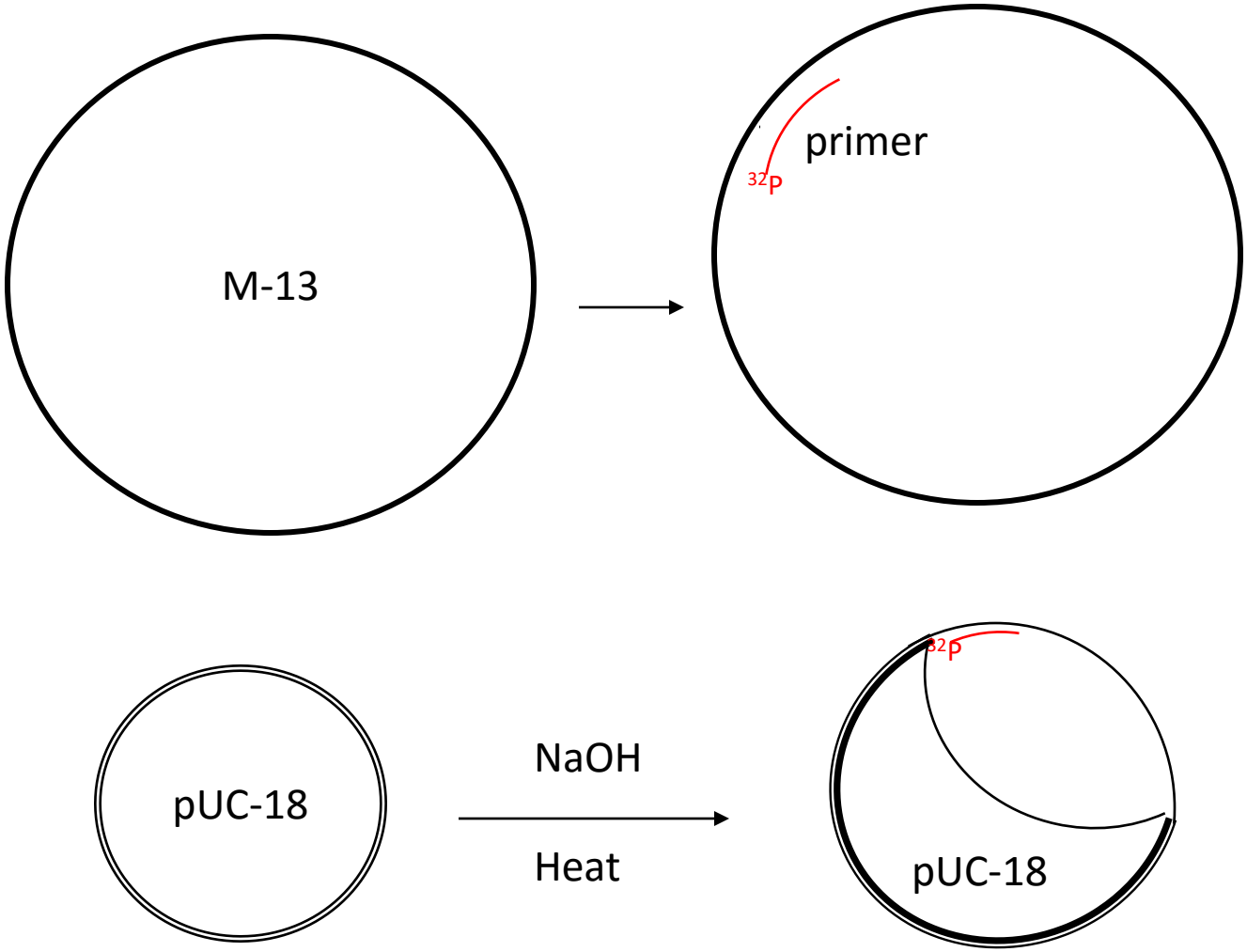
3'→ 5' exonuclease will ruin the sequencing reaction.

3'→ 5' exonuclease sense incorporated ddNMP is not correct nucleotide
So it removes this ddNMP as proof reading effect.

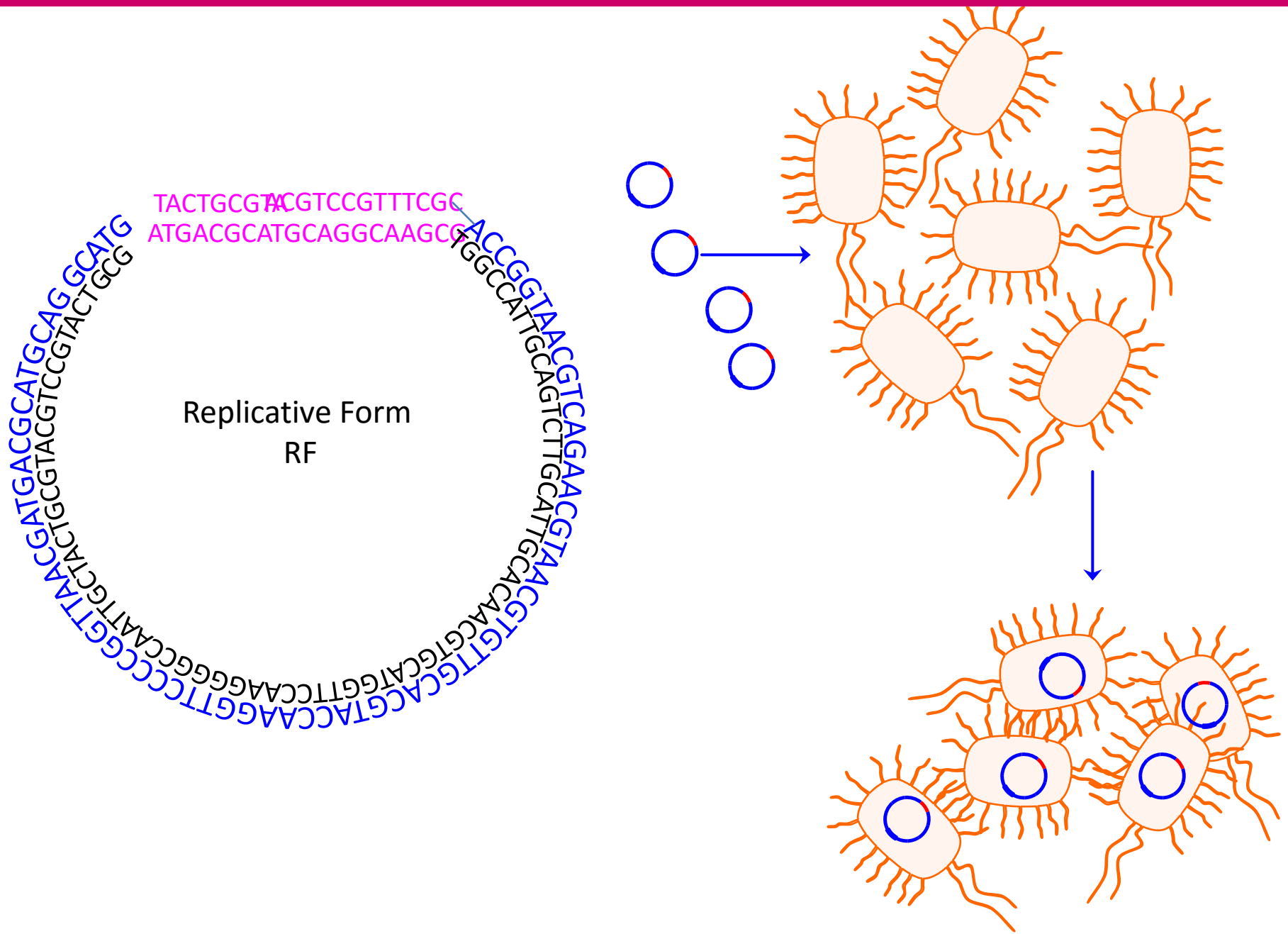
Mechanism of DNA chain formation



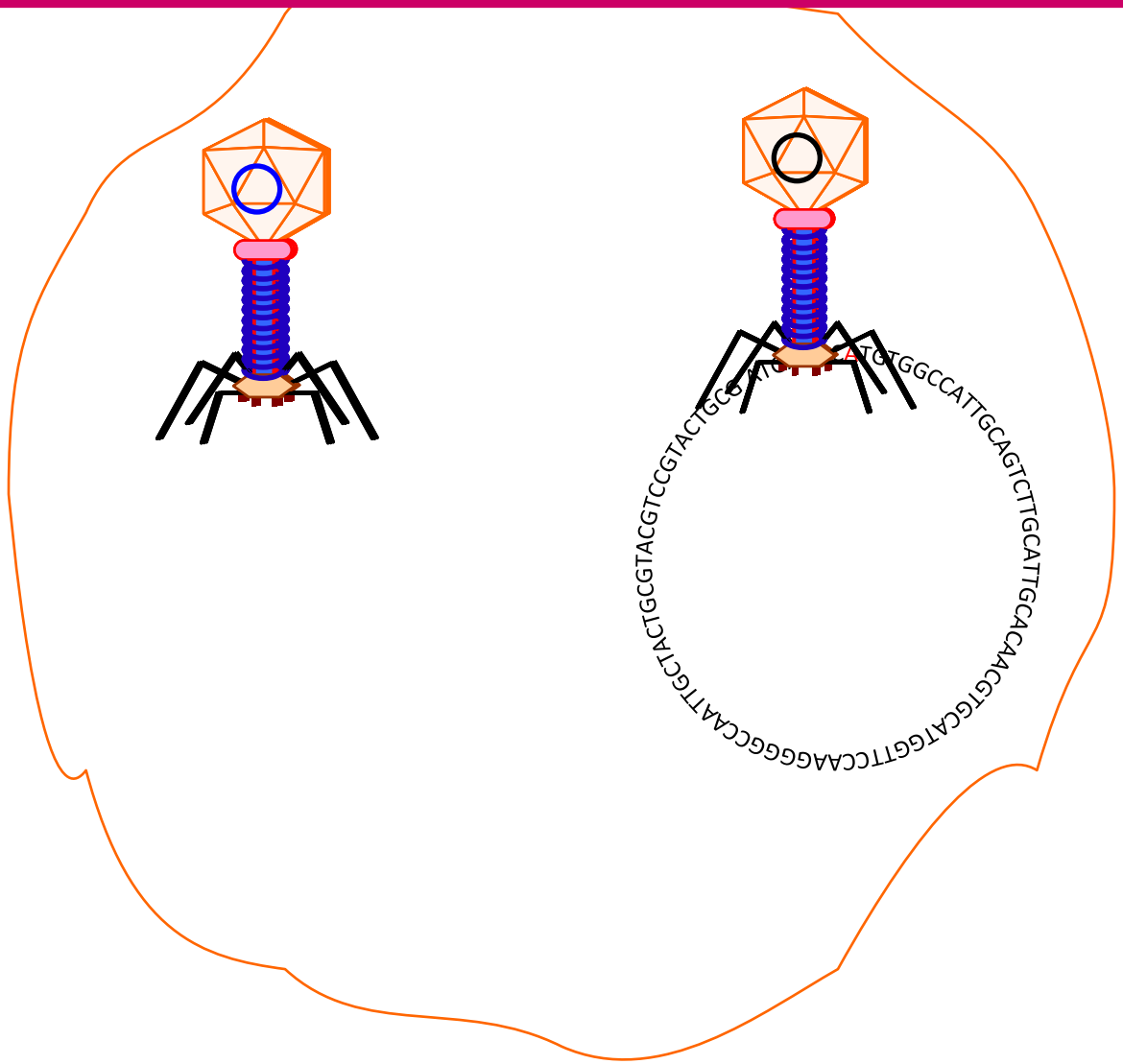
Making Single strand DNA template



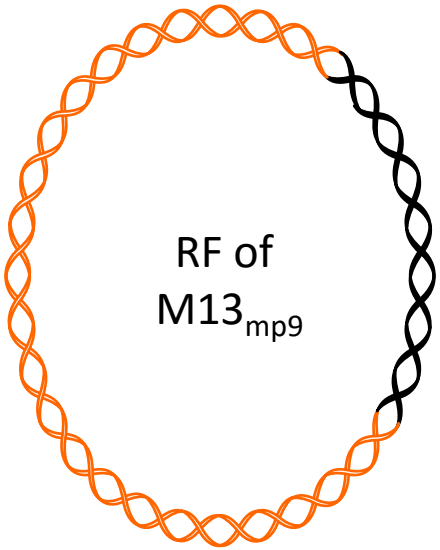
Making Single strand DNA template

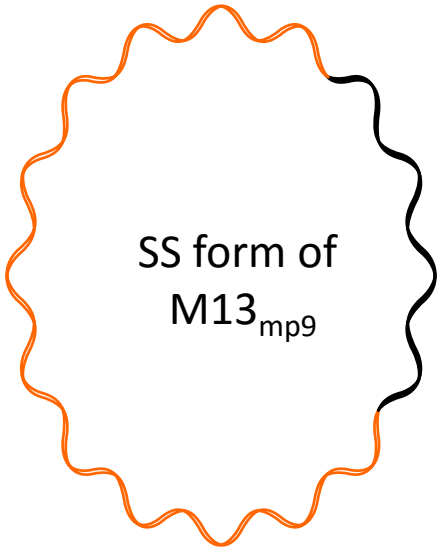


Making Single strand DNA template



Making Single strand DNA template

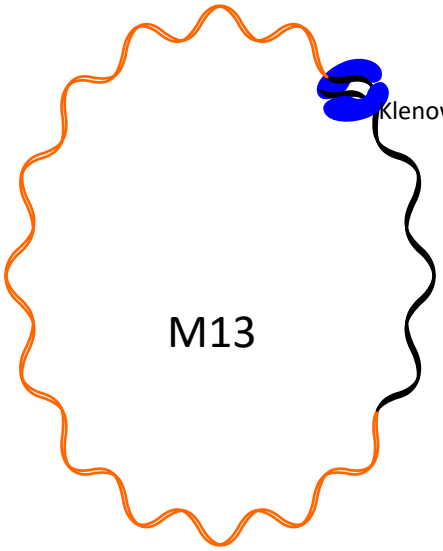




Sanger's Method: Molecular Mechanism

Mechanism of DNA sequencing by Sanger's Method

ACTGCGTACGTACGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGCTGACGTACGTT
TGACGCATGCA TGACGCATGCAGGCACTGCACTAGCTGCACGTTGGCCAATCGACTGCATGCAA

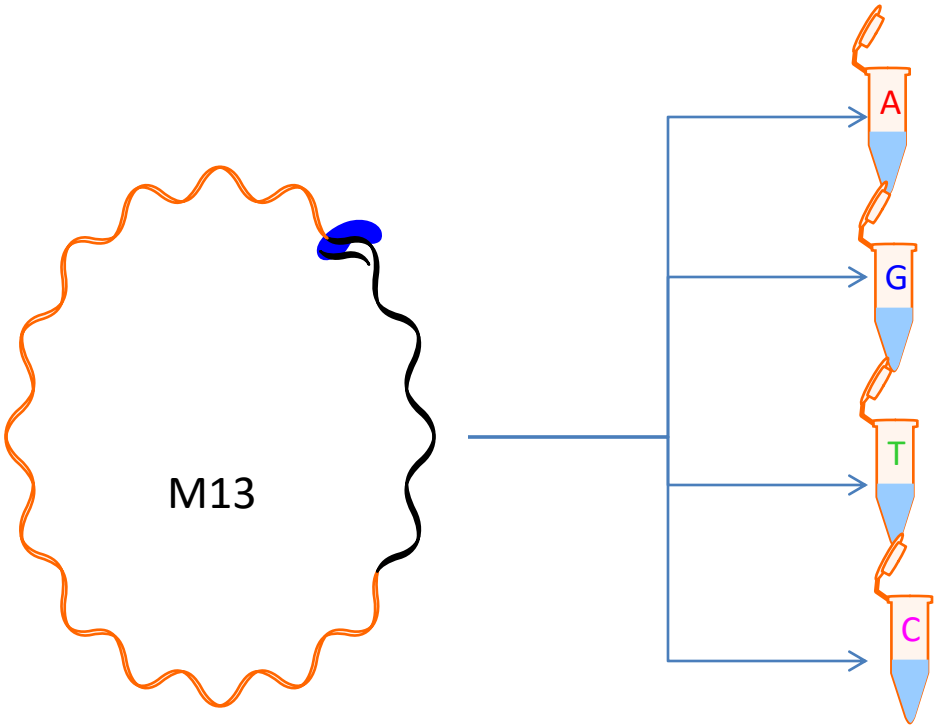


Primer base pair (hybridize or binding) with its complementary region in plasmid.

Primer is a short DNA nucleotide sequence called oligonucleotides simply called "Oligo"

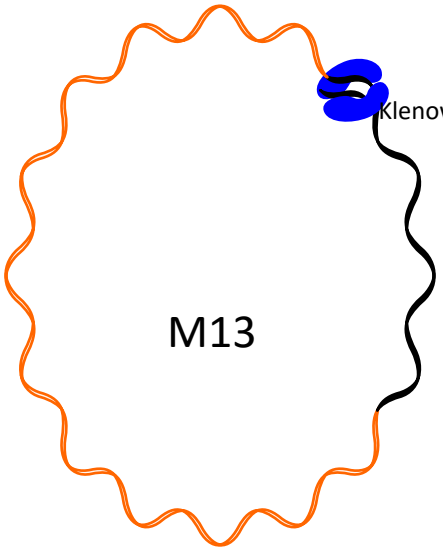
Base pairing primer with plasmid will make single strand plasmid DNA into double strand region where DNA polymerase can bind and can able to synthesis new DNA strand which is complementary to its lagging strand.

4 separate reactions



Mechanism of DNA sequencing by Sanger's Method

ACTGCGTACGTACGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGCTGACGTACGTT
TGACGCATGCA TGACGCATGCAGGCACTGCACTAGCTGCACGTTGGCCAATCGACTGCATGCAA

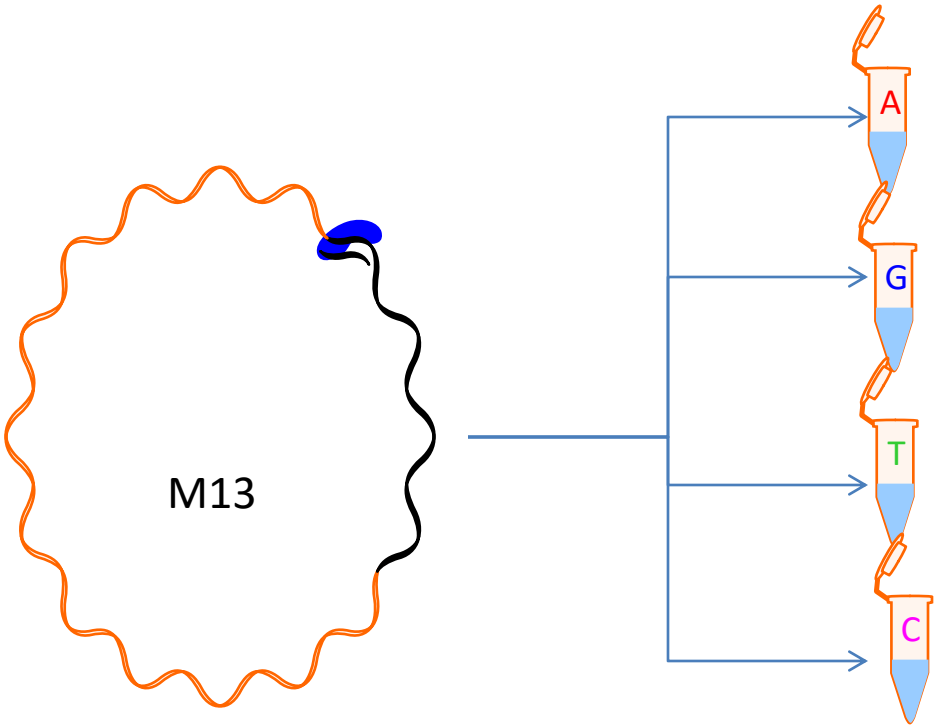


Primer base pair (hybridize or binding) with its complementary region in plasmid.

Primer is a short DNA nucleotide sequence called oligonucleotides simply called "Oligo"

Base pairing primer with plasmid will make single strand plasmid DNA into double strand region where DNA polymerase can bind and can able to synthesis new DNA strand which is complementary to its lagging strand.

4 separate reactions



Tube-A

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGCTGA
TGACGCATGCAGGCA TG**A**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGCTGAC
TGACGCATGCAGGCA TGACGC**A**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAT
TGACGCATGCAGGCA TGACGCATGC**A**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAT
TGACGCATGCAGGCA TGACGCATGCAGGC**A**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAT
TGACGCATGCAGGCA TGACGCATGCAGGCACTGC**A**$_{dd}$

When DNA polymerase or sequenase encounter dideoxyribonucleotide (ddNTPs)
New DNA synthesis cannot happened because of lack of 3'OH moiety in ddNTP

Tube-G

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGC
TGACGCATGCAGGCA T**G**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAG
TGACGCATGCAGGCA TGAC**G**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTA
TGACGCATGCAGGCA TGACGCAT**G**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTA
TGACGCATGCAGGCA TGACGCATGCA**G**$_{dd}$

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTA
TGACGCATGCAGGCA TGACGCATGCAG**G**$_{dd}$

Tube-T

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGC
TGACGCATGCAGGCA T_{dd}

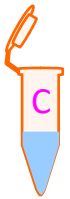
ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACTACGTT
TGACGCATGCAGGCA TGACGCAT_{dd}

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAT
TGACGCATGCAGGCA TGACGCATGCAGGCACTdd

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAT
TGACGCATGCAGGCA TGACGCATGCAGGCACTGCACT_{dd}

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAT
TGACGCATGCAGGCA TGACGCATGCAGGCACTGCACTAGCT_{dd}

Tube-C



ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAGC
TGACGCATGCAGGCA T GAC_{dd}

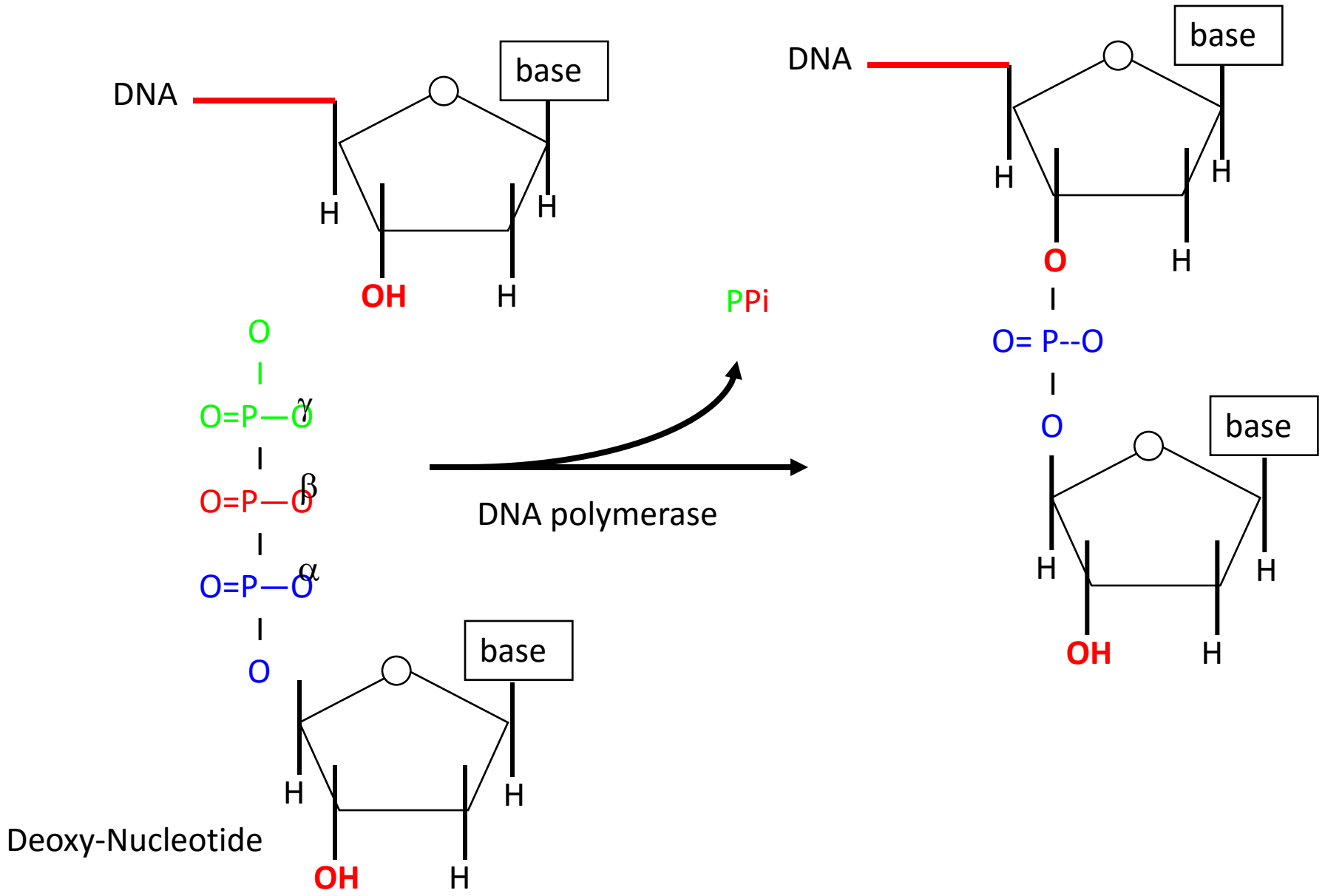
ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAG
TGACGCATGCAGGCA TGACGC_{dd}

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTA
TGACGCATGCAGGCA TGACGCATGC_{dd}

ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTAT
TGACGCATGCAGGCA TGACGCATGCAGGC_{dd}

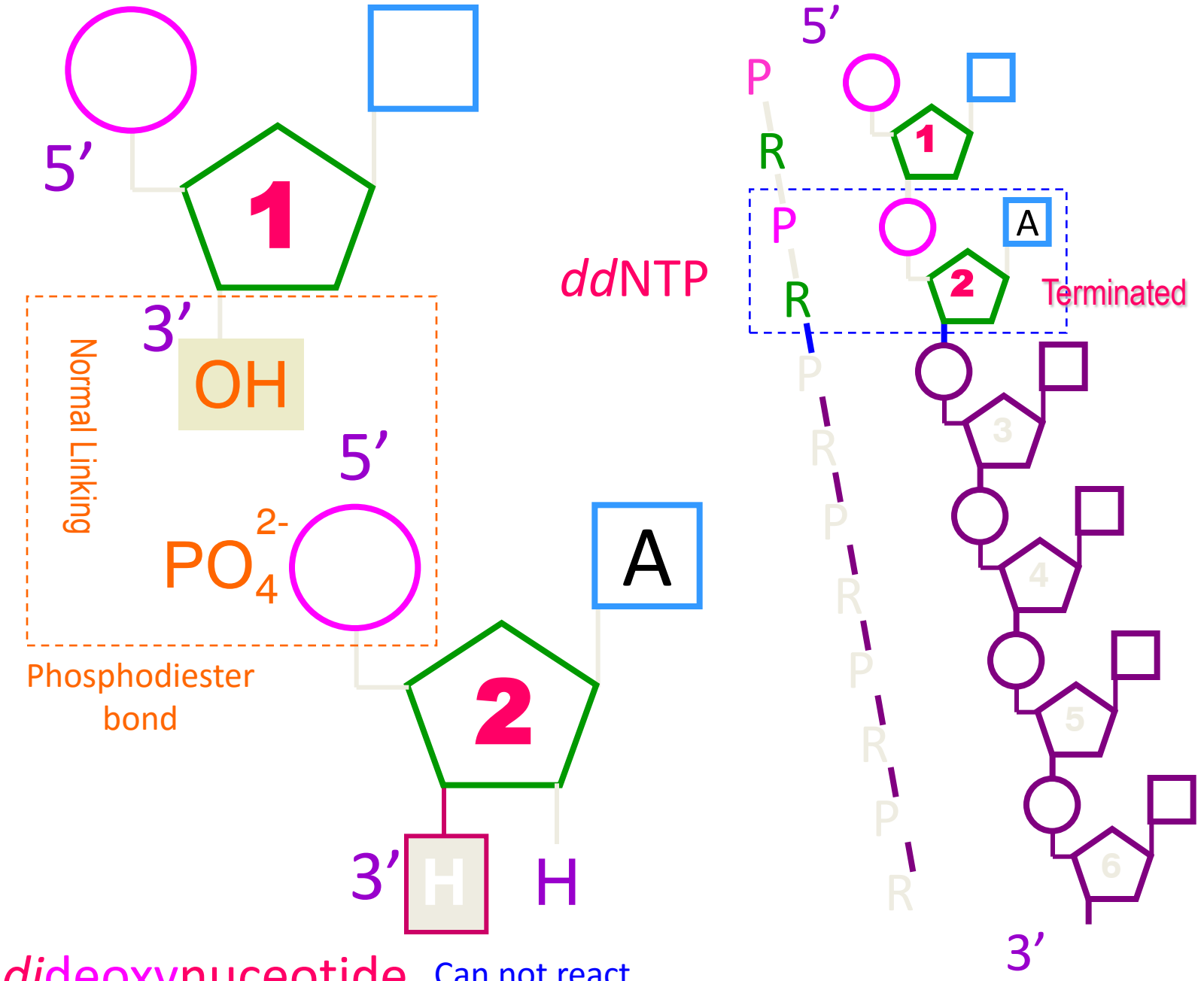
ACTGCGTACGTCCGTACTGCGTACGTCCGTGACGTGATCGACGTGCAACCGGTTA
TGACGCATGCAGGCA TGACGCATGCAGGCACTGC_{dd}

Mechanism of DNA chain formation



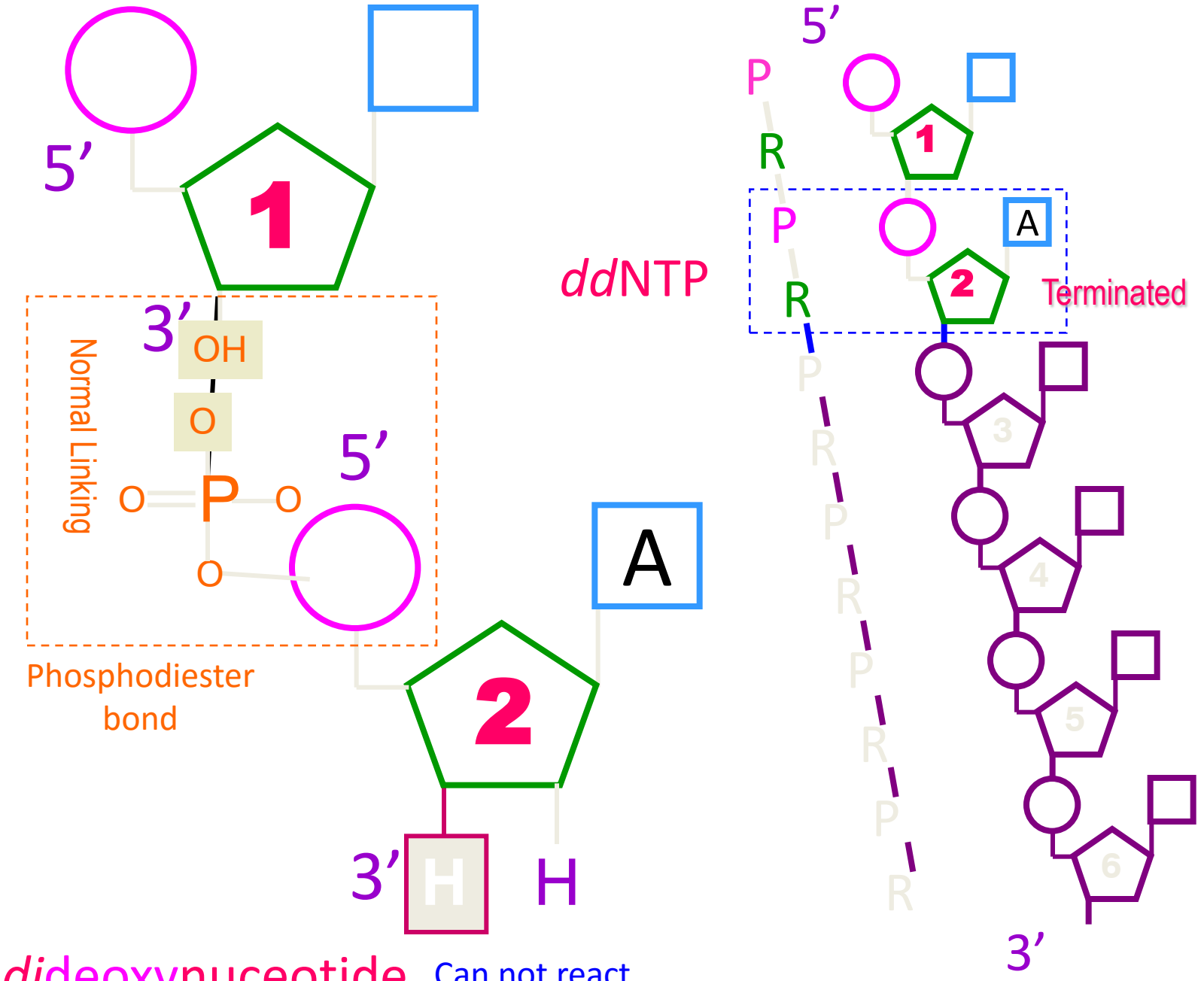
Deoxy-Nucleotide

Sanger's Method: Molecular Mechanism



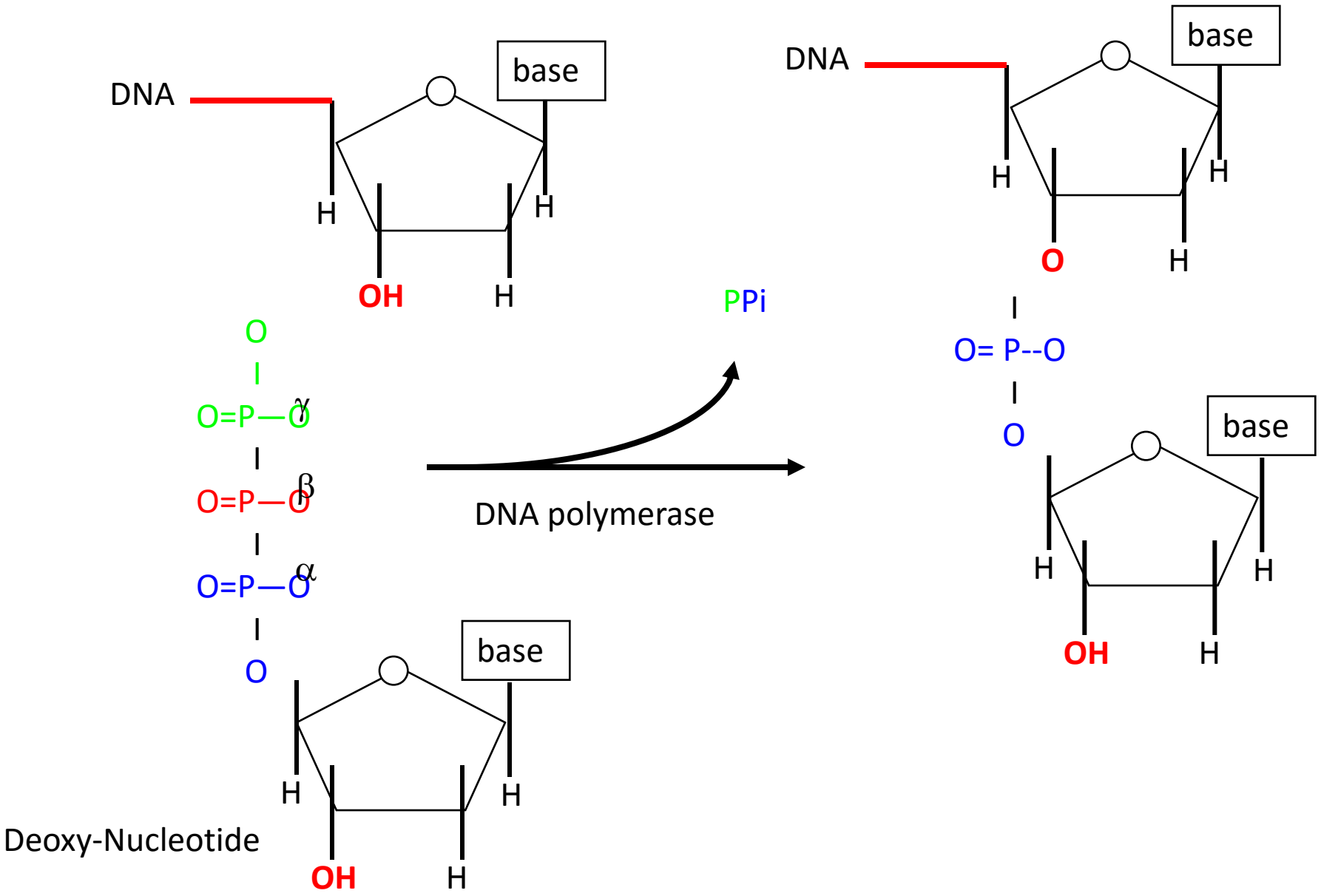
Juang RH (2004) BCbasics

Sanger's Method: How Synthesized

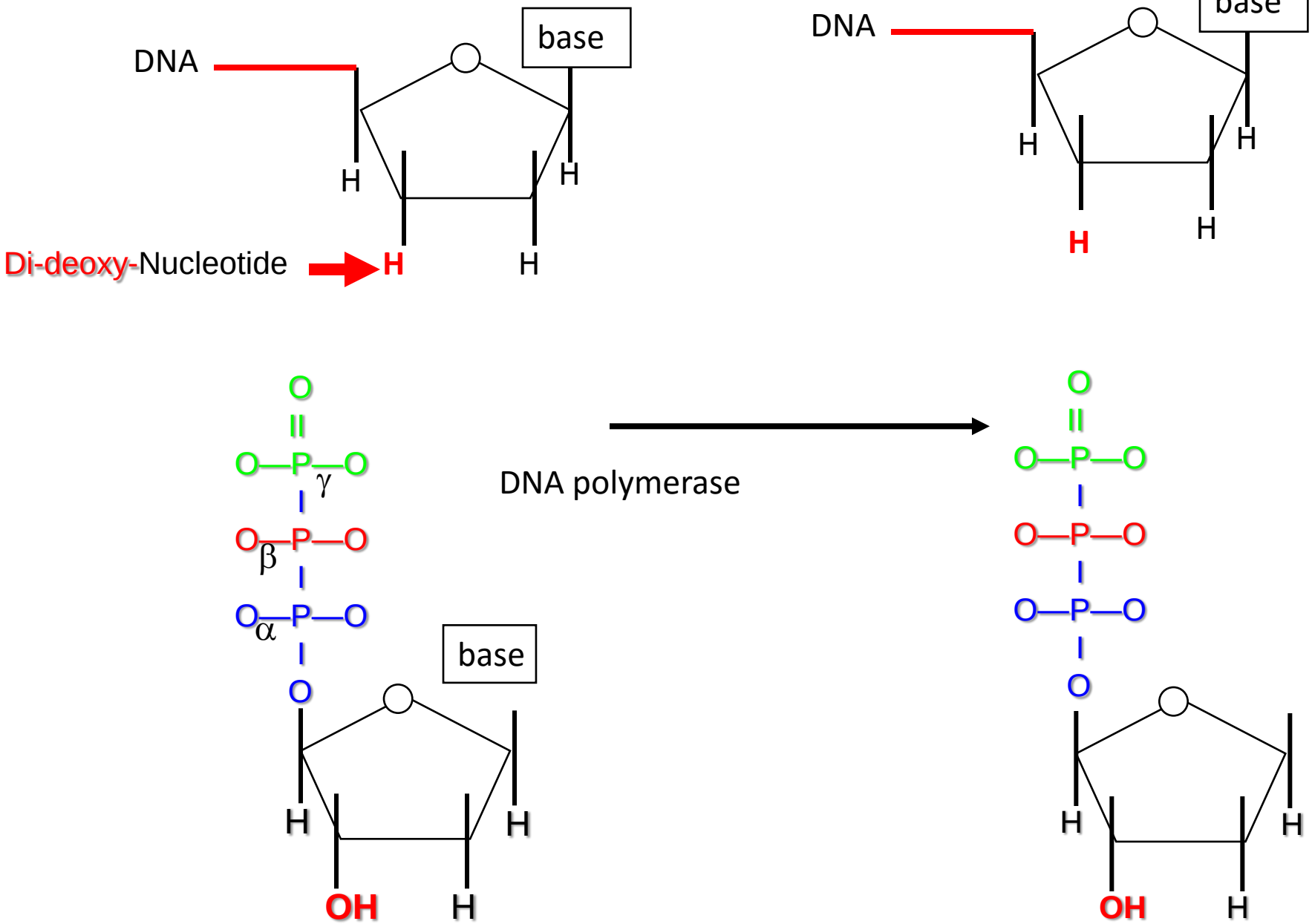


Juang RH (2004) BCbasics

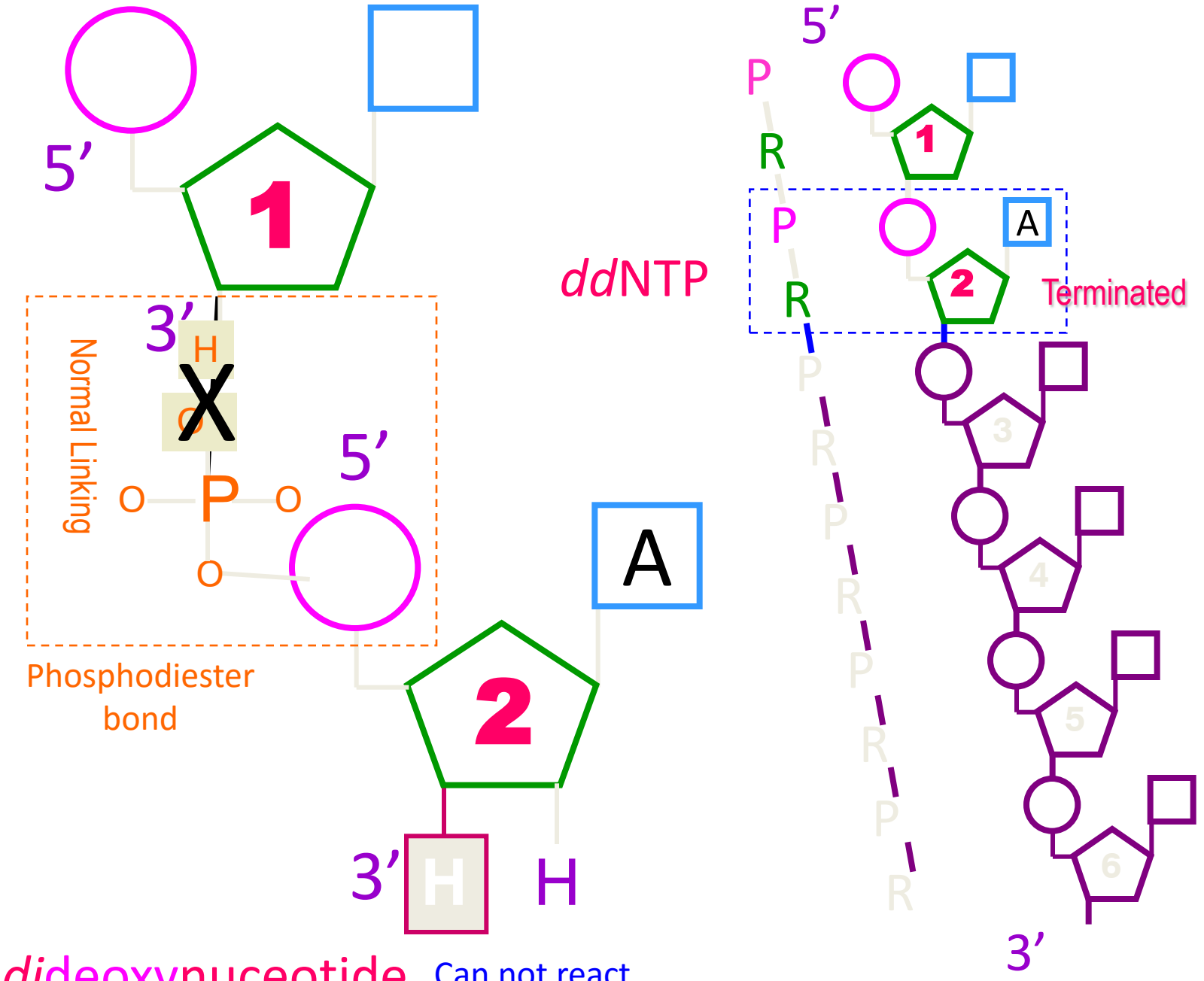
Mechanism of DNA chain formation



Mechanism of Dideoxy chain termination



Sanger's Method: How Terminated



Juang RH (2004) BCbasics

Tube-A

^{32}P-TGACGCATGCAGGCATGA_{dd}

^{32}P-TGACGCATGCAGGCATGACGCA_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCA_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCAGGCA_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCAGGCACTGCA_{dd}

Tube-G

^{32}P-TGACGCATGCAGGCATGG_{dd}

^{32}P-TGACGCATGCAGGCATGACG_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATG_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCAG_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCAGG_{dd}

Tube-C

^{32}P-TGACGCATGCAGGCAT GAC_{dd}

^{32}P-TGACGCATGCAGGCATGACGC_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGC_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCAGGC_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCAGGCACTGC_{dd}

Tube-T

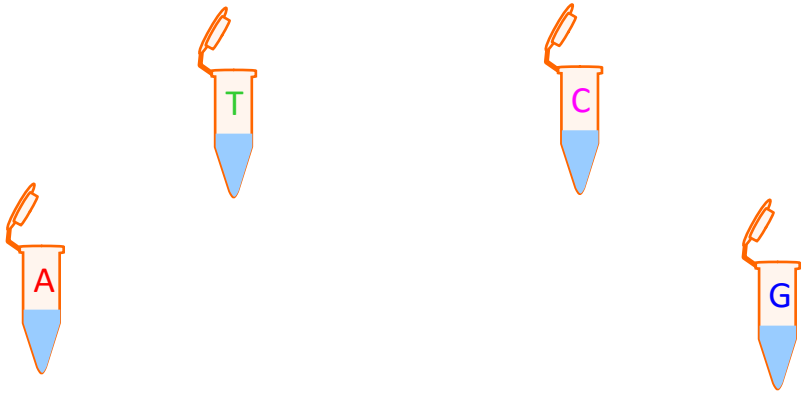
^{32}P-TGACGCATGCAGGCAT_{dd}

^{32}P-TGACGCATGCAGGCATGACGCAT_{dd}

^{32}P-TGACGCATGCAGGCATGACGCATGCAGGCACTdd

^{32}P-TGACGCATGCAGGCATGACGCATGCAGGCACTGCACT_{dd}

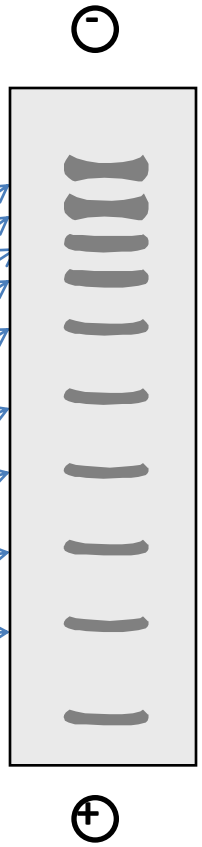
^{32}P-TGACGCATGCAGGCATGACGCATGCAGGCACTGCACTAGCT_{dd}



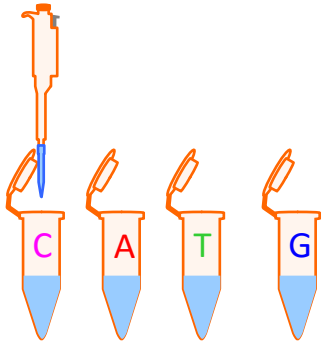
How DNA Sequence Is Determined?

How DNA Sequence Is Determined?

TGACGCATGCAGGCA TGACGCATGCAGGCACTGCACT_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGCACTGCAC_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGCACTGCA_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGCACTGC_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGCACTG_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGCACT_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGCAC_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGCA_{dd}
TGACGCATGCAGGCA TGACGCATGCAGGC_{dd}
TGACGCATGCAGGCA TGACGCATGCAGG_{dd}
TGACGCATGCAGGCA TGACGCATGCAG_{dd}
TGACGCATGCAGGCA TGACGCATGCA_{dd}
TGACGCATGCAGGCA TGACGCATGC_{dd}
TGACGCATGCAGGCA TGACGCATG_{dd}
TGACGCATGCAGGCA TGACGCAT_{dd}
TGACGCATGCAGGCA TGACGCA_{dd}
TGACGCATGCAGGCA TGACGC_{dd}
TGACGCATGCAGGCA TGACG_{dd}
TGACGCATGCAGGCA T GAC_{dd}
TGACGCATGCAGGCA TGA_{dd}
TGACGCATGCAGGCA TG_{dd}
TGACGCATGCAGGCA T_{dd}



If you load all the sequence reac
In one lane, you cannot determin
the sequence of the DNA



Gel electrophoresis techniques for separating DNA molecules by size.

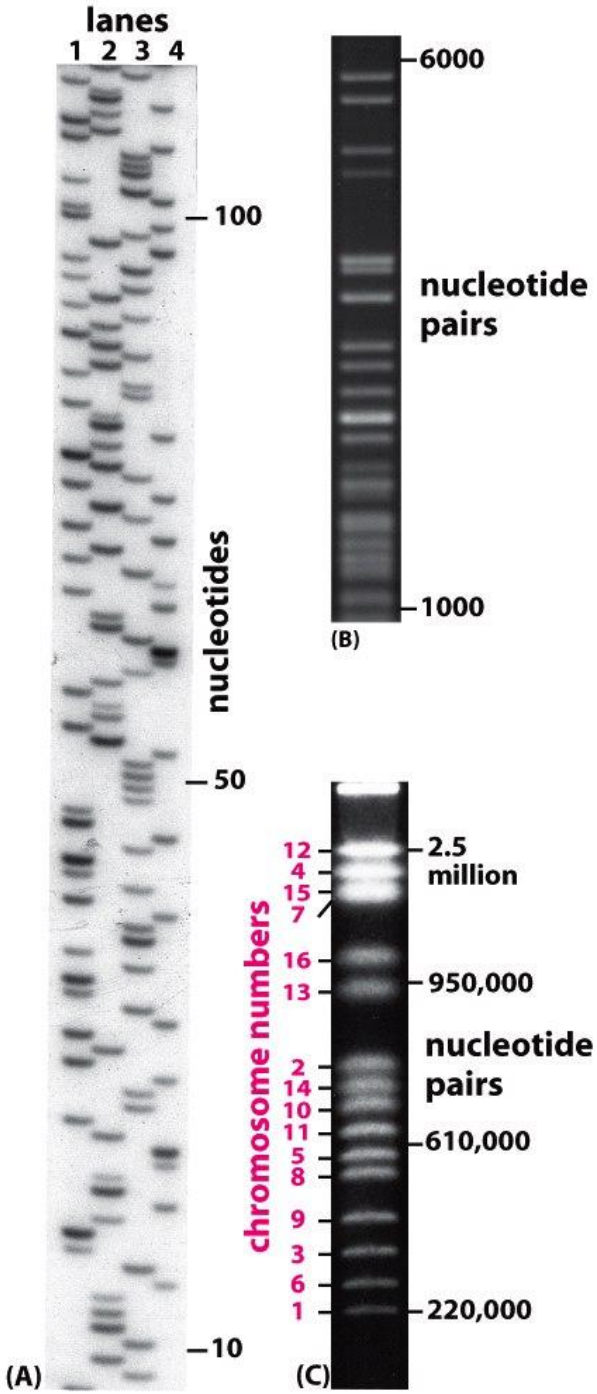
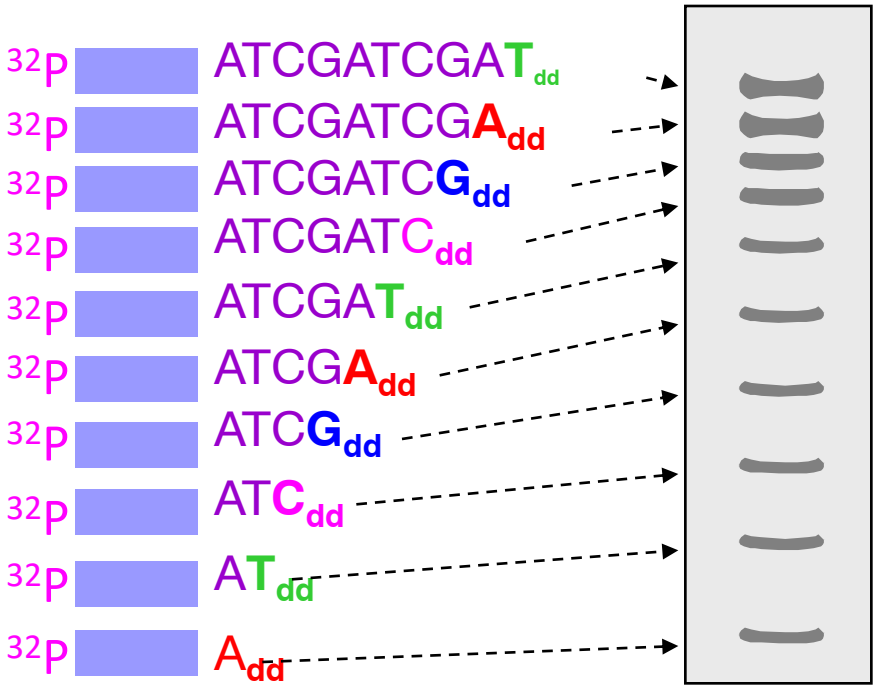
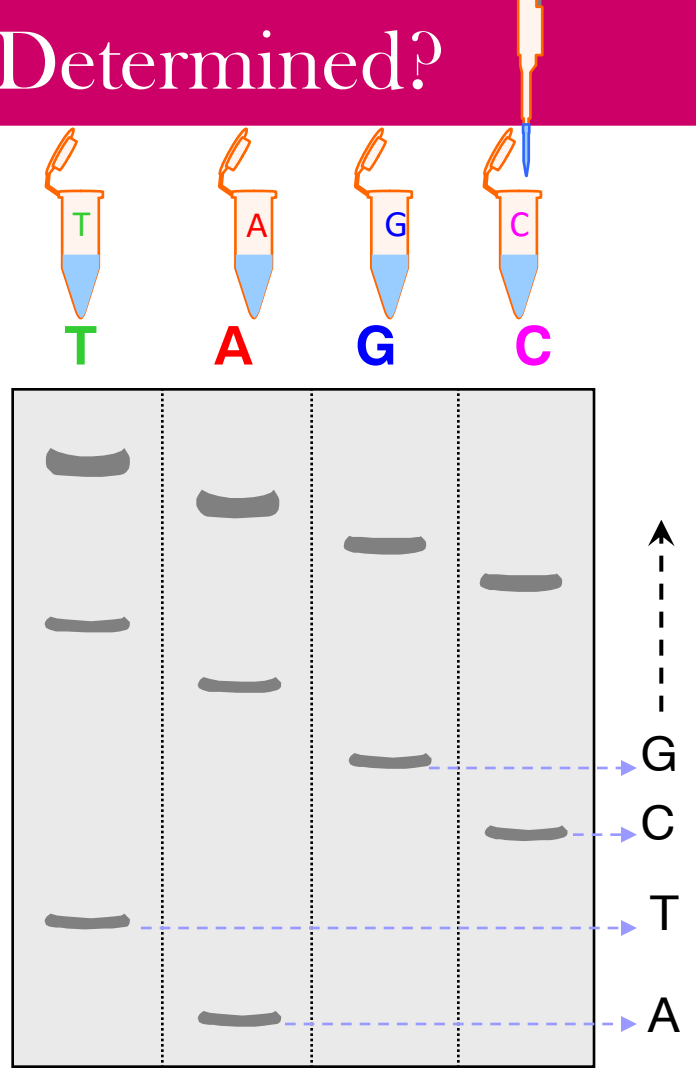


Figure 8-33 Molecular Biology of the Cell 5/e (© Garland Science 2008)

How DNA Sequence Is Determined?

DNA fragments having a difference of one nucleotide can be separated on gel electrophoresis





But these bands can't tell us the identity of the terminal nucleotides

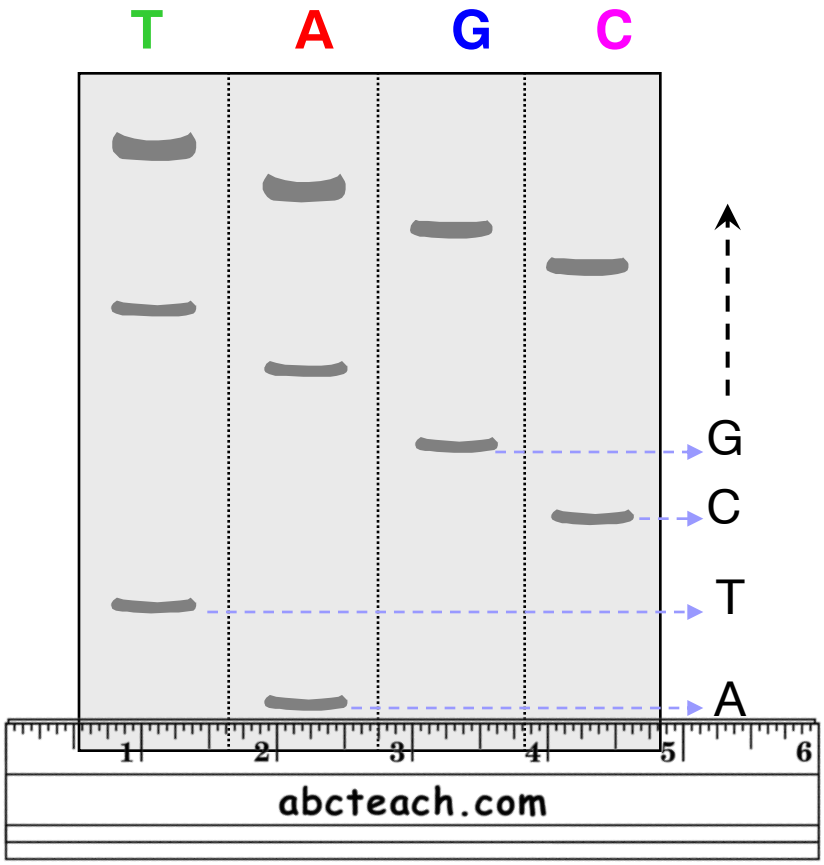
If those band with the same terminal nucleotide can be grouped, then it is possible to read the whole sequence

Polyacrylamide Gel Electrophoresis

Juang RH (2004) BCbasics

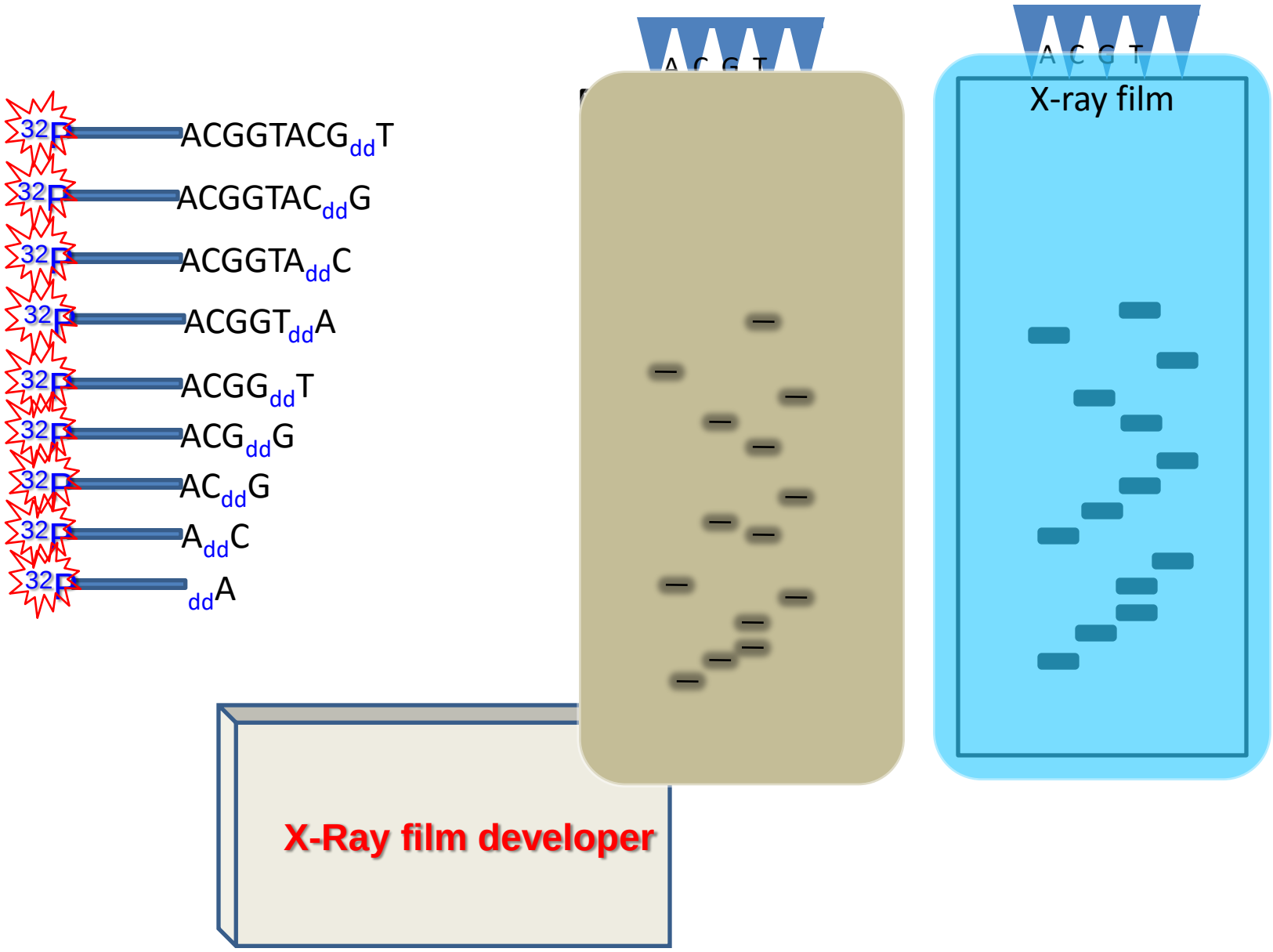
How DNA Sequence Is Determined?

Polyacrylamide Gel Electrophoresis

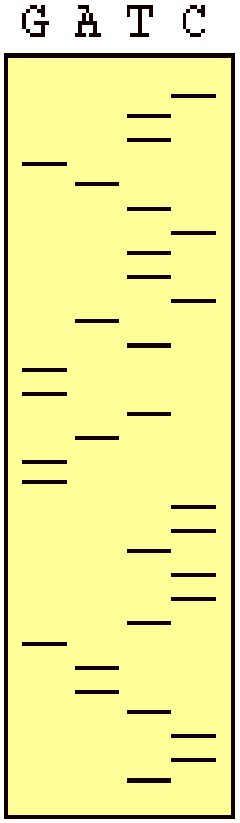


Reading sequence from Autoradiogram by manually

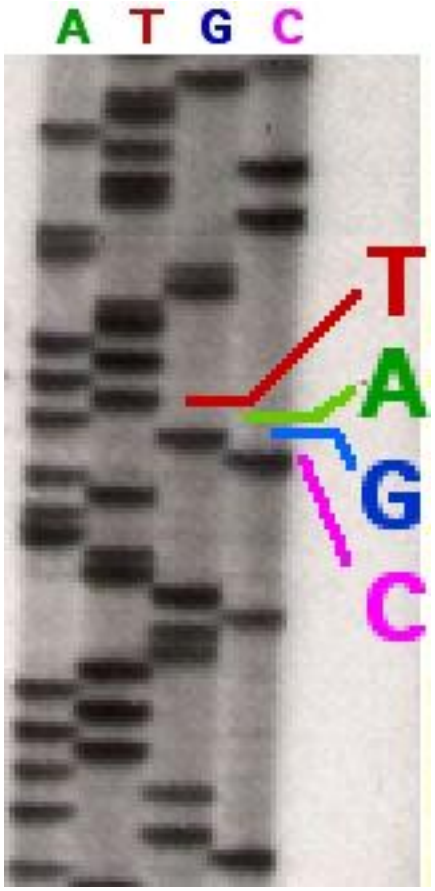
Autoradiogram



Autoradiogram



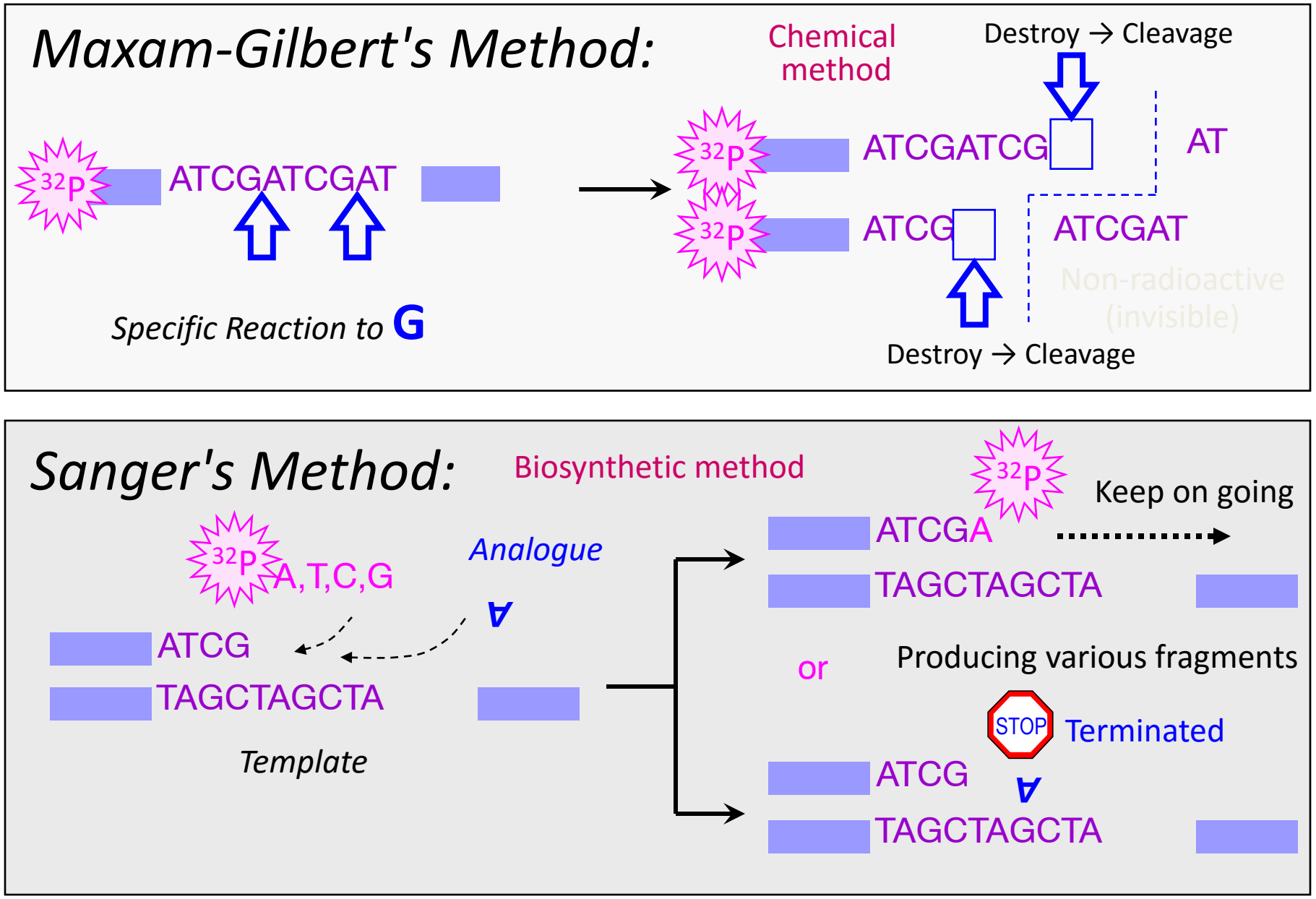
Line up all four reactions, and you can "read" the sequence ladder 5' to 3' as TCCTAAG...etc. in this example:



MAXAM-GILBERT SEQUENCING

- A method of DNA sequencing, based on the “controlled degradation of a DNA fragment in a set of independent, nucleotide specific reactions”
- The resulting fragments have characteristic sizes depending on the sequence of the template, that can be resolved on a sequencing gel.
- Although no longer the main protocol, Maxam Gilbert sequencing still has advantages, for example: oligonucleotides or covalently modified DNA sequencing.
- This chemical cleavage method uses double-stranded DNA samples and so does not require cloning of DNA into an M13 phage vector to produce single-stranded DNA as is the case with the Sanger-Coulson method. It involves modification of the bases in DNA followed by chemical base-specific cleavage.

How to Obtain DNA Fragments



Juang RH (2004) BCbasics

DNA SEQUENCING

1) Double-stranded (Ds) DNA to be sequenced is labelled by attaching a radioactive phosphorus (^{32}P) group to the 5' end.
2) Using dimethyl sulphoxide (DMSO) and heating to 90°C, the two strands of the DNA are separated and purified (*e.g.* using gel electrophoresis and the principle that one of the strands is likely to be heavier than the other due to the fact that it contains more purine nucleotides (A and G) than pyrimidines (C and T) which are lighter).
3) Single-stranded sample is split into separate samples and each is treated with one of the cleavage reagents.

   This part of the process involves **alteration** of bases (*e.g.* dimethylsulphate methylates guanine)

   Followed by **removal** of altered bases.

   Lastly, piperidine is used for **cleavage** of the strand at the points where bases are missing.

DNA SEQUENCING

Chemicals used in Maxam-Gilbert method

	alteration	base removal	cleavage
G	DMSO	Piperidine	Piperidine
A+G	Acid	Acid	Piperidine
C+T	Hydrazine	Piperidine	Piperidine
C	Hydrazine + 1.5M NaCl	Piperidine	Piperidine
A>C	1.2N NaOH	Piperidine	Piperidine

Methylation of N7 with DMSO (dimethyl sulfate or sulfoxide) make C8-C9 bond specifically susceptible to cleavage.

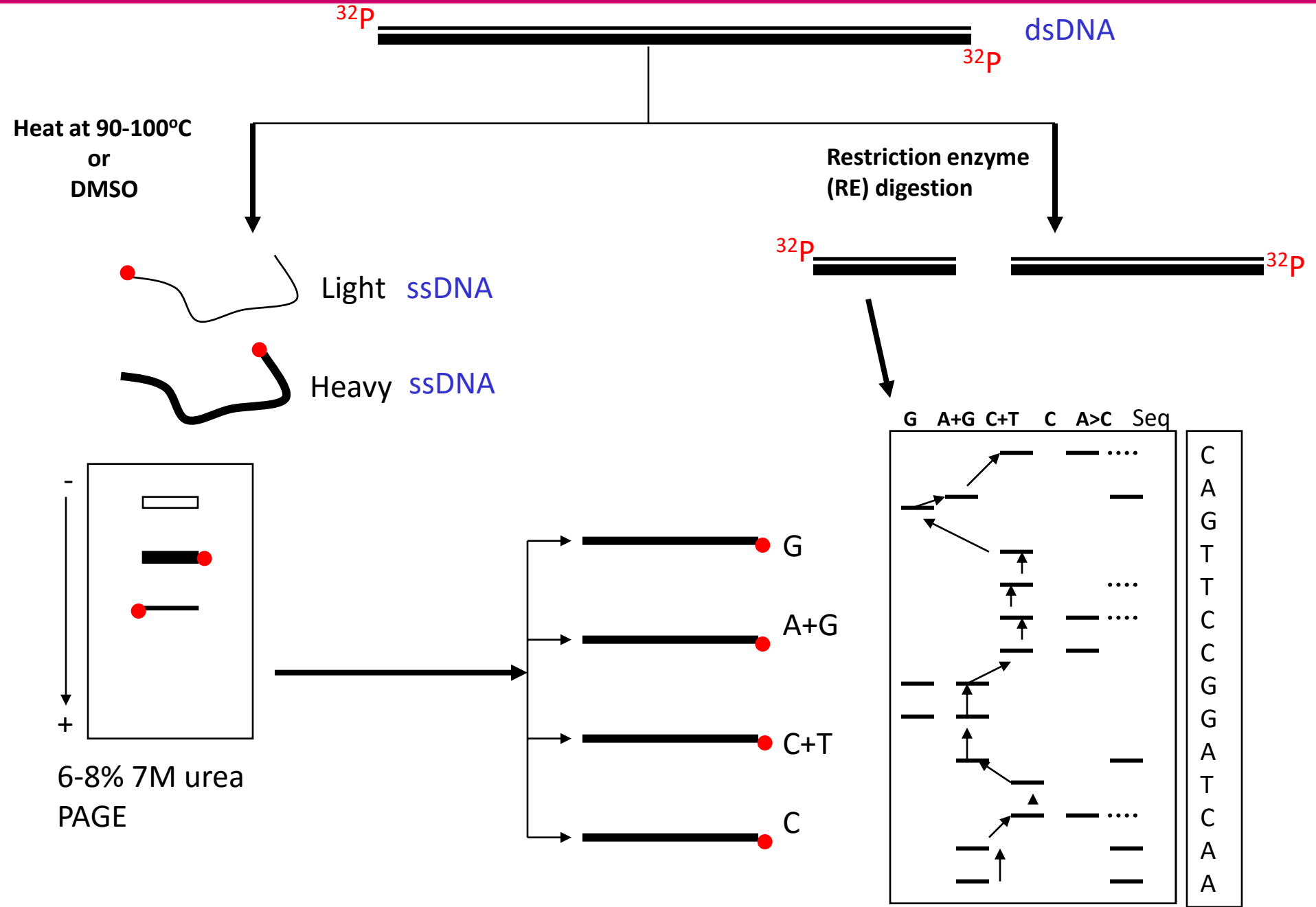
Piperidine formate weakens the glycosidic bond of purines by protonating N atoms

Hydrazine splits the pyrimidine rings

In the presence of 1.5M NaCl only cytosine reacts with hydrazine.

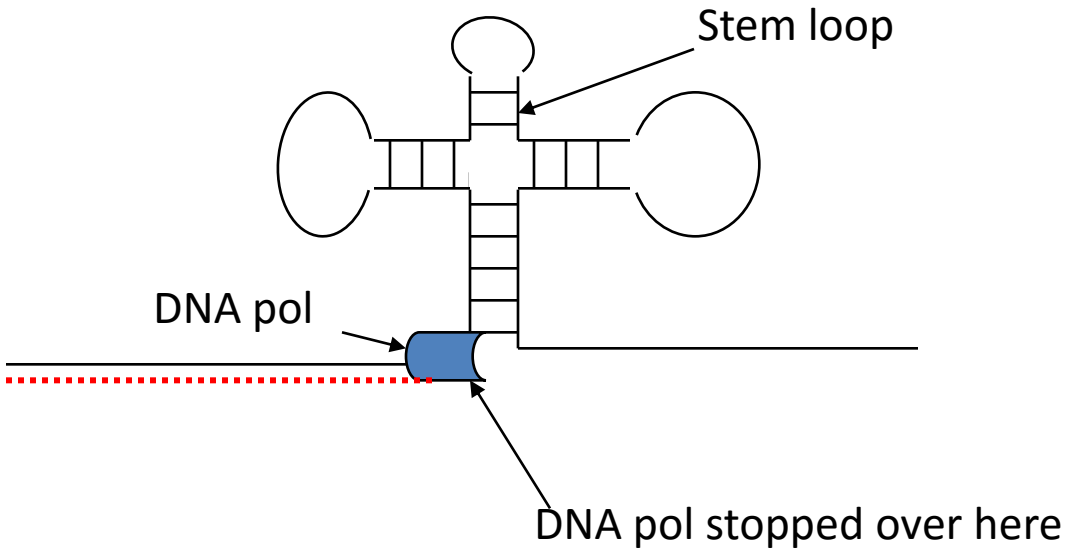
1.2N NaOH at 90oC result in strong cleavage at adenine and weaker cleavage at cytosine.

DNA SEQUENCING



Advantage of Maxam-Glibert Method

- Short DNAs like oligonucleotides can be sequenced
- DNAs that form Secondatory structure can be sequenced



DNA SEQUENCING

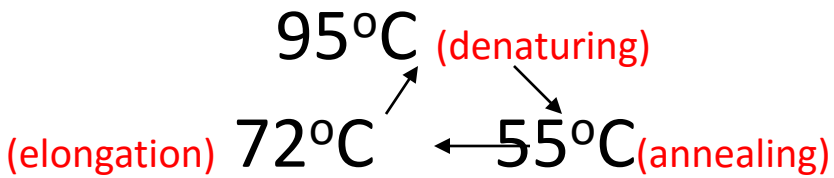
Comparison

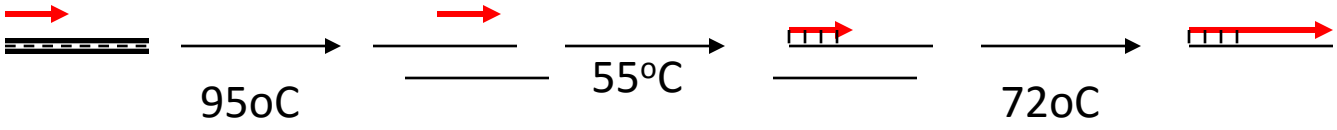
- Sanger's method
- Additional cloning was needed to make ssDNA (time consuming)
- Primers need to be synthesized
- 2^o structures give problems
- Short oligonucleotides cannot be sequenced

- Maxam-Glibert method
- No additional cloning needed
- No primer involved
- No 2^o structures problems
- Short oligos can be sequenced by this methods

Cycle (PCR) sequencing

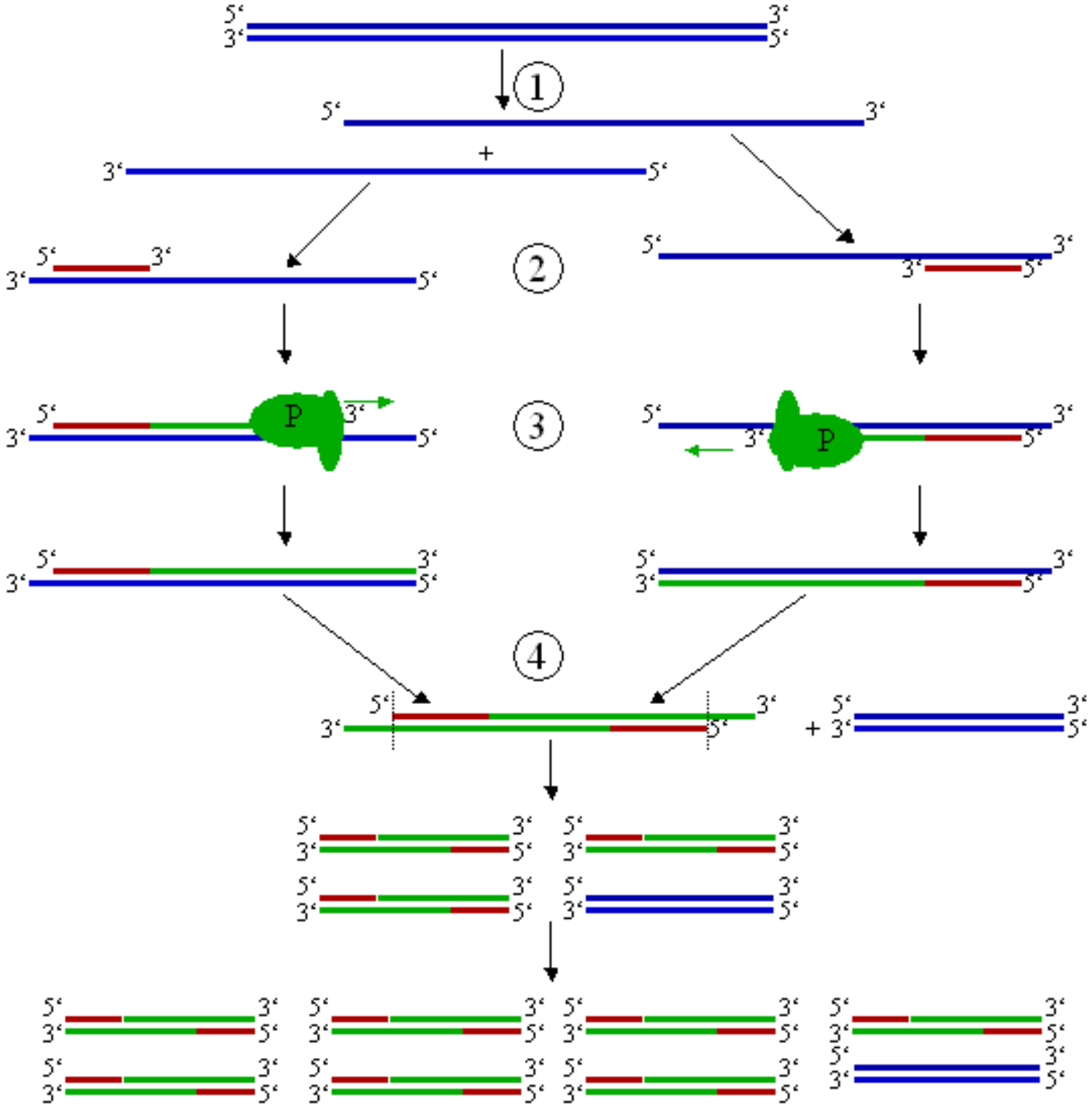
- Same Sanger's method used, but PCR was used to perform the reactions.
- PCR has three steps
  - 1. denaturing
  - 2. annealing
  - 3. elongation or synthesis





'5-32P-TAGCTGACTCAddG*-3'
3'-ATCGACTGAGTCAAGAACTATTGGGCTTAA-5'

5'-32P-TAGCTGACTCddA*-3'
3'-ATCGACTGAGTCAAGAACTATTGGGCTTAA-5'

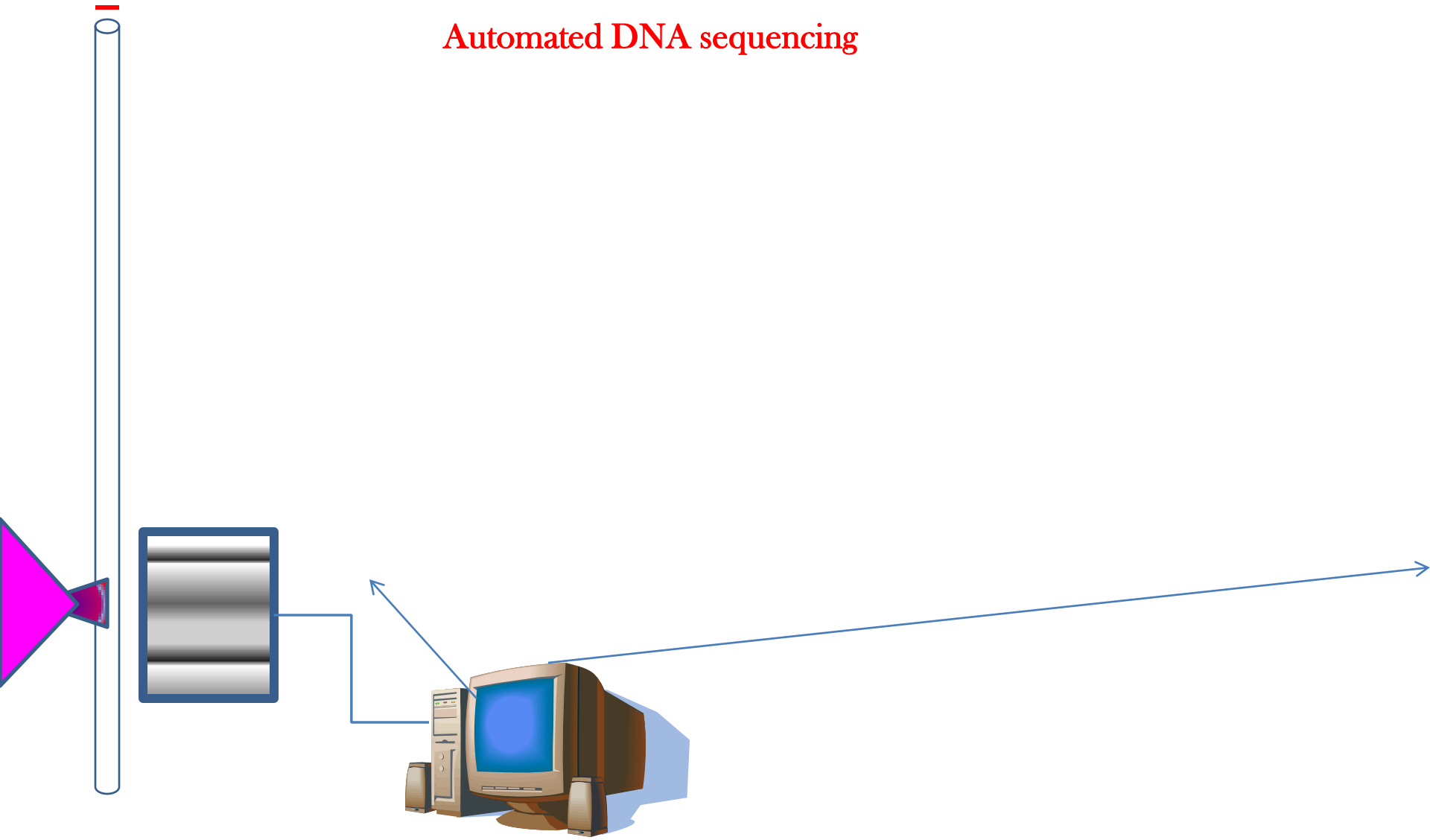

5'
3'
1
+
2
P
3
4

DNA SEQUENCING

Automated sequence

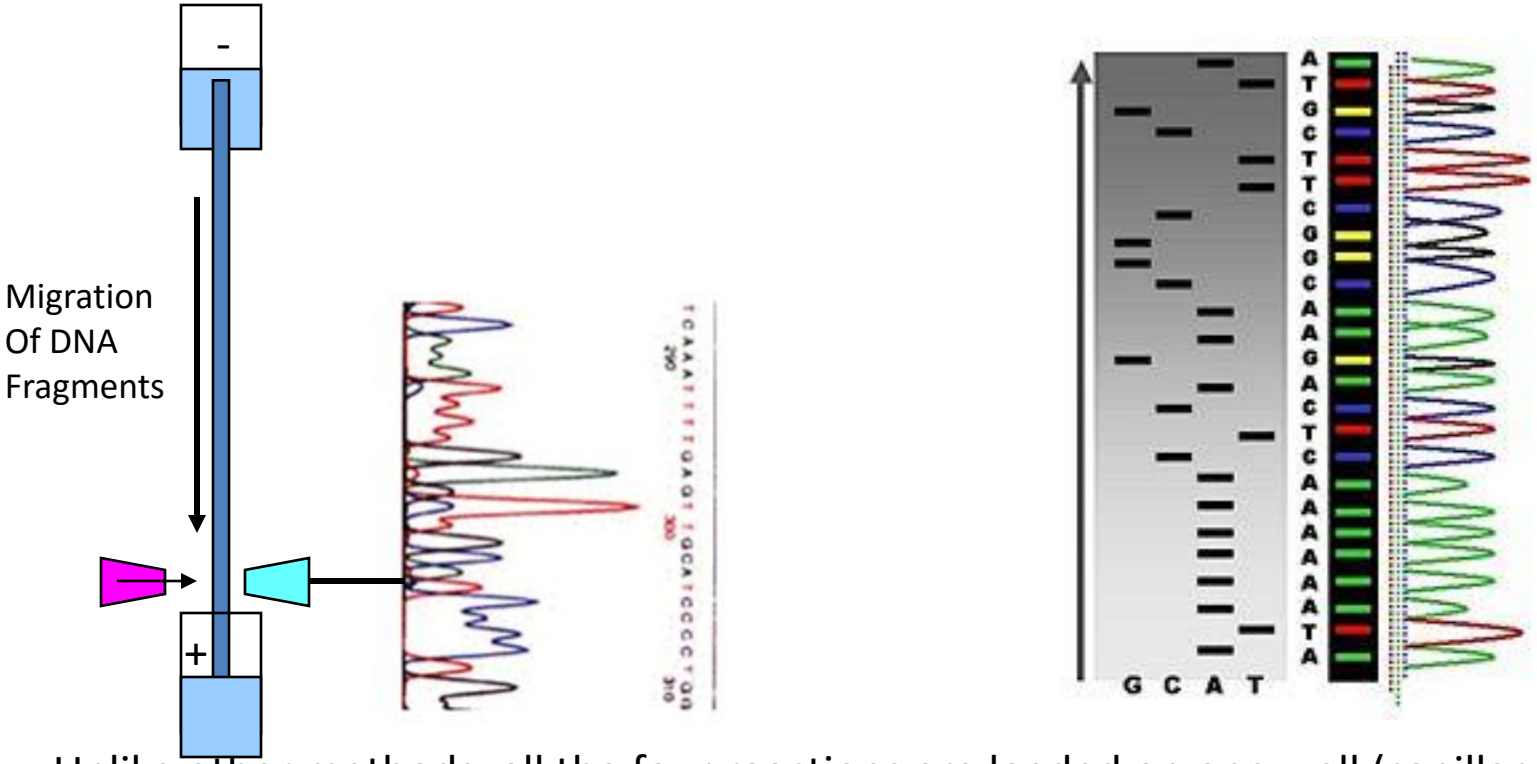
How DNA Sequence Is Determined?

Automated DNA sequencing



DNA SEQUENCING

Automated sequence



Unlike other methods, all the four reactions are loaded on one well (capillary gel)
At the end, dye was excited by laser light and fluorescence was detected.

SEQUENCING LARGE MOLECULES OF DNA

- Both the Maxam-Gilbert and Sanger-Coulson methods can only produce about 300-500 bases of sequence at a time.

  Most genes are larger than this. To sequence a large DNA molecule it is cut up into different fragments and each fragment is sequenced in turn, including overlaps.

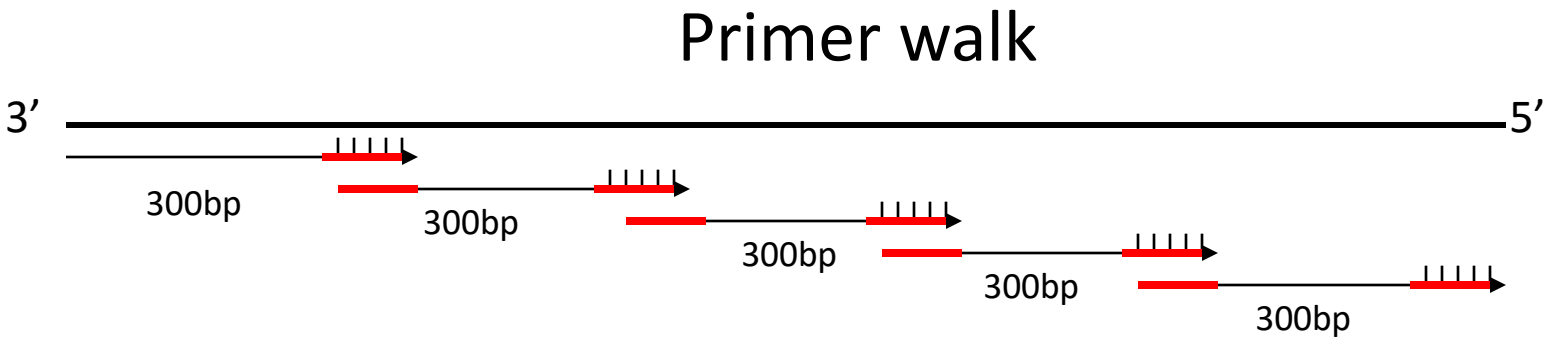
  The full sequence can then be determined.

DNA SEQUENCING

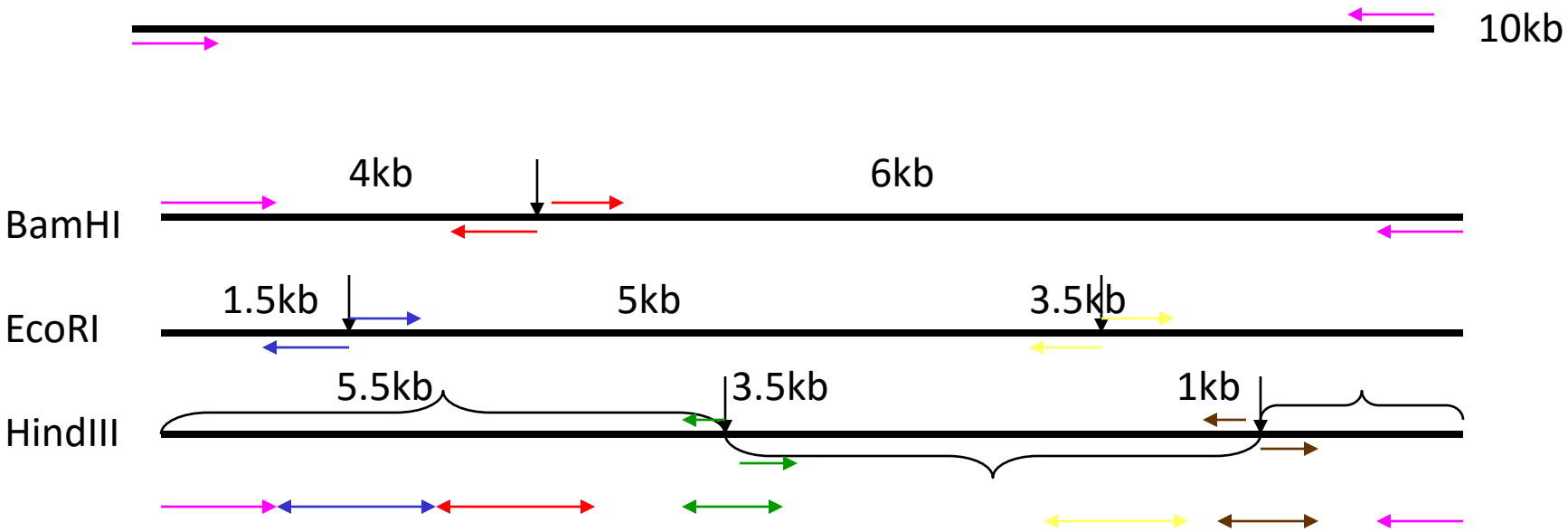
Two methods

Primer walk

Restriction enzymes (RE) map.



RE Mapping



New sequencing methods

- human genome is about 3 billion (3,000,000,000) bp long.
- if the average fragment length is 500 bases, it would take a minimum of six million to sequence the human genome (1-fold coverage)

Major landmarks in DNA sequencing

- 1953 Discovery of the structure of the DNA double helix.
- 1972 Development of recombinant DNA technology, which permits isolation of defined fragments of DNA; prior to this, the only accessible samples for sequencing were from bacteriophage or virus DNA.
- 1975 The first complete DNA genome to be sequenced is bacteriophage ϕX174
- 1977
  - Allan Maxam and Walter Gilbert publish "DNA sequencing by chemical degradation".
  - Fred Sanger, independently, publishes "DNA sequencing by enzymatic synthesis".
- 1980 Fred Sanger and Wally Gilbert receive the Nobel Prize in Chemistry
- 1982 Genbank starts as a public repository of DNA sequences.

DNA SEQUENCING

- 1984 Medical Research Council scientists decipher the complete DNA sequence of the Epstein-Barr virus, 170 kb.
- 1985 Kary Mullis and colleagues develop the polymerase chain reaction, a technique to replicate small fragments of DNA
- 1986 Leroy E. Hood's laboratory at the Cal Tech and Smith announce the first semi-automated DNA sequencing machine.
- 1987 Applied Biosystems markets first automated sequencing machine, the model ABI 370.
- 1990 The U.S. National Institutes of Health begins large-scale sequencing trials on
    - *Mycoplasma capricolum*,
    - *Escherichia coli*,
    - *Caenorhabditis elegans*, and
    - *Saccharomyces cerevisiae*.
  - Barry Karger, Lloyd Smith and Norman Dovichi publish on capillary electrophoresis.
- 1991 Craig Venter develops strategy to find expressed genes with ESTs (Expressed Sequence Tags).

DNA SEQUENCING

1995 Venter, Fraser and Smith publish first sequence of free-living organism, *Haemophilus influenzae* (genome size of 1.8 Mb).

- Richard Mathies et al. publish on sequencing dyes (PNAS, May)
- Michael Reeve and Carl Fuller, thermostable polymerase for sequencing

- 1996 International Human Genome Project partners agree to release sequence data into public databases within 24 hours.

- 1997 Blattner, Plunkett et al. publish the sequence of E. coli (genome size of 5 Mb)

- 1999 NIH moves up completion date for rough draft, to spring 2000.
  - NIH launches the mouse genome sequencing project.
  - First sequence of human chromosome 22 published.

- 2000 Celera and collaborators sequence fruit fly *Drosophila melanogaster* (genome size of 180Mb)

DNA SEQUENCING

- HGP consortium publishes sequence of chromosome 21.
- HGP & Celera jointly announce working drafts of HG sequence, promise joint publication.
  - Estimates for the number of genes in the human genome range from 35,000 to 120,000.
  - International consortium completes first plant sequence, *Arabidopsis thaliana* (genome size of 125 Mb).

- 2001 HGP consortium publishes Human Genome Sequence draft in Nature (15 Feb)
  - Celera publishes the Human Genome sequence

- 2007 For the first time, a set of closely related species (12 Drosophilidae) are sequenced, launching the era of phylogenomics.
  - Craig Venter publishes his full diploid genome: the first human genome to be sequenced completely.

- 2008 An international consortium launches The 1000 Genomes Project, aimed to study human genetic variability.
- 2008 Leiden University Medical Center scientists decipher the first complete DNA sequence of a woman.

Core course
BMS361N
Genetic Engineering

The End

Prof. Narkunaraja Shanmugam
Dept. Of Biomedical Science
School of Basic Medical Sciences
Bharathidasan University