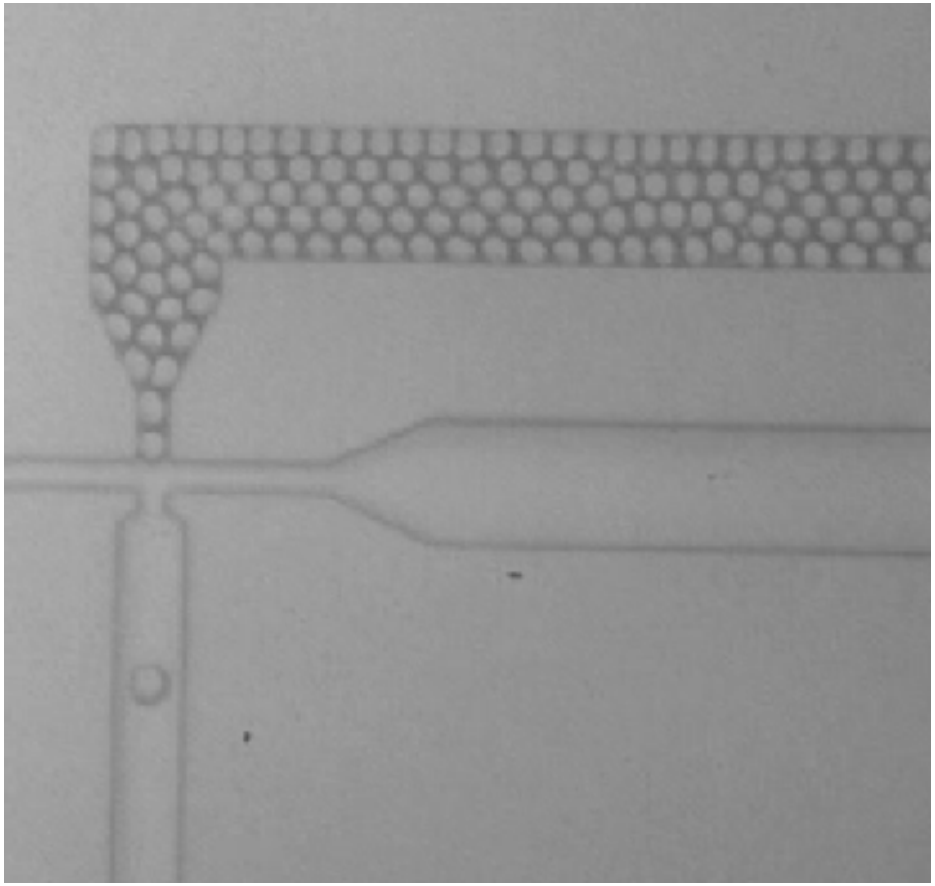
Biotechnologie Moléculaire

3ème année

Andrew Griffiths

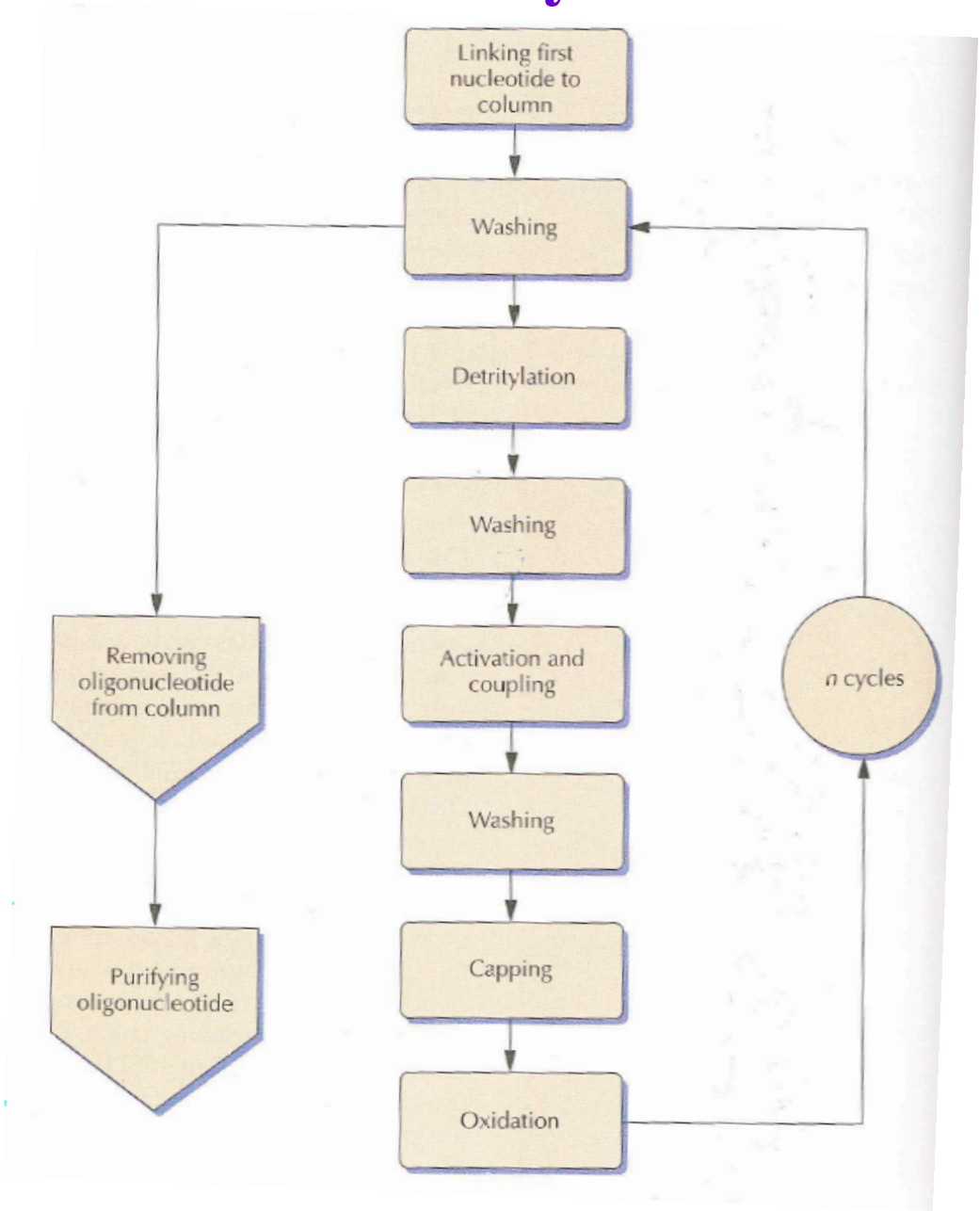
Laboratoire de Biochimie

Room B 231
Tel: 01 40 79 45 39
Email: andrew.griffiths@espci.fr



Chemical Synthesis, Sequencing and Amplification of DNA

Chemical Synthesis of DNA Oligonucleotides

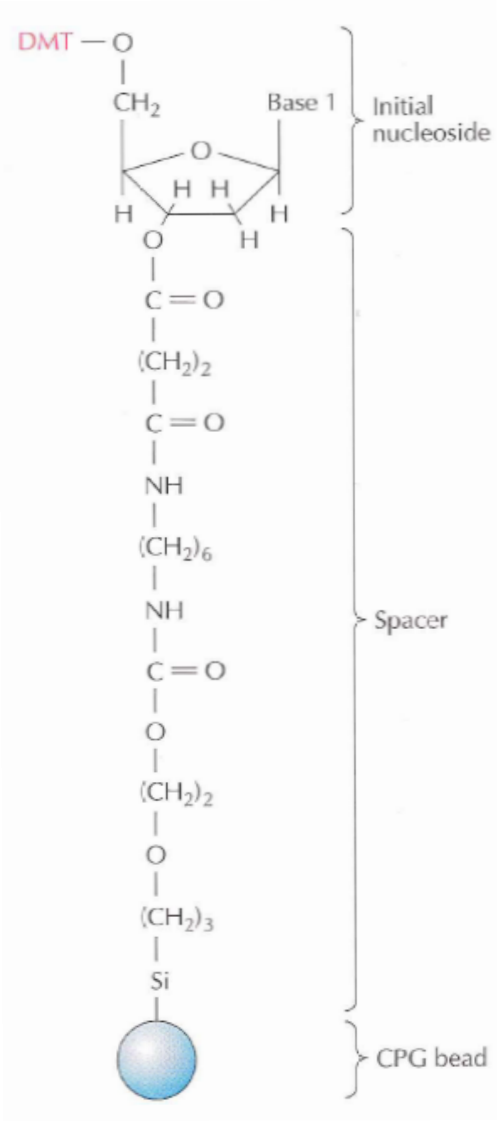


Synthesis is performed on an automated machine

Synthesis is from 3' to 5' (not 5' to 3' as in nature)

After *n* coupling reactions (cycles) a single-stranded DNA with $n + 1$ nucleotides is produced

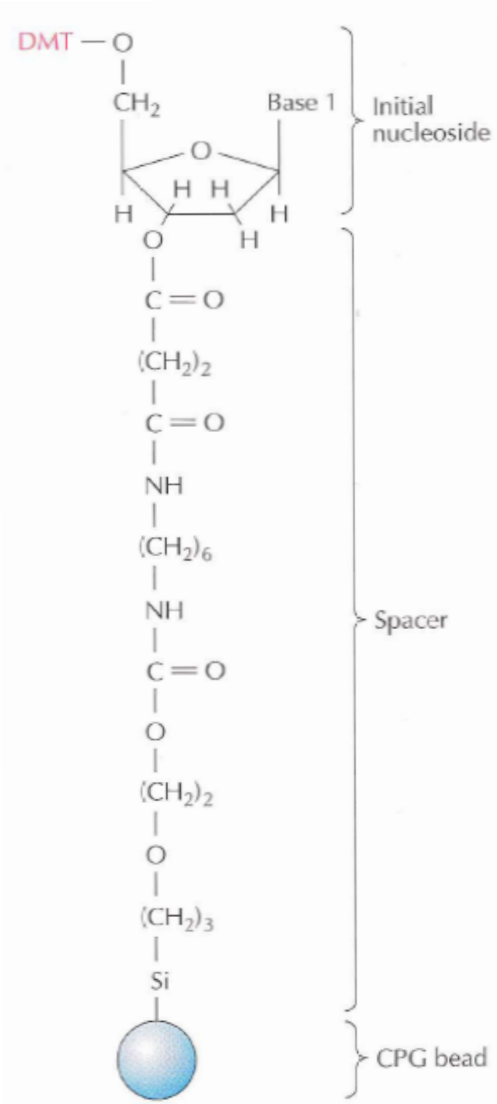
Chemical Synthesis of DNA Oligonucleotides



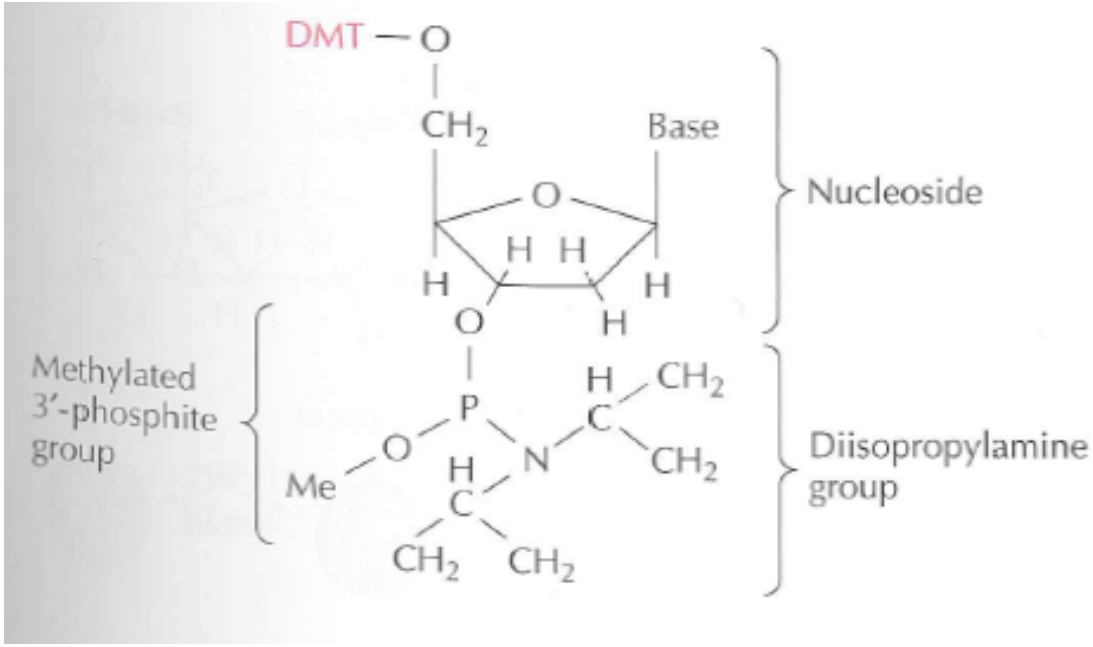
Starting complex

DMT = dimethoxytrityl

Chemical Synthesis of DNA Oligonucleotides



Starting complex
DMT = dimethoxytrityl

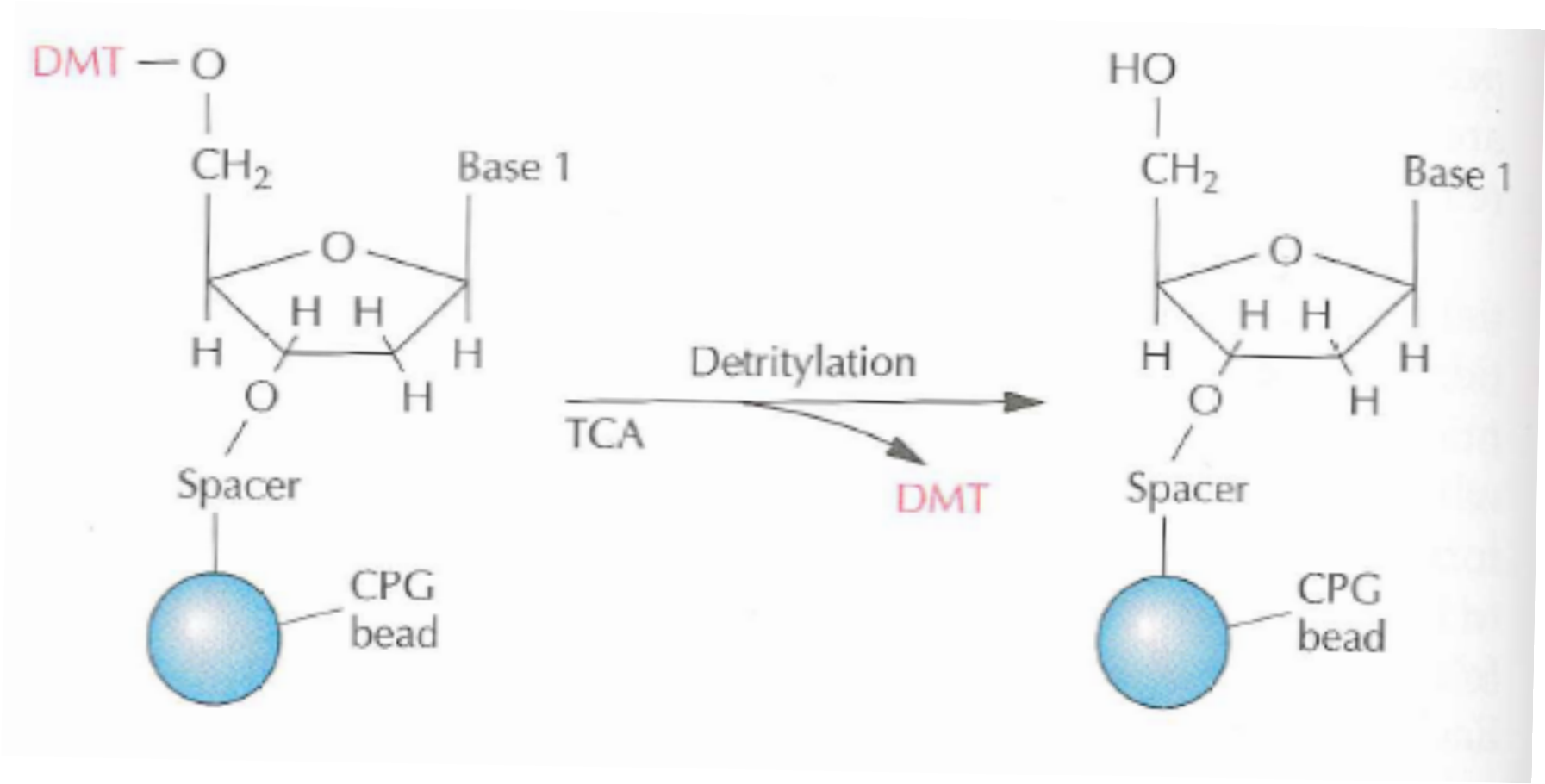


The amino groups of the bases adenine, guanine and cytosine are protected with benzoyl, isobutyryl and benzoyl groups respectively (thymine has no amino group).

Structure of a phosphoramidite

Chemical Synthesis of DNA Oligonucleotides

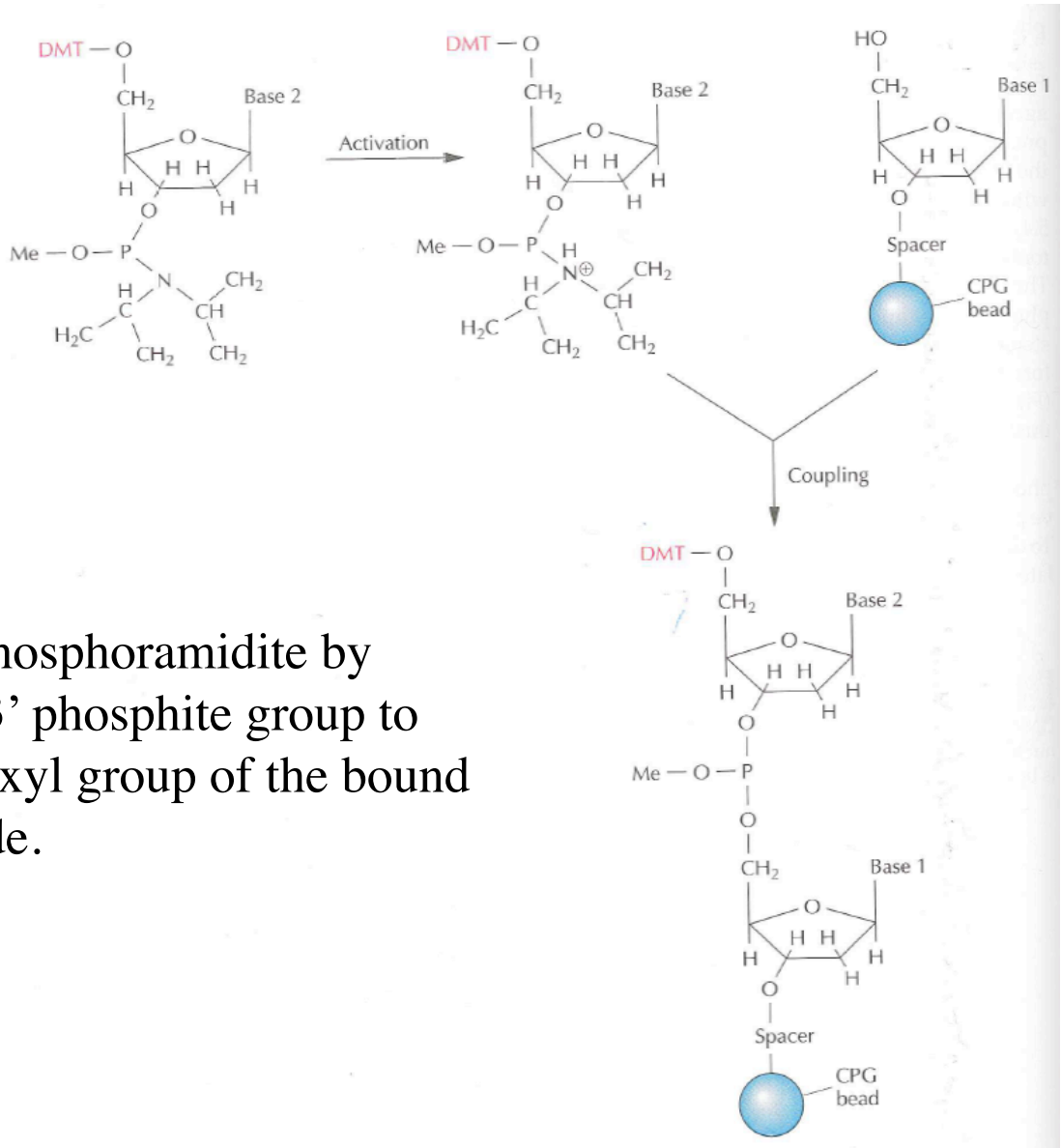
Detritylation



The 5’ DMT group is removed by treatment with tricarboxylic acid (TCA)

Chemical Synthesis of DNA Oligonucleotides

Activation and coupling

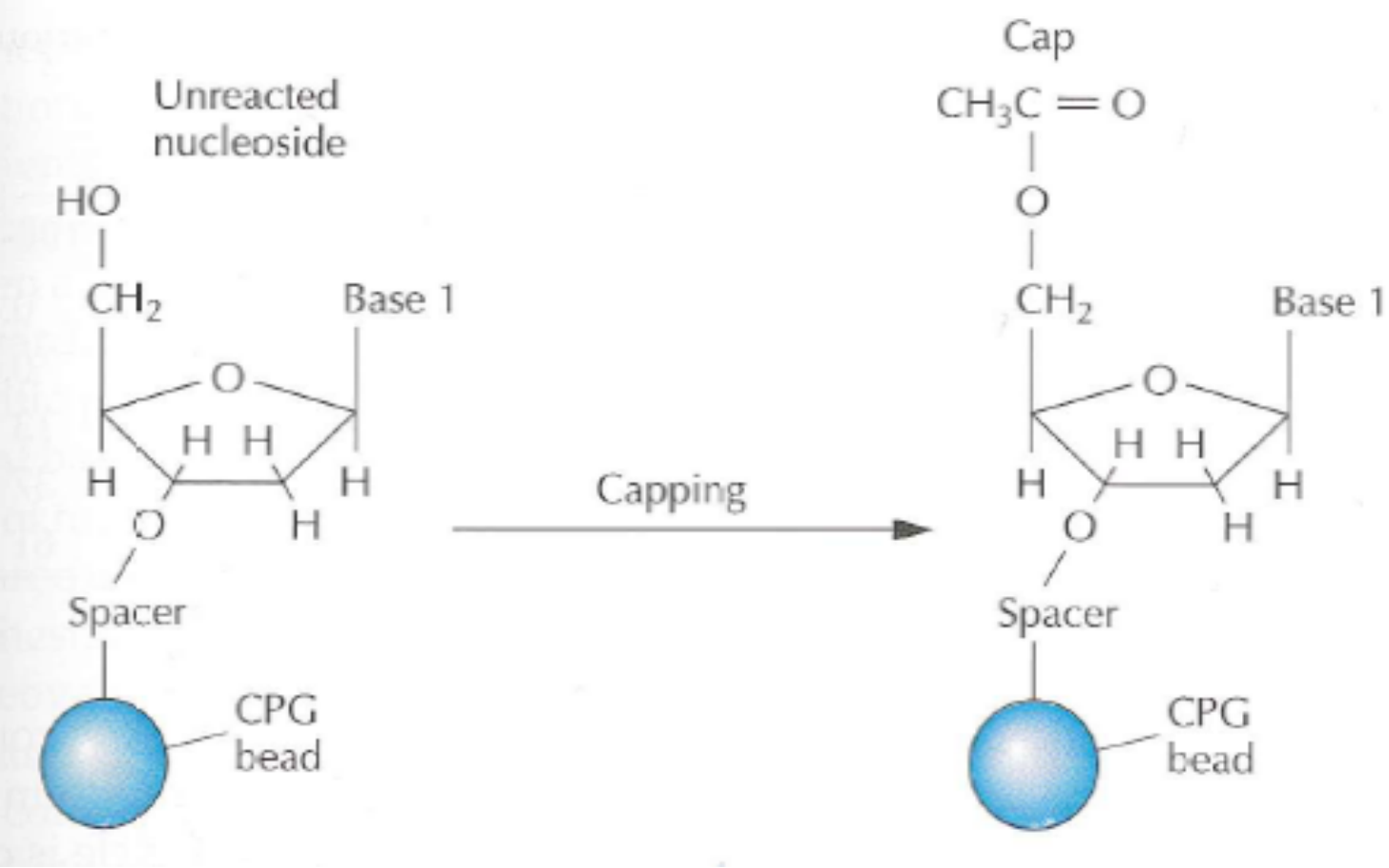


The activation of a phosphoramidite by tetrazole enables its 3' phosphite group to attach to the 5' hydroxyl group of the bound detritylated nucleoside.

Me = methyl

Chemical Synthesis of DNA Oligonucleotides

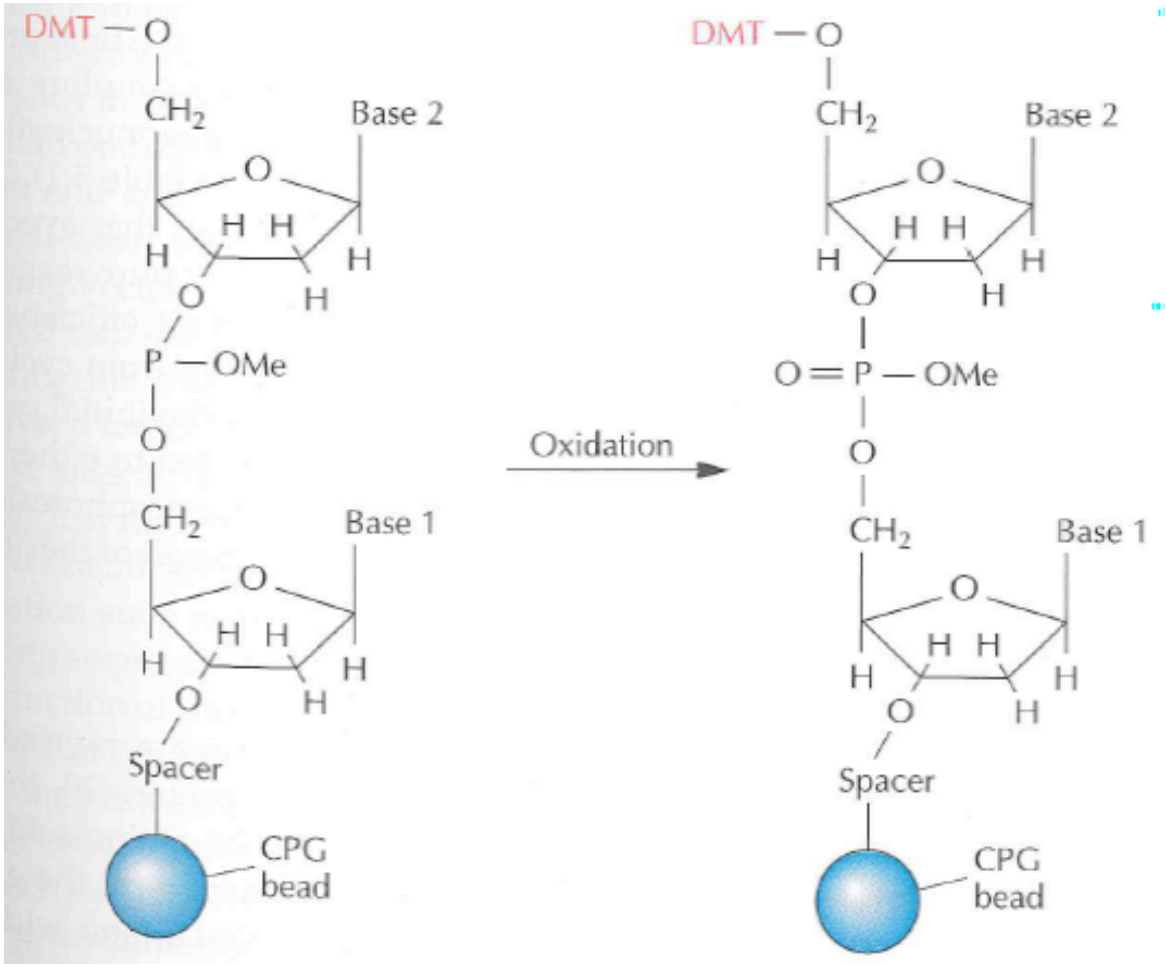
Capping



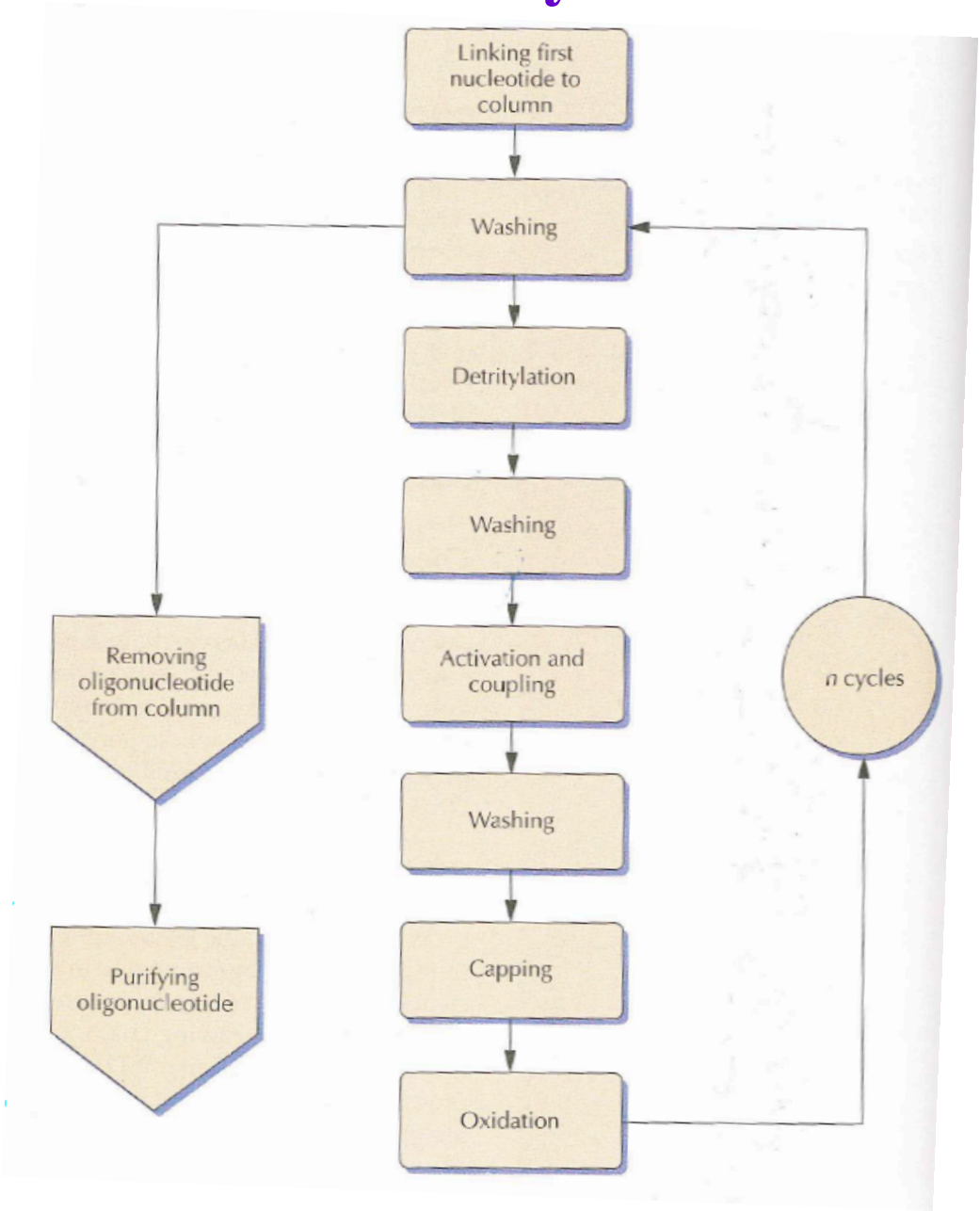
The available 5' hydroxyl groups of unreacted detritylated nucleosides are acetylated to prevent them from participating in the coupling reaction of the next cycle

Chemical Synthesis of DNA Oligonucleotides

Oxidation



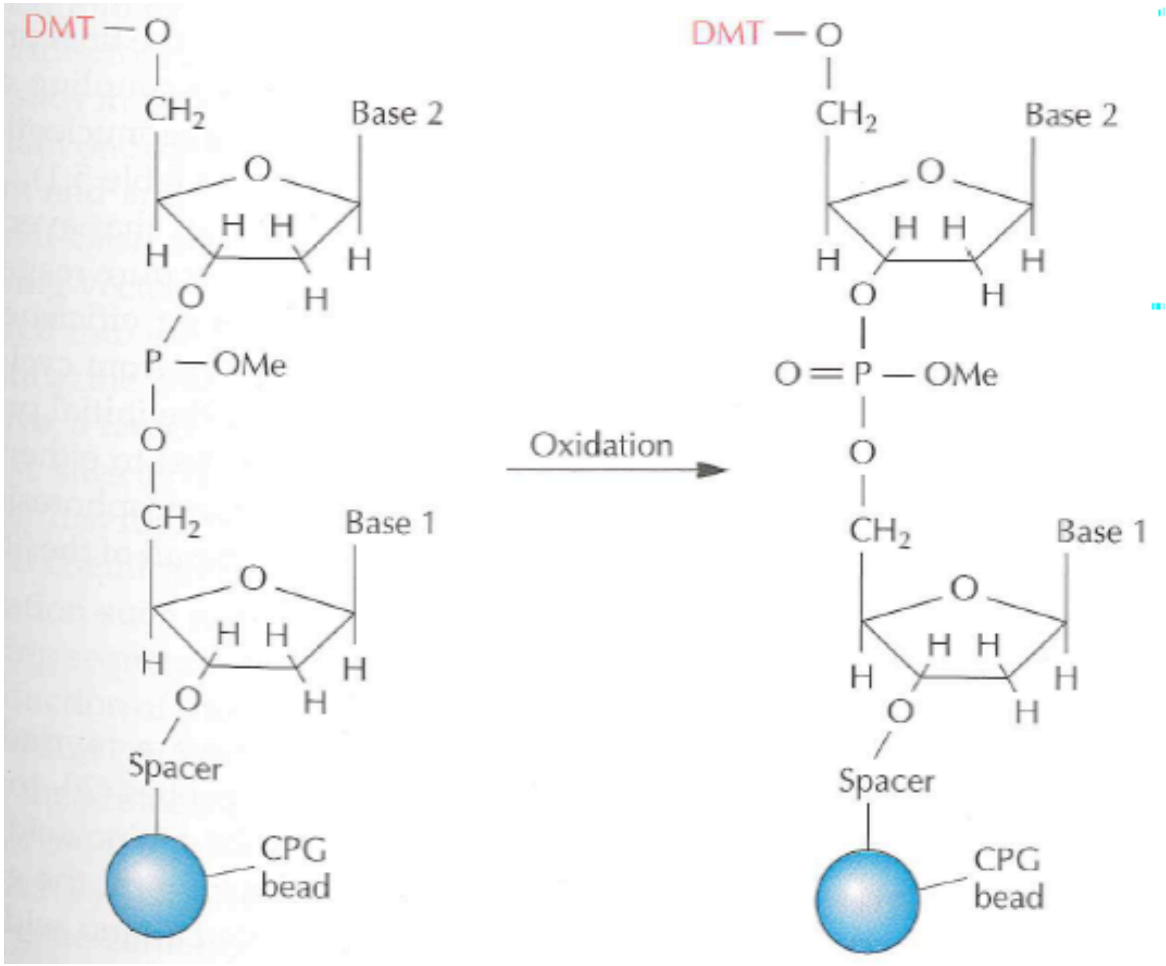
Chemical Synthesis of DNA Oligonucleotides



Repeat Cycles

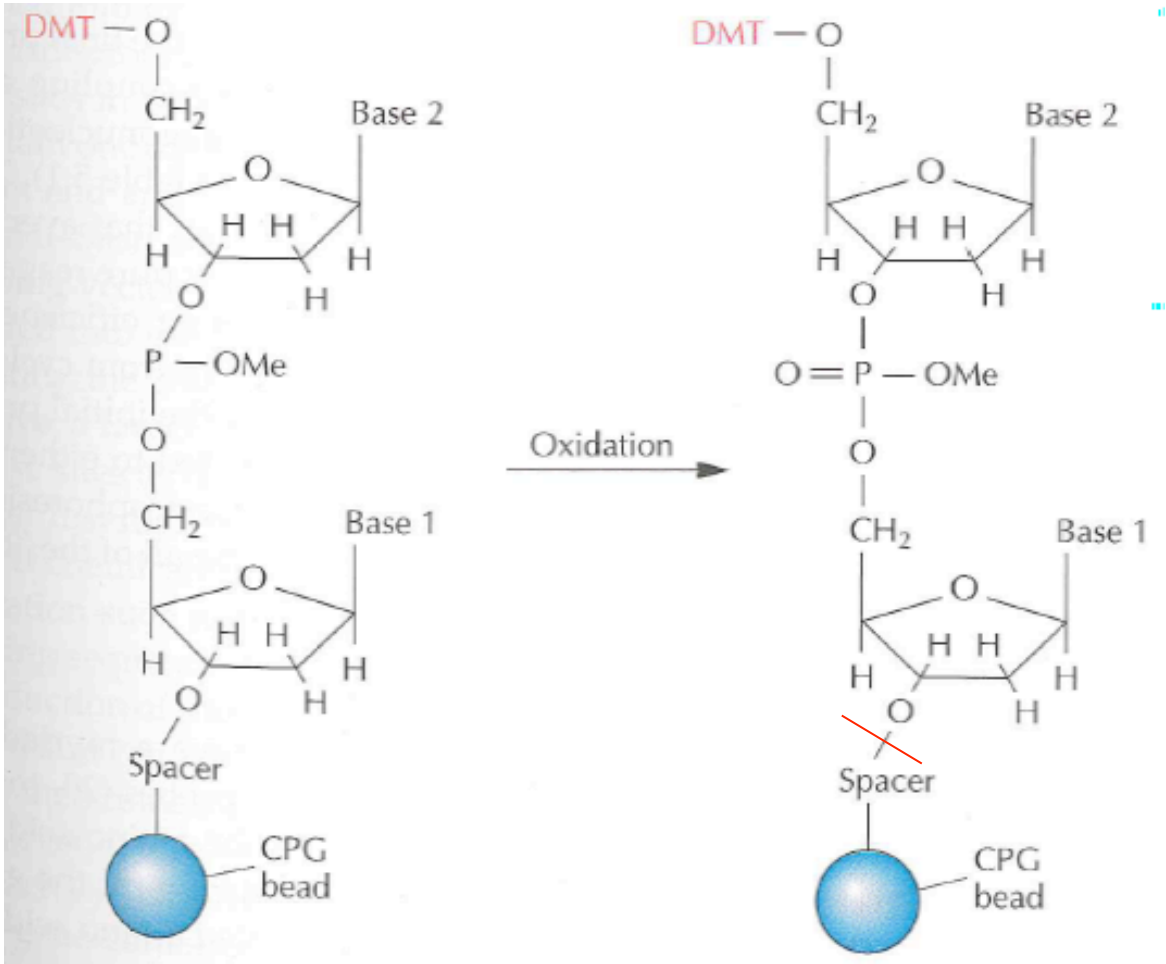
Chemical Synthesis of DNA Oligonucleotides

Deprotection, detritylation and release from bead



Chemical Synthesis of DNA Oligonucleotides

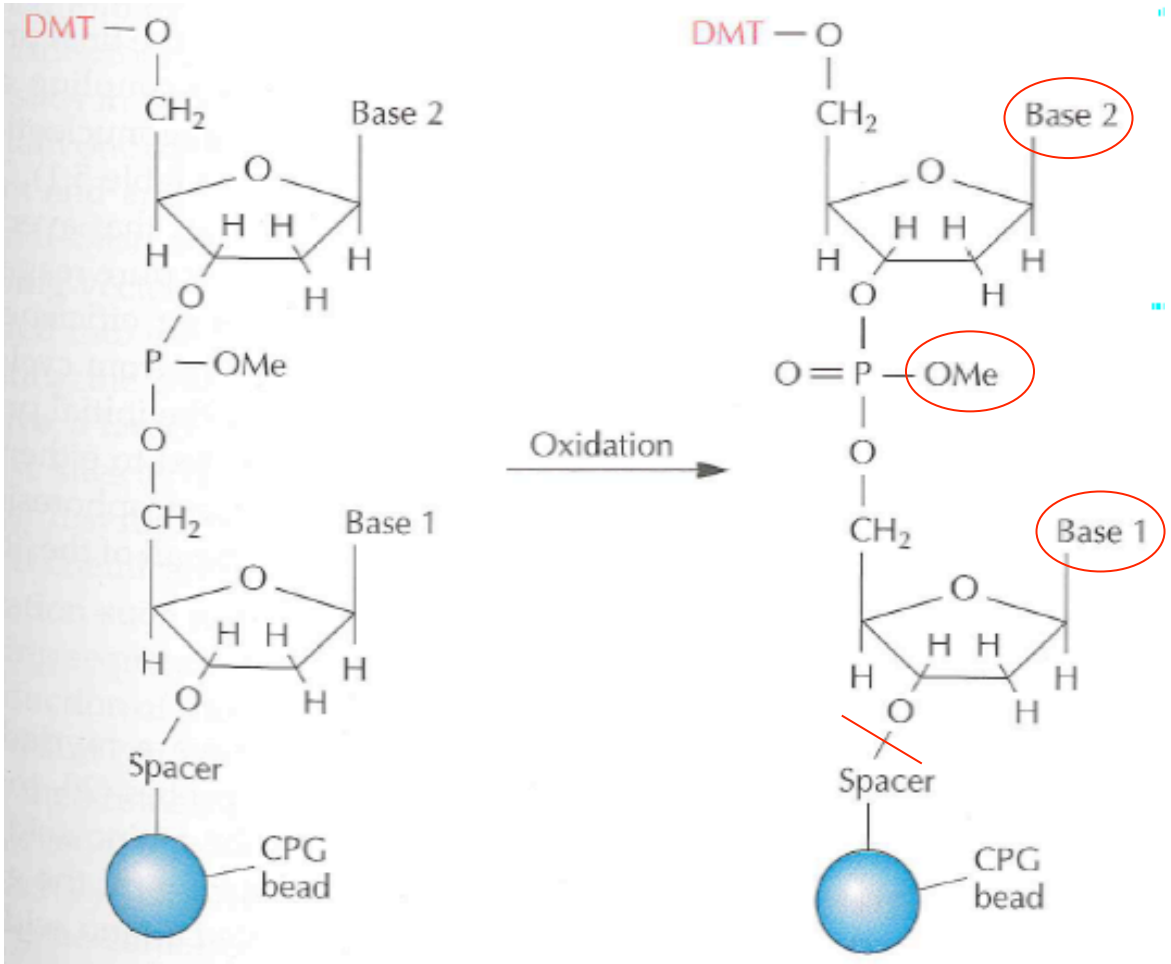
Deprotection, detritylation and release from bead



The DNA is eluted from the column

Chemical Synthesis of DNA Oligonucleotides

Deprotection, detritylation and release from bead

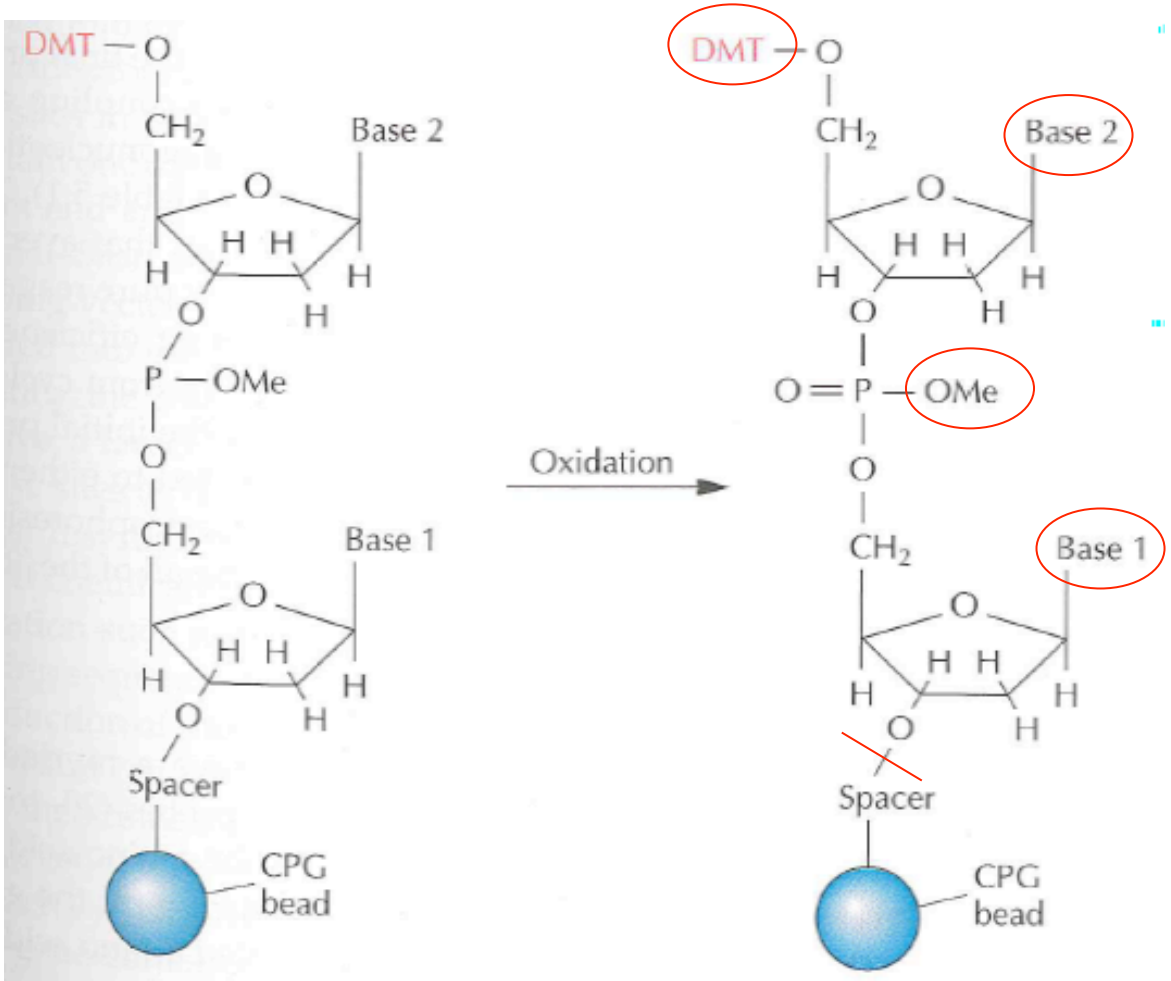


The DNA is eluted from the column

The protecting groups are removed from the bases and the backbone

Chemical Synthesis of DNA Oligonucleotides

Deprotection, detritylation and release from bead



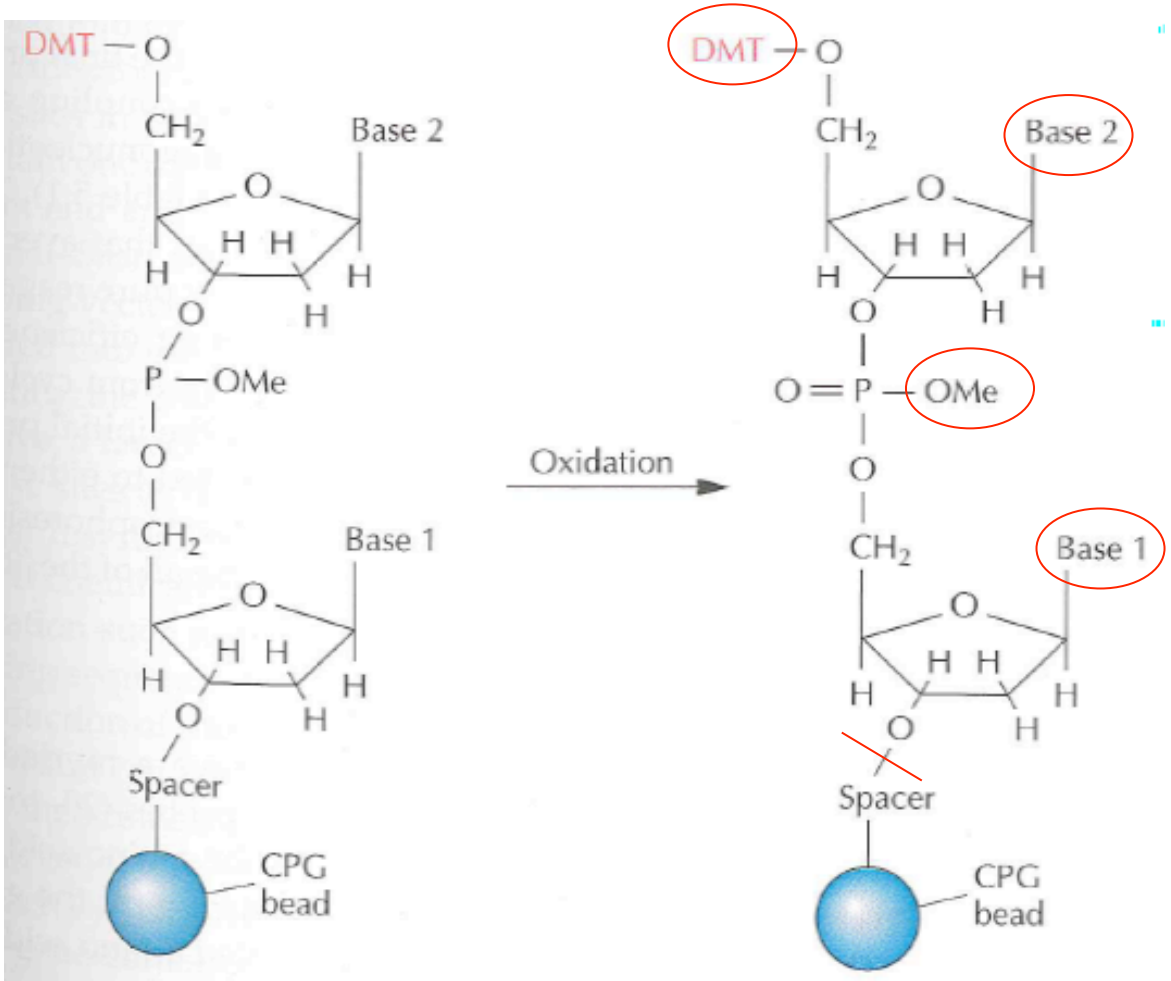
The DNA is eluted from the column

The protecting groups are removed from the bases and the backbone

The DNA is detritylated

Chemical Synthesis of DNA Oligonucleotides

Deprotection, detritylation and release from bead



The DNA is eluted from the column

The protecting groups are removed from the bases and the backbone

The DNA is detritylated

5'-phosphorylation can optionally be performed after detritylation while the oligo is still on the bead

Chemical Synthesis of DNA Oligonucleotides

Yield

Coupling efficiency (%)	Overall yield of oligonucleotide (%)				
	20-mer	40-mer	60-mer	80-mer	100-mer
90	12	1.5	0.18	0.02	0.003
95	36	13	4.6	1.7	0.6
98	67	45	30	20	13
99	82	67	55	45	37
99.5	90	82	74	67	61

Overall yield for various coupling efficiencies for each cycle

Use of Synthesised Oligonucleotides as Probes

DNA hybridization probes deduced from protein sequence to screen genomic or cDNA libraries

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC TAA **Ter** TAG **Ter**	TGT Cys (C) TGC TGA **Ter** TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " **ATG** Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

Problem of redundancy of the genetic code - a single amino acid can be encoded by several different sequences

Use of Synthesised Oligonucleotides as Probes

DNA hybridization probes deduced from protein sequence to screen genomic or cDNA libraries

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC TAA **Ter** TAG **Ter**	TGT Cys (C) TGC TGA **Ter** TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " **ATG** Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

Problem of redundancy of the genetic code - a single amino acid can be encoded by several different sequences

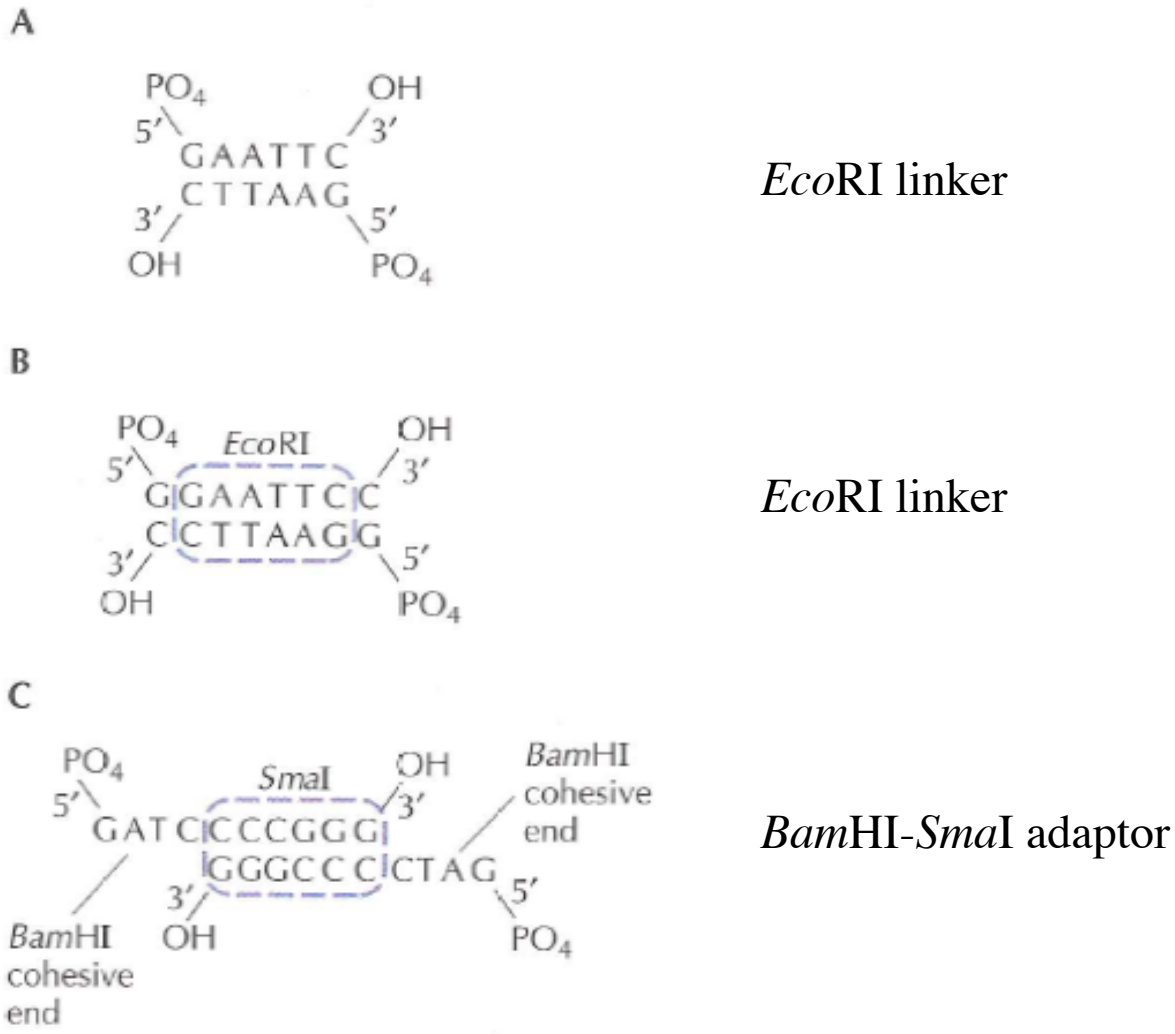
Uses of Synthesised Oligonucleotides

DNA hybridization probes deduced from protein sequence to screen genomic or cDNA libraries

The third “wobble” position of the codon is synthesized using equal concentrations of the phosphoramidites for each of the 4 bases (A, C, G, T)

Uses of Synthesised Oligonucleotides

Linkers and adaptors



*Eco*RI linker

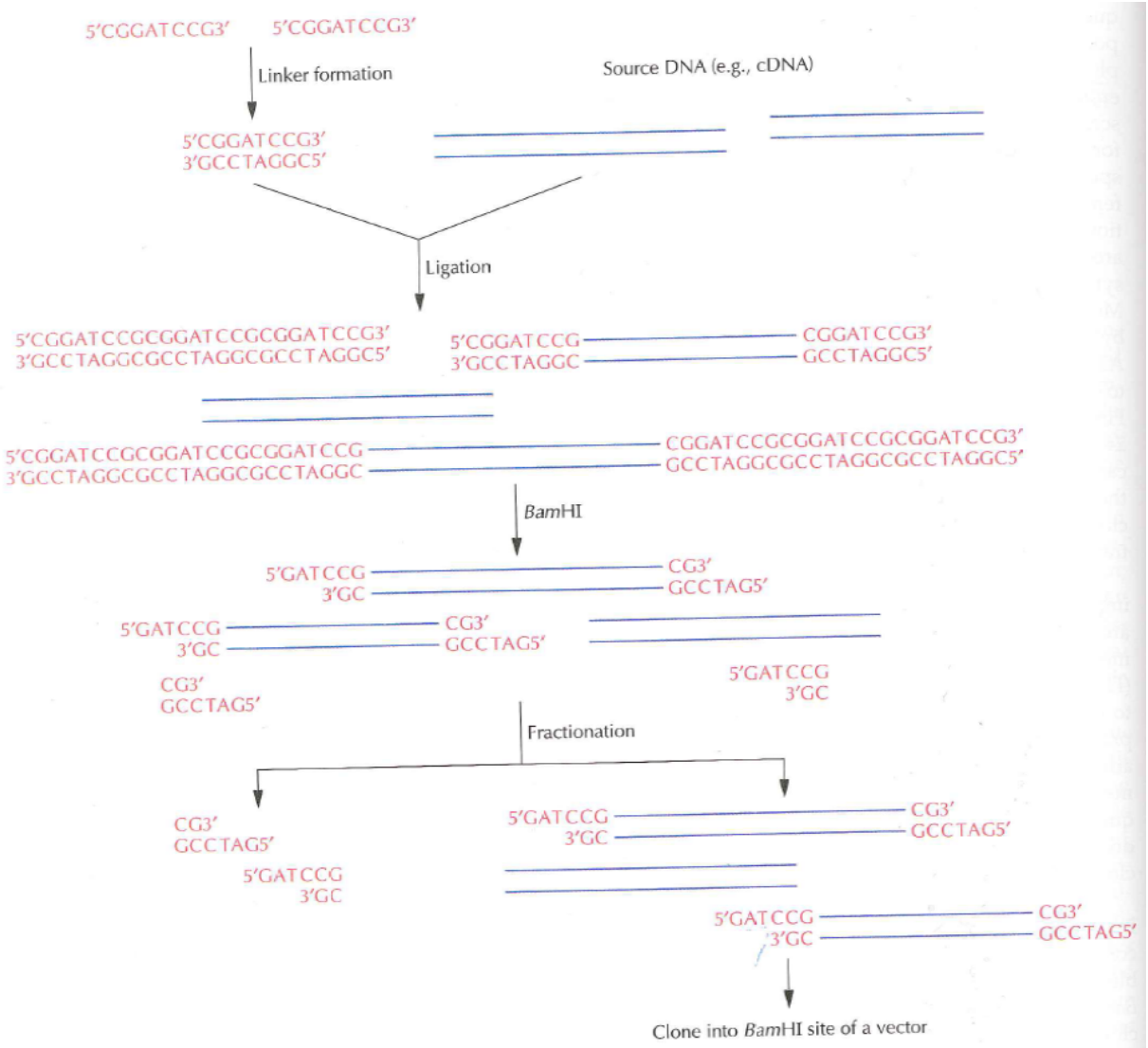
*Eco*RI linker

*Bam*HI-*Sma*I adaptor

Typical linker and adaptor sequences

Uses of Synthesised Oligonucleotides

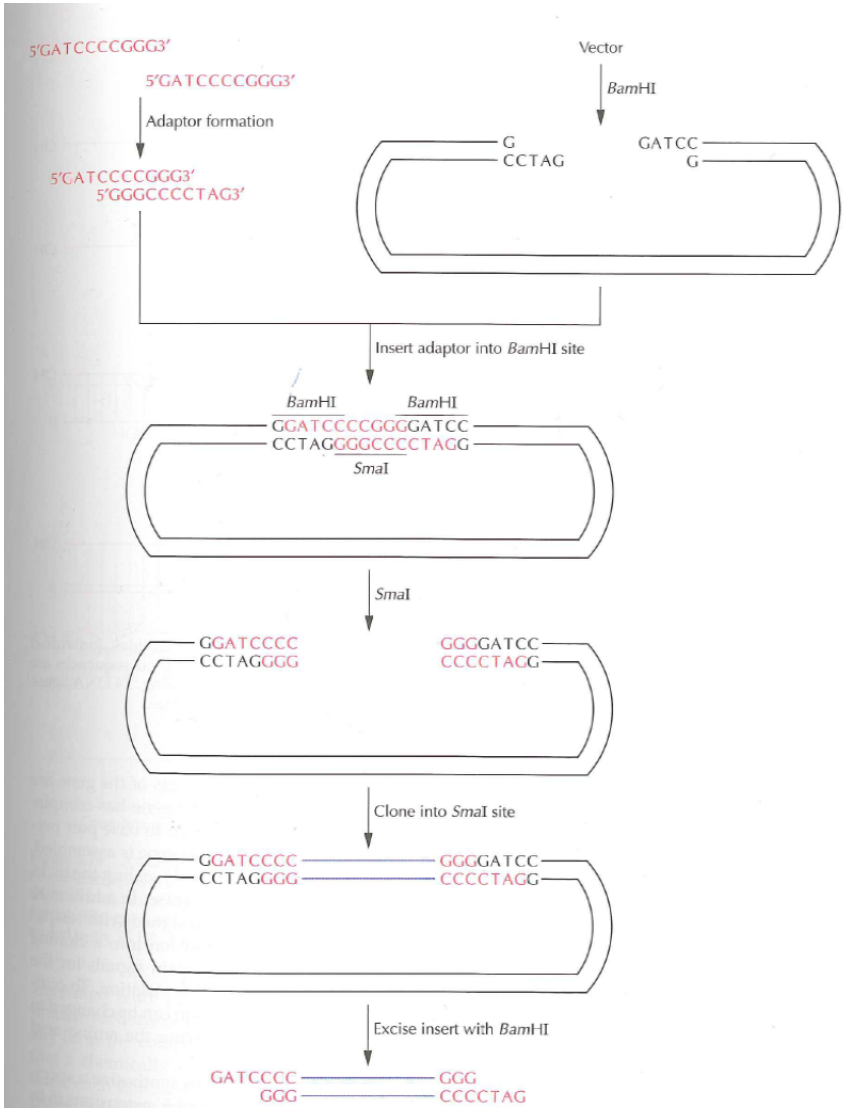
Cloning with a linker



*Bam*HI Linker

Uses of Synthesised Oligonucleotides

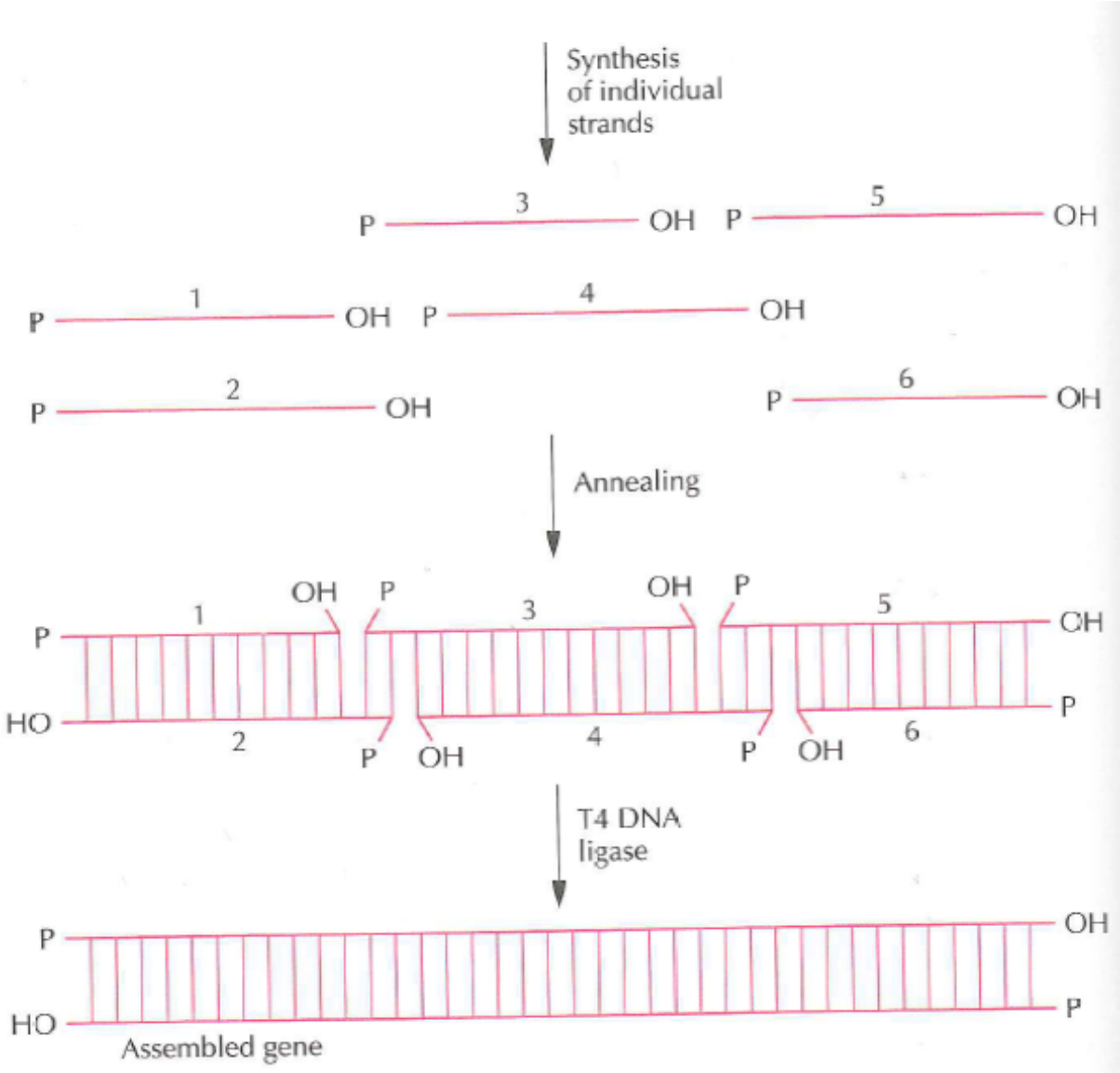
Creating a restriction endonuclease site in a vector with an adaptor



Creating a *Sma*I cloning site

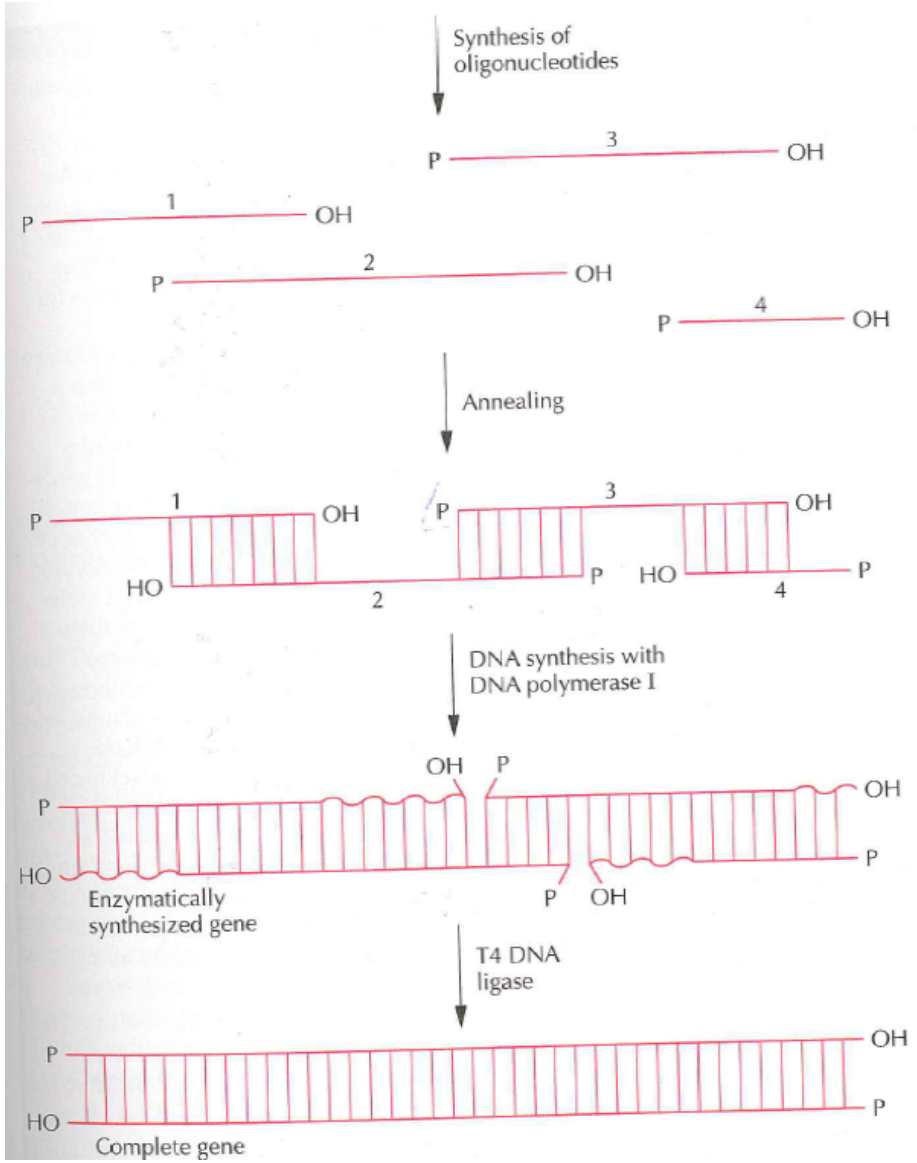
Uses of Synthesised Oligonucleotides

Assembly of a synthetic gene from short oligonucleotides



Uses of Synthesised Oligonucleotides

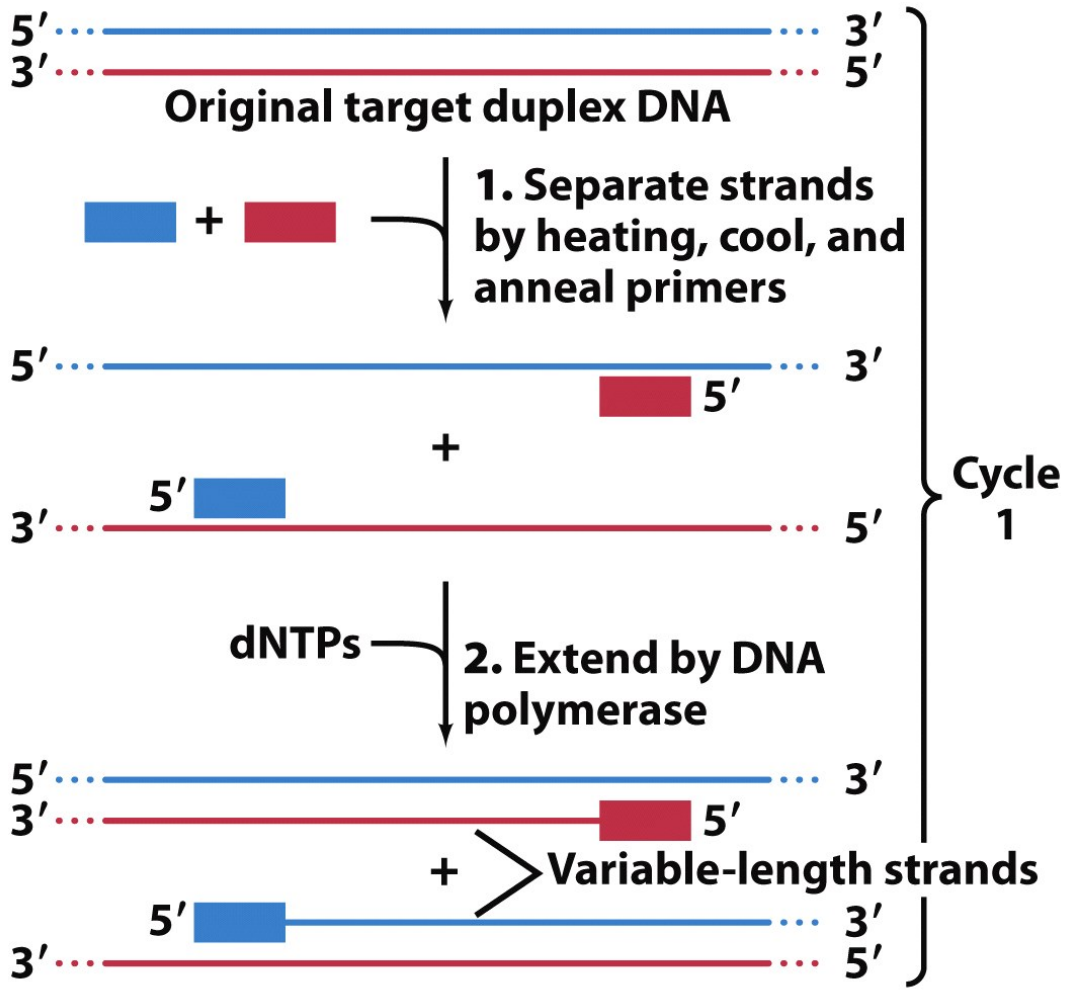
Assembly and in vitro enzymatic DNA synthesis of a gene



Polymerase Chain Reaction
PCR

Recombinant DNA

Amplifying DNA using the polymerase chain reaction (PCR)



The polymerase chain reaction (PCR) was devised by Kerry Mullis in 1985

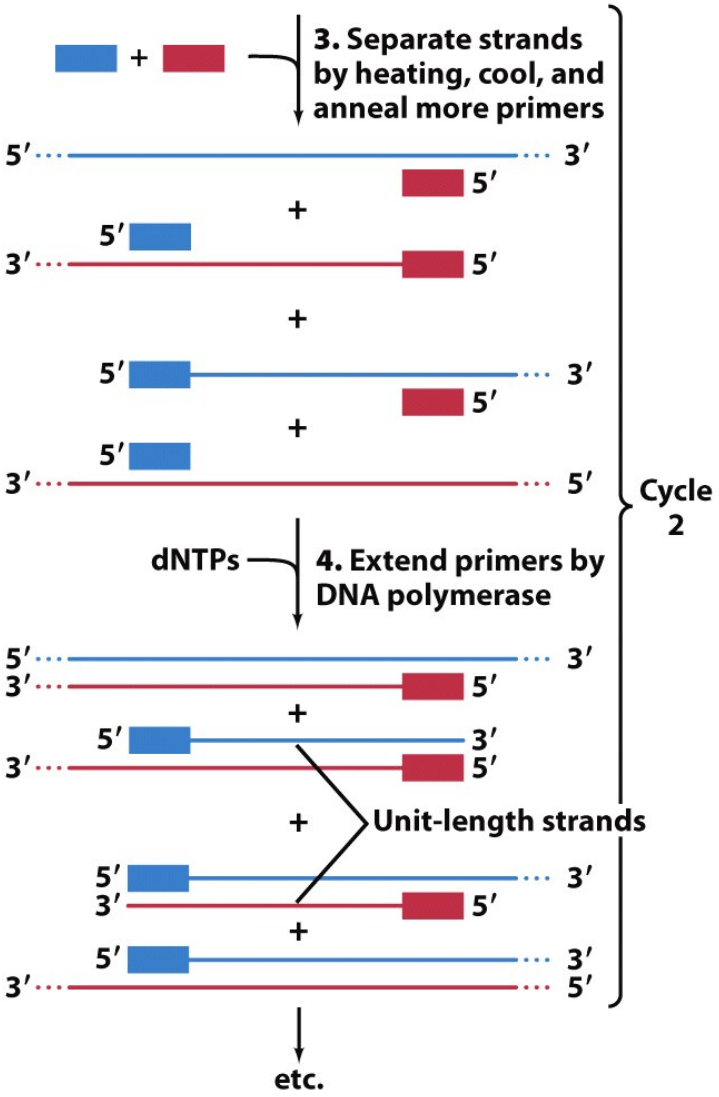
It allows exponential amplification of any defined fragments of DNA

20 cycles of PCR increases the amount of target DNA by ~1 million fold (~2^{20})

Even single DNA molecules can be amplified

Recombinant DNA

Amplifying DNA using the polymerase chain reaction (PCR)



The polymerase chain reaction (PCR) was devised by Kerry Mullis in 1985

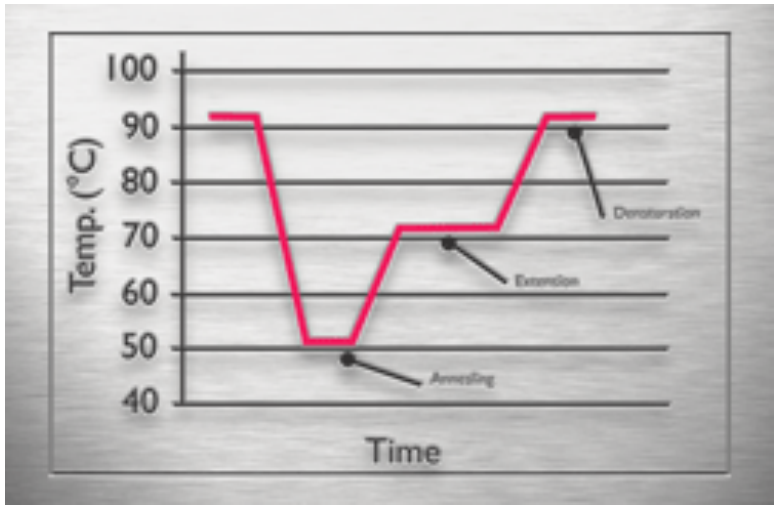
It allows exponential amplification of any defined fragments of DNA

20 cycles of PCR increases the amount of target DNA by ~1 million fold (~2^{20})

Even single DNA molecules can be amplified

Recombinant DNA

Amplifying DNA using the polymerase chain reaction (PCR)



The reaction is thermocycled to allow denaturation, annealing and extension

Using a thermostable polymerase such as *Taq* polymerase avoids the need to add fresh enzyme after each cycle

Can introduce mutations and restriction sites in the primer sequence allowing simple mutation and cloning of any sequence

The polymerase chain reaction (PCR) was devised by Kerry Mullis in 1985

It allows exponential amplification of any defined fragments of DNA

20 cycles of PCR increases the amount of target DNA by ~1 million fold (~2^{20})

Even single DNA molecules can be amplified

Recombinant DNA

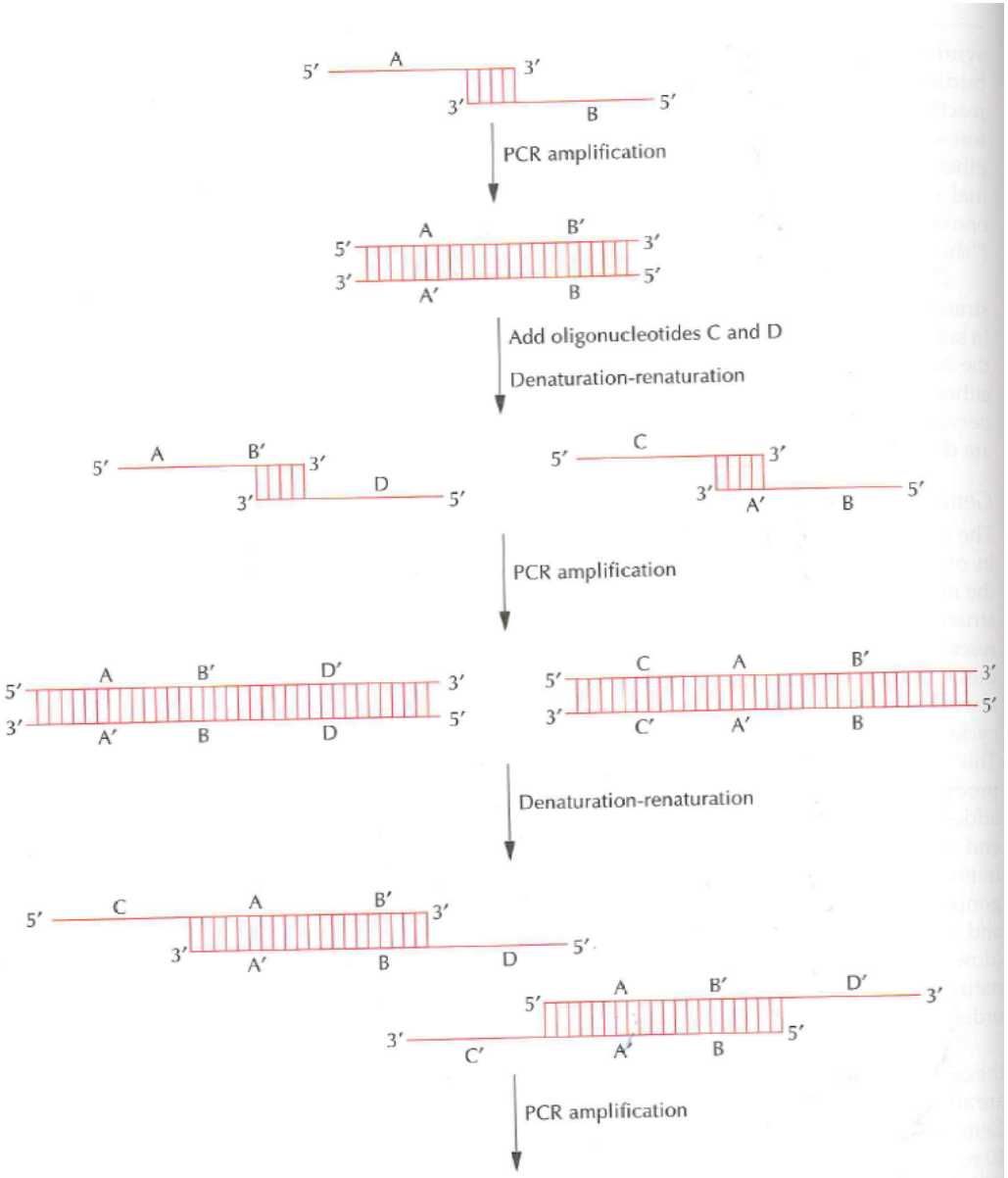
Amplifying DNA using the polymerase chain reaction (PCR)

Recombinant DNA

Amplifying DNA using the polymerase chain reaction (PCR)

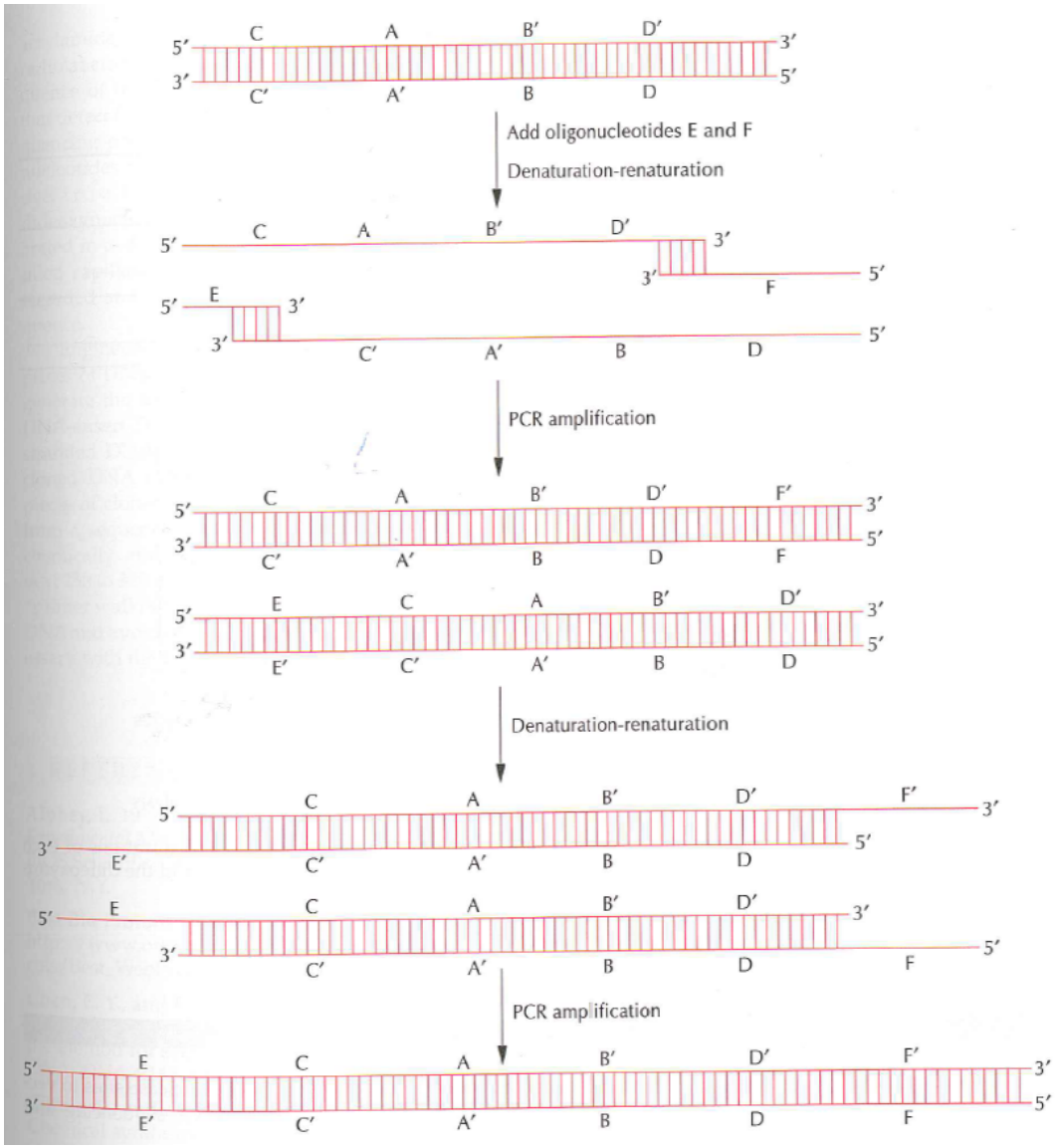
PCR

Gene synthesis by PCR



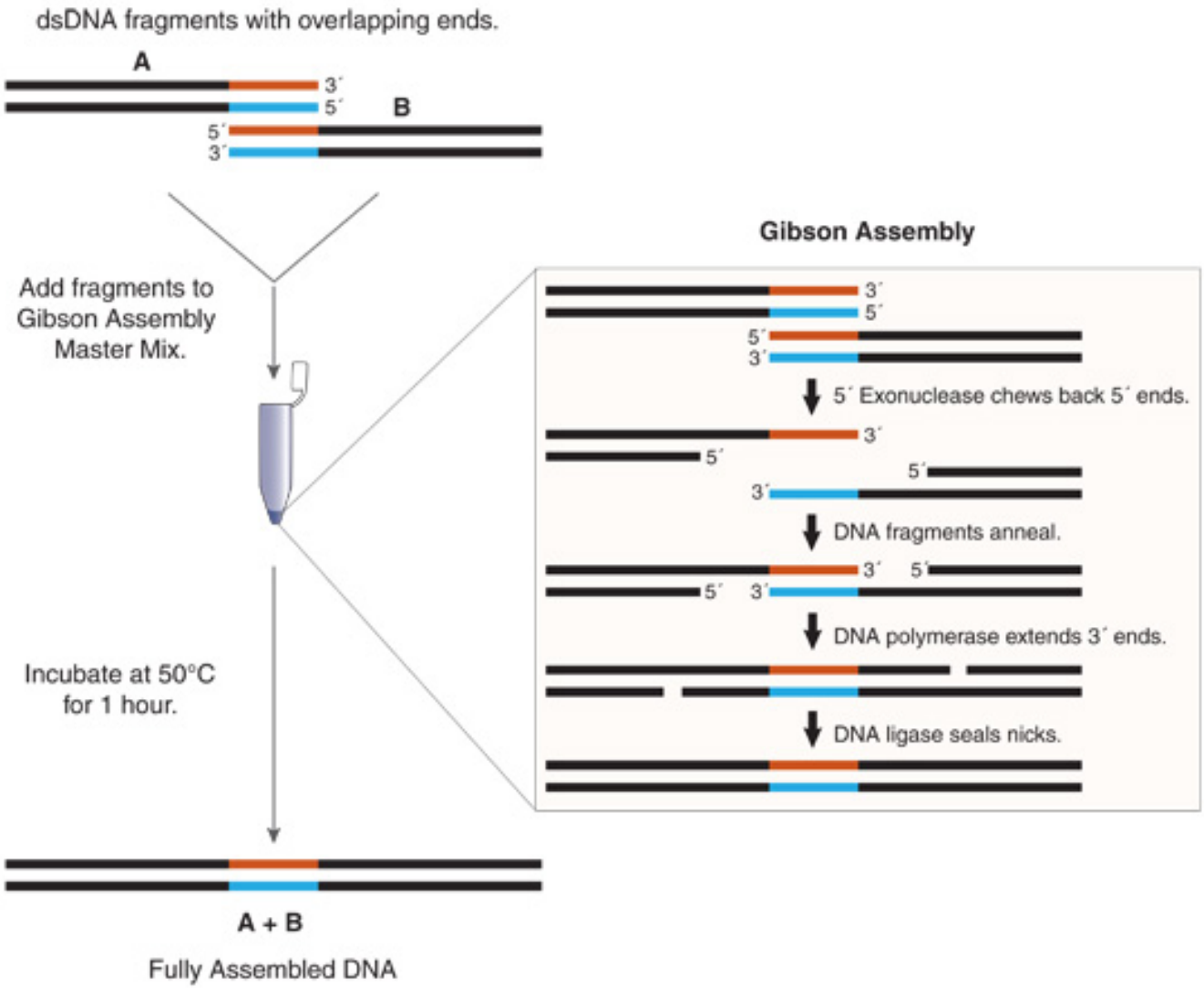
PCR

Gene synthesis by PCR (continued)



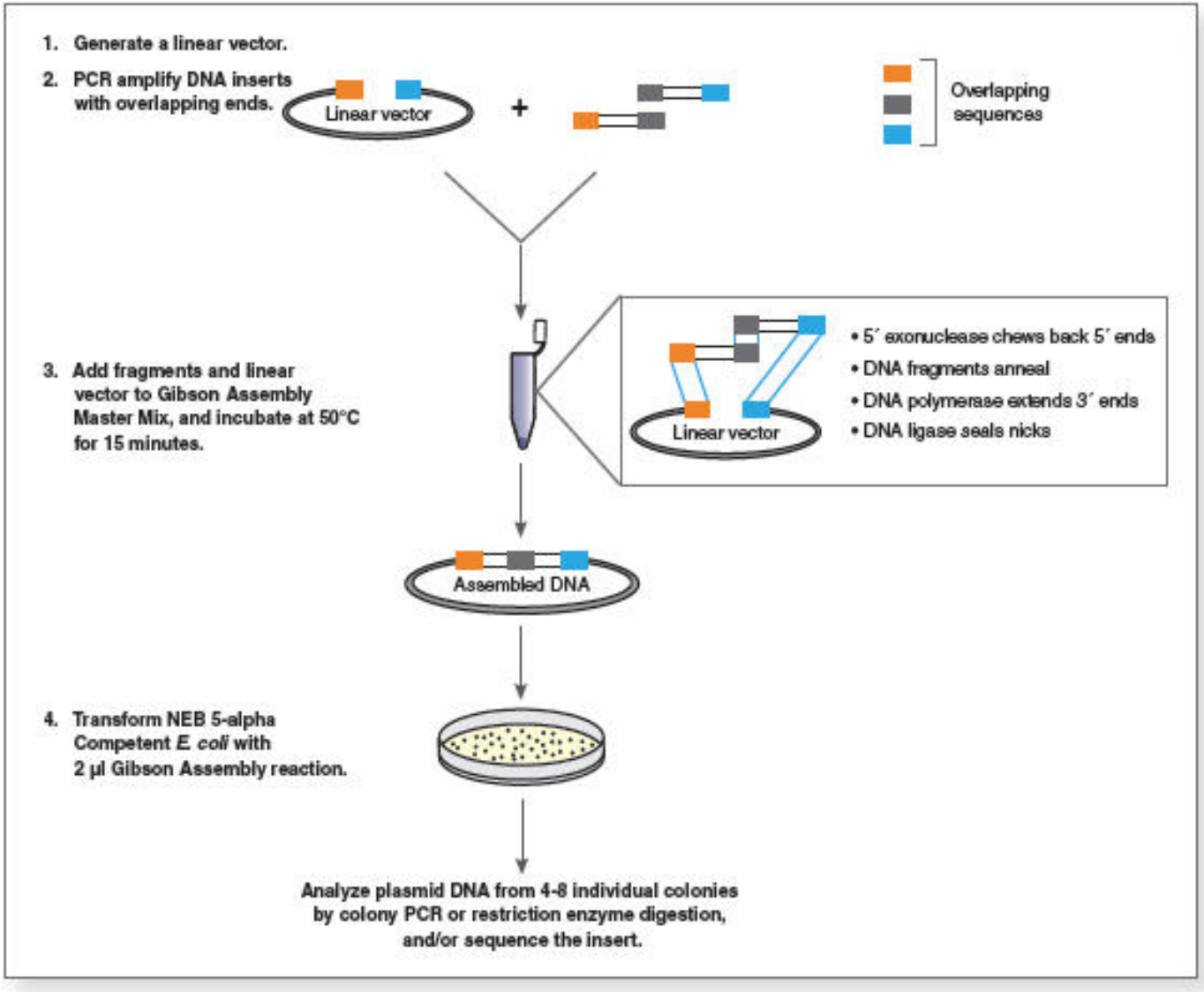
Gibson Assembly

Seamless one-pot assembly of DNA fragments



Gibson Assembly

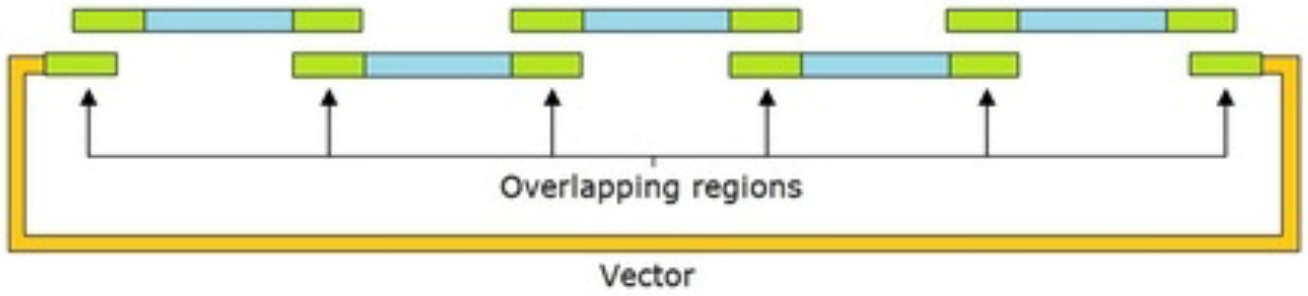
Seamless one-pot assembly of DNA fragments



Can assemble **multiple DNA fragments** directly into a **plasmid vector**

Gibson Assembly

Seamless one-pot assembly of DNA fragments

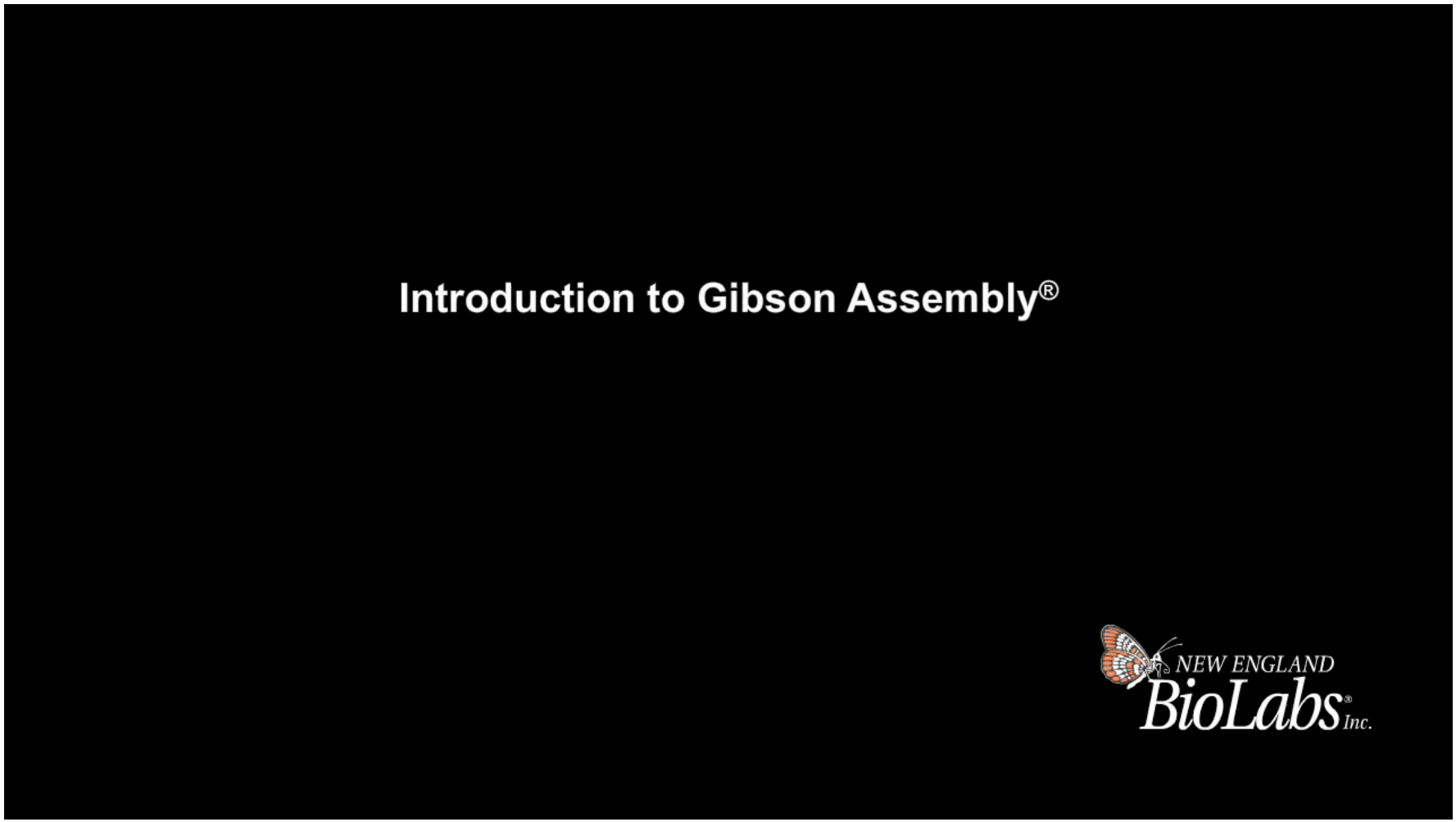


Can assemble **multiple DNA fragments** directly into a plasmid vector

All you need are **overlapping regions** at the end of each DNA fragment

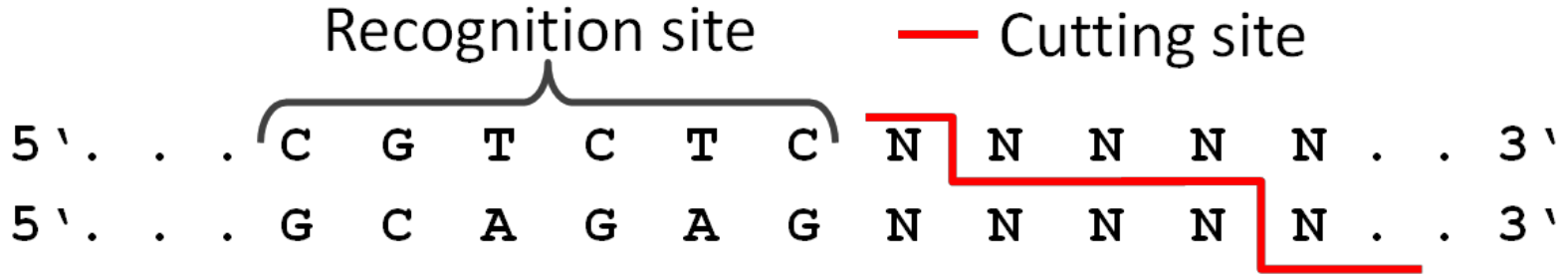
Gibson Assembly

Seamless one-pot assembly of DNA fragments



Golden Gate Assembly

Seamless assembly of DNA fragments



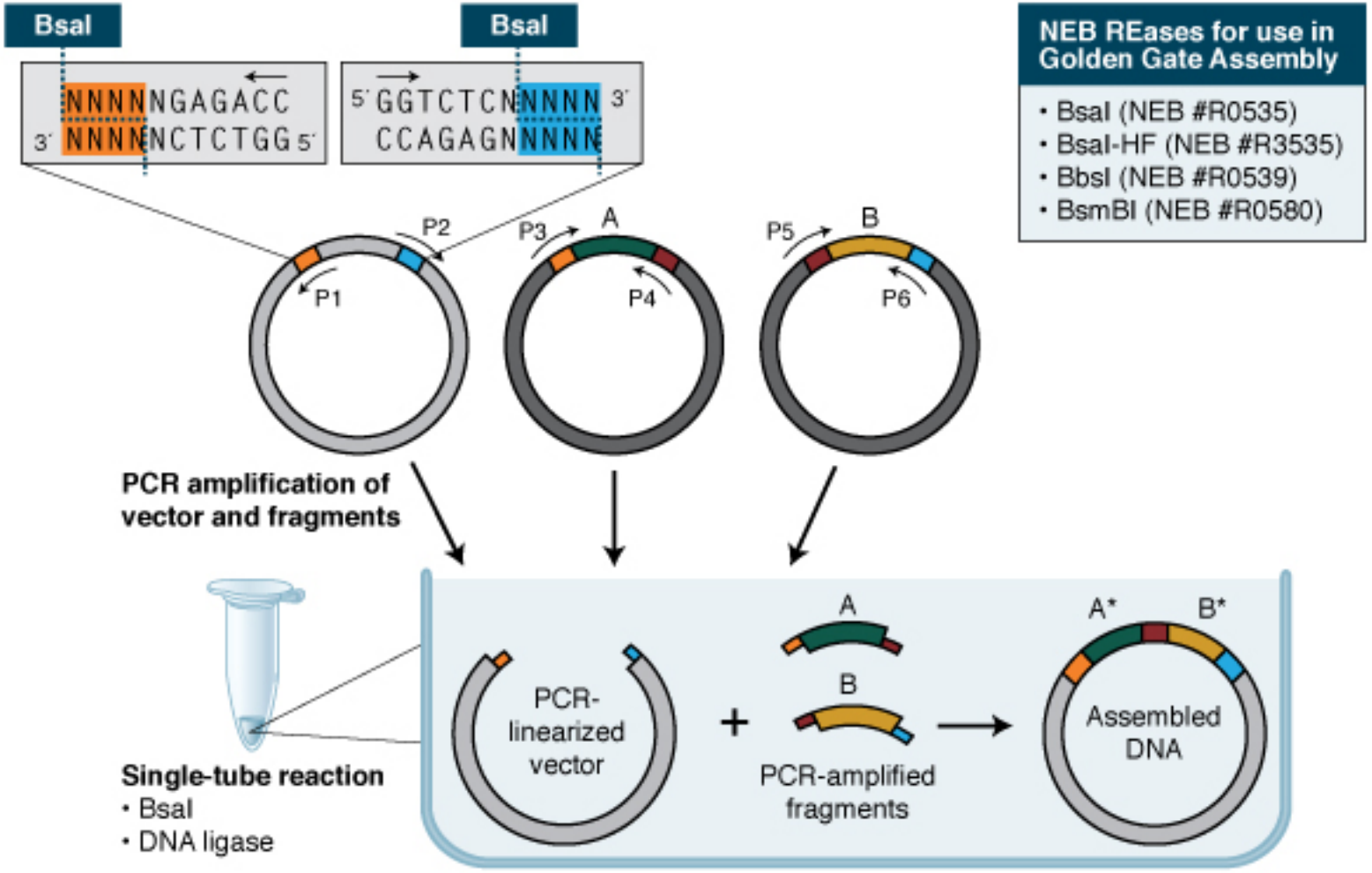
BbsI	5′	...GAAGAC(N)$_2$... 3′
	3′	...CTTCTG(N)$_6$... 5′
BsaI	5′	...GGTCTC(N)$_1$... 3′
	3′	...CCAGAG(N)$_5$... 5′
BsmBI	5′	...CGTCTC(N)$_1$... 3′
	3′	...GCAGAG(N)$_5$... 5′
SapI	5′	...GCTCTTC(N)$_1$... 3′
	3′	...CGAGAAG(N)$_4$... 5′

Based on using **type IIs restriction endonucleases**

These endonucleases cut **adjacent to non-palindromic recognition sites** enabling **seamless cloning**

Golden Gate Assembly

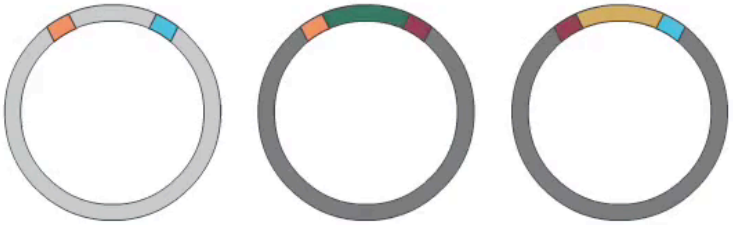
Seamless assembly of DNA fragments



* While A and B insert sequences involved in 4-base overlaps are shown in separate colors for clarity, the actual assembly is seamless; 4-base overlaps are insert derived.

Golden Gate Assembly

Seamless assembly of DNA fragments

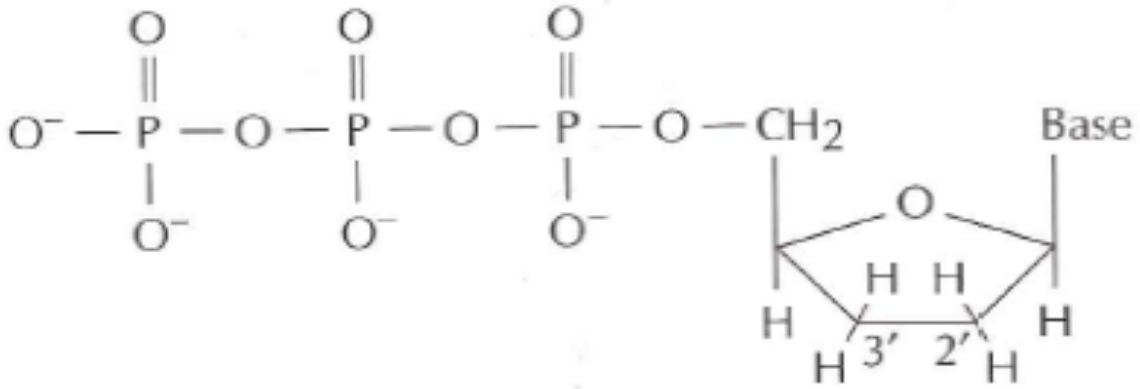


DNA Sequencing

DNA Sequencing

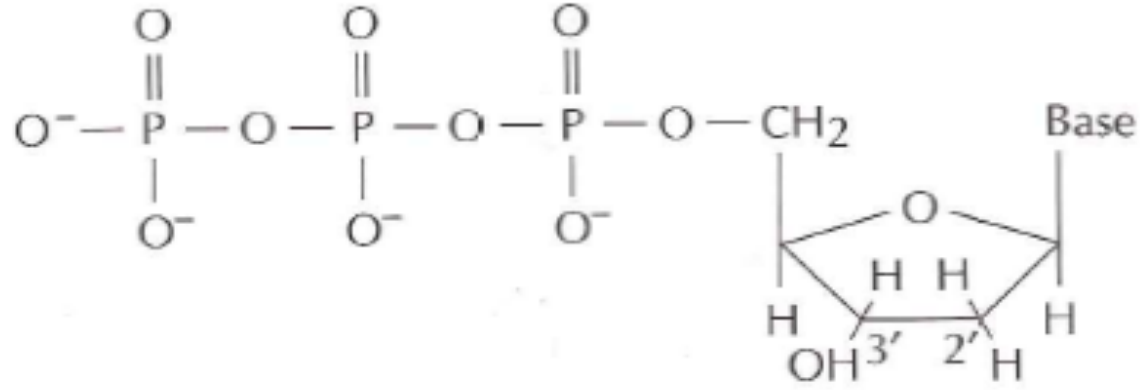
Dideoxynucleotide (Sanger) Sequencing

A



A dideoxyribonucleoside triphosphate

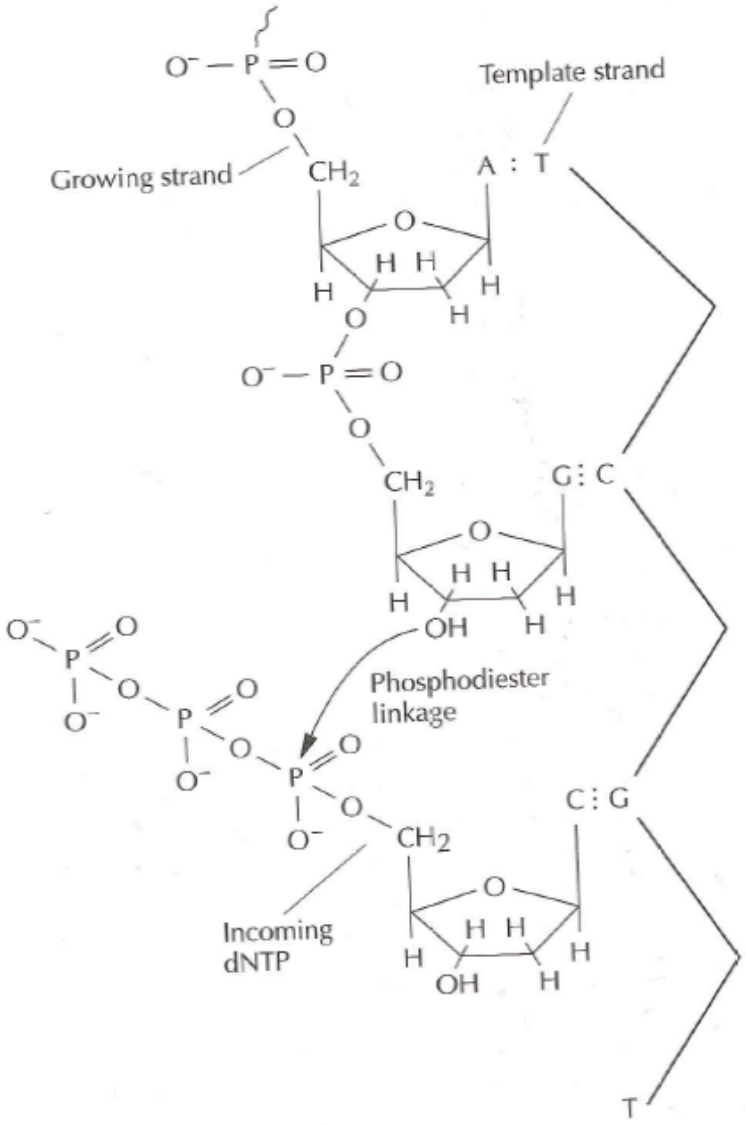
B



A deoxyribonucleoside triphosphate

DNA Sequencing

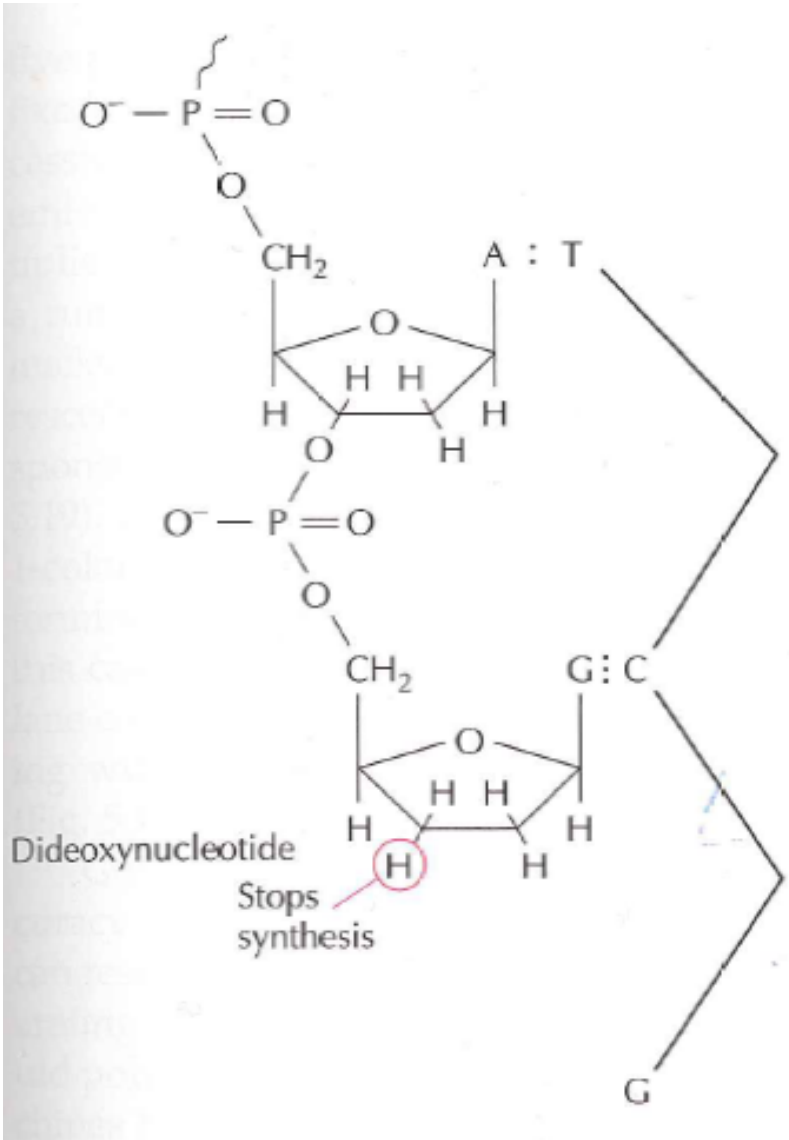
Dideoxynucleotide (Sanger) Sequencing



Normal DNA synthesis

DNA Sequencing

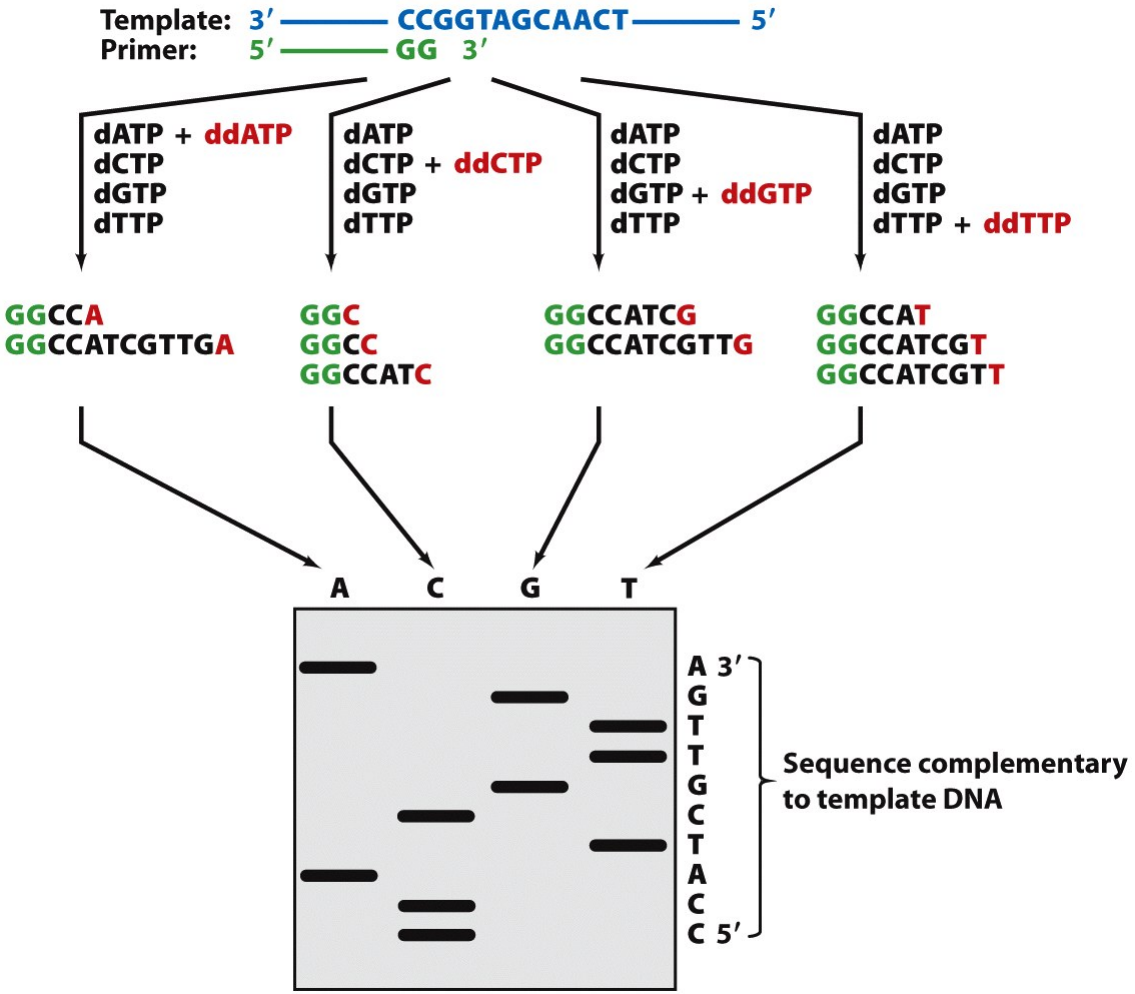
Dideoxynucleotide (Sanger) Sequencing



Blocked DNA synthesis with a dideoxynucleotide chain terminator

DNA Sequencing

Sanger sequencing – the chain terminator (dideoxy) method



The technique which Fred Sanger developed:

A short synthetic DNA ‘primer’ is base-paired to single stranded DNA

The reaction is split into 4 tubes containing a small amount of either ddATP, ddCTP, ddGTP or ddTTP

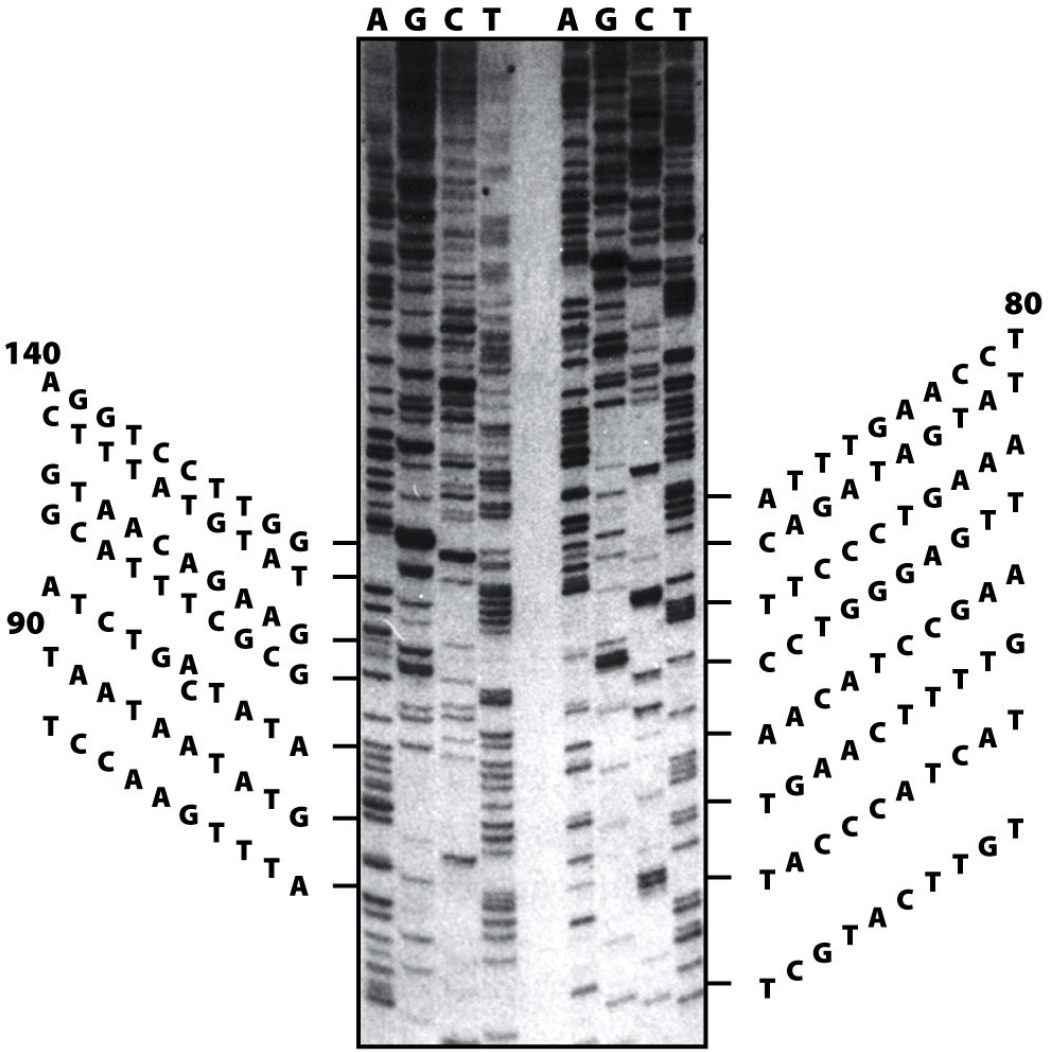
The polymerase generates a nested a series of DNA terminated by the dideoxy analog occupied by the corresponding base

The DNA fragments are separated by size on a polyacrylamide gel

The DNA is read bottom to top

DNA Sequencing

Sanger sequencing – the chain terminator (dideoxy) method



From Hindley, J., DNA sequencing, *in* Work, T.S. and Burdon, R.H. (Eds.), *Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 10, p. 82, Elsevier (1983). Used by permission.

The DNA is labelled by polymerizing with a radioactive dNTP

Exposing an X-ray film laid over the gel generates an autoradiograph

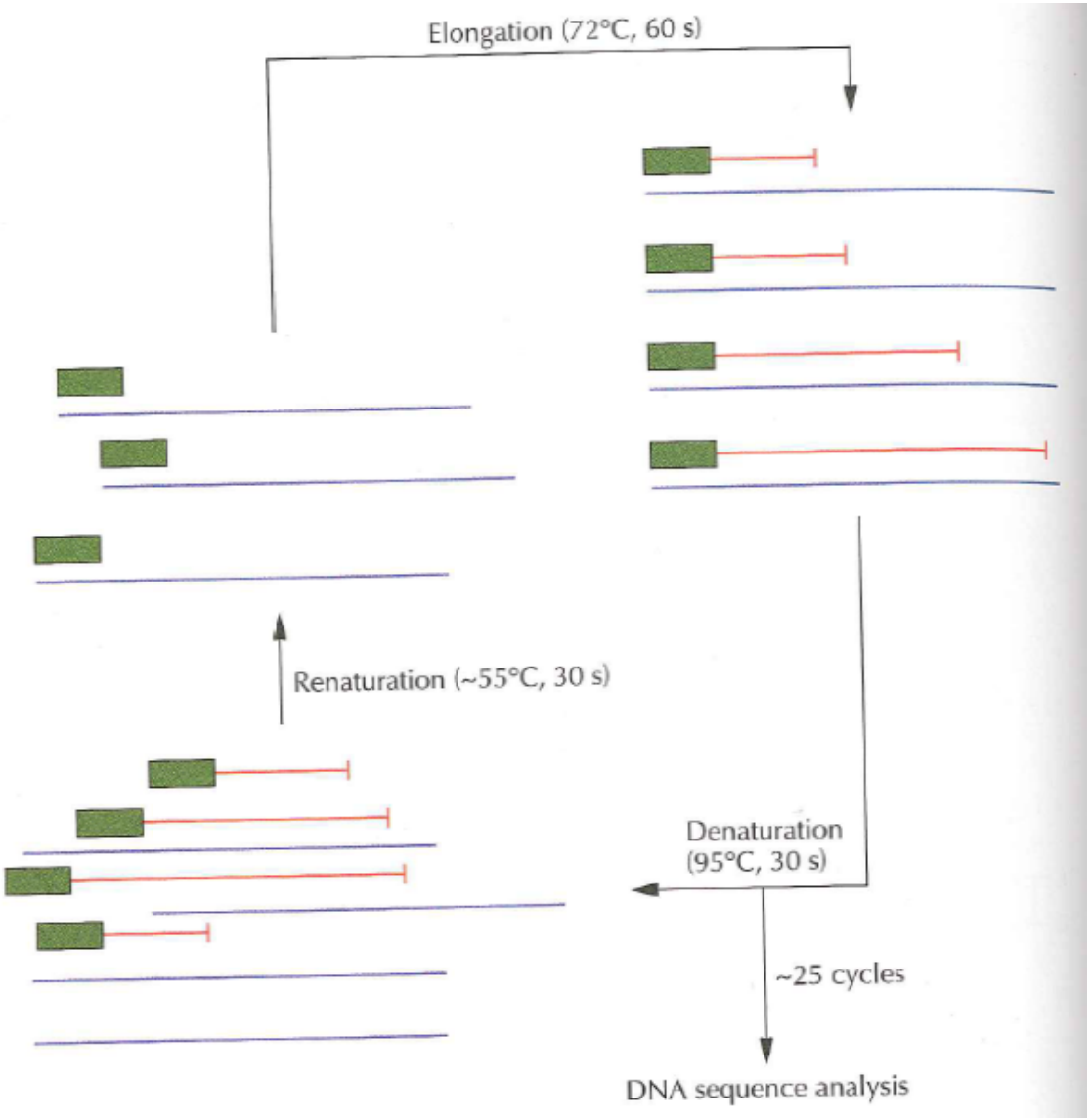
The sequence can be read from the autoradiograph

DNA Sequencing

Dideoxynucleotide (Sanger) Sequencing

Cycle Sequencing

Thermocycling using a thermostable DNA polymerase (e.g. Taq Polymerase)



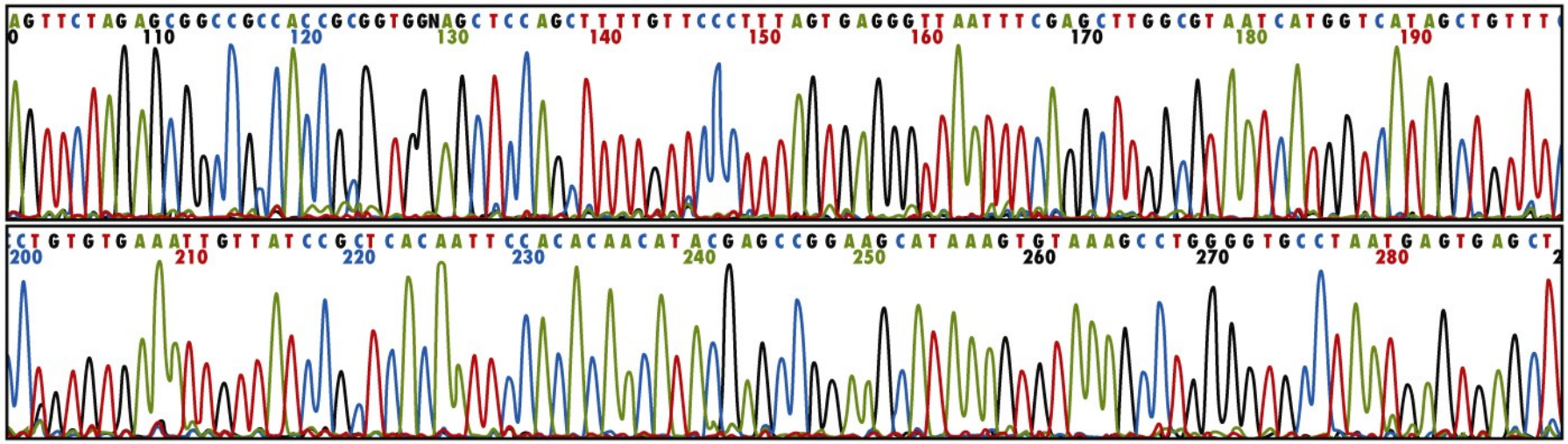
Primer oligonucleotide:
green

Template DNA:
blue

Dideoxynucleotide terminated fragments:
red

DNA Sequencing

Automated Sanger sequencing using fluorescent dye terminators



Courtesy of Mark Adams, The Institute for Genomic Research, Rockville, Maryland

Today Sanger sequencing is automated.

A single polymerization reaction is performed using a fluorescent chain-terminator with a different colour for each base.

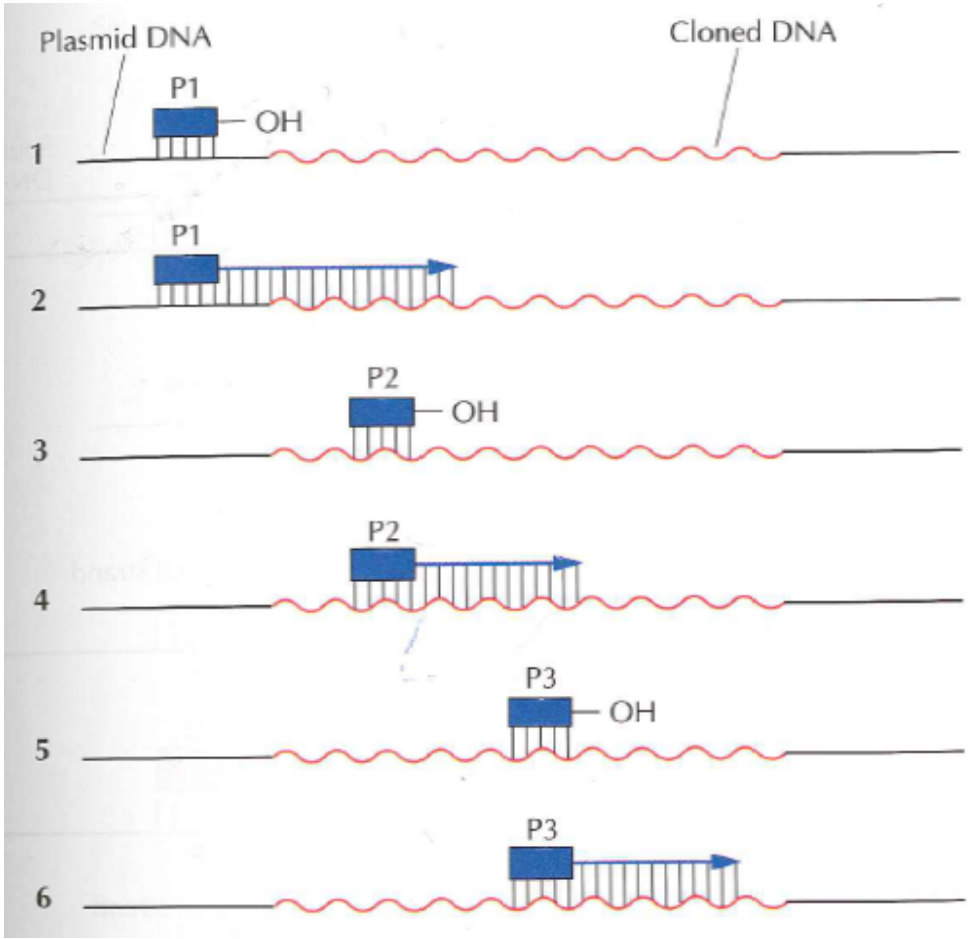
The DNA fragments are separated by capillary electrophoresis

The fluorescence of the fragment is measured as they leave the capillary producing an electrophoregram

Green, red, black and blue peaks correspond to A, T, G and C, respectively

DNA Sequencing

DNA sequencing by primer walking



Can only sequence up to **~1000 bases** using Sanger sequencing

Longer sequences can be sequenced **by primer walking**

Sequencing starts with a primer (P1) complementary to a site on the plasmid
Based on the sequence from P1 a second primer (P2) is designed
Based on the sequence from P2 a third primer (P") is designed, etc…

DNA Sequencing

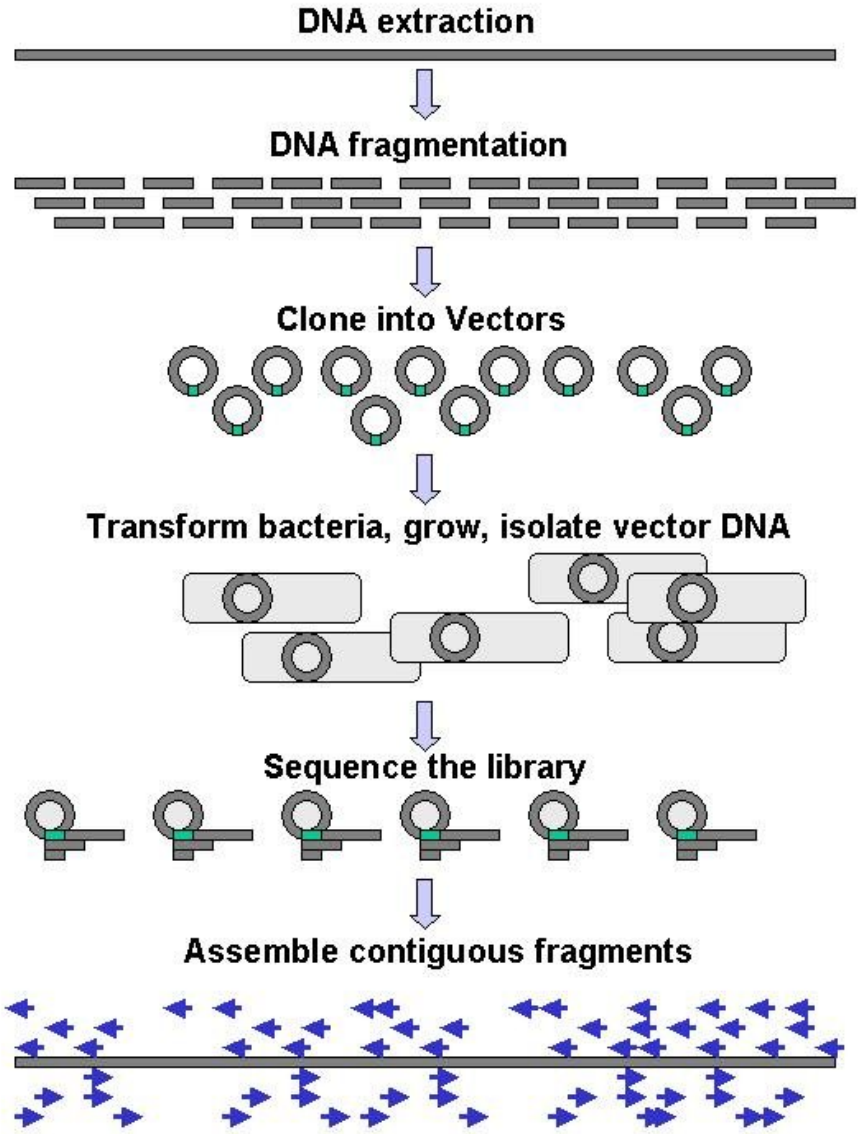
Automated sequencing has allowed entire genomes to be sequenced

Table 3-3 Some Sequenced Genomes

Organism	Genome Size (kb)	Number of Chromosomes
Mycoplasma genitalium (human parasite)	580	1
Rickettsia prowazekii (putative relative of mitochondria)	1,112	1
Haemophilus influenzae (human pathogen)	1,830	1
Escherichia coli (human symbiont)	4,639	1
Saccharomyces cerevisiae (baker's yeast)	11,700	16
Plasmodium falciparum (protozoan that causes malaria)	30,000	14
Caenorhabditis elegans (nematode)	97,000	6
Arabidopsis thaliana (dicotyledonous plant)	117,000	5
Drosophila melanogaster (fruit fly)	137,000	4
Oryza sativa (rice)	390,000	12
Danio rerio (zebra fish)	1,700,000	25
Gallus gallus (chicken)	1,200,000	40
Mus musculus (mouse)	2,500,000	20
Homo sapiens	3,200,000	23

DNA Sequencing

Shotgun sequencing of whole genomes or chromosomes



Genomic DNA is fragmented using restriction enzymes or shearing

The fragmented DNA is cloned into a DNA vector and amplified in *E. coli*.

The short DNA fragments from individual bacteria are individually sequenced and assembled electronically into one long contiguous sequence

Primer walking can be used to fill gaps

DNA Sequencing

Public human genome sequencing

DNA Sequencing

Private human genome sequencing

Next Generation Sequencing Systems

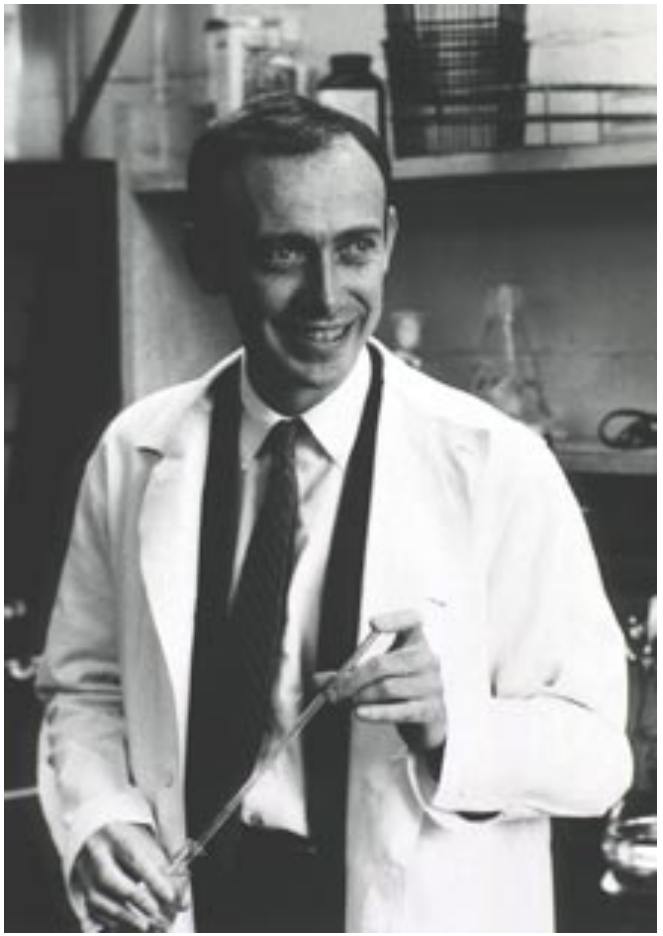
Using novel non-Sanger systems

- Roche Genome Sequencer
- Applied Biosystems SOLiD™
- LifeTech PGM™ sequencer
- Complete Genomics
- Illumina
- PacBio
- Oxford Nanopore

A whole bacterial genome can be
Sequenced in a single experiment in 4 hours

Used to sequence the genome of Jim Watson!

$1.5 million and 4 months -
compared to $3 billion and over 10 years
for the Human Genome Project's reference genome

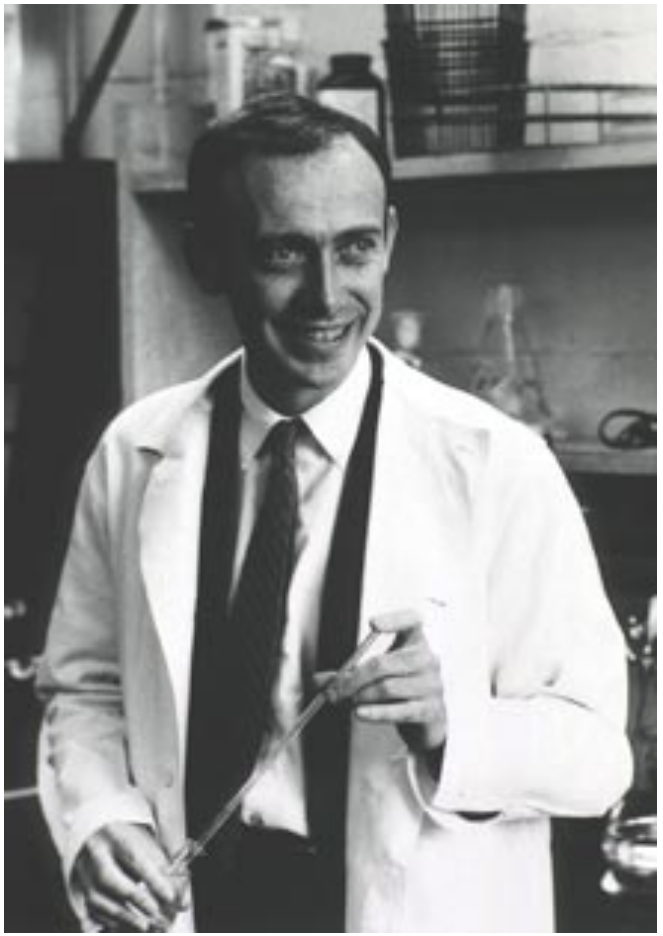


Next Generation Sequencing Systems

Using novel non-Sanger systems

- Roche Genome Sequencer
- Applied Biosystems SOLiD™
- LifeTech PGM™ sequencer
- Complete Genomics
- Illumina
- PacBio
- Oxford Nanopore

A whole bacterial genome can be
Sequenced in a single experiment in 4 hours



Used to sequence the genome of Jim Watson!

$1.5 million and 4 months -
compared to $3 billion and over 10 years
for the Human Genome Project's reference genome

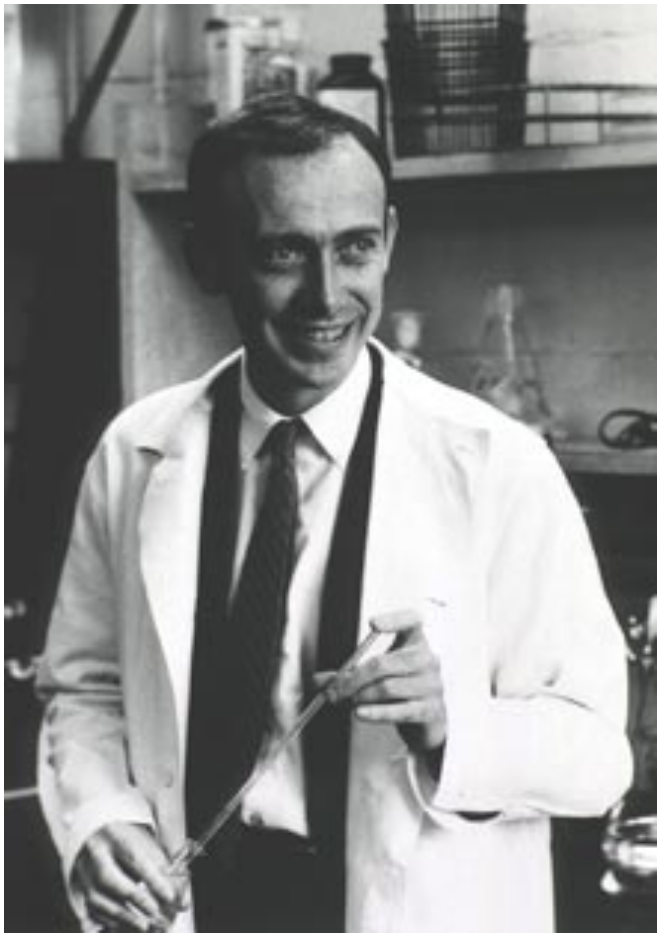
A human genome can now be sequenced for ~$8,000 in ~2 days

Next Generation Sequencing Systems

Using novel non-Sanger systems

- Roche Genome Sequencer
- Applied Biosystems SOLiD™
- LifeTech PGM™ sequencer
- Complete Genomics
- Illumina
- PacBio
- Oxford Nanopore

A whole bacterial genome can be
Sequenced in a single experiment in 4 hours



Used to sequence the genome of Jim Watson!

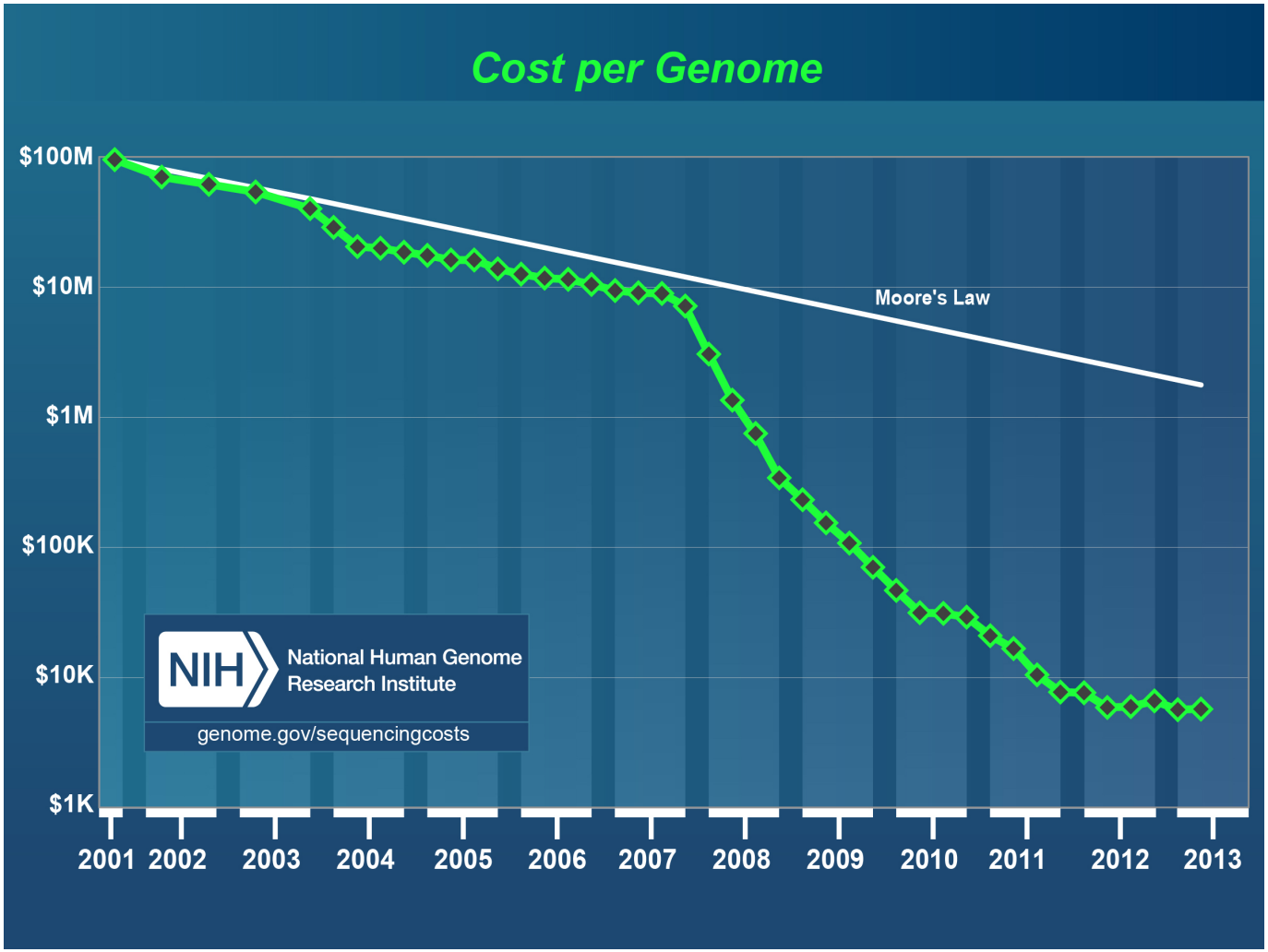
$1.5 million and 4 months -
compared to $3 billion and over 10 years
for the Human Genome Project's reference genome

A human genome can now be sequenced for ~$8,000 in ~2 days

16% of his genes were likely to have come from a black ancestor of African descent

Next-Generation Sequencing

Using novel non-Sanger systems



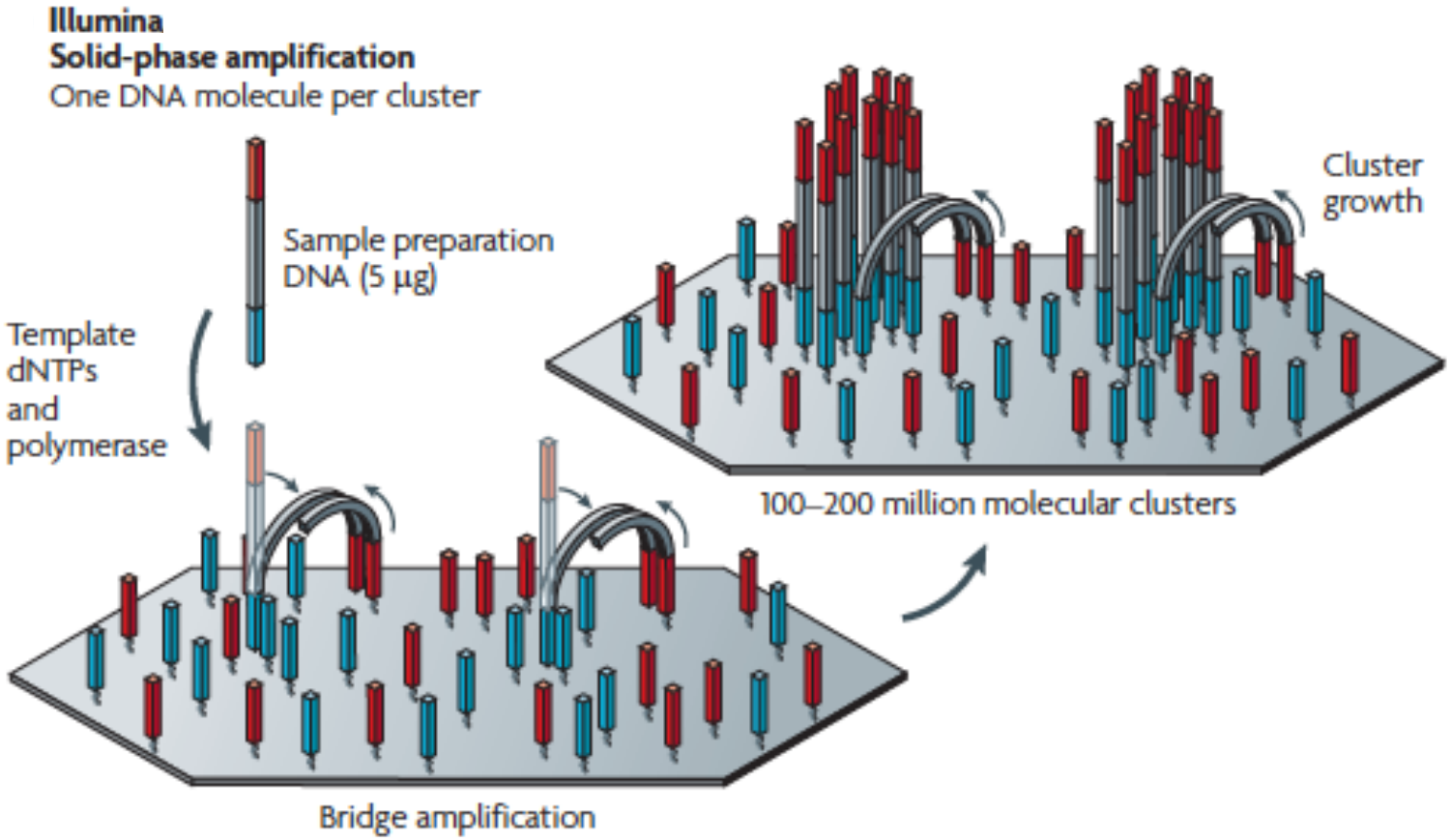
Human Genome Project's reference genome:

- $3 billion >10 years

A human genome can now be sequenced for:

- ~$8,000 in ~2 days

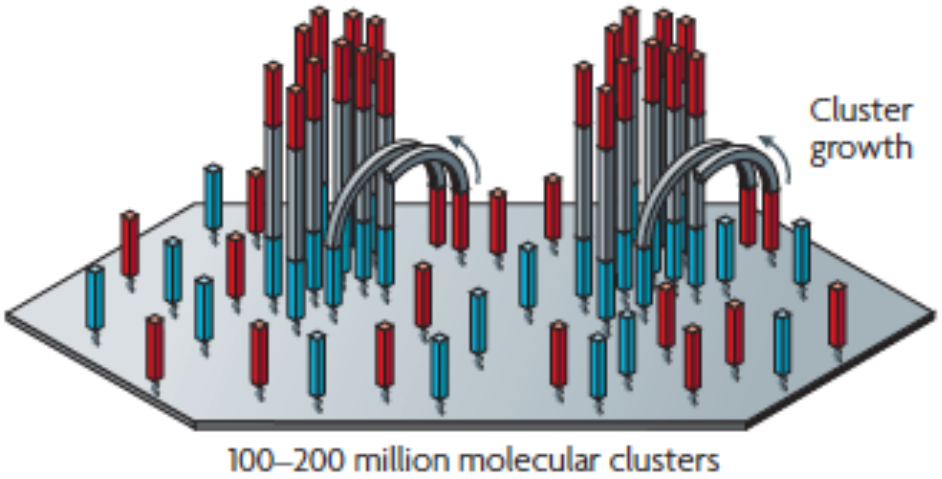
Next-Generation Sequencing
Illumina



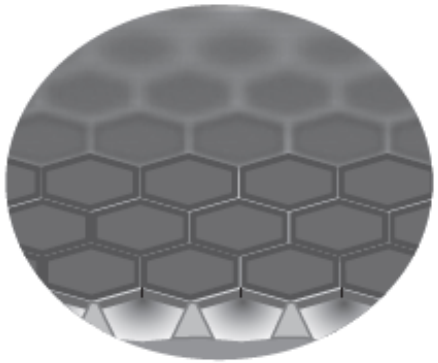
DNA amplification to form 'clusters'

The DNA molecules in each cluster contains the same sequence

Next-Generation Sequencing
Illumina



Patterned flow cell
Microwells on flow cell direct cluster generation, increasing cluster density

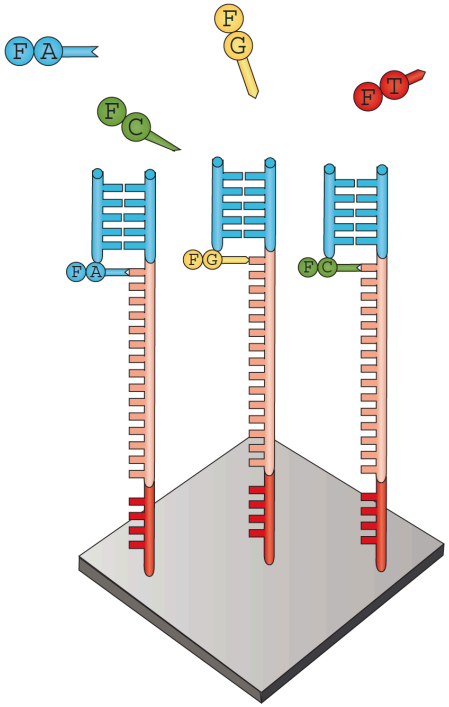


DNA amplification to form 'clusters'

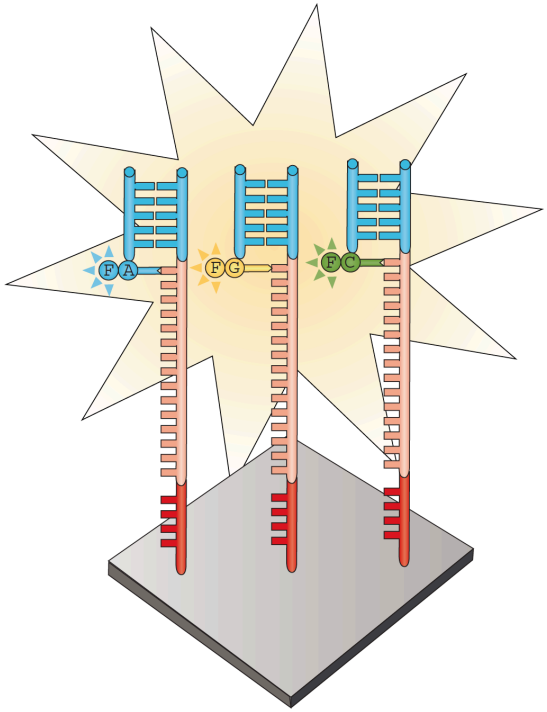
The DNA molecules in each cluster have the same sequence

Next-Generation Sequencing

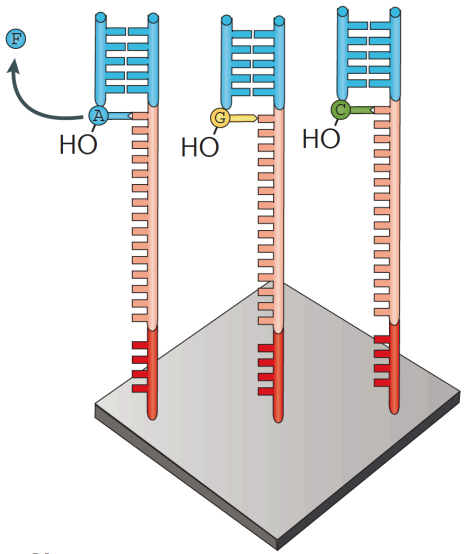
Illumina



Nucleotide addition
Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.



Imaging
Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.



Cleavage
Fluorophores are cleaved and washed from flow cells and the 3′-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

Sequencing by synthesis using 3’-blocked reversible chain terminators

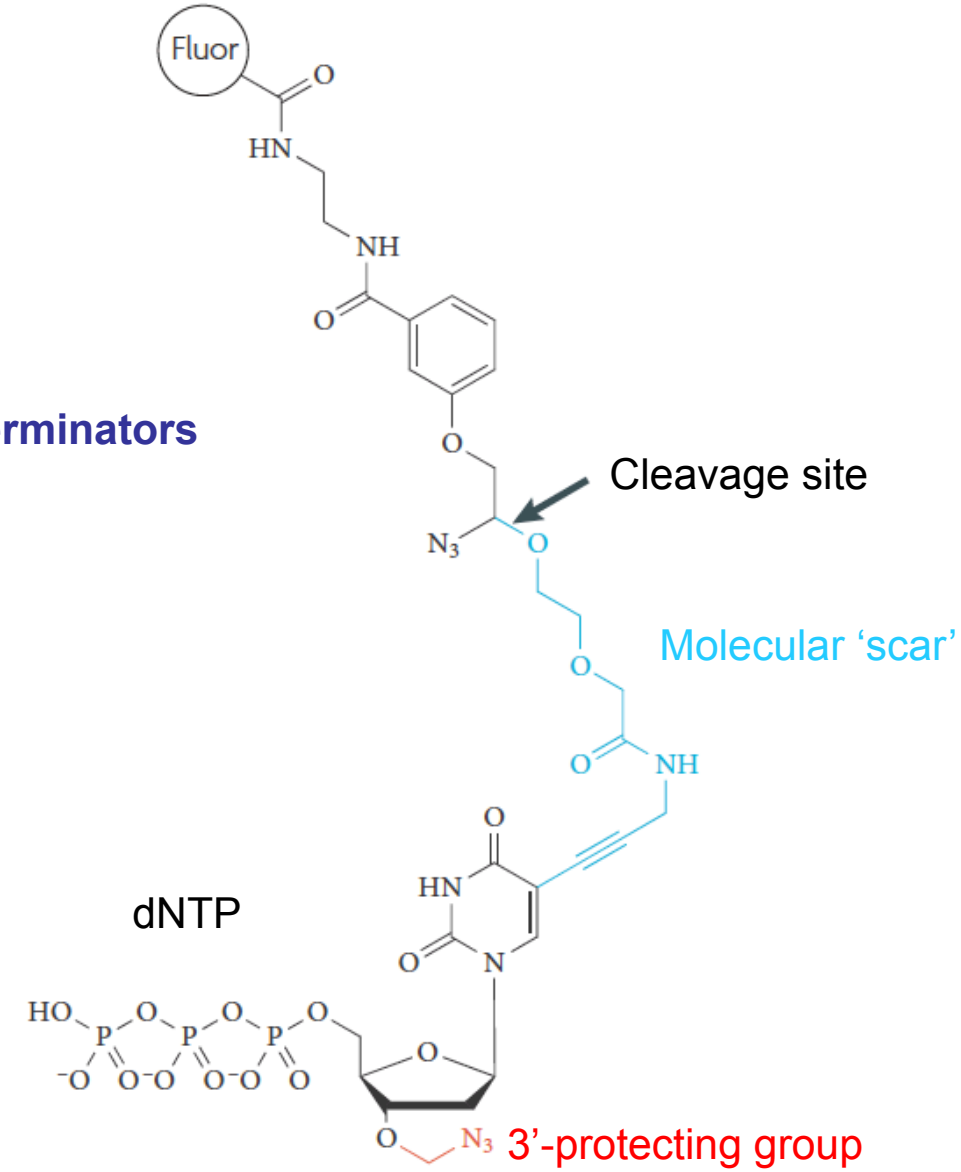
Each modified dNTP carries a different coloured fluorescent group on the base and a protecting group on the 3’-OH

Both are removed after each cycle

Next-Generation Sequencing

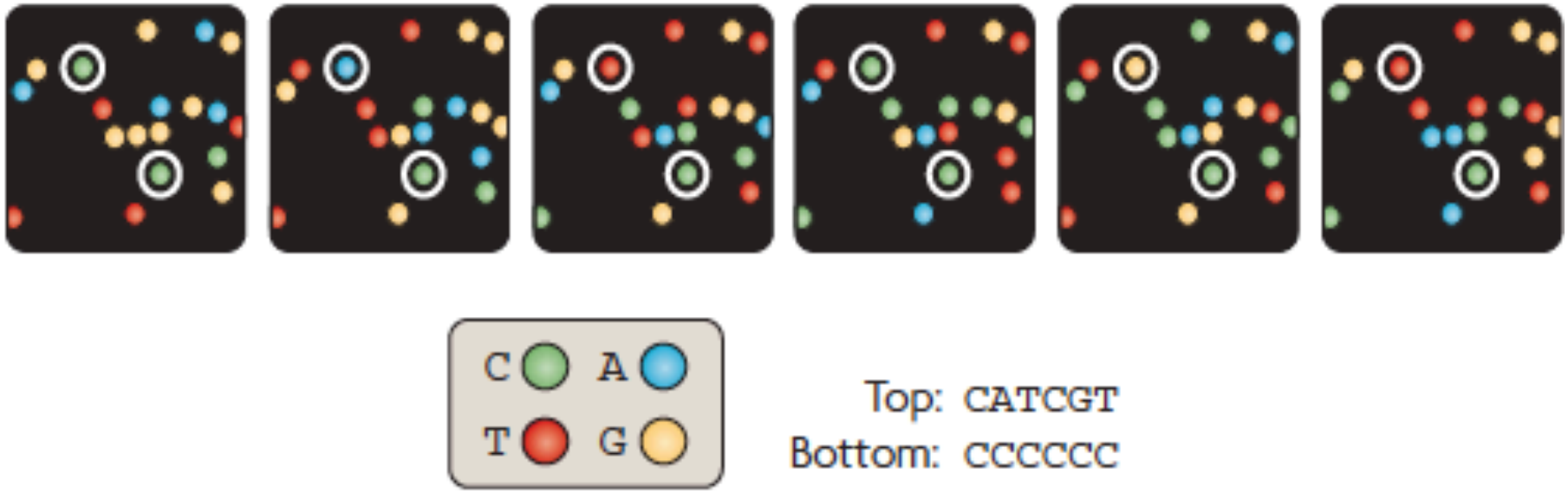
Illumina

3’-blocked reversible terminators



Next-Generation Sequencing

Illumina

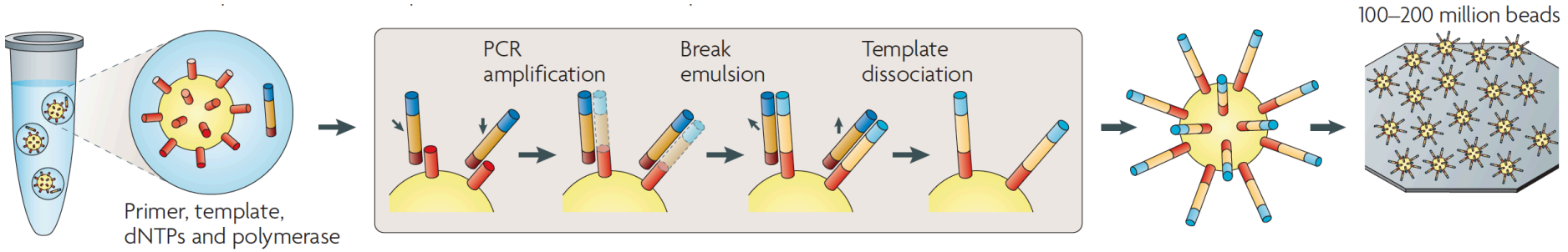


Next-Generation Sequencing

Illumina

Next-Generation Sequencing

Ion-Torrent



DNA amplification by emulsion PCR

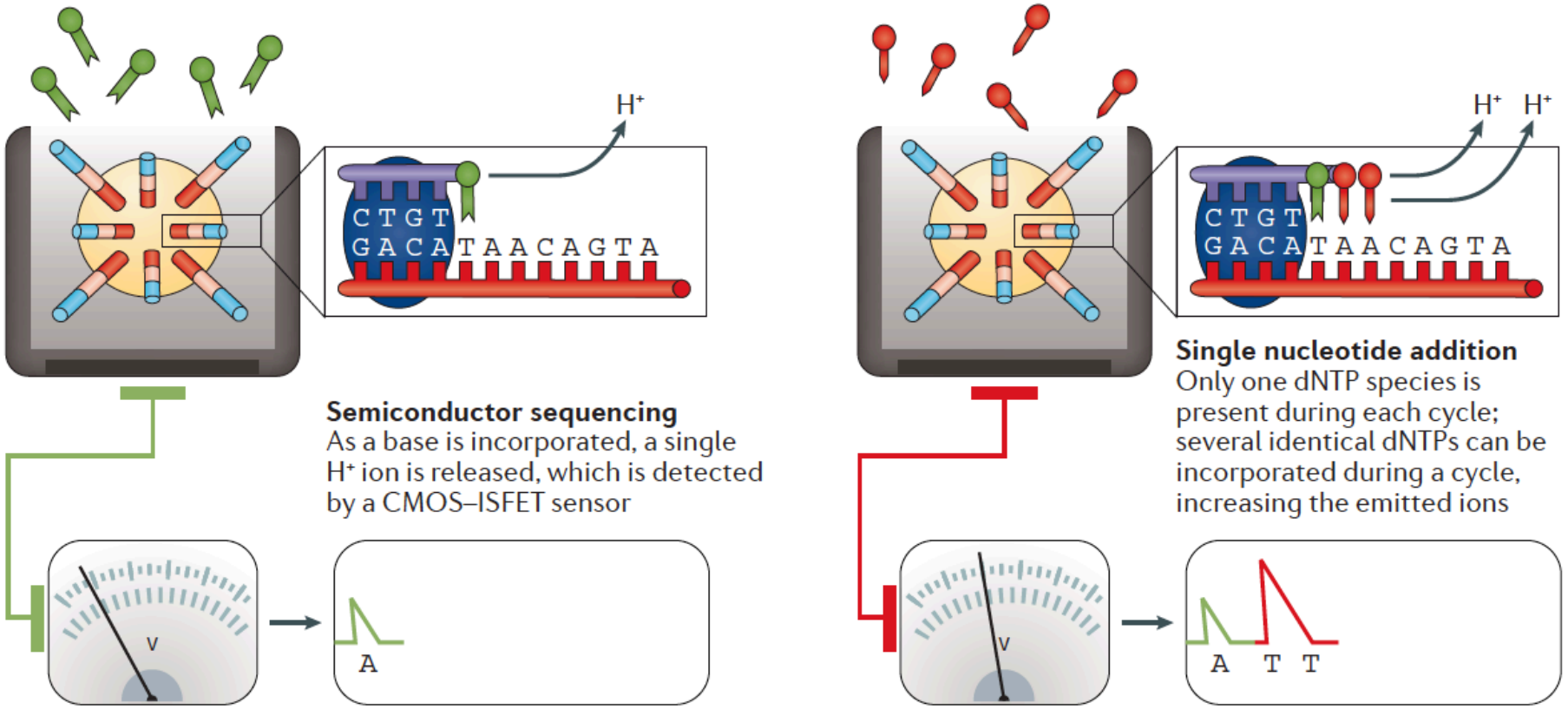
An emulsion is made comprising water droplets containing single beads and single template DNA molecules

The template DNA molecules are amplified in drops by PCR and captured on the beads (one of the PCR primers is on the beads)

The beads then carry many DNA molecules and the DNA molecules on each bead have the same sequence

Next-Generation Sequencing

Ion-Torrent



Sequencing by synthesis using semiconductor chips

Beads carrying many identical copies of the DNA to be sequenced are loaded into microfabricated wells on a semiconductor chip (one bead per well)

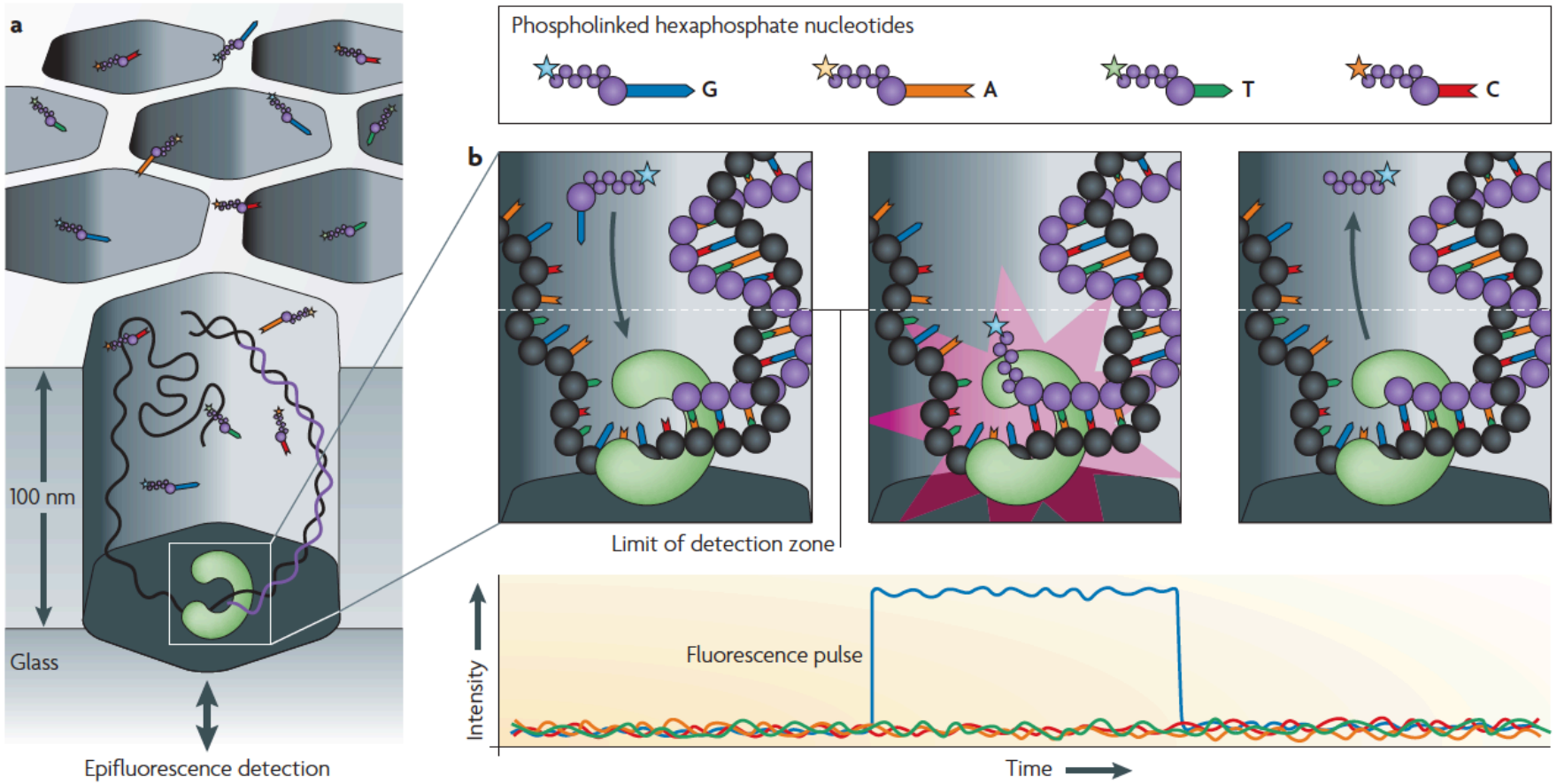
In each cycle one unmodified dNTP is added (C, A, G or T) and incorporation is measured by detecting the change in pH

Next-Generation Sequencing

Ion-Torrent

Next-Generation Sequencing

Pacific Biosciences



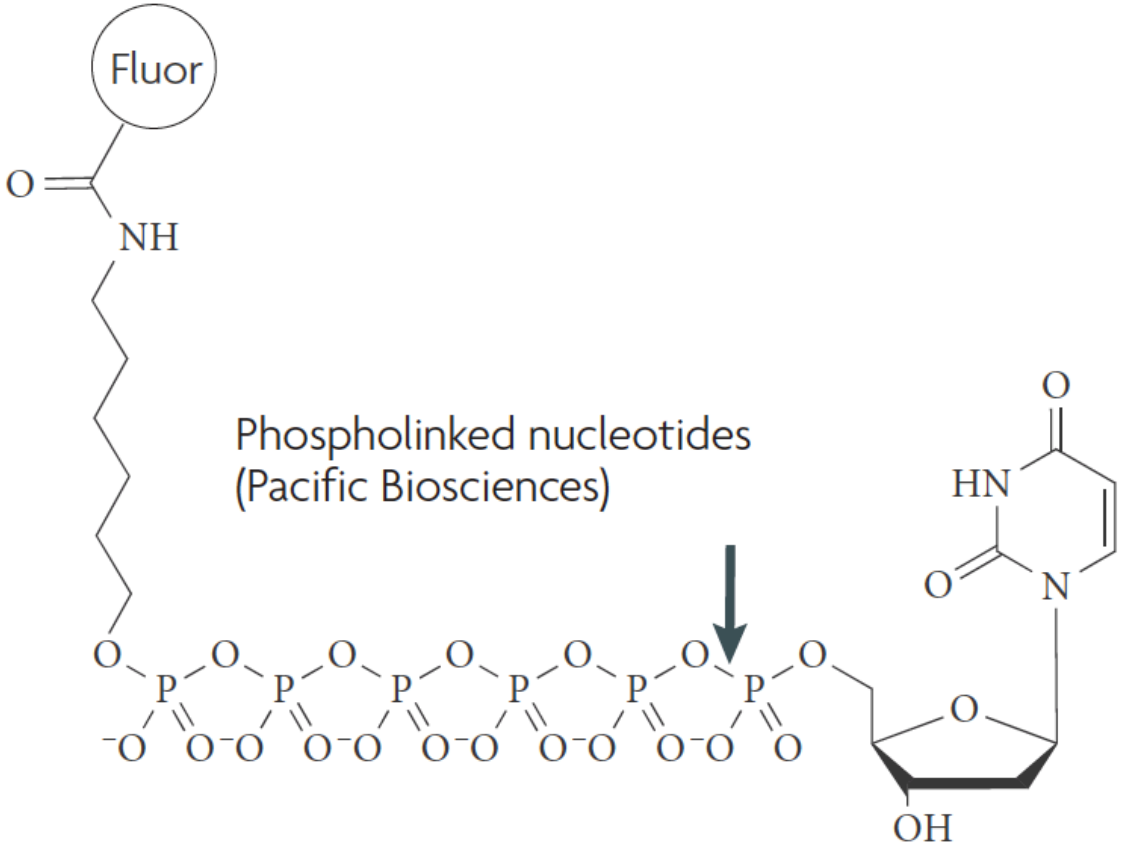
Single molecule sequencing using phospholinked hexaphosphate nucleotides

Four modified nucleotides are used (C, A, G and T), each with a different coloured fluorophore

When a nucleotide is incorporated it stays longer in the active site of the polymerase at the bottom a a zero-mode waveguide (ZMW) and can be detected

Next-Generation Sequencing

Pacific Biosciences



Single molecule sequencing using phospholinked hexaphosphate nucleotides

Four modified nucleotides are used (C, A, G and T), each with a different coloured fluorophore

When a nucleotide is incorporated it stays longer in the active site of the polymerase at the bottom a a zero-mode waveguide (ZMW) and can be detected

Next-Generation Sequencing

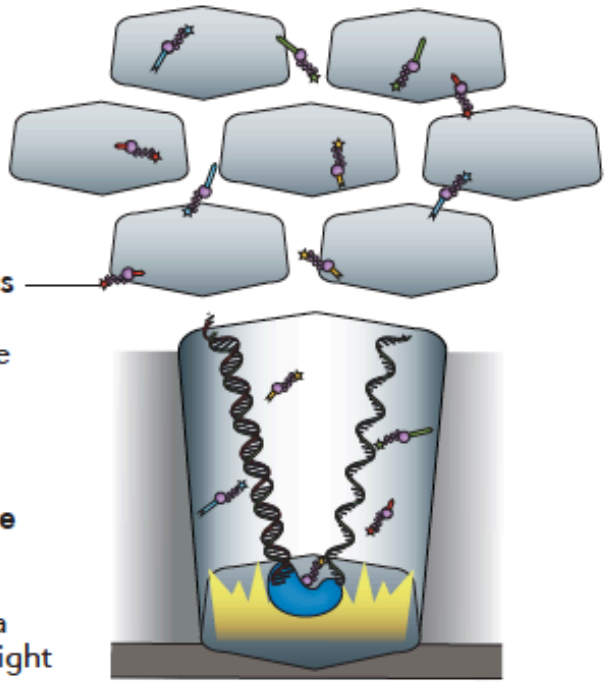
Pacific Biosciences

SMRTbell template
Two hairpin adapters allow continuous circular sequencing

ZMW wells
Sites where sequencing takes place

Labelled nucleotides
All four dNTPs are labelled and available for incorporation

Modified polymerase
As a nucleotide is incorporated by the polymerase, a camera records the emitted light



PacBio output
A camera records the changing colours from all ZMWs; each colour change corresponds to one base

Single molecule sequencing using phospholinked hexaphosphate nucleotides

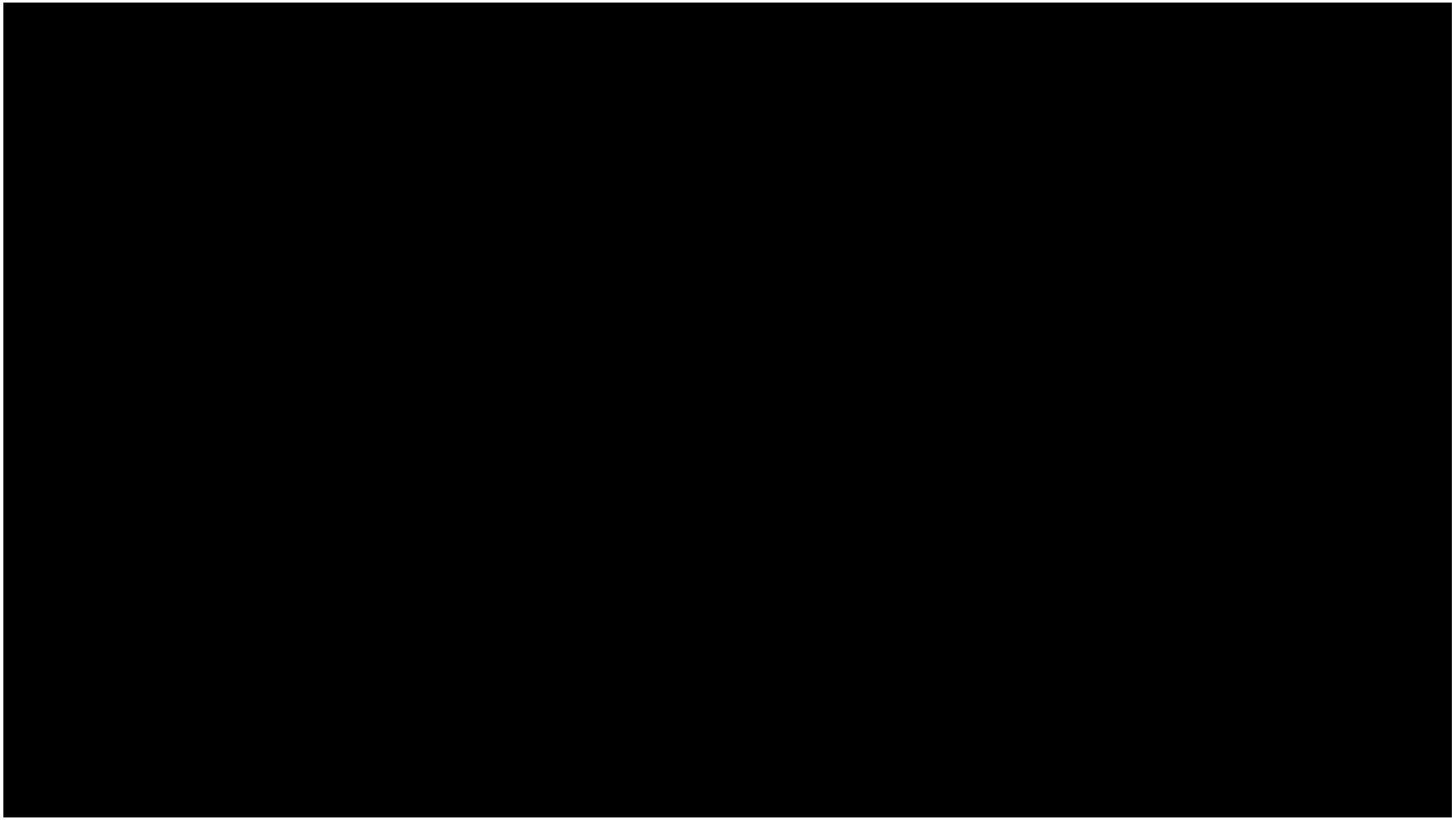
This single-molecule technique allows much longer reads than techniques such as Illumina and Ion Torrent (~20 Kb versus ≤0.3 kb Illumina and ≤1 Kb Ion Torrent)

But is more error prone (13% indel versus 0.1% substitution Illumina and 1% indel Ion Torrent)

However, the use of SMRTbell templates allows the same DNA to be sequenced many times, greatly increasing fidelity (to ≤1% indel)

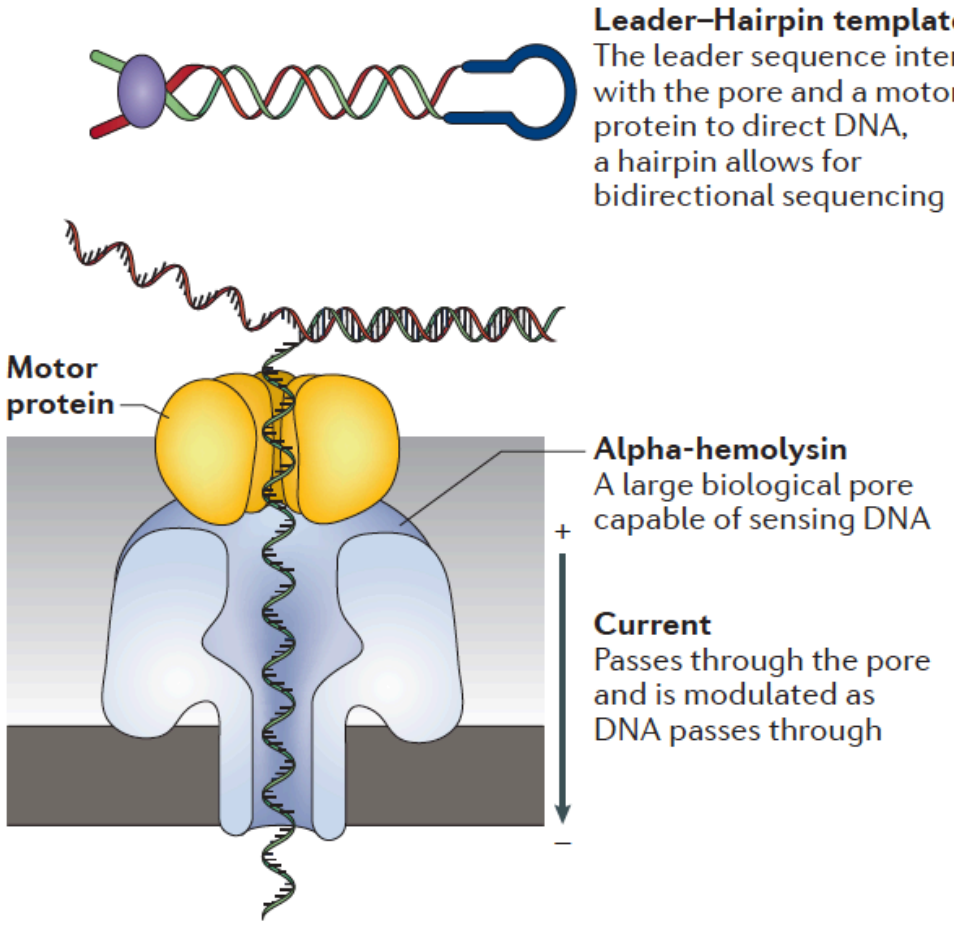
Next-Generation Sequencing

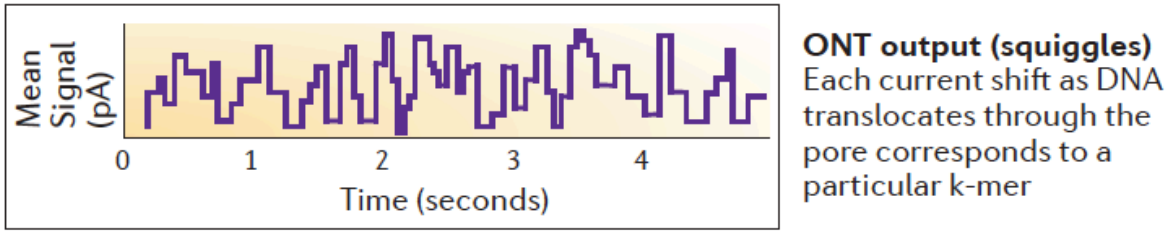
Pacific Biosciences



Next-Generation Sequencing

Oxford Nanopore





Single molecule sequencing using a membrane pore protein – 'strand sequencing'

As the DNA is passed through the pore by the motor protein it modifies the current passing through the pore.

The temporal voltage trace (squiggle space) is characteristic of the DNA sequence

The machine detects >1000 k-mers - 3-6 bases long (rather than individual nucleotides).

Allows very long reads up to 200 Kb

But with high error rate ~12% indel

Next-Generation Sequencing

Oxford Nanopore

Biotechnologie Moléculaire

3ème année

URL for films from Cold Spring Harbor Dolan DNA learning Center:

http://www.dnalc.org/resources/animations/

You can download the films for PC or for Mac