Introduction to biology
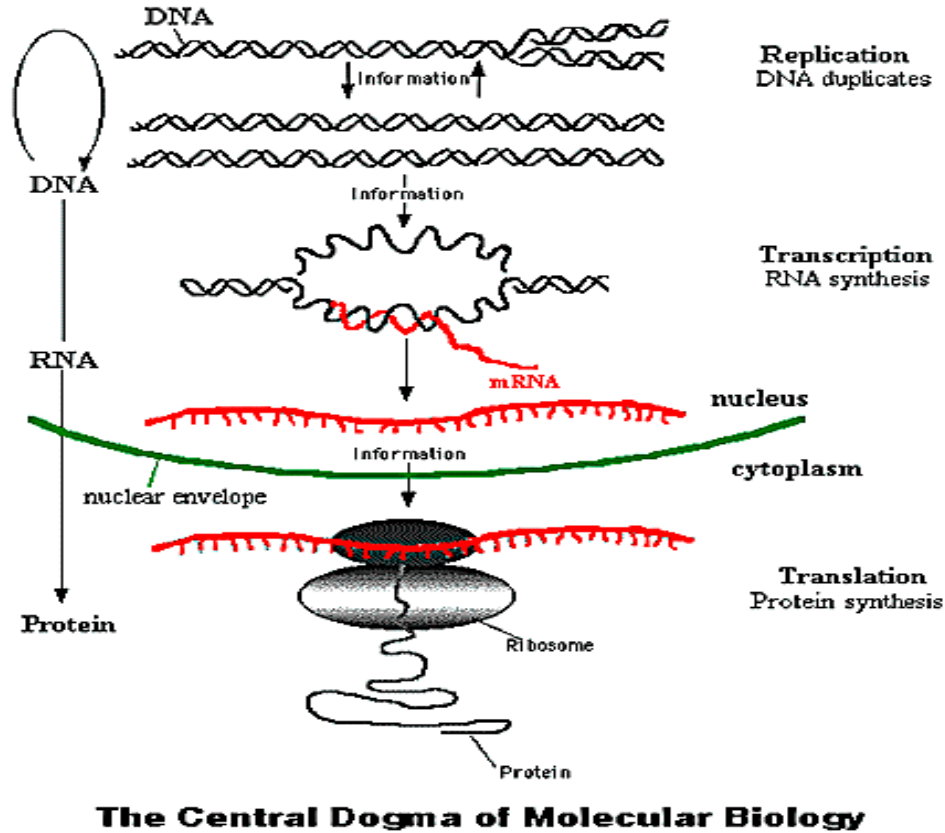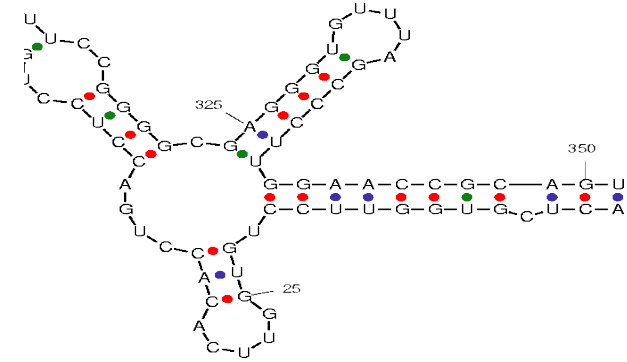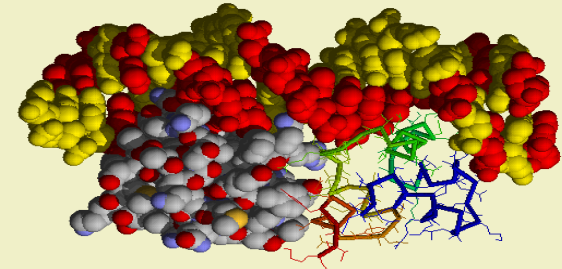## CMSC423

Central dogma of molecular biology

AGGTACGCGTACCTGACAGG

Phage CRO Repressor on DNA. Andrew Coulson & Roger Sayle with RasMol, University of Edinburgh, 1993

DNA – the code of life

- Purines A, G, caffeine
- Pyrimidines C, T
- Sugar backbone (ticker tape)
- Double-stranded – allows replication

Adenine
Thymine

Guanine
Cytosine

phosphate -deoxyribose backbone

3' end

5' end

H-bond

Template Strands

Replication Fork

DNA Polymerase

Okazaki Fragments

DNA Ligase

Lagging Strand

Leading Strand

pictures from wikipedia
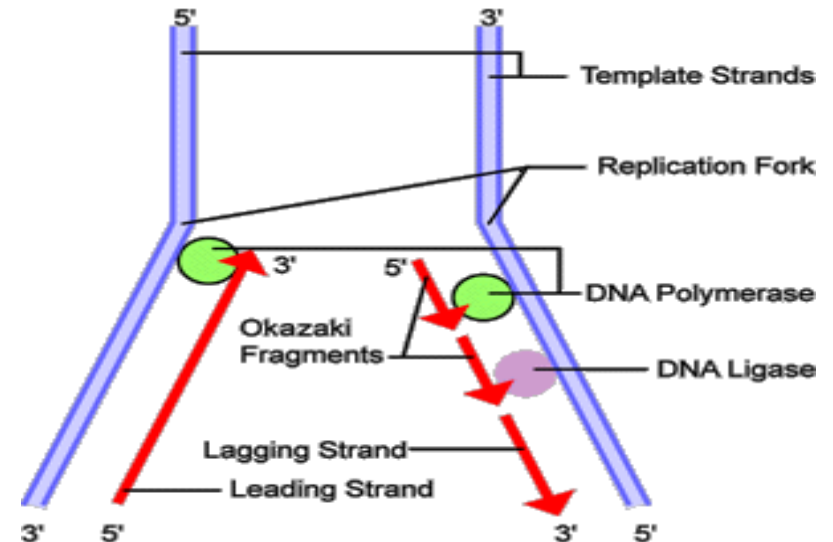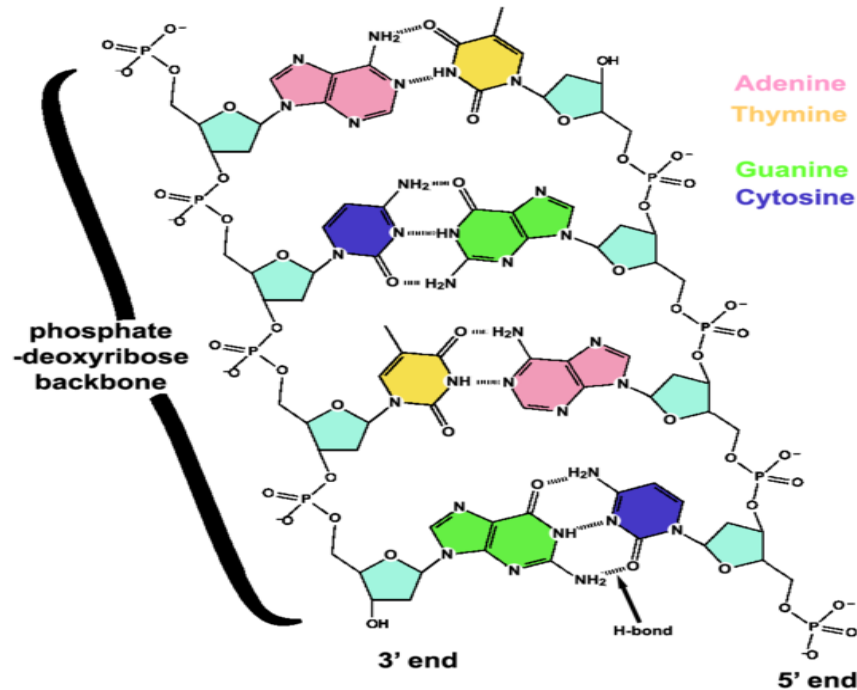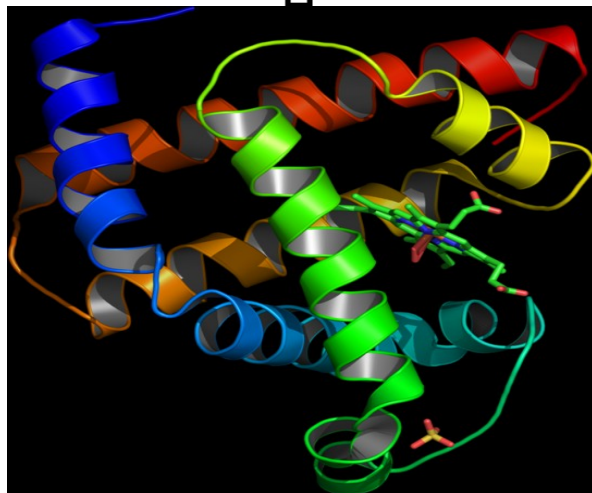
Protein strings

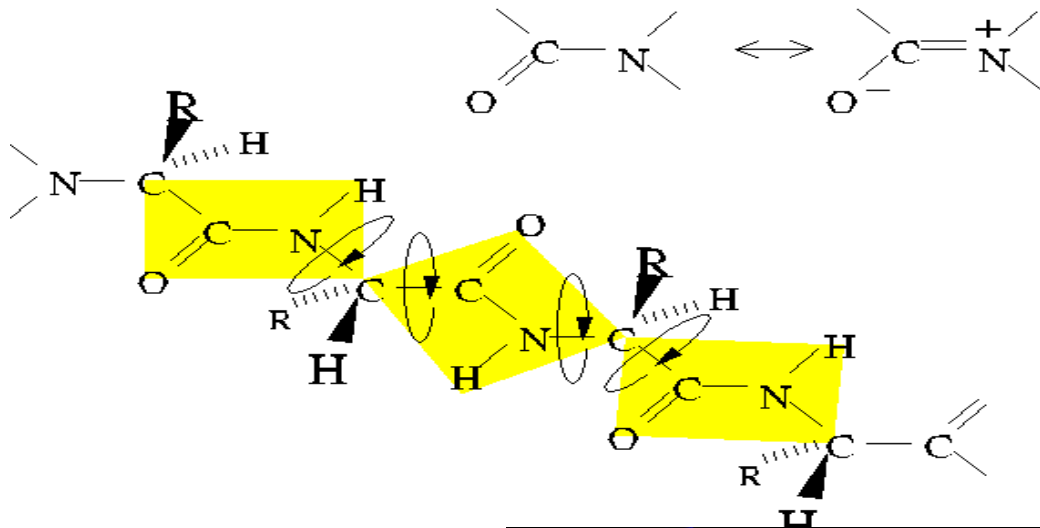Phage CRO Repressor on DNA. Andrew Coulson & Roger Sayle with RasMol, University of Edinburgh, 1993
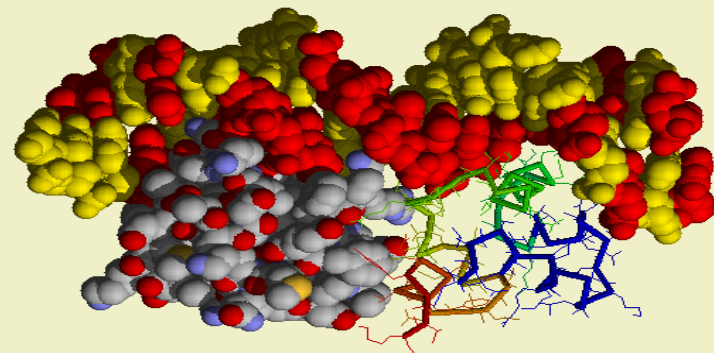
Genes, transcription, translation

- DNA – RNA - Thymine replaced by Uracil (T-U)
- The transcribed segments are called genes

  ACCGUACC**AUG**UUA...AUAGGC**UGA**GCA

- AUG – start codon (also amino-acid Methionine)
- UAA, UAG, UGA – stop codons
- Genes are read in sets of 3 nucleotides during translation – $4^3 = 64$ possible combinations
- Each combination codes for one of 20 amino-acids – the building blocks for proteins

Amino-acid translation table

Second letter

| First letter	U	C	A	G	Third letter
U	UUU ⎤ Phe UUC ⎦ UUA ⎤ Leu UUG ⎦	UCU ⎤ UCC ⎥ Ser UCA ⎥ UCG ⎦	UAU ⎤ Tyr UAC ⎦ UAA　Stop UAG　Stop	UGU ⎤ Cys UGC ⎦ UGA　Stop UGG　Trp	U C A G
C	CUU ⎤ CUC ⎥ Leu CUA ⎥ CUG ⎦	CCU ⎤ CCC ⎥ Pro CCA ⎥ CCG ⎦	CAU ⎤ His CAC ⎦ CAA ⎤ Gln CAG ⎦	CGU ⎤ CGC ⎥ Arg CGA ⎥ CGG ⎦	U C A G
A	AUU ⎤ AUC ⎥ Ile AUA ⎦ AUG　Met	ACU ⎤ ACC ⎥ Thr ACA ⎥ ACG ⎦	AAU ⎤ Asn AAC ⎦ AAA ⎤ Lys AAG ⎦	AGU ⎤ Ser AGC ⎦ AGA ⎤ Arg AGG ⎦	U C A G
G	GUU ⎤ GUC ⎥ Val GUA ⎥ GUG ⎦	GCU ⎤ GCC ⎥ Ala GCA ⎥ GCG ⎦	GAU ⎤ Asp GAC ⎦ GAA ⎤ Glu GAG ⎦	GGU ⎤ GGC ⎥ Gly GGA ⎥ GGG ⎦	U C A G

Alternative splicing

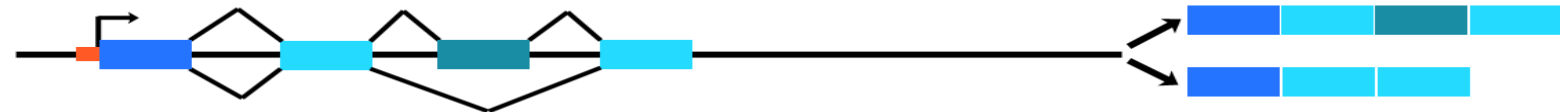(a) Alternative selection of promoters (e.g., *myosin* primary transcript)

P_1 P_2
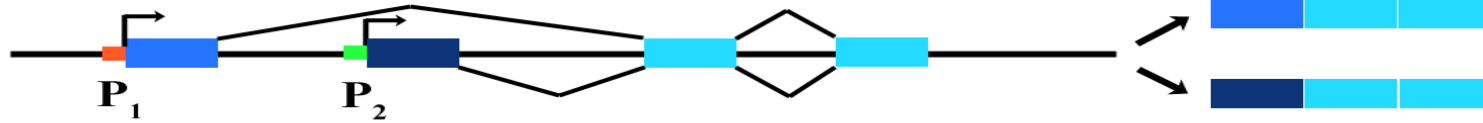
(b) Alternative selection of cleavage/polyadenylation sites (e.g., tropo*myosin* transcript)

Polyadenylation sites

(c) Intron retaining mode (e.g., *transposase* primary transcript)
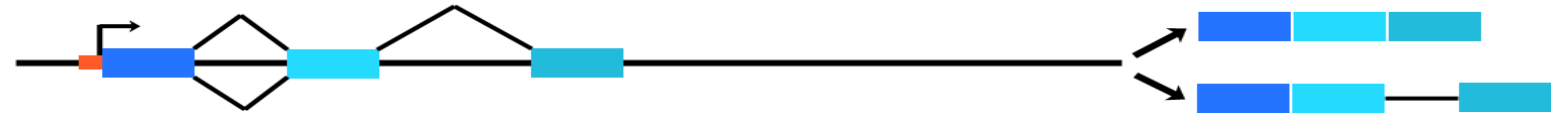
(d) Exon cassette mode (e.g., *troponin* primary transcript)

Stop and think...

- What is common to these two DNA sequences?

ATTACAGTCTGCCGCACTGC

GCAGTGCGGCAGACTGTAAT

Stop and think...

- What is common to these two DNA sequences?

5' – ATTACAGTCTGCCGCACTGC – 3'

5' – GCAGTGCGGCAGACTGTAAT – 3' (reverse complement)

5' – ATTACAGTCTGCCGCACTGC – 3'

3' – TAATGTCAGACGGCGTGACG – 5'

Stop and think

- Which of these mutations is more likely to be harmful to an organism?

  Original: GACTGATGTCCAGACTGCACT
  Mutant1: GACTGATG**A**CCAGACTGCACT
  Mutant2: GACTGATGTC**A**AGACTGCACT
  Mutant3: GACTGATGTCC_GACTGCACT (deletion)
  Mutant4: GACTGATGTCC**A**AGACTGCACT (insertion)

Stop and think

- Which of these mutations is more likely to be harmful to an organism?

Start codon

Original: GACTG**ATG** TCC AGA CTG CAC T
　　　　　　　　　　 S　　 R　　 L　　 H

Mutant1: GACTG**ATG** **A**CC AGA CTG CAC T
　　　　　　　　　　 T　　 R　　 L　　 H

Mutant2: GACTG**ATG** TC**A** AGA CTG CAC T
　　　　　　　　　　 S　　 R　　 L　　 H

Mutant3: GACTG**ATG** TCC_GAC TGC ACT (deletion)
　　　　　　　　　　 S　　 D　　 S　　 T

Mutant4: GACTG**ATG** TCC **A**AG ACT GCA CT (insertion)
　　　　　　　　　　 S　　 K　　 T　　 A

Representing sequences computationally

- FASTA

```
>seq1 some more information
ACCGGTAGCATAGA
CGGATAGACTTAGT
```

- FASTQ

```
@seqid other info
ACCAGTACGTCCGTG
+seqid other info (optional)
!+30qr-130!@+-@
```
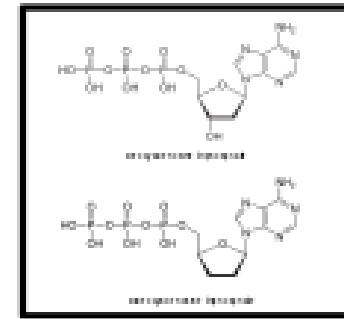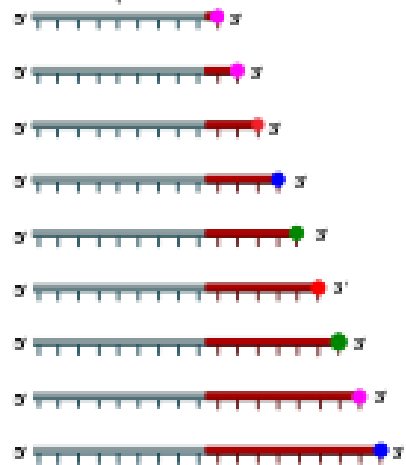
DNA sequencing

- Sanger (sorting by size)

- 454 (luminescence)

- Illumina (a different type of luminescence)

- Pacific Biosciences (trapped polymerase)

- Oxford Nanopore (nanopore)
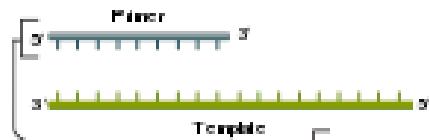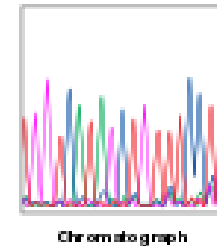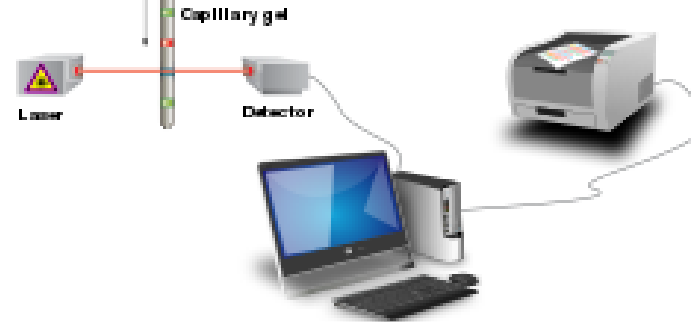
By Estevezj [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], from Wikimedia Commons

Next-generation DNA sequencing

1. Library preparation
2. Clonal amplification
3. Cyclic array sequencing

1. DNA fragmentation and in vitro adaptor ligation

emulsion PCR

bridge PCR

2.

3. Pyrosequencing — 454 sequencing

Sequencing-by-ligation — SOLiD platform

Sequencing-by-synthesis — Solexa technology
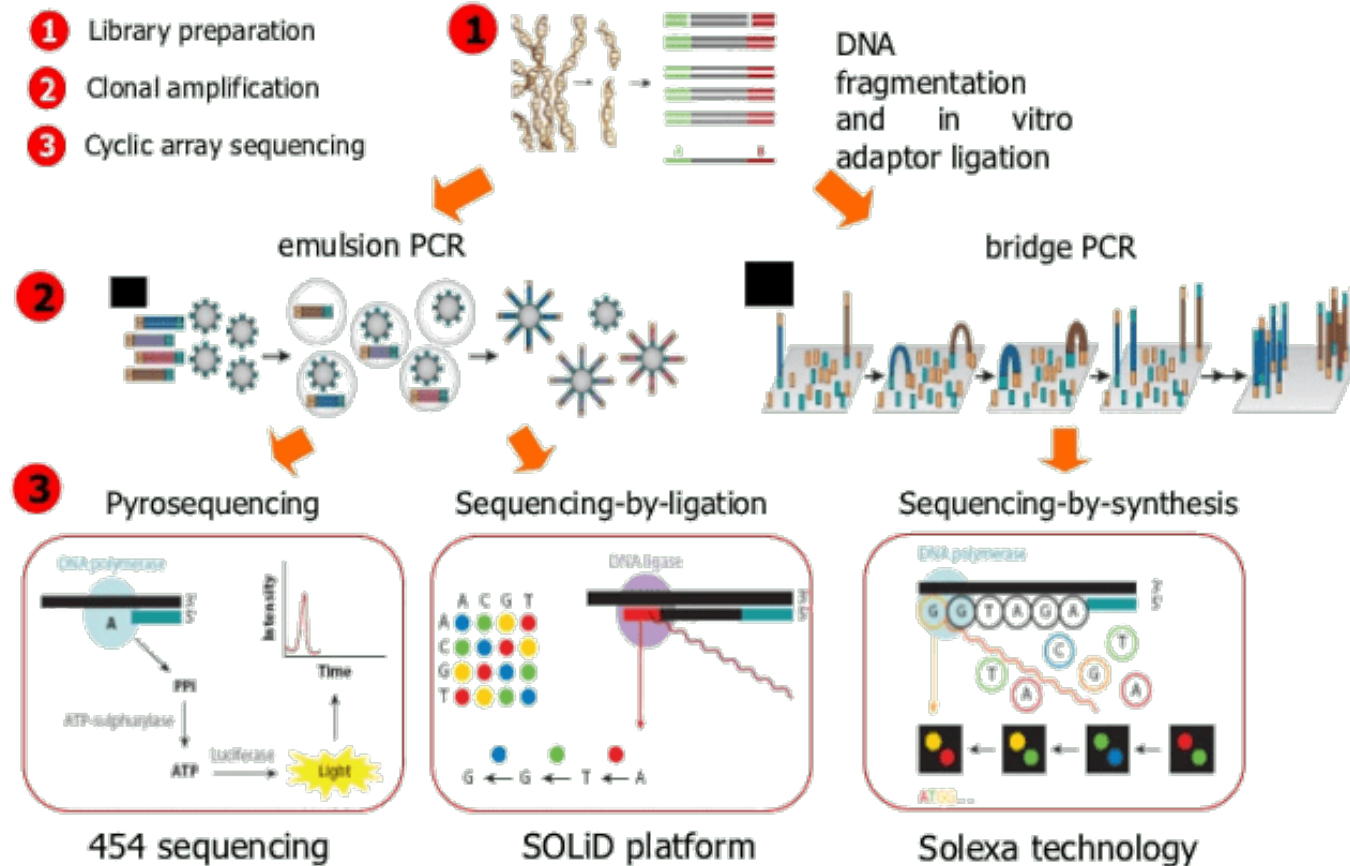
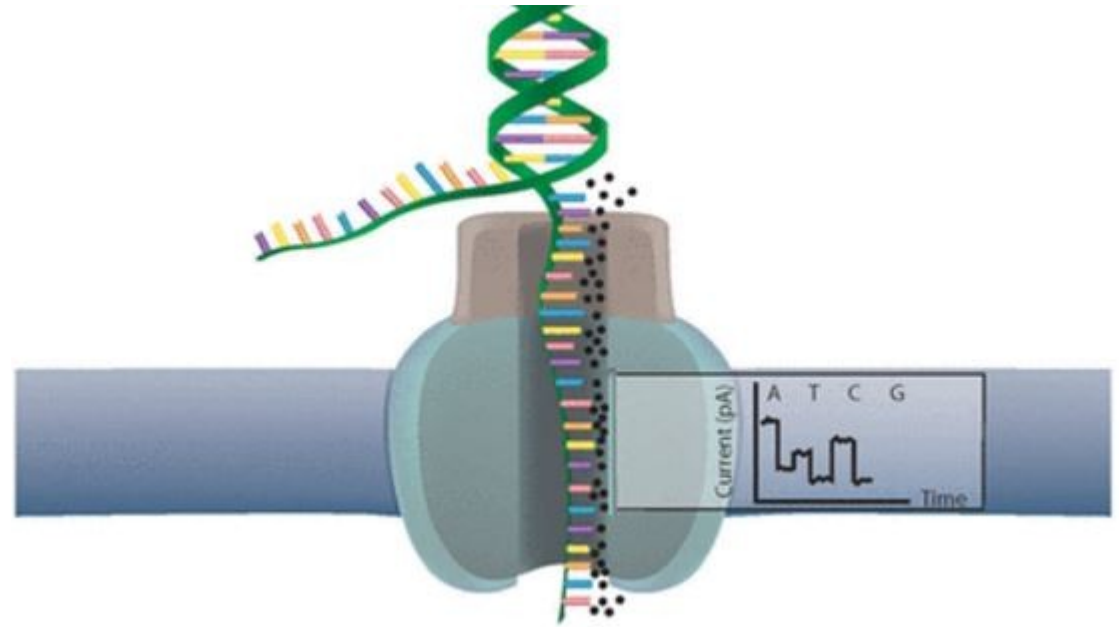# Pacific Biosciences
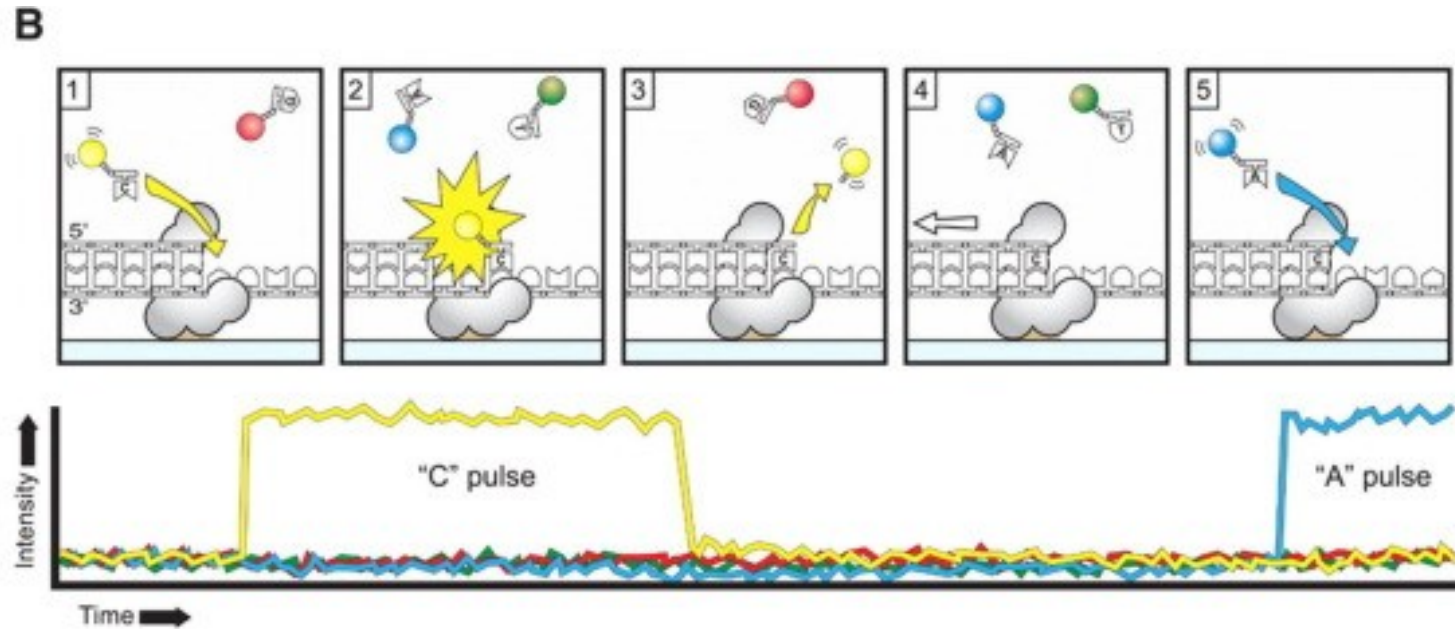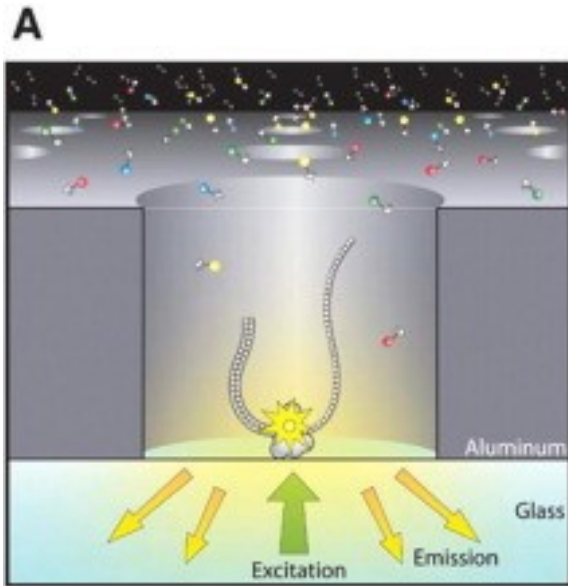
Oxford Nanopore

from labbiotech.eu