Update in Molecular Biology

Basis and molecular tools used in the study of DNA

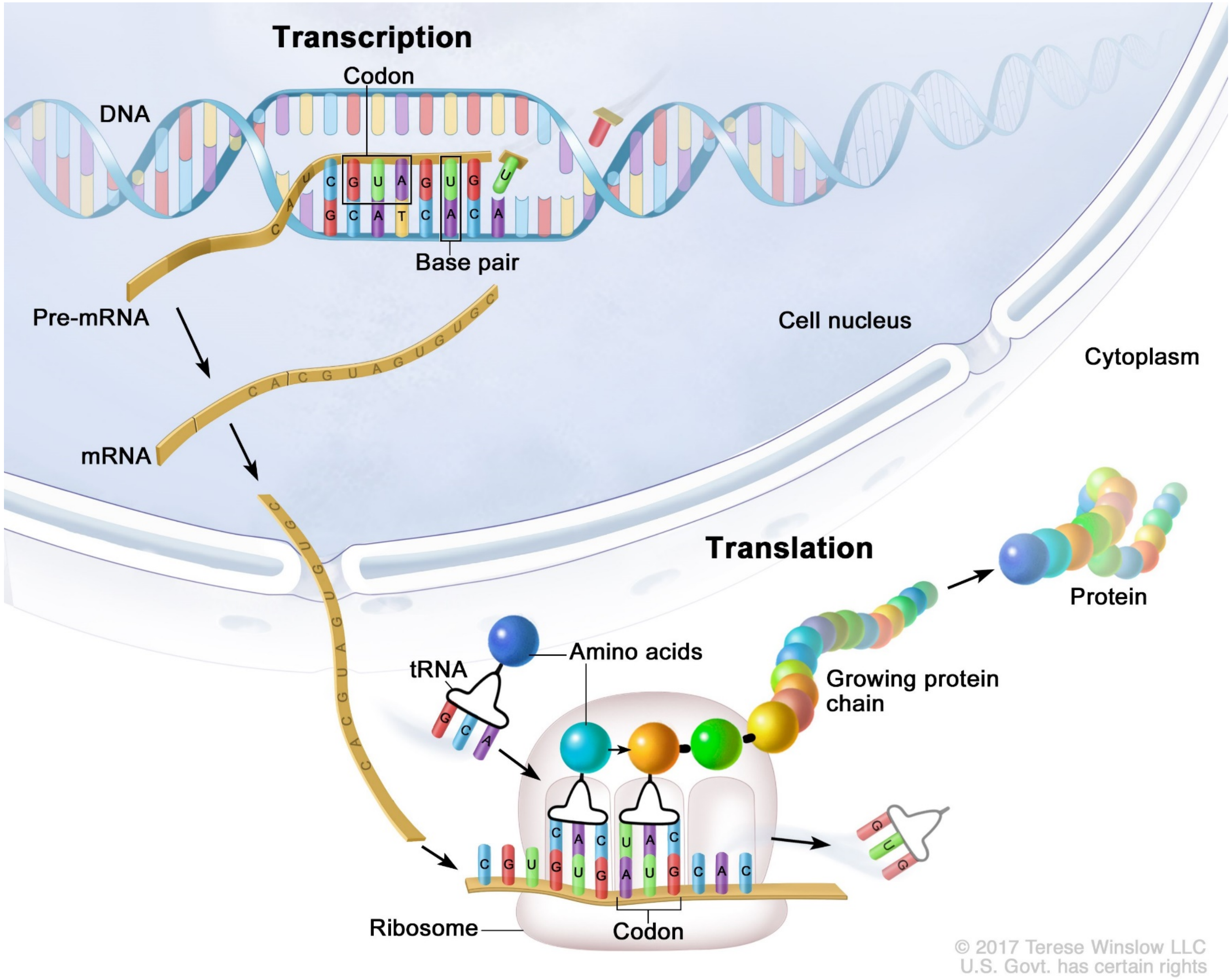
Guillaume Beauclair

guillaume.beauclair@universite-paris-saclay.fr

https://app.wooclap.com/2024D2HPBM

M1 D²HP Development of Drugs and Health Products - September 27th 2024

Plan



I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

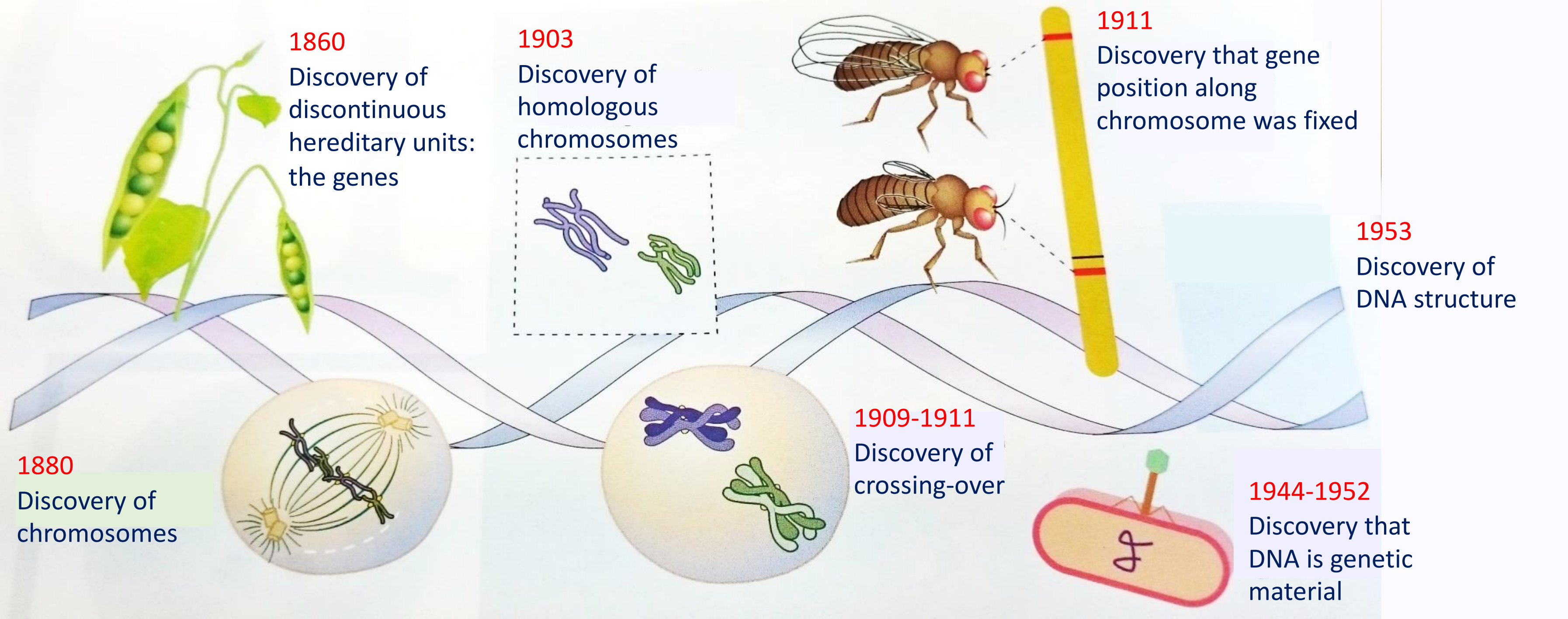
- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

Historical background : *Important findings on the nature of the DNA*



1860
Discovery of discontinuous hereditary units: the genes

1903
Discovery of homologous chromosomes

1911
Discovery that gene position along chromosome was fixed

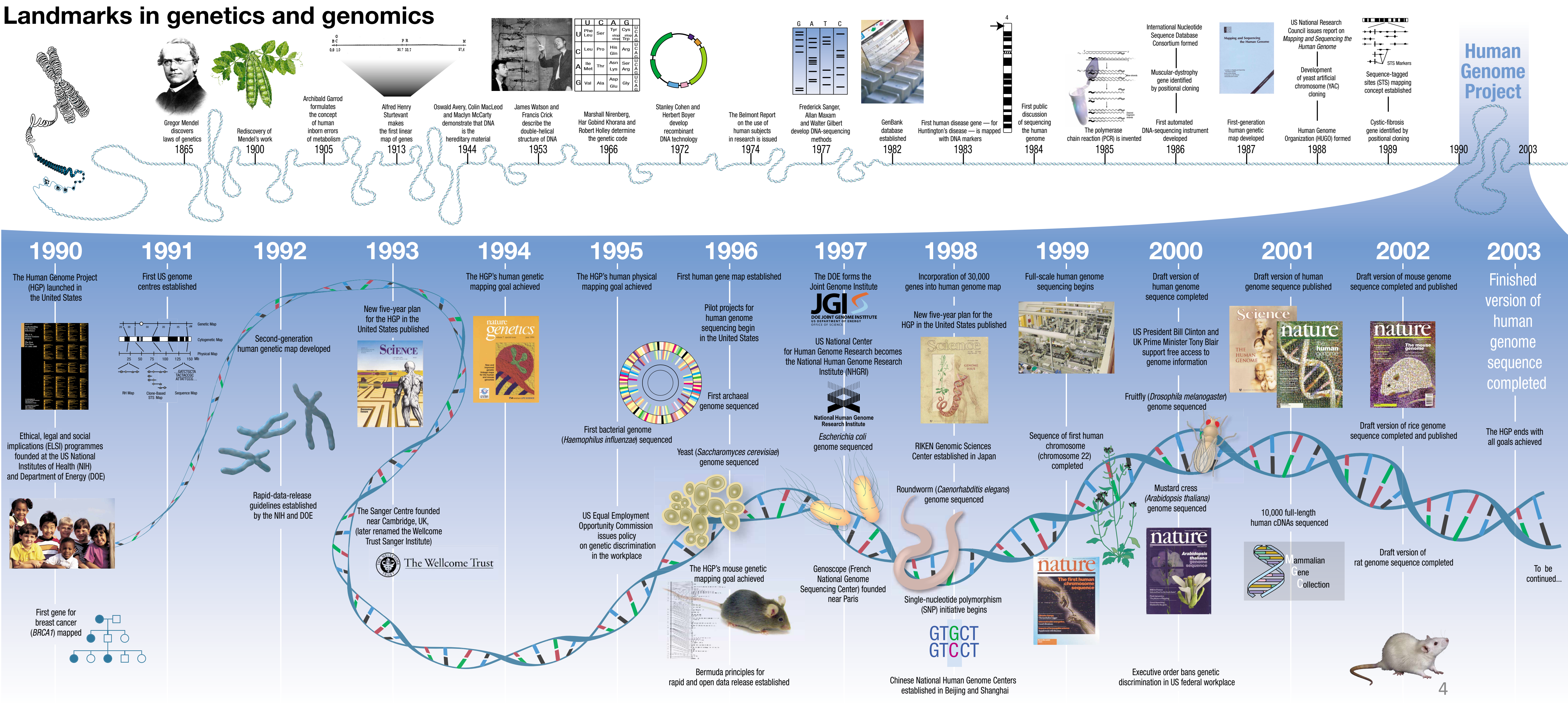
1953
Discovery of DNA structure

1880
Discovery of chromosomes

1909-1911
Discovery of crossing-over

1944-1952
Discovery that DNA is genetic material

Historical background : *Important findings on the nature of the DNA*



Landmarks in genetics and genomics

1865
Gregor Mendel discovers laws of genetics

1900
Rediscovery of Mendel's work

1905
Archibald Garrod formulates the concept of human inborn errors of metabolism

1913
Alfred Henry Sturtevant makes the first linear map of genes

1944
Oswald Avery, Colin MacLeod and Maclyn McCarty demonstrate that DNA is the hereditary material

1953
James Watson and Francis Crick describe the double-helical structure of DNA

1966
Marshall Nirenberg, Har Gobind Khorana and Robert Holley determine the genetic code

1972
Stanley Cohen and Herbert Boyer develop recombinant DNA technology

1974
The Belmont Report on the use of human subjects in research is issued

1977
Frederick Sanger, Allan Maxam and Walter Gilbert develop DNA-sequencing methods

1982
GenBank database established

1983
First human disease gene — for Huntington's disease — is mapped with DNA markers

1984
First public discussion of sequencing the human genome

1985
The polymerase chain reaction (PCR) is invented

1986
International Nucleotide Sequence Database Consortium formed
Muscular-dystrophy gene identified by positional cloning
First automated DNA-sequencing instrument developed

1987
First-generation human genetic map developed

1988
US National Research Council issues report on *Mapping and Sequencing the Human Genome*
Development of yeast artificial chromosome (YAC) cloning
Human Genome Organization (HUGO) formed

1989
Sequence-tagged sites (STS) mapping concept established
Cystic-fibrosis gene identified by positional cloning

1990
2003

Human Genome Project

1990
The Human Genome Project (HGP) launched in the United States

Ethical, legal and social implications (ELSI) programmes founded at the US National Institutes of Health (NIH) and Department of Energy (DOE)

First gene for breast cancer (*BRCA1*) mapped

1991
First US genome centres established

1992
Second-generation human genetic map developed

Rapid-data-release guidelines established by the NIH and DOE

1993
New five-year plan for the HGP in the United States published

The Sanger Centre founded near Cambridge, UK, (later renamed the Wellcome Trust Sanger Institute)

The Wellcome Trust

1994
The HGP's human genetic mapping goal achieved

1995
The HGP's human physical mapping goal achieved

First bacterial genome (*Haemophilus influenzae*) sequenced

US Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace

1996
First human gene map established

Pilot projects for human genome sequencing begin in the United States

First archaeal genome sequenced

Yeast (*Saccharomyces cerevisiae*) genome sequenced

The HGP's mouse genetic mapping goal achieved

Bermuda principles for rapid and open data release established

1997
The DOE forms the Joint Genome Institute

JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

US National Center for Human Genome Research becomes the National Human Genome Research Institute (NHGRI)

National Human Genome Research Institute

Escherichia coli genome sequenced

Genoscope (French National Genome Sequencing Center) founded near Paris

1998
Incorporation of 30,000 genes into human genome map

New five-year plan for the HGP in the United States published

RIKEN Genomic Sciences Center established in Japan

Roundworm (*Caenorhabditis elegans*) genome sequenced

Single-nucleotide polymorphism (SNP) initiative begins

GTGCT
GTCCT

Chinese National Human Genome Centers established in Beijing and Shanghai

1999
Full-scale human genome sequencing begins

Sequence of first human chromosome (chromosome 22) completed

2000
Draft version of human genome sequence completed

US President Bill Clinton and UK Prime Minister Tony Blair support free access to genome information

Fruitfly (*Drosophila melanogaster*) genome sequenced

Mustard cress (*Arabidopsis thaliana*) genome sequenced

Executive order bans genetic discrimination in US federal workplace

2001
Draft version of human genome sequence published

10,000 full-length human cDNAs sequenced

Mammalian
Gene
Collection

2002
Draft version of mouse genome sequence completed and published

Draft version of rice genome sequence completed and published

Draft version of rat genome sequence completed

2003
Finished version of human genome sequence completed

The HGP ends with all goals achieved

To be continued...

PEAS COURTESY J. BLAMIRE, CITY UNIV. NEW YORK; WATSON & CRICK COURTESY A. BARRINGTON BROWN/SPL; *SCIENCE* COVERS COURTESY AAAS

Historical background : *Important findings on the nature of the DNA*

2003

The Human Genome Project is completed, confirming humans have approx. 20,000 to 25,000 genes. It read and recorded more than 92% of the genome – as much as was technologically possible at the time.

2005

The first report from HapMap (Map of Human Genetic Variation) is published, which aimed to produce a 'catalogue' of common human genetic variations and where they are found in the genome.

The first video is uploaded to YouTube.

2007

A new DNA sequencing technology (microarray hybridisation) is introduced, increasing the output of DNA sequencing by 70-fold.

Apple introduces the iPhone.

2012

The ENCODE study confirms that the human genome contains more than 20,600 protein-coding genes.

The Olympic Games are held in London, UK.

2013

The gene editing tool CRISPR-Cas9 is discovered. Later, Jennifer Doudna and Emmanuelle Charpentier win the Nobel prize for its discovery – the first time in history the prize is awarded to two women.

The US Supreme Court rules that naturally occurring DNA cannot be patented.

2017

The first gene editing of human embryos takes place, using CRISPR-Cas9, to remove DNA responsible for a hereditary heart condition.

SpaceX builds and flies the first reusable rocket.

2018

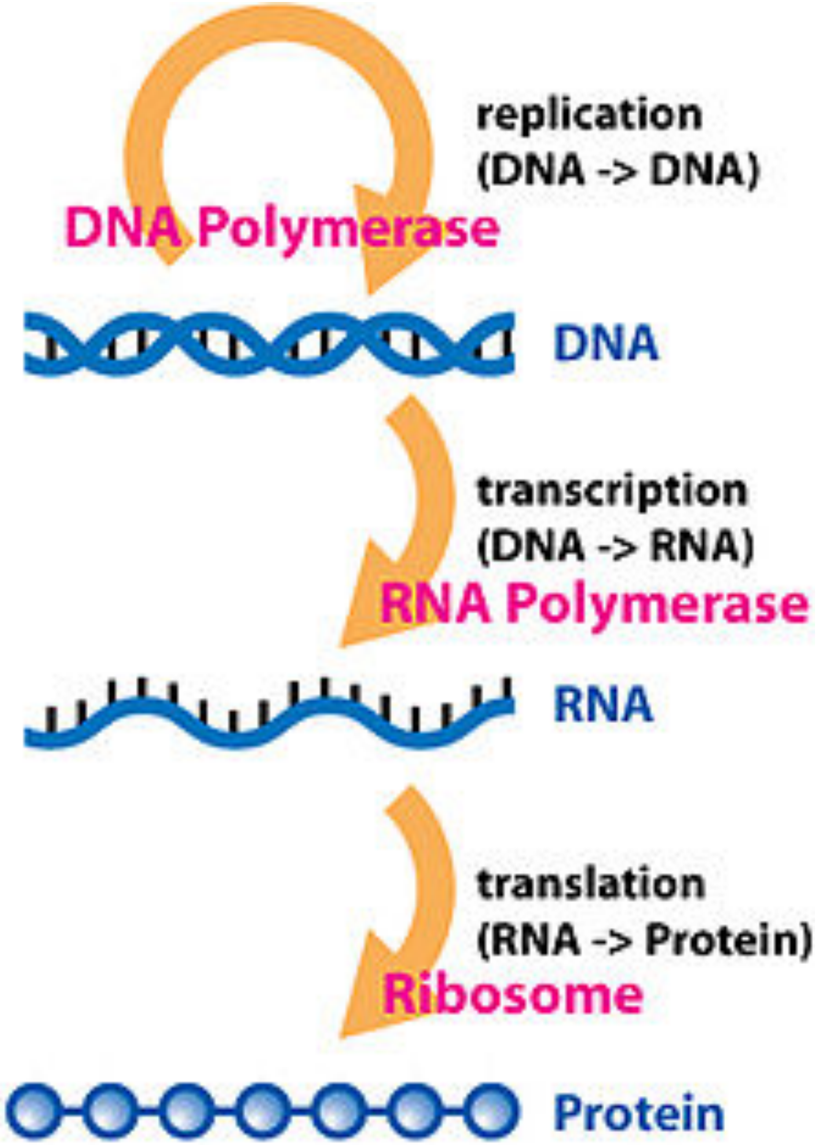
The 100K Genomes project is completed, sequencing 100,000 genomes from people affected by a rare genetic disease or cancer.

NASA launches its first mission to the sun.

2022

The entirety of the human genome sequence is completed. Various project have slowly filled the remaining gaps since the first 92% was published in 2003.

From DNA to proteins in Eukaryotes

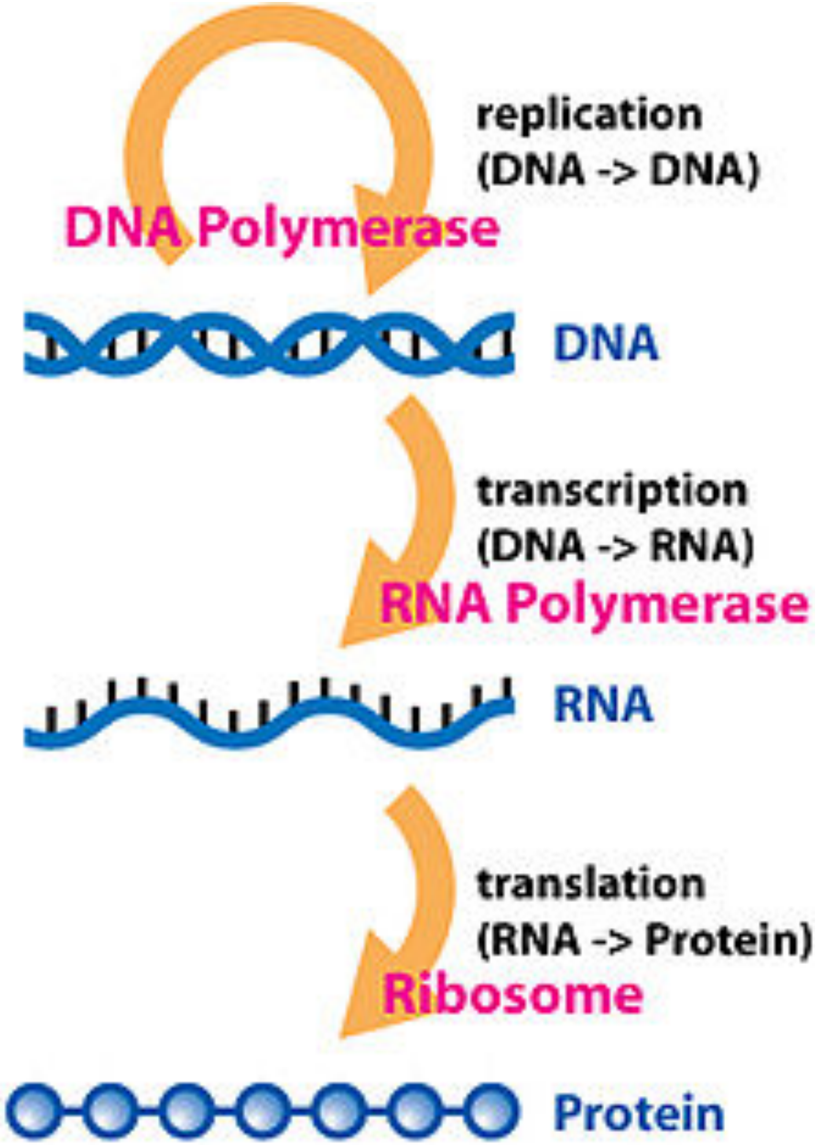


Replication and heredity

Storage of genetic information

Expression of the genetic message

From DNA to proteins in Eukaryotes



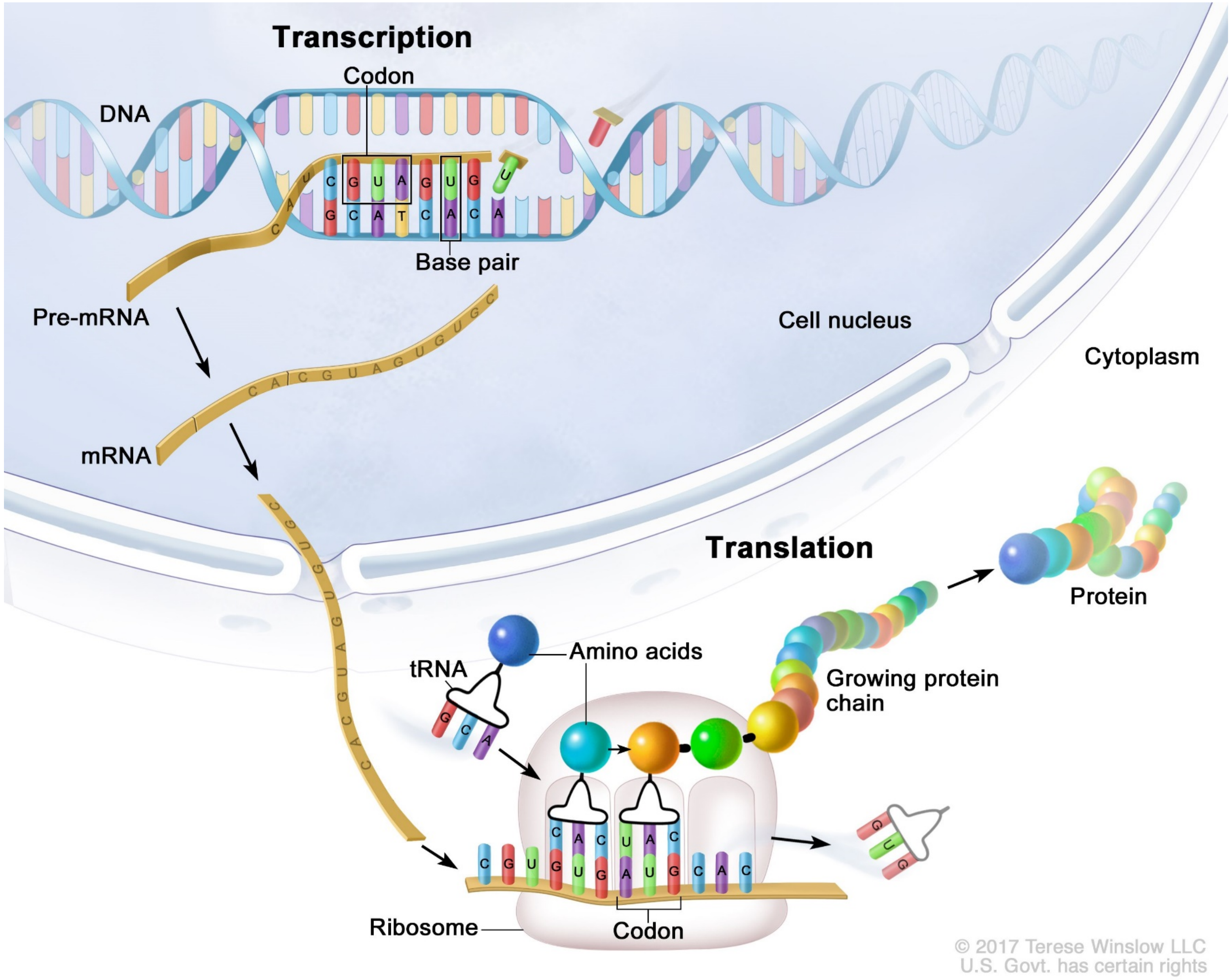
Replication and heredity

Storage of genetic information

also non-coding RNA, …

Expression of the genetic message

Plan



© 2017 Terese Winslow LLC
U.S. Govt. has certain rights

I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

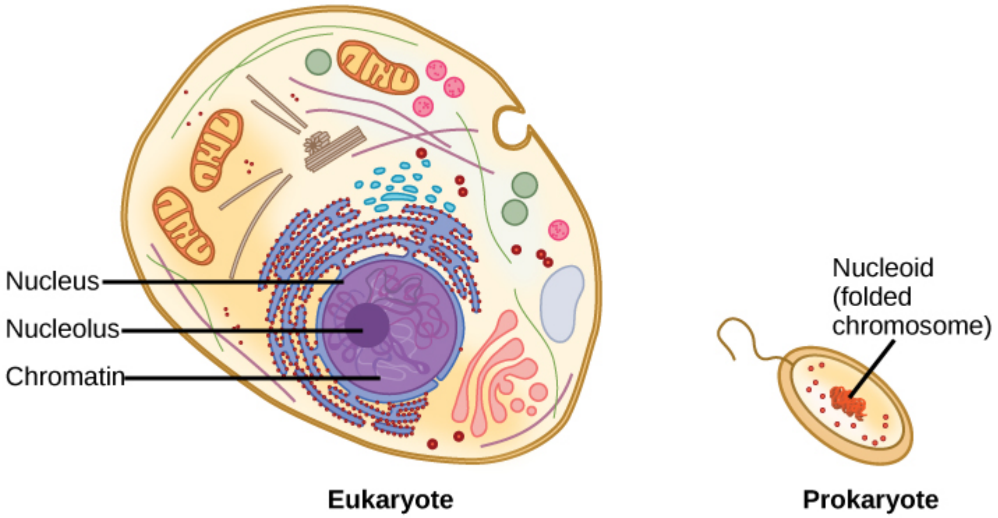
IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

Two types of nucleic acids

DNA = DeoxyriboNucleic Acid

- * Double-stranded structure
- * Location
  - in the nucleus of Eukaryotes
  - in the cytoplasm in Procaryotes
- * It carries the genetic information needed for protein synthesis



A eukaryote contains a well-defined nucleus, whereas in prokaryotes, the chromosome lies in the cytoplasm in an area called the nucleoid

RNA = RiboNucleic Acid

- * Single strand structure
- * Several types of RNA
- * They allow the execution of protein synthesis
- * Some regulate gene expression

Nucleotides

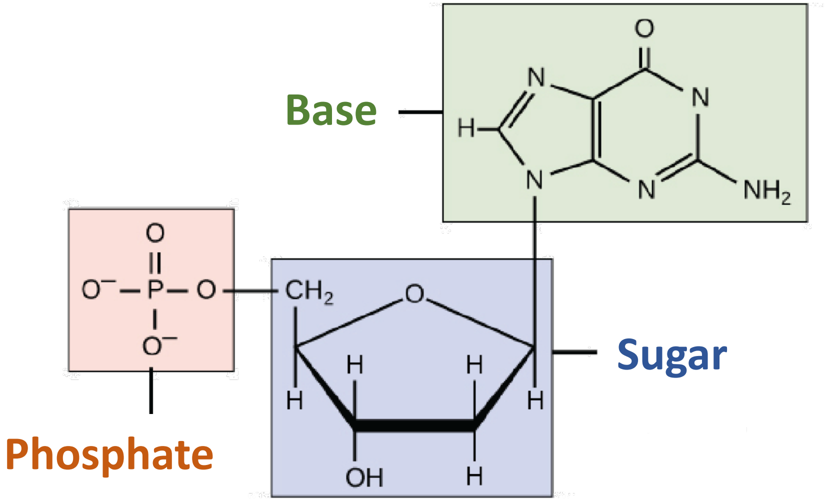
Nucleotide = elemental component of nucleic acids.
Nucleic acids are polynucleotides

Composed of :

Nucleotides

Nucleotide = elemental component of nucleic acids.
Nucleic acids are polynucleotides

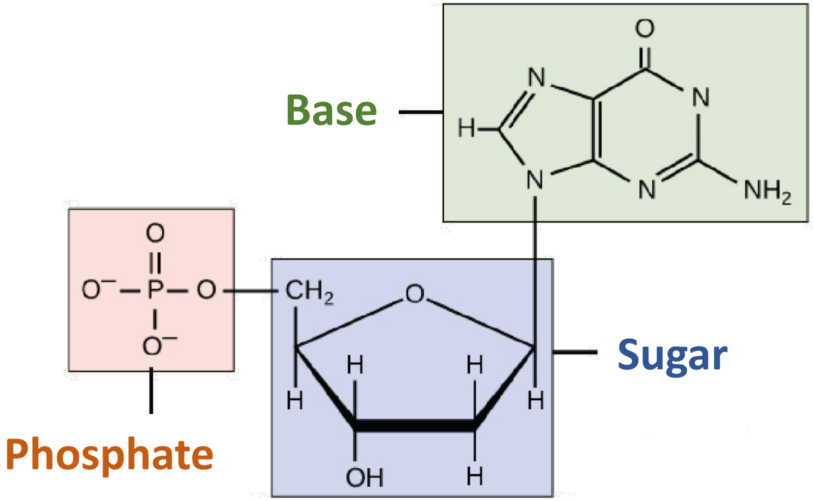
Composed of :



Two types of base

Nucleotide = elemental component of nucleic acids.
Nucleic acids are polynucleotides

Composed of :



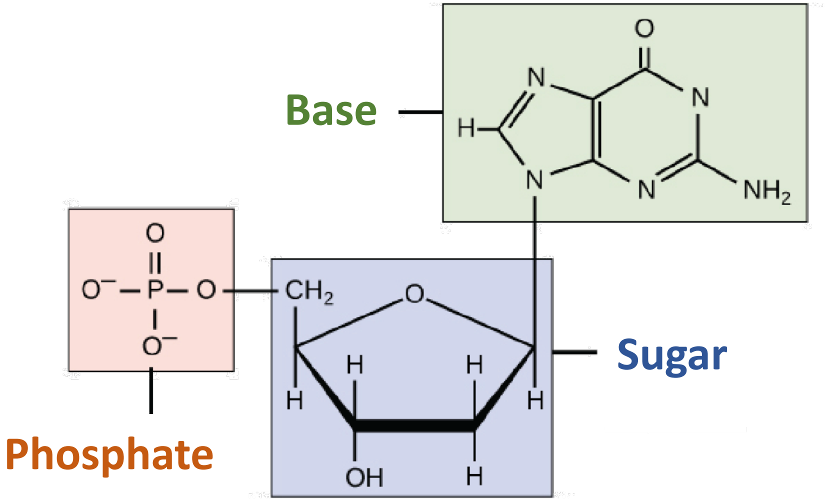
Pyrimidines

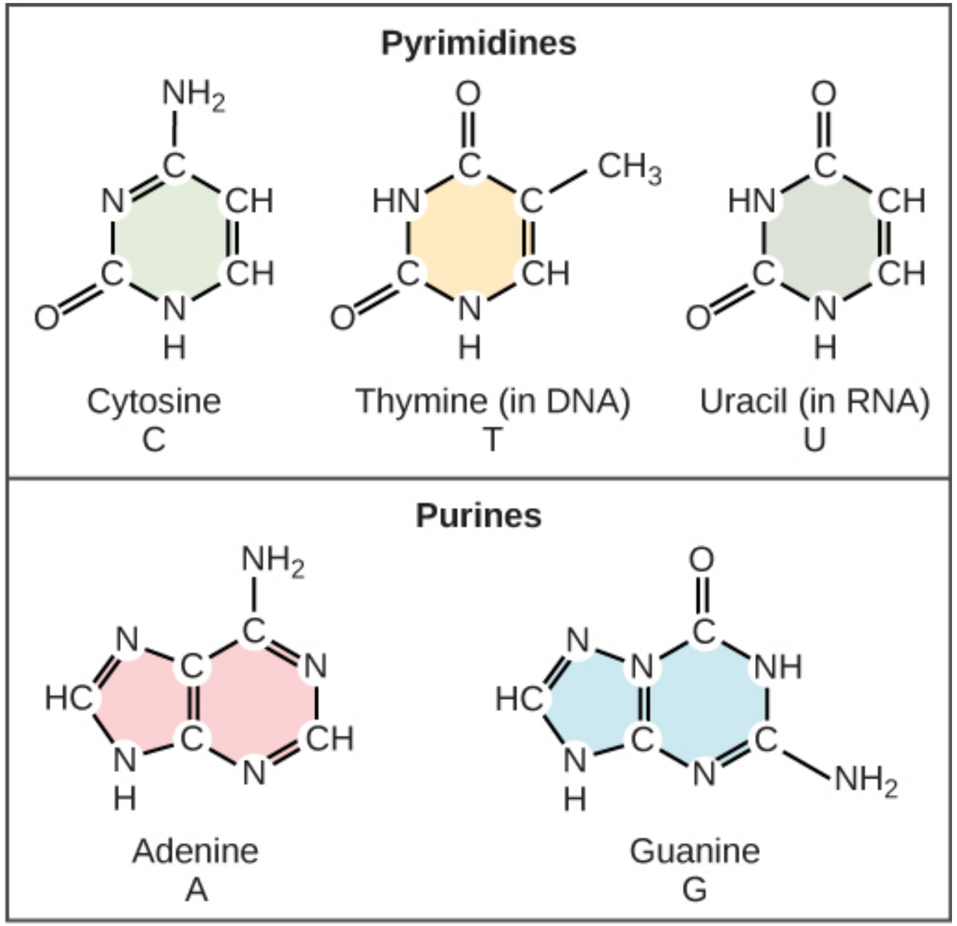
Purines

Two types of base : Pyrimidine and Purine

Nucleotide = elemental component of nucleic acids.
Nucleic acids are polynucleotides

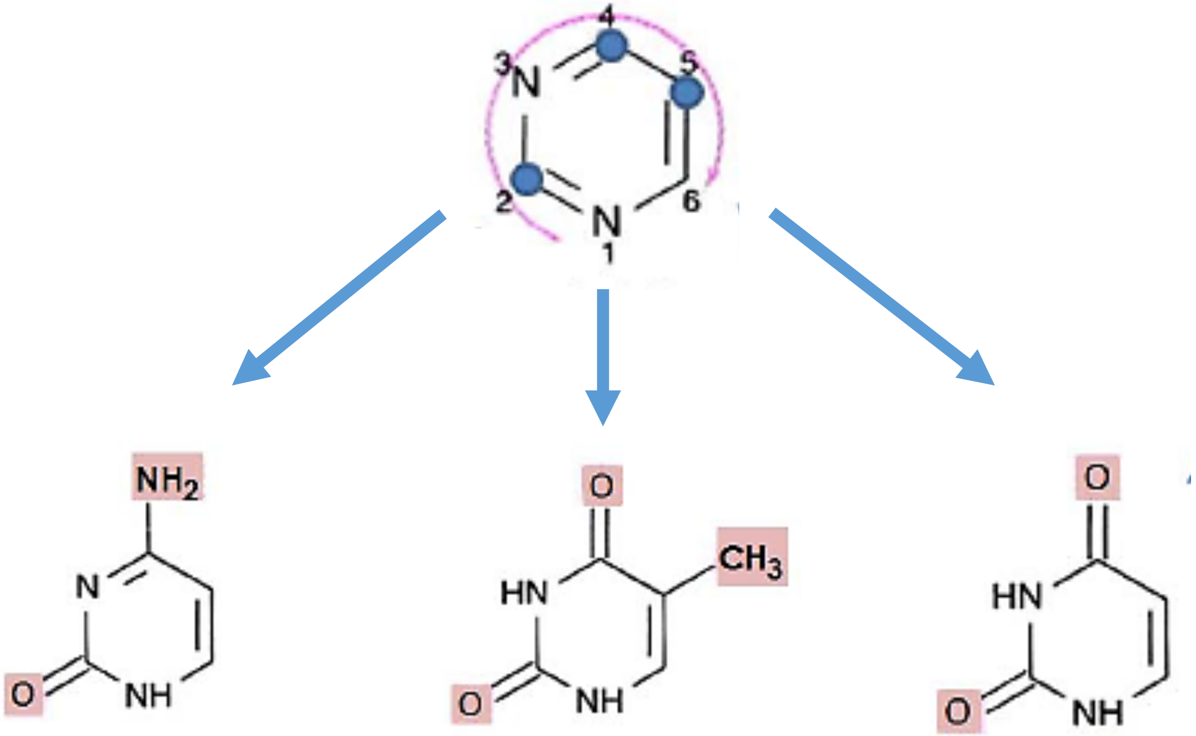
Composed of :





The pyrimidine bases

Pyrimidine ring (x6)



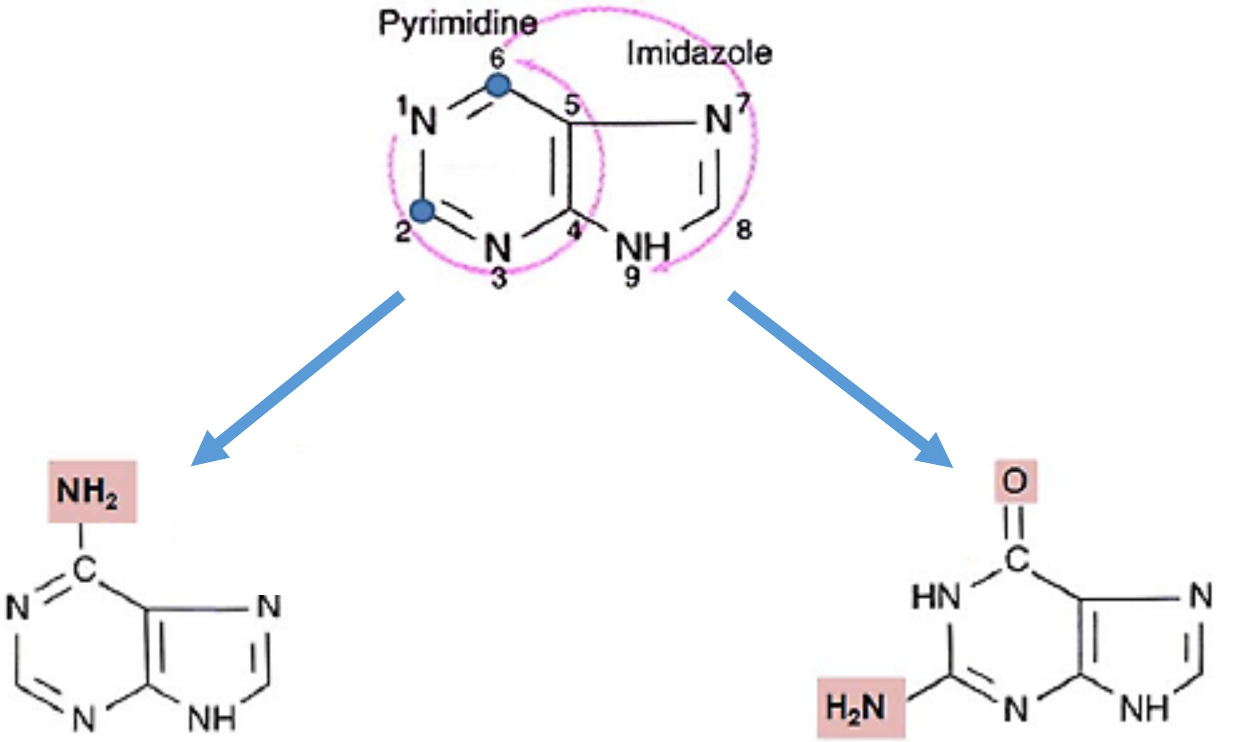
Cytosine
C

Thymine
T

Uracil
U

The purine bases

Pyrimidine ring (x6) + Imidazole ring (x5)

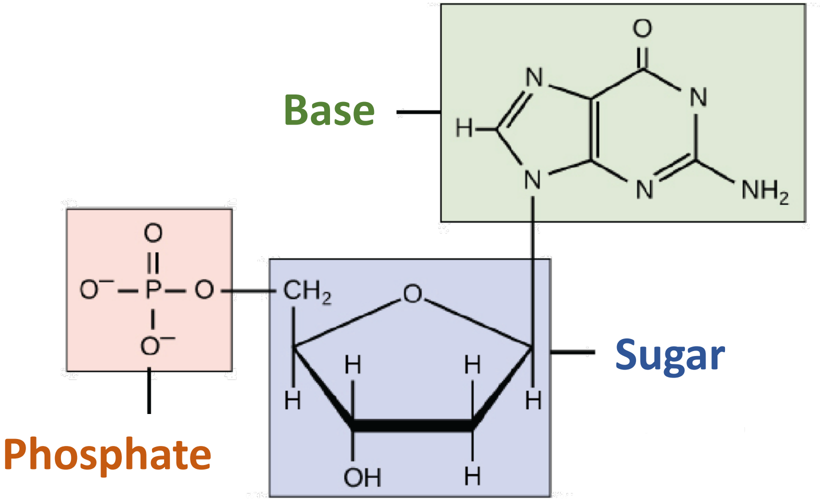


Adenine
A

Guanine
G

Nucleotides

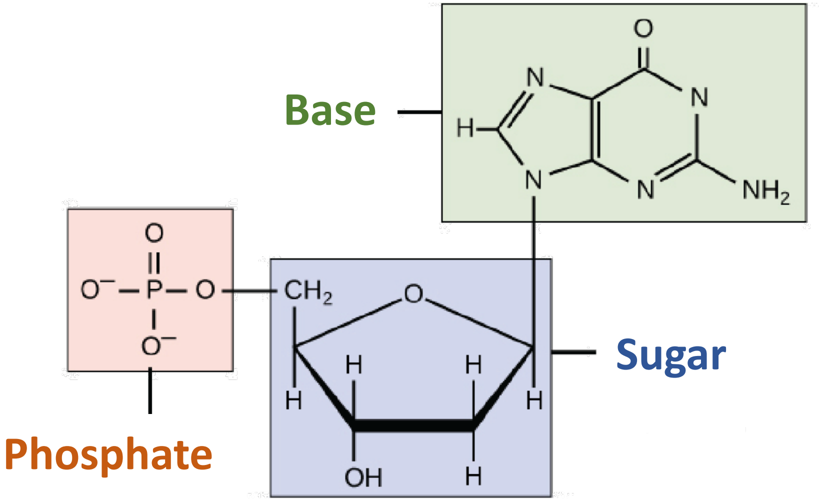
Nucleotide = elemental component of nucleic acids.
Nucleic acids are polynucleotides

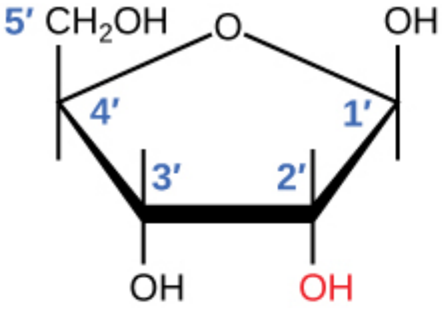


Nucleotides

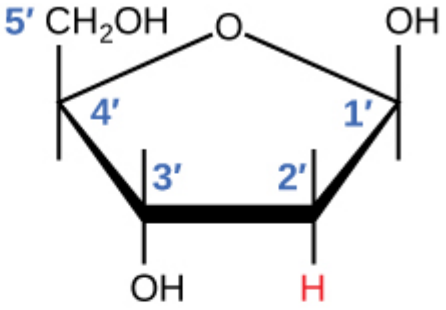
Nucleotide = elemental component of nucleic acids.
Nucleic acids are polynucleotides

Adenine + Sugar = Adenosine
Guanine + Sugar = Guanosine
Thymine + Sugar = Thymidine
Cytosine + Sugar = Cytidine
Uracyl + Sugar = Uridine



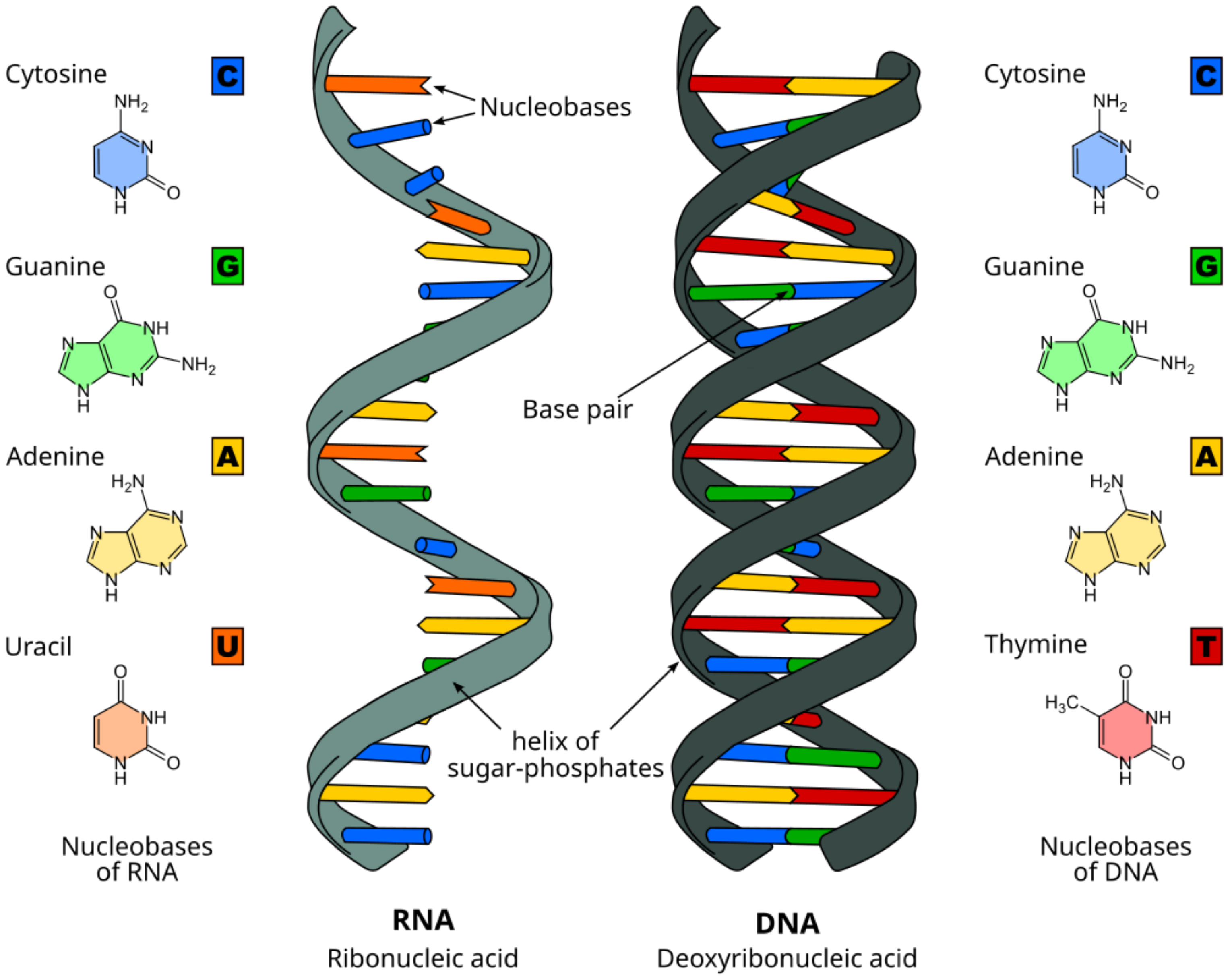


Ribose



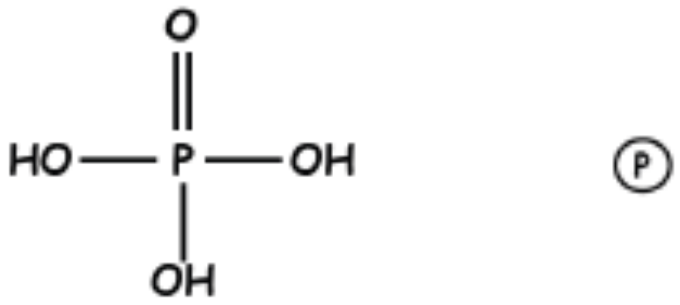
Deoxyribose

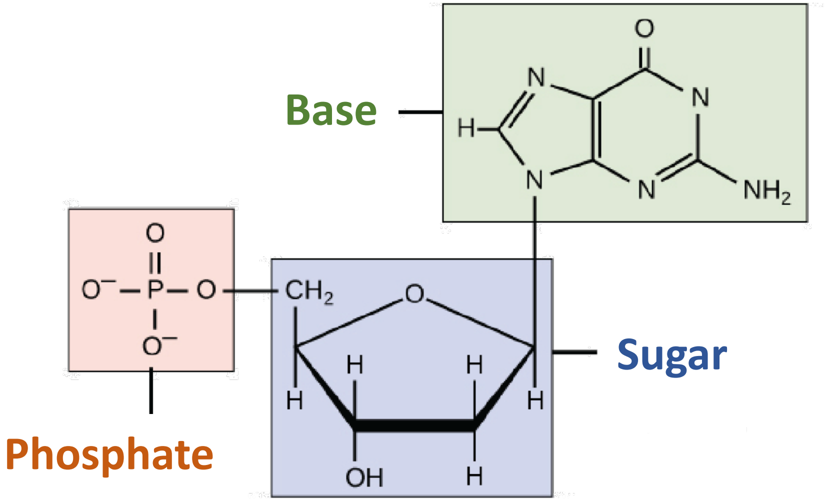
RNA vs DNA


Cytosine
C
NH_2
N
N
H
O
Guanine
G
O
NH
N
NH_2
N
N
H
Adenine
A
H_2N
N
N
N
N
H
Uracil
U
O
NH
N
H
O
Nucleobases of RNA
Nucleobases
Base pair
helix of sugar-phosphates
RNA
Ribonucleic acid
DNA
Deoxyribonucleic acid
Cytosine
C
Guanine
G
Adenine
A
Thymine
T
H_3C
Nucleobases of DNA

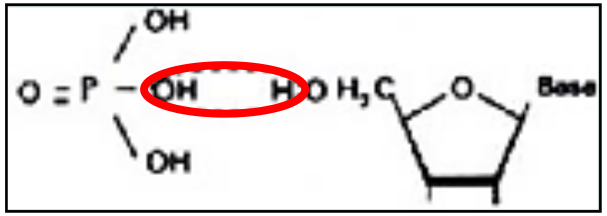
Phosphate group

Nucleotide = elemental component of nucleic acids.
Nucleic acids are polynucleotides





- Acid
- Negatively charged
- bound to sugar by removing one H2O molecule

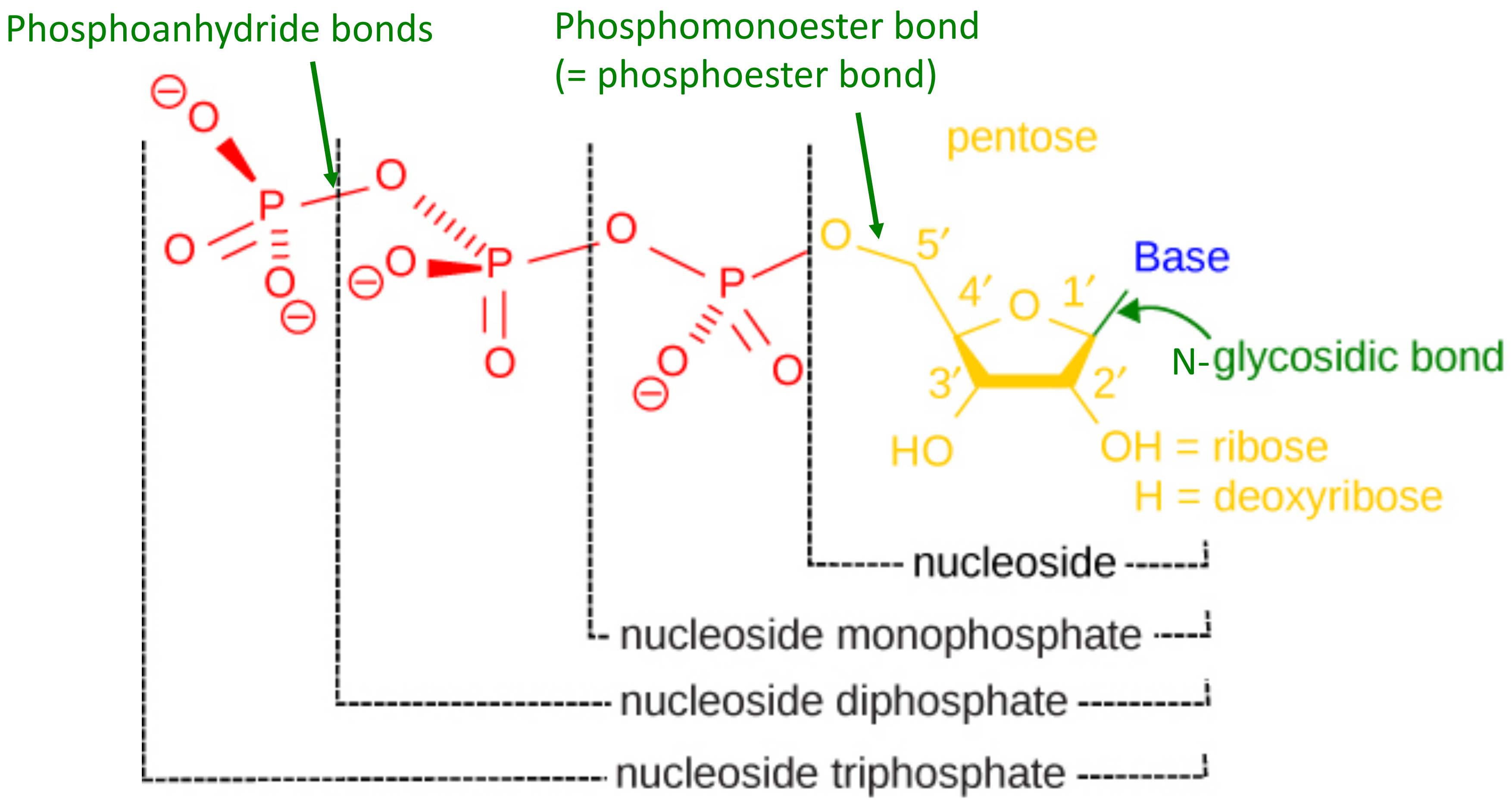


Nucleotide vs Nucleoside vs Nitrogenous bases

Nucleoside =
Nucleotide =

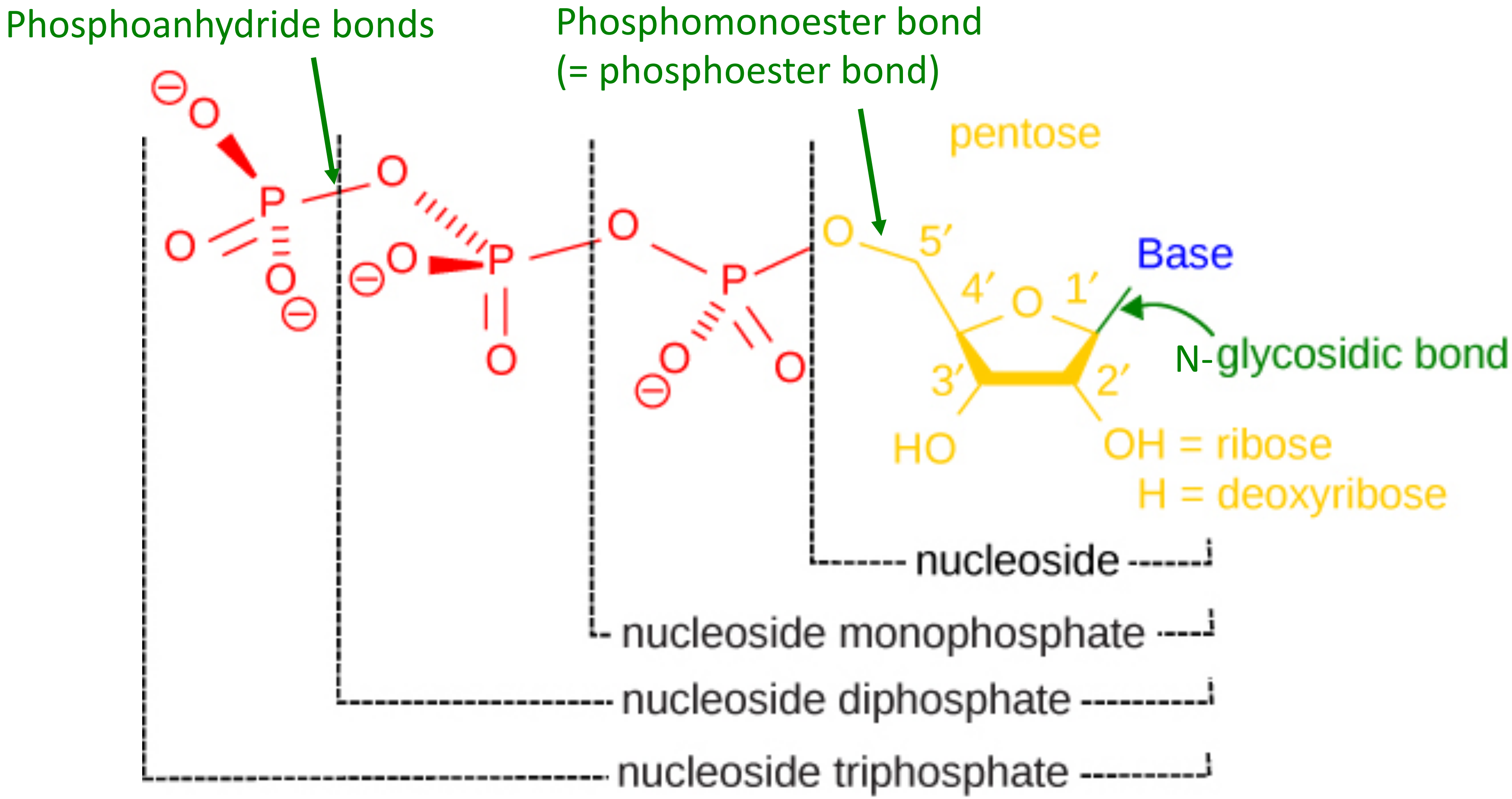
Nucleotide vs Nucleoside vs Nitrogenous bases

Nucleoside = Base + Sugar
Nucleotide = Base + Sugar + phosphate(s)



Nucleotide vs Nucleoside vs Nitrogenous bases

Nucleoside = Base + Sugar
Nucleotide = Base + Sugar + phosphate(s)



Phosphoanhydride bonds

Phosphomonoester bond
(= phosphoester bond)

pentose

Base

N-glycosidic bond

OH = ribose
H = deoxyribose

nucleoside

nucleoside monophosphate

nucleoside diphosphate

nucleoside triphosphate

Adenine

Adenosine

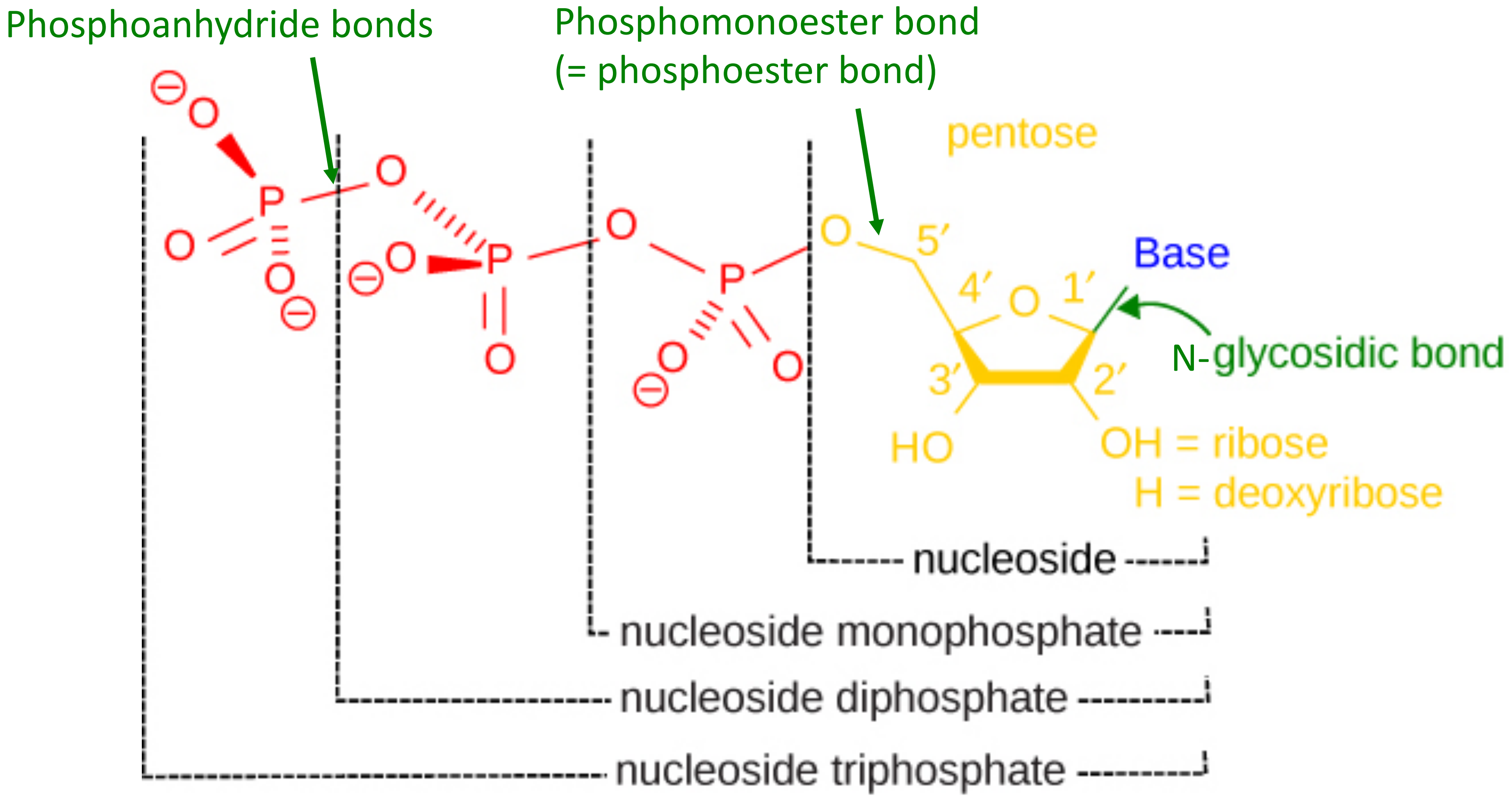
AMP

ADP

ATP

Nucleotide vs Nucleoside vs Nitrogenous bases

Nucleoside = Base + Sugar
Nucleotide = Base + Sugar + phosphate(s)



Adenine

d-Adenosine

dAMP

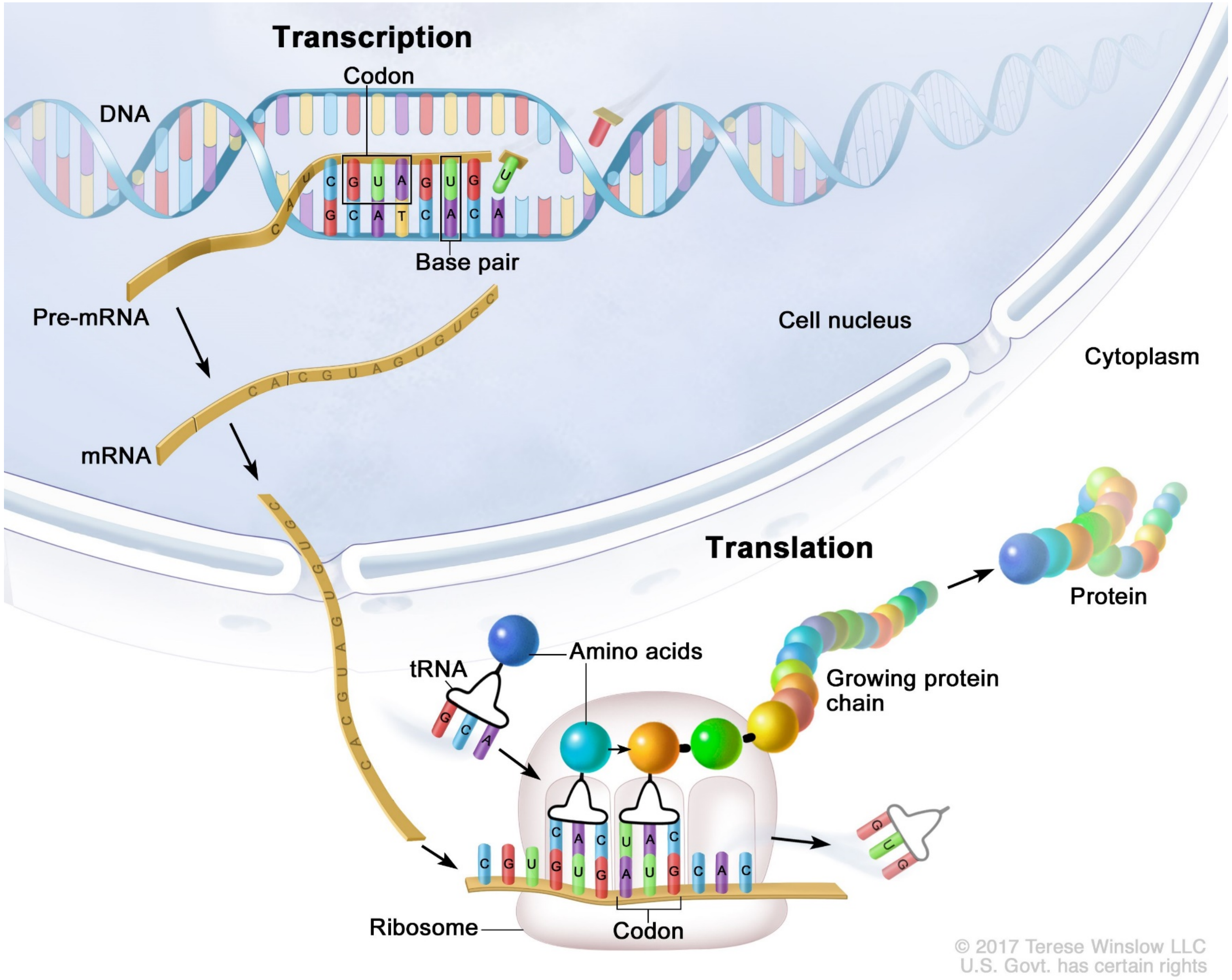
dADP

dATP

Summary on bases, nucleosides and nucleotides

		Ribose		Desoxyribose	
Bases		**Nucleosides**	**Nucleotides**	**Nucleosides**	**Nucleotides**
Purines	Adenine	Aden**osine**	AMP	d-Aden**osine**	dAMP
	Guanine	Guan**osine**	GMP	d-Guan**osine**	dGMP
Pyrimidines	Uracil	Ur**idine**	**UMP**	-	-
	Thymine	-	-	d-Thym**idine**	**dTMP**
	Cytosine	Cyt**idine**	CMP	d-Cyt**idine**	dCMP

Plan



I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

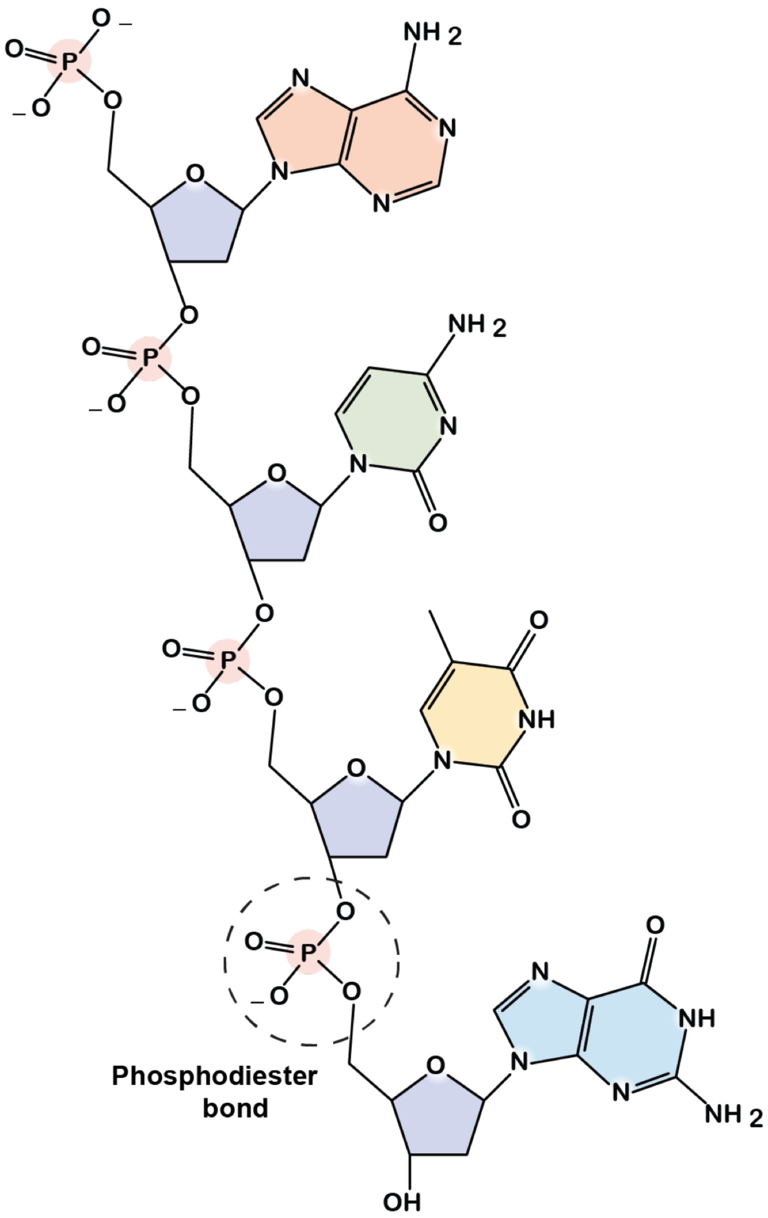
- Basic mechanism
- Maturation of mRNA

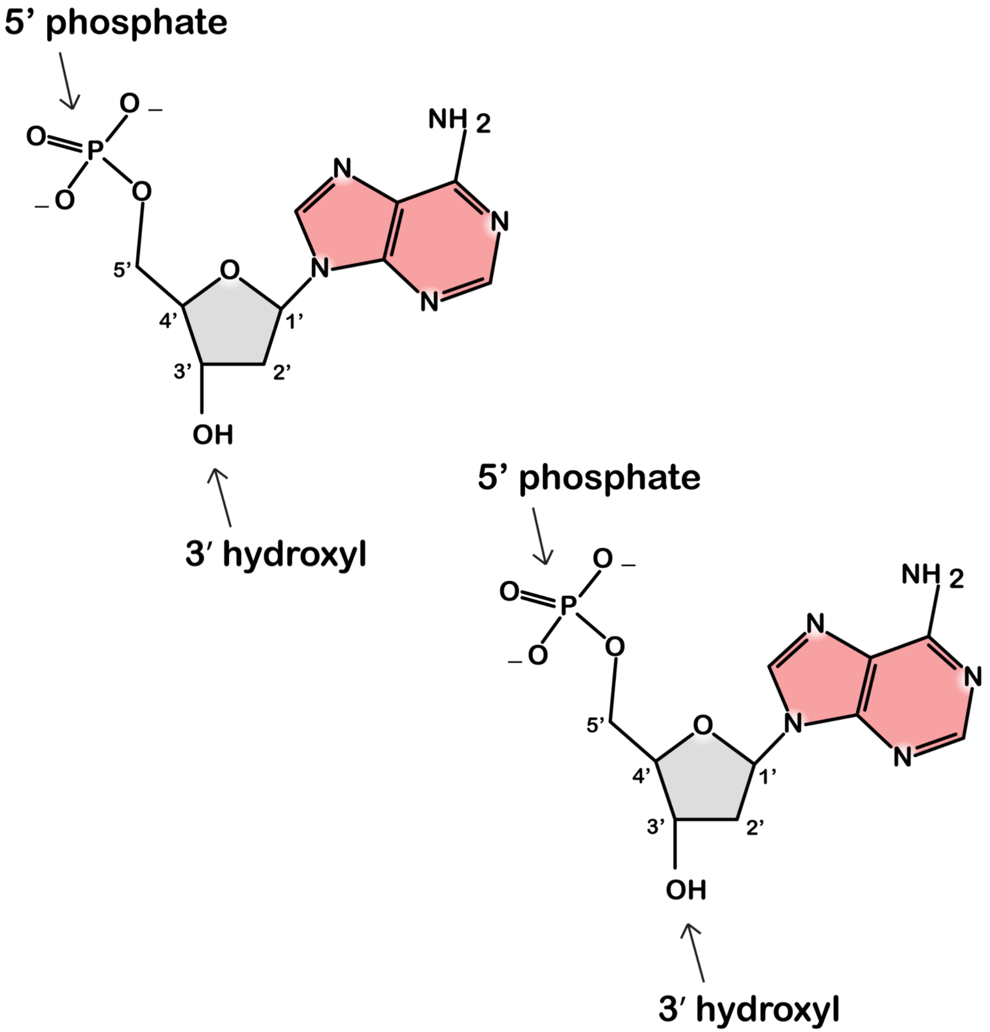
III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

Nucleotide assembly : Phosphodiester bond


Phosphodiester bond

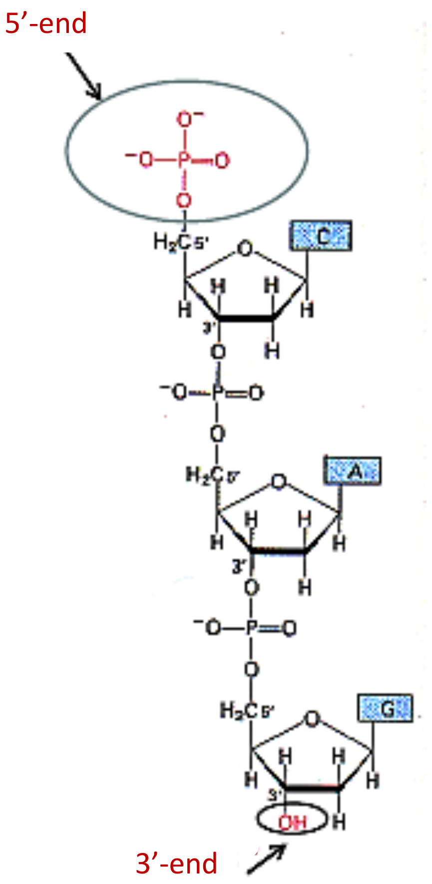

5’ phosphate

3′ hydroxyl

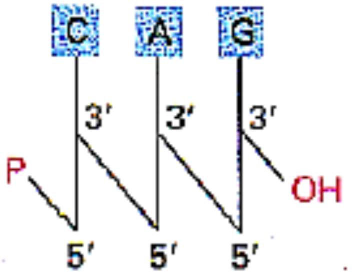
5’ phosphate

3′ hydroxyl

Representation of polynucleotides



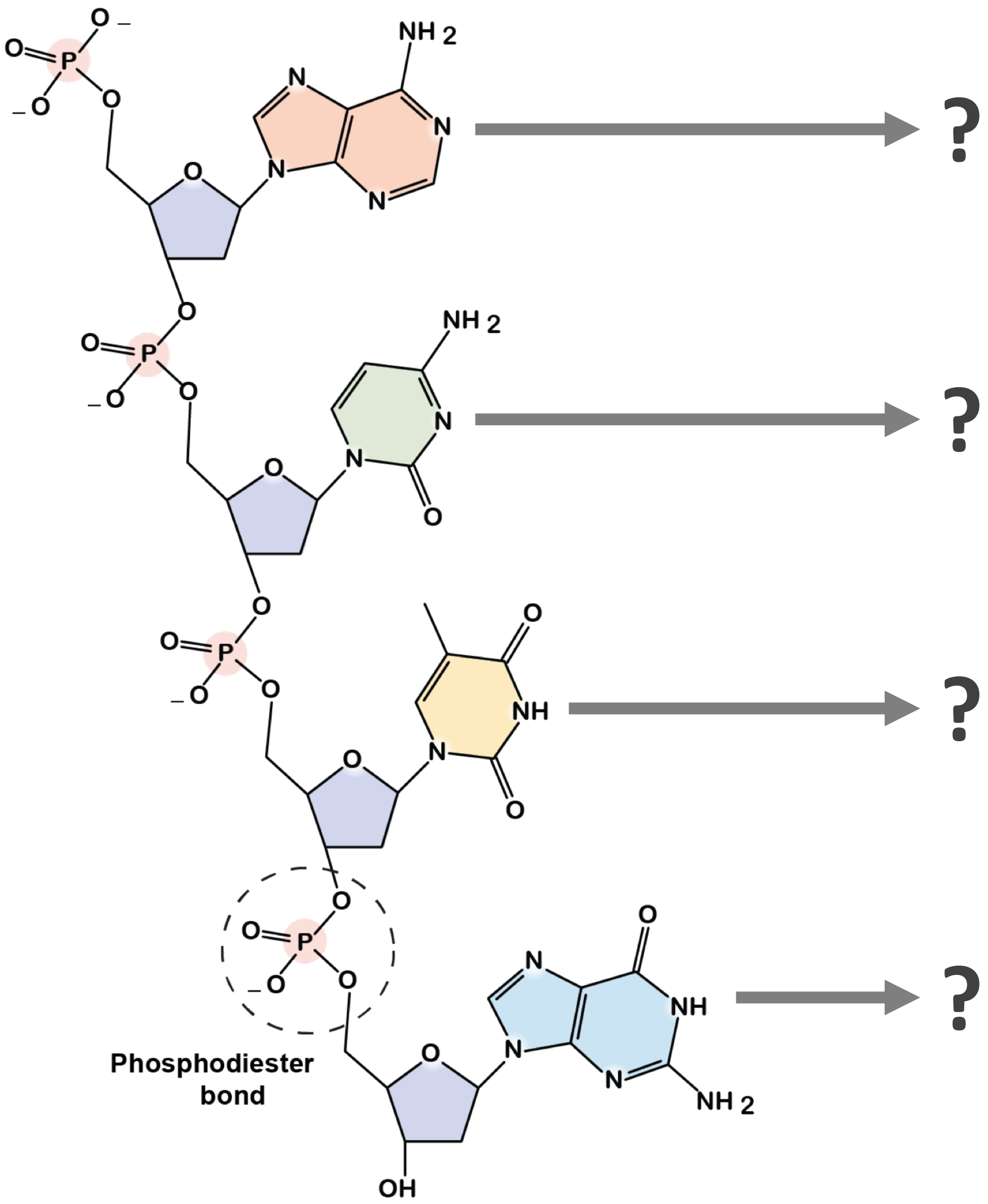
- Two common patterns used to represent polynucleotides



OR

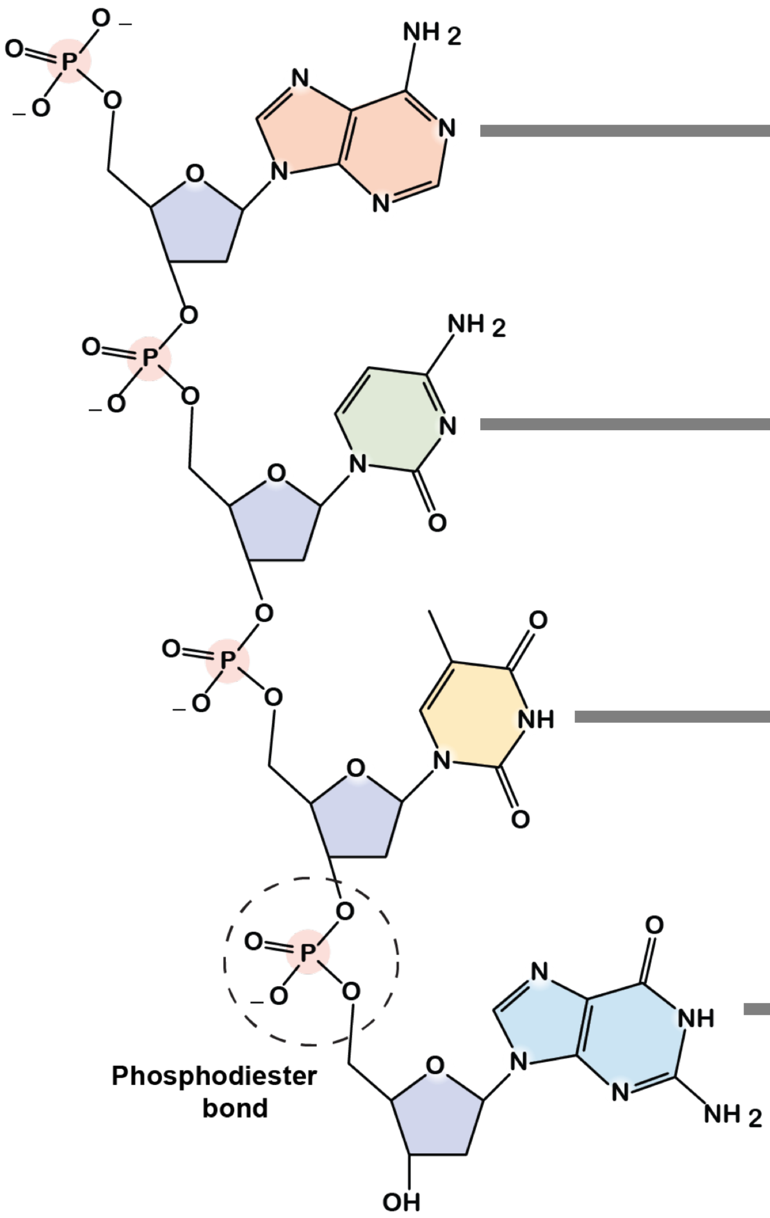
5'C-A-G 3'

Nucleotide assembly : Phosphodiester bond



Phosphodiester bond

Nucleotide assembly : Phosphodiester bond



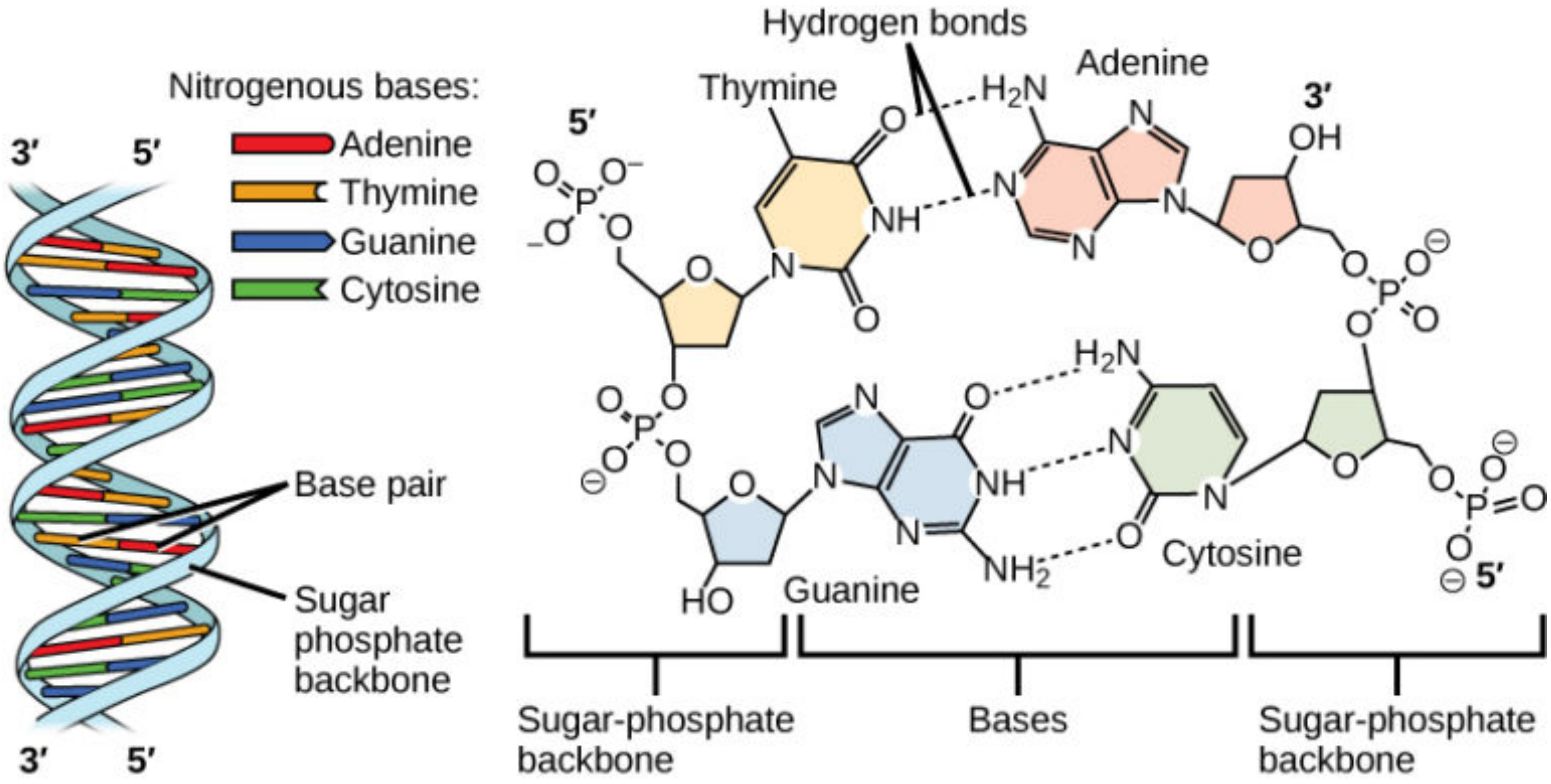
d-Adenosine monophosphate (dAMP)

d-Cytidine monophosphate (dCMP)

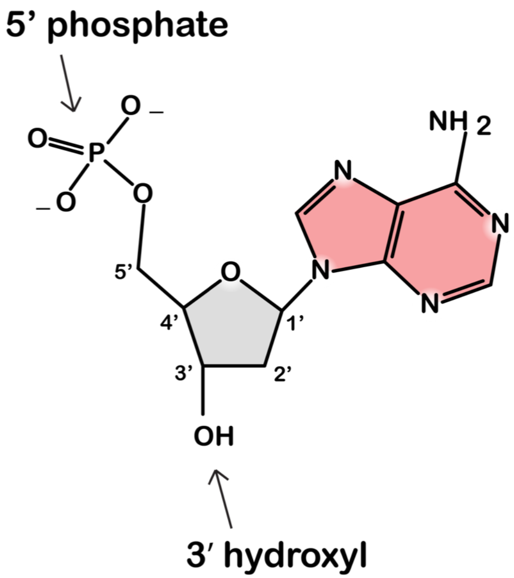
d-Thymidine monophosphate (dCTP)

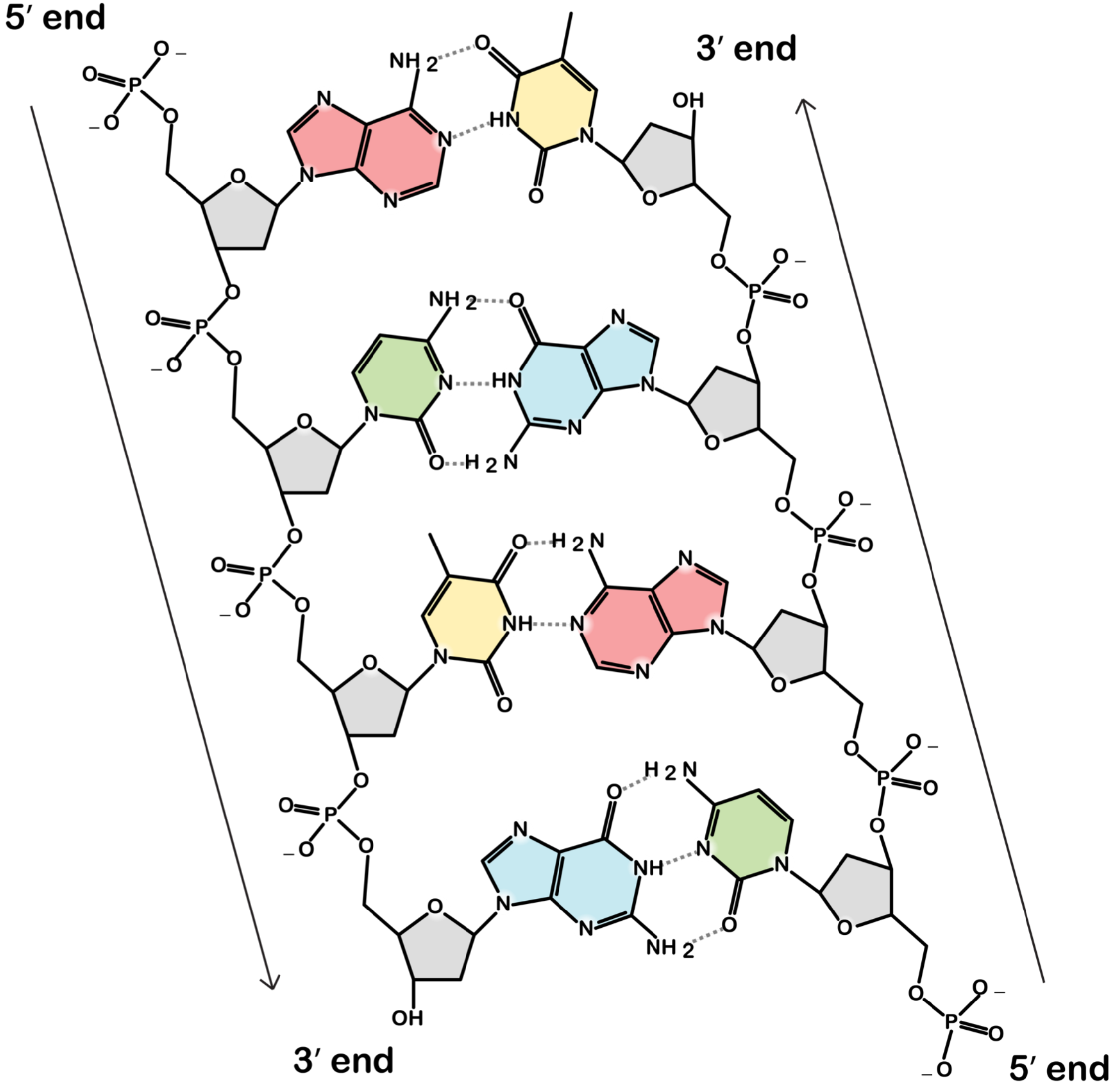
d-Guanosine monophosphate (dCGP)

Nucleotide assembly : Hydrogen bonds

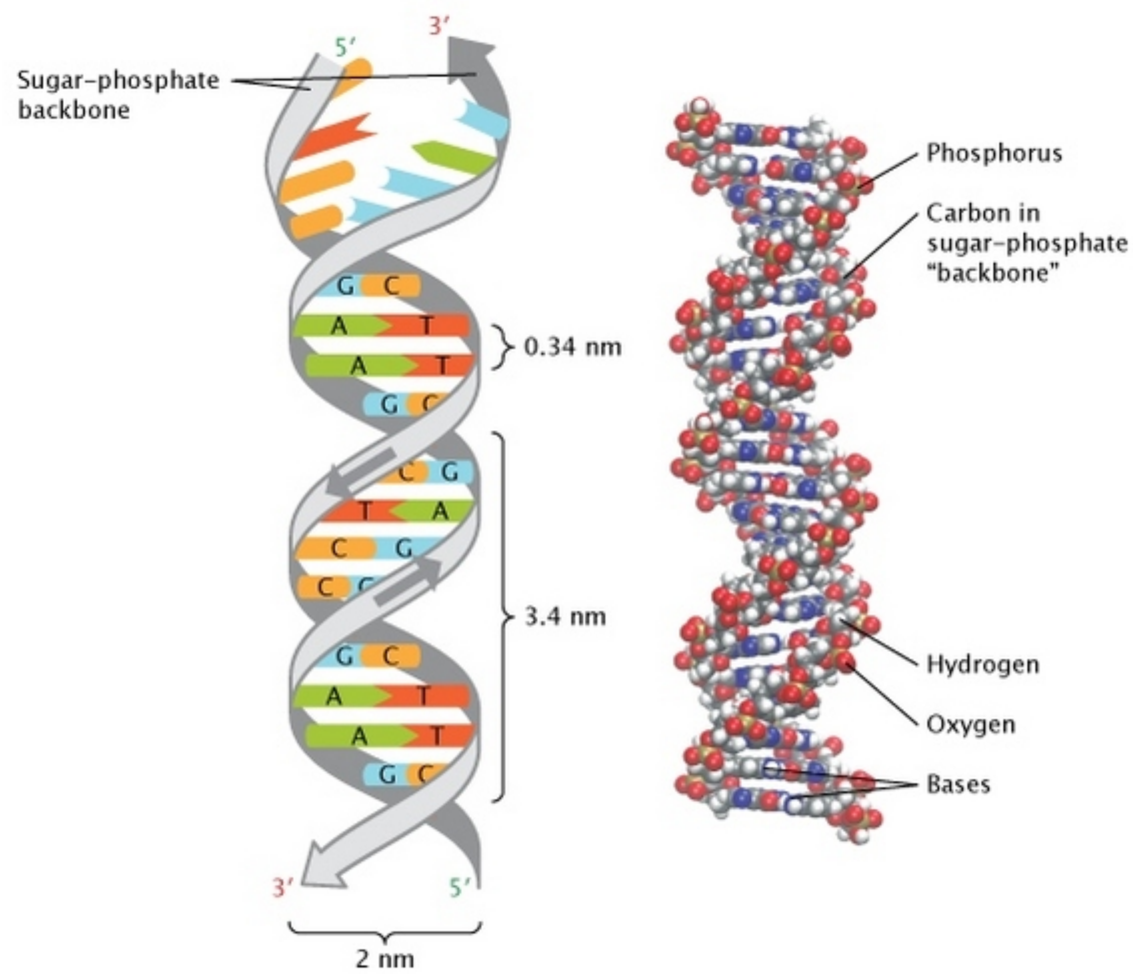


5‘ and 3’ ends





Native DNA shape = double helix

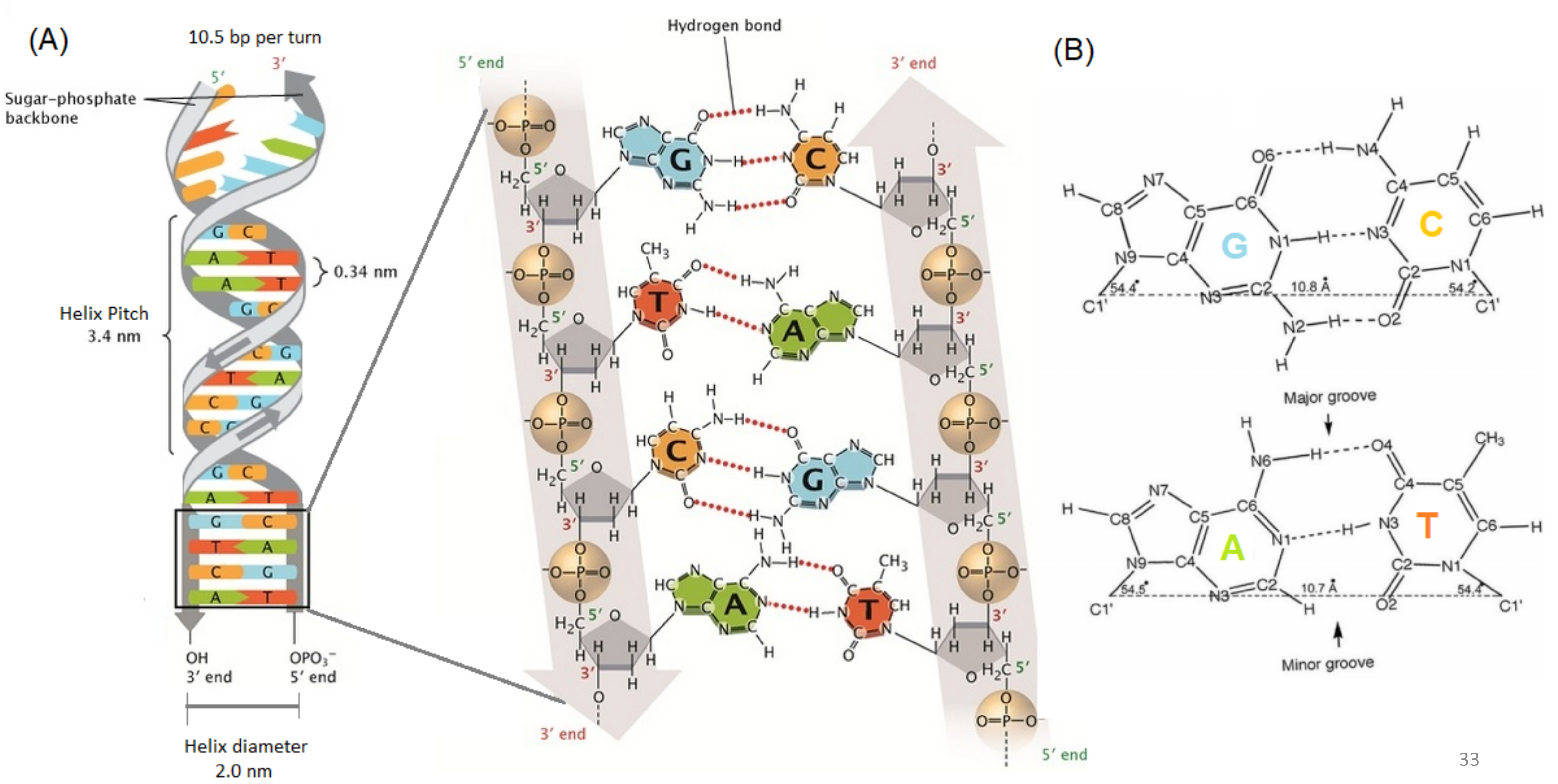


1) The 2 chains form a right hand double helix

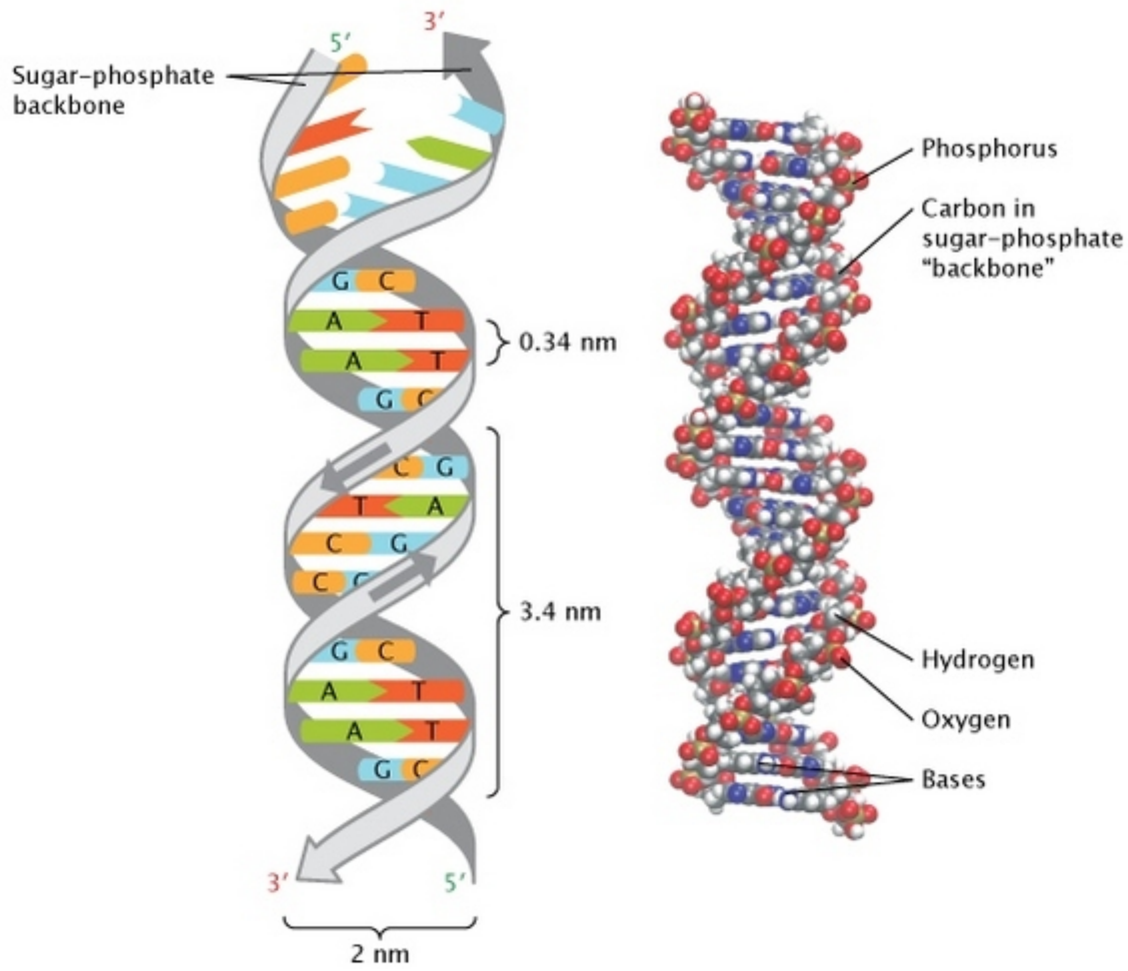
2) The sugar-phosphate backbone is located on the outside of the molecule

3) The bases occupy planes roughly perpendicular to the major axis of the molecule and are located inside

Native DNA shape = double helix



Native DNA shape = double helix



1) The 2 chains form a right hand double helix

2) The sugar-phosphate backbone is located on the outside of the molecule

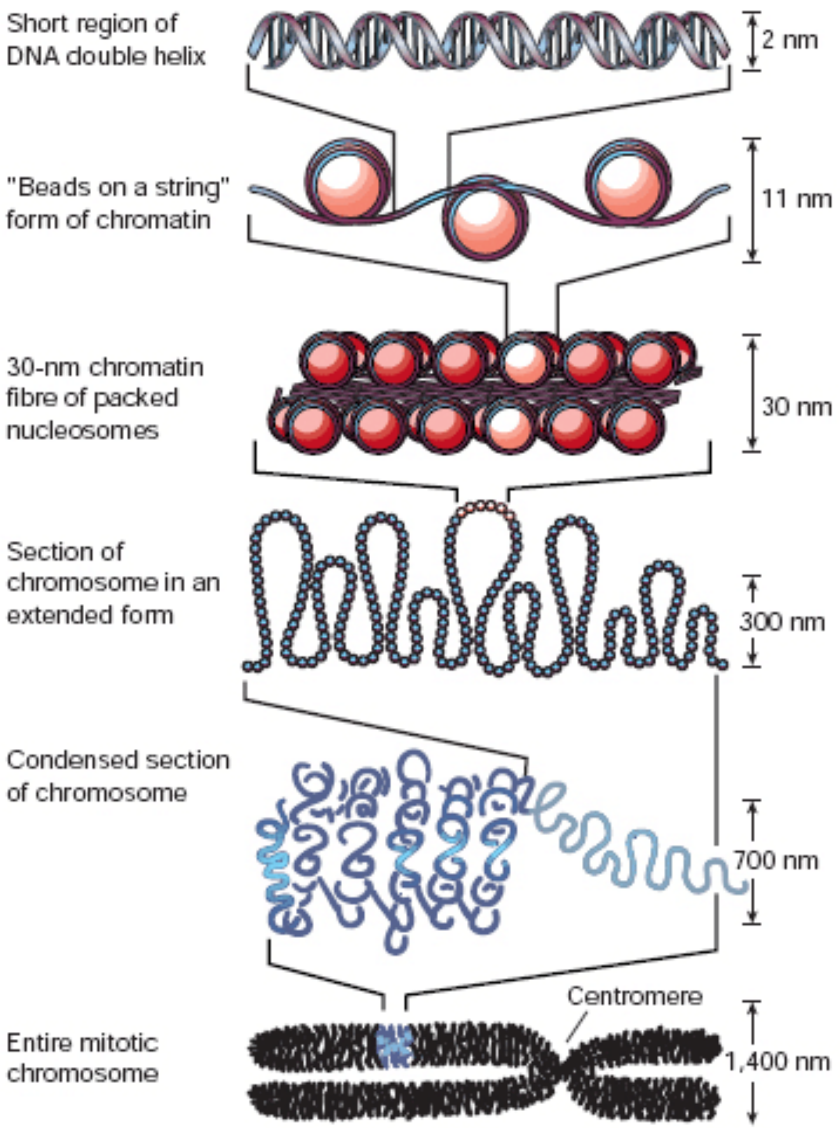
3) The bases occupy planes roughly perpendicular to the major axis of the molecule and are located inside

4) The 2 chains are held together by hydrogen bonds (H)

5) The only possible base pairs are A-T and G-C

6) Both chains are anti-parallel

Necessary DNA compaction in eukaryotic cell



In Human genome

☞ 6.4 billion base pairs

  ⇘ 3.4 Å between each base

    ⇘ ≈ 2 meters long

☞ In each eukaryotic cell

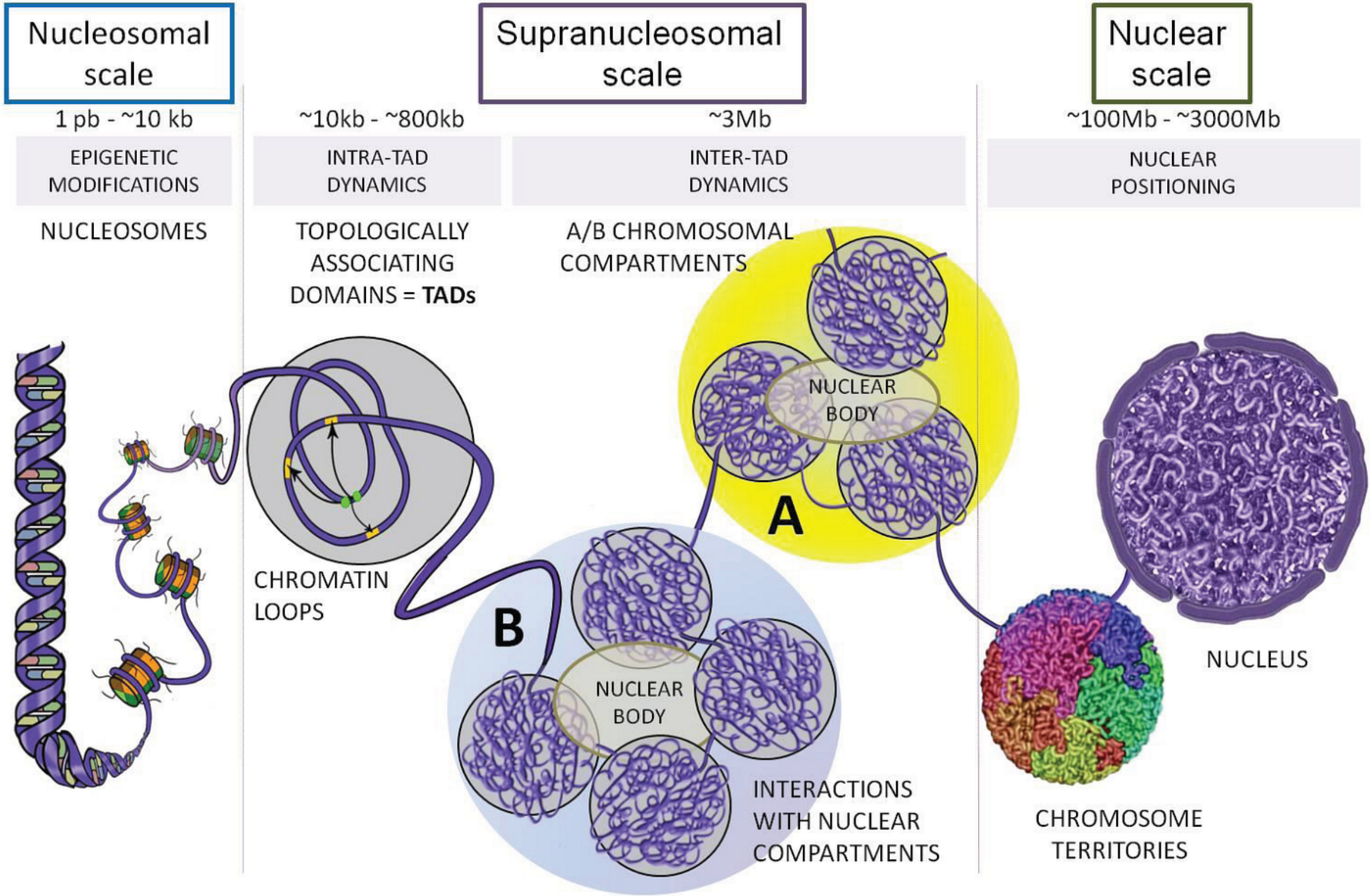
  ⇘ need to fit into an about 6 µm diameter nucleus

⇨ Possible compaction by proteins (histones, etc.)

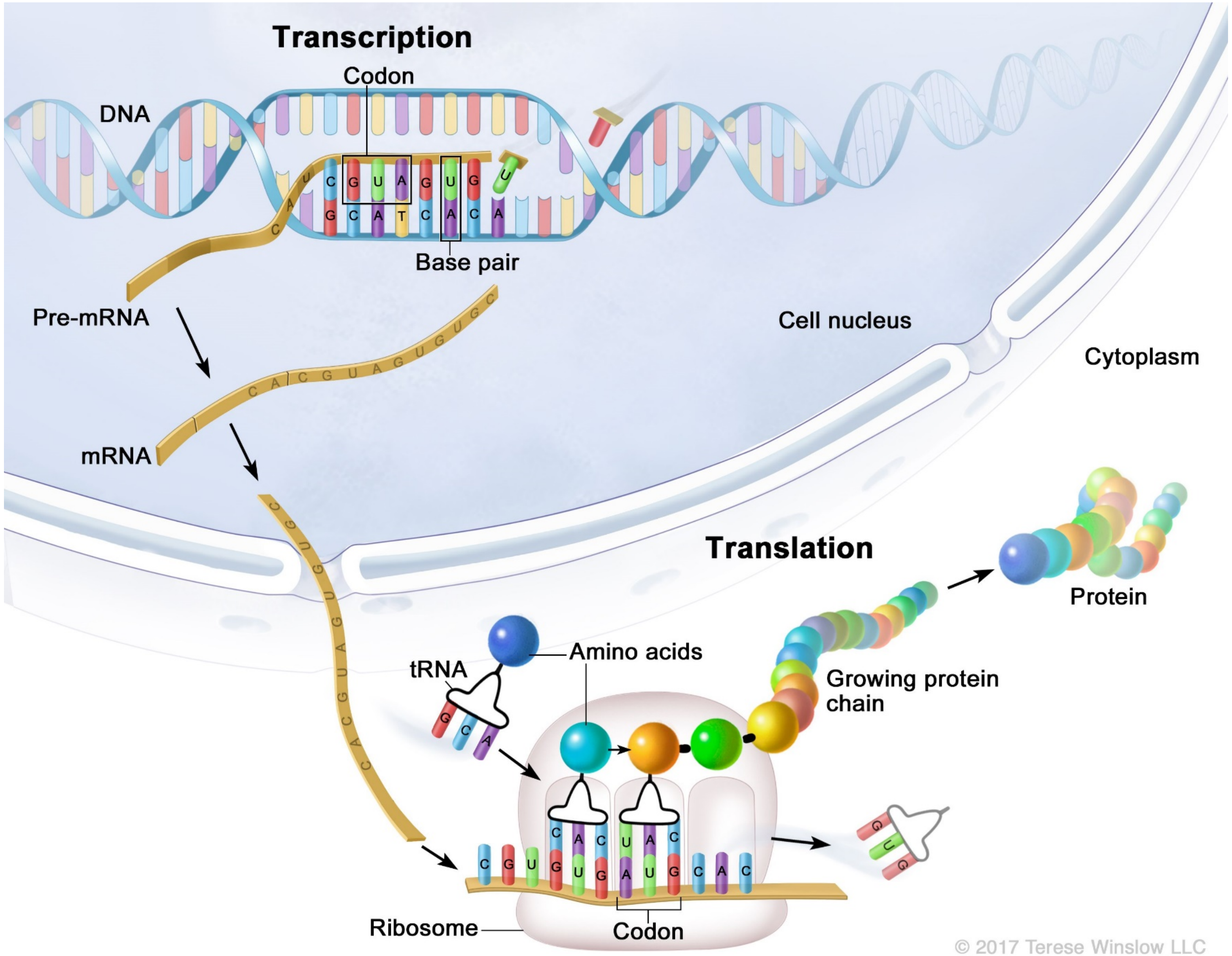
If all the DNA in your body was put end to end, it would reach to the sun and back over 600 times

Maximal compaction…

Necessary DNA compaction in eukaryotic cell



Plan



I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

Genome

☞ Each species has a specific set of genetic information stored in chromosomes that forms its genome

  * number of specific chromosomes

  * specific genetic information (DNA sequence)

☞ In Prokaryotes

  The genome consists of a single circular chromosome

    +/- extrachromosomal structures: the plasmids

☞ In Eukaryotes

  The genome is

    * in the nucleus

    * represented by a number of chromosomes specific to each species

  Chromosomal structure = DNA + various proteins

    + DNA in cellular organelles

Genome

☞ In Eukaryotes

Most genomes are diploid

chromosomes are present in pairs

* several pairs of homologous chromosomes

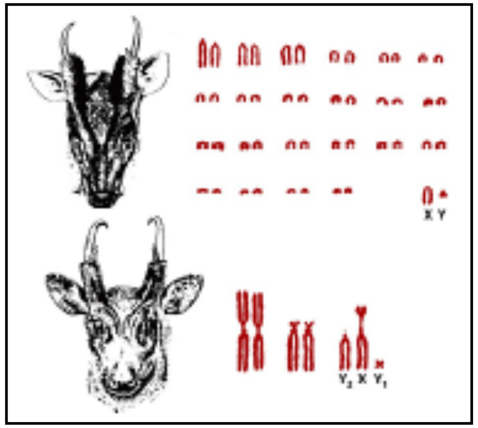
* 2 sex chromosomes

No correlation between

the number of chromosomes

the degree of evolution of the species

and the total size of the genome



⇨ The number of genes depends on the complexity of the species

Genome size of some organisms

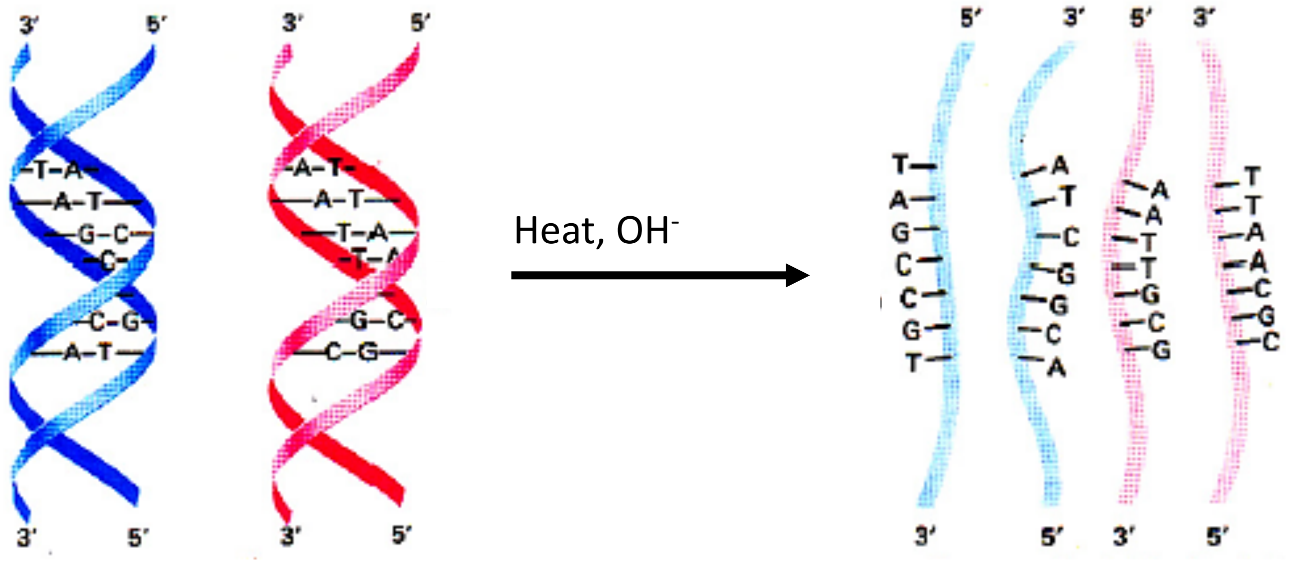
Organisms	Size of the haploid genome (bp)
Viruses	10^3 to 10^6
Bacteria	10^6 to 10^7
Yeasts	5×10^7
Mammals	10^8 to 10^{10}
Human	3×10^9
Plants	10^{10} to 10^{11}
Amœba dubia	$\approx 10^{11}$

⇨ No direct relationship between the degree of evolution of an organism and the size of its genome

Techniques for studying the complexity of genomes: denaturation of DNA

☞ When heating a DNA solution, the strands of the helix separate

↳ denaturation or melting

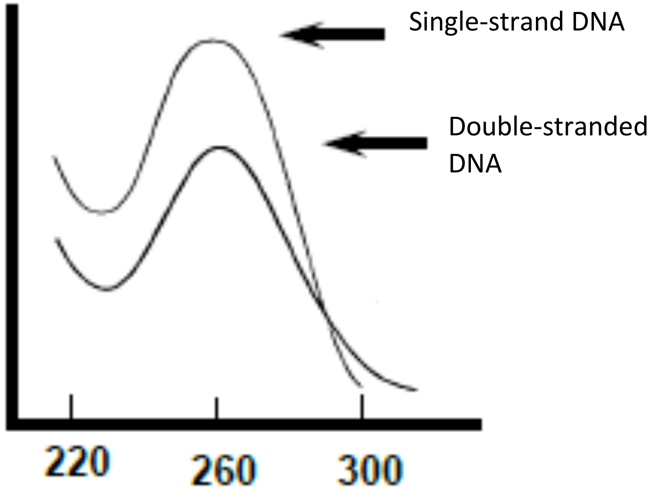


Native state

Single strand shape
= denatured form

Techniques for studying the complexity of genomes: denaturation of DNA

☞ Denaturation or fusion of DNA is studied by measuring the OD (absorbance) in UV (260 nm)



UV absorption spectrum of single and double-stranded DNA solutions

⇨ Hyperchrome effect

☞ Nitrogen bases absorb UV with maximum absorbance at 260 nm

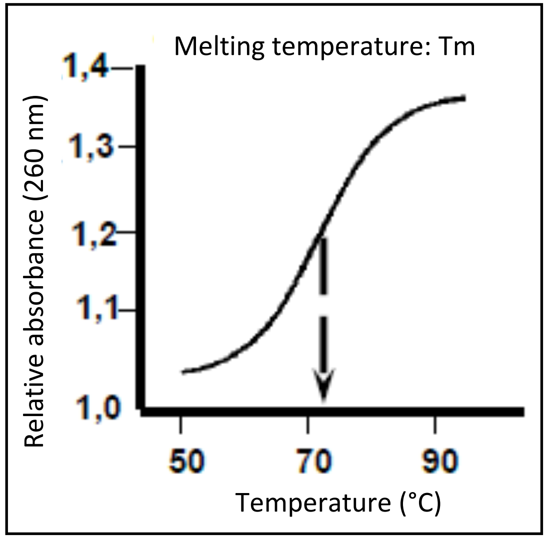
☞ As soon as the 2 strands of DNA are separated (denaturation)

↳ there is an increase in absorbance at 260 nm

Techniques for studying the complexity of genomes: denaturation of DNA

☞ The temperature at which the absorbance is increased by half

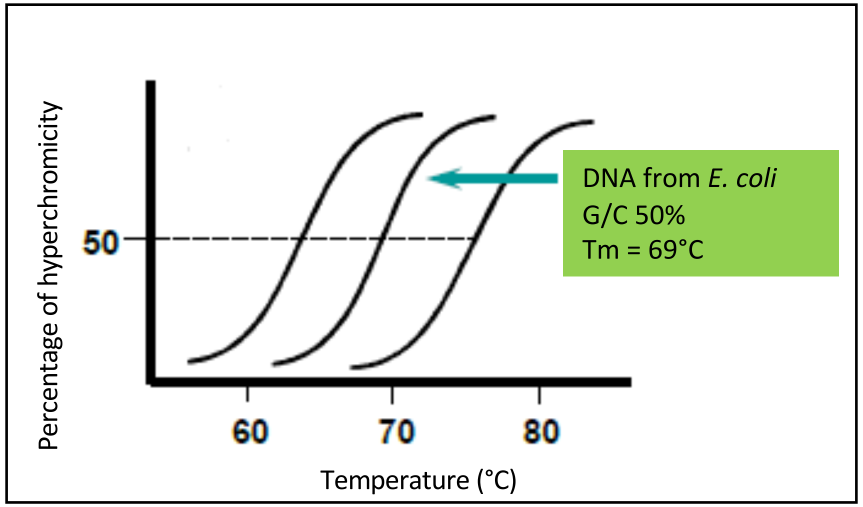
= melting temperature or Tm



Tm = temperature at which half of the DNA bases have lost their pairing

↳ 1 base not paired out of 2

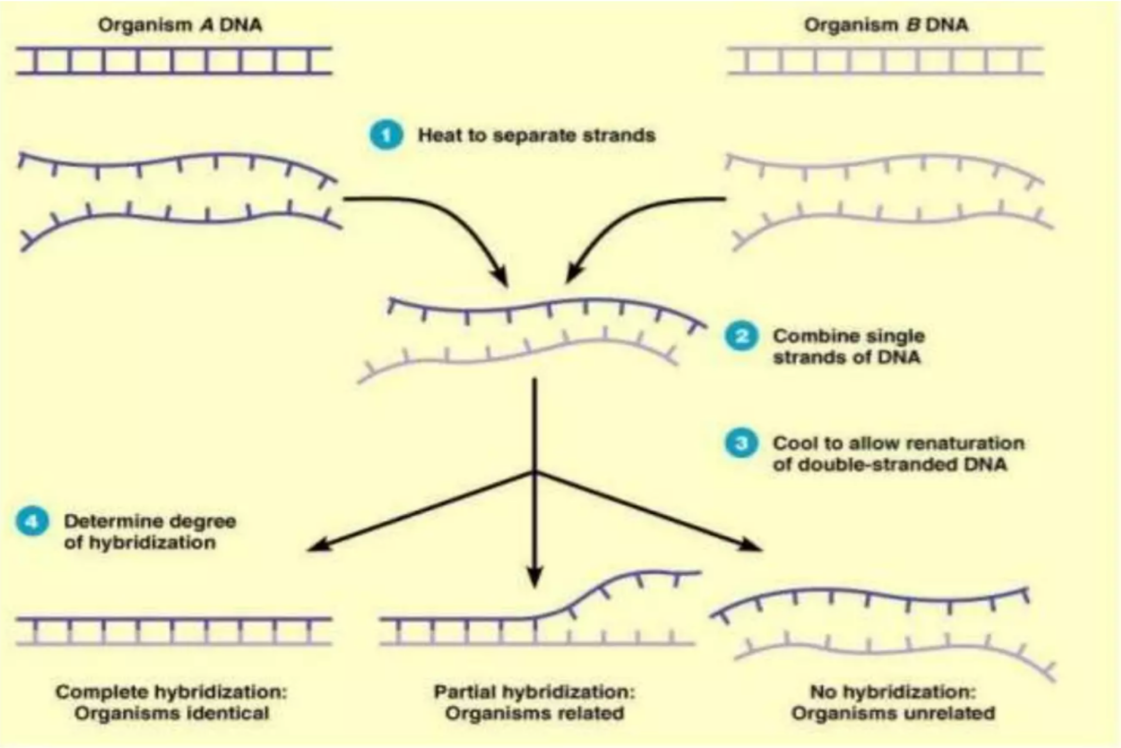
☞ The Tm of a genomic DNA extract depends on the G/C percentage



Techniques for studying the complexity of genomes: renaturation of DNA

☞ Methodology called "nucleic acid hybridization"

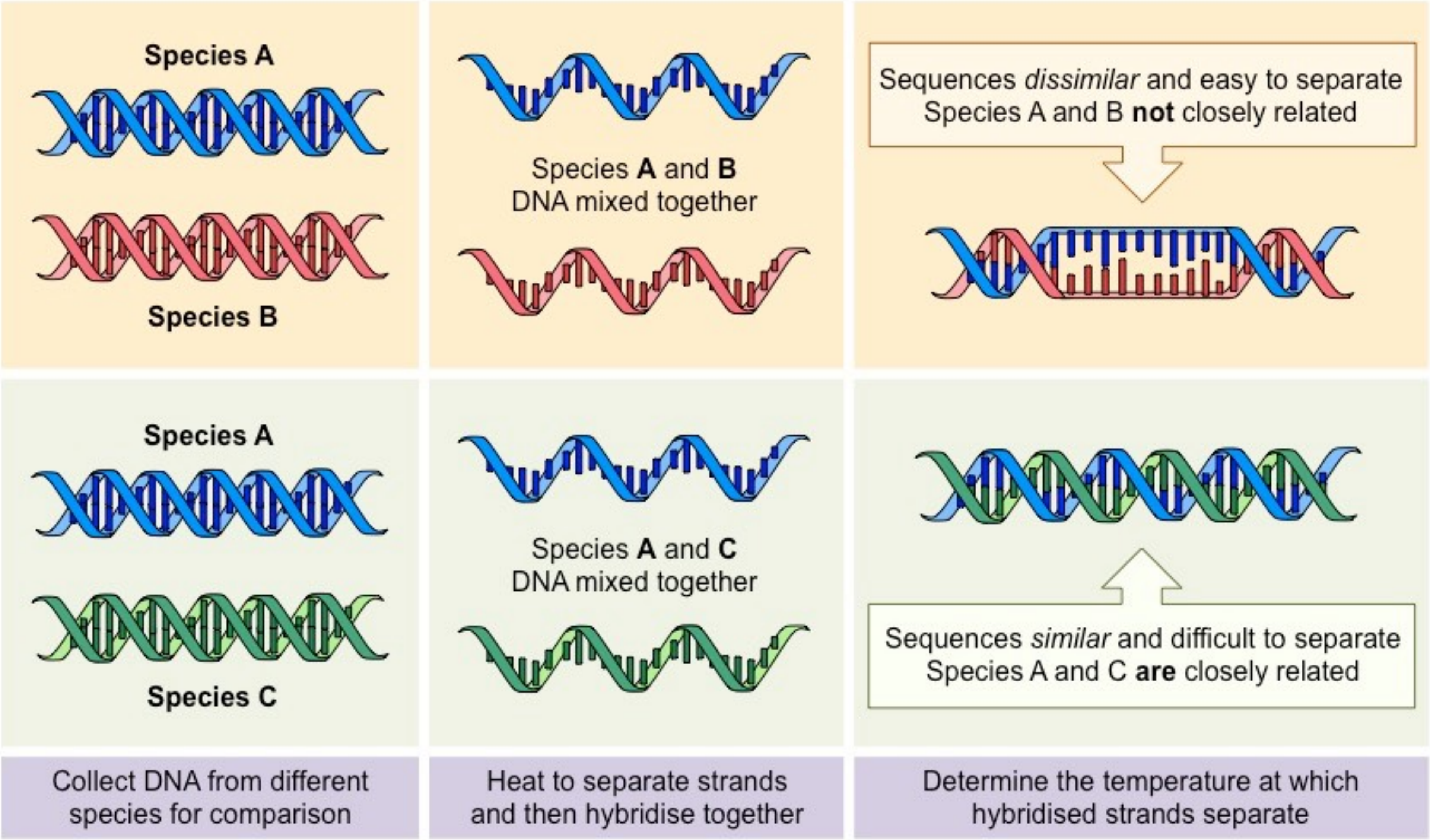
- ↳ estimation of the relatedness of 2 DNA (sequence)



☞ The rate of renaturation depends on

- * the ionic strength
- * the incubation temperature
- * the DNA concentration
- * the incubation period
- * the size of the molecules that interact

Techniques for studying the complexity of genomes: renaturation of DNA



Sequences *dissimilar* and easy to separate
Species A and B **not** closely related

Sequences *similar* and difficult to separate
Species A and C **are** closely related

Collect DNA from different species for comparison

Heat to separate strands and then hybridise together

Determine the temperature at which hybridised strands separate

Techniques for studying the complexity of genomes: renaturation of DNA

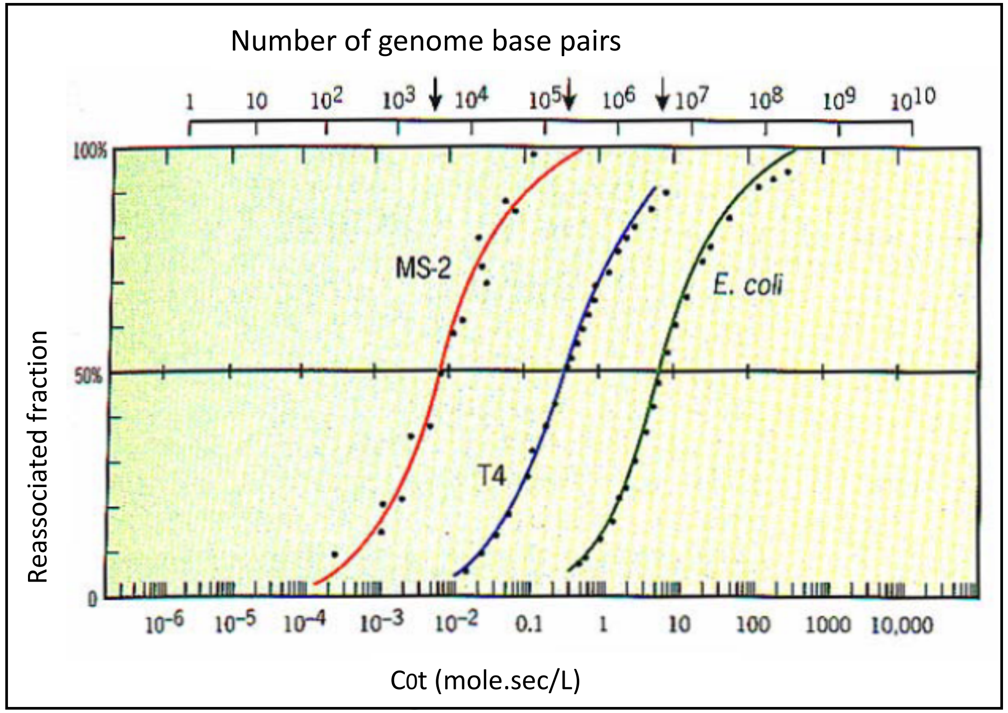
☞ Renaturation parameters C_0t

C_0 = molar concentration of DNA (mol/L)

↳ depends on the size of the genome

t = incubation time (seconds)

Complexity of viral and bacterial genomes



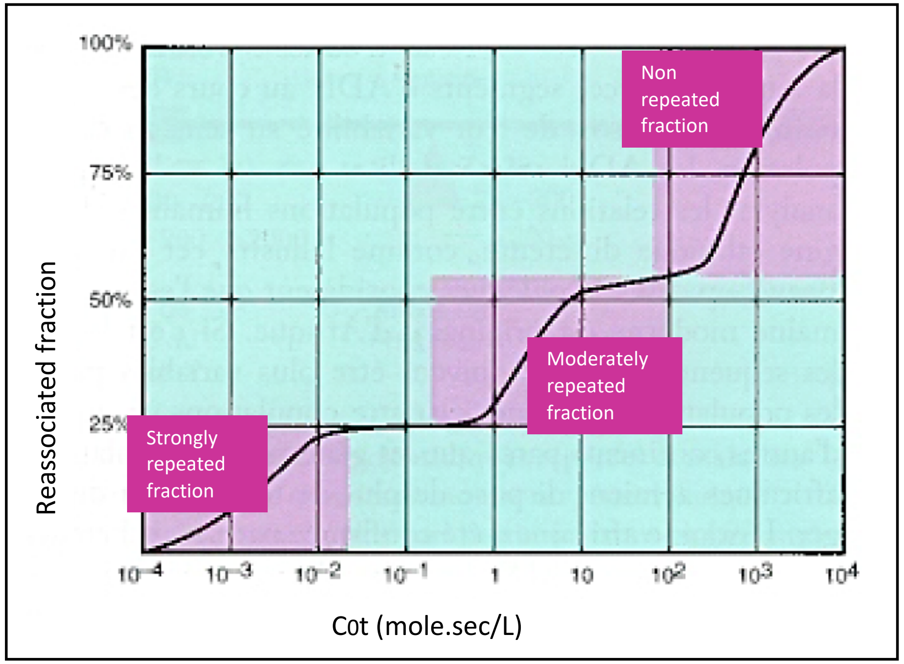
☞ The time required for renaturation depends on the concentration of the complementary fragments

⇨ Renaturation is as faster as the genome is smaller

Techniques for studying the complexity of genomes: renaturation of DNA

☞ C_0t curves : 3 distinct steps for eukaryotic genomes

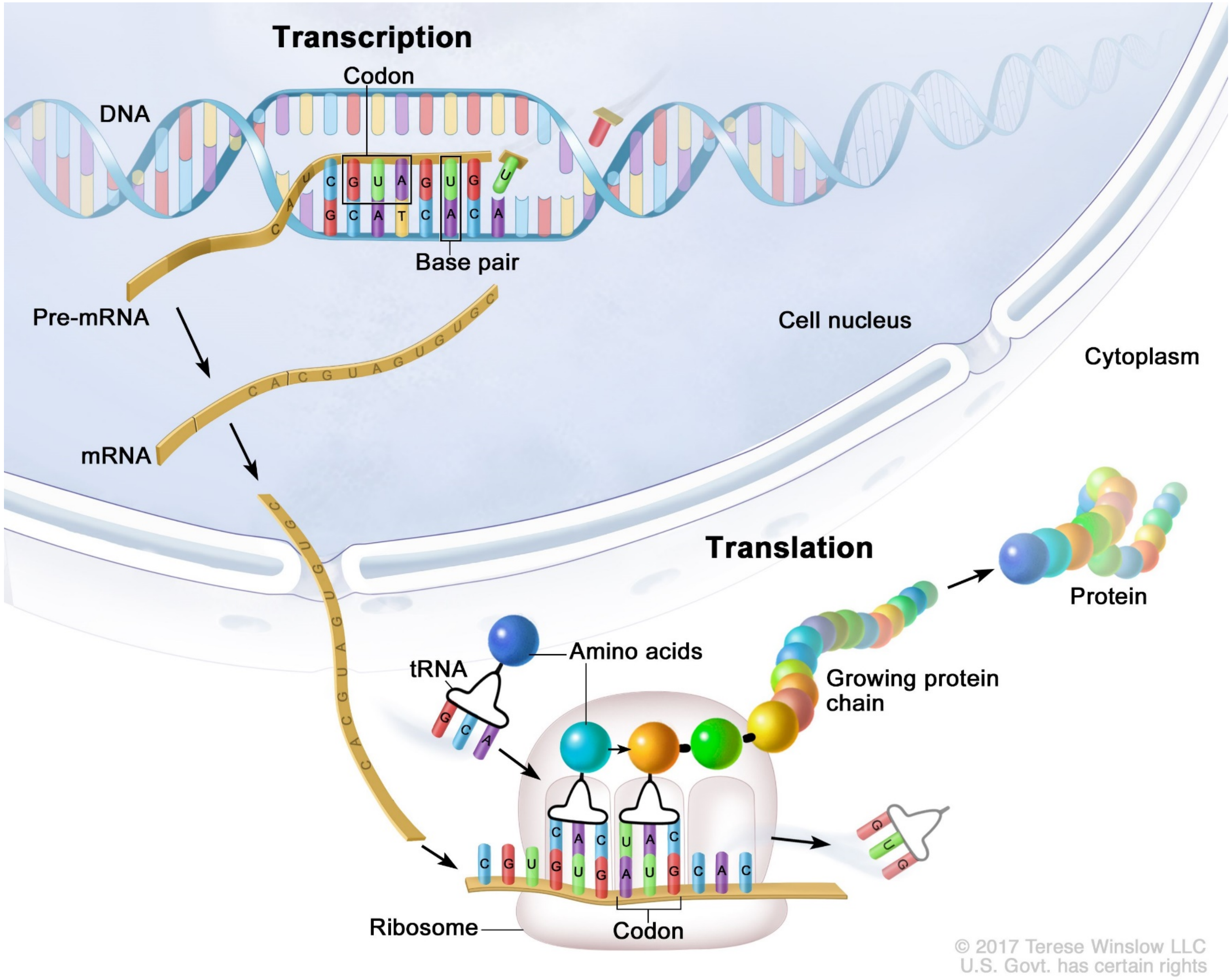
↳ reappearance of the 3 major classes of DNA sequences



☞ The 3 classes/ renaturation speeds

↳ function of the number of repetitions of DNA sequences

Plan



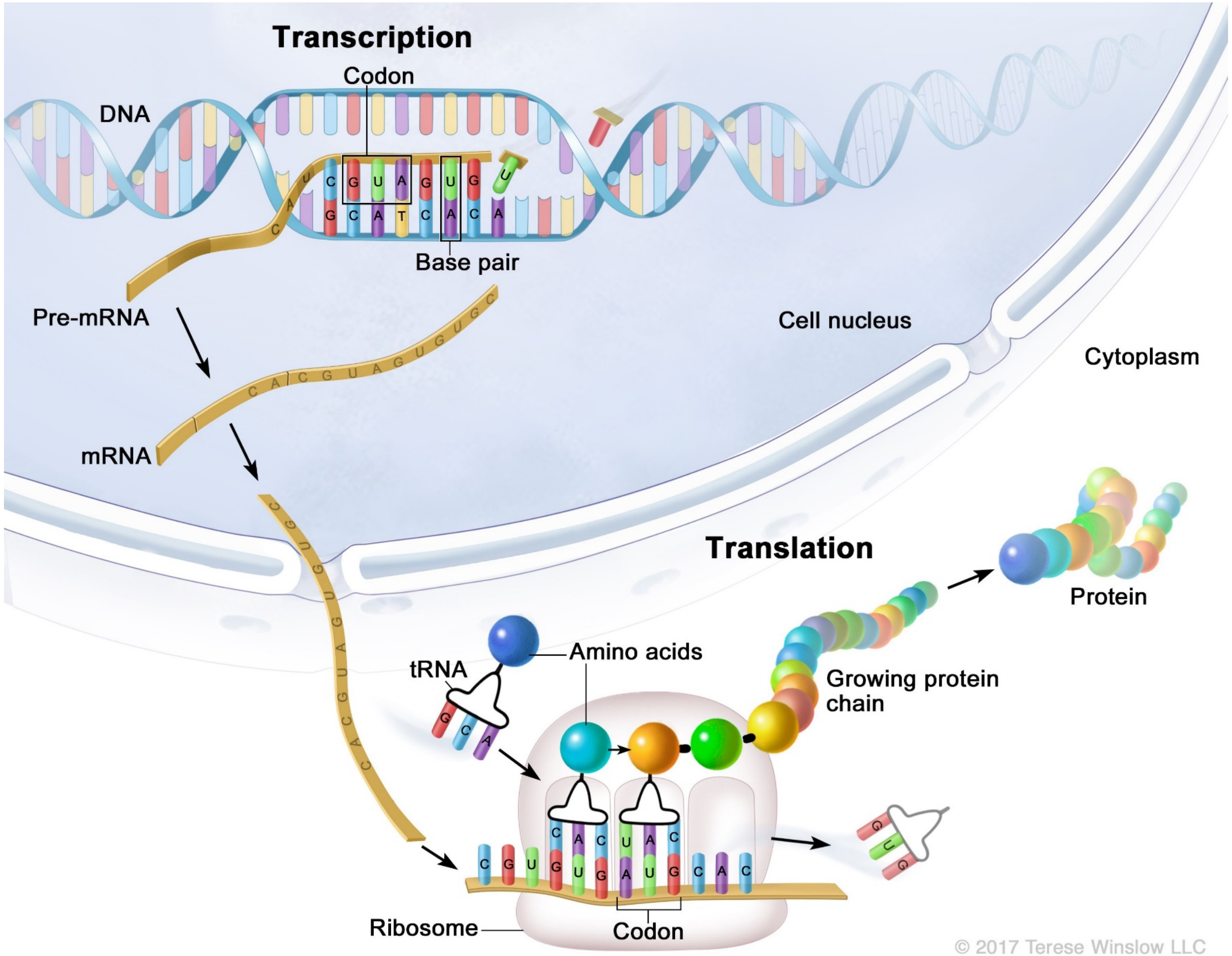
I - DNA and genome
- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA
- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA
- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

From nucleotide sequence to amino acid sequence



The nucleotide sequence of a gene determines the amino acid sequence of a protein

1) Messenger RNA (mRNA) = intermediate between a gene (DNA) and its corresponding polypeptide

2) Transcription = production of mRNA from DNA as a model

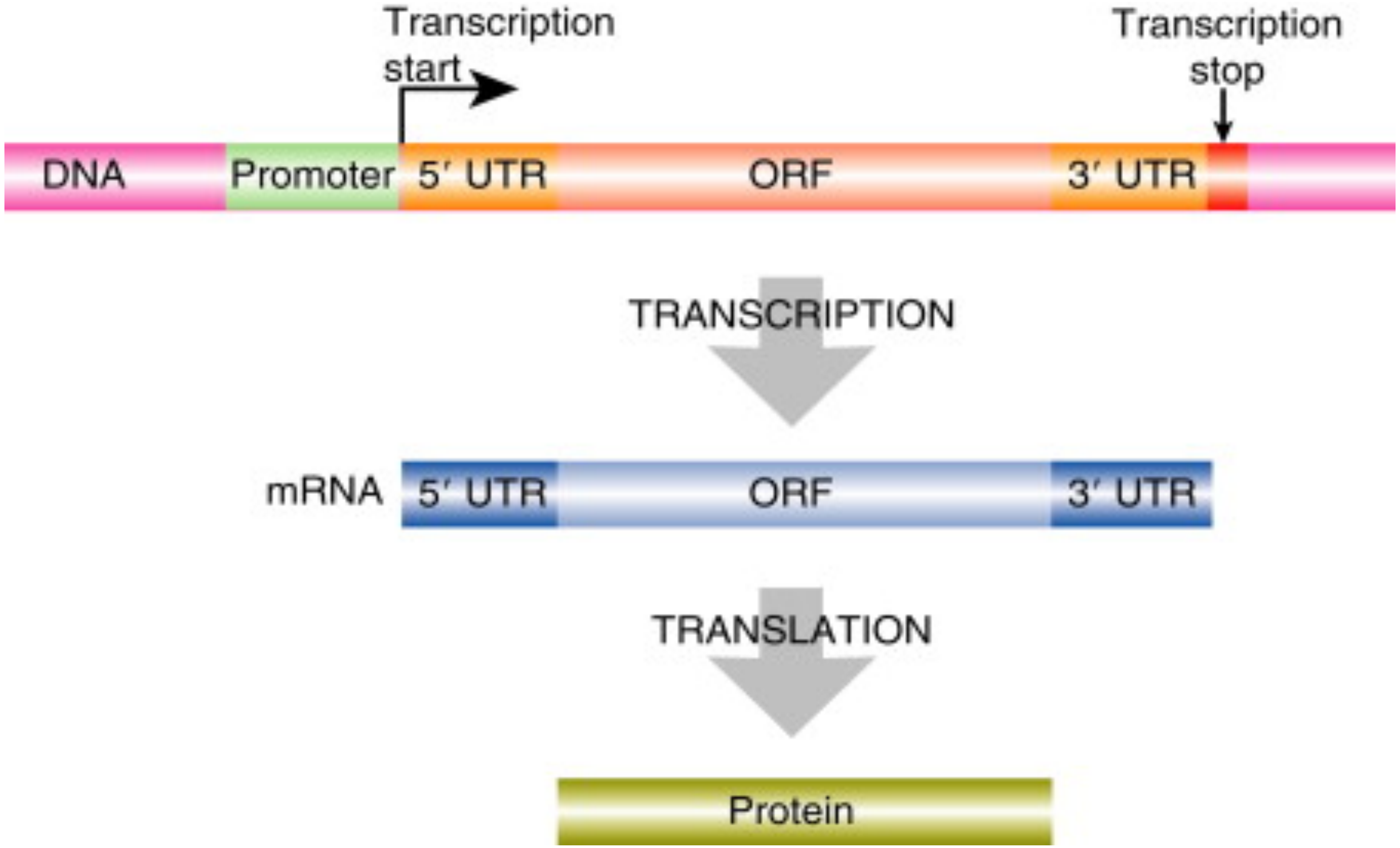
3) What are the benefits of using a mRNA?

- separate hereditary storage (DNA) from active information (mRNA)
- strongly amplify the synthesis activity

Different types of RNA with different functions

RNA	**Functions**
Messenger RNA = **mRNA**	Encode the proteins
Ribosomal RNA = **rRNA**	Form part of the ribosome structure and participate in protein synthesis
Transfer RNA = **tRNA**	Used in protein synthesis as adapters between mRNA and amino acids
Small nuclear RNA = **snRNA**	Used in many nuclear processes, including splicing of pre-mRNA
Small nucleolar RNA= **snoRNA**	Chemically modify rRNAs
microRNA = **miRNA**	Small inhibitors of mRNA translation with which they are complementary
Small interfering RNA = **siRNA**	Cause degradation of the mRNA of which they are complementary
Piwi-interacting RNA = **piwiRNA**	Protect germ cells from mobile elements
Non-coding RNA = **ncRNA**	Many thousands
Long non-coding RNA = **lncRNA**	Group of non-coding transcripts larger than 200 nt in size

Gene



☞ Promoter

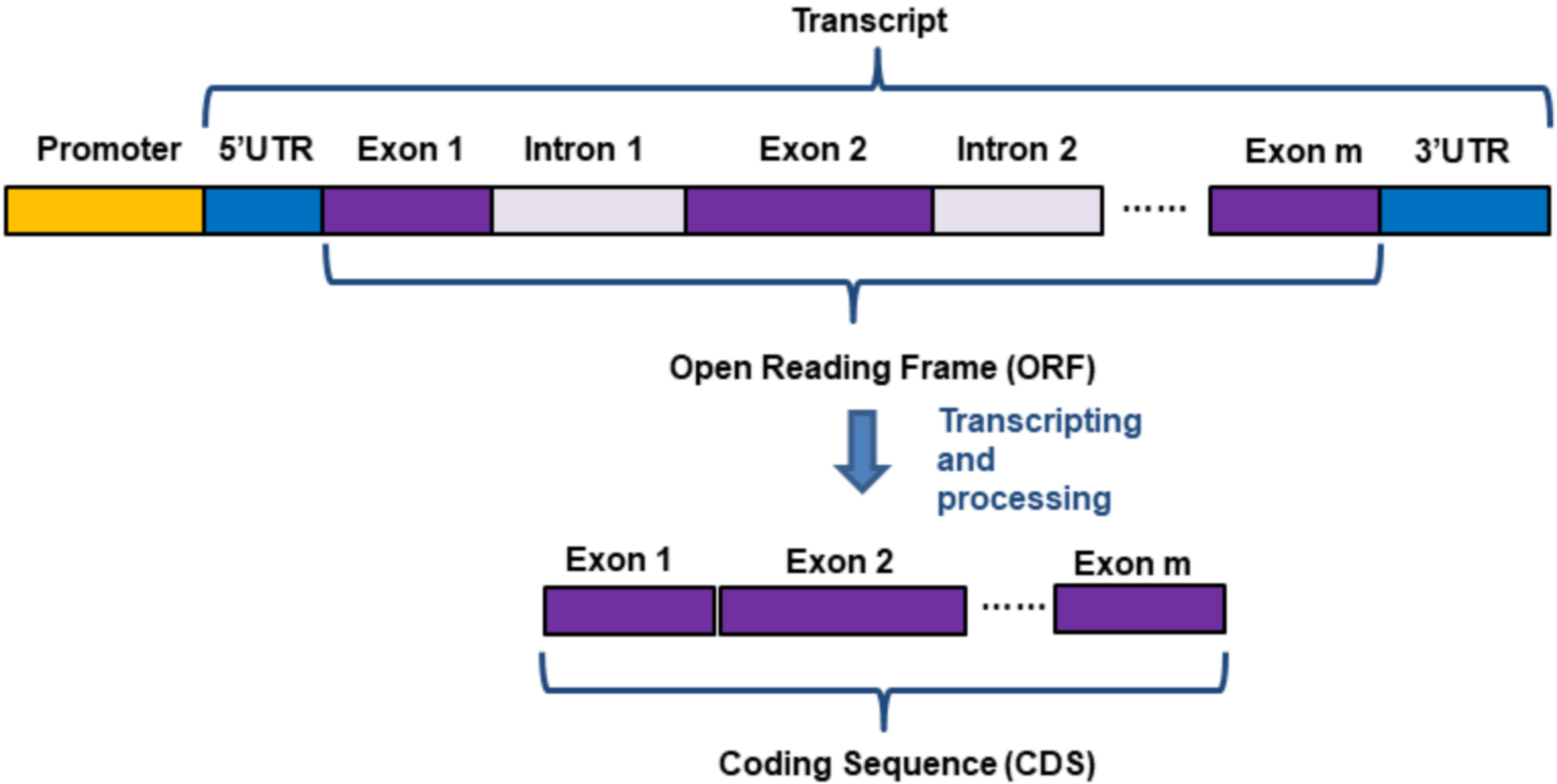
= start site of the DNA transcription

= site where RNA polymerase binds

= oriented DNA sequence

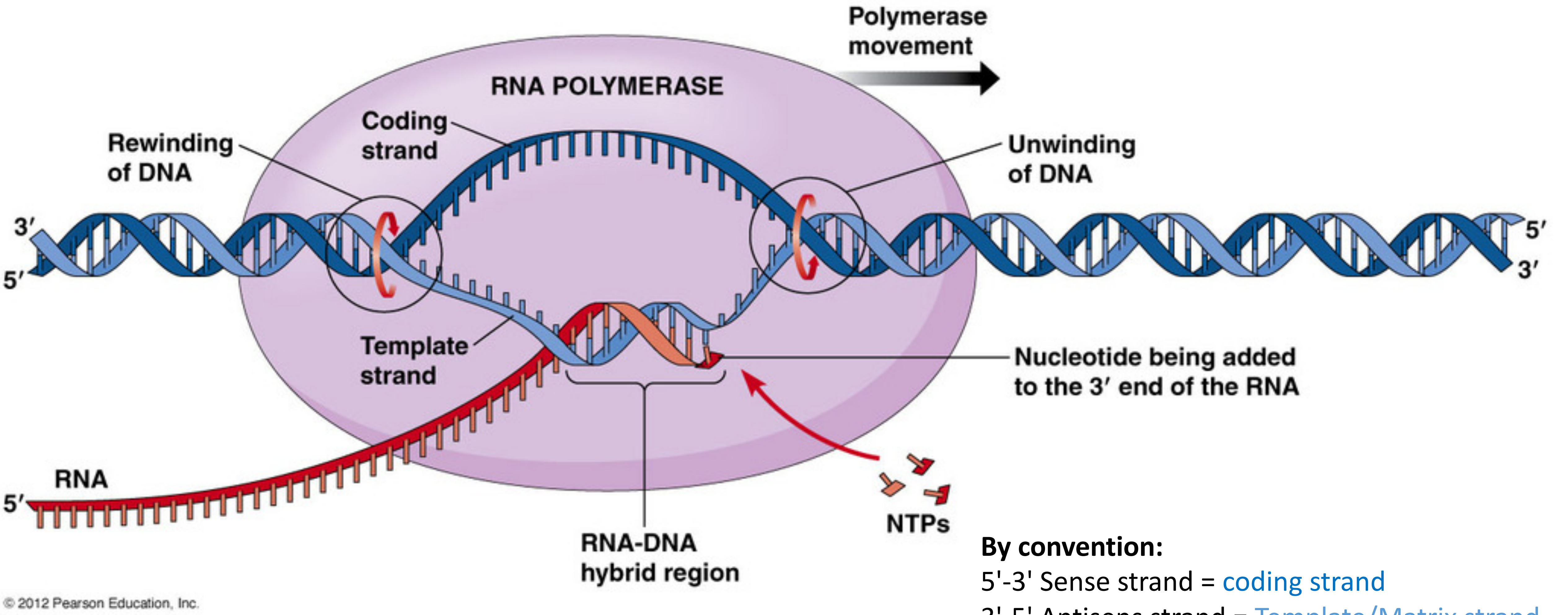
↳ determines the copied DNA strand (strand + or strand -)

Gene vs Transcript vs ORF vs CDS



Mechanism of transcription

Transcription = RNA synthesis from a DNA matrix



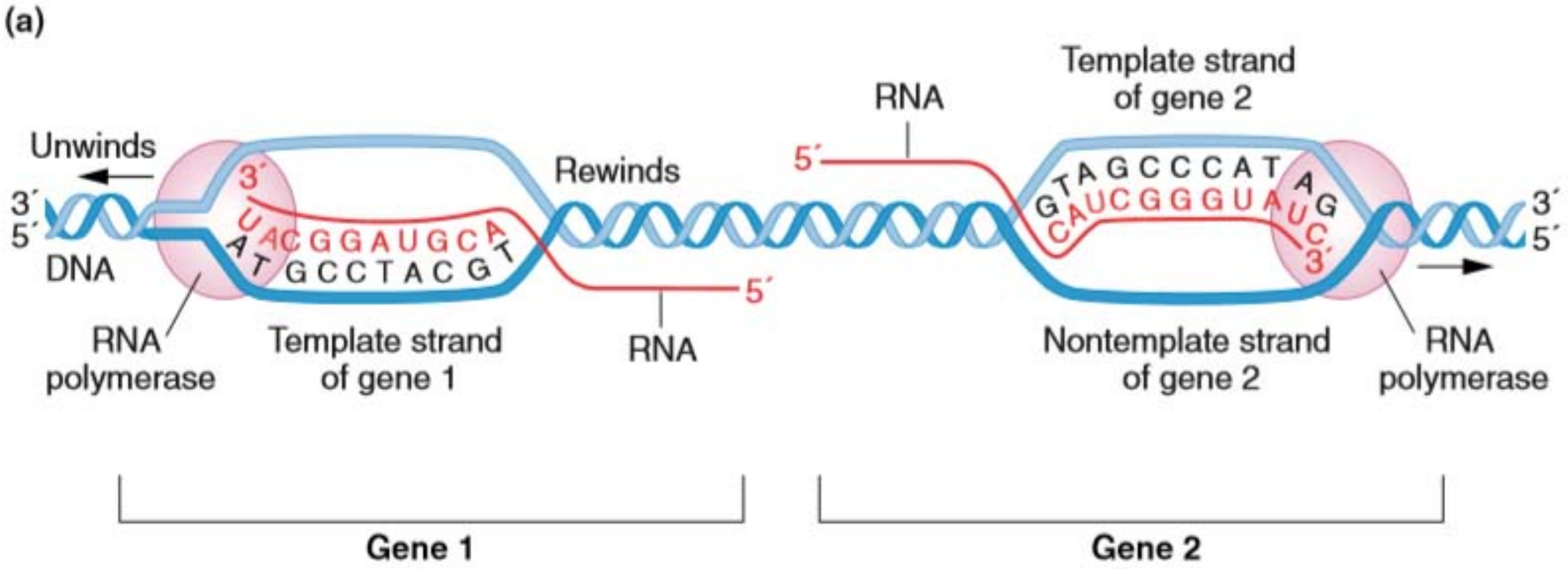
By convention:

5'-3' Sense strand = coding strand

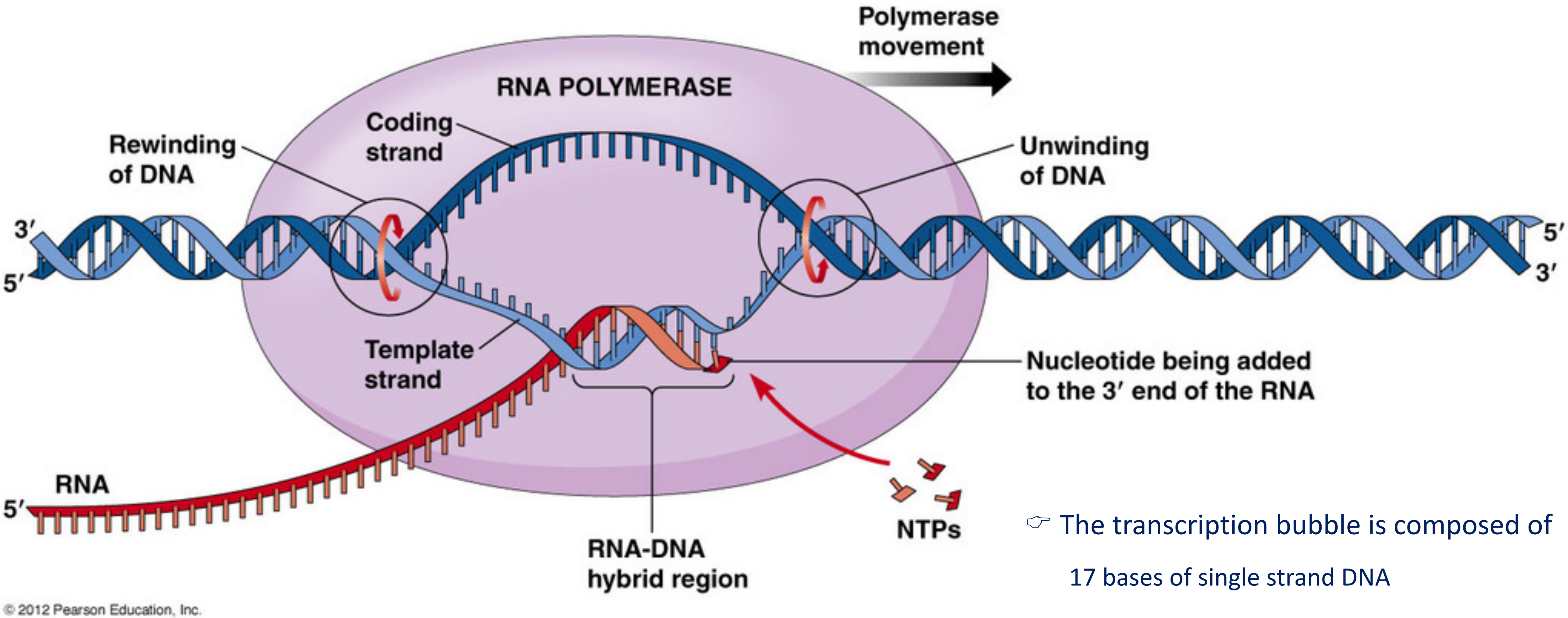
3'-5' Antisens strand = Template/Matrix strand

5'-3' Transcript

The two strands of DNA can be used as a matrix



Transcription bubble



☞ The transcription bubble is composed of

17 bases of single strand DNA

8 bp of DNA-RNA hybrid

Eukaryotic RNA polymerases

☞ 3 different RNA polymerases to synthesize RNA in Eukaryotes

Types of RNA polymerase	Transcribed DNA regions
RNA polymerase I	rRNA 5.8S, 18S, 28S
RNA polymerase II	Protein-coding genes, most nuclear RNA (snRNA and snoRNA)
RNA polymerase III	tRNA, rRNA 5S

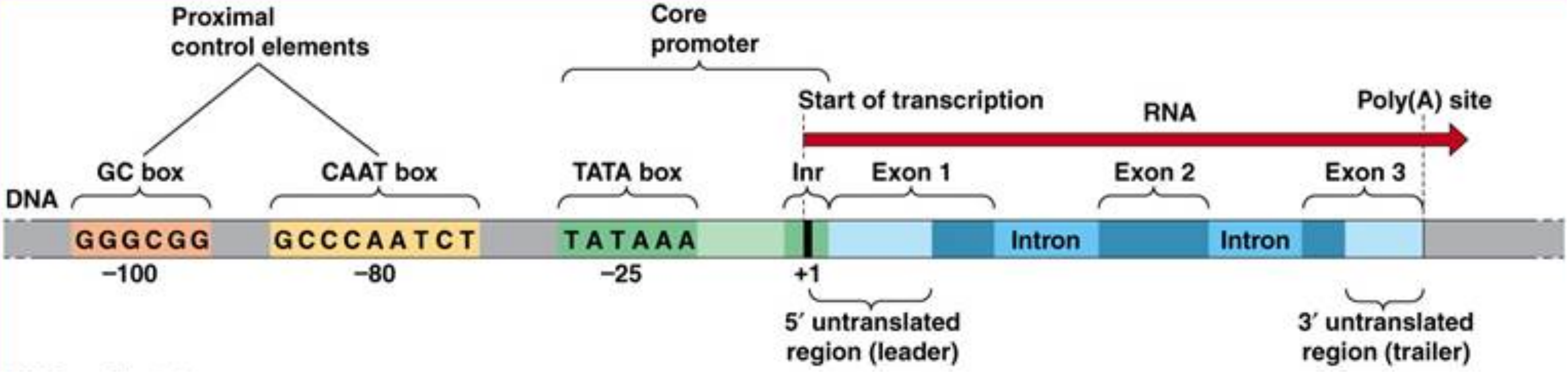
☞ RNA polymerases are formed from 8 to 14 distinct polypeptides

☞ They need transcription factors for their action

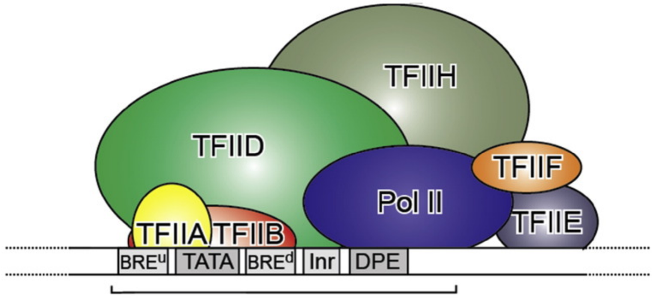
* General Transcription Factors (GTF)

* Specific transcription factors or gene regulation proteins

Mechanism for initiation of mRNA transcription



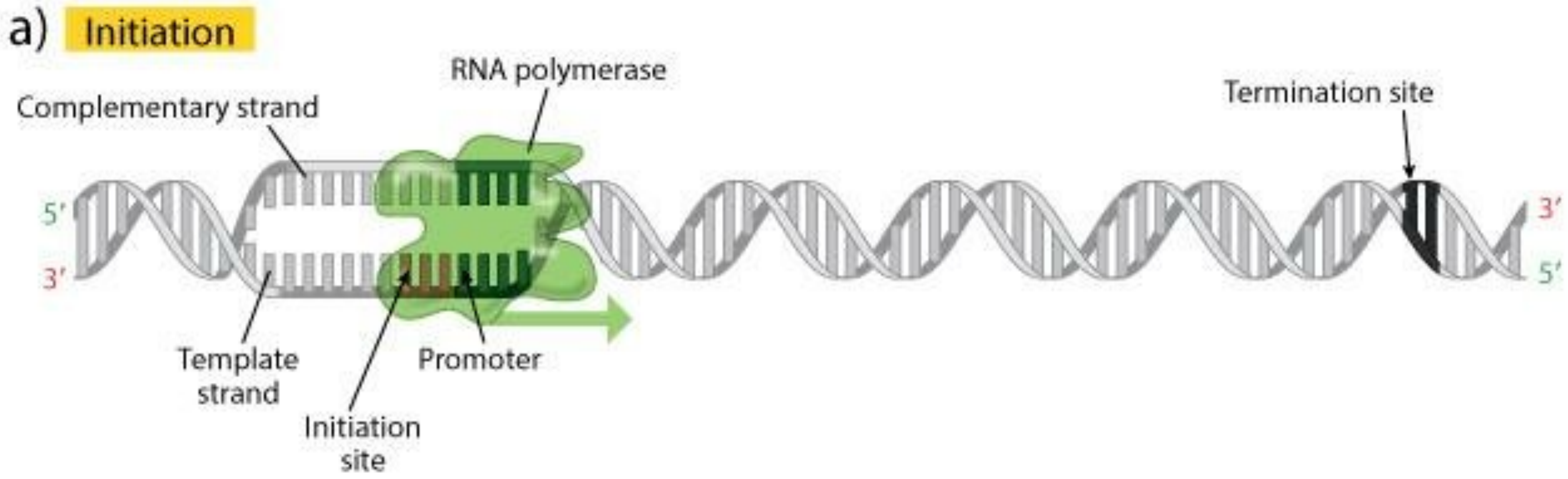
- ☞ Precursors of mRNAs synthesized by RNA polymerase II
- ☞ Initiation by general transcription factors, including TFIIA
- ☞ Promoter located in 5’ of each transcription unit
- ☞ TATA box = «classic» assembly site of the pre-initiation complex



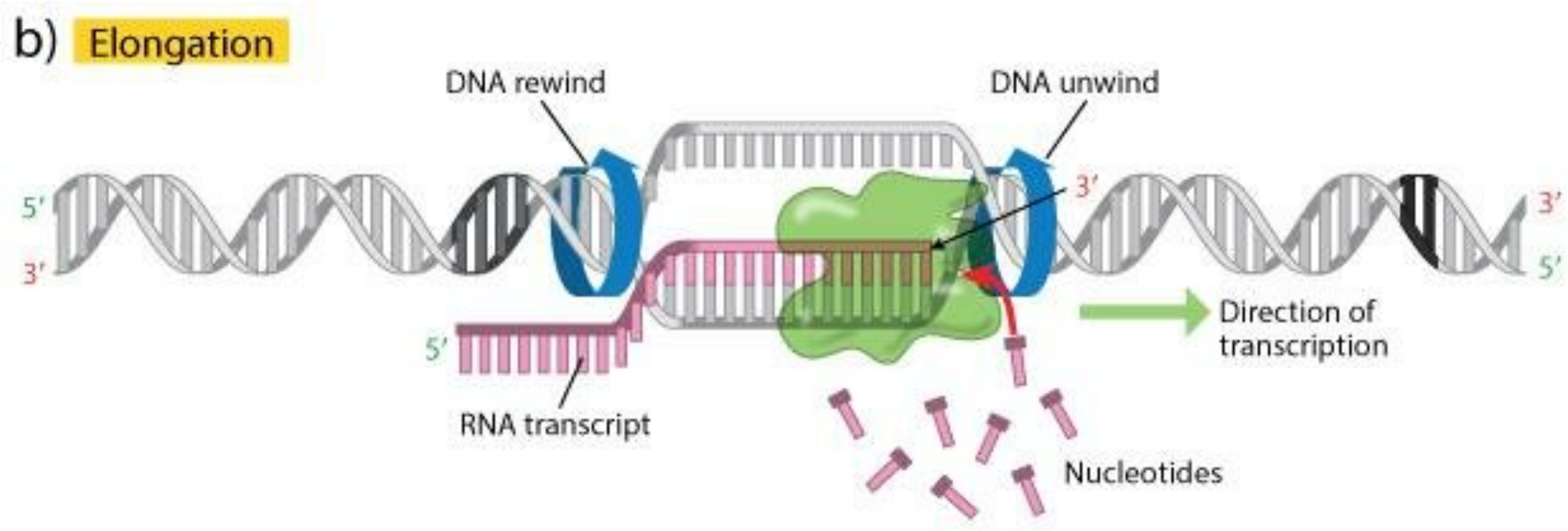
core promotor DNA

TFIID — TFIIA — TFIIB — Pol II/TFIIF — TFIIE — TFIIH
Sequential Assembly Pathway

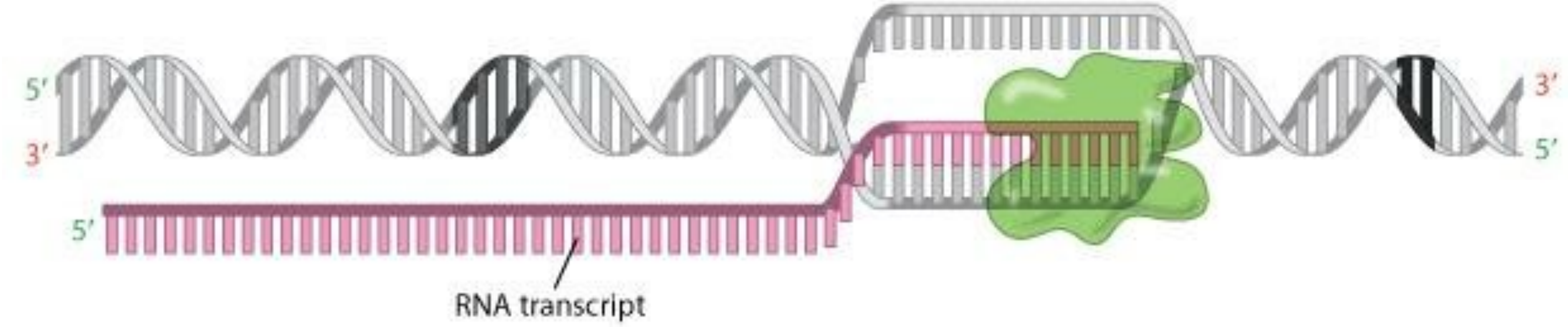
Main steps in transcription



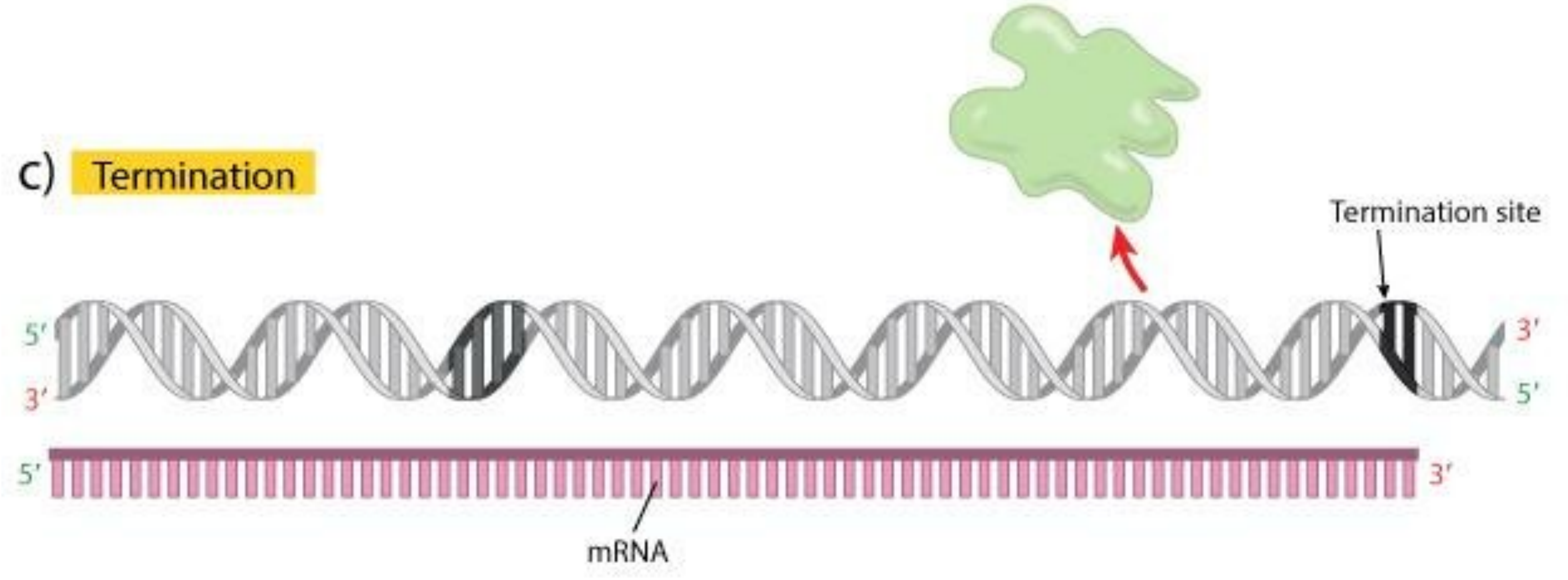
Main steps in transcription



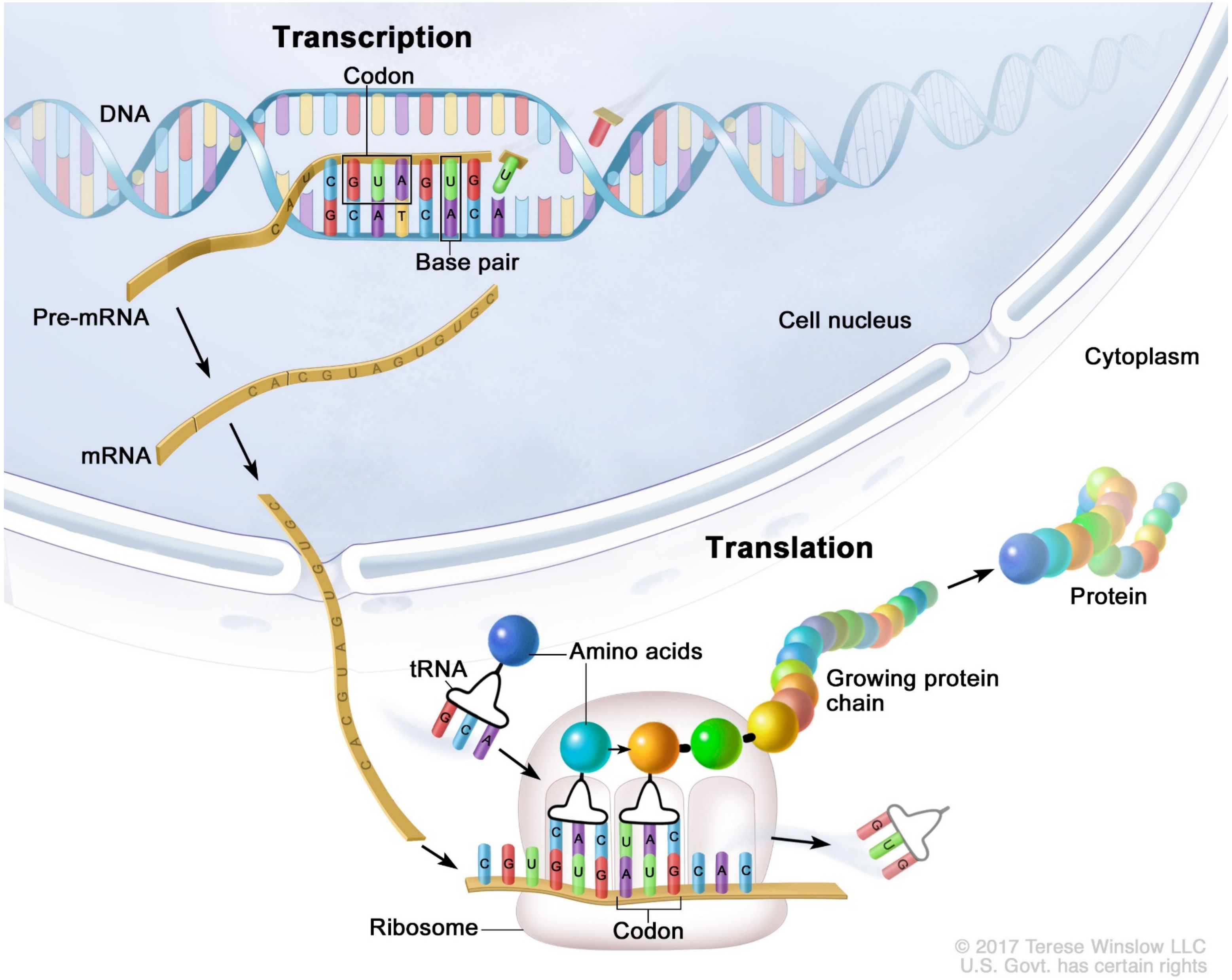
Main steps in transcription



Main steps in transcription



Plan



I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

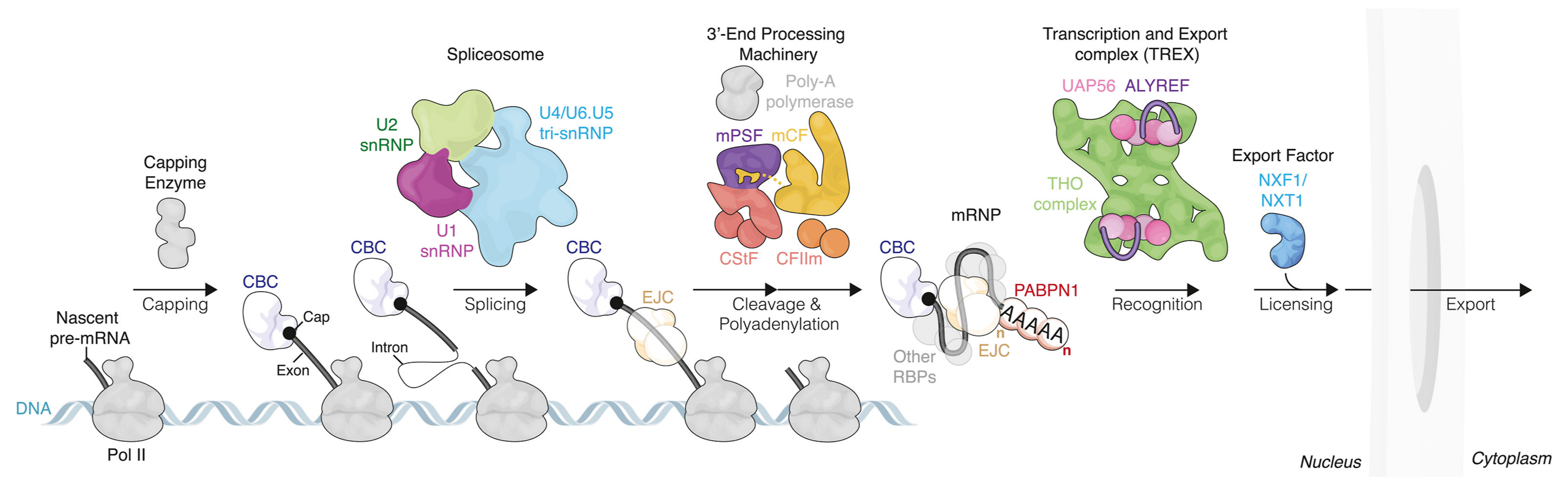
- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

Maturation of mRNA



Current Opinion in Structural Biology

Maturation of mRNA

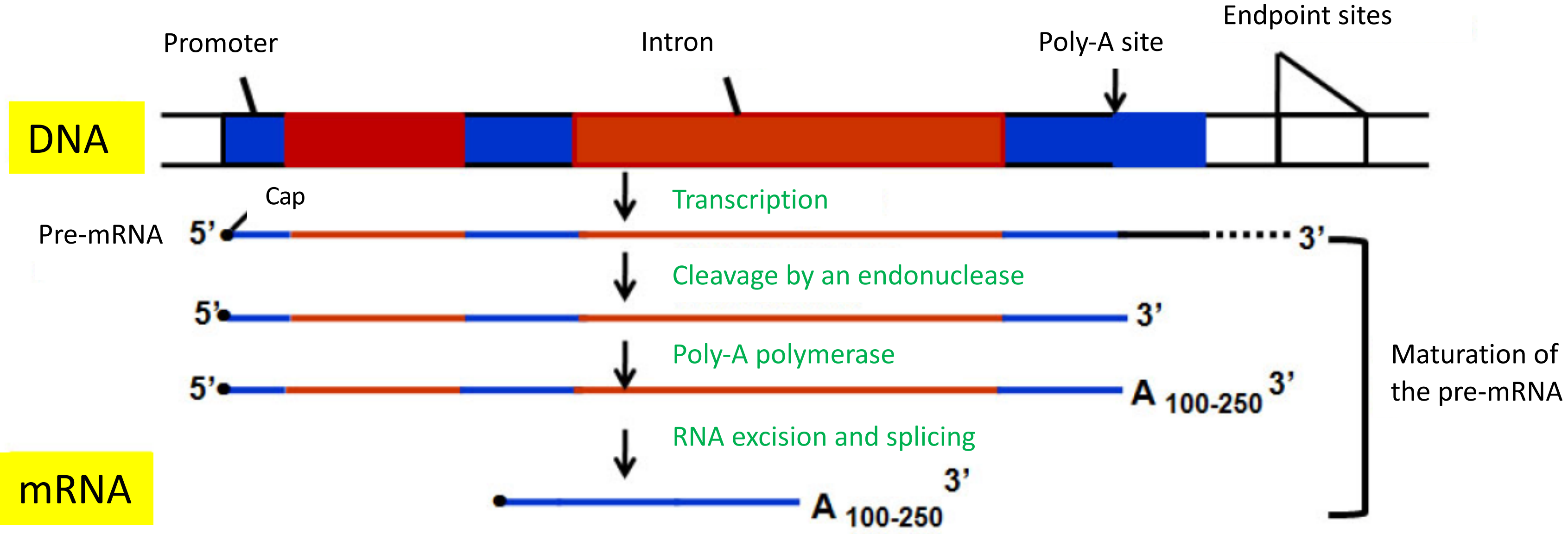
- RNA with length = length of transcribed DNA

  = primary transcript or pre-mRNA

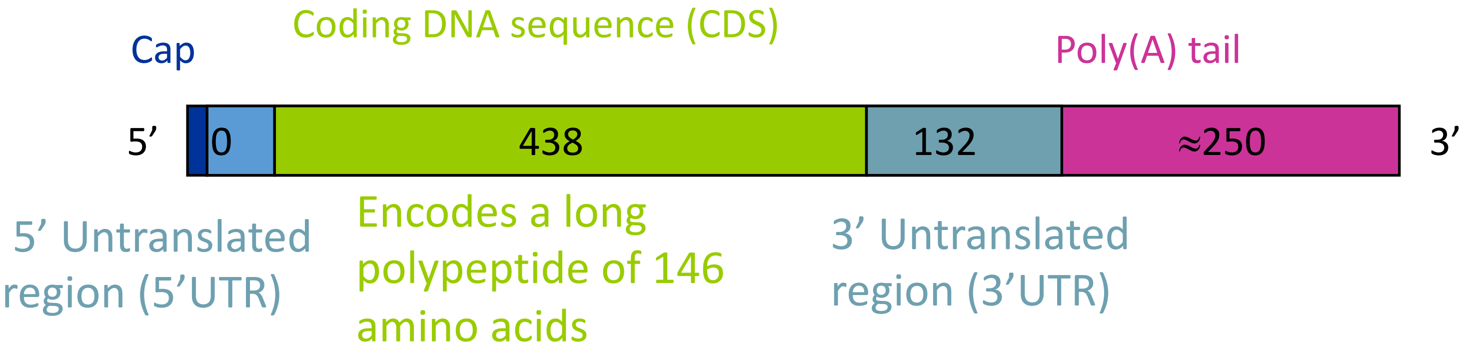
- DNA segment that gives the primary transcript

  = transcription unit

- Transcription unit



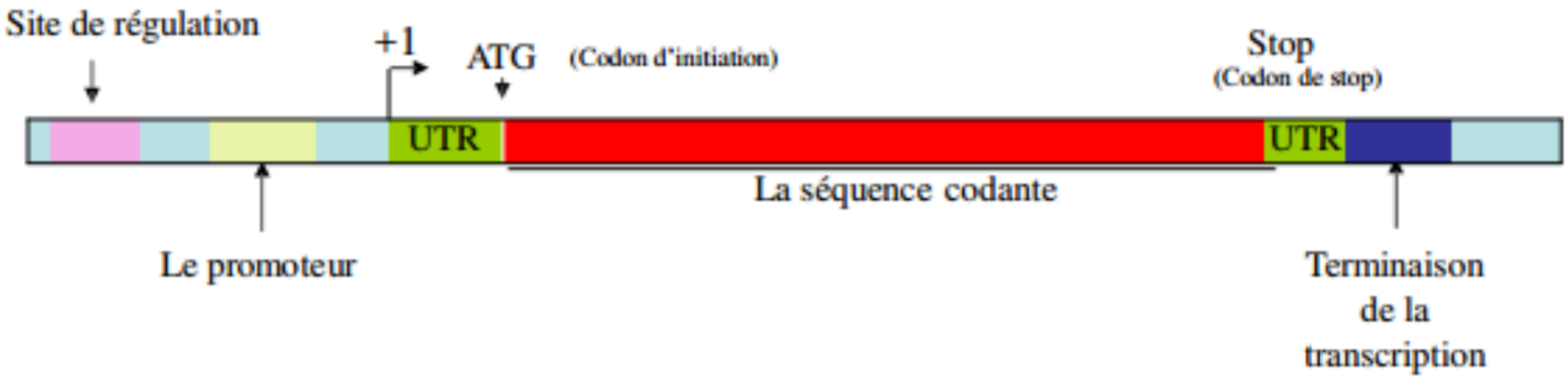
Structure of a eukaryotic mRNA



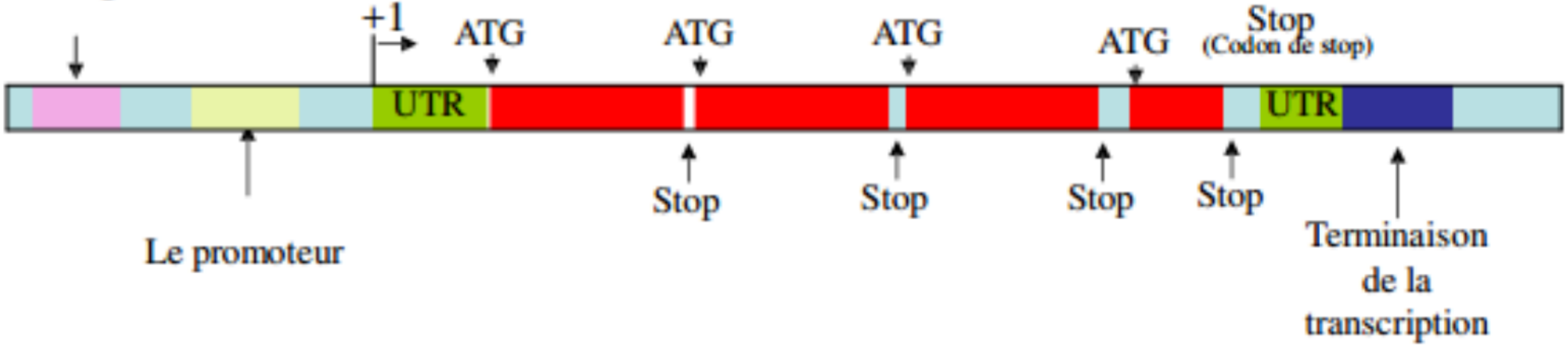
- ☞ Continuous sequence including CDS sequence encoding a specific polypeptide
- ☞ Non-coding regions at 5’ and 3’ ends = 5’UTR and 3’UTR
- ☞ Cap at 5’ end
- ☞ Presence of 50 - 250 adenosine residues at 3’ end = poly(A) tail
- ☞ Localized in the cytoplasm

Mono- vs Polycistronique RNA (Procaryotes)

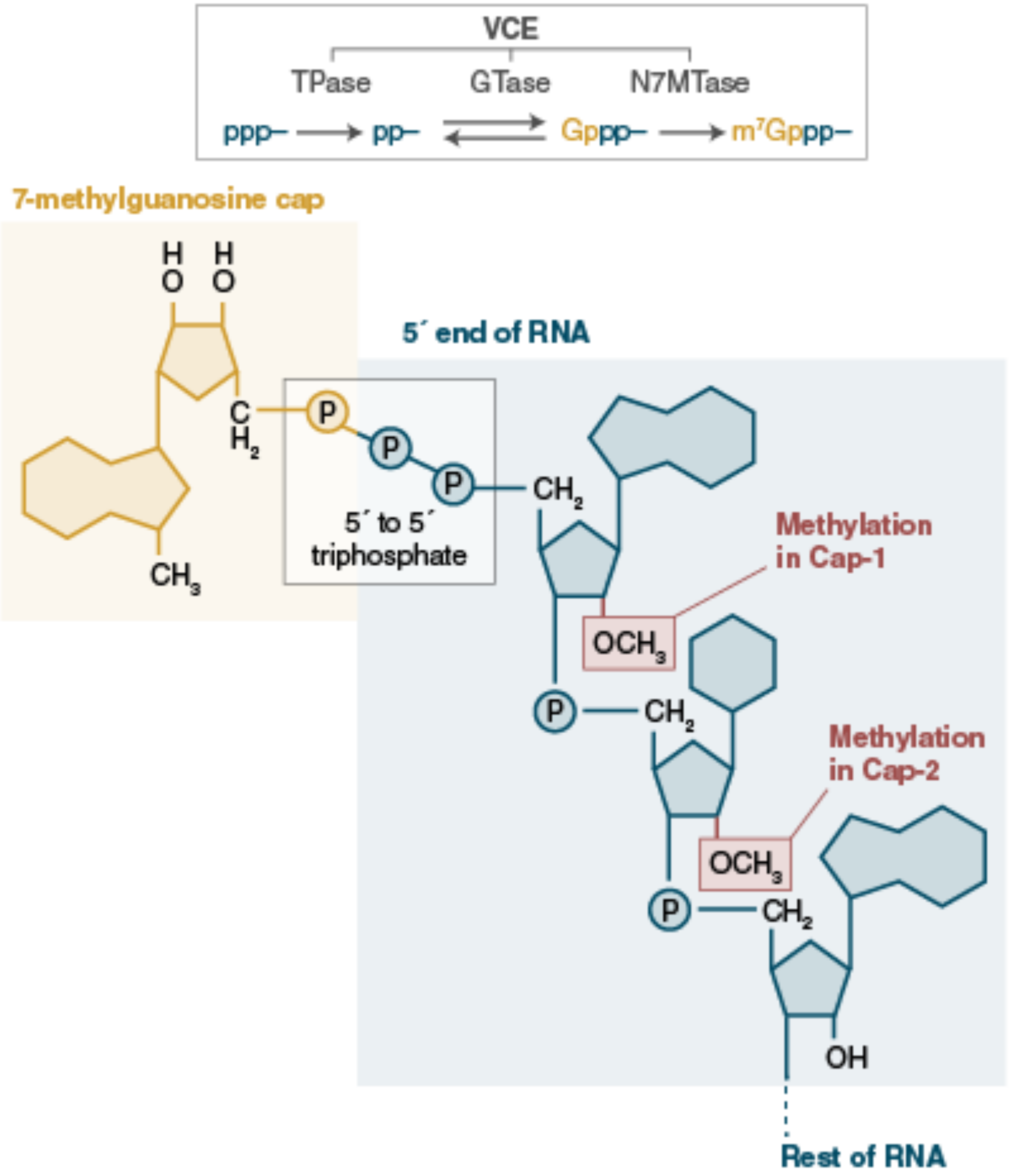
Monocistroniques:



Polycistroniques: (Opéron)



Maturation of the pre-mRNA: cap at 5’ end



☞ Roles of the cap

* It prevents the 5’ end from being digested by the nucleases

* It is used to identify mRNA

* It facilitates transport of mRNA

* It plays an important role in the initiation of translation

Maturation of the pre-mRNA: Poly(A) tail at 3’ end

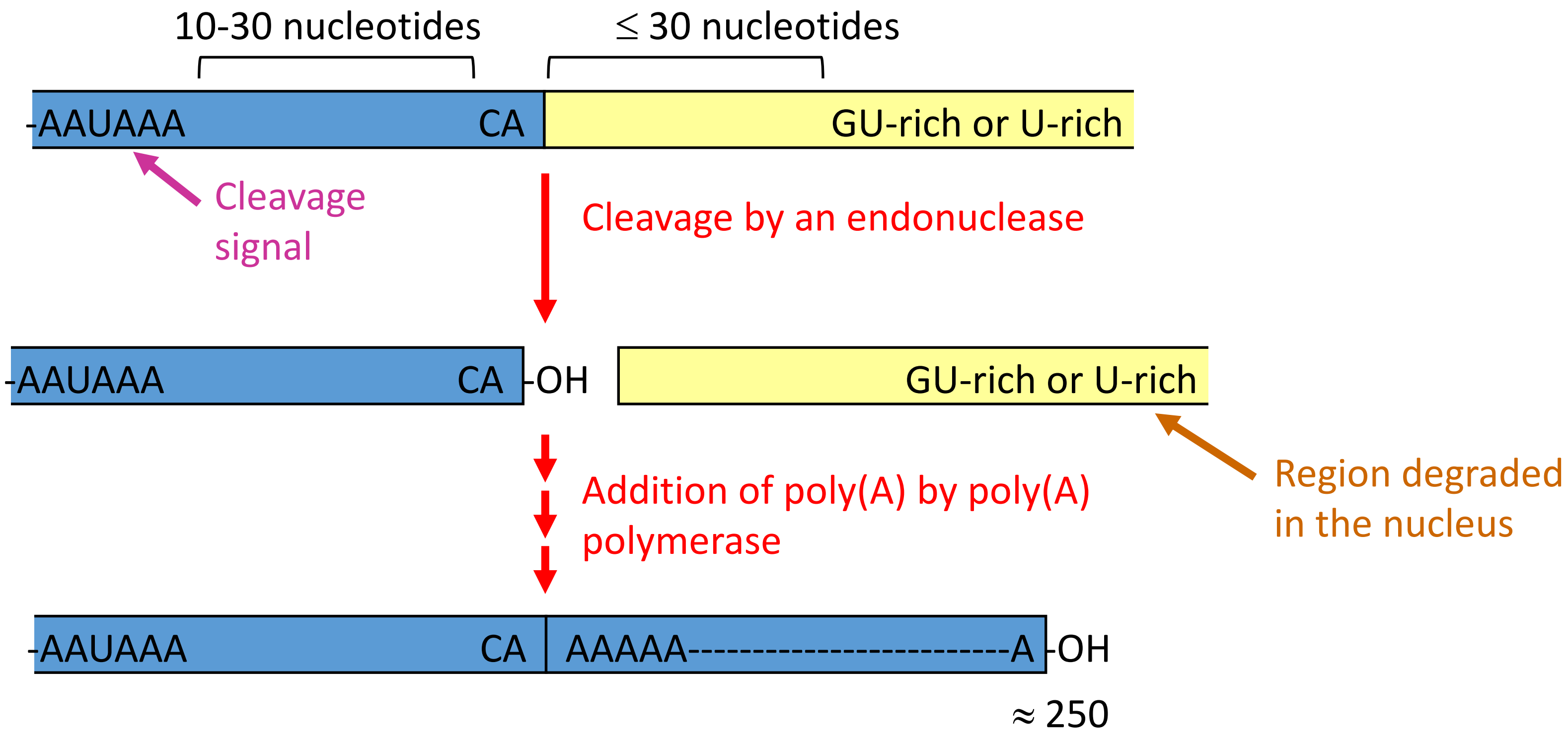
☞ The 3’ end is specified

by consensus sequences present in the transcription unit

☞ These signals are recognized by

RNA binding proteins

and RNA maturation proteins



Maturation of the pre-mRNA: Splicing

☞ The pre-mRNA contains exons and introns

Introns = sequences present in DNA, absent from mRNA

Exons = unique sequences present in mRNAs

↳ excision of the introns of the primary transcript

☞ Coding RNA sequences located on either side of an intron must be linked to each other

= RNA splicing

☞ To splice an RNA, cuts occur at the 5’ and 3’ edges of each intron

= splicing sites

Maturation of the pre-mRNA: Splicing

- ☞ Splicing is done by small nuclear RNA molecules = the ribozymes

  = snRNA

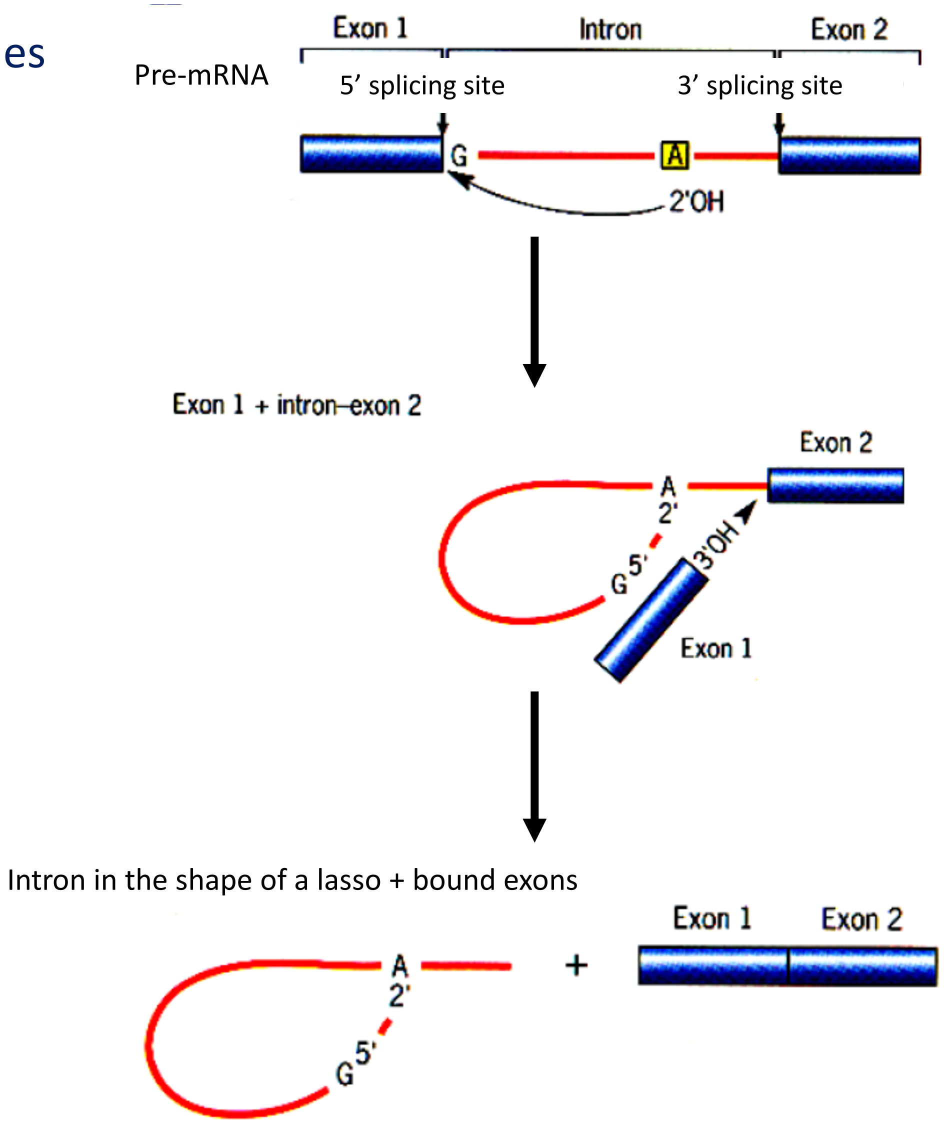
- ☞ snRNA « ribozymes »

  U1, U2, U4, U5, and U6

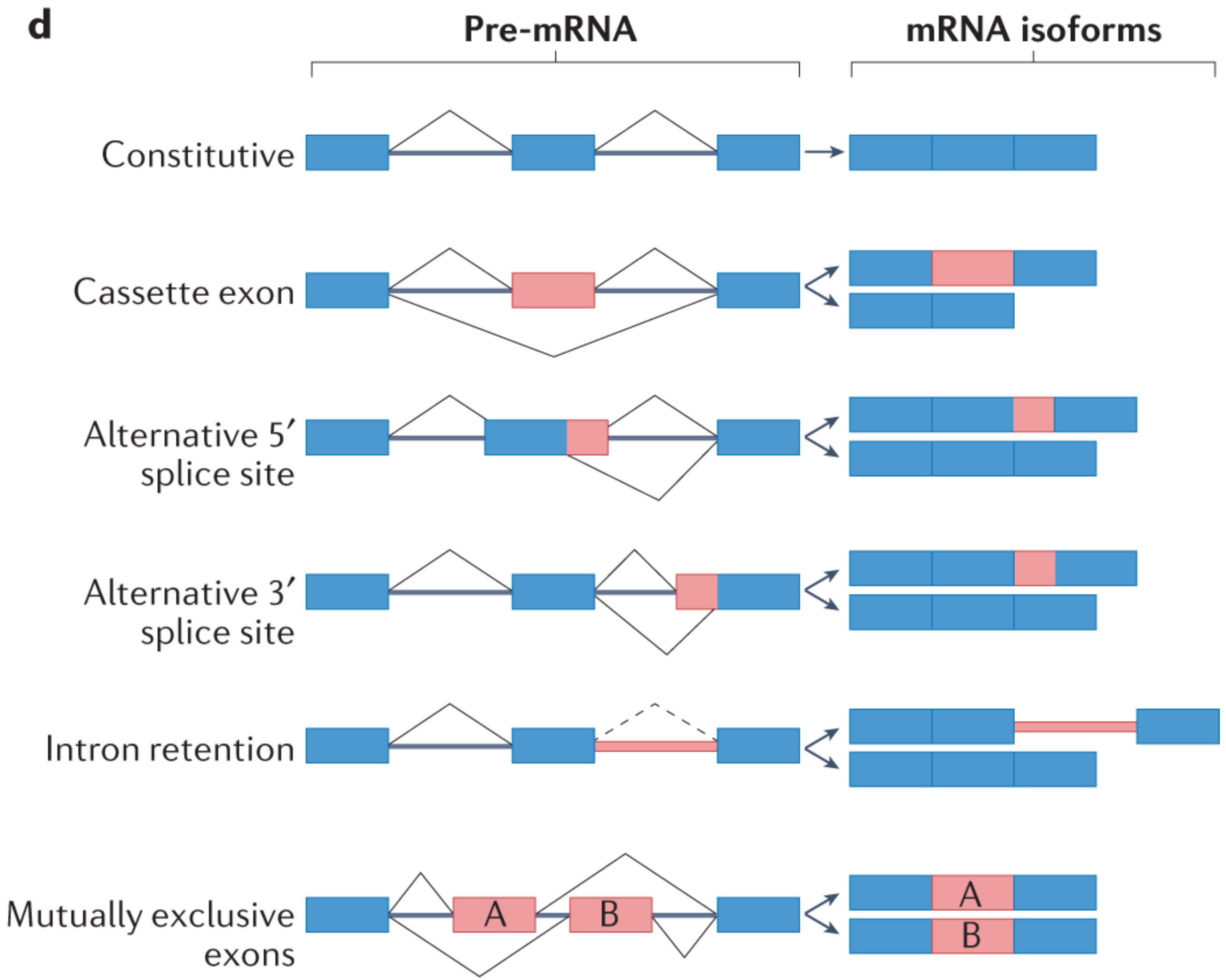
  + macromolecular complex

  = Spliceosome

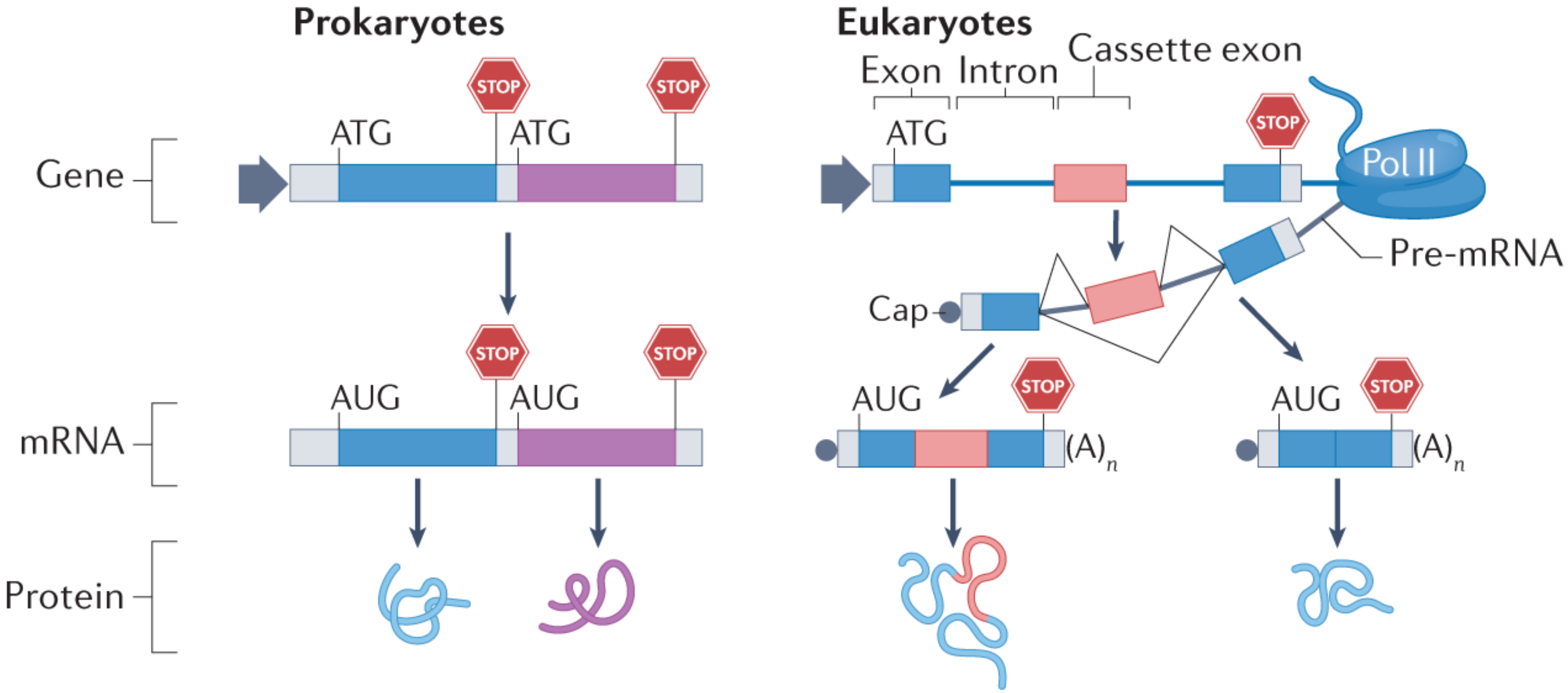
- ☞ Three steps are needed to remove introns from pre-mRNA molecules



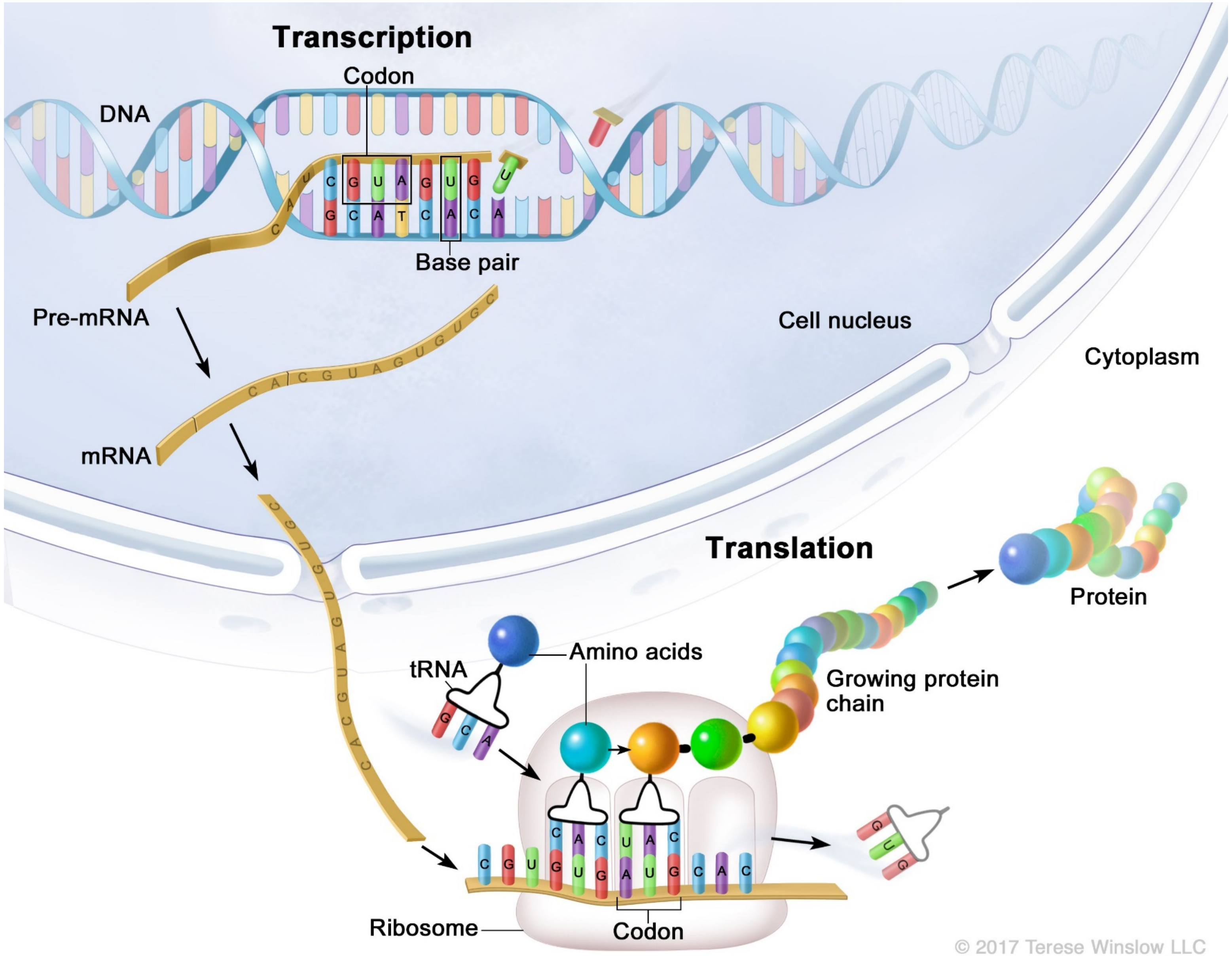
Alternative splicing



Prokaryotes vs Eukaryotes



Plan



I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

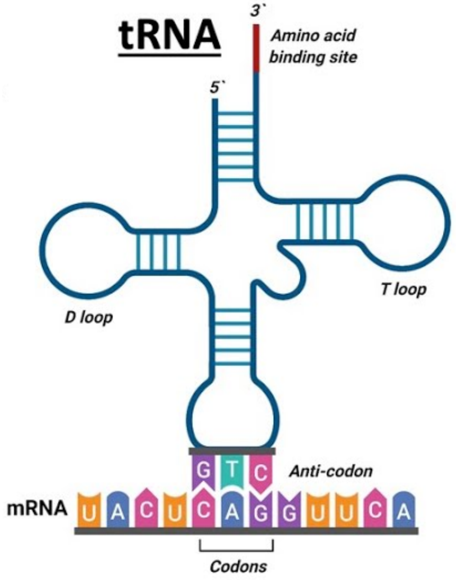
- Basic mechanism
- Maturation of mRNA

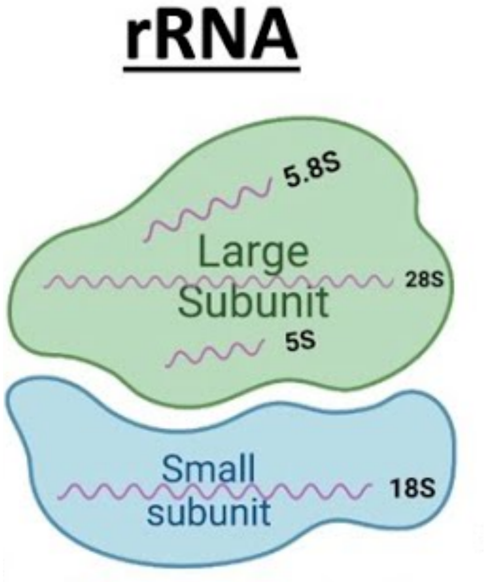
III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

Different RNA are involved…

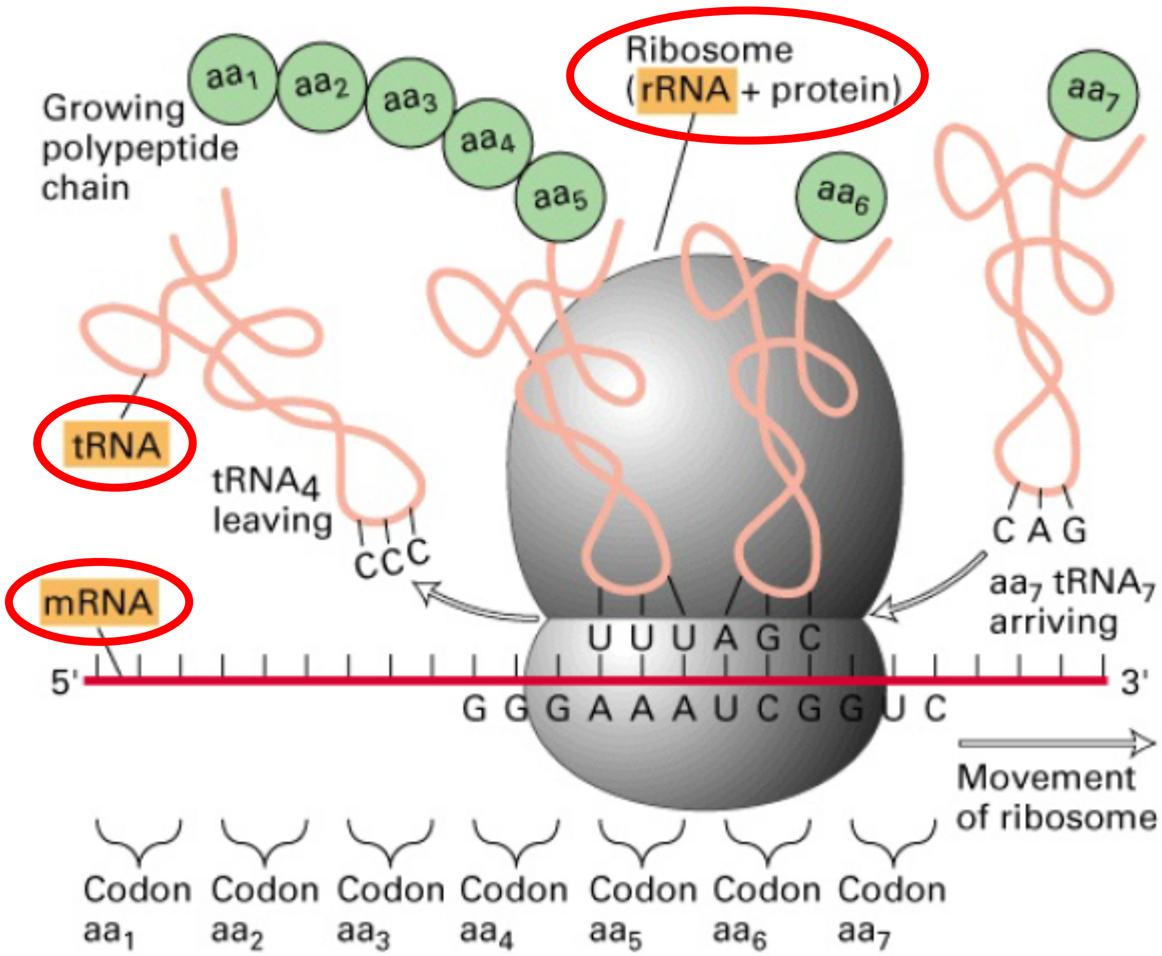




Ribosomal subunits

- ☞ Translation needs ribosomes
  - ↳ composed of ribosomal RNA (rRNA) and proteins
- ☞ And transfer RNA (tRNA)
  - ↳ reading of the genetic code

Different RNA are involved…



Transmission of genetic information

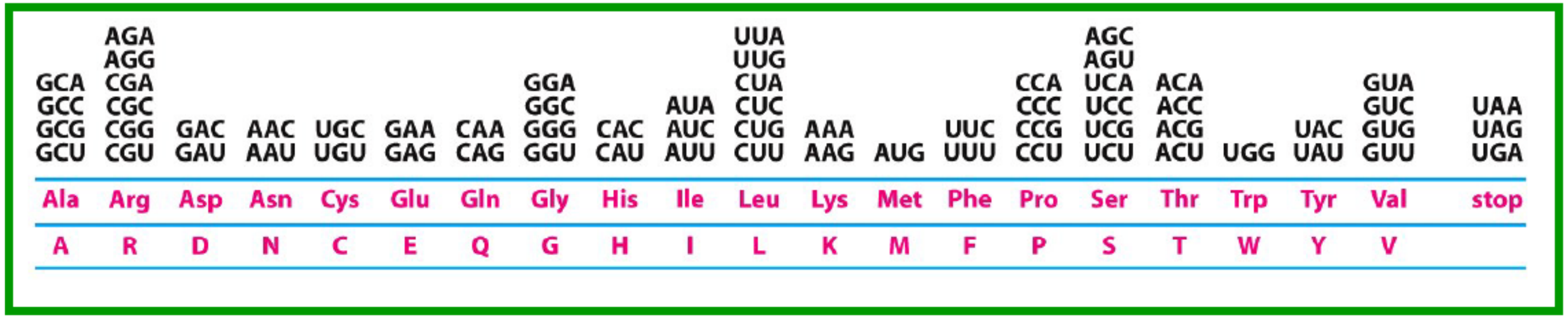
- Translation is done according to specific rules

  = genetic code

- The mRNA nucleotide sequence is read in groups of 3 nucleotides

  each nucleotide triplet = codon

  1 codon ⇨ 1 amino acid



GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

- Most amino acids are determined by several codons = synonymous codons

  ⇘ the genetic code is degenerate

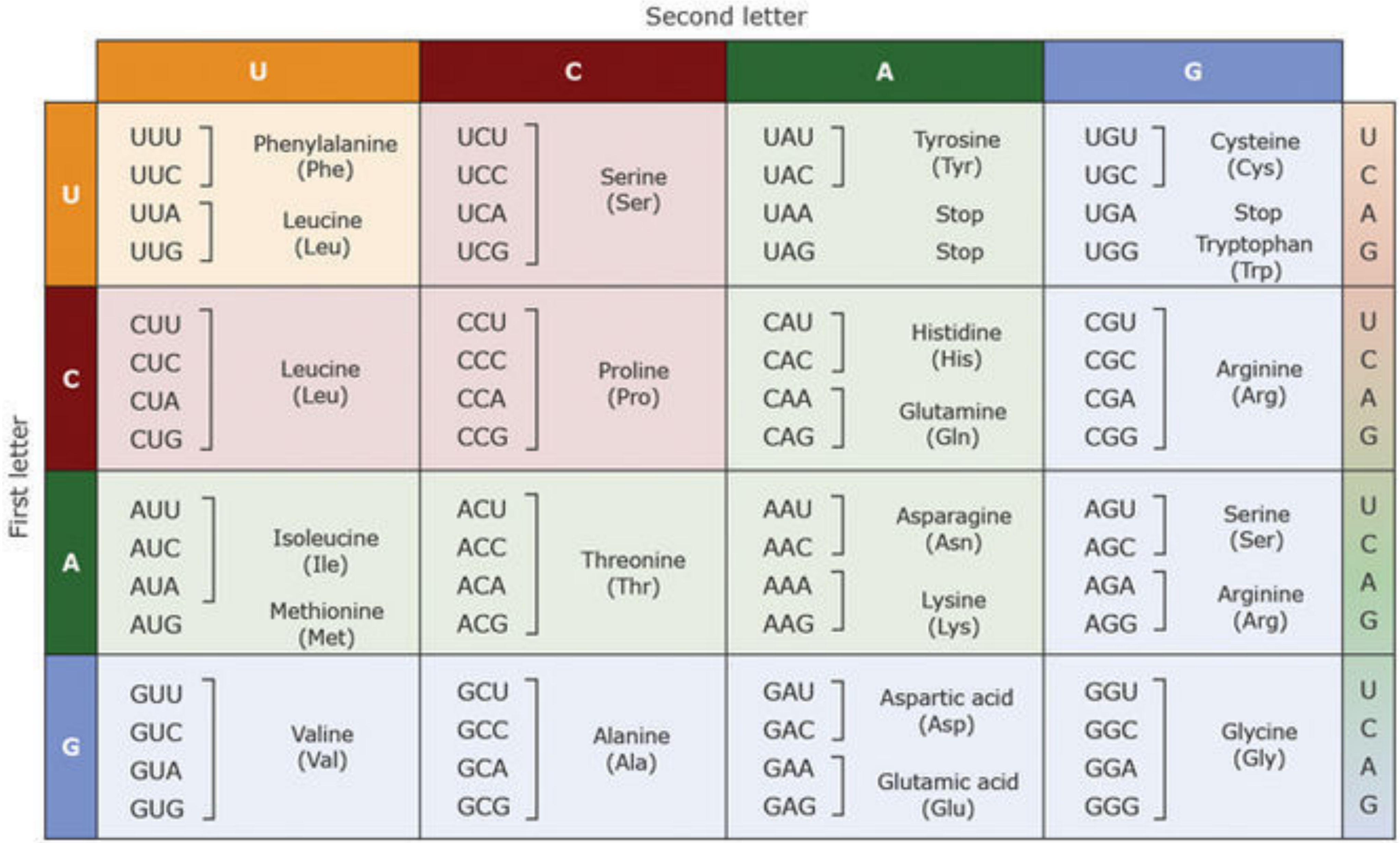
- Codons that specify the same amino acid have common characteristics

- 3 codons do not specify any amino acid = stop codons

- Codon AUG = initiation codon

  = codon specifying methionine (Met)

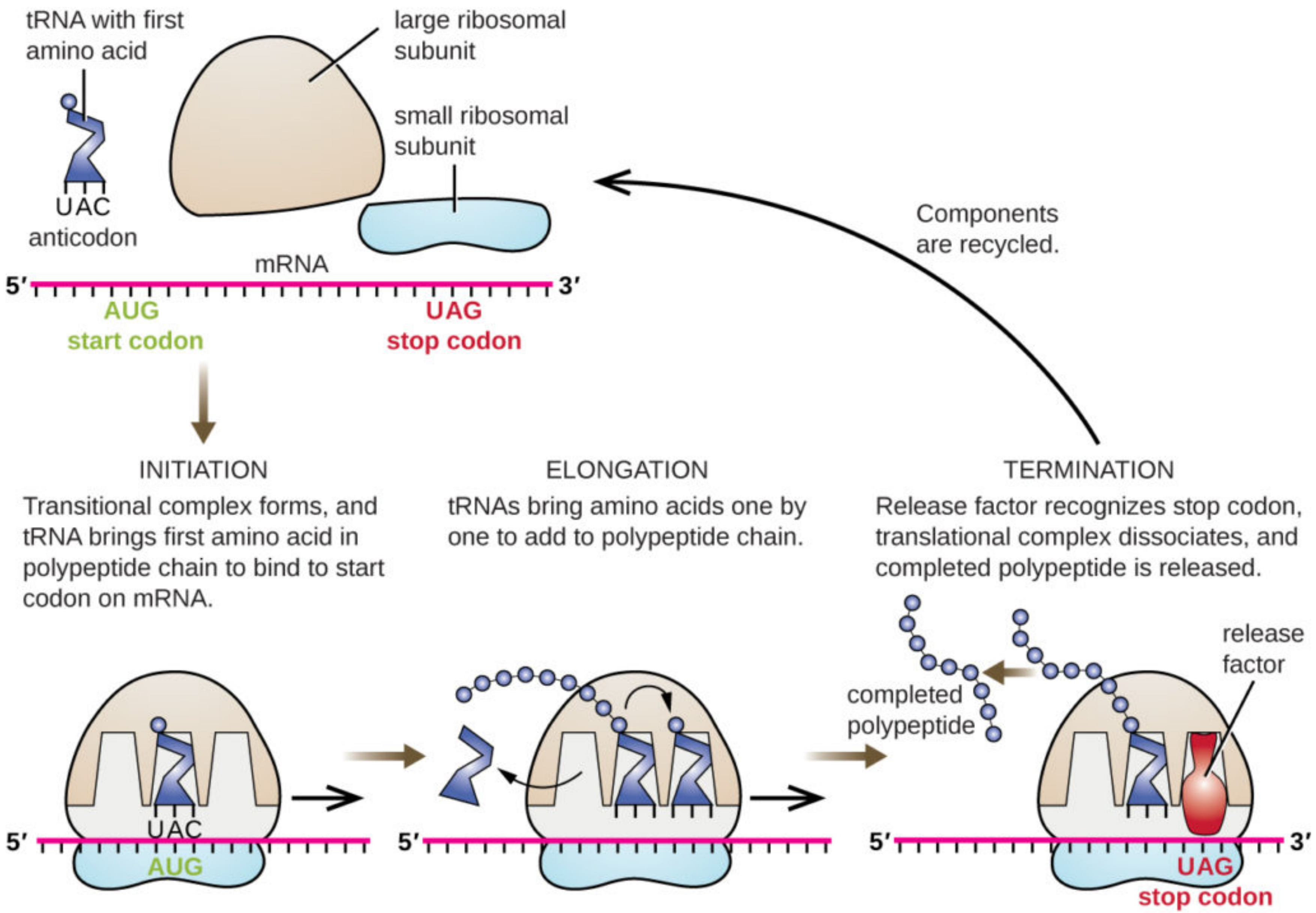
Genetic code



Second letter

First letter	U		C		A		G		
U	UUU	Phenylalanine (Phe)	UCU	Serine (Ser)	UAU	Tyrosine (Tyr)	UGU	Cysteine (Cys)	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine (Leu)	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Tryptophan (Trp)	G
C	CUU	Leucine (Leu)	CCU	Proline (Pro)	CAU	Histidine (His)	CGU	Arginine (Arg)	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine (Gln)	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine (Ile)	ACU	Threonine (Thr)	AAU	Asparagine (Asn)	AGU	Serine (Ser)	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine (Lys)	AGA	Arginine (Arg)	A
	AUG	Methionine (Met)	ACG		AAG		AGG		G
G	GUU	Valine (Val)	GCU	Alanine (Ala)	GAU	Aspartic acid (Asp)	GGU	Glycine (Gly)	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid (Glu)	GGA		A
	GUG		GCG		GAG		GGG		G

Translation



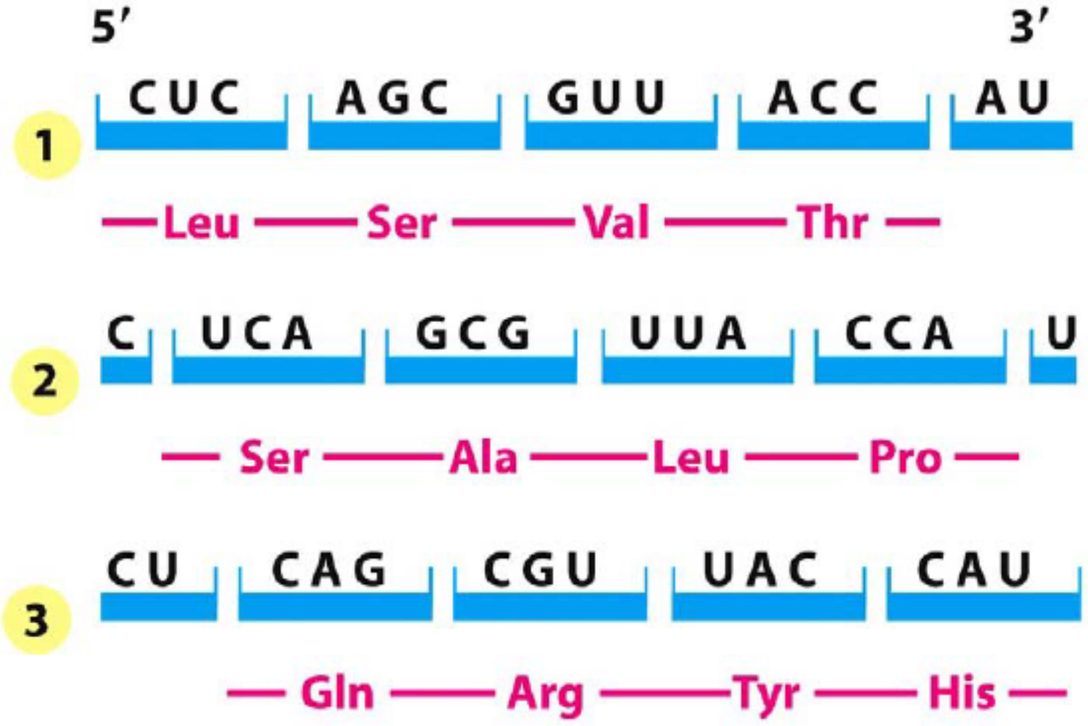
Reading frame

☞ The nucleotide sequence is read

from the 5’ end to the 3’ end

by group of 3 nucleotides

☞ The RNA sequence can be translated into 3 open reading phases

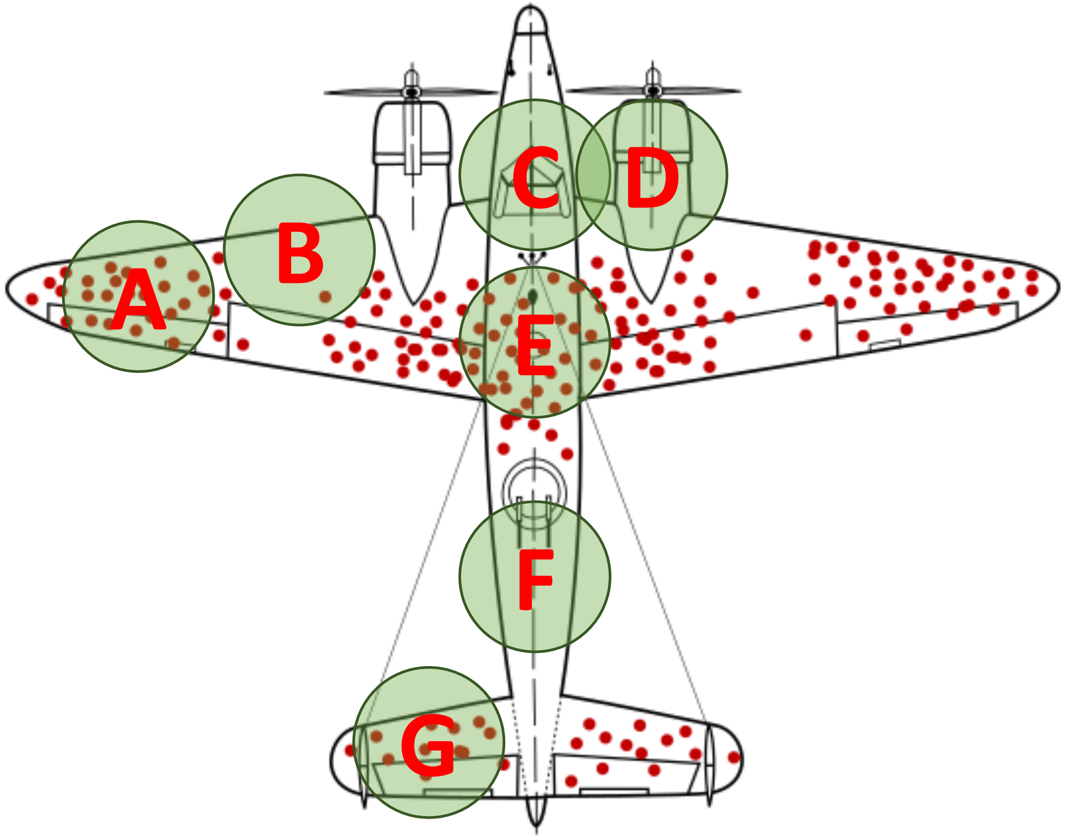


The effective reading frame (1 of 3) is set when the mRNA translation is initiated

☞ Frame-shift

= shift of the reading frame

Pause…

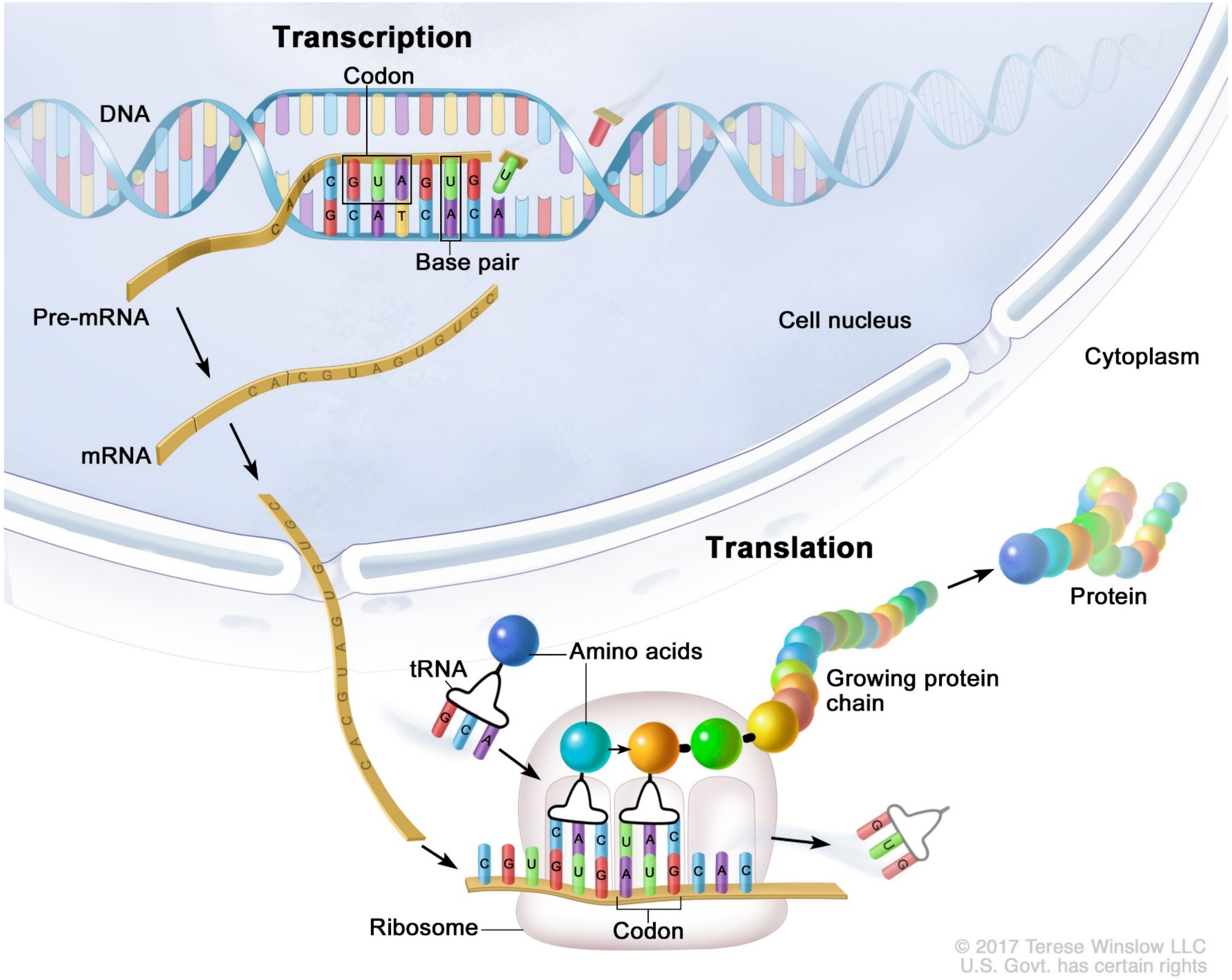


This diagram shows where returning WW2-era planes were hit.

Which part(s) should be reinforced ?

A hunter sees 6 birds in a tree and shot 1. How many birds left?

Plan



I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

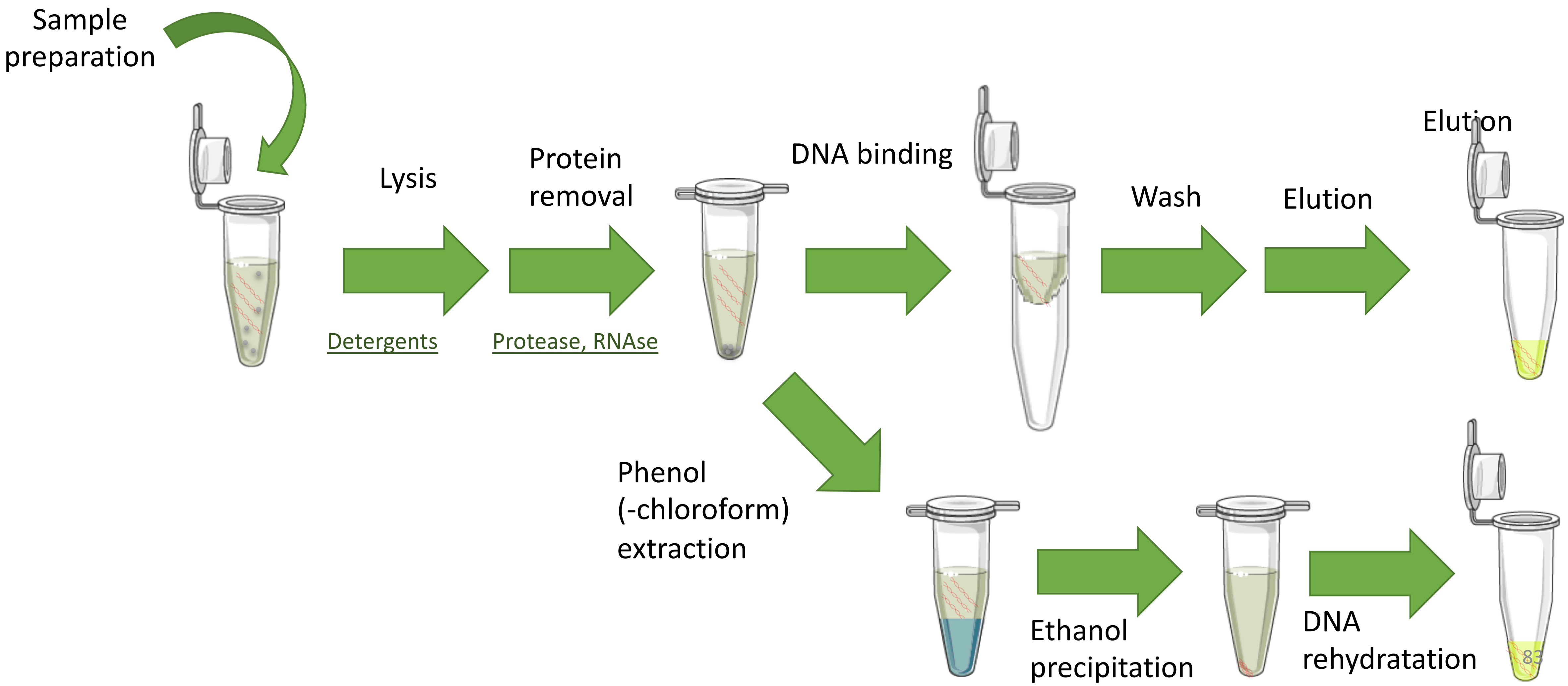
- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

DNA purification



Sample preparation

Lysis

Detergents

Protein removal

Protease, RNAse

DNA binding

Wash

Elution

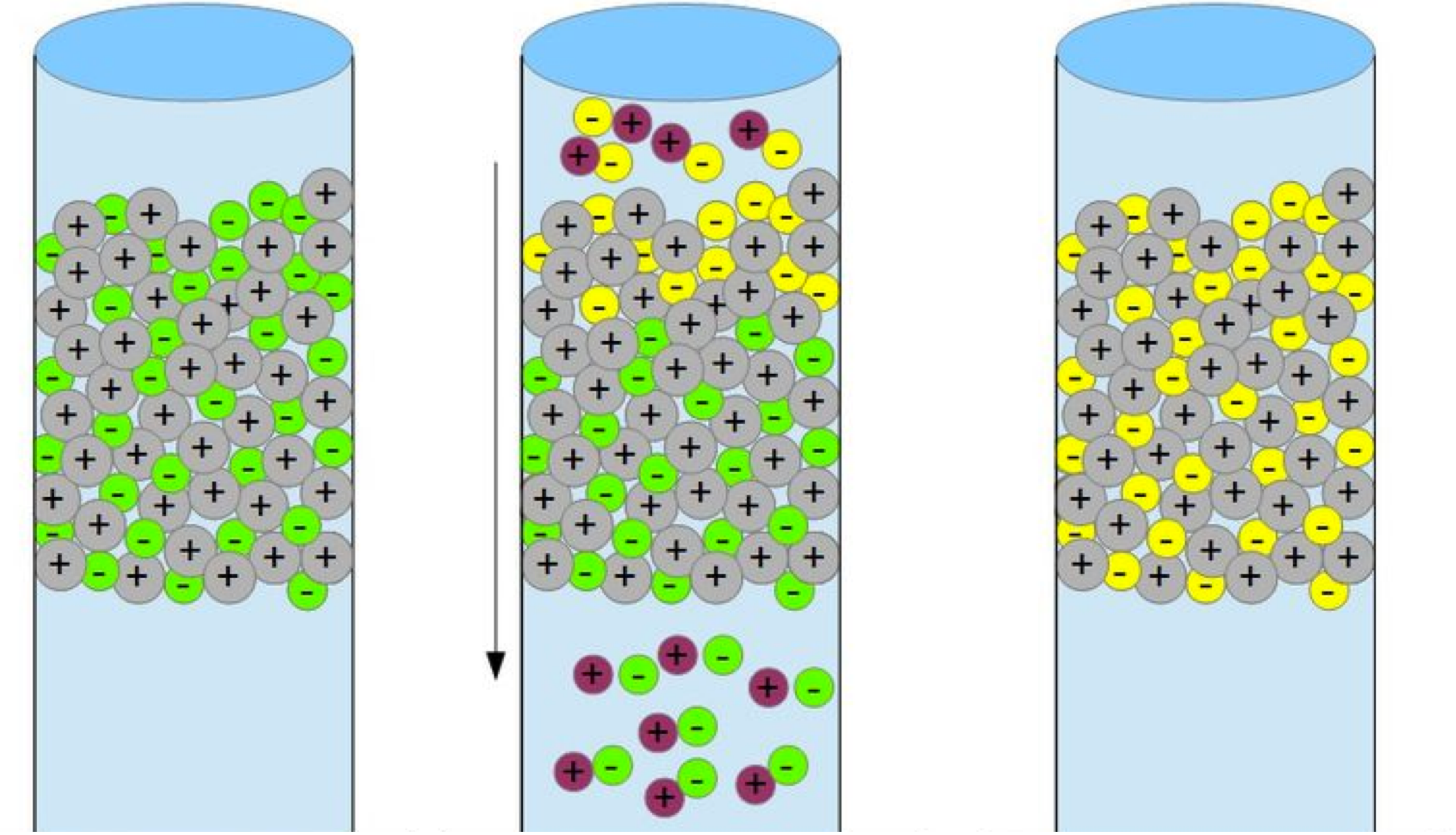
Elution

Phenol (-chloroform) extraction

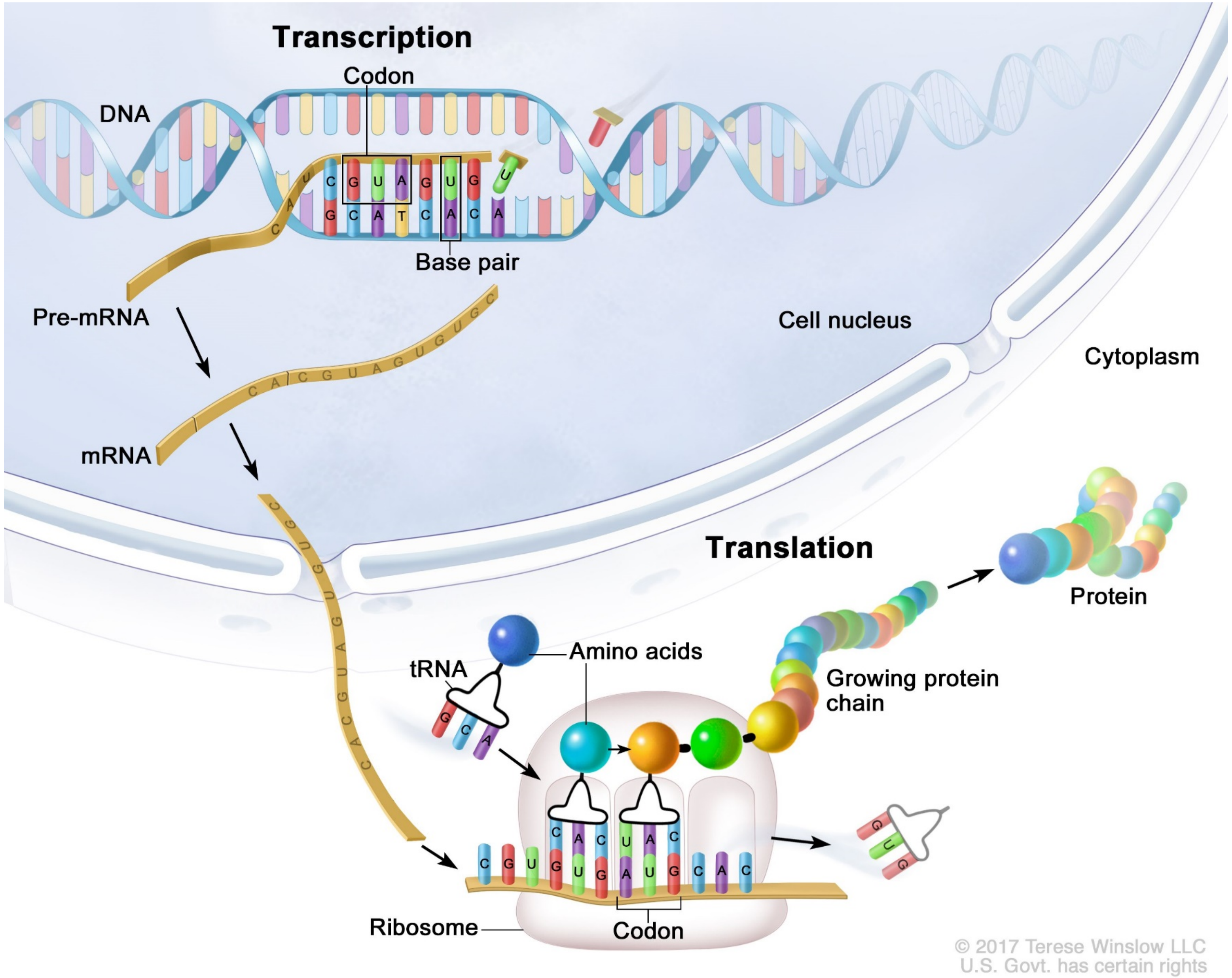
Ethanol precipitation

DNA rehydratation

DNA binding on resin



Plan



I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

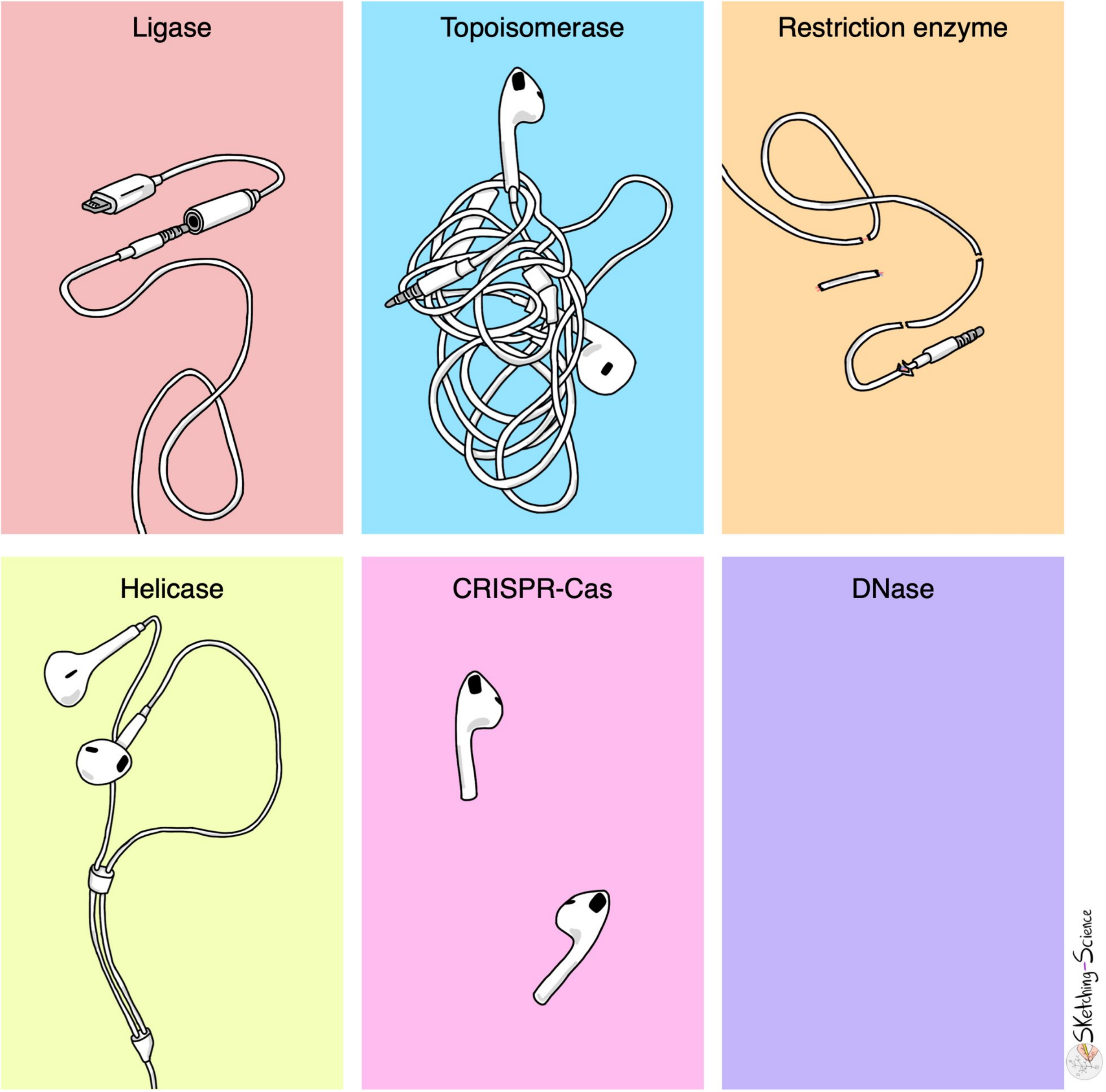
- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

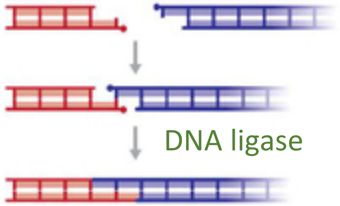
- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

Enzymatic tools for the study of nucleic acids

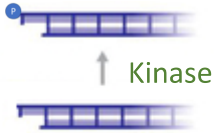


Enzymatic tools for the study of nucleic acids

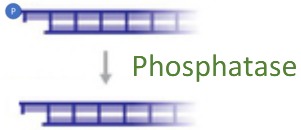
- ☞ Nucleases
  - to digest
- ☞ Ligases
  - to ligate



- ☞ Polymerases
  - to polymerize
- ☞ Kinases
  - to phosphorylate



- ☞ Phosphatases
  - to dephosphorylate



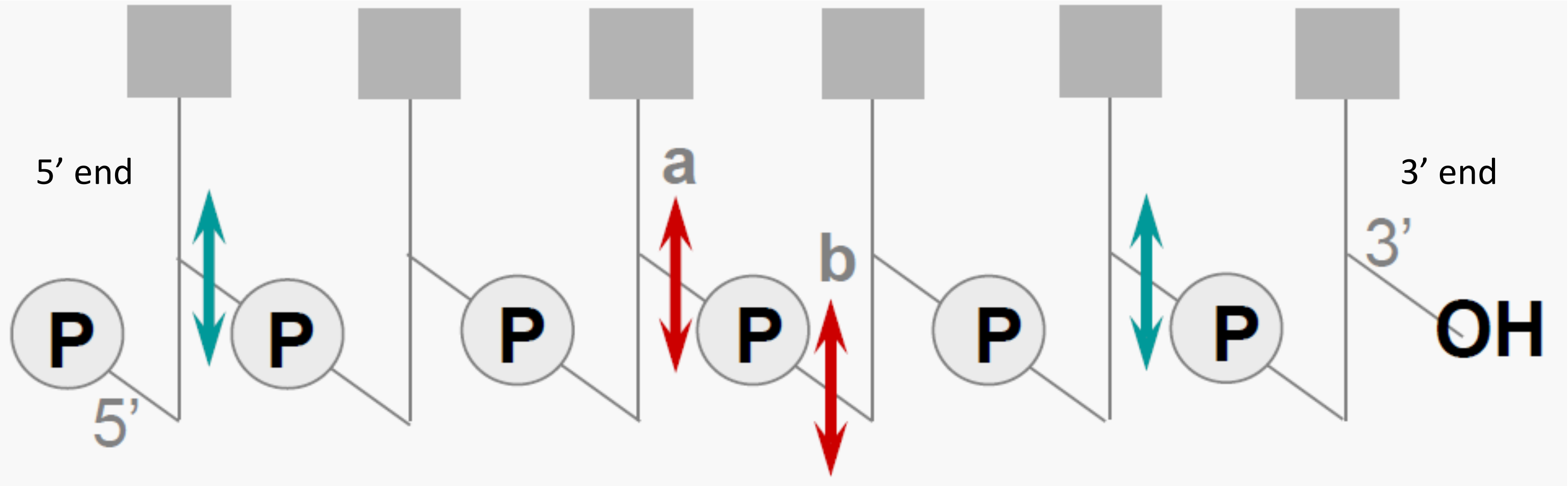
Nucleases

- ☞ Exonucleases

  = they release the nucleotide at the 5’ or 3’ end

- ☞ Endonucleases

  = they hydrolyse an internal phosphodiester bond



Exonuclease Endonuclease Exonuclease

Restriction enzymes

☞ These are endonucleases

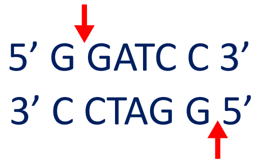
They recognize a specific base sequence on double-stranded DNA

They cut the 2 strands of the duplex

☞ Names with 3 or 4 letters

↳ origin of the microorganism

Microorganism	Abbreviation	Sequence
Bacillus amyloliquefaciens H	*Bam*HI	5’ GGATCC 3’ 3’ CCTAGG 5’



5’ G GATC C 3’
3’ C CTAG G 5’

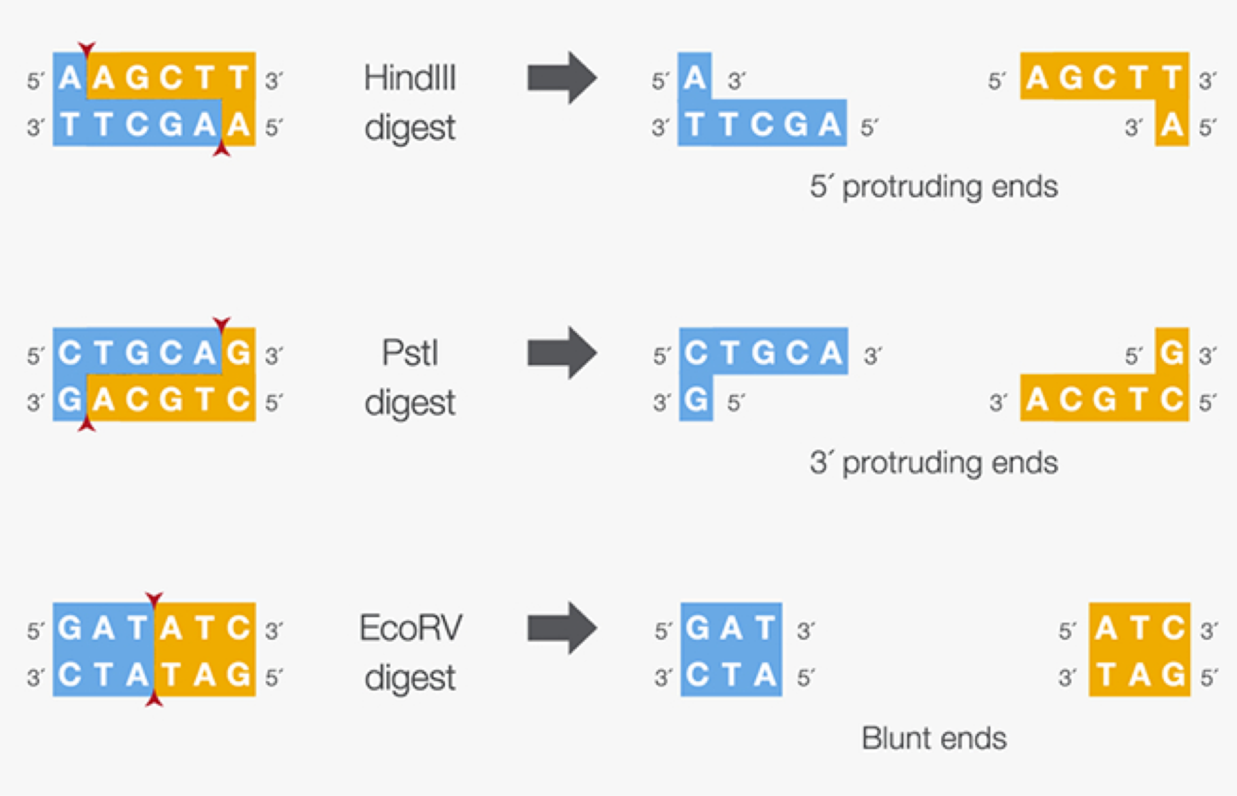
☞ « Restriction »

Refers to the function of these enzymes

↳ destroy / restrict foreign DNA

Restriction enzymes

- Many restriction enzymes recognize specific sequences
  - from 4 to 6 base pairs
- This sequence has a rotational axis of symmetry
  - called “palindrome”
- Restriction enzymes produce DNA fragments
  - with cohesive or straight ends



Restriction enzymes

- ☞ The products of genome digestion by restriction enzymes

  = restriction fragments

- ☞ A given restriction enzyme cuts a given DNA molecule

  always at the same sites = restriction sites

- ☞ The frequency of target sequences is variable

  Site of 4 bp ⇨ 1 for 4^4 ⇨ 256 bp

  Site of 6 bp ⇨ 1 for 4^6 ⇨ 4,096 bp

- ☞ The average size of DNA fragments produced by different restriction enzymes is variable

Restriction maps

☞ It is possible

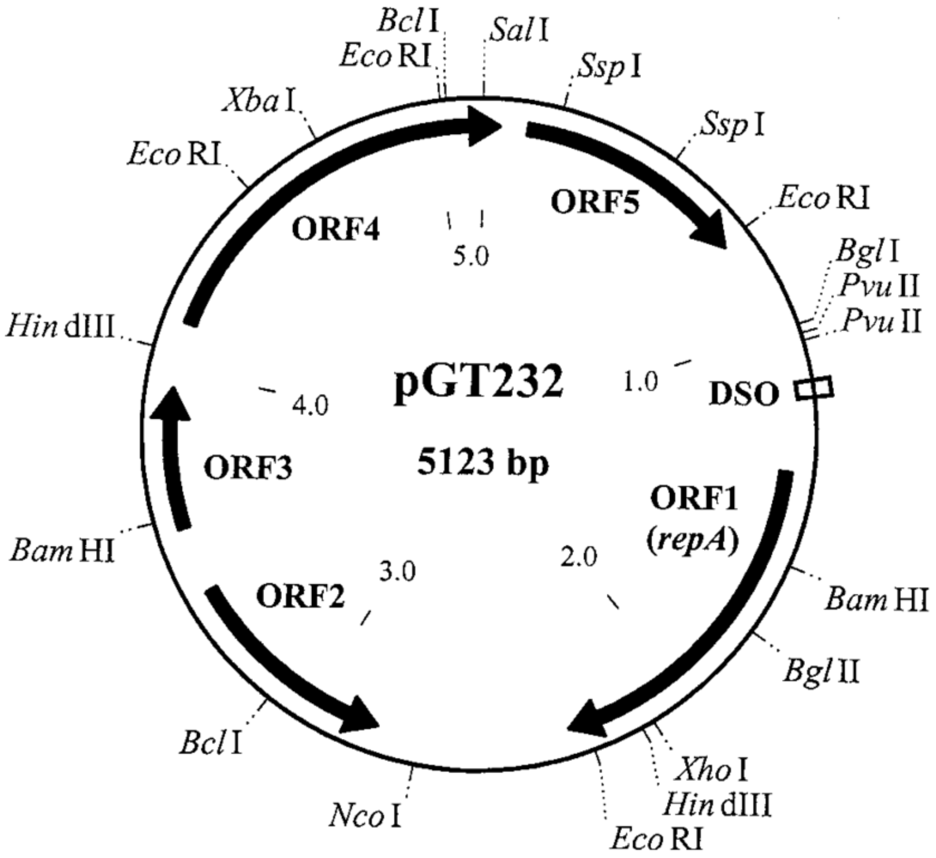
to order the relative position of restriction sites on a DNA molecule

↳ this is the restriction map of a particular genetic region

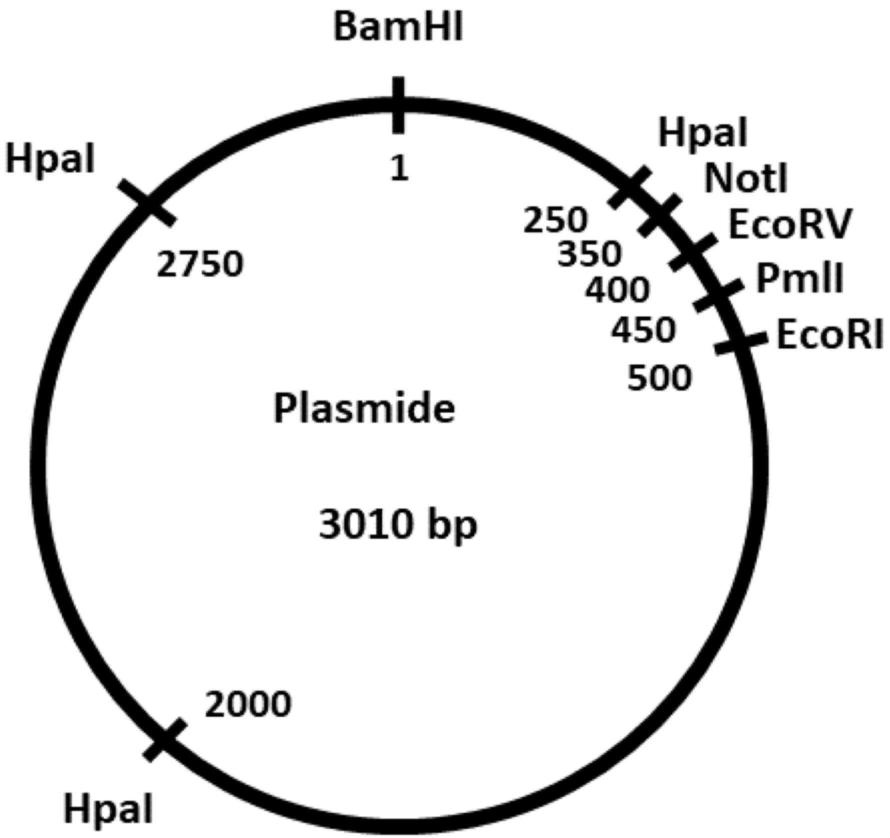
☞ Restriction maps permit the location of a gene and make it easier to clone it

Usefull for Functional analysis of genes

- Directed mutagenesis
- Gene knock-out
- Recombinant proteins
- Two-hybrid screening
- …



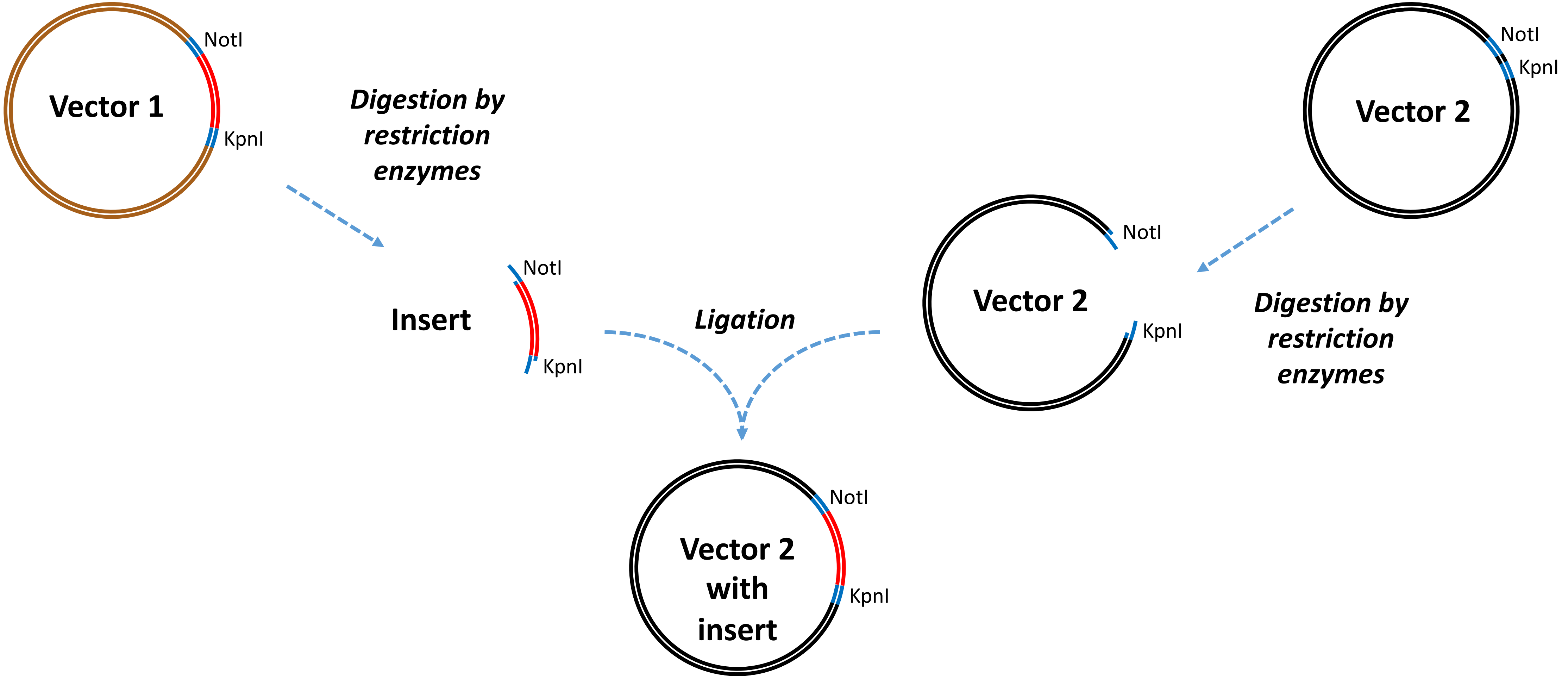
Your turn…



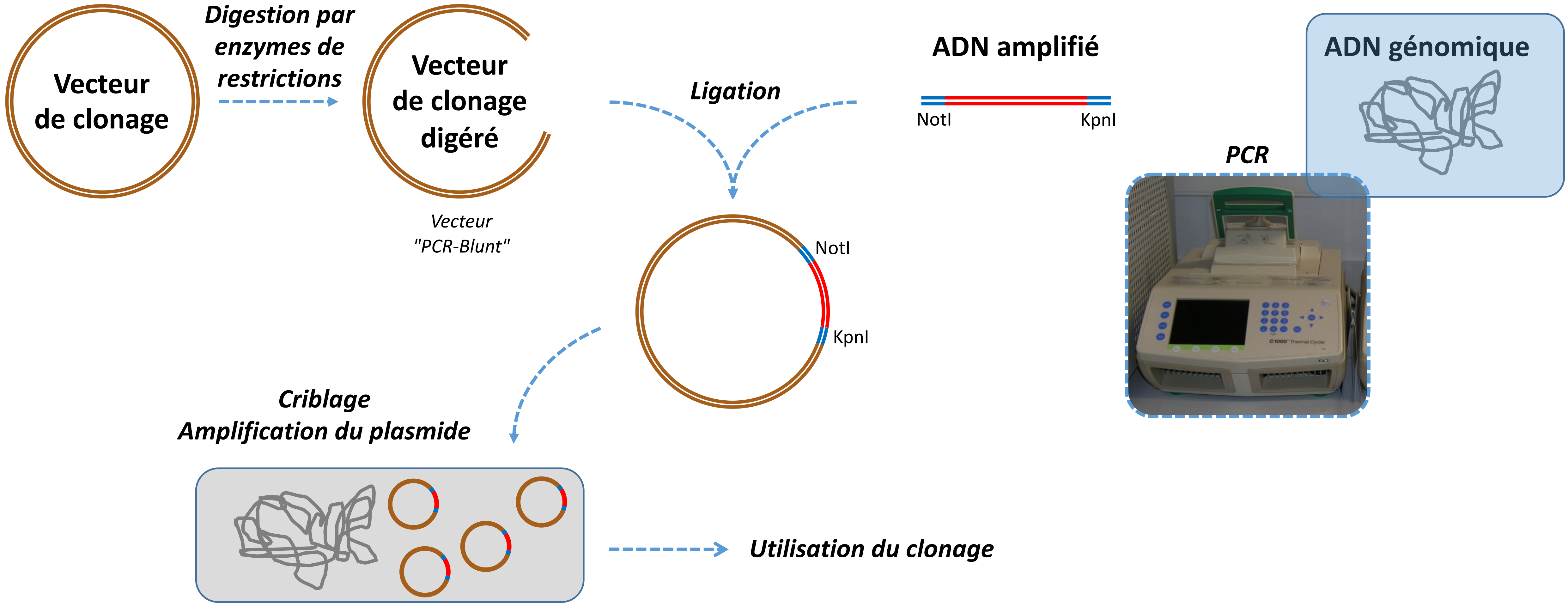
Size of the fragment(s) after digestion by :

A) EcoRI
B) EcoRV
C) BamHI + NotI
D) HpaI
E) HpaI + BamHI + EcoRI

Sous-clonage



Sous-clonage



Vecteur de clonage

Digestion par enzymes de restrictions

Vecteur de clonage digéré

Vecteur "PCR-Blunt"

Ligation

ADN amplifié

NotI

KpnI

ADN génomique

PCR

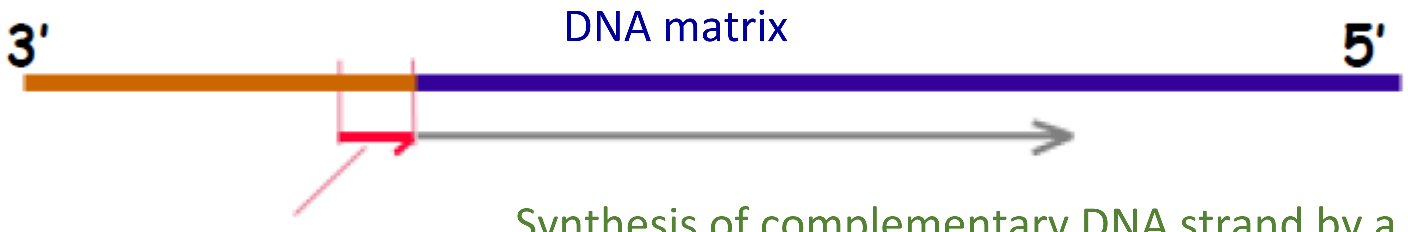
NotI

KpnI

Criblage
Amplification du plasmide

Utilisation du clonage

DNA polymerases



3'

DNA matrix

5'

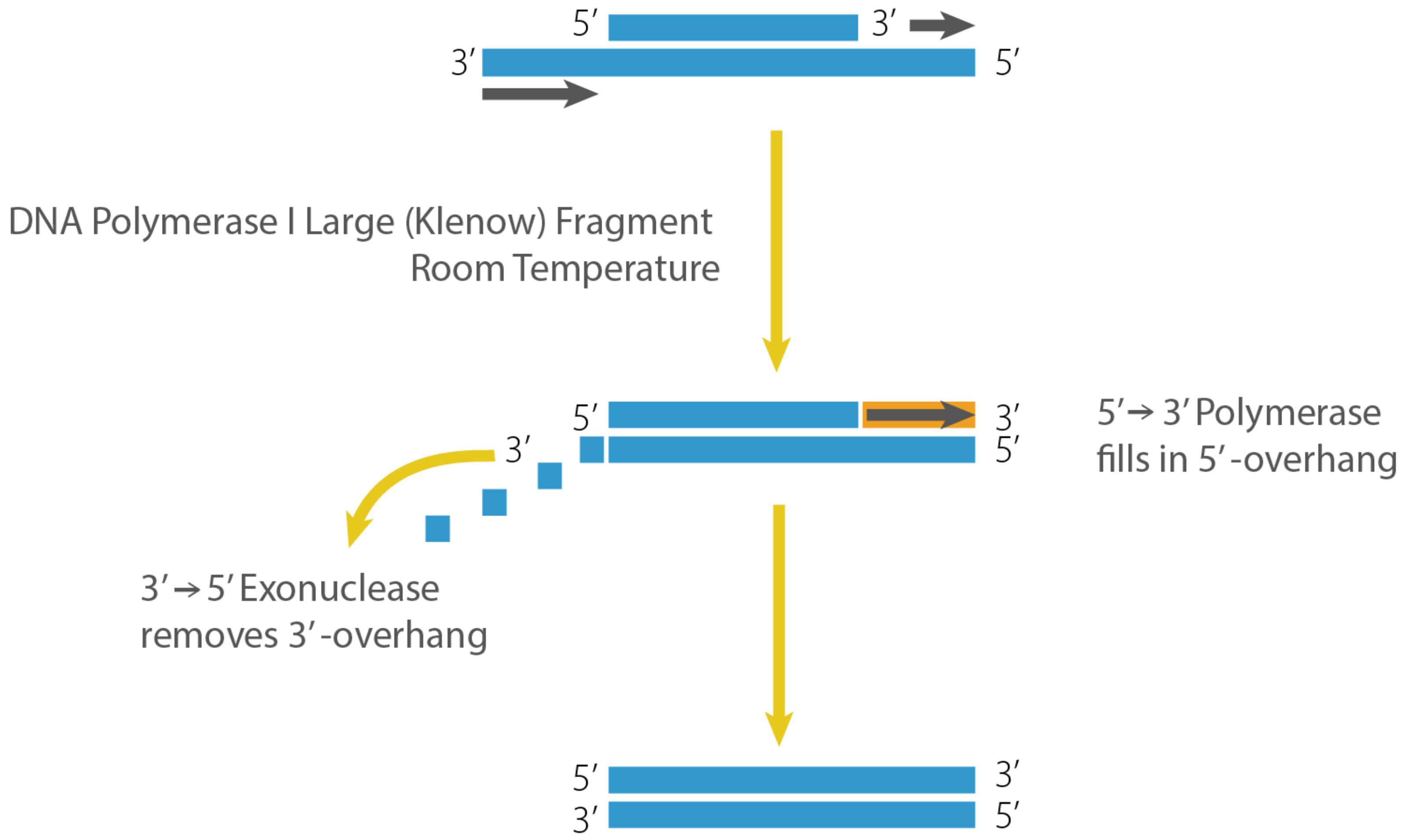
Primer

Synthesis of complementary DNA strand by a DNA polymerase

☞ DNA polymerase I (*E. coli*)	DNA dependent
☞ Taq polymerase	DNA dependent
☞ Reverse transcriptase	RNA dependent

DNA polymerase I / Klenow fragment

- *The Klenow fragment is a large protein fragment produced when DNA polymerase I from E. coli is enzymatically cleaved*
- *It retains the 5' → 3' polymerase activity and the 3’ → 5’ exonuclease activity for removal of precoding nucleotides and proofreading, but loses its 5' → 3' exonuclease activity*



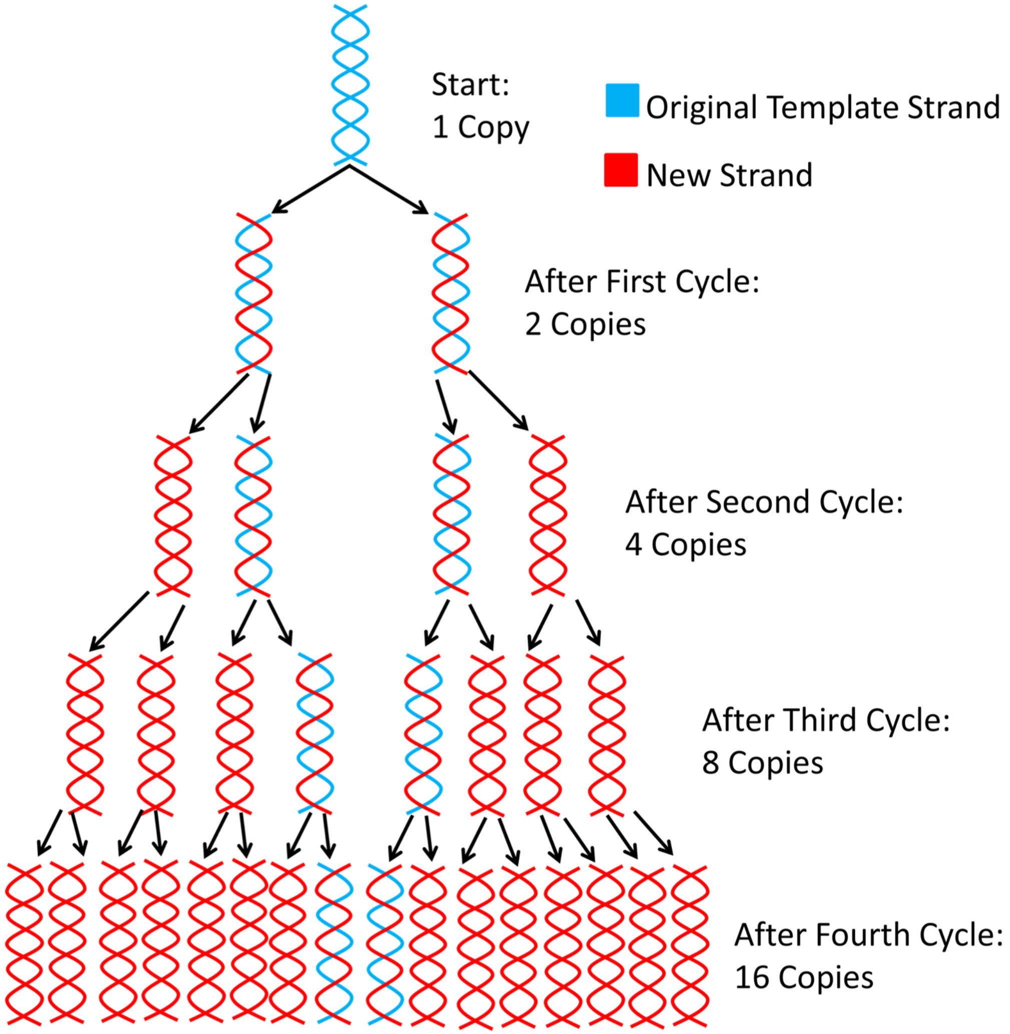
DNA polymerase I / Klenow fragment

☞ *In vitro*, the Klenow fragment is used for:

- the synthesis of the second strand, complementary to a cDNA
- the labeling of the 5' outer ends of the double-stranded DNA
- the DNA labeling by random primer technique
- the DNA sequencing by didesoxynucleotide technique
- the directed mutagenesis from synthetic oligonucleotides

Taq Polymerase and PCR

Thermus aquaticus



🟦 Original Template Strand

🟥 New Strand

$2^0 = 1$

Start:
1 Copy

$2^1 = 2$

After First Cycle:
2 Copies

$2^2 = 4$

After Second Cycle:
4 Copies

$2^3 = 8$

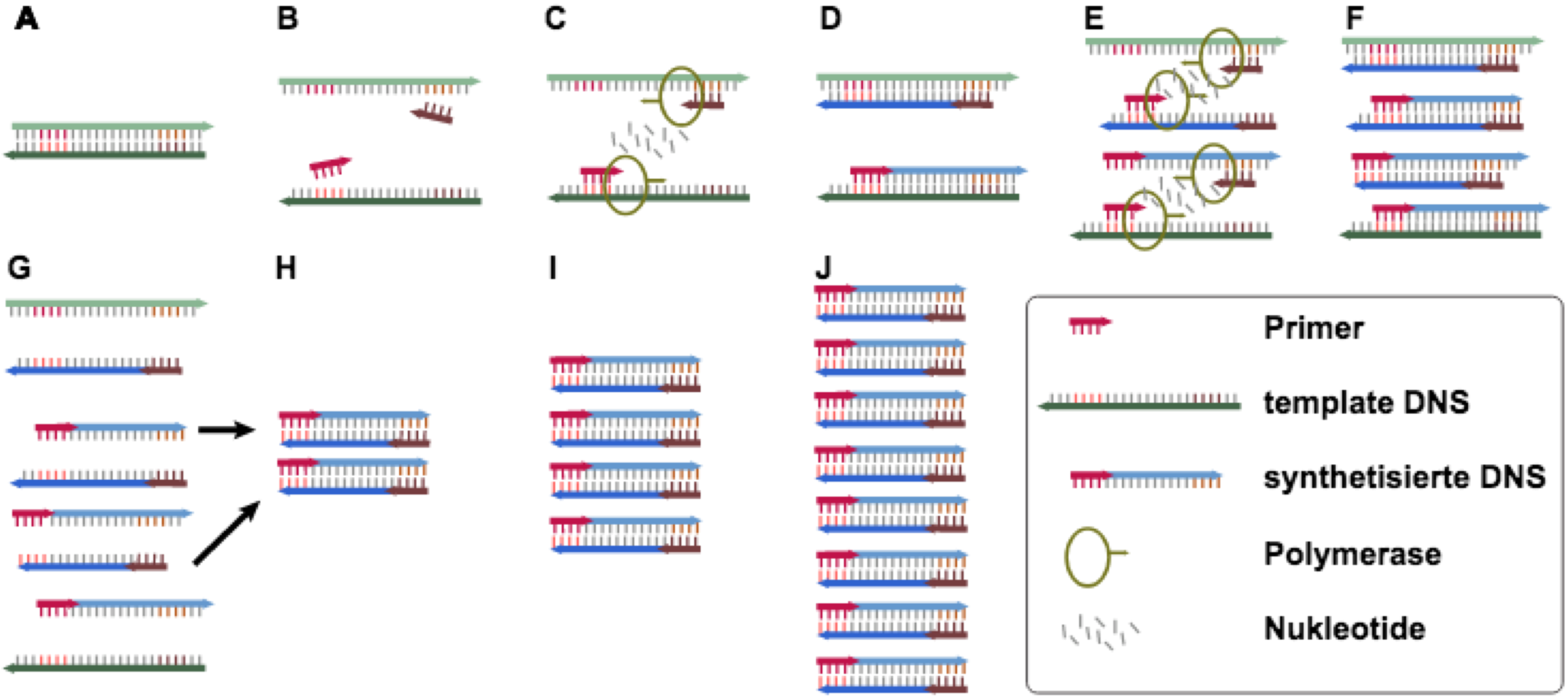
After Third Cycle:
8 Copies

$2^4 = 16$

After Fourth Cycle:
16 Copies

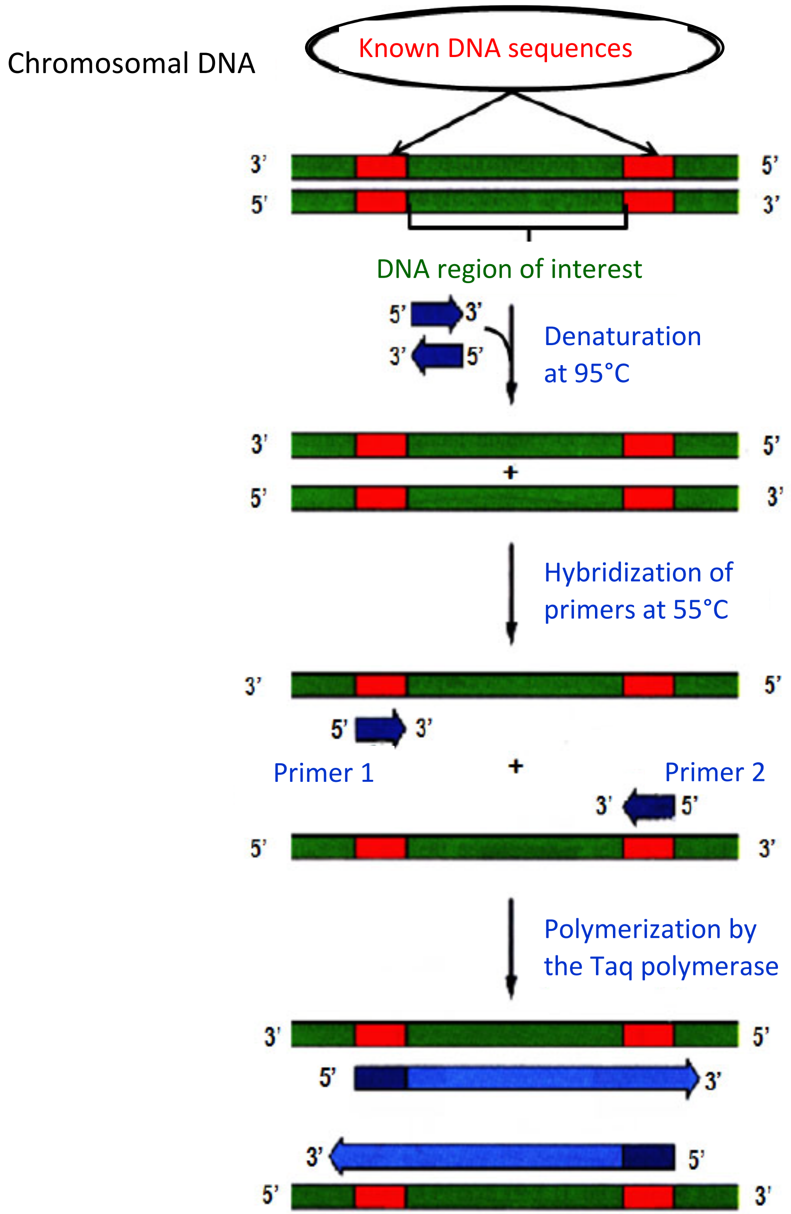
2^{20} = 1 million de copies
2^{30} = 1 milliard de copies

PCR : Polymerase Chain reaction



PCR : Polymerase Chain reaction

- PCR = Polymerase Chain reaction
  - production of a large number of copies of a specific DNA sequence
    - N cycles of PCR
    - 2^N copies of initial DNA
    - General case: N = 30 cycles
- Synthesis in direction 5' to 3'
- Reaction mix
  - Taq polymerase
  - DNA matrix
  - 2 primers forward and reverse
  - Buffer
  - dNTPs
  - MgCl2

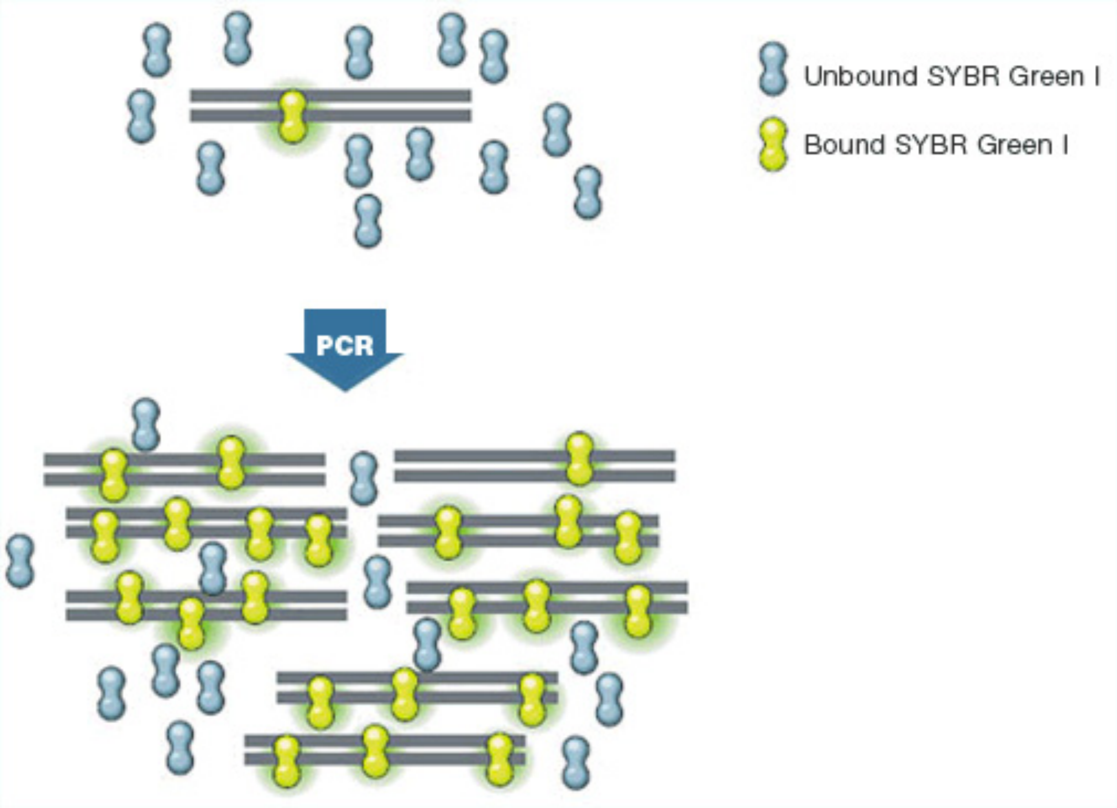


How to detect PCR products?

2 types, depending on DNA interactions

Sequence non-specific:
intercalating agent

SYBR Green



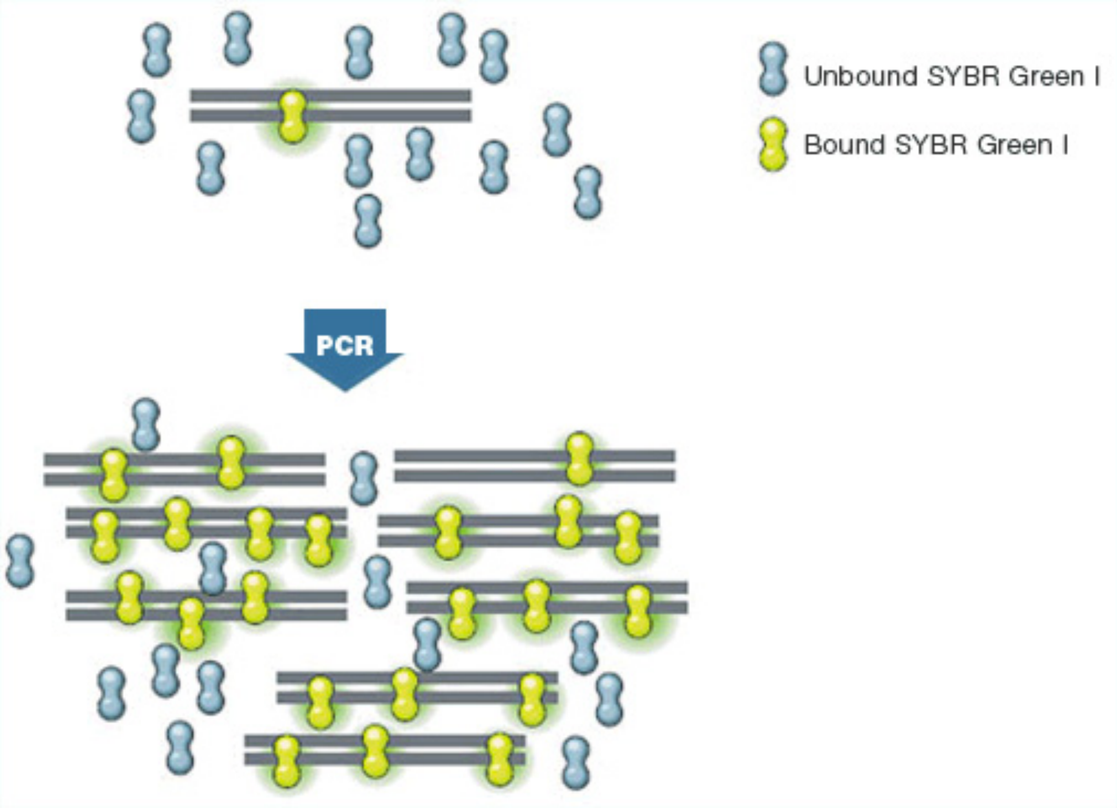
How to detect PCR products?

2 types, depending on DNA interactions

Sequence non-specific:
intercalating agent

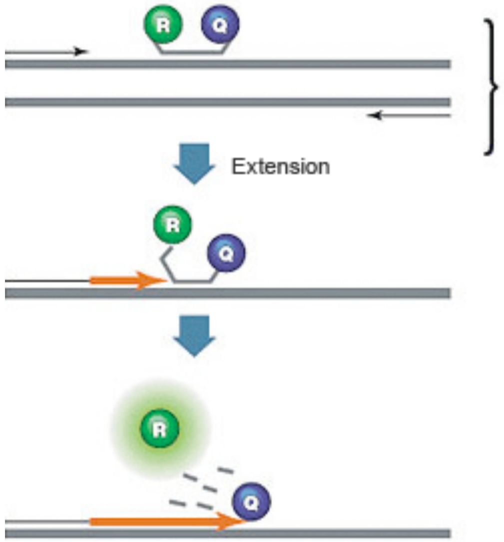
Sequence specific:
probe

SYBR Green



How to detect PCR products?

Hydrolysis (TaqMan) Probes

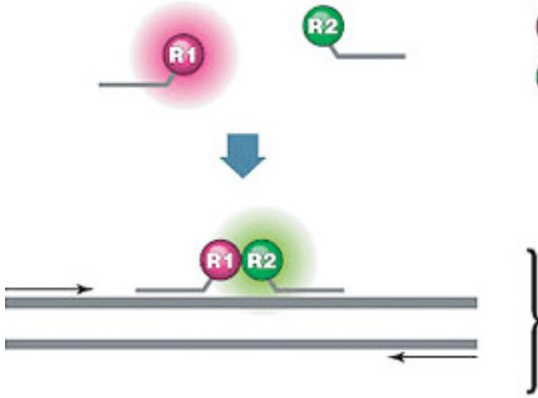


During annealing, the hydrolysis probe binds to the target sequence

During extension, the probe is partially displaced and the reporter is cleaved. The free reporter fluoresces

R Reporter
Q Quencher

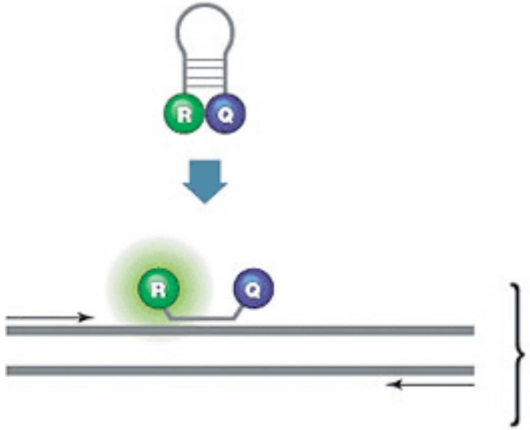
Dual Hybridization Probes



R1 Donor fluorophore
R2 Acceptor fluorophore

During annealing, the two probes bind to the target in a head-to-tail orientation. The acceptor fluorophore fluoresces

Molecular Beacons

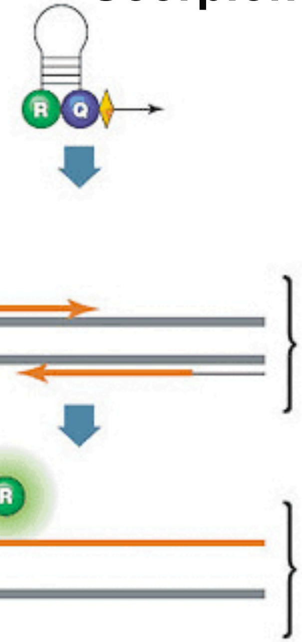


Molecular beacons are hairpin probes with reporter and quencher

During annealing, the probe binds to the target sequence to separate reporter and quencher. The reporter fluoresces

Scorpions PCR Primers

R Reporter
Q Quencher
◆ PCR blocker



The Scorpions primer acts as a probe. The intact primer forms a hairpin so that the quenched reporter does not fluoresce

During annealing, the hairpin primer binds to the template, and is then extended

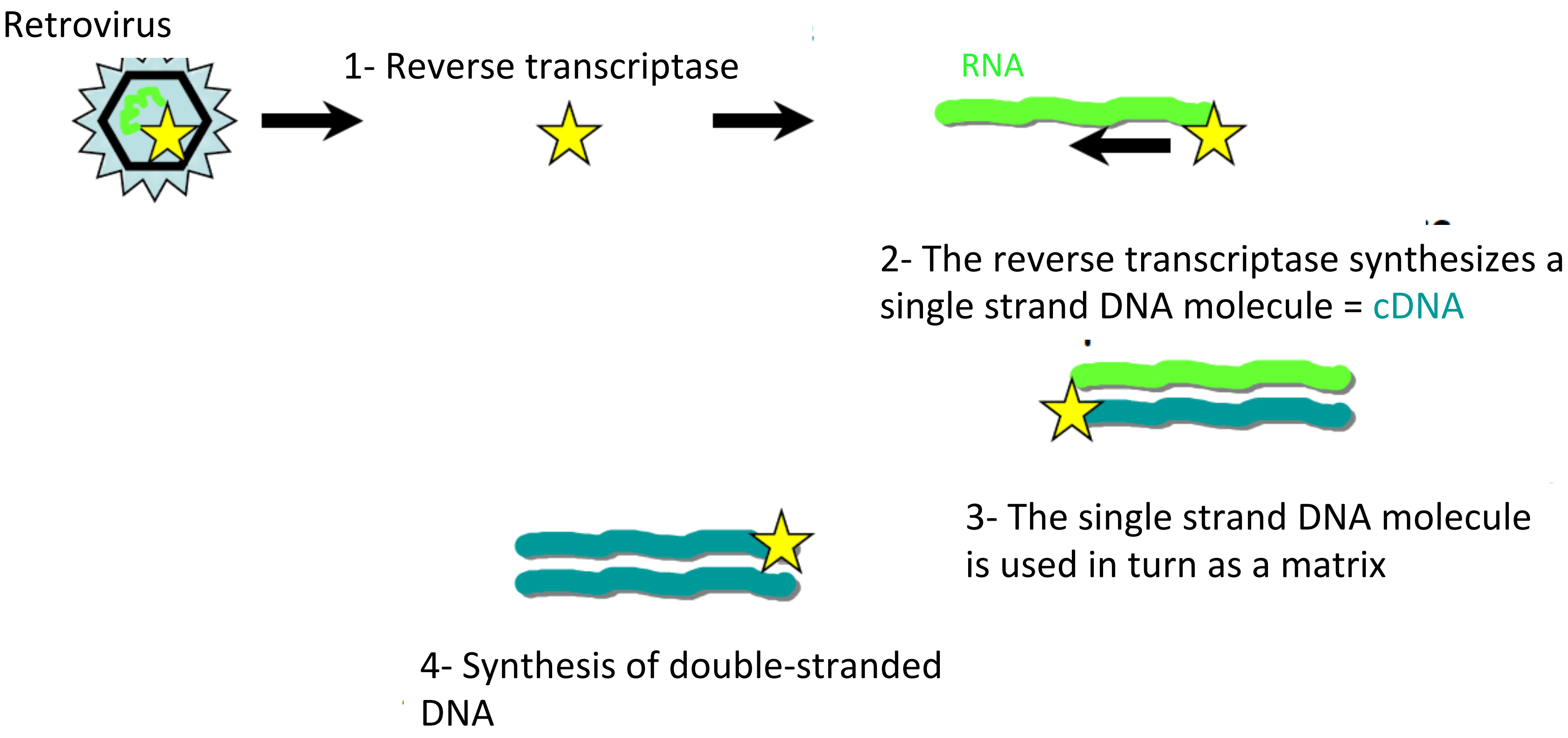
During subsequent denaturation, the reporter separates from the quencher, and the loop sequence binds to the internal target sequence. The reporter on the extended Scorpions primer fluoresces

https://www.bio-rad.com/

RT-PCR / RT-qPCR

☞ Reverse transcriptase synthesizes single strand DNA

↳ using an RNA as a matrix



Retrovirus

1- Reverse transcriptase

RNA

2- The reverse transcriptase synthesizes a single strand DNA molecule = cDNA

3- The single strand DNA molecule is used in turn as a matrix

4- Synthesis of double-stranded DNA

⇨ The complementary DNA (cDNA) is then amplifiable by PCR

A

Reverse transcription-PCR

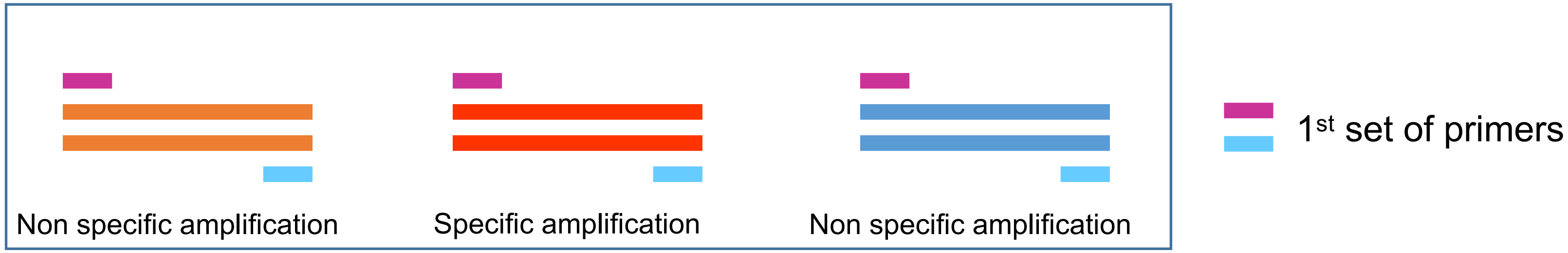
1. Isolate RNA

DOI: 10.1042/BIO20200034

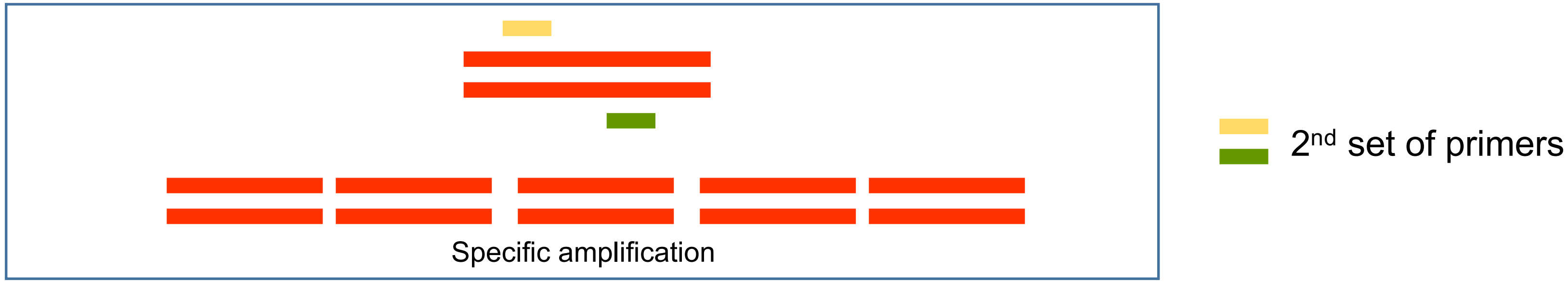
Nested PCR

☞ 2 successive PCR intended to amplify specifically one DNA sequence

1st amplification (10 cycles)



2nd amplification within the amplicon of the 1st amplification (25 cycles)

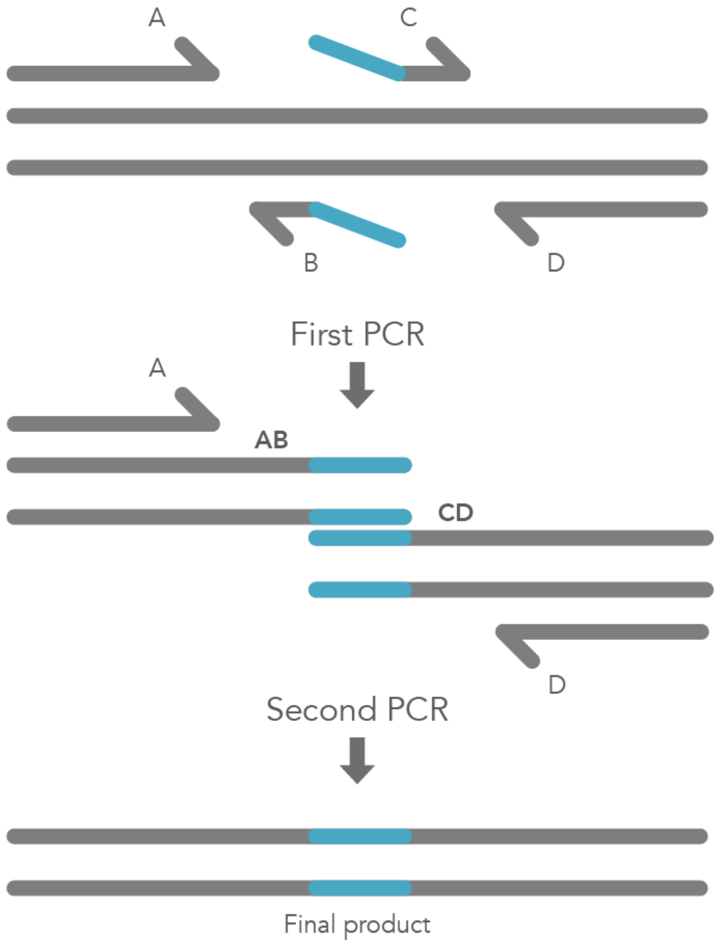


⇨ Possible to detect:

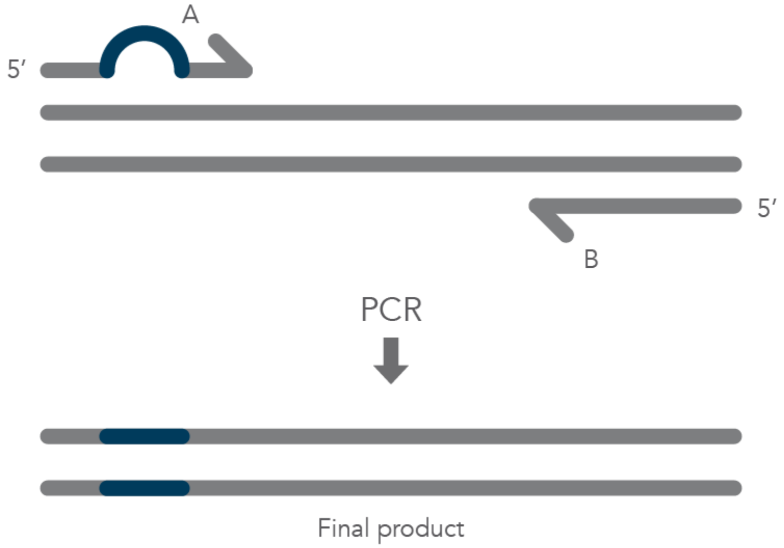
- 1 abnormal cell among 135 normal cells
- up to 3 fg of starting DNA

Site-directed mutagenesis by PCR

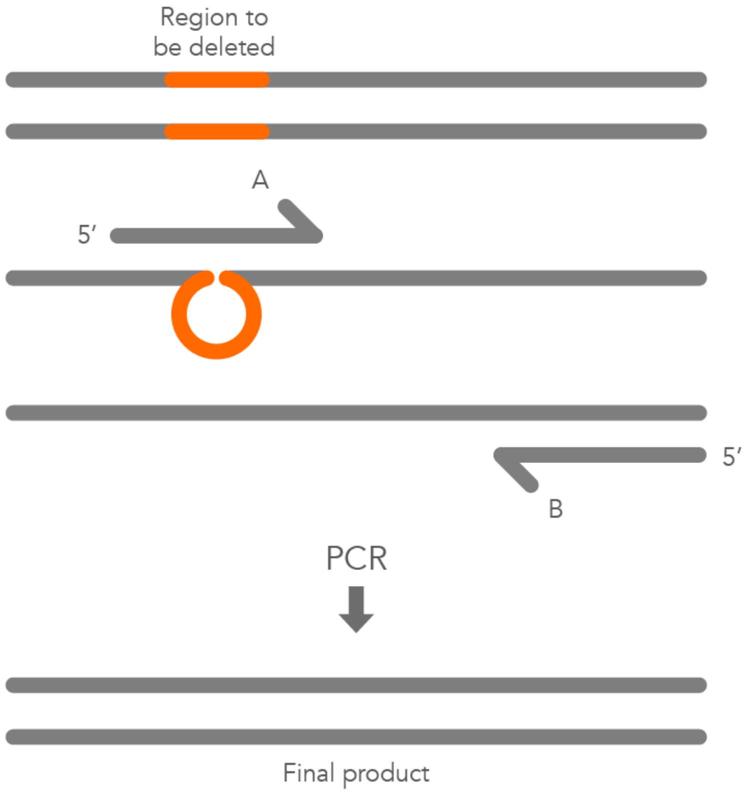
Insertion



Mutation

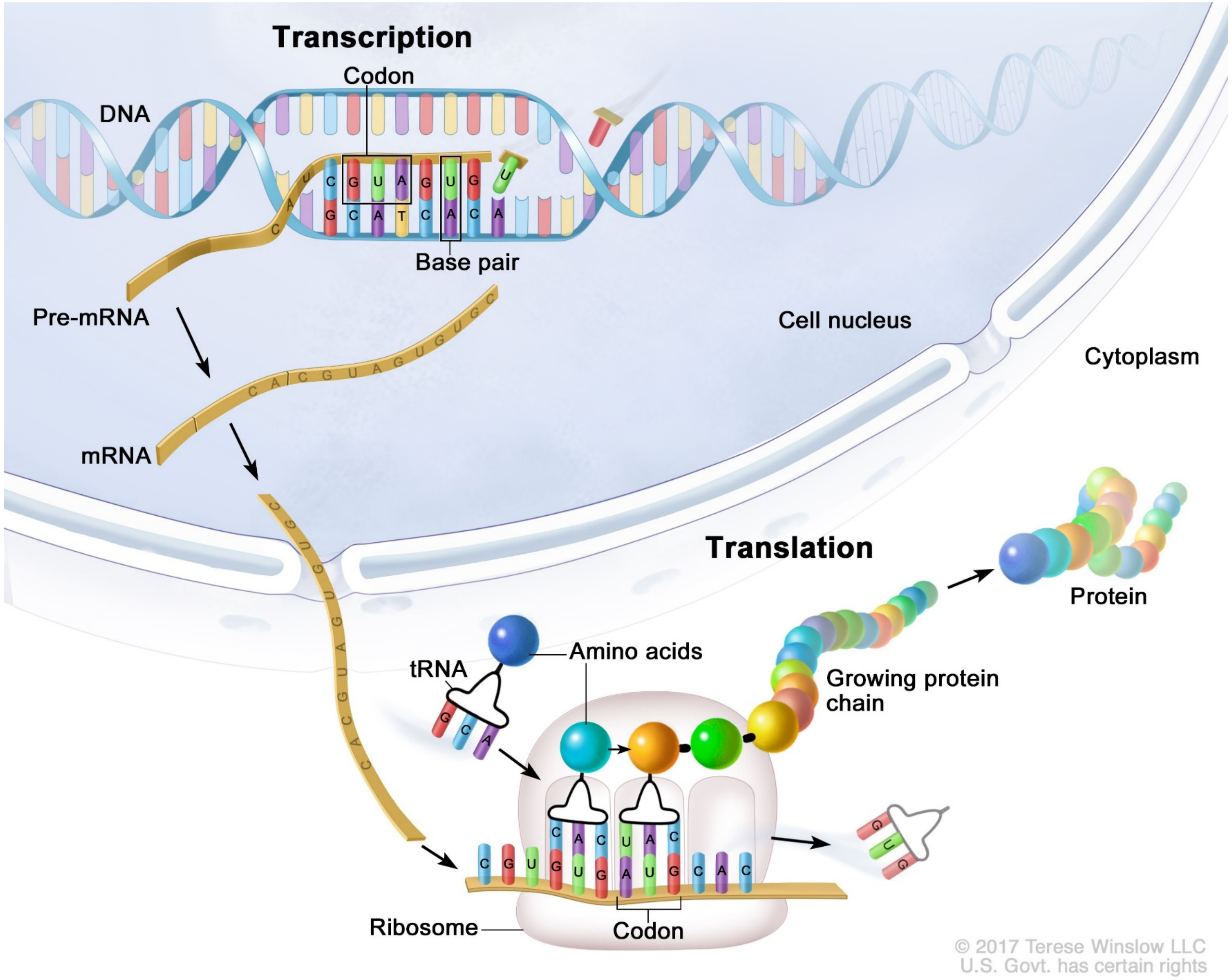


Deletion



Site-directed mutagenesis by PCR

Plan



© 2017 Terese Winslow LLC
U.S. Govt. has certain rights

I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

- Basic mechanism
- Maturation of mRNA

III- Translation: from RNA to proteins

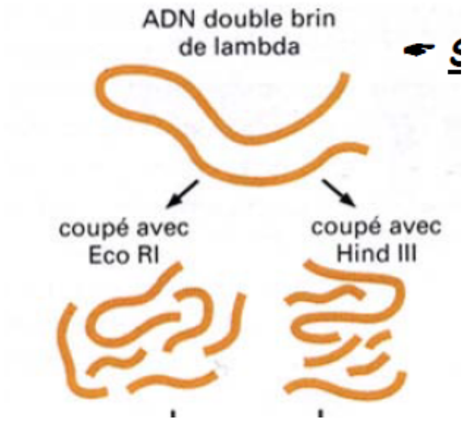
IV- Molecular tools in the study of DNA

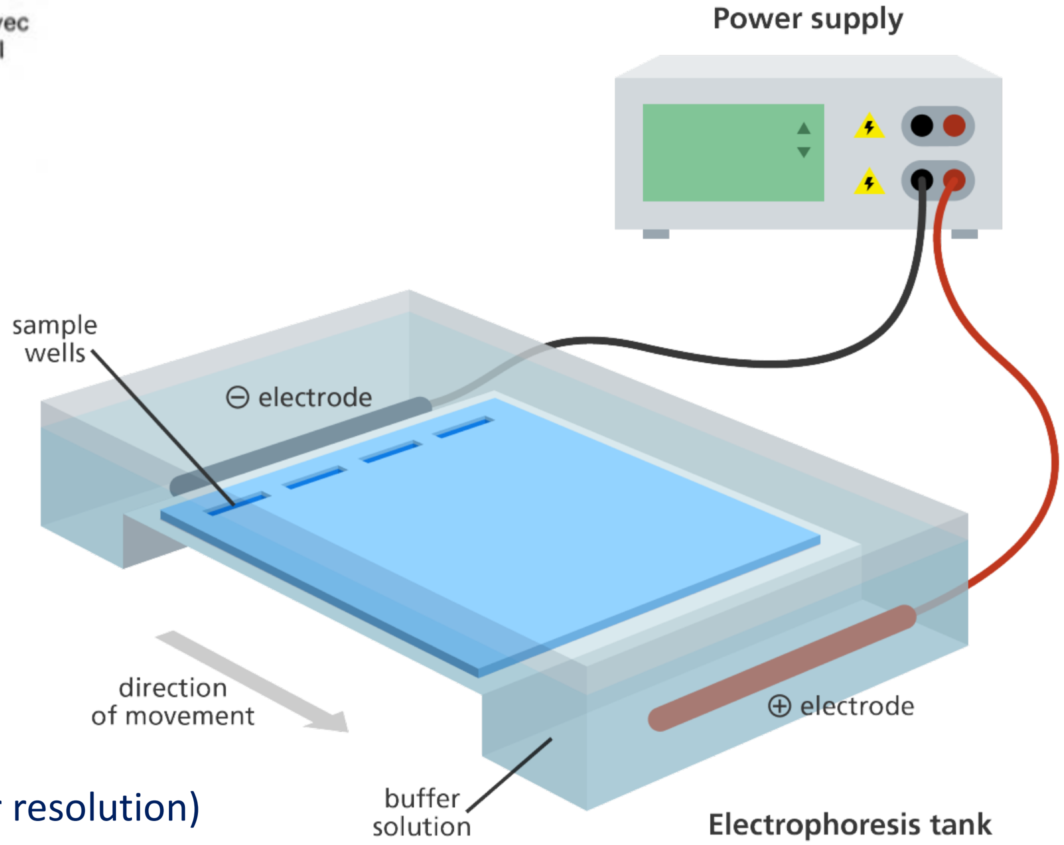
- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

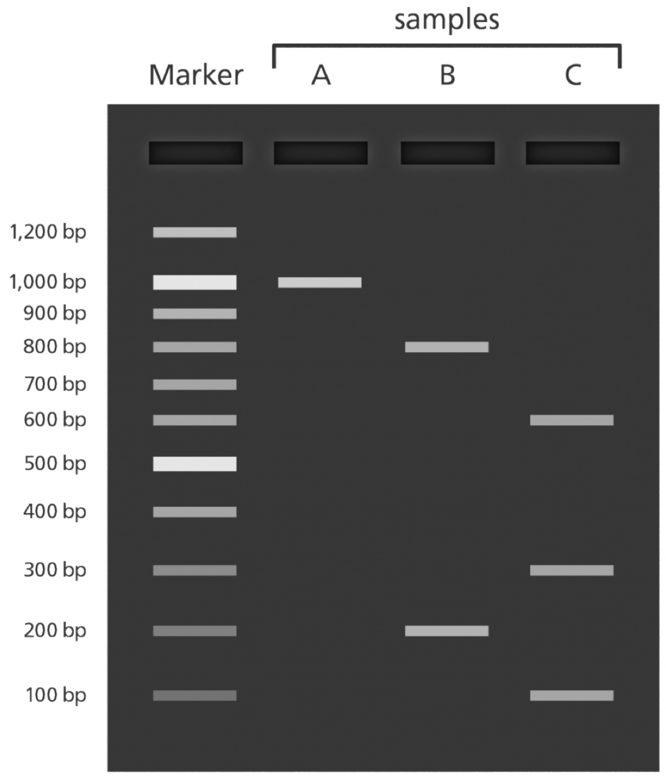
Electrophoresis : Separation and detection of DNA molecules by size

⇨ Phosphate groups of the DNA are ionized and polynucleotide chains are polyanions (-)

⇨ Migration to the positive pole (+)







Detection with DNA intercaling agent

☞ Separation on gel

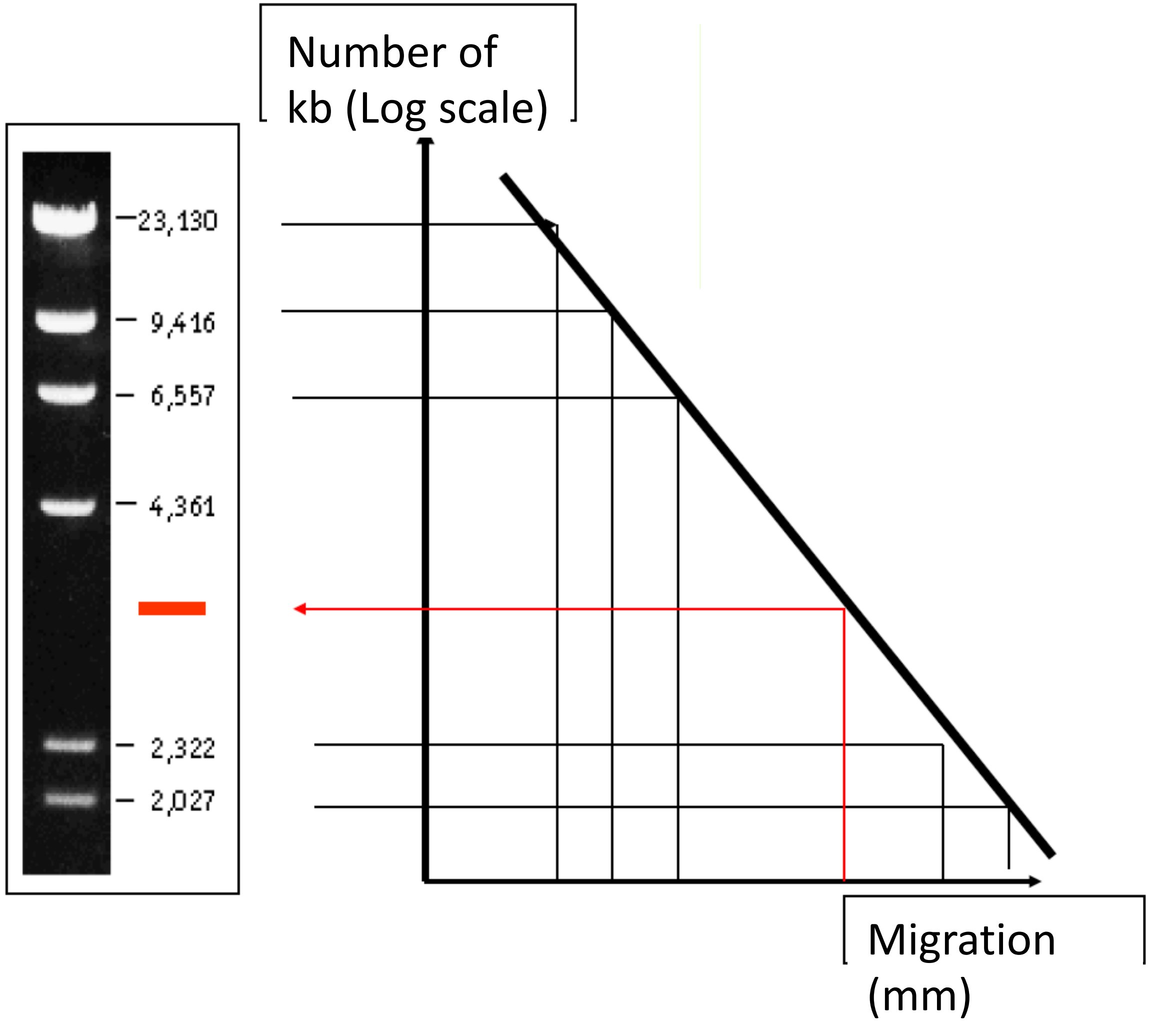
of agarose

of polyacrylamide (better resolution)

Electrophoresis : Separation and detection of DNA molecules by size

☞ The mobility of a DNA fragment

is inversely proportional to the logarithm of the number of base pairs



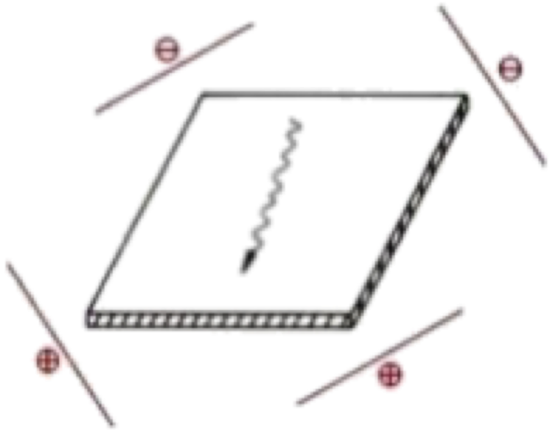
Electrophoresis : Separation and detection of DNA molecules by size

☞ Examples of resolution depending on the types of gel and migration

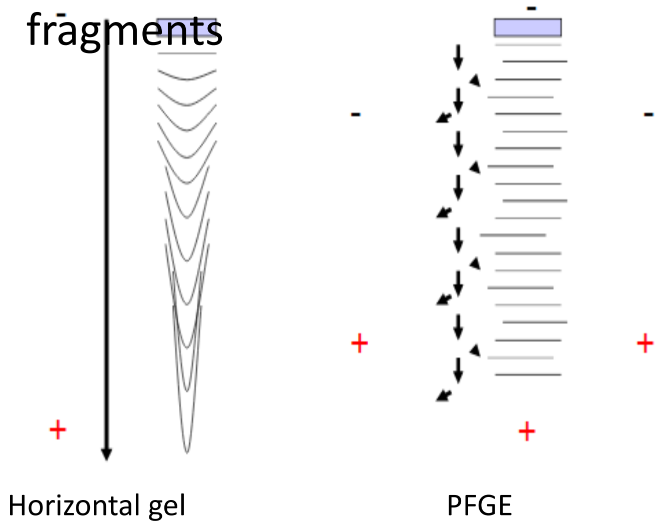
Agarose %	DNA length
0.5	1-30 kb
0.7	0.8-12 kb
1.0	0,5-10 kb
1.2	0.4-7 kb
1.5	0.2-3 kb

Types of electrophoresis	DNA length
Acrylamide gel	1-100 bp
Agarose	0.1-10 kb
PFGE	10 kb-10 Mb

☞ Pulse field gel electrophoresis (PFGE)

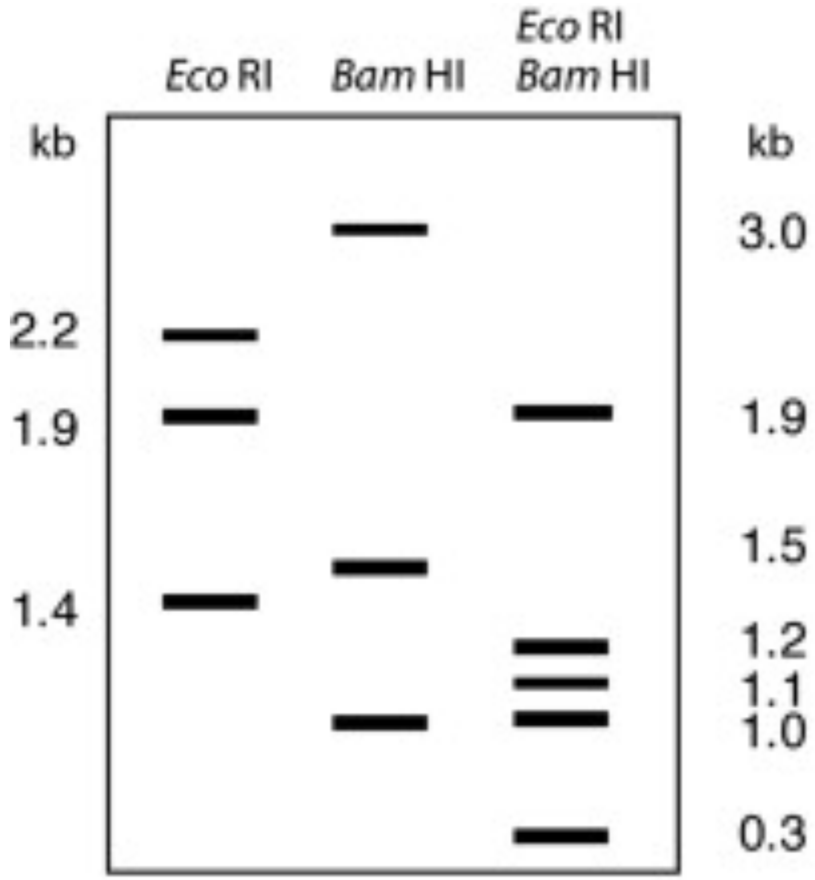


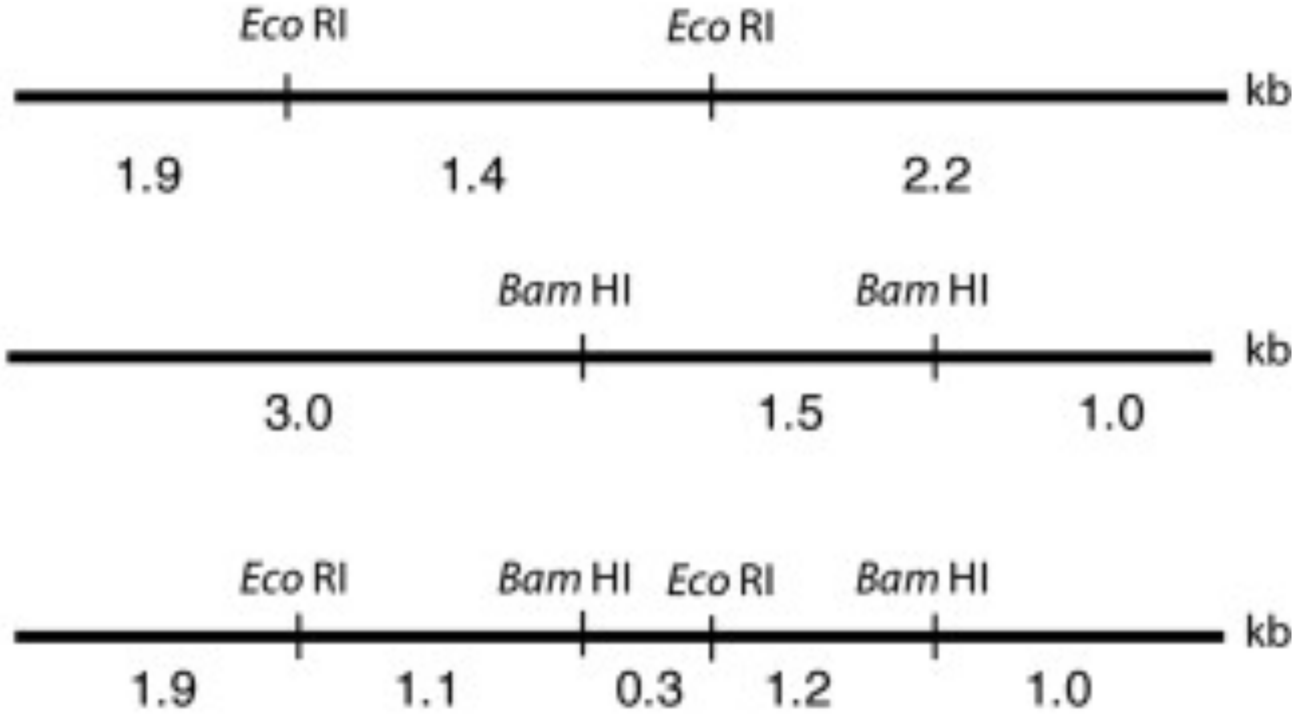
Separation of large DNA fragments



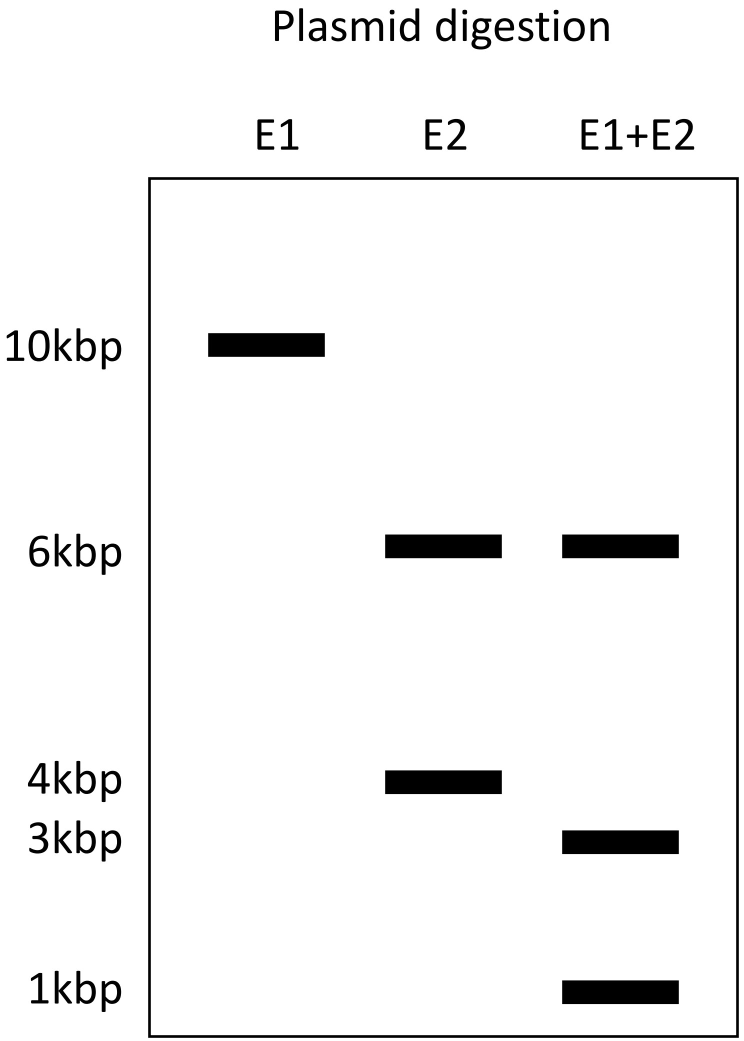
Restriction maps : Linear DNA

- Example of establishing a restriction map

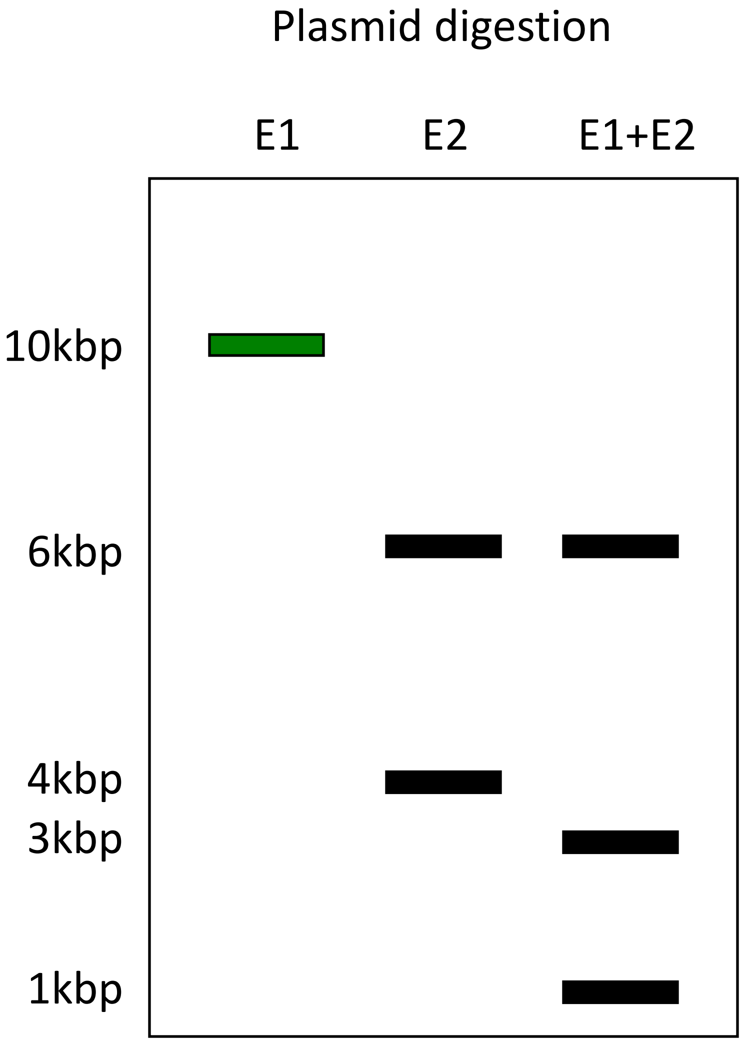


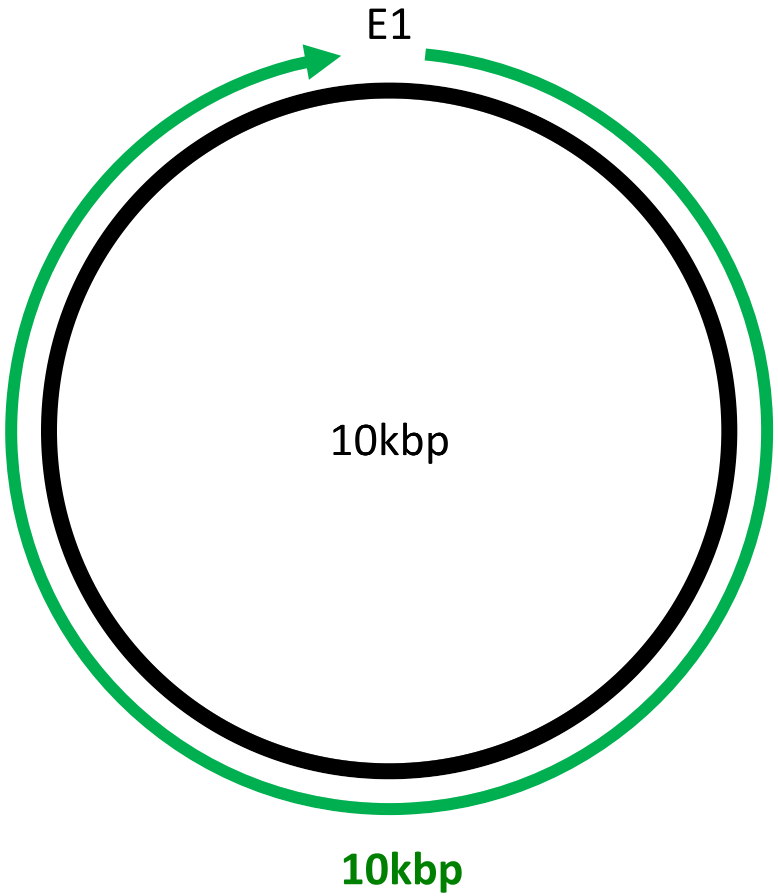


Your turn…

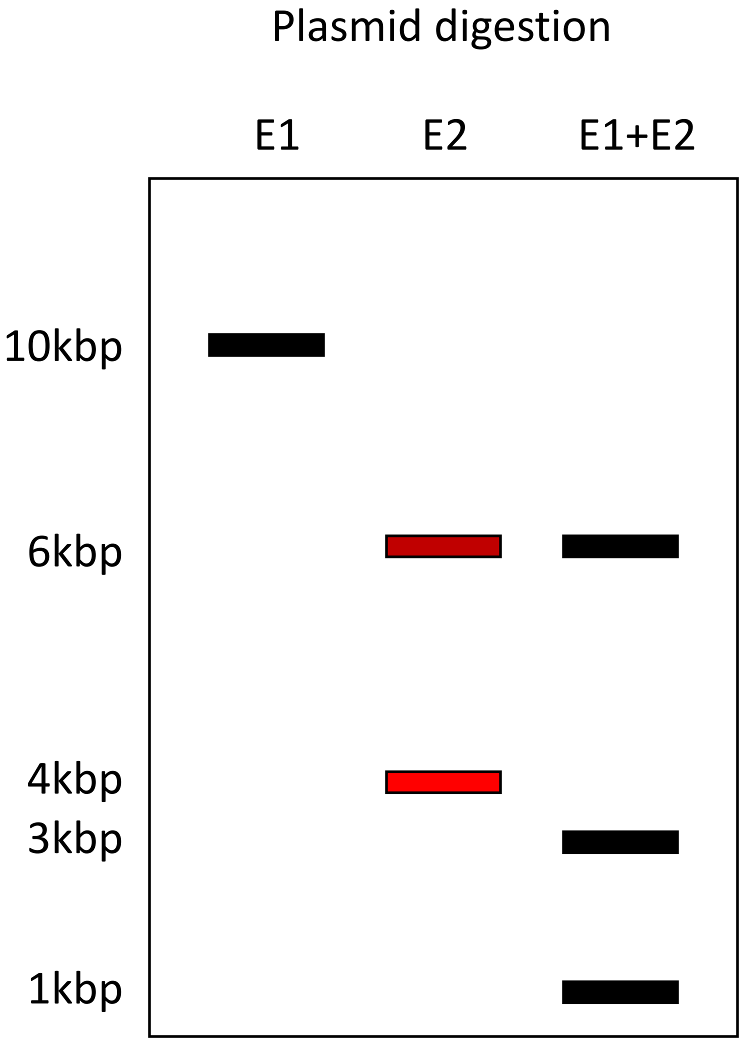


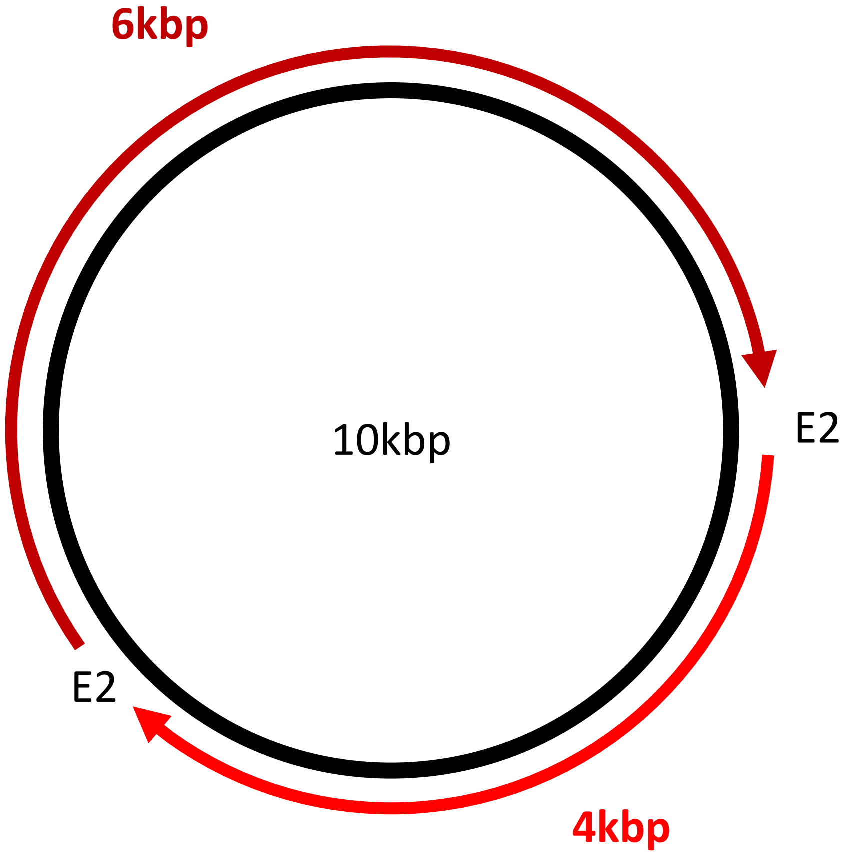
Your turn…


Plasmid digestion
E1
E2
E1+E2
10kbp
6kbp
4kbp
3kbp
1kbp

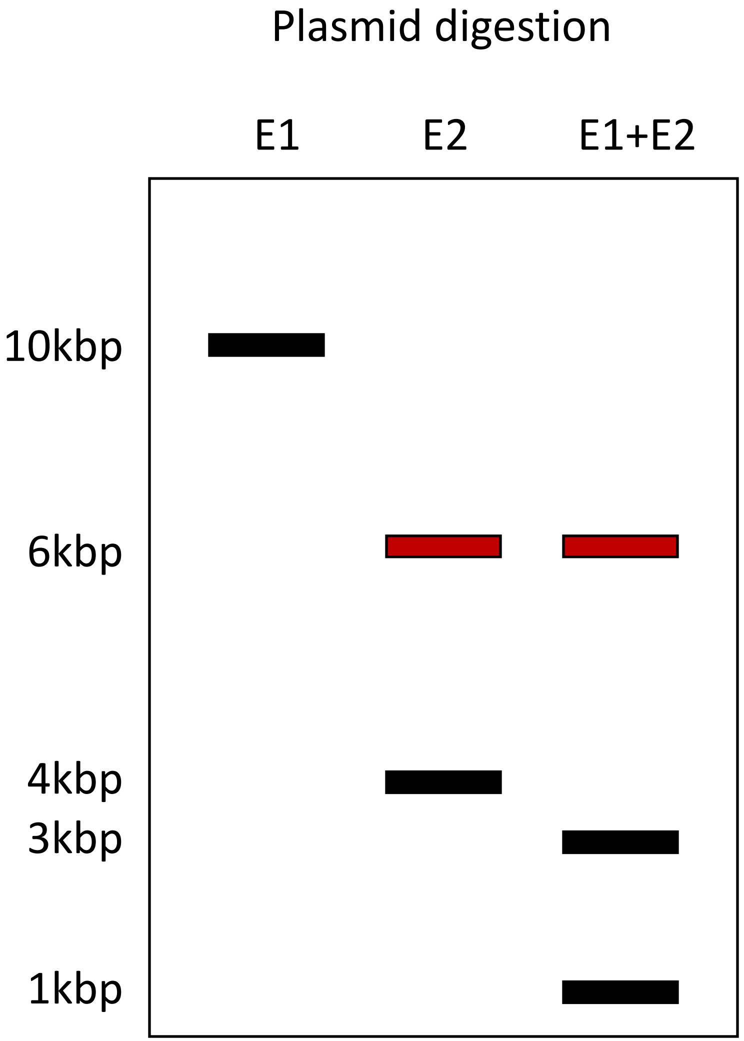

E1
10kbp
10kbp

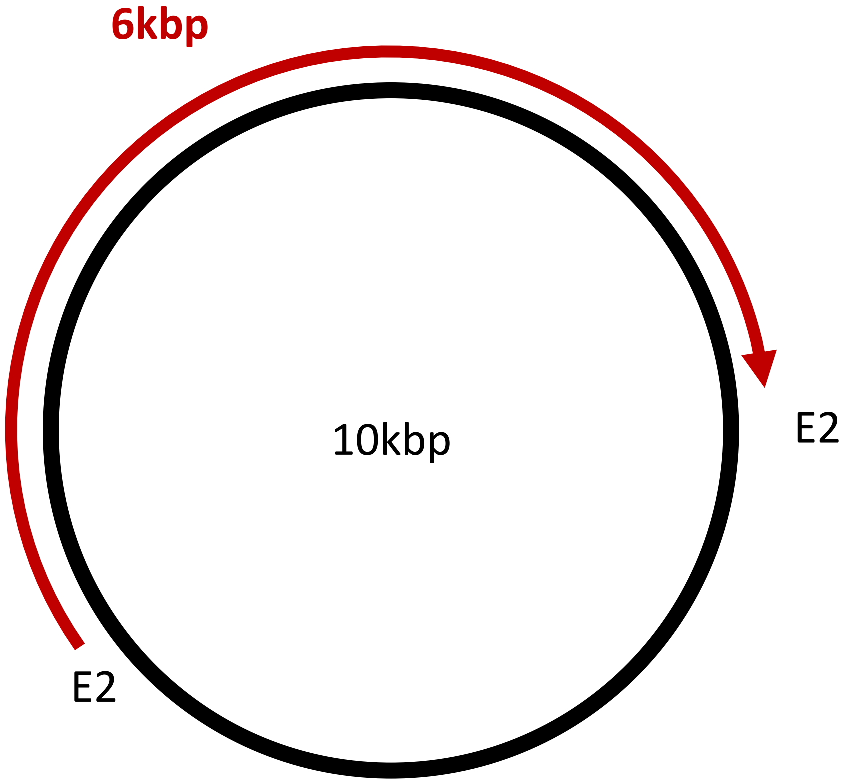
Your turn…


Plasmid digestion
E1
E2
E1+E2
10kbp
6kbp
4kbp
3kbp
1kbp

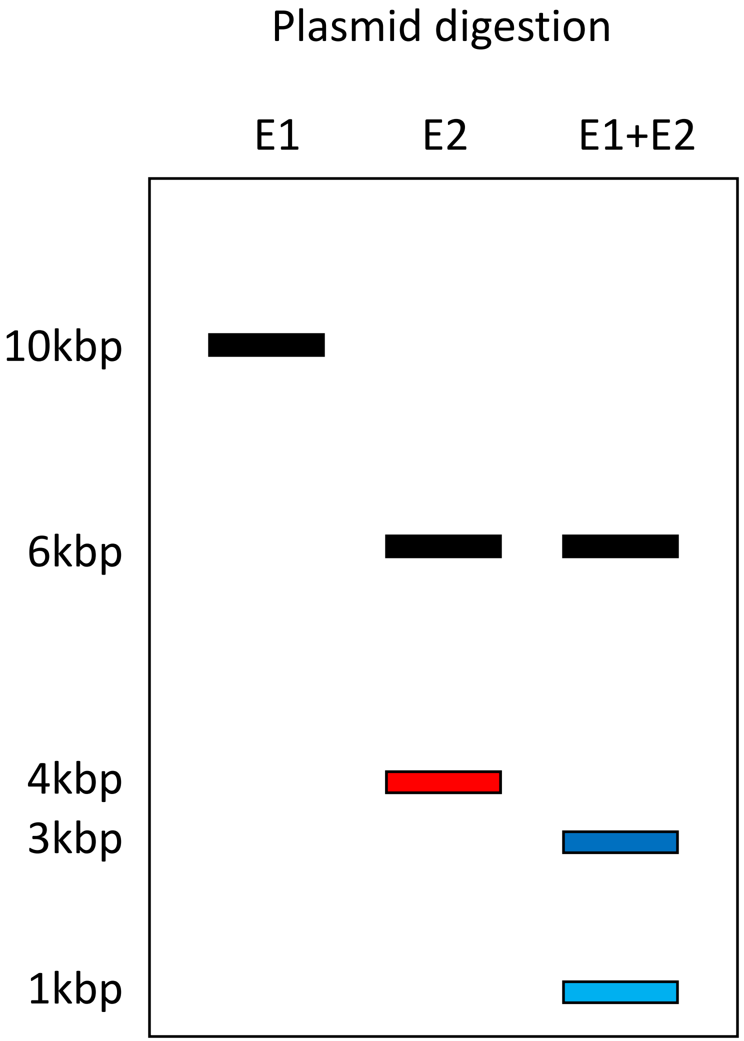

6kbp
10kbp
E2
E2
4kbp

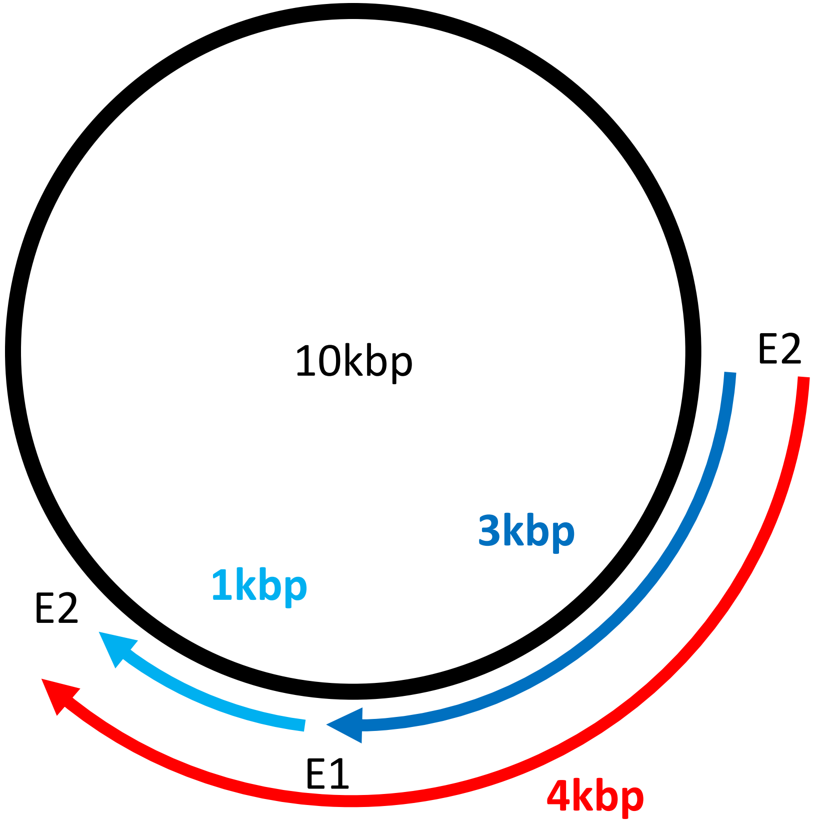
Your turn…


Plasmid digestion
E1
E2
E1+E2
10kbp
6kbp
4kbp
3kbp
1kbp

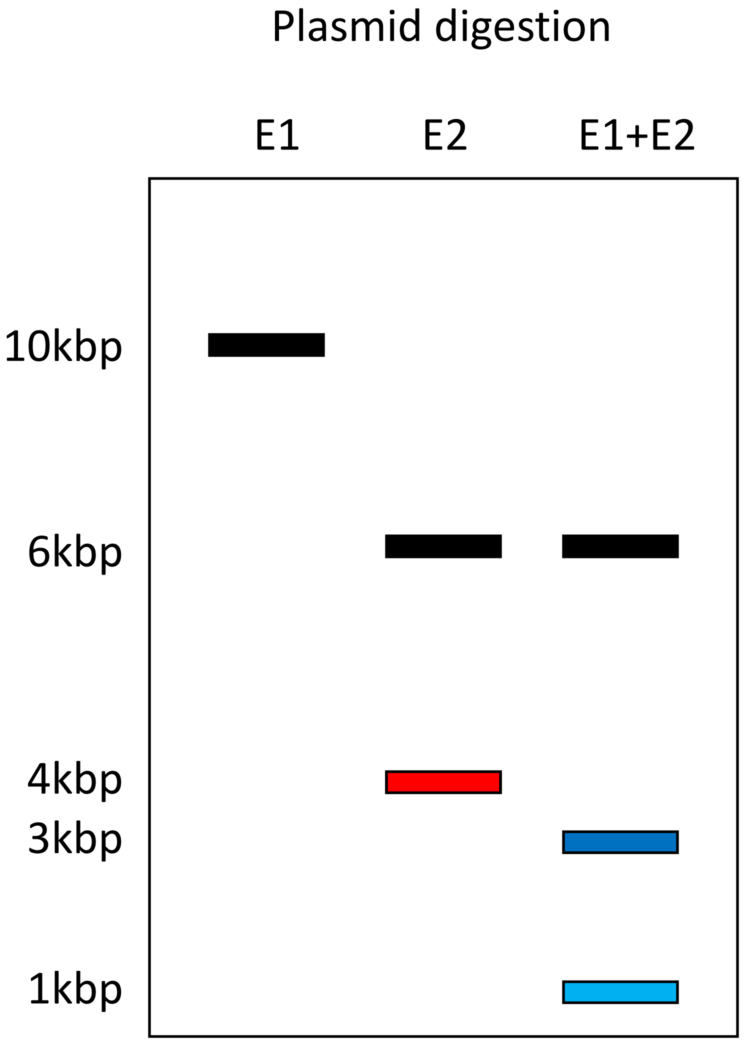

6kbp
E2
10kbp
E2

Your turn…

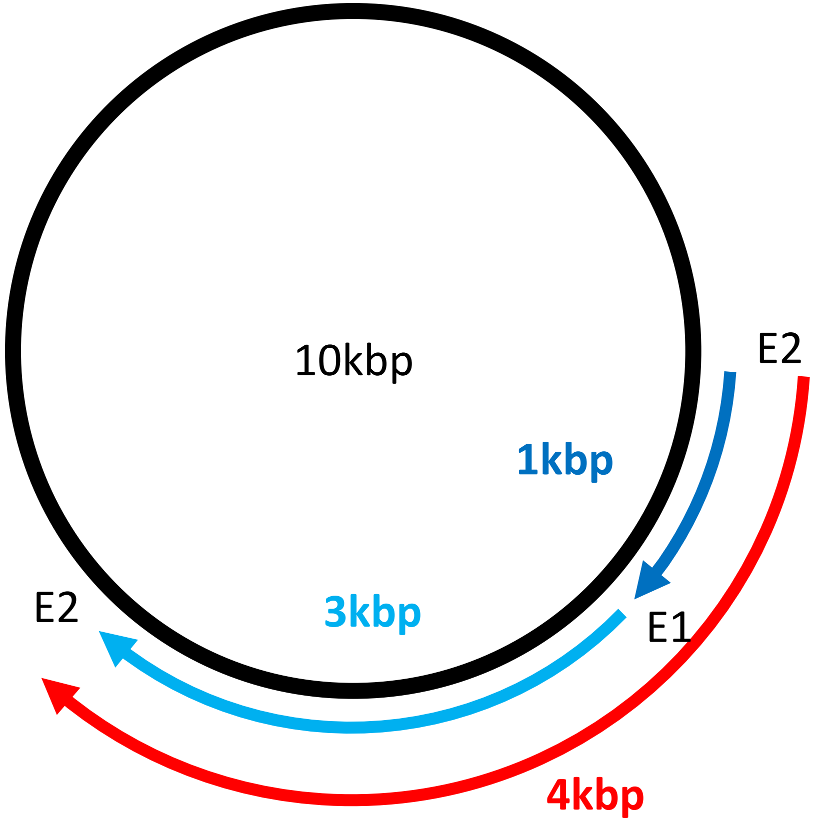

Plasmid digestion
E1
E2
E1+E2
10kbp
6kbp
4kbp
3kbp
1kbp


10kbp
E2
3kbp
1kbp
E2
E1
4kbp

Your turn...



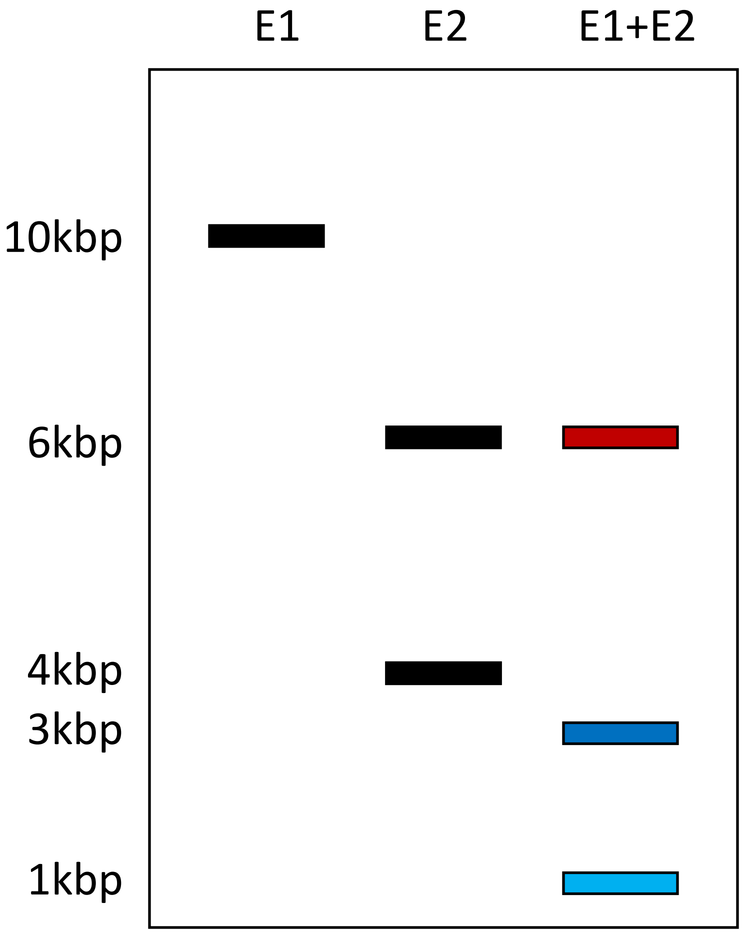
Plasmid digestion
E1
E2
E1+E2
10kbp
6kbp
4kbp
3kbp
1kbp

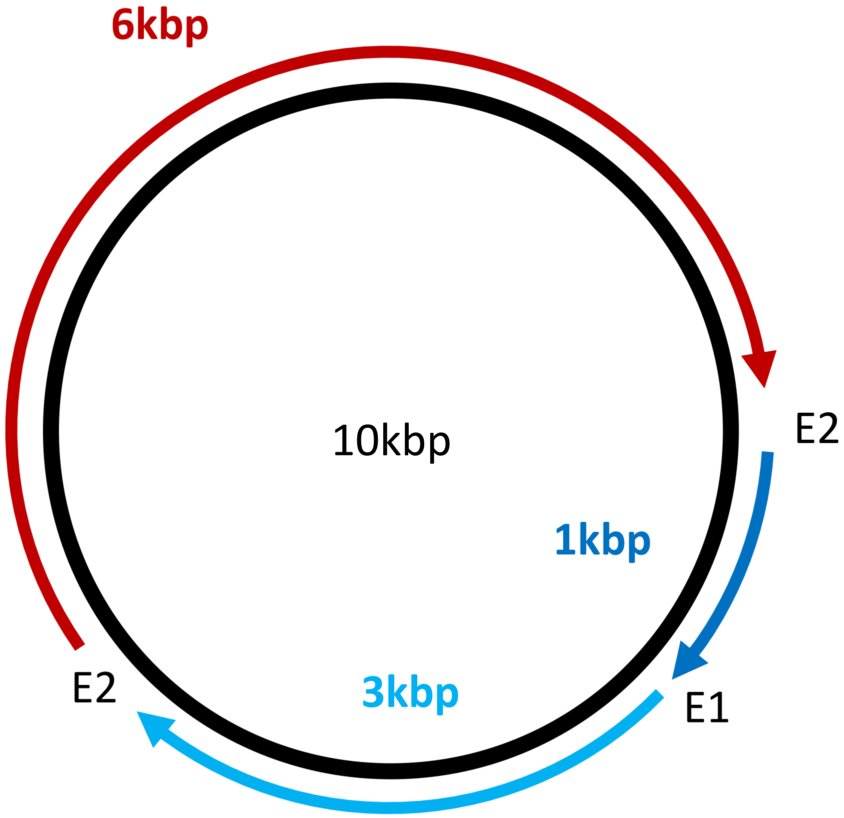


10kbp
E2
1kbp
E2
3kbp
E1
4kbp

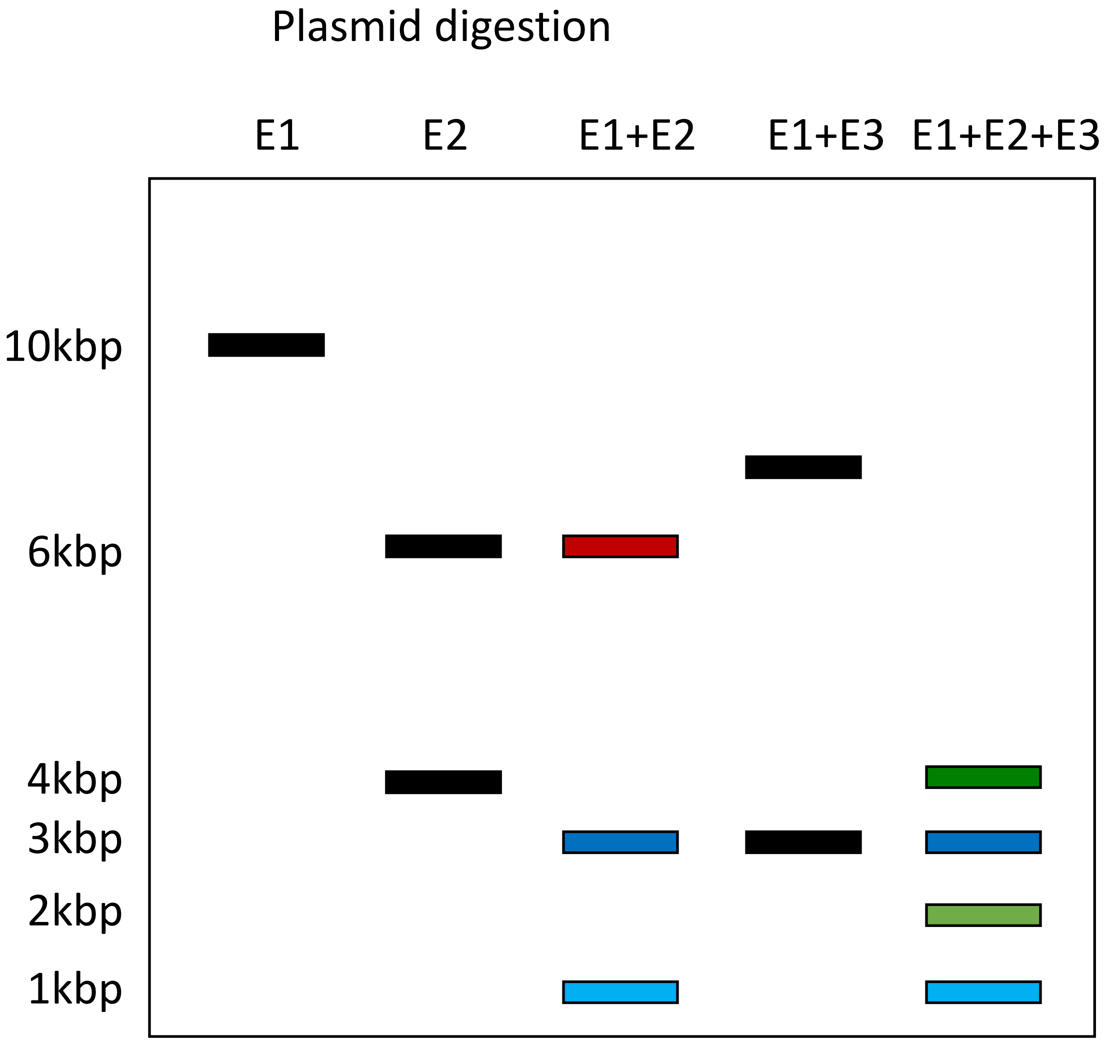
Your turn…

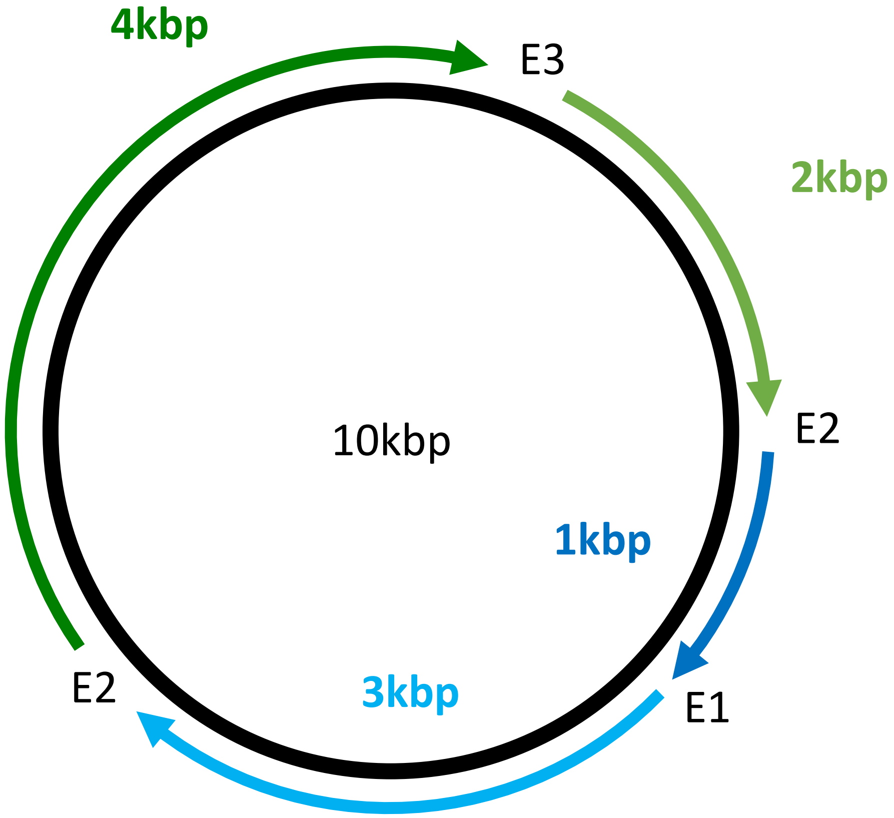
Plasmid digestion





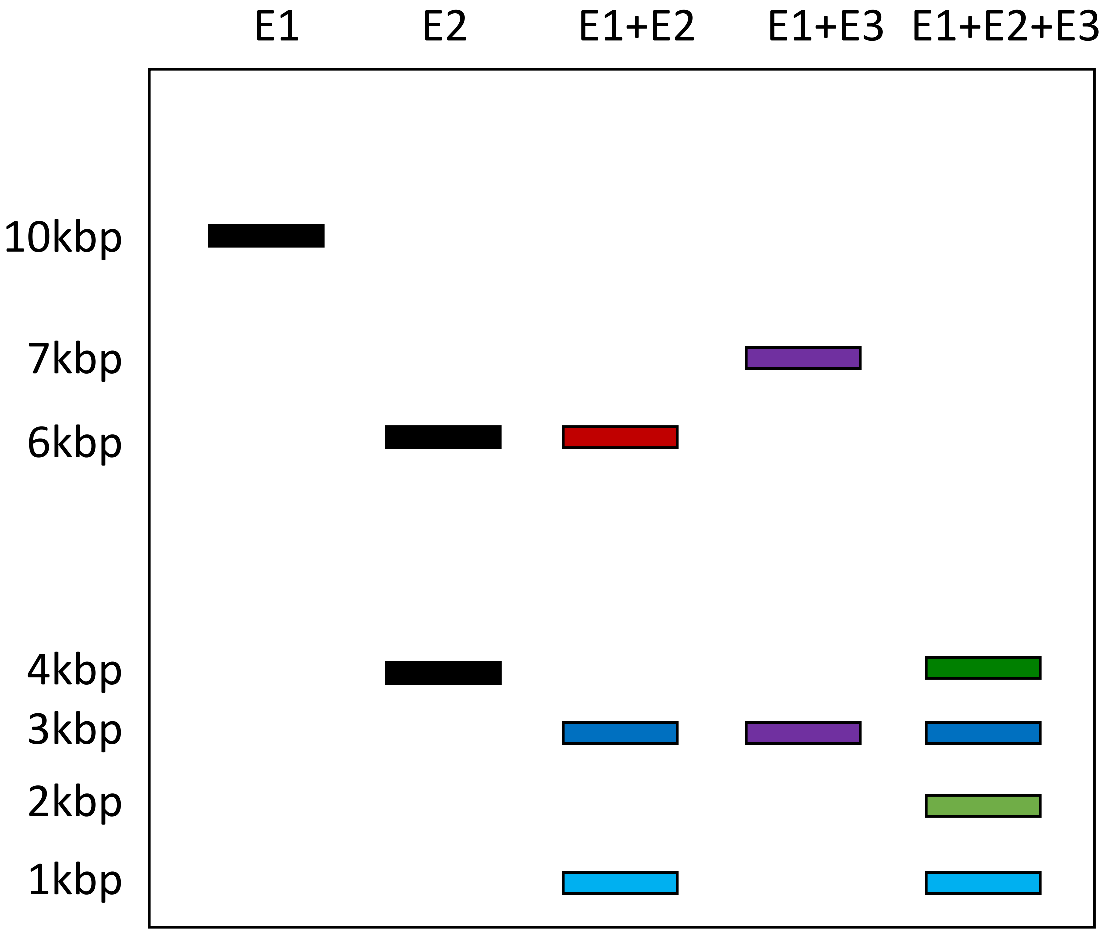
Your turn…

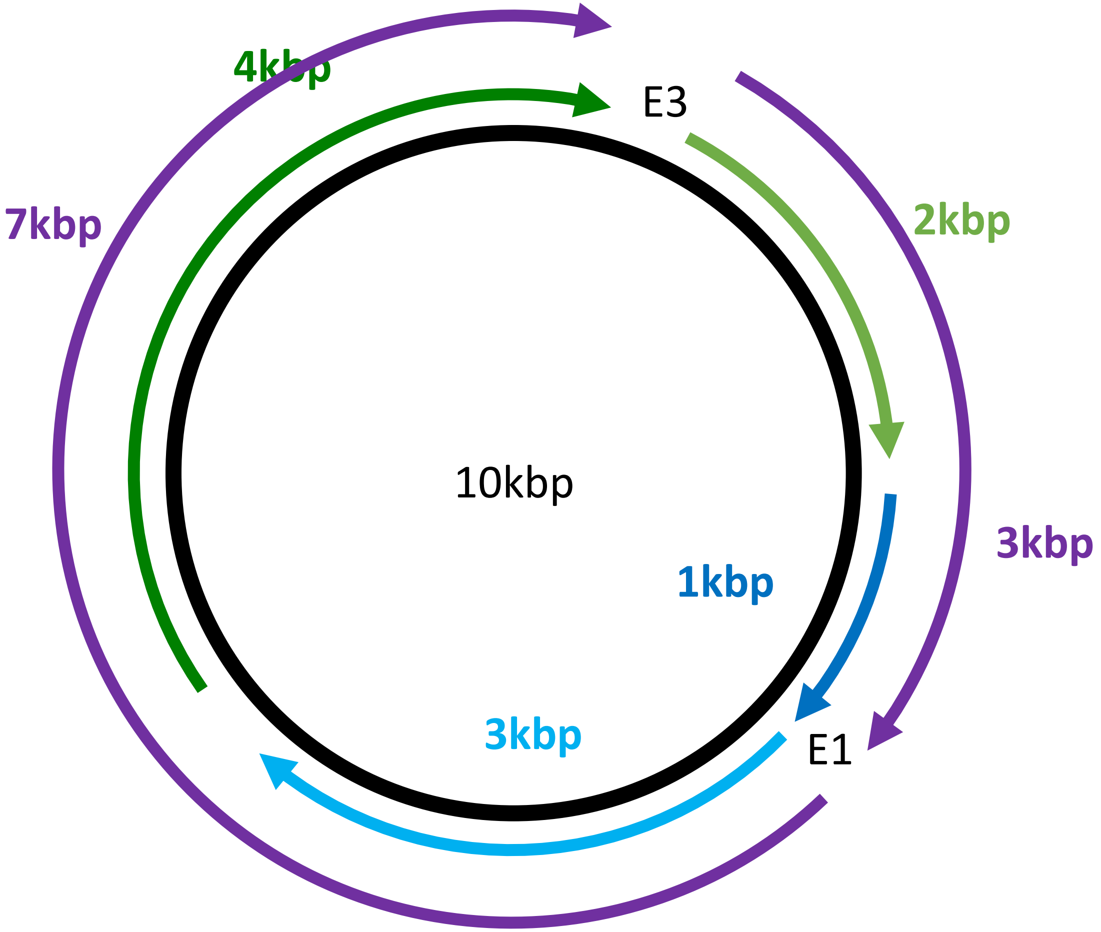

Plasmid digestion
E1
E2
E1+E2
E1+E3
E1+E2+E3
10kbp
6kbp
4kbp
3kbp
2kbp
1kbp


4kbp
E3
2kbp
E2
10kbp
1kbp
E2
3kbp
E1

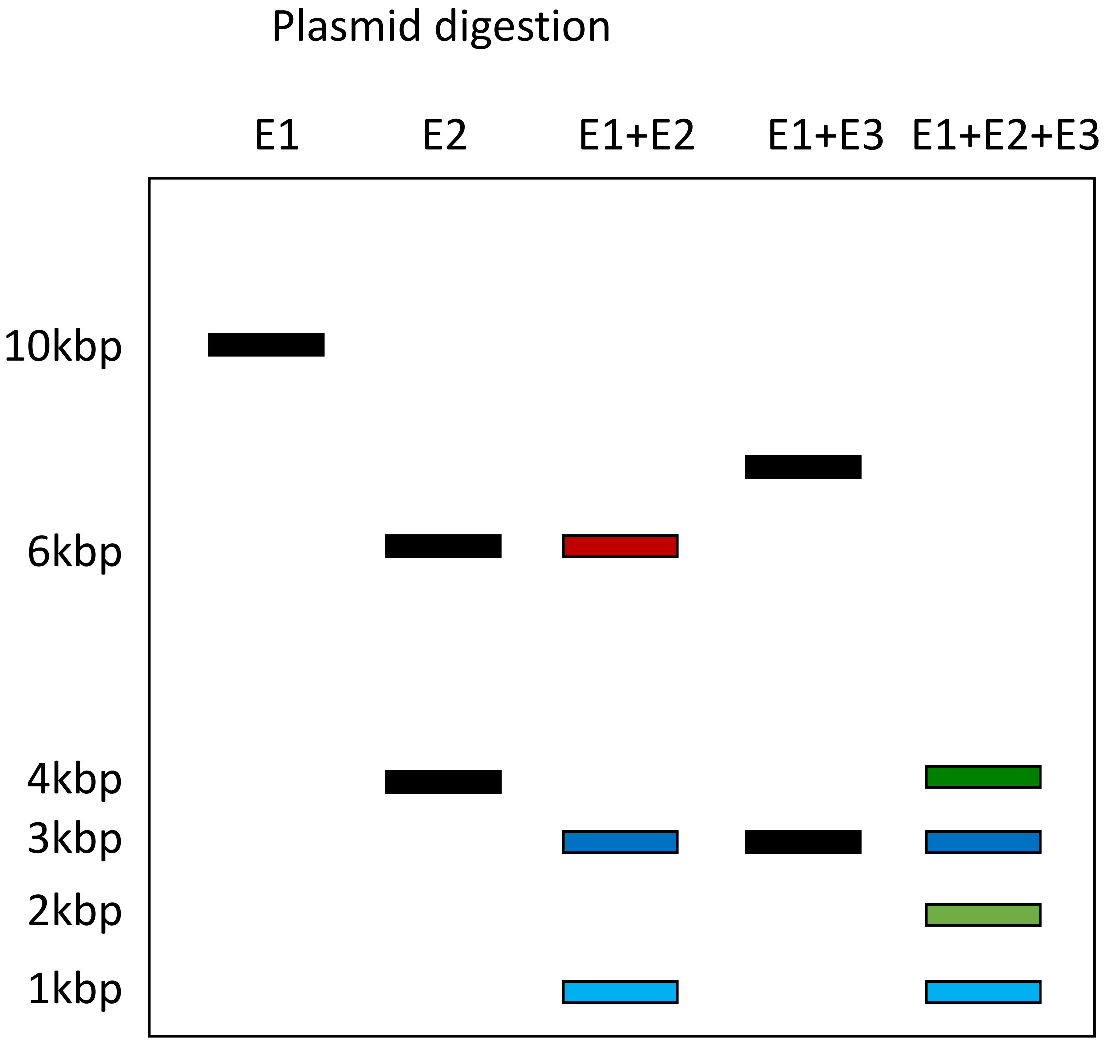
Your turn…

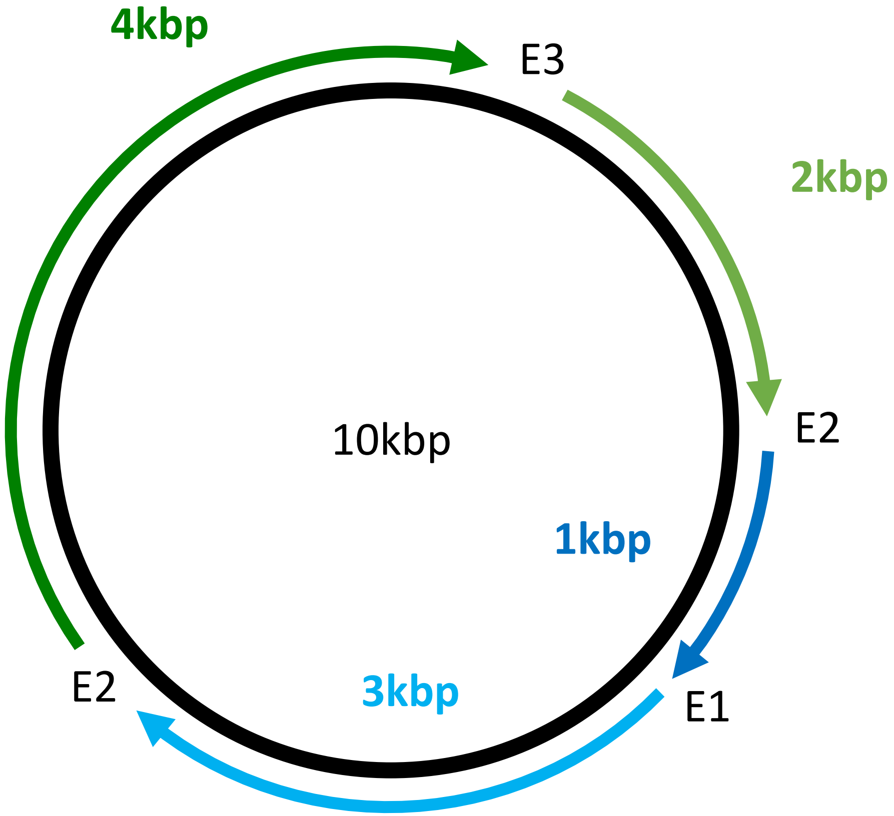
Plasmid digestion



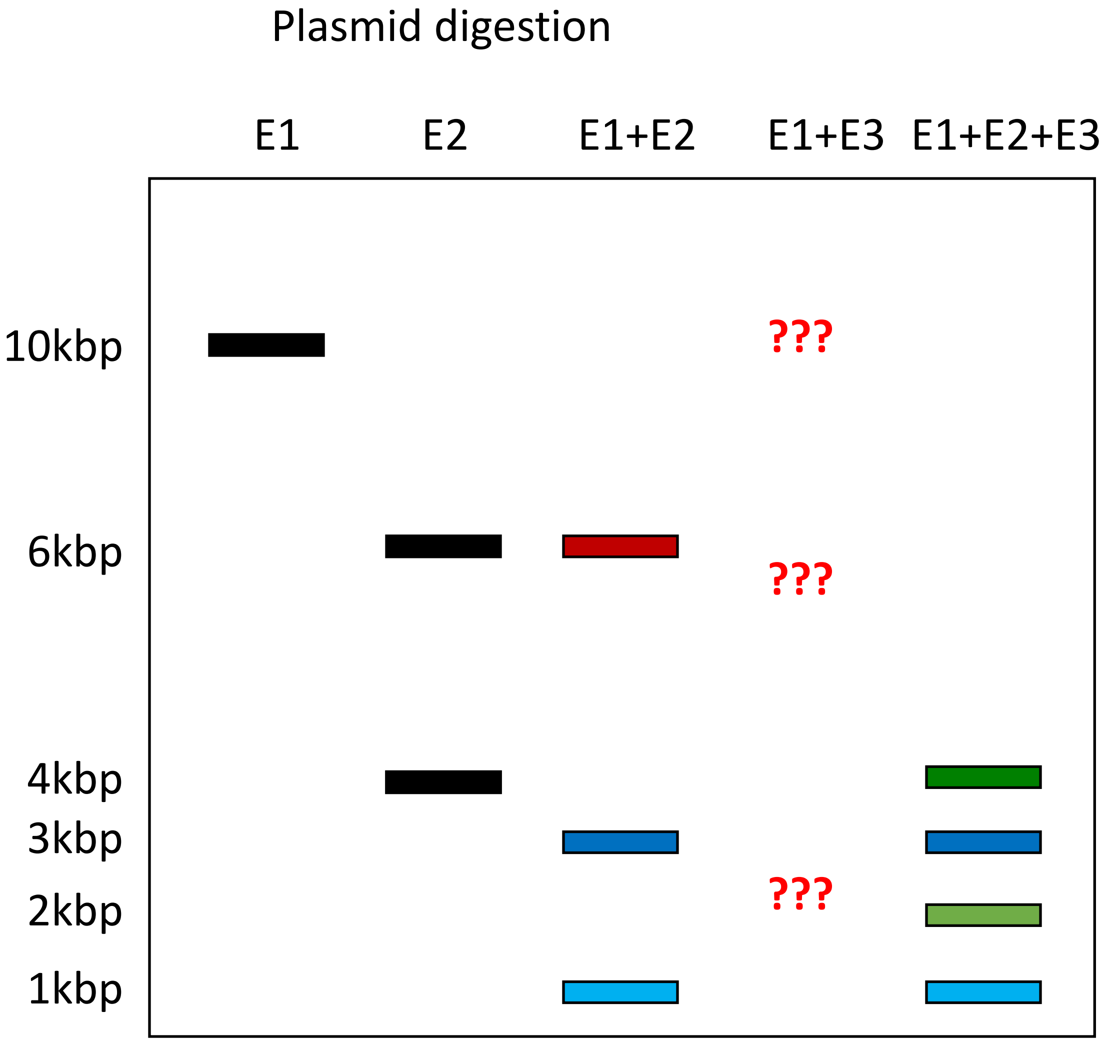


Your turn...

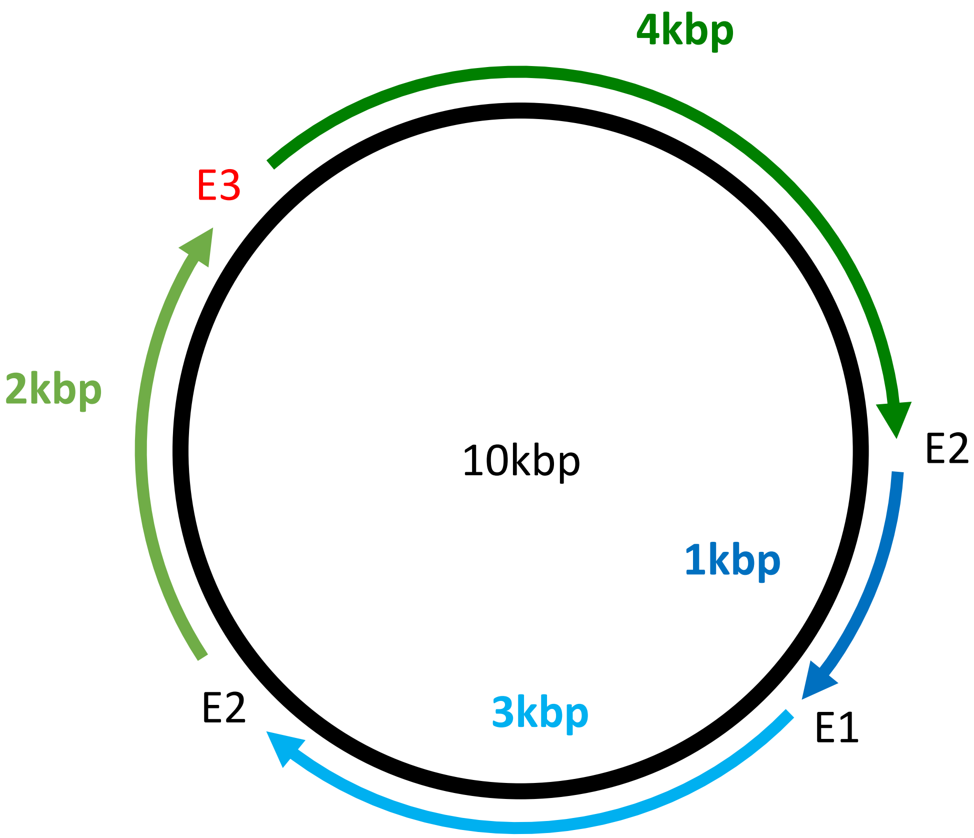

Plasmid digestion
E1
E2
E1+E2
E1+E3
E1+E2+E3
10kbp
6kbp
4kbp
3kbp
2kbp
1kbp


4kbp
E3
2kbp
E2
10kbp
1kbp
E2
3kbp
E1

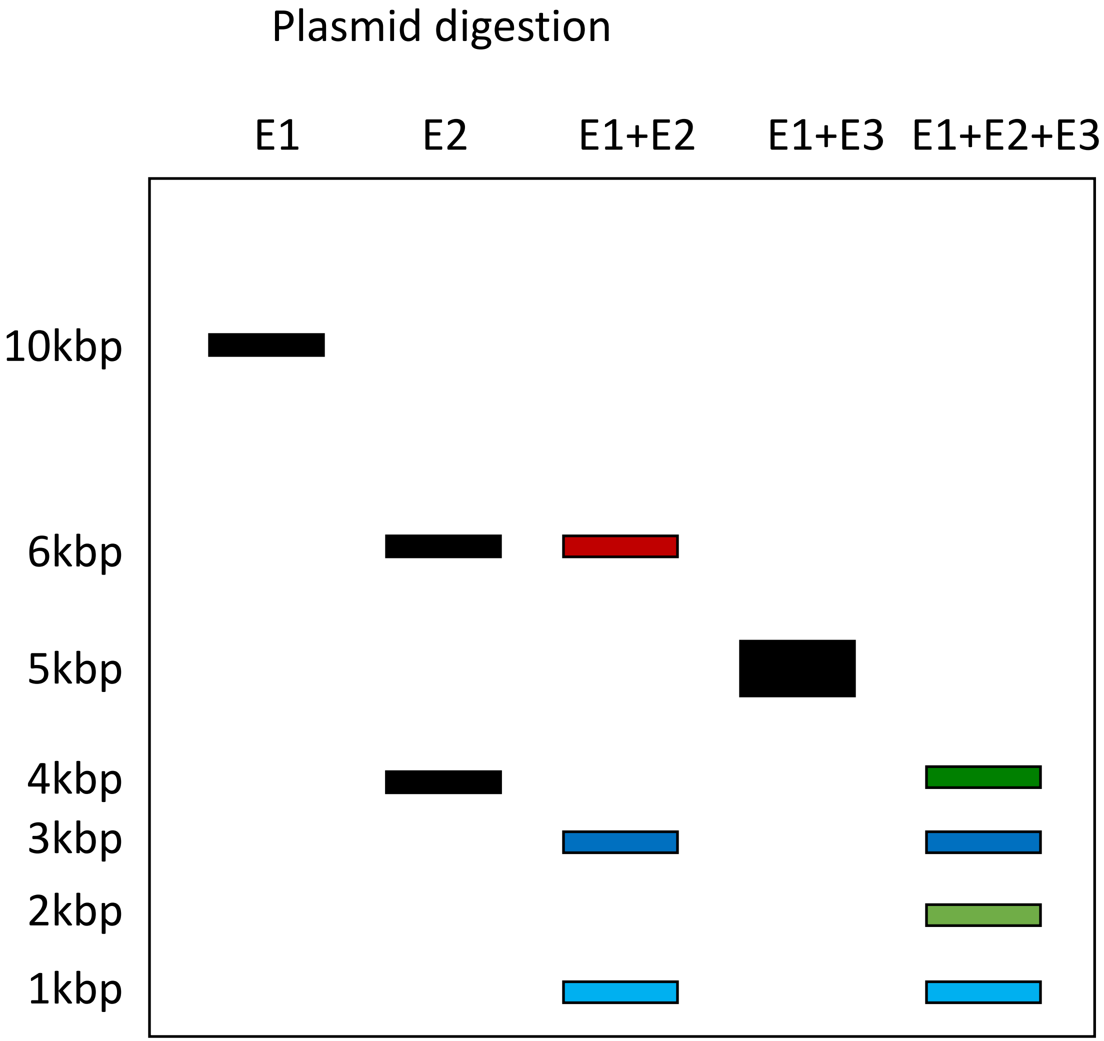
Your turn...



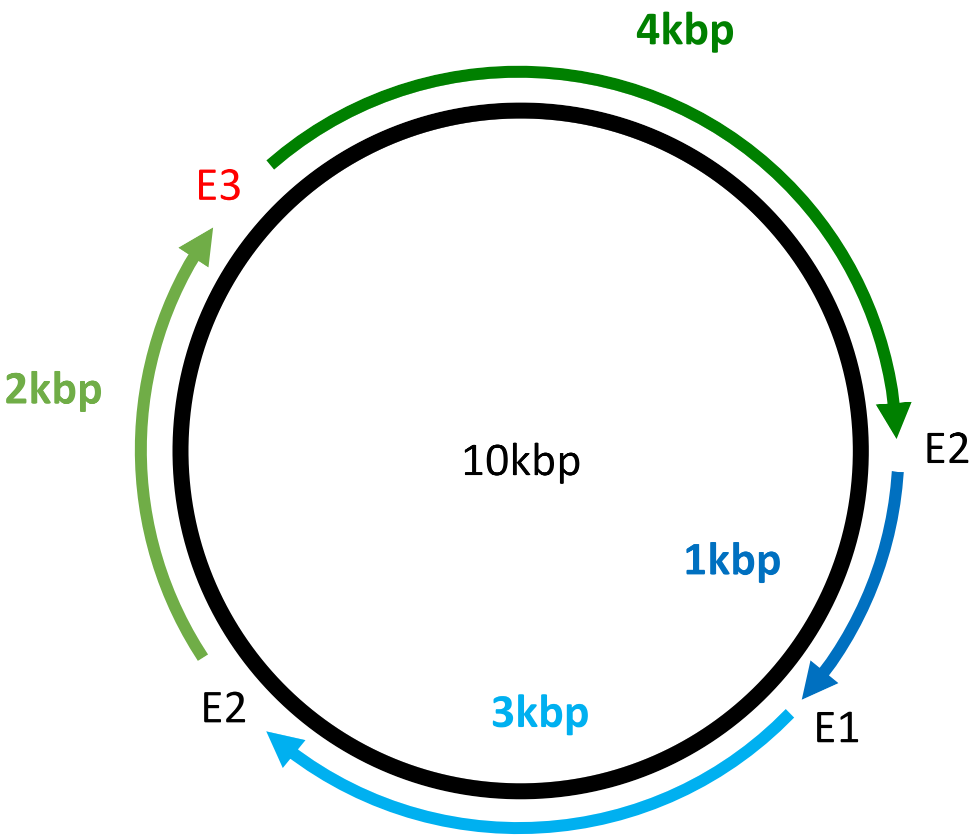
E1+E3 = ???



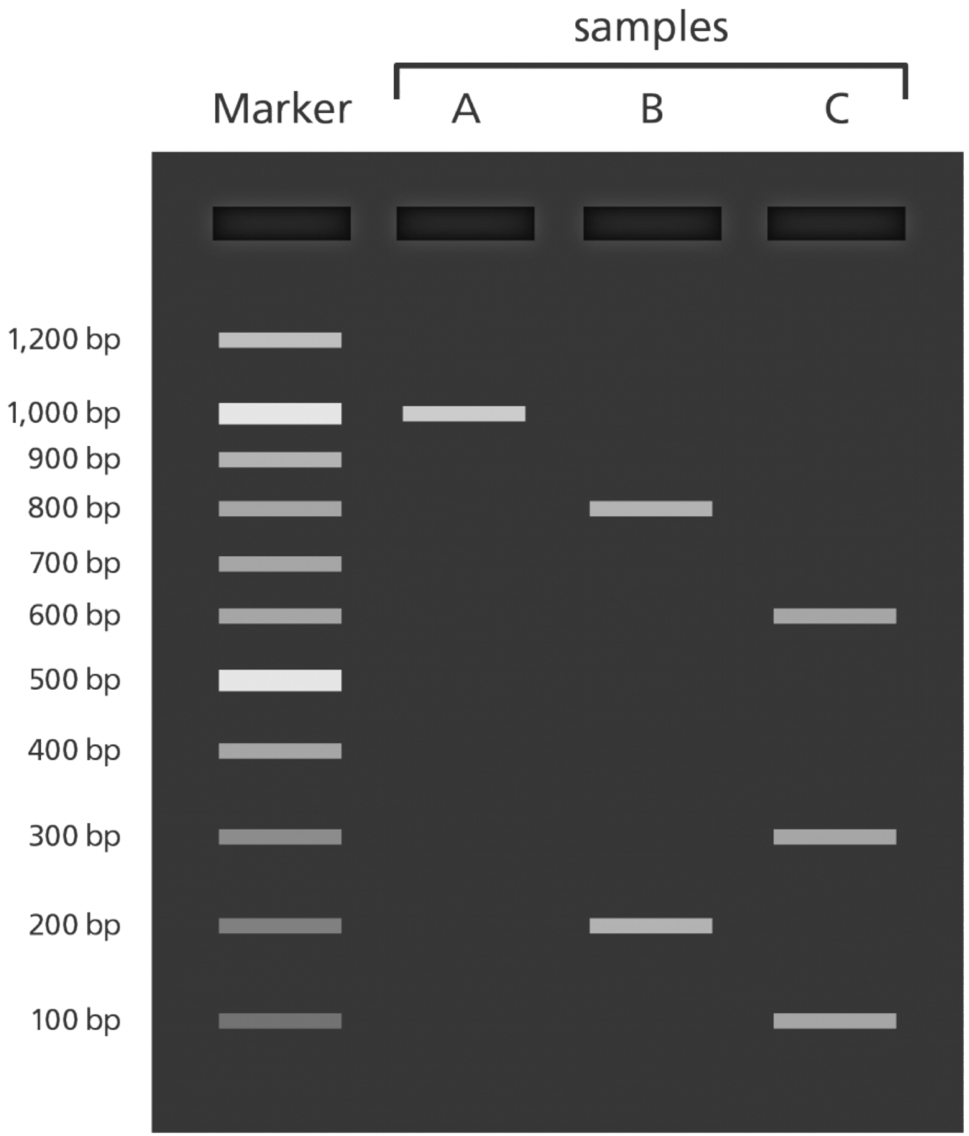
Your turn…



E1+E3 = ???



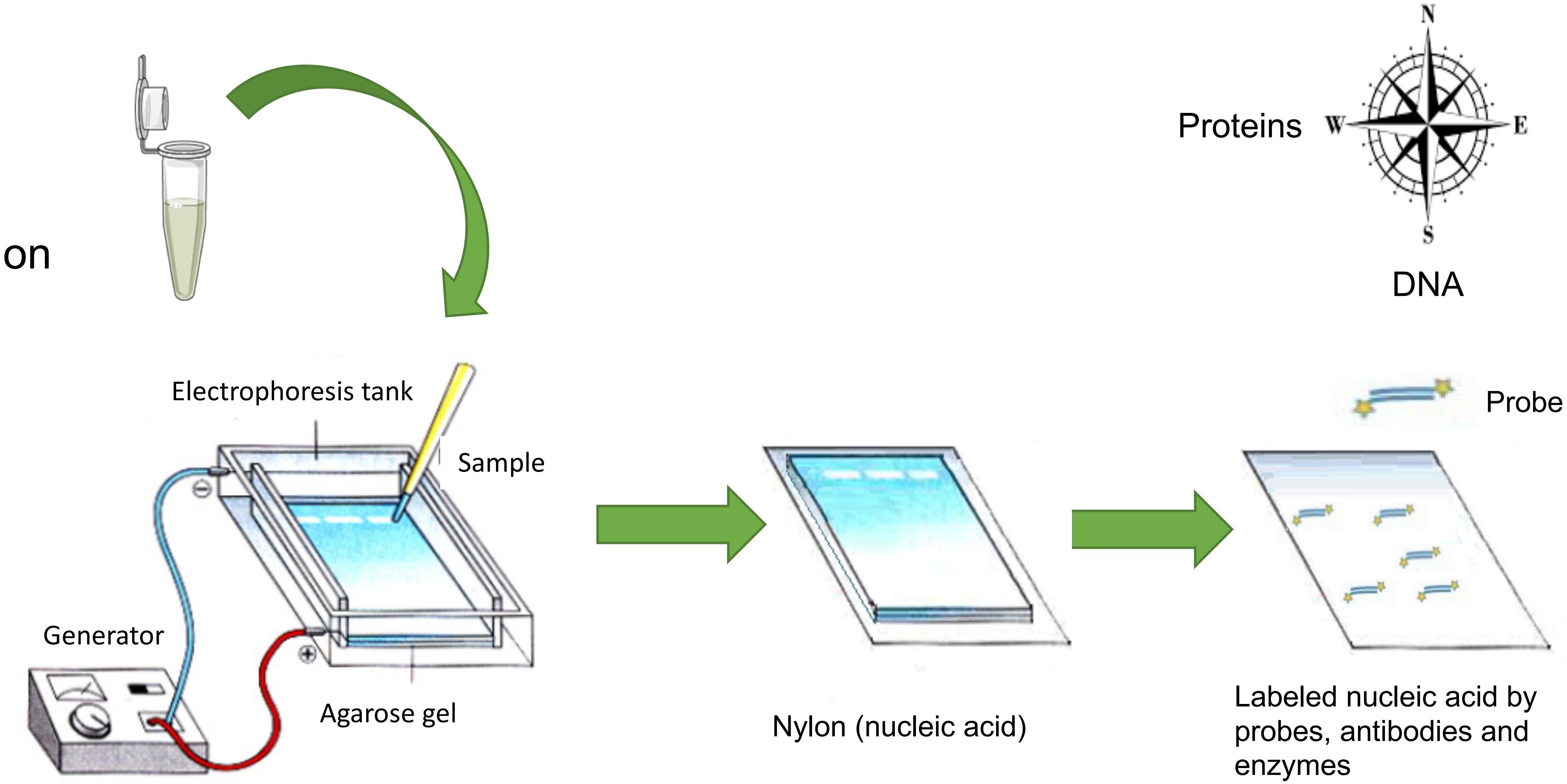
Pause…



True of False

- Condition C can contain only 1 enzymes
- Condition C can be a combination of 2 enzymes
- Condition C can be a combination of 3 enzymes
- Condition C can be a combination of A+B
- The same DNA was potentially used in A, B and C
- The DNA can be circular
- The DNA can be linear
- If the DNA is linear the enzyme in A doesn't cut
- If the DNA is circular, enzymes cut 2 times in B
- If the DNA is circular, enzymes cut 3 times in C

Blotting techniques



RNA

Proteins

DNA

Sample preparation

Electrophoresis tank

Sample

Generator

Agarose gel

Probe

Nylon (nucleic acid)

Labeled nucleic acid by probes, antibodies and enzymes

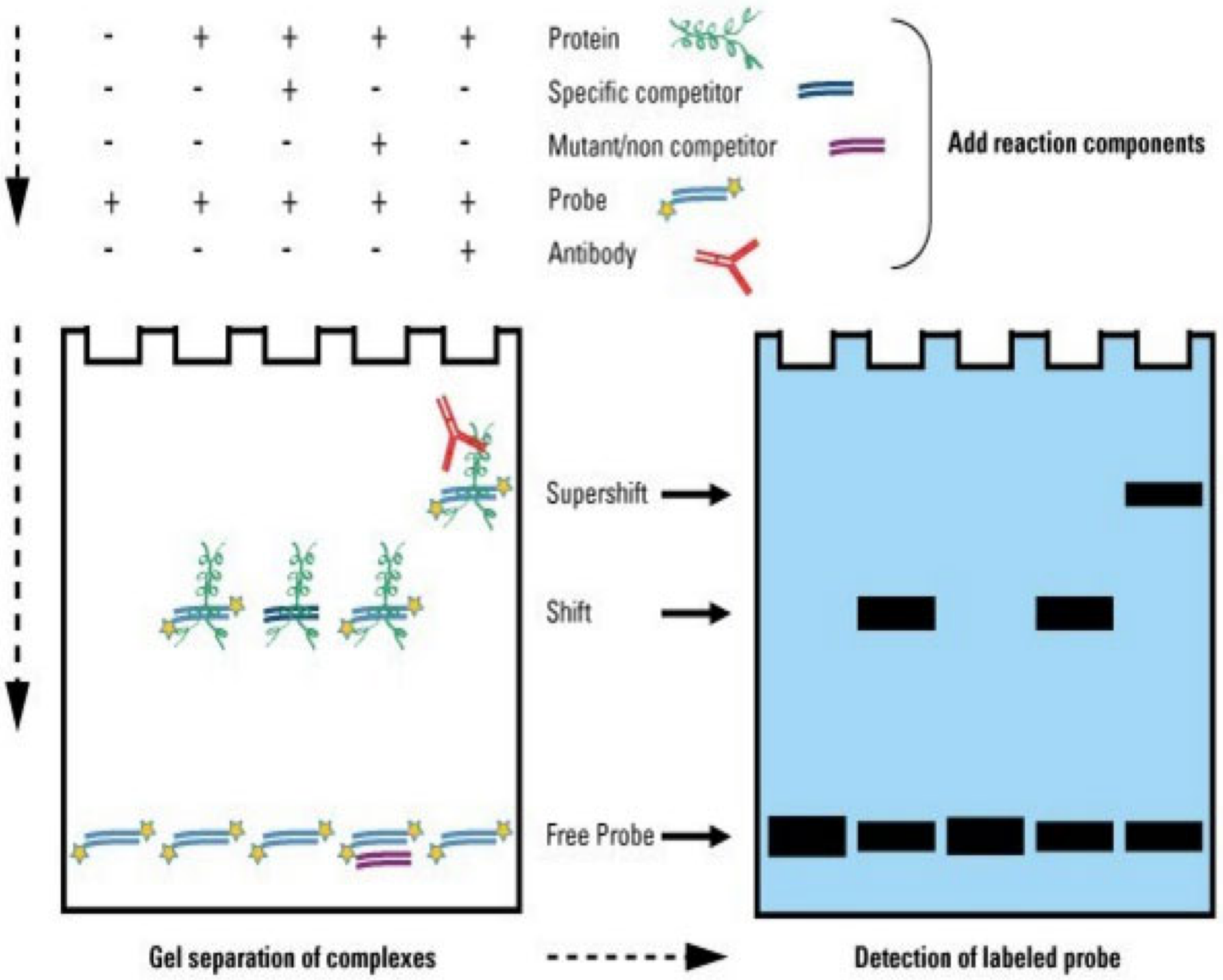
Electrophoresis

Membrane transfer

Revelation

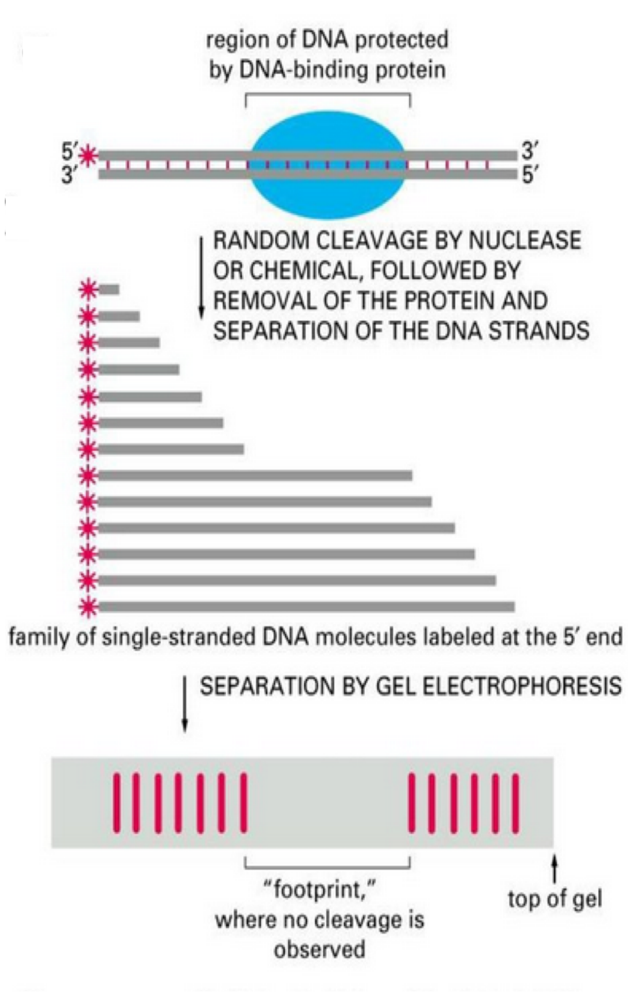
Electrophoretic Mobility Shift Assay (EMSA)

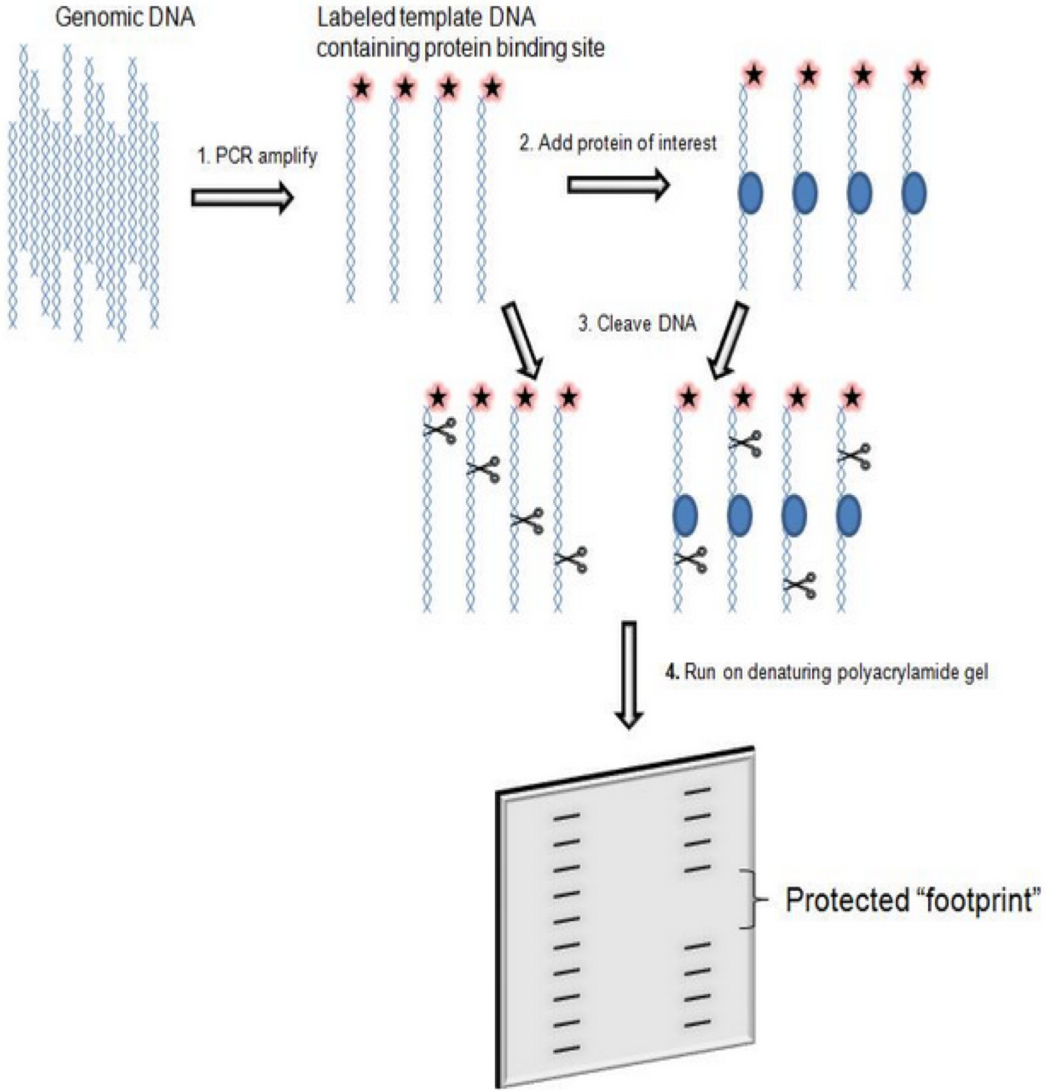
☞ Study of protein-DNA and protein-RNA interactions



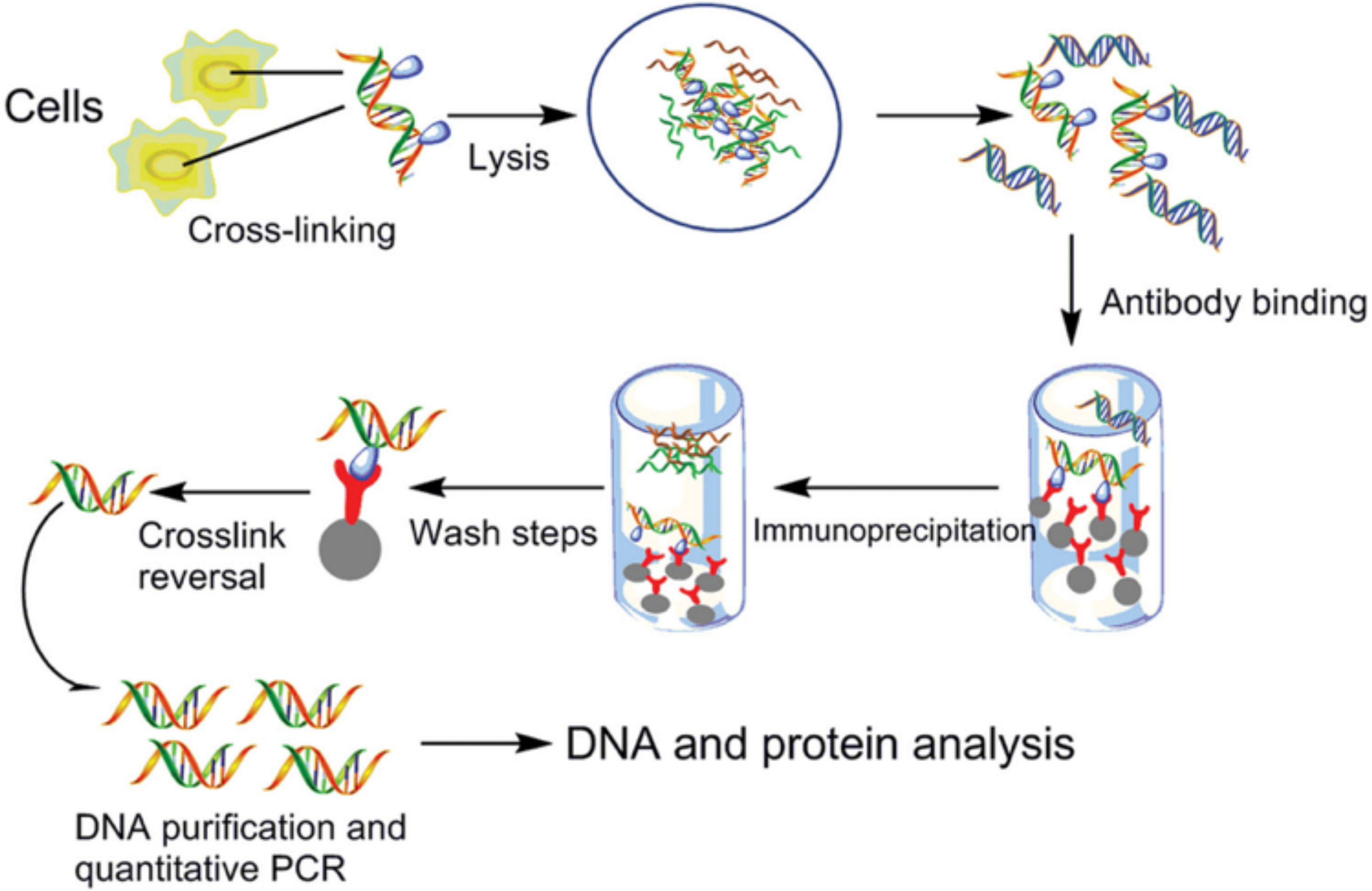
DNA footprinting

- Characterization of cis-acting sequences (DNA-protein interactions)

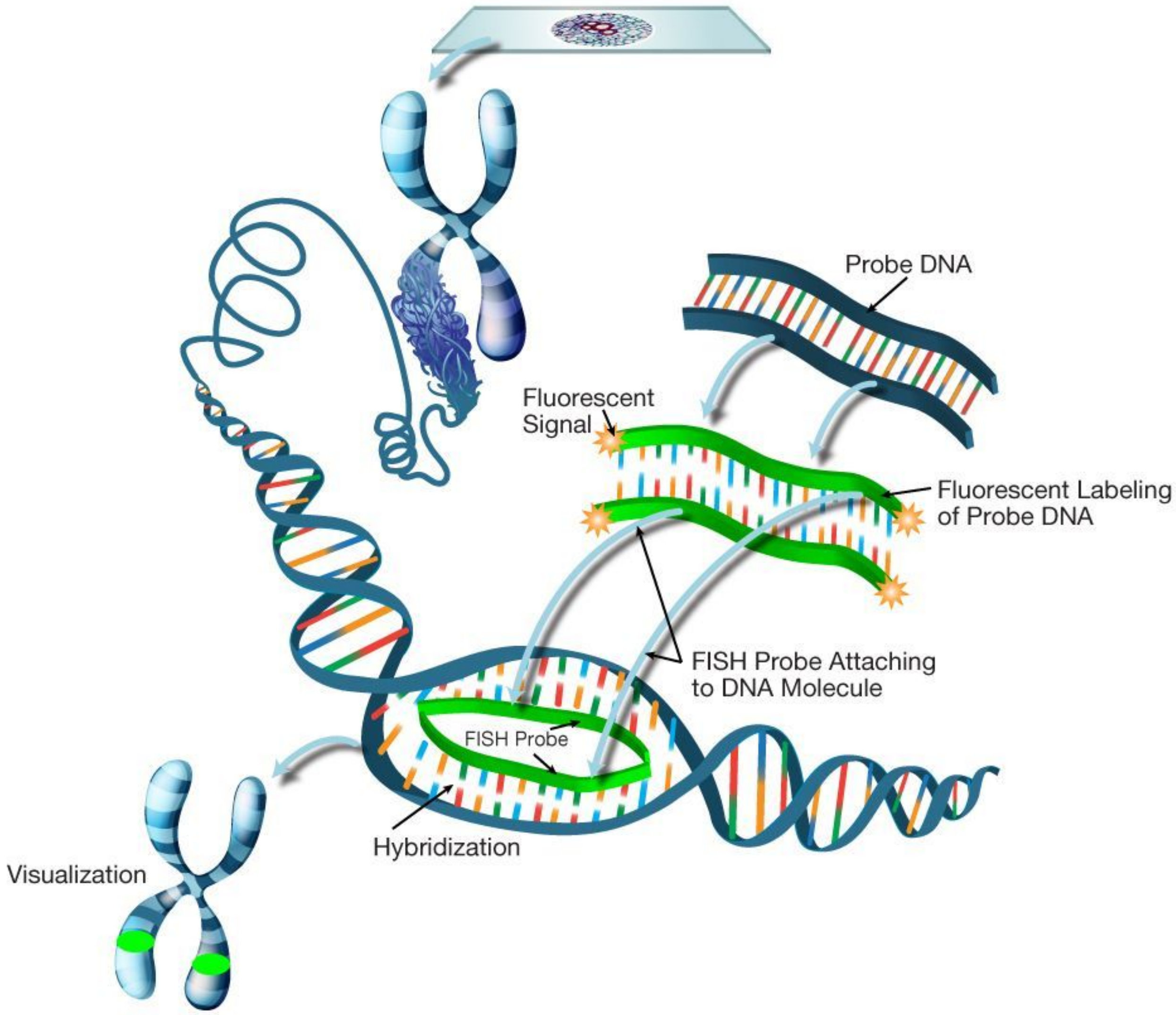


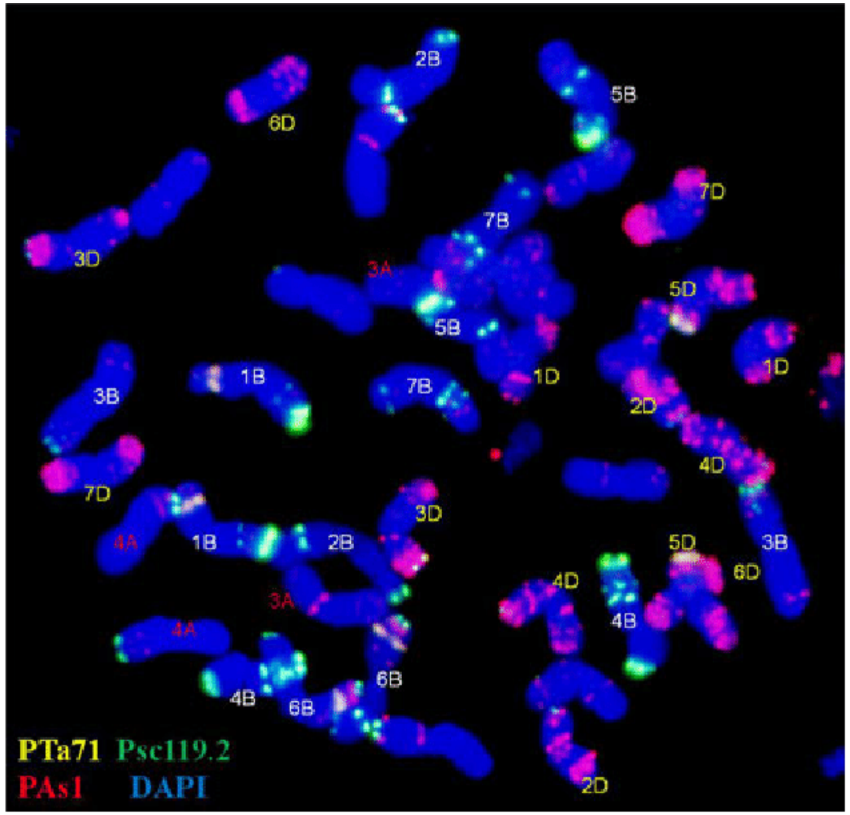


ChIP and ChIP-Seq

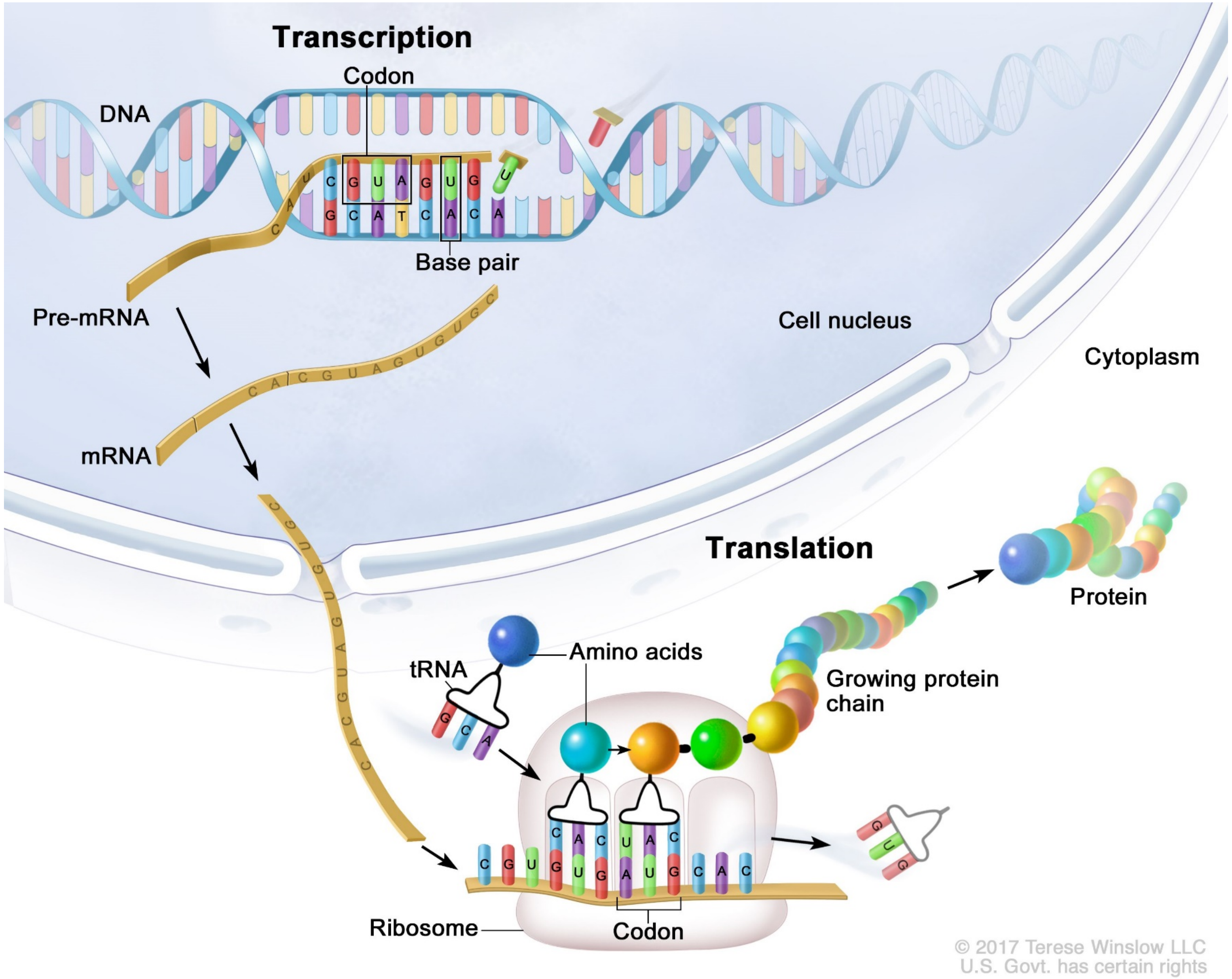


FISH : Fluorescence in situ hybridization





Plan



© 2017 Terese Winslow LLC
U.S. Govt. has certain rights

I - DNA and genome

- Nucleic acids: generalities
- DNA structure
- Organization of genomes

II- Transcription: from DNA to RNA

- Basic mechanism
- Maturation of mRNA

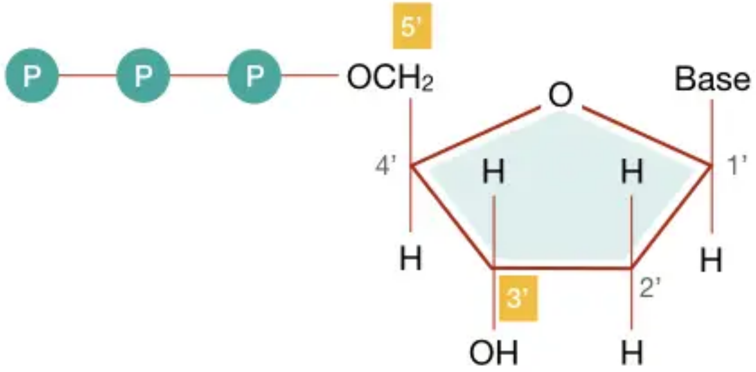
III- Translation: from RNA to proteins

IV- Molecular tools in the study of DNA

- DNA extraction
- Enzymes used in molecular biology
- Electrophoresis
- DNA sequencing

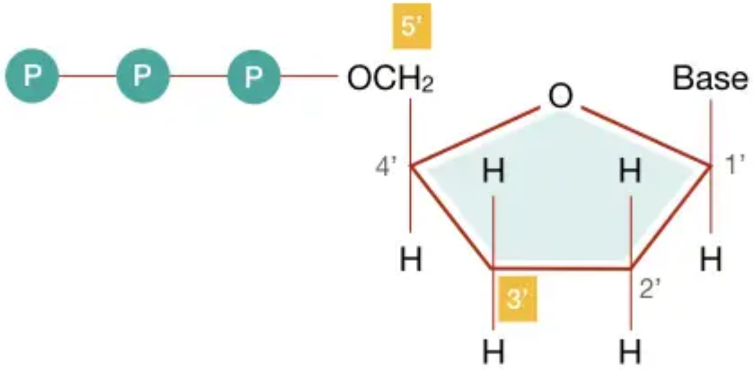
DNA sequencing using the Sanger method

PCR in presence of fluorescent, chain-terminating nucleotides



dNTP- deoxynucletide

© Genetic Education Inc.



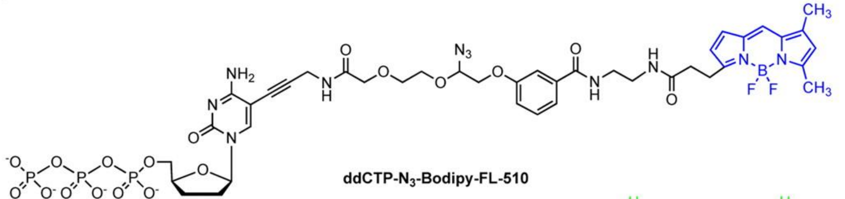
ddNTP- dideoxynucletide

DNA synthesis in the presence of didesoxynucleotides (= ddNTP)

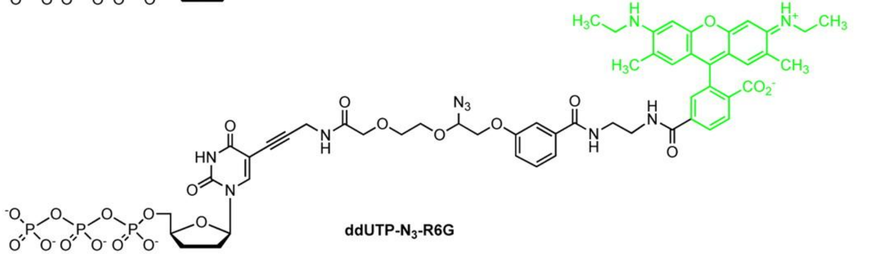
ddNTP can be incorporated into the chain during synthesis
> stop of the synthesis (chain-terminating nucleotides)

DNA sequencing using the Sanger method

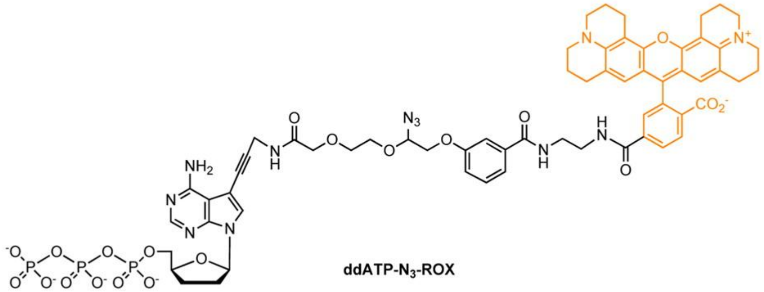
PCR in presence of fluorescent, chain-terminating nucleotides



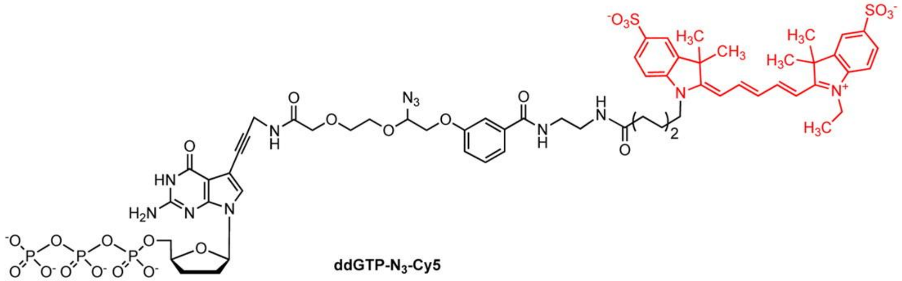
ddCTP-N_3-Bodipy-FL-510



ddUTP-N_3-R6G



ddATP-N_3-ROX



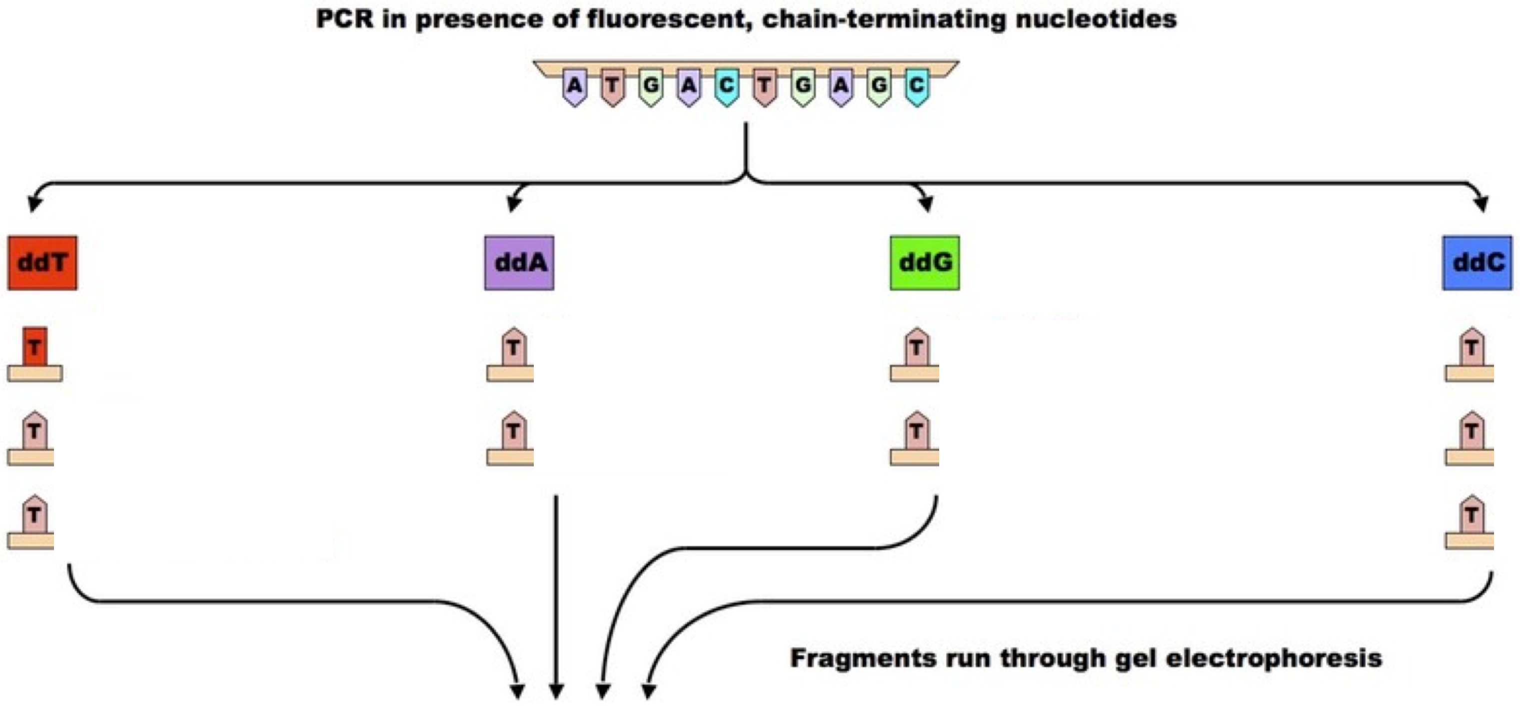
ddGTP-N_3-Cy5

DNA synthesis in the presence of didesoxynucleotides (= ddNTP)

ddNTP can be incorporated into the chain during synthesis
> stop of the synthesis (chain-terminating nucleotides)

Fluorescent ddNTP

DNA sequencing using the Sanger method (1987)



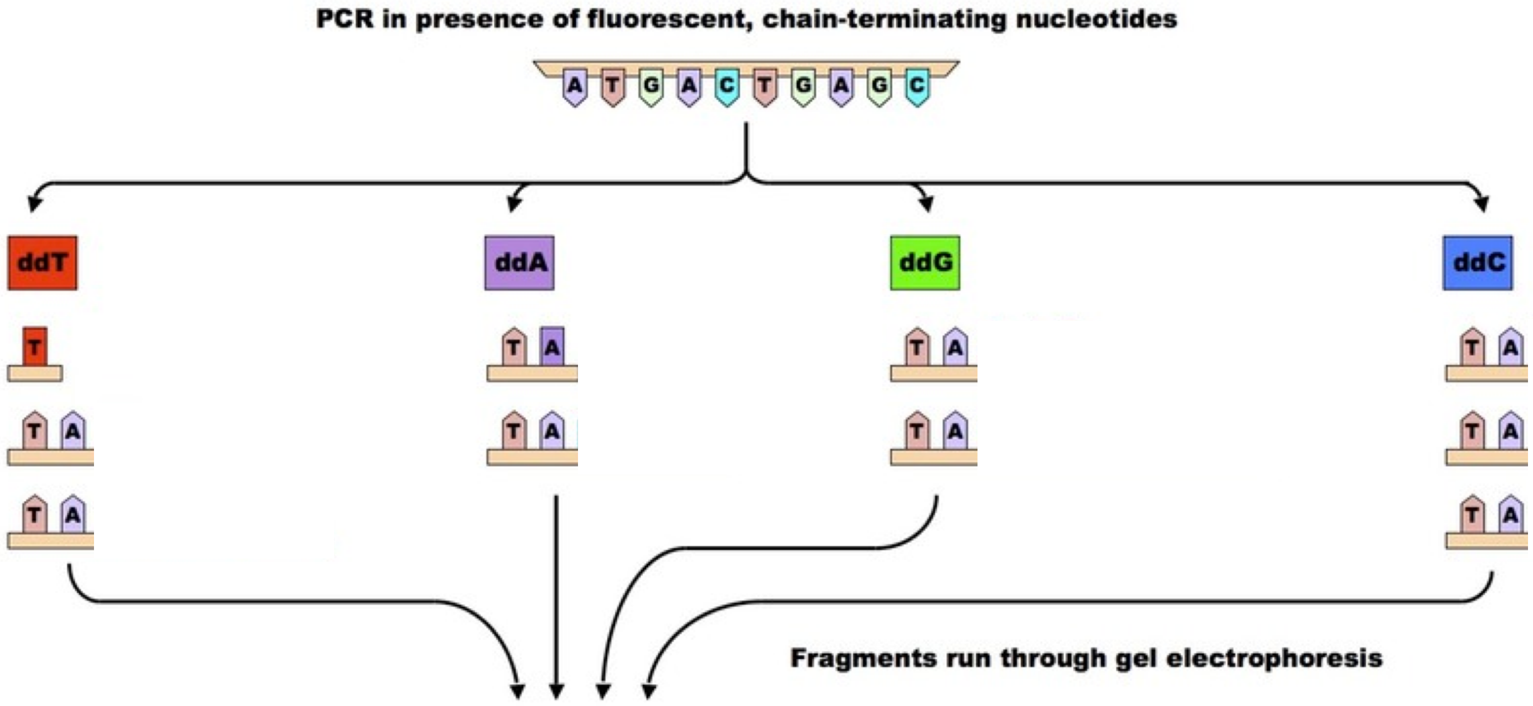
ddTTP

dTTP

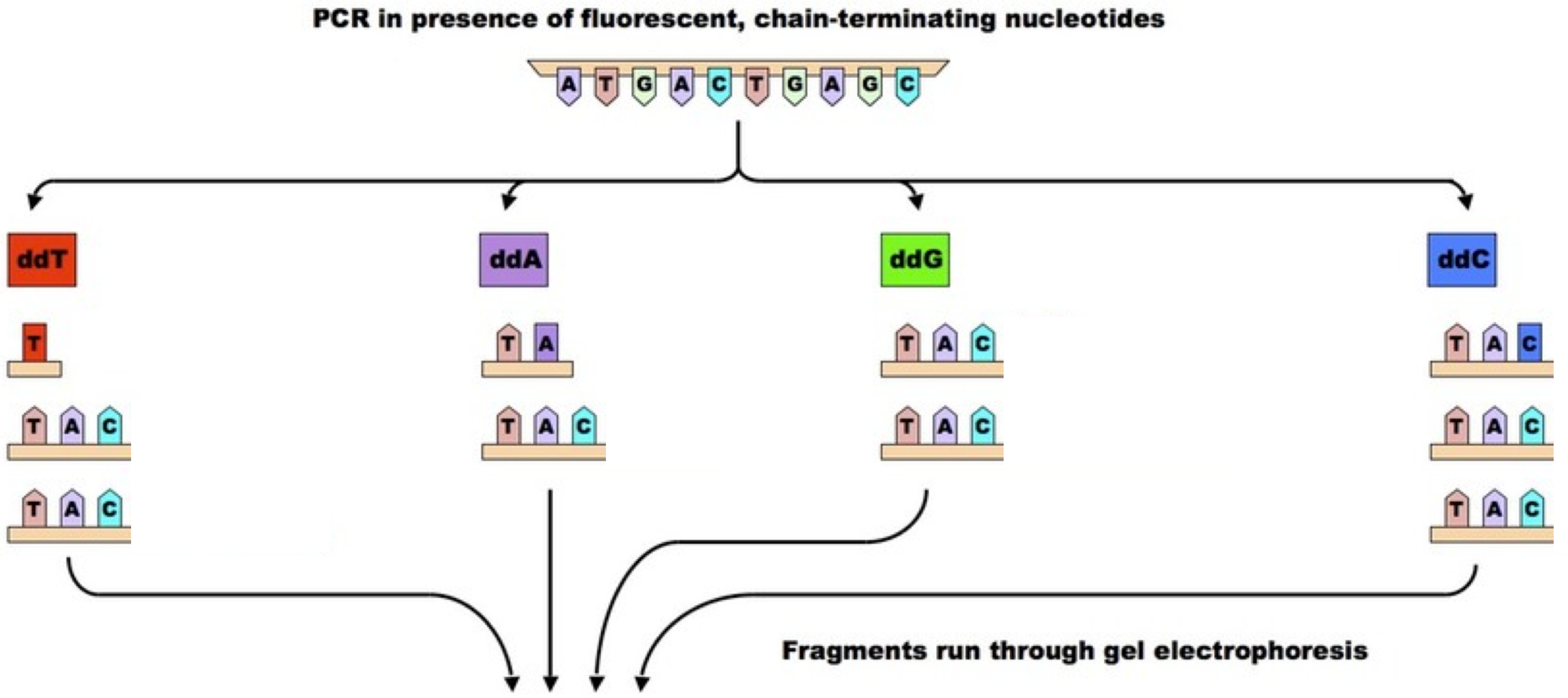
DNA sequencing using the Sanger method

ddATP

dATP



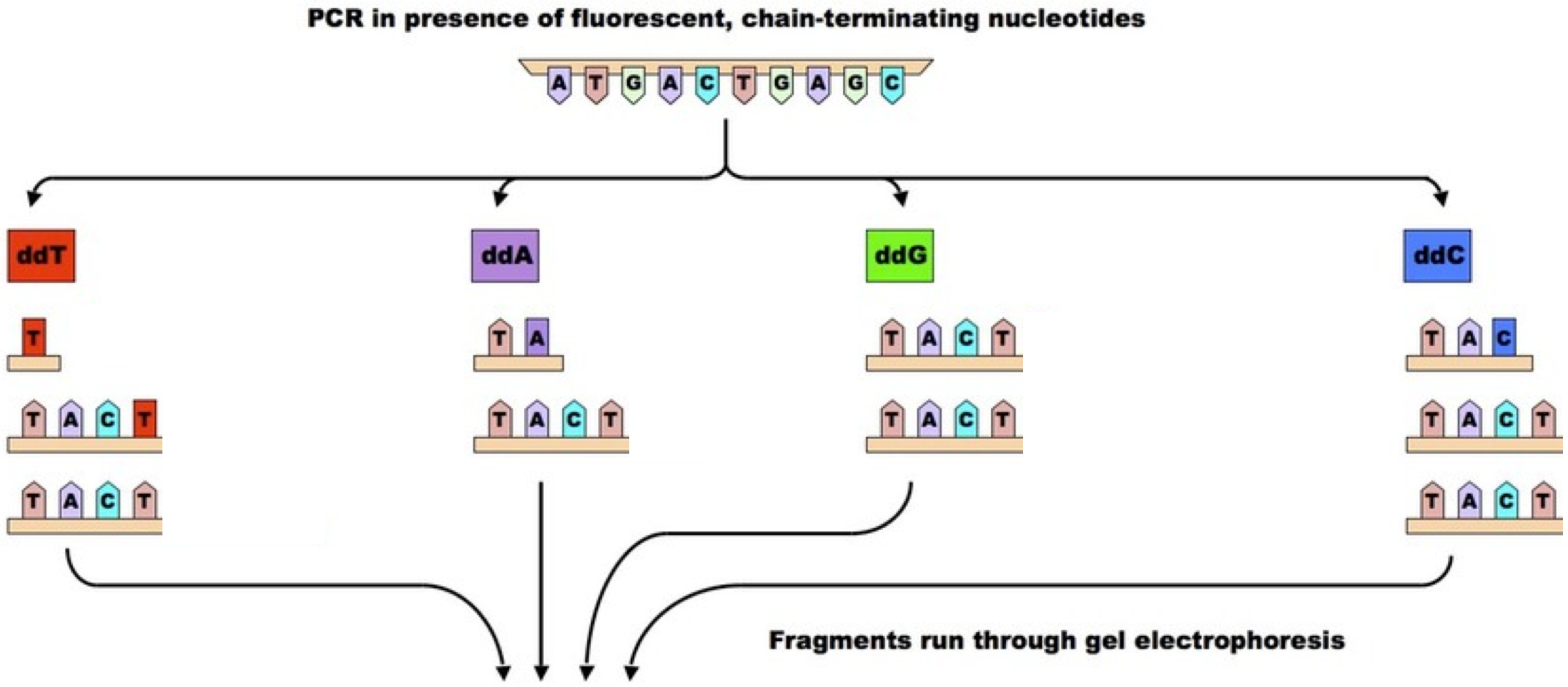
DNA sequencing using the Sanger method



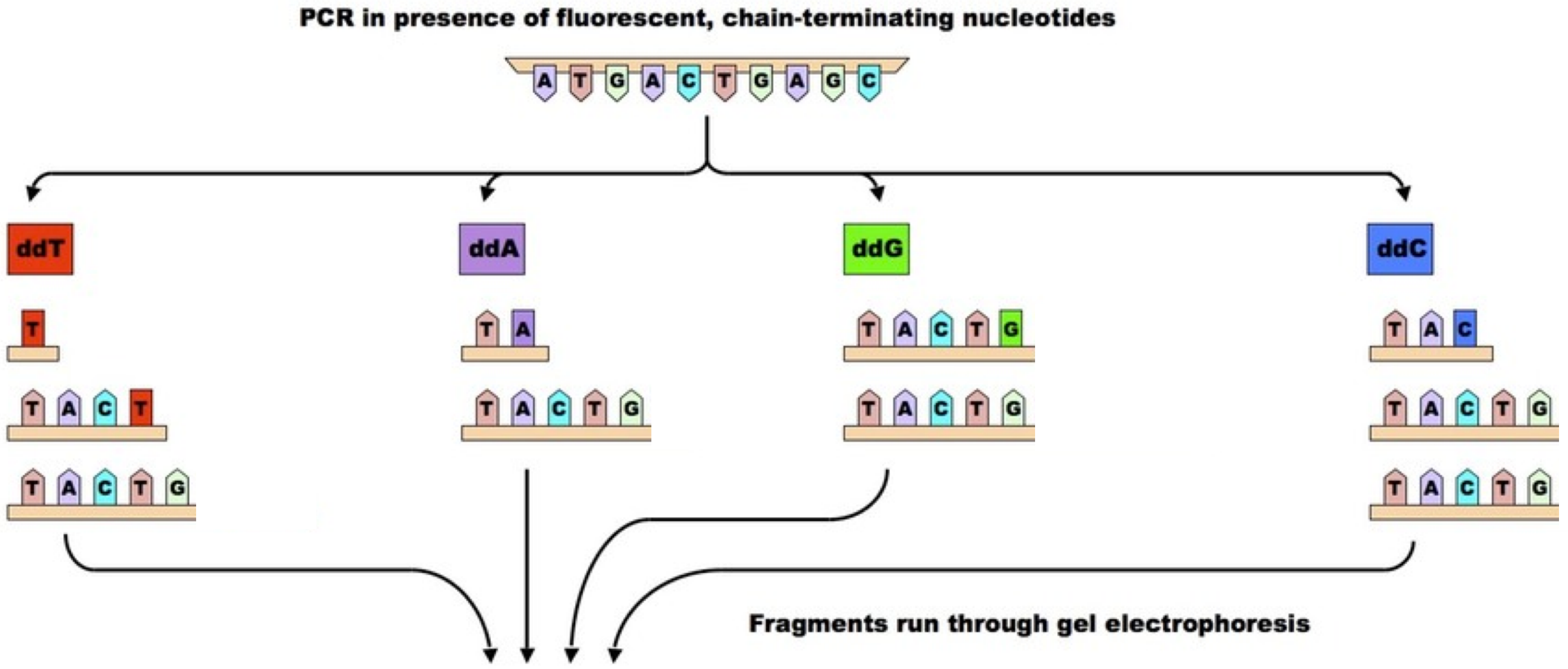
ddCTP

dCTP

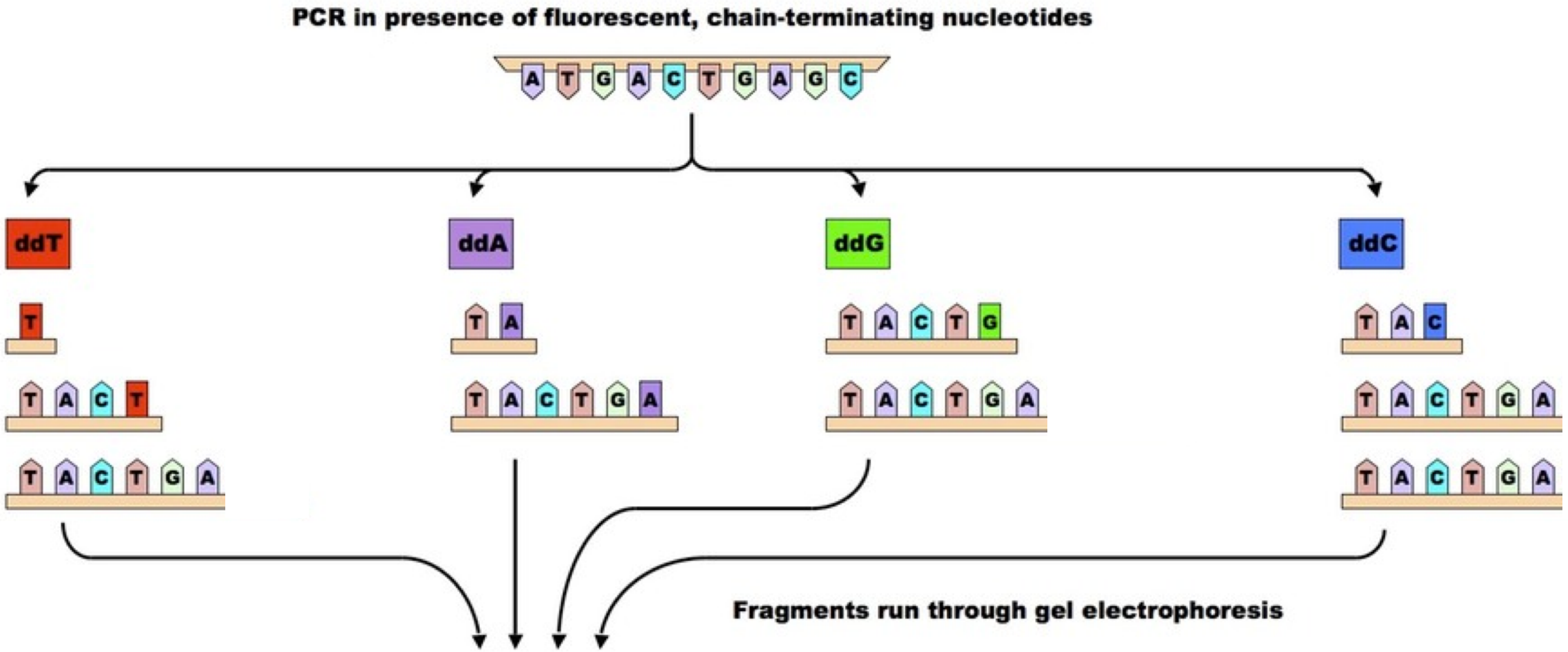
DNA sequencing using the Sanger method



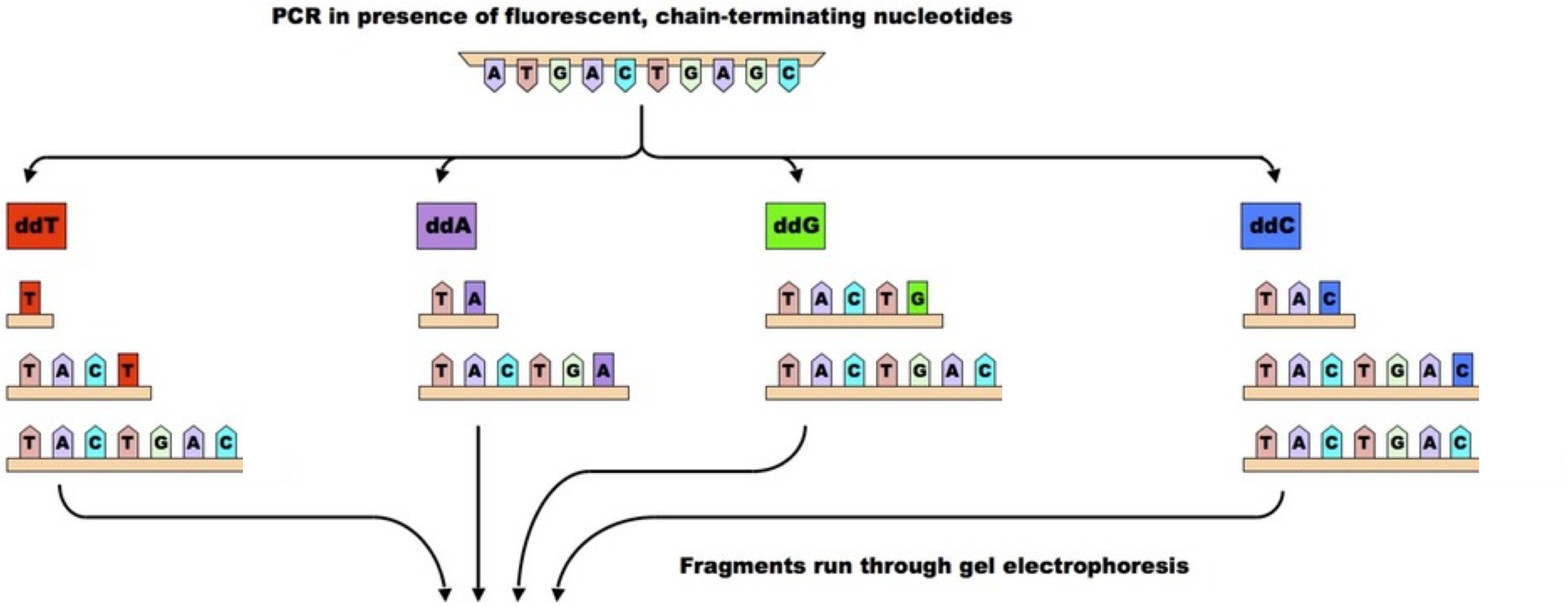
DNA sequencing using the Sanger method



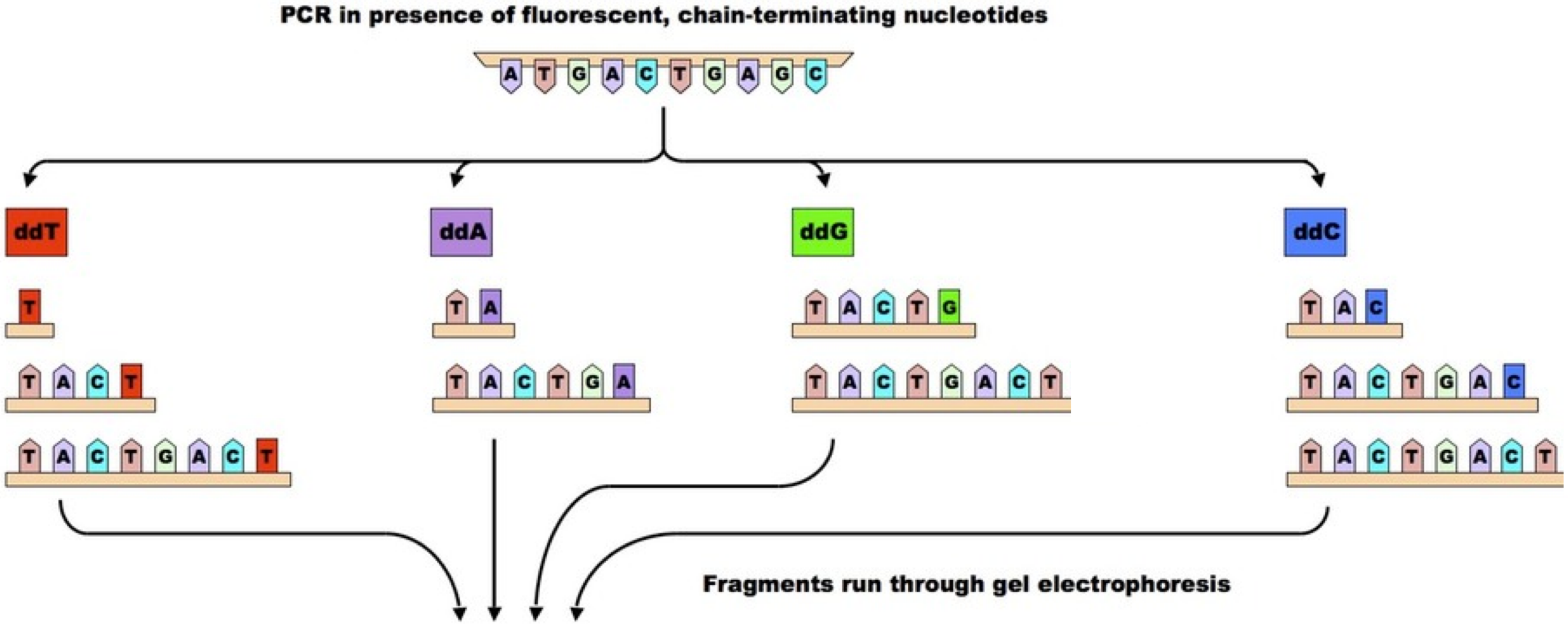
DNA sequencing using the Sanger method



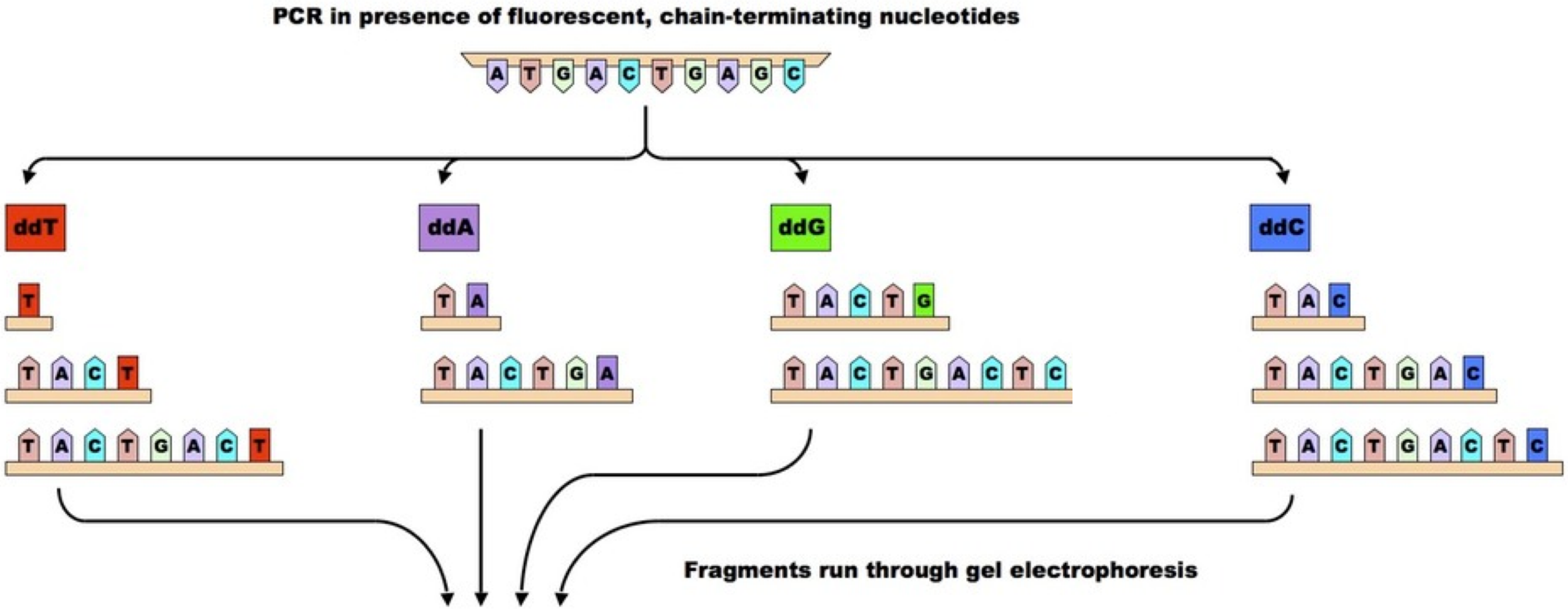
DNA sequencing using the Sanger method



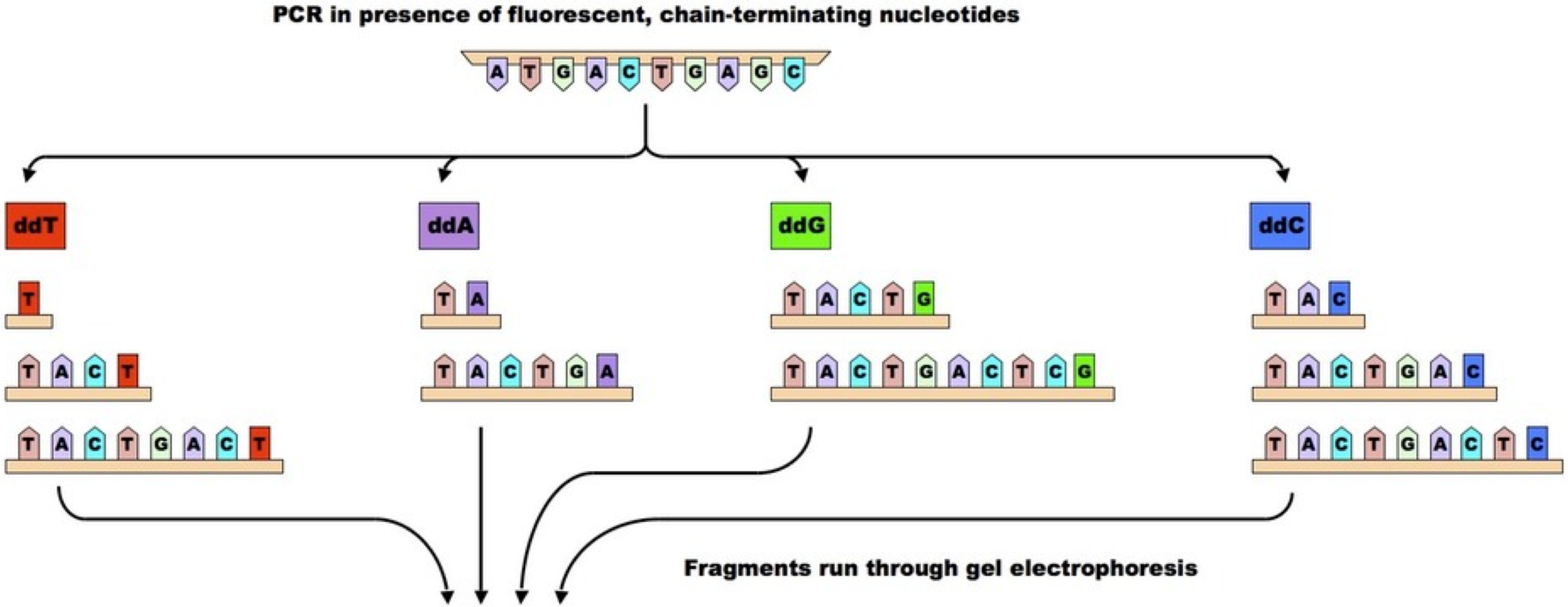
DNA sequencing using the Sanger method



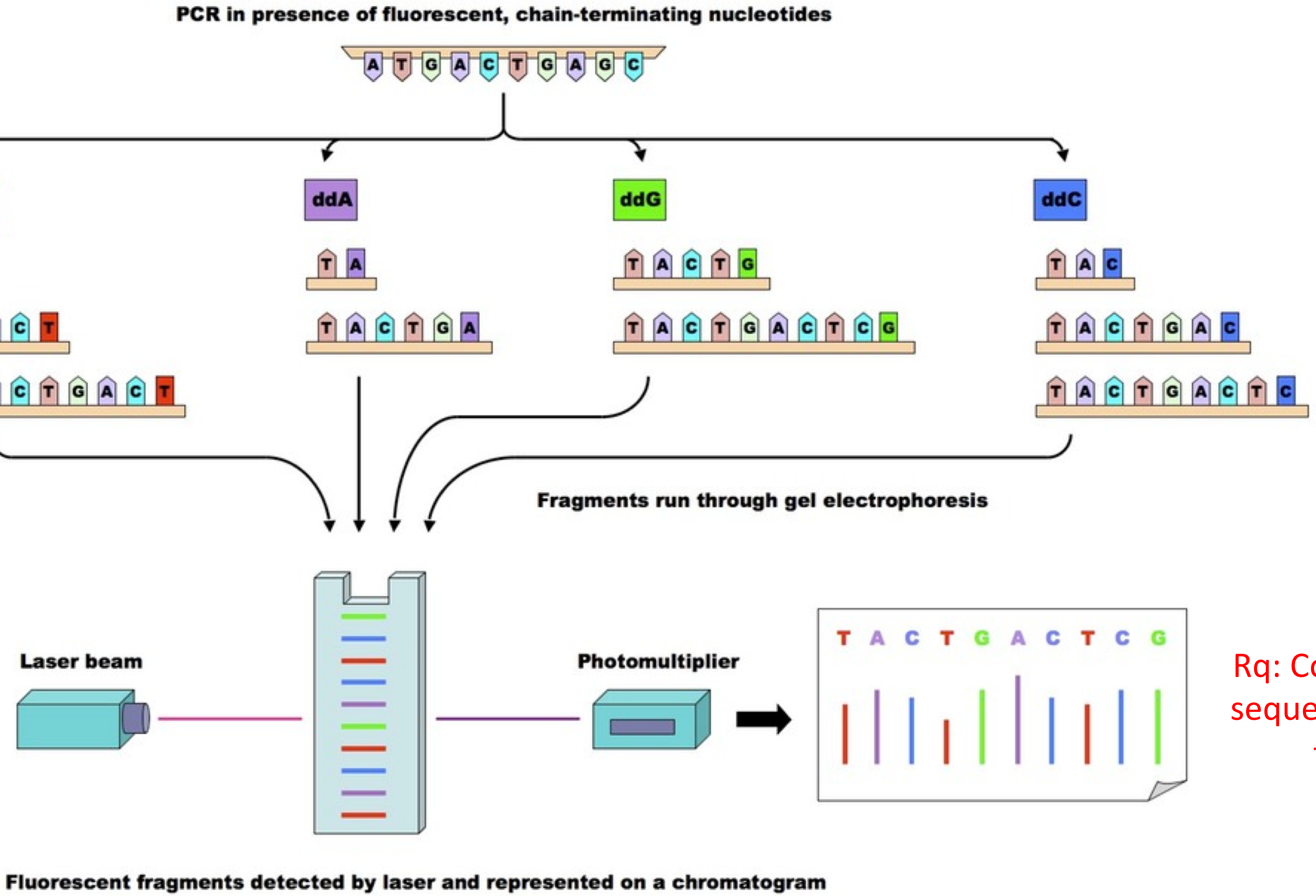
DNA sequencing using the Sanger method



DNA sequencing using the Sanger method

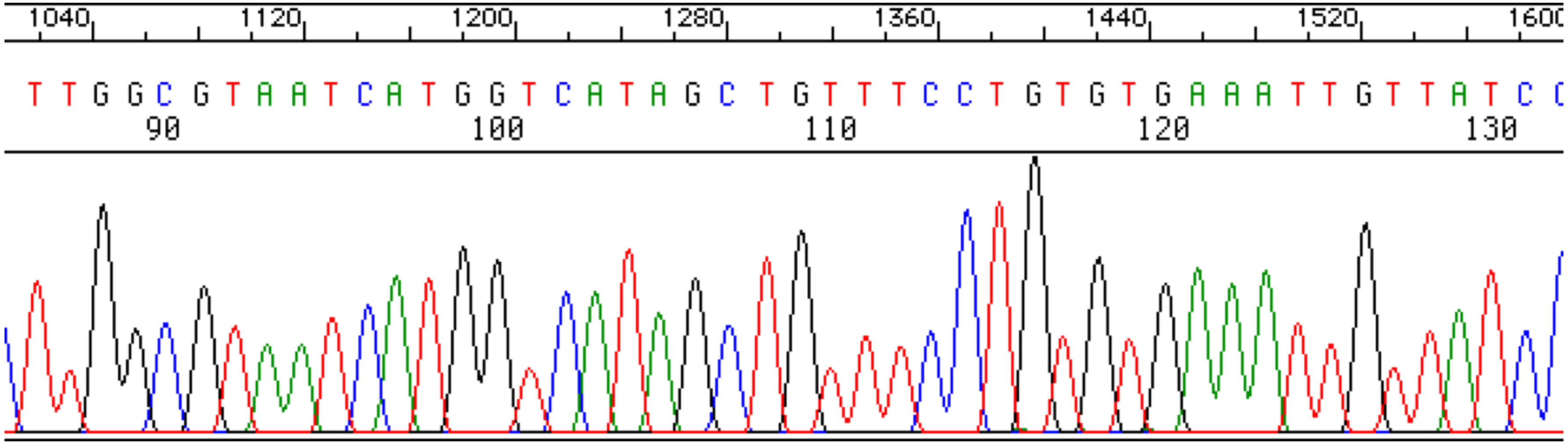


DNA sequencing using the Sanger method



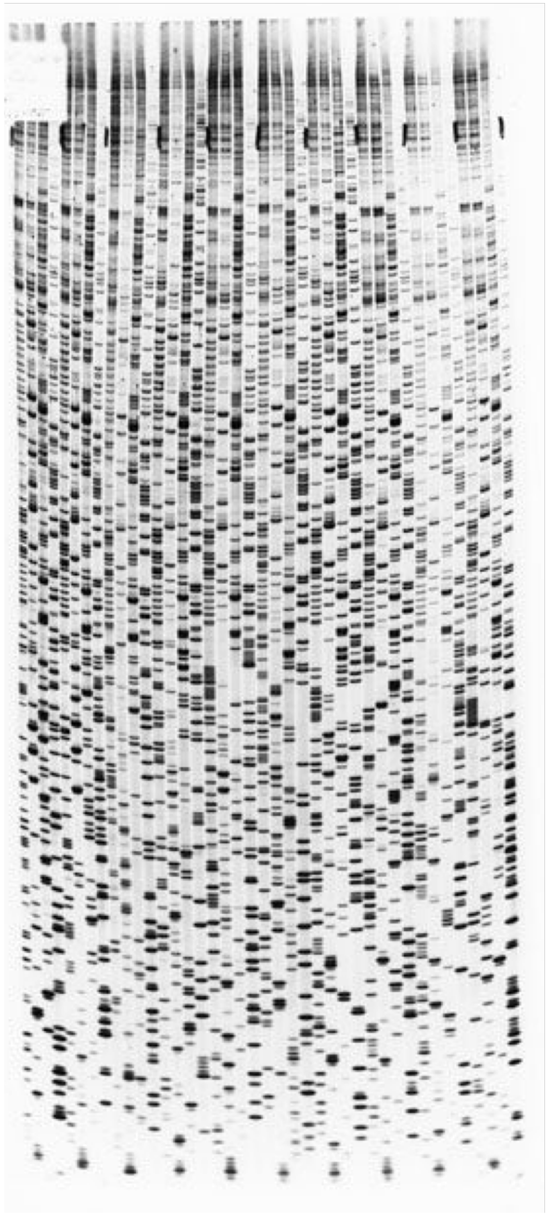
Rq: Complementary sequence compared to matrix

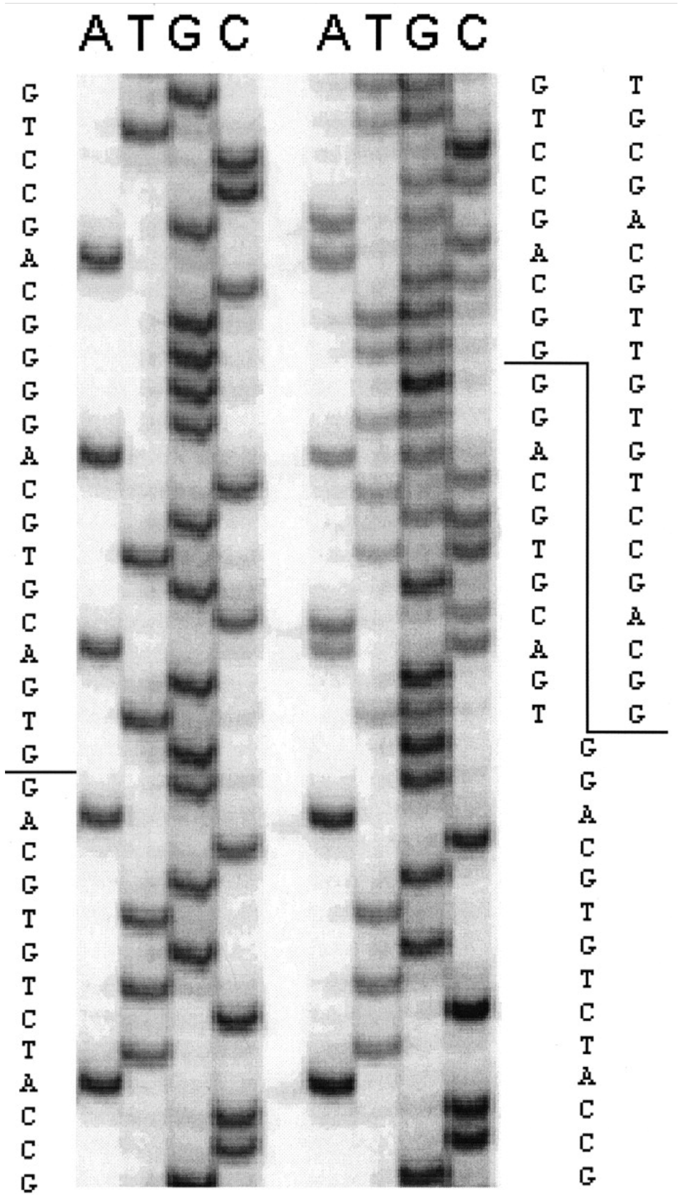
DNA sequencing using the Sanger method



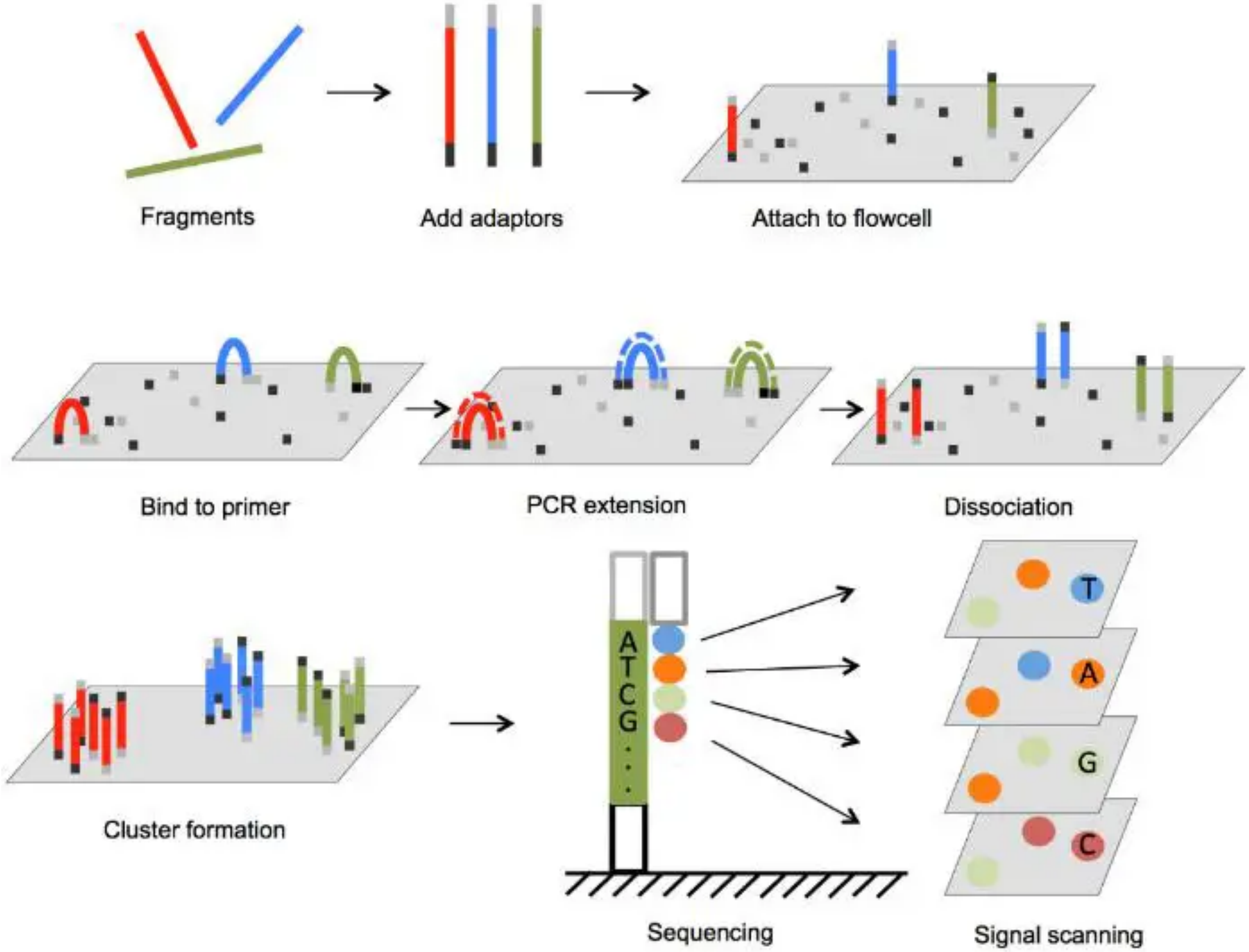
Non fluorescent Sanger method (1977)

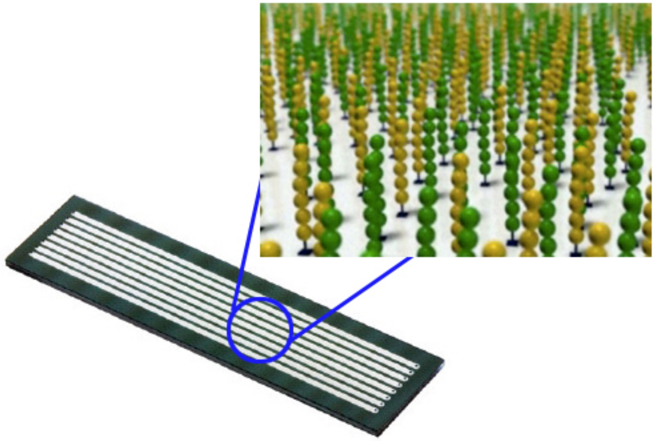
ddNTPs are radioactively labbeled



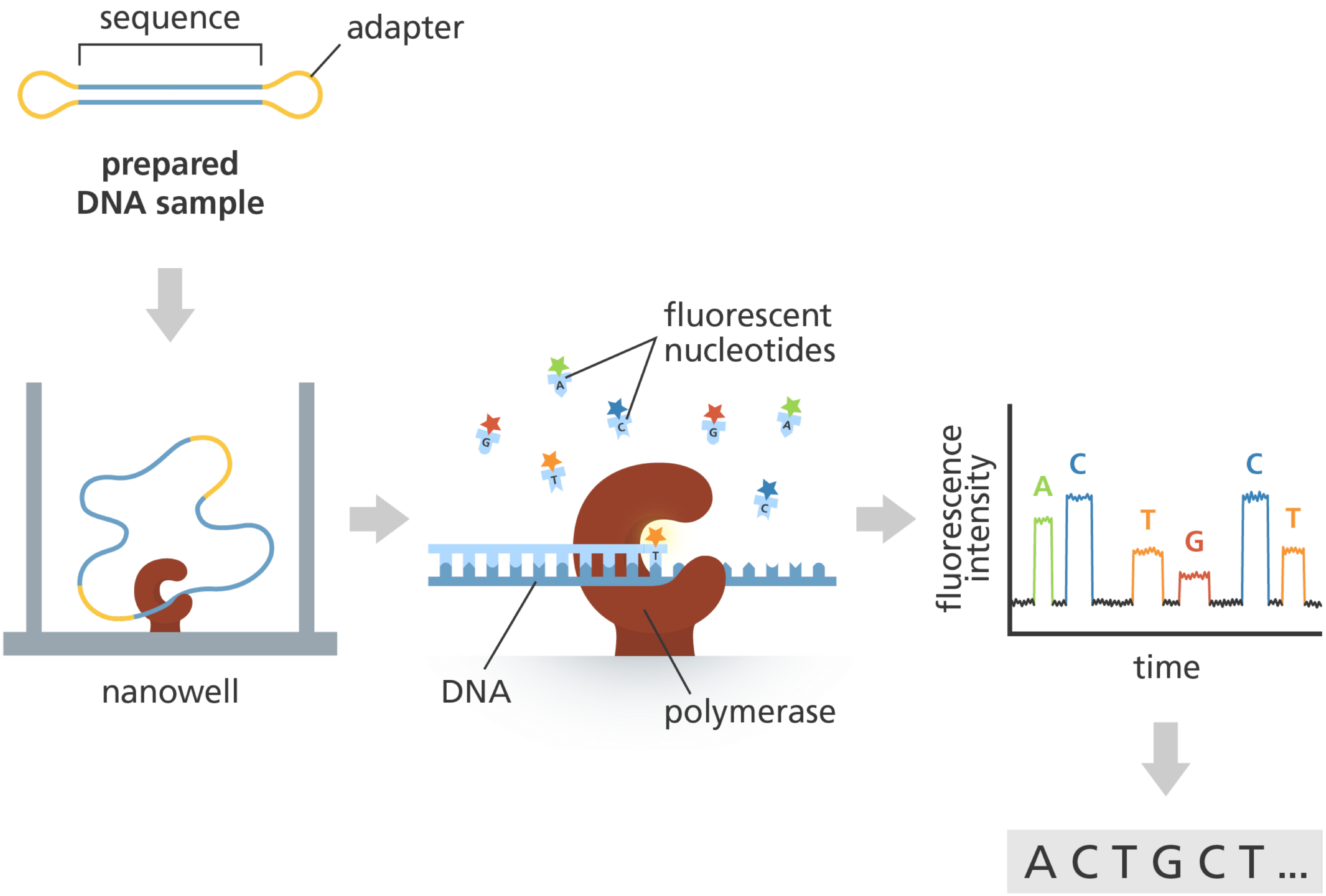


2nd Generation : Illumina technology (2007)

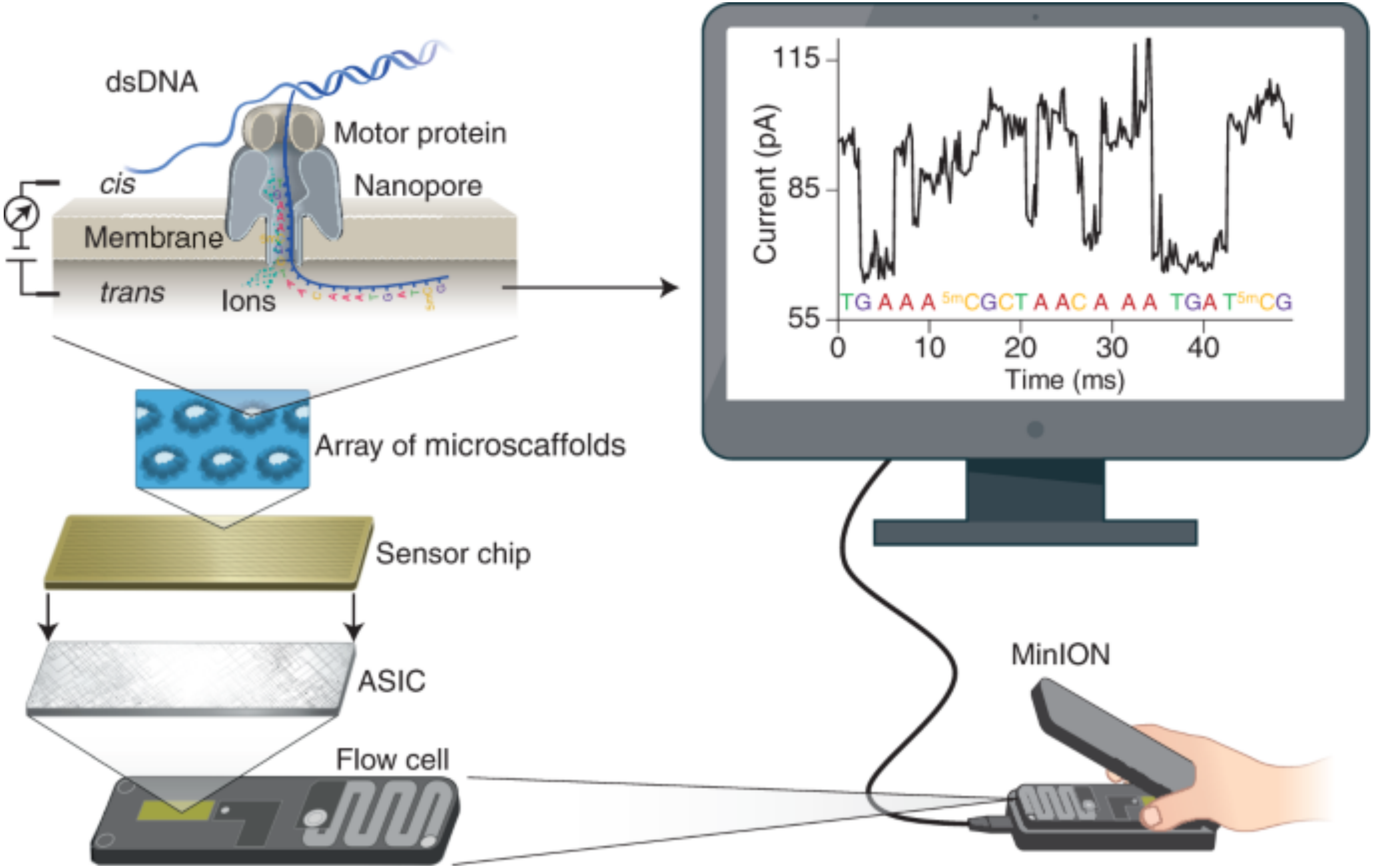




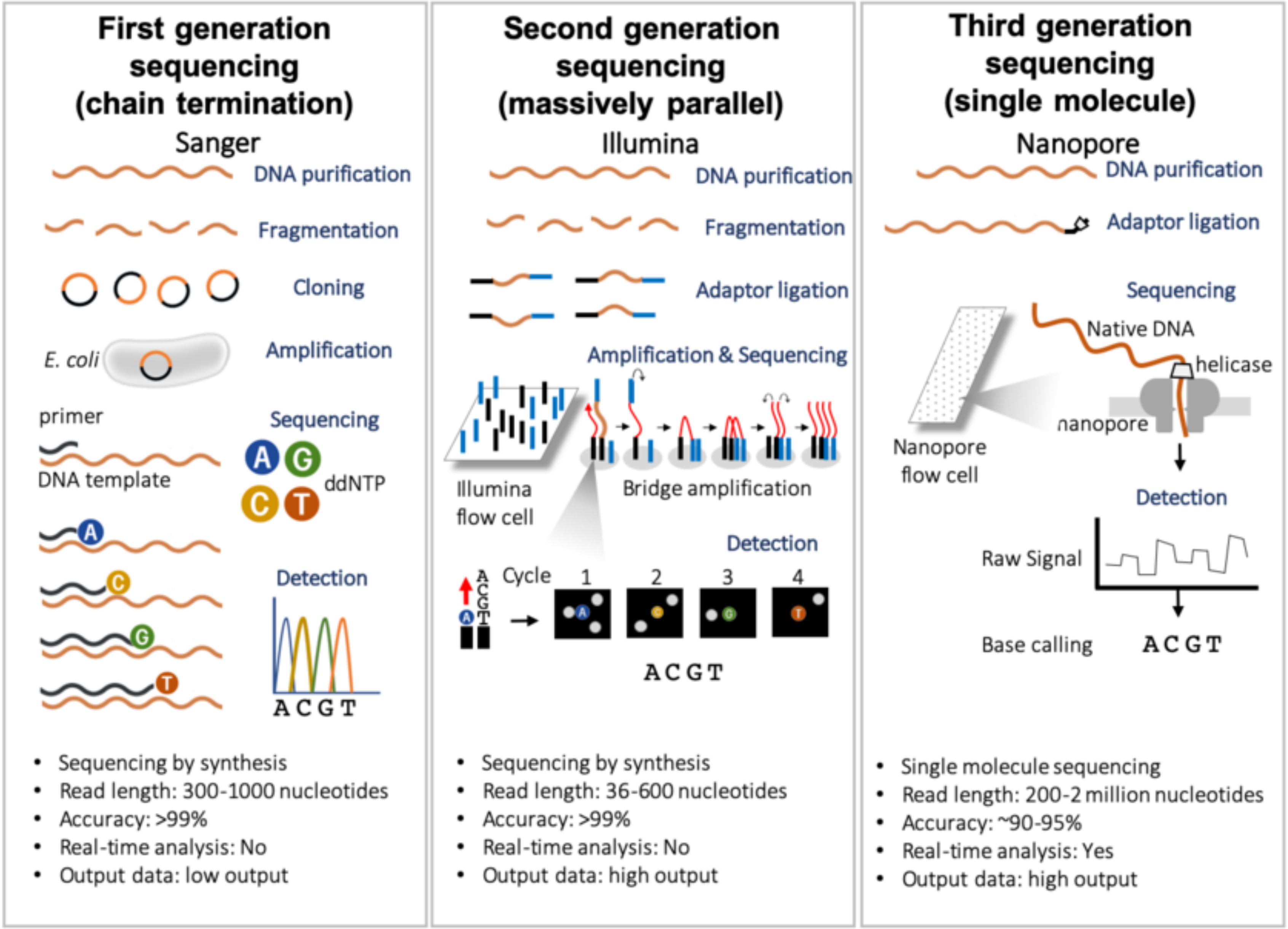
3rd géneration : PacBio technology (2011)



3rd géneration : Nanopore technology (2014)



Sanger / Illumina / Nanopore sequencing…



First generation sequencing (chain termination) – Sanger

- Sequencing by synthesis
- Read length: 300-1000 nucleotides
- Accuracy: >99%
- Real-time analysis: No
- Output data: low output

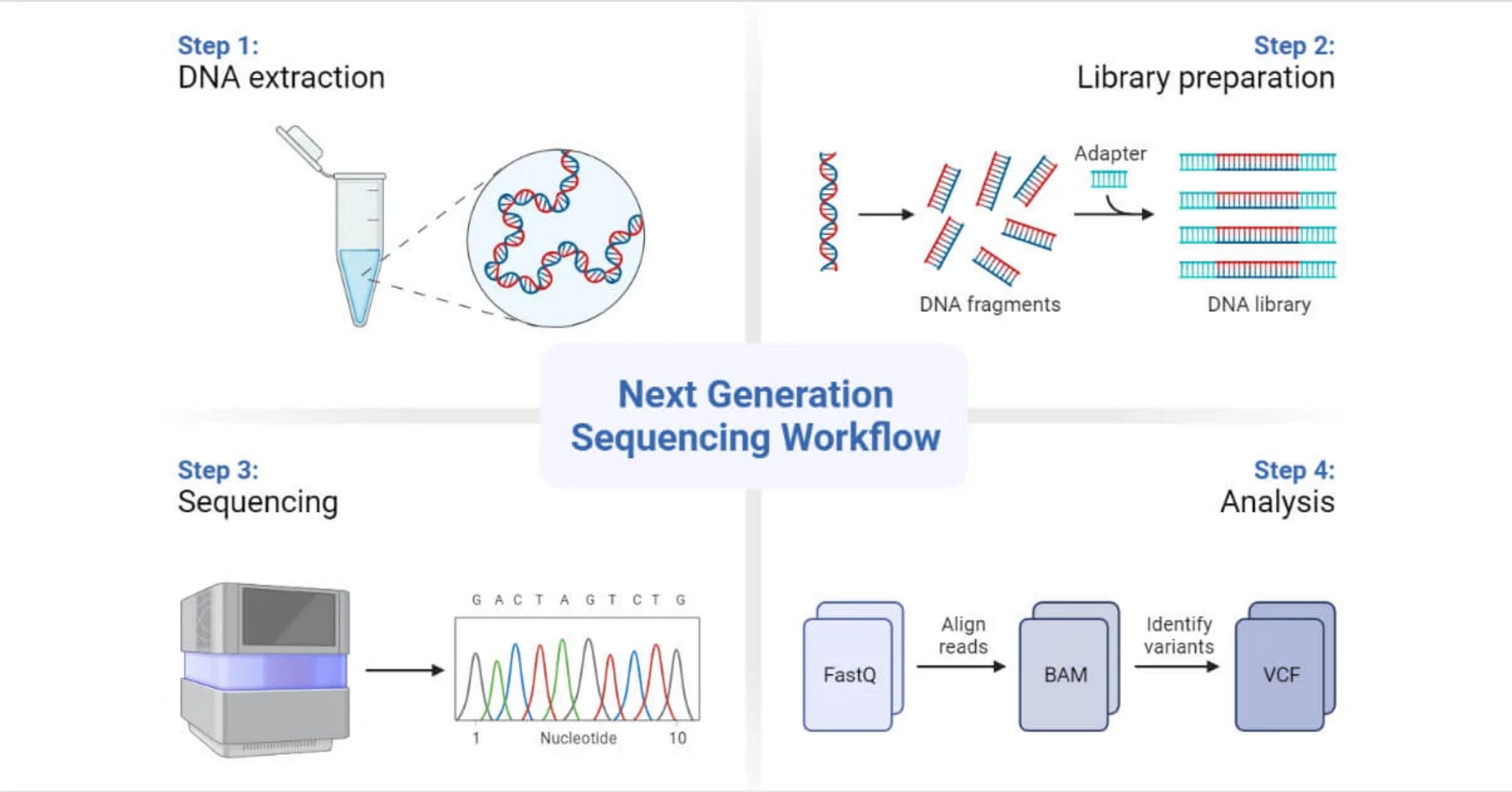
Second generation sequencing (massively parallel) – Illumina

- Sequencing by synthesis
- Read length: 36-600 nucleotides
- Accuracy: >99%
- Real-time analysis: No
- Output data: high output

Third generation sequencing (single molecule) – Nanopore

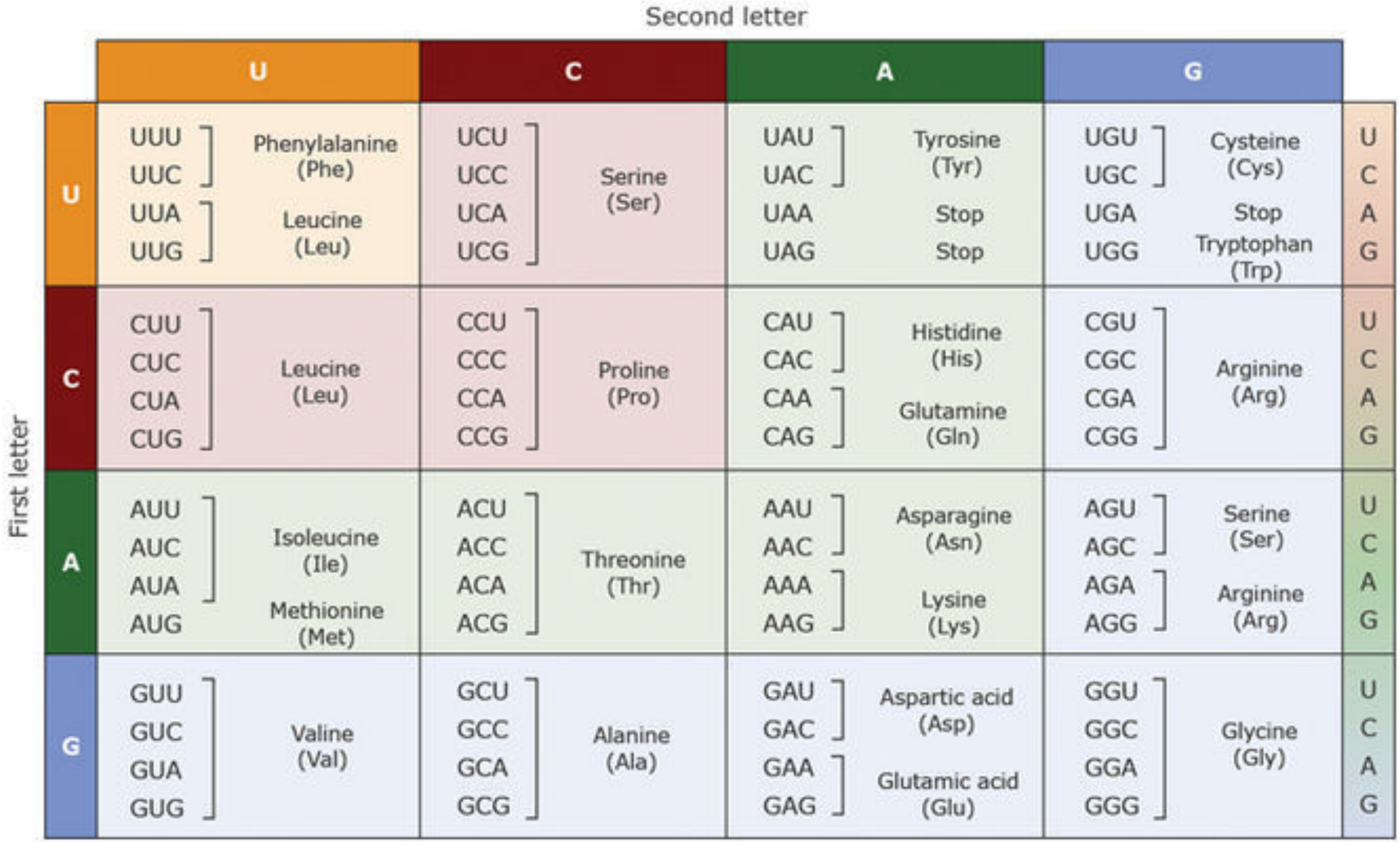
- Single molecule sequencing
- Read length: 200-2 million nucleotides
- Accuracy: ~90-95%
- Real-time analysis: Yes
- Output data: high output

Commun steps…



…

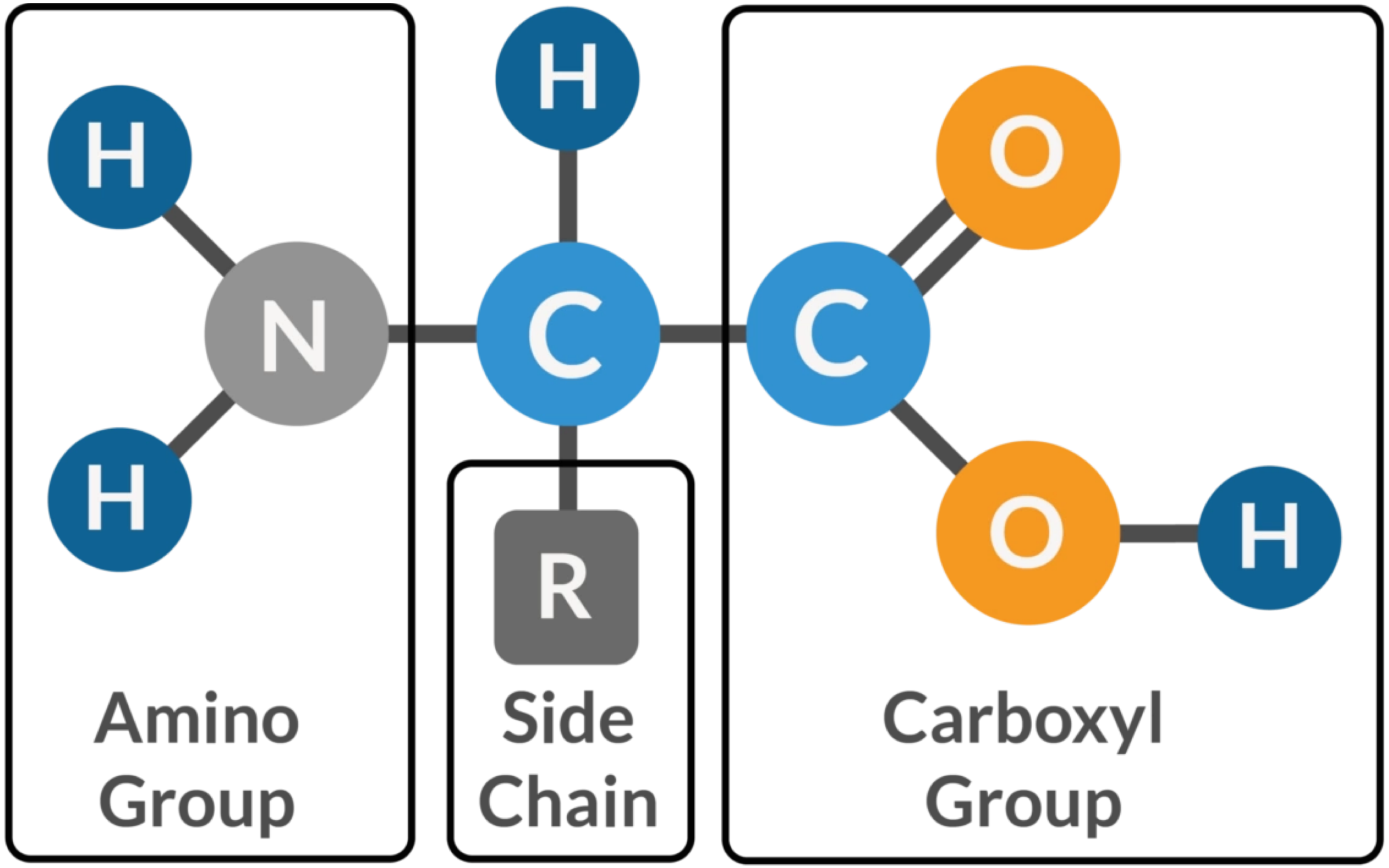
Genetic code



Second letter

First letter	U		C		A		G		
U	UUU	Phenylalanine (Phe)	UCU	Serine (Ser)	UAU	Tyrosine (Tyr)	UGU	Cysteine (Cys)	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine (Leu)	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Tryptophan (Trp)	G
C	CUU	Leucine (Leu)	CCU	Proline (Pro)	CAU	Histidine (His)	CGU	Arginine (Arg)	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine (Gln)	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine (Ile)	ACU	Threonine (Thr)	AAU	Asparagine (Asn)	AGU	Serine (Ser)	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine (Lys)	AGA	Arginine (Arg)	A
	AUG	Methionine (Met)	ACG		AAG		AGG		G
G	GUU	Valine (Val)	GCU	Alanine (Ala)	GAU	Aspartic acid (Asp)	GGU	Glycine (Gly)	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid (Glu)	GGA		A
	GUG		GCG		GAG		GGG		G

Amino acids structure

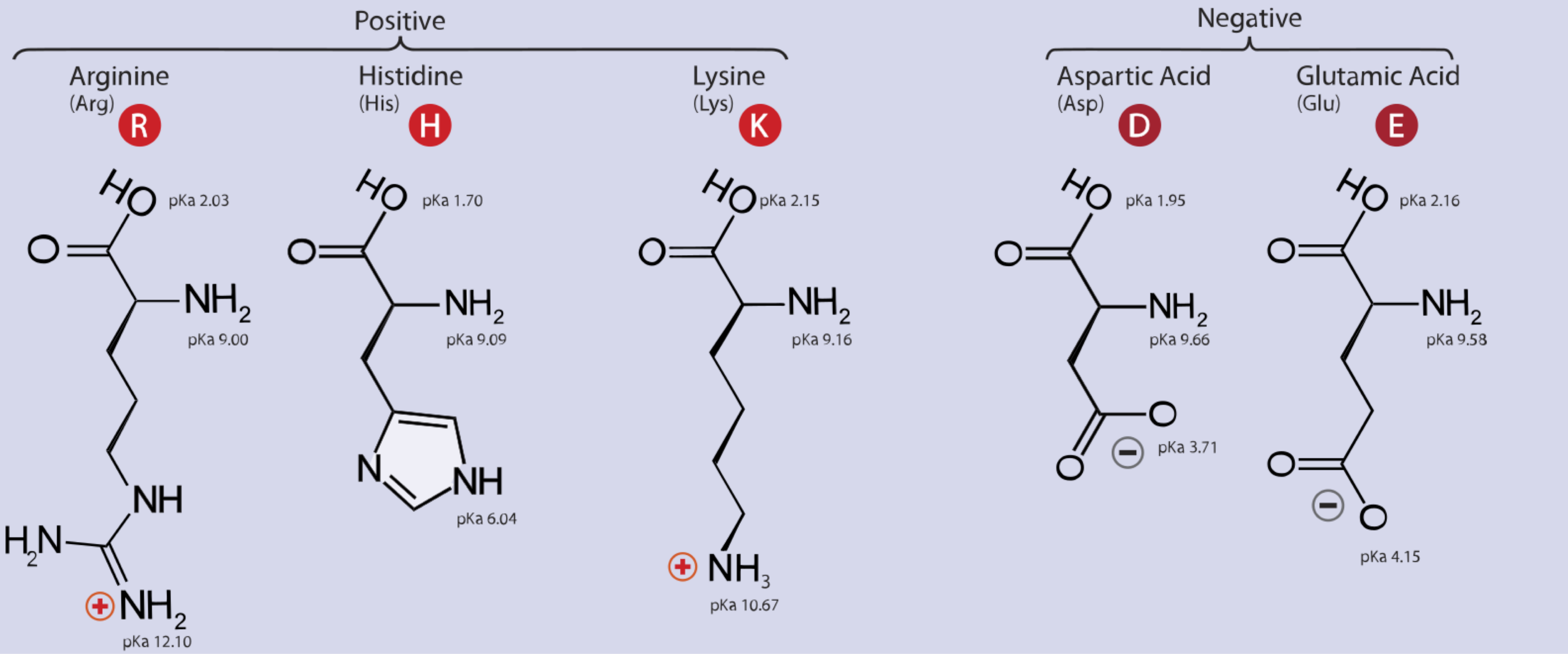


Amino acids

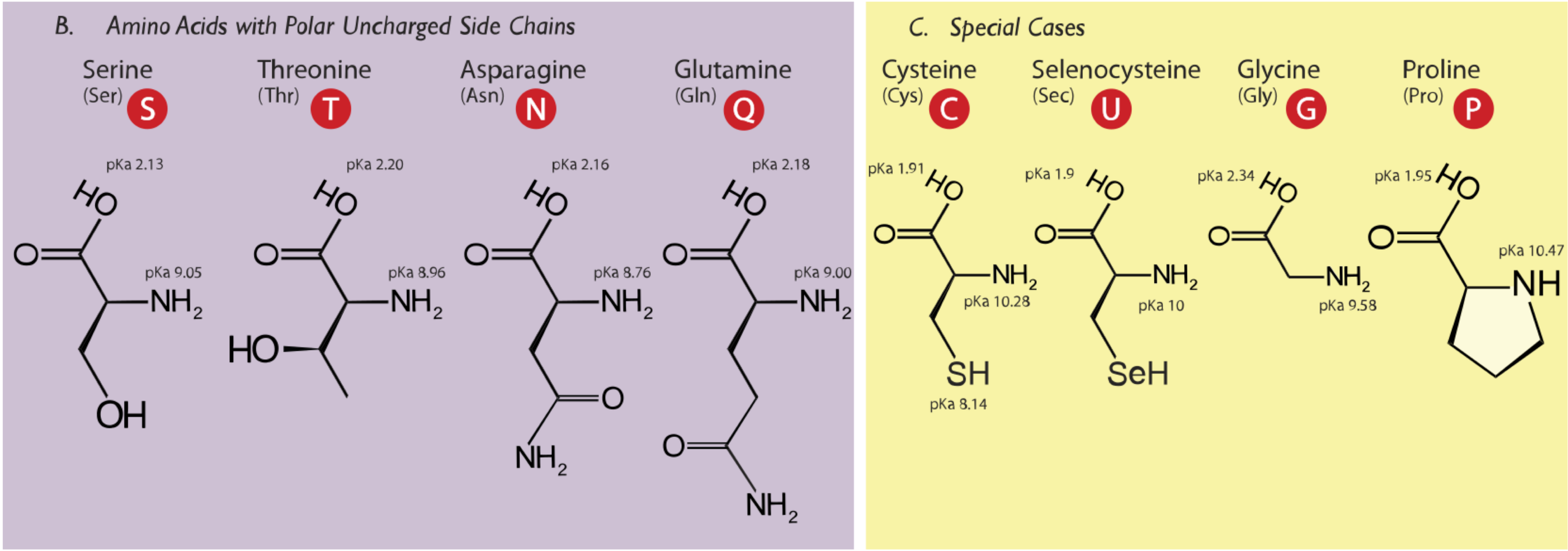
Twenty-One Amino Acids

⊕ Positive ⊖ Negative
- Side chain charge at physiological **pH 7.4**

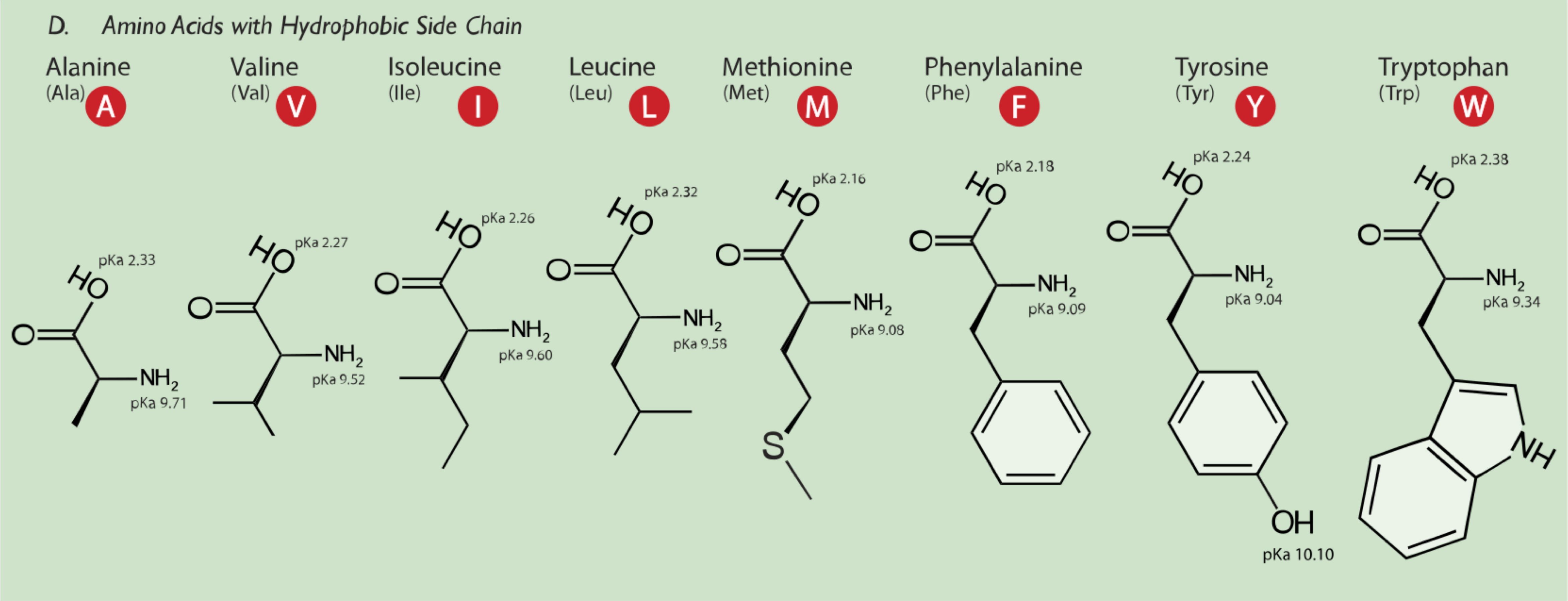
A. *Amino Acids with Electrically Charged Side Chains*



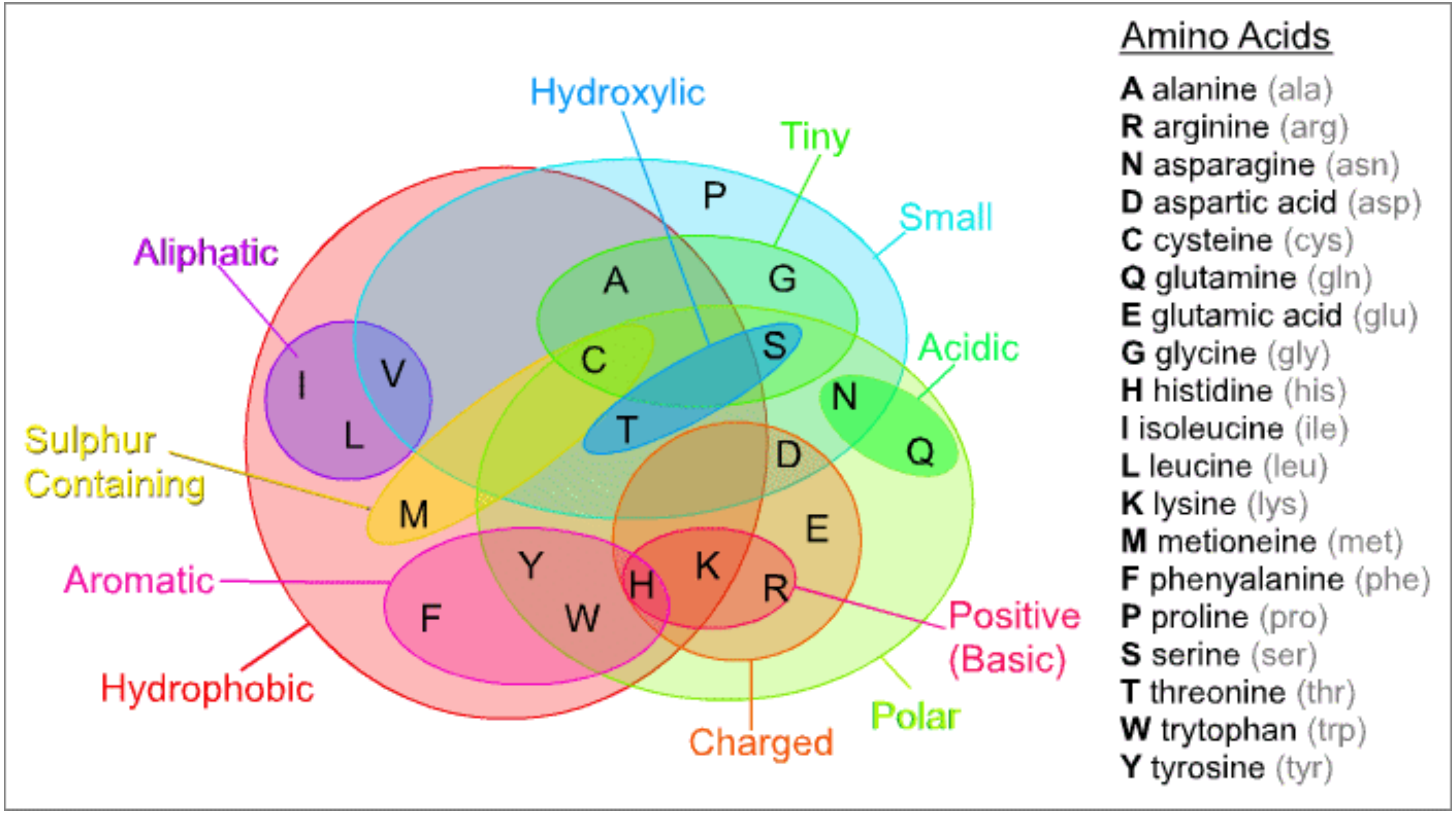
Amino acids


B. Amino Acids with Polar Uncharged Side Chains
Serine (Ser) S
pKa 2.13
HO
O
pKa 9.05
NH$_2$
OH
Threonine (Thr) T
pKa 2.20
HO
O
pKa 8.96
NH$_2$
HO
Asparagine (Asn) N
pKa 2.16
HO
O
pKa 8.76
NH$_2$
O
NH$_2$
Glutamine (Gln) Q
pKa 2.18
HO
O
pKa 9.00
NH$_2$
O
NH$_2$
C. Special Cases
Cysteine (Cys) C
pKa 1.91
HO
O
NH$_2$
pKa 10.28
SH
pKa 8.14
Selenocysteine (Sec) U
pKa 1.9
HO
O
NH$_2$
pKa 10
SeH
Glycine (Gly) G
pKa 2.34
HO
O
NH$_2$
pKa 9.58
Proline (Pro) P
pKa 1.95
HO
O
pKa 10.47
NH

Amino acids


D. Amino Acids with Hydrophobic Side Chain
Alanine
(Ala)
A
HO
pKa 2.33
O
NH2
pKa 9.71
Valine
(Val)
V
HO
pKa 2.27
O
NH2
pKa 9.52
Isoleucine
(Ile)
I
HO
pKa 2.26
O
NH2
pKa 9.60
Leucine
(Leu)
L
HO
pKa 2.32
O
NH2
pKa 9.58
Methionine
(Met)
M
HO
pKa 2.16
O
NH2
pKa 9.08
S
Phenylalanine
(Phe)
F
HO
pKa 2.18
O
NH2
pKa 9.09
Tyrosine
(Tyr)
Y
HO
pKa 2.24
O
NH2
pKa 9.04
OH
pKa 10.10
Tryptophan
(Trp)
W
HO
pKa 2.38
O
NH2
pKa 9.34
NH

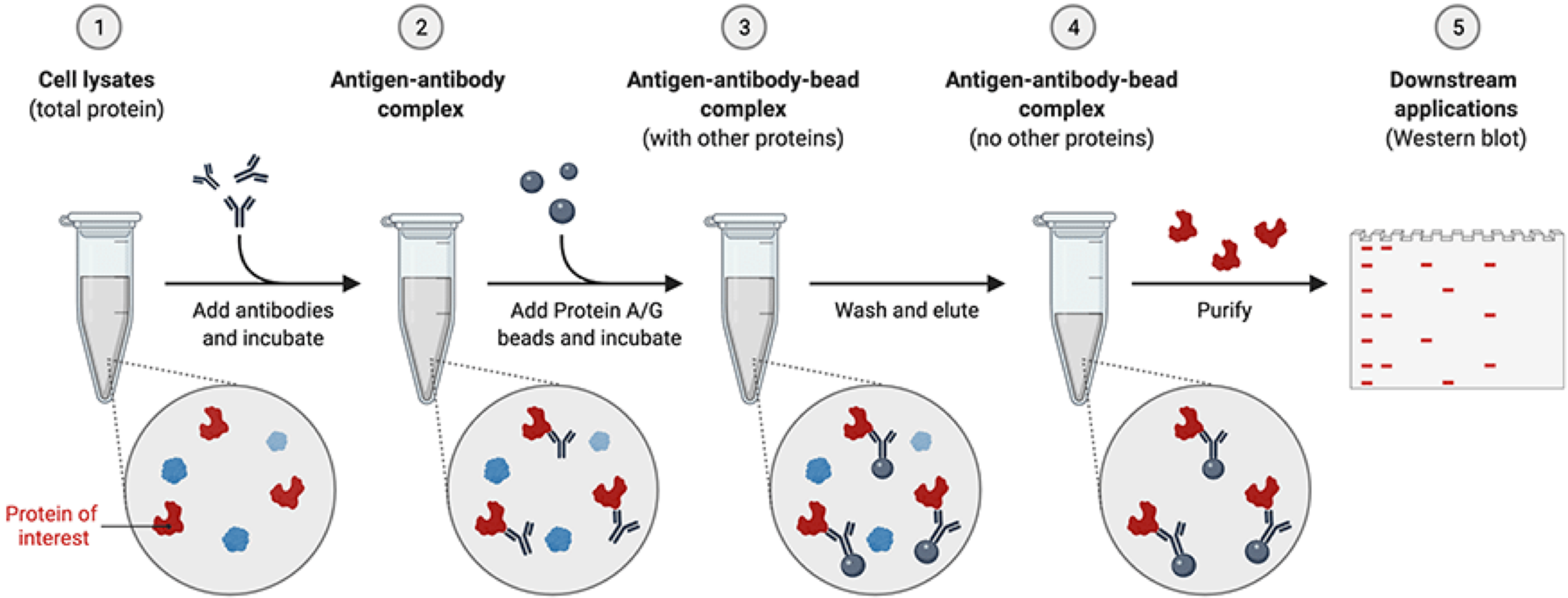
Amino acids



Amino Acids

A	alanine	(ala)
R	arginine	(arg)
N	asparagine	(asn)
D	aspartic acid	(asp)
C	cysteine	(cys)
Q	glutamine	(gln)
E	glutamic acid	(glu)
G	glycine	(gly)
H	histidine	(his)
I	isoleucine	(ile)
L	leucine	(leu)
K	lysine	(lys)
M	metioneine	(met)
F	phenyalanine	(phe)
P	proline	(pro)
S	serine	(ser)
T	threonine	(thr)
W	trytophan	(trp)
Y	tyrosine	(tyr)

Immunoprecipitation



① Cell lysates (total protein)

② Antigen-antibody complex

③ Antigen-antibody-bead complex (with other proteins)

④ Antigen-antibody-bead complex (no other proteins)

⑤ Downstream applications (Western blot)

Add antibodies and incubate

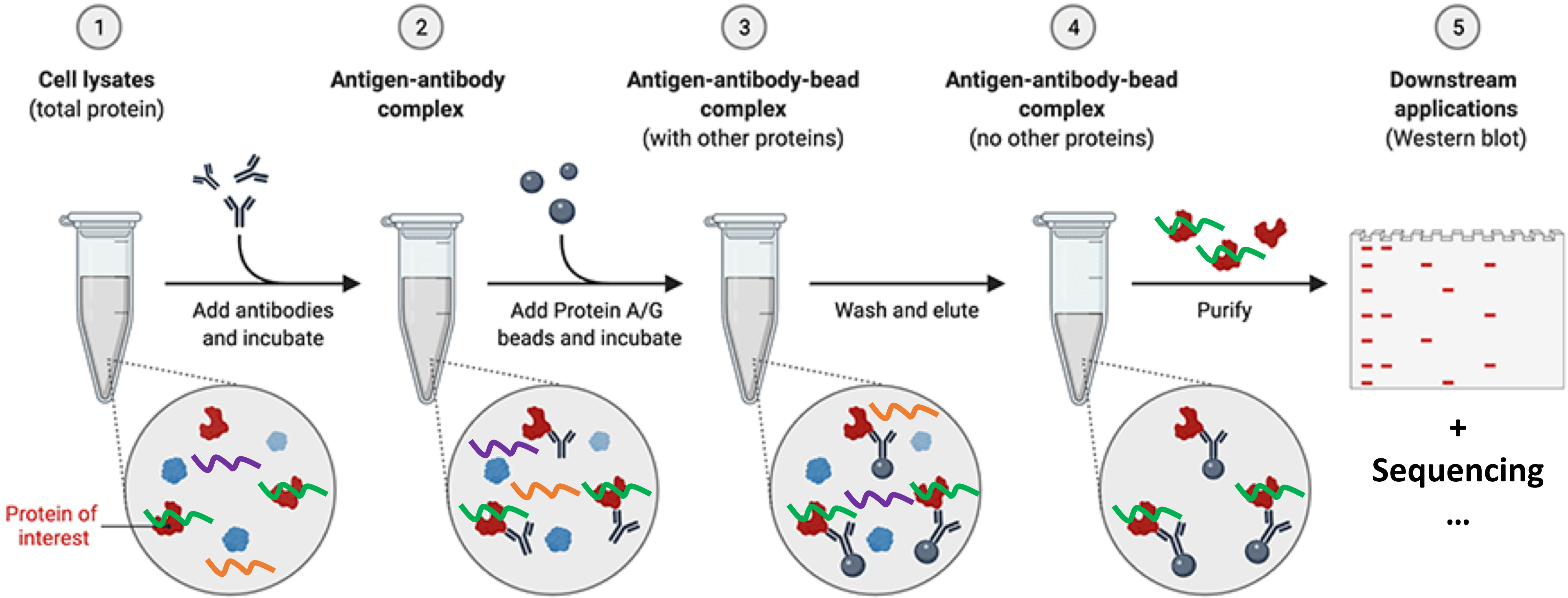
Add Protein A/G beads and incubate

Wash and elute

Purify

Protein of interest

Immunoprecipitation and nucleic acid purification

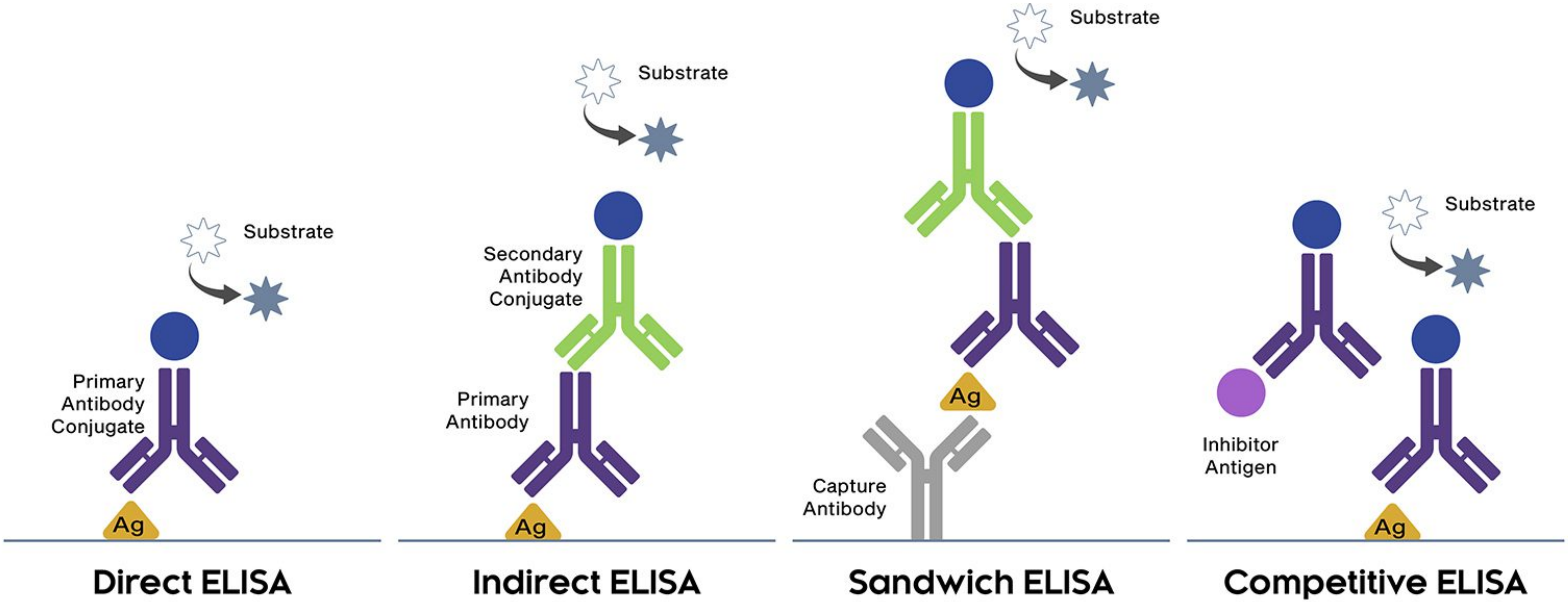


+

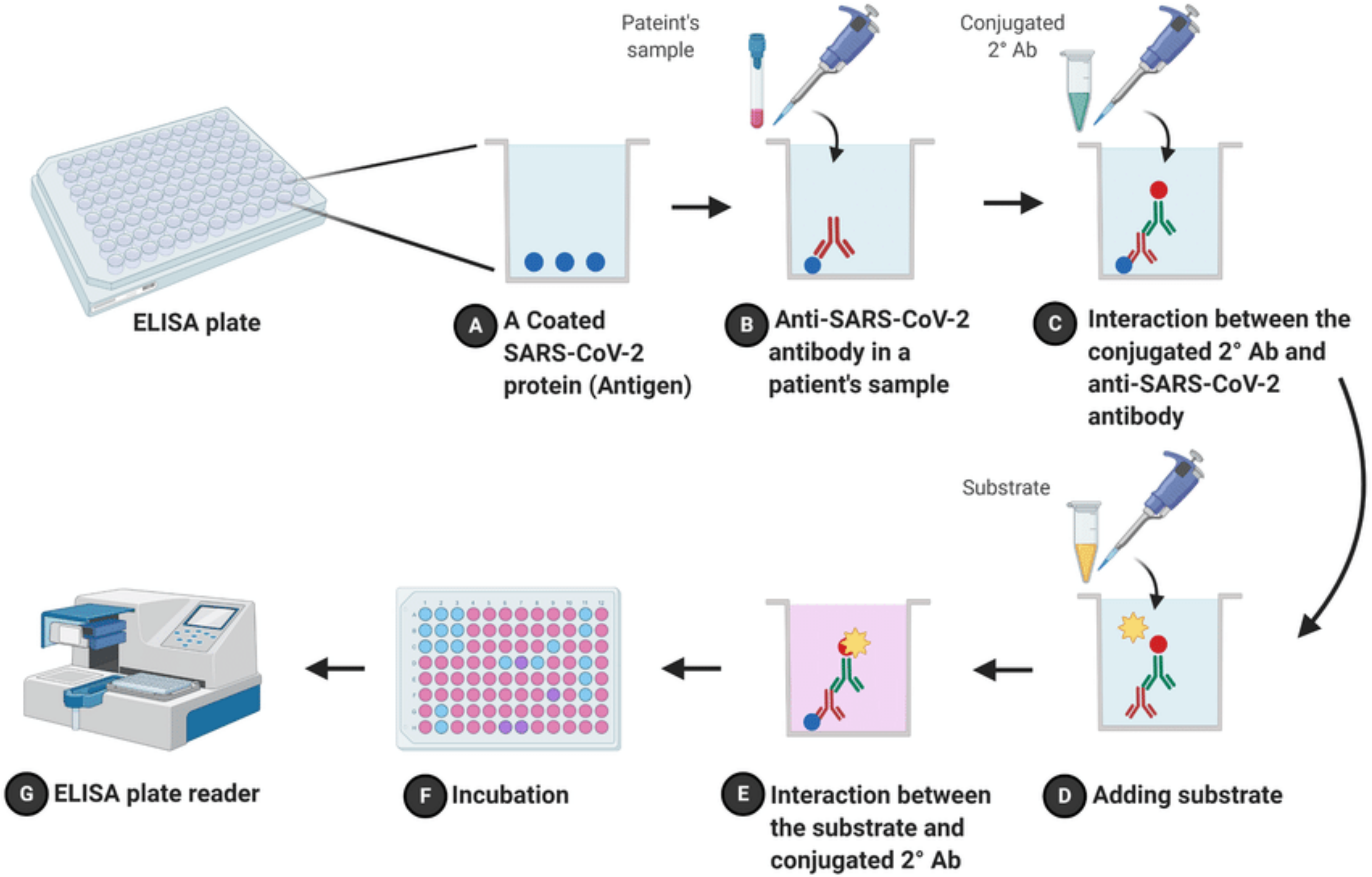
Sequencing

…

Types of ELISA



Example of SARS-CoV-2 ELISA



Ⓐ A Coated SARS-CoV-2 protein (Antigen)

Ⓑ Anti-SARS-CoV-2 antibody in a patient's sample

Ⓒ Interaction between the conjugated 2° Ab and anti-SARS-CoV-2 antibody

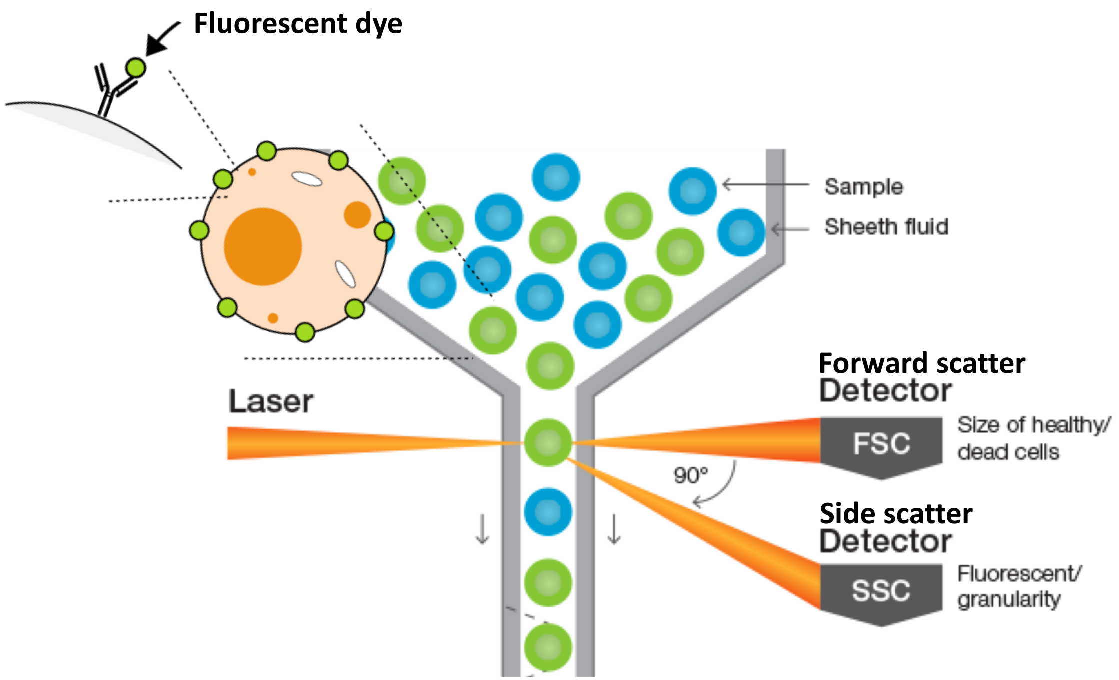
Ⓓ Adding substrate

Ⓔ Interaction between the substrate and conjugated 2° Ab

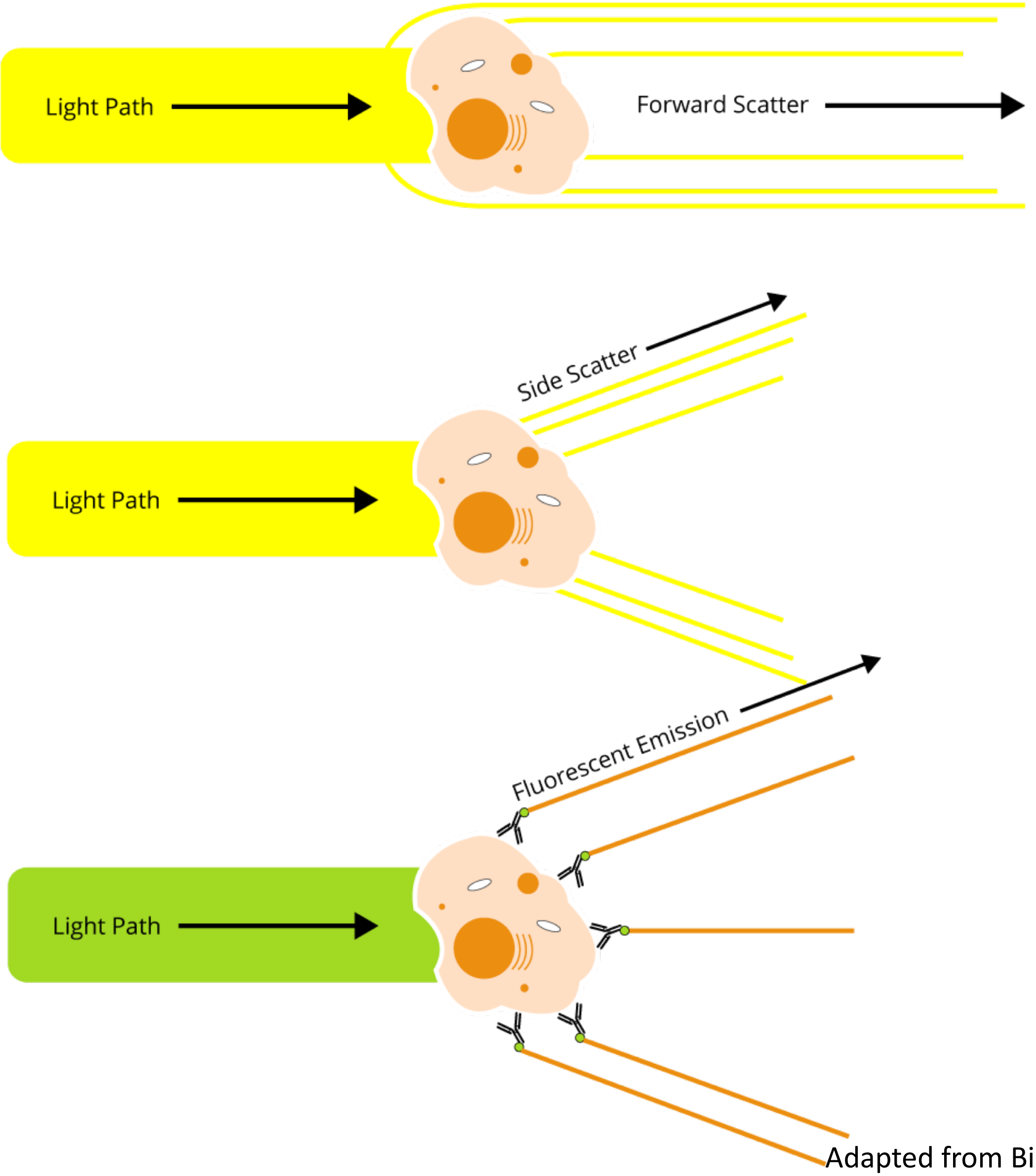
Ⓕ Incubation

Ⓖ ELISA plate reader

Flow Cytometry

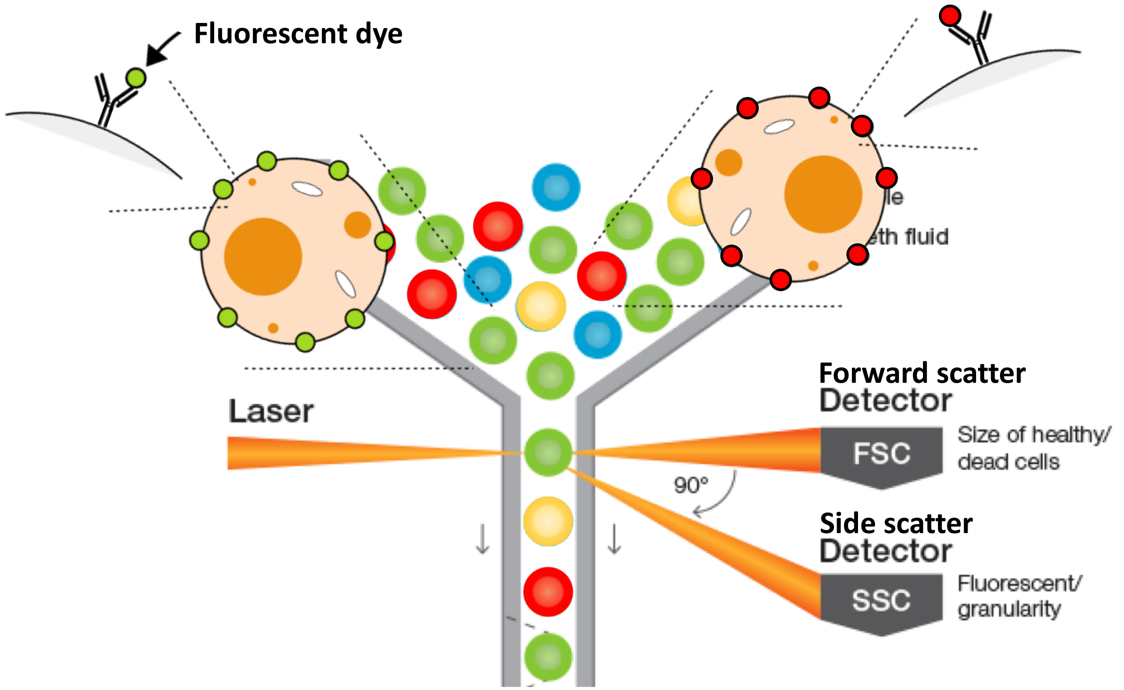


Flow cytometry

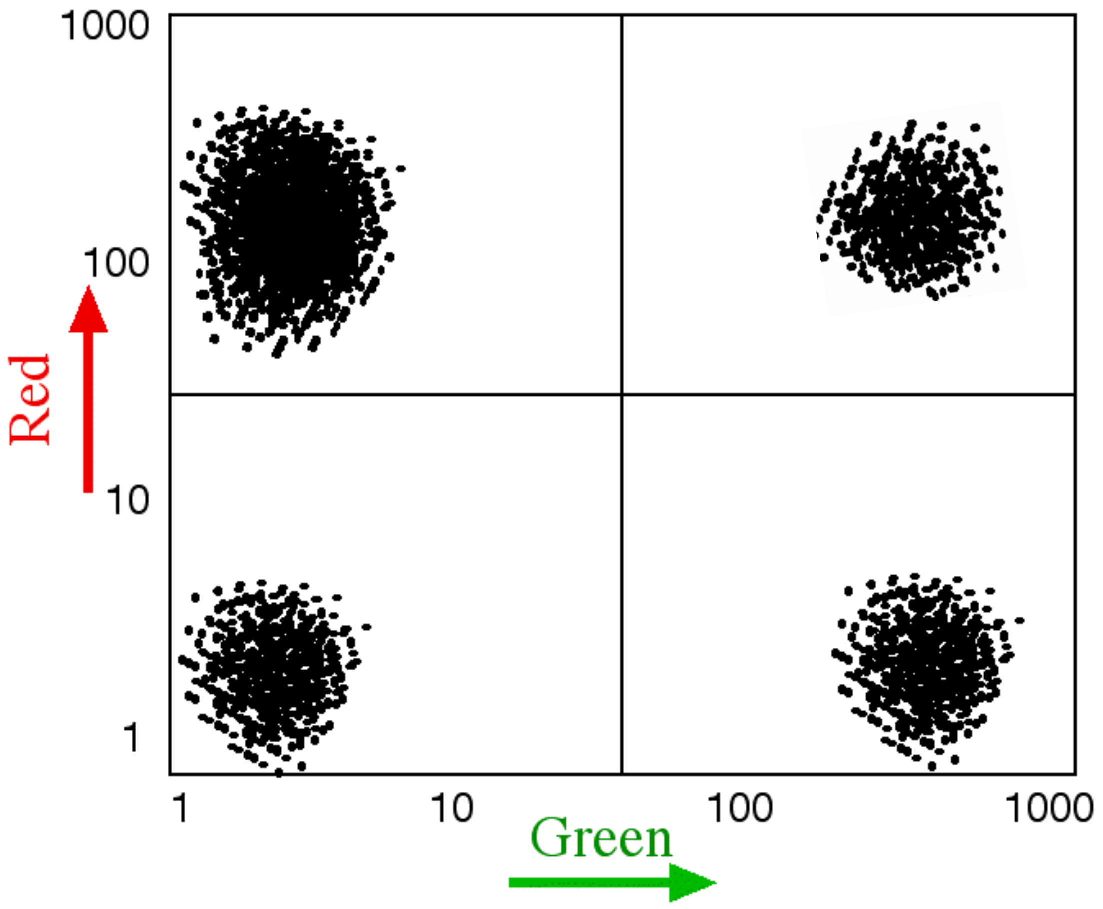


Adapted from BioRad and BosterBio

Flow Cytometry

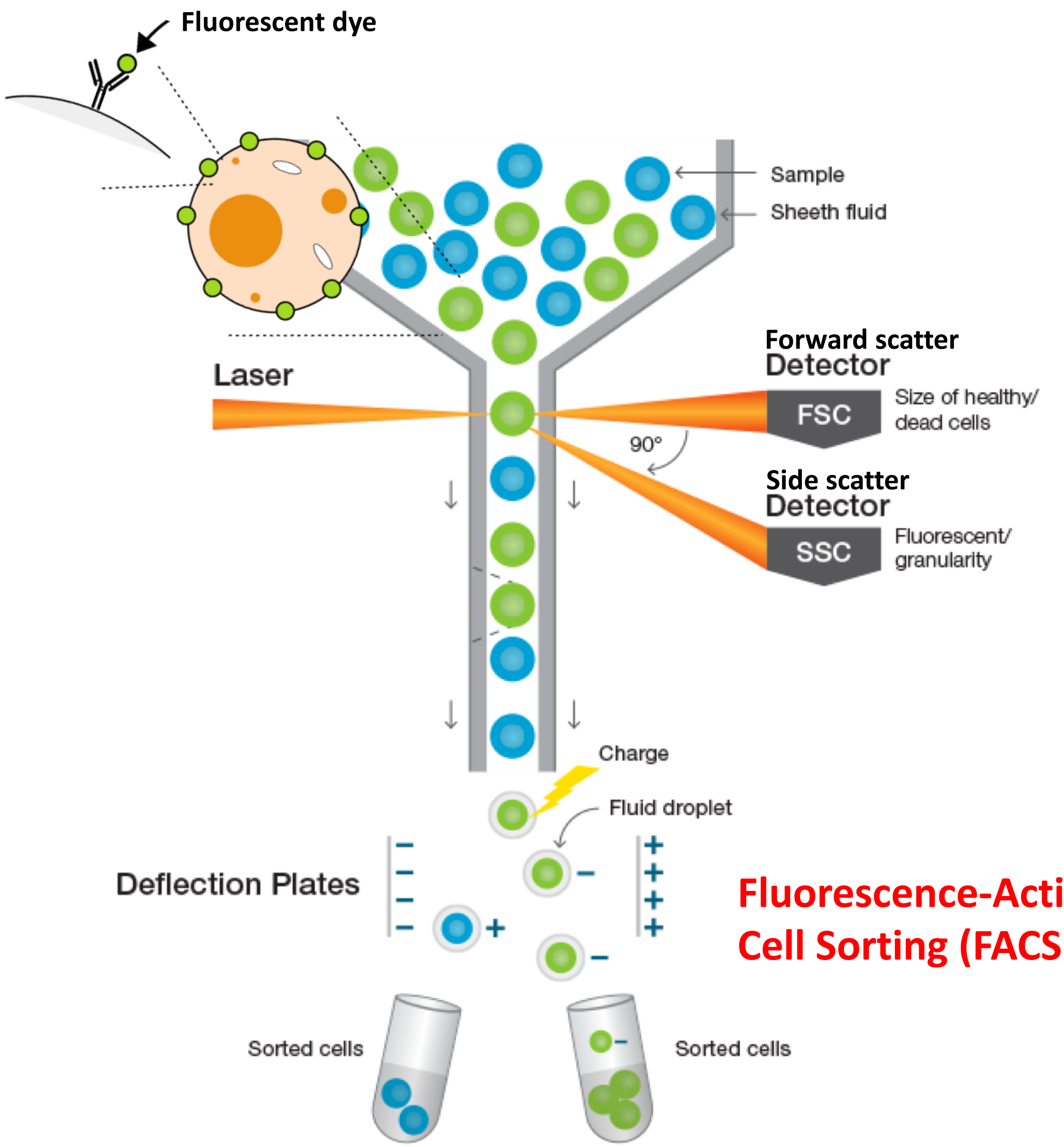


Flow cytometry



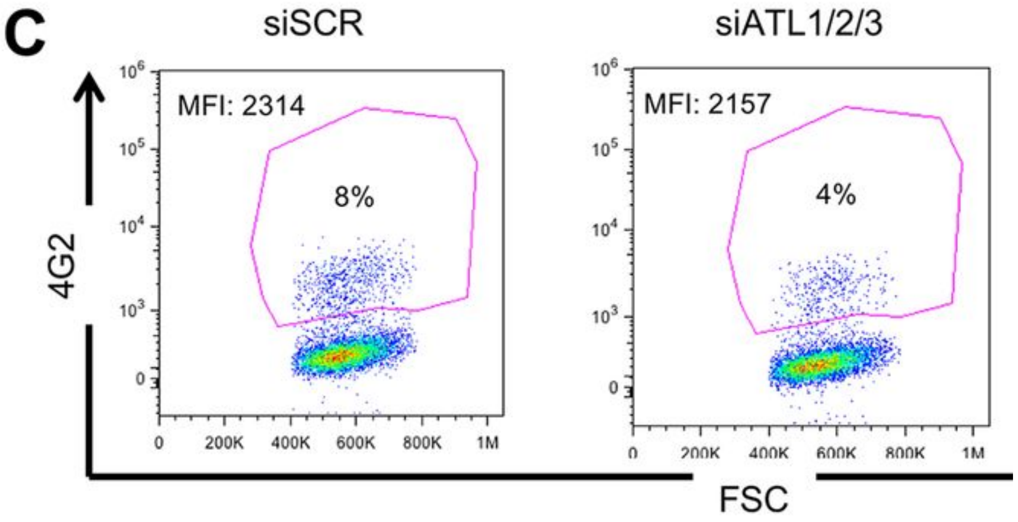
Adapted from BioRad and BosterBio

FACS : Fluorescence-Activated Cell Sorting



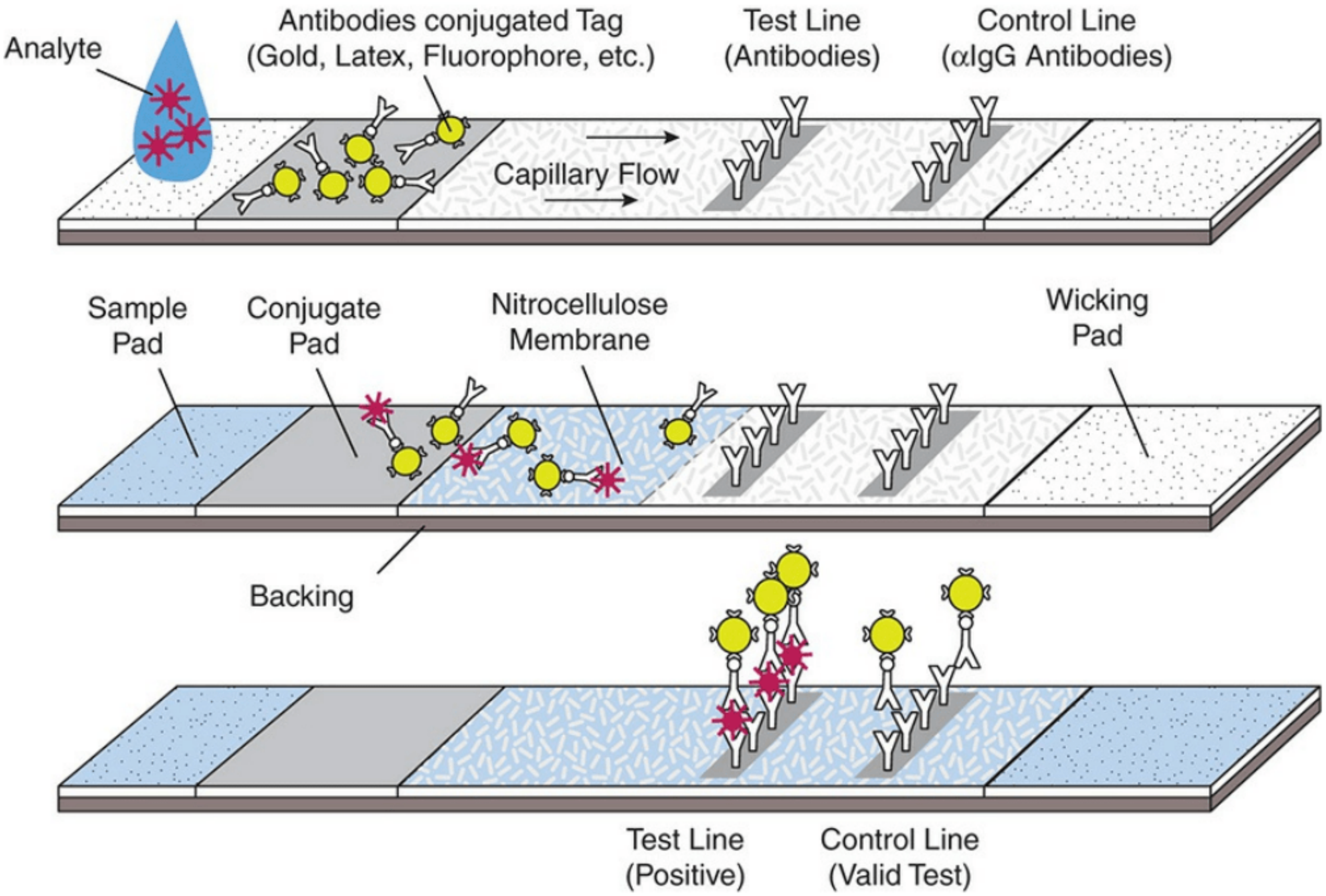
Fluorescence-Activated Cell Sorting (FACS)

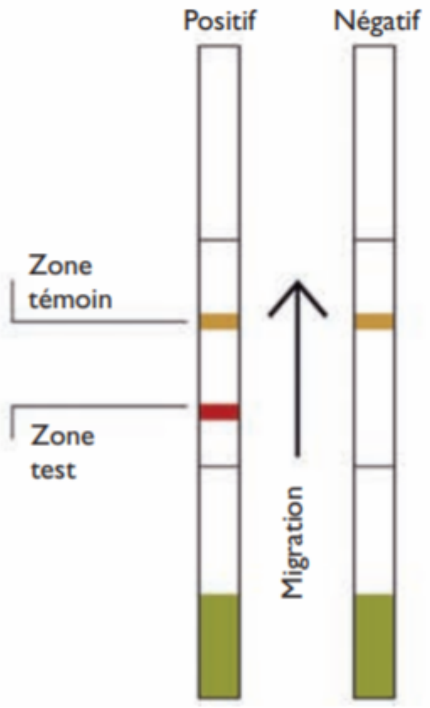
MFI = Mean Fluorescent Intensity



Adapted from BioRad and BosterBio

Immunochromatography





XXX