Cellular and Molecular Biology I

BIO-205-10

Camille Goemans

V. Molecular and Cellular Biology in the lab

1. Model organisms

2. Cell cultures

3. Studying proteins

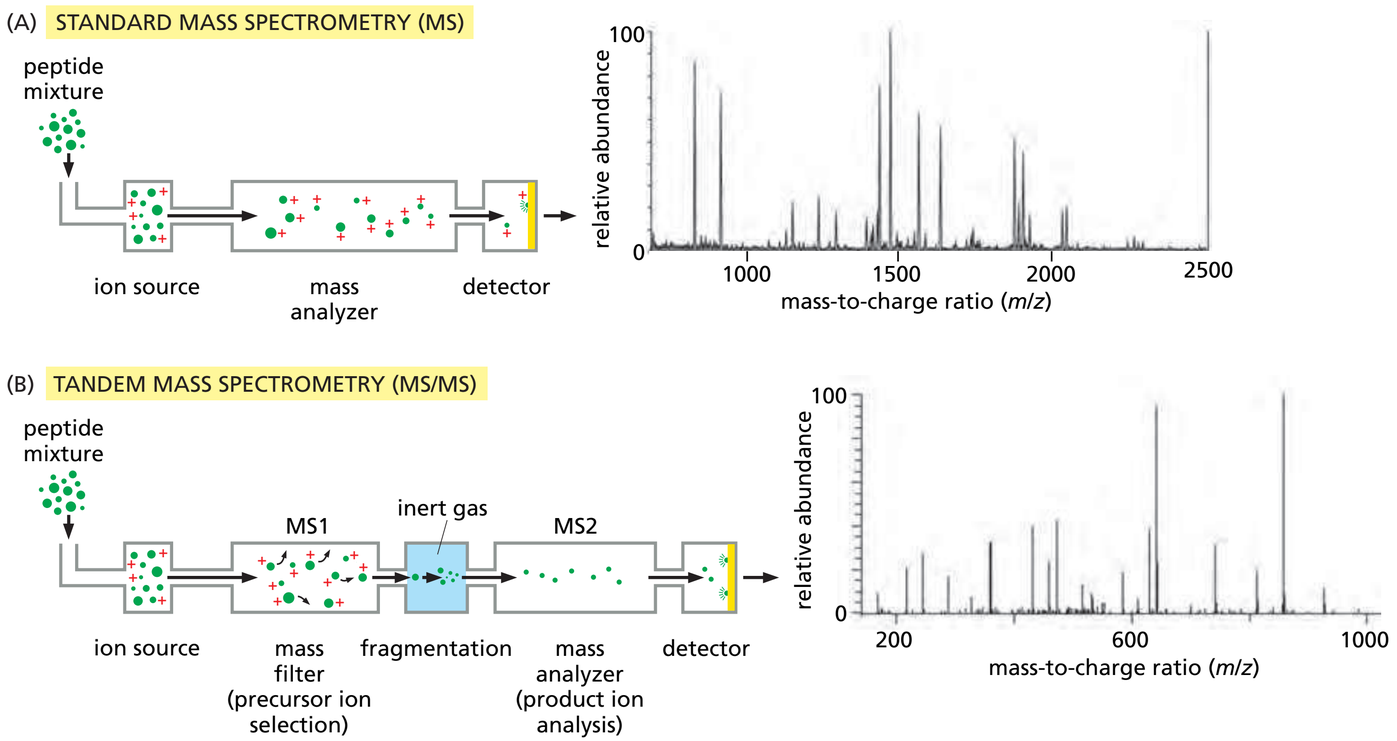
- Protein sequence
- Protein purification
- Protein visualization
- Protein structure
- **Mass spectrometry**
- Additional methods

Mass spectrometry to identify unknown proteins

- How to **identify** a protein obtained from one of the purification methods or all the proteins in a sample?
- **Genomes** of most experimentally used organisms are known - **catalogue of possible proteins** is known
- Use of **mass spectrometry** combined with **searches of databases**
- Charged particles have **very precise dynamics** when subjected to **electric and magnetic fields in a vacuum**
- **Mass spectrometry** separates ions according to their mass-to-charge ratio (m/z)
- Extremely **sensitive** method

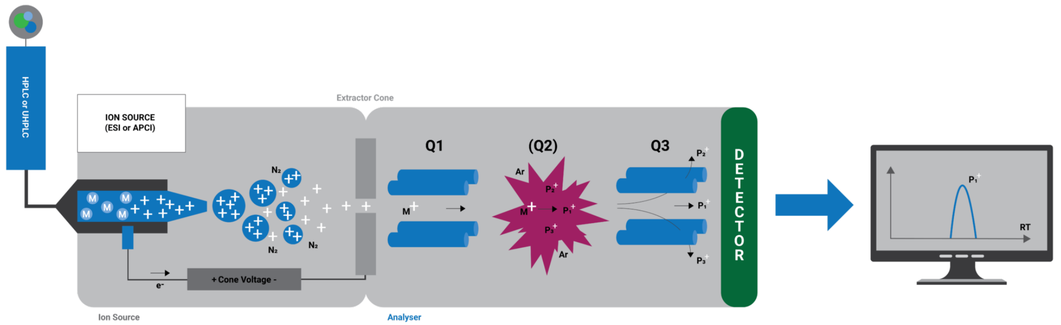
Mass spectrometry to identify unknown proteins

- **Ion source** transforms tiny amounts of peptide into a gas containing individual charged peptides
- **Mass analyzer** where ions are accelerated by electric or magnetic fields in which ions are separated according to their mass-to-charge ratio
- **Detector** on which the ions collide which generate a mass spectrum
- **Different types** of mass spectrometer, with different ion sources and mass analyzers (e.g. MALDI-TOF)



Mass spectrometry to identify unknown proteins

- Use of **LC-MS/MS** for **complex protein mixtures**
  - LC (Liquid Chromatography):
    - Separates components of a mixture based on their chemical properties (like polarity).
    - Helps reduce interference before detection.
  - MS/MS (Tandem Mass Spectrometry):
    - The first mass spectrometer isolates a specific molecule (the precursor ion)
    - The molecule is then broken into fragments.
    - The second mass spectrometer measures these fragments, providing high-specificity identification and quantification



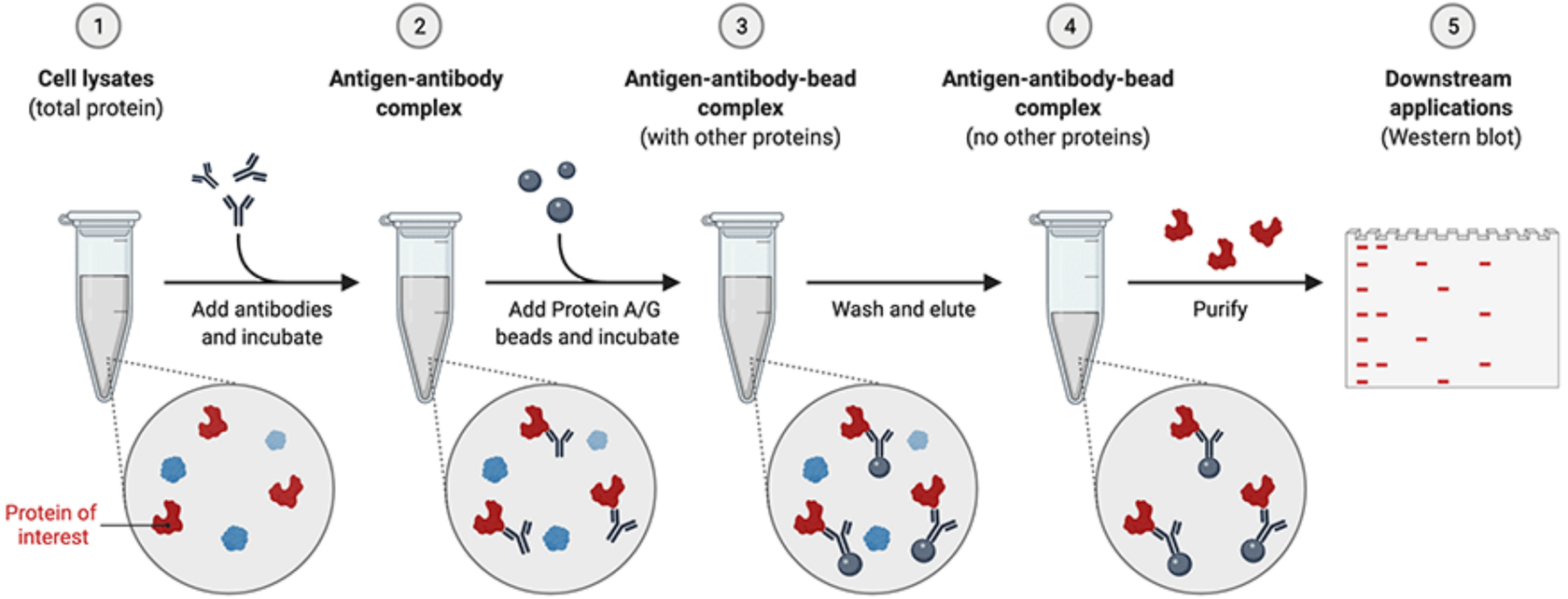
https://blog.perkinelmer.com/posts/mass-spectrometry-explained-how-lc-ms-ms-works/

V. Molecular and Cellular Biology in the lab

1. Model organisms
2. Cell cultures
3. **Studying proteins**
    - Protein sequence
    - Protein purification
    - Protein visualization
    - Protein structure
    - Mass spectrometry
    - **Additional methods**

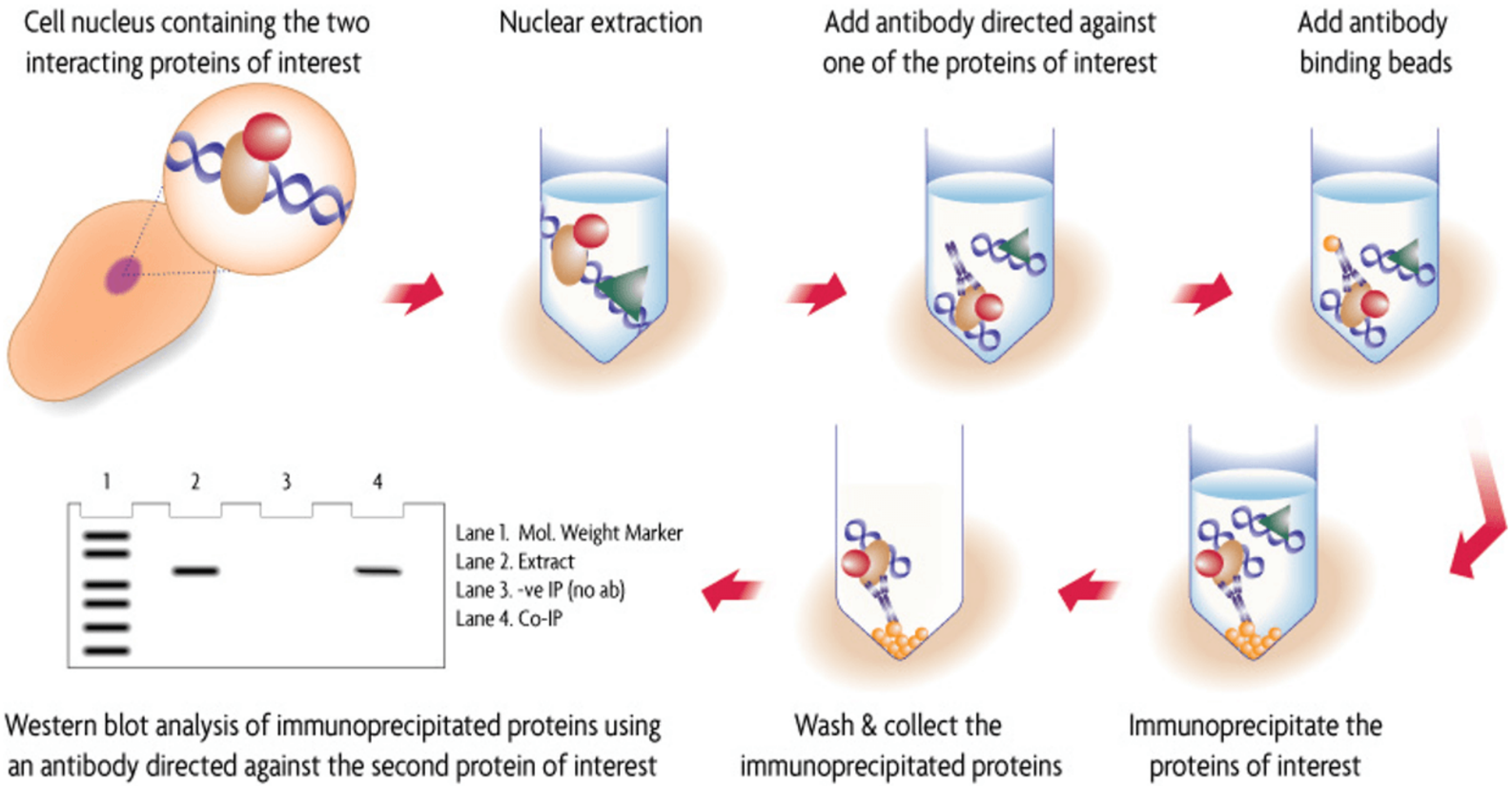
Immunoprecipitation (IP)

- **Immunoprecipitation (IP)** as a rapid affinity purification



https://www.rockland.com/resources/immunoprecipitation-technique/

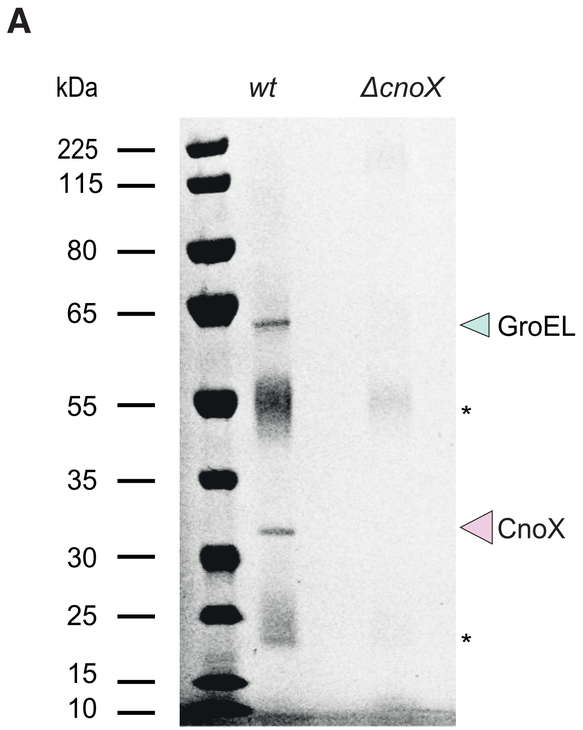
Finding proteins' partners: Co-immunoprecipitation (co-IP)



https://www.activemotif.com/catalog/25/nuclear-complex-co-ip-kit

Real-life example: CnoX

Find the partners: Co-IP +SDS-PAGE

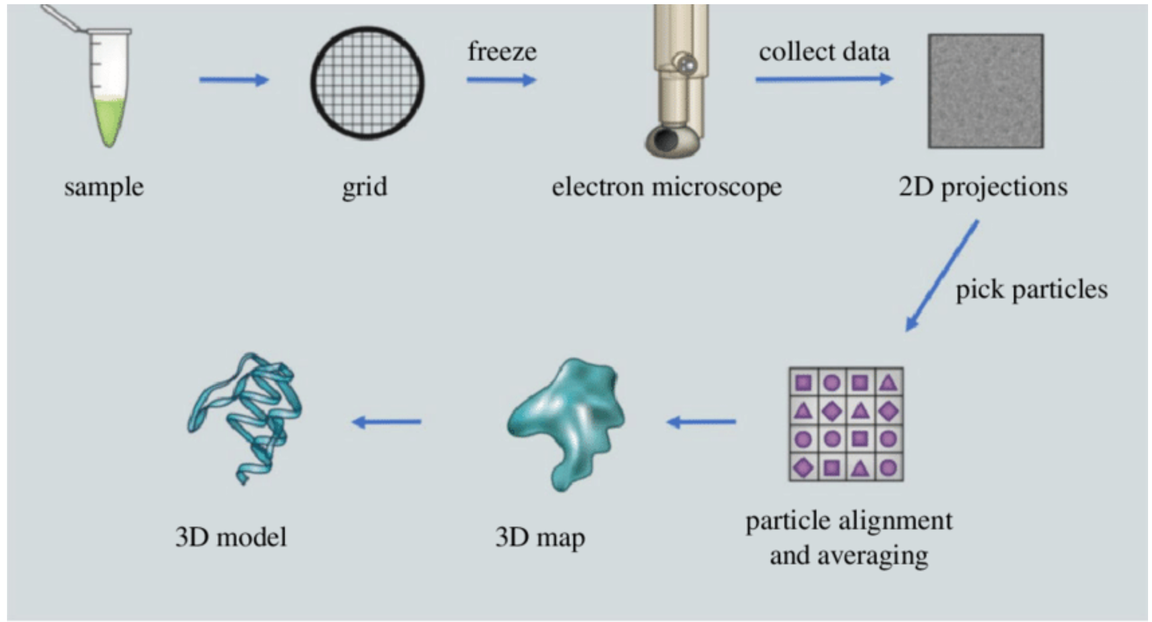


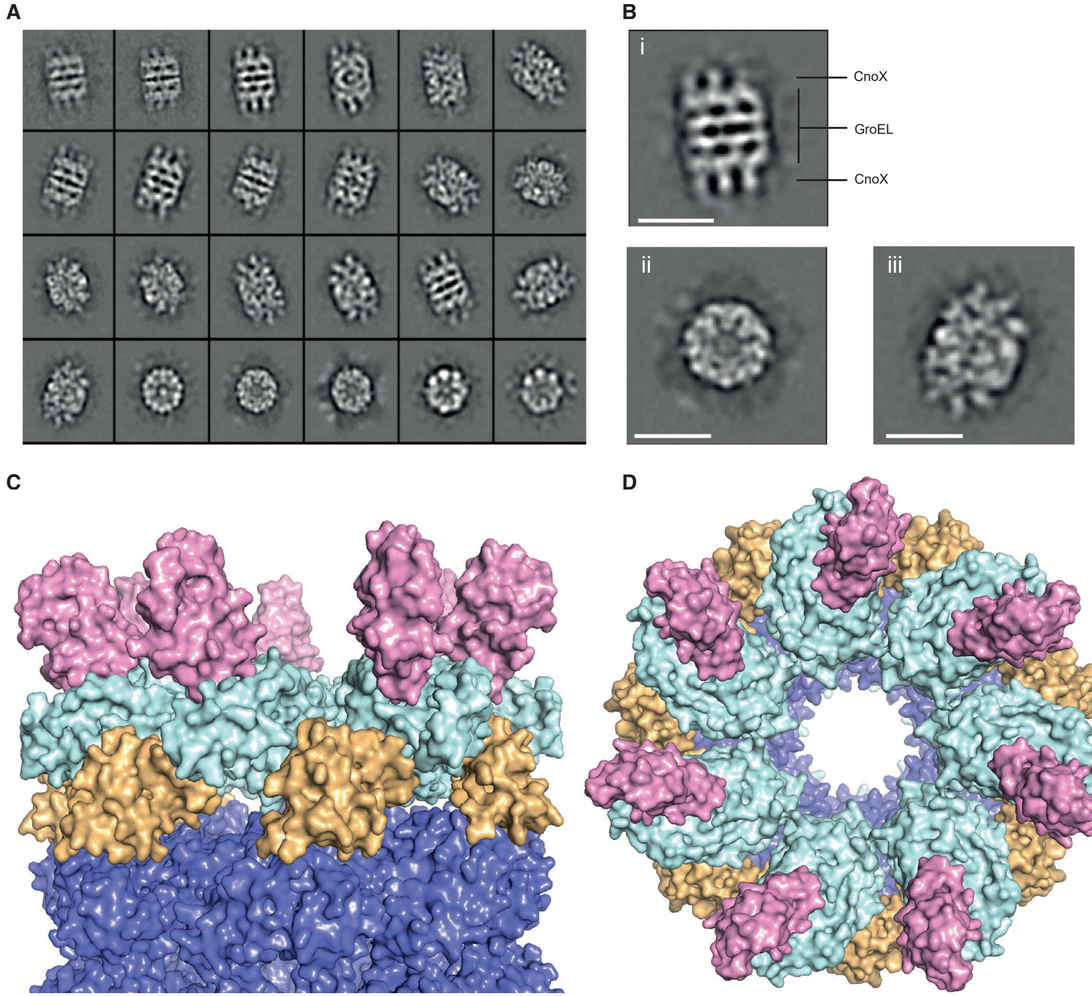
- co-IP using a-CnoX antibodies
- SDS-PAGE: we only detect one other band
- Mass spectrometry: this band is GroEL

Why don’t we do a western blot here?

Real-life example: CnoX

- CryoEM





Finding proteins' partners: Tandem affinity purification (TAP-Tag)

- The TAP-TAG method refers to the **Tandem Affinity Purification** (TAP) tag system — a biochemical technique used to isolate protein complexes with very high purity so researchers can study protein–protein interactions.

- It is a **two-step purification** strategy in which a protein of interest is genetically fused to a dual affinity tag. This allows the protein and its interacting partners to be pulled out of a cell extract with minimal contaminants.

- The traditional **TAP tag** contains:
    - Protein A domains (bind to IgG resin)
    - A TEV protease cleavage site
    - Calmodulin-binding peptide (CBP) (binds to calmodulin resin in the presence of calcium)

Finding proteins' partners: Tandem affinity purification (TAP-Tag)

1. Create a tagged protein

- The gene for your protein is fused with the TAP tag.
- The fusion protein is expressed in the cell.

2. First purification

- Cell lysate is passed over IgG resin.
- *Protein A* portion binds strongly.
- Non-specific proteins are washed away.

3. TEV cleavage

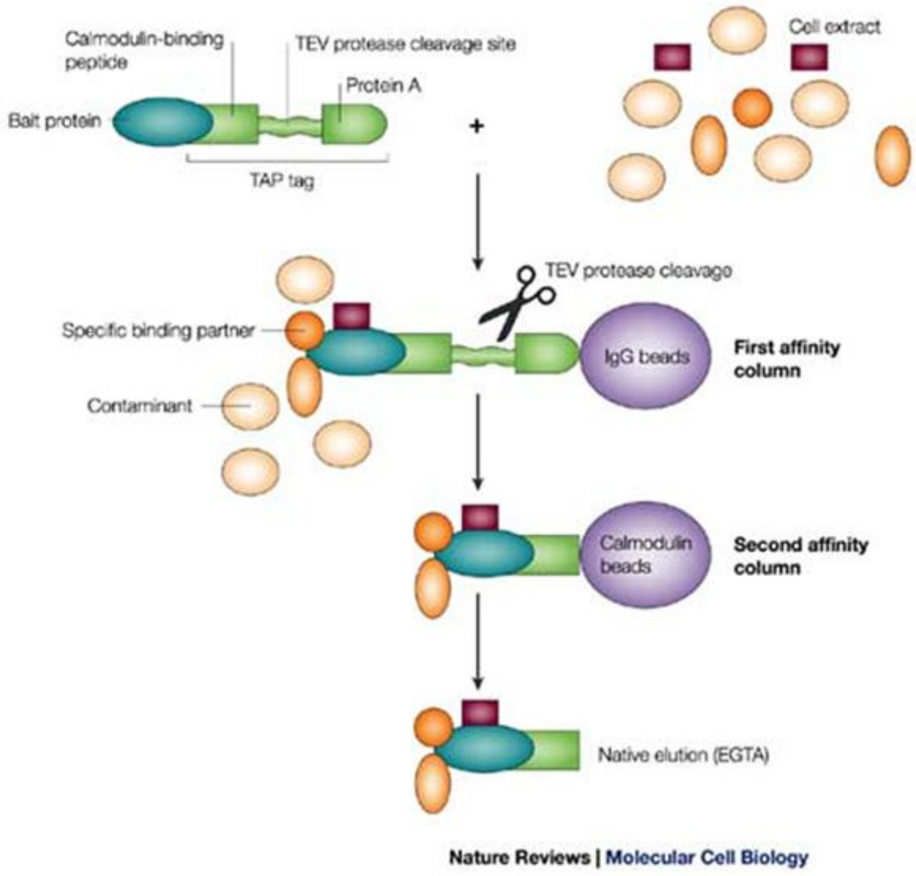
- TEV protease releases the protein + its binding partners from the first resin.
- Protein A is left behind.

4. Second purification

- The eluate binds to calmodulin resin (via CBP).
- Additional washing removes contaminants.

5. Final elution

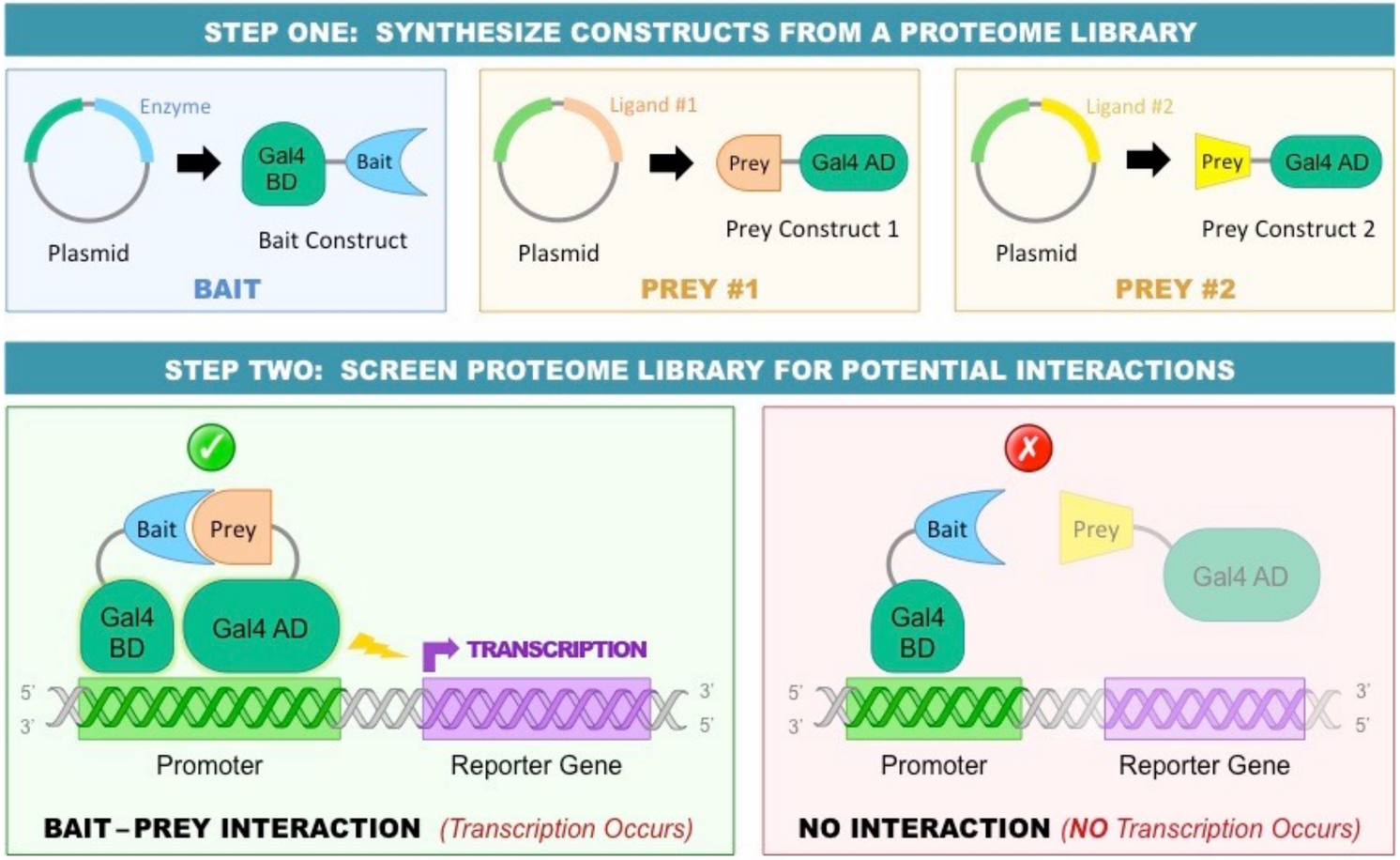
- Removing calcium releases the highly purified protein complex.



*__Protein A__ is a bacterial protein (from Staphylococcus aureus) that binds strongly and specifically to the Fc region of antibodies, particularly:

Finding proteins' partners: two-hybrid

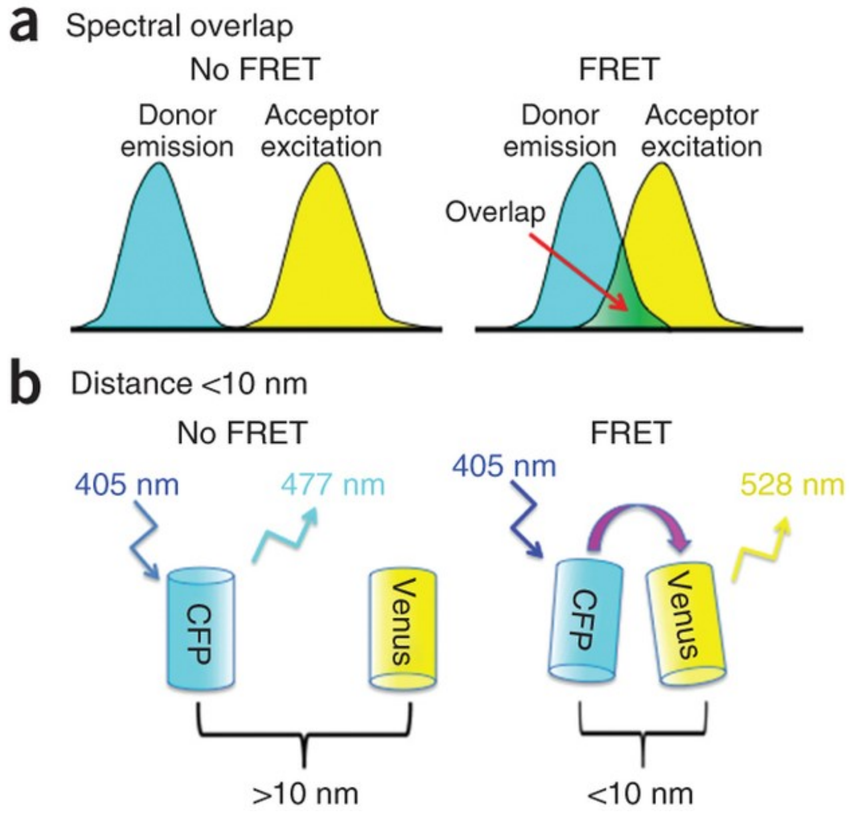
- Yeast/Bacterial two-hybrid



https://old-ib.bioninja.com.au/options/untitled/b4-medicine/yeast-2-hybrid-system.html

Finding proteins' partners: FRET

- FRET = fluorescence resonance energy transfer



https://www.photometrics.com/learn/physics-and-biophysics/fret

Fluorescence, phosphorescence, luminescence

- Fluorescence and phosphorescence are both **photoluminescence** (i.e. glow is triggered by light) whereas in chemiluminescence, glow is triggered by a **chemical reaction**
- Fluorescence and phosphorescence both **absorb light and emit light of a longer wavelength (and lower energy)**
- Fluorescence is **immediate**
- In phosphorescence, absorbed light can be **stored and emitted later on**

V. Molecular and Cellular Biology in the lab

1. Model organisms
2. Cell cultures
3. Studying proteins
4. **Studying DNA**
    - **DNA sequencing**
    - DNA extraction
    - DNA amplification
    - DNA cloning

Building a DNA toolbox

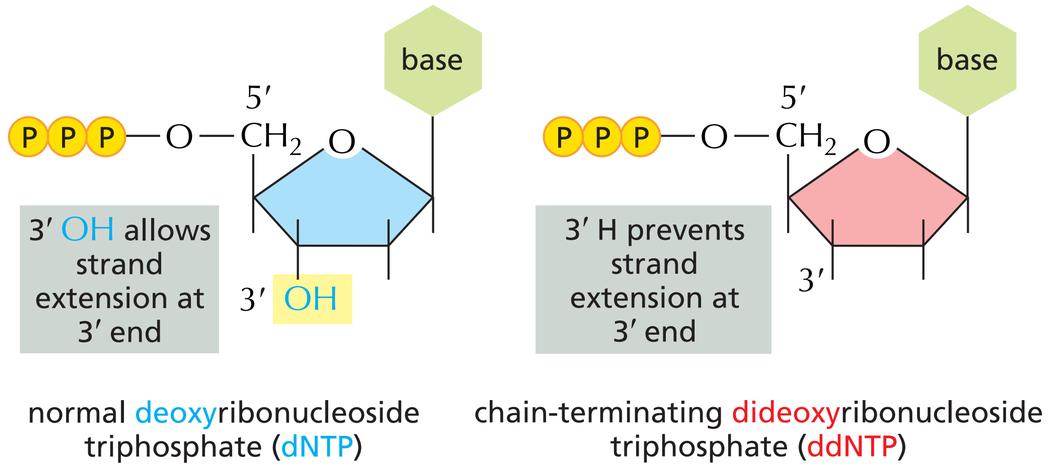
- Based on **DNA sequencing**
- DNA sequencing has allowed advances in **technology**
- Use of **recombinant DNA**, i.e DNA from different sources that is combined
- This is useful for **genetic engineering**, i.e. manipulating genes for practical purposes
- DNA technologies have an **impact** on research, medicine, forensics, agriculture, …

DNA sequencing

- exploits **complementary base pairing**
- developed in the **1970s by Sanger** (Nobel Prize in 1980)
- in 2000s, development of **next-generation sequencing**, which is faster and cheaper: the DNA fragments are amplified, then one strand is immobilized and the complementary strand is synthetized, one nucleotide at a time —> real-time identification of the added nucleotide
- recently, development of **third-generation sequencing**. In some methods, long stretches of DNA are sequenced without cutting or amplifying (e.g. nanopore)

DNA sequencing

- **Sanger or dideoxy sequencing** relies on dideoxy nucleotides that terminate elongation

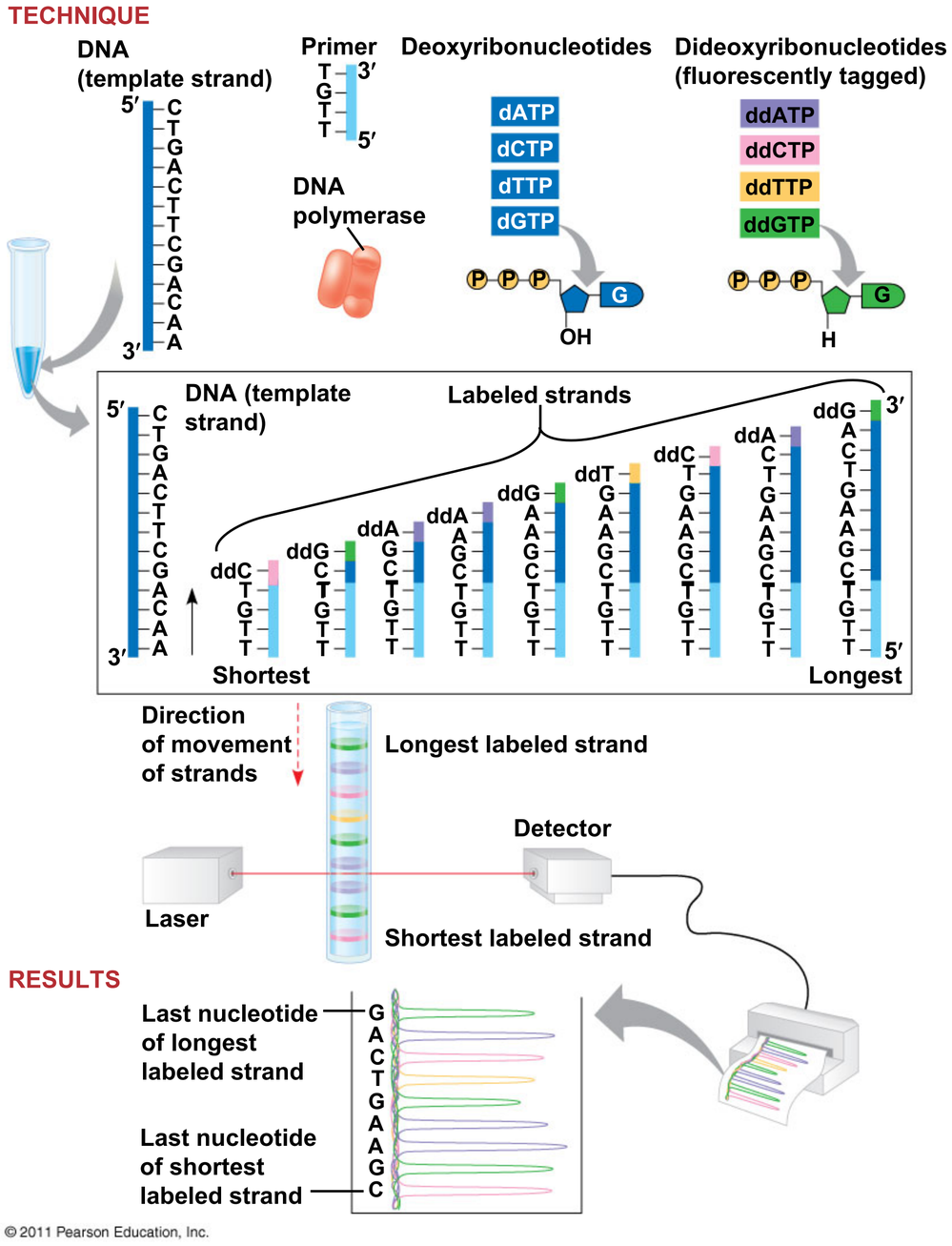


DNA sequencing

- Sanger sequencing is a DNA sequencing method that determines the order of nucleotides (A, T, C, G) in a DNA fragment.
- It relies on **dideoxy nucleotides** that terminate elongation
- Low cost, small scale (short DNA fragments)

Method:

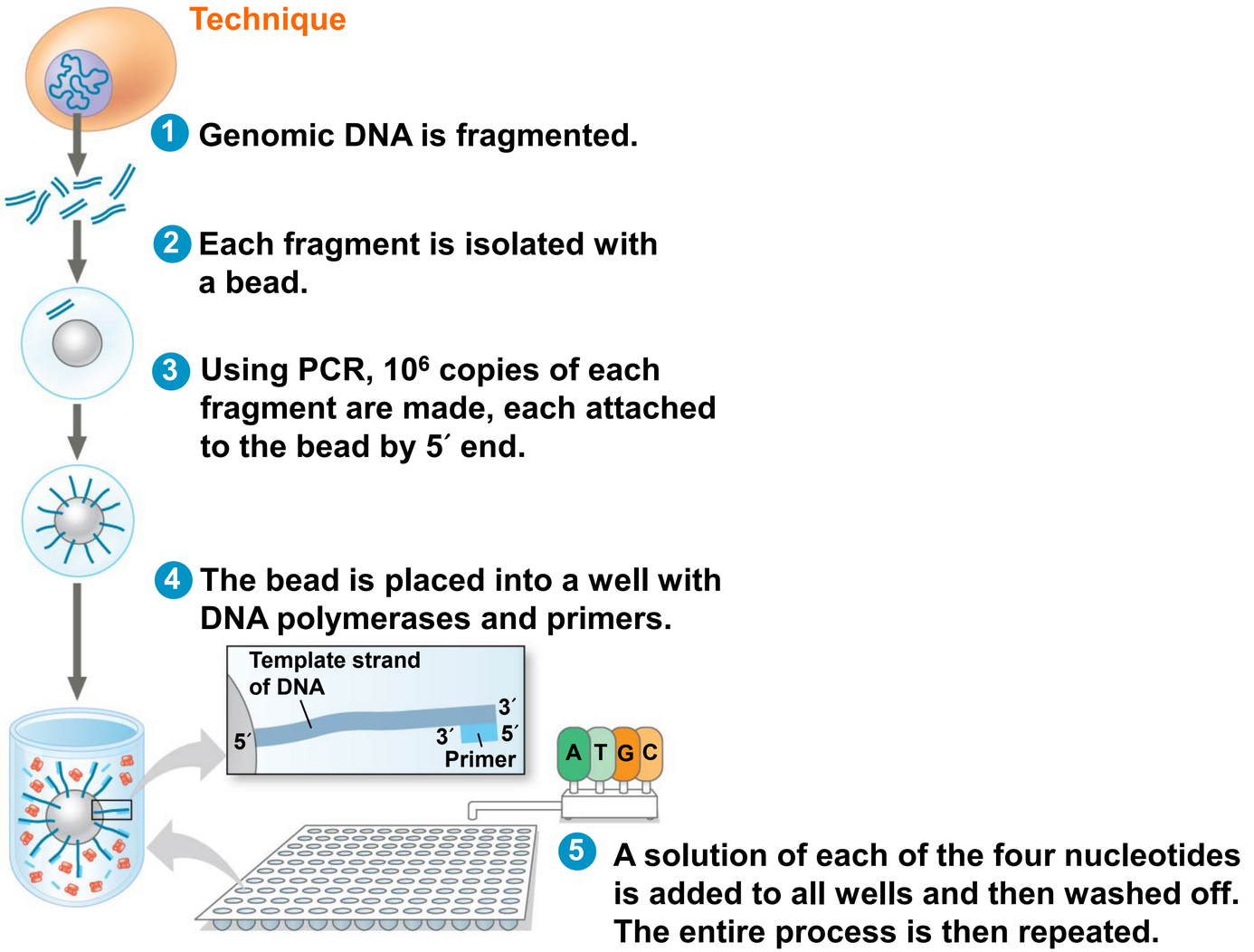
1. A DNA template is copied using normal nucleotides plus special dideoxynucleotides (ddNTPs).
2. When a ddNTP is added, the chain stops because it lacks the 3′-OH group needed for extension.
3. This creates DNA fragments of different lengths, each ending with a labeled ddNTP.
4. The fragments are separated by size (usually by capillary electrophoresis).
5. A detector reads the colored labels on the ddNTPs, revealing the DNA sequence.



DNA sequencing

- **Next-generation sequencing**
  - since **2005**
  - allow **large-scale sequencing**
  - most common is **Illumina Sequencing**
  - **short** DNA sequences (few hundreds nt)
  - **bioinformatic** analysis

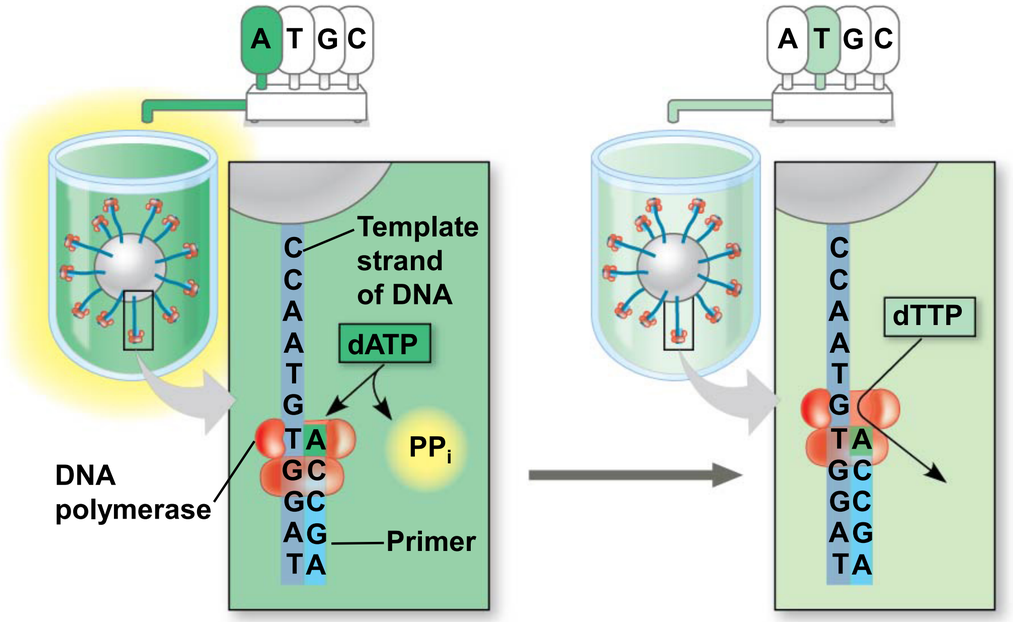
Figure 19.4a



DNA sequencing

- **Next-generation sequencing**
  - since **2005**
  - allow **large-scale sequencing**
  - most common is **Illumina Sequencing**
  - **short** DNA sequences (few hundreds nt)
  - **bioinformatic** analysis

Technique



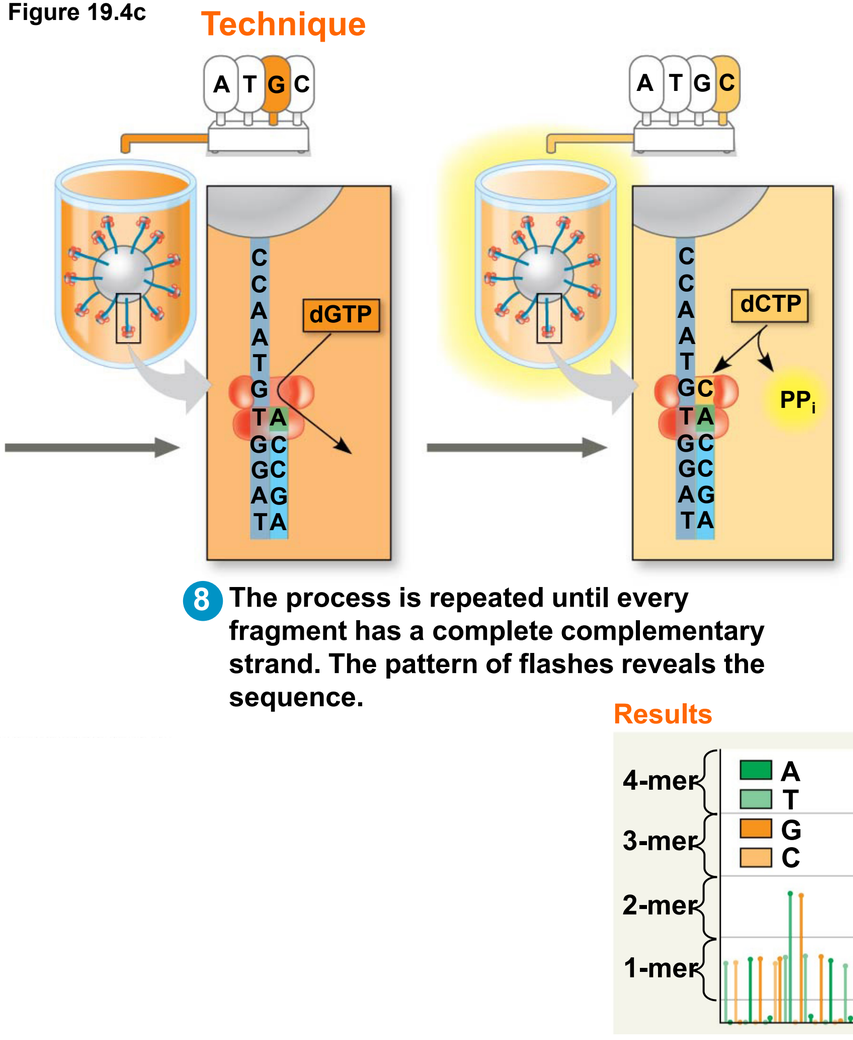
6 **If a nucleotide is joined to a growing strand, PP_i is released, causing a flash of light that is recorded.**

7 **If a nucleotide is not complementary to the next template base, no PP_i is released, and no flash of light is recorded.**

DNA sequencing

- **Next-generation sequencing**
  - since **2005**
  - allow **large-scale sequencing**
  - most common is **Illumina Sequencing**
  - **short** DNA sequences (few hundreds nt)
  - **bioinformatic** analysis

Figure 19.4c



DNA sequencing

- **Third-generation sequencing** refers to DNA sequencing technologies that read **single DNA molecules in real time**, without needing amplification (PCR). This makes them faster and able to produce ultra-long reads.

  - Single-molecule sequencing (no PCR amplification).
  - Real-time detection of nucleotide addition.
  - Very long read lengths (tens to hundreds of kilobases).
  - Useful for analyzing structural variants, repetitive regions, and whole genomes with fewer gaps.

DNA sequencing

- **Third-generation sequencing** refers to DNA sequencing technologies that read **single DNA molecules in real time**, without needing amplification (PCR). This makes them faster and able to produce ultra-long reads.

Main technologies

A. PacBio SMRT sequencing —> Detects fluorescently labeled nucleotides as a polymerase incorporates them.

B. Oxford Nanopore sequencing —> DNA passes through a nanopore and changes in ionic current identify each nucleotide.

- Ultra-long reads (can exceed 1 Mb).
- Portable devices (e.g., MinION).

Advantages

- Very long reads.
- Faster sequencing.
- Better for genome assembly, structural variants, epigenetics (e.g., methylation detection).

Limitations

- Higher raw error rates (though improving).
- More expensive per read than some short-read methods.

Gene annotations

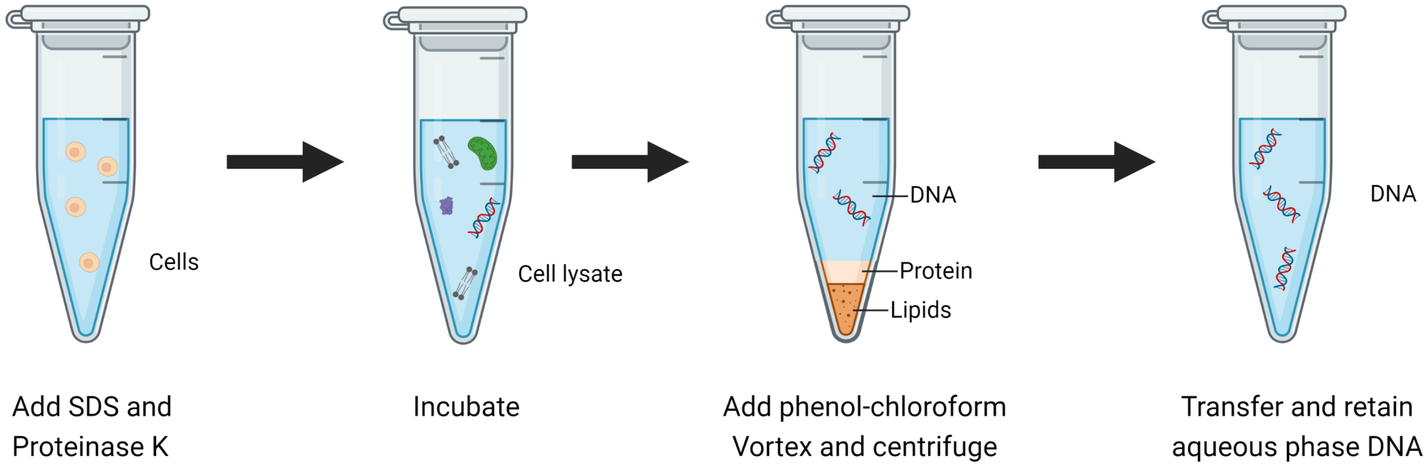
- **Mark** the genes
- Assign possible **roles**

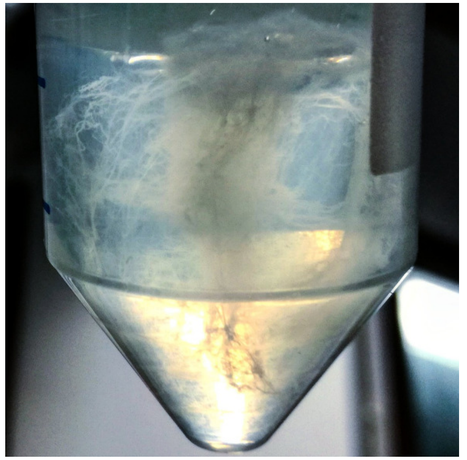
V. Molecular and Cellular Biology in the lab

1. Model organisms
2. Cell cultures
3. Studying proteins
4. **Studying DNA**
    - DNA sequencing
    - **DNA extraction**
    - DNA amplification
    - DNA cloning

How is DNA extracted?

- **Goal:** isolate the DNA without its associated proteins

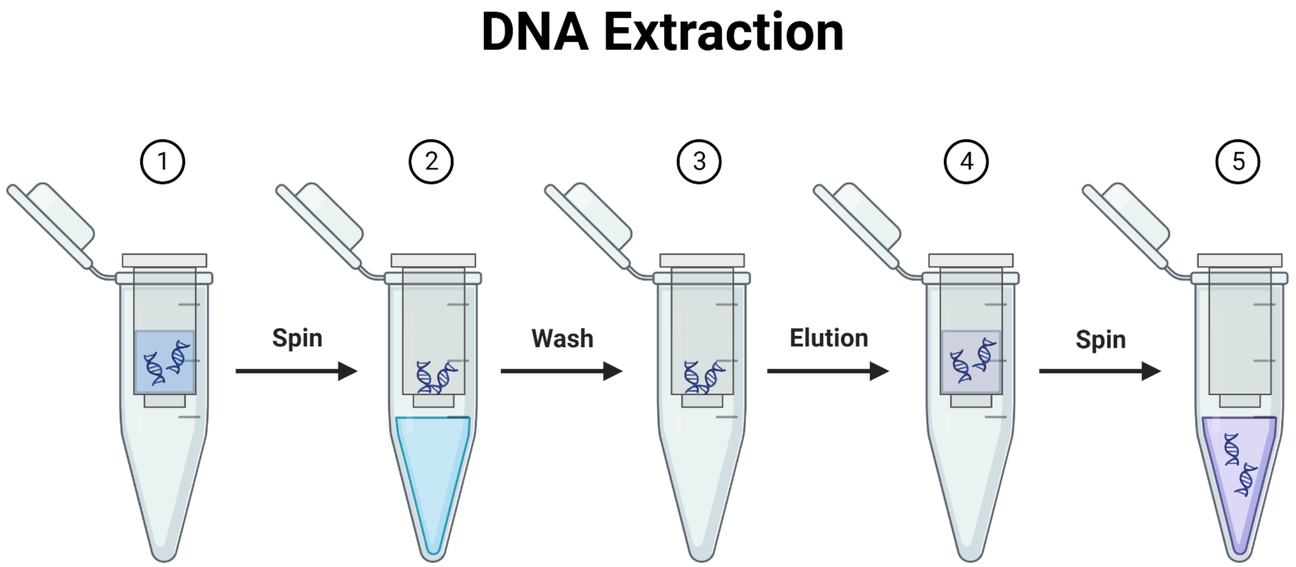




- Phenol-chloroform is less polar than water and induces protein aggregation
- DNA is further precipitated with ethanol

But how is DNA extracted?

- **Goal:** isolate the DNA without its associated proteins

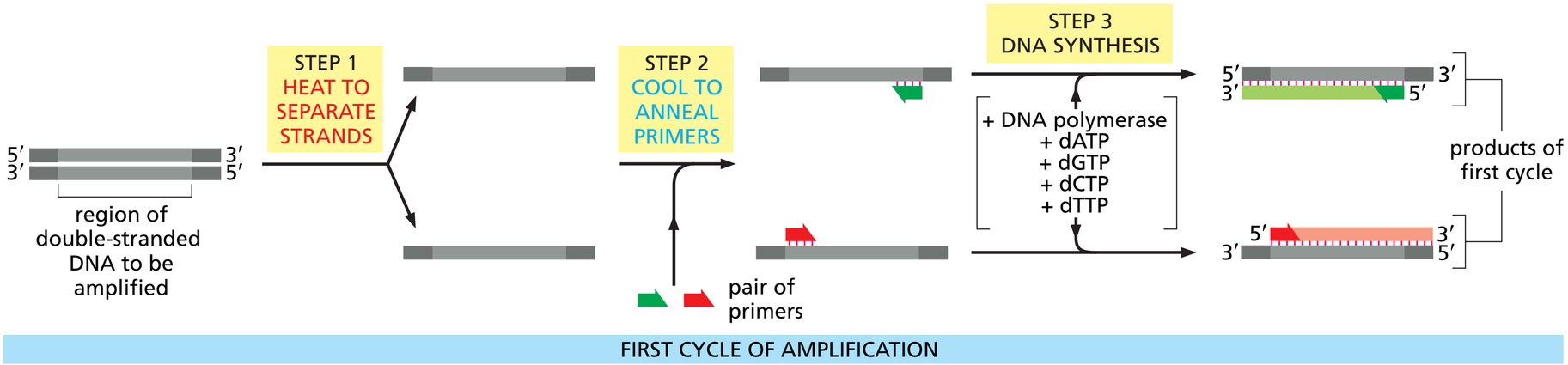


V. Molecular and Cellular Biology in the lab

1. Model organisms
2. Cell cultures
3. Studying proteins
4. **Studying DNA**
    - DNA sequencing
    - DNA extraction
    - **DNA amplification**
    - DNA cloning

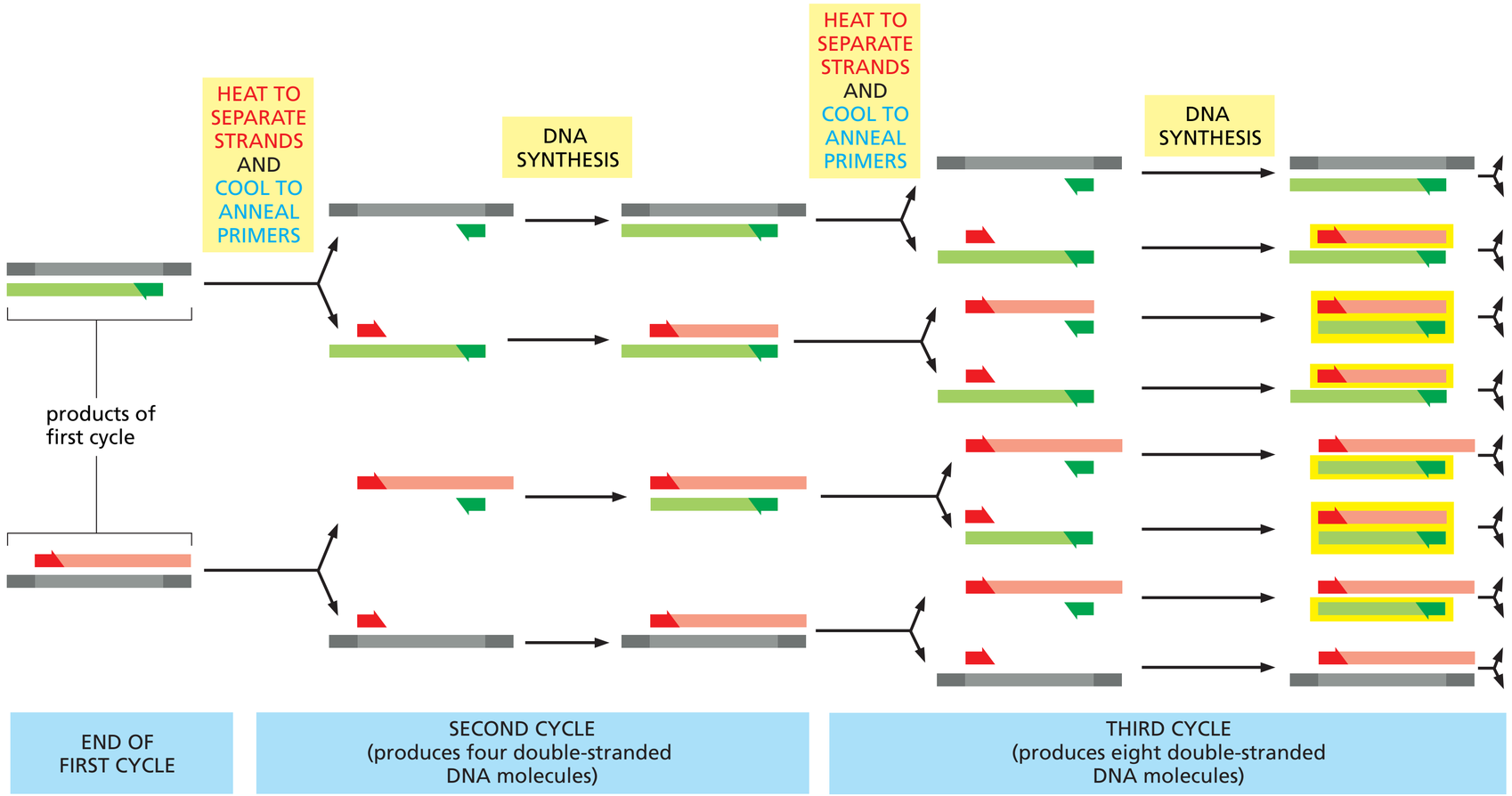
Amplifying specific regions of DNA

- To work directly on **specific regions/genes**, we amplify these regions to obtain multiple identical copies
- Polymerase Chain Reaction (**PCR**) for specific **DNA region amplification**
    - design the **DNA primers** needed by the **DNA polymerase**
    - need **nucleotides**
    - get **billions of copies** of the original sequences after **20-30 cycles**



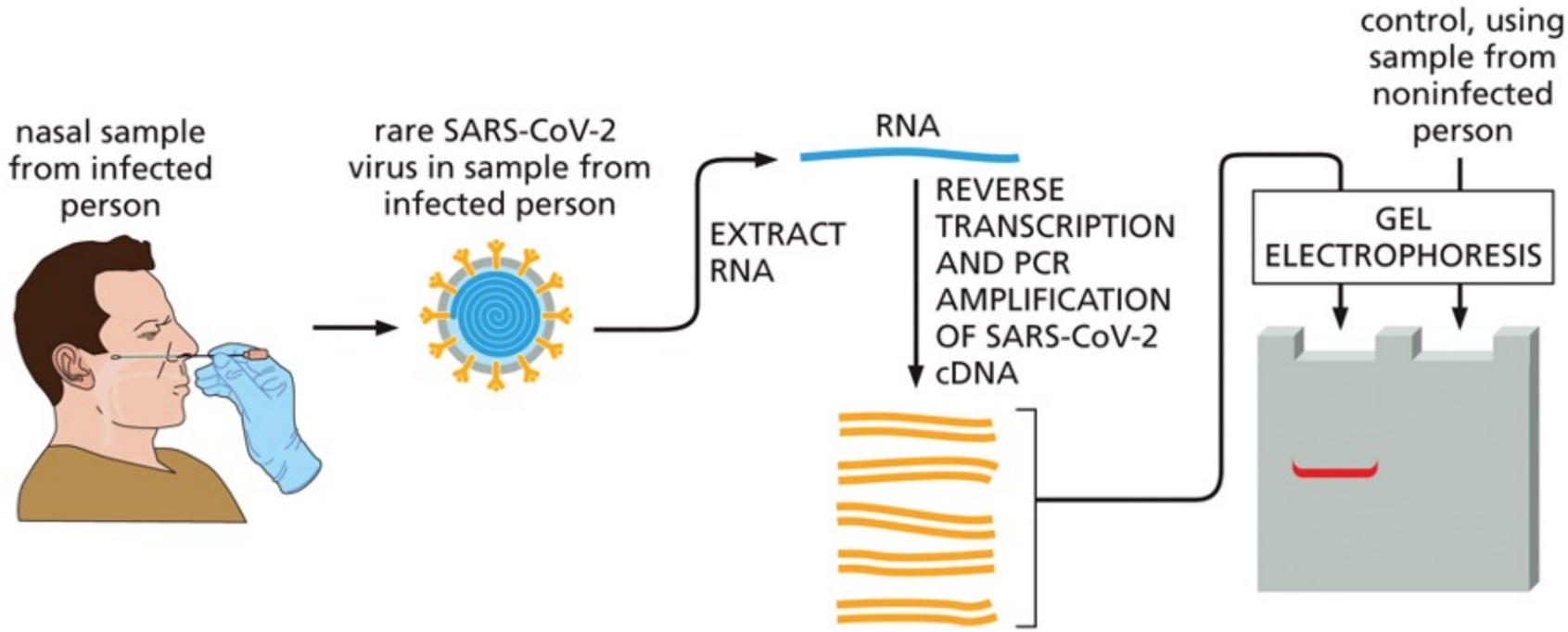
Amplifying specific regions of DNA

- Polymerase Chain Reaction (**PCR**) for specific **DNA region amplification**



Amplifying specific regions of DNA

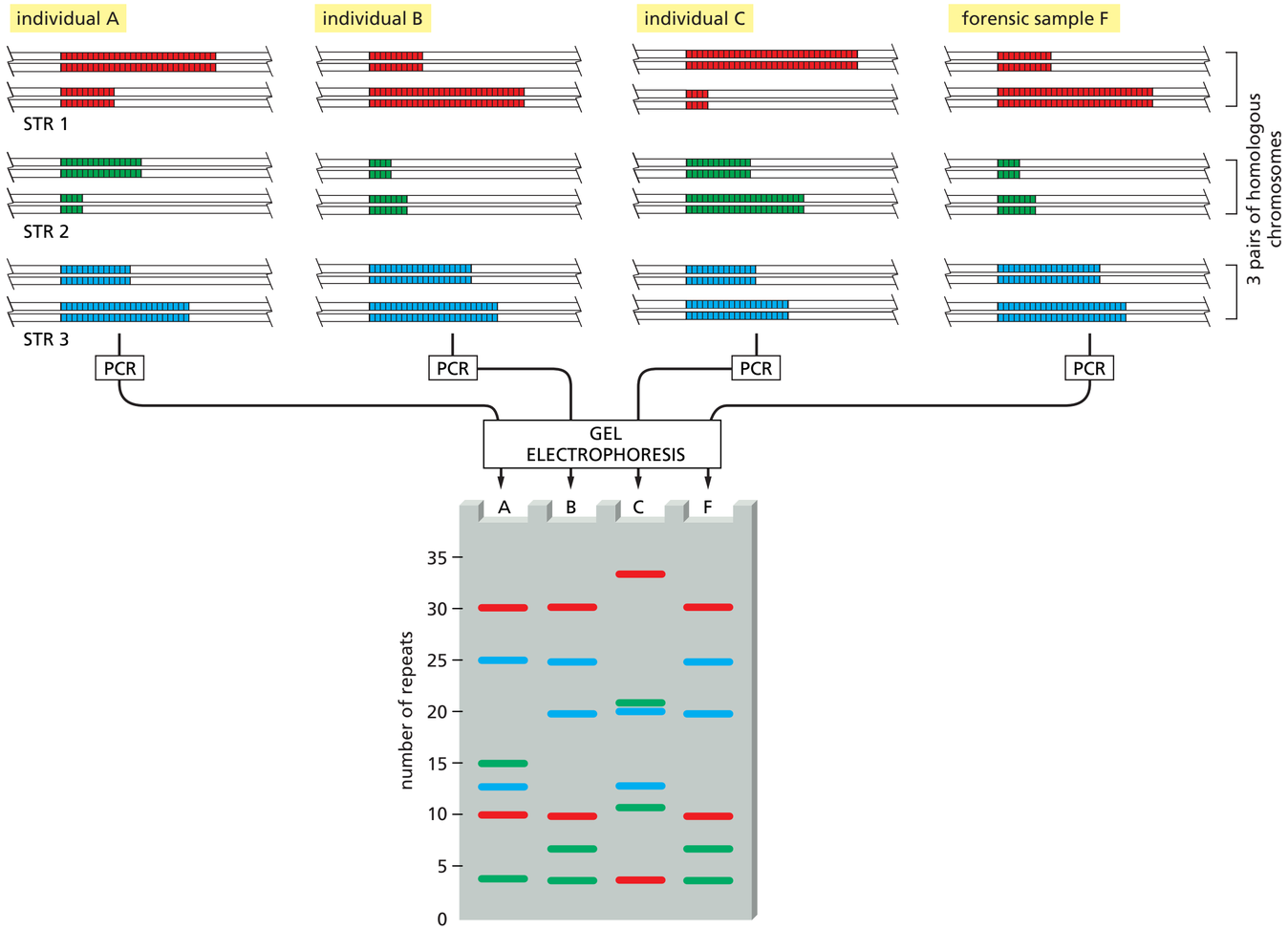
- Polymerase Chain Reaction (**PCR**) for specific **DNA region amplification - diagnostic or forensics**



Amplifying specific regions of DNA

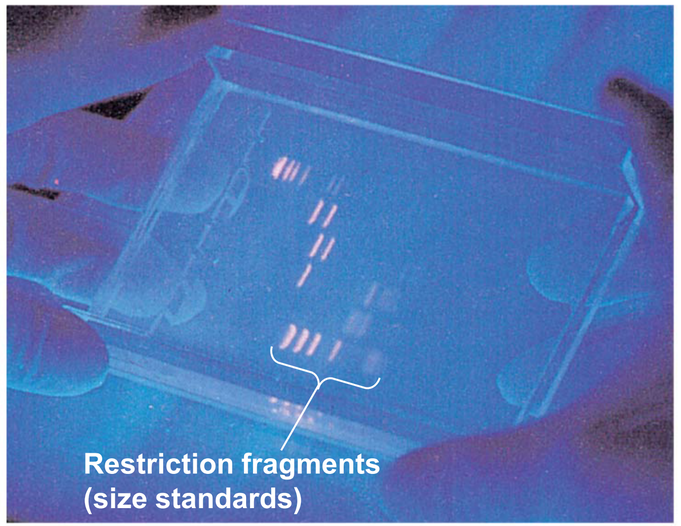
- Polymerase Chain Reaction (**PCR**) for specific **DNA region amplification - diagnostic or forensics**

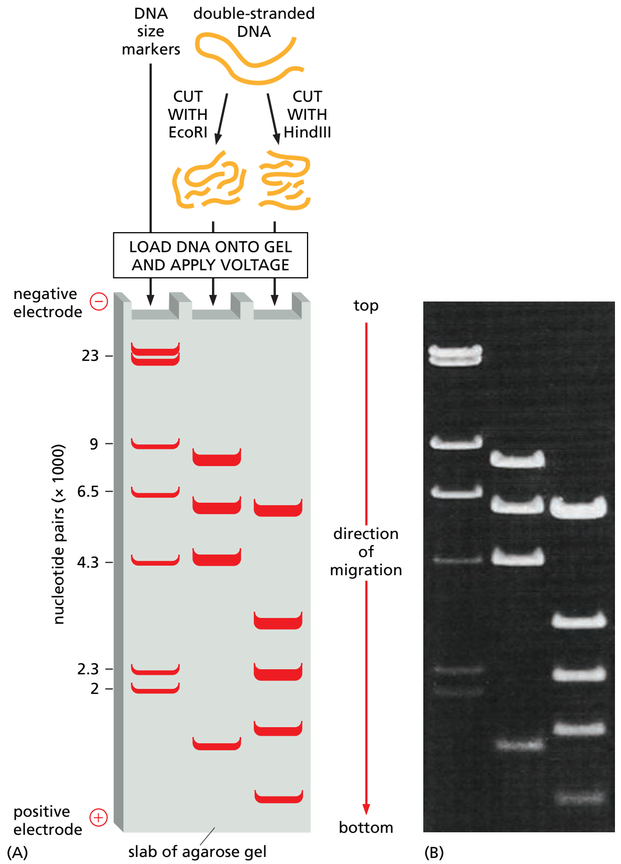
STR = short tandem repeats



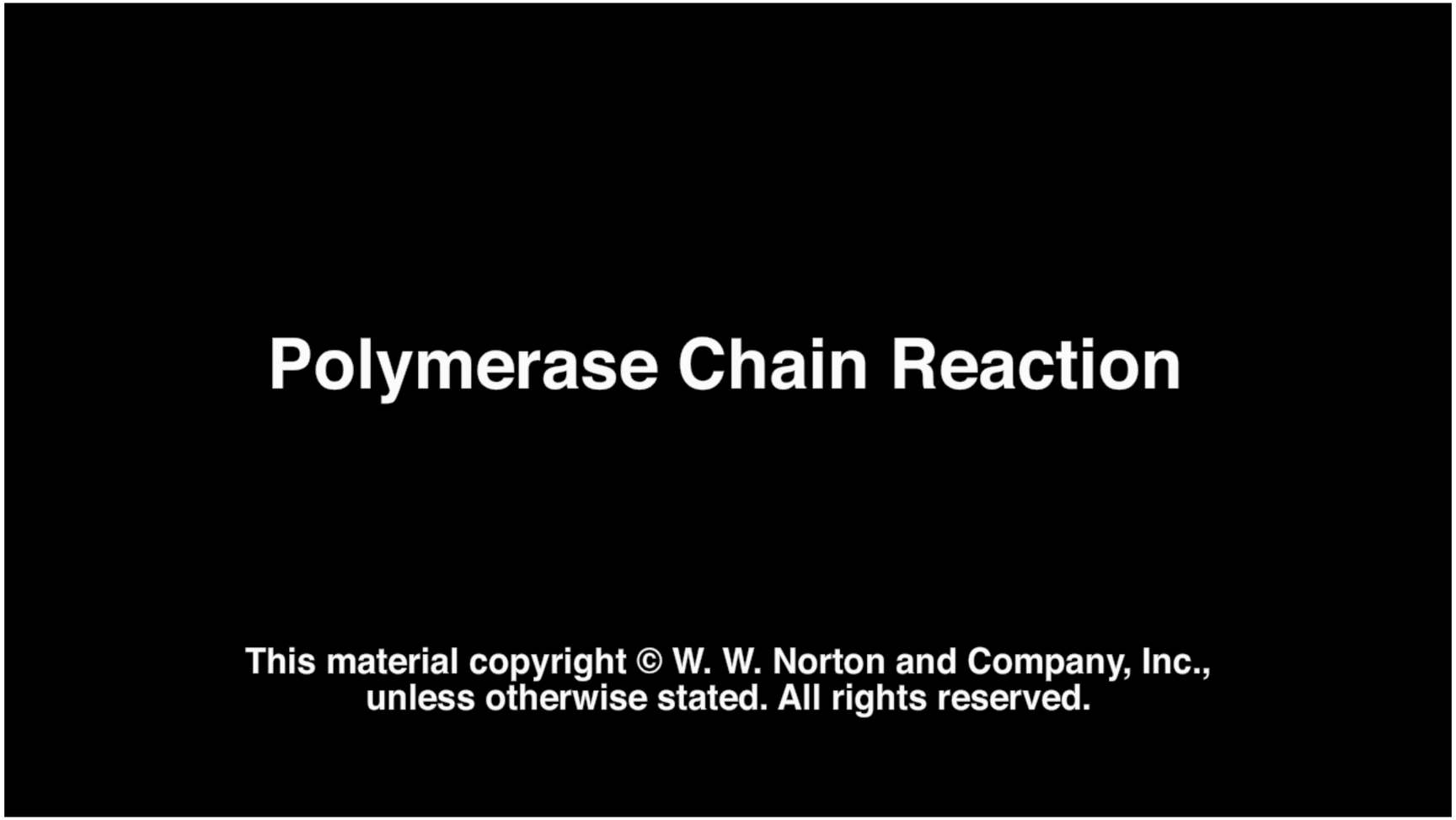
Visualizing DNA

- **Gel electrophoresis** separates DNA fragment by size
  - As for proteins, but each nucleotide already carries a **negative charge** on the phosphate group
  - Agarose gel
  - Use of an **intercalating agent** that **binds DNA** and fluoresces under **UV**





Amplifying specific regions of DNA

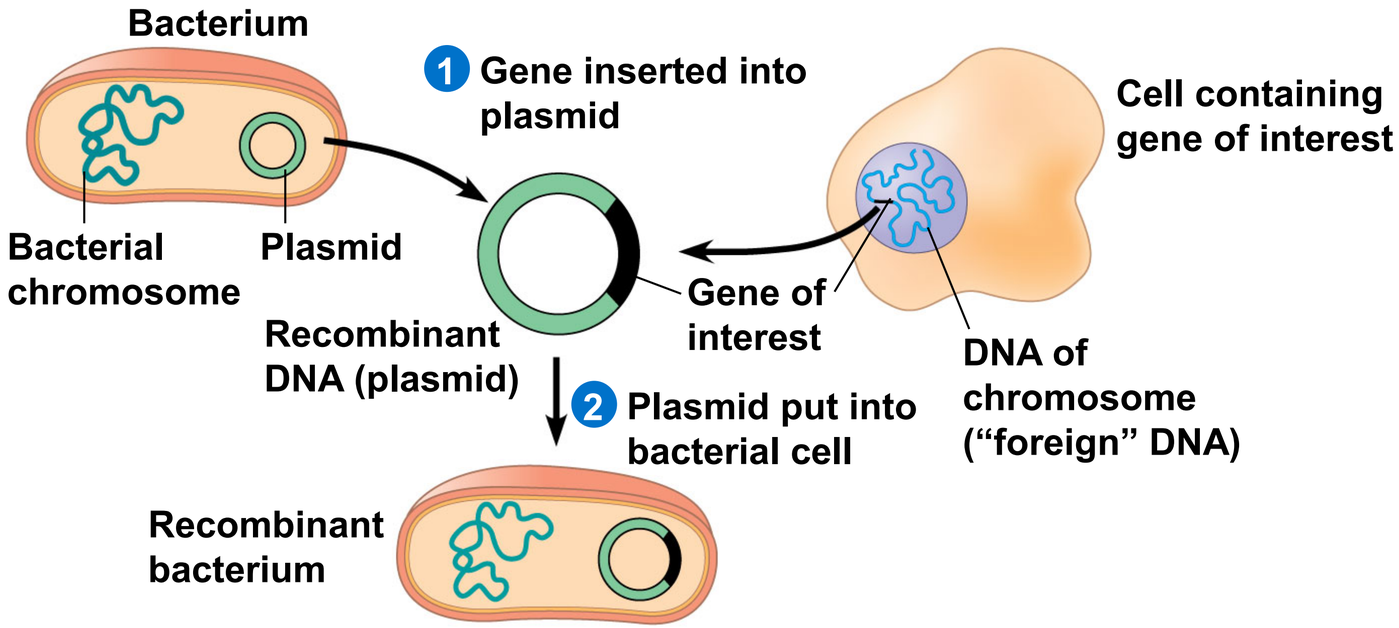


V. Molecular and Cellular Biology in the lab

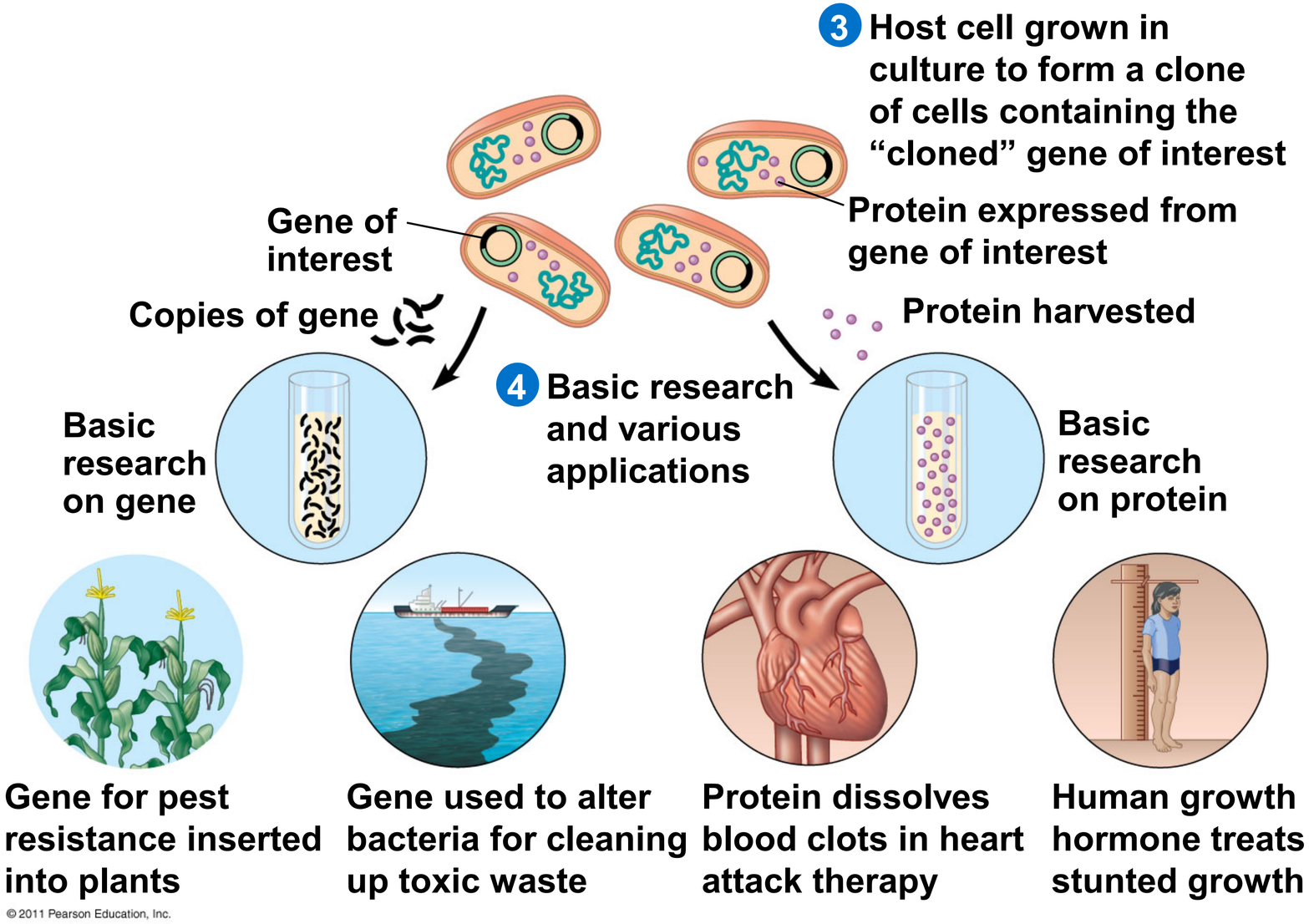
1. Model organisms
2. Cell cultures
3. Studying proteins
4. **Studying DNA**
    - DNA sequencing
    - DNA extraction
    - DNA amplification
    - **DNA cloning**

DNA cloning

- **Plasmids** are small circular DNA molecules that are present in bacteria and replicate separately from the bacterial chromosome
- We can insert **DNA into plasmids** to produce recombinant DNA
- This is then **inserted into bacteria** to make copies of the plasmid, to express a protein, etc.

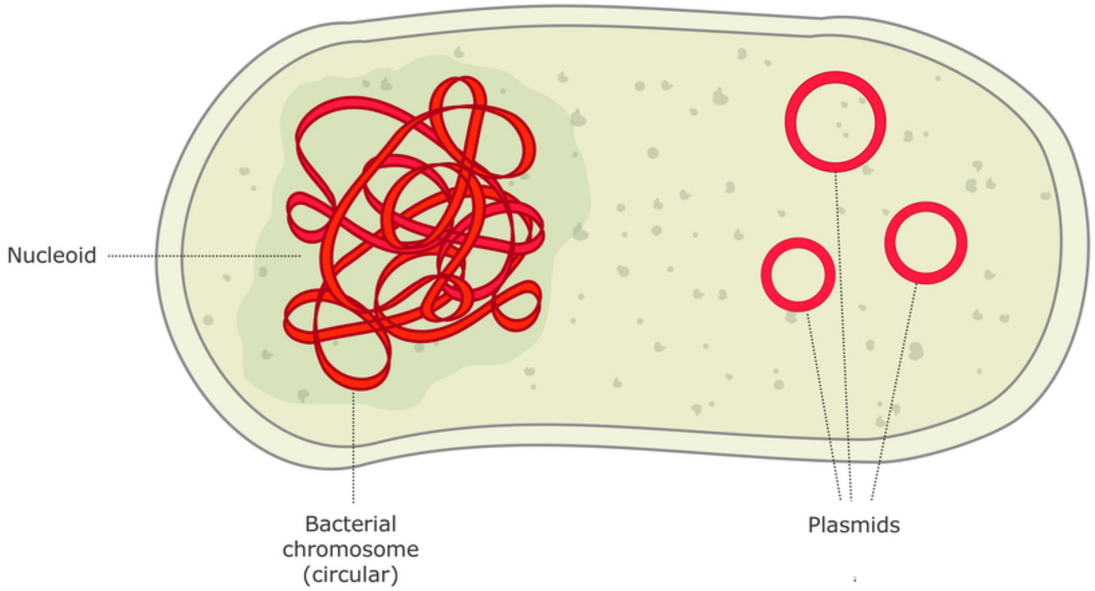


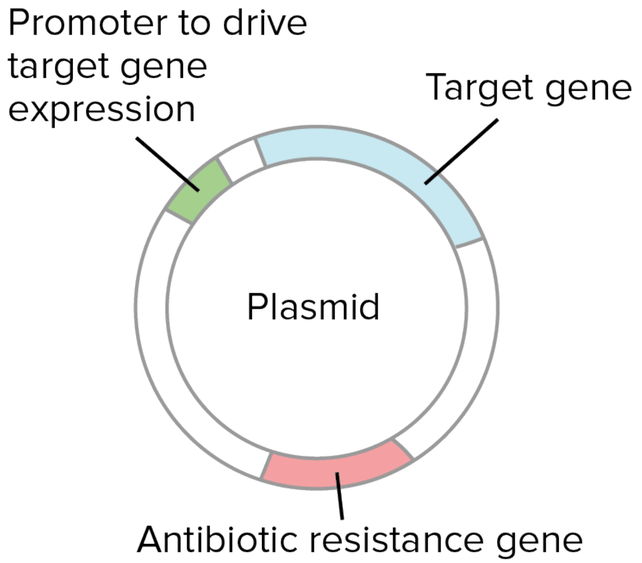
DNA cloning



Preparing plasmids

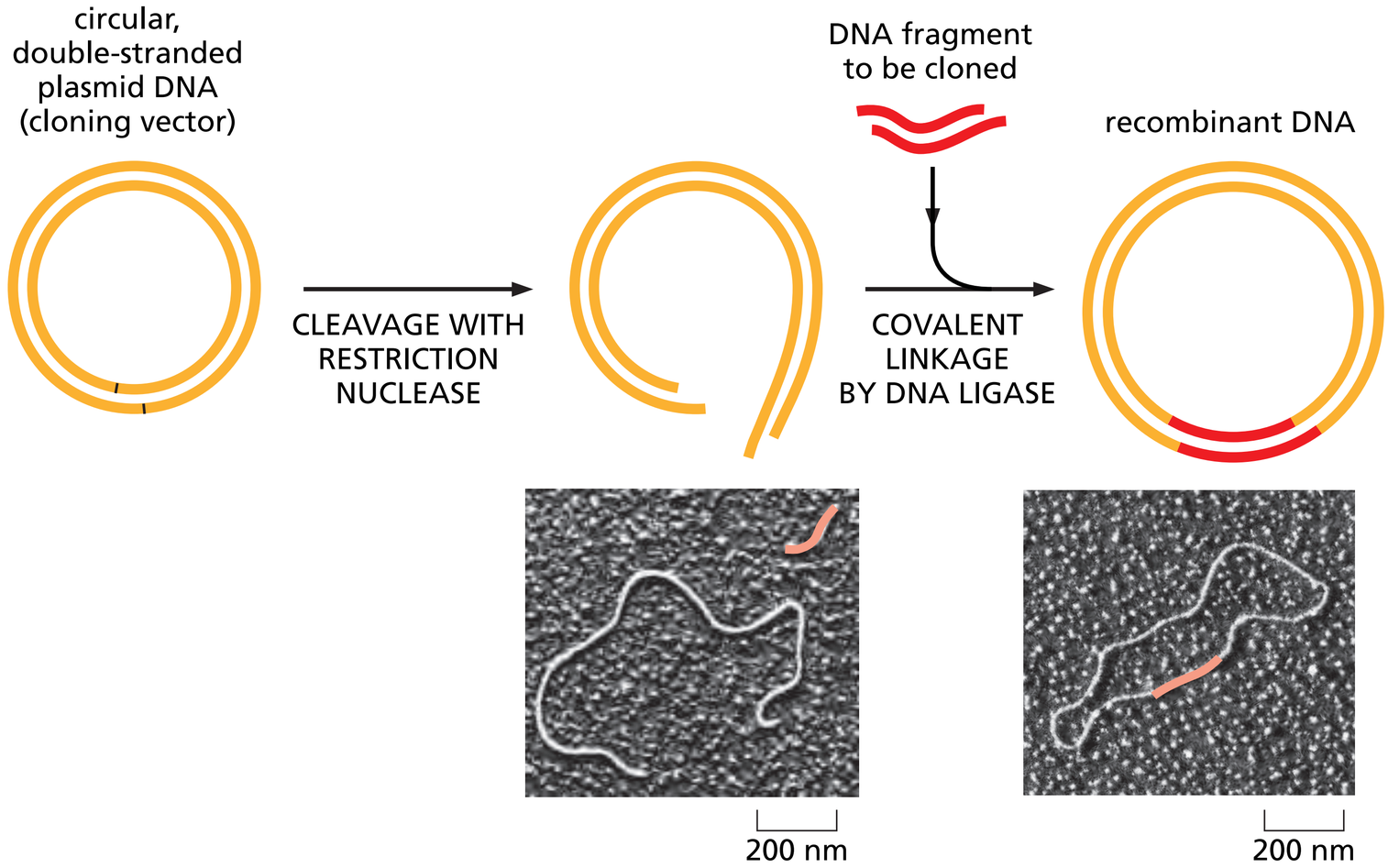
- **Cloning** genes using **bacteria** and **plasmids** (self-replicating circular dsDNA)





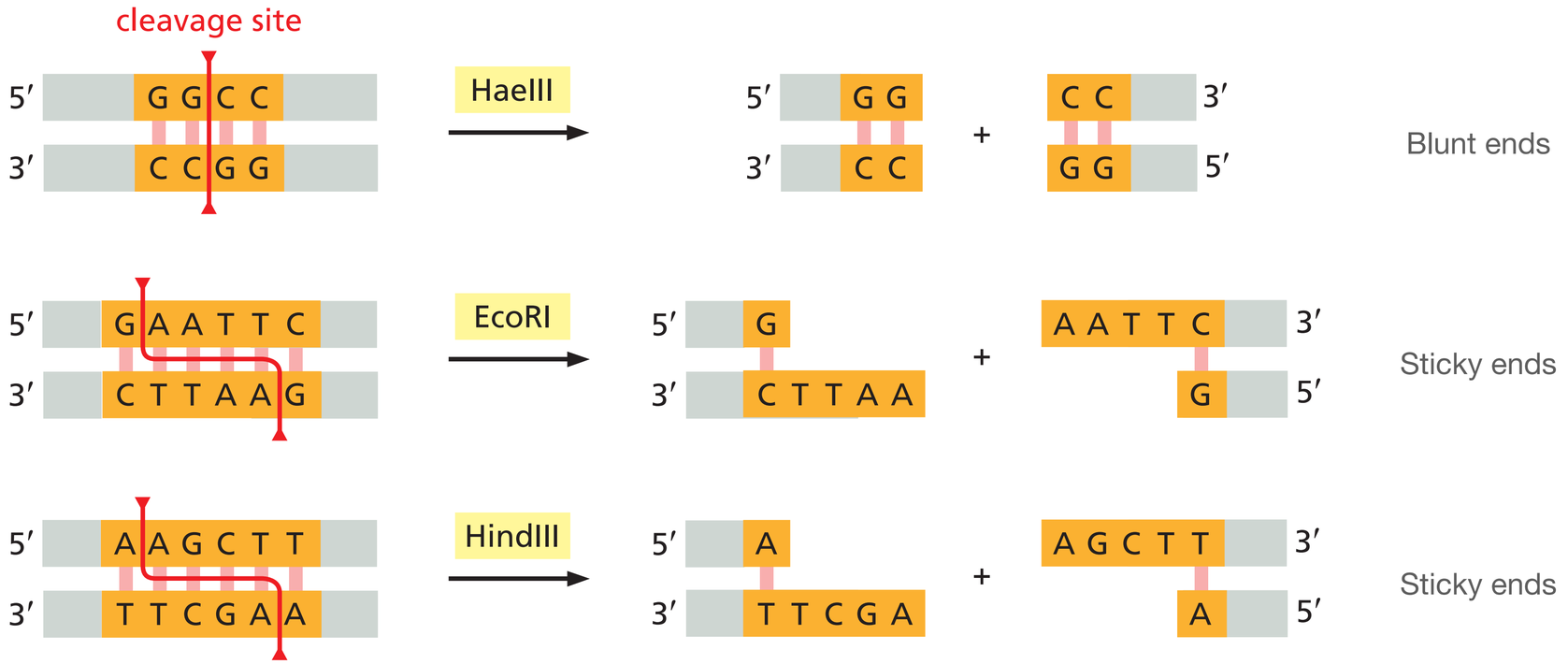
Preparing plasmids

- **Cloning** genes using **bacteria** and **plasmids**



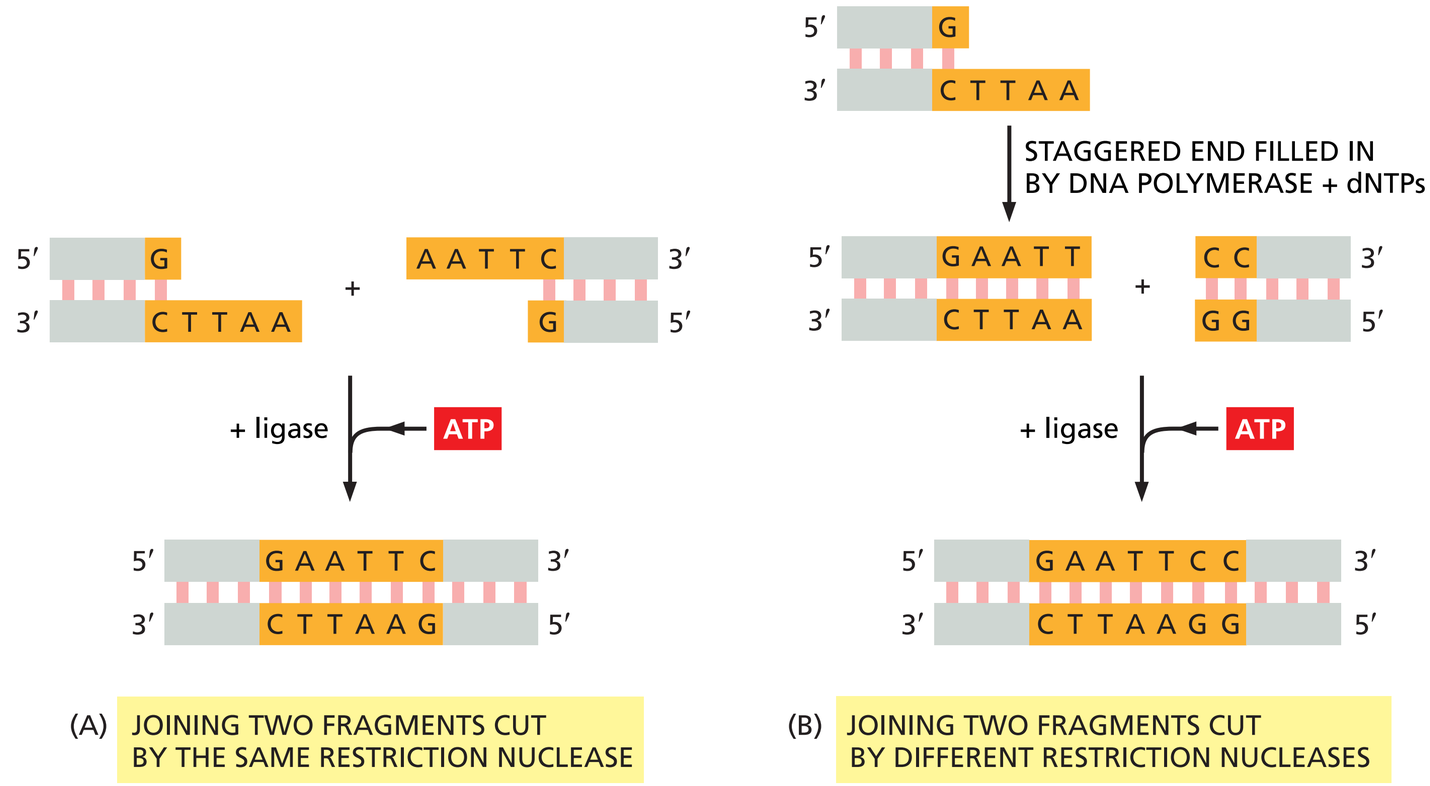
1. Cloning with restriction enzymes

- **Restriction nucleases** are enzymes that **cut** DNA at specific sequences
  - Purified from bacteria where they serve as **foreign DNA defense systems**

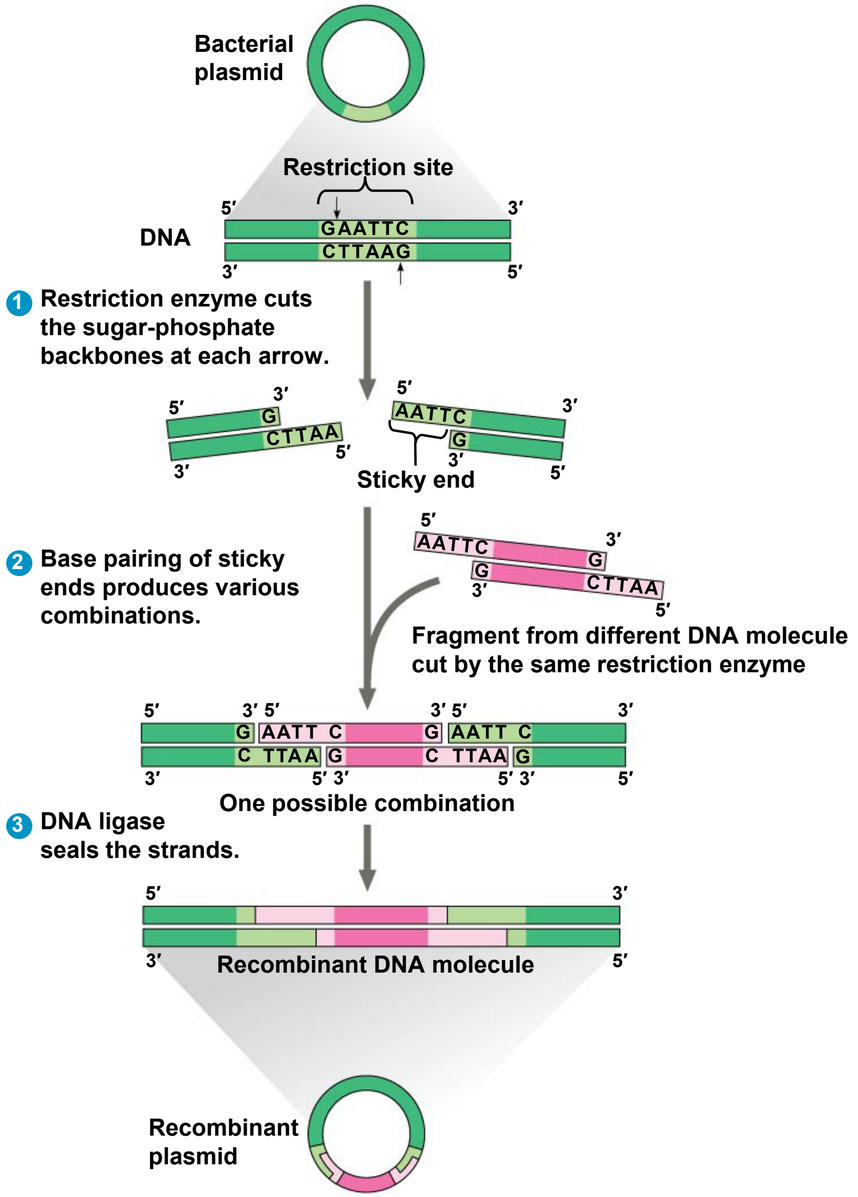


1. Cloning with restriction enzymes

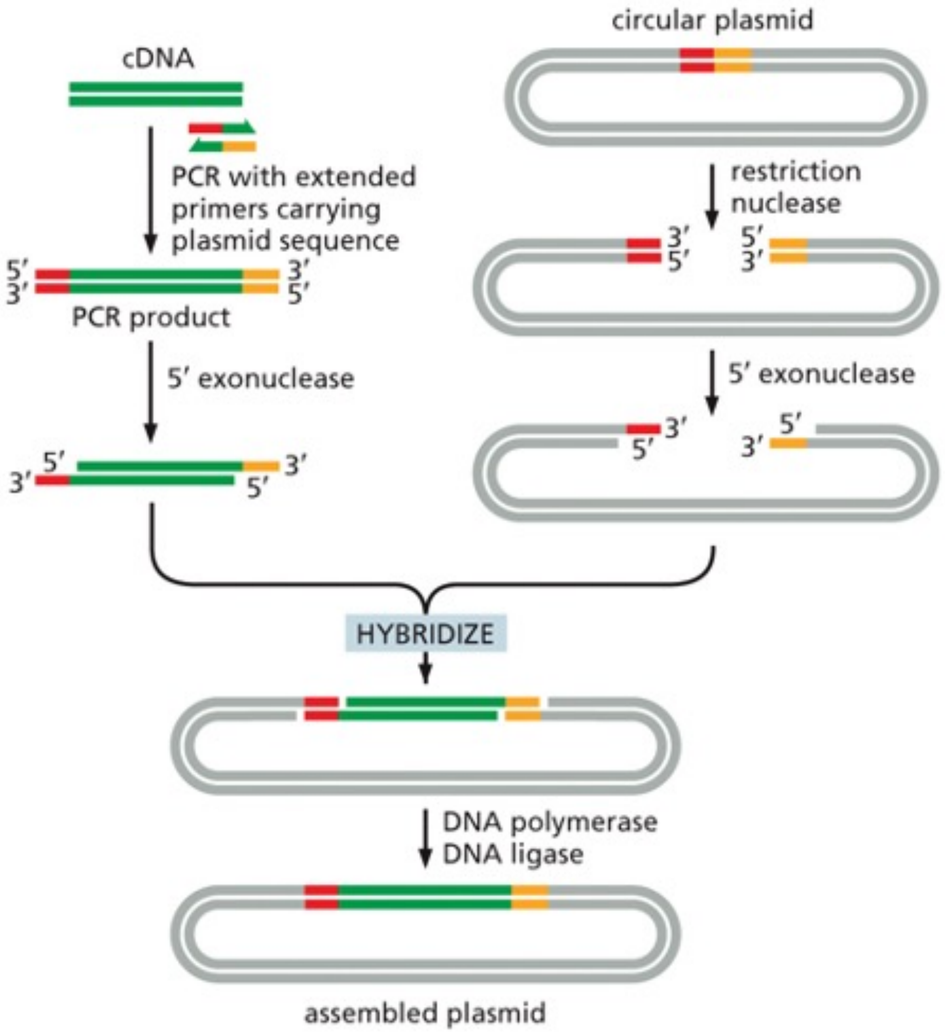
- DNA **ligation**



1. Cloning with restriction enzymes

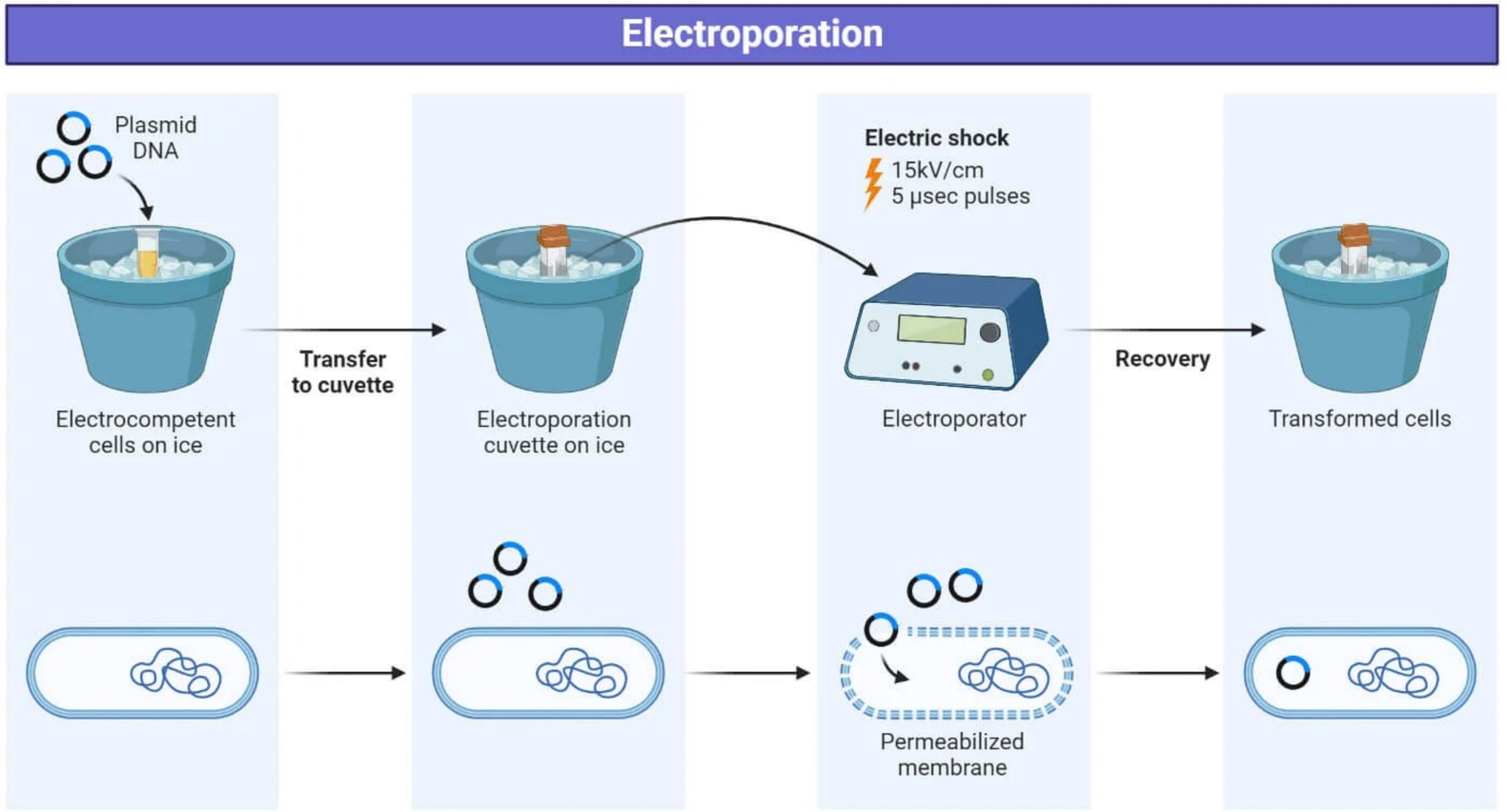


2. Cloning using Gibson Assembly



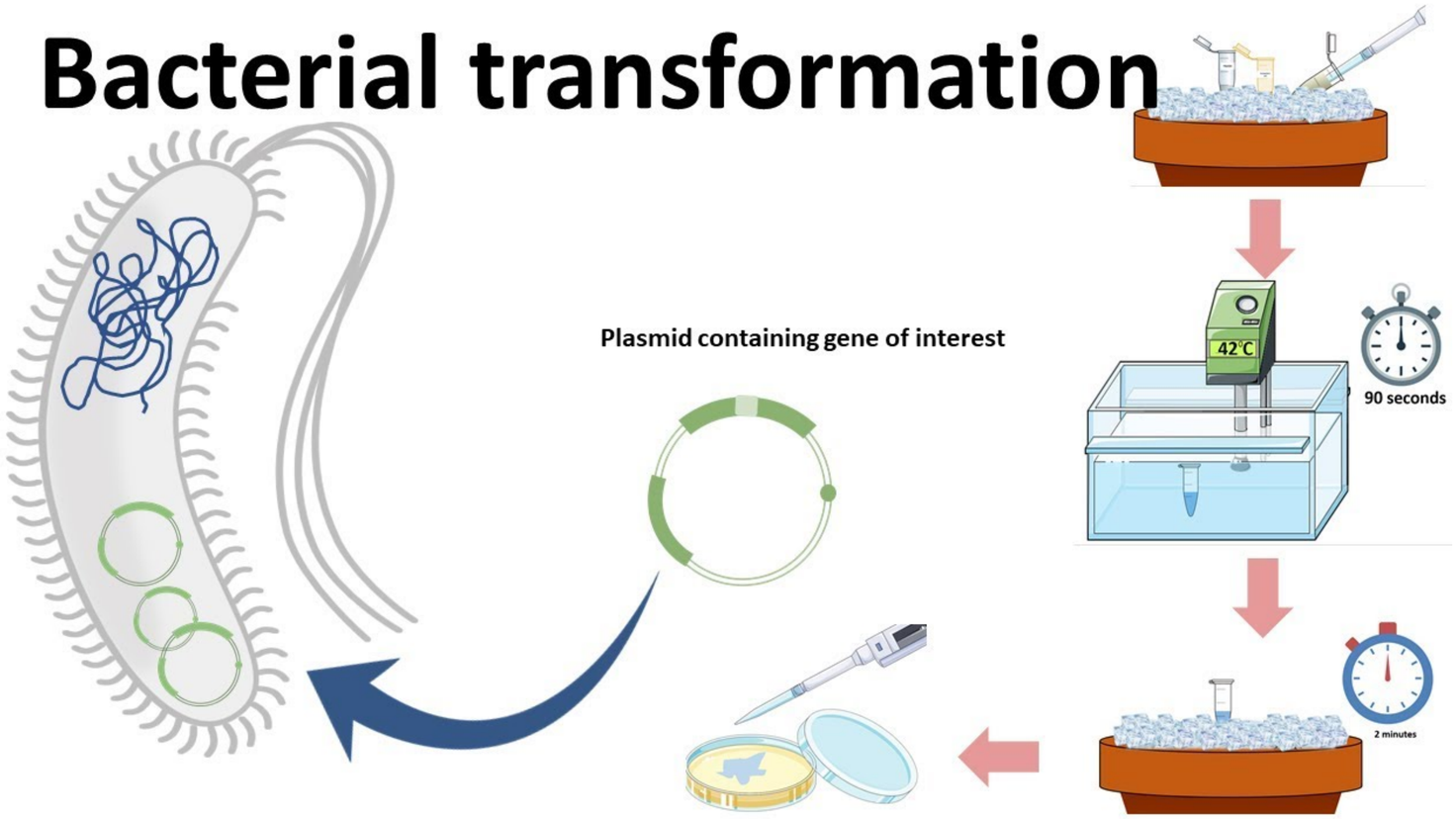
Transforming plasmids into bacteria

- using electroporation



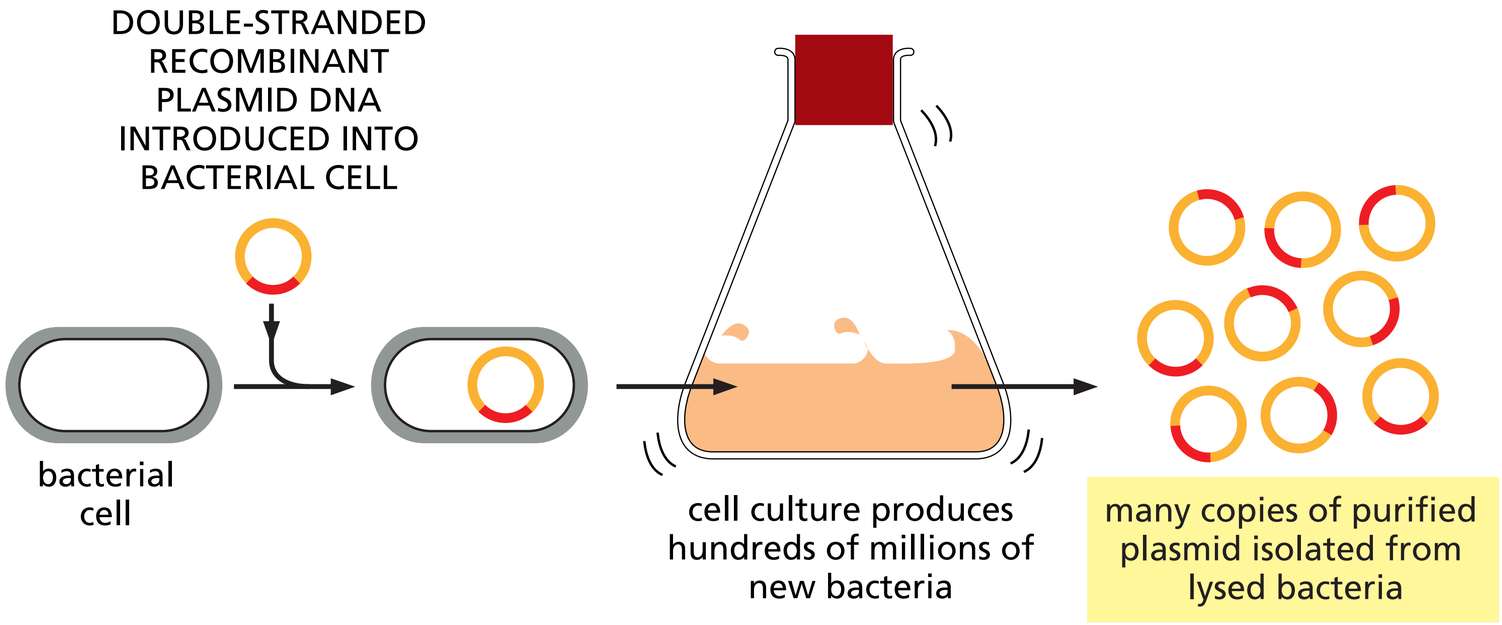
Transforming plasmids into bacteria

- using heat shock



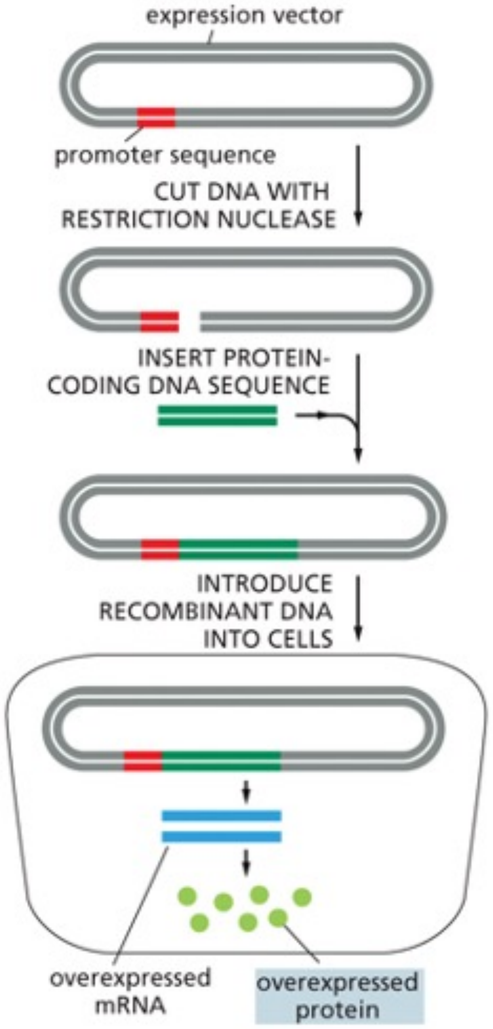
Growing the transformed bacteria

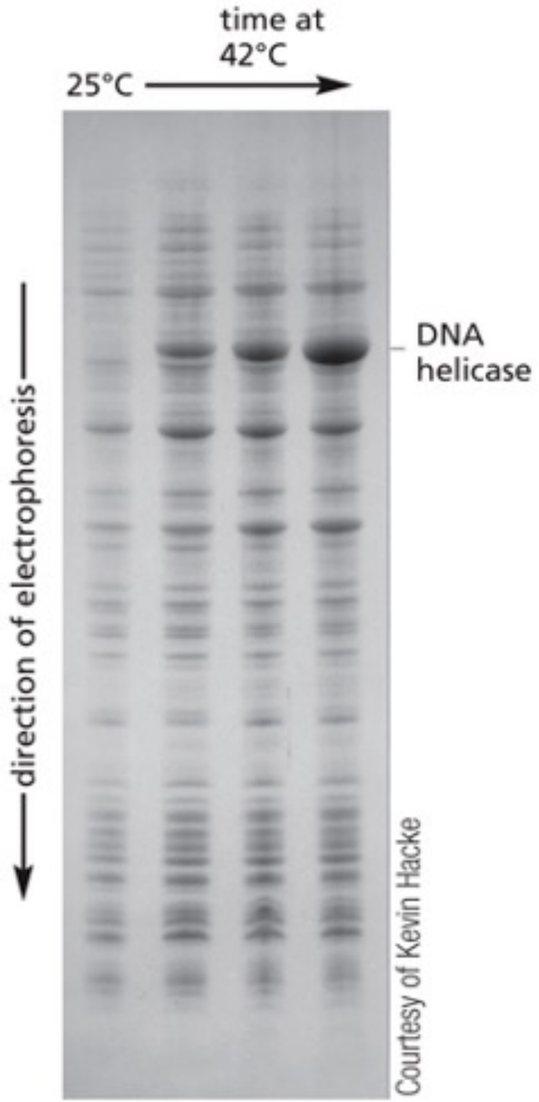
- **Growing** the bacterial culture containing the plasmid of interest



- Purify the plasmid
- Express a protein of interest
- Overexpress a protein for purification
- . . .

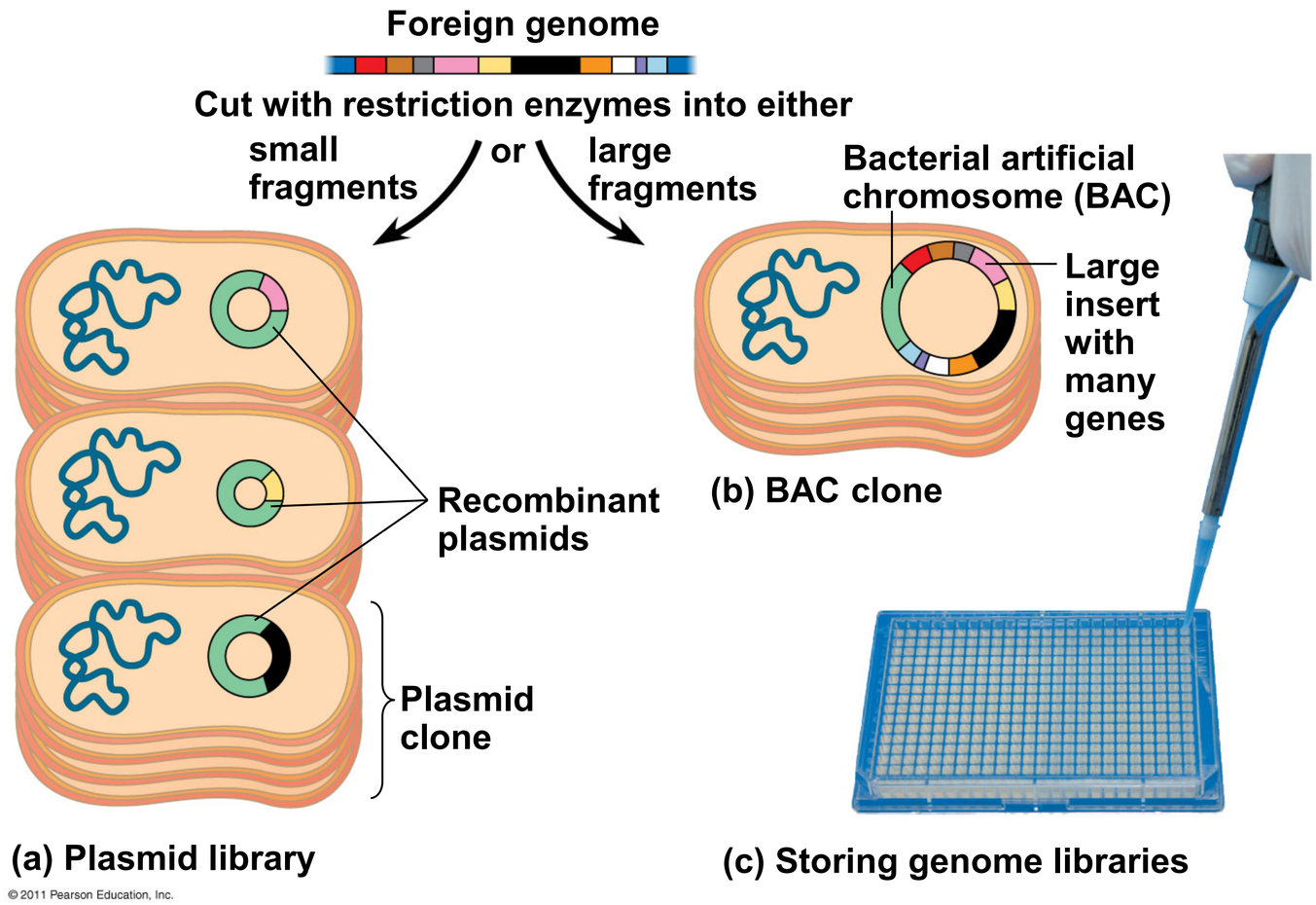
Protein overexpression





DNA libraries

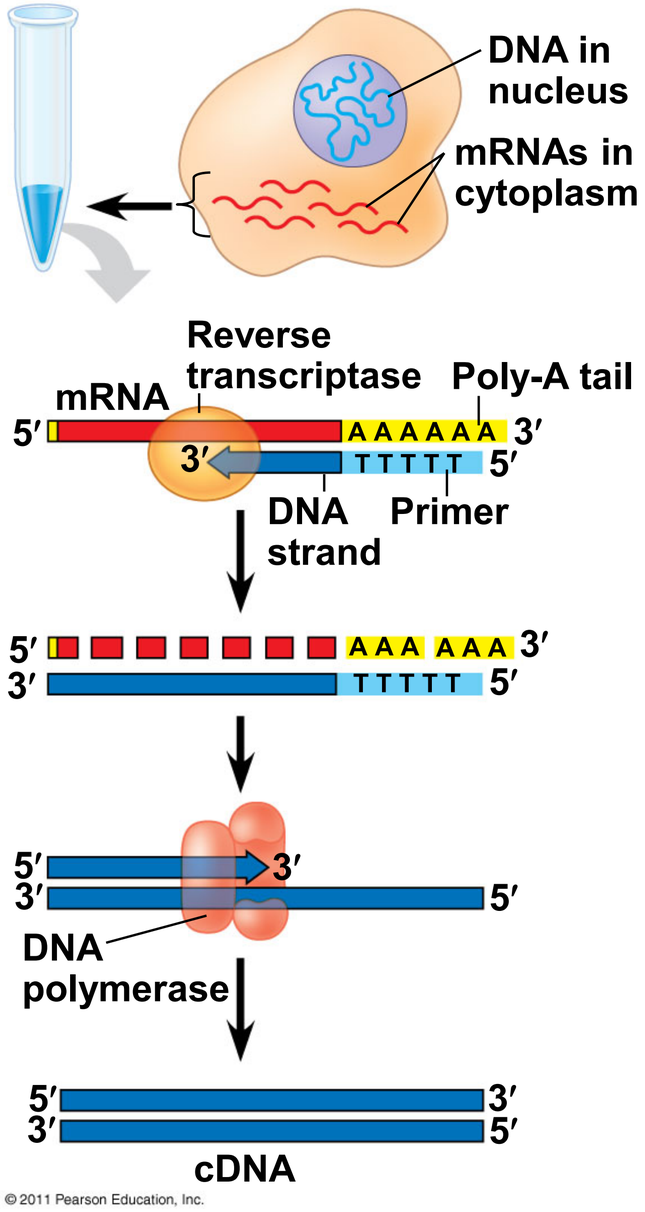
- A **genomic library** is a collection of bacteria each carrying a plasmid with one DNA fragment coming from an entire genome
- The library aims at covering the **whole genome**
- A **bacterial artificial chromosome** (BAC) is a large plasmid that can carry large DNA inserts



DNA libraries

- A **cDNA library** is made by cloning DNA made *in vitro* by reverse transcription of all the mRNA produced in the cell

- It represents **only one part of the genome**, the part being transcribed



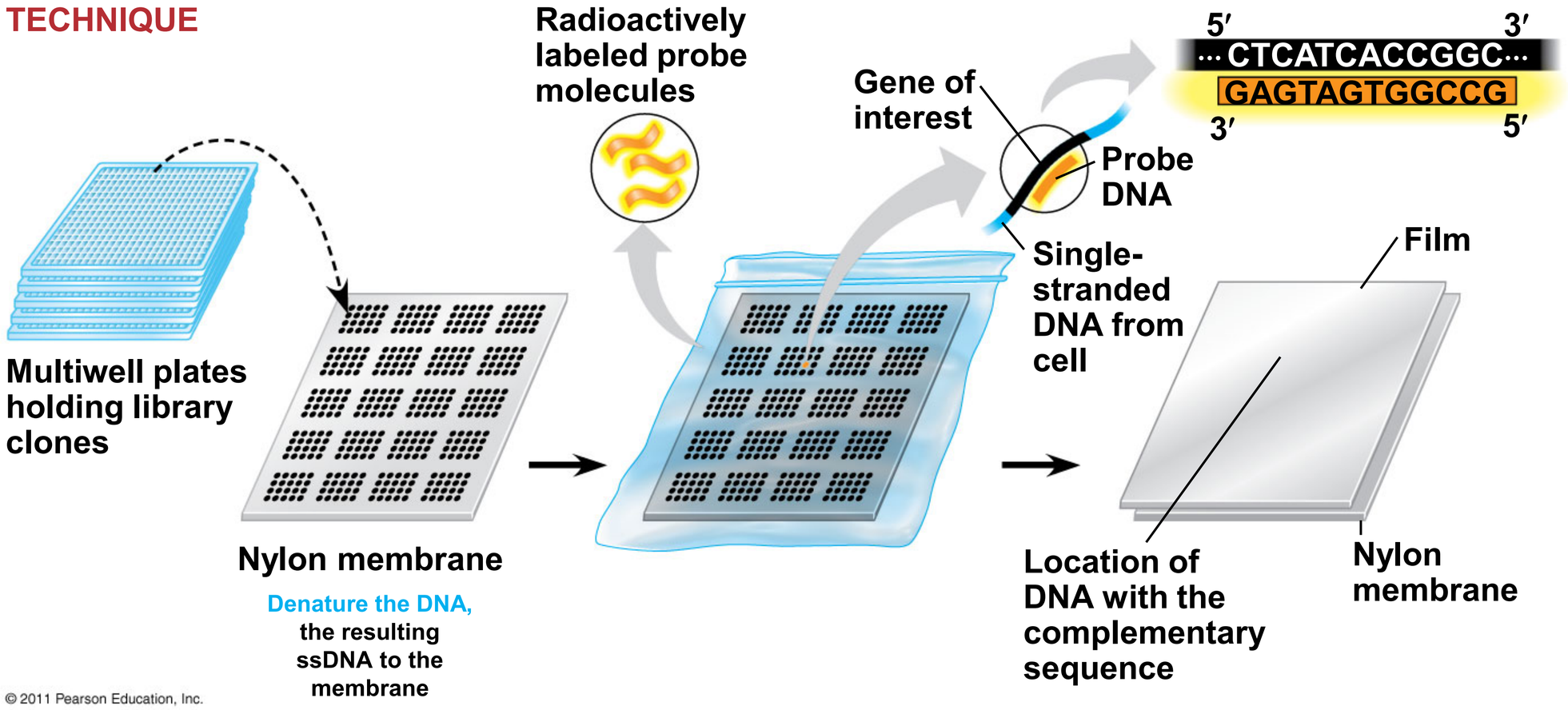
DNA libraries

- Libraries can be screened to identify clones carrying a gene of interest using **nucleic acid probes**
- This based on nucleic acid **hybridization**

Sequence in the gene of interest

Probe

DNA libraries



Expressing eukaryotic genes

- After cloning a gene, its **protein product** can be produced in **large amounts**
- They can be expressed in **bacteria or eukaryotic cells**

- What are the **difficulties** when expressing a eukaryotic protein in bacterial cells?
    - Differences in **promoters** -> use of expression vectors with strong promoters
    - Presence of **introns** -> use of cDNA
    - Protein **glycosylation**

Eukaryotic cloning systems

- Eukaryotic genes may **not be expressed properly** in bacteria (introns, post-translational modifications, …)
- Researchers have then to use eukaryotic cells, like **yeasts**
- In some cases, yeasts do not have proteins required to modify a mammalian protein properly, in which case **insect cells** or **cultured mammalian cells** are used

- Recombinant DNA is typically introduced by **electroporation** or by **injection** (with microscopically thin needles)
- Once inside the cell, DNA is incorporated by **natural recombination**

Have a nice day!