

Summary: DNA

Basics of DNA structure:

- Double helix hold together by H-bond between G-C, A-T
- B-form helix is the most stable conformation
- Other possible conformations: A-form, Z-form, Triplex, Hairpin, Cruciform (Holliday Junction) or unwound DNA

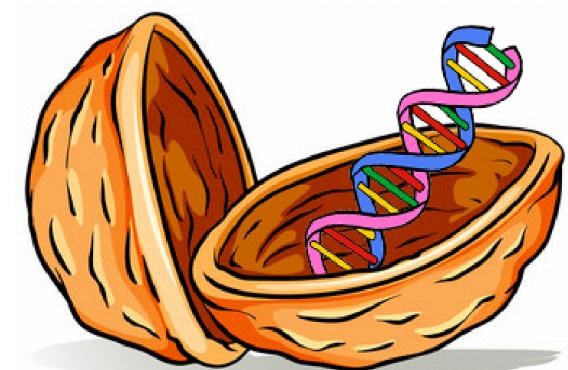
Building with DNA:

- Based on strain hybridization and stably branched DNA
- Tile-based or scaffold based
- Building blocks: tensegrity triangles, modular DNA (rigid organic vertex), DNA tiles, bundle DNA tiles, DNA blocks
- DNA origami

Dynamic DNA devices:

- Strand displacement by ‘toehold’ mechanism (speed related to length + G-C content)
- DNA tweezers
- DNA walkers (non-autonomous and autonomous)
- Computing with DNA (OR and AND gates)

} New articles in Toledo!



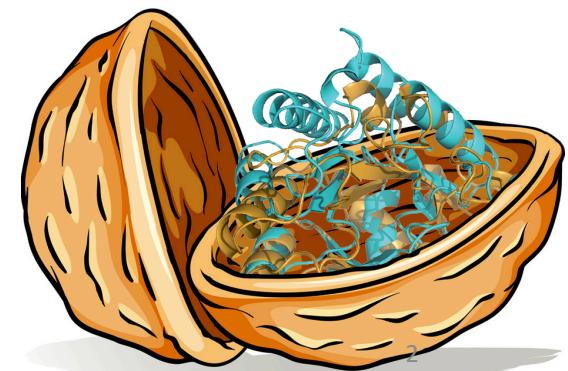
Summary: Protein Structure

Basics of Protein structure:

- Primary structure hold by peptide bonds (peptide plane, angles, Ramachandran plot)
- Secondary structures hold by H-bonds (alpha helix, beta sheet, beta turn)
- Tertiary and quaternary structures hold by non-covalent bonds (H-bonds, electrostatic interactions, van der Waals forces, hydrophobic interactions) AND S-S bonds
- Weak bonds allow dynamic changes
- Protein domains

Protein classification:

- Three main classes: fibrous, globular and membrane proteins
- Fibrous proteins – repeats of amino acids
- Different structures (collagen, alpha keratin, beta keratin)
- Globular protein – hydrophobic core, complex surface with hydrophilic residues
- Membrane protein – peripheral (extrinsic) or integral (intrinsic)
- Peripheral membrane proteins are bound through ionic/H-bond interactions, amphipathic alpha helix, hydrophobic loop, association with integral protein
- Integral proteins are bound through helical transmembrane domains, membrane spanning alpha-sheets or lipid anchors



Learning objectives

Protein production:

- Gene editing
- Cell-based systems
- Case study: insulin

Nanopores

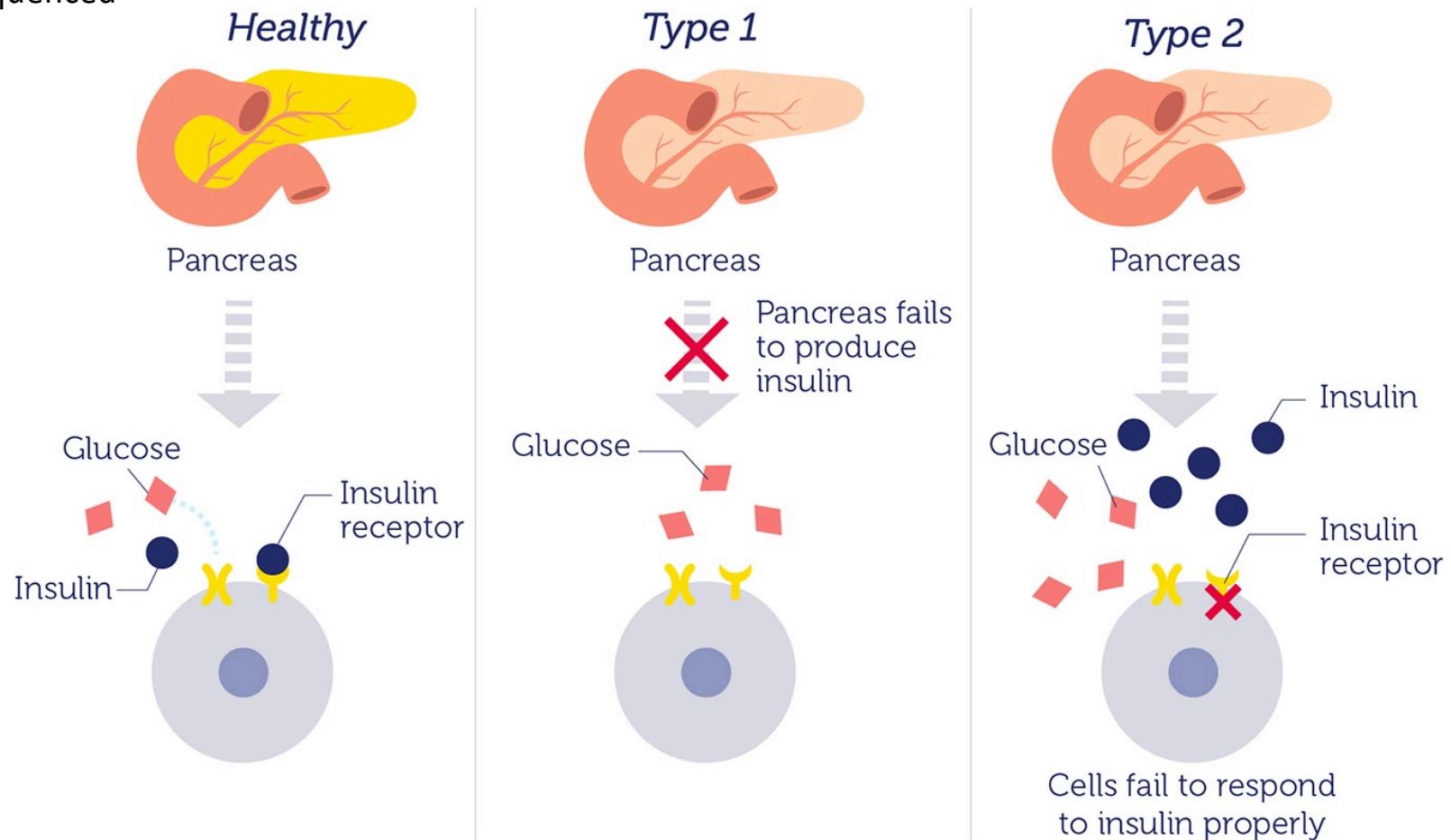
- What are nanopores
- How do nanopores work
- What can nanopores be used for



Diabetes and insulin

Why insulin?

- First protein to be sequenced
- Several Nobel prizes
- First using genetic modified organisms



Diabetes and insulin

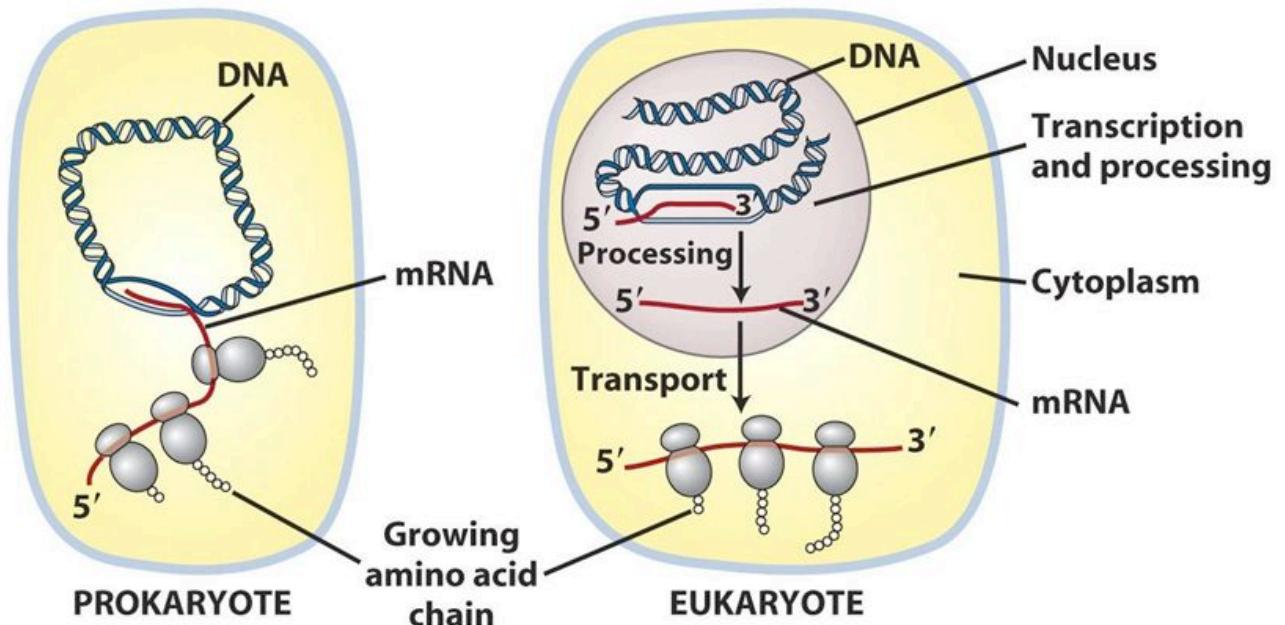
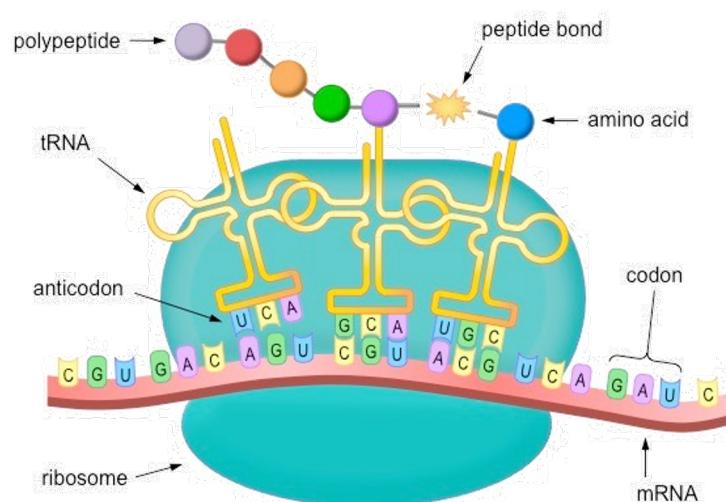
How is insulin produced?

1. Chemically synthesized
2. Extracted from bacteria
3. Extracted from pancreas from dogs
4. Extracted from pancreas from pigs

Production of recombinant proteins

1. Which organism to use?
2. Which plasmid should be chosen?

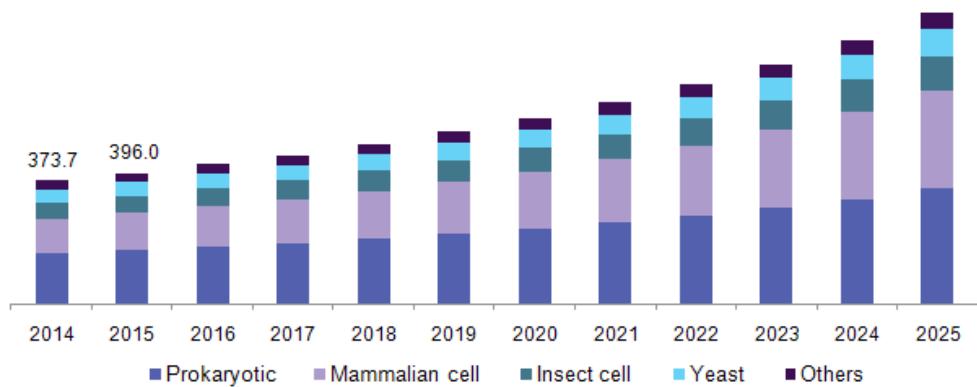
Protein synthesis



Protein folding and post-translational modifications depend on the system used!

Protein production

1. Which organism to use?



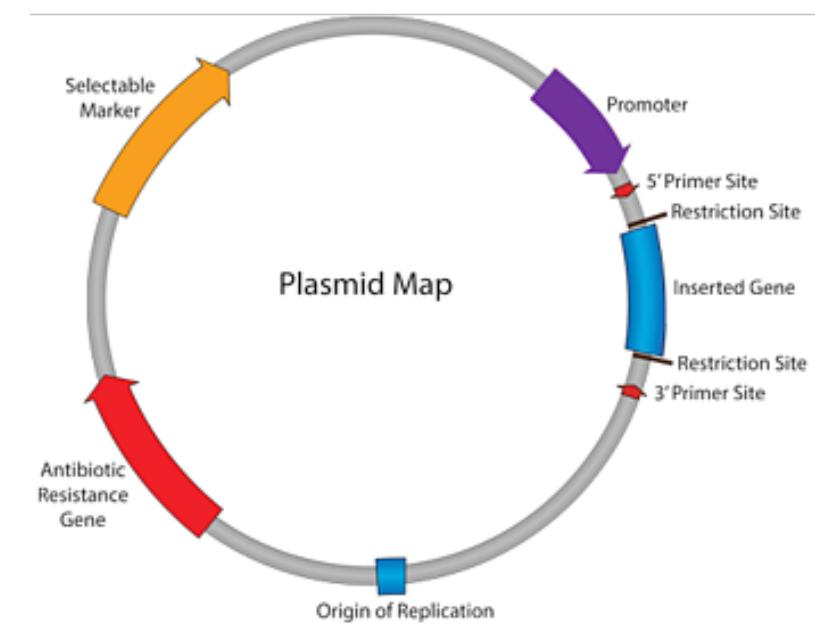
2. Which plasmid should be chosen?

Expression system	Most common application	Advantages	Challenges
Mammalian 	<ul style="list-style-type: none"> Functional assays Structural analysis Antibody production Expression of complex proteins Protein interactions Virus production 	<ul style="list-style-type: none"> Highest-level protein processing Can produce proteins either transiently, or by stable expression Robust optimized transient systems for rapid, ultrahigh-yield protein production 	<ul style="list-style-type: none"> Gram-per-liter yields only possible in suspension cultures More demanding culture conditions
Insect 	<ul style="list-style-type: none"> Functional assays Structural analysis Expression of intracellular proteins Expression of protein complexes Virus production 	<ul style="list-style-type: none"> Similar to mammalian protein processing Can be used in static or suspension culture 	<ul style="list-style-type: none"> More demanding culture conditions than prokaryotic systems Production of recombinant baculovirus vectors is time consuming
Yeast 	<ul style="list-style-type: none"> Structural analysis Antibody generation Functional analysis Protein interactions 	<ul style="list-style-type: none"> Eukaryotic protein processing Scalable up to fermentation (grams per liter) Simple media requirements 	<ul style="list-style-type: none"> Fermentation required for very high yields Growth conditions may require optimization
Bacterial 	<ul style="list-style-type: none"> Structural analysis Antibody generation Functional assays Protein interactions 	<ul style="list-style-type: none"> Scalable Low cost Simple culture conditions 	<ul style="list-style-type: none"> Protein solubility May require protein-specific optimization May be difficult to express some mammalian proteins
Algal 	<ul style="list-style-type: none"> Studying photosynthesis, plant biology, lipid metabolism Genetic engineering Biofuel production 	<ul style="list-style-type: none"> Genetic modification and expression systems for photosynthetic microalgae Superb experimental control for biofuel, nutraceuticals, and specialty chemical production Optimized system for robust selection and expression 	<ul style="list-style-type: none"> Nascent technology Less developed compared to other host platforms
Cell-free 	<ul style="list-style-type: none"> Toxic proteins Incorporation of unnatural label or amino acids Functional assays Protein interactions Translational inhibitor screening 	<ul style="list-style-type: none"> Open system; able to add unnatural components Fast expression Simple format 	<ul style="list-style-type: none"> Scaling above multigram quantities may not be costly

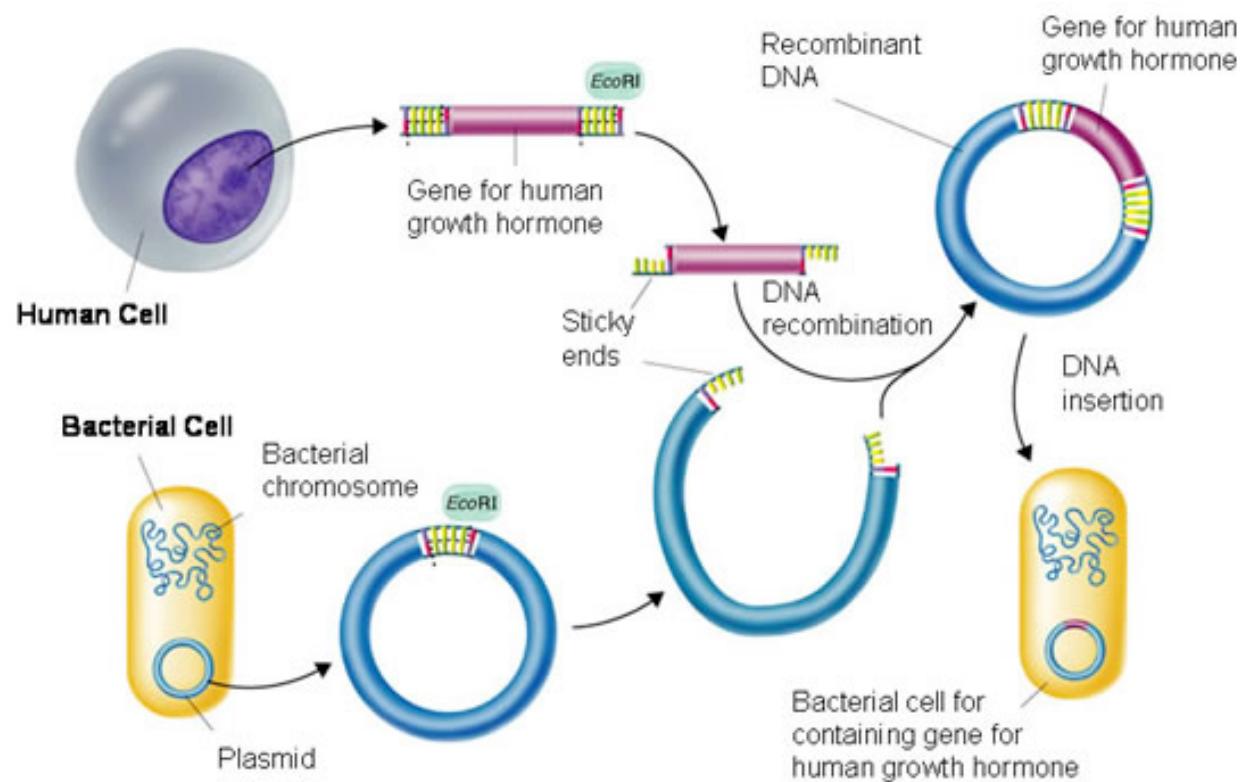
What is a plasmid?

Small circular piece of DNA found in bacterial cells

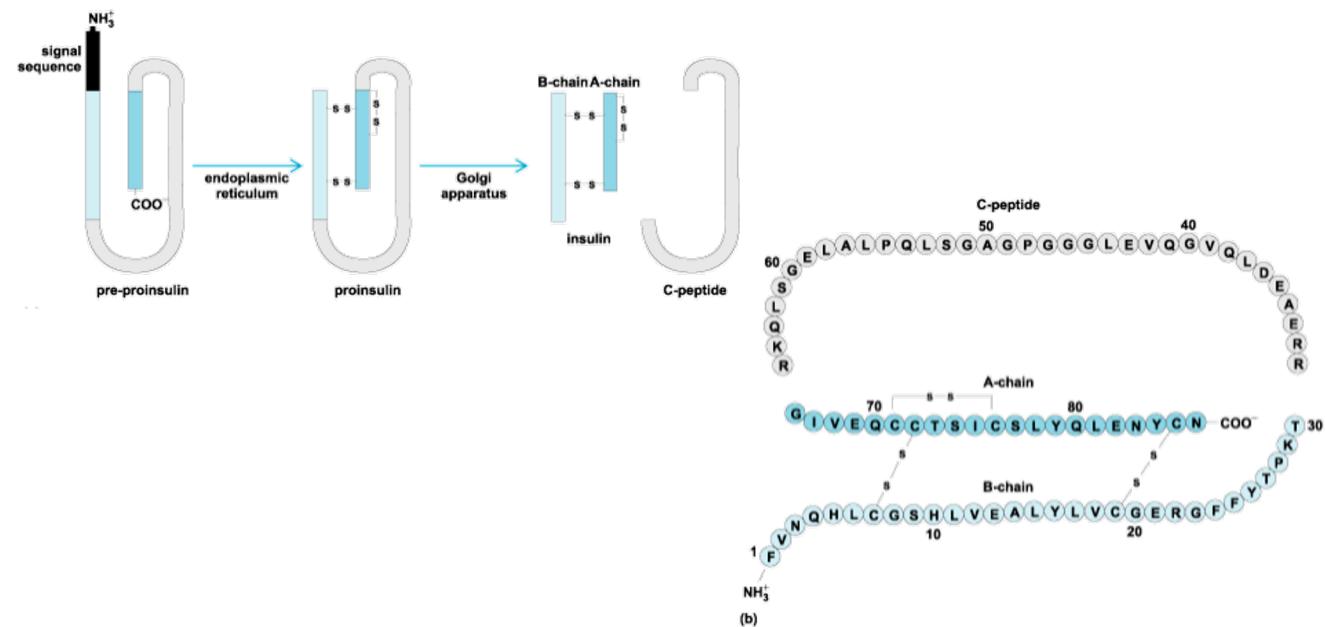
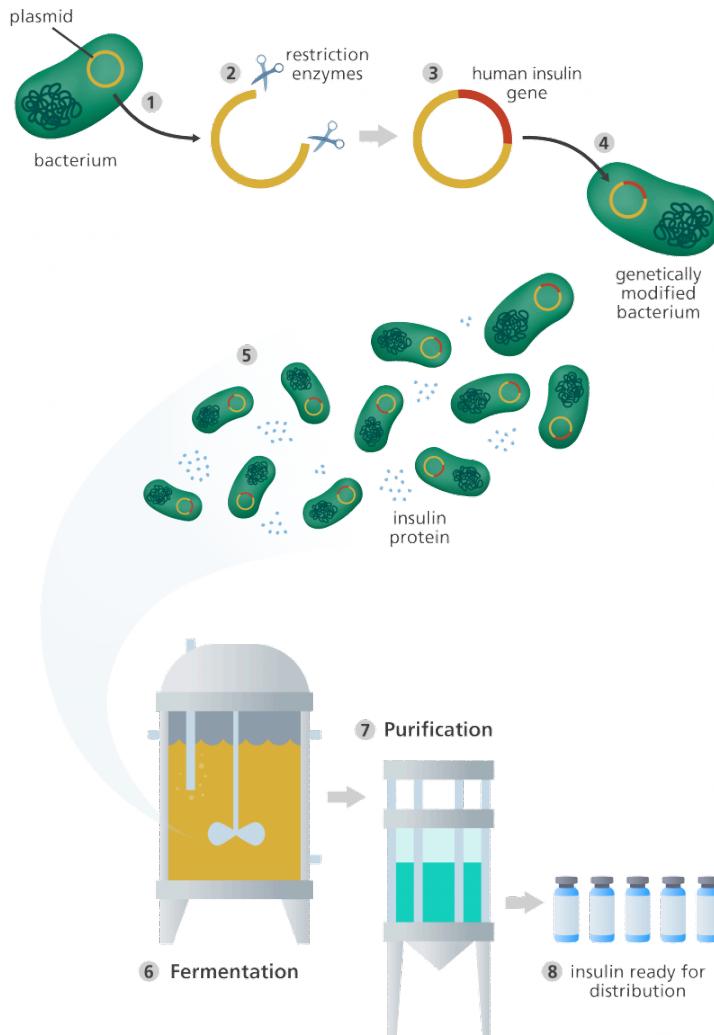
Vector Element	Description
<u>Origin of Replication (ORI)</u>	DNA sequence which allows initiation of replication within a plasmid by recruiting transcriptional machinery proteins
<u>Antibiotic Resistance Gene</u>	Allows for selection of plasmid-containing bacteria.
Multiple Cloning Site (MCS)	Short segment of DNA which contains several restriction sites allowing for the easy insertion of DNA. In expression plasmids, the MCS is often downstream from a promoter.
Insert	Gene, promoter or other DNA fragment cloned into the MCS for further study.
<u>Promoter Region</u>	Drives transcription of the target gene. Vital component for expression vectors: determines which cell types the gene is expressed in and amount of recombinant protein obtained.
Selectable Marker	The antibiotic resistance gene allows for selection in bacteria. However, many plasmids also have selectable markers for use in other cell types.
Primer Binding Site	A short single-stranded DNA sequence used as an initiation point for PCR amplification or sequencing. Primers can be exploited for sequence verification of plasmids.



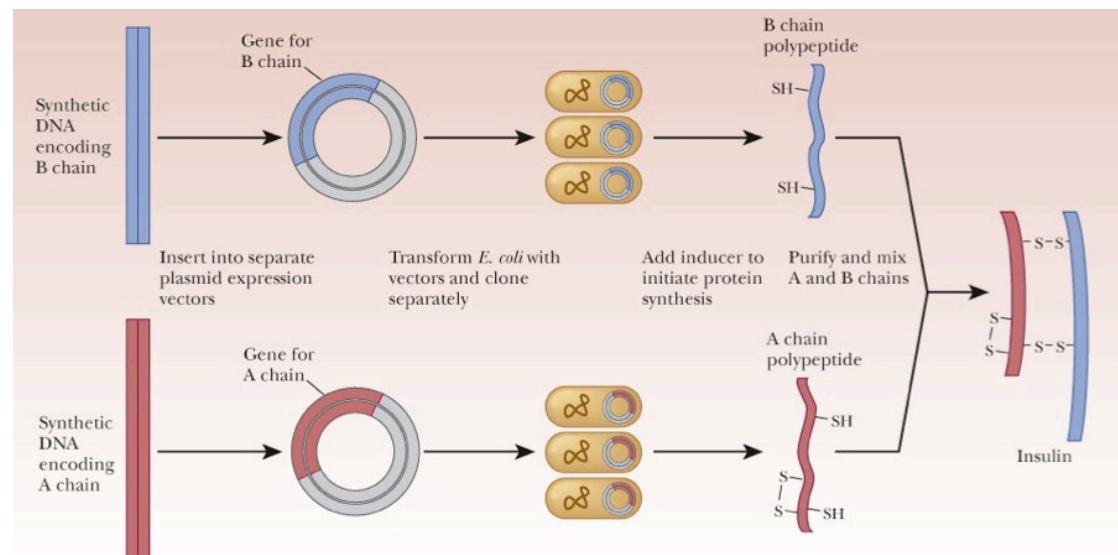
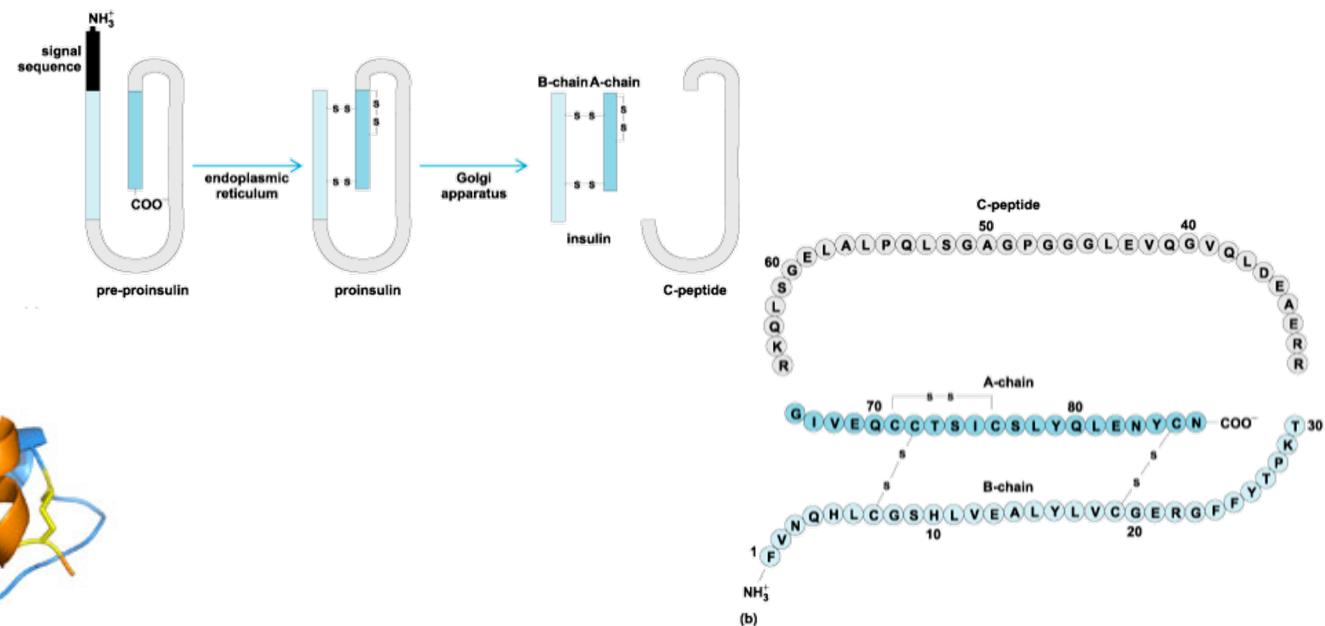
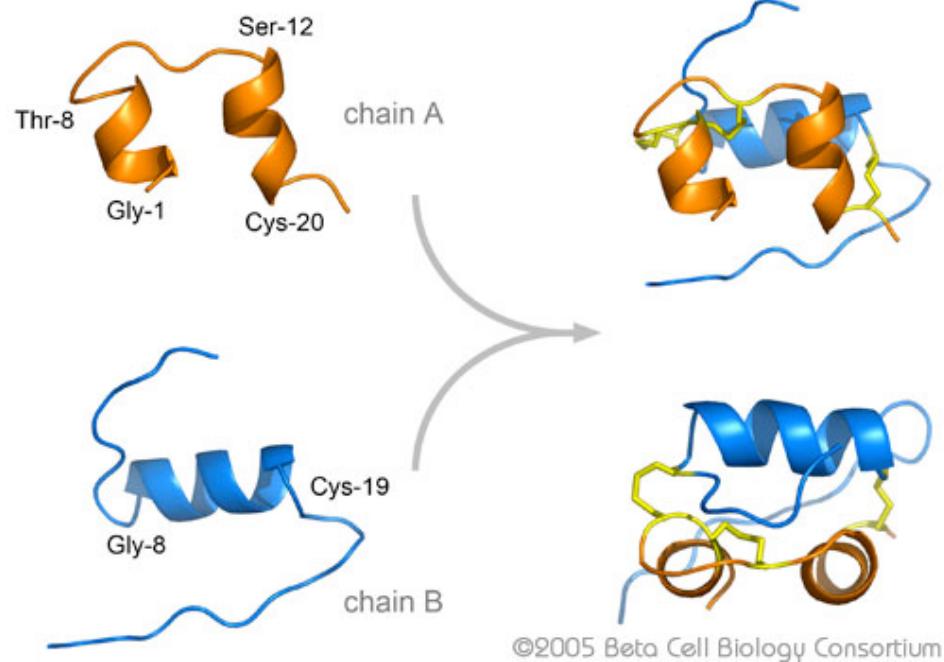
Recombinant proteins



Case study: insulin



Case study: insulin



Steps for protein production

1. Get protein or DNA sequence
2. Select expression system (prokaryotic,eukaryotic...)
3. Insert DNA sequence coding for the protein in plasmid DNA
4. Deliver plasmid DNA into expression system
5. Let the cells multiply – and synthesize the protein
6. Extract and purify your protein

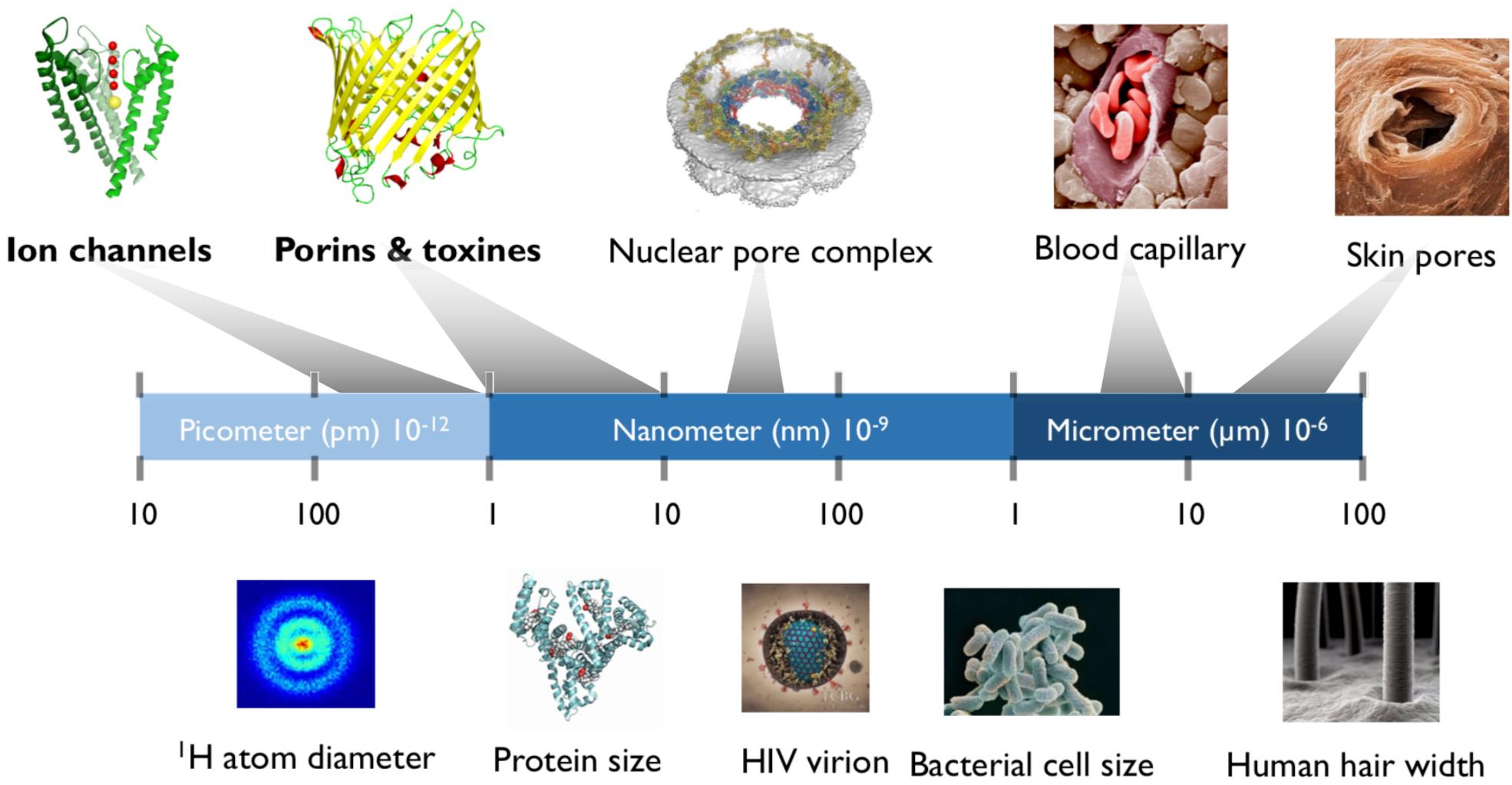
What are nanopores?

from Greek πόρος (*poros*)
meaning 'passage'

nano

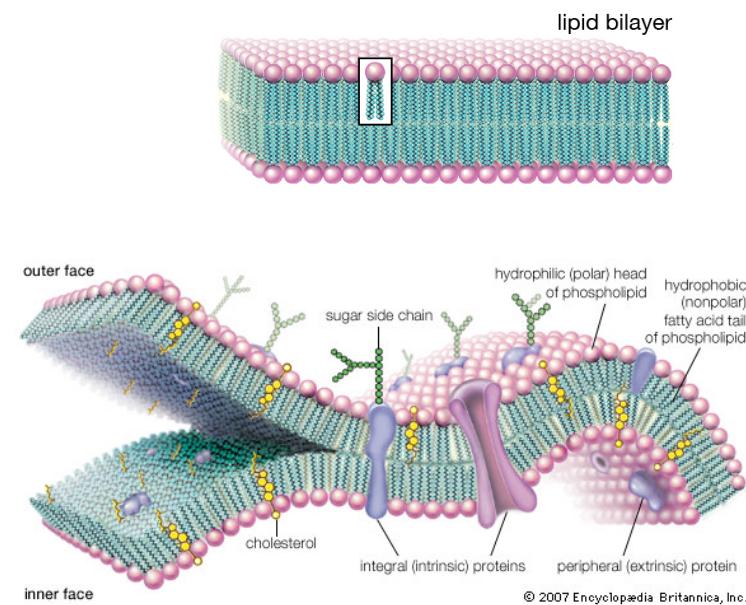
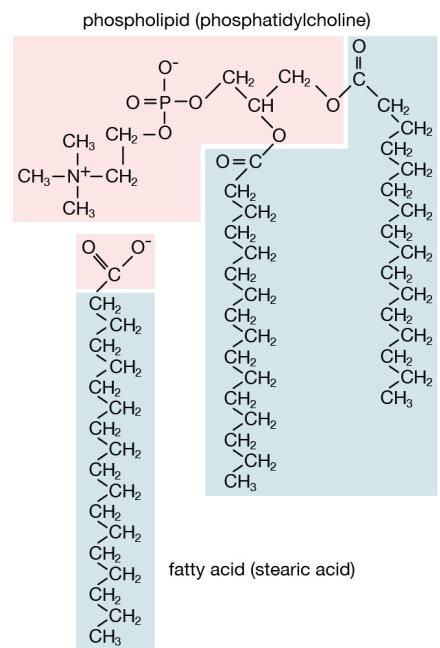
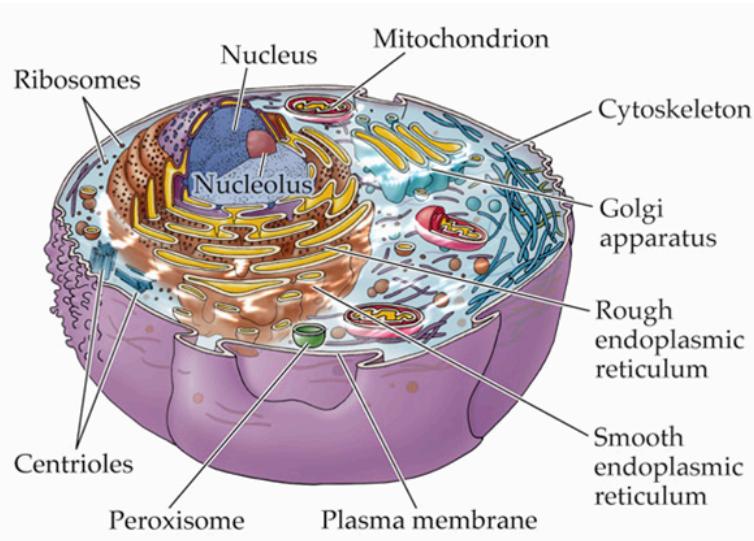
from Greek νάνος (*nanos*)
meaning 'dwarf'

Pores exist at different length scales



In the cell, the lipid bilayer is the ‘border wall’...

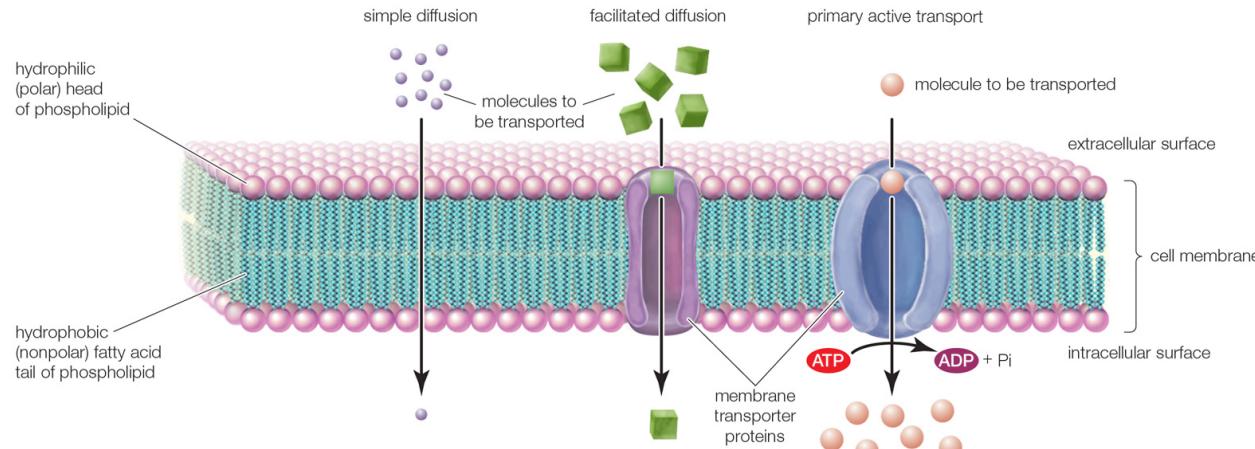
- Membrane separates the inside from the outside of the cell
 - Consists of a lipid bilayer with hydrophobic core
 - Impermeable for charged or large molecules
 - Used for protection, energy storage or signal transduction



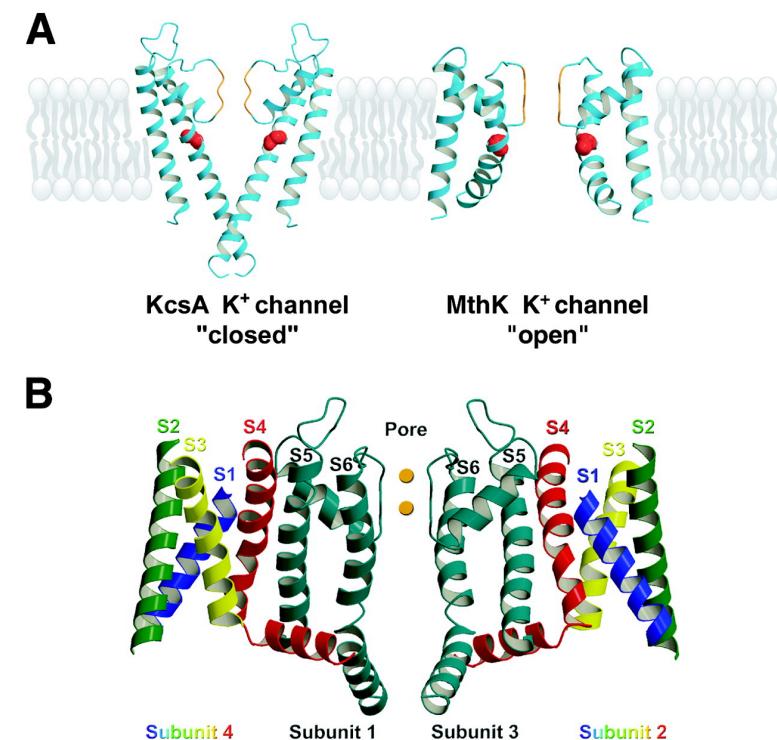
... and nanopores are its ‘immigration officers’

- Transport across membrane is regulated by protein pores
- Channels for ions, water, nutrients and waste products
- Can also be toxins: secreted by bacteria to perforate and kill host cells

Different types of membrane transport

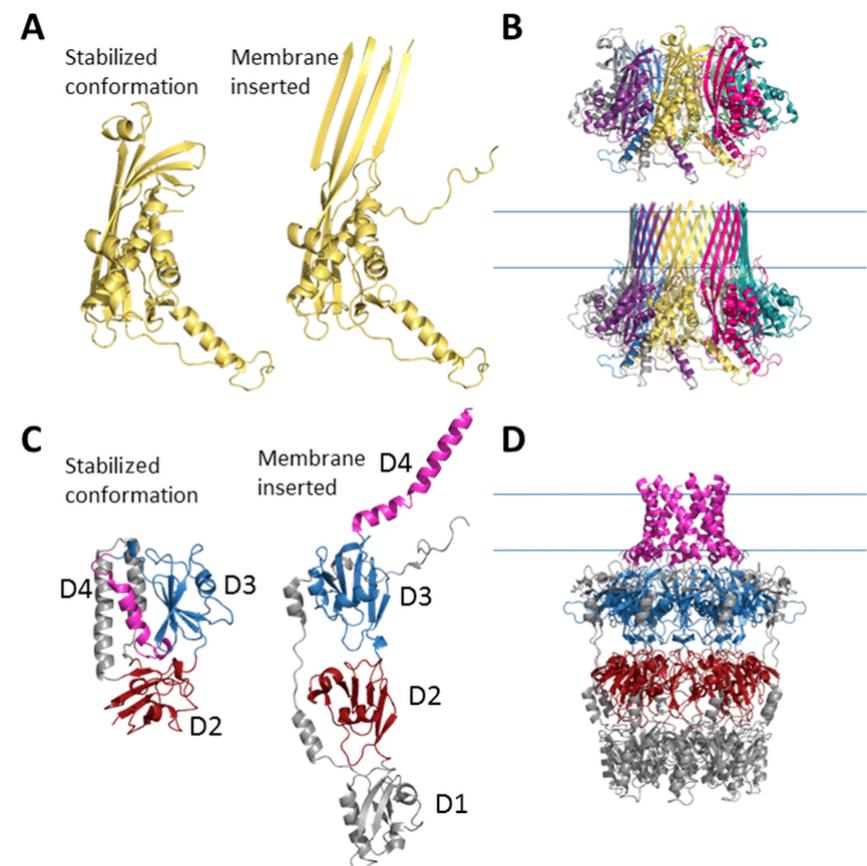
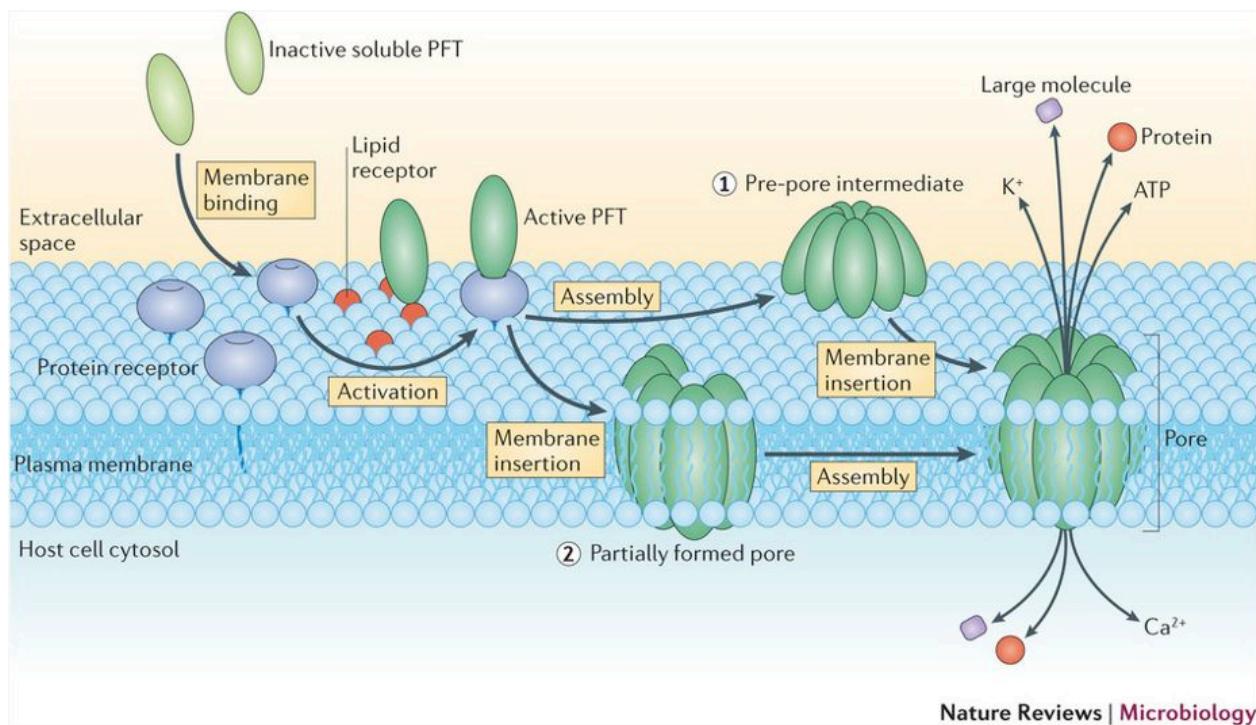


© 2010 Encyclopædia Britannica, Inc.

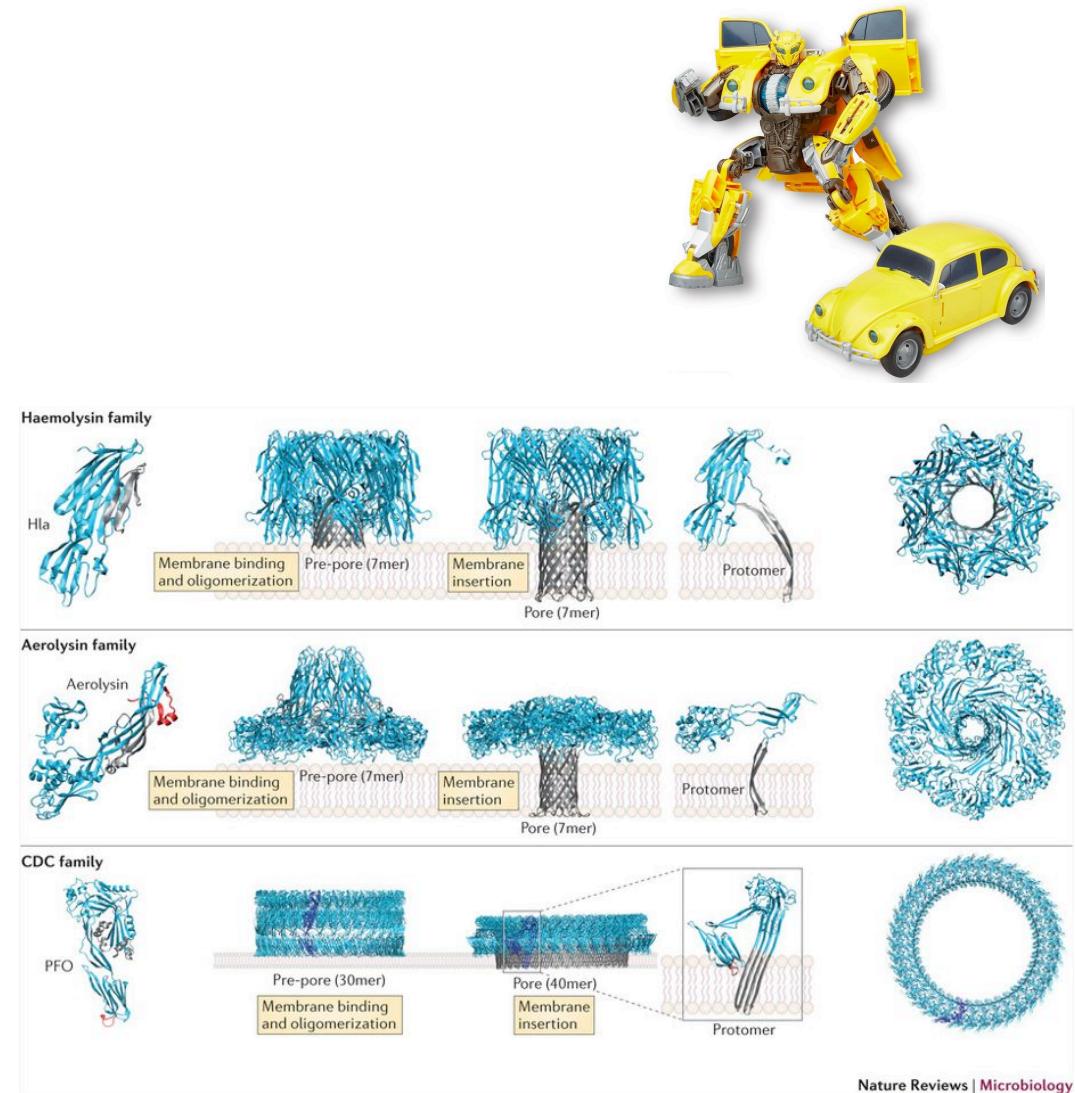
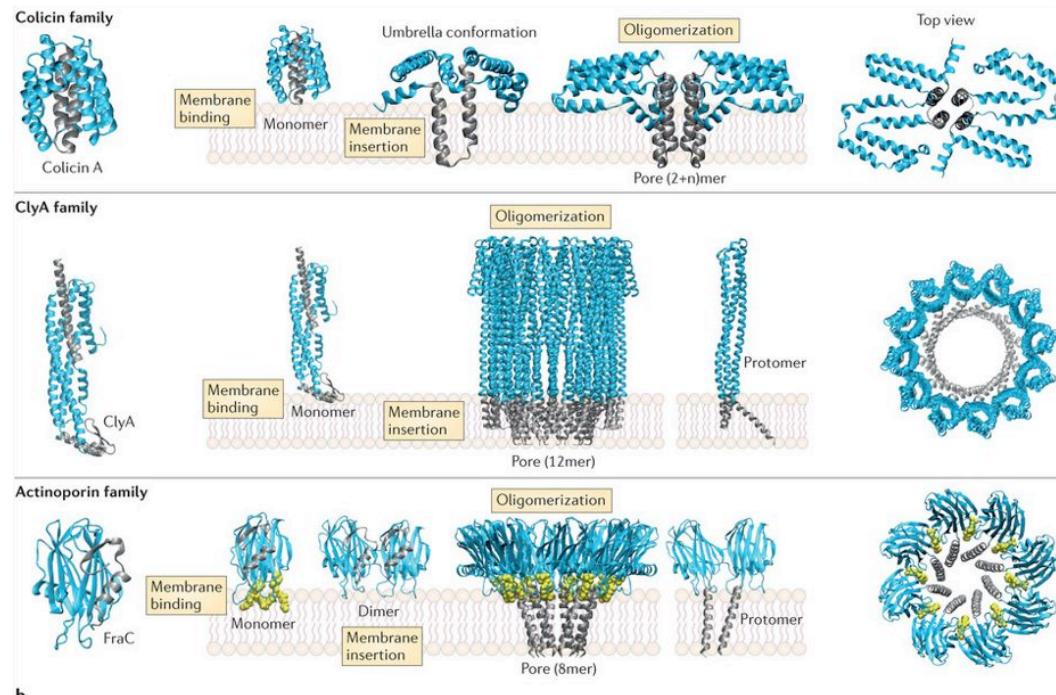


Most nanopores in nanotechnology are bacterial cytotoxins

- Soluble PFTs are recruited to the host membrane by protein receptors and/or specific interactions with lipids
- Upon membrane binding, the toxins concentrate and start the oligomerization process
- Protein ring stabilizes and punctures the cell membrane

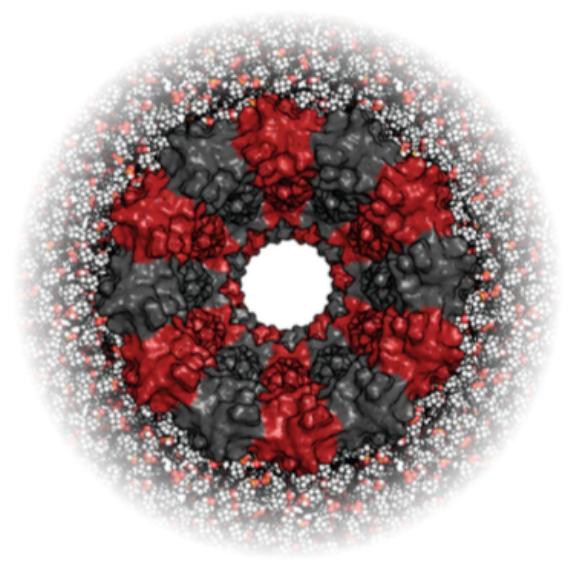
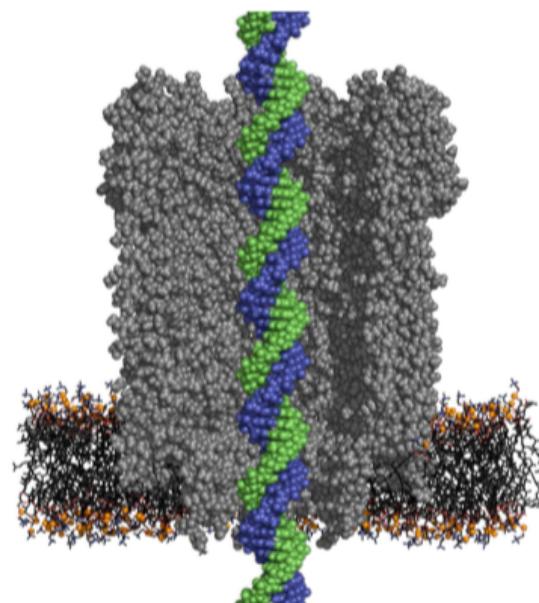


Structural architectures and pore formation mechanisms of pore-forming toxin families



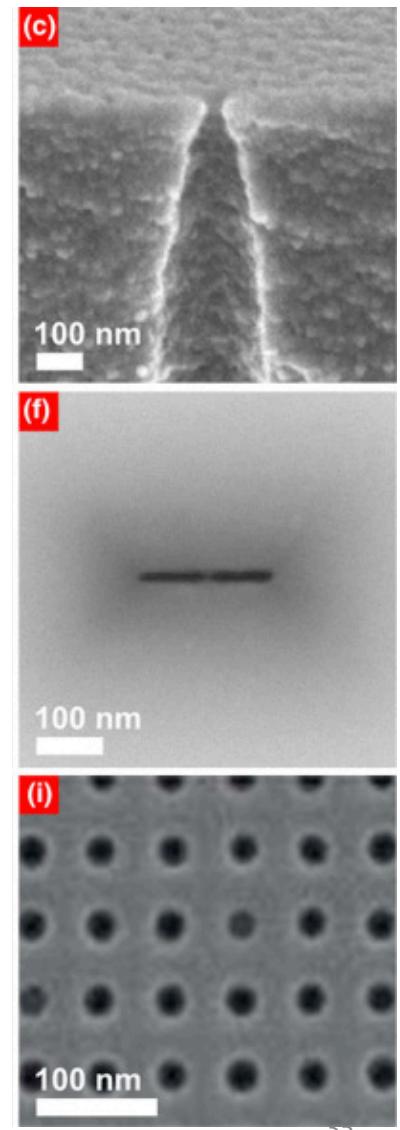
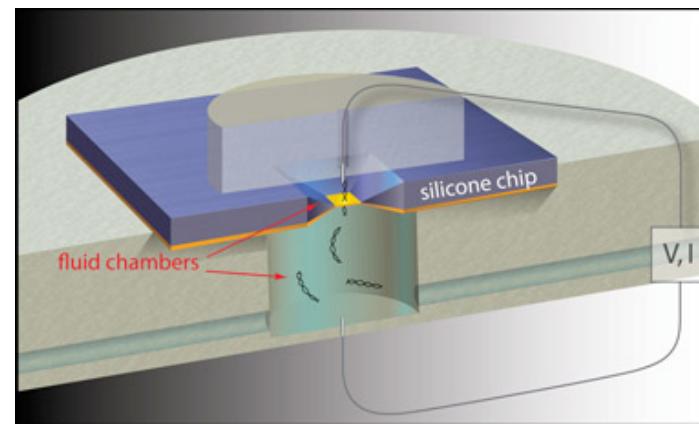
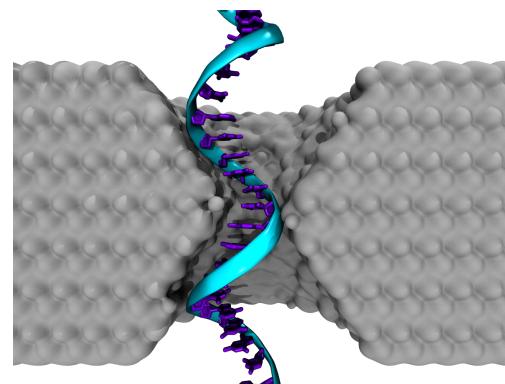
Nanopores come in “biological” flavors ...

- Bottom up fabrication (self-assembly), protein or DNA-based
- Advantages
 - ✓ Atomic level precision
 - ✓ Highly reproducible dimensions
 - ✓ High bio-compatibility
 - ✓ Cheap (protein- or DNA-based)
- Disadvantages
 - ✗ Fragile due to lipid bilayer
 - ✗ Difficult to parallelize (stochastic)

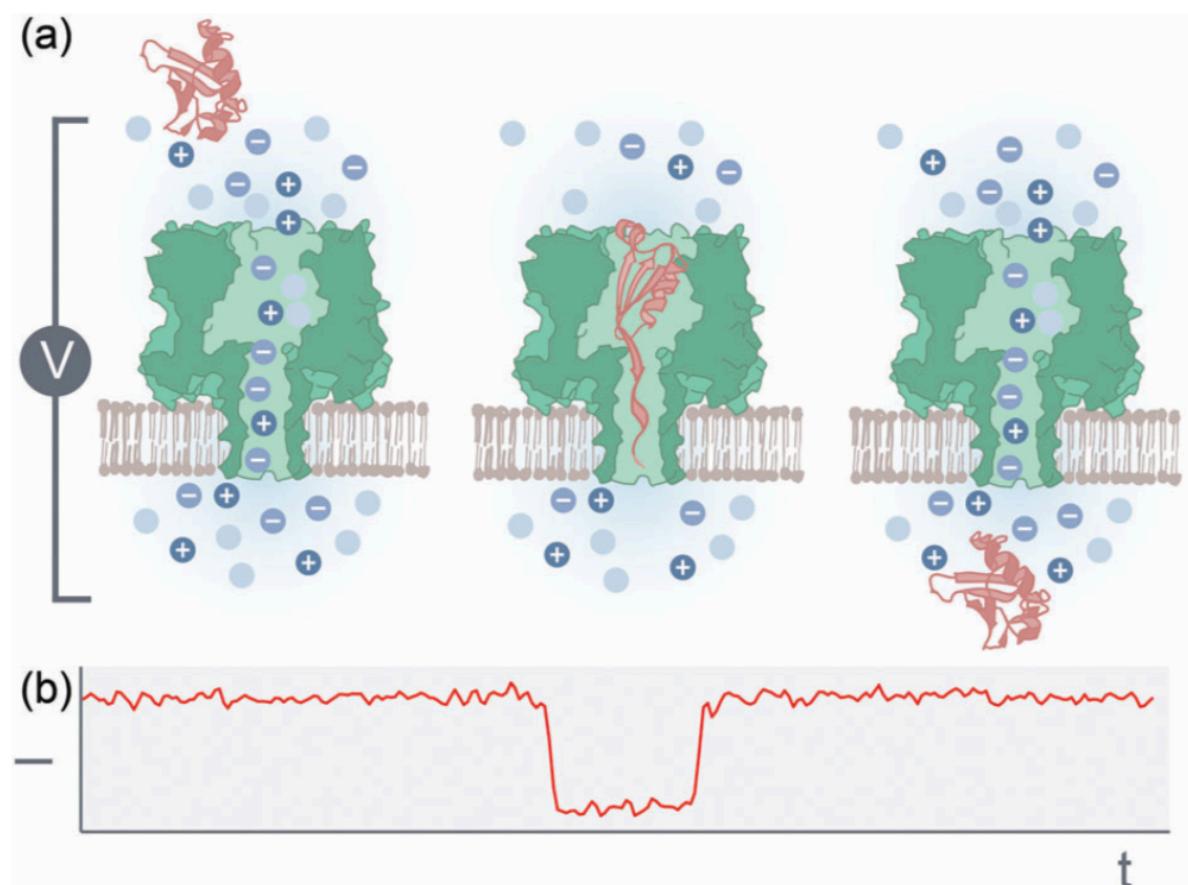
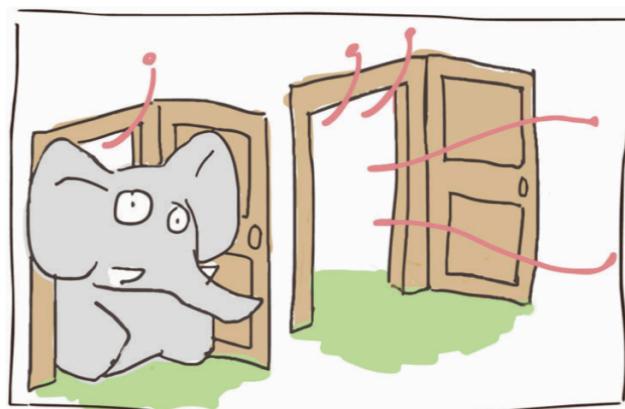
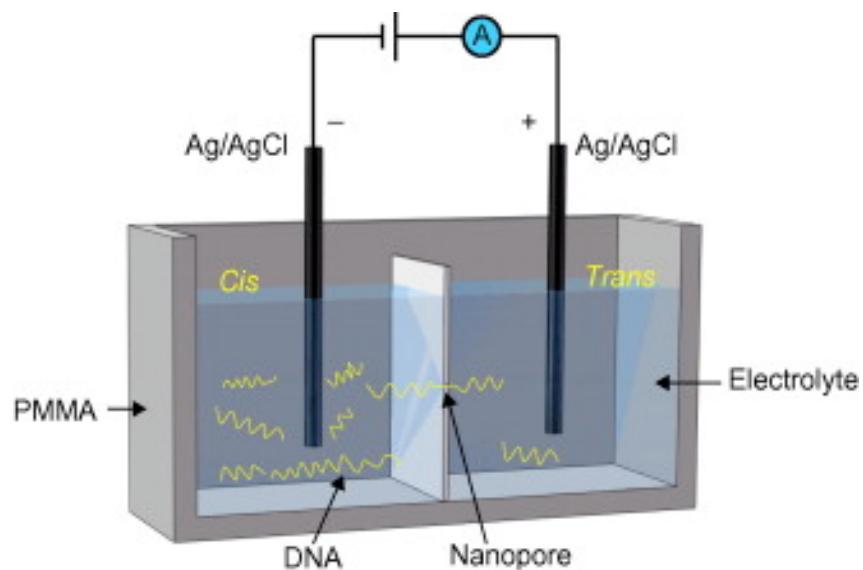


... but also exist in “solid-state” materials

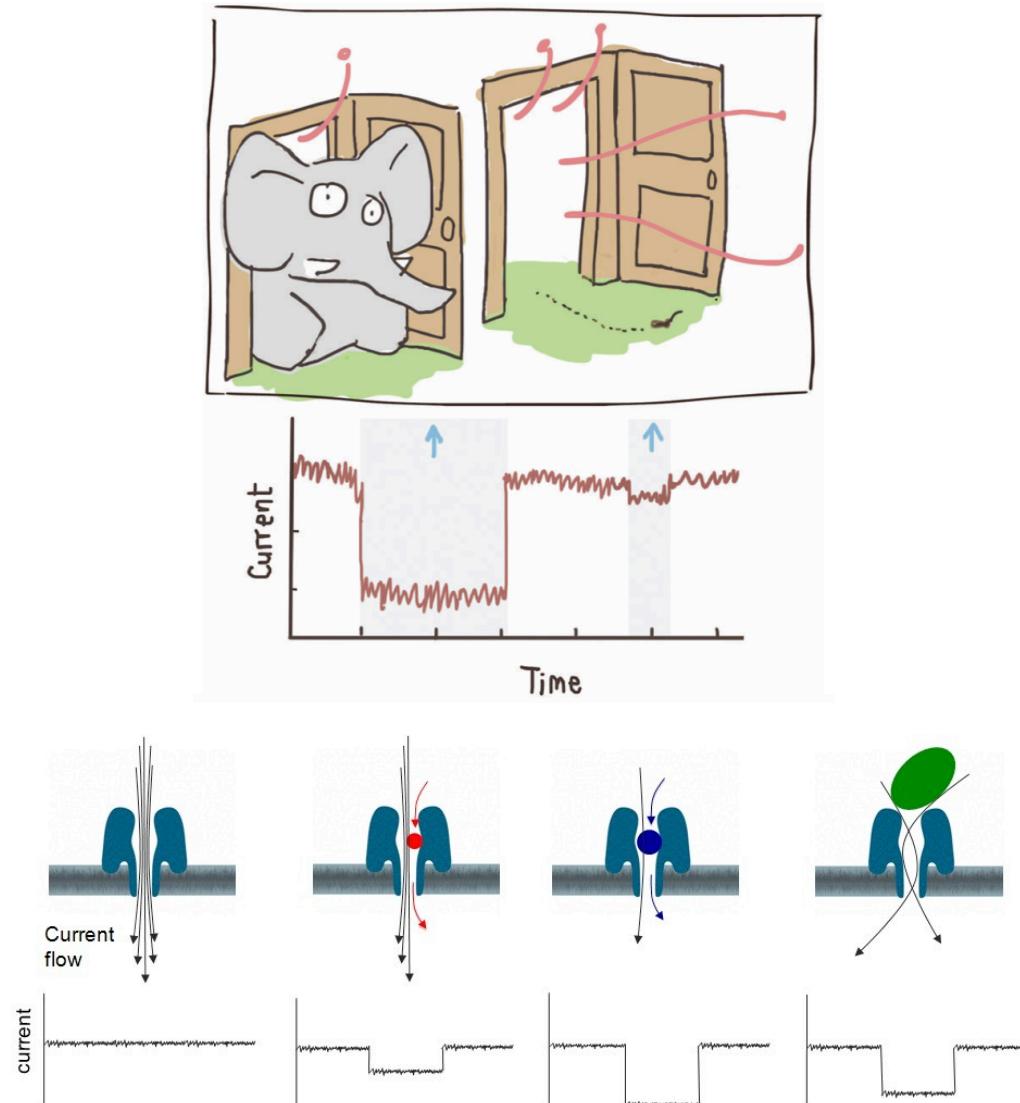
- Top down fabrication (micromachining), silicon-based
- Advantages
 - ✓ Robust due to “solid” materials
 - ✓ Adaptable shape, size and length
 - ✓ Straightforward to create arrays
- Disadvantages
 - ✗ Poor size reproducibility
 - ✗ Atomic level modifications are difficult
 - ✗ Expensive to fabricate



How do nanopores work?



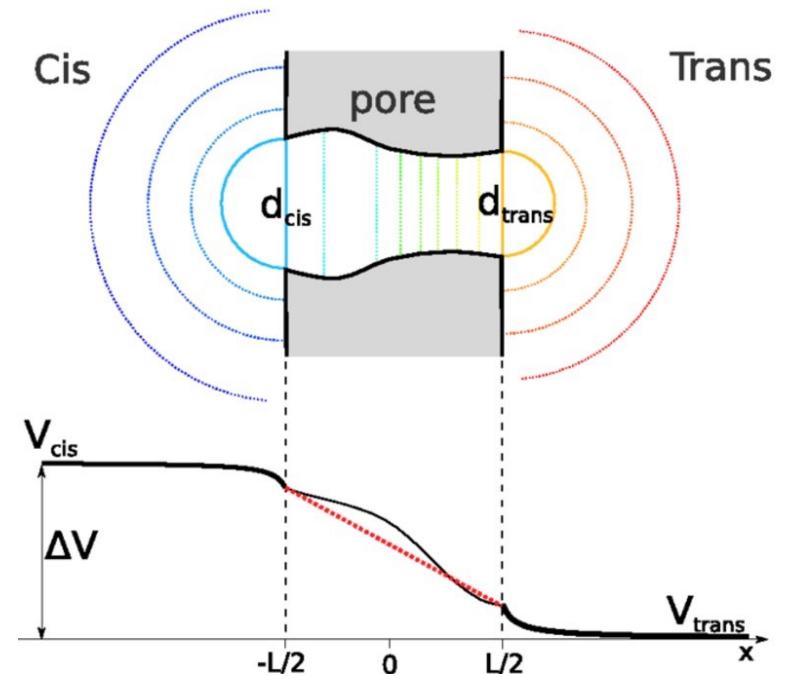
Size matters!



Biological Nanopore	Structure Side View	Structure Top View	Critical Dimension
α -HL	A complex, multi-subunit structure with a central channel.	A circular structure with a central hole.	1.4 nm
OmpG	A more compact, elongated structure.	A circular structure with a central hole.	1.3 nm
MspA	A large, branched structure.	A circular structure with a central hole.	1.2 nm
AeL	A large, branched structure.	A circular structure with a central hole.	1.0 nm
Phi29 Motor	A large, complex structure.	A circular structure with a central hole.	3.6 nm
ClyA	A large, branched structure.	A circular structure with a central hole.	3.3 nm

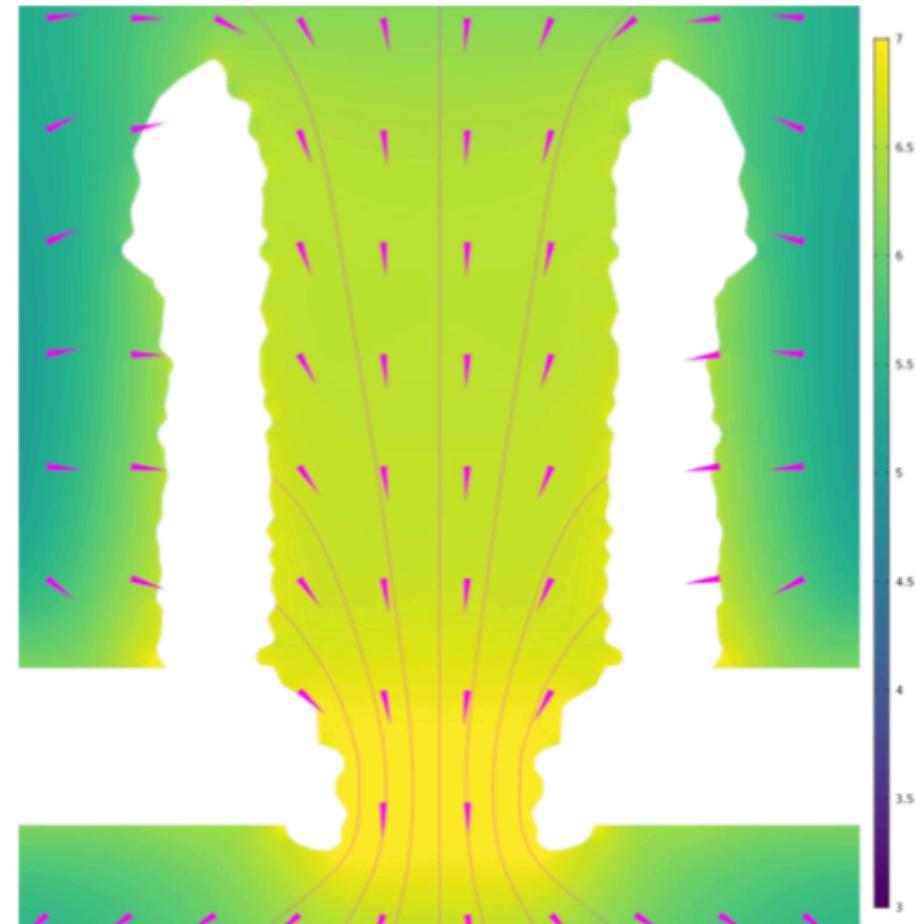
The application of a bias voltage results in a strong electrical field in the nanopore

- Nanopores have very high electrical ('ionic') resistances ($\cong 0.1$ to $1\text{ G}\Omega$)
- Most of the voltage change (ΔV) occurs near and within the nanopore, resulting in very high electrical fields (10^6 to 10^8 V/m)
- Example: $\Delta V = 100\text{ mV}$ over a nanopore with length $l = 10\text{ nm}$:
 $E_z = \Delta V/l = 10\text{ mV/nm} = 10^7\text{ V/m}$



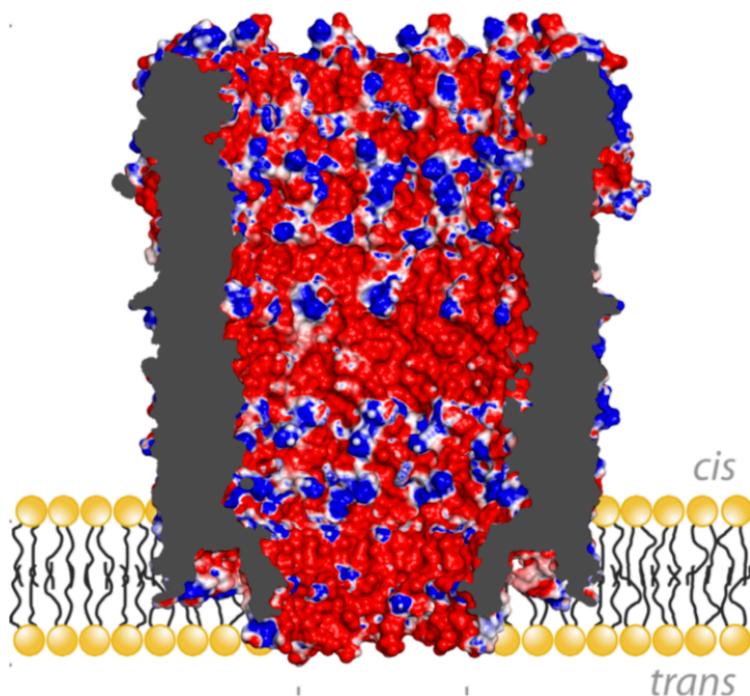
The electrical field drives ions and charged molecules through the pore

- Ions in solution migrate along electrical field lines
 - current depends on mobility of ions (size and charge)
 - a highly charged pore is more selective for its counter-ion
- For highly charged molecules (e.g. DNA), electromigration typically dominates
 - effective charge (~ionic strength) determines mobility

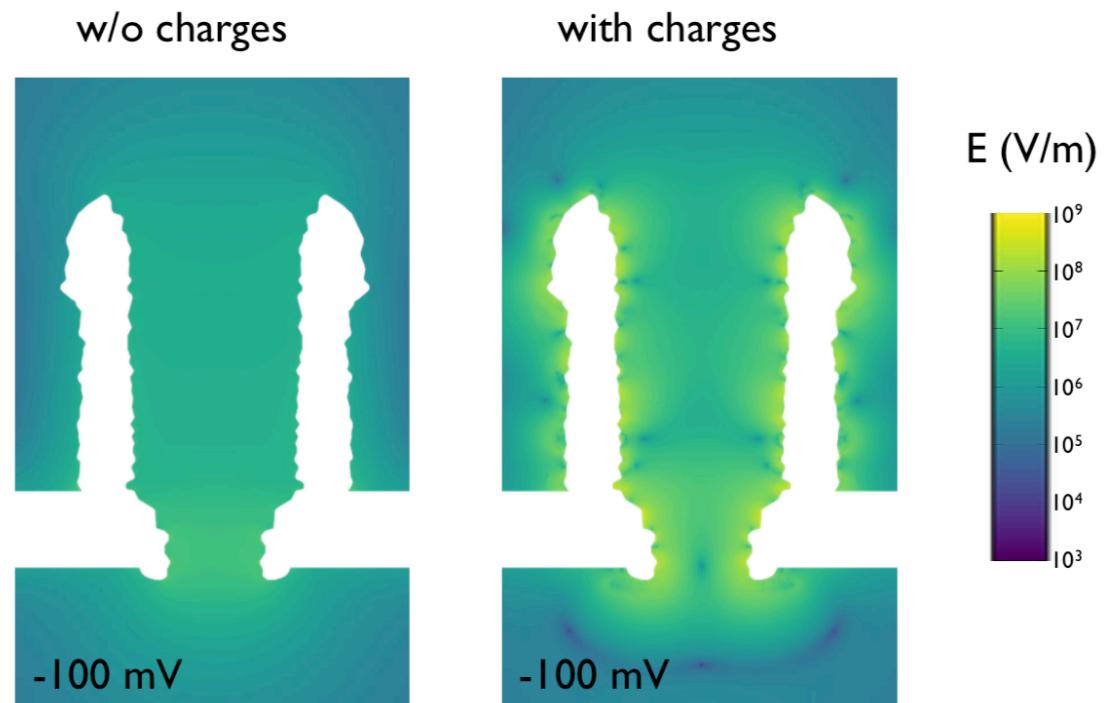


Electrical field is heavily influenced by fixed charges inside the pore walls.

The cytolysin A (ClyA) protein nanopore contains a highly negatively charged interior

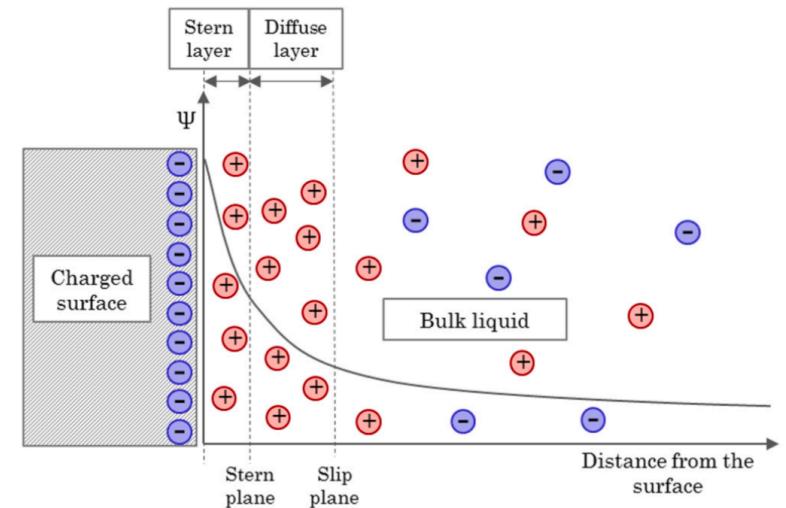
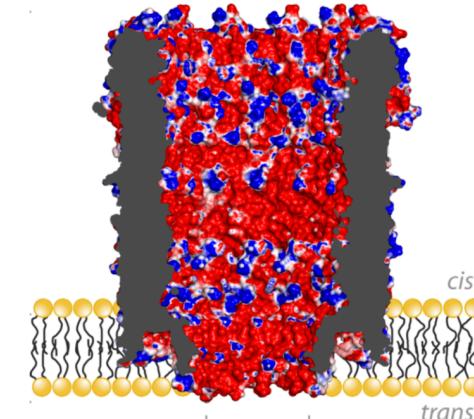


Simulation of electrical field inside ClyA



Many nanopores have fixed charges in their walls, resulting in an electrical double layer (EDL)

- **Bulk electrolyte**
 - ions are mobile charges
 - locally electroneutral ($+ == -$)
- **Electrolyte near charged surface**
 - fixed charges attract layer of oppositely charged counter-ions (= electrical double layer, EDL)
 - EDL hides the presence of the surface charges from the bulk solution ('screening')
- **Electrical double layer**
 - thickness depends on the ionic strength (cfr. Debye length)

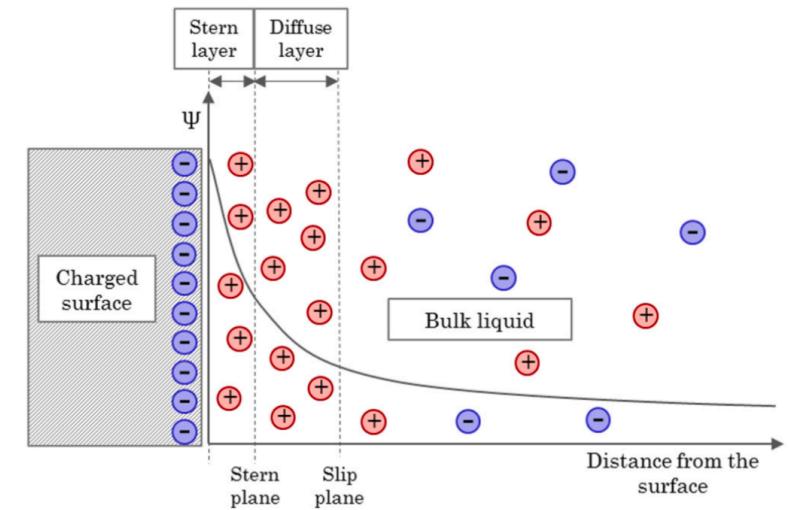
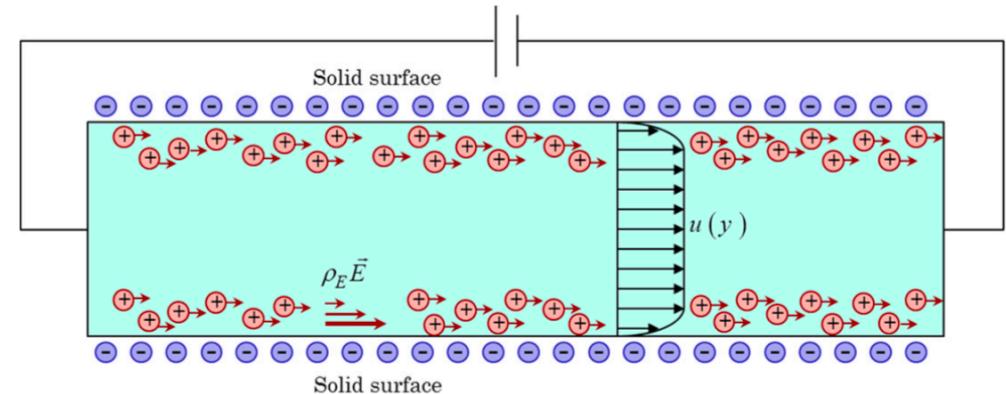
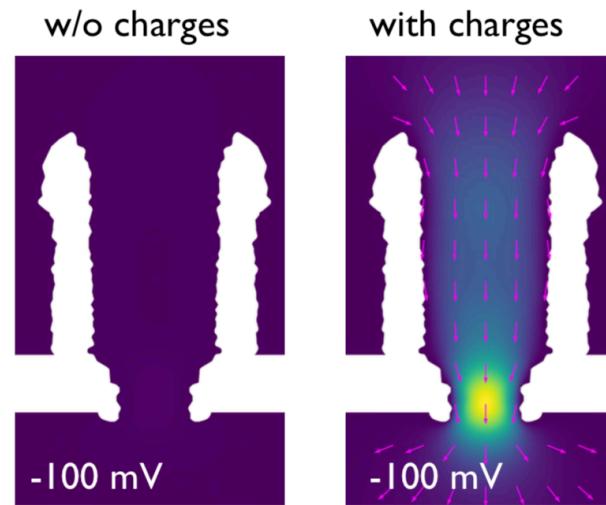


The drag on the fluid by the movement of the EDL results in a net electro-osmotic flow

- Electrical field exerts force on ions
- Ions exert force on their hydration shell
- Hydration shell exerts force on other water molecules

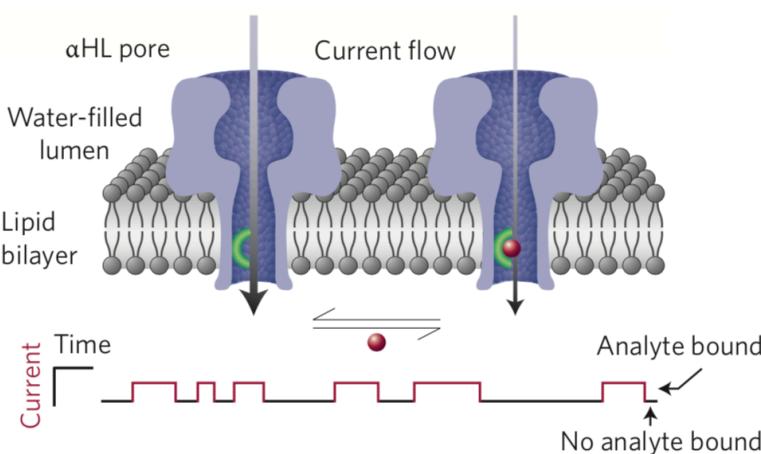
Example: simulation water velocity inside ClyA

- Typical velocity: 100 mm/s
- Flux: $10^{-18} \text{ m}^3/\text{s}$
or $3.3 \times 10^9 \text{ water mol/s}$

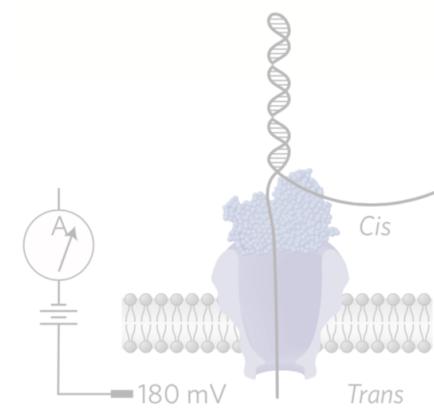


Applications of protein pores

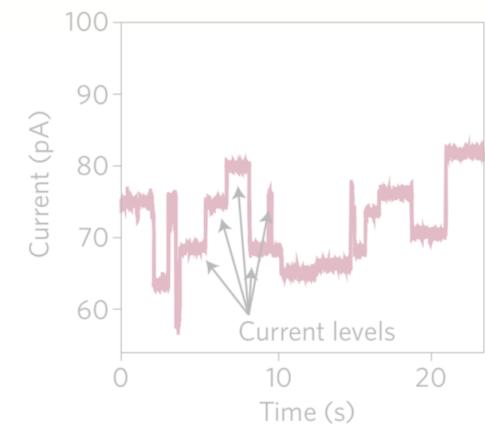
- Detecting a wide range of substances including organic species, ions and proteins
- Studying biomolecular folding and unfolding
- Investigating covalent and non-covalent interactions
- Probing enzyme kinetics
- Sequencing DNA molecules



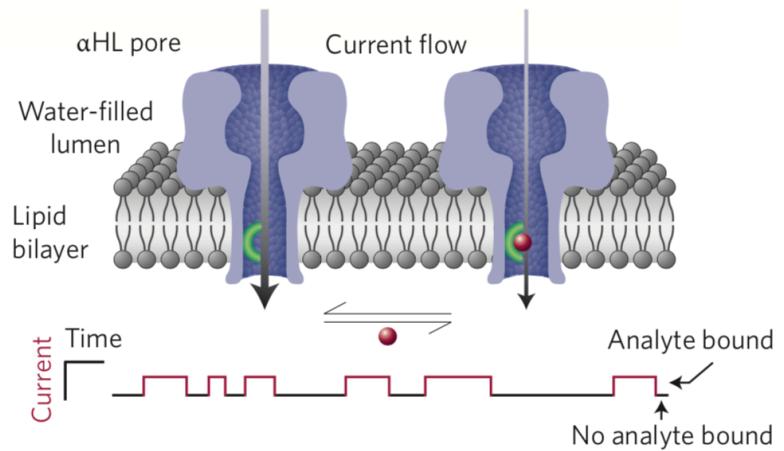
Molecular recognition/binding



Polymer sequencing



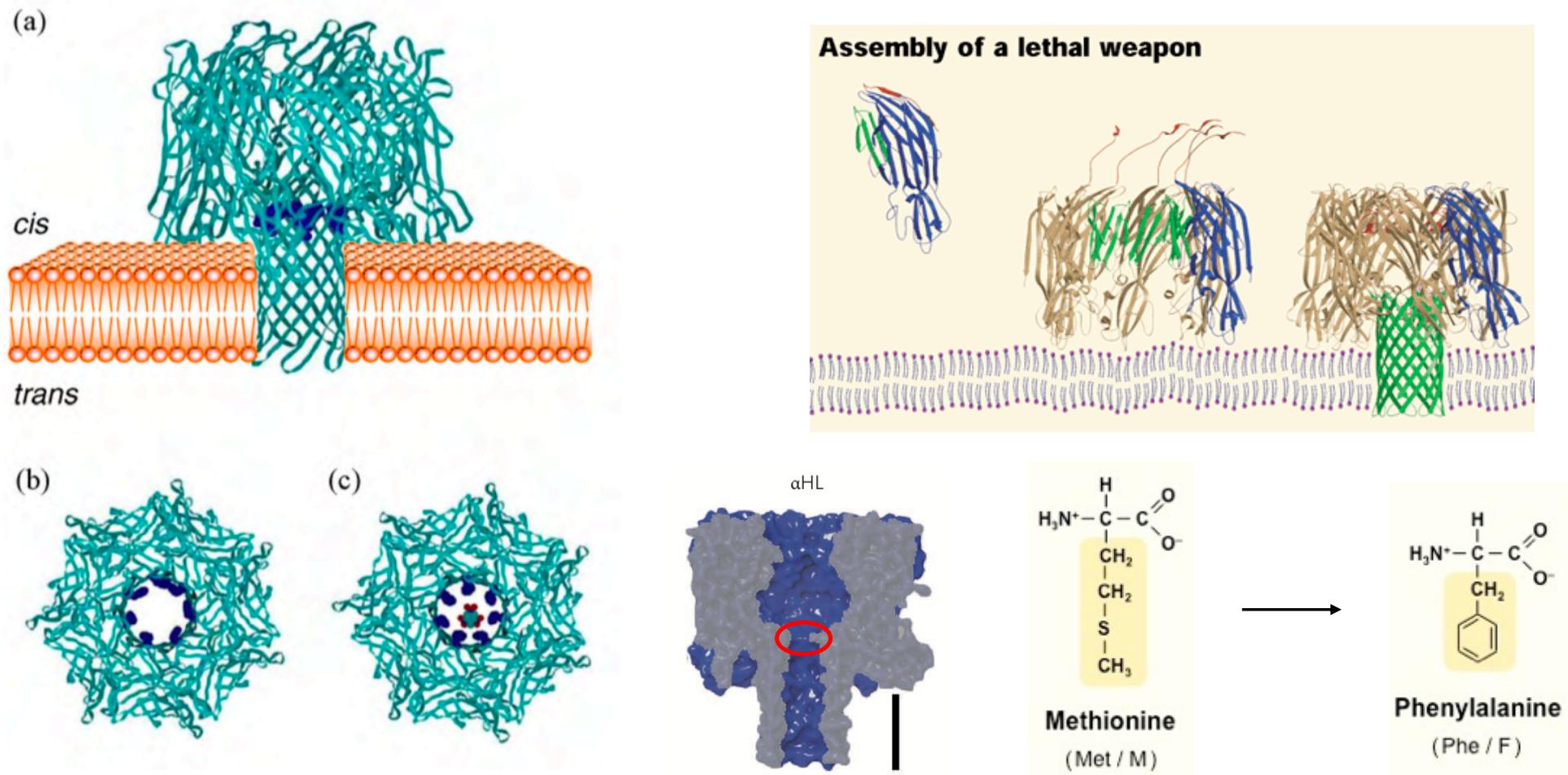
Exploring molecular interactions



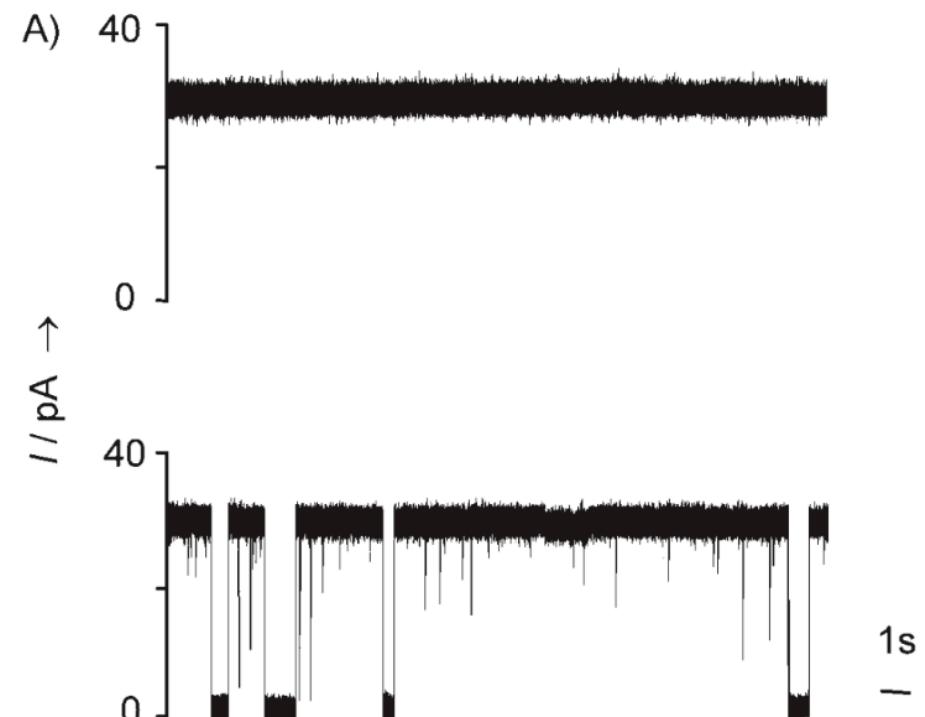
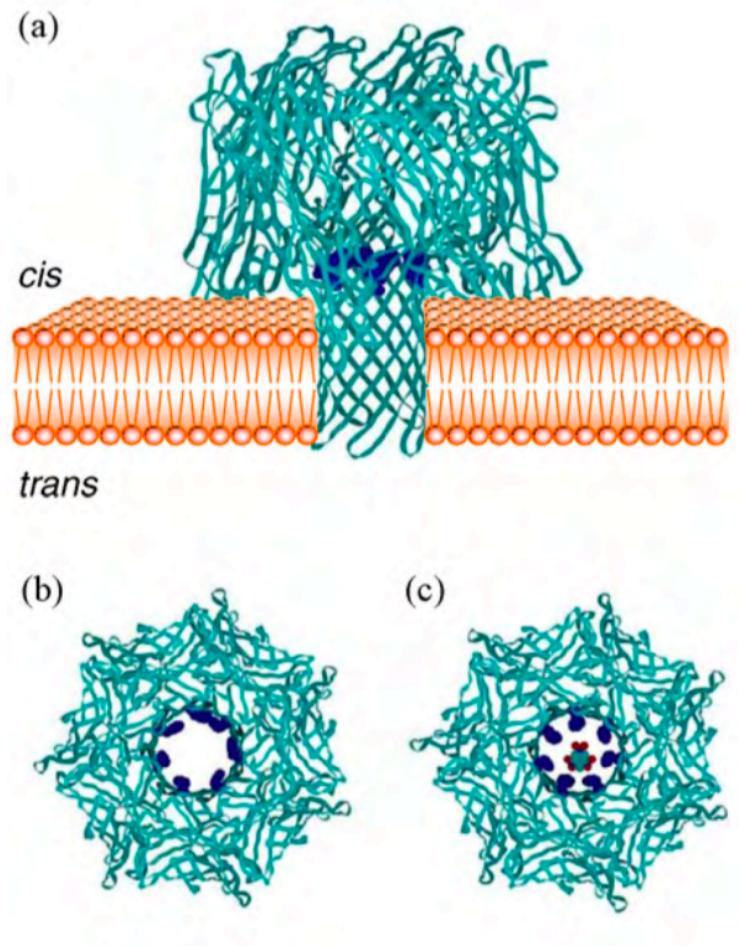
- Pulse amplitude → volume/charge of the translocated molecule
- Frequency → concentration in solution
- Translocation time (too fast → not detected)

Chemical or genetic modification

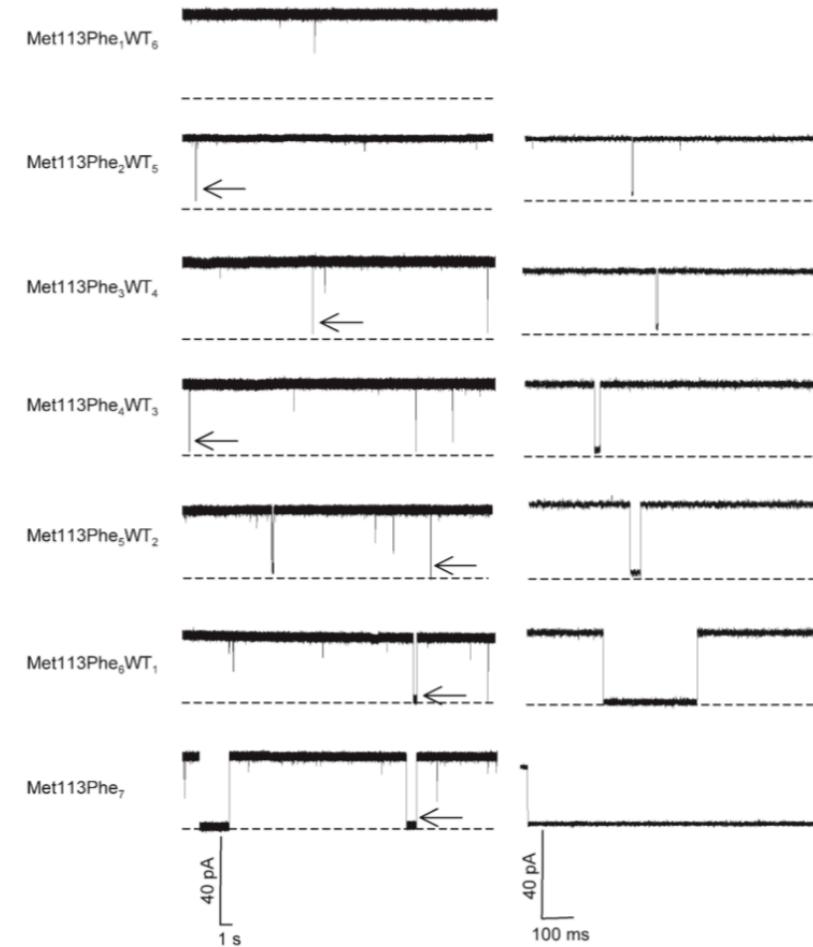
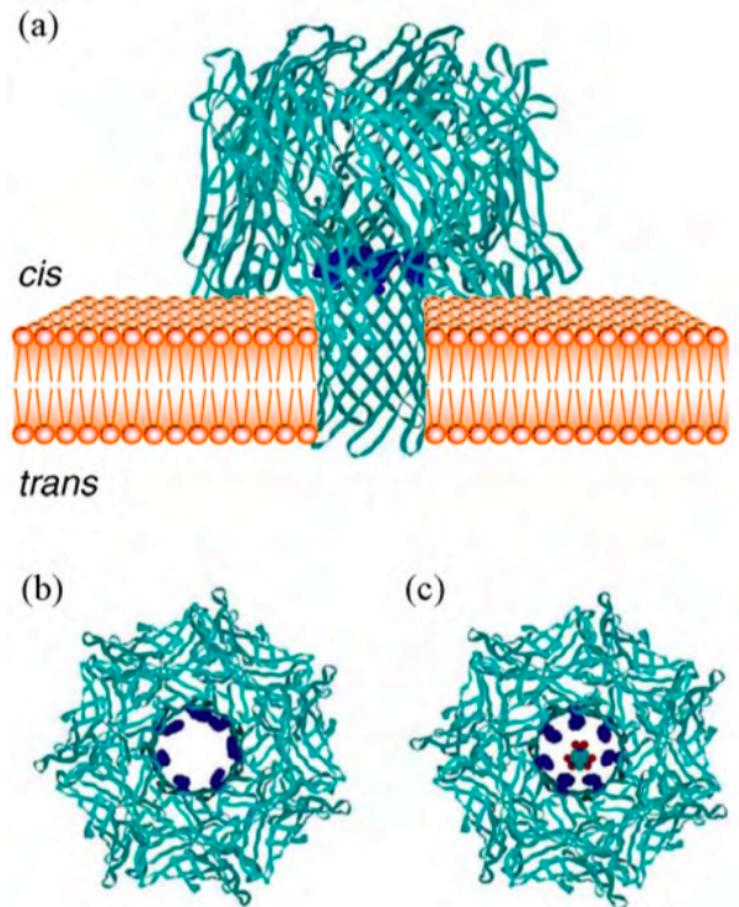
Stochastic nanopore sensors for the detection of terrorist agents



Stochastic nanopore sensors for the detection of terrorist agents



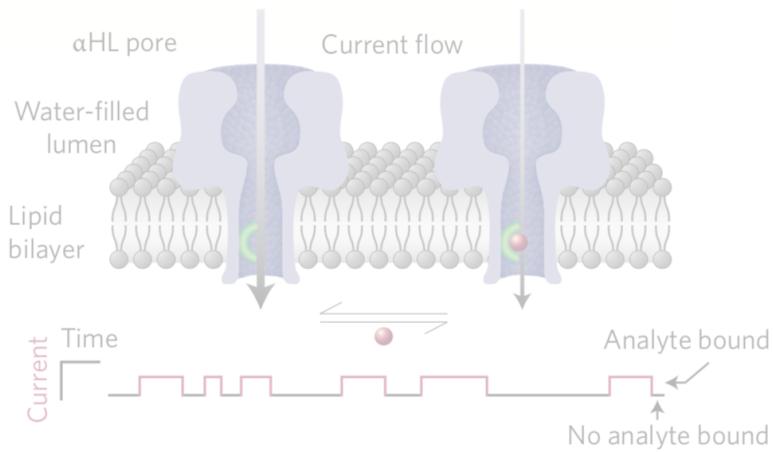
Stochastic nanopore sensors for the detection of terrorist agents



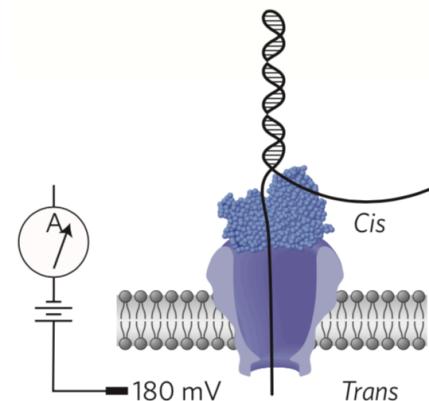
Further reading: Stochastic Sensing of TNT with a Genetically Engineered Pore, DOI: 10.1002/cbic.200500064

Stochastic nanopore sensors for the detection of terrorist agents: Current status and challenges, doi:10.1016/j.aca.2010.07.001
(both uploaded in Toledo)

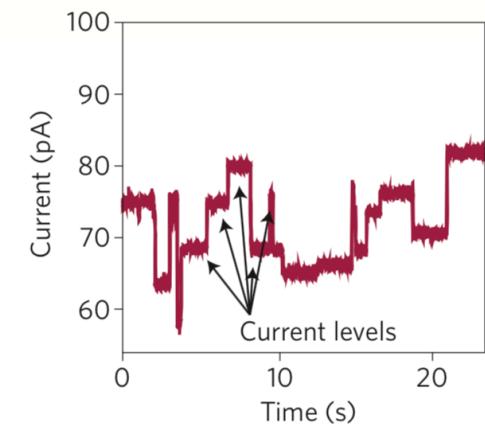
Applications of protein pores



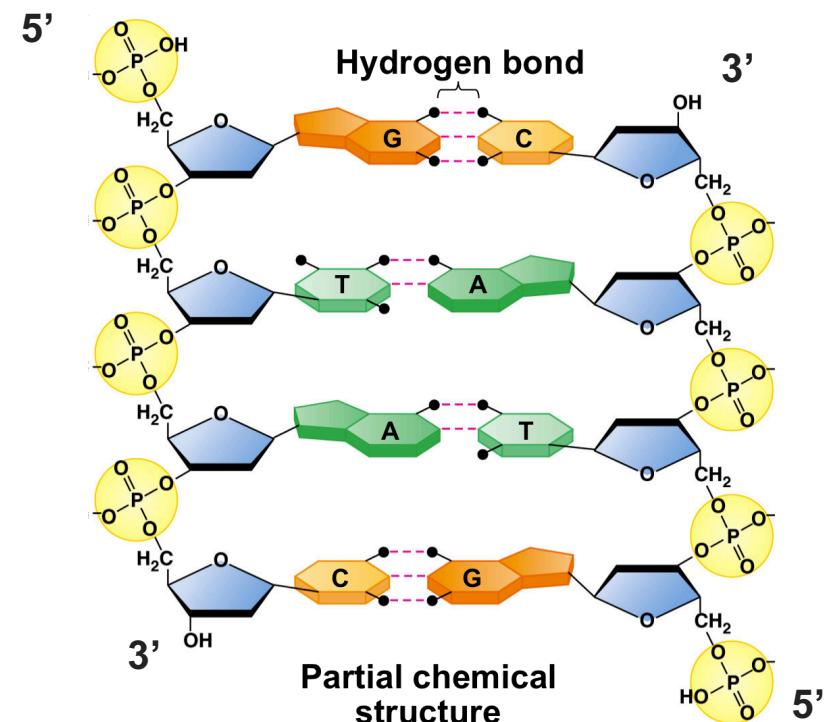
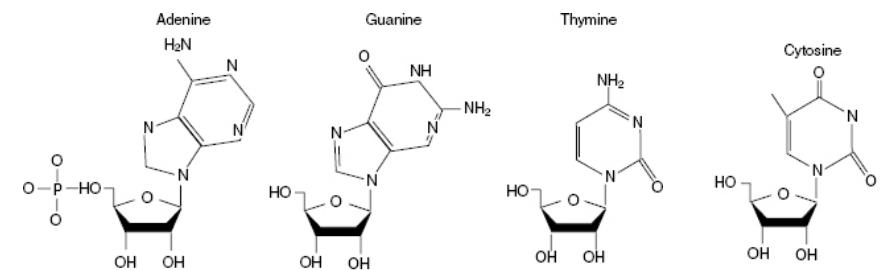
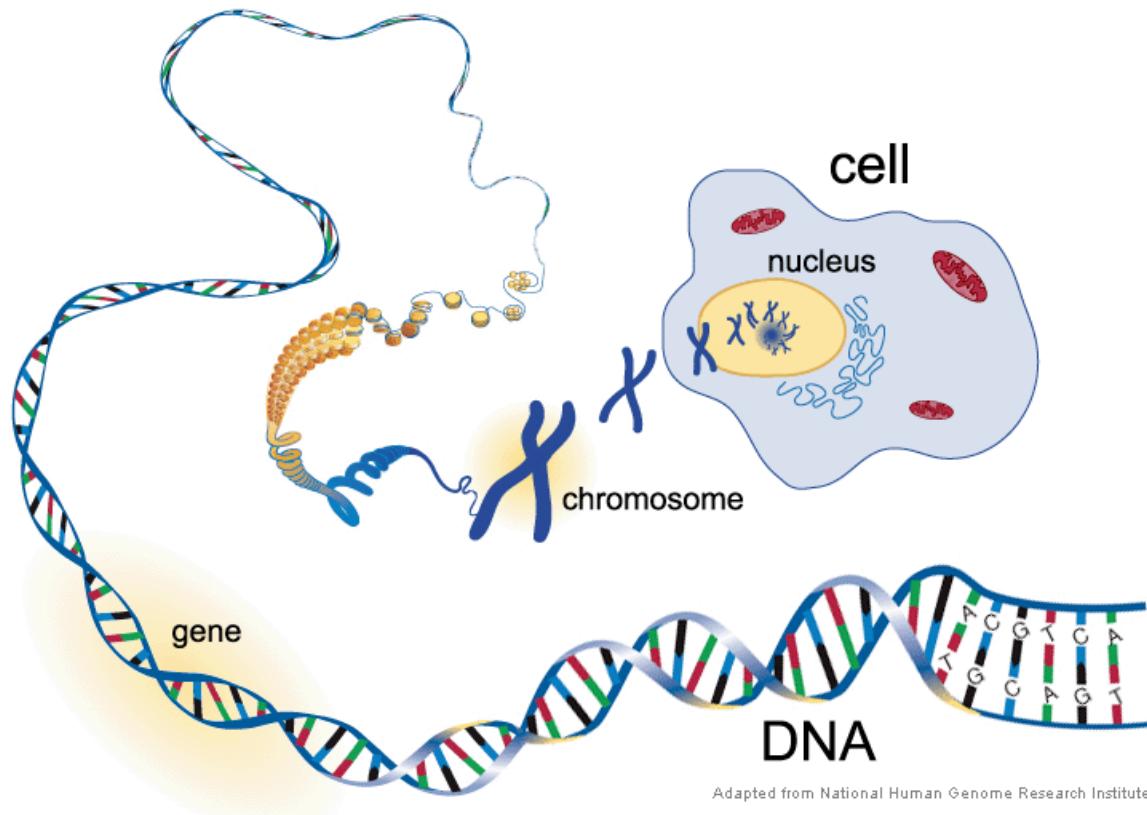
Molecular recognition/binding



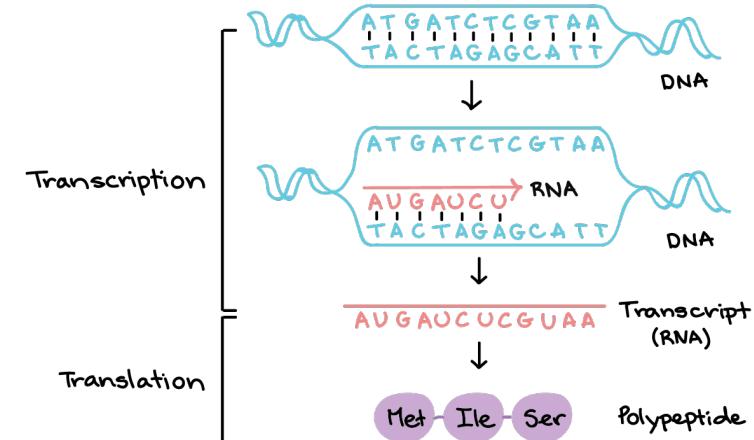
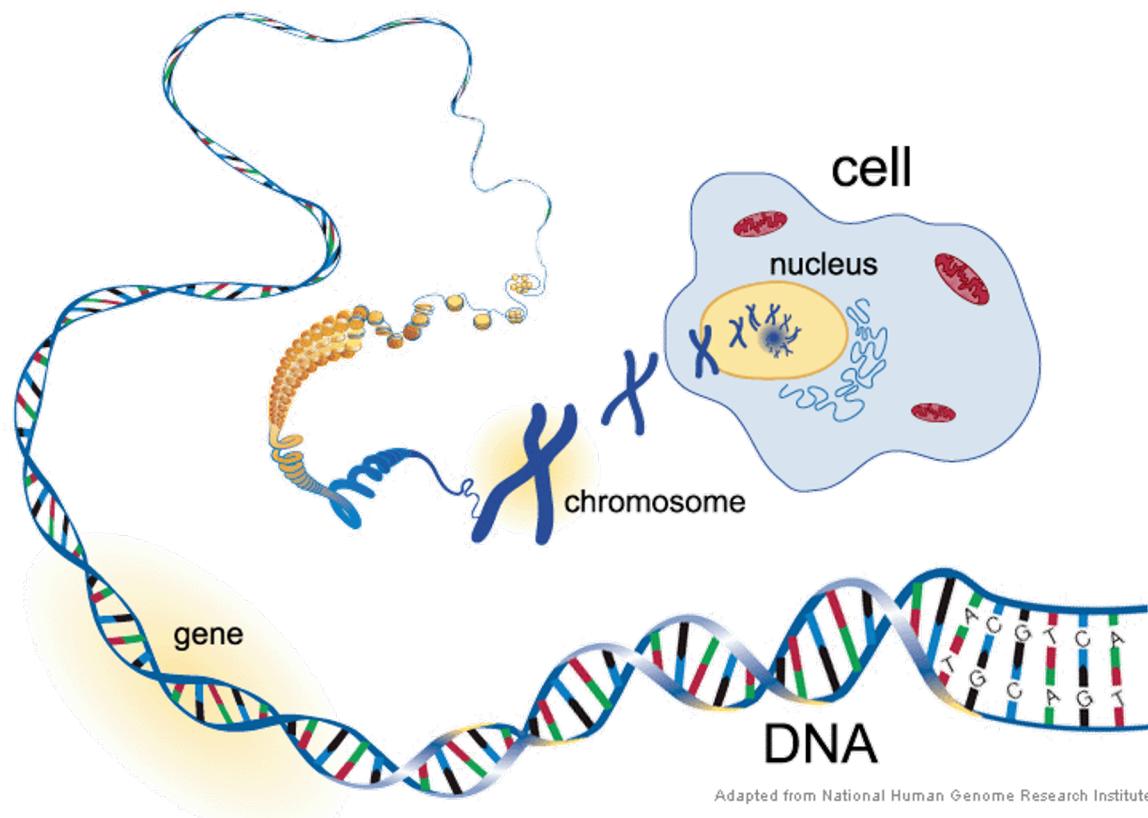
Polymer sequencing
DNA



Basics of DNA



Basics of DNA



Only about 1% of genome directly codes for proteins

About 25% make up genes and their regulatory elements

Human Genome Project (1990-2003)

\$2.7 billion

~3 billion basepairs

How much does it cost to sequence the genome of one person today?

- > US\$ 10?
- > US\$ 100?
- > US\$ 1.000?
- > US\$ 10.000?
- > US\$ 100.000?

How much would it cost today?

Get the **most comprehensive** genetic testing service there is.
Now for **\$599.**

ORDER NOW

Human Genome Project

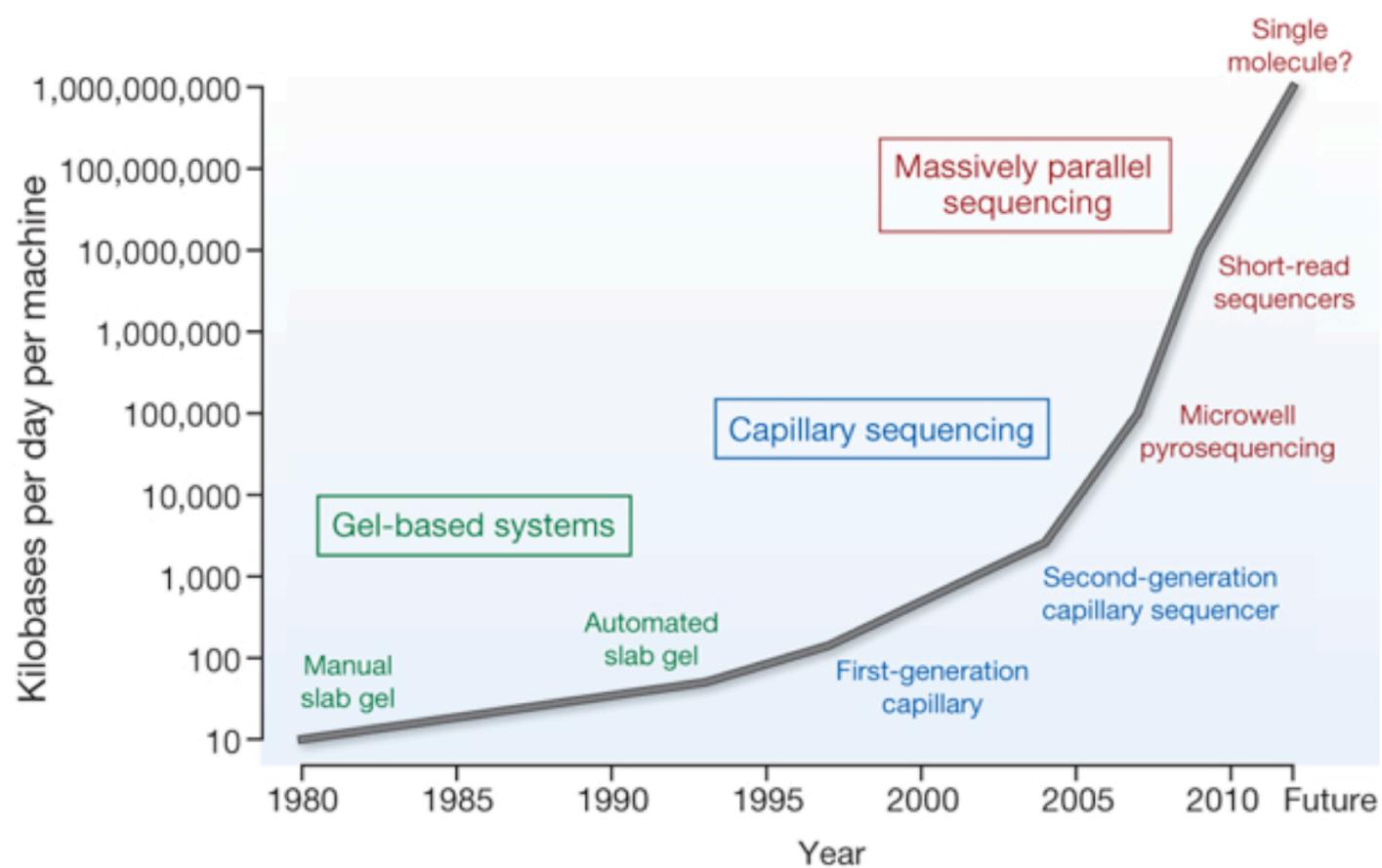
\$2.7 billion
13 years

Veritas
\$600
12 weeks

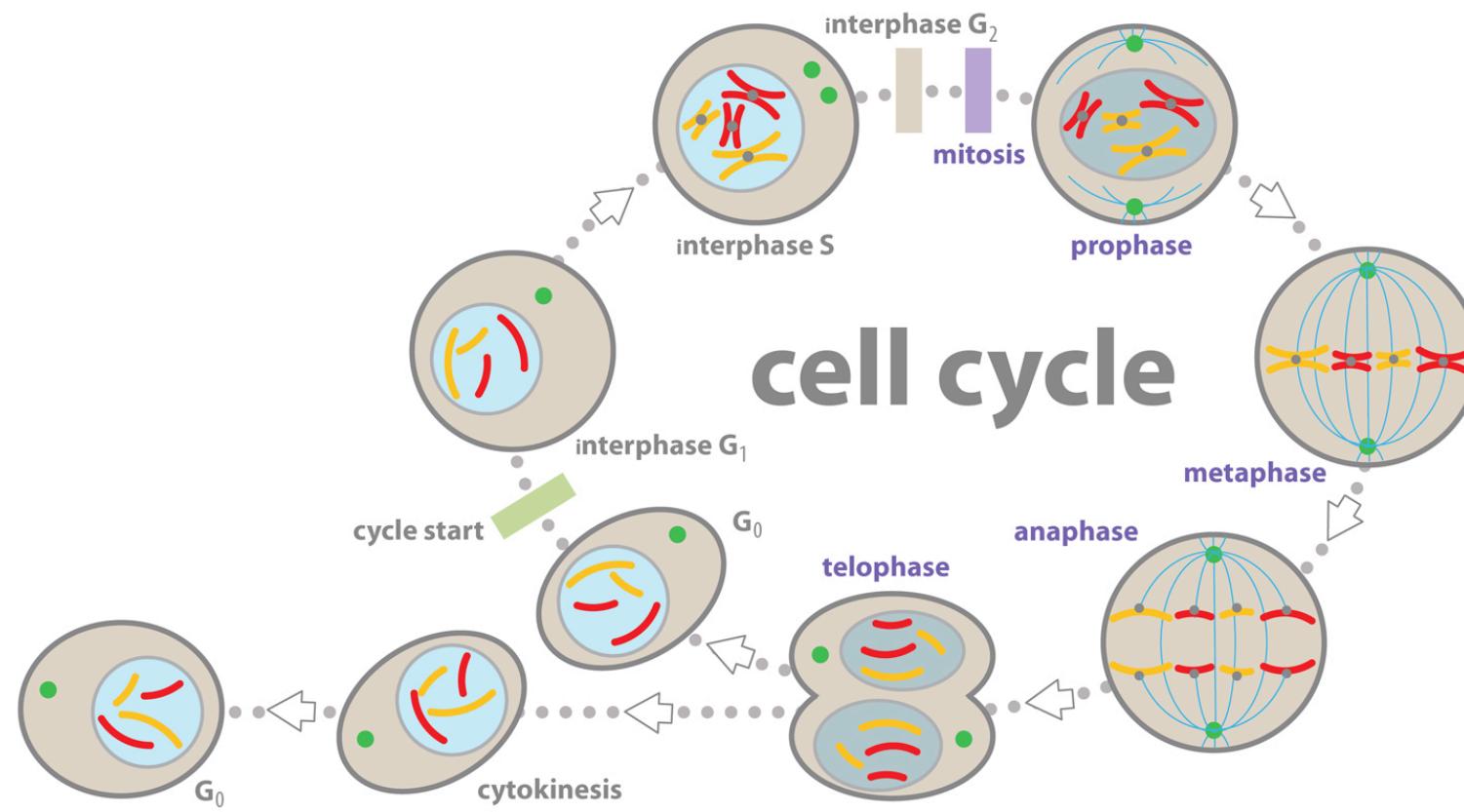


€783.99
Apple iPhone 11 64GB
A2223 Dual Sim

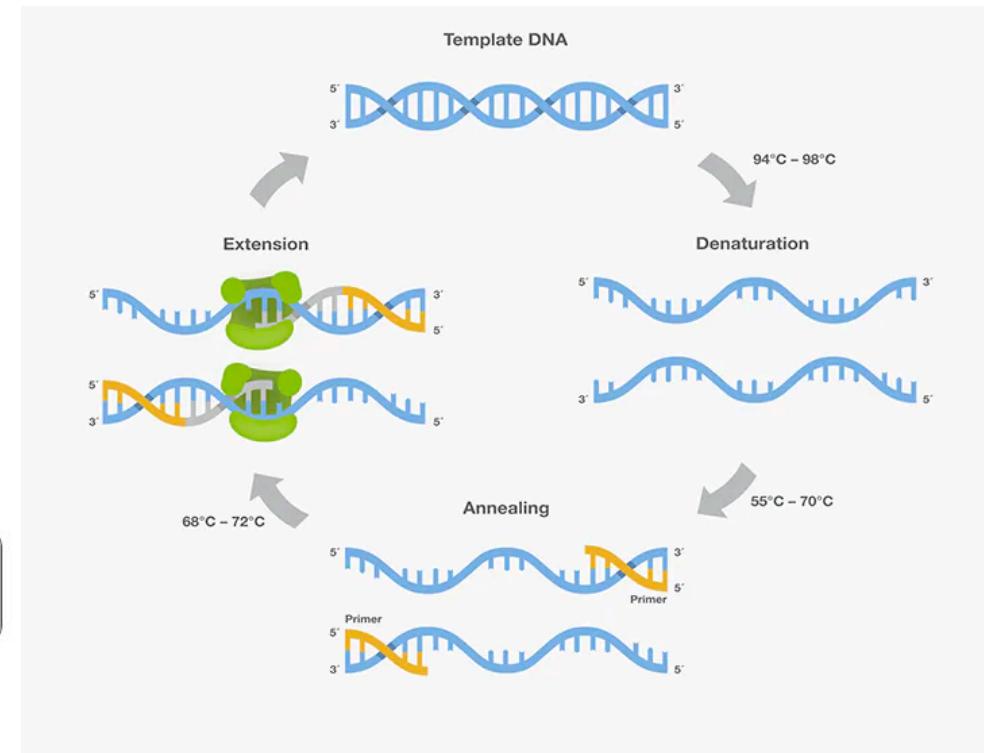
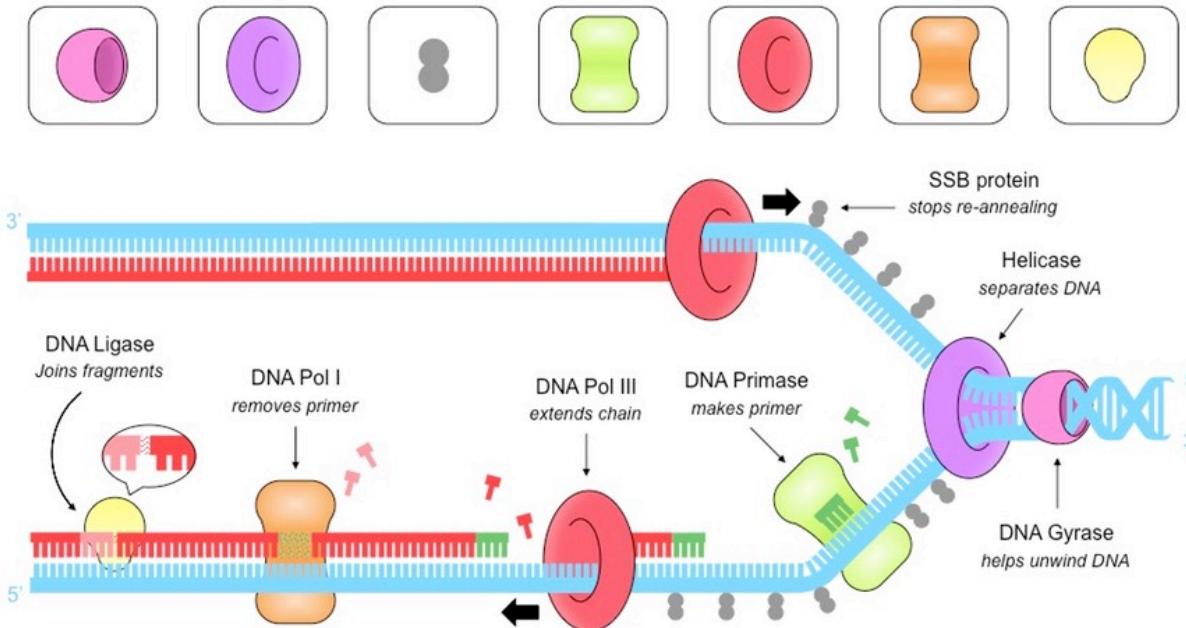
Evolution DNA sequencing techniques



DNA replication

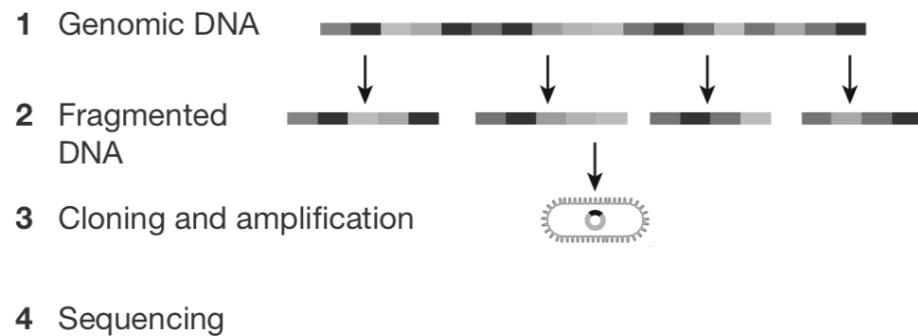


DNA replication



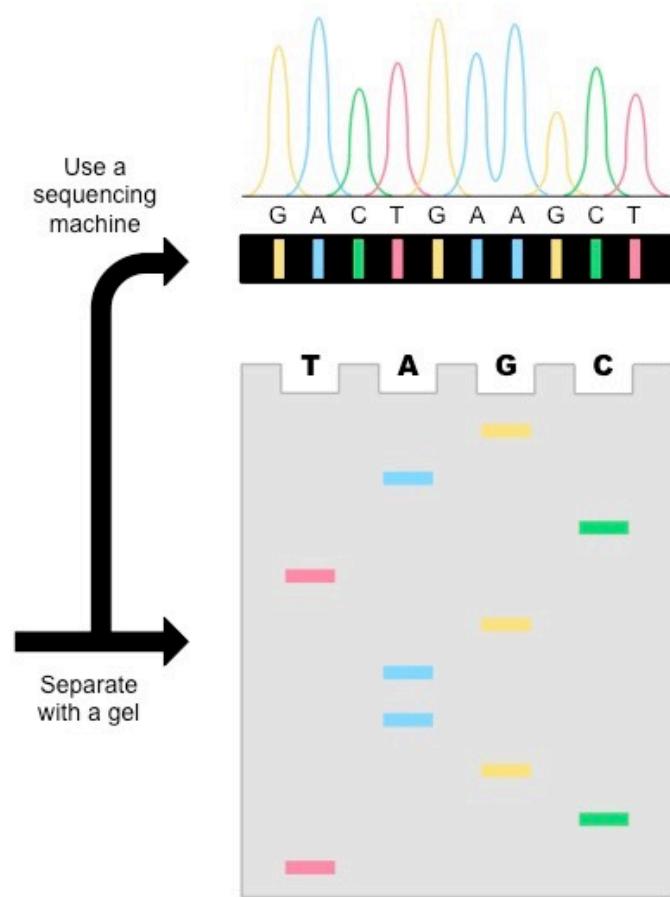
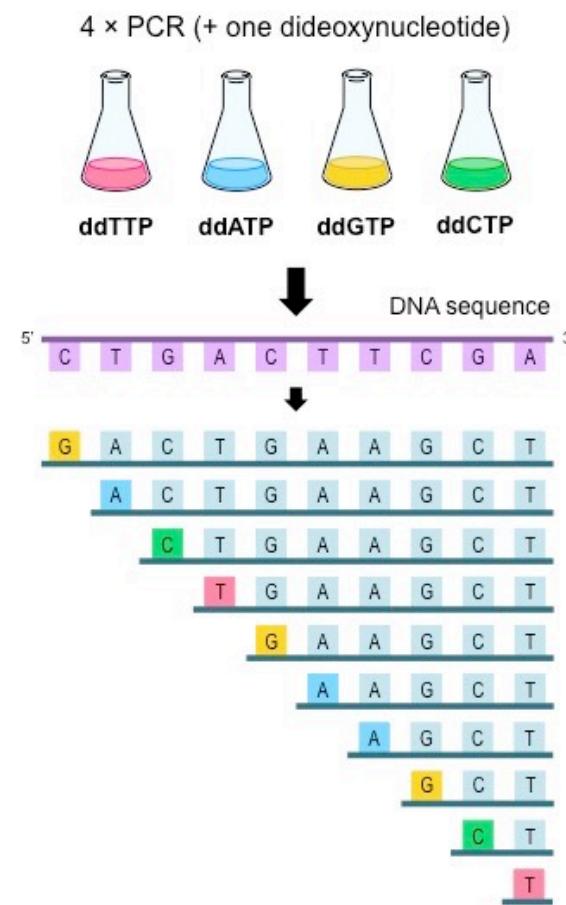
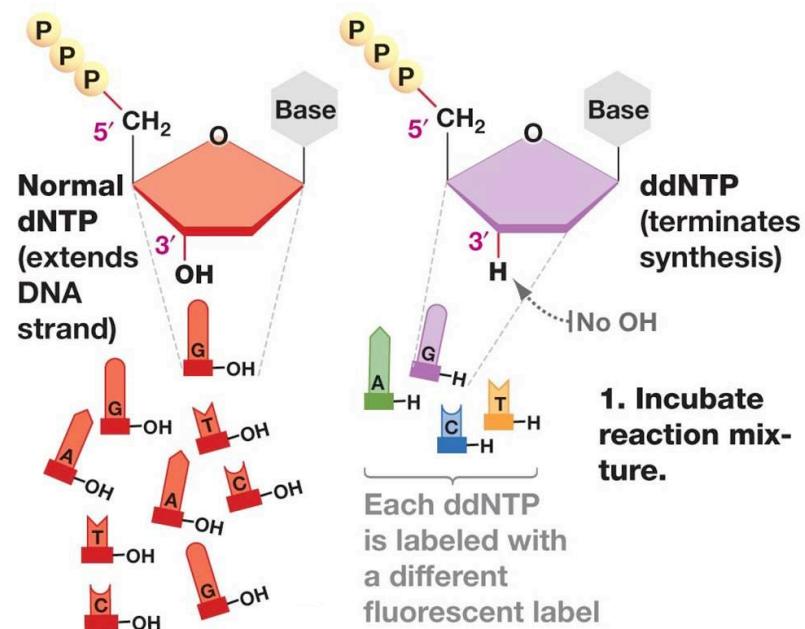
PCR reaction
Polymerase chain reaction

First generation DNA sequencing



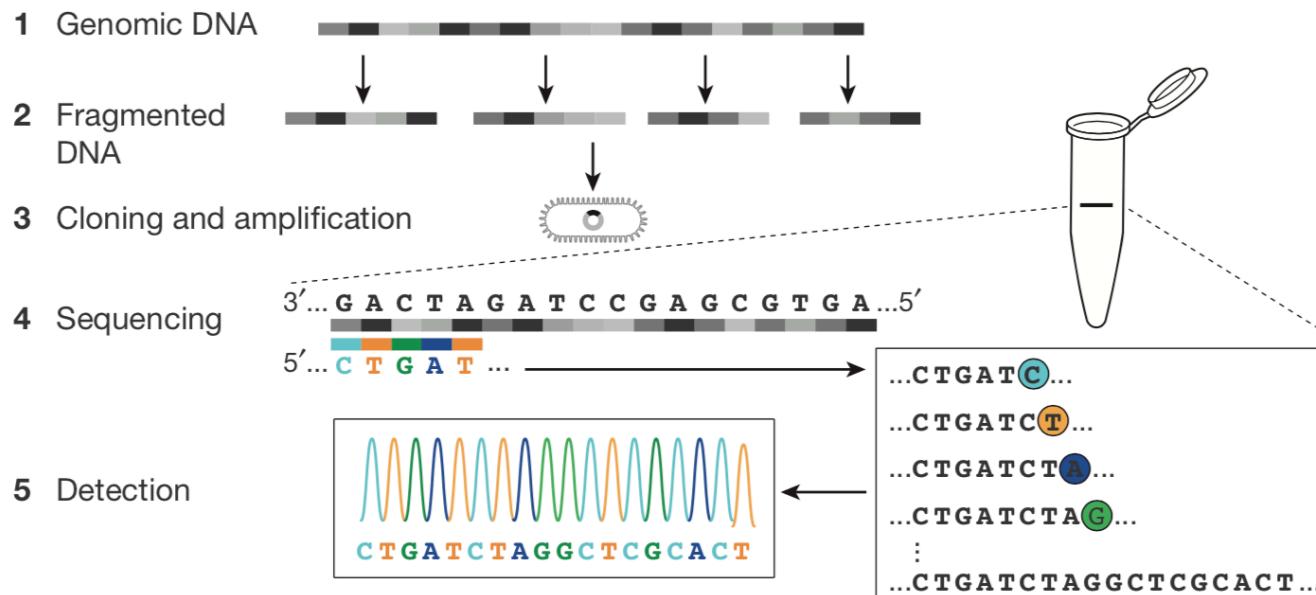
Polymerase chain reaction with special nucleotides

First generation DNA sequencing



Polymerase chain reaction with special nucleotides

First generation DNA sequencing



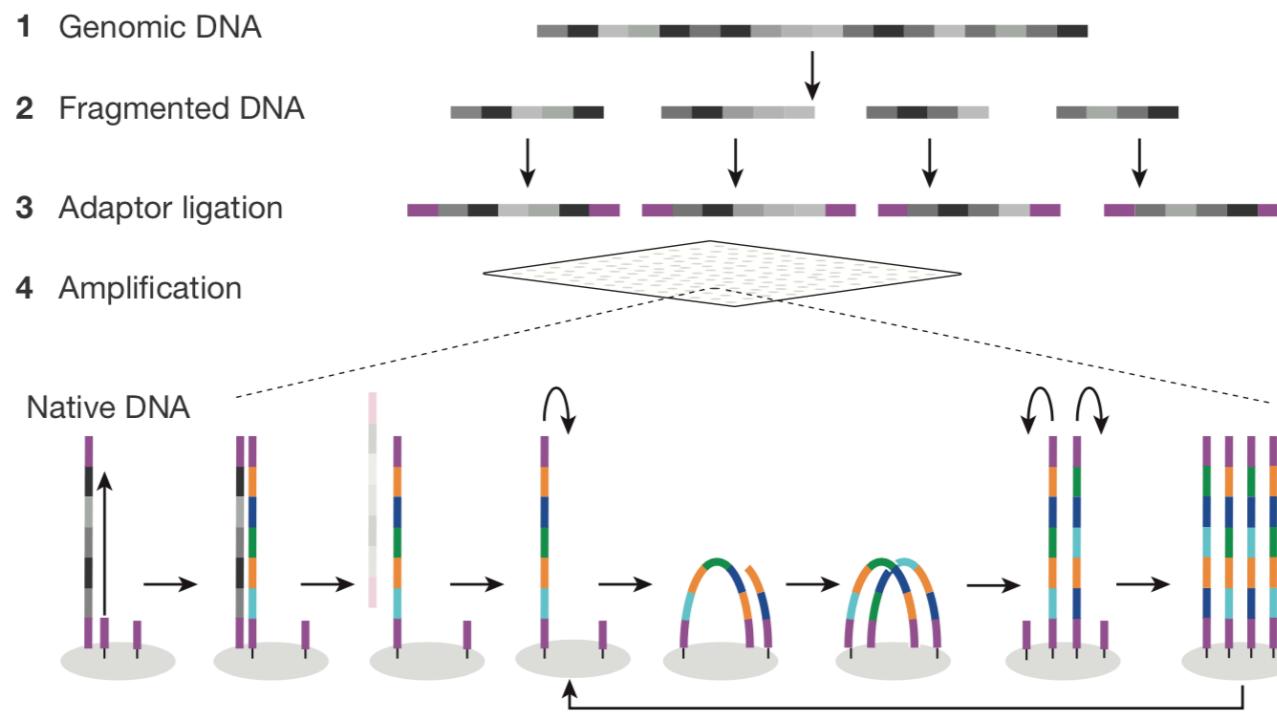
Sanger sequencing

- chain termination method
- last nucleotide has fluorescent dye
- separation by capillary
- electrophoresis
- max read length ~1000 bp

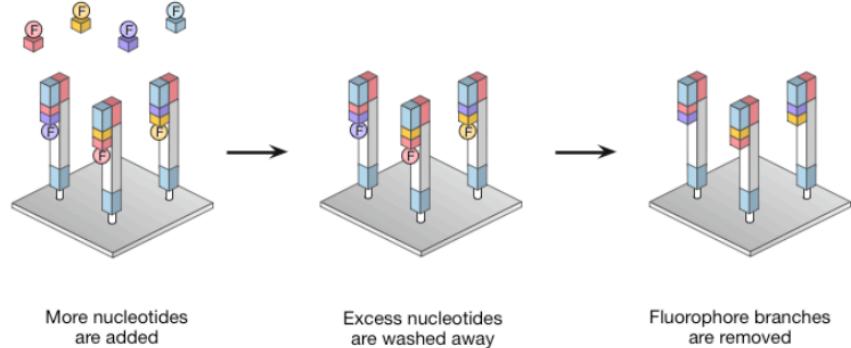
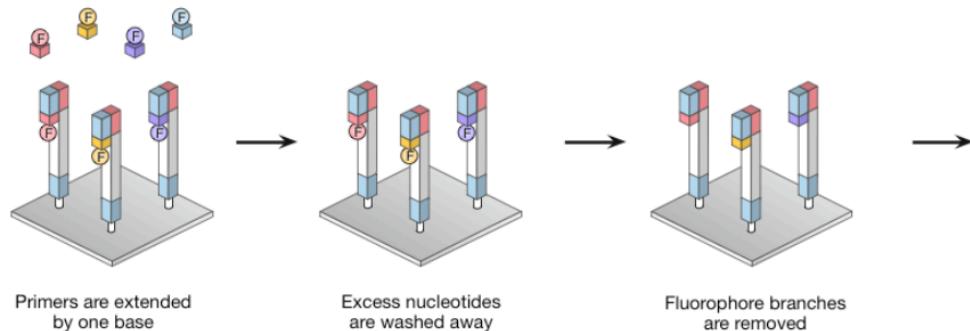
Cost: 2400 USD / 1M bp

Robust, but not cost-effective for entire genomes

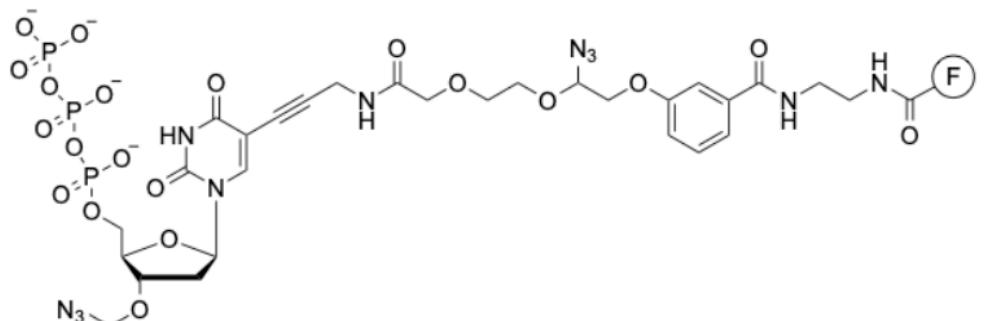
Second generation (next generation sequencing, NGS)



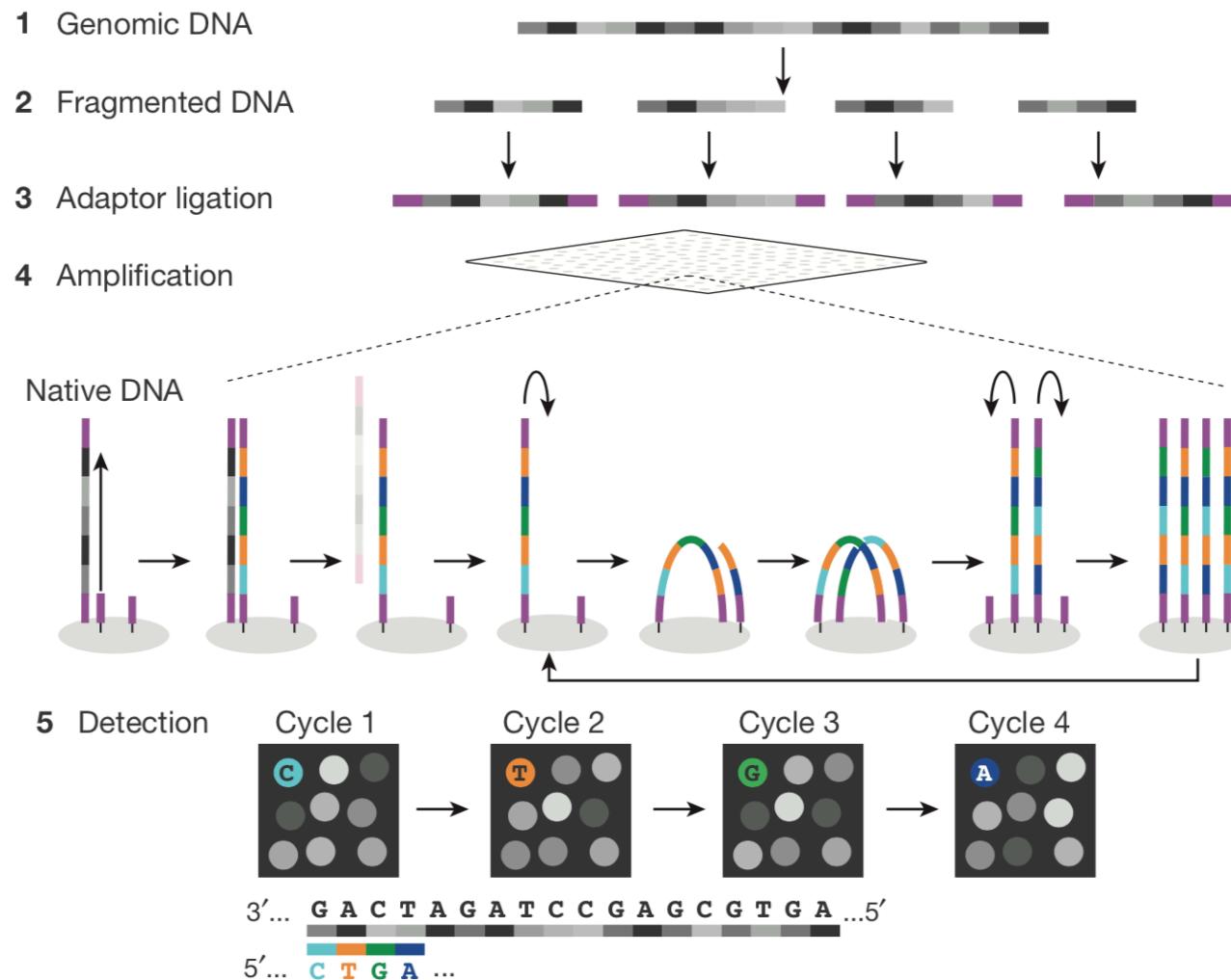
Second generation (next generation sequencing, NGS)



■ A ■ C ■ F Fluorophore
■ G ■ T



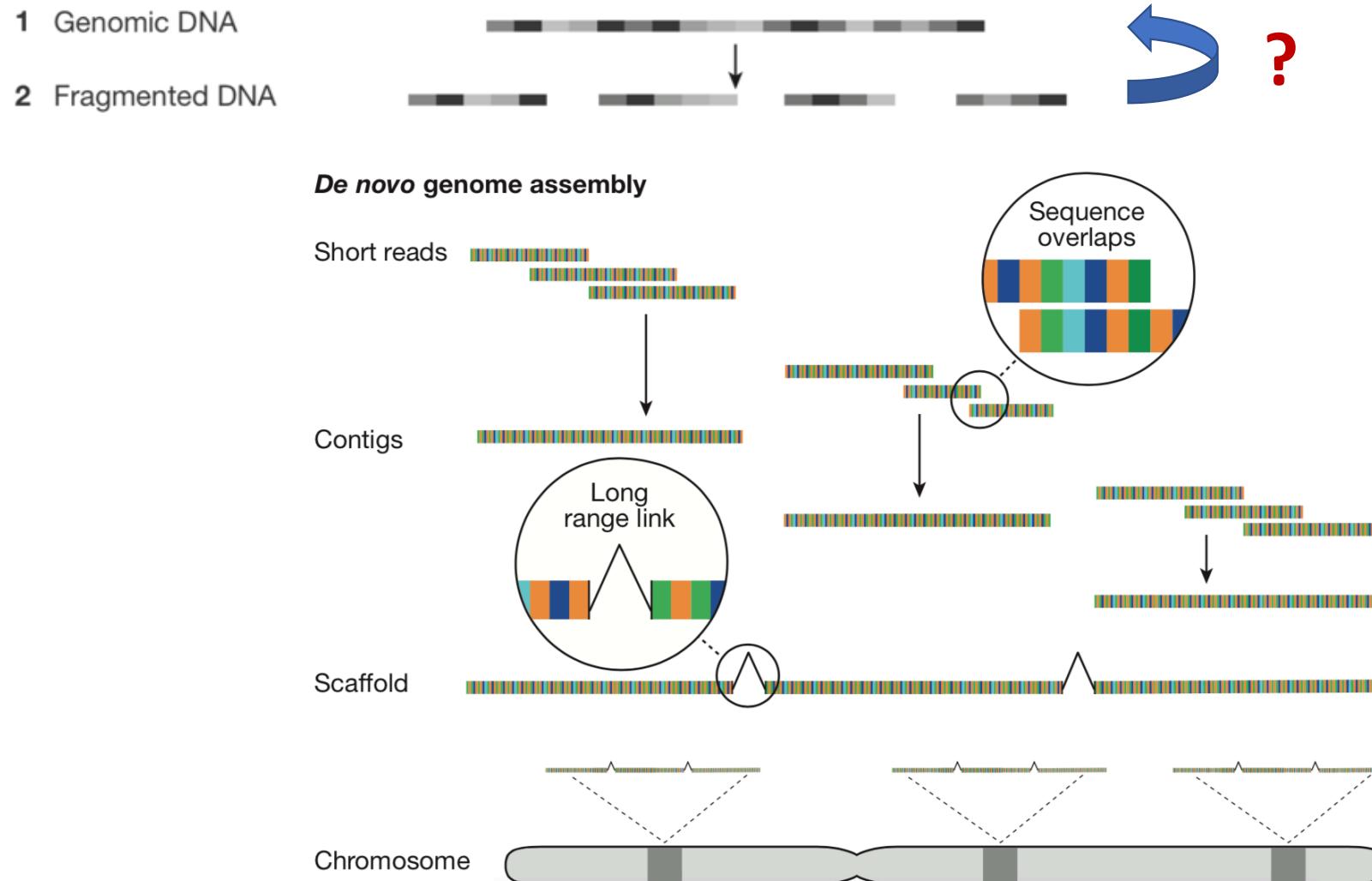
Second generation (next generation sequencing, NGS)



Solexa (Illumina)

- Max read length: ~50-500 bp
- Market leader

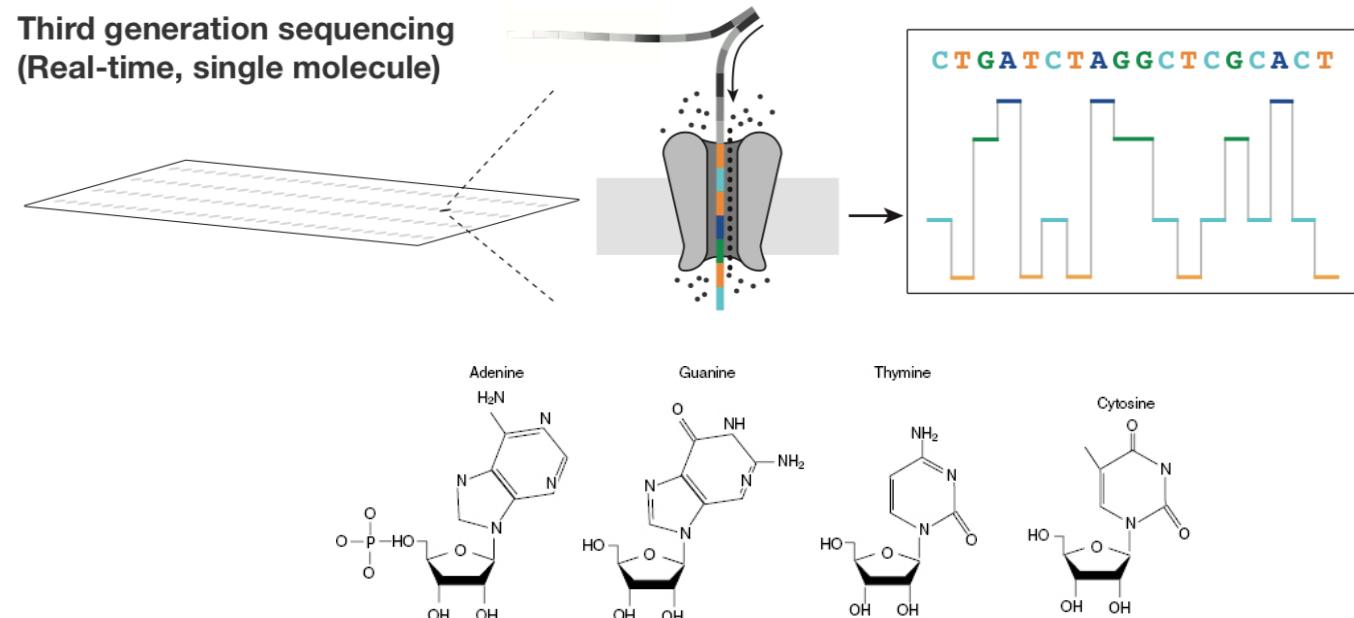
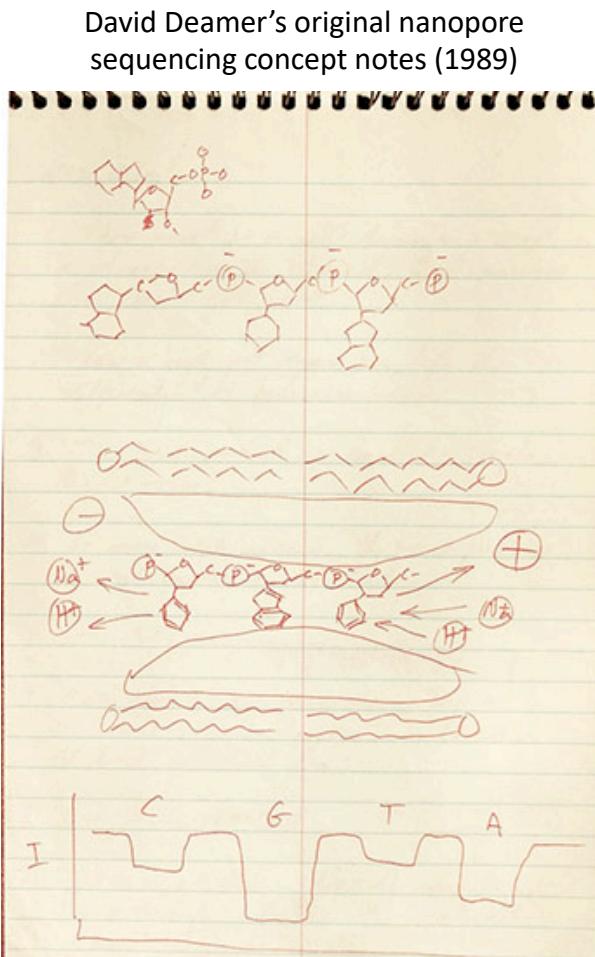
De novo genome assembly



Next generation methods - Illumina

https://www.youtube.com/watch?time_continue=2&v=fCd6B5HRaZ8

Third generation – real time, single molecule



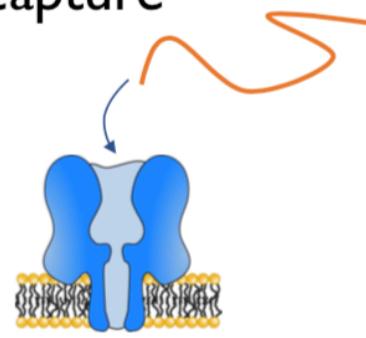
Third generation technologies obtain very long, single-molecule reads

Single-molecule reads

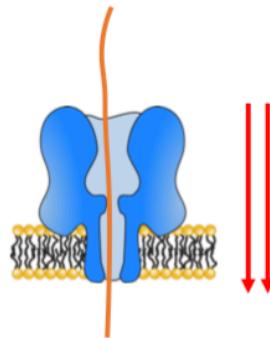
- Low sample concentration possible
- No amplification bias!

The challenges of strand sequencing

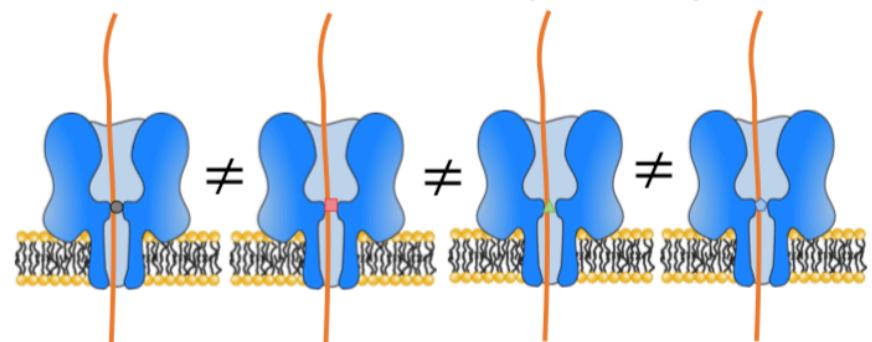
I. DNA capture



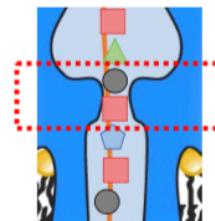
2. Translocation speed



3. Base discrimination (GATC)



4. Single base resolution



DNA capture

Single stranded DNA is readily captured due to its high negative charge

- DNA is a highly negatively charged polymer (-1 e / 0.3 nm)
- High electrical field in and around nanopore automatically captures DNA and translocates it
- Single stranded DNA is balled up in solution, double stranded DNA is more rigid!

Translocation speed

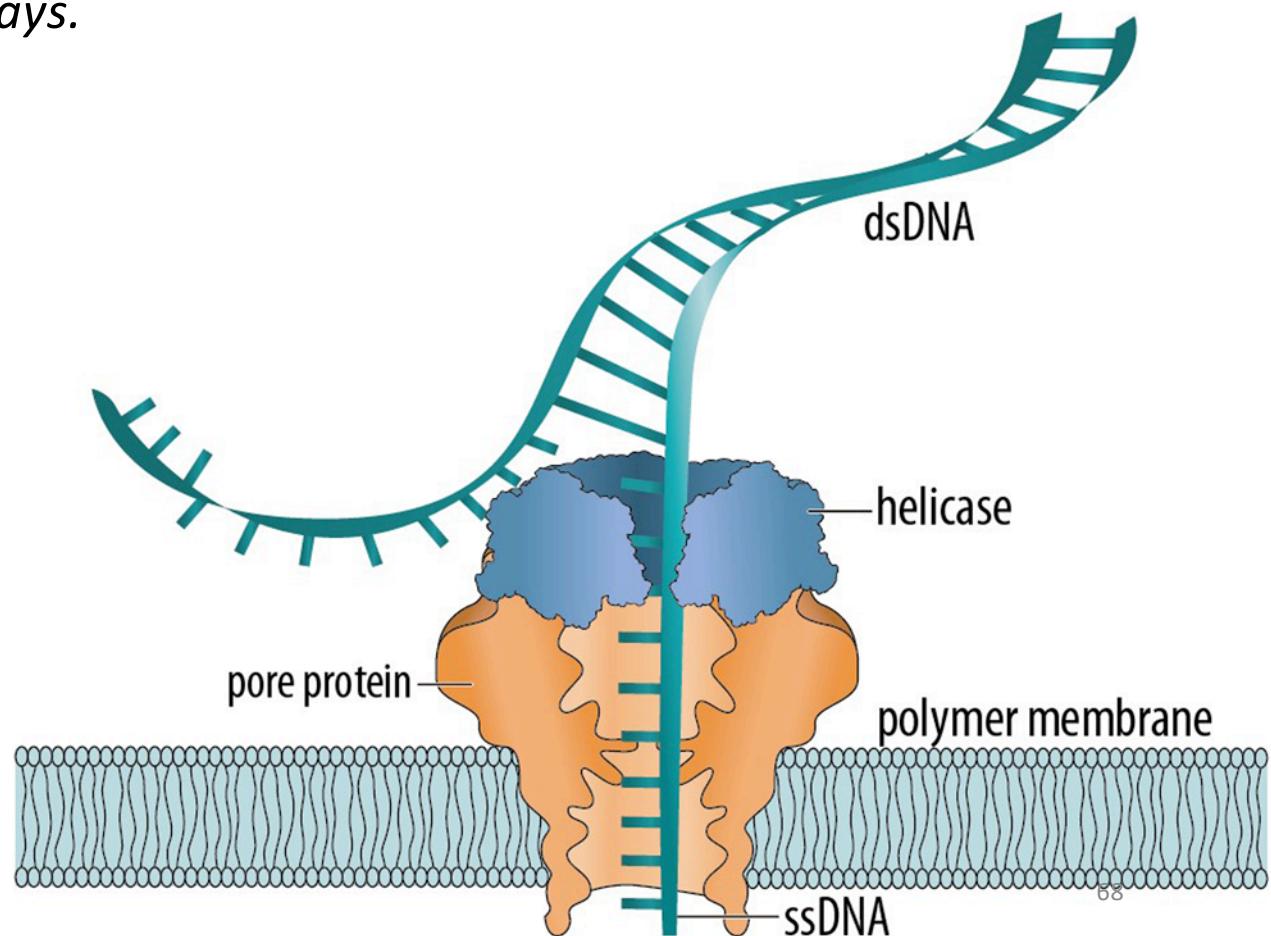
Translocation speed of DNA is generally too fast, but it can be controlled in various ways.

Translocation rate

- naked DNA: $< 1 \mu\text{s}/\text{base}$
- ideally: $> 1 \text{ ms}/\text{base}$

Non-enzymatic control

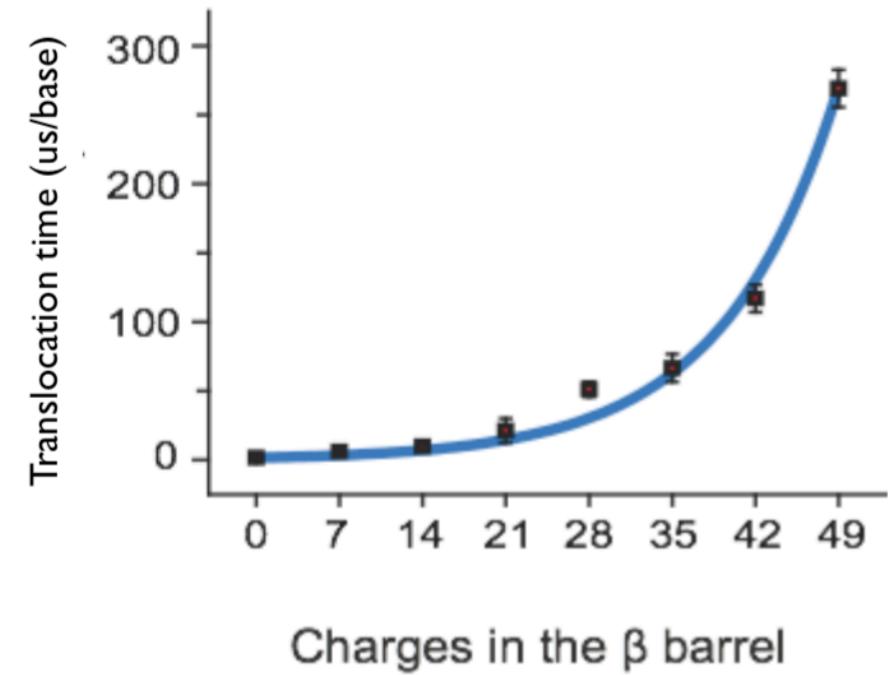
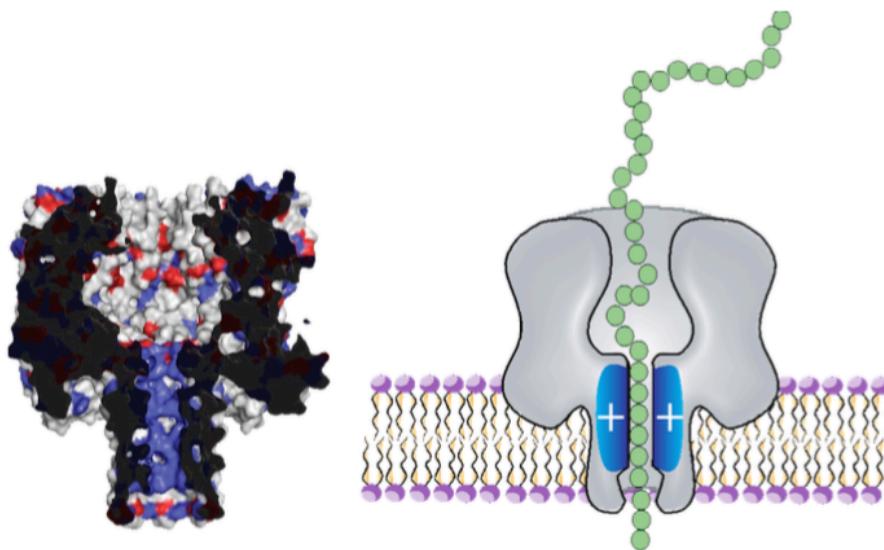
- Modification of pore
- Modification of DNA
- Modification of electrolyte



Electrostatic interactions can dramatically reduce DNA translocation speed

Electrostatic 'brakes'

- additional positive charges
- slowdown of DNA translocation



Translocation speed

Translocation speed of DNA is generally too fast, but it can be controlled in various ways.

Translocation rate

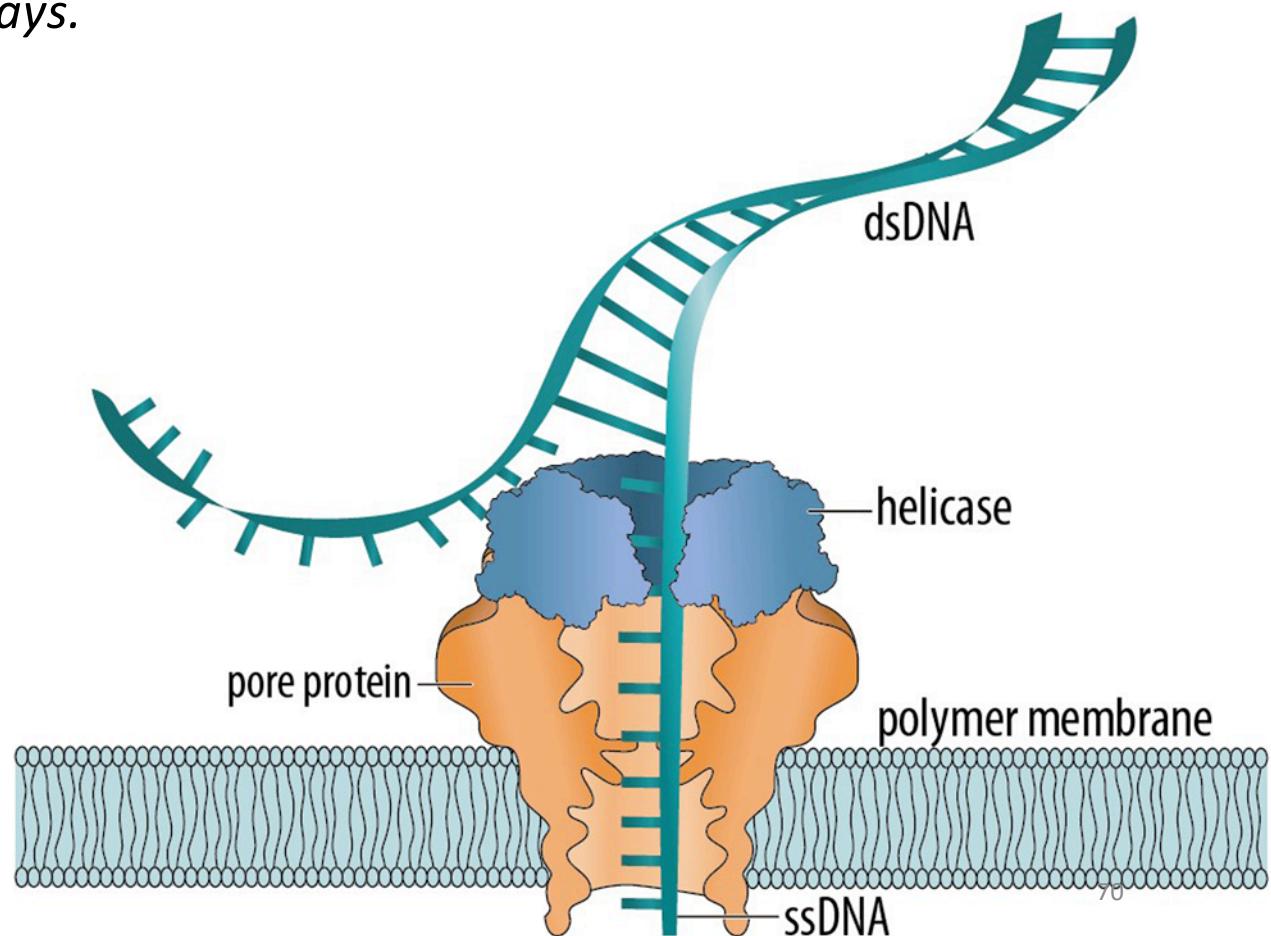
- naked DNA: $< 1 \mu\text{s}/\text{base}$
- ideally: $> 1 \text{ ms}/\text{base}$

Non-enzymatic control

- Modification of pore
- Modification of DNA
- Modification of electrolyte

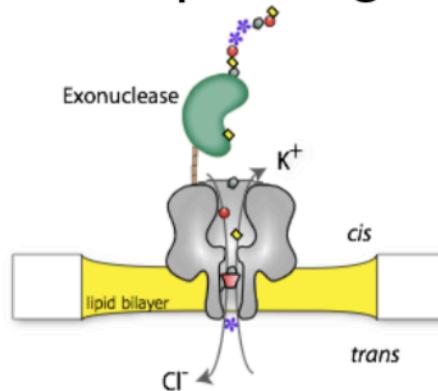
Enzymatic control

- DNA polymerase
- Processive exonuclease
- Helicase
- Single strand binding protein



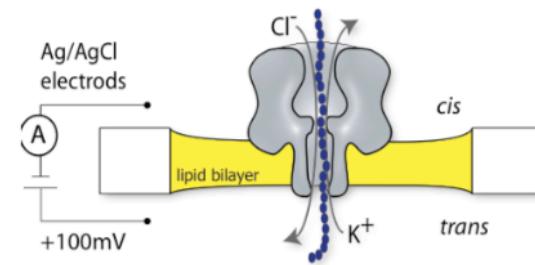
DNA sequencing with (biological) nanopores

Exonuclease sequencing



- Exonuclease enzyme cleaves DNA one base at a time
- Each cleaved base is captured by the nanopore and identified by its unique ionic current disruption

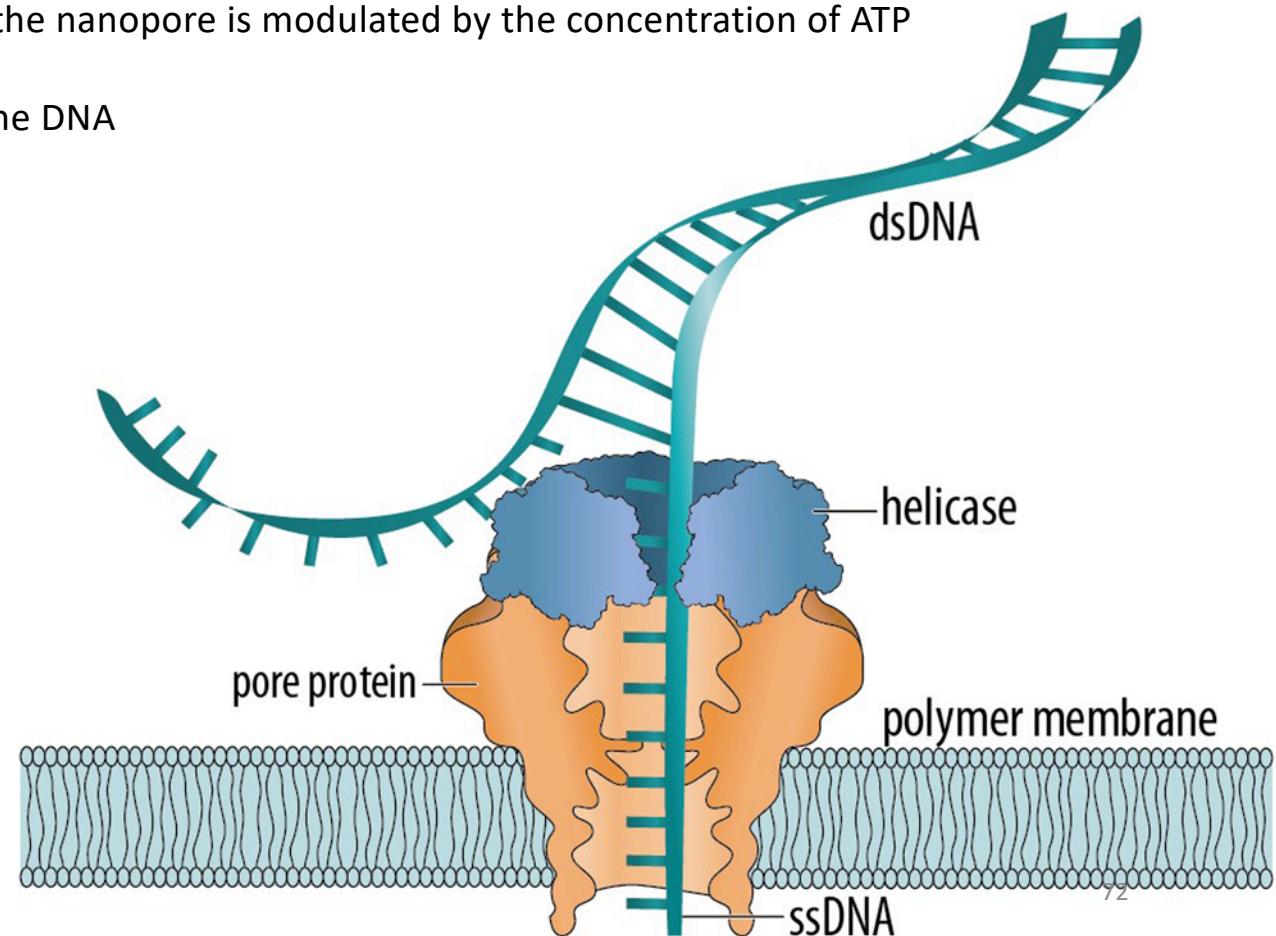
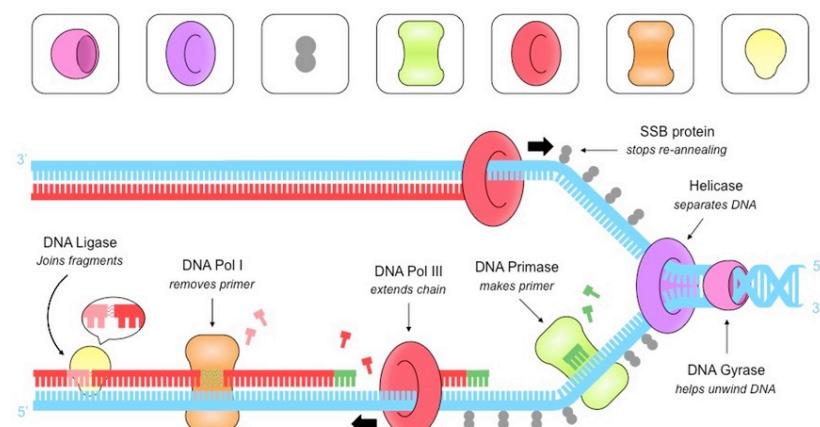
Strand sequencing



- Entire ssDNA strand is driven through the pore via electrophoresis
- The base at the narrowest location in the pore determines the ionic current disruption

Enzymatic control using DNA helicase

- Enzymes are used to control the speed of translocation across the nanopore
- Using an helicase, the speed of DNA across the nanopore is modulated by the concentration of ATP
- DNA is recognized during the unzipping of the DNA



Commercial devices

Oxford Nanopore technologies

- First commercial nanopore sequencing devices
- From handheld to large scale

Core technologies

- Strand sequencing with biological pore (CsgG) and enzymatic translocation control
- 512 microfluidics ‘wells’ with a polymerized block-copolymer bilayer (stability!)
- Disposable chamber

MinION



GridION



PromethION



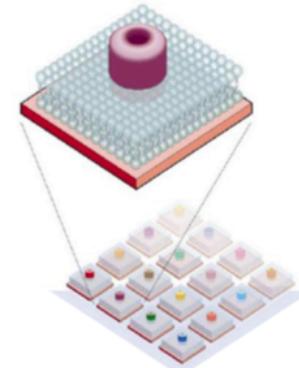
Block-copolymer



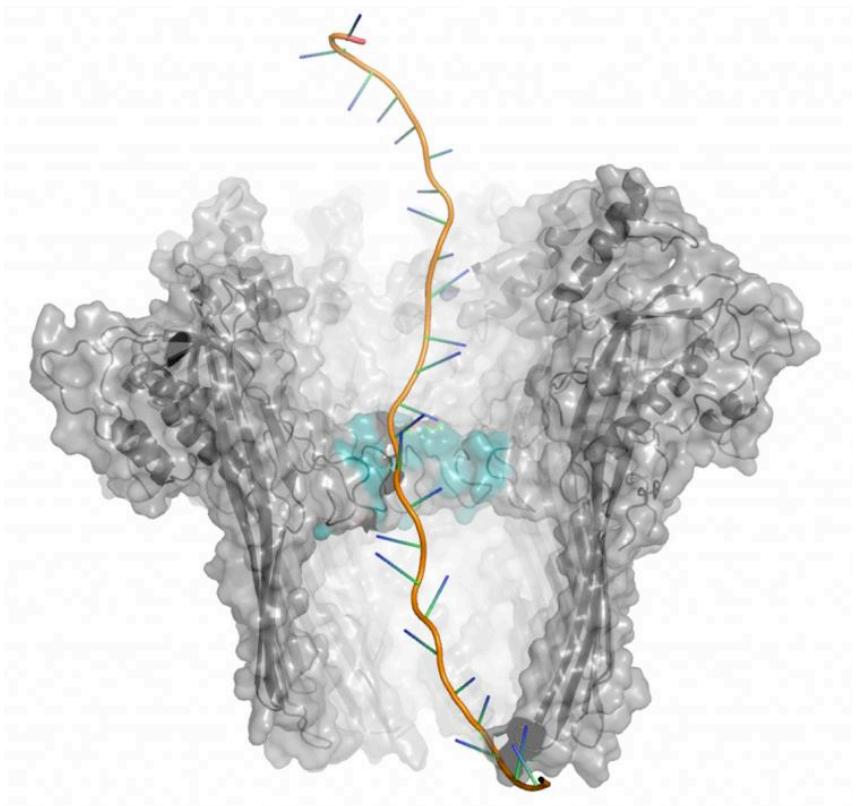
Hydrophilic

Hydrophobic

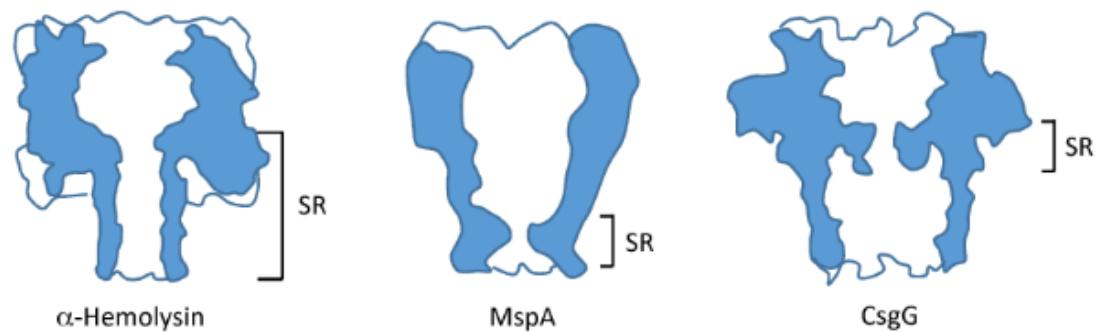
Hydrophilic



Nanopore used



very narrow and well-defined passage for a DNA strand



Learning objectives 2nd lesson

Protein production:

- Gene editing
- Cell-based systems
- Case study: insulin

Nanopores

- What are nanopores
- How do nanopores work
- What can nanopores be used for (determination chemicals, DNA sequencing)