Towards ultra-fast sequencing of single DNA molecules with biological nanopores

Andrew Heron

Oxford University
Department of Chemistry

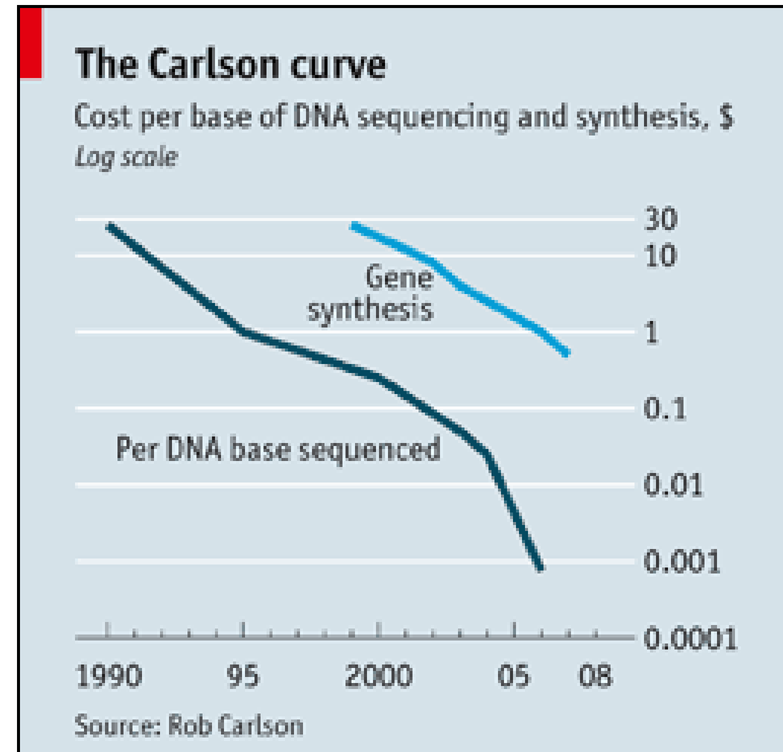
New and Developing Technologies for Genetics Diagnostics
5-6th July 2010, Salisbury

Today's Talk

- rationale for DNA sequencing with nanopores
- single-molecule detection with protein nanopores
- approaches to sequencing with protein pores
  - focus on base identification
- Future work
  - Enzymes to slow DNA
  - arrays for nanopore sequencing

Next generation is already here…

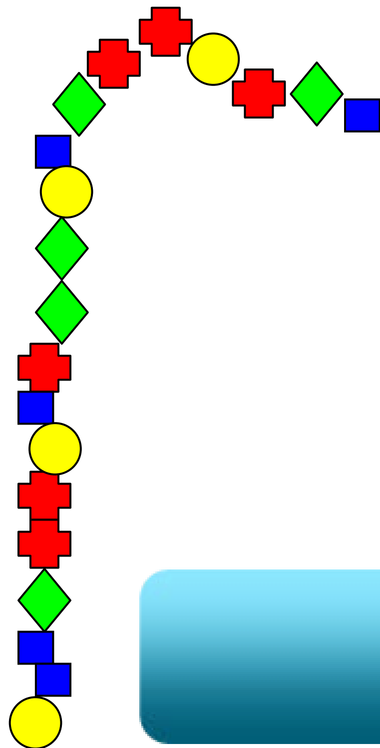
- NIH "Human Genome Project" 1990-2000 (human genome sequence released in 2003, cost: $3 billion)
- NIH "$100,000 genome" 2004-2009
- NIH "$1000 genome" 2009-2013
- on the way (Illumina, Roche, ABI/Helicos/PacBio)
- Speed, long reads and modified bases will be the focus

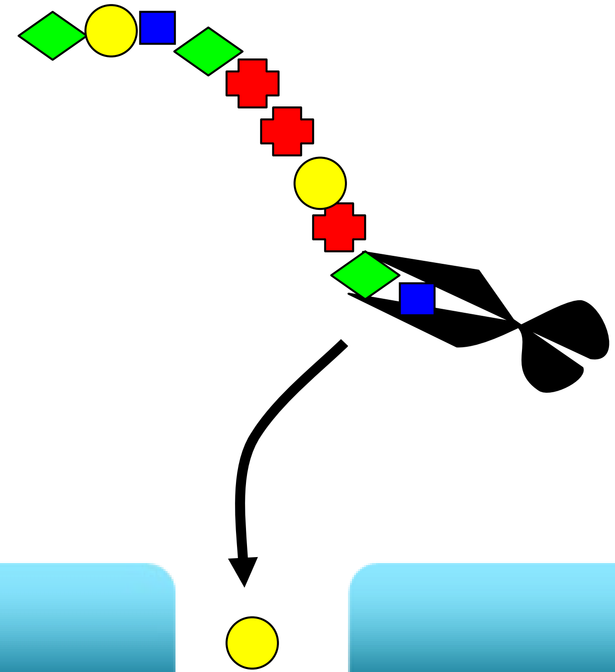


What is Nanopore Sequencing?

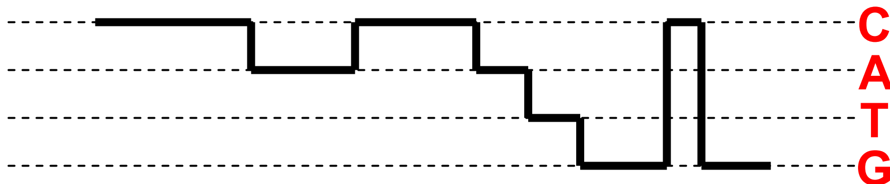
Strand Sequencing

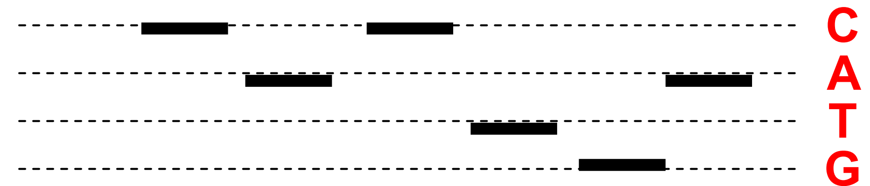
Base Sequencing





Membrane

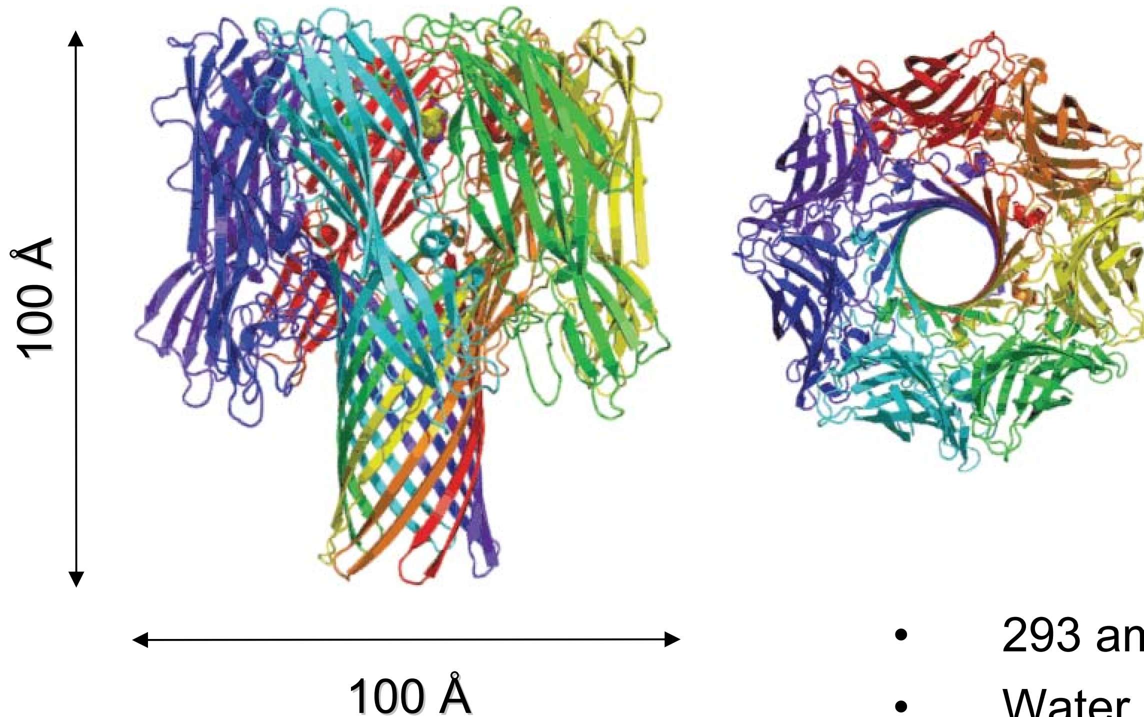




Potential advantages of Nanopore Sequencing (Next, next generation…)

- Single Molecule technique
- Native nucleobase identification (no need for labeling)
- Direct identification of modified bases (e.g. MeC, hMeC)
  - epigenetics
- No need for DNA amplification (Single cells?)
- Speed (up to 10-1000nt/s, x10^6 parallel pores)
  - The 10minute genome?
- Long sequencing reads (>10 kb?)
- Adaptable to direct RNA sequencing

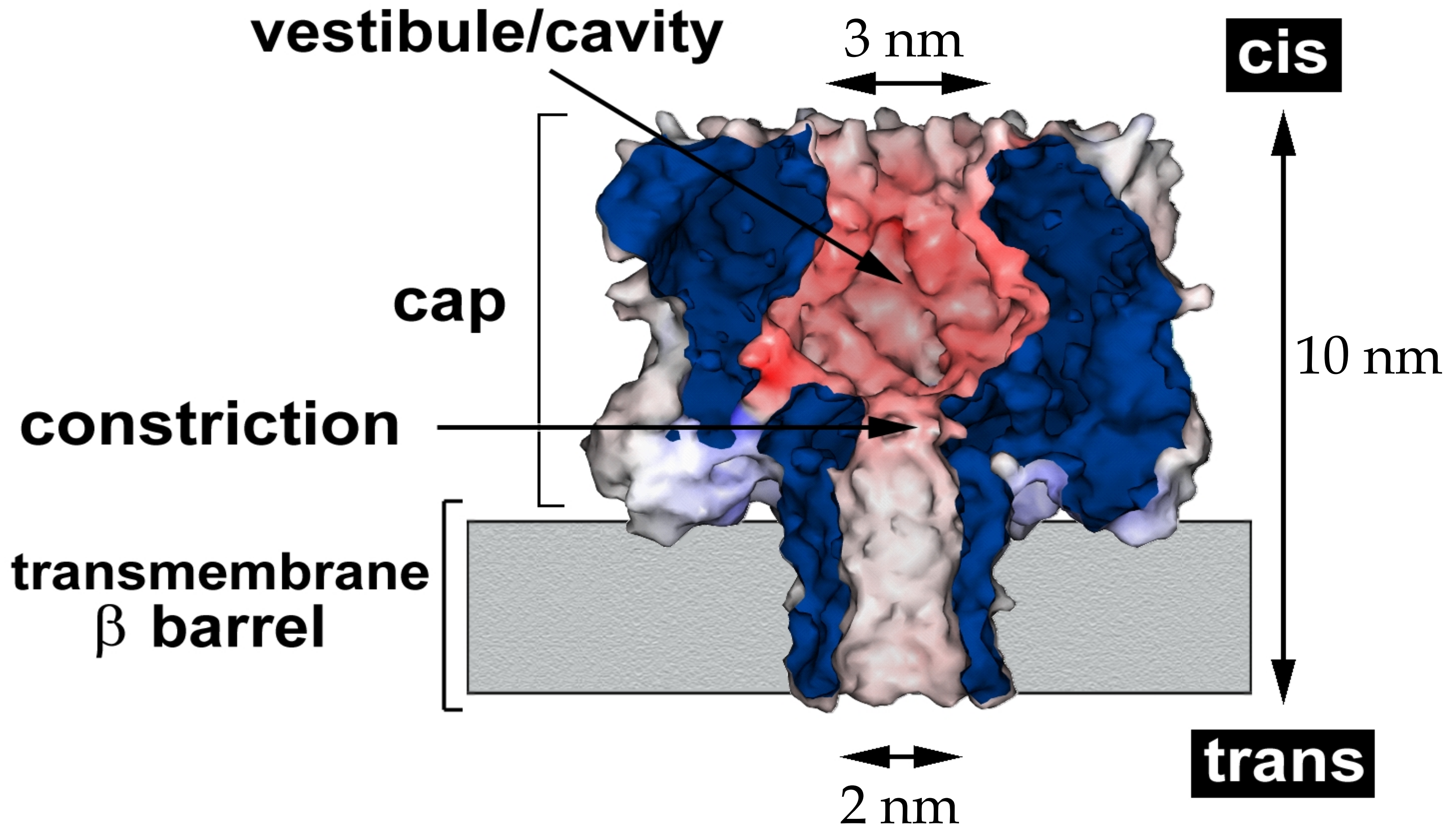
α-hemolysin



Science *274*, 1859-1865 (1996)

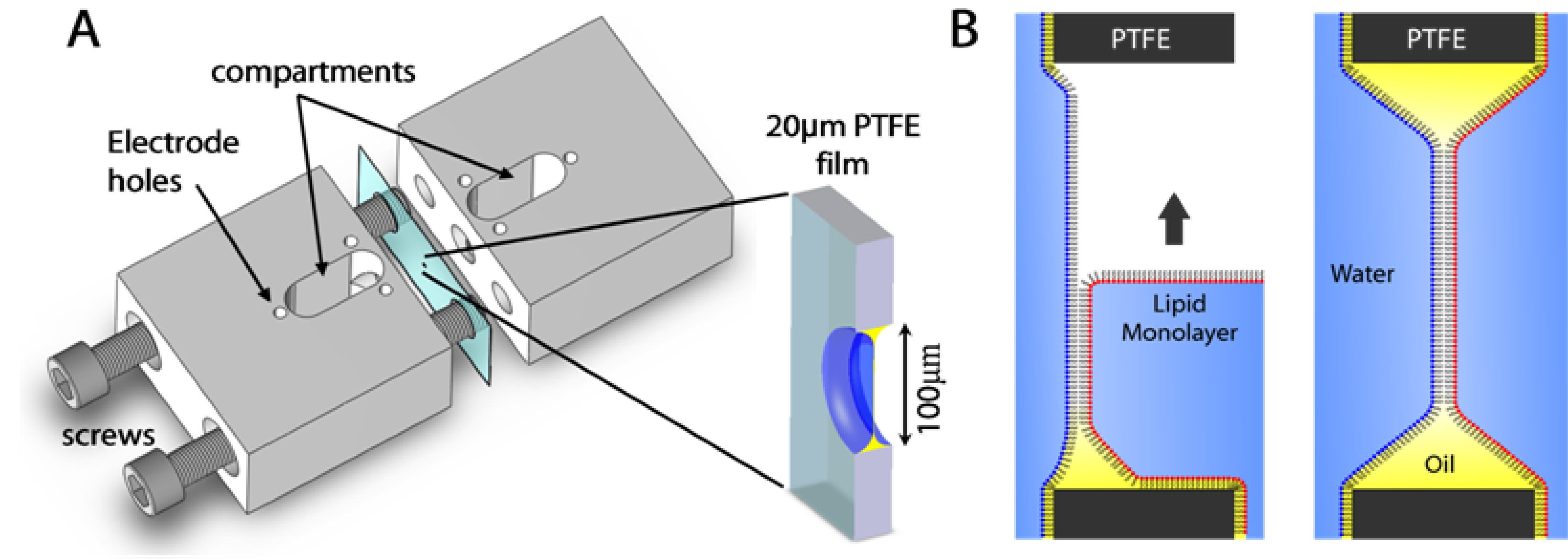
- 293 amino acids per monomer
- Water soluble Monomer
- Heptameric transmembrane pore
- Open at high ionic strength
- Open at high potential
- High current ($7x10^8$ ion / sec)
- Low level background noise

α-hemolysin



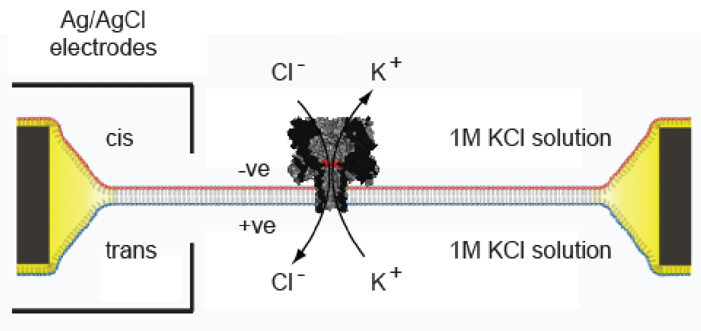
Electrical Detection

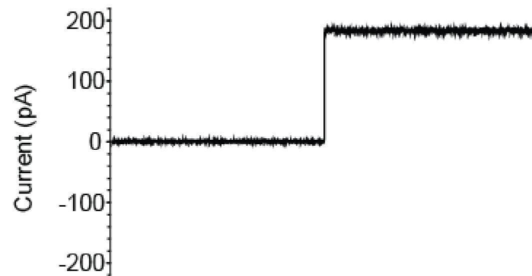
- Electrical detection is cheap and fast (>10 kHz) electrical detection
- Nanopores inserted into planar lipid bilayers (~100µm)

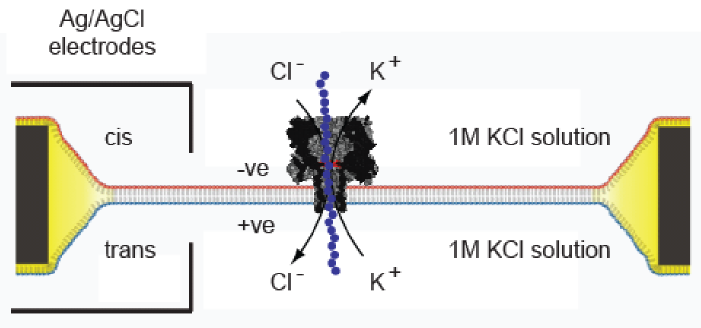


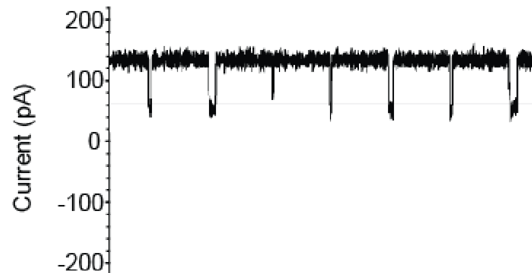
Maglia, G, Heron, A. J., Stoddart, D., Japrung, D., Bayley, H., **Methods in Enzymology.** (2010) in press

Electrical Detection







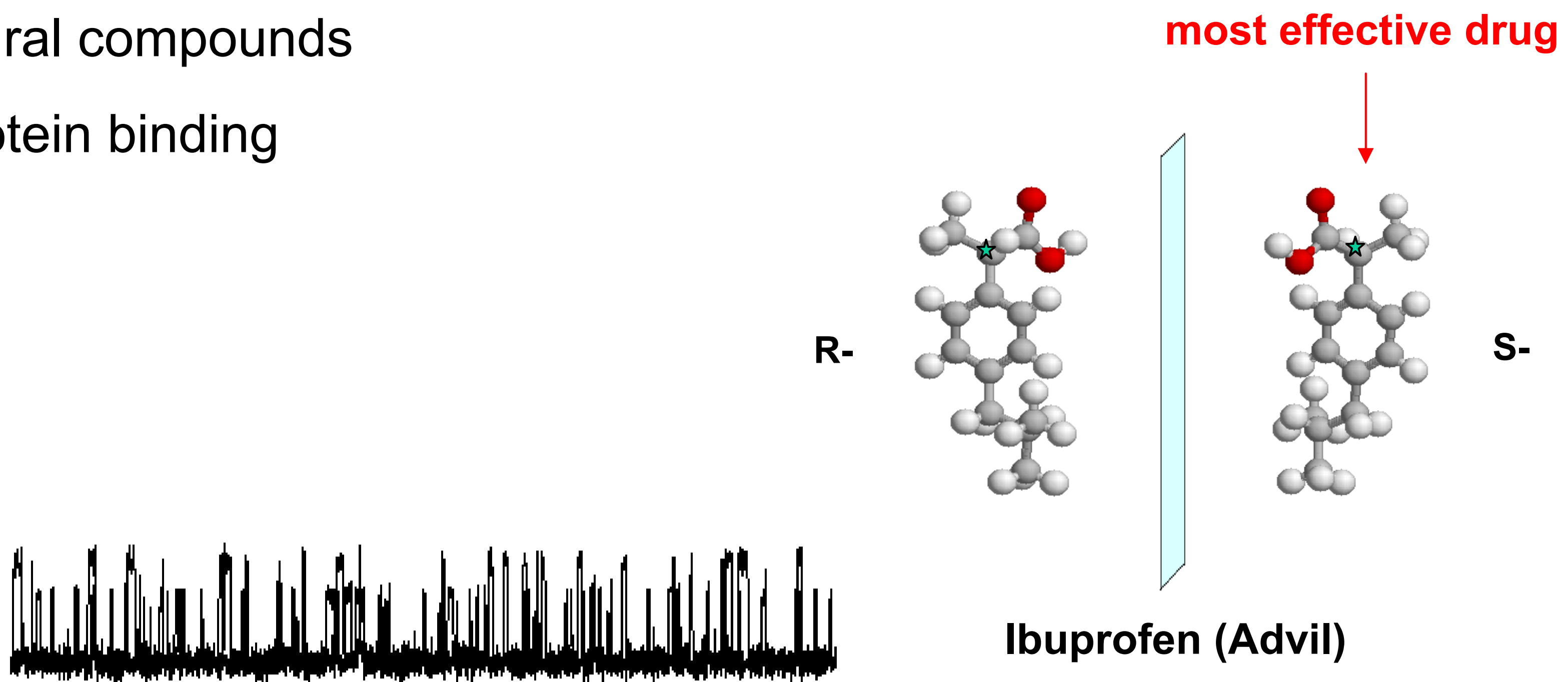


- perturbations of the current through a protein pore at a fixed applied potential tell us what is happening inside the pore
- we see individual interactions or reaction steps for single molecules

Exquisite Sensitivity

αHL has been used to detect:

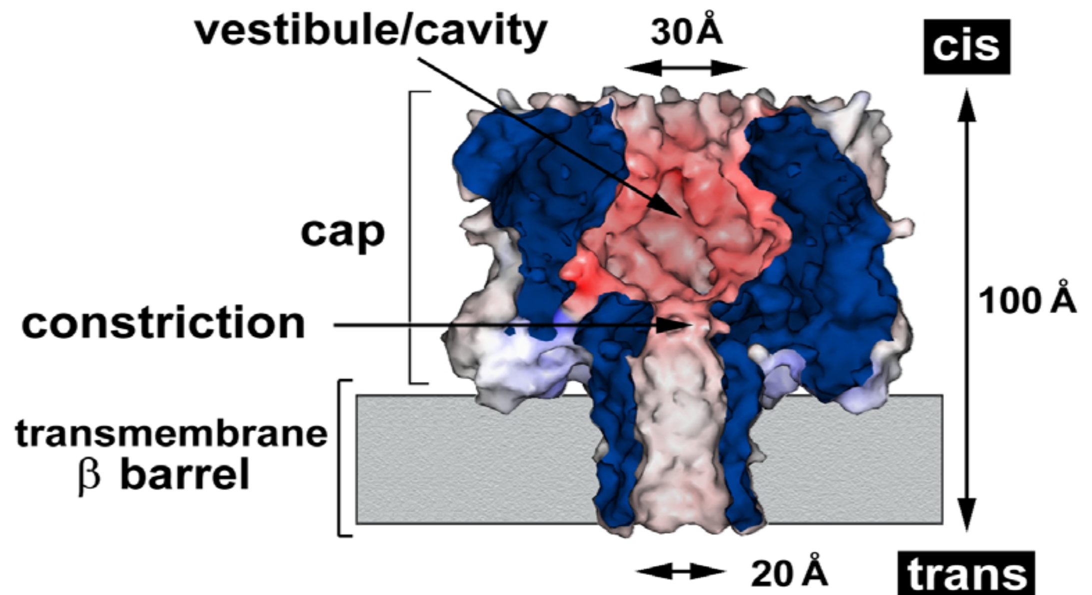
- a wide range of single molecules (ATP, IP_3, drugs, TNT)
- Reaction intermediates
- Chiral compounds
- Protein binding



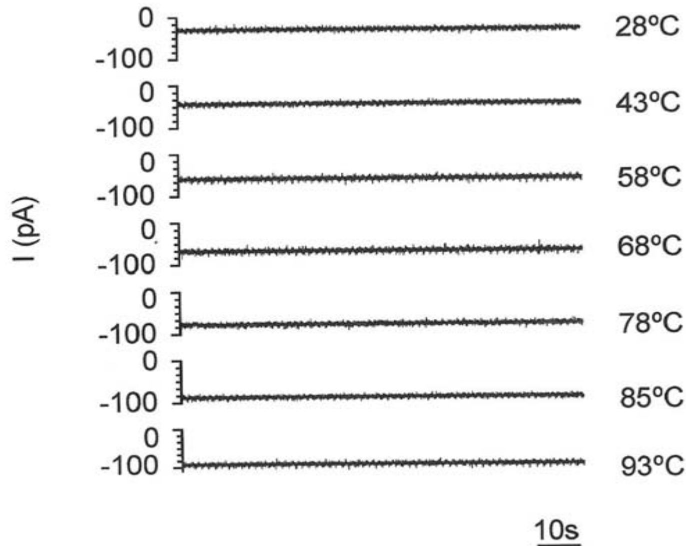
Kang, X.-f., et al. **J. Am. Chem. Soc.** 128, 10684-10685 (2006)

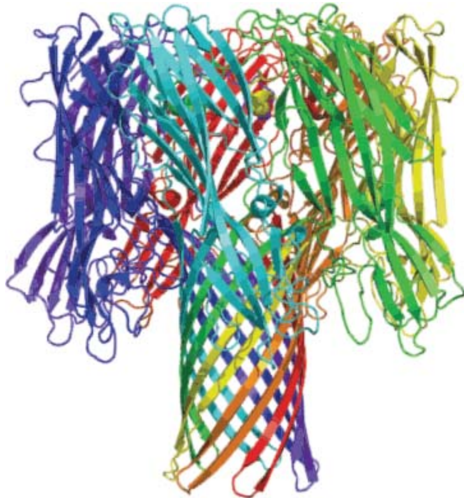
Protein engineering

- The WT pore is a “blank-state”: mutagenesis or chemical modification can be used to alter the structure and function
- Gives us atomic level control that is simply not possible to achieve using synthetic nanopores



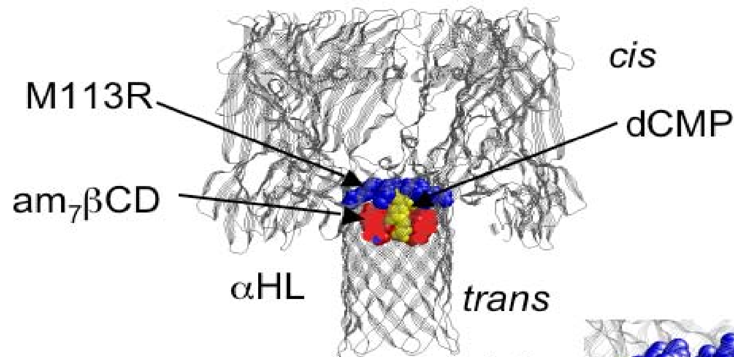
Protein Stability



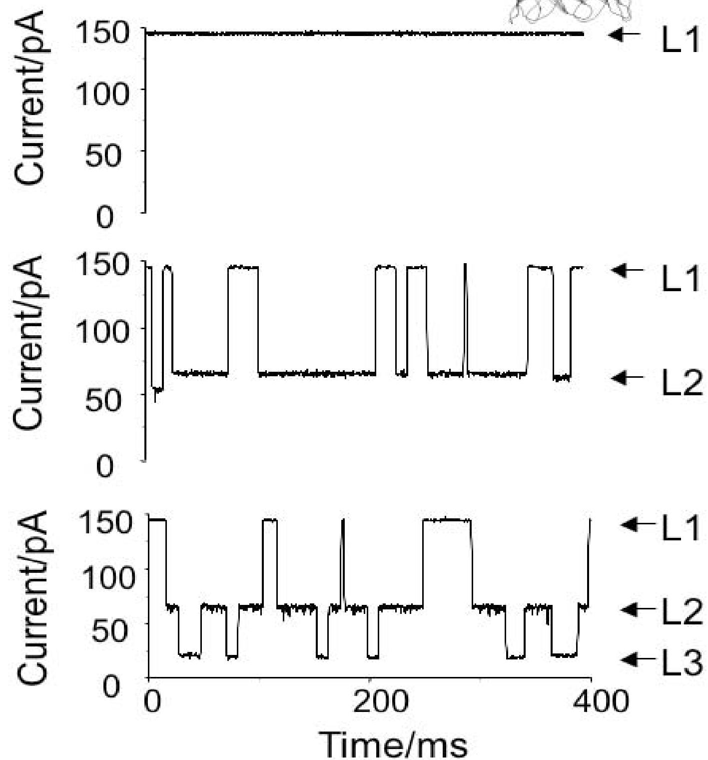


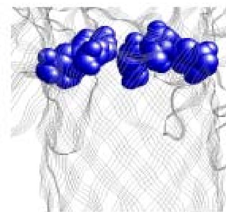
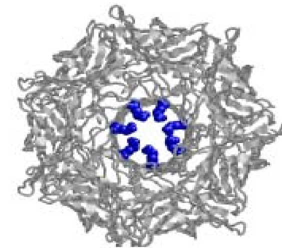
- It is a myth that protein pores are unstable
- **Temperature:** α-Hemolysin, leukocidin and OmpG are stable at high temperatures and show no sharp transitions in conductance
- **pH:** up to pH ~12.5
- **Urea:** up to 8M
- **Detergents:** SDS
- **Long term storage:** weeks

Base Sequencing

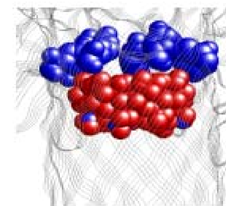
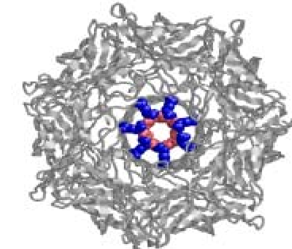


Astier, Y., Braha, O. and Bayley, H. **J. Am. Chem. Soc.** 128, 1705-1710 (2006).

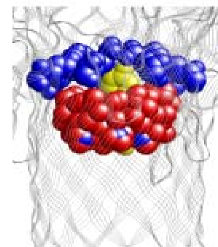
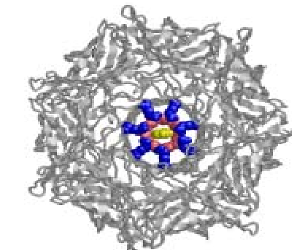


Empty pore

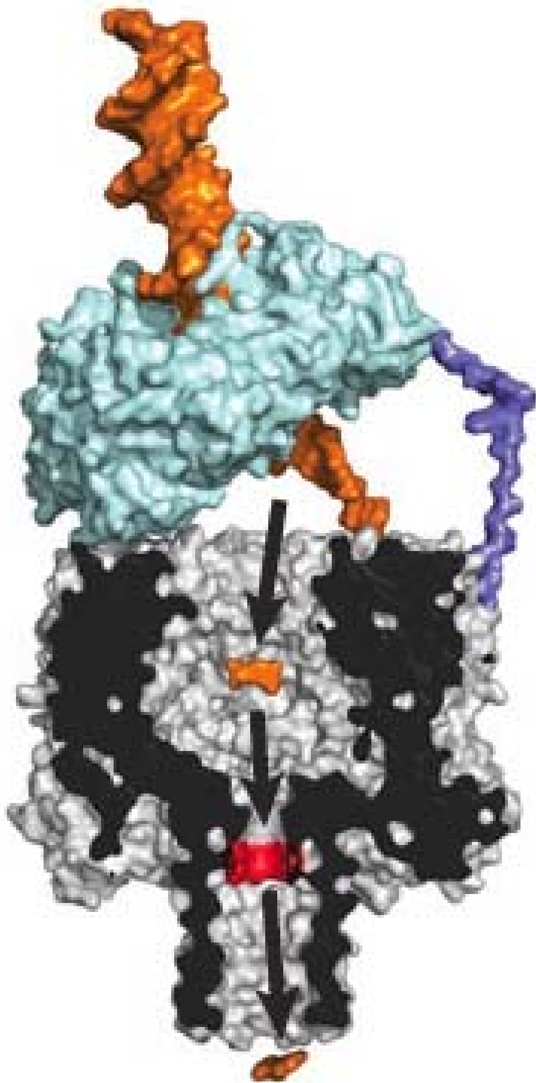



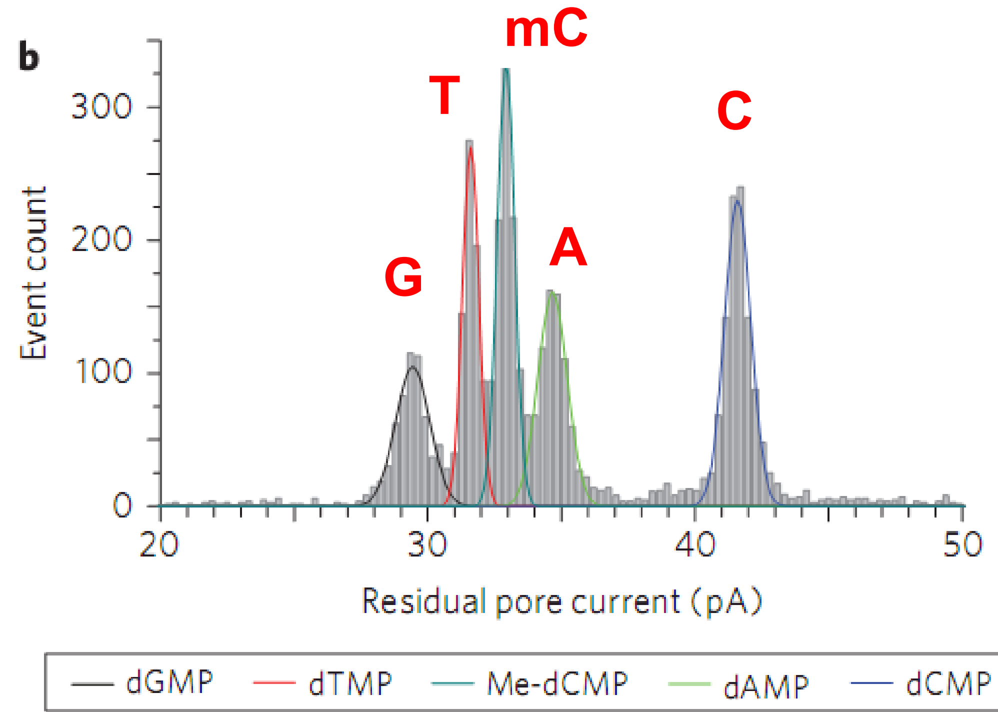
The cyclodextrin adapter binds

dCMP inserts in the adapter and is identified

Base Sequencing

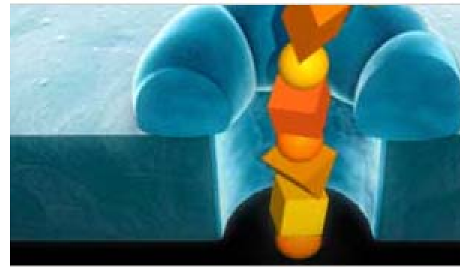




- H. Wu et al., J. Am. Chem. Soc. 129, 16142-16148 (2007)
- J. Clarke et al., Nature Nanotechnology 4, 265-270 (2009)

Oxford NANOPORE Technologies®

Oxford NANOPORE Technologies®



Nanopores.

Label-free, single molecule analysis

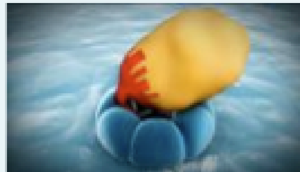
+ View Nanopore sequencing video

Application Specific

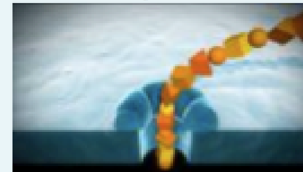
Adaptable protein nanopore:

DNA Sequencing

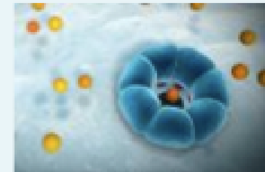
Proteins



Polymers

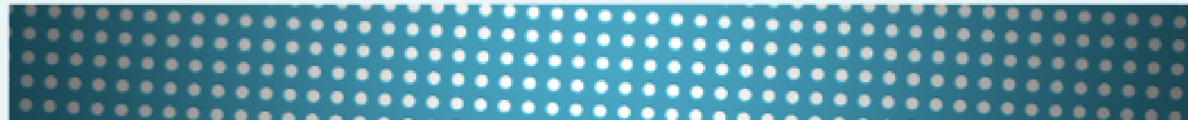


Small Molecules



Generic Platform

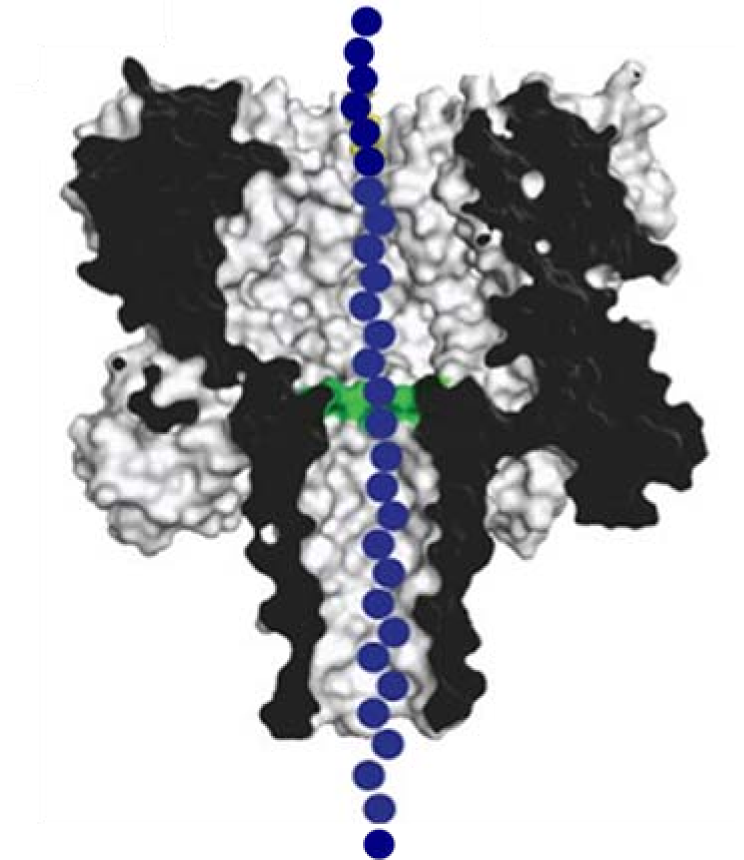
Sensor array chip: many nanopores in parallel



Electronic, real-time read-out system

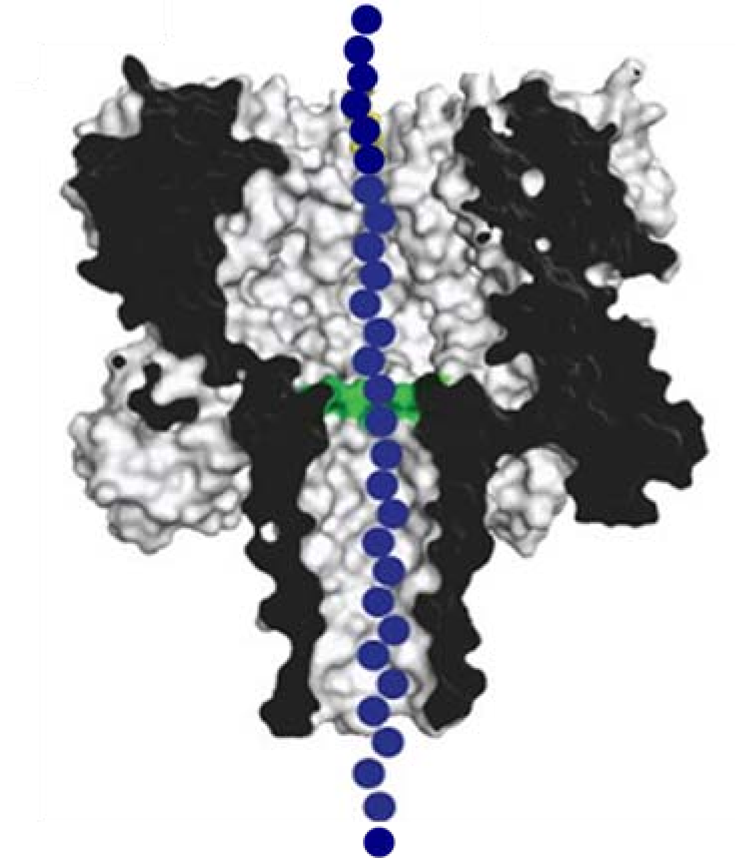
Strand Sequencing - Key Aims

- DNA binding and capture
- DNA strand translocation
- Base identification

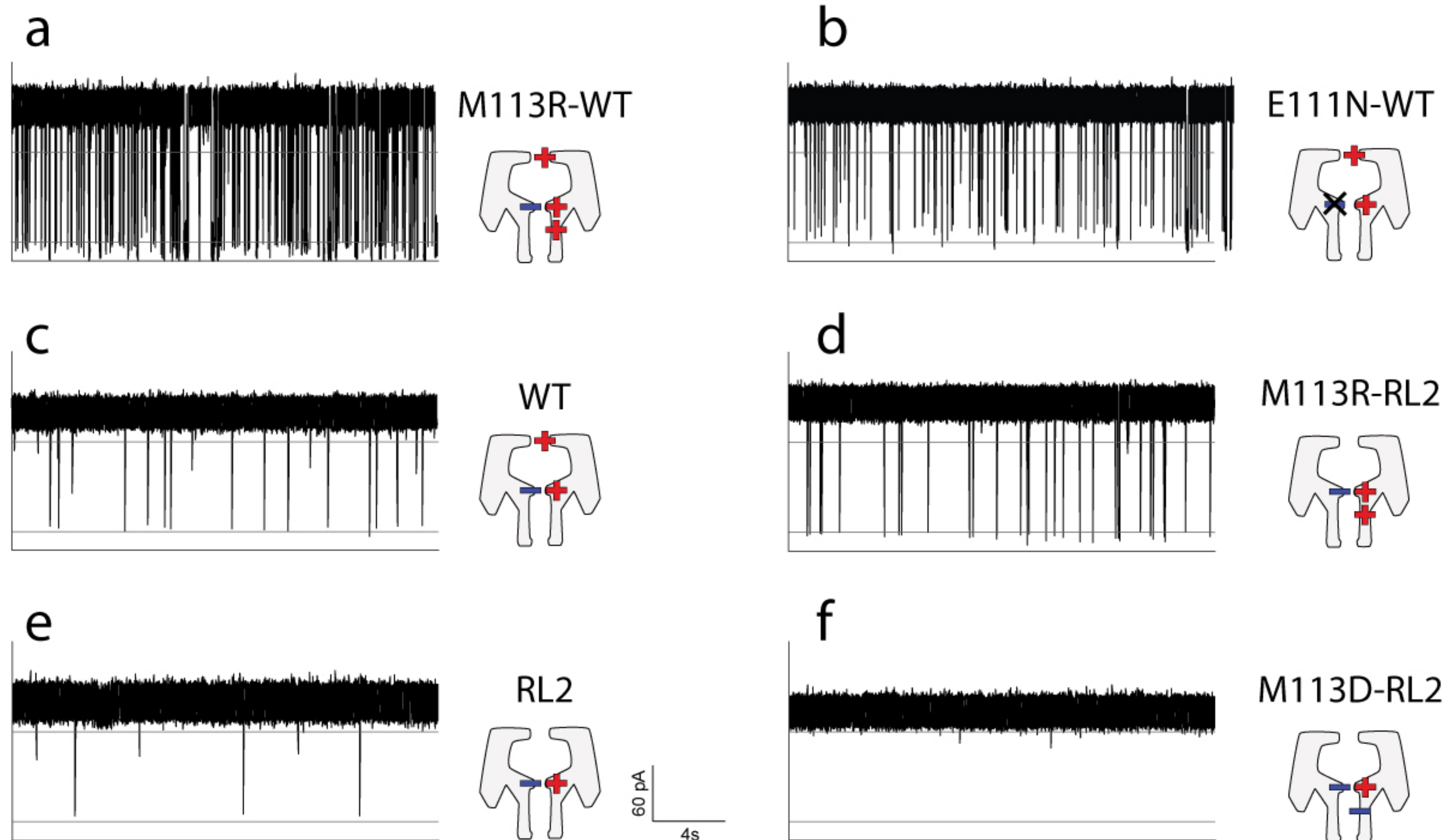


Strand Sequencing - Key Aims

- **DNA binding and capture**
- DNA strand translocation
- Base identification



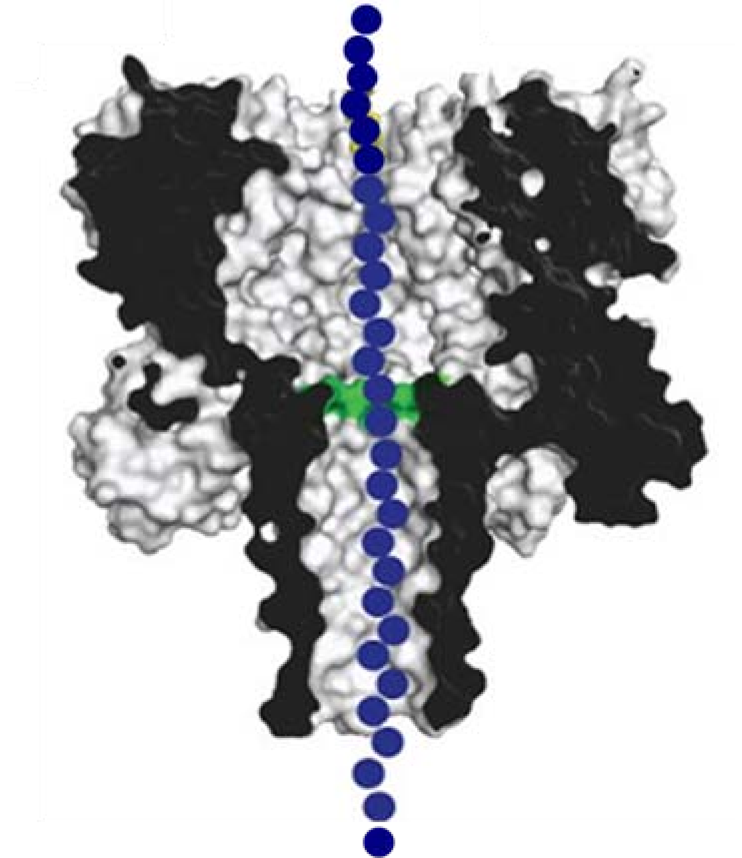
Improving DNA capture frequency



Maglia, G., Rincon Restrepo, M., Mikhailova, E. and Bayley, H. **Proc. Natl. Acad. Sci. USA** 105, 19720-19725 (2008)

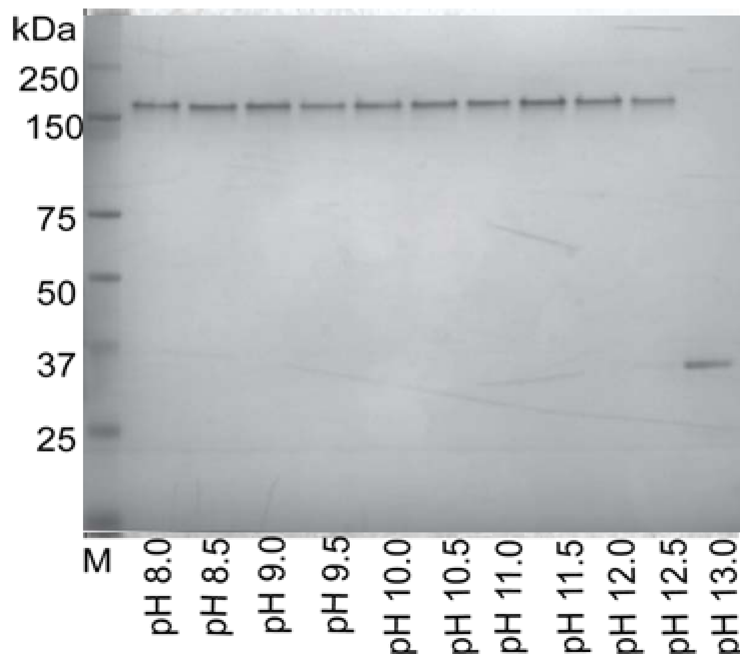
Strand Sequencing - Key Aims

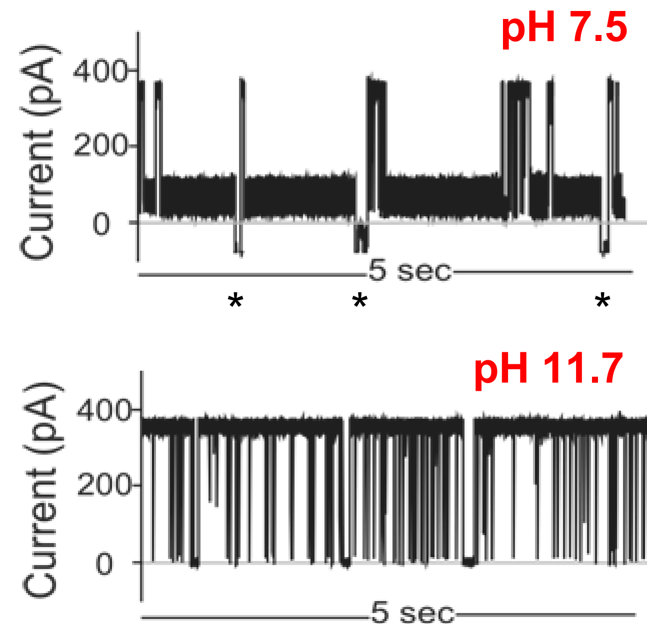
- DNA binding and capture
- **DNA strand translocation**
- Base identification



Threading ssDNA with secondary structure

- It would be preferable to work with double-stranded DNA
- dsDNA can be translocated through an engineered αHL at pH 11.7

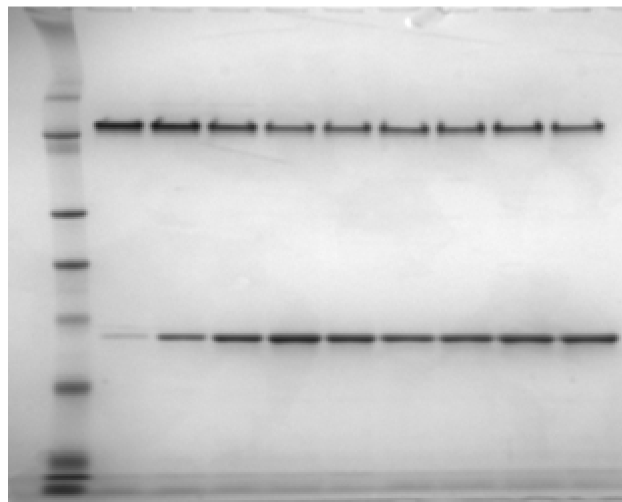


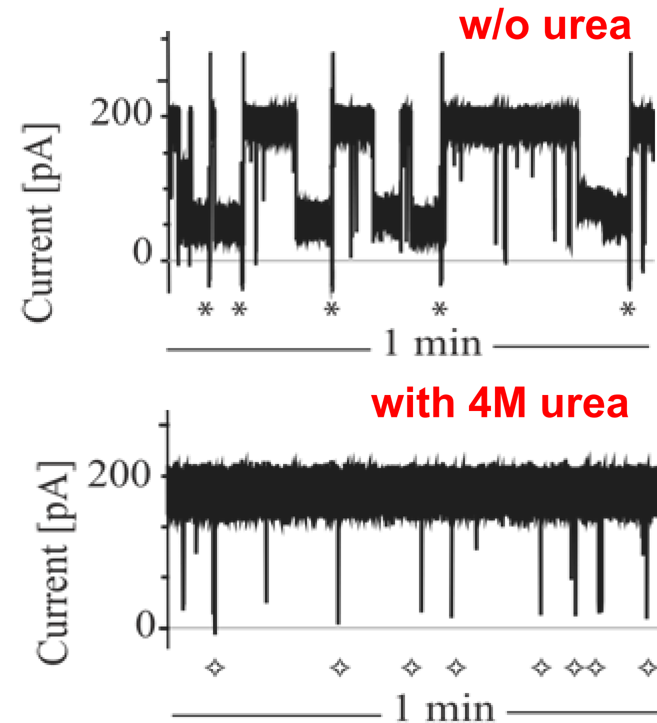


Maglia, M., Henricus, M., Wyss, R., Li, Q., Cheley, S. and Bayley, H. **Nano Letters** 9, 3831-3836 (2009).

Threading long RNA with secondary structure

- It is important to capture long and complex DNA strands
- Long ssRNA (up to 3Kb) can be translocated through an engineered αHL using 4M urea

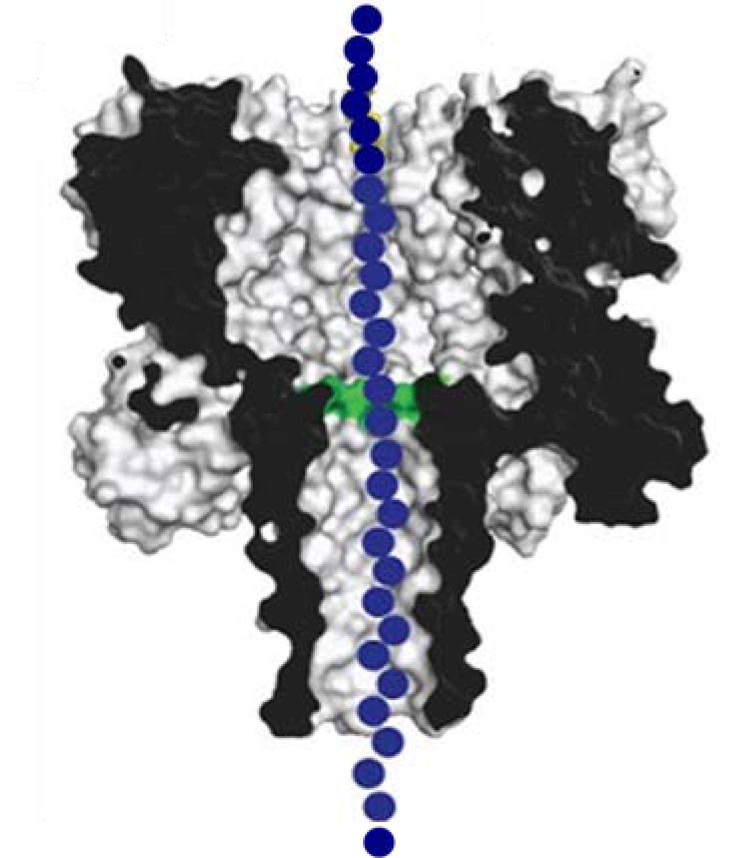




Japrung, D., Henricus, M., Li, Q., Maglia, G. and Bayley, H. Biophysical Journal, in press (2010).

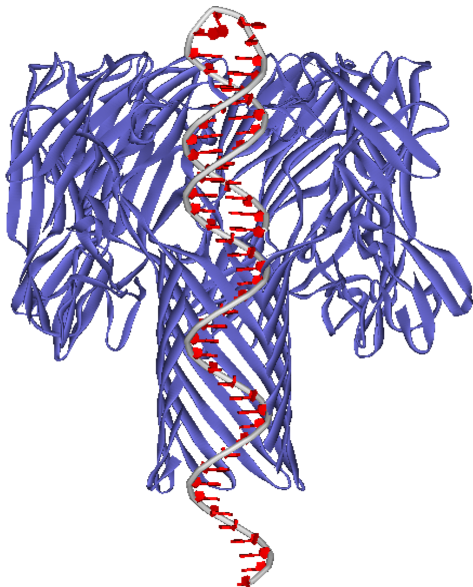
Strand Sequencing - Key Aims

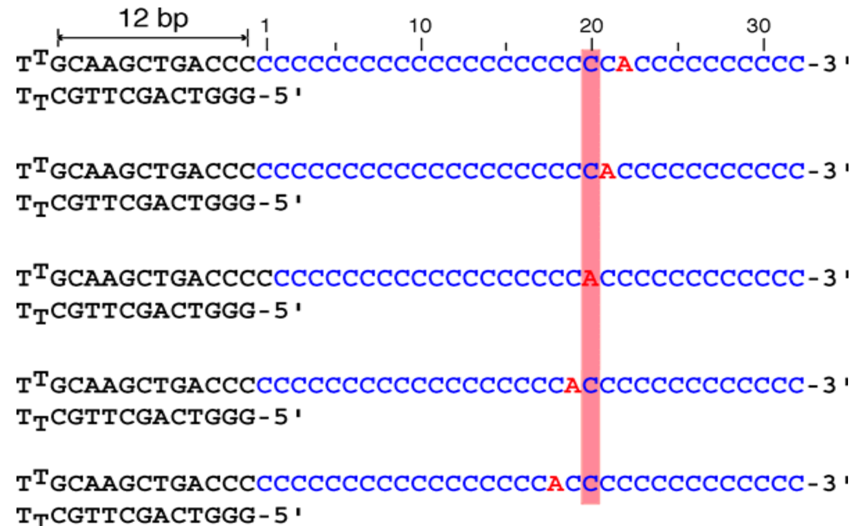
- DNA binding and capture
- DNA strand translocation
- **Base identification**



Recognition of DNA bases in strands

- DNA needs to be immobilized (e.g. dsDNA Hairpin)
    - Freely translocating DNA moves too fast to discriminate
- Recognition found at the exit of the pore
    - Recognition poor: considered to be due to hairpin interactions at the constriction

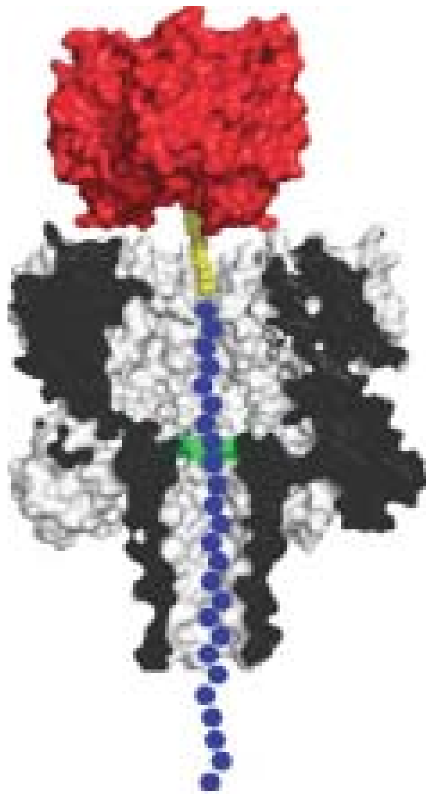


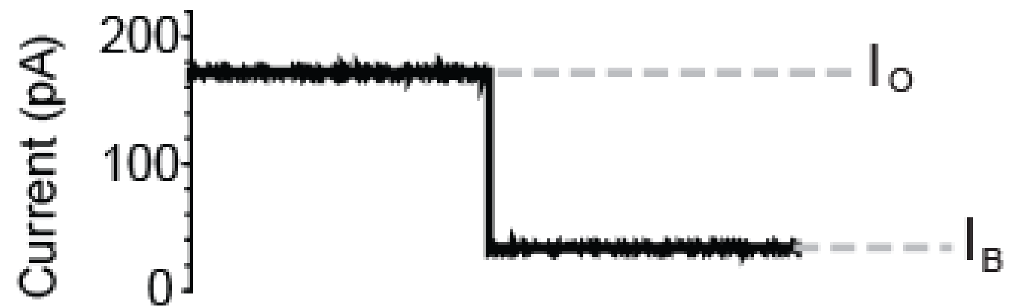


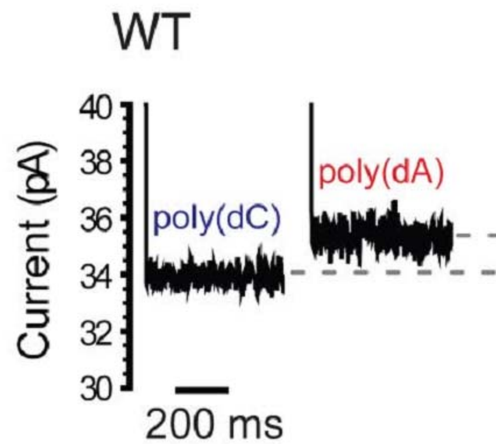
Ashkenasy, N., Sanchez-Quesada, J., Bayley, H. and Ghadiri, M.R. **Angew. Chem. Int. Ed.** 44, 1401-1404 (2005)

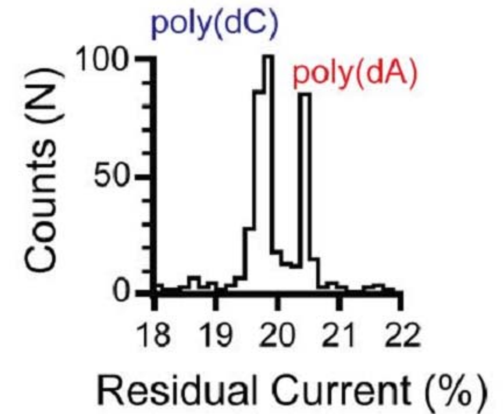
Recognition of DNA bases in strands

- Streptavidin immobilized homopolymers can be resolved in αHL





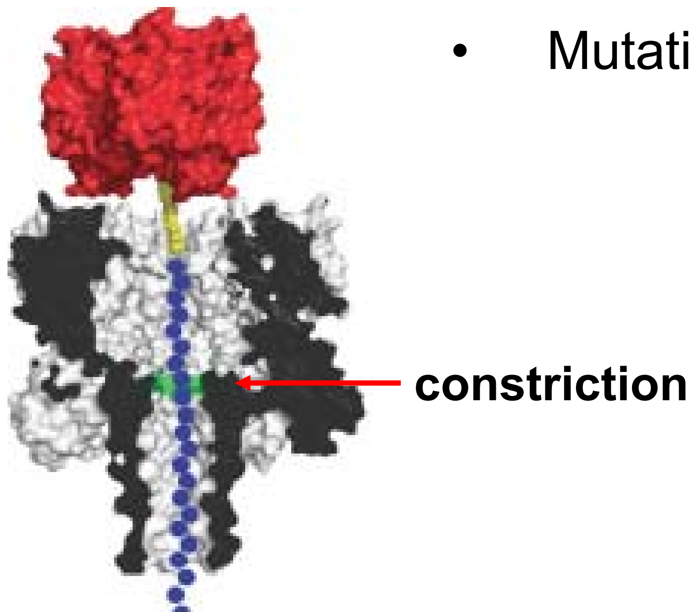


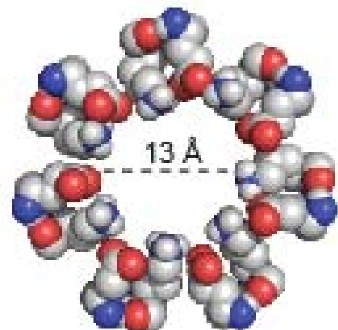


Stoddart, D., Heron, A., Mikhailova, E., Maglia, G. & Bayley, H. **Proc. Natl. Acad. Sci. USA** 106, 7702-7707 (2009).

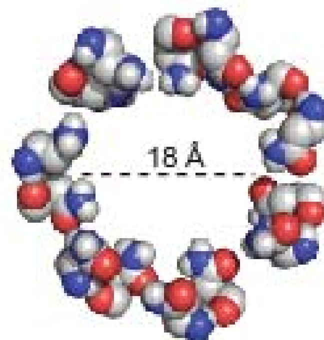
Mutagenesis affects recognition

- Mutation at the constriction alters DNA recognition



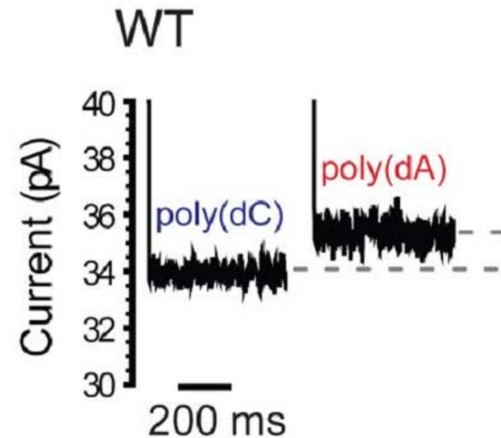


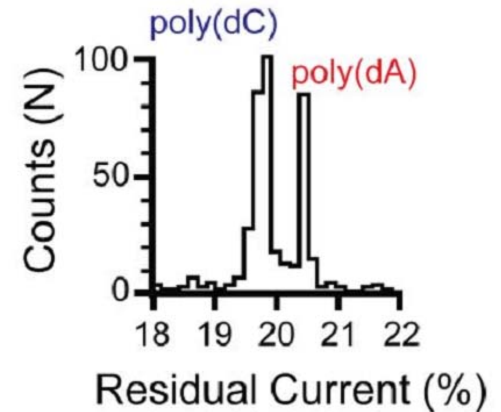
WT



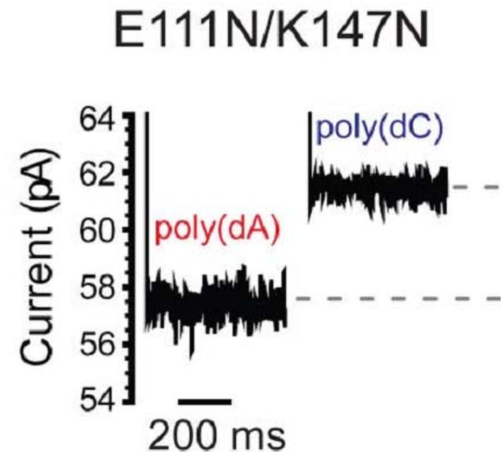
E111N/K147N

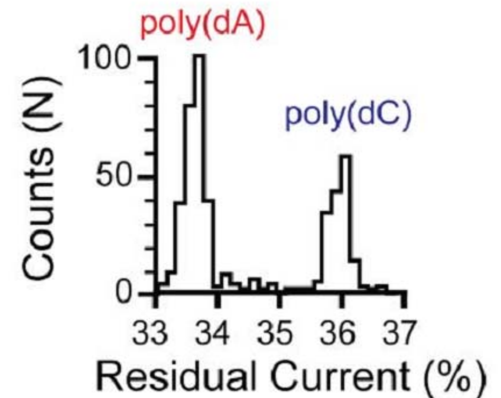
WT





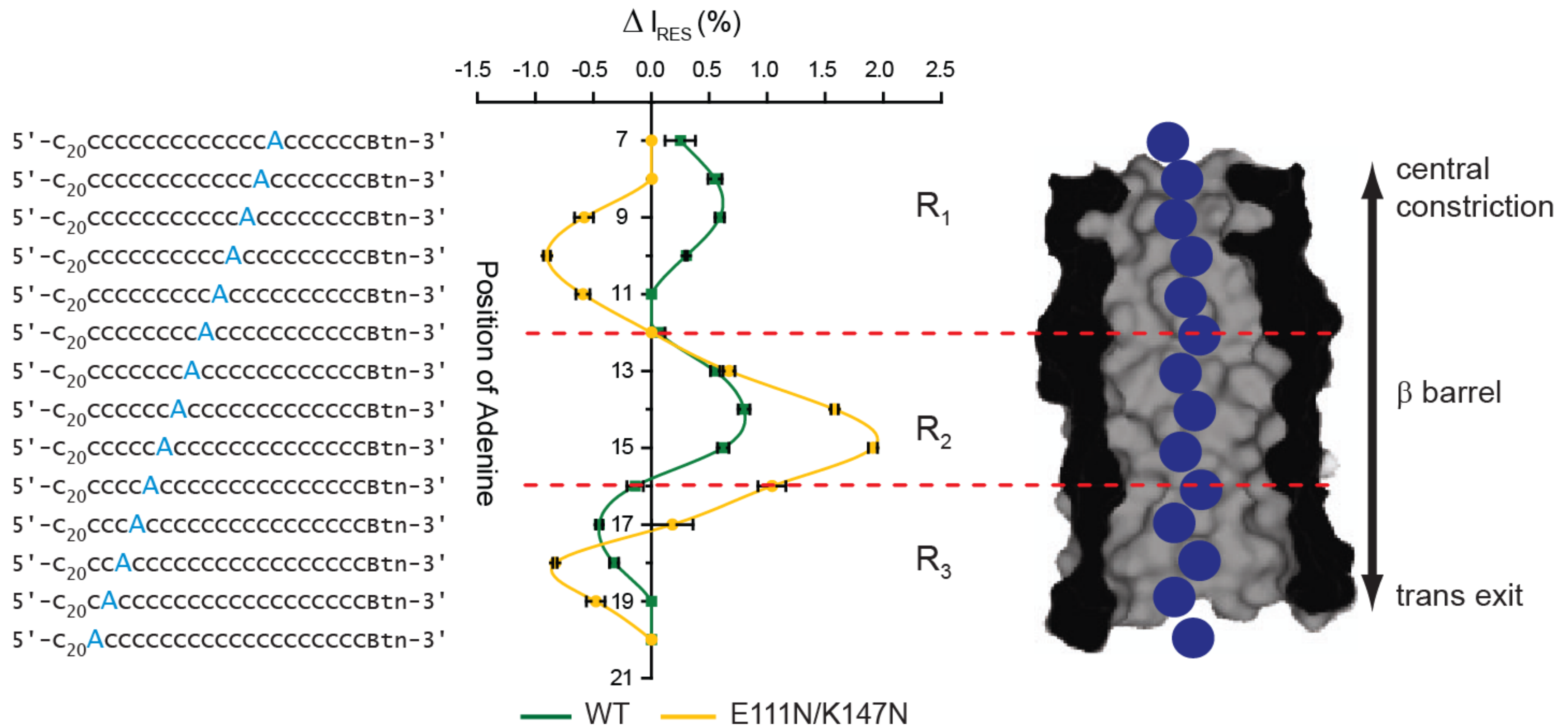
E111N/K147N





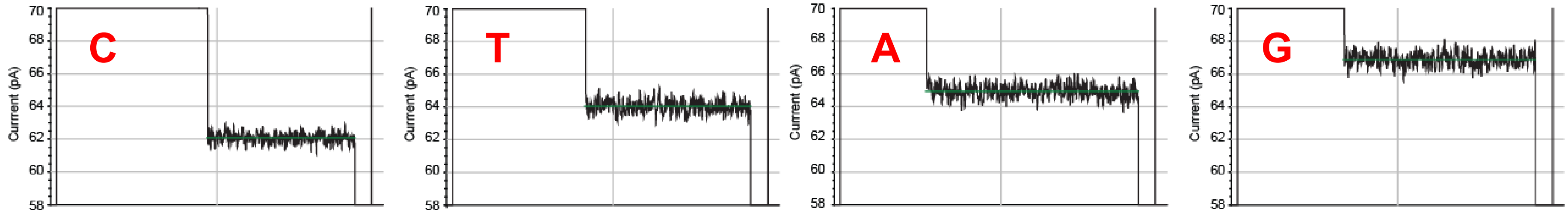
Recognition of DNA bases in strands

- There are **three recognition elements** in the β–barrel of α-HL

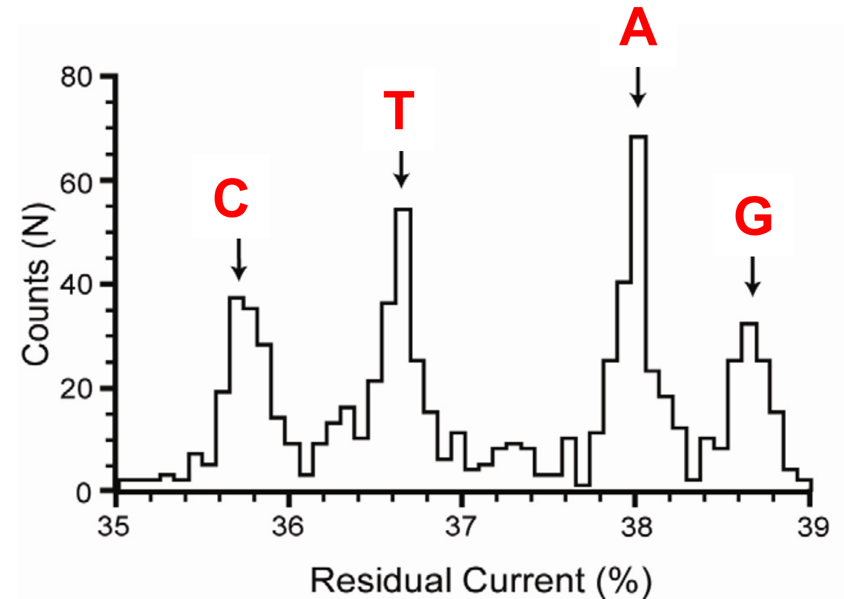


Stoddart, D., Heron, A., Mikhailova, E., Maglia, G. & Bayley, H. **Proc. Natl. Acad. Sci. USA** 106, 7702-7707 (2009).

Recognition of DNA bases in strands



- **All 4 DNA bases are easily resolved**
- Single DNA bases can be resolved in both homopolymeric and 'natural' heteropolymeric DNA backgrounds
- Base recognition is robust under a wide variety of conditions



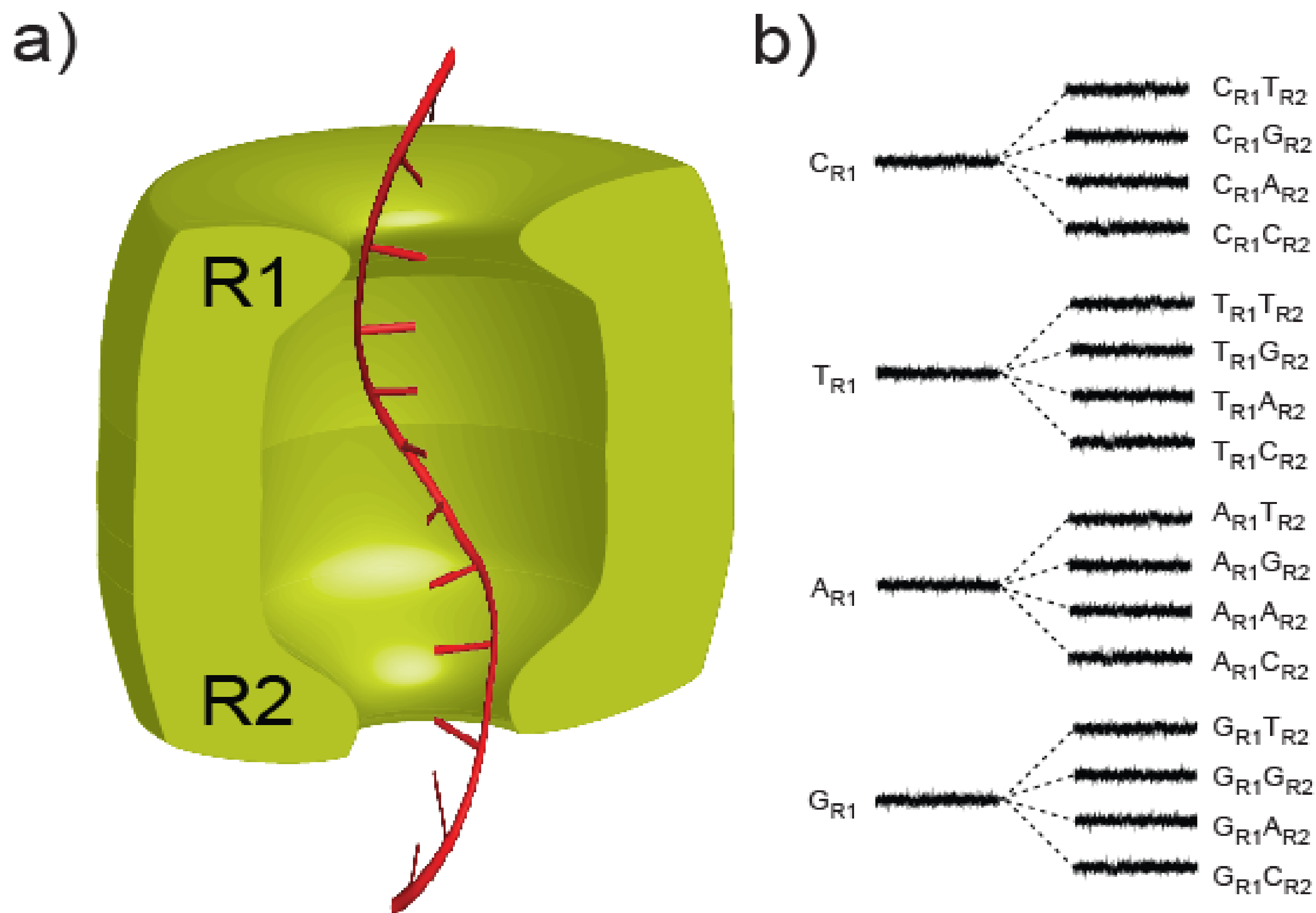
5'-ACTACCTAGTTTACGTAATCCATCTGCACAATGCAGCATTBtn-3'

5'-ACTACCTAGTTTACGTAATCCATCTGTACAATGCAGCATTBtn-3'

5'-ACTACCTAGTTTACGTAATCCATCTGAACAATGCAGCATTBtn-3'

5'-ACTACCTAGTTTACGTAATCCATCTGGACAATGCAGCATTBtn-3'

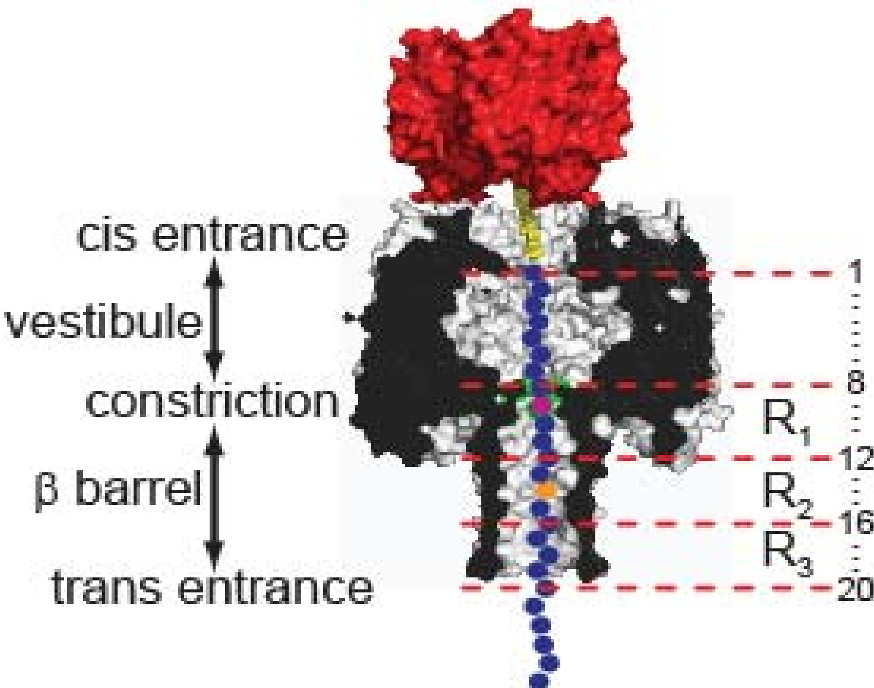
Are two heads better than one?

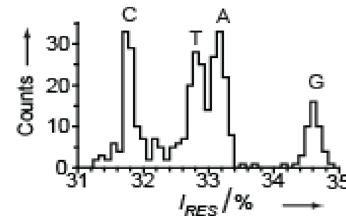


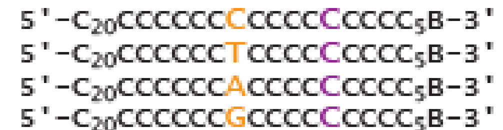
Stoddart, D., Maglia, G., Mikhailova, E., Heron, A. and Bayley, H. **Angew. Chem. Int. Ed.** 49, 556-559 (2010)

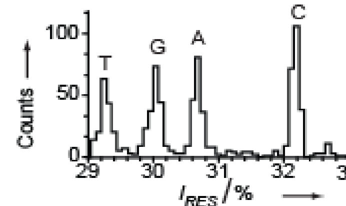
Are two heads better than one?

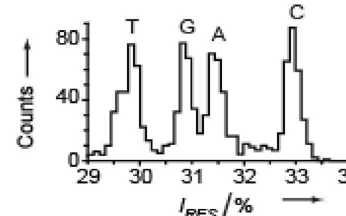
- E111N/K147N/M113Y pore has 2 strong recognition points (R_1 and R_2)
- Each capable of 4-base discrimination- little "crosstalk"

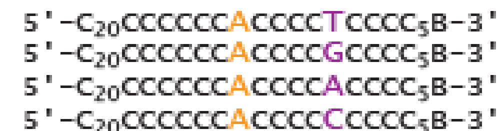


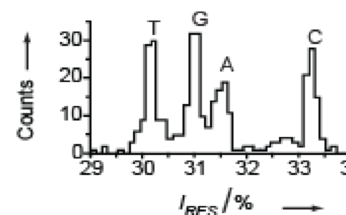


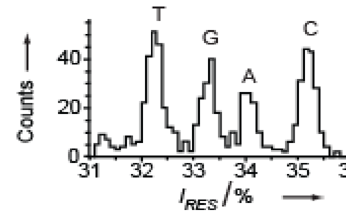






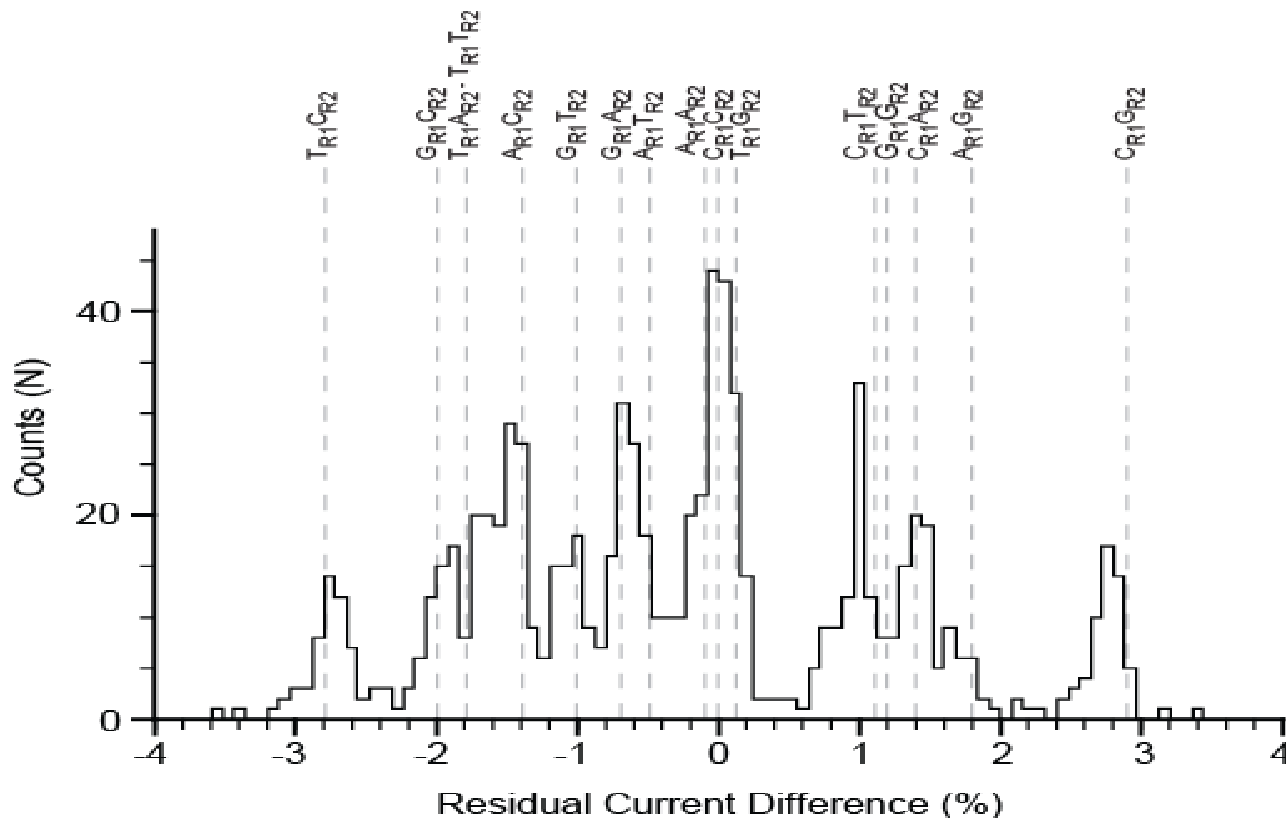






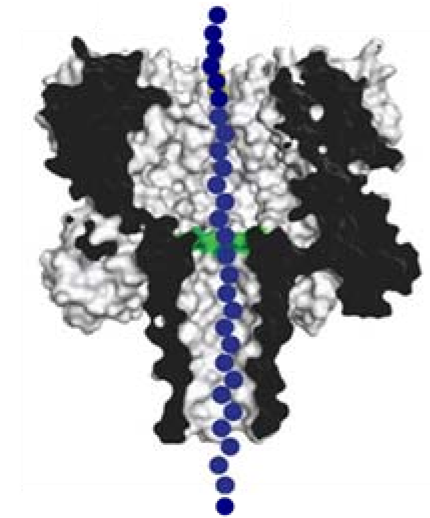
Are two heads better than one?

- 11 levels resolved, not 16, due to degeneracy
- More information obtained (but perhaps a bioinformatic nightmare)
- Base pairs to be read twice - reduce errors

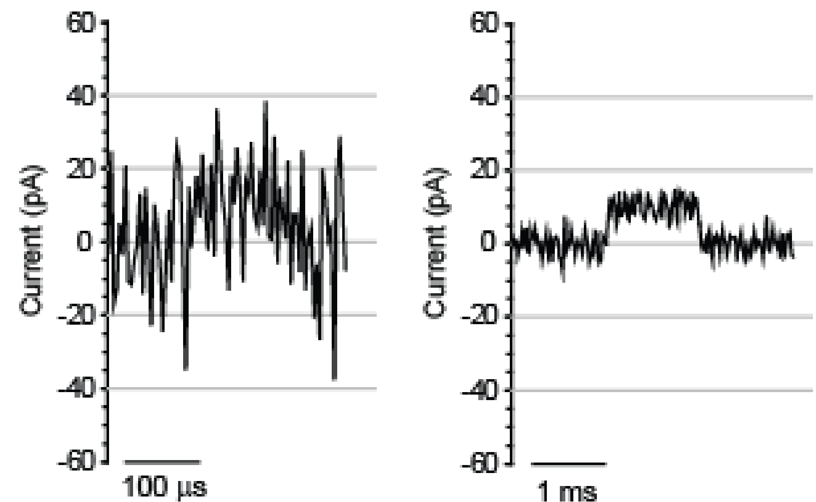


So where do we stand?

- DNA binding and capture √
- DNA strand translocation √
- Base identification √
  - Mutagenesis, unnatural amino-acids, different pores

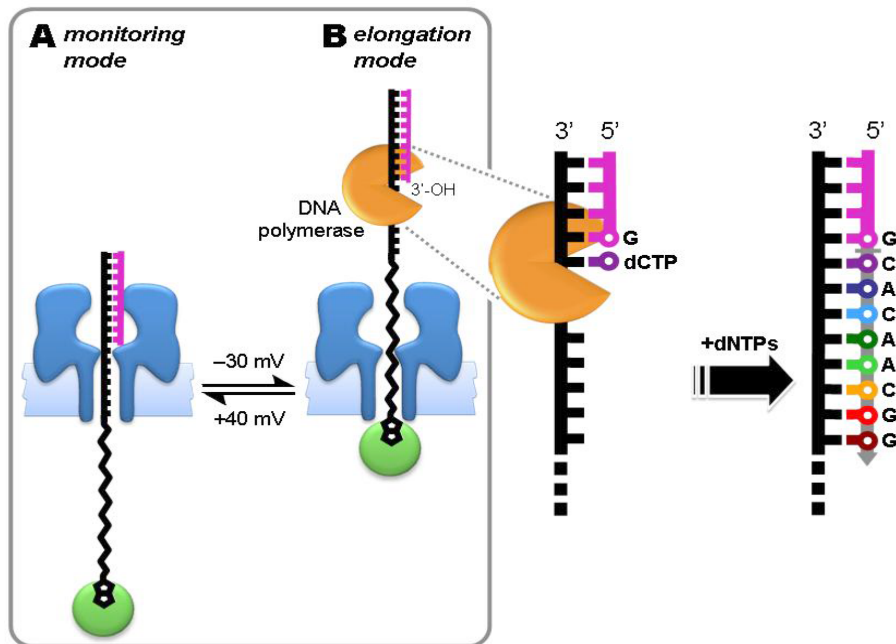


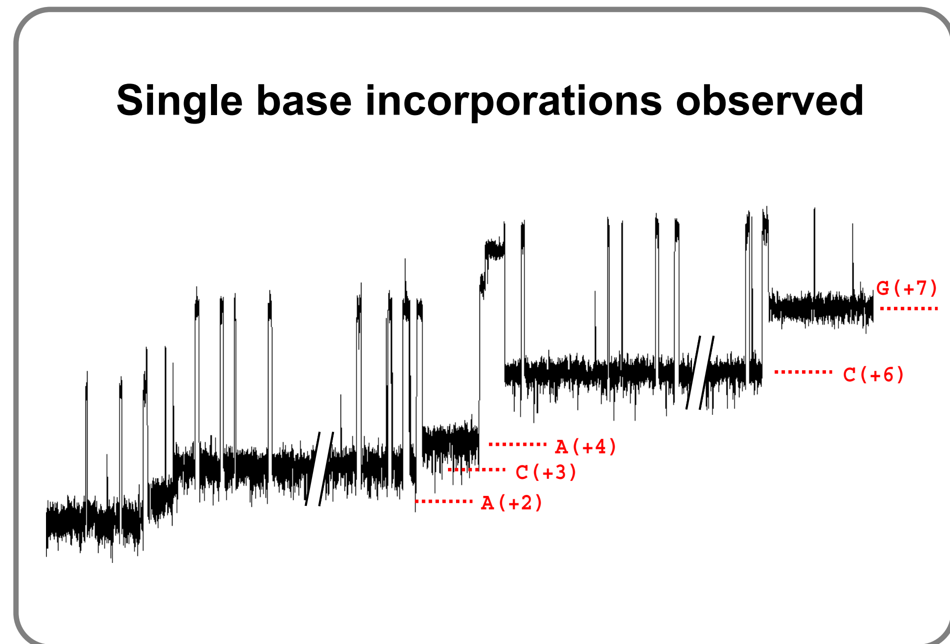
- **Towards a full platform:**
  - Bilayer arrays
  - Slow down the DNA



Enzymes to process DNA

- Enzymes (e.g. polymerases, exonucleases) proven to move DNA through αHL
- Key challenges: processivity, salt tolerance, high force binding

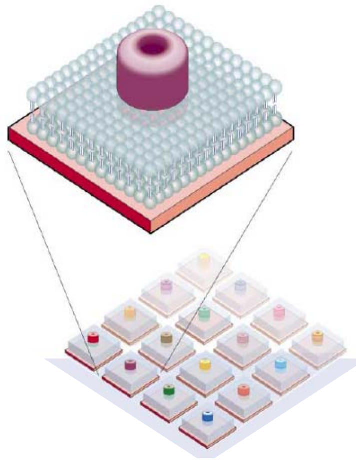




Cockroft SL, Chu J, Amorin M, Ghadiri MR: ***J Am Chem Soc*** 2008, **130**:818-820.

Towards Bilayer Arrays

Bilayers on solid-supports

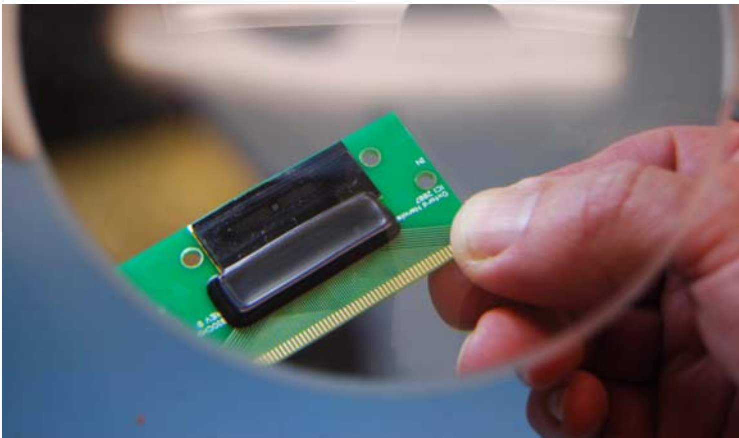


Bayley, H. and Cremer, P.S., **Nature** 413, 226-230 (2001)

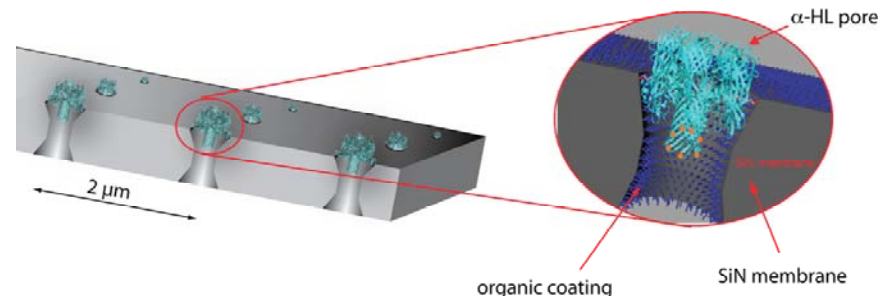
Speed is key:

- We need a 10^6 array for the 10 minute genome
- Progression of sequencing technology back into the labs and at clinical interfaces

ONT 128-Bilayer Chip (2007)

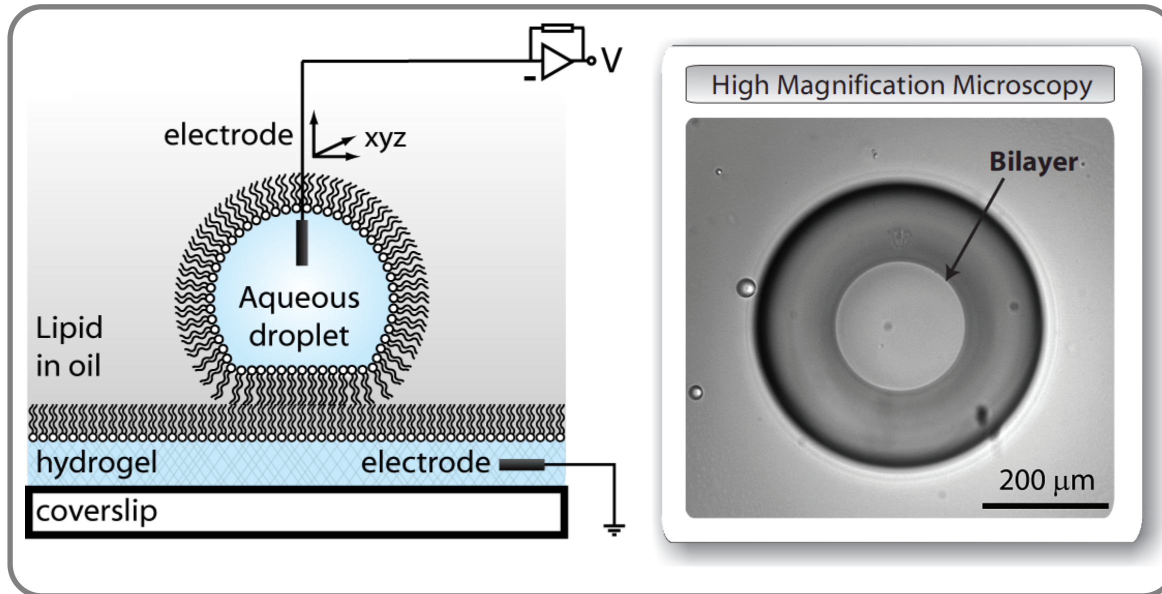


Biological pores in solid-state membranes

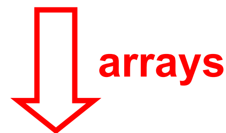


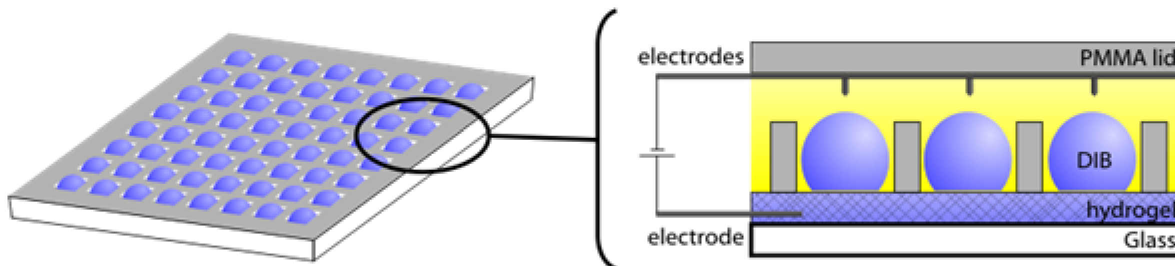
Soni G. V., Meller A., **Clin Chem**, 53, 1996-2001 (2007)

Droplet interface bilayers



- Water-in-oil droplets
- Nanolitre volumes
- Easily arrayable
- Complementary to existing technologies (e.g. Raindance)
- Ideal for protein pores





Heron, A. *et al*.. **J. Am. Chem. Soc.**, 129, 16042-16047 (2007)

Bayley, H. *et al.*, **Molecular Biosystems**, 4, 1191-1208 (2008)

Acknowledgements

Prof. Hagan Bayley

DNA sequencing Team:

- Andrew Heron
- Giovanni Maglia
- David Stoddart
- Deanpen Japrung
- Ellina Mikhailova
- Marcela Rincon-Restrepo

- HB group

Funding: