Introduction to Nanopore DNA sequencing

Prof. Dr. Abdul Haque

PhD (Osaka Univ. Japan)

Post. Doc. (Osaka Univ. Japan; Imperial College, UK)

Rector/Director

Akhuwat FIRST University

Faisalabad, Pakistan

- Ex- Deputy Chief Scientist, NIBGE, PAEC, Pakistan
- Ex- Prof. of Pathology/Director Advanced Studies and Research, The University of Faisalabad, Faisalabad

Note:

This lecture is meant only to provide basic concepts of Nanopore DNA sequencing, a revolutionary technique. It is assumed that the reader is already well versed with other types of DNA sequencing.

- There are many DNA sequencing techniques developed during last 50 years.

- In all of these techniques:

  the genome is broken into many tiny pieces;

  followed by *creation of vast copies of those pieces;*

  sequencing them;

  and then *putting the resulting data back together* with computers.

- **Nanopore DNA sequencing**, however, *threads one DNA strand through a really tiny nanopore* and then *detects very small changes in electric current* to determine the different nucleotides of DNA.

- As the DNA moves through the pore, it *creates signals that can be converted to read each base.*

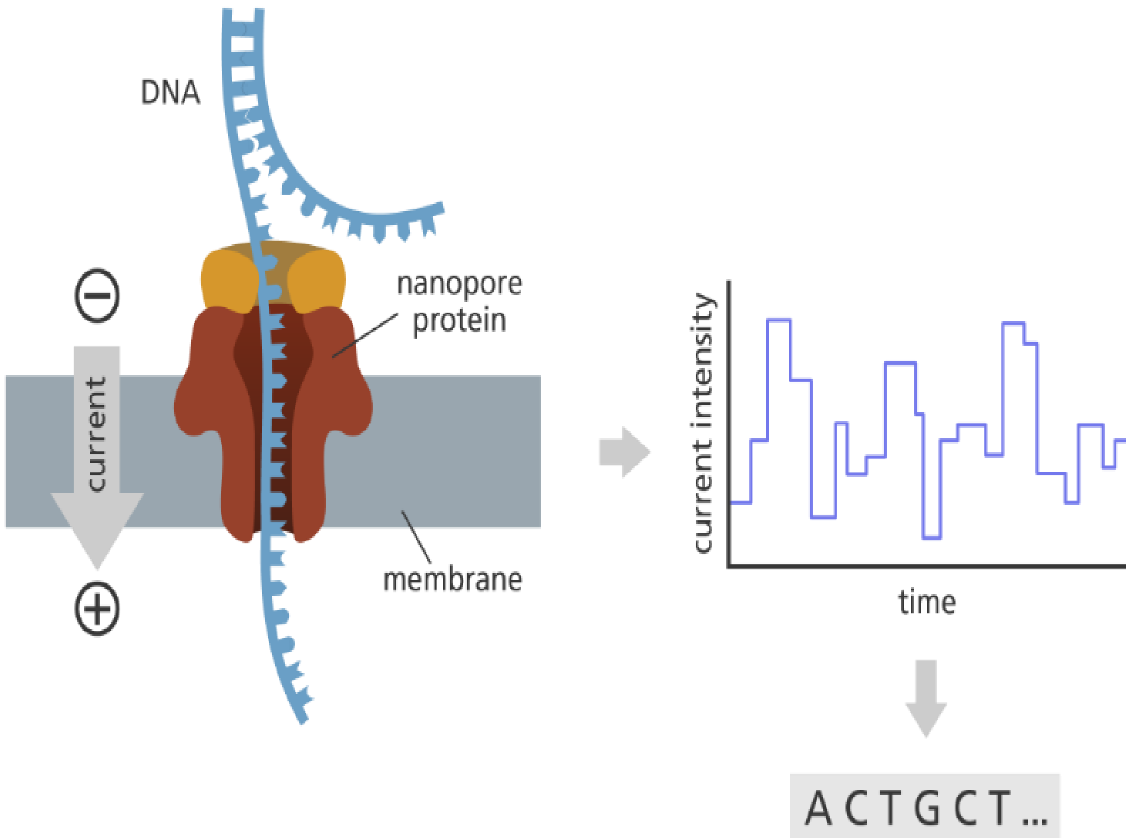
- The results are obtained in *real time*.

- A **major advantage** of nanopore DNA sequencing is the *ability to produce ultra-long reads* — much, much longer than that produced by earlier techniques.

- It also *detects the nucleotides from a single DNA molecule* rather than many.

- This approach offers a *low-cost, rapid process* for studying long stretches of DNA.

- These advantages have helped scientists work out complex regions of the human genome.



Nanopore proteins are embedded into an artificial membrane inside the sequencing flow cell. A current is applied across the membrane forcing the negatively charged DNA strands to move through the nanopores. The obstruction of a nanopore by a DNA fragment leads to a change in the current that is measured continuously by an electronics chip integrated within the flow cell. A base-calling algorithm converts the current variation back into the original DNA sequence. Image credit: Genome Research Limited.

- Using nanopore sequencing, a *single molecule of DNA or RNA* can be sequenced *without the need for PCR amplification or chemical labeling of the sample*.

- Nanopore sequencing has the potential to offer relatively low-cost genotyping, high mobility for testing, and rapid processing of samples with the ability to display results in real-time.

This method has already been used for:

1. Rapid identification of viral pathogens e.g., Ebola,
2. Environmental monitoring,
3. Food safety monitoring
4. Human genome sequencing
5. Plant genome sequencing
6. Monitoring of antibiotic resistance
7. Haplotyping

- It has been used in the COVID-19 pandemic.
- A quarter of all the world's SARS-CoV-2 viral genomes have now been sequenced with nanopore devices.

History

- Nanopore sequencing took 25 years to fully materialize.

- One of the first people to put forward the idea for nanopore sequencing was David Deamer, University of California, USA.

- In 1999, Deamer and his colleagues published the first paper using the term 'nanopore sequencing' and two years later produced an image capturing a hairpin of DNA passing through a nanopore in real time.

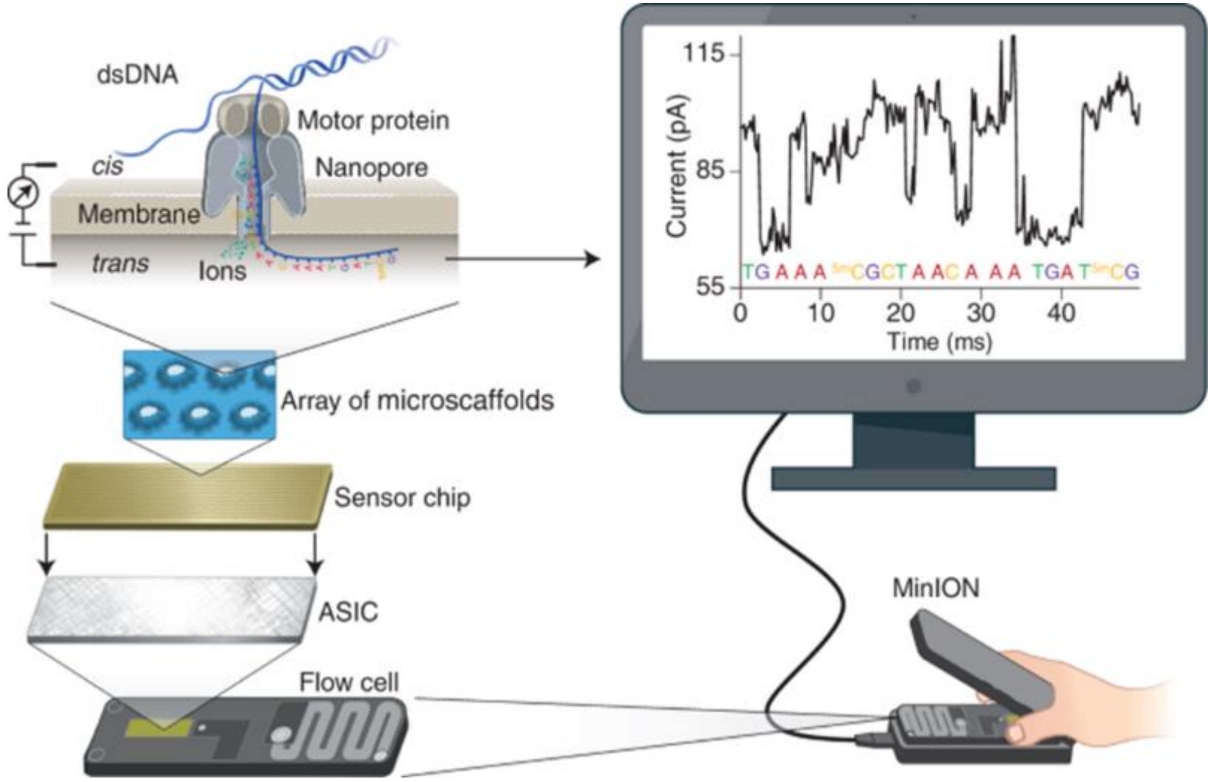
- Another important name is Hagan Bayley, Professor of Chemical Biology, Oxford University, UK.

- In 1990s, he independently developed ***stochastic sensing***, a technique that measures the change in an ionic current passing through a nanopore to determine the concentration and identity of a substance.

- By 2005, Bayley had made substantial progress with the method to sequence DNA and co-founded Oxford Nanopore to help push the technology further.

- In 2014 the company released its first portable nanopore sequencing device **MinION**.

- This made it possible for DNA sequencing to be carried out almost anywhere, even in remote areas with limited resources.



Principles for detection

- The membrane used is *either biological or solid-state* having nanopores.
- The region where the nanopore is found, is surrounded by electrolyte solution.
- At suitably high concentrations, the electrolyte solution is well distributed and all the voltage drop *concentrates* near and inside the nanopore.

- This means charged particles in the solution only feel a force from the electric field when they are near the pore region. This region is often referred as the ***capture region***.

- Inside the capture region, ions have a *directed motion* that can be recorded as a steady ionic current by placing electrodes near the membrane.

- Once inside the nanopore, the DNA molecule *translocate through* via a combination of electro-phoretic, electro-osmotic and sometimes thermo-phoretic forces.

- Inside the pore, the molecule occupies a volume that partially restricts the flow of ions, observed as an ***ionic current drop***.

- Based on various factors such as geometry, size and chemical composition, the *change in magnitude of the ionic current* and the *duration of the translocation* will vary.

- Different molecules can then be sensed and potentially identified based on this modulation in ionic current.

Base identification

- The magnitude of the electric current density across a nanopore surface depends on the *nanopore's dimensions* and the *composition of DNA or RNA* that is occupying the nanopore.

- Sequencing was made possible because, passing through the channel of the nanopore, the samples cause *characteristic changes in the density of the electric current* flowing through the nanopore *according to each of four bases*.

Types of nanopore sequencing

There are two types of nanopore sequencing:

1. **Biological nanopore sequencing**

2. **Solid state nanopore sequencing**

Biological nanopore sequencing

- Biological nanopore sequencing relies on the *use of transmembrane proteins*, called *protein nanopores*.
- These proteins are usually *bacterial toxins*.
- These toxins are embedded in *artificial lipid membranes*.
- They *create size dependent porous surfaces* - with *nanometer scale "holes"* distributed across the membranes.

- Normally a DNA or RNA strand has *high translocation velocity.*

- It means that it will travel *very fast* through these pores making it difficult to record data.

- *Sufficiently low translocation velocity* can be attained through the incorporation of various proteins that facilitate the movement of DNA or RNA through the pores of the lipid membranes.

Alpha hemolysin

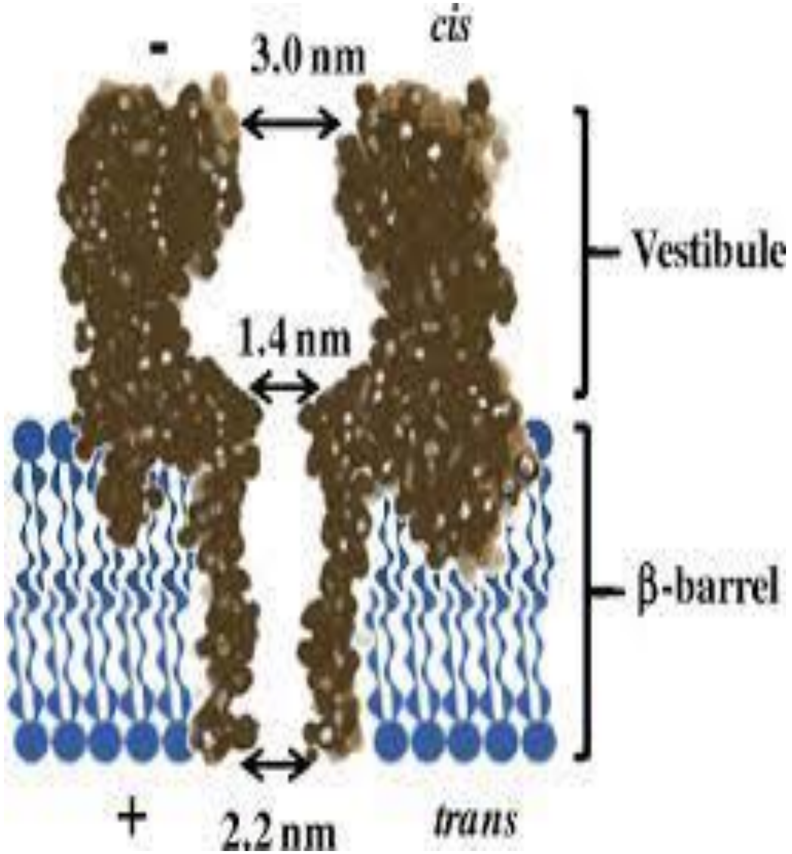
- Alpha hemolysin (αHL), a *nanopore forming toxin* from bacteria (e.g. *Staphylococcus aureus*) that causes lysis of red blood cells, has been studied for over 15 years.

- To this point, studies have shown that *all four bases can be identified* using ionic current measured across the αHL pore.

- The structure of αHL is advantageous to identify specific bases moving through the pore.

The αHL pore is ~10 nm long, with two distinct 5 nm sections.

The *upper section* consists of a larger, vestibule-like structure and the *lower section* consists of three possible recognition sites (R1, R2, R3), and is able to discriminate between each base.



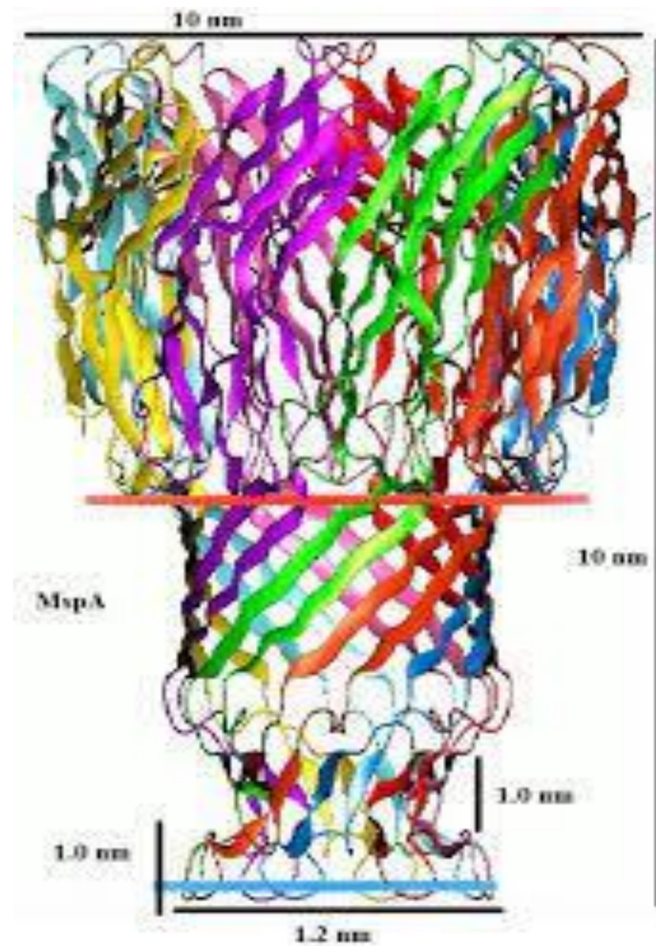
MspA

- *Mycobacterium smegmatis* porin A (MspA) is the second biological nanopore currently being investigated for DNA sequencing.

- The MspA pore has been identified as a potential improvement over αHL due to a *more favorable structure*.

The pore is described as a *goblet* with a thick rim and a diameter of 1.2 nm at the bottom of the pore.

The electric current detection of nucleotides across the membrane has been shown to be *tenfold more specific than αHL* for identifying bases.



Advantages

- **Uniform pore structure**

- The **precise control of sample translocation** through pore channels.

  Normally a DNA strand (ssDNS) moves rapidly at the rate of 1 to 5μs per base through the nanopore resulting in *low resolution* and *increased error rate*. The speed of DNA movement can be better controlled in Biological systems.

- The **detection of individual nucleotides** in samples can be facilitated by *unique proteins* from a variety of organisms.

Disadvantages

- The *sensitivity of the proteins* in these systems *negatively* affect longevity of the units, overall.

- *High pH sensitivity* is another issue.

Solid state nanopore sequencing

- Solid state nanopore sequencing approaches *do not incorporate proteins* into their systems.

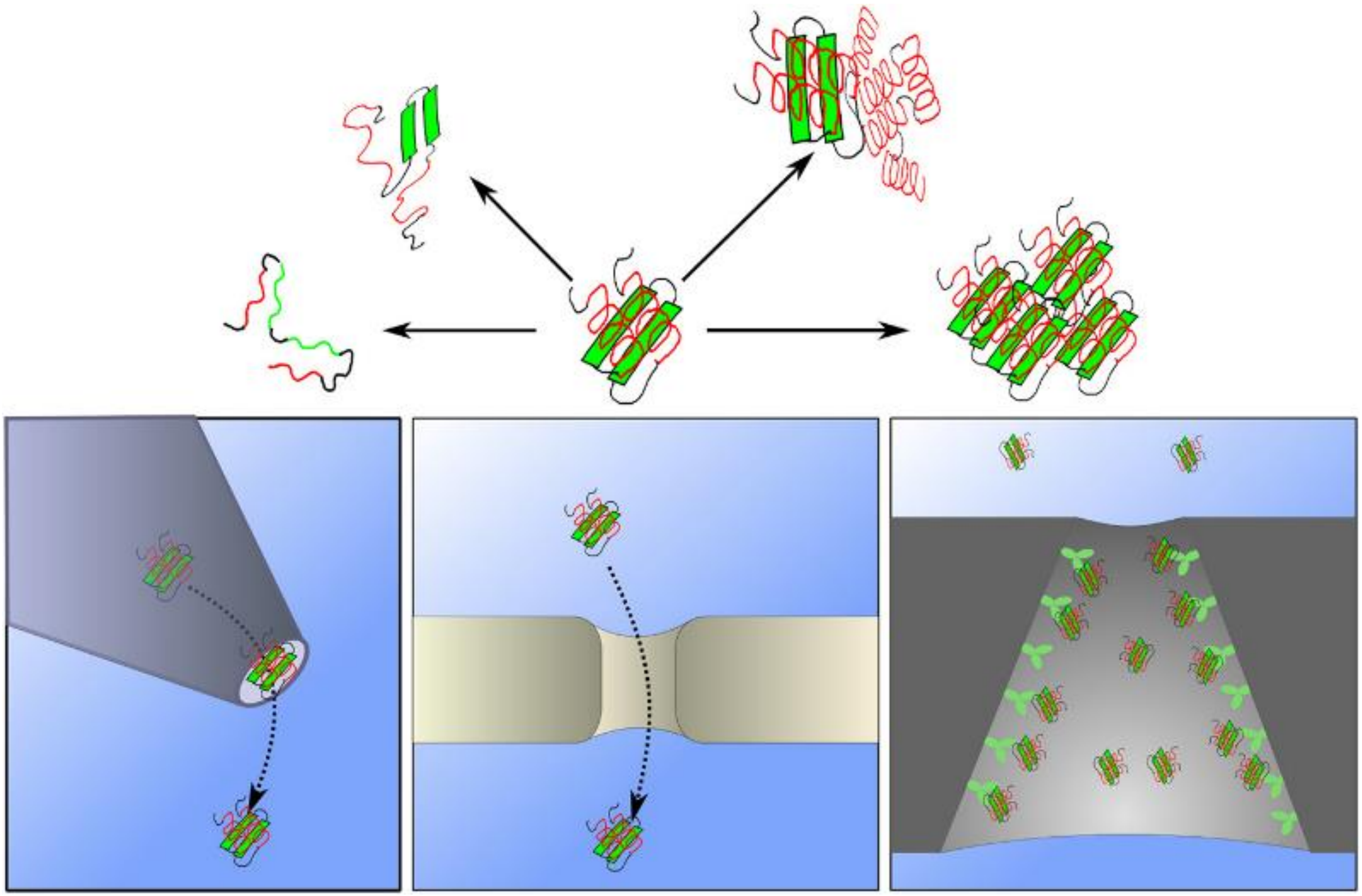
- Instead, various *metal or metal alloy substrates* with *nanometer sized pores* are used that allow DNA or RNA to pass through.

- Solid-state nanopores provide *superior material robustness* and *large-scale integrability with on-chip electronics*.
- So they have the potential to surpass the limitations of their biological counterparts.

- These substrates allow the *sequence recognition* of nucleic acids as they translocate through the channels along the substrates.

- Solid-state nanopores are generally made in **silicon compound membranes**, one of the most common being **silicon nitride**.

- The second type of widely used solid-state nanopores are **glass nanopores** fabricated by laser-assisted pulling of glass capillary.



Comparison of Biological and Solid State Nanopore Sequencing Systems

	Biological	Solid State
Low Translocation Velocity	√	
Dimensional Reproducibility	√	
Stress Tolerance		√
Longevity		√
Ease of Fabrication		√


Thank You!