GENOMICS

Nanopores read long genomic DNA

An ion-current map characteristic of four nucleotides traversing a protein nanopore allows resequencing of long DNA reads.

For over a decade, the promise of sequencing long stretches of DNA with minimal up-front sample preparation in real time through a nanopore has excited researchers. Jens Gundlach and his team at the University of Washington have a long history of contributing to the realization of such a nanopore sequencer.

The basic idea of nanopore sequencing is simple: a single strand of DNA passes through a small opening in a membrane that conducts an ionic current. As each nucleotide passes through, it reduces the current to a degree that is characteristic for this nucleotide; and through readouts of the residual current, the DNA can, in theory, be decoded. There are several requirements: a stable pore, a way to thread the DNA through this pore at a speed that allows recording of current changes, and an efficient way to decode these changes. Over the years Gundlach's team has helped to address all of these issues.

Nanopores can be made of various materials, but protein-based pores have proven the most suitable for sequencing purposes. For Gundlach, a physicist by training, this initially took a bit of refocusing. He says, "I actually had to learn that the better nanotech materials are proteins. Nature uses this nanotechnology all the time. Maybe nature would have chosen silicon if it was easier—it is an abundant element—but nature chose a carbon-based system." One of the big advantages of protein pores is their atomistic reproducibility, robustness and stability in the face of changing temperatures or pH.

Early protein pores were made of α -hemolysin, but, says Gundlach, such pores do not have the short, narrow constriction required for sequencing and, consequently, probe 10–12 nucleotides

MspA nanopore in a membrane. The green phi29 DNA polymerase threads single-stranded DNA through the pore, where quadromers in the pore constriction elicit current changes. Figure from Laszlo *et al.*, Nature Publishing Group.

(nt) at the same time. Thus, individual nucleotides are not decipherable. His team showed in 2010 that an engineered version of *Mycobacterium smegmatis* porin A (MspA) with a very short constriction that is only 1.2 nanometers wide provided discrimination between the four nucleotides. In subsequent work they showed that coupling phi29 DNA polymerase with MspA could control the rate of DNA translocation through the pore, and it allowed them to read short, ~50-nt stretches of DNA with single-nucleotide resolution.

This was an important breakthrough, but a practical nanopore sequencer would have to handle much longer sequence reads. Andrew Laszlo, a graduate student in Gundlach's group, together with post-doctoral fellow Ian Derrington, tackled the challenge of extending read length from a few tens to thousands of bases.

They observed that at any given time, the current through the pore was controlled by 4 nt passing through the constriction of an MspA pore. Each of these groups of 4 nt, or quadromers, triggers unique current values. The researchers decided to systematically map the value of each quadromer. They made a 256-nt-long stretch of DNA that contained all possible combinations of the quadromers

and measured the current as the DNA went through the pore.

After many iterative measurements, the team had a current reference map and used it to interpret the DNA of the 5-kilobase phi X virus. "The sequence of phi X is well known," says Gundlach, "so we first interpreted it with our original map, then took the phi X data to go back and correct errors in the original map. This process is still ongoing: the more we read, the more precisely we can define the quadromers."

With the ability to characterize long stretches of DNA, the pore is poised for certain applications, such as resequencing, that are important for clinical genomics. The MspA pore can also potentially be used to identify infections in patients. Gundlach's team showed that when they compared the sequence obtained for phi X against a database of 156 megabases of other viral DNA, they could identify it with high probability.

At present, Gundlach's team has not yet achieved high-accuracy *de novo* sequencing, but the team is actively working on eliminating error modes that are inherent in the system. Until they have an improved algorithm for *de novo* sequencing, Gundlach does not want to state the calling accuracy of the pore: "When you align something, you shift things around; and while you hit a lot of the levels right, you don't know how many deletes or inserts there were."

The team will tackle *de novo* sequencing on the algorithmic and also the experimental front. Switching from the phi29 DNA polymerase, which shows some erratic motion as it threads the DNA through the pore, to an enzyme with less stochasticity, such as a helicase, could produce a less noisy ion-current readout.

Nicole Rusk

RESEARCH PAPERS

Laszlo, A.H. *et al.* Decoding long nanopore sequencing reads of natural DNA. *Nat. Biotechnol.* **32**, 829–833 (2014).