
Identifying Individual Proteins Using Nanopores and Supercomputers

2021-11-17

New proof-of-concept study in Science shows capacity to identify proteins with single-amino acid resolution.

The amount and types of proteins our cells produce tell us important details about our health and how our bodies work. But the methods we have of identifying and quantifying individual proteins are inadequate to the task. Not only is the diversity of proteins unknown, but often, amino acids are changed after synthesis through post-translational modifications.

In recent years, much progress has been made in DNA reading using nanopores — minute membranes large enough to let an unspooled DNA strand through, but just barely. By carefully measuring the ionic voltage of the nanopore as DNA crosses over, biologists have been able to rapidly identify the order of base pairs in the sequence. In fact, this year, nanopores were used to finally sequence the entire human genome — something that was not previously possible with other technologies.

In new research out in [Science](#) magazine, researchers from [Delft University of Technology](#) in the [Netherlands](#) and the [University of Illinois](#) at Urbana-Champaign (UIUC) in the U.S. have extended these DNA nanopore successes and provided a proof-of-concept that the same method is possible for single protein identification, characterizing proteins with single-amino acid resolution and a vanishingly small (10^{-6} or 1 in a million) margins of error.



High-fidelity reading of single protein composition by pulling the same protein through the nanopore multiple times.

"This nanopore peptide reader provides site-specific information about the peptide's primary sequence that may find applications in single-molecule protein fingerprinting and variant identification," the authors wrote.

The workhorses of our cells, proteins are a long peptide strings made of 20 different types of amino acids. The researchers utilized an enzyme called helicase Hel308 that can attach to DNA-peptide hybrids and pull them, in a controlled way, through a biological nanopore known as MspA (mycobacterium smegmatis porin A). They chose the Hel308 DNA helicase because it can pull peptides through the pore in half-nucleotide observable steps, which correspond closely to single- amino acids.

Each step through the narrow gate theoretically produces a unique current signal as the amino acid partially blocks an electrical current carried by ions through the nanopore.

Lead author Henry Brinkerhoff, who pioneered this work as a postdoc in physicist Cees Dekker's lab, likens the protein to a necklace with different-sized beads. "Imagine you turn on the tap as you slowly move that necklace down the drain, which in this case is the nanopore," he said. "If a big bead is blocking the drain, the water flowing through will only be a trickle; if you have smaller beads in the necklace right at the drain, more water can flow through."

With their technique, the researchers can measure the amount of ion current very precisely — but not exactly, because the step-wise passage through the pore is irregular. However, by loading the liquid medium with helicases, the researchers can get many separate, overlapping reads of the same molecule, or in their terms, they can "rewind" the protein and read its amino acid sequence again. Doing so, reduced the errors from 13% to practically zero.

Their approach allowed the researchers to discriminate peptide variants that differed by only a single-amino acid — something they proved by creating synthetic peptides with only one amino acid changed and showing the system could discriminate among them.

But to read out the individual amino acids, they first had to know what sort of signal each one produces as it travels through the pore. Some of these signals may be counterintuitive, the researchers found.

For instance, when the bulky tryptophan amino acid moved through the constriction, the ion current first decreased and then, counterintuitively, increased relative to the small and medium-sized variants.

To understand the origin of these patterns, the team relied on supercomputer simulations by computational biologist Aleksei Aksimentiev (UIUC), performed on several of the fastest supercomputers available to academic researchers in the world: Frontera, at the Texas Advanced Computing Center; Blue Waters, at the National Center for Supercomputing Applications; and Expanse, at the San Diego Supercomputer Center.

Aksimentiev's team used a method called molecular dynamics simulation to recreate the behavior of the nanopore, proteins, and the surrounding medium, with atomic resolution. Such simulations cannot fully capture the true timescale of the nanopore activity, which extends to seconds. But by generating 40 to 50 initial states at different positions, and then running 70 simulations in parallel, the team was able to derive statistics for different conformations of peptides. From those, they computed the current and compared it to experiments. This computational work was led by Jingqian Liu, a biophysics graduate student in Aksimentiev's lab.

The simulations included 30,000 atoms interacting over 200 to 500 nanosecond and were able to match experimental results. More importantly, they showed why certain amino acids produce counterintuitive signals as they pass through the nanopore. In the case of the tryptophan variant, the signal could be traced back to a binding of the peptide side chain to the nanopore surface above the constriction.

"For each specific conformation, we could see what happened to the sidechain, whether it interacts with the surface or remains inside of the pore," said Aksimentiev, professor of Physics at UIUC. "Then we could establish directly that the binding of the sidechain enhanced

the current."

The simulations took weeks to generate on Frontera, currently the 10th fastest supercomputer in the world and the most powerful at any university. But they would have taken years with the type of computing cluster available on most campuses. The single protein identification research — for which there is a global race for success — was published online by Science as a "First Release" on November 4, 2021. The research was supported by the Dutch Research Council, U.S. National Institutes of Health, and U.S. National Science Foundation, among others.

"There's tremendous opportunities to develop diagnostics by reading individual protein using this nanopore approach," Aksimentiev said. "The computation will play a big role in developing these technologies. It's amazing that with computer models we can reproduce experiments and tell what sort of interactions are going on on the nano-scale."

Not only that, computer models provide a different modality for design, allowing researchers to test nanopores of different size or with strategically placed residues that can produce enhanced signals.

More work is required to perform reads longer than 20 amino acids and to identify amino acids that are heterogeneously charged, but Aksimentiev ventures that in three to five years it may be possible to develop a working model.

"We think that our new approach will allow us to detect post-translational changes," said Dekker, "and thus shine some light on the proteins that we carry with us."

Read the [original article](#) on The University of Texas at Austin.