Molecular Cell

CellPress

Spotlight

# Are nanopore technologies ready for the proteomic challenge primetime?

Amit Meller[1,2,*]
[1]Faculty of Biomedical Engineering, Technion – IIT, Haifa, Israel 32000
[2]Present address: The WYSS Institute, Harvard University, Boston, MA 02115, USA
*Correspondence: ameller@technion.ac.il
https://doi.org/10.1016/j.molcel.2021.12.018

**Novel techniques for single-protein molecule sequencing are rapidly becoming the focus of contemporary biomedical research. Here, Brinkerhoff et al. (2021) report a significant progress in nanopore-based rereading of DNA-peptide conjugates.**

High-throughput and low-cost DNA sequencing technologies have profoundly transformed basic biomedical research and are currently considered cornerstone tools in clinical diagnostics and drug discovery. Modern DNA sequencing technologies have not only enabled genomic analyses of individual cells, hence playing an instrumental role in uncovering cell to cell heterogeneity, they have also offered unprecedented sequencing speed and accuracy. These factors have played a crucial role in the timely mapping of the genomic makeup of thousands of organisms, including emerging viruses such as the SARS-Cov-2. Looking forward, the scientific community has identified high-throughput, single-cell proteomics as a major frontier (Alfaro et al., 2021). A comprehensive snapshot analysis of billions of proteins and their posttranslational modifications in a mammalian cell is considered the "holy-grail" of molecular biology, anticipated to shed light on numerous biological processes and open vast biomedical avenues for clinical breakthroughs. But unlike DNA, proteins cannot be amplified, hence putting a focus on single-molecule techniques that are able to quantify low copy-number, as well as abundant proteins with high precision. To this end, novel methods tackling single-protein molecule identification and protein sequencing are highly sought after. In a timely publication, Brinkerhoff et al., 2021 report a major advancement toward this goal, using a nanopore-based protein sequencing strategy parallel to an already successful DNA strand sequencing.

Among presently emerging DNA sequencing technologies, nanopores have carved their own niche, offering single-molecule DNA sequencing with extremely long reads in a lightweight, portable device (Lu et al., 2016). Unlike most other DNA sequencing methods, nanopore sequencing distinguishes among the four DNA nucleotides based on angstrom-level differences in their physical dimensions. These differences induce slight variations in the degree these nucleotides block the ion current flowing through the pore as the DNA strand is threaded through it. Importantly, the most successful to date nanopore strand sequencing method does not provide a base-by-base sequence readout. Rather, the nanopore ionic current traces are influenced by short sequences (called "*k*-mers") in the DNA strand, which are unidirectionally "ratcheted" through the nanopore in single-nucleotide steps. A sequential readout of the *k*-mers is essential to uncover the underlying DNA sequence. It is mediated by a processive motor enzyme capable of stepping the DNA strand through the nanopore.

Taking a similar approach, Brinkerhoff et al. (2021) progressively stepped short linear polypeptides through a genetically modified MspA pore embedded in a lipid bilayer. To achieve the stepwise motion required for protein sensing, they have chemically attached a DNA oligonucleotide to the C terminus of the peptide and used it as a handle to control the motion of a 26-amino-acid-long peptide through the nanopore (Figure 1A). This approach resembles a recently published work (Yan et al., 2021) but instead utilizes the Hel308 helicase known to step DNA in discrete half-nucleotide steps (roughly 0.33 nm) that better match the amino-acid spacing. It was estimated that about 8 amino acids within and around the MspA nanopore constriction affect the ion current. Still, due to the fine unidirectional stepping motion of the peptide, Brinkerhoff et al. were able to sense a single amino acid substitution (from aspartic acid to glycine or tryptophan).

Currently, the maximum length of peptide is limited by the total height of the MspA nanopore (about 10 nm), but interestingly the accuracy of the read could be significantly improved by rereading the same peptide repeatedly. Upon reaching the nucleic acids' amino-acid linker, the helicase fell off, and the biopolymer quickly slid back in through the nanopore under the influence of the electrical force (Figures 1B and 1C). Due to the high concentration of Hel308 in the solution, another helicase was already bound at its designated binding site, ready to pull back the biopolymer, thus facilitating repeated reads of the same molecule. While each individual peptide read involves significant noise, the accumulated analysis of several rereads dramatically improved the method's accuracy.

Brinkerhoff's report presents a bold step toward the ultimate goal of nanopore-based whole-proteome sequencing. This capability, when fully realized, would complement, or in some cases, fully replace well-established protein sensing methods such as mass-spectroscopy (MS) and enzyme-linked immunosorbent assay (ELISA). However, the road ahead for nanopore technologies is long: first, the nanopore method should be capable of identifying all 20 canonical amino acids in any (or at least most) of the possible

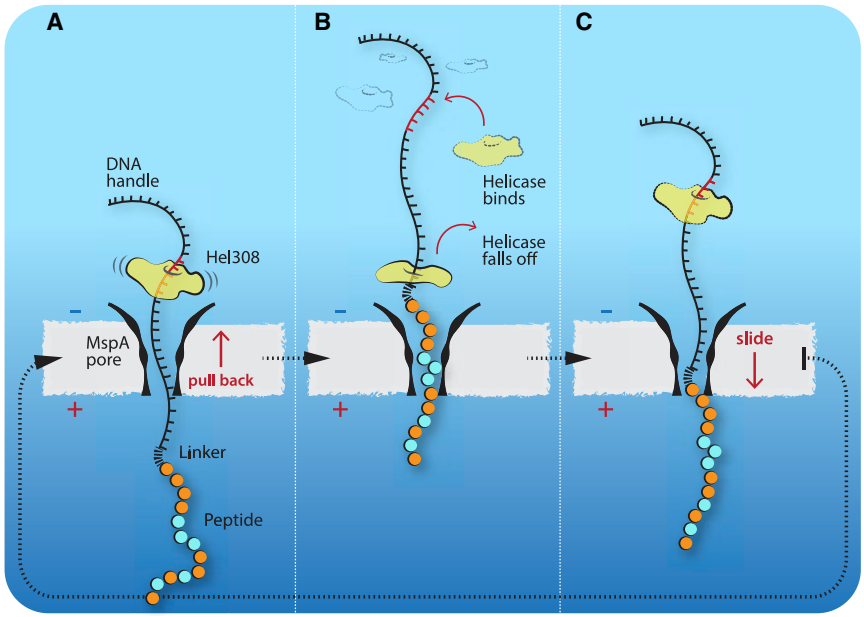


**Figure 1. Repetitive rereading of a peptide using a modified MspA pore**
(A) The Hel308 helicase pulls a DNA-peptide conjugate against the electrical force (pointing downward), allowing a stepwise reading of the ion current during the passage of the DNA handle and the peptide through the MspA pore.
(B) The helicase falls off the DNA-peptide conjugate when reaching the linker, and another helicase can be loaded onto the Hel308 binding fragment (red).
(C) After the first helicase falls off, the DNA-peptide conjugate quickly slides down under the influence of the electrical force, allowing the process to be repeated.
Figure reproduced from Brinkerhoff et al., 2021.

"*k*-mer" sequences, which obviously include a vast number of combinations. Second, the technique must be massively upscaled to meet the goal of whole-proteome sequencing in a reasonable timescale. A single human cell is estimated to contain billions of proteins, and even a prokaryotic proteome from *E. coli* contains about a million proteins. Third, to address the cell's vast proteoforms, the method must be able to sense and identify posttranslational amino-acid modifications with a single-amino-acid resolution.

However, the good news is that several groups have already made significant headway in addressing these challenges. For example, Ouldali et al. demonstrated the identification of all 20 amino acids individually via connection to cationic carriers containing 7 arginine residues using a different protein: aerolysin (Ouldali et al., 2020). Independently, Oxford Nanopore's MinIon is capable of reading the signals from hundreds of protein nanopores simultaneously fixed in a custom polymeric membrane (Lu et al., 2016). Another nanopore-based approach applies optical rather than pure electrical sensing to map the order of appearance of just two or three amino acids along each protein (Ohayon et al., 2019). Computer simulations of this protein reading process analyzed by machine-learning technology suggested that it can potentially be used to identify the vast majority of the human proteome based on single passages of the proteins through a nanopore.

Three decades ago, the journey for nanopore-based single-DNA-molecule sequencing appeared to be elusive (Deamer et al., 2016). But the vision, perseverance, and imagination of scientists around the globe showed that nothing is impossible. Today, profiling the entire protein repertoire of individual cells is considered one of the ultimate challenges in proteomics. Solving this challenge would likely transform proteomics research, just as contemporary sequencing techniques transformed genomics. Implementation of such a technology would enable the scientific community to explore a myriad of biological specimens with an unprecedented level of detail, from protein synthesis and degradation to the dynamics of signaling pathways involved in all cellular functions.

## ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) grant no. 833399 (NanoProt-ID) under the European Union's Horizon 2020 research and innovation program grant agreement.

## REFERENCES

Alfaro, J.A., Bohländer, P., Dai, M., Filius, M., Howard, C.J., van Kooten, X.F., Ohayon, S., Pomorski, A., Schmid, S., Aksimentiev, A., et al. (2021). The emerging landscape of single-molecule protein sequencing technologies. Nat. Methods *18*, 604–617.

Brinkerhoff, H., Kang, A.S.W., Liu, J., Aksimentiev, A., and Dekker, C. (2021). Multiple rereads of single proteins at single–amino acid resolution using nanopores. Science *374*, 1509–1513.

Deamer, D., Akeson, M., and Branton, D. (2016). Three decades of nanopore sequencing. Nat. Biotechnol. *34*, 518–524.

Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics Proteomics Bioinformatics *14*, 265–279.

Ohayon, S., Girsault, A., Nasser, M., Shen-Orr, S., and Meller, A. (2019). Simulation of single-protein nanopore sensing shows feasibility for whole-proteome identification. PLoS Comput. Biol. *15*, e1007067.

Ouldali, H., Sarthak, K., Ensslen, T., Piguet, F., Manivet, P., Pelta, J., Behrends, J.C., Aksimentiev, A., and Oukhaled, A. (2020). Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nat. Biotechnol. *38*, 176–181.

Yan, S., Zhang, J., Wang, Y., Guo, W., Zhang, S., Liu, Y., Cao, J., Wang, Y., Wang, L., Ma, F., et al. (2021). Single Molecule Ratcheting Motion of Peptides in a *Mycobacterium smegmatis* Porin A (MspA) Nanopore. Nano Lett. *21*, 6703–6710.