<Strapline> Nanopore sequencing

<Title>Unfolding the path to nanopore protein sequencing

<Standfirst> A modified nanopore enables enzyme-free threading of single polypeptides to detect post-translational modifications

<Authors> Adam Dorey and Stefan Howorka

DNA sequencing with nanopores has had a considerable impact during the COVID-19 pandemic by identifying SARS-CoV-2 varients fast. More generally, nanopores provide step-changing portable and long-read sequencing[1,2], simply by threading DNA strands through ring-like nanopores. Can nanopores also be used for protein sequencing? This is challenging, as the compactly folded polypeptides are more difficult to thread through a nanopore than elongated DNA. Now writing in *Nature Nanotechnology*, Martin-Baniandres et al.[3] describe how protein unravelling and threading helps determine biologically relevant post-translational modifications in proteins at single-molecule resolution.

In classical nanopore sequencing, individual, single-stranded, and negatively charged DNA molecules are electrophoretically threaded through a 1 nm-wide lumen of a protein pore[2,4,5]. This causes distinct electrical blockades for the bases. To facilitate sequential reading, a motor enzyme feeds the DNA strand base-by-base into the pore's lumen[4,5]. Replicating this for reading proteins would be a critical scientific advance. But protein sequencing is no easy feat[5]. Proteins are not only more complexly folded than DNA, deciphering 20 amino acids is also more challenging than reading four nucleobases. Furthermore, mature proteins contain chemically complex and extensive post-translational modifications (PTMs).

Turning protein sequencing with nanopores into reality requires researchers to achieve two aims: Resolving the 20 amino acid residues, including post-translational modifications, and controlling the unfolding and threading of polypeptides through the pore. Regarding the first aim, 20 amino acids could be discriminated when single residues were attached to a carrier peptide[6]. Nanopores are sensitive enough to distinguish the often-minuscule differences in residue size and structure.

Martin-Baniandres et al.[3] advance the field by threading proteins across a nanopore and detecting post-translational modifications within the chain (Figure 1). Crucially, the author use electroosmosis (EO) to thread proteins. The traditional electrophoretic pulling -as used for negatively charged DNA- does not work for polypeptides due to their charge-neutral backbone and the heterogeneously charged residues. In EO, an electric field drives the flux of buffer solution, which co-transports the protein chain through the nanopore.

Martin-Baniandres et al.[3] achieved EO with a engineered protein nanopore carrying a ring of positively charged amino acids at its narrowest part. The electrostatic filter caused strong flux of electrolyte anions and the surrounding water which co-transports the polypeptide chain. Protein unfolding and threading was enhanced by adding protein-denaturing agent guanidium hydrochloride (GdHCl) at 0.75 M. EO-mediated threading was demonstrated with engineered proteins carrying multiple interlinked globular

repeats resembling beads on a string. As threading of a single repeat causes a characteristic long unfolding signature, multiple corresponding signatures were expected for proteins with 2, 4, 6 and 8 repeats. Indeed, this was largely the case, even though a few repeats were not detected.

To determine whether PTMs can also be detected, the authors produced a protein with a single modification in between two central repeats. Three modification types were tested: phosphorylation, glycosylation, and glutathionylation, the most common PTMs. Upon pore threading, each modification caused a reduction in electrical signal proportional to the mass of the PTM. Significantly, the resolution was high enough to discriminate a mixture of proteins with different PTMs.

How practical is this approach -shown for one engineered protein- to proteomic analysis? As native proteins vary significantly in their resistance to unfolding, it is challenging to apply universal conditions to capture and thread all proteins within a cell. A potential solution is utilising voltage sweeps in combination with denaturants to promote protein capture, unravelling and threading. Furthermore, small PTMs that cause insufficient electrical signatures may be amplified with cognate antibodies or chemical binders.

How does the approach relate to other techniques? EO has been used previously for polypeptide transport through a non-engineered nanopore, but with a higher GdHCl concentration of up to 2 M[7]. GdHCl alone can causes EO as it is charged and adsorbs to the pore wall to induce flow of counterions. Using this system, longer proteins threaded at 10 µs per residue which is slower than in electrophoresis and may help improve resolution. However, the sensing of PTMs was not tested.

Alternatively, protein threading can be driven by enzymes. For example, a nanomachine was engineered by fusing an unfoldase motor protein or the 20S proteasome onto a nanopore[8]. The enzymatic activities allowed for threading long polypeptides or proteasome-fragmented shorter peptides, respectively. The strands were detected, yet individual amino acids were not resolved.

Another enzyme-driven method did, however, resolve single amino acids in a short peptide. This technology cleverly harnessed a DNA-processing motor protein in combination with DNA sequencing[9]. The peptide was fused to the oligonucleotide, so that the motor protein threaded the DNA as well as the linked peptide strand. Re-threading the strand increased sequencing accuracies to over 99.99%. This is a significant leap forward but currently limited to peptides <25 residues in length and peptides of predominantely negative charge.

So far, Edman degradation and mass spectrometry dominate the protein sequencing market. With these techniques, sequencing of peptides with 30-40 residues is often sufficient to correctly predict the protein's identity from a database, as does detecting the position and nature of any post-translational modification. While the detection level is at picomoles to femtomoles of protein, the sensitivity is insufficient for analyzing the proteome of single cells, which is common for DNA and RNA sequencing. This is likely an area where single-molecule methods can contribute[10].

It is currently difficult to predict which of the enzyme-driven or enzyme-free nanopore technologies, or combinations thereof, will come to market for protein sequencing. Likely, techniques will be tailored for applications where ease of preparation, accuracy, cost-effectiveness, read length, portability and scalability are case-specific criteria. These developments will also take advantage of existing DNA sequencing platforms and new commercial entrants[5].

Once, nanopore DNA sequencing was regarded as an outlandish idea and deemed impossible by several leading scientists. Yet, interdisciplinary research and iterative improvements -underpinned by vison and dogged determination- have yielded this powerful, commercially available, and widely used sequencing platform. We are now in a similar phase of exciting technological developments for nanopore protein sequencing and other single-molecule approaches[11].

References

1. Marx, V. Method of the year: long-read sequencing. *Nat. Methods.* **20**, 6–11 (2023).
2. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature.* **530**, 228–232 (2016).
3. Martin-Baniandres, P. *et al.* Enzyme-less nanopore detection of post-translational modifications within long polypeptides. *Nat. Nanotechnol.* (2023).
4. Wang, Y. Y., Zhao, Y., Bollas, A., Wang, Y. Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
5. Dorey, A. & Howorka, S. Nanopore-based nucleic acid sequencing technologies and their applications towards single-molecule proteomics. *Nat. Chem.* (2023).
6. Ouldali, H. *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
7. Yu, L. *et al.* Unidirectional single-file transport of full-length proteins through a nanopore. *Nat. Biotechnol.* (2023).
8. Zhang, S. *et al.* Bottom-up fabrication of a proteasome–nanopore that unravels and processes single proteins. *Nat. Chem.* **13**, 1192–1199 (2021).
9. Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A. & Dekker, C. Multiple rereads of single proteins at single–amino acid resolution using nanopores. *Science.* **374**, 1509–1513 (2021).
10. Reed, B. D. *et al.* Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science.* **378**, 186–192 (2022).
11. Eisenstein, M. Seven technologies to watch in 2023. *Nature.* **613**, 794–797 (2023).

Adam Dorey and Stefan Howorka

Department of Chemistry, Institute of Structural and Molecular Biology, University College London, London, UK. e-mail: s.howorka@ucl.ac.uk

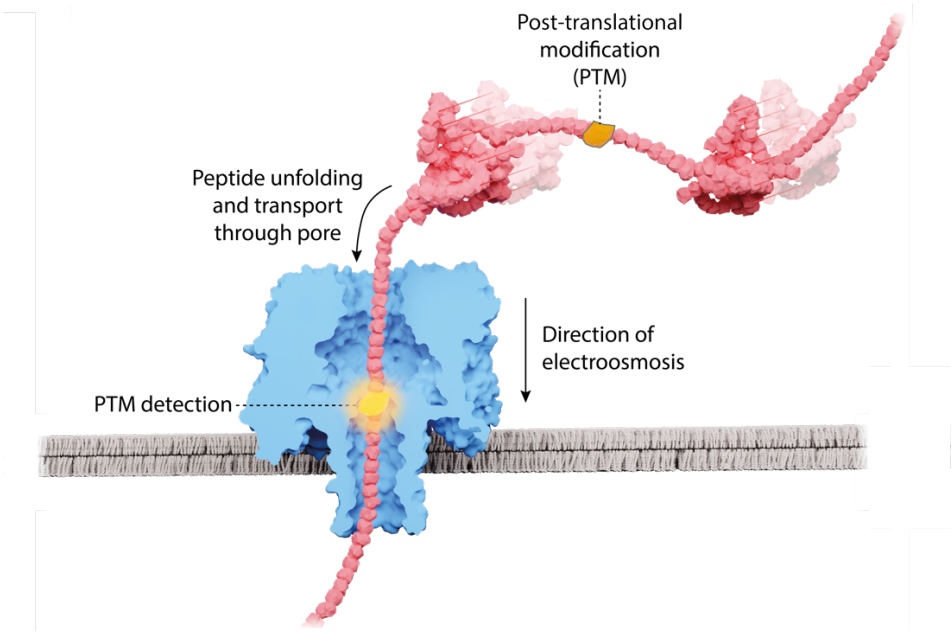

Figure 1: A peptide (pink) with globular repeats is driven through a protein pore (blue) embedded in a bilayer membrane (grey). Electroosmosis (arrow) drives the threading and heps detect post-translational modifications (yellow) in the polypeptide.