N&V SUBHEAD: SEQUENCING TECHNOLOGY

Reading amino acids in a nanopore

Stefan Howorka & Zuzanna S. Siwy

Stefan Howorka is in the Department of Chemistry, Institute of Structural and Molecular Biology, University College London, London, UK, and Zuzanna S. Siwy is in the Department of Physics and Astronomy, University of California, Irvine, Irvine, California, USA. e-mail: s.howorka@ucl.ac.uk and zsiwy@uci.edu

In a step toward nanopore sequencing of proteins, an aerolysin pore discriminates the twenty amino acids.

[AU: A few references were moved due to editing. Please check/revise the references as needed. Reduce total number of references to 10.] We have removed one. Could we please stick to 11? All the of them are essential. The N&V template you sent us has 13 references.

Nanopore sequencing of DNA works by threading single-stranded DNA through a narrow pore and measuring electrical signatures of the four bases—tiny perturbations, unique to each base, in the current flowing through the pore[1, 2]. Could a similar approach be used to sequence proteins? Perhaps, but the problem is more intractable. Proteins are chemically more complex than DNA, with 20 amino acids rather than 4 bases. In addition, proteins are more compact and heterogeneously charged, and need to be unfolded for pore threading[3]. **[AU: Revise/elaborate as needed.]** In this issue, Ouldali *et al*.[4] describe an important step toward single-molecule protein sequencing with nanopores. The authors electrically detect all 20 proteinogenic amino acids attached to a short peptide carrier [**AU: Add one-sentence summary of the paper.]**

Traditionally, protein sequencing has relied on either Edman degradation or mass spectrometry. A read-length of 30–40 residues is sufficient to retrieve the protein's identity from a database. These methods require only picomoles to femtomoles of protein, but this is not sensitive enough to sequence the tiny amounts of protein in small cellular samples[3] [**AU:OK**? **OK**]. A recent single-molecule optical approach to Edman degradation, which detects a small number of fluorescently tagged residues in single peptides on a massively parallel scale, works in the zeptomolar range[5]. However, this method has not been optimized for cost and accuracy, it requires cutting proteins into small peptides and labelling at least two amino acids [**AU: OK? Any other limitations worth mentioning (e.g., the limit on the length of the sequence that can be analysed)? Yes, two are mentioned**]. If nanopore sequencing of proteins can be achieved, it has the potential to outperform chip-based, optical sequencing with respect to easy of sample preparation, accuracy, cost and read length. **[AU: OK? Other potential advantages? Yes, one is mentioned]**

In nanopore sequencing of DNA, individual ssDNA molecules are pulled through membrane-embedded, pore-forming proteins, such as CsgG , in an array format[2]. Only one strand fits inside the pore, which has an inner width of around 1 nm. DNA bases are deciphered when they pass one-by-one through the pore's internal reading head, resulting in base-specific blockades in ionic current. Some pores have two reading heads to re-read the DNA and improve accuracy **[AU need reference, this has not been published yet, only as patent]**. DNA movement is unidirectional because an external electric field is applied that electrophoretically pulls the negatively charged strands. DNA translocation must be slowed down in order for bases to be read individually, so an enzyme is attached to the pore entrance to feed the DNA base-by-base into the reading head[1].

To adapt nanopore sequencing technology to proteins, researchers must solve two main problems: moving the protein through the pore and identifying 20 amino acid residues. Unfolding and transporting non-homogeneously charged polypeptides through a solid-state nanopore has

been accomplished using adhering negative ionic detergents **[AU:OK? YES]**[6]. Whole proteins have also been moved electrophoretically through a protein nanopore using an attached oligonucleotide[7]. Another method, which is electrophoresis-free, pulls proteins through a protein nanopore with an unfoldase[8]. Despite this progress on moving the protein through a pore, [**AU tell the reader the remaining challenges as far as moving the protein through the pore, Done**], it is not yet possible to thread polypetides in a ratchet-like motion to resolve each residue, similar to DNA sequencing.

Reading and differentiating all 20 amino acids is vastly more demanding than reading four DNA bases. In addition, the average amino acid residue is 2-3 times smaller than a monophosphate nucleotide making it harder to detect the smaller current blockade [**AU is there any size comparison you can make with DNA bases to relate this back to the increased difficulty of the protein sequencing problem**? **Done**] However, amino acid residues vary up to 2.5-fold in size which aides their distinction [**AU insert examples, Done**]. Isolated amino acids including proline, histidine, glutamic and aspartic acid, isoleucine, lysine, cysteine and methionine **[AU name them please, Done]** have been distinguished with metal nanoscale gaps [**AU explain and say if this is in a solid state nanopore, Done**] via the alternative read-out method of transversal tunnelling current[9, 10]. However, distinction of all 20 amino acids within a peptide sequence **[AU: this suggests to me that the authors have sequenced actual peptides. Delete the highlighted part? The wording has been changed to show that the authors have not sequenced peptides]** has not been shown until now.

Ouldali *et al*.[4] report detection of all 20 proteinogenic amino acids with a biological nanopore, bringing single-protein sequencing with nanopores closer to fruition. Two factors contributed to the success of their experiments. The first was their choice of nanopore. In order to detect all 20 amino acids, the authors chose to use the aerolysin nanopore. The lumen of this nanopore is wide enough, at ~2.5 nm [**AU insert dimension, Done**], to accommodate peptides of several residues The sensing zone of the wild-type aerolysin pore [**AU naturally? Without**

engineering? If it just happens to be that the peptide moves slowly through this zone and if the reason is known points to it here please, Done] slows down the movement of the peptide, which aids in amino acid detection. The pore's sensing zone is about 2 nm long, about 3-4 time **[AU: how much? Done]** longer than the reading head for DNA sequencing. Sensing of a peptide relies on it residing inside the pore for at least a few milliseconds. **[AU: Can you comment on how many residues contribute to the signal at a given moment? No]**

The second factor underlying the progress reported was that the authors linked each of the 20 residues to a carrier peptide comprising seven arginines [**AU important to tell the reader how they hit upon a 7-arg tag. Did they try other tags**? **Done**]. In previous work, homopolymeric peptides led to well-defined ion current blockages that differed depending on single-residue changes. The net positive charge ensures unidirectional [**AU: correct? Yes**] electrophoretic transport of the peptides across the pore. **[SJ: We are talking about just one AA residue linked to 7 Arg residues, for a total of 8 residues per peptide, right? If not, the structure should be clarified. The edits and the new Fig. 1a clarify this]**

When the 20 carrier constructs were individually passed through the aerolysin pore, 13 amino acids were accurately distinguished by ion current blockage. Reassuringly, the molecular volume of these residues correlated with the magnitude of the blockade. Importantly, the 13 individual peptides species were also discriminated in mixtures, with the sensitivity being sufficient to distinguish leucine from isoleucine. The remaining 7 amino acids clustered into two groups with different current blockages. Using computer simulations, the authors showed that these amino acids could be distinguished by a combination of increasing residence time in the pore and chemically modifying the aerolysin pore or some amino acids. In preliminary experiments, chemically modifying methionine and tyrosine [**AU in the pore? Or by putting a different tag on the amino acids met and tyr. Please explain. Done]** significantly changed ion current blockades.

What are the limitations of the approach? While 13 amino acids can be reliably distinguished, each must be attached to a carrier peptide. In addition, peptides with a single residue substitution are distinguished by current blockade but this is different to sequencing where one residue after each other is identified in a threading strand. The pore would also require engineering to discern all 20 amino acids.

Looking ahead, how can tagged amino acids and an aerolysin pore be used to obtain sequence information from proteins? The authors propose that one could sequentially cleave amino acids from an analyte protein and then link released residues to a peptide carrier for analysis. A simpler solution would not require a carrier but rather pull analyte proteins through the pore with an unfoldase motor, assuming that many amino acids in a non-repetitive peptide sequence could be distinguished [**AU would these need to be labelled with the 7 arg peptide? No**]. Until now, nanopore sensing of proteins has succeeded only in identifying bulky post-translational modifications[6], discriminating unphosphorylated, monophosphorylated and diphosphorylated proteins (ref. 85 in [3]). For protein sequencing, translocation of the protein would have to be slow enough to enable reading of multiple residues. Helpful strategies may include applying a pressure difference across the pore to assure controlled, unidirectional motion of the protein, and dual identification of residues by blockade amplitude and duration, as well as combining ion and tunnelling read-out, as explored for DNA sequencing[11]. Another component might be adoption of nanopores with subnano reading heads, related to those used in DNA sequencing [**AU how would this be better than aerolysin and do you have a reference – is this indeed published? Explained, and it has not been published**]. Smaller reading heads could detect separate amino acids in translocating polypeptides.

The report by Ouldali *et al*.[4] marks an important step toward single-polypeptide reading with nanopores. With further refinements, it might be applied to fingerprint proteins with a few easy-to-distinguish amino acids, as has already been achieved via fluorescence[3]. Direct *de-novo* sequencing of proteins remains a formidable challenge. But nanopore sequencing of DNA also faced several hurdles in its past. The study by Ouldali *et al*.[4] illustrates the benefit of catalysing progress in protein sensing by learning lessons from DNA sequencing in order to advance biological discovery, biomedicine, and diagnosis.

AU please tell me where the figure is from: a cartoon would be better, do you have one? An improved version of Fig. 1 has been provided by email on 18 Dec at 15:45.

Figure 1. Nanopore-based sensing of peptides carrying all 20 different amino acid residues. (A) A cationic carrier peptide of seven arginine amino acids is chemically linked at the C-terminus to each of the 20 different amino acids. (B) The aerolysin pore features a sensing zone that experimentally differentiates 13 out of 20 amino acids by the magnitude of current blockade. (B) An example read-out trace of pore current showing different current blockades for amino acids-Arg_7 peptides in a mixture.

Reference list cannot be more than 10. Please remove the review and one other, apologies.

1. Manrao, E. A. *et al. Nat. Biotechnol.* **30**, 349-353 (2012).
2. Quick, J. *et al. Nature* **530**, 228-232 (2016).
3. Restrepo-Perez, L., Joo, C., Dekker, C. *Nat. Nanotechnol.* **13**, 786-796 (2018).
4. Ouldali, H. *et al. Nat. Nanotechnol.*, (2020).
5. Swaminathan, J. *et al. Nat Biotechnol*, (2018).
6. Kennedy, E., Dong, Z., Tennant, C., Timp, G. *Nat. Nanotechnol.* **11**, 968-976 (2016).
7. Rodriguez-Larrea, D., Bayley, H. *Nat. Nanotechnol.* **8**, 288-295 (2013).
8. Nivala, J., Marks, D. B., Akeson, M. *Nat Biotechnol* **31**, 247-250 (2013).
9. Zhao, Y. *et al. Nat. Nanotechnol.* **9**, 466-473 (2014).
10. Ohshiro, T. *et al. Nat. Nanotechnol.* **9**, 835-840 (2014).

11. Fanget, A. *et al. Nano Lett.* **14**, 244-249 (2014).