

# Toward single-molecule protein sequencing using nanopores

Received: 2 August 2024

Accepted: 31 January 2025

Published online: 17 March 2025

 Check for updates

Chunzhe Lu<sup>1,2</sup>, Andrea Bonini<sup>1,2</sup>, Jakob H. Viel<sup>1,2</sup> & Giovanni Maglia<sup>1,2</sup> 

Over the past three decades, biological nanopore sequencing has grown from a research curiosity to a mature technology to sequence nucleic acids at the single-molecule level. Now, recent achievements suggest that nanopores might be able to sequence proteins soon. In this Perspective, we analyze the different approaches that have been proposed to measure proteins and peptides using nanopores. We predict that, more likely than not, nanopores will be capable of identifying full-length proteins at the single-molecule level and with single-amino acid resolution, paving the way to single-molecule protein sequencing. This would allow several applications in proteomics that are at present challenging, including measuring the heterogeneity of post-translational modifications, quantifying low-abundance proteins and characterizing protein splicing.

The sequencing of the human genome has been a scientific landmark achievement, and it has had major implications for our understanding of human health<sup>1</sup>. However, excluding contributing factors from the epigenetic state of the genome, all 200 cell types in our body have the same DNA. What finally distinguishes them are the proteins they produce. Proteins determine the correct functioning of our cells, playing a pivotal role in health, disease, diagnoses and potential cures. For this reason, it is essential to understand not only our genome but also all the proteins that are expressed in our body, our proteome. However, accessing our proteome like we did for our genome poses a considerable challenge.

Compared to DNA, proteins are more complex in many ways. Instead of having four bases and a uniform charge as DNA, proteins are made of 20 amino acids with diverse charge contents and chemical compositions. Moreover, multiple proteins may be synthesized from a single gene through alternative messenger RNA splicing and post-translational modifications (PTMs). These so-called proteoforms and their distribution are crucial for functionality, and they substantially increase the challenge of protein sequencing. Effectively, identifying just part of a protein, or even its whole amino acid sequence, does not accurately represent the set of proteoforms produced from a specific gene. Finally, some proteins and proteoforms in our cells are several orders of magnitude more abundant than others. This dynamic range complicates the detection of proteins that are present in small amounts. A protein-sequencing method that can overcome all these challenges would lay the groundwork for many new fields of study.

Currently, most proteomic analysis is done with bottom-up mass spectrometry (MS). In this method, protein samples from cell extracts are enzymatically digested, the resulting peptides are separated by chromatography and the masses in the resulting peptide mixture are determined<sup>2</sup>. With genetic or proteomic information of the protein sample available, the predicted *in silico* digestion products are compared with the detected fragments and the proteins from which the fragments originated are identified.

To sequence proteins without reference data, *de novo* protein sequencing, a combination of Edman degradation and tandem MS (MS/MS), is used<sup>2,3</sup>. In Edman degradation, a peptide sequence is determined one residue at the time by the chemical removal and subsequent identification of amino acids from the N terminus<sup>4</sup>. It requires purified protein samples with concentrations in the low picomolar range and can sequence up to ~30 amino acids in length<sup>5</sup>. With MS/MS *de novo* sequencing, the amino acid sequence of a protein can be identified through fragmentation and subsequent fragment mass determination<sup>6</sup>. Here, a bottom-up approach can also be used<sup>7</sup>. A purified protein is predigested, after which the resulting peptides are separated by chromatography. The peptides are measured by a first analyzer (MS<sup>1</sup>) and then fragmented and finally sequenced by a second analyzer (MS<sup>2</sup>).

The methods described above give an incredible insight into our proteome, but they generally require purification and fragmentation methods that cause proteoform information to be convoluted or lost<sup>8</sup>. To tap into this extra layer of information, methods to address an entire

<sup>1</sup>University of Groningen, Chemical Biology Department, Groningen Biomolecular Sciences and Biotechnology, Groningen, Netherlands. <sup>2</sup>These authors contributed equally: Chunzhe Lu, Andrea Bonini, Jakob H. Viel, Giovanni Maglia.  e-mail: [giovanni.maglia@rug.nl](mailto:giovanni.maglia@rug.nl)

protein are being developed, like top-down MS/MS, where intact rather than predigested protein samples are measured<sup>9</sup>. Although this method is a major improvement for proteoform identification, locating PTMs to specific amino acids for each proteoform remains for now challenging<sup>9</sup>. Top-down MS/MS, already employed for single-cell proteomics, still faces challenges with dynamic range, resolution and coverage<sup>9,10</sup>.

Even though developments in the field of top-down proteomics are promising, considering the clinical relevance of de novo protein sequencing, an alternative approach is warranted. A major challenge in proteomic analysis is that proteins cannot be amplified. This is often a drawback, as many relevant proteins are present in heterogeneous mixtures and in low abundance. These proteins are typically missed by the ensemble methods used in discovery-based proteomics. Here, single-molecule nanopore sequencing, especially when coupled to high-throughput analysis, can play an important role.

## Nanopores used for sequencing

A nanopore is a small water-filled nanoscale aperture on an insulating membrane. Nanopores may be drilled in solid-state membranes<sup>11</sup> or can be made of proteins<sup>12,13</sup>, DNA<sup>14</sup> or peptides<sup>15</sup>. To date, nanopores that are successfully used in DNA-sequencing applications are made of proteins. These ‘biological nanopores’ are proteins that play a crucial role in a variety of biological processes, facilitating the regulated translocation of ions, water, small molecules and other substances. Advantages of biological nanopores are that they can be made cheaply and reproducibly, while not changing over the time of the measurement, and they can be engineered with atomic precision. Therefore, here we only consider biological nanopores (hereafter nanopores) for sequencing.

The size of nanopores can vary considerably, ranging from a few angstroms to tens of nanometers. Typically, nanopores with a narrow constriction such as *Mycobacterium smegmatis* porin A (MspA) and curlin sigma S-dependent growth subunit G pore (CsgG) are used for polymer sequencing, whereas nanopores with an extended barrel, such as α-hemolysin (α-HL), aerolysin and cytotoxin K (CytK), and α-helical nanopores, such as fragaceatoxin C (FraC), are used for peptide analysis (Fig. 1a–f). Although at present only a limited number of nanopores can be found in nature, computer-designed nanopores might also be generated<sup>16–18</sup>, suggesting that nanopores with bespoke size and shapes will soon become available.

Over the past decades, nanopores have been developed to identify molecules, to study chemical and enzymatic reactions and, most notably, to sequence nucleic acids at the single-molecule level (Box 1). In nanopore analysis, a membrane separates two compartments, named *cis* and *trans*. A potential is applied across the membrane, which generates an electric field and a resulting current of hydrated ions across the nanopore (Fig. 1g,h). Molecules lodged inside the nanopore generate ‘blockades’, providing information about the molecule’s structure, translocation dynamics and chemical identity. This current change highly depends on the type of nanopore used and the analyte investigated.

## Challenges of using nanopores to address single proteins

Taking DNA sequencing as a benchmark (Box 1), the sequencing of proteins with nanopores requires us to overcome two fundamental challenges: identifying the 20 natural amino acids and the plethora of PTMs, and transporting polypeptides unidirectionally in a linear manner across nanopores at a speed that is compatible with reading individual amino acids.

### The issue of identifying 20 amino acids

Over the past three decades, many studies revealed that nanopores can differentiate among molecules, including single amino acids in peptides<sup>19–23</sup>, unfolded proteins<sup>24</sup> or folded proteins<sup>25</sup>, and PTMs<sup>24,26–32</sup>. Most strikingly, differences between two enantiomers, the smallest

chemical difference in two molecules, have been reported using several nanopores<sup>26,33–35</sup>. This suggests that nanopore currents should be capable of identifying any difference between two amino acids in protein.

Several studies have investigated whether all 20 amino acids can be distinguished simultaneously by the same nanopore. Initial work used a wild-type aerolysin nanopore and sampled the 20 proteinogenic amino acids attached at the N terminus of a poly-R7 sequence<sup>36</sup>. It was found that, assuming a residence time of 200 ms, 16 peptide species could be identified with a probability of 90% or higher. Some amino acids (such as methionine and tyrosine) could be further separated through selective chemical modifications<sup>36</sup>. In addition, two recent studies demonstrated that both Ni<sup>2+</sup>-modified MspA<sup>37</sup> and copper(II)-functionalized MspA<sup>38</sup> can differentiate among the 20 individual amino acids, albeit many amino acids showed overlapping Gaussian distributions. Immobilized strands might be used to mimic the pausing of enzymes between steps. A recent work with α-HL nanopores used host–guest interaction to immobilize a set of 20 peptides containing a polyanionic tail differing by the 20 proteinogenic amino acids at a specific position. Despite some amino acids showing overlap in their Gaussian distributions, the authors found that 14 amino acids could be differentiated by the wild-type pore, with the remaining six distinguished using an engineered pore<sup>39</sup>.

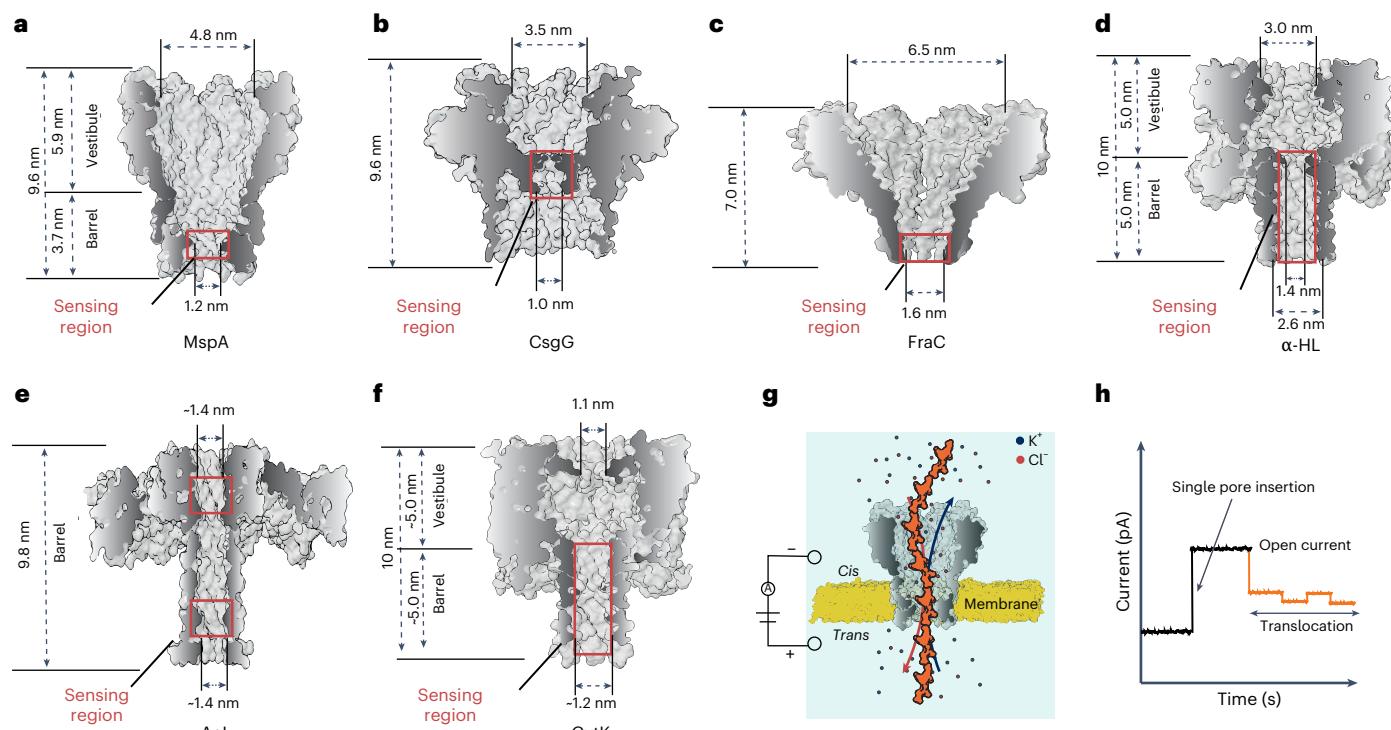
Another research focus is whether all individual amino acids of a polypeptide can be sequenced as it moves through the pore. The first report of amino acid resolution in moving polypeptide strands used idealized peptides attached to single-stranded DNA ratcheted by a helicase and revealed that at least some amino acids (for example, aspartic acid, glycine and tryptophan)<sup>40</sup> can be addressed by MspA-M2 nanopores. More recent work using idealized polypeptides moved by an unfoldase across a CsgG nanopore showed that differences in individual amino acids are based on charge or steric exclusion<sup>41</sup>. Using a rereading strategy, the authors achieved 61% accuracy in the 20-amino acid classification.

These works showed that, overall, there might be enough bandwidth in the nanopore signal to uniquely identify most amino acids by a single nanopore. However, it is worth noting that differences between similar amino acids are not always significant during single reads and signal overlapping occurs, which may affect the measurement accuracy. In addition, in several studies, amino acids could be distinguished using the mean values from multiple measurements, with overlap in the Gaussian distributions. Hence, if rereading approaches are not used, as discussed below, the precise identification of amino acids during a single pass might be challenging. Furthermore, a yet bigger challenge might be the presence of multiple amino acids within the constriction of the nanopore (about 17 in CsgG<sup>42</sup>). This complicates the signal notably. In DNA sequencing (Box 1), the deconvolution of the signal deals with four canonical nucleobases. In proteins, the signal must account for 20 amino acids (plus modifications). Thus, de novo sequencing of proteins provides a more difficult combinatorial problem than sequencing DNA.

However, it is also worth noting that, with a step size of two amino acids as measured for the unfoldase ClpX<sup>41</sup>, the same amino acid is read multiple times, which should help signal deconvolution. Furthermore, separation and deconvolution of signals from individual amino acids might not be required. This is because a crucial advantage of proteins compared to DNA is that a proteome has a limited number of genes (~20,000 in humans), which limits the number of possible amino acid combinations. Therefore, protein sequencing might be achieved by solving three problems: (1) finding a nanopore that provides unique signals for all proteins (or peptides) in a proteome, (2) establishing a correlation between the signal and the protein sequence and (3) identifying the variations from the reference signal with single-amino acid resolution.

### The problem of delivery to the nanopore

A second main challenge in nanopore analysis is the delivery of a molecule to the nanopore sensing region, which is often near or around the



**Fig. 1 | Nanopore analysis.** **a–f**, Structure of MspA (**a**), CsgG (**b**), FraC (**c**),  $\alpha$ -HL (**d**), aerolysin (AeL) (**e**) and CytK (**f**) nanopores, which are widely used in DNA and protein sequencing applications. **g**, Schematic representation of a polypeptide translocating across a nanopore (gray) embedded into an amphiphatic membrane (yellow). When a potential is applied, ions (for example,  $K^+$  and  $Cl^-$ )

are transported across the nanopore. **h**, An idealized example of an ionic current signal generated by a single nanopore inserted into the amphiphatic membrane. When a polymer translocates across the nanopore, fluctuation of the signal can provide information about the polymer.

narrower part of the nanopore, the constriction (Fig. 1a–f). A charged molecule experiences an electrophoretic force ( $F_{EP}$ ) that dominates the diffusion of charged molecules across the nanopore. In the case of DNA, electrophoretic forces are crucial for initial DNA threading into the nanopore and to linearize the DNA strand inside the nanopore. Polypeptides have a weak and variable charge. Hence, the  $F_{EP}$  on the polypeptides cannot be used to thread proteins across nanopores. This has long been thought to pose a fundamental limitation to nanopore protein sequencing. However, it is also known that, in ion-selective nanopores, the external potential creates a directional fluid motion, which results in an electroosmotic flow ( $F_{EOF}$ ) on the fluid and nearby molecules. This  $F_{EOF}$  affects the diffusion and capture of neutral molecules<sup>43</sup>, folded proteins<sup>44,45</sup> and (poly)peptides<sup>19,29,46</sup> inside a nanopore. Therefore, the  $F_{EOF}$  might be used to transport polypeptides across nanopores, providing that, for charged polypeptides, the  $F_{EOF}$  is stronger than the  $F_{EP}$ .

During our initial work with FraC nanopores, we recognized that peptides are captured by an  $F_{EOF}$  (ref. 19). However, subsequent work with aerolysin<sup>46–48</sup> showed that peptides might not translocate through the nanopore against an applied potential. Meanwhile, using FraC nanopores<sup>49</sup>, we demonstrated that the translocation of charged peptides against an applied potential was only achievable by reducing the peptide's charge by lowering the pH to 3.8 (ref. 49). This raised questions about whether an  $F_{EOF}$  could overcome the  $F_{EP}$  during polypeptide translocation in sequencing applications.

Full-length proteins unfolded by guanidine hydrochloride have been initially studied using weakly selective  $\alpha$ -HL or aerolysin nanopores<sup>50,51</sup>. Recently, it was discovered that adding a charged polymer (such as DNA or polyaspartate) at the protein terminus is required for most proteins to translocate across the nanopore<sup>52</sup>. Interestingly, molecular dynamic simulations suggested that the binding of guanidine hydrochloride cations to the lumen of the nanopore induces an

$F_{EOF}$ , facilitating protein transport<sup>52</sup>. Another recent study showed that an  $F_{EOF}$  generated by introducing a ring of positive charges into the barrel of an  $\alpha$ -HL nanopore allows the capture of a folded protein elongated with a neutral polypeptide<sup>29</sup>. However, the translocation of the captured and unfolded protein, which was also observed with the weakly selective  $\alpha$ -HL nanopore<sup>24,53–55</sup>, was promoted by both the  $F_{EP}$  and the  $F_{EOF}$ . In a study published at around the same time, we set out to investigate whether unraveled proteins without additional tags could be transported across a nanopore against an  $F_{EP}$ . We found that, if a CytK nanopore's inner surface is lined with at least three rings of charges, an  $F_{EOF}$  can be engineered that is strong enough to transport unstructured polypeptides and urea-unfolded proteins against relatively strong  $F_{EP}$ <sup>56</sup>. This work proved that the  $F_{EOF}$  can be used on proteins in a similar way as the  $F_{EP}$  is used on DNA, indicating that no fundamental limitation exists for the sequencing of proteins using nanopores.

## Approaches to sequence proteins using nanopores

Several approaches have been formulated for the identification of proteins using nanopores. First, a distinction might be made whether information from single proteins is retrieved. This is an important distinction, because, despite nanopores being single-molecule sensors, the information extracted is often aggregated over many reads. Extracting sequence information from single reads is relevant, for example, to establish the heterogeneity of a protein sample or to measure complex mixtures. Nanopore devices for ensemble protein analysis could provide portable, low-maintenance, low-cost and rapid proteomic measurements when MS is not available. Here, we will focus on approaches that allow the identification of single proteins or peptides, an important focus of proteomic analysis.

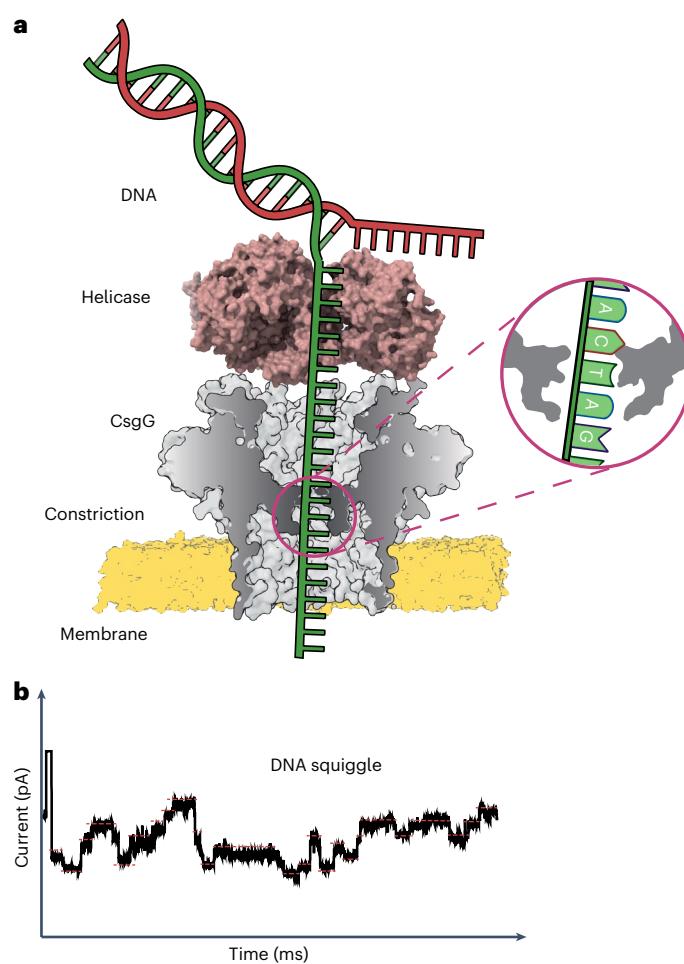
A second distinction might be made whether full-length proteins or peptides are identified. Peptides are important in many biological

**BOX1**

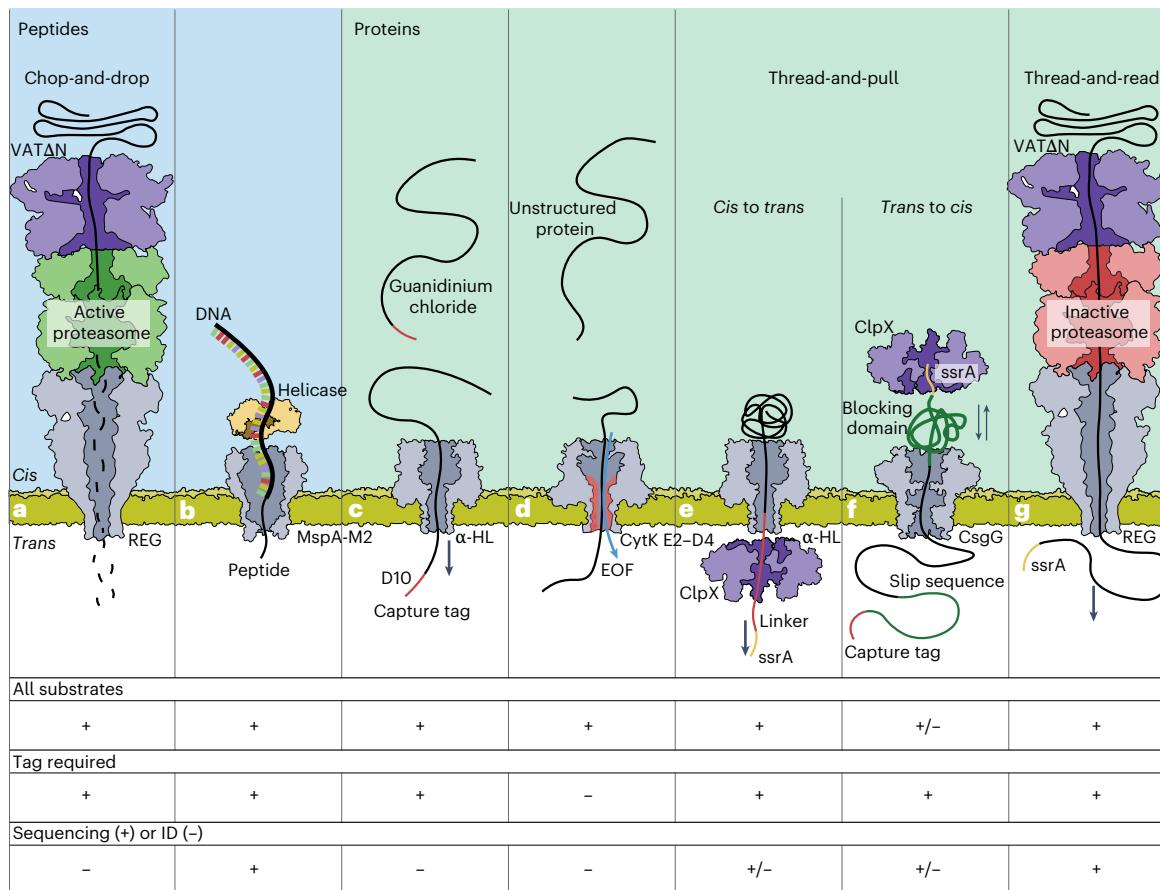
## Nanopore DNA sequencing

Arguably the biggest success of nanopore analysis has been the sequencing of single DNA molecules. From its initial conceptualization in 1989 (ref. 76) to the commercialization of the first portable nanopore DNA-sequencing device (Oxford Nanopore Technologies' MinION) in 2014 (ref. 77), nanopore DNA sequencing has taken 25 years to be realized. Crucial discoveries were the ability of the applied potential to induce the translocation of single-strand DNA across a nanopore<sup>78</sup>, the recognition of two<sup>79</sup> and then four<sup>80</sup> DNA bases in immobilized DNA strands (and four individual deoxyribonucleoside 5'-monophosphates<sup>59</sup>) and the stepwise motion of DNA by enzymes<sup>81,82</sup>. Aside from the above milestone achievements, however, the success of nanopore DNA sequencing relies on several key elements. First, a high-resolution nanopore is required. For example, nanopores such as MspA and CsgG have a relatively narrow constriction (~1nm), which allow large differences between nucleobases to be detected. Second, a driving force must

be used to facilitate the threading of the DNA strand and to keep the strand linearized throughout the nanopore. Owing to the uniform and negative charge properties of DNA, this force can be given by an externally applied potential. Third, a double-stranded DNA strand must first be unwound, and a resulting single strand must be passed through a nanopore in a base-by-base manner. This task is provided by molecular motor proteins such as DNA polymerase Φ29 (ref. 82), helicase 308 (ref. 83) and helicase Dda<sup>76</sup>. A complication is that multiple bases are read simultaneously at the constriction of nanopores (that is, >4 by MspA)<sup>84</sup>. Hence, a final challenge has been the deconvolution of the signal into a sequence. Key to this process is the unidirectional stepwise motion of DNA across the nanopore, allowing the same base to be addressed multiple times. Using this method, single DNA molecules can now be sequenced with precision over 99% during single nanopore passes<sup>85</sup>.



**Box Fig. 1 | CsgG nanopore used for DNA sequencing.** **a**, Schematic diagram of DNA sequencing using the CsgG nanopore (PDB 4UV3) and the ATP-dependent helicase Dda (PDB 3UPU). The helicase unwinds a double-stranded DNA molecule and delivers the resulting single DNA strand into the CsgG nanopore. The sensing region of this nanopore covers up to five consecutive nucleotides. **b**, Example current readout of nanopore DNA-sequencing signal.



**Fig. 2 | Approaches of nanopore peptide and protein sequencing.**

**a**, Single-molecule nanopore peptide spectrometry using a VAT unfoldase–proteasome–nanopore-based approach by chop-and-drop. **b**, Helicase-assisted peptide sequencing. **c**, Agent-assisted unfolding and translocation for protein identification. A capture tag is required for threading the nanopore. **d**, EOF-induced transport of unfolded proteins. No tag is required to induce initial transport. Unstructured polypeptides or protein unfolded by using denaturants can be used. **e**, Protein identification by a *cis*-to-*trans* thread-and-pull approach.

Proteins are electrophoretically captured and remain folded on the *cis* side. A ClpX unfoldase pulls the substrate from *trans* across  $\alpha$ -HL nanopores. **f**, Protein sequencing by a *trans*-to-*cis* thread-and-pull approach. Substrates are first electrophoretically translocated through CsgG to the *trans* side and then pulled out by ClpX to the *cis* side. **g**, Protein sequencing by a thread-and-read approach using VAT unfoldase and proteasome–nanopore. Unfolded substrates are immediately captured by the nanopore.

processes (for example, in immunology), and sequencing them by MS analysis is often challenging. A single-molecule peptide-sequencing approach would also provide a complementary method to address the limitation of MS-based bottom-up proteomics, the workhorse of proteomics analysis. The sequencing of full-length proteins, however, is likely to deliver the most complete information about a protein. Full-length sequencing methods will preserve information about the heterogeneity in the polypeptide chain, allowing researchers to address complex information such as the exact distribution of PTMs or information about spliced proteins. This information is often obscured by sample fragmentation or inherently convoluted in bottom-up MS analysis<sup>9</sup>.

#### Chop-and-drop analysis of full-length proteins

A starting point for protein identification is through protein fingerprinting. We<sup>57</sup> and others<sup>58</sup> showed that proteins predigested with specific peptidases (for example, trypsin) are identified according to their nanopore peptide spectrum or fingerprint. This approach, named nanopore peptide spectrometry<sup>21</sup>, may require a robust database if broadly applied. However, it was also shown by others<sup>21</sup> and our group that nanopore signals correlate with specific properties of peptides such as their volume. Therefore, when genetic or proteomic information of the protein sample is available, predicted *in silico* digestion products might be used to identify proteins de novo.

An alternative method called exopeptidase sequencing, which is the equivalent of a method proposed for DNA sequencing<sup>59</sup>, has been proposed in which amino acids are sequentially released from a peptide C or N terminus using an exopeptidase and subsequently identified. This method, which was initially conceptualized to be used with MS<sup>60</sup>, might also be used with nanopore technology<sup>61</sup>. Exopeptidase nanopore sequencing is an enzymatic nanopore equivalent of Edman degradation, albeit with potential for higher speed and throughput and lower detection limits.

Both identification methods might be turned into single-molecule approaches by attaching a peptidase above the nanopore by a chop-and-drop mechanism. Early steps have been made by our laboratory in nanopore spectrometry, in which an archaeal proteasome has been attached above a designed REG (proteasome activator 28 $\alpha$ ) nanopore (Fig. 2a)<sup>62</sup>. A VAT unfoldase has been shown to deliver proteins to the nanopore, which are gradually ‘chopped up’ into peptides. This work is a stepping stone, showing the capture of peptides from individual proteins. Bioinformatic analysis revealed that 97% of proteins in the human proteome can be identified using this approach<sup>63</sup>.

#### Strand sequencing: peptides

One approach might identify polypeptides as they are transported in single file across a nanopore. In 2021, three independent groups attached a single-stranded DNA molecule to a peptide terminus. The

DNA was then used by enzymes (Hel308 helicase<sup>40</sup>,  $\Phi$ 29 DNA polymerase<sup>64</sup> or MTA helicase<sup>65</sup>) to move the polypeptide stepwise through an MspA-M2 nanopore. These methods have shown differences in individual amino acid substitutions on negatively charged peptides (Fig. 2b). For example, using these DNA-peptide conjugate substrates, an MspA-M2 nanopore can discriminate a single-amino acid (aspartic acid, glycine and tryptophan) difference in the peptide sequence using either Hel308 (ref. 40) or  $\Phi$ 29 DNA polymerase<sup>64</sup>. Hel308 is a particularly suitable motor protein because it moves DNA by half a nucleotide during ATP binding and hydrolysis, which is roughly the step size required to address individual amino acids. PTMs such as phosphorylation and sulfation in weakly charged peptides have been detected<sup>30,66</sup>. One group<sup>65</sup> used a ‘sandwich’ strategy, whereas a 23-amino acid neutral peptide sequence was embedded between two DNA strands, allowing them to overcome the issue of using a negative peptide strand. In this configuration, using MTA helicase, single-amino acid substitutions, such as aspartic acid, glutamic acid, lysine, glutamine and phenylalanine, as well as PTMs like phosphorylation within a neutral glycine homopolymer environment were observed. One challenge of this approach is that positive or neutral peptides could not be stretched by the electric field, and they tangled up inside the nanopore<sup>65</sup>. Another limitation is the reading length for peptides, at present, ~25 amino acids, which depends on the length of the nanopore (MspA and CsgG are ~10 nm; Fig. 1a). Using a nanopore with longer length should allow the reading of longer peptides.

### Strand sequencing: full-length proteins

The ultimate goal of proteomic analysis is the ability to sequence full-length proteins. For this, proteins must be unfolded and linearly transported across a nanopore at a speed that is compatible with single-amino acid analysis. Two approaches have been proposed, one of which uses an enzyme to unfold the target protein and the other does not.

**Enzyme-free characterization of a protein by consensus identification.** In one approach, proteins are unfolded and threaded across nanopores by electrophoretic and/or electroosmotic forces. Early work showed that, in the presence of guanidine hydrochloride, proteins may be translocated across an  $\alpha$ -HL nanopore<sup>50,51</sup>. Later, it was shown that an electrophoretic tag attached at a protein terminus might be required to obtain protein translocation (Fig. 2c)<sup>52</sup>. The threading orientation and identity of two proteins extended with polyaspartate extensions could be distinguished with accuracies greater than 90%<sup>52</sup>, suggesting that single-translocation events may contain sufficient information to fingerprint proteins. We recently showed that a nanopore with a strong  $F_{EOF}$  allowed us to characterize urea-destabilized proteins without the need to introduce electrophoretic appendices at the N or C terminus<sup>56</sup>, providing a simplified path to protein analysis (Fig. 2d).

Native proteins may be elongated with a long unstructured tag to allow capture by the nanopore and may be unfolded spontaneously or by using a small amount of denaturant and transported across nanopores<sup>24,29,53,54</sup>. Differences in phosphorylation, glutathionylation and glycosylation could be observed during the pauses of domain unfolding<sup>24,29</sup>. However, in freely translocating strands, phosphorylated residues required chemical modifications to enhance the signal<sup>32</sup>. Free translocation could provide a fast and efficient way to measure PTMs in proteins. However, these findings suggest that methods to slow down molecular transport across the nanopore (for example, by engineering the nanopore’s inner surface<sup>67</sup>) and effectively stretching polypeptides may be required to identify differences in individual amino acids.

It is possible, nonetheless, that free transport across the nanopore will be too erratic to allow single-pass resolution, and, possibly, many of these methods will require the compilation of a consensus signal from the transport of multiple copies of the same protein (Fig. 3a). Although this might limit their use for the characterization of simple proteomic

mixtures, rereading approaches, as described below, might increase accuracy on single-molecule reads. Furthermore, this approach has the advantage of being rapid, not requiring sample preparation and being potentially low cost.

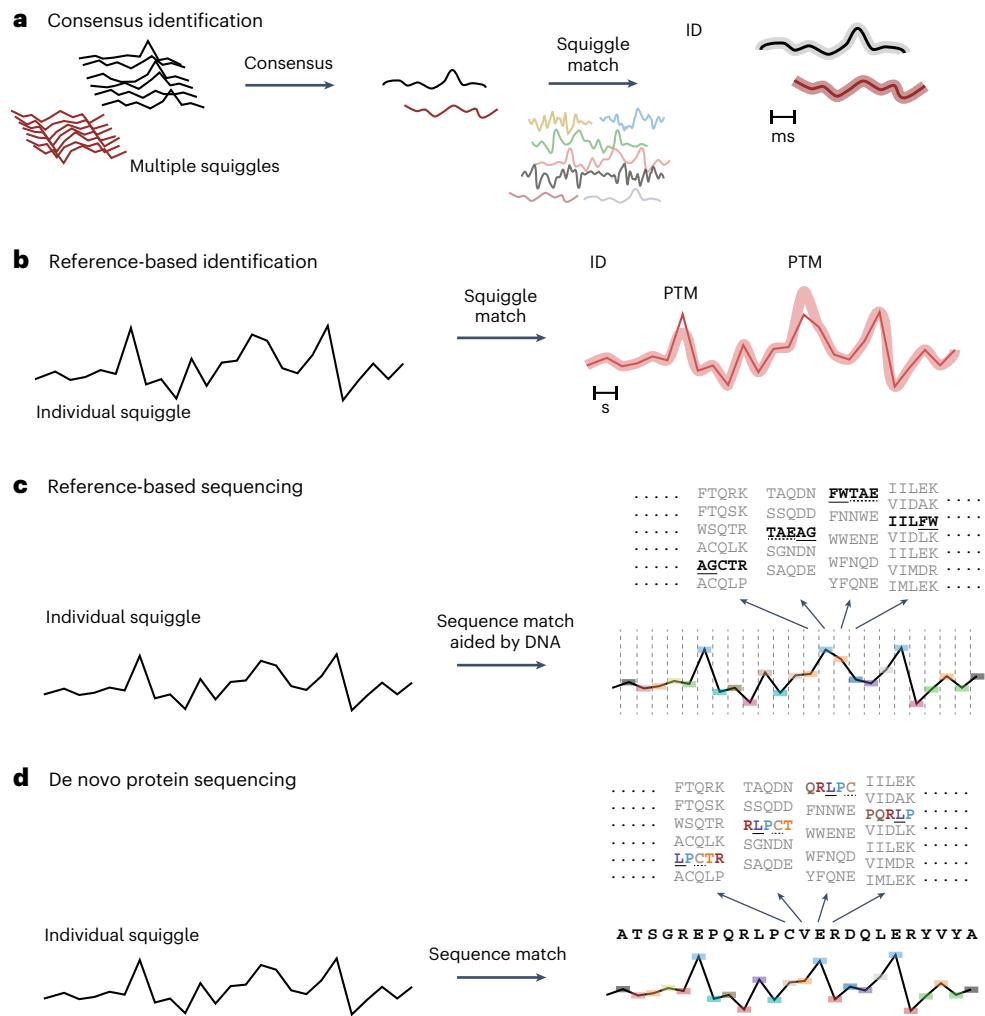
**Enzyme-based full-length analysis during a single pass.** The identification of proteins during a single pass will allow the single-molecule characterization and sequencing of proteins in mixtures. As observed for DNA, enzymes might provide the favorable solution for controlling polypeptide translocation speed to allow single-amino acid resolution.

In a first example, a protein containing a C-terminal ssrA tag (for recognition by an unfoldase) and a 65-amino acid-long charged polypeptide was electrophoretically threaded across an  $\alpha$ -HL nanopore from the *cis* side (Fig. 2e)<sup>68</sup>. Full translocation was prevented by a folded protein introduced after the charged polypeptide. Next, the unfoldase ClpX placed on the *trans* side was used to pull and unfold the protein across the nanopore. A follow-up work demonstrated the ability to discriminate between different proteins and their variant domains<sup>69</sup>. This *cis*-to-*trans* thread-and-pull approach used the chemical energy of ATP to pull the polypeptide through the nanopore. The narrow entry of the nanopore was then used to mechanically unfold the protein. However, if the unfolding of the proteins rather than the polypeptide passing the constriction of the nanopore dominates the signal, this approach might not allow the sequencing of proteins.

Another issue is that the unfoldase and the substrate protein are introduced on opposite sides of the membrane, which limits their use in commercial nanopore platforms where the *trans* side is sealed. This issue was tackled by electrophoretically threading an unfolded polypeptide from the *cis* side extended by a blocking protein containing an ssrA tag (Fig. 2f)<sup>41</sup>. Next, a ClpX unfoldase subsequently added from the *cis* side was used to pull the unfolded polypeptide back out of the nanopore. Single point mutations were tested using 59-amino acid-repeating sequence blocks (named PASTOR), each containing a single-amino acid substitution. Five consecutive PASTOR regions allowed the analysis of up to five different amino acids in a single read. These idealized polypeptide sequences allowed the characterization of many important features. Ionic currents from all 20 amino acids revealed that both the size and the charge of amino acids contribute to the nanopore signal. The PASTOR design allowed measurement of the step size of the unfoldase (ClpX), which was found to be two amino acids in accordance with previous studies<sup>70</sup>. Furthermore, the authors showed the detection of monophosphorylation and polyphosphorylation, an important proof-of-concept application of nanopore analysis.

Finally, the authors tested two PASTOR substrates containing non-idealized proteins: titin I27<sup>V15P</sup> and I27<sup>V15P,C47E,C63E</sup> domains, which were destabilized by the V15P substitution<sup>71</sup>, and an amyloid  $\beta$  peptide 1–42 ( $A\beta$  peptide) and its shorter derivative amyloid  $\beta$  peptide 1–15. The protein domains could be transported to the *trans* side of the nanopore, where they refolded, and then pulled back out by ClpX to the *cis* side. For titin, ClpX-mediated unfolding was observed by deep current blockades most likely reflecting the occupancy of titin within the nanopore constriction, followed by a sequence-specific signal for the transport of unfolded titin domains. The nanopore signals of the different peptides and protein domains were similar, but distinct features in the unfolding states could be observed.

This work strongly suggests that the transport of different proteins and their modifications across a nanopore is likely to produce unique signals. However, the present form of this *trans*-to-*cis* thread-and-pull method presents limitations. Notably, only unstructured or weakly folded proteins can be addressed and the unfolding of the target protein against the nanopore might mask the sequence-specific current signal. Furthermore, because the action of the unfoldase cannot be started at the nanopore, subsequent steps of loading, washing and unfolding and reading are required to address each single protein, reducing throughput compared to free-flowing systems.



**Fig. 3 | Approaches to full-length nanopore protein identification.** **a**, In consensus identification, protein signals are grouped, averaged and then matched to known signals. Differences in the matched squiggles (for example, because of PTMs or mutations) might be observed by variation of the signal. **b**, In reference-based identification, individual protein signals are of good enough quality to be matched after a single pass. Deviations from the predicted signal are used to identify, for example, PTMs. **c**, In reference-based sequencing, segments of the signal are matched to known protein signals. Although multiple

possible sequences are possible, the correct sequence is selected by analyzing genomic data. Stepwise motion would help recognition, as groups of amino acids are read multiple times. In the example, the step size is three amino acids, and the constriction reads five amino acids at a time. **d**, In de novo sequencing, high-quality signals allow the sequencing of individual amino acids. Amino acid-by-amino acid motion would help recognition, as each amino acid is read multiple times.

The nanopore–enzymatic machine described earlier (chop-and-drop)<sup>62</sup>, in which the proteasome is inactivated or removed, might provide a solution for many of these issues. In the thread-and-read mode (Fig. 2g), the VAT unfoldase was shown to deliver proteins to the nanopore–proteasome and allow multiple unraveled translocations of folded proteins from *cis* to *trans* (Fig. 2g) in a free-flowing manner. However, several challenges must still be overcome. The signal from the translocating protein was too short for full protein characterization. Most likely, the length of the proteasome–nanopore was too long to characterize the entire protein, the  $F_{EOF}$  was too weak to allow smooth translocation and the nanopore was too wide to show features in the translocating protein. If these issues can be solved, this approach should be compatible with commercial nanopore sequencing platforms and it should be able to address full-length proteins.

#### The power of resequencing

The ability of sequencing or recognizing proteins or peptides during a single pass might be challenging, as thermal fluctuations during protein translocation could make the signal noisy. Work with peptides

showed that, by increasing the concentration of the helicase (that is, to  $>1\text{ }\mu\text{M}$ ), the same peptide can be reread multiple times<sup>40</sup>. Similarly, the incorporation of a ‘slippery’ amino acid sequence near the N terminus of a PASTOR induced the unfoldase to ‘lose grip’ on the strand, allowing the protein to reread into the pore by electrophoresis. By combining data from multiple rereads, identification accuracy can be substantially increased, allowing for low single-molecule error rates, even when single-pass accuracy is limited, thus greatly improving the accuracy of amino acid classification. Furthermore, if single proteins can be recaptured multiple times after their first translocation, for example, by reversing the applied potential as already done for DNA<sup>72</sup>, then proteins might be identified even without an enzyme controlling the transport speed across the nanopore.

#### En route to protein sequencing

Although important first steps have been taken, several key technical problems must be solved to allow nanopore protein sequencing. Most enzymes that can transport proteins across a nanopore recognize a specific sequence of amino acids or require a DNA strand at the C or

N terminus. Hence, these moieties must be attached to the protein of interest. Chemical or enzymatic methods that can modify the N or C terminus of proteins have been described in the literature<sup>73,74</sup>. However, it is unknown whether they act on all proteins and/or peptides or whether they are efficient or selective enough for nanopore characterization.

Chop-and-drop methods should prove whether all generated fragments can be captured in sequence by the nanopore and whether the signal from a single blockade can identify the molecule. Theoretical work assessing a similar method proposed for DNA highlighted the challenge of capturing individual molecules<sup>75</sup>. Most likely, an enclosing peptidase–nanopore will be required to aid the capture of protein fragments. In this respect, exo-sequencing has the additional challenges that no exopeptidase is known to enclose its substrate and that the  $F_{EOF}$  is weaker on smaller molecules. Finally, no exopeptidase is known to unfold proteins, suggesting that only peptides might be sequenced.

Strand-sequencing approaches might be at a higher readiness level. They have shown the ability to move (poly)peptides across nanopores and to recognize at least a few amino acids in moving strands. In peptide sequencing, it will be important to show that peptides can be stretched by the  $F_{EOF}$ . A more challenging task will be decoding the signal of the peptide translocating through the nanopore. As peptides are short (-25 amino acids can be read) and multiple amino acids are lodged within the nanopore constriction simultaneously, most likely large training sets will be required to learn the peptide signals.

Full-length strand-sequencing methods are the ultimate goal of nanopore sequencing. Initial work using the thread-and-pull approach with idealized polypeptides showed that the different amino acids induce significant and specific changes to the nanopore signal. Furthermore, sequence-specific signals (squiggles) from natural domains (amyloid  $\beta$  and titin I27<sup>V1SP</sup>) have been shown. Although, in the present state, the thread-and-pull approach might only address partially structured domains and the signal from the unfolding of the protein against the nanopore might interfere with the squiggle, methods using enzymes to unfold proteins and feed the unstructured polypeptide *cis* to *trans* might overcome these issues. Nanopores with a strong  $F_{EOF}$  have been shown to capture and thread unraveled proteins and unstructured polypeptides. Work with DNA indicated that the enzymes above the nanopore usually do not significantly affect the current signal from the translocating polymer, suggesting that the signal should only reflect polypeptides passing the nanopore constriction. The challenge is to ensure that the enzyme starts the unfolding process at the nanopore rather than in solution. An important question is whether a motor protein can provide consistent and stepwise movement of the polymer to enable high-accuracy protein sequencing.

#### De novo sequencing versus identification with single-amino acid resolution

Compared to a genome, which has an astronomical number of possible nucleobase arrangements, typically, a proteome only has a few thousand unique protein sequences. Therefore, many proteomic approaches simply require identifying proteins, ideally with single-amino acid resolution (that is, all amino acids should contribute to the signal). Hence, we expect that the simplest approach to nanopore proteomics will be a reference-based identification approach, in which a database of squiggles from known purified proteins will be first collected and then matched to the protein analyzed (Fig. 3b).

A range of applications can be envisaged. Single-molecule resolution combined with high-throughput analysis will allow characterization of low-abundance proteins, a challenge in MS analysis. Full-length protein analysis will reveal information not easily accessible to bottom-up proteomics, such as profiling the heterogeneity of PTMs in proteins. Portable, low-cost and high-throughput devices will allow, for example, assessment of the quality of protein-based formulations, such as antibodies, in real time at the site of protein production, at the clinic or at home.

## BOX 2

### Industry efforts focusing on single-molecule protein sequencing

Commercial efforts are underway to offer single-molecule sequencing of proteins and peptides using nanopores. Portal Biotech is developing full-length protein-sequencing methods (<https://www.portalbiotech.com/>) based on the published technologies from our laboratory<sup>56,62</sup>. Oxford Nanopore Technology is exploring peptide- and protein-sequencing methods<sup>86</sup> like the ones recently published for peptide<sup>40,64,65</sup> and full-length protein analysis<sup>41</sup>.

A few other single-molecule approaches, based on different reading technologies, are already offering partial peptide sequencing. Quantum-Si technology offers a solution that relies on the identification of the N terminus of immobilized peptides using three different fluorescently labeled recognizers (typically binding one or more amino acids at time). Recognition is obtained by decoding parameters such as their temporal order and on/off binding kinetics, allowing the identification of several amino acids and some PTMs. Exopeptidases are then used to remove the last amino acid in the peptide<sup>87</sup>. In a similar fashion, Encodia uses affinity reagent binders to recognize the N terminus of an immobilized peptide. The binders, however, introduce a DNA barcode tag to create a peptide–DNA chimera molecule that is eventually sequenced after the last amino acid to be identified is removed. Erisyon technology chemically labels a subset of amino acids with fluorophores. The entire peptide's fluorescence is read before a step of Edman degradation is performed to remove one amino acid at the N terminus. These approaches take advantage of the massively parallel platforms developed for second-generation DNA-sequencing technologies but have a main limitation that only peptides can be sequenced. The Nautilus platform offers the possibility to identify full-length proteins using a short-epitope mapping (PrISM) labeling technique<sup>88</sup>, in which arrays of billions of single proteins are immobilized on DNA pads<sup>89</sup> and different cycles of the multiaffinity PrISM library probe are then probed. The platform translates single interaction events into protein counts and can recognize billions of different single-molecule proteins in one measuring run. For sequencing proteins, the development of binders that identify single amino acids within an intact and unfolded polypeptide strand might be required.

If the nanopore signal can be learned, then de novo recognition might become possible, expanding the application to all proteins that cannot be easily or cheaply preanalyzed. In the first instance, reference-based (de novo) sequencing is envisaged in which a genetic reference will simplify the recognition of protein sequences by limiting the combinatorial space of the number of possible amino acids (Fig. 3c). Eventually, if all 20 amino acids provide distinguishable signals in a moving strand and an unfoldase is found to move processively ideally one amino acid at the time (a possible requirement to reduce signal complexity), de novo protein sequencing should also become possible even without a genetic blueprint (Fig. 3d).

#### Conclusion

The development of DNA-sequencing technologies that allowed low-cost sequencing of the human genome has brought a revolution in science and medicine. The ability to understand the proteome at

the single-amino acid level will likely bring a similar outcome. Industry efforts are now underway to make single-molecule proteomics a reality (Box 2). Nanopores have advantages compared to other techniques, especially those based on fluorescence, as they process native and intact molecules, they do not require costly antibody recognition and they might eventually be capable of de novo sequencing of proteins. Furthermore, the nanopore signal can be easily interfaced with low-cost silicon-based devices, and arrays of hundreds of thousands of nanopores allow high-throughput analysis.

Single-molecule protein identification appears to be on the cusp of realization, already promising a wealth of applications in proteomics, including full-length PTM profiling. De novo protein sequencing may require overcoming additional challenges. First, the rich nanopore signal should be translated into sequence information, a task for which artificial intelligence tools might prove especially suited. Second, it might be necessary to bioengineer the molecular components required for sequencing proteins. Existing nanopores appear to already be capable of detecting the subtle chemical differences among the 20 amino acids, but bespoke engineering might be required to measure polypeptides as they move through nanopores. Under electroosmotic forces, unfoldases are expected to move substrates unidirectionally through nanopores. However, it is yet unknown whether they can move polypeptides single amino acid by single amino acid, which might be necessary to simplify the deconvolution of the nanopore signal.

## References

- Hood, L. & Rowen, L. The Human Genome Project: big science transforms biology and medicine. *Genome Med.* **5**, 79 (2013).
- Dupree, E. J. et al. A critical review of bottom-up proteomics: the good, the bad, and the future of this field. *Proteomes* **8**, 14 (2020).
- König, S., Obermann, W. M. J. & Eble, J. A. The current state-of-the-art identification of unknown proteins using mass spectrometry exemplified on de novo sequencing of a venom protease from *Bothrops moojeni*. *Molecules* **27**, 4976 (2022).
- Edman, P. & Begg, G. A protein sequenator. *Eur. J. Biochem.* **1**, 80–91 (1967).
- Chang, E. et al. N-terminal amino acid sequence determination of proteins by N-terminal dimethyl labeling: pitfalls and advantages when compared with Edman degradation sequence analysis. *J. Biomol. Tech.* **27**, 61–74 (2016).
- Vyatkina, K. De novo sequencing of top-down tandem mass spectra: a next step towards retrieving a complete protein sequence. *Proteomes* **5**, 6 (2017).
- Miller, R. M. & Smith, L. M. Overview and considerations in bottom-up proteomics. *Analyst* **148**, 475–486 (2023).
- Su, T., Hollas, M. A. R., Fellers, R. T. & Kelleher, N. L. Identification of splice variants and isoforms in transcriptomics and proteomics. *Annu. Rev. Biomed. Data Sci.* **6**, 357–376 (2023).
- Po, A. & Eyers, C. E. Top-down proteomics and the challenges of true proteoform characterization. *J. Proteome Res.* **22**, 3663–3675 (2023).
- Brown, K. A., Melby, J. A., Roberts, D. S. & Ge, Y. Top-down proteomics: challenges, innovations, and applications in basic and clinical research. *Expert Rev. Proteomics* **17**, 719–733 (2020).
- Xue, L. et al. Solid-state nanopore sensors. *Nat. Rev. Mater.* **5**, 931–951 (2020).
- Mayer, S. F., Cao, C. & Dal Peraro, M. Biological nanopores for single-molecule sensing. *iScience* **25**, 104145 (2022).
- Ayub, M. & Bayley, H. Engineered transmembrane pores. *Curr. Opin. Chem. Biol.* **34**, 117–126 (2016).
- Howorka, S. Building membrane nanopores. *Nat. Nanotechnol.* **12**, 619–630 (2017).
- Scott, A. J. et al. Constructing ion channels from water-soluble  $\alpha$ -helical barrels. *Nat. Chem.* **13**, 643–650 (2021).
- Xu, C. et al. Computational design of transmembrane pores. *Nature* **585**, 129–134 (2020).
- Berhanu, S. et al. Sculpting conducting nanopore size and shape through de novo protein design. *Science* **385**, 282–288 (2024).
- Goverde, C. A. et al. Computational design of soluble and functional membrane protein analogues. *Nature* **631**, 449–458 (2024).
- Huang, G., Willems, K., Soskine, M., Wloka, C. & Maglia, G. Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores. *Nat. Commun.* **8**, 935 (2017).
- Piguet, F. et al. Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* **9**, 966 (2018).
- Chavis, A. E. et al. Single molecule nanopore spectrometry for peptide detection. *ACS Sens.* **2**, 1319–1328 (2017).
- Afshar Bakshloo, M. et al. Discrimination between  $\alpha$ -synuclein protein variants with a single nanometer-scale pore. *ACS Chem. Neurosci.* **14**, 2517–2526 (2023).
- Asandei, A., Rossini, A. E., Chinappi, M., Park, Y. & Luchian, T. Protein nanopore-based discrimination between selected neutral amino acids from polypeptides. *Langmuir* **33**, 14451–14459 (2017).
- Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* **32**, 179–181 (2014).
- Huang, G. et al. PlyAB nanopores detect single amino acid differences in folded haemoglobin from blood. *Angew. Chem. Int. Ed. Engl.* **134**, e202206227 (2022).
- Ensslen, T., Sarthak, K., Aksimentiev, A. & Behrends, J. C. Resolving isomeric posttranslational modifications using a biological nanopore as a sensor of molecular shape. *J. Am. Chem. Soc.* **144**, 16060–16068 (2022).
- Versloot, R. C. A. et al. Quantification of protein glycosylation using nanopores. *Nano Lett.* **22**, 5357–5364 (2022).
- Restrepo-Pérez, L., Wong, C. H., Maglia, G., Dekker, C. & Joo, C. Label-free detection of post-translational modifications with a nanopore. *Nano Lett.* **19**, 7957–7964 (2019).
- Martin-Baniandres, P. et al. Enzyme-less nanopore detection of post-translational modifications within long polypeptides. *Nat. Nanotechnol.* **18**, 1335–1340 (2023).
- Nova, I. C. et al. Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nat. Biotechnol.* **42**, 710–714 (2023).
- Huo, M., Hu, Z., Ying, Y. & Long, Y. Enhanced identification of tau acetylation and phosphorylation with an engineered aerolysin nanopore. *Proteomics* **22**, e2100041 (2021).
- Lan, W.-H., He, H., Bayley, H. & Qing, Y. Location of phosphorylation sites within long polypeptide chains by binder-assisted nanopore detection. *J. Am. Chem. Soc.* **146**, 24265–24270 (2024).
- Kang, X.-F., Cheley, S., Guan, X. & Bayley, H. Stochastic detection of enantiomers. *J. Am. Chem. Soc.* **128**, 10684–10685 (2006).
- Abraham Versloot, R. C. et al. Seeing the invisibles: detection of peptide enantiomers, diastereomers, and isobaric ring formation in lanthipeptides using nanopores. *J. Am. Chem. Soc.* **145**, 18355–18365 (2023).
- Wang, J. et al. Identification of single amino acid chiral and positional isomers using an electrostatically asymmetric nanopore. *J. Am. Chem. Soc.* **144**, 15072–15078 (2022).
- Ouldali, H. et al. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
- Wang, K. et al. Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat. Methods* **21**, 92–101 (2024).

38. Zhang, M. et al. Real-time detection of 20 amino acids and discrimination of pathologically relevant peptides with functionalized nanopore. *Nat. Methods* **21**, 609–618 (2024).
39. Zhang, Y. et al. Peptide sequencing based on host-guest interaction-assisted nanopore sensing. *Nat. Methods* **21**, 102–109 (2024).
40. 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).
41. Motone, K. et al. Multi-pass, single-molecule nanopore reading of long protein strands. *Nature* **633**, 662–669 (2024).
42. Cardozo, N. et al. Multiplexed direct detection of barcoded protein reporters on a nanopore array. *Nat. Biotechnol.* **40**, 42–46 (2022).
43. Gu, L.-Q., Cheley, S. & Bayley, H. Electroosmotic enhancement of the binding of a neutral molecule to a transmembrane pore. *Proc. Natl Acad. Sci. USA* **100**, 15498–15503 (2003).
44. Soskine, M. et al. An engineered ClyA nanopore detects folded target proteins by selective external association and pore entry. *Nano Lett.* **12**, 4895–4900 (2012).
45. Firnkes, M. et al. Electrically facilitated translocations of proteins through silicon nitride nanopores: conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* **10**, 2162–2167 (2010).
46. Li, S., Cao, C., Yang, J. & Long, Y. Detection of peptides with different charges and lengths by using the aerolysin nanopore. *ChemElectroChem* **6**, 126–129 (2019).
47. Bakshloo, M. A. et al. Polypeptide analysis for nanopore-based protein identification. *Nano Res.* **15**, 9831–9842 (2022).
48. Versloot, R. C. A., Straathof, S. A. P., Stouwie, G., Tadema, M. J. & Maglia, G.  $\beta$ -barrel nanopores with an acidic-aromatic sensing region identify proteinogenic peptides at low pH. *ACS Nano* **16**, 7258–7268 (2022).
49. Huang, G., Voet, A. & Maglia, G. FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution. *Nat. Commun.* **10**, 835 (2019).
50. Pastoriza-Gallego, M. et al. Dynamics of unfolded protein transport through an aerolysin pore. *J. Am. Chem. Soc.* **133**, 2923–2931 (2011).
51. Oukhaled, G. et al. Unfolding of proteins and long transient conformations detected by single nanopore recording. *Phys. Rev. Lett.* **98**, 158101 (2007).
52. Yu, L. et al. Unidirectional single-file transport of full-length proteins through a nanopore. *Nat. Biotechnol.* **41**, 1130–1139 (2023).
53. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* **8**, 288–295 (2013).
54. Rodriguez-Larrea, D. & Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat. Commun.* **5**, 4841 (2014).
55. Pastoriza-Gallego, M. et al. Evidence of unfolded protein translocation through a protein nanopore. *ACS Nano* **8**, 11350–11360 (2014).
56. Sauciuc, A., Morozzo della Rocca, B., Tadema, M. J., Chinappi, M. & Maglia, G. Translocation of linearized full-length proteins through an engineered nanopore under opposing electrophoretic force. *Nat. Biotechnol.* **42**, 1275–1281 (2023).
57. Lucas, F. L. R., Versloot, R. C. A., Yakovlieva, L., Walvoort, M. T. C. & Maglia, G. Protein identification by nanopore peptide profiling. *Nat. Commun.* **12**, 5795 (2021).
58. Afshar Bakshloo, M. et al. Nanopore-based protein identification. *J. Am. Chem. Soc.* **144**, 2716–2725 (2022).
59. Astier, Y., Braha, O. & Bayley, H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J. Am. Chem. Soc.* **128**, 1705–1710 (2006).
60. Helbig, A. O. & Tholey, A. Exopeptidase assisted N- and C-terminal proteome sequencing. *Anal. Chem.* **92**, 5023–5032 (2020).
61. Bonini, A., Sauciuc, A. & Maglia, G. Engineered nanopores for exopeptidase protein sequencing. *Nat. Methods* **21**, 16–17 (2024).
62. Zhang, S. et al. Bottom-up fabrication of a proteasome-nanopore that unravels and processes single proteins. *Nat. Chem.* **13**, 1192–1199 (2021).
63. de Lannoy, C., Lucas, F. L. R., Maglia, G. & de Ridder, D. In silico assessment of a novel single-molecule protein fingerprinting method employing fragmentation and nanopore detection. *iScience* **24**, 103202 (2021).
64. Yan, S. et al. Single molecule ratcheting motion of peptides in a *Mycobacterium smegmatis* porin A (MspA) nanopore. *Nano Lett.* **21**, 6703–6710 (2021).
65. Chen, Z. et al. Controlled movement of ssDNA conjugated peptide through *Mycobacterium smegmatis* porin A (MspA) nanopore by a helicase motor for peptide sequencing application. *Chem. Sci.* **12**, 15750–15756 (2021).
66. Chen, X. et al. Resolving sulfation PTMs on a plant peptide hormone using nanopore sequencing. *ACS Nano* **18**, 28999–29007 (2024).
67. Sauciuc, A. & Maglia, G. Controlled translocation of proteins through a biological nanopore for single-protein fingerprint identification. *Nano Lett.* **24**, 14118–14124 (2024).
68. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an  $\alpha$ -hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–250 (2013).
69. Nivala, J., Mulroney, L., Li, G., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* **8**, 12365–12375 (2014).
70. Fei, X. et al. Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *eLife* **9**, e52774 (2020).
71. Oguro, T. et al. Structural stabilities of different regions of the titin I27 domain contribute differently to unfolding upon mitochondrial protein import. *J. Mol. Biol.* **385**, 811–819 (2009).
72. Gershoff, M. & Golovchenko, J. A. Recapturing and trapping single molecules with a solid-state nanopore. *Nat. Nanotechnol.* **2**, 775–779 (2007).
73. Shadish, J. A. & DeForest, C. A. Site-selective protein modification: from functionalized proteins to functional biomaterials. *Matter* **2**, 50–77 (2020).
74. Toplak, A. et al. From thiol-subtilisin to omniligase: design and structure of a broadly applicable peptide ligase. *Comput. Struct. Biotechnol. J.* **19**, 1277–1287 (2021).
75. Reiner, J. E. et al. The effects of diffusion on an exonuclease/nanopore-based DNA sequencing engine. *J. Chem. Phys.* **137**, 214903 (2012).
76. Branton, D. & Deamer, D. *Nanopore Sequencing* (World Scientific, 2018).
77. Oxford Nanopore Technologies. *Company History* [nanoporetech.com/about/history](http://nanoporetech.com/about/history) (2025).
78. Kasianowicz, J. J., Brandin, E., Branton, D., Deamer, D. W. & David, W. D. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).

79. Ashkenasy, N., Sanchez-Quesada, J., Bayley, H. & Ghadiri, M. R. Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopores. *Angew. Chem. Int. Ed. Engl.* **44**, 1401–1404 (2005).
80. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl Acad. Sci. USA* **106**, 7702–7707 (2009).
81. Cockroft, S. L., Chu, J., Amorin, M. & Ghadiri, M. R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J. Am. Chem. Soc.* **130**, 818–820 (2008).
82. Lieberman, K. R. et al. Processive replication of single DNA molecules in a nanopore catalyzed by Φ29 DNA polymerase. *J. Am. Chem. Soc.* **132**, 17961–17972 (2010).
83. Craig, J. M. et al. Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers. *Proc. Natl Acad. Sci. USA* **114**, 11932–11937 (2017).
84. Derrington, I. M. et al. Nanopore DNA sequencing with MspA. *Proc. Natl Acad. Sci. USA* **107**, 16060–16065 (2010).
85. Oxford Nanopore Technologies. *Nanopore Sequencing Accuracy* <https://nanoporetech.com/platform/accuracy> (2025).
86. Oxford Nanopore Technologies. *Proteomic Methods at Oxford Nanopore Technology* <https://nanoporetech.com/oxford-nanopore-proteomics> (2025).
87. Reed, B. D. et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science* **378**, 186–192 (2022).
88. Egertson, J. D. et al. A theoretical framework for proteome-scale single-molecule protein identification using multi-affinity protein binding reagents. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.10.11.463967> (2021).
89. Aksel, T. et al. Highly dense and scalable protein arrays for single-molecule studies. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.05.02.490328> (2022).

## Acknowledgements

C.L., A.B., J.H.V. and G.M. acknowledge the funders NWO-VICI (grant number 192.068), NWO-XS (grant number OCENW.XS24.1.119) and NHGRI of the NIH (grant number 1R01HG012554) for support of this work.

## Author contributions

All authors contributed equally to this work.

## Competing interests

G.M. is a founder, director and shareholder of Portal Biotech, a company engaged in the development of nanopore technologies.

## Additional information

**Correspondence** should be addressed to Giovanni Maglia.

**Peer review information** *Nature Biotechnology* thanks Jeff Nivala, Abdelghani Oukhaled and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2025