Single-molecule protein sequencing with nanopores

Ritmejeris, Justas; Chen, Xiuqi; Dekker, Cees

DOI
10.1038/s44222-024-00260-8

Publication date
2024

Document Version
Final published version

Published in
Nature Reviews Bioengineering

nature reviews bioengineering

https://doi.org/10.1038/s44222-024-00260-8

Review article

Check for updates

Single-molecule protein sequencing with nanopores

Justas Ritmejeris [1,2], Xiuqi Chen [1,2] & Cees Dekker [1]

Abstract

Protein sequencing and the identification of post-translational modifications are key to understanding cellular signalling pathways and metabolic processes in health and disease. Nanopores, that is, nanometre-sized holes in a membrane, were previously put to use for DNA and RNA sequencing, offering single-molecule sensitivity and long read lengths. This prompted efforts to engineer nanopores for the high-throughput sequencing of peptides and proteins. In this Review, we discuss research towards single-molecule protein sequencing using biological nanopores, investigating how their sensitivity allows the discrimination of all 20 amino acids. We outline how fingerprinting of proteins is facilitated by using motor proteins and electro-osmotic flow to promote the slow translocation of proteins through nanopores. Moreover, we examine applications of nanopores to the identification of post-translational modifications, highlighting the potential of this technology for fundamental and clinical proteomic studies. Finally, we outline the advantages and limitations of nanopore systems for protein sequencing and the challenges that remain to be overcome for realizing de novo nanopore protein sequencing.

Sections

[1]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft, the Netherlands. [2]These authors contributed equally: Justas Ritmejeris, Xiuqi Chen. e-mail: C.Dekker@tudelft.nl

Key points

- Twenty individual amino acids can be distinguished by biological nanopores.
- Nanopores allow the discrimination of single-amino acid substitutions within proteins.
- Electro-osmotic flow allows uncharged and heterogeneously charged proteins to translocate through nanopores.
- Motor proteins reduce protein translocation speed, resulting in high-resolution signals.
- Various post-translational modifications and their locations can be identified with nanopores.

Introduction

Nanopore sequencing is a single-molecule technique that was first introduced for the reading of nucleic acids[1–8]. Its high sensitivity is rooted in the detection of partial ion-flow disruptions caused by local molecular differences as a linear biopolymer, such as DNA, slowly passes through a nanopore. This principle can be leveraged in multiple ways to achieve precise discrimination of the sequence, for example, through the identification of cleaved nucleobases[9], DNA-by-synthesis[10,11] and base-by-base sequencing of DNA or RNA traversing through the nanopore at a controlled slow speed[12,13].

The pioneering achievements of DNA sequencing with nanopore technology catalysed the exploration of similar sequencing efforts for proteomics[14–16]. Nanopore-based protein sequencing approaches face distinct challenges owing to the intricate structural characteristics of proteins[17]. In contrast to distinguishing four nucleotides, protein sequencing requires the discrimination of 20 amino acids with considerable variability in local properties, such as charge and hydrophobicity,[18] and a 3.4-fold volume difference between the largest and the smallest amino acid[19]. Their diverse chemical nature allows various non-covalent interactions within the protein as well as with the nanopore[8,20]. In addition, a huge repertoire of post-translational modifications (PTMs) are added to proteins in cells, which contributes to their functional versatility[21] but also pose additional challenges for nanopore-based analysis.

Protein amino acid sequencing and determination of PTMs are typically achieved by mass spectrometry[22]. Here, after purification and fractionation of a protein into peptide fragments, the mass-to-charge ratios of ionized peptide fragments can be sensitively measured; from these data, amino acid sequences can be inferred using tandem mass spectrometry. However, these bulk measurements require a sizeable amount of material. Furthermore, isobaric peptides (that is, peptides with identical mass) carrying combinatorial PTMs are technically challenging, if not impossible, to distinguish using this approach, and delicate PTMs, such as glycosylation or sulfation, can be labile during the ionization process[23–25].

Alternatively, nanopore-based protein sequencing affords sensitivity at the single-molecule level, with the potential for high-throughput analysis[26,27]. Nanopores come in a variety of types, from drilled holes in solid-state membranes[28] to protein nanopores in a lipid membrane[29]. In particular, advances in solid-state and synthetic nanopore technology have contributed to protein sequencing by offering enhanced stability, tunability and durability[30–32]. Materials such as glass[33], silicon nitride[34], molybdenum disulfide[35], graphene[36] and DNA origami[37] have been used to fabricate nanopores capable of detecting various analytes[38–42]. However, the size and shape of such solid-state nanopores is less defined and less reproducible than those of their biological counterparts, and they often exhibit higher noise levels[43,44]. Therefore, mainly biological nanopores have been explored for high-resolution single-molecule protein sequencing, owing to their consistent atomic-scale-defined geometry, high signal-to-noise ratio and ability to precisely control analyte movement[45].

In this Review, we discuss protein sequencing by biological nanopores, outlining how nanopores can be implemented to enable individual peptides or proteins to translocate in an increasingly consistent and discernible manner. For example, charges on the inner walls of the nanopore can induce an electro-osmotic flow that moves a liquid containing an analyte through the pore. This facilitates the translocation of neutrally or oppositely charged proteins, and overcomes limitations of electrophoretic force approaches[46–48], in which charged molecules are directly driven by an applied electric field to move through the nanopore. Furthermore, motor proteins can be integrated in nanopore systems for controlled, slow and high-precision protein translocation[49–52]. Finally, we survey key applications of nanopore protein sequencing, including the detection of PTMs at the single-molecule level.

Nanopore-based protein sequencing

Nanopore measurements

In a typical nanopore measurement, a biological nanopore in a lipid bilayer connects two isolated compartments (commonly referred to as cis and trans) filled with electrolyte solutions. Under an electrical field, an active ion flow through the nanopore can be recorded as an electrical signal in the picoampere (pA) range (Fig. 1a). If the cis compartment is supplemented with analyte molecules, spikes in the electrical signal appear, owing to analyte biopolymers partially blocking the ion flow through the nanopore after their traversal (Fig. 1b), yielding characteristic current levels and dwell times (that is, the length of time during which a molecule or a segment remains within the nanopore during translocation). The translocation of the molecule depends on its charge properties, as charged molecules experience an electrophoretic force that attracts them towards the corresponding electrode, enabling translocation of the analyte through the pore. Negatively charged DNA molecules translocate with a speed of ~1 nt μs^{-1} (refs. 53,54), which is too fast to resolve individual bases. Therefore, high-precision DNA sequencing necessitates slower translocation speeds to improve the signal-to-noise ratio[5,43], which can be achieved using motor enzymes that slow down DNA translocation to a time scale of (sub)seconds[6,12,13,55]. In protein sequencing, the heterogeneous charge distribution and the presence of secondary and tertiary structures complicate translocation[31], requiring strategies such as capturing methods, application of an electro-osmotic flow or motor enzymes (Fig. 1c).

Nanopore proteins

Various biological membrane proteins can be engineered to aid in nanopore protein sequencing[49,52,56] (Box 1). Typically, hexameric to nonameric homo-oligomers constitute ion channels, which are 5–10 nm in height, with their narrowest pore constriction measuring 1–2 nm in diameter[57–60]. To add specificity for protein sequencing, modifications on both the analytes and the nanopore proteins can boost capture rates and foster specific chemical interactions, thereby prolonging the

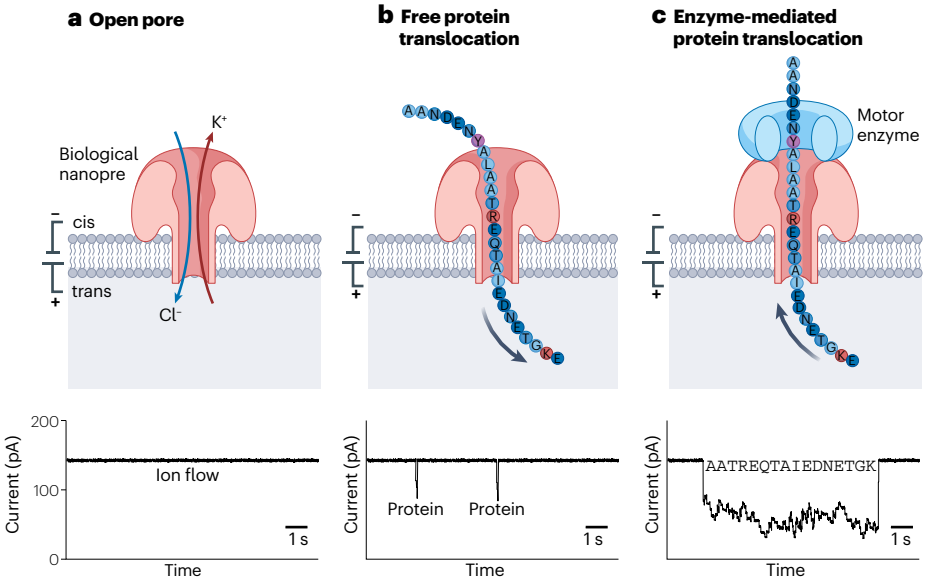


Fig. 1 | Nanopore-based protein sequencing. **a**, A biological nanopore is inserted in a synthetic or biological membrane that separates two electrolyte compartments. A voltage is applied across the membrane, causing an active ion flow through the nanopore. **b**, Protein translocations result in sudden ion current dips as proteins pass through the pore. **c**, A motor enzyme capable of walking on the protein strand slows down protein movement through the pore, resulting in a longer-duration signal with distinct ion-current changes, which can be assigned to specific parts of the protein sequence.

dwell time of the analyte within the pore and improving the detection fidelity[47,54,61,62]. Biological pores are typically composed of multiple subunits. Introducing a single mutation results in changes in all subunits, which considerably alters the pore's dimensions and biophysical characteristics[63,64]. For example, a *Mycobacterium smegmatis* porin A (MspA)-M2 pore mutant with six targeted mutations in each of its eight monomeric subunits is characterized by considerable improvements in pore stability and in the detection rate of single-stranded DNA[54]. Precise control of subunits can also be beneficial for the detection of other analytes; for example, by varying the numbers of Fragaceatoxin C (FraC) nanopore subunits, single-amino acid differences can be measured in an angiotensin peptide at low pH (ref. 65). A fusion MspA-M2 protein that encodes all pore subunits within a single gene benefits from precise control of the number of subunits and mutations across the protein[66]. This fusion protein approach enables the design of nanopores with heteromeric composition by precisely introducing point mutations in separate nanopore subunits.

Identifying 20 proteinogenic amino acids

The sensitivity of nanopores can be used to discriminate all 20 proteinogenic amino acids. For example, individual amino acids bound to an arginine heptapeptide carrier can be detected using a wild-type aerolysin pore; here, the carrier prolongs entrapment of the peptide in the nanopore to a length of time that is necessary for measuring the blocking current, which is slightly different for different amino acids[67] (Fig. 2a). According to an electrostatic potential map computed by all-atom molecular dynamics simulations, a constant potential compartment in the middle of the aerolysin stem allows the charged carrier to reside inside the pore for an extended period of time. Using this approach, 13 of the 20 amino acids were directly distinguished, whereas the remaining seven amino acids appear in two clusters (Q, M, H, I, Y and E, V). Of note, the positive correlation between the magnitude of the ionic current reduction and the volume of the analyte is in agreement with other studies[68–70]. Several amino acids, however, do not comply with this trend; for example, phenylalanine (F), tyrosine (Y) and arginine (R) have nearly identical volumes but distinguishably different current levels. Vice versa, amino acids with different volumes produce nearly indistinguishable current signals (Q, M, H, I and Y). Similar to DNA sequencing[71], these observations indicate that the ion current is not only affected by volume exclusion but also by non-covalent interactions with the nanopore or conformational changes of the analyte within the pore[19].

The formation of a host–guest complex to a peptide carrier with an individual amino acid of interest allows accurate recognition of the 20 proteinogenic amino acids using the α-hemolysin (αHL) nanopore[72] (Fig. 2b); here, owing to the FG-X-D_8 peptide, in which the N-terminal phenylalanine (F) forms a complex with a cucurbit[7]uril moiety in the pore constriction, the amino acid of interest can be oriented at the constriction site, resulting in its prolonged capture. Using a wild-type αHL pore, this method enables the discrimination of nine amino acids, whereas the remaining 11 amino acids are overlapping in five different subgroups. The distinguishing power within subgroups was improved by applying an M113F αHL mutant, allowing the separation of the remaining 11 amino acids. Furthermore, peptide identification was achieved through stepwise digestion and detection of constituent amino acids of the peptide. Terminal amino acids from the peptide were cleaved and incorporated into a modified carrier FGGC(-X)D_8, in which the amino acid of interest was measured and identified. The sequence of the peptide can thus be deduced after multiple rounds of this process, requiring a new conjugate to be prepared for each amino acid. Using the FGGC(-X)D_8 carrier, all 20 amino acids can be identified by taking advantage of the structural and biophysical properties of wild-type αHL, two additional nanopore mutants (M113G and M113R), and different pH conditions for specific amino acids. Although this approach achieves nearly 100% amino acid detection accuracy during the stepwise digestion and measurement of several model peptides, it is cumbersome to implement for nanopore-based protein analysis.

Individual amino acids also were directly measured using a heteromeric MspA nanopore mutant consisting of seven M2 subunits and a single N90C mutant monomer, which introduces a strategically placed cysteine residue at the constriction site[62]. The cysteine is modified with nitrilotriacetic acid, which chelates Ni^{2+} ions using

cysteine–maleimide conjugation (Fig. 2c). Individual amino acids that contain both an amino and a carboxyl group are chelating ligands that reversibly interact with Ni^{2+} at the constriction site[62]. The interaction between the MspA-nitrilotriacetic acid-Ni complex and single amino acid molecules results in distinct current signals for different amino acids, achieving a detection accuracy of 96% for all 20 amino acids using a machine learning-based classifier, based on the mean and standard deviation of the signals. To further improve the accuracy, a five-dimensional matrix consisting of mean ion current, standard deviation, skewness (a measure of the asymmetry of the probability distribution), kurtosis (a measure of the 'tailedness' of the probability distribution) and dwell time of the interaction signals was tested with different classifiers. The best-performing classifier can distinguish the 20 amino acids with 98.8% accuracy. This approach was applied to analyse commercially available food supplement tablets with nine amino acids. In addition, the amino acid composition of synthetic peptides (GHK and TLEIYNRF) was identified through stepwise digestion and measurement. Although limited to single-amino acid sensing, this method provides unambiguous discrimination of all 20 amino acids, establishing a foundation for a range of analytical applications. Moreover, such machine learning-based data analysis highlights the amount of information that can be extracted from ion current measurements with nanopores.

In a similar approach, an N91H substitution was introduced in MspA-M2, resulting in the incorporation of Cu^{2+} ions in eight histidine residues at the constriction site[61], which interact with individual amino acids at both their amino and carboxyl groups. By implementing a machine-learning-based classifier, a high identification accuracy (>95%) was achieved among the 20 amino acids, whereas positively charged arginine and lysine residues were more difficult to distinguish. This nanopore also allows the 'real-time' detection of amino acids in a model peptide using carboxypeptidase A1, which cleaves individual amino acids from the C terminus in situ; here, amino acids more proximate to the C terminus are (on average) detected earlier. Consequently, this strategy can be implemented for the identification of peptides; for example, point mutations in a melanoma neoantigen and Alzheimer's disease-related point mutations in a β-amyloid peptide were recognized by identifying new amino acids arising from the mutation[61].

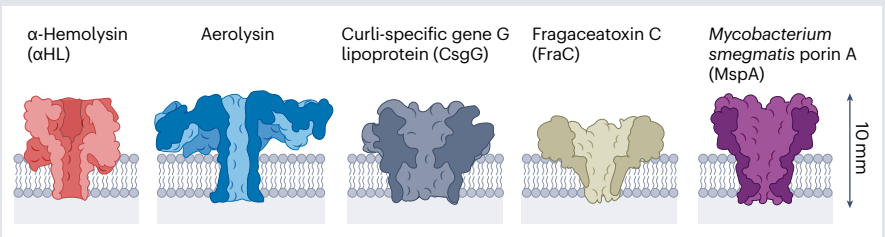
Therefore, nanopores can be engineered and their specificity improved using molecular dynamics simulations and machine-learning classifiers to discriminate all proteinogenic amino acids, for example, for peptide fingerprinting. However, protein sequencing remains challenging and requires carrier-assisted identification and thus additional steps to tag each amino acid, which is limited by varying coupling efficiencies and may introduce detection bias[73–75]. Moreover, for single-amino acid approaches, digesting, translocating and detecting individual amino acids in the correct order remains difficult, because hydrolysis of amino acids occurs in bulk, and their sequential translocation remains impossible, making the identification of complex samples with diverse peptides impossible. More precise control over amino acid cleavage and translocation processes will be crucial to advance this approach to protein sequencing.

Translocation of proteins by electro-osmotic flow

A key challenge for nanopore sequencing of proteins is their heterogeneous charge distribution. Under a consistent voltage configuration, only positively charged or only negatively charged proteins translocate, based on the electrophoretic force[76]. However, regions of opposing charge, mixed or neutral charge can considerably hinder sample translocation[76]. The translocation of short peptides with different charges can be promoted by introducing a charged threading tail (DNA, polyD or polyR) that drags the peptide through the pore[46,51,67,72]. Translocation of longer polypeptides and proteins can be established by adding charged amino acids across the entire protein[52]. However, a general strategy to translocate molecules independent of their native charge remains elusive.

Instead of relying on the electrophoretic driving force, proteins with neutral or opposing charges can also be translocated using electro-osmotic flow, that is, by taking advantage of fluid transport across a nanopore. Here, the inner charged nanopore surface attracts

Box 1 | The structure of biological nanopores in sequencing applications



The efficacy of nanopore sequencing is intimately linked to the architecture of the nanopore. Biological nanopores should maintain a stable, non-gating structure, with a defined sensing region that is situated in the narrowest part of the nanopore, often referred to as the constriction site. As the translocating biopolymer alters the ion current at the constriction site, the geometry of this area is crucial for sequencing sensitivity[5]. The diameter of the constriction site should thus range between 1.2 nm and 1.6 nm, and the height of the nanopore should be ~10 nm. The overall shape of nanopores typically varies depending on the type of pore (see figure). For example, the α-hemolysin (αHL)[57] nanopore has the form of a 5 nm-long barrel, whereas the aerolysin nanopore[58] has an elongated 10 nm-long cylindrical profile. By contrast, the funnel-shaped *Mycobacterium smegmatis* porin A (MspA) nanopore features a 0.6 nm-long constriction[59] and the curli-specific gene G lipoprotein (CsgG) nanopore has an hourglass-shaped pore with a 2 nm-long constriction[60]. The V-shaped Fragaceatoxin C (FraC) nanopore has a slightly more gradual profile with a ~3 nm-long constriction. Within longer constrictions, the ion current signal is modulated by ~20–40 amino acids[53,133], whereas, in short constrictions, the ion current signal is modulated by ~6–12 amino acids[6,12]. Notably, a longer sensing region accommodates a larger biopolymer segment, which enables higher sensitivity and potentially extends dwell times, but also complicates signal interpretation. Therefore, an abrupt short constriction is typically preferred. Unravelling convolution patterns of translocating proteins is often challenging, owing to their irregular charge distribution and additional interactions with the pore[110]. The figure is adapted with permission from ref. 29, Cell Press.

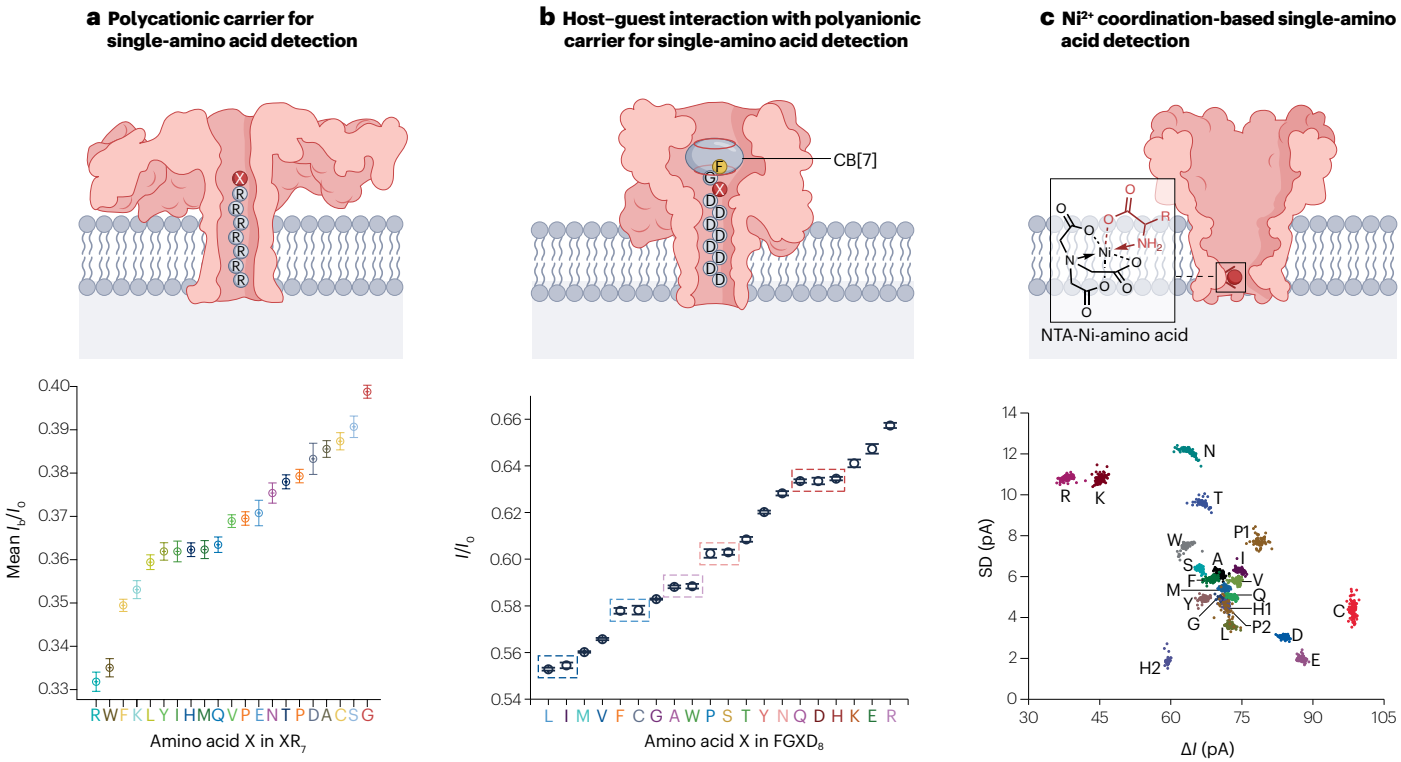


Fig. 2 | Single amino acids can be discriminated using nanopores. a, Amino acids connected to a polycationic carrier can be inserted into a wild-type aerolysin nanopore. The mean I_b/I_0 (ratio of the ionic current (I_b) when a molecule is inside the pore to the open-pore current (I_0) in an empty nanopore) values of all 20 amino acids can be measured with this system[67]. **b**, In an α-hemolysin (αHL) nanopore, an amino acid connected to a polyanionic carrier engages in a host–guest interaction with cucurbit[7]uril (CB[7]) at the constriction site. Mean I/I_0 values of all 20 amino acids can be measured with this system[72]. **c**, A *Mycobacterium smegmatis* porin A (MspA) mutant with a nitrilotriacetic acid (NTA) modification forms an NTA-Ni^{2+} complex, in which Ni^{2+} interacts with amino and carboxyl groups on the amino acid. The standard deviation (SD) of the signal over ion current change (ΔI) is shown for all 20 amino acids[62]. Part **a** adapted from ref. 67, Springer Nature Limited. Part **b** adapted from ref. 72, Springer Nature Limited. Part **c** adapted from ref. 62, Springer Nature Limited.

counterions that are driven by the electrical field and that drag the surrounding fluid along, resulting in a net flow of the liquid in the direction of the moving ions[77–79], thereby facilitating directional translocation of analyte proteins regardless of their charges. Some naturally occurring nanopore proteins (for example, the wild-type MspA nanopore) exhibit a base level of electro-osmotic flow but are limited by low stability, molecular selectivity and high noise levels[54]. Alternatively, pores with a strong electro-osmotic flow can be designed to translocate proteins even against the electrophoretic force[46–48], for example, by covering the pore surface with 'sticky ions'[46,80,81] (Fig. 3a) or by introducing charged amino acid mutations at the inside walls of the pore[47,48,82,83] (Fig. 3b,c). Both approaches result in ion charge accumulation inside the nanopore, which subsequently causes a net flux of water through the nanopore[78].

To discriminate different proteins, an electro-osmotic trapping approach was applied inside an MspA-M2 mutant nanopore[80]. Although the mutant nanopore exhibits weak electro-osmotic flow in the opposite direction to the electric field[78], the directionality of the electro-osmotic flow can be reversed by applying asymmetric KCl/$CaCl_2$ buffer conditions, resulting in a tenfold increase in protein capture rate[80]. Accumulation of Ca^{2+} cations influences the charge profile of the MspA lumen, generating a strong electro-osmotic flow to temporarily trap folded proteins inside the nanopore's vestibule. Interestingly, this approach only works if Ca^{2+} ions are added to the opposite side of the analyte, as Ca^{2+} may also coat the surface of the analyte, which could cause an opposing electrophoretic force that prevents analyte translocation. Using this method, acidic proteins (thyroid hormone and retinoid receptors, as well as the nuclear co-activator binding domain complex), basic proteins (lysozyme) and neutral proteins (apo-myoglobin and holo-myoglobin) were discriminated, achieving high protein identification accuracy (99.9%), assisted by machine learning[80]. Although only shown for distinct proteins, this high identification accuracy is promising for detecting other protein biomarkers.

Guanidinium ion (Gdm^+)-induced charge selectivity was applied to generate an electro-osmotic flow in an αHL nanopore owing to the binding of Gdm^+ ions to negatively charged amino acids, which results in a layer of positive charges in the nanopore lumen. In denaturing conditions, this approach allows directional translocation of large proteins[46] (Fig. 3a); for example, a heterogeneously charged maltose-binding protein and green fluorescent protein were detected. However, analyte capture requires the addition of a charged tail made of 10 aspartate residues (D_{10}) or 20 deoxythymidine residues (dT_{20}) to initiate the threading process. By tagging the tail to the C terminus or N terminus, the orientation of the translocating molecule can be controlled. Gdm^+ accumulation occurs at the inner nanopore surface, near the αHL vestibule and the end of the stem[46], as shown by all-atom molecular dynamics simulations. Molecular dynamics simulations of the

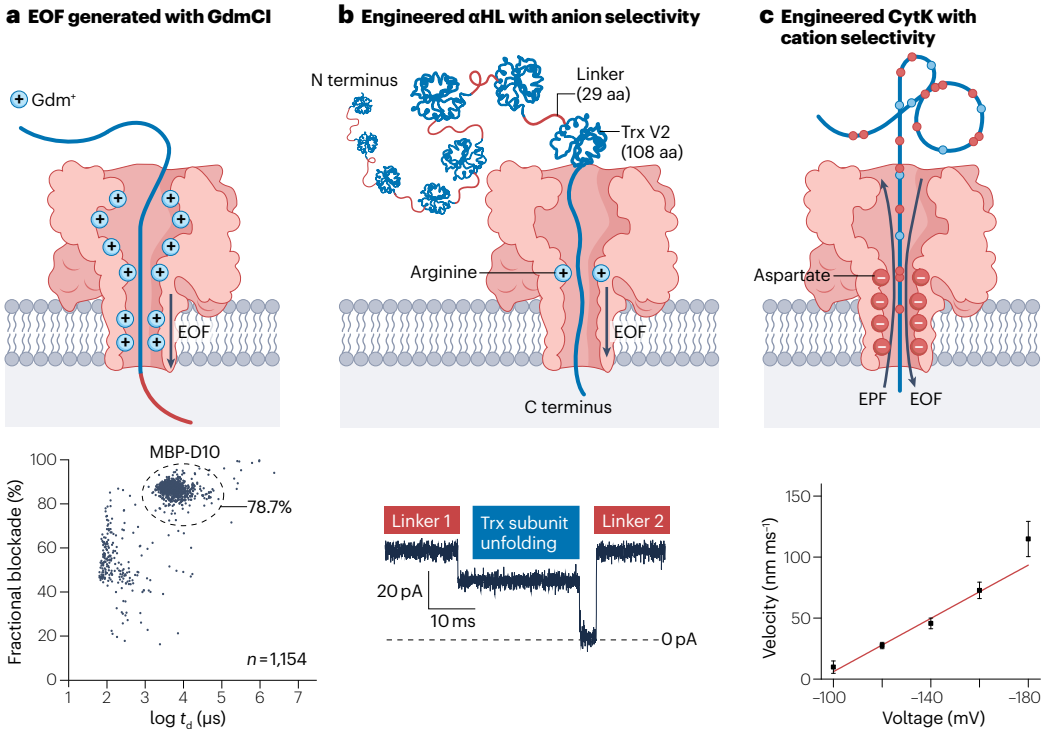


Fig. 3 | Strategies to introduce electroosmotic flow in nanopores. a, An electro-osmotic flow (EOF) can be generated in an α-hemolysin (αHL) nanopore using guanidinium chloride (GdmCl). Guanidinium (Gdm^+) ions cover the vestibule of the pore and the end of the stem as 'sticky ions'. The resulting EOF and denaturing conditions allow the translocation of heterogeneously charged unfolded proteins. The scatter plot shows the protein-dependent blockade level versus the dwell time (t_d), in microseconds (μs), of maltose-binding protein conjugated with ten aspartate residues (MBP-D10)[46]. **b**, By mutating the constriction site in an αHL nanopore with positively charged arginine, an EOF can be generated to unfold and translocate mixed-charged thioredoxin (Trx) concatemers. The graph shows the ion current signals corresponding to translocation of the linker, and unfolding and translocation of the Trx subunit, as well as translocation of the next linker[47]. **c**, By mutating the β-barrel of cytotoxin K (CytK) with negatively charged aspartate groups, an EOF can be generated to translocate proteins that are opposing the electrophoretic flow (EPF). The graph shows the linear dependence of the translocation velocity on the applied voltage[48]. aa, amino acids. Part **a** adapted from ref. 46, Springer Nature Limited. Part **b** adapted from ref. 47, Springer Nature Limited. Part **c** adapted from ref. 48, Springer Nature Limited.

guanidinium chloride (GdmCl)-induced electro-osmotic force effect in different nanopores (αHL, curli-specific gene G lipoprotein (CsgG), MspA and aerolysin) have further shown that Gdm^+ ions strongly bind to negatively charged amino acids in the vestibule and engage in a weaker interaction with other amino acids. Notably, not only do high concentrations of GdmCl serve to generate an electro-osmotic force but also as a protein denaturant[81].

An electro-osmotic force can also be induced by strategically placing charged amino acid mutations inside the nanopores. For example, by placing a positively charged amino acid along the interior surface of a mutant αHL nanopore (NN-113R variant), an anion-selective nanopore was engineered with enhanced electro-osmosis[84]. This mutant nanopore enables the translocation of a mixed-charge polypeptide and its concatemers to discriminate different PTMs (phosphate, glutathione and 6′-sialyllactosamine)[47] (Fig. 3b), demonstrating a 25 times higher capture rate than a wild-type αHL nanopore. In addition, chaotropic agents (for example, GdmCl and urea for disrupting hydrogen bonds) were applied at non-denaturing concentrations to facilitate the unfolding of protein analytes during translocation[47].

Similarly, sets of negatively charged residues were introduced, spaced approximately 1 nm apart, inside a cytotoxin K (CytK) nanopore (structurally similar to αHL) to enhance the electro-osmotic force owing to strong cation selectivity under denaturing conditions. This force allows the directional translocation of differently charged unstructured natural polypeptides (Fig. 3c) and unfolded native proteins[48]. According to molecular dynamics simulations, the electro-osmotic force in these engineered nanopores is similar in strength to the electric force on single-stranded DNA (amounting to ~30 pN)[55], highlighting the effectiveness of this approach[48].

Although electro-osmotic force-driven translocation has the potential to translocate different types of proteins through nanopores, a quantitative understanding of this phenomenon in biological nanopores is required for rational nanopore protein engineering. For example, electro-osmotic forces have been theoretically compared in three different biological pores (αHL, CsgG and MspA) by solving the Navier–Stokes equation in the Poisson–Nernst–Planck formalism[78], showing that, above certain applied voltages, the electro-osmotic force velocity decreases owing to internal fluid instability. Similarly, it was experimentally observed that the dwell time of proteins increases above a certain voltage[82]. Electro-osmotic forces in engineered nanopores can typically be generated in mild salt concentrations, in which motor enzymes, used for nanopore sequencing, can

function[47,48]. In this environment, electro-osmotic force could also be combined with motor-assisted protein translocation to achieve better control of stretching and translocating the analyte[49,51,52,85]. Of note, electro-osmotic forces generated by high concentrations of Ca^{2+} or Gdm^{+} (refs. 46,80) create denaturing conditions that may prevent the function of motor enzymes.

Motor-mediated protein translocation

To measure protein signals at high resolution, protein translocation can be slowed down by motor enzymes that induce a stepwise motion of the protein through the pore, thereby extending the dwell time at each translocation step and enabling the discrimination of minor changes in the protein sequence[49,52,86]. The ideal motor enzyme would feature a slow stepping rate, a small step size (ideally, ~0.3 nm per step, that is, the distance between amino acids), no backstepping, a stable functionality in a range of operating conditions (buffer and pH), and a stable and strong force to disrupt folded protein structures.

DNA motor enzymes for peptide sequencing

Several motor enzymes have been explored to control peptide translocation. For example, DNA motor enzymes[49–51], such as helicases[49,50] and polymerases[51] (in particular, Hel308 and phi29 DNA polymerase[12,87]), enable sequential peptide translocation by walking along a DNA strand that is covalently linked to the end of the peptide, thereby slowly pulling the peptide through the nanopore (Fig. 4a). During the translocation of this composite polymer, a DNA signal is first obtained (as in nanopore-based DNA sequencing), followed by a signal from the peptide. Hel308 and phi29 DNA polymerase[12,87] show robust and precise stepping and have previously been optimized for DNA sequencing, with Hel308 having the advantage of a small ~0.3 nm step size[55]; by contrast, the step size of phi29 DNA polymerase is twice as large[12].

DNA helicase was applied to detect single-amino acid substitutions of glycine, aspartate or tryptophan in a negatively charged 25 amino acid-long peptide linked to single-stranded DNA[49] (Fig. 4a). Such single-site variations affect approximately eight ion current levels (four nucleotides or eight amino acids) owing to the finite constriction height and stochastic up-and-down thermal displacements of the peptide through the nanopore[88]. Therefore, individual peptide variants result in distinct current patterns, which can be discriminated with 87% single-read accuracy. In particular, the glycine variant leads to higher ion current levels because it occupies less volume at the constriction site. The variant with the bulkiest amino acid, tryptophan, initially has low current levels, indicating blockage at the constriction site; however, after the residue passes the constriction site, higher current levels appear than it is the case for the glycine and aspartate variants. All-atom molecular dynamics simulations suggest that the interaction between a hydrophobic tryptophan and the nanopore surface just above the constriction site is responsible for drawing the tryptophan variant closer to the nanopore wall, thus opening up the pore for higher ion conduction[49].

This approach can be expanded beyond single reads of the peptide. Helicase-driven peptide readings show a repeated signal pattern that resembles the last part of the read, indicating that a particular segment is read multiple times in a row[49,50,86]. This rereading phenomenon, which can be studied by adding an excess of the enzyme[49] or with dual-helicase constructs[50], occurs when the DNA-translocating motor enzyme detaches at the DNA–peptide junction as it cannot travel further along the attached peptide. If abundant amounts of motor enzymes are present, multiple motor enzymes are sequentially stacked onto the DNA, allowing the next helicase to restart the translocation process once the initial enzyme detaches. In the process, the molecule shifts down with a read length equal to the size of the motor footprint on the DNA, that is, approximately 17 amino acids. This rereading can be repeated more than 100 times, resulting in substantially improved signal quality by eliminating random errors that lead to inaccuracies. Even for poor single-read accuracy, samples can be resolved with a virtually undetectably low error rate (<1 in 10^6) when using >30 rereads of the same molecule[49].

Using DNA-translocating motors, the read length of the peptide is limited to the distance between the upper rim and the sensing region of the nanopore, which can vary between ~5 nm in CsgG[60] and αHL[57] pores to ~9 nm in MspA[59]. Therefore, the maximum sequencing length of a peptide is ~25 amino acids in an MspA nanopore. Although shorter than the length of typical proteins, this sequencing resolution allows the measurement of various peptides, including cancer neoantigens[89], immunopeptides[90] and antimicrobial peptides[91], which are typically ~10–20 amino acids-long. The use of DNA-translocating motors further provides the possibility to use the lead DNA as a barcode (a unique DNA sequence that distinguishes the individual molecule of interest) to measure different analytes in parallel, based on the sequence of the connected DNA[92]. Furthermore, similar to DNA sequencing schemes, the capture rate can be increased 100–1000-fold using complementary DNA strands containing a cholesterol group[55,93]. After inserting cholesterol into a bilayer, the analyte is being concentrated closer to the membrane, thus substantially improving the capture rate of DNA–peptide constructs[49–51,86].

Protein motor enzymes for protein sequencing

Instead of using a DNA translocase that walks on a DNA–peptide construct, a peptide translocase motor enzyme can be applied that walks directly along the peptide backbone. However, only few such protein motor enzymes have been identified thus far. For example, a 900 kDa multi-protein proteasome–nanopore complex was engineered to control the unfolding and threading of individual proteins through a nanopore[94] (Fig. 4b). This complex features a low-noise β-barrel nanopore sensor that is linked to a mammalian proteasome, and an ssrA (short peptide) tag at the C terminus of the proteins captures the analyte for translocation. This approach allows unravelling and linearly transporting a protein analyte across the nanopore using two different modes: in the 'chop-and-drop' mode, proteins are fragmented into short peptides by an active proteasome variant, and, in the 'thread-and-read' mode, the nanopore protein continuously translocates intact polypeptides with an inactive proteasome. Furthermore, this enzyme–nanopore complex can accommodate a high salt concentration (1 M NaCl), thus providing a good signal-to-noise ratio. However, its application in protein sequencing may be limited owing to the high experimental complexity and poorly understood signals. Nevertheless, combined motor and nanopore complex engineering may be promising for protein sequencing.

Motor-mediated unfolding and threading of polypeptides was first achieved with the molecular chaperone caseinolytic protease X (ClpX)[15,16]. ClpX is a hexameric ATPase that forms a ring structure around its protein substrate and acts as an unfoldase and translocase by walking on the substrate strand[95]. ClpX was used to control protein translocation through a CsgG nanopore[52] (Fig. 4c) to unfold and thread a protein for sequential analysis. In the commercially available portable nanopore sequencing device MinION, which allows high-throughput sequencing by running hundreds of pores in parallel, proteins are first electrophoretically driven into the nanopore owing

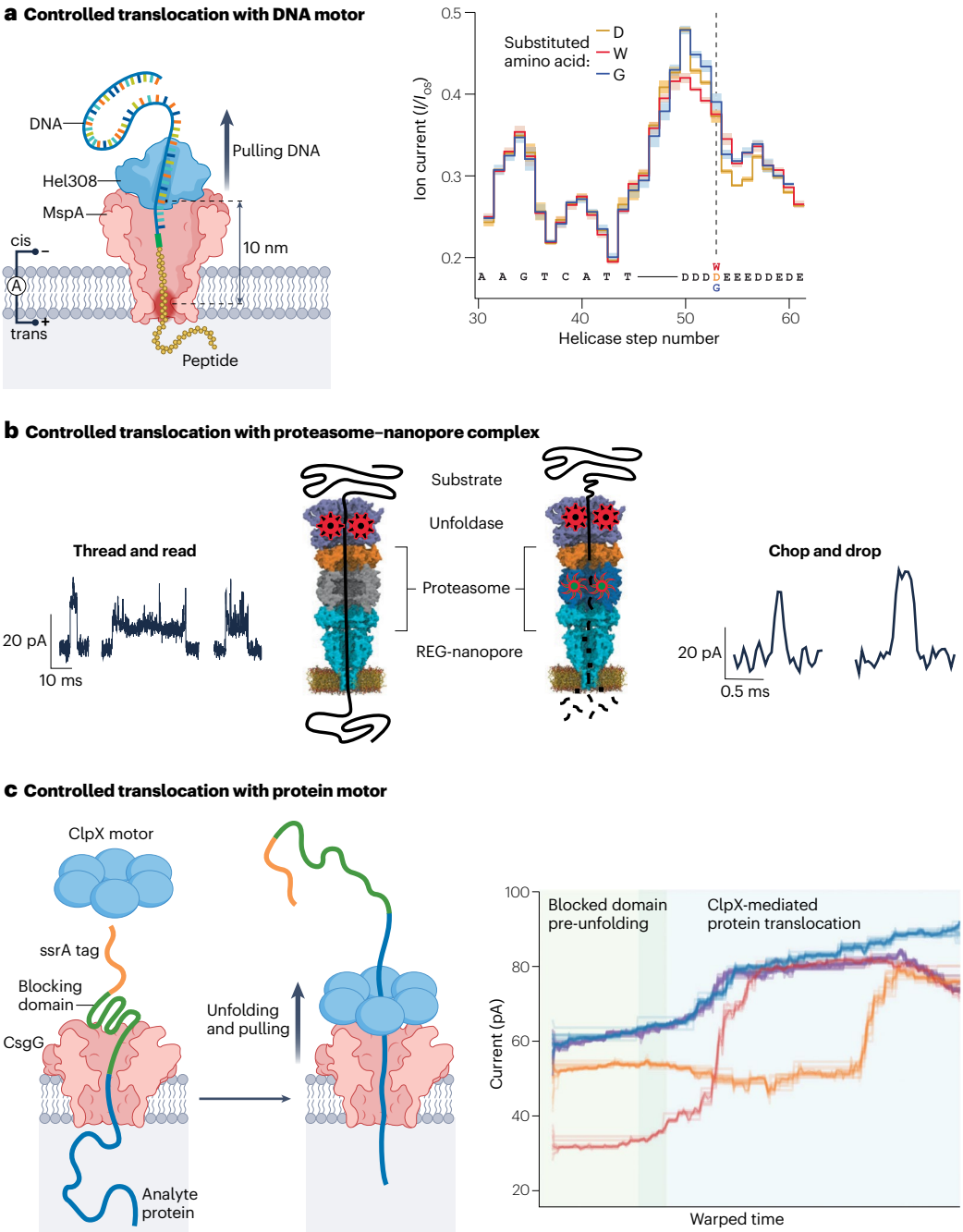


Fig. 4 | Motor enzymes can control protein translocation speed. a, Controlled peptide translocation through an *Mycobacterium smegmatis* porin A (MspA) nanopore can be achieved using the DNA motor enzyme Hel308. As this helicase walks along DNA, it pulls the peptide through the nanopore. The circled A at bottom left indicates applied voltage. The graph shows consensus patterns of theion current ratio of the blocked state and open state (I/I_{OS}) of three DNA–peptide variants with a single amino acid substitution (D, W, G), showing identical ion current signals at the DNA section and noticeable deviation within the variant site (dashed line)[49]. **b**, Engineered protein complex consisting of an inactive or an active proteasome fused with an 11S regulator (REG) nanopore and a detachable unfoldase. In the 'thread and read' mode, full proteins containing an ssrA recognition tag are unfolded and translocated, resulting in 10–100 ms-long events. In the 'chop and drop' mode, proteins are digested into smaller peptides that translocate the pore, resulting in <0.5 ms-long events[94]. **c**, Proteins can be translocated through a curli-specific gene G lipoprotein (CsgG) nanopore that is controlled by the peptide translocase caseinolytic protease X (ClpX). A negatively charged protein is threaded inside the pore, and the blocking domain prevents full translocation. ClpX recognizes the ssrA tag on the C terminus and starts unfolding the blocking domain, subsequently pulling the protein back through the pore. The graph shows ion current signal patterns of four polypeptide variants, each containing several substitutions inside the polyGSD tail[52]. Part **a** adapted with permission from ref. 49, American Association for the Advancement of Science. Part **b** adapted from ref. 94, Springer Nature Limited. Part **c** adapted from ref. 52, Springer Nature Limited.

to their negatively charged polyGSD tail. A 'blocking domain'—a strongly folded protein region—then prevents the protein from complete translocation. Subsequently, the ClpX unfoldase, attached to an ssrA tag on top of the protein, pulls the protein back through the pore, thereby allowing the nanopore to read the sequence of the protein as it translocates. Between the blocking domain and the threading tail, repeating sequence blocks with point mutations are introduced, for which signatures of all 20 amino acids were measured, providing distinctive and reproducible signal variation between different amino acids. The magnitude of the blockade depth for neutral amino acid variants shows a positive correlation with the volume of the amino acid, whereas charged variants diverge from the volume-exclusion model. In addition, a mechanism of protein rereading can be added in this system by inserting a proline-rich $(EPPPP)_5$ 'slip sequence' near the N terminus that disrupts the interaction between ClpX and the protein, causing the rethreading of the protein through electrophoretic force. This sets the stage for another round of reading when the same ClpX re-engages and pulls the protein through the nanopore again. The back-slipping occurs over a short range (50–100 amino acids) with 40% probability or over a long range (>300 amino acids) with 30% probability, and the number of rereads shows a negative correlation with the ClpX concentration used in the sequencing experiments. Therefore, the ClpX system enables the unfolding and translocation of intact proteins with high signal resolution. Nevertheless, its application in protein sequencing remains challenging because the step size of ClpX is irregular, ranging from 1 nm to bursts reaching 4 nm, corresponding to translocation distances of 3–10 amino acids per step[96]. Furthermore, this approach requires an 11-amino acid ssrA tag on the C terminus[97], a negatively charged polyGSD tail, a blocking domain and, potentially, an $(EPPPP)_5$ slip sequence[52].

In general, motor enzyme-assisted translocation for protein sequencing requires the addition of specific tags to the terminal amino acids for the motor enzymes to work effectively. Therefore, terminal modification strategies need to be improved, especially for the analysis of naturally occurring proteins. In addition, the irregular behaviour of motor enzymes can lead to inconsistent ion current readings. This issue might be addressed by protein translocases that feature small and constant step sizes and that can directly interact with the analyte of interest. Moreover, the length of the peptide that can be sequenced in DNA–peptide conjugates remains limited. Longer peptide sections could be sequenced by increasing the distance between the nanopore's constriction site and the motor enzyme, for example, through protein engineering. Rereading strategies, alongside parallelized high-throughput sequencing systems, may further improve precision by averaging out random reading errors in multiple nanopores simultaneously. Finally, a combination of electro-osmotic force-driven protein threading and motor-assisted protein translocation would enable high-fidelity charge-independent protein reading.

Identifying post-translational modifications

Accurate identification of the various types of PTMs to proteins is a major application of proteomic analysis[21]. With only ~20,000 protein coding genes in humans, variants on the RNA level increase the diversity in the transcriptome (and thus in the associated protein isoforms) by an order of magnitude. Protein PTMs, however, expand the proteome exponentially as they typically occur as combinations of different PTM types, multiple target sites and in various degrees for each protein template. These PTMs often act as switches that modify the function of target proteins, which is essential in cell signalling and metabolism[98]. Methodologies such as mass spectrometry have greatly contributed to the understanding of PTMs, but assessing the combinatorial nature of native PTMs remains challenging; for example, distinguishing isobaric peptide fragments from lysine methylations on neighbouring sites in histones remains difficult[99]. Nanopore sequencing offers high sensitivity to PTM variants at the single-molecule level, and combinatorial modifications can, in principle, be resolved as the peptide sequentially translocates through the pore. Furthermore, some PTMs, such as phosphorylation and sulfation, can be labile during sample preparation for mass spectrometry[24]. Nanopores, by contrast, require only mild conditions for sample preparation.

Phosphorylation and glycosylation are among the most abundant PTMs, and their misregulation is often involved in human disease[100,101]. Phosphorylation and glycosylation PTMs in an engineered model peptide have been detected using a FraC nanopore[102]; here, the model peptide has a long negatively charged tail at the N terminus and a positively charged tail at the C terminus. After its translocation with a voltage bias, the polarized charge distribution enables a sufficiently long dwell time (>1 ms) for detection, thereby allowing the differentiation of phosphorylated and glycosylated peptides from nonmodified peptides (Fig. 5a). Phosphorylation on serine residues in the peptide increases ion current blockade and the dwell time in the nanopore. *O*-glycosylation causes a similar increase in current blockade (despite its larger molecular size), but the dwell time decreases. Therefore, PTM detection is possible using nanopores; however, a multitude of factors influence the ion current during peptide translocation, and a single PTM on a peptide can considerably shift the peptide signal.

Peptide dwell times in FraC nanopores were extended by engineering an additional hydrophobic residue at the constriction site (G13F)[103]. Such a FraC mutant enables the identification of protein rhamnosylation—a type of bacterial glycosylation[104]—under high ionic strengths (3 M LiCl) and low pH (pH 3)[105]. Rhamnosylation on an arginine residue noticeably increases the current blockade and dwell time during translocation, reflecting the large steric effect from the modification. By digesting the elongation factor P with a protease, the native rhamnosylated protein fragment was identified. This approach may also be adapted for the detection of glycosylation in native full-length proteins.

Phosphorylation may be assessed using an aerolysin mutant (T232K/K238Q) that modulates the peptide–nanopore interaction[106]. For example, this mutant was used to study a tau fragment that carries two consecutive phosphorylation targets at the serine and threonine sites[106]; here, the extended peptide–nanopore interaction in the mutant aerolysin enables dwell times of hundreds of milliseconds and ion current blockades for the four permutations of phosphorylation on two residues, highlighting the significance of charge interaction during peptide translocation. Wild-type aerolysin can also detect charged mutations or PTMs in proteolytic fragments, albeit with shorter dwell times[107]. Aerolysin was also engineered to further increase dwell times; in particular, the R220S aerolysin mutant can be applied to study histone H4 acetylation and methylation[56]. Testing H4 fragments with three lysine residues, seven permutations of acetylation on three lysine residues were distinguished (Fig. 5b), and the added PTM masses show a linear relationship with the current blockade levels. Although the positions of the smaller methylation on lysine cannot be fully resolved, the large sensing region of aerolysin provides high resolution to the presence and positions of acetylation. Similarly, a K238A mutant of

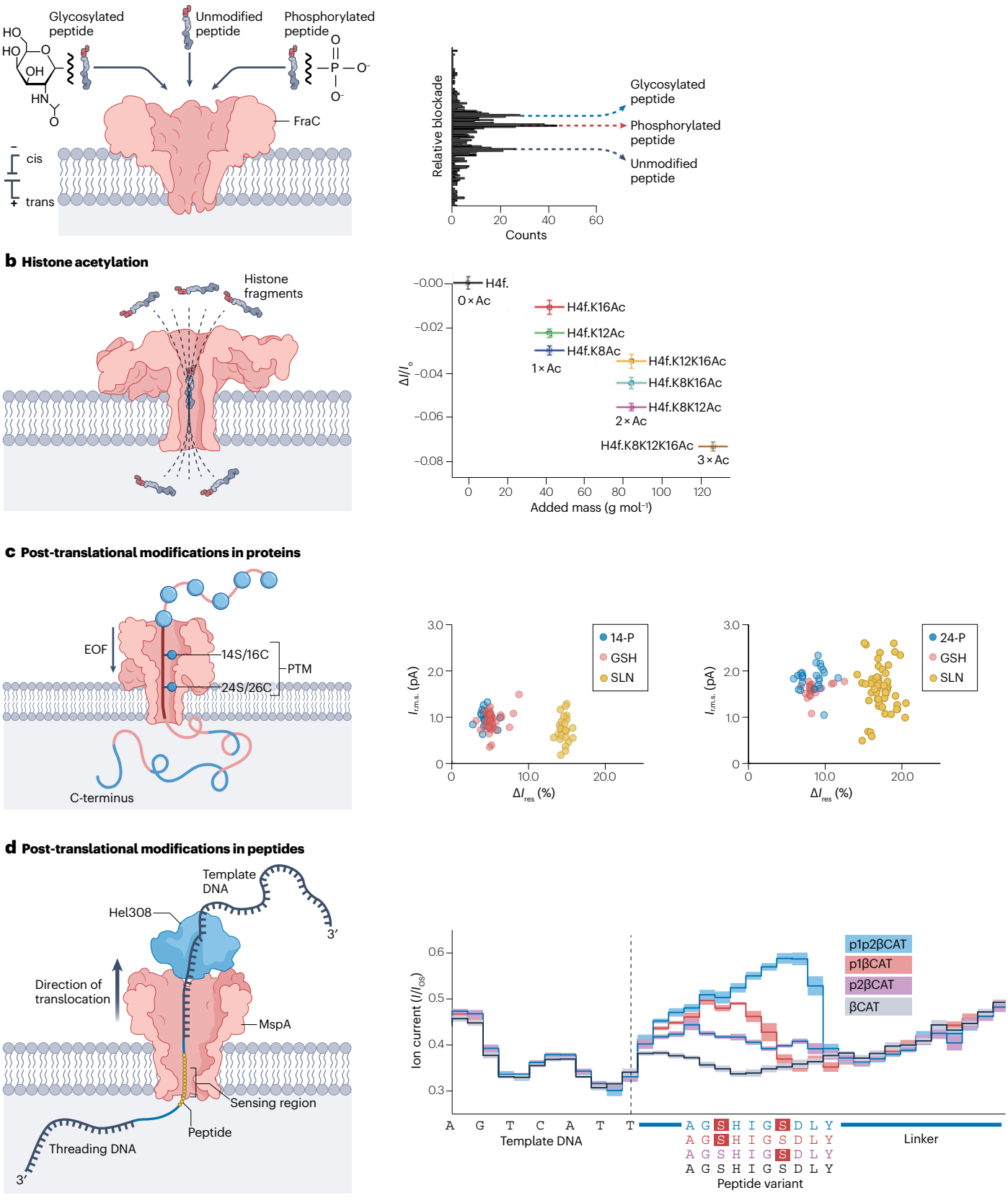

a Glycosilation and phosphorylation
Glycosylated peptide
Unmodified peptide
Phosphorylated peptide
FraC
cis
trans
Relative blockade
Glycosylated peptide
Phosphorylated peptide
Unmodified peptide
Counts
b Histone acetylation
Histone fragments
H4f. 0×Ac
H4f.K16Ac
H4f.K12Ac
H4f.K8Ac
1×Ac
H4f.K12K16Ac
H4f.K8K16Ac
H4f.K8K12Ac
2×Ac
H4f.K8K12K16Ac
3×Ac
ΔI/I0
Added mass (g mol-1)
c Post-translational modifications in proteins
EOF
14S/16C
24S/26C
PTM
C-terminus
14-P
GSH
SLN
24-P
GSH
SLN
Irms (pA)
ΔIres (%)
d Post-translational modifications in peptides
Template DNA
Hel308
3′
Direction of translocation
MspA
Sensing region
Peptide
Threading DNA
3′
Ion current (I/I0S)
p1p2βCAT
p1βCAT
p2βCAT
βCAT
A G T C A T T
Template DNA
A G S H I G S D L Y
A G S H I G S D L Y
A G S H I G S D L Y
A G S H I G S D L Y
Peptide variant
Linker

Fig. 5 | Detection of post-translational modifications with nanopores. **a**, Glycosylated and phosphorylated peptides can be distinguished after free-peptide translocation through a Fragaceatoxin C (FraC) nanopore[102]. **b**, Acetylation variants of H4 histone tails show a linear correlation between ion current and added mass using an aerolysin nanopore[56]. **c**, Translocation of full-length proteins under denaturing conditions through an α-hemolysin (αHL) nanopore enables the detection of different post-translational modifications (PTMs)—glutathione (GSH), 6′-sialyllactosamine (SLN), phosphate (P)—attached on different residues of the protein strand: 14-serine, 16-cysteine, 24-serine and 26-cysteine amino acids[47]. **d**, Motor-driven translocation of a peptide through a *Mycobacterium smegmatis* porin A (MspA) nanopore allows the discrimination of multiple PTMs on immunopeptides[86]. βCAT, β-catenin; $\Delta I/I_o$, mean relative residual current value; ΔI_{res} (%), percentage residual current; EOF, electro-osmotic flow; H4f, histone protein 4 fragment; K8Ac, lysine 8 acetylation site; $I_{r.m.s.}$ (pA), root mean square noise, in picoamperes; p1, phosphorylation site 1; p2, phosphorylation site 2; I/I_{OS}, ratio of blocked-state ion current to open-state ion current. Part **a** reprinted with permission from ref. 102, American Chemical Society. Part **b** reprinted with permission from ref. 56, American Chemical Society. Part **c** adapted from ref. 47, Springer Nature Limited. Part **d** adapted from ref. 86, Springer Nature Limited.

aerolysin was used to examine the C-terminal domain of α-synuclein, which carries multiple sites for phosphorylation, nitration and oxidation[108]. Moreover, short ion-current blockade signals from peptide translocations were analysed with a deep-learning protocol[109] to distinguish PTM variants, allowing the identification of modifications with different masses and charges with positional sensitivity.

The detection of PTMs within long proteins requires the unfolding of proteins and threading of the entire peptide chain through a pore. For example, an αHL mutant that features enhanced electro-osmosis flow was applied to investigate concatemers of the thioredoxin protein[47] with up to 1,247 amino acids. The stepwise unfolding of the thioredoxin concatemers by electro-osmotic flow and a GdmCl denaturant reveals the stages of peptide translocation and enables identification of the region carrying the PTM. In particular, phosphorylation, glutathionylation and glycosylation at two target sites result in distinct signals (Fig. 5c). Therefore, label-free full-length protein translocation and PTM detection in native proteins are possible, albeit with a large measurement error owing to limited translocation control.

The high sensitivity and distinguishing power of stepwise peptide translocation (with second-long recordings) can also be utilized for the characterization of clinically relevant immunopeptides. For example, four permutations of serine phosphorylation on two serine residues in a β-catenin immunopeptide produce distinct ion current signals using the MspA-M2 nanopore[86] (Fig. 5d). The larger size of phospho-serine counter-intuitively results in substantially elevated currents in comparison with an unphosphorylated serine residue, which may be related to charge-mediated polymer stretching and local salt concentration increase owing to the residence of the charged PTM near the pore mouth[110]. Importantly, the long and stepwise measurement allows an in-depth dissection of the contributing factors to the ion current signals from the peptide PTM, which promises sensitive detection of diverse PTM types in the proteome. Highly similar PTMs, such as sulfation and phosphorylation, can be accurately distinguished by this method[111] and by aerolysin nanopore measurements[112].

Outlook

The genome provides indispensable information for understanding the protein composition in human cells. However, translational regulation[113] and PTMs[21] exponentially expand the number of protein isoforms present in a cell. Therefore, quantitative measurement of proteins and their PTMs is key to understanding cellular signalling pathways and metabolic processes, which requires accurate quantification techniques to fingerprint and sequence proteins, including their PTMs. In addition, sequence analysis of functional proteins in biological samples might be applied in diagnostic applications.

The sensitivity of mass spectrometry has greatly improved, with various hardware and data analysis advancements, as well as specialized sample preparation workflows[114–117] and compatibility with labile PTMs[118–120]. Alternatively, single-molecule imaging methods using specific fluorescent dyes[121,122] enable high-throughput peptide fingerprinting analysis. Nanopore methodologies have demonstrated the potential to become a high-throughput multi-omics tool for genomics, transcriptomics and proteomics on one platform. Importantly, nanopore methods do not require large and expensive hardware configurations, and they can be designed portable for applications in remote areas and in space[93,123,124].

However, de novo nanopore protein sequencing will require an understanding of the contributing factors that determine the ion current signals. A combination of steric, charge and hydrophobic effects of amino acids, signal convolution from multiple residues, polymer stretching, local salt modulation by charged residues near the constriction site, and peptide–pore interactions influence ion current signals[49,86,110]. Accumulation of more experimental data for building a comprehensive biophysical model of nanopore sequencing will be required. Furthermore, the sequencing of folded proteins with secondary and tertiary structures necessitates additional treatment for capture, linearization and translocation. For example, enzyme-free electro-osmotic force-driven protein translocation[46–48,78] may enable consistent and controlled sequencing of native proteins. By improving the stability of membranes, nanopore sequencing will further allow harsh conditions to fully linearize structured proteins[125,126].

Protein translocation with motor enzymes currently provides the most detailed ion current signals, but it is inherently limited by the stochastic behaviour of the molecular machinery. Although DNA motor enzymes exhibit consistent behaviour in controlling protein translocation, the read length in the DNA–peptide sequencing process is restricted to the initial ~25 amino acids. To extend the read length, taller pores may be engineered to increase the distance between the nanopore's sensing region and the motor enzyme[94,127,128]. Furthermore, functional protein design[129,130] could improve motor protein capabilities. Moreover, taking advantage of the single-molecule sensitivity of nanopore measurements, sample enrichment protocols will have to be developed to avoid saturating amounts of contaminant molecules during experiments.

Even with only empirical understanding of ion current signals, nanopore sequencing can be applied to identify point mutations and PTMs in model peptides and proteins. Known as 'reference sequencing', nanopore measurements can be used to detect variants from a template peptide, once ion current signatures are established a priori. In particular, nanopore-based protein fingerprinting can be useful for detecting enzymatically cleaved peptide fragments, and for facilitating detailed protein analysis and precise identification, thereby increasing our knowledge of protein abundance[131,132].

Box 2 | Potential clinical applications of nanopore protein sequencing

Disease diagnosis

Protein mutations and abnormal post-translational modifications can be sensitively detected implementing appropriate sample preparation protocols. The single-molecule nature of the measurement enables early diagnosis of low-abundance biomarkers and disease-associated proteins in patient samples, for example, for the diagnosis of neurodegenerative disorders[134] and cancer[61].

Pathogen identification

Nanopore devices benefit from portability and accessibility. Therefore, samples can be extracted on site from various origins, for example, during infectious disease outbreaks[135]. This allows early screening to follow pathogen progression and treatment efficacy.

Biomarker discovery

The sensitive nature of nanopore sequencing might enable the identification of new post-translational modification-related protein biomarkers, associated with various diseases, which can be used for diagnostic, prognostic and predictive purposes.

Personalized medicine

By analysing an individual's proteome profile, nanopore sequencing can help tailor treatment strategies based on an individual's specific molecular characteristics, beyond DNA and RNA, which might allow more comprehensive and effective therapies.

Prenatal and newborn screening

Nanopore protein sequencing can be used for non-invasive prenatal testing and newborn screening. Owing to the blood exchange between the mother and the fetus, nanopore protein sequencing of maternal blood might enable the early detection of genetic disorders and congenital conditions.

In addition, nanopore sequencing, especially PTM detection, may be valuable for clinical applications (Box 2). A better understanding of ion current signals from PTM variants will further contribute to building a comprehensive nanoscale model of nanopore peptide measurements.

The different nanopore-based protein sequencing approaches may be suitable for distinct applications. The sequencing of small peptides may particularly benefit from aerolysin-based tag-free measurements, whereas heavily modified peptides may be better analysed using a DNA-motor approach to disentangle modification states and locations. The sequencing of folded proteins requires protein motors or denaturants. Importantly, understanding the advantages and disadvantages of the various translocation and detection strategies will be key to develop a de novo nanopore protein sequencing platform. Beyond DNA and RNA sequencing, nanopore sequencing is particularly promising for proteomics analysis, which can be applied to various fundamental research questions and integrated into a bioengineer's analytical toolbox.

Published online: 21 November 2024

References

1. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93**, 13770–13773 (1996).
2. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* **77**, 3227–3233 (1999).
3. Meller, A., Nivon, L., Brandin, E., Golovchenko, J. & Branton, D. Rapid nanopore discrimination between single polynucleotide molecules. *Proc. Natl Acad. Sci. USA* **97**, 1079–1084 (2000).
4. Deamer, D. W. & Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* **35**, 817–825 (2002).
5. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
6. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
7. Chen, P. et al. Portable nanopore-sequencing technology: trends in development and applications. *Front. Microbiol.* **14**, 1043967 (2023).
8. Dorey, A. & Howorka, S. Nanopore DNA sequencing technologies and their applications towards single-molecule proteomics. *Nat. Chem.* **16**, 314–334 (2024).
9. Clarke, J. et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
10. 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).
11. Kumar, S. et al. PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. *Sci. Rep.* **2**, 684 (2012).
12. Manrao, E. A. et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
13. Cherf, G. M. et al. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* **30**, 344–348 (2012).
14. Branton, D. et al. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26**, 1146–1153 (2008).
15. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–250 (2013).
16. 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).
17. Wei, X. et al. Engineering biological nanopore approaches toward protein sequencing. *ACS Nano* **17**, 16369–16395 (2023).
18. Pugh, G. C., Burns, J. R. & Howorka, S. Comparing proteins and nucleic acids for next-generation biomolecular engineering. *Nat. Rev. Chem.* **2**, 113–130 (2018).
19. Huo, M.-Z., Li, M.-Y., Ying, Y.-L. & Long, Y.-T. Is the volume exclusion model practicable for nanopore protein sequencing? *Anal. Chem.* **93**, 11364–11369 (2021).
20. Jayasinghe, L. & Wallace, J. E. Mutant pore. US patent 11186868 B2 (2021).
21. Aebersold, R. et al. How many human proteoforms are there? *Nat. Chem. Biol.* **14**, 206–214 (2018).
22. Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. *Nat. Biotechnol.* **28**, 695–709 (2010).
23. Polasky, D. A., Yu, F., Teo, G. C. & Nesvizhskii, A. I. Fast and comprehensive *N*- and *O*-glycoproteomics analysis with MSFragger-Glyco. *Nat. Methods* **17**, 1125–1132 (2020).
24. Daly, L. A., Clarke, C. J., Po, A., Oswald, S. O. & Eyers, C. E. Considerations for defining +80 Da mass shifts in mass spectrometry-based proteomics: phosphorylation and beyond. *Chem. Commun.* **59**, 11484–11499 (2023).
25. Xu, T., Wang, Q., Wang, Q. & Sun, L. Mass spectrometry-intensive top-down proteomics: an update on technology advancements and biomedical applications. *Anal. Methods* **16**, 4664–4682 (2024).
26. Nicholson, J. A nanopore distance away from next-generation protein sequencing. *Chem* **8**, 17–19 (2022).
27. Motone, K. & Nivala, J. Not if but when nanopore protein sequencing meets single-cell proteomics. *Nat. Methods* **20**, 336–338 (2023).
28. Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* **2**, 209–215 (2007).
29. Mayer, S. F., Cao, C. & Dal Peraro, M. Biological nanopores for single-molecule sensing. *iScience* **25**, 104145 (2022).
30. Plesa, C. et al. Fast translocation of proteins through solid state nanopores. *Nano Lett.* **13**, 658–663 (2013).
31. Restrepo-Pérez, L., Joo, C. & Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **13**, 786–796 (2018).
32. Xue, L. et al. Solid-state nanopore sensors. *Nat. Rev. Mater.* **5**, 931–951 (2020).
33. Wang, H., Tang, H., Qiu, X. & Li, Y. Solid-state glass nanopipettes: functionalization and applications. *Chem. Eur. J.* **30**, e202400281 (2024).
34. Chou, Y.-C., Masih Das, P., Monos, D. S. & Drndić, M. Lifetime and stability of silicon nitride nanopores and nanopore arrays for ionic measurements. *ACS Nano* **14**, 6715–6728 (2020).
35. Graf, M. et al. Fabrication and practical applications of molybdenum disulfide nanopores. *Nat. Protoc.* **14**, 1130–1168 (2019).
36. Wilson, J., Sloman, L., He, Z. & Aksimentiev, A. Graphene nanopores for protein sequencing. *Adv. Funct. Mater.* **26**, 4830–4838 (2016).

37. Thomsen, R. P. et al. A large size-selective DNA nanopore with sensing applications. *Nat. Commun.* **10**, 5655 (2019).
38. Tripathi, P. et al. Electrical unfolding of cytochrome c during translocation through a nanopore constriction. *Proc. Natl Acad. Sci. USA* **118**, e2016262118 (2021).
39. Xing, Y., Dorey, A., Jayasinghe, L. & Howorka, S. Highly shape- and size-tunable membrane nanopores made with DNA. *Nat. Nanotechnol.* **17**, 708–713 (2022).
40. Bošković, F. & Keyser, U. F. Nanopore microscope identifies RNA isoforms with structural colours. *Nat. Chem.* **14**, 1258–1264 (2022).
41. Wang, F. et al. MoS_2 nanopore identifies single amino acids with sub-1 Dalton resolution. *Nat. Commun.* **14**, 2895 (2023).
42. Shen, Q. et al. Functionalized DNA-origami-protein nanopores generate large transmembrane channels with programmable size-selectivity. *J. Am. Chem. Soc.* **145**, 1292–1300 (2023).
43. Fragasso, A., Schmid, S. & Dekker, C. Comparing current noise in biological and solid-state nanopores. *ACS Nano* **14**, 1338–1349 (2020).
44. Mohammadi, M. M. & Bavi, O. DNA sequencing: an overview of solid-state and biological nanopore-based methods. *Biophys. Rev.* **14**, 99–110 (2021).
45. Alfaro, J. A. et al. The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **18**, 604–617 (2021).
46. Yu, L. et al. Unidirectional single-file transport of full-length proteins through a nanopore. *Nat. Biotechnol.* **41**, 1130–1139 (2023).
47. Martin-Baniandres, P. et al. Enzyme-less nanopore detection of post-translational modifications within long polypeptides. *Nat. Nanotechnol.* **18**, 1335–1340 (2023).
48. 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 (2024).
49. 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).
50. 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).
51. Yan, S. et al. Single molecule ratcheting motion of peptides in a *Mycobacterium smegmatis* porin A (MspA) nanopore. *Nano Lett.* **21**, 6703–6710 (2021).
52. Motone, K. et al. Multi-pass, single-molecule nanopore reading of long protein strands. *Nature* **633**, 662–669 (2024).
53. Meller, A., Nivon, L. & Branton, D. Voltage-driven DNA translocations through a nanopore. *Phys. Rev. Lett.* **86**, 3435–3438 (2001).
54. Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl Acad. Sci. USA* **105**, 20647–20652 (2008).
55. Derrington, I. M. et al. Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat. Biotechnol.* **33**, 1073–1075 (2015).
56. 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).
57. Song, L. et al. Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859–1865 (1996).
58. Iacovache, I. et al. Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. *Nat. Commun.* **7**, 12062 (2016).
59. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science* **303**, 1189–1192 (2004).
60. Goyal, P. et al. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* **516**, 250–253 (2014).
61. 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).
62. Wang, K. et al. Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat. Methods* **21**, 92–101 (2024).
63. Zhou, W., Qiu, H., Guo, Y. & Guo, W. Molecular insights into distinct detection properties of α-hemolysin, MspA, CsgG, and aerolysin nanopore sensors. *J. Phys. Chem. B* **124**, 1611–1618 (2020).
64. Crnković, A., Srnko, M. & Anderluh, G. Biological nanopores: engineering on demand. *Life* **11**, 27 (2021).
65. 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).
66. Pavlenok, M., Yu, L., Herrmann, D., Wanunu, M. & Niederweis, M. Control of subunit stoichiometry in single-chain MspA nanopores. *Biophys. J.* **121**, 742–754 (2022).
67. Ouldali, H. et al. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
68. Piguet, F. et al. Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* **9**, 966 (2018).
69. Baaken, G. et al. High-resolution size-discrimination of single nonionic synthetic polymers with a highly charged biological nanopore. *ACS Nano* **9**, 6443–6449 (2015).
70. Chavis, A. E. et al. Single molecule nanopore spectrometry for peptide detection. *ACS Sens.* **2**, 1319–1328 (2017).
71. Delahaye, C. & Nicolas, J. Sequencing DNA with nanopores: troubles and biases. *PLoS ONE* **16**, e0257521 (2021).
72. Zhang, Y. et al. Peptide sequencing based on host–guest interaction-assisted nanopore sensing. *Nat. Methods* **21**, 102–109 (2024).
73. Peeters, J. M., Hazendonk, T. G., Beuvery, E. C. & Tesser, G. I. Comparison of four bifunctional reagents for coupling peptides to proteins and the effect of the three moieties on the immunogenicity of the conjugates. *J. Immunol. Methods* **120**, 133–143 (1989).
74. Zhang, L. et al. Photoredox-catalyzed decarboxylative C-terminal differentiation for bulk- and single-molecule proteomics. *ACS Chem. Biol.* **16**, 2595–2603 (2021).
75. De Rosa, L., Di Stasi, R., Romanelli, A. & D'Andrea, L. D. Exploiting protein N-terminus for site-specific bioconjugation. *Molecules* **26**, 3521 (2021).
76. Afshar Bakshloo, M. et al. Polypeptide analysis for nanopore-based protein identification. *Nano Res.* **15**, 9831–9842 (2022).
77. Boukhet, M. et al. Probing driving forces in aerolysin and α-hemolysin biological nanopores: electrophoresis versus electroosmosis. *Nanoscale* **8**, 18352–18359 (2016).
78. Li, M. & Muthukumar, M. Electro-osmotic flow in nanoconfinement: solid-state and protein nanopores. *J. Chem. Phys.* **160**, 084905 (2024).
79. Wen, C., Schmid, S. & Dekker, C. Understanding electrophoresis and electroosmosis in nanopore sensing with the help of the nanopore electro-osmotic trap. *ACS Nano* **18**, 20449–20458 (2024).
80. Liu, Y. et al. Machine learning assisted simultaneous structural profiling of differently charged proteins in a *Mycobacterium smegmatis* porin A (MspA) electroosmotic trap. *J. Am. Chem. Soc.* **144**, 757–768 (2022).
81. Mehrafrooz, B. et al. Electro-osmotic flow generation via a sticky ion action. *ACS Nano* **18**, 17521–17533 (2024).
82. Sauciuc, A. et al. Blobs form during the single-file transport of proteins across nanopores. *Proc. Natl Acad. Sci. USA* **121**, e2405018121 (2024).
83. Baldelli, M. et al. Controlling electroosmosis in nanopores without altering the nanopore sensing region. *Adv. Mater.* **36**, 2401761 (2024).
84. 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).
85. Nova, I. C. et al. Investigating asymmetric salt profiles for nanopore DNA sequencing with biological porin MspA. *PLoS ONE* **12**, e0181599 (2017).
86. Nova, I. C. et al. Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nat. Biotechnol.* **42**, 710–714 (2023).
87. Craig, J. M. et al. Determining the effects of DNA sequence on Hel308 helicase translocation along single-stranded DNA using nanopore tweezers. *Nucleic Acids Res.* **47**, 2506–2513 (2019).
88. Bhattacharya, S., Yoo, J. & Aksimentiev, A. Water mediates recognition of DNA sequence via ionic current blockade in a biological nanopore. *ACS Nano* **10**, 4644–4651 (2016).
89. Katsikis, P. D., Ishii, K. J. & Schliehe, C. Challenges in developing personalized neoantigen cancer vaccines. *Nat. Rev. Immunol.* **24**, 213–227 (2024).
90. Yewdell, J. W. MHC class I immunopeptidome: past, present, and future. *Mol. Cell. Proteom.* **21**, 100230 (2022).
91. Boparai, J. K. & Sharma, P. K. Mini review on antimicrobial peptides, sources, mechanism and recent applications. *Protein Pept. Lett.* **27**, 4–16 (2020).
92. Koch, C. et al. Nanopore sequencing of DNA-barcoded probes for highly multiplexed detection of microRNA, proteins and small biomarkers. *Nat. Nanotechnol.* **18**, 1483–1491 (2023).
93. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
94. Zhang, S. et al. Bottom-up fabrication of a proteasome–nanopore that unravels and processes single proteins. *Nat. Chem.* **13**, 1192–1199 (2021).
95. Barkow, S. R., Levchenko, I., Baker, T. A. & Sauer, R. T. Polypeptide translocation by the AAA+ ClpXP protease machine. *Chem. Biol.* **16**, 605–612 (2009).
96. Cordova, J. C. et al. Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158**, 647–658 (2014).
97. Flynn, J. M. et al. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc. Natl Acad. Sci. USA* **98**, 10584–10589 (2001).
98. Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* **17**, 666–672 (2010).
99. El Kennani, S., Crespo, M., Govin, J. & Pflieger, D. Proteomic analysis of histone variants and their PTMs: strategies and pitfalls. *Proteomes* **6**, 29 (2018).
100. Park, J. S. et al. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun.* **10**, 3090 (2019).
101. Costa, A. F., Campos, D., Reis, C. A. & Gomes, C. Targeting glycosylation: a new road for cancer drug discovery. *Trends Cancer* **6**, 757–766 (2020).
102. 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).
103. Lucas, F. L. R. et al. The manipulation of the internal hydrophobicity of FraC nanopores augments peptide capture and recognition. *ACS Nano* **15**, 9600–9613 (2021).
104. Rajkovic, A. et al. Cyclic rhamnosylated elongation factor P establishes antibiotic resistance in *Pseudomonas aeruginosa*. *mBio* **6**, e00823-15 (2015).
105. Versloot, R. C. A. et al. Quantification of protein glycosylation using nanopores. *Nano Lett.* **22**, 5357–5364 (2022).
106. Li, S. et al. T232K/K238Q aerolysin nanopore for mapping adjacent phosphorylation sites of a single tau peptide. *Small Methods* **4**, 2000014 (2020).
107. Afshar Bakshloo, M. et al. Discrimination between alpha-synuclein protein variants with a single nanometer-scale pore. *ACS Chem. Neurosci.* **14**, 2517–2526 (2023).

108. Cao, C. et al. Deep learning-assisted single-molecule detection of protein post-translational modifications with a biological nanopore. *ACS Nano* **18**, 1504–1515 (2024).
109. Cao, C. et al. Aerolysin nanopores decode digital information stored in tailored macromolecular analytes. *Sci. Adv.* **6**, eabc2661 (2020).
110. Liu, J. & Aksimentiev, A. Molecular determinants of current blockade produced by peptide transport through a nanopore. *ACS Nanosci. Au* **4**, 21–29 (2024).
111. Chen, X. et al. Resolving sulfation posttranslational modifications on a peptide hormone using nanopores. *ACS Nano* **18**, 28999–29007 (2024).
112. Niu, H. et al. Direct mapping of tyrosine sulfation states in native peptides by nanopore. *Nat. Chem. Biol*. https://doi.org/10.1038/s41589-024-01734-x (2024).
113. Franks, A., Airoldi, E. & Slavov, N. Post-transcriptional regulation across human tissues. *PLoS Comput. Biol*. **13**, e1005535 (2017).
114. Sinitcyn, P. et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat. Biotechnol.* **39**, 1563–1573 (2021).
115. Cox, J. Prediction of peptide mass spectral libraries with machine learning. *Nat. Biotechnol.* **41**, 33–43 (2023).
116. Deslignière, E. et al. Ultralong transients enhance sensitivity and resolution in Orbitrap-based single-ion mass spectrometry. *Nat. Methods* **21**, 619–622 (2024).
117. Ye, Z. et al. One-Tip enables comprehensive proteome coverage in minimal cells and single zygotes. *Nat. Commun.* **15**, 2474 (2024).
118. Daly, L. A. et al. Custom workflow for the confident identification of sulfotyrosine-containing peptides and their discrimination from phosphopeptides. *J. Proteome Res.* **22**, 3754–3772 (2023).
119. Schjoldager, K. T., Narimatsu, Y., Joshi, H. J. & Clausen, H. Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol*. **21**, 729–749 (2020).
120. Bagdonaite, I. et al. Glycoproteomics. *Nat. Rev. Methods Primers* **2**, 48 (2022).
121. Reed, B. D. et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science* **378**, 186–192 (2022).
122. Filius, M. et al. Full-length single-molecule protein fingerprinting. *Nat. Nanotechnol.* **19**, 652–659 (2024).
123. Castro-Wallace, S. L. et al. Nanopore DNA sequencing and genome assembly on the international space station. *Sci. Rep.* **7**, 18022 (2017).
124. Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
125. Yu, L. et al. Stable polymer bilayers for protein channel recordings at high guanidinium chloride concentrations. *Biophys. J.* **120**, 1537–1541 (2021).
126. Vreeker, E. et al. Hybrid lipid-block copolymer membranes enable stable reconstitution of a wide range of nanopores and robust sampling of serum. Preprint at *bioRxiv* https://doi.org/10.1101/2024.05.16.594548 (2024).
127. Lutz, I. D. et al. Top-down design of protein architectures with reinforcement learning. *Science* **380**, 266–273 (2023).
128. Xu, C. et al. Computational design of transmembrane pores. *Nature* **585**, 129–134 (2020).
129. Notin, P., Rollins, N., Gal, Y., Sander, C. & Marks, D. Machine learning for functional protein design. *Nat. Biotechnol.* **42**, 216–228 (2024).
130. Eisenstein, M. Seven technologies to watch in 2024. *Nature* **625**, 844–848 (2024).
131. 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).
132. Afshar Bakshloo, M. et al. Nanopore-based protein identification. *J. Am. Chem. Soc.* **144**, 2716–2725 (2022).
133. Cao, C. et al. Mapping the sensing spots of aerolysin for single oligonucleotides analysis. *Nat. Commun.* **9**, 2823 (2018).
134. Wu, J., Yamashita, T., Hamilton, A. D., Thompson, S. & Luo, J. Single-molecule nanopore dielectrophoretic trapping of α-synuclein with lipid membranes. *Cell Rep. Phys. Sci.* **4**, 101243 (2023).
135. Bošković, F. et al. Simultaneous identification of viruses and viral variants with programmable DNA nanobait. *Nat. Nanotechnol.* **18**, 290–298 (2023).

Acknowledgements
We thank B. Albada, E. Bertosin, E. van der Sluis and L. Yu for a critical reading of the manuscript. This work was supported by funding from the Dutch Research Council (NWO) project NWO-I680 (SMPS), European Research Council Advanced Grant 883684 and US National Institutes of Health National Human Genome Research Institute project HG012544-01.

Author contributions
J.R. and X.C. surveyed the literature for this article. All authors contributed to writing the manuscript. C.D. supervised the work.

Competing interests
C.D. is named inventor on a patent on protein sequencing with nanopores, which is licensed to Oxford Nanopore Technologies. Beyond that, the authors declare no competing interests.

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

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