iScience

CellPress
OPEN ACCESS

Review

Herding cats: Label-based approaches in protein translocation through nanopore sensors for single-molecule protein sequence analysis

Keisuke Motone,[1] Nicolas Cardozo,[1] and Jeff Nivala[1,*]

SUMMARY

Proteins carry out life's essential functions. Comprehensive proteome analysis technologies are thus required for a full understanding of the operating principles of biological systems. While current proteomics techniques suffer from limitations in sensitivity and/or throughput, nanopore technology has the potential to enable de novo protein identification through single-molecule sequencing. However, a significant barrier to achieving this goal is controlling protein/peptide translocation through the nanopore sensor for processive strand analysis. Here, we review recent approaches that use a range of techniques, from oligonucleotide conjugation to molecular motors, aimed at driving protein strands and peptides through protein nanopores. We further discuss site-specific protein conjugation chemistry that could be combined with these translocation approaches as future directions to achieve single-molecule protein detection and sequencing of native proteins.

INTRODUCTION

Proteins are the major functional molecules involved in essentially every biological process, such as building the cellular structure, regulating gene expression, and powering the immune system. Although genomics and transcriptomics provide fundamental information about cellular history and basal activity, proteomics plays a crucial role in filling the gap between genotype and phenotype as protein activities are more directly related to phenotype (Figure 1). Thus, protein analysis provides valuable information for understanding biological phenomena and disease. Unfortunately, unlike the remarkable technological improvements in DNA and RNA sequencing in recent years, the development of highly sensitive, high-throughput protein sequencing techniques have not yet been realized. There are two principal methods currently available for protein sequencing/identification that do not use affinity reagents such as antibodies: Edman degradation and mass spectrometry (Edman et al., 1950; Steen and Mann, 2004). Edman degradation is a useful technique for de novo sequencing, but it is limited to the analysis of homogenous protein samples and read lengths typically <50 amino acids, which are far shorter than the median protein length of eukaryotic (361 amino acid long), bacterial (267 amino acid long), and archaeal organisms (247 amino acid long) (Brocchieri, 2005). Mass spectrometry allows the analysis of protein mixtures and currently dominates proteomics research. Mass spectrometry has undergone significant improvements in instrumentation and sample preparation over the decades (Budnik et al., 2018; Specht et al., 2021; Zhu et al., 2018a, 2018b), although it still faces limitations in terms of detection sensitivity, dynamic range, analytical throughput, and instrumentation cost (Zubarev, 2013).

To address these challenges, complementary or potentially disruptive platforms for next-generation protein analysis and sequencing have been envisioned recently (Alfaro et al., 2021; Asandei et al., 2020; Cressiot et al., 2020; Hu et al., 2021; Restrepo-Pérez et al., 2018; Timp and Timp, 2020). These emerging techniques include tunneling currents, single-molecule fluorescence, and nanopores. In 2014, Ohshiro et al. identified 12 different amino acids and phosphotyrosine using tunneling currents measured as the individual molecules threaded through a nanoscale electrode gap (Ohshiro et al., 2014). Tunneling currents have shown the remarkable sensitivity to discriminate even enantiomers and isobaric amino acids (Zhao et al., 2014). In more recent years, several fluorescence-based studies have demonstrated experimental proof-of-principle toward single-molecule protein fingerprinting (de Lannoy et al., 2021; Swaminathan et al., 2018; van Ginkel et al., 2018). For example, Swaminathan et al. established a fluorosequencing

[1]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA

*Correspondence: jmdn@uw.edu

https://doi.org/10.1016/j.isci.2021.103032

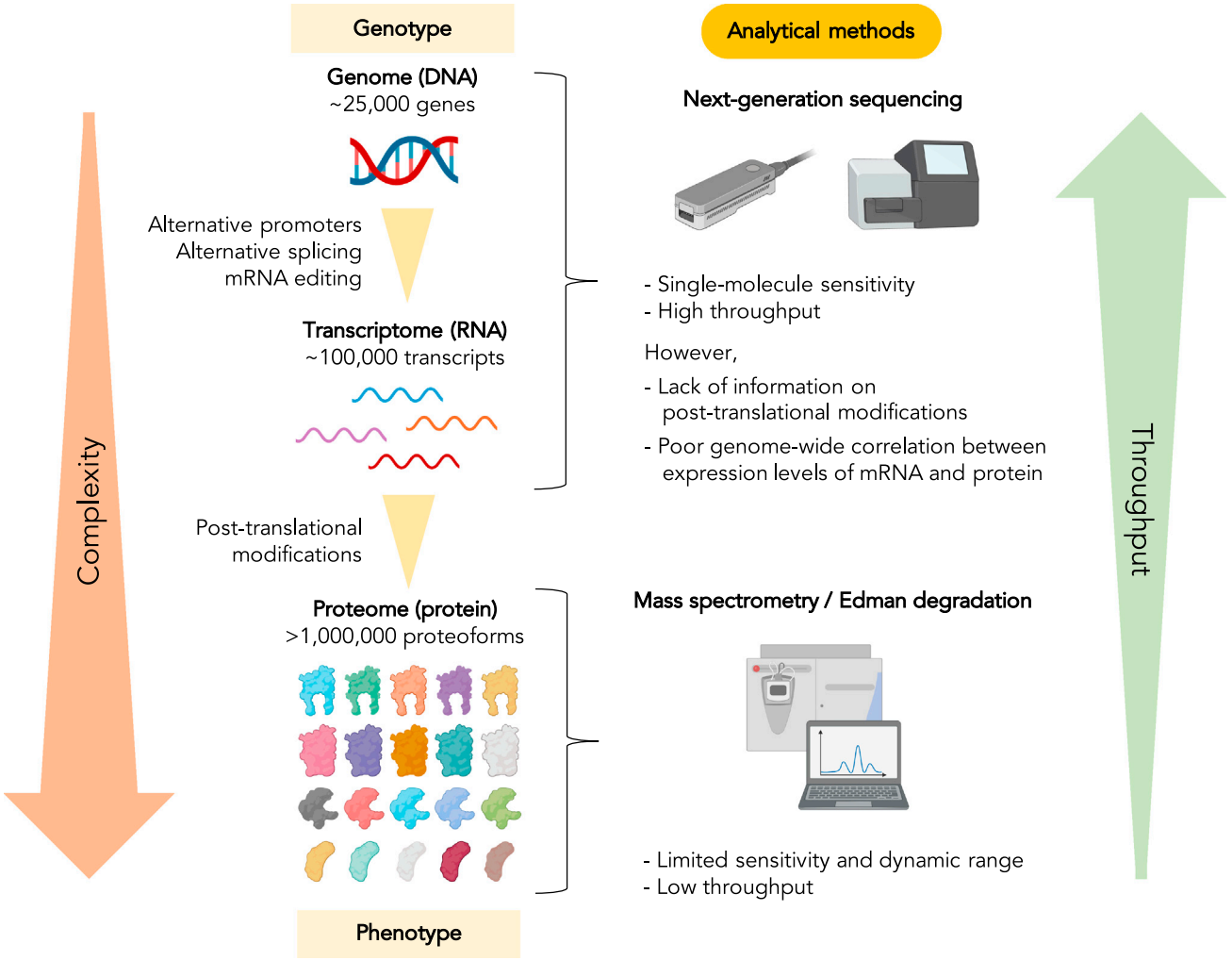

Figure 1. Current analytical techniques for omics studies

(Left) The proteome is orders of magnitude more complex than the genome and transcriptome (Aebersold et al., 2018). (Right) Second and third-generation sequencing techniques can be used for high throughput genomic and transcriptomic analysis, while protein sequencing/identification is performed by relatively-low throughput methods such as mass spectrometry or Edman degradation.

technology that couples Edman degradation with fluorescent labeling of specific residues (e.g., cysteines and lysines). Total internal reflection fluorescence (TIRF) microscopy allows for millions of fluorescently labeled peptide molecules to be identified in parallel based on the sparse fluorescent sequence (i.e., fingerprint) of each molecule determined by recurrent rounds of Edman degradation (Swaminathan et al, 2015, 2018). Thus, those innovative approaches based on tunneling currents or fluorescence have opened new avenues to revolutionize conventional proteomic technologies. However, they may face limitations compared to nanopore technology in terms of the potential to achieve de novo full-length protein sequencing. Approaches using tunneling currents, for instance, lack experimental demonstrations of controlling translocation of polypeptides through the gap and are currently limited to the analysis of amino acids or short peptides. Fluorescence-based methods are so far incompatible with de novo sequencing owing to constraint in the chemical repertoire to orthogonally label the 20 amino acids and the number of uniquely distinguishable fluorescent labels. In the following section, we present an overview of protein analysis and sequencing using nanopore technology.

Recent advances and challenges in nanopore sequencing

Nanopore technology relies on a nanometer-sized pore (i.e., nanopore) within an insulating membrane that separates two electrolyte-filled wells. A voltage applied across the membrane drives ionic current flow through the

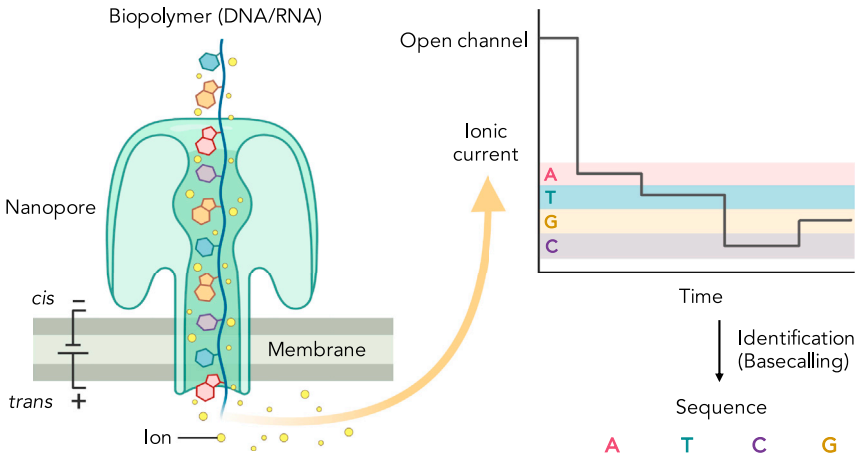

Figure 2. Schematic representation of single-molecule sequencing with a nanopore

In nanopore sequencing, a nanometer-sized protein pore is embedded in an insulating membrane that separates two electrolyte-filled wells. Voltage is applied between the wells, causing ionic current flow through the pore. As single biopolymer molecules translocate through the channel, they generate sequence-specific ionic current signals that are diagnostic of the polymer sequence. Nanopore sequencing has been established as a commercial technology to read single DNA and RNA molecules (Garalde et al., 2018; Jain et al., 2016).

nanopore sensor between the *cis* and *trans* wells. As individual molecules translocate through or otherwise interact with the nanopore, they can rectify this ionic current flow and cause an observable signal change that provides structural insights into the molecules (Deamer and Akeson, 2000). Although nanopore sensing was initially proposed as a technique for the sequencing of nucleic acid strands (Kasianowicz et al., 1996) (Figure 2), it also has great potential for protein analysis. Single-molecule sensitivity, full-length readout, real-time measurement, and device portability is just as, if not more, crucial for proteomics than it is for genomics and transcriptomics (Shi et al., 2017). Towards this end, nanopore sensors have been used for discrimination of peptides and proteins (Asandei et al., 2017; Cardozo et al., 2021; Huang et al, 2017, 2019; Nivala et al., 2014; Piguet et al., 2018; Robertson and Reiner, 2018), real-time measurement of protein–protein (Thakur and Movileanu, 2019) and protein–ligand interactions (Harrington et al., 2013; Movileanu et al., 2000), antigen−antibody binding assays (Han et al., 2008; Madampage et al., 2010; Sexton et al., 2007), and aptamer-mediated protein detection (Rotem et al., 2012; Soskine et al., 2012; Sze et al., 2017). Moreover, protein nanopores have shown promise in identifying amino acids and post-translational modifications (PTMs), taking a major step toward single-molecule protein sequencing. For example, Ouldali et al. have recently shown that 13 of the 20 standard amino acids are distinguishable based on their current signals using an aerolysin nanopore (Ouldali et al., 2020). The study also has proposed a way toward the identification of the remaining 7 amino acids by instrumentation advances and nanopore engineering. The detection of PTMs such as phosphorylation and glycosylation, which serve as biomarkers of cell states and diseases (Aebersold et al., 2018; Pagel et al., 2015), has also been achieved with protein nanopore sensors (Restrepo-Pérez et al., 2019b; Rosen et al., 2014b; Wloka et al., 2017; Ying et al., 2019; Zhang et al., 2021b).

Despite these promising results, protein sequencing of intact, full-length protein strands using nanopores has been hindered, in part, because of the difficulty in controlling protein translocation through the sensor. This challenge exists because of two major reasons (Figure 3): First, the polypeptide backbone is neutrally charged and amino acid side chains can vary in charge state. Thus, electrophoresis-driven unidirectional translocation of peptides or proteins through nanopores cannot be as effectively employed, as it can for uniformly negatively-charged polymers like nucleic acids. Second, most proteins adopt a stable 3-dimensional fold. Thus, disruption of this tertiary structure is required for proteins to translocate through a narrow nanopore constriction for primary sequence analysis. In this review, we summarize the recent advances and remaining obstacles in controlling protein translocation through a nanopore and highlight label-based approaches that potentially address these challenges.

APPROACHES FOR PROTEIN/PEPTIDE TRANSLOCATION THROUGH A NANOPORE

In 2004, peptide translocation through a nanopore was demonstrated for the first time (Sutherland et al., 2004). This study showed the translocation of short repeats of the collagen-like sequence (GPP) through an

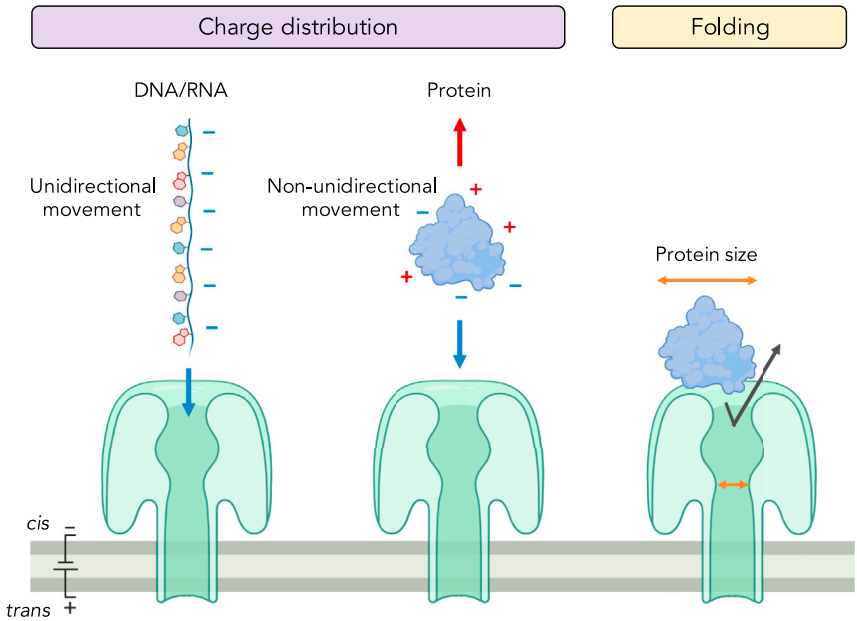

Figure 3. Challenges in protein translocation through a nanopore

(Left) While DNA/RNA is uniformly negatively charged, (Middle) proteins can contain both negatively-charged (glutamate and aspartate) and positively-charged (arginine, lysine, and histidine) residues at physiological pH. Unidirectional translocation of proteins in the electric field must be achieved despite their nonuniform charge. (Right) The diameters of folded proteins are typically larger than the constriction of protein nanopores that would be suitable for protein sequencing application (~1–2 nm) (Steinbock et al., 2014). For processive strand analysis, proteins must be unfolded to allow the denatured protein strand to thread through the nanopore with amino acid residues in single-file order.

α-hemolysin nanopore. Another pioneering study investigated the interactions of an α-hemolysin nanopore with helical peptides containing the $(AAKAA)_n$ sequence (Movileanu et al., 2005). These works laid the foundation for peptide analysis using nanopores, but general approaches for translocation of native peptides/proteins through nanopores are required for the development of single-molecule protein sequencing.

First efforts for protein translocation involved physical and chemical denaturants. Several groups have shown that the use of high temperature (Payet et al., 2012), chaotropic agents (Cressiot et al, 2012, 2015; Merstorf et al., 2012; Oukhaled et al., 2007; Pastoriza-Gallego et al., 2011; Talaga and Li, 2009), or detergents (Kennedy et al., 2016; Restrepo-Pérez et al., 2017), facilitates protein translocation through nanopores. For example, Timp and colleagues have demonstrated protein translocation through solid-state pores using sodium dodecyl sulfate (SDS) as a denaturant (Kennedy et al., 2016). SDS further provides a near-uniform negative charge to denatured proteins and promotes the electrical control of the translocation kinetics, though it is unclear if the protein-bound SDS could interfere with the nanopore signal's sensitivity to amino acid sequence. Although such denaturation methods are compatible with solid-state nanopores and show great promise for protein translocation, they cannot be as readily applied to protein nanopore systems that include lipid or lipid-like membranes, which are susceptible to harsh conditions required to completely unfold stable proteins (e.g. high temperature or a high concentration of denaturants). To overcome this barrier, several label-based translocation strategies that are compatible with protein nanopores have been explored (Figure 4).

Oligonucleotide-assisted translocation

In nanopore-based DNA and RNA sensing, nucleic acid strands can be electrophoretically-driven into and through a nanopore unidirectionally by an applied voltage as their phosphodiester backbone is intrinsically negatively charged. Hence, attaching an oligonucleotide strand to a protein is a straightforward way to facilitate electrophoresis-driven protein translocation (Figure 4A, Left). Using this approach, Bayley and colleagues have demonstrated controlled directional translocation of proteins through an α-hemolysin nanopore (Rodriguez-Larrea and Bayley, 2013, 2014; Rosen et al., 2020). The group used genetically engineered thioredoxin as a

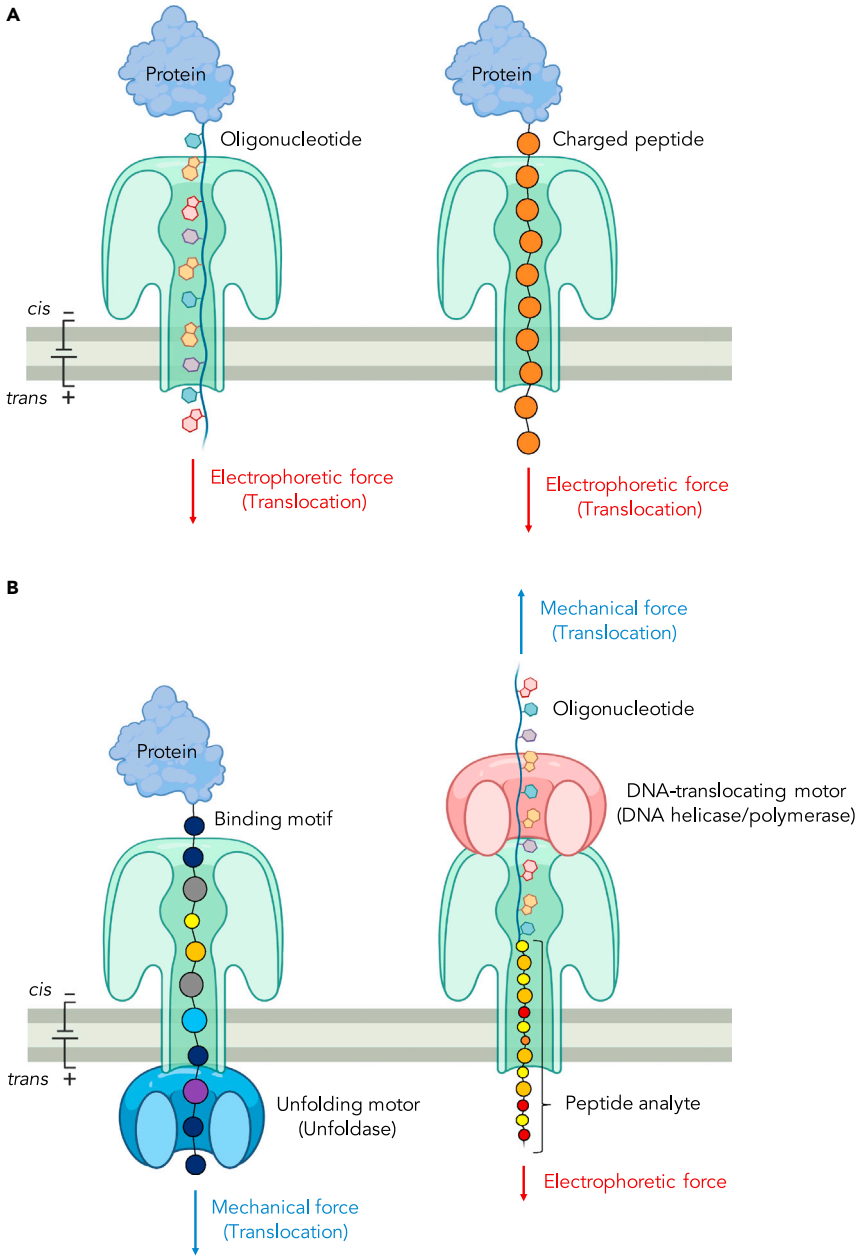

Figure 4. Approaches for protein/peptide translocation through a nanopore

(A) Electrophoresis-driven translocation. Left) The negative charges of oligonucleotides enable electrophoresis-driven unidirectional movement of proteins. Right) Positively-charged (arginine, lysine, and histidine) or negatively-charged (glutamate and aspartate) residues promote unidirectional protein translocation under a negative or positive voltage bias, respectively.

(B) Motor-driven translocation. Left) An unfoldase present in the *trans* solution binds a recognition motif on the threaded protein (e.g., ssrA tag) and generates mechanical force sufficient to denature and processively pull the protein through the pore. Right) A DNA-processive motor bound to an oligonucleotide strand pulls the oligonucleotide-peptide conjugate from the *trans* to *cis* side against electrophoretic force. This allows for the peptide strand to pass through the nanopore constriction with a regular stepping size and yields sequence-dependent signals derived from the peptide strand in the pore's sensitive region.

model protein and linked a C-terminal cysteine side chain to a 30-mer oligo(dC) nucleotide via a disulfide bond. The study demonstrated a co-translocational unfolding mechanism; first, the protein-tethered oligonucleotide is captured by the nanopore. Second, the oligonucleotide is pulled through the nanopore and the C-terminus

of the protein locally unfolds because of the pulling force. Third, the remaining protein structure unfolds spontaneously after being critically destabilized as a result of the C-terminal local structure unfolding. Fourth, the unfolded protein completely translocates through the nanopore. A recent follow-up work to this has also shown that simple neural network-based classifiers can distinguish single point mutations in the protein that result in altered unfolding characteristics (Rodriguez-Larrea, 2021). Other groups have also demonstrated successful translocation of proteins and peptides using tethered oligonucleotides (Biswas et al., 2015; Pastoriza-Gallego et al., 2014), confirming the effectiveness of this translocation approach. This method, however, generates fast translocation events (<1 ms) that may cause poor signal-to-noise ratios and thus make this method less sensitive to protein sequence-level changes. This method may also be ineffective at translocating larger, multi-domain proteins, as the electrophoretic pulling force is largely absent after the oligonucleotide has completely translocated through the nanopore. As molecular motors and nanopore engineering have been employed to reduce the velocity of translocating DNA and RNA in nanopore nucleic acid sequencing (Cherf et al., 2012; Garalde et al., 2018; Manrao et al., 2012; Rincon-Restrepo et al., 2011), techniques to regulate the rate of protein translocation may be required for the acquisition of well-resolved and reproducible current signals.

Charged peptide-assisted translocation

Similar to the oligonucleotide-assisted approach, the fusion of charged peptides promotes unidirectional protein translocation (Figure 4A, Right). Ouldali et al. have demonstrated translocation of XRRRRRRR peptides through an aerolysin nanopore, where X represents the 20 standard amino acids (Ouldali et al., 2020). The translocation experiment was run under a −50 mV bias applied to the trans side to enable the polycationic peptide to move toward the trans compartment by electrophoretic force. Moreover, Restrepo-Pérez et al. have shown translocation of 30 amino acid long peptides, containing 10 glutamates at the N-terminus and 10 arginines at the C-terminus (Restrepo-Pérez et al, 2019a, 2019b). The "tug-of-war" state created by the terminal charged residues enabled slow translocation (>1 ms) of peptides through a FraC nanopore to read out PTMs and 6 distinct chemical tags on the peptides. Our group has recently shown that an Smt3 protein genetically tagged at its C-terminus with a polyGSD peptide can promote capture of the protein in a CsgG nanopore but does not readily drive unfolding and complete translocation (Cardozo et al., 2021). This observation supports the conclusion that a more highly charged peptide or the addition of other mechanisms would be necessary to generate enough force to completely unfold and translocate stably folded proteins through a narrow nanopore.

Molecular motor-assisted translocation

Another approach for protein translocation is based on an unfoldase that enables enzyme-mediated unfolding and translocation of tagged proteins (Figure 4B, Left). We previously employed the AAA+ unfoldase ClpX, which specifically unfolds proteins bearing a C-terminal ssrA peptide tag (AANDENYALAA), for processive unfolding of large proteins (Nivala et al, 2013, 2014). ClpX generates sufficient mechanical force (~20 pN) to denature stable protein folds and translocates proteins at a rate suitable for nanopore sequencing (up to 80 amino acids per second) (Maillard et al., 2011). This approach has demonstrated ClpX-mediated translocation of proteins over 700 amino acids in length, including a variety of protein domains, that are genetically fused with the ssrA tag and a polyanion peptide linker designed to promote protein capture and retention in the nanopore electric field. Distinct protein domains as well as specific point mutations, proteolytic cleavage, and sequence rearrangements in those domains resulted in detectable ionic current pattern changes and single-molecule classification accuracies of 86–99% (Nivala et al., 2014).

While ClpX is capable of unfolding many different types of proteins even with very high stabilities (Kenniston et al., 2004), it likely does not generate sufficient force for some protein folds (Hoskins et al., 2002). In addition, disulfide bonds, which commonly occur in proteins (Bosnjak et al., 2014; Wiedemann et al., 2020), may interrupt protein translocation. The force exerted on protein strands by ClpX or by an electric field (tens to several hundred pN) in a typical experimental setup (Ouldali et al., 2020) is not able to break the covalent disulfide bond (Baldus and Gräter, 2012; Wiita et al., 2007). Although ClpX's ring-like structure is flexible enough to translocate a disulfide-linked beta hairpin into the proteolytic chamber of ClpP (Burton et al., 2001), it is unlikely that a more narrow, rigid nanopore protein would accommodate such a structure. Thus, the use of reducing agents would assist the linear translocation of proteins with disulfide bonds. Another consideration to the unfoldase approach is the large and variable translocation step size of ClpX. Although the fundamental step size of ClpX is ~1 nm, this distance corresponds to an irregular number of amino acids that is dependent on the confirmation of the peptide backbone (typically 5–8 amino acids per 1 nm step). ClpX stepping can also occur in quick bursts of up to 4 nm (Cordova et al., 2014).

Although this bursting activity is critical to its unfolding activity, it could complicate sequencing with single-amino-acid resolution. Efforts to explore alternative unfoldase motors that have a more well-defined step size, such as ClpA (Miller et al., 2013; Olivares et al., 2014), or proteasome systems (Olszewski et al., 2019; Zhang et al., 2020) may be necessary for building a more robust translocation system with optimal resolution.

As an alternative to the unfoldase approach, recent work has explored controlled peptide translocation through nanopores using DNA processive motors. This was accomplished by conjugating peptides to DNA oligonucleotides. A DNA processive motor can then ratchet the peptide-oligonucleotide conjugate through the pore (Figure 4B, Right). This technique has been demonstrated with two different types of DNA processive motors, a helicase (Brinkerhoff et al., 2021) or a polymerase (Yan et al., 2021). The process yields a series of discrete ionic current steps derived from enzyme-mediated peptide-oligonucleotide strand translocation. As the motors are not able to move along the peptide strand itself, the read length is limited to less than the ~30 amino acids that typical nanopore vestibules, such as MspA, can accommodate. This length is still longer than the typical peptide analyzed in bottom-up mass spectrometry and could be further extended by nanopore engineering. While this novel approach overcomes the irregular stepping of ClpX and allows identification of peptide variants in some contexts with single-amino-acid resolution, applications are currently limited to the analysis of synthetic peptide strands with highly negative charges responsive to electrophoretic forces that stretch and trap the peptides within the nanopore (Brinkerhoff et al., 2021); otherwise peptide strands would escape from the nanopore (Yan et al., 2021). A sandwich structure with a negatively-charged strand linked to the other free terminus of a peptide may allow for the peptide to remain captured in the nanopore regardless of its charge distribution.

In summary, key advances have been made toward facilitated translocation of proteins and peptides through narrow protein nanopore sensors for the realization of single-molecule protein sequencing. The development of protein translocation systems has been an exciting and active research area and further improvements are anticipated in the near future.

SITE-SPECIFIC PROTEIN MODIFICATION AND CONJUGATION CHEMISTRY

Nanopore sequencing is an interdisciplinary technology where a wide variety of fields intersect, from protein science and biophysics to chemistry and computer science (Alfaro et al., 2021; Robertson et al., 2021). As the translocation methods discussed above involve the conjugation of translocation-assisting molecules (e.g., oligonucleotides and peptides) to a protein terminus, protein modification and conjugation chemistry are a key step to allow analysis of native proteins, analogous to how nanopore DNA/RNA sequencing requires sample preparation steps such as adapter ligation (Garalde et al., 2018; Jain et al., 2016). To this end, robust bioconjugation molecules such as the azide and alkyne groups can be incorporated into proteins in a site-specific manner (Shadish and DeForest, 2020; Sletten and Bertozzi, 2009). This section reviews conventional amine-specific *N*-hydroxysuccinimide (NHS) ester chemistry, emerging terminal-specific protein modification chemistry, and protein bioconjugation with translocation-assisting molecules (Figure 5).

NHS ester

NHS esters are reagents that target primary amines and form stable amide bonds with both the N-terminal amine and lysine side chains. As the basicity of N-terminal α-amines (pK_a = 6–8) is lower than that of lysine ε-amines (pK_a = ~10.5), a lower pH offers a higher degree of selectivity toward the N-terminal amine at the cost of decreased reactivity (Rosen and Francis, 2017; Sereda et al., 1993). NHS esters are widely used to modify proteins with a variety of functional molecules, such as an azide group and HaloTag ligands (Los et al., 2008), that are then used in subsequent conjugation steps. By preparing translocation-assisting molecules linked to the NHS ester, it would be also possible to directly conjugate proteins with them. The main drawback of NHS esters for N-terminal specific labeling is minimizing the cross reactivity with lysines. For example, RNase A, which has 10 lysine residues, had 2 and 5 simultaneous modifications at pH 6.5 and 7.5, respectively, when reacted with 2-azidoacetic acid NHS ester (Inoue et al., 2019).

N-terminal specific modification

Because it is challenging to achieve complete selectivity toward the N-terminus using NHS esters, much effort has been expended to explore more selective reagents and a growing number of studies are now demonstrating efficient targeting (Hoyt et al., 2019; Rosen and Francis, 2017; Shadish and DeForest,

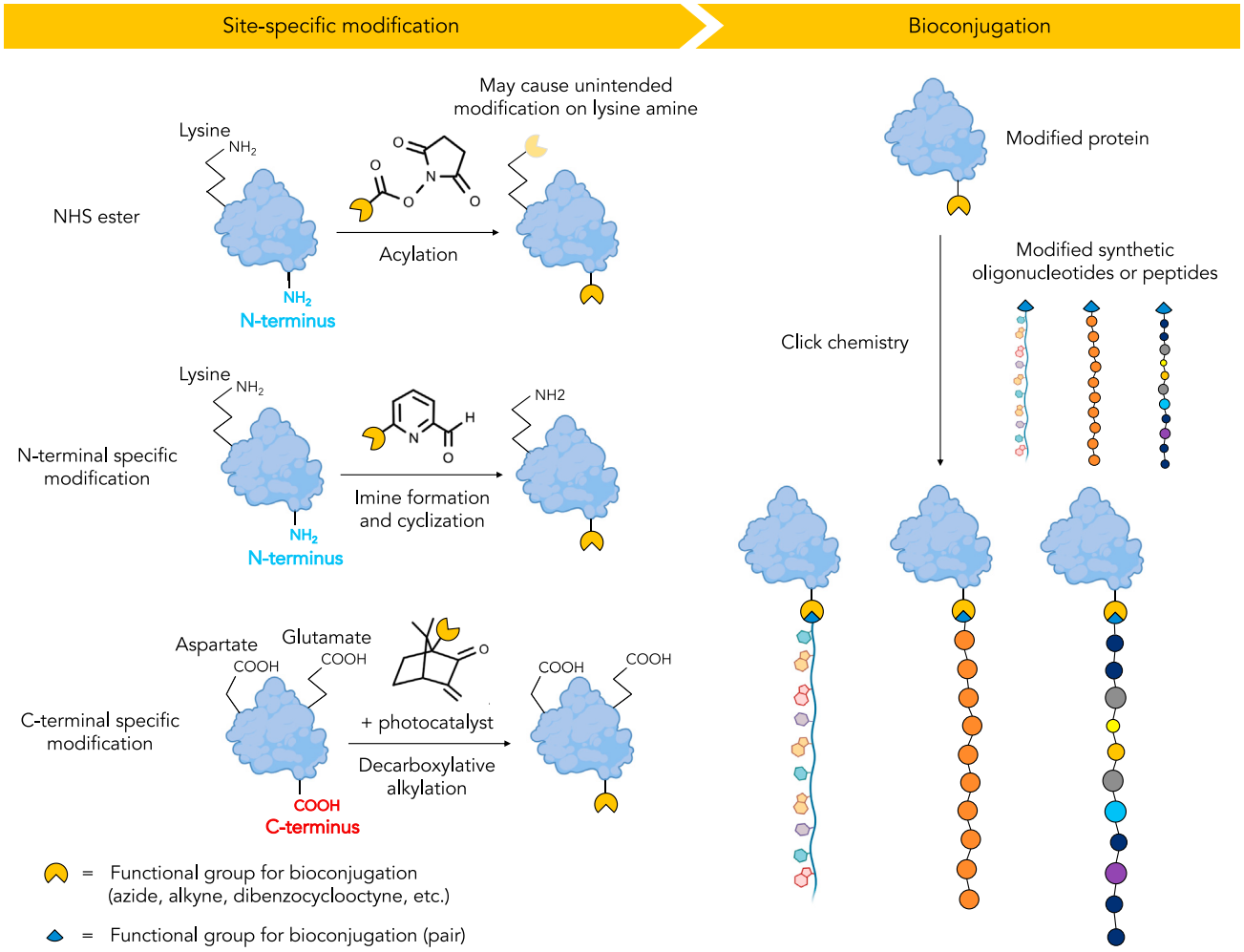

Figure 5. Schematic representation of site-specific protein modification and conjugation

In the site-specific modification step, proteins are modified with functional groups for subsequent bioconjugation. While *N*-hydroxysuccinimide (NHS) esters and 2-pyridinecarboxaldehyde derivatives target the protein N-terminus, photoredox-catalyzed decarboxylation enables C-terminal specific reaction. In the bioconjugation step, the modified proteins are conjugated with nanopore-targeting molecules such as oligonucleotides or peptides via click chemistry.

2020). Notably, the emergence of facile and versatile N-terminal specific modification methods have enabled protein modification at an excellent conversion rate, relying solely on commercially available reagents and thereby circumventing the requirement of complex chemical synthesis steps. In 2015, Francis and colleagues reported one-step modification using 2-pyridinecarboxaldehyde (2PCA) (MacDonald et al., 2015). They found that 2PCA specifically reacts in moderate to excellent conversion rates (43–96%) with the N-termini of peptides and proteins in physiological conditions (37°C, pH 7.5), and is compatible with most N-terminal sequences. The study further demonstrated incorporation of functional molecules such as biotin and fluorescein onto the N-terminus. The reaction depends on the characteristics of 2PCA, which specifically reacts with the N-terminus through the formation of cyclic imidazolidinone. The ε-amino groups of lysine residues are unreactive in this pathway owing to their higher basicity compared to the α-amino group of the N-terminus and the lack of a neighboring amide group suitable for cyclization, yielding no detectable side reaction on the ε-amino groups of lysines. One drawback of this method is that proteins with a proline in their second N-terminal position cannot be modified because of cyclization blocking.

More recently, other groups have also performed similar or revised approaches for N-terminal specific modification (Chen et al., 2017; Deng et al., 2020; Inoue et al., 2019; Li et al., 2018; Onoda et al., 2020).

In 2019, Inoue et al. reported one-step N-terminal specific modification using 6-(azidomethyl)-2-pyridinecarbaldehyde (6AMPC) derivatives (Inoue et al., 2019). This study showed that 6AMPC provides excellent conversions toward various peptides and proteins (e.g., 95% for Angiotensin I and >90% for RNase A). The azide group added to the N-terminus of RNase A was subsequently used for conjugation with fluorescein via copper-catalyzed azide–alkyne cycloaddition (CuAAC; 92% conversion in 60 min) and strain-promoted azide–alkyne cycloaddition (SPAAC; 59% in 60 min), demonstrating the feasibility of linking alkyne-modified molecules to native proteins without the requirement of genetic engineering. In 2020, Onoda et al. developed a more general strategy for one-step modification of the protein N-terminus using triazolecarbaldehyde reagents (Onoda et al., 2020). In principle, this method is applicable for protein modification with any functional molecule containing a primary amine. Indeed, the N-terminus of RNase A was modified with a variety of functional molecules such as biotin, polyethylene glycol, azide, and alkyne groups with conversion rates ranging from 50% to 85%.

C-terminal specific modification

C-terminal specific modification methods have also been explored, though not as thoroughly as N-terminus targeting reactions. In 2018, MacMillan and co-workers developed a method based on light-mediated single-electron transfer as a mechanism enabling site- and chemoselective bioconjugation (Bloom et al., 2018). This approach exploited the innate difference in oxidation potentials between side chains (glutamate and aspartate) and C-terminal α-amino carboxylates (Galicia and González, 2002; Zuo and MacMillan, 2014). In their work, insulin was modified at its C-terminus with an alkyne group with a conversion rate of 41% in 8 h incubation at room temperature and pH 3.5. Although this approach did not show a comparable conversion rate to the N-terminal modification strategies, C-terminal specific modification is useful for targeting native proteins with N-terminals that have been post-translationally modified. While N-terminal acetylation rarely occurs in prokaryotes or archaea, 80–90% of cytosolic mammalian proteins and 50% of yeast proteins are N-terminally acetylated (Polevoda and Sherman, 2003). Thus, N-terminal specific modification needs an additional step for removal of the blocked N-terminus (Gheorghe et al., 1997) when targeting post-translationally N-terminally modified proteins, whereas the C-terminal modification approach can circumvent this requirement.

Protein conjugation chemistry

Protein conjugation is followed by site-specific modification in order to prepare protein conjugates linked to translocation-assisting biomolecules such as oligonucleotides and peptides (Figure 5). SPAAC is one of the most widely used click chemistry reactions for biomolecule conjugation (Khatwani et al., 2012). SPAAC utilizes a pair of reagents, azides and cyclooctynes, that exclusively and efficiently react with each other while remaining inert to naturally occurring functional groups such as amines. SPAAC enables protein labeling with a wide variety of biomolecules without any auxiliary reagents in an aqueous and low temperature (e.g. 4°C) environment that is suitable for protein stability. Using the SPAAC chemistry, Biswas et al. have demonstrated translocation of three different peptides ranging between 8 and 12 amino acids long through a solid-state nanopore (Biswas et al., 2015). In their work, protein N-termini were selectively modified with an azide group and subsequently conjugated with a dibenzocyclooctyne-modified 20-mer oligo(dT) nucleotide via SPAAC for oligo-mediated nanopore translocation. In a similar way, CuAAC is also useful for such tagging reactions. The CuAAC reaction requires the addition of a copper catalyst and a reductive reagent such as sodium ascorbate (Inoue et al., 2019). In addition, there are a variety of other toolkits available for bioconjugation (Hatlem et al., 2019; Rosen et al., 2014a; Rosen and Francis, 2017; Shadish and DeForest, 2020; Sletten and Bertozzi, 2009; Stephanopoulos and Francis, 2011). These bioconjugation approaches are customizable by simply changing chemical groups introduced into proteins in the site-specific modification step. HaloTag, for instance, is a mutant haloalkane dehalogenase that specifically binds synthetic ligands comprising a chloroalkane linker attached to various functional molecules (Los et al., 2008). Covalent bond formation between HaloTag and the chloroalkane linker is fast and essentially irreversible under physiological conditions, enabling versatile protein conjugation with a variety of molecules (England et al., 2015; Yazaki et al., 2020). Thus, these current, and potentially future, conjugation strategies will facilitate protein conjugation with translocation-assisting molecules, enabling controlled protein translocation through a nanopore without the need of genetic engineering.

CONCLUSION AND PERSPECTIVE

This review highlighted the intersection of nanopore sensing, protein chemistry, and bioconjugation strategies. The site-specific bioconjugation chemistries presented here are useful for protein sequencing and identification using nanopore technology as well for other single-molecule approaches (de Lannoy et al., 2021; Howard et al.,

2020; Zhang et al., 2021a). Single-molecule protein sequencing is a highly interdisciplinary endeavor and will require the integration of a wide array of techniques on top of bioconjugation chemistry for successful realization. While advances in bioconjugation chemistry hold great promise to resolve the critical issue of driving protein translocation in nanopore sequencing, the grand challenge to be addressed is accurate decoding of the ionic current signals into amino acid sequences. Indeed, unfoldase-mediated linear translocation of full-length proteins through an α-hemolysin nanopore did not demonstrate reading of amino acids with single-residue resolution (Nivala et al, 2013, 2014). However, recent progress in synergistic areas of research has increased the feasibility of nanopore-based protein sequencing. First, signal-to-noise ratios can be improved by exploring and engineering nanopores suitable for the detection of amino acids. As the average amino acid residue is around half the size of a monophosphate nucleotide, high sensitivity and resolution would likely be achieved by a nanopore with a sharp and narrow constriction, for example, recently developed CsgG or MspA rather than α-hemolysin (Van der Verren et al., 2020; Wang et al., 2018). Second, protein fingerprinting, in which a small subset of residues are orthogonally labeled with chemical tags that produce readily distinguishable nanopore signals, could improve the accuracy of amino acid read-out and protein identification (Ohayon et al., 2019; Yao et al., 2015). This strategy has been adopted in efforts toward nanopore-based protein/peptide identification (Restrepo-Pérez et al., 2019a). Third, the emergence of commercial nanopore sequencers has enabled highly parallel nanopore analysis and high-throughput data collection on a nanopore array (Cardozo et al., 2021; Zhang et al., 2021b), which otherwise would not be realized by single-channel experiments. A large set of training data, coupled with state-of-the-art machine learning models, would boost accurate translation of complex raw signals into amino acid sequences and PTMs, analogous to the example of highly improved basecalling accuracy in nanopore DNA sequencing using deep learning (Rang et al., 2018; Teng et al., 2018; Wick et al., 2019).

Nanopore protein sequencing is a challenging frontier that has yet to be realized, but there is great potential for this nascent technology to revolutionize current proteomics studies, which often suffer from limited sensitivity and throughput in contrast to genomics and transcriptomics technologies. As nanopore sensors provide single-molecule sensitivity, the development of nanopore-based de novo protein sequencing would enable comprehensive identification and quantification of proteins having low abundance or heterogenous PTMs, opening the door to single-molecule proteomics, novel protein biomarker and drug target discovery, and early diagnosis of disease. Further, the low cost and portability of nanopore sensing platforms, in contrast to mass spectrometry instrumentation, open new avenues for more ubiquitous proteomic analyses, for example in the field or in other low resource environments.

ACKNOWLEDGMENTS

K.M. thanks the Japan Society for the Promotion of Science for the Overseas Research Fellowships and BioRender for helping create figures. This work was supported in part by NSF EAGER Award no. 1841188, NSF CCF Award no. 2006864, NIH/NCI Cancer Center Support Grant (no. P30 CA015704) Pilot Award and NSF Award 2021552 to J.N.

AUTHOR CONTRIBUTIONS

K.M. and N.C. wrote the first article draft. K.M., N.C., and J.N. edited the article and conceived of the review topic.

DECLARATION OF INTERESTS

J.N. is a consultant to Oxford Nanopore Technologies. K.M. and N.C. declare no competing interests.

REFERENCES

Aebersold, R., Agar, J.N., Amster, I.J., Baker, M.S., Bertozzi, C.R., Boja, E.S., Costello, C.E., Cravatt, B.F., Fenselau, C., Garcia, B.A., et al. (2018). How many human proteoforms are there? Nat. Chem. Biol. *14*, 206–214. https://doi.org/10.1038/nchembio.2576.

Alfaro, J.A., Bohländer, P., Dai, M., Filius, M., Howard, C.J., van Kooten, X.F., Ohayon, S., Pomorski, A., Schmid, S., Aksimentiev, A., et al. (2021). The emerging landscape of single-molecule protein sequencing technologies. Nat. Methods *18*, 604–617. https://doi.org/10.1038/s41592-021-01143-1.

Asandei, A., Di Muccio, G., Schiopu, I., Mereuta, L., Dragomir, I.S., Chinappi, M., and Luchian, T. (2020). Nanopore-based protein sequencing using biopores: current achievements and open challenges. Small Methods *4*, 1900595. https://doi.org/10.1002/smtd.201900595.

Asandei, A., Rossini, A.E., Chinappi, M., Park, Y., and Luchian, T. (2017). Protein nanopore-based discrimination between selected neutral amino acids from polypeptides. Langmuir *33*, 14451–14459. https://doi.org/10.1021/acs.langmuir.7b03163.

Baldus, I.B., and Gräter, F. (2012). Mechanical force can fine-tune redox potentials of disulfide bonds. Biophys. J. *102*, 622–629. https://doi.org/10.1016/j.bpj.2011.12.039.

Biswas, S., Song, W., Borges, C., Lindsay, S., and Zhang, P. (2015). Click addition of a DNA thread to the N-termini of peptides for their translocation through solid-state nanopores. ACS

Nano *9*, 9652–9664. https://doi.org/10.1021/acsnano.5b04984.

Bloom, S., Liu, C., Kölmel, D.K., Qiao, J.X., Zhang, Y., Poss, M.A., Ewing, W.R., and MacMillan, D.W.C. (2018). Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. Nat. Chem. *10*, 205–211. https://doi.org/10.1038/nchem.2888.

Bosnjak, I., Bojovic, V., Segvic-Bubic, T., and Bielen, A. (2014). Occurrence of protein disulfide bonds in different domains of life: a comparison of proteins from the Protein Data Bank. Protein Eng. Des. Sel. *27*, 65–72. https://doi.org/10.1093/protein/gzt063.

Brinkerhoff, H., Kang, A.S.W., Liu, J., Aksimentiev, A., and Dekker, C. (2021). Infinite re-reading of single proteins at single-amino-acid resolution using nanopore sequencing. bioRxiv. https://doi.org/10.1101/2021.07.13.452225.

Brocchieri, L. (2005). Protein length in eukaryotic and prokaryotic proteomes. Nucl. Acids Res. *33*, 3390–3400. https://doi.org/10.1093/nar/gki615.

Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. *19*, 161. https://doi.org/10.1186/s13059-018-1547-5.

Burton, R.E., Siddiqui, S.M., Kim, Y.I., Baker, T.A., and Sauer, R.T. (2001). Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. EMBO J. *20*, 3092–3100. https://doi.org/10.1093/emboj/20.12.3092.

Cardozo, N., Zhang, K., Doroschak, K., Nguyen, A., Siddiqui, Z., Bogard, N., Strauss, K., Ceze, L., and Nivala, J. (2021). Multiplexed direct detection of barcoded protein reporters on a nanopore array. Nat. Biotechnol. https://doi.org/10.1038/s41587-021-01002-6.

Chen, D., Disotuar, M.M., Xiong, X., Wang, Y., and Chou, D.H.-C. (2017). Selective N-terminal functionalization of native peptides and proteins. Chem. Sci. *8*, 2717–2722. https://doi.org/10.1039/C6SC04744K.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., and Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. Nat. Biotechnol. *30*, 344–348. https://doi.org/10.1038/nbt.2147.

Cordova, J.C., Olivares, A.O., Shin, Y., Stinson, B.M., Calmat, S., Schmitz, K.R., Aubin-Tam, M.-E., Baker, T.A., Lang, M.J., and Sauer, R.T. (2014). Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. Cell *158*, 647–658. https://doi.org/10.1016/j.cell.2014.05.043.

Cressiot, B., Bacri, L., and Pelta, J. (2020). The promise of nanopore technology: advances in the discrimination of protein sequences and chemical modifications. Small Methods *4*, 2000090. https://doi.org/10.1002/smtd.202000090.

Cressiot, B., Braselmann, E., Oukhaled, A., Elcock, A.H., Pelta, J., and Clark, P.L. (2015). Dynamics and energy contributions for transport of unfolded pertactin through a protein nanopore. ACS Nano *9*, 9050–9061. https://doi.org/10.1021/acsnano.5b03053.

Cressiot, B., Oukhaled, A., Patriarche, G., Pastoriza-Gallego, M., Betton, J.-M., Auvray, L., Muthukumar, M., Bacri, L., and Pelta, J. (2012). Protein transport through a narrow solid-state nanopore at high voltage: experiments and theory. ACS Nano *6*, 6236–6243. https://doi.org/10.1021/nn301672g.

de Lannoy, C., Filius, M., van Wee, R., Joo, C., and de Ridder, D. (2021). Evaluation of FRET X for single-molecule protein fingerprinting. bioRxiv. https://doi.org/10.1101/2021.06.30.450512.

Deamer, D.W., and Akeson, M. (2000). Nanopores and nucleic acids: prospects for ultrarapid sequencing. Trends Biotechnol. *18*, 147–151. https://doi.org/10.1016/S0167-7799(00)01426-8.

Deng, J.-R., Lai, N.C.-H., Kung, K.K.-Y., Yang, B., Chung, S.-F., Leung, A.S.-L., Choi, M.-C., Leung, Y.-C., and Wong, M.-K. (2020). N-terminal selective modification of peptides and proteins using 2-ethynylbenzaldehydes. Commun. Chem. *3*, 67. https://doi.org/10.1038/s42004-020-0309-y.

Edman, P., Högfeldt, E., Sillén, L.G., and Kinell, P.-O. (1950). Method for determination of the amino acid sequence in peptides. Acta Chem. Scand. *4*, 283–293. https://doi.org/10.3891/acta.chem.scand.04-0283.

England, C.G., Luo, H., and Cai, W. (2015). HaloTag technology: a versatile platform for biomedical applications. Bioconjug. Chem. *26*, 975–986. https://doi.org/10.1021/acs.bioconjchem.5b00191.

Galicia, M., and González, F.J. (2002). Electrochemical oxidation of tetrabutylammonium salts of aliphatic carboxylic acids in acetonitrile. J. Electrochem. Soc. *149*, D46. https://doi.org/10.1149/1.1450616.

Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods *15*, 201–206. https://doi.org/10.1038/nmeth.4577.

Gheorghe, M.T., Jörnvall, H., and Bergman, T. (1997). Optimized alcoholytic deacetylation of N-acetyl-blocked polypeptides for subsequent Edman degradation. Anal. Biochem. *254*, 119–125. https://doi.org/10.1006/abio.1997.2380.

Han, A., Creus, M., Schürmann, G., Linder, V., Ward, T.R., de Rooij, N.F., and Staufer, U. (2008). Label-free detection of single protein molecules and protein−protein interactions using synthetic nanopores. Anal. Chem. *80*, 4651–4658. https://doi.org/10.1021/ac7025207.

Harrington, L., Cheley, S., Alexander, L.T., Knapp, S., and Bayley, H. (2013). Stochastic detection of Pim protein kinases reveals electrostatically enhanced association of a peptide substrate. Proc. Natl. Acad. Sci.U S A *110*, E4417–E4426. https://doi.org/10.1073/pnas.1312739110.

Hatlem, D., Trunk, T., Linke, D., and Leo, J.C. (2019). Catching a SPY: using the SpyCatcher-SpyTag and related systems for labeling and localizing bacterial proteins. Int. J. Mol. Sci. *20*, 2129. https://doi.org/10.3390/ijms20092129.

Hoskins, J.R., Yanagihara, K., Mizuuchi, K., and Wickner, S. (2002). ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. Proc. Natl. Acad. Sci. U S A *99*, 11037–11042. https://doi.org/10.1073/pnas.172378899.

Howard, C.J., Floyd, B.M., Bardo, A.M., Swaminathan, J., Marcotte, E.M., and Anslyn, E.V. (2020). Solid-phase peptide capture and release for bulk and single-molecule proteomics. ACS Chem. Biol. *15*, 1401–1407. https://doi.org/10.1021/acschembio.0c00040.

Hoyt, E.A., Cal, P.M.S.D., Oliveira, B.L., and Bernardes, G.J.L. (2019). Contemporary approaches to site-selective protein modification. Nat. Rev. Chem. *3*, 147–171. https://doi.org/10.1038/s41570-019-0079-1.

Hu, Z., Huo, M., Ying, Y., and Long, Y. (2021). Biological nanopore approach for single-molecule protein sequencing. Angew. Chem. *133*, 14862–14873. https://doi.org/10.1002/ange.202013462.

Huang, G., Voet, A., and Maglia, G. (2019). FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution. Nat. Commun. *10*, 835. https://doi.org/10.1038/s41467-019-08761-6.

Huang, G., Willems, K., Soskine, M., Wloka, C., and Maglia, G. (2017). Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores. Nat. Commun. *8*, 935. https://doi.org/10.1038/s41467-017-01006-4.

Inoue, N., Onoda, A., and Hayashi, T. (2019). Site-specific modification of proteins through N-terminal Azide labeling and a chelation-assisted CuAAC reaction. Bioconjug. Chem. *30*, 2427–2434. https://doi.org/10.1021/acs.bioconjchem.9b00515.

Jain, M., Olsen, H.E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. *17*, 239. https://doi.org/10.1186/s13059-016-1103-0.

Kasianowicz, J.J., Brandin, E., Branton, D., and Deamer, D.W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. Proc. Natl. Acad. Sci. U S A *93*, 13770–13773. https://doi.org/10.1073/pnas.93.24.13770.

Kennedy, E., Dong, Z., Tennant, C., and Timp, G. (2016). Reading the primary structure of a protein with 0.07 nm3 resolution using a subnanometre-diameter pore. Nat. Nanotechnol. *11*, 968–976. https://doi.org/10.1038/nnano.2016.120.

Kenniston, J.A., Burton, R.E., Siddiqui, S.M., Baker, T.A., and Sauer, R.T. (2004). Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. J. Struct. Biol. *146*, 130–140. https://doi.org/10.1016/j.jsb.2003.10.023.

Khatwani, S.L., Kang, J.S., Mullen, D.G., Hast, M.A., Beese, L.S., Distefano, M.D., and Taton, T.A. (2012). Covalent protein–oligonucleotide

conjugates by copper-free click reaction. Bioorg. Med. Chem. *20*, 4532–4539. https://doi.org/10.1016/j.bmc.2012.05.017.

Li, D., Han, B., Wei, R., Yao, G., Chen, Z., Liu, J., Poon, T.C.W., Su, W., Zhu, Z., Dimitrov, D.S., et al. (2018). N-terminal α-amino group modification of antibodies using a site-selective click chemistry method. mAbs *10*, 712–719. https://doi.org/10.1080/19420862.2018.1463122.

Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., et al. (2008). HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem. Biol. *3*, 373–382. https://doi.org/10.1021/cb800025k.

MacDonald, J.I., Munch, H.K., Moore, T., and Francis, M.B. (2015). One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. Nat. Chem. Biol. *11*, 326–331. https://doi.org/10.1038/nchembio.1792.

Madampage, C.A., Andrievskaia, O., and Lee, J.S. (2010). Nanopore detection of antibody prion interactions. Anal. Biochem. *396*, 36–41. https://doi.org/10.1016/j.ab.2009.08.028.

Maillard, R.A., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C.M., Hodges, C., Martin, A., and Bustamante, C. (2011). ClpX(P) generates mechanical force to unfold and translocate its protein substrates. Cell *145*, 459–469. https://doi.org/10.1016/j.cell.2011.04.010.

Manrao, E.A., Derrington, I.M., Laszlo, A.H., Langford, K.W., Hopper, M.K., Gillgren, N., Pavlenok, M., Niederweis, M., and Gundlach, J.H. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nat. Biotechnol. *30*, 349–353. https://doi.org/10.1038/nbt.2171.

Merstorf, C., Cressiot, B., Pastoriza-Gallego, M., Oukhaled, A., Betton, J.-M., Auvray, L., and Pelta, J. (2012). Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording. ACS Chem. Biol. *7*, 652–658. https://doi.org/10.1021/cb2004737.

Miller, J.M., Lin, J., Li, T., and Lucius, A.L. (2013). E. coli ClpA catalyzed polypeptide translocation is allosterically controlled by the protease ClpP. J. Mol. Biol. *425*, 2795–2812. https://doi.org/10.1016/j.jmb.2013.04.019.

Movileanu, L., Howorka, S., Braha, O., and Bayley, H. (2000). Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. Nat. Biotechnol. *18*, 1091–1095. https://doi.org/10.1038/80295.

Movileanu, L., Schmittschmitt, J.P., Martin Scholtz, J., and Bayley, H. (2005). Interactions of peptides with a protein pore. Biophys. J. *89*, 1030–1045. https://doi.org/10.1529/biophysj.104.057406.

Nivala, J., Marks, D.B., and Akeson, M. (2013). Unfoldase-mediated protein translocation through an α-hemolysin nanopore. Nat. Biotechnol. *31*, 247–250. https://doi.org/10.1038/nbt.2503.

Nivala, J., Mulroney, L., Li, G., Schreiber, J., and Akeson, M. (2014). Discrimination among protein variants using an unfoldase-coupled nanopore. ACS Nano *8*, 12365–12375. https://doi.org/10.1021/nn5049987.

Ohayon, S., Girsault, A., Nasser, M., Shen-Orr, S., and Meller, A. (2019). Simulation of single-protein nanopore sensing shows feasibility for whole-proteome identification. PLOS Comput. Biol. *15*, e1007067. https://doi.org/10.1371/journal.pcbi.1007067.

Ohshiro, T., Tsutsui, M., Yokota, K., Furuhashi, M., Taniguchi, M., and Kawai, T. (2014). Detection of post-translational modifications in single peptides using electron tunnelling currents. Nat. Nanotechnol. *9*, 835–840. https://doi.org/10.1038/nnano.2014.193.

Olivares, A.O., Nager, A.R., Iosefson, O., Sauer, R.T., and Baker, T.A. (2014). Mechanochemical basis of protein degradation by a double-ring AAA+ machine. Nat. Struct. Mol. Biol. *21*, 871–875. https://doi.org/10.1038/nsmb.2885.

Olszewski, M.M., Williams, C., Dong, K.C., and Martin, A. (2019). The Cdc48 unfoldase prepares well-folded protein substrates for degradation by the 26S proteasome. Commun. Biol. *2*, 29. https://doi.org/10.1038/s42003-019-0283-z.

Onoda, A., Inoue, N., Sumiyoshi, E., and Hayashi, T. (2020). Triazolecarbaldehyde reagents for one-step N-terminal protein modification. ChemBioChem *21*, 1274–1278. https://doi.org/10.1002/cbic.201900692.

Oukhaled, G., Mathé, J., Biance, A.-L., Bacri, L., Betton, J.-M., Lairez, D., Pelta, J., and Auvray, L. (2007). Unfolding of proteins and long transient conformations detected by single nanopore recording. Phys. Rev. Lett. *98*, 158101. https://doi.org/10.1103/PhysRevLett.98.158101.

Ouldali, H., Sarthak, K., Ensslen, T., Piguet, F., Manivet, P., Pelta, J., Behrends, J.C., Aksimentiev, A., and Oukhaled, A. (2020). Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nat. Biotechnol. *38*, 176–181. https://doi.org/10.1038/s41587-019-0345-2.

Pagel, O., Loroch, S., Sickmann, A., and Zahedi, R.P. (2015). Current strategies and findings in clinically relevant post-translational modification-specific proteomics. Expert Rev. Proteomics *12*, 235–253. https://doi.org/10.1586/14789450.2015.1042867.

Pastoriza-Gallego, M., Breton, M.-F., Discala, F., Auvray, L., Betton, J.-M., and Pelta, J. (2014). Evidence of unfolded protein translocation through a protein nanopore. ACS Nano *8*, 11350–11360. https://doi.org/10.1021/nn5042398.

Pastoriza-Gallego, M., Rabah, L., Gibrat, G., Thiebot, B., van der Goot, F.G., Auvray, L., Betton, J.-M., and Pelta, J. (2011). Dynamics of unfolded protein transport through an aerolysin pore. J. Am. Chem. Soc. *133*, 2923–2931. https://doi.org/10.1021/ja1073245.

Payet, L., Martinho, M., Pastoriza-Gallego, M., Betton, J.-M., Auvray, L., Pelta, J., and Mathé, J. (2012). Thermal unfolding of proteins probed at the single molecule level using nanopores. Anal. Chem. *84*, 4071–4076. https://doi.org/10.1021/ac300129e.

Piguet, F., Ouldali, H., Pastoriza-Gallego, M., Manivet, P., Pelta, J., and Oukhaled, A. (2018). Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. Nat. Commun. *9*, 966. https://doi.org/10.1038/s41467-018-03418-2.

Polevoda, B., and Sherman, F. (2003). N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J. Mol. Biol. *325*, 595–622. https://doi.org/10.1016/S0022-2836(02)01269-X.

Rang, F.J., Kloosterman, W.P., and de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol. *19*, 90. https://doi.org/10.1186/s13059-018-1462-9.

Restrepo-Pérez, L., Huang, G., Bohländer, P.R., Worp, N., Eelkema, R., Maglia, G., Joo, C., and Dekker, C. (2019a). Resolving chemical modifications to a single amino acid within a peptide using a biological nanopore. ACS Nano *13*, 13668–13676. https://doi.org/10.1021/acsnano.9b05156.

Restrepo-Pérez, L., Wong, C.H., Maglia, G., Dekker, C., and Joo, C. (2019b). Label-free detection of post-translational modifications with a nanopore. Nano Lett. *19*, 7957–7964. https://doi.org/10.1021/acs.nanolett.9b03134.

Restrepo-Pérez, L., John, S., Aksimentiev, A., Joo, C., and Dekker, C. (2017). SDS-assisted protein transport through solid-state nanopores. Nanoscale *9*, 11685–11693. https://doi.org/10.1039/C7NR02450A.

Restrepo-Pérez, L., Joo, C., and Dekker, C. (2018). Paving the way to single-molecule protein sequencing. Nat. Nanotechnol. *13*, 786–796. https://doi.org/10.1038/s41565-018-0236-6.

Rincon-Restrepo, M., Mikhailova, E., Bayley, H., and Maglia, G. (2011). Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes. Nano Lett. *11*, 746–750. https://doi.org/10.1021/nl1038874.

Robertson, J.W.F., Ghimire, M.L., and Reiner, J.E. (2021). Nanopore sensing: a physical-chemical approach. Biochim. Biophys. Acta BBA Biomembr, 183644. https://doi.org/10.1016/j.bbamem.2021.183644.

Robertson, J.W.F., and Reiner, J.E. (2018). The utility of nanopore technology for protein and peptide sensing. Proteomics *18*, 1800026. https://doi.org/10.1002/pmic.201800026.

Rodriguez-Larrea, D. (2021). Single-aminoacid discrimination in proteins with homogeneous nanopore sensors and neural networks. Biosens. Bioelectron. *180*, 113108. https://doi.org/10.1016/j.bios.2021.113108.

Rodriguez-Larrea, D., and Bayley, H. (2014). Protein co-translocational unfolding depends on the direction of pulling. Nat. Commun. *5*, 4841. https://doi.org/10.1038/ncomms5841.

Rodriguez-Larrea, D., and Bayley, H. (2013). Multistep protein unfolding during nanopore translocation. Nat. Nanotechnol. *8*, 288–295. https://doi.org/10.1038/nnano.2013.22.

Rosen, C.B., Bayley, H., and Rodriguez-Larrea, D. (2020). Free-energy landscapes of membrane co-translocational protein unfolding. Commun. Biol. *3*, 160. https://doi.org/10.1038/s42003-020-0841-4.

Rosen, C.B., and Francis, M.B. (2017). Targeting the N terminus for site-selective protein modification. Nat. Chem. Biol. *13*, 697–705. https://doi.org/10.1038/nchembio.2416.

Rosen, C.B., Kodal, A.L.B., Nielsen, J.S., Schaffert, D.H., Scavenius, C., Okholm, A.H., Voigt, N.V., Enghild, J.J., Kjems, J., Tørring, T., et al. (2014a). Template-directed covalent conjugation of DNA to native antibodies, transferrin and other metal-binding proteins. Nat. Chem. *6*, 804–809. https://doi.org/10.1038/nchem.2003.

Rosen, C.B., Rodriguez-Larrea, D., and Bayley, H. (2014b). Single-molecule site-specific detection of protein phosphorylation with a nanopore. Nat. Biotechnol. *32*, 179–181. https://doi.org/10.1038/nbt.2799.

Rotem, D., Jayasinghe, L., Salichou, M., and Bayley, H. (2012). Protein detection by nanopores equipped with aptamers. J. Am. Chem. Soc. *134*, 2781–2787. https://doi.org/10.1021/ja2105653.

Sereda, T.J., Mant, C.T., Quinn, A.M., and Hodges, R.S. (1993). Effect of the α-amino group on peptide retention behaviour in reversed-phase chromatography determination of the pK_a values of the α-amino group of 19 different N-terminal amino acid residues. J. Chromatogr. A *646*, 17–30. https://doi.org/10.1016/S0021-9673(99)87003-4.

Sexton, L.T., Horne, L.P., Sherrill, S.A., Bishop, G.W., Baker, L.A., and Martin, C.R. (2007). Resistive-pulse studies of proteins and protein/antibody complexes using a conical nanotube sensor. J. Am. Chem. Soc. *129*, 13144–13152. https://doi.org/10.1021/ja0739943.

Shadish, J.A., and DeForest, C.A. (2020). Site-selective protein modification: from functionalized proteins to functional biomaterials. Matter *2*, 50–77. https://doi.org/10.1016/j.matt.2019.11.011.

Shi, W., Friedman, A.K., and Baker, L.A. (2017). Nanopore sensing. Anal. Chem. *89*, 157–188. https://doi.org/10.1021/acs.analchem.6b04260.

Sletten, E.M., and Bertozzi, C.R. (2009). Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. Angew. Chem. Int. Ed. *48*, 6974–6998. https://doi.org/10.1002/anie.200900942.

Soskine, M., Biesemans, A., Moeyaert, B., Cheley, S., Bayley, H., and Maglia, G. (2012). An engineered ClyA nanopore detects folded target proteins by selective external association and pore entry. Nano Lett. *12*, 4895–4900. https://doi.org/10.1021/nl3024438.

Specht, H., Emmott, E., Petelski, A.A., Huffman, R.G., Perlman, D.H., Serra, M., Kharchenko, P., Koller, A., and Slavov, N. (2021). Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome Biol. *22*, 50. https://doi.org/10.1186/s13059-021-02267-5.

Steen, H., and Mann, M. (2004). The abc's (and xyz's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. *5*, 699–711. https://doi.org/10.1038/nrm1468.

Steinbock, L.J., Krishnan, S., Bulushev, R.D., Borgeaud, S., Blokesch, M., Feletti, L., and Radenovic, A. (2014). Probing the size of proteins with glass nanopores. Nanoscale *6*, 14380–14387. https://doi.org/10.1039/C4NR05001K.

Stephanopoulos, N., and Francis, M.B. (2011). Choosing an effective protein bioconjugation strategy. Nat. Chem. Biol. *7*, 876–884. https://doi.org/10.1038/nchembio.720.

Sutherland, T.C., Long, Y.-T., Stefureac, R.-I., Bediako-Amoa, I., Kraatz, H.-B., and Lee, J.S. (2004). Structure of peptides investigated by nanopore analysis. Nano Lett. *4*, 1273–1277. https://doi.org/10.1021/nl049413e.

Swaminathan, J., Boulgakov, A.A., Hernandez, E.T., Bardo, A.M., Bachman, J.L., Marotta, J., Johnson, A.M., Anslyn, E.V., and Marcotte, E.M. (2018). Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. Nat. Biotechnol. *36*, 1076–1082. https://doi.org/10.1038/nbt.4278.

Swaminathan, J., Boulgakov, A.A., and Marcotte, E.M. (2015). A theoretical justification for single molecule peptide sequencing. PLOS Comput. Biol. *11*, e1004080. https://doi.org/10.1371/journal.pcbi.1004080.

Sze, J.Y.Y., Ivanov, A.P., Cass, A.E.G., and Edel, J.B. (2017). Single molecule multiplexed nanopore protein screening in human serum using aptamer modified DNA carriers. Nat. Commun. *8*, 1552. https://doi.org/10.1038/s41467-017-01584-3.

Talaga, D.S., and Li, J. (2009). Single-molecule protein unfolding in solid state nanopores. J. Am. Chem. Soc. *131*, 9287–9297. https://doi.org/10.1021/ja901088b.

Teng, H., Cao, M.D., Hall, M.B., Duarte, T., Wang, S., and Coin, L.J.M. (2018). Chiron: translating nanopore raw signal directly into nucleotide sequence using deep learning. GigaScience *7*, giy037. https://doi.org/10.1093/gigascience/giy037.

Thakur, A.K., and Movileanu, L. (2019). Real-time measurement of protein–protein interactions at single-molecule resolution using a biological nanopore. Nat. Biotechnol. *37*, 96–101. https://doi.org/10.1038/nbt.4316.

Timp, W., and Timp, G. (2020). Beyond mass spectrometry, the next step in proteomics. Sci. Adv. *6*, eaax8978. https://doi.org/10.1126/sciadv.aax8978.

Van der Verren, S.E., Van Gerven, N., Jonckheere, W., Hambley, R., Singh, P., Kilgour, J., Jordan, M., Wallace, E.J., Jayasinghe, L., and Remaut, H. (2020). A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. Nat. Biotechnol. *38*, 1415–1420. https://doi.org/10.1038/s41587-020-0570-8.

van Ginkel, J., Filius, M., Szczepaniak, M., Tulinski, P., Meyer, A.S., and Joo, C. (2018). Single-molecule peptide fingerprinting. Proc. Natl. Acad. Sci. U S A *115*, 3338–3343. https://doi.org/10.1073/pnas.1707207115.

Wang, S., Zhao, Z., Haque, F., and Guo, P. (2018). Engineering of protein nanopores for sequencing, chemical or protein sensing and disease diagnosis. Curr. Opin. Biotechnol. *51*, 80–89. https://doi.org/10.1016/j.copbio.2017.11.006.

Wick, R.R., Judd, L.M., and Holt, K.E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. *20*, 129. https://doi.org/10.1186/s13059-019-1727-y.

Wiedemann, C., Kumar, A., Lang, A., and Ohlenschläger, O. (2020). Cysteines and disulfide bonds as structure-forming units: insights from different domains of life and the potential for characterization by NMR. Front. Chem. *8*, 280. https://doi.org/10.3389/fchem.2020.00280.

Wiita, A.P., Perez-Jimenez, R., Walther, K.A., Gräter, F., Berne, B.J., Holmgren, A., Sanchez-Ruiz, J.M., and Fernandez, J.M. (2007). Probing the chemistry of thioredoxin catalysis with force. Nature *450*, 124–127. https://doi.org/10.1038/nature06231.

Wloka, C., Van Meervelt, V., van Gelder, D., Danda, N., Jager, N., Williams, C.P., and Maglia, G. (2017). Label-free and real-time detection of protein ubiquitination with a biological nanopore. ACS Nano *11*, 4387–4394. https://doi.org/10.1021/acsnano.6b07760.

Yan, S., Zhang, J., Wang, Y., Guo, W., Zhang, S., Liu, Y., Cao, J., Wang, Y., Wang, L., Ma, F., et al. (2021). Single molecule ratcheting motion of peptides in a *Mycobacterium smegmatis* Porin A (MspA) nanopore. Nano Lett. https://doi.org/10.1021/acs.nanolett.1c02371.

Yao, Y., Docter, M., van Ginkel, J., de Ridder, D., and Joo, C. (2015). Single-molecule protein sequencing through fingerprinting: computational assessment. Phys. Biol. *12*, 055003. https://doi.org/10.1088/1478-3975/12/5/055003.

Yazaki, J., Kawashima, Y., Ogawa, T., Kobayashi, A., Okoshi, M., Watanabe, T., Yoshida, S., Kii, I., Egami, S., Amagai, M., et al. (2020). HaloTag-based conjugation of proteins to barcoding-oligonucleotides. Nucl. Acids Res. *48*, e8. https://doi.org/10.1093/nar/gkz1086.

Ying, Y.-L., Yang, J., Meng, F.-N., Li, S., Li, M.-Y., and Long, Y.-T. (2019). A nanopore phosphorylation sensor for single oligonucleotides and peptides. Research *2019*, 1–8. https://doi.org/10.34133/2019/1050735.

Zhang, L., Floyd, B.M., Chilamari, M., Mapes, J., Swaminathan, J., Bloom, S., Marcotte, E.M., and Anslyn, E.V. (2021a). Photoredox-catalyzed decarboxylative *C*-terminal differentiation for bulk and single molecule proteomics. bioRxiv. https://doi.org/10.1101/2021.07.08.451692.

Zhang, L., Gardner, M.L., Jayasinghe, L., Jordan, M., Aldana, J., Burns, N., Freitas, M.A., and Guo, P. (2021b). Detection of single peptide with only one amino acid modification via electronic fingerprinting using reengineered durable channel of Phi29 DNA packaging motor. Biomaterials, 121022. https://doi.org/10.1016/j.biomaterials.2021.121022.

Zhang, S., Huang, G., Versloot, R., Herwig, B.M., de Souza, P.C.T., Marrink, S.-J., and Maglia, G. (2020). Bottom-up fabrication of a multi-component nanopore sensor that unfolds, processes and recognizes single proteins. bioRxiv. https://doi.org/10.1101/2020.12.04.411884.

Zhao, Y., Ashcroft, B., Zhang, P., Liu, H., Sen, S., Song, W., Im, J., Gyarfas, B., Manna, S., Biswas, S., et al. (2014). Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. Nat. Nanotechnol. *9*, 466–473. https://doi.org/10.1038/nnano.2014.54.

Zhu, Y., Clair, G., Chrisler, W.B., Shen, Y., Zhao, R., Shukla, A.K., Moore, R.J., Misra, R.S., Pryhuber, G.S., Smith, R.D., et al. (2018a). Proteomic analysis of single mammalian cells enabled by microfluidic nanodroplet sample preparation and ultrasensitive NanoLC-MS. Angew. Chem. Int. Ed. *57*, 12370–12374. https://doi.org/10.1002/anie.201802843.

Zhu, Y., Piehowski, P.D., Zhao, R., Chen, J., Shen, Y., Moore, R.J., Shukla, A.K., Petyuk, V.A., Campbell-Thompson, M., Mathews, C.E., et al. (2018b). Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. Nat. Commun. *9*, 882. https://doi.org/10.1038/s41467-018-03367-w.

Zubarev, R.A. (2013). The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics *13*, 723–726. https://doi.org/10.1002/pmic.201200451.

Zuo, Z., and MacMillan, D.W.C. (2014). Decarboxylative arylation of α-amino acids via photoredox catalysis: a one-step conversion of biomass to drug pharmacophore. J. Am. Chem. Soc. *136*, 5257–5260. https://doi.org/10.1021/ja501621q.