

*Annual Review of Biophysics*

# Protein Sequencing, One Molecule at a Time

Brendan M. Floyd and Edward M. Marcotte

Department of Molecular Biosciences, Center for Systems and Synthetic Biology, University of Texas, Austin, Texas, USA; email: bmfloyd@utexas.edu, marcotte@utexas.edu

Annu. Rev. Biophys. 2022. 51:181–200

First published as a Review in Advance on January 5, 2022

The *Annual Review of Biophysics* is online at [biophys.annualreviews.org](http://biophys.annualreviews.org)

<https://doi.org/10.1146/annurev-biophys-102121-103615>

Copyright © 2022 by Annual Reviews.  
All rights reserved

## Keywords

single-molecule protein sequencing, SMPS, proteomics, fluorescence, fluorosequencing, nanopores, DNA nanotechnology

## Abstract

Despite tremendous gains over the past decade, methods for characterizing proteins have generally lagged behind those for nucleic acids, which are characterized by extremely high sensitivity, dynamic range, and throughput. However, the ability to directly characterize proteins at nucleic acid levels would address critical biological challenges such as more sensitive medical diagnostics, deeper protein quantification, large-scale measurement, and discovery of alternate protein isoforms and modifications and would open new paths to single-cell proteomics. In response to this need, there has been a push to radically improve protein sequencing technologies by taking inspiration from high-throughput nucleic acid sequencing, with a particular focus on developing practical methods for single-molecule protein sequencing (SMPS). SMPS technologies fall generally into three categories: sequencing by degradation (e.g., mass spectrometry or fluorosequencing), sequencing by transit (e.g., nanopores or quantum tunneling), and sequencing by affinity (as in DNA hybridization-based approaches). We describe these diverse approaches, which range from those that are already experimentally well-supported to the merely speculative, in this nascent field striving to reformulate proteomics.

## ANNUAL REVIEWS CONNECT

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

## Contents

1. INTRODUCTION .....	182
2. SEQUENCING BY DEGRADATION .....	183
2.1. Fluorosequencing .....	184
2.2. N-Terminal-Specific Amino Acid Binders .....	186
2.3. Single-Molecule Mass Spectrometry .....	187
3. SEQUENCING BY TRANSIT .....	187
3.1. Nanopore Sequencing .....	187
3.2. Electron Tunneling .....	191
3.3. ClpXP-Based Fluorescence Resonance Energy Transfer .....	192
4. SEQUENCING BY AFFINITY .....	192
5. OUTLOOK .....	193

## 1. INTRODUCTION

It is no exaggeration to say that next-generation DNA and RNA sequencing have revolutionized the study of biology. The ability to sequence DNA and RNA in many different cellular contexts has led to novel discoveries about the underlying causes of genetic diseases (58), organismal development (36), and tracing cell lineages (102), among innumerable others. These methods were invented to sequence nucleic acids at both the bulk and single-molecule levels and capitalized on many technical advances ranging in scope from replicating nucleic acids with polymerases in a manner that provides a fluorescent readout to threading nucleic acids through nanopores and measuring the resulting perturbations to electric current (6, 24, 30). Not only have these new methods increased the throughput and accuracy of nucleic acid sequencing, they have also massively decreased costs. While this substantial growth in nucleic acid sequencing approaches has occurred at a rapid clip, sequencing methods for proteins, especially at the single-molecule level, have largely lagged. As a consequence, knowledge about proteins' abundances, processing, modifications, localizations, and interactions has not benefited to the same degree as has parallel knowledge about nucleic acids, motivating a recent, significant push by researchers to develop entirely new modes of protein sequencing.

The primary amino acid sequence of a protein was first obtained by Frederic Sanger and his colleagues (87, 88) in the early 1950s using the combination of difluoronitrobenzene, protein hydrolysis, and paper chromatography. This early work was quickly superseded by a method first presented by Pehr Edman (29) that used phenylisothiocyanate (PITC) as a reagent for the step-wise degradation of a protein, accompanied by chromatographic detection of each subsequent detached amino acid derivative. Despite several issues, such as the need for a free amino terminus, automated Edman sequencers remained the gold standard of proteomics until the 1990s and the rise of mass spectrometry-based proteomics (93). Early mass spectrometry experiments for proteomics relied heavily on a bottom-up approach, which involved digesting proteins into shorter peptides and identifying them using tandem mass spectrometry (i.e., liquid chromatography with tandem mass spectrometry) (107). Mass spectrometry has remained a standard method for protein identification and quantification up to the present and has seen remarkable increases in its sensitivity and applications in recent years (12). However, limitations in the dynamic range and sensitivity of conventional protein mass spectrometry, particularly when compared to what is achievable by nucleic acid sequencing methods, are serving to stimulate efforts to invent new methods for

**Table 1 Analogies between available nucleic acid sequencing technologies and single-molecule protein sequencing methods**

Sequencing category	DNA sequencing method	Protein sequencing method
Sequencing by synthesis	Polymerase-based sequencing (e.g., Sanger, Illumina, Pacific Biosciences)	None, as no practical chemical or enzymatic copying mechanisms are known for proteins
Sequencing by degradation	Enzymatic digest sequencing (28, 54)	Fluorosequencing NAAB binding Nanopore mass spectrometry
Sequencing by transit	Nanopore sequencing	Nanopore sequencing Electron tunneling ClpXP-based fingerprinting
Sequencing by affinity	DNA microarrays Fluorescence in situ hybridization	DNA nanotechnology (DNA-PAINT, FRET X, DNA nanoscope)

Abbreviations: DNA-PAINT, DNA point accumulation in nanoscale topology; FRET X, fluorescence resonance energy transfer using DNA eXchange; NAAB, N-terminal-specific amino acid binder.

sequencing proteins (44, 115). Single-molecule protein sequencing (SMPS) has emerged as a viable approach for identifying and quantifying proteins at the ultimate possible (single-molecule) sensitivity. Importantly for both methods that might rely on matching to a reference sequence database and *de novo* (i.e., reference-free) methods for protein sequencing, even gaining partial sequence information can often be sufficiently informative to identify many proteins.

Although the field is still nascent, researchers of SMPS have already made significant achievements and developed a strikingly diverse repertoire of technologies (2, 16, 78, 98). Some incorporate aspects previously developed for proteomics or nucleic acid sequencing, such as Pehr Edman’s sequencing chemistry or nanopores akin to those used for DNA sequencing. Others depart radically from tradition and take advantage of entirely new detection modalities, such as DNA hybridization–based methods. Nonetheless, there are clear parallels with approaches developed for nucleic acid sequencing, and it is helpful to consider the developing proteomics technologies by analogy to four broad categories of DNA sequencing methods: sequencing by synthesis, sequencing by degradation, sequencing by transit, and sequencing by affinity (**Table 1**). It is particularly worth noting that the dominant technique used for nucleic acids (sequencing by synthesis)—the basis, for example, for Illumina’s, Sanger’s, and Pacific Biosciences’ sequencing methods—has no obvious parallel for proteins. There are as yet no practical methods, neither enzymatic nor chemical, for copying amino acid polymers molecule by molecule. Without protein-dependent protein polymerases, reverse ribosomes, or polymerase chain reaction for proteins, for example, sequencing by synthesis has yet to be explored as a viable single-molecule proteomics technology.

Thus, in this review, we explore the approaches that currently appear viable. We discuss the remaining broad sequencing categories, describing the most recent developments for SMPS approaches in each category. There are increasingly many proposed approaches for sequencing proteins and peptides at the single-molecule level, and several remain speculative, with the only associated data found in patent applications. To focus this review, we opt to restrict discussion to those methods for which a theoretical justification or proof-of-principle experiment has been reported in a preprint or published, with an emphasis on recent developments for each technology.

## 2. SEQUENCING BY DEGRADATION

In contrast to nucleic acid sequencing, proteomics is historically rooted in sequencing-by-degradation approaches, from Frederick Sanger and colleagues’ (87, 88) degradative methods for

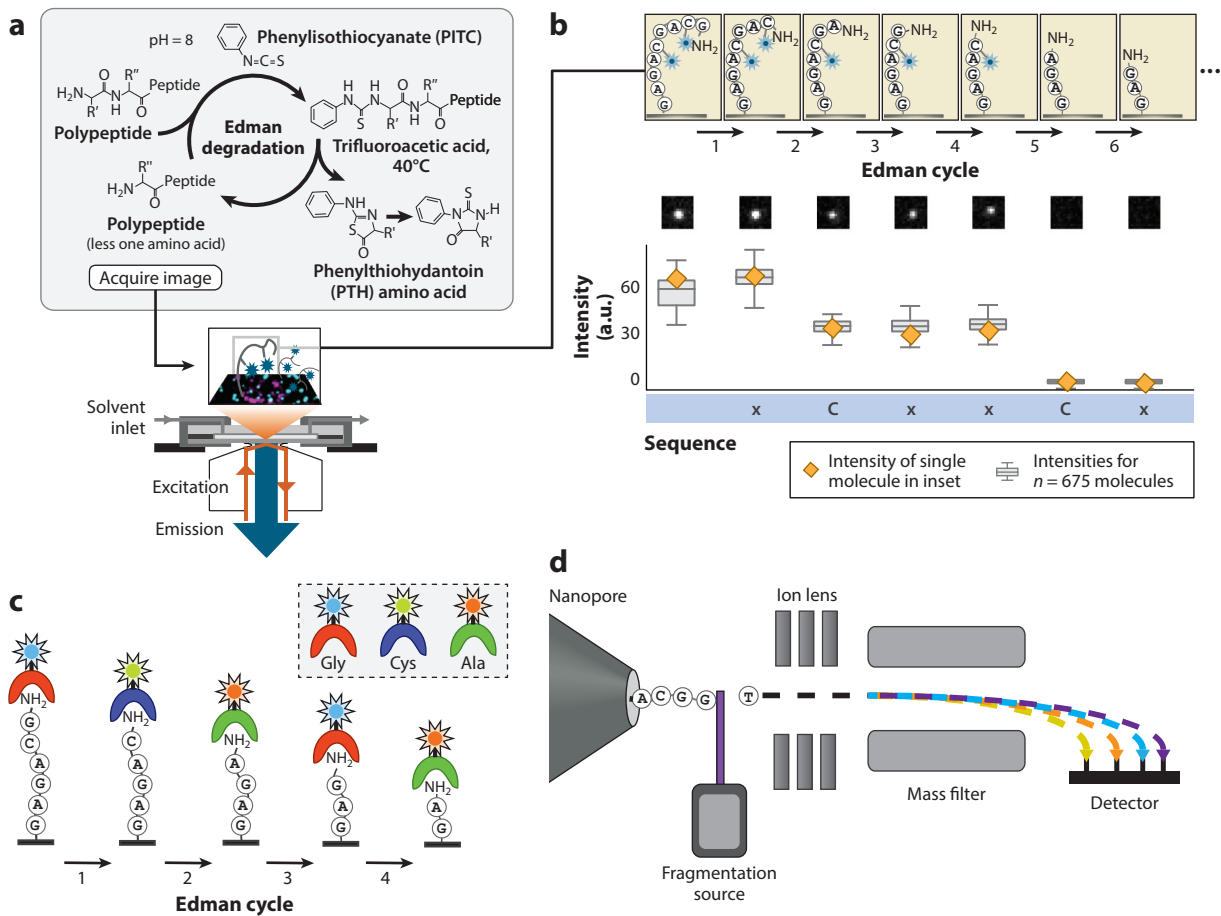
identifying the primary sequences of the A and B chains of insulin to the subsequent methods of Edman degradation (see 29) and tandem mass spectrometry (62). Each of the SMPS methods covered in this section borrows aspects of these older, tried-and-true proteomics methods but extends them to the sensitivity of individual molecules and implements them with an eye to scalability to full proteomes.

## 2.1. Fluorosequencing

One of the first high-throughput SMPS methods developed, named fluorosequencing (96), combines single-molecule microscopy of fluorescent peptides with Pehr Edman's degradative chemistry, enabling many individual peptide molecules to be monitored in parallel in a sequencing flow cell as their consecutive N-terminal amino acids are removed in a cyclic fashion (**Figure 1a**). Edman degradation relies on the reactivity of PITC with the  $\alpha$ -amino group of a protein or peptide (29). Following this reaction, the first amino acid can be removed from the protein in the presence of a strong acid, leaving behind a protein with a fresh amino terminus that is one amino acid shorter in length (**Figure 1a**). Historically, Edman sequencers were applied only to purified (homogenous) proteins, so the detached amino acid derivatives could then be identified by bulk chromatography. A protein's primary sequence was ultimately obtained by repeating this chemistry and subsequent chromatography in a cyclical fashion (55, 97). The fluorosequencing method reported by Swaminathan et al. (95) relies instead on proteolytic digestion of proteins into peptides, followed by chemically labeling select amino acid types on the peptides with amino acid-specific fluorophores and subjecting the labeled peptides to successive rounds of Edman degradation. Rather than monitoring the released amino acid derivatives, fluorosequencing analyzes the retained parent peptides, thus enabling analysis of protein mixtures, with the sequence of each molecule independently recorded by total internal reflectance fluorescence microscopy across cycles of Edman degradation. The resulting fluorosequence for each peptide molecule indicates the sequence positions at which the fluorescently labeled amino acids were removed (**Figure 1b**).

Fluorosequencing is an example of a bottom-up proteomics approach, determining partial sequences of individual proteolytic peptides or protein fragments that are subsequently matched to a reference database for protein identification, taking advantage of a known proteome or genome sequence. It has been shown that, instead of the full *de novo* identification of each amino acid on a protein being required, identifying the sequence positions of a few amino acids (fingerprinting) can often be sufficient to identify proteins from a reference proteome (73, 96, 109). In practice, fluorosequencing is a hybrid method, identifying the sequence positions of labeled amino acids *de novo*, then matching those partial sequences to a reference database to identify proteins. A simulation study showed that minimal labeling schemes (comprising 2–4 labeled amino acid types) in combination with various endoprotease choices can in principle be used to identify a large proportion of the proteome (96).

Bioconjugation strategies with high specificity are thus important for many SMPS fingerprinting approaches to identify proteins successfully. Lysine and cysteine have both been covalently labeled with fluorophores at high efficiency and sequenced using fluorosequencing (95), as well as being detected using fluorescence (also known as Förster) resonance energy transfer (FRET) using DNA eXchange (FRET X) (31) and ClpXP FRET (101), techniques discussed further below. Additional labeling strategies have been demonstrated for labeling up to six amino acids (KDYWEC), with glutamate and aspartate being indistinguishable (39). In concert with side-chain labeling, SMPS may require an N-terminal labeling strategy for reversibly blocking the N terminus (required for lysine labeling) and a C-terminal labeling strategy, which is necessary to conjugate a linker for slide coupling. In a recent study, Howard et al. (41) applied the reagent 2-pyridinyl carboxaldehyde (PCA), which had previously been shown to specifically form a stable



**Figure 1**

Sequencing-by-degradation approaches. (a) Schematic of fluorosequencing, which uses single-molecule microscopy to monitor the reduction in fluorescence from immobilized, fluorescently labeled peptides following consecutive rounds of Edman degradation. Panel a adapted with permission from Reference 95. (b) Example data from Reference 95 showing a peptide with labeled cysteines (blue stars) undergoing fluorosequencing. (Top) Peptide sequence after each successive Edman cycle. (Middle) Fluorescence emitted from one peptide molecule across Edman cycles. (Bottom) Fluorescence quantified across 675 individual molecules, with the characteristic stairstep decreases indicating positions of the fluorescently labeled cysteines. (c) A proposed strategy to combine Edman degradation with amino acid detection using N-terminal-specific amino acid binders (NAABs) (8, 83, 100). In one possible implementation, a palette of NAABs might be chosen that have weak but varying affinity for different amino acids, allowing for peptide sequencing based on binding kinetics (83). (d) Schematic representation of the nanopore mass spectrometer under development by Stein and colleagues (14). As single peptides or proteins are fed through the pore, a fragmentation source releases each amino acid, which will ultimately be identified by an ion detector. It has been speculated that a UV-laser fragmentation source could be used to dissociate amino acids.

covalent attachment with the N terminus of proteins, to fluorosequencing (see 59). They showed that peptides captured onto beads coated in PCA could be labeled with fluorescent dyes released from the solid support and sequenced by fluorosequencing with efficiencies comparable to that observed when the common protecting group fluorenylmethoxycarbonyl was used for solution-phase labeling. For C-terminal labeling, Bloom et al. (7) reported a covalent reaction, which successfully differentiates between the C terminus and the side chains of aspartate and glutamate, that could be adapted to this approach. This approach was optimized for proteomics by Zhang

et al. (112) and found to be 76% efficient for labeling human cell extract tryptic peptides. Furthermore, using this approach, peptides were successfully immobilized and fluorosequenced following conjugation of an alkyne linker to the C terminus of the peptide.

While fluorosequencing has been successfully implemented to determine sparse amino acid sequence information for individual protein molecules for thousands to millions of molecules in parallel (95), there are several technical hurdles that still must be overcome for widespread adoption of this sequencing approach. The first is the stability of the fluorophores used as labels. The propensity of the harsh Edman reagents to inactivate the fluorophores on the peptides requires a palette of fluorophores that are highly chemically stable (95). The length of the Edman cycle (approximately 1 h/cycle) has also given rise to some concern, especially when considering sequencing peptides of longer length, although it is worth noting that the approach is not unlike many DNA sequencing approaches in being highly parallel in nature. Borgo & Havranek (9) proposed an alternative approach of using a designed cysteine protease, called an Edmanase, that would cleave the N-terminal amino acid and would diminish the need for the harsh reagents used for the Edman degradation reaction. Regardless, while sequencing of full-length proteins has yet to be shown, quantification of diagnostic peptides in samples of reduced complexity should be possible with this approach at the present.

Finally, in addition to protein identification, another goal of SMPS technologies is to identify post-translational modifications (PTMs) such as phosphorylation and glycosylation. PTM identification using tandem mass spectrometry often requires enrichment for modified peptides from larger amounts of starting sample, and this process can result in biases in the detected PTMs (65, 80). SMPS methods offer a potentially powerful approach to identifying and quantifying PTMs directly by sequencing. Initial experiments have confirmed the ability of single-molecule approaches to localize PTMs: Phosphoserines were fluorosequenced after covalently labeling them with fluorophores (95), and, as discussed further below, electron tunneling (71) and nanopores (56, 79) have each been successfully employed to distinguish phosphorylated proteoforms.

## 2.2. N-Terminal-Specific Amino Acid Binders

Another potential strategy for sequencing peptides by degradation considers N-terminal-specific amino acid binders (NAABs) as an alternative to chemical labeling. NAABs are members of a class of proteins whose affinity depends on the N-terminal amino acid. NAAB-based sequencing therefore relies on the sequential removal of the N-terminal amino acid to obtain sequence information (8, 83), alternating removal with NAAB-based identification of the new terminal amino acid (**Figure 1c**). One looming question for this approach is how to generate a palette of NAABs with high specificity for a majority of the proteinogenic amino acids. One strategy presented by Tullman et al. (100) is to use directed evolution of an enzyme, in this case the substrate recognition protein ClpS, to generate a protein with high specificity for specific N-terminal amino acids. Using successive rounds of a yeast display screen, they were able to obtain multiple variants with increased binding affinity to Phe, Tyr, or Trp. While some background affinity for the wild-type biases (Phe, Trp, Tyr, Leu) still remains, increasing the affinity for these amino acids over the background amino acids is a promising step for future applications to sequencing methods. To show how a NAAB-based SMPS platform could work in practice, Rodriques et al. (83) studied whether the weak binding affinity of NAABs could provide sufficient information for protein identification. They proposed the use of kinetic measurements on a set of NAABs, measuring their association and dissociation rates at the single-molecule level, to assign identities to the N-terminal amino acid of an immobilized peptide. In a series of simulations that accounted for various error modes, they found that high accuracies of amino acid identification were in principle achievable by such a

method. The use of NAABs in practice for protein sequencing has yet to be experimentally shown, but it is an interesting alternative to covalent labeling strategies, such as those in fluorosequencing, for which chemical destruction of dyes has previously been observed (95).

### 2.3. Single-Molecule Mass Spectrometry

Mass spectrometry has been the gold standard for protein identification for at least the past two decades (1, 110). Improved instrumentation and sampling over the years has increased the sensitivity of mass spectrometers to the point that single-cell proteomics via mass spectrometry has become experimentally possible (21, 64, 74). Mass spectrometry utilizes the charging of peptides or proteins—often by electrospray ionization—to detect and identify the ion based on its mass-per-charge ratio ( $m/z$ ) (93). Single-ion detection was originally made possible by Fourier-transform ion cyclotron resonance instruments but was limited to the characterization of a molecule of interest, rather than sequencing (49, 51, 84, 92). It has been posited that the placement of a nanopore as the ion source would allow for the sequencing of single proteins using mass spectrometry (14). This approach would rely on the threading of proteins through the pore and subsequent dissociation of individual amino acids at the exit of the pore, which would release an amino acid for detection (**Figure 1d**). The sequential readout of the dissociated amino acids will allow for the sequencing of proteins through passage (66). While a dissociation method for this approach has not yet been shown, it has been proposed that a method such as UV-induced dissociation could perform the task (2, 11). Due to the sequential nature of the nanopore mass spectrometer, throughput could potentially be challenging, and it remains to be seen if the process could be parallelized or, indeed, what throughput might be feasible.

## 3. SEQUENCING BY TRANSIT

Sequencing by transit has become an increasingly popular approach for single-molecule nucleic acid sequencing, in large part due to the ongoing technological advances in nanopore sequencing. Commercialized by Oxford Nanopore Technologies, the nanopore DNA sequencer has proved to be a powerful platform for long-read single-molecule DNA sequencing while also affording high portability and field applications (47). Not surprisingly, nanopore sequencing has also promisingly been applied to SMPS, as have other sequencing-by-transit methods such as electron tunneling and a protease-based FRET sequencing approach. Each of these approaches threads proteins or peptides of interest through a sensor, which differs according to the specific approach, e.g., a nanometer-sized pore, gold electrodes, or a fluorescently labeled AAA+ protease. This section explores recent developments and applications of each of these different sequencing-by-transit approaches for protein sequencing.

### 3.1. Nanopore Sequencing

The passage of a biopolymer such as DNA through a nanometer-sized pore as detected by an electric current has rapidly moved from a laboratory proof-of-concept to successful commercialization by Oxford Nanopore Technologies over the past two and a half decades (24, 37). Nanopore sequencing works by separating two compartments containing an ionic buffer with an organic, inorganic, or hybrid pore (50). A current is established across the cell containing the separated compartments and pore. The passage of a molecule through the pore from the *cis* to the *trans* compartment results in a blockaded current that corresponds to the volume of the transiting molecule. In the case of DNA sequencing, the current blockade corresponds to the volume of the individual



nucleobases, and strings of bases can be threaded through the pore to give the DNA sequence in a length-independent manner.

The combination of long-read capability and portability of nanopore sequencers is currently energizing the genomics field, and there has been a great push to apply this same technology to protein sequencing (78). However, some clear challenges emerge for applying this technology to protein sequencing. First among these is that proteins and peptides are far more heterogeneous in composition than are nucleic acids, with some amino acids such as leucine and isoleucine being largely indistinguishable by charge or occluded volume. Additionally, unlike DNA, proteins do not have a homogenous charge, which complicates the capture of proteins and their translocation across a pore. Many of the technical challenges for nanopore-based protein sequencing have been well reviewed (23, 43), so this section largely focuses on recent advances in applying this technology to protein sequencing. Some of the main challenges in the field include how to construct pores to support efficient capture and translocation of peptides and proteins, how to unfold proteins prior to their passing through the pore, and how to best read out the amino acid sequence from a transiting protein.

**3.1.1. Capture and translocation.** One of the main requirements for nanopore protein sequencing is the capture and translocation of the protein through the pore. In early nanopore sequencing experiments using silicon nitride pores, only several dozen translocation events were recorded over the course of several days, revealing challenges with the capture and passage of proteins (16, 52). These observed rates of passage would both seriously hinder read depth and increase experimental run time, so extensive effort has been put into developing new methods for capturing and translocating proteins through the pore. Unlike in the case of sequencing DNA, where electrophoretic forces can be used to draw molecules from the *cis* to the *trans* side of a nanopore (94), the nonuniform charge distribution of proteins makes this approach untenable on unmodified proteins (43). One of the initial methods developed for getting around this translocation issue was to attach a highly charged polymer such as DNA or a peptide composed of charged amino acids (e.g., glutamate or arginine) to one or both of the proteins' termini to support electrophoretically driven translocation (79, 81, 82). This approach may prove challenging for whole proteomes, as it requires either extensive chemical modification or genetic tagging, motivating researchers to seek other means of translocating proteins across pores.

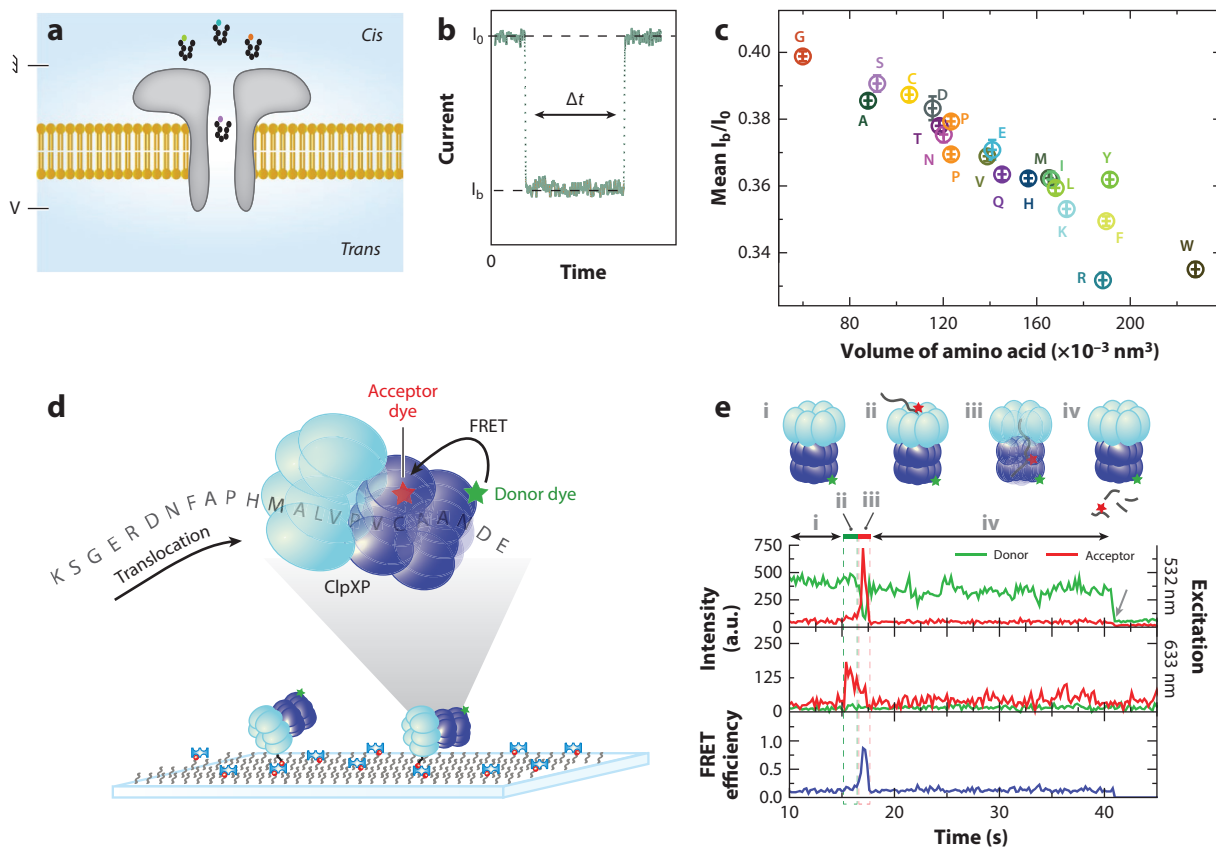
One promising alternative approach is the use of electroosmotic flow (EOF) as a capture and translocation mechanism that can overcome the heterogeneous charge distribution of polypeptides. EOF is the flow generated by ions moving in solution under the influence of an electric field, inducing the surrounding water molecules to also move toward the cathode or anode (57). In an early study showing the use of EOF for capturing peptides, the rate of current blockages in an  $\alpha$ -hemolysin pore was directly related to a lower buffer pH (67), interpreted as the increased presence of cations in the lower pH buffers providing more ions to drive liquid movement toward the *trans* anode. In a follow-up study using an  $\alpha$ -hemolysin pore at pH 2.8, EOF could overcome the electrophoretic effect on a cationic peptide, supporting the viability of this approach for translocation (3). Peptides could be retained and recaptured into the center, or lumen, of the nanopore with this approach, offering a means to repeatedly resample a peptide with the nanopore. More recently, it was demonstrated that it is possible to engineer a nanopore to generate an EOF and capture folded proteins (46). Rather than relying on the flow generated by a low pH buffer, this study used site-specific mutagenesis to incorporate positively charged residues inside the pore to help to generate an EOF at pH 7.5. Ren and colleagues (76) also developed field-effect transistors (FETs) coupled with nanopores, where the gate medium for the FET controls the flow of ions, thus adding selectivity to the capture and translocation of proteins. In a follow-up study, adding



protein-specific (thrombin-binding) aptamers to the shell of the nanopipette used for the FET further increased selectivity in sampling. While EOF has been shown to translocate peptides and folded proteins through pores, it has not yet been determined whether full-length proteins can be linearized in the nanopore lumen using only these forces. It is possible that, to overcome the entropic barriers of protein unfolding, a detergent or additional enzyme, such as an unfoldase, may be necessary.

Enzyme-driven unfolding and translocation have thus been seen as promising approaches for overcoming several of the technical hurdles to nanopore use. For DNA sequencing, the use of a modified version of the polymerase phi29 as a motor to ratchet DNA strands through a nanopore in a controlled manner greatly improved base calling accuracy (20, 63). Deliberately slowing down pore traversal is also likely to benefit protein sequencing, and the use of exoproteases as molecular ratchets has been viewed as a feasible strategy for doing so. In one of the earliest examples of nanopore sequencing with proteins, Nivala et al. (68) used a variant of the ATP-dependent chaperone ClpX that was placed on the *trans* side of an  $\alpha$ -hemolysin pore to unfold and thread proteins containing an ssrA tag across the pore. In a follow-up study, the Akeson group showed that a nanopore used in conjunction with ClpXP (which contains ClpX as well as the proteolytic ClpP assembly) could discriminate variants of a titin fragment contained in a modified protein (69). It should be noted that the motor enzymes in these studies were placed on the *trans* side of the pores, so the step sizes taken by proteins transiting through the pores were dictated by the activities and structures of the motor enzymes. This use of motor enzymes as molecular ratchets speaks to the promise of exploring analogs of DNA sequencing approaches for next-generation proteomics platforms.

**3.1.2. Amino acid identification using nanopores.** Great strides have been made in identifying amino acids using nanopore sequencing, with a particularly clear example of the ability of nanopores to discriminate the current blockades of the proteinogenic amino acids described by Ouldali and colleagues (72). Using a polyarginine peptide with a variable C-terminal amino acid, they showed that most of the 20 canonical amino acids could be differentiated based on residual current and blockage volume in a wild-type aerolysin pore (**Figure 2a–c**). These results, in conjunction with the differential estimates of peptide length recorded in a previous study (75), support the promise of nanopores for discriminating peptide sequences to a high degree. To date, nanopore protein sequencing experiments have predominantly used wild-type  $\alpha$ -hemolysin and aerolysin nanopores, but there has also been a push to engineer biological pores with improved characteristics. Cao et al. (17) showed that the aerolysin pore sensitivity can be increased by rationally engineering its charge and diameter. In a more radical application of rational engineering to nanopore design, Zhang et al. (113) constructed a proteasome-coupled nanopore that is capable of translocating unfolded proteins when the proteasome component is inactive. As an alternative to this thread-and-read approach, when the proteasome component was active, this pore could digest proteins into 6- to 10-amino-acid-long peptides, which might then—at least, in principle—be identified in the nanopore using a chop-and-drop strategy. A somewhat related strategy for identifying individual amino acids was demonstrated by Wei et al. (105). By using the aromatic reagents 2,3-naphthalenedicarboxaldehyde and 2-naphthylisothiocyanate as coupling reagents to the N termini of free amino acids, they showed that these additional tags boosted the distinguishability of the nine amino acids that were tested as they translocated across an  $\alpha$ -hemolysin pore. Since free amino acids are required for identification with this approach, actual application would presumably require a protease or chemical reagent that cleaves amino acids processively so they could then be consecutively identified; such a combined approach has not yet been reduced to practice.



**Figure 2**

Sequencing-by-transit approaches. (a) Schematic of a nanopore peptide sequencing experiment. An external voltage (V) is applied on the *trans* side of the pore, and a voltage ground is applied on the *cis* side, represented by the symbol on the top left side of the panel. (b) Depiction of a typical current blockade.  $I_0$  symbolizes the open pore ionic current, and  $I_b$  symbolizes the blocked current. (c) Mean relative residual current ( $I_b/I_0$ ) and its standard deviation from the measurement of  $X_{R7}$  (X being any amino acid) peptides. (d) Schematic of the ClpX single-molecule protein sequencing approach. Donor dye-labeled ClpX is immobilized on a slide surface, and an ssrA-conjugated peptide labeled with an acceptor dye is threaded through the complex. (e) A fluorescence time trace that captures the different moments of translocation. (i) The donor signal from Cy3-labeled ClpX excited at 532 nm (top green trace). (ii) Appearance of acceptor signal due to acceptor-directed excitation at 633 nm (middle red trace) indicates the binding of the acceptor-labeled peptide to ClpX. (iii) The fluorescence resonance energy transfer (FRET) signal (bottom blue trace) reports on the presence of the peptide in ClpX. (iv) Loss of fluorescent and FRET signals indicates the release of the peptide. Panels a–c adapted with permission from Reference 72. Panels d and e adapted with permission from Reference 101.

In a strategy aimed at a more direct sequence readout, Brinkerhoff and colleagues (10) attached a peptide to a DNA strand and pulled the peptide–DNA conjugate through an MspA nanopore via the action of a helicase acting on the DNA segment. The result was clear stepwise ion current traces such as are more typically seen in DNA nanopore sequencing experiments. Importantly, Brinkerhoff and colleagues demonstrated that peptides differing at individual amino acids gave rise to distinguishable step traces, supporting the notion that direct amino acid sequencing might be possible with such an approach, and that they could repeatedly resample the peptide, which should help in achieving low sequencing error rates. Such approaches and others (86) offer

intriguing potential routes forward for fingerprinting and identifying proteins that would leverage nanopores' abilities to discriminate amino acid types and potentially even modifications (45, 79).

In fact, protein fingerprinting can be achieved in different ways, and some current nanopore implementations rely on chemical labeling of specific residues such as lysine and cysteine. Simulations by Ohayon et al. (70) showed that three labels could allow for discrimination of up to 97% of the human proteome by monitoring fluorescence intensity and protein translocation through the nanopore. Promisingly, multiple groups have demonstrated the feasibility of identifying labeled amino acids with a nanopore. Wang et al. (104) showed that fluorescence microscopy and TiO<sub>2</sub> nanopore detection could discriminate identical polypeptides with different fluorescent labels. This approach could, in principle, allow for the quantitation and localization of labels on a polypeptide. In another example of fingerprinting with a nanopore, Restrepo-Pérez et al. (77) used a fragaceatoxin C nanopore to identify six different labels that, when conjugated to cysteine at various positions, could be resolved site specifically. Fingerprinting proteins using nanopores may serve as a springboard for nanopore sequencing while full de novo sequencing strategies continue to develop.

Work done with nanopores has also shown the potential of PTM identification without any enrichment or labeling step. In one such study, Restrepo-Pérez et al. (79) showed that, on a dipolar peptide, phosphorylated and glycosylated modifforms could be distinguished from the unmodified peptide. Importantly, they observed that modified peptides tended to have an extended dwell time within the pore, which could result in better characterization of the modified peptide. In a later study, Li et al. (56) used an engineered aerolysin nanopore to identify adjacent phosphorylation sites on the medically relevant Tau peptide. Notably, three variants of the phosphorylated Tau peptide were added to the same nanopore chamber sequentially and were ably distinguished, indicating that, even in samples with multiple proteoforms, discrimination will still be possible (at least in low-complexity samples).

SMPS using nanopores is one of the most mature of the technologies covered in this review, but there are still challenges to its widespread adoption, among them the need for better computational strategies for discriminating amino acids. In one interesting approach, Rodriguez-Larrea (81) trained a neural network to identify single-amino-acid differences on a point mutant of thioredoxin. This study focused on a limited set of nine amino acids, but future studies will need to focus on discriminating at least the 20 proteinogenic amino acids. Label-free identification of amino acids and their modified variants is very promising for the future of proteomics. In addition to sequencing, nanopores are also being used to profile folded proteins and obtain signatures based on the different orientations adopted by the protein in the pore (40, 90, 111). It can be imagined that the nanopore sensor will turn into a multitool for single-molecule protein analysis.

### 3.2. Electron Tunneling

Whereas nanopore-based sequencing relies on the electrical properties of biomolecules in the longitudinal direction (i.e., down the strand of the polymer), electron, or quantum, tunneling surveys the electrical properties in the transverse direction (27). In the case of DNA, this would mean the probing of individual bases perpendicular to the DNA strand (116), and in the case of peptides and proteins, this would mean probing amino acid side chains perpendicular to the peptide bond. Measurements for electron tunneling involve measuring the electron tunneling current of a subunit sandwiched between two electrodes (typically gold electrodes) with a set gap distance. For proteins and peptides, it is hoped that a distinct tunneling signature will be observed for each amino acid as the polypeptide passes through a pore containing the electrodes.

Two pioneering studies have shown the feasibility of applying electron tunneling to SMPS. The first, by Zhao et al. (114), used a 2-nm electrode gap to distinguish among modified forms of the same amino acid, among enantiomers of asparagine, and leucine from isoleucine to show the effectiveness of the approach. Using a support vector machine classifier that was trained on the electronic signature of each amino acid sampled, this strategy for electron tunneling performed well enough to detect distinct signals between the enantiomers L- and D-asparagine. This study sampled only monomeric amino acids, so when considering how to sequence proteins, Zhao et al. proposed a microreactor with an *exo*-peptidase that can release a terminal amino acid that can then be sampled. A following study by Ohshiro et al. (71) showed that electron tunneling could be used to discriminate among 12 amino acids, including distinguishing the phosphorylated form of tyrosine from the unphosphorylated variant. This required the use of two sizes of nanogaps—the distance between the sandwiching electrodes—of 0.55 and 0.7 nm. Of the 12 amino acids sampled, there was a clear discriminatory power for approximately half of the amino acids with the two nanogap sizes when both dwelling time and conductance were used to form an electronic signature. In addition, when sampling a mixed solution of tyrosine and phosphotyrosine at different ratios, it was possible to distinguish between the two based on their individual signals. Importantly, by using two peptides that had matching sequences except for a single residue, Ohshiro et al. were able to show that peptides can be measured using this approach—all of the previous measurements were with monomeric amino acids—and that single-residue differences can result in distinguishable electronic signatures. It is worth noting that distinguishable residues such as tyrosine and phenylalanine could be localized on the peptide with site specificity, offering a clear protein fingerprinting option.

### 3.3. ClpXP-Based Fluorescence Resonance Energy Transfer

A unique protein fingerprinting method, first introduced in a simulation study by Yao et al. (109), relies on the enzyme ClpX to gather sequential information on labeled amino acids. ClpX is an AAA+ protease that functions as a molecular motor that processively unfolds and degrades proteins (4, 61). By tethering ClpX labeled with a donor dye to a slide surface, which can then unfold and translocate a protein labeled on cysteines and lysines with acceptor dyes, a linear fingerprint of these residues can be observed by FRET (101) (**Figure 2d,e**). Van Ginkel and colleagues (101) demonstrated that peptides labeled with either one or two fluorophores and containing an N-terminal *ssrA* tag for ClpP recognition could be differentiated based on FRET signals. They further showed that differences in the wild-type and mutant variants of the protein titin could be distinguished based on their FRET signal. They noted that titin's tertiary structure was disrupted by the labeling process, suggesting that labels might potentially be applied even to sites buried in the native 3D structure, and moreover, that the labels would sufficiently disrupt the structure for subsequent ClpX fingerprinting. It is worth noting that the requirement of an *ssrA* tag for translocation through ClpX makes scaling this method to the full proteome difficult in practice.

## 4. SEQUENCING BY AFFINITY

Affinity-based strategies have been used for decades to identify and quantify DNA in mixtures, along with methods such as DNA microarray technologies (35, 38, 103). DNA microarrays use the hybridization of sample DNA to thousands of DNA strands coupled onto a slide surface to identify the sample sequence. The corresponding fluorescent signals give a rough estimate of gene expression in the cell (53). For proteins, affinity-based identification is extremely common thanks to the widespread use of antibodies as tools for protein assays and imaging (5, 13, 32, 48, 99, 106,

108). Antibody-based protein profiling assays, such as those for the Human Protein Atlas project, and other affinity-based methods, such as the aptamer-based approach used by Somalogic, are being successfully applied to complex samples and at large scale, but as they profile proteins rather than sequence them, they are excluded from this section (33, 60, 85). The technologies covered in this section largely rely on the hybridization of DNA to DNA strands covalently attached to a peptide or protein to read out amino acid composition.

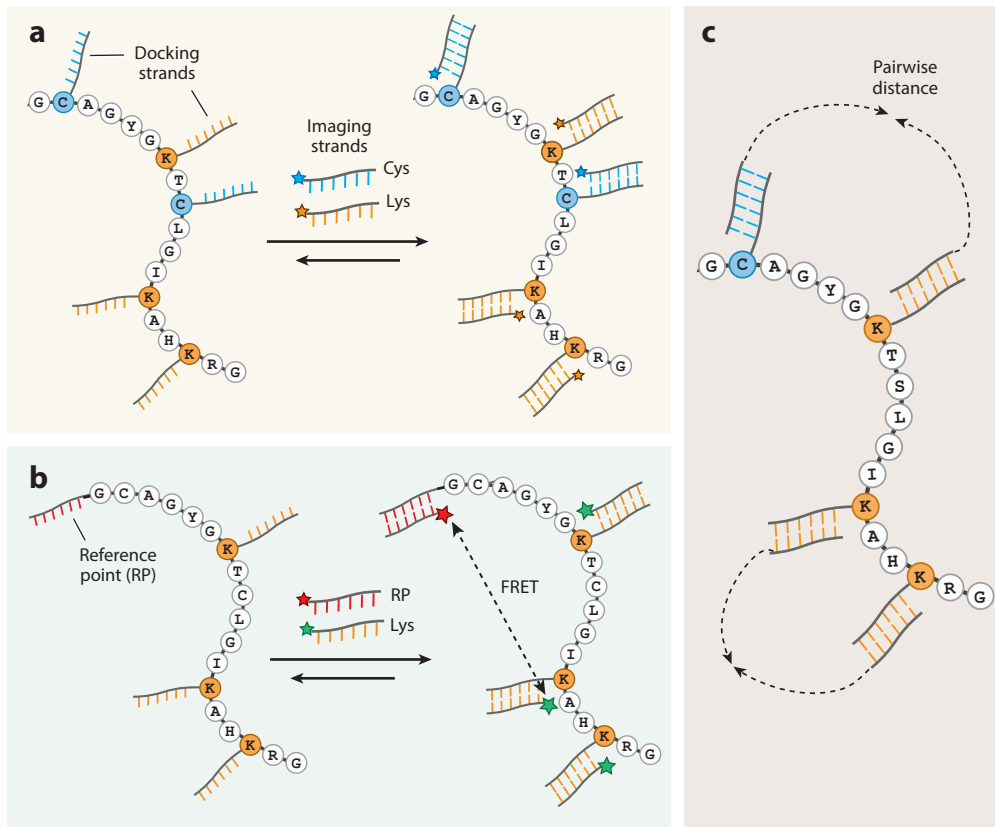
DNA nanotechnology has been applied for use in fields ranging from drug delivery to disease diagnostics (18, 19, 22, 42) and, in the field of imaging, to perform nanoscale topography. Super-resolution microscopy of DNA probes has been achieved by accumulating the centroids of observed fluorescence events from the DNA probes [DNA point accumulation in nanoscale topology (DNA-PAINT)] (91). DNA-PAINT relies on the programmable nature of DNA strand hybridization between docking and imaging strands to gain fine-spatial-resolution compositional information about a molecule of interest (**Figure 3a**). This spatial resolution can reach below 5 nm. It has been proposed that, if individual protein chains could be linearized, then the labeling of lysines or cysteines (or both) with docking strands and imaging with imager strands could be used to obtain a protein fingerprint (25). By one estimate, localizing lysines in proteins within 5-nm imaging resolution might allow more than half of the proteome to be uniquely identified, extending to three-quarters if lysines and cysteines were labeled with unique strands and similarly localized (2). In practice, it has yet to be shown how linearization will be achieved for this approach, although approaches akin to optical tweezers might possibly be applied (15).

Another strategy that uses DNA probes to measure amino acid composition and spatial organization is termed FRET X. Rather than directly localizing DNA imaging strands, as in DNA-PAINT, FRET X measures dye–dye distances via FRET as a readout (31). In FRET X, a single reference DNA strand is attached to either the C or the N terminus of the protein, and multiple docking strands are conjugated onto accessible reactive amino acids such as lysines or cysteines. One imaging strand containing an acceptor fluorophore is hybridized to the reference point DNA. An imaging strand containing a donor fluorophore can then be hybridized to one of the amino acid–conjugated DNA strands, and the distance between the reference point (C or N terminus) and the labeled amino acid position can be measured based on the acceptor–donor fluorophore FRET efficiency (**Figure 3b**). By sampling across multiple such amino acid–conjugated strands, a histogram of FRET distances can be collected and used as a fingerprint of the peptide or protein of interest. De Lannoy et al. (26) determined by simulation that, with cysteine, lysine, and arginine labels, FRET X could, in principle, be used to identify proteins with 95% accuracy in a sample of approximately 300 human proteins. This study also experimentally demonstrated the capability of FRET X to measure the distance of an amino acid (cysteine in this case) at variable distances along a peptide chain from the N terminus of the peptide. Importantly, unlike DNA-PAINT, FRET X is applicable to folded proteins, removing the need to linearize proteins prior to fingerprinting.

Finally, DNA nanoscopy represents a DNA nanotechnology approach that relies on DNA amplification rather than docking and imager strands. The DNA nanoscope uses a process termed autocycling proximity recording to essentially build a molecular ruler out of DNA that can be used to measure the distance between conjugated molecules such as amino acids (34, 89) (**Figure 3c**). Proof-of-principle experiments have shown that this approach can work to reconstruct DNA nanostructure organization, but it has yet to be demonstrated on a protein sample.

## 5. OUTLOOK

The recent ongoing development in SMPS technologies in many ways parallels advances made for DNA and RNA sequencing. In addition to the technologies covered in this review, there have



**Figure 3**

Sequencing-by-affinity approaches. (a) Depiction of the DNA point accumulation in nanoscale topology (DNA-PAINT) method and its proposed application to protein fingerprinting (25, 91). Proteins are first covalently labeled with docking DNA strands that contain sequences specific to the amino acids being labeled. Imaging DNA strands conjugated with fluorophores hybridize with docking strands for specific amino acids to provide compositional information about a peptide or protein. In principle, by linearizing a protein and visualizing it by super-resolution microscopy, relative positional information for the labeled amino acids can also be obtained. (b) Depiction of the fluorescence resonance energy transfer (FRET) using DNA eXchange (FRET X) method (31) and an application to protein fingerprinting (26). Proteins are labeled with docking DNA strands in addition to an acceptor dye on the protein terminus, the latter serving as a fixed reference point. Imaging strands containing a donor dye are then hybridized onto the docking strands. The resulting FRET signals are proportional to distance, providing a distance histogram characteristic of the protein's sequence and fold. (c) Depiction of the proposed DNA nanoscope protein fingerprinting method (34, 89). DNA strands are amplified in such a way as to record distances between pairs of oligonucleotide-conjugated amino acids, with the resulting histogram of pairwise distances giving information about the order and position of the labeled amino acids.

been numerous additional proposals and patent applications for more speculative technologies on which there is as yet no published data or information. In fact, there seems to be a perfect storm of factors propelling the SMPS field forward: a focused push by the technology community to develop new proteomics approaches, the strong progress made to date helping to spur new entrants to the field, the potential applicability of many interesting and powerful strategies from nucleic acid sequencing, and a drive to commercialize SMPS technologies. We expect that the coming years will witness even more diverse protein sequencing technologies, creating a rich ecosystem of complementary SMPS technologies.

In light of this, it is worth noting that, in many instances, the goal of protein sequencing will often not be to obtain the sequences themselves, but rather to identify and quantify the proteins. In fact, we can expect the technologies under development to have very different strengths for sequencing versus identification versus quantification. In the world of nucleic acid sequencing, short-read, high-throughput approaches (as are common for, e.g., Illumina sequencing) stand in contrast to long-read, moderate-throughput approaches (e.g., those for Pacific Biosciences or Oxford Nanopore sequencing), and specialized applications leverage these strengths, just as RNA sequencing benefits from the deep read counts of short-read sequencing and metagenomics or genome assembly benefits from longer reads. We similarly expect short-read, highly parallelized SMPS technologies (e.g., fluorosequencing) to offer deep quantification, while long-read, moderate-throughput SMPS approaches (e.g., nanopores) promise characterization of long proteoforms or combinatorial mapping of PTMs. SMPS technologies will almost certainly be differentiated along other axes as well, including their abilities to perform absolute versus relative quantification, their abilities to identify novel modifications versus the detection and cataloging of known ones, their abilities to quantify distinct proteoforms, and—a major factor that has not received the attention it deserves—the complexity of their sample preparations and their applicability to low-abundance samples.

Nonetheless, the field of SMPS is still young, and most of the technologies with published proof-of-principle results have yet to demonstrate substantial results on complex protein samples, much less in studies of the most challenging proteomics samples, such as single-cell analysis of proteins or high-throughput PTM discovery. We are very excited to see these bleeding-edge SMPS approaches make the leap from technology development to broader applications, fostered by a growing community effort to push back the boundaries of the field.

## DISCLOSURE STATEMENT

E.M.M. is a cofounder and shareholder of Erisyon, Inc., and serves on the scientific advisory board. B.M.F. and E.M.M. are coinventors on granted patents or pending patent applications related to single-molecule protein sequencing.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge Jagannath Swaminathan and Eric Anslyn for helpful discussion, as well as research support from the Welch Foundation (F-1515); Erisyon, Inc.; and the National Institutes of Health (grants R35 GM122480, R01 HD085901, and R21 HD103588).

## LITERATURE CITED

1. Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. *Nature* 422(6928):198–207
2. Alfaro JA, Bohländer P, Dai M, Filius M, Howard CJ, et al. 2021. The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* 18(6):604–17
3. Asandei A, Schiopu I, Chinappi M, Seo CH, Park Y, Luchian T. 2016. Electroosmotic trap against the electrophoretic force near a protein nanopore reveals peptide dynamics during capture and translocation. *ACS Appl. Mater. Interfaces* 8(20):13166–79
4. Aubin-Tam M-E, Olivares AO, Sauer RT, Baker TA, Lang MJ. 2011. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* 145(2):257–67
5. Avrameas S. 1969. Coupling of enzymes to proteins with glutaraldehyde: use of the conjugates for the detection of antigens and antibodies. *Immunochemistry* 6(1):43–52
6. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218):53–59



7. Bloom S, Liu C, Kölmel DK, Qiao JX, Zhang Y, et al. 2018. Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* 10(2):205–11
8. Borgo B, Havranek JJ. 2014. Motif-directed redesign of enzyme specificity. *Protein Sci.* 23(3):312–20
9. Borgo B, Havranek JJ. 2015. Computer-aided design of a catalyst for Edman degradation utilizing substrate-assisted catalysis. *Protein Sci.* 24(4):571–79
10. Brinkerhoff H, Kang ASW, Liu J, Aksimentiev A, Dekker C. 2021. Infinite re-reading of single proteins at single-amino-acid resolution using nanopore sequencing. bioRxiv 2021.07.13.452225. <https://doi.org/10.1101/2021.07.13.452225>
11. Brodbelt JS. 2014. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* 43(8):2757–83
12. Brunner A-D, Thielert M, Vasilopoulou CG, Ammar C, Coscia F, et al. 2021. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. bioRxiv 2020.12.22.423933. <https://doi.org/10.1101/2020.12.22.423933>
13. Burnette WN. 1981. “Western blotting”: electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* 112(2):195–203
14. Bush J, Maulbetsch W, Lepoitevin M, Wiener B, Mihovilovic Skanata M, et al. 2017. The nanopore mass spectrometer. *Rev. Sci. Instrum.* 88(11):113307
15. Bustamante C, Alexander L, Maciuba K, Kaiser CM. 2020. Single-molecule studies of protein folding with optical tweezers. *Annu. Rev. Biochem.* 89:443–70
16. Callahan N, Tullman J, Kelman Z, Marino J. 2020. Strategies for development of a next-generation protein sequencing platform. *Trends Biochem. Sci.* 45(1):76–89
17. Cao C, Cirauqui N, Marcaida MJ, Buglakova E, Duperrex A, et al. 2019. Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores. *Nat. Commun.* 10:4918
18. Chandrasekaran AR, Punnoose JA, Zhou L, Dey P, Dey BK, Halvorsen K. 2019. DNA nanotechnology approaches for microRNA detection and diagnosis. *Nucleic Acids Res.* 47(20):10489–505
19. Chen T, Ren L, Liu X, Zhou M, Li L, et al. 2018. DNA nanotechnology for cancer diagnosis and therapy. *Int. J. Mol. Sci.* 19(6):1671
20. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. 2012. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* 30(4):344–48
21. Cheung TK, Lee C-Y, Bayer FP, McCoy A, Kuster B, Rose CM. 2021. Defining the carrier proteome limit for single-cell proteomics. *Nat. Methods* 18(1):76–83
22. Chi Q, Yang Z, Xu K, Wang C, Liang H. 2020. DNA nanostructure as an efficient drug delivery platform for immunotherapy. *Front. Pharmacol.* 10:1585
23. Chinappi M, Cecconi F. 2018. Protein sequencing via nanopore based devices: a nanofluidics perspective. *J. Phys. Condens. Matter* 30(20):204002
24. Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S, Bayley H. 2009. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* 4(4):265–70
25. Dai M, Jungmann R, Yin P. 2016. Optical imaging of individual biomolecules in densely packed clusters. *Nat. Nanotechnol.* 11(9):798–807
26. de Lannoy C, Filius M, van Wee R, Joo C, de Ridder D. 2021. Evaluation of FRET X for single-molecule protein fingerprinting. bioRxiv 2021.06.30.450512. <https://doi.org/10.1101/2021.06.30.450512>
27. Di Ventra M, Taniguchi M. 2016. Decoding DNA, RNA and peptides with quantum tunnelling. *Nat. Nanotechnol.* 11(2):117–26
28. Donis-Keller H, Maxam AM, Gilbert W. 1977. Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res.* 4(8):2527–38
29. Edman P. 1949. A method for the determination of amino acid sequence in peptides. *Arch. Biochem.* 22(3):475
30. Eid J, Fehr A, Gray J, Luong K, Lyle J, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323(5910):133–38
31. Filius M, Kim SH, Severins I, Joo C. 2021. High-resolution single-molecule FRET via DNA eXchange (FRET X). *Nano Lett.* 21(7):3295–301

32. Free RB, Hazelwood LA, Sibley DR. 2009. Identifying novel protein-protein interactions using co-immunoprecipitation and mass spectroscopy. *Curr. Protoc. Neurosci.* 46:5.28.1–14
33. Gold L, Ayers D, Bertino J, Bock C, Bock A, et al. 2010. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLOS ONE* 5(12):e15004
34. Gopalkrishnan N, Punthambaker S, Schaus TE, Church GM, Yin P. 2020. A DNA nanoscope that identifies and precisely localizes over a hundred unique molecular features with nanometer accuracy. *bioRxiv* 2020.08.27.271072. <https://doi.org/10.1101/2020.08.27.271072>
35. Govindarajan R, Duraiyan J, Kaliyappan K, Palanisamy M. 2012. Microarray and its applications. *J. Pharm. Bioallied Sci.* 4(Suppl. 2):S310–12
36. He P, Williams BA, Trout D, Marinov GK, Amrhein H, et al. 2020. The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Nature* 583(7818):760–67
37. Heather JM, Chain B. 2016. The sequence of sequencers: the history of sequencing DNA. *Genomics* 107(1):1–8
38. Heller MJ. 2002. DNA microarray technology: devices, systems, and applications. *Annu. Rev. Biomed. Eng.* 4:129–53
39. Hernandez ET, Swaminathan J, Marcotte EM, Anslyn EV. 2017. Solution-phase and solid-phase sequential, selective modification of side chains in KDYWEC and KDYWE as models for usage in single-molecule protein sequencing. *New J. Chem.* 41(2):462–69
40. Houghtaling J, Ying C, Eggenberger OM, Fennouri A, Nandivada S, et al. 2019. Estimation of shape, volume, and dipole moment of individual proteins freely transiting a synthetic nanopore. *ACS Nano* 13(5):5231–42
41. Howard CJ, Floyd BM, Bardo AM, Swaminathan J, Marcotte EM, Anslyn EV. 2020. Solid-phase peptide capture and release for bulk and single-molecule proteomics. *ACS Chem. Biol.* 15(6):1401–7
42. Hu Q, Li H, Wang L, Gu H, Fan C. 2019. DNA nanotechnology-enabled drug delivery systems. *Chem. Rev.* 119(10):6459–506
43. Hu Z-L, Huo M-Z, Ying Y-L, Long Y-T. 2021. Biological nanopore approach for single-molecule protein sequencing. *Angew. Chem. Int. Ed.* 60(27):14738–49
44. Huang B, Wu H, Bhaya D, Grossman A, Granier S, et al. 2007. Counting low-copy number proteins in a single cell. *Science* 315(5808):81–84
45. Huang G, Voet A, Maglia G. 2019. FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution. *Nat. Commun.* 10:835
46. Huang G, Willems K, Bartelds M, van Dorpe P, Soskine M, Maglia G. 2020. Electro-osmotic vortices promote the capture of folded proteins by PlyAB nanopores. *Nano Lett.* 20(5):3819–27
47. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17(1):239
48. Kaboord B, Perr M. 2008. Isolation of proteins and protein complexes by immunoprecipitation. *Methods Mol. Biol.* 424:349–64
49. Kafader JO, Melani RD, Senko MW, Makarov AA, Kelleher NL, Compton PD. 2019. Measurement of individual ions sharply increases the resolution of Orbitrap mass spectra of proteins. *Anal. Chem.* 91(4):2776–83
50. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *PNAS* 93(24):13770–73
51. Keifer DZ, Jarrold MF. 2017. Single-molecule mass spectrometry. *Mass Spectrom. Rev.* 36(6):715–33
52. Kennedy E, Dong Z, Tennant C, Timp G. 2016. Reading the primary structure of a protein with 0.07 nm<sup>3</sup> resolution using a subnanometre-diameter pore. *Nat. Nanotechnol.* 11(11):968–76
53. Kothapalli R, Yoder SJ, Mane S, Loughran TP. 2002. Microarray results: How accurate are they? *BMC Bioinform.* 3:22
54. Kuchino Y, Nishimura S. 1989. Enzymatic RNA sequencing. *Methods Enzymol.* 180:154–63
55. Laursen RA. 1971. Solid-phase Edman degradation. *Eur. J. Biochem.* 20(1):89–102
56. Li S, Wu X-Y, Li M-Y, Liu S-C, Ying Y-L, Long Y-T. 2020. T232K/K238Q aerolysin nanopore for mapping adjacent phosphorylation sites of a single tau peptide. *Small Methods* 4(11):2000014
57. Li SFY, Wu YS. 2000. Capillary electrophoresis. In *Encyclopedia of Separation Science*, ed ID Wilson, pp. 1176–87. Oxford, UK: Academic

58. Luo L, Boerwinkle E, Xiong M. 2011. Association studies for next-generation sequencing. *Genome Res.* 21(7):1099–108
59. MacDonald JI, Munch HK, Moore T, Francis MB. 2015. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* 11(5):326–31
60. Mahdessian D, Cesnik AJ, Gnann C, Danielsson F, Stenström L, et al. 2021. Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* 590(7847):649–54
61. Maillard RA, Chistol G, Sen M, Righini M, Tan J, et al. 2011. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* 145(3):459–69
62. Mann M. 2016. The rise of mass spectrometry and the fall of Edman degradation. *Clin. Chem.* 62(1):293–94
63. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, et al. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* 30(4):349–53
64. Marx V. 2019. A dream of single-cell proteomics. *Nat. Methods* 16(9):809–12
65. Matheron L, van den Toorn H, Heck AJR, Mohammed S. 2014. Characterization of biases in phosphopeptide enrichment by  $\text{Ti}^{4+}$ -immobilized metal affinity chromatography and  $\text{TiO}_2$  using a massive synthetic library and human cell digests. *Anal. Chem.* 86(16):8312–20
66. Maulbetsch W, Wiener B, Poole W, Bush J, Stein D. 2016. Preserving the sequence of a biopolymer's monomers as they enter an electrospray mass spectrometer. *Phys. Rev. Appl.* 6(5):054006
67. Mereuta L, Roy M, Asandei A, Lee JK, Park Y, et al. 2014. Slowing down single-molecule trafficking through a protein nanopore reveals intermediates for peptide translocation. *Sci. Rep.* 4:3885
68. Nivala J, Marks DB, Akeson M. 2013. Unfoldase-mediated protein translocation through an  $\alpha$ -hemolysin nanopore. *Nat. Biotechnol.* 31(3):247–50
69. Nivala J, Mulroney L, Li G, Schreiber J, Akeson M. 2014. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* 8(12):12365–75
70. Ohayon S, Girsault A, Nasser M, Shen-Orr S, Meller A. 2019. Simulation of single-protein nanopore sensing shows feasibility for whole-proteome identification. *PLOS Comput. Biol.* 15(5):e1007067
71. Ohshiro T, Tsutsui M, Yokota K, Furuhashi M, Taniguchi M, Kawai T. 2014. Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat. Nanotechnol.* 9(10):835–40
72. Ouldali H, Sarthak K, Ensslen T, Piguet F, Manivet P, et al. 2020. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* 38(2):176–81
73. Palmblad M. 2021. Theoretical considerations for next-generation proteomics. *J. Proteome Res.* 20(6):3395–99
74. Petelski AA, Emmott E, Leduc A, Huffman RG, Specht H, et al. 2021. Multiplexed single-cell proteomics using SCoPE2. bioRxiv 2021.03.12.435034. <https://doi.org/10.1101/2021.03.12.435034>
75. Piguet F, Ouldali H, Pastoriza-Gallego M, Manivet P, Pelta J, Oukhaled A. 2018. Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* 9:966
76. Ren R, Zhang Y, Nadappuram BP, Akpinar B, Klenerman D, et al. 2017. Nanopore extended field-effect transistor for selective single-molecule biosensing. *Nat. Commun.* 8:586
77. Restrepo-Pérez L, Huang G, Bohländer PR, Worp N, Eelkema R, et al. 2019. Resolving chemical modifications to a single amino acid within a peptide using a biological nanopore. *ACS Nano* 13(12):13668–76
78. Restrepo-Pérez L, Joo C, Dekker C. 2018. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* 13(9):786–96
79. Restrepo-Pérez L, Wong CH, Maglia G, Dekker C, Joo C. 2019. Label-free detection of post-translational modifications with a nanopore. *Nano Lett.* 19(11):7957–64
80. Riley NM, Coon JJ. 2016. Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal. Chem.* 88(1):74–94
81. Rodriguez-Larrea D. 2021. Single-amino acid discrimination in proteins with homogeneous nanopore sensors and neural networks. *Biosens. Bioelectron.* 180:113108
82. Rodriguez-Larrea D, Bayley H. 2013. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* 8(4):288–95

83. Rodriques SG, Marblestone AH, Boyden ES. 2019. A theoretical analysis of single molecule protein sequencing via weak binding spectra. *PLOS ONE* 14(3):e0212868
84. Rose RJ, Damoc E, Denisov E, Makarov A, Heck AJR. 2012. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* 9(11):1084–86
85. Rozenblatt-Rosen O, Shin JW, Rood JE, Hupalowska A, Regev A, Heyn H. 2021. Building a high-quality Human Cell Atlas. *Nat. Biotechnol.* 39(2):149–53
86. Sampath G. 2015. Amino acid discrimination in a nanopore and the feasibility of sequencing peptides with a tandem cell and exopeptidase. *RSC Adv.* 5:30694–700
87. Sanger F, Thompson EOP. 1953. The amino-acid sequence in the glycyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochem. J.* 53(3):366–74
88. Sanger F, Tuppy H. 1951. The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochem. J.* 49(4):463–81
89. Schaus TE, Woo S, Xuan F, Chen X, Yin P. 2017. A DNA nanoscope via auto-cycling proximity recording. *Nat. Commun.* 8:696
90. Schmid S, Stömmers P, Dietz H, Dekker C. 2021. Nanopore electro-osmotic trap for the label-free study of single proteins and their conformations. *bioRxiv* 2021.03.09.434634. <https://doi.org/10.1101/2021.03.09.434634>
91. Schnitzbauer J, Strauss MT, Schlichthaerle T, Schueder F, Jungmann R. 2017. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* 12(6):1198–228
92. Smith RD, Cheng X, Brace JE, Hofstadler SA, Anderson GA. 1994. Trapping, detection and reaction of very large single molecular ions by mass spectrometry. *Nature* 369(6476):137–39
93. Steen H, Mann M. 2004. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 5(9):699–711
94. Storm AJ, Storm C, Chen J, Zandbergen H, Joanny J-F, Dekker C. 2005. Fast DNA translocation through a solid-state nanopore. *Nano Lett.* 5(7):1193–97
95. Swaminathan J, Boulgakov AA, Hernandez ET, Bardo AM, Bachman JL, et al. 2018. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* 36(11):1076–82
96. Swaminathan J, Boulgakov AA, Marcotte EM. 2015. A theoretical justification for single molecule peptide sequencing. *PLOS Comput. Biol.* 11(2):e1004080
97. Takagi T, Suzuki M, Baba T, Minegishi K, Sasaki S. 1984. Complete amino acid sequence of amelogenin in developing bovine enamel. *Biochem. Biophys. Res. Commun.* 121(2):592–97
98. Timp W, Timp G. 2020. Beyond mass spectrometry, the next step in proteomics. *Sci. Adv.* 6(2):eaax8978
99. Towbin H, Staehelin T, Gordon J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *PNAS* 76(9):4350–54
100. Tullman J, Callahan N, Ellington B, Kelman Z, Marino JP. 2019. Engineering ClpS for selective and enhanced N-terminal amino acid binding. *Appl. Microbiol. Biotechnol.* 103(6):2621–33
101. van Ginkel J, Filius M, Szczepaniak M, Tulinski P, Meyer AS, Joo C. 2018. Single-molecule peptide fingerprinting. *PNAS* 115(13):3338–43
102. Wagner DE, Klein AM. 2020. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* 21(7):410–27
103. Wang L, Li PCH. 2011. Microfluidic DNA microarray analysis: a review. *Anal. Chim. Acta* 687(1):12–27
104. Wang R, Gilboa T, Song J, Huttner D, Grinstaff MW, Meller A. 2018. Single-molecule discrimination of labeled DNAs and polypeptides using photoluminescent-free TiO<sub>2</sub> nanopores. *ACS Nano* 12(11):11648–56
105. Wei X, Ma D, Jing L, Wang LY, Wang X, et al. 2020. Enabling nanopore technology for sensing individual amino acids by a derivatization strategy. *J. Mater. Chem. B* 8(31):6792–97
106. Wingren C. 2016. Antibody-based proteomics. In *Proteogenomics*, ed. Á Végvári, pp. 163–79. Berlin: Springer
107. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nat. Methods* 6(5):359–62
108. Yalow RS, Berson SA. 1960. Immunoassay of endogenous plasma insulin in man. *J. Clin. Investig.* 39:1157–75

109. Yao Y, Docter M, van Ginkel J, de Ridder D, Joo C. 2015. Single-molecule protein sequencing through fingerprinting: computational assessment. *Phys. Biol.* 12(5):055003
110. Yates JR, Ruse CI, Nakorchevsky A. 2009. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* 11:49–79
111. Yusko EC, Bruhn BR, Eggenberger OM, Houghtaling J, Rollings RC, et al. 2017. Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* 12(4):360–67
112. Zhang L, Floyd BM, Chilamari M, Mapes J, Swaminathan J, et al. 2021. Photoredox-catalyzed decarboxylative C-terminal differentiation for bulk and single molecule proteomics. *bioRxiv* 2021.07.08.451692. <https://doi.org/10.1101/2021.07.08.451692>
113. Zhang S, Huang G, Versloot R, Herwig BM, de Souza PCT, et al. 2020. Bottom-up fabrication of a multi-component nanopore sensor that unfolds, processes and recognizes single proteins. *bioRxiv* 2020.12.04.411884. <https://doi.org/10.1101/2020.12.04.411884>
114. Zhao Y, Ashcroft B, Zhang P, Liu H, Sen S, et al. 2014. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotechnol.* 9(6):466–73
115. Zubarev RA. 2013. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 13(5):723–26
116. Zwolak M, Di Ventra M. 2005. Electronic signature of DNA nucleotides via transverse transport. *Nano Lett.* 5(3):421–24