# Nanopore Protein Sequencing: Principles, Strategies, and Challenges toward Single-Amino-Acid Resolution


## Abstract

Nanopore-based single-molecule protein sequencing promises direct readout of amino acid sequences without amplification. This review examines three core nanopore strategies: (1) direct sequencing using aerolysin and Cu(II)-functionalized MspA nanopores that electrically discriminate all twenty proteinogenic amino acids with up to 99.1% classification accuracy; (2) label-assisted sensing that converts protein identity into DNA barcode or chemical tag signals for multiplexed detection; and (3) native protein sensing via electroosmotic flow through engineered CytK pores, enabling translocation of full-length proteins under opposing electrophoretic force. Deep learning-assisted post-translational modification detection at single-residue resolution is also discussed. Key challenges in translocation control, signal resolution, and bioinformatic infrastructure are evaluated.



***Keywords:*** *nanopore sequencing; single-molecule proteomics; amino acid discrimination; post-translational modification; electroosmotic flow*


## Introduction

The human proteome encompasses millions of proteoforms arising from alternative splicing and post-translational modifications (PTMs), far exceeding the approximately 20,000 protein-coding genes [1]. Many clinically significant biomarker proteins exist at femtomolar concentrations, well below the detection limits of conventional platforms such as two-dimensional gel electrophoresis and liquid chromatography-tandem mass spectrometry (LC-MS/MS) [1, 2]. A fundamental constraint distinguishes protein analysis from genomics: proteins cannot be amplified. Unlike DNA, which can be exponentially copied via polymerase chain reaction, proteins must be detected directly, making sensitivity improvements entirely dependent on the detection technology itself.

Bottom-up mass spectrometry (MS), the current workhorse of proteomics, identifies proteins indirectly by enzymatic digestion into peptides followed by LC-MS/MS analysis and database searching [2, 3]. Despite its success, this indirect approach suffers from inherent limitations: PTM information is partially lost during digestion, low-abundance peptide signals are masked by high-abundance species, and

complete proteoform characterization remains extremely difficult [3]. These constraints have motivated the pursuit of single-molecule approaches capable of directly reading protein sequences without amplification.

Beyond MS, other emerging single-molecule protein sequencing concepts include fluorosequencing-which identifies amino acids by sequential fluorescent labeling and imaging-and single-molecule Edman degradation. While these approaches show promise, nanopore-based strategies have advanced furthest toward practical implementation and are the focus of this review. Three principal nanopore strategies have emerged: direct sequencing, label-assisted sensing, and native protein direct sensing. This review examines these strategies, their achievements in amino acid discrimination and PTM detection, and the remaining challenges toward practical single-molecule protein sequencing.

## 1. Direct Sequencing: Amino Acid Discrimination via Biological Nanopores

The direct sequencing strategy envisions a linearized polypeptide threading through a nanopore such that successive amino acid residues pass through the sensing zone, generating residue-specific ionic current blockades. Two landmark studies have demonstrated the feasibility of this approach for all twenty proteinogenic amino acids.

Ouldali et al. (2020) used a wild-type aerolysin nanopore-a heptameric beta-barrel pore from *Aeromonas* bacteria that contains a built-in single-molecule trap-to achieve electrical recognition of amino acids [4]. Each amino acid was covalently linked to a polycationic carrier peptide of seven arginines (R7), which electrostatically drove the conjugate into the aerolysin trap. The resulting current blockade ($I_b/I_0$) correlated with the hydrodynamic volume rather than molecular mass of each amino acid. Thirteen of twenty amino acids were directly distinguishable; chemical modification of overlapping species (e.g., nitro-tyrosine, sulfoxide-methionine) extended identification to fifteen. Statistical modeling predicted that sixteen of twenty amino acids could be identified with greater than 90% probability at a 200 ms residence time. Notably, the isomeric pair leucine and isoleucine-indistinguishable by MS-was discriminated with 99% accuracy at 125 ms residence time [4].

Zhang et al. (2024) advanced this paradigm with a Cu(II)-functionalized MspA (*Mycobacterium smegmatis* porin A) nanopore [5]. An N91H substitution at the pore

constriction, together with the native N90 residue, created a histidine-brace motif that reversibly coordinated one Cu(II) ion and one amino acid molecule via its alpha-carboxyl and alpha-amine groups. Using a Random Forest classifier trained on 1,003 signal features-including blockade depth, dwell time, and normalized current density profiles-all twenty amino acids were discriminated with a validation AUC of 0.989 and 99.1% accuracy at high-confidence filtering. The detection limit reached below 100 nM for glycine, approximately 500-fold improvement over prior Ni(II)-MspA systems. Critically, the authors demonstrated real-time detection of pathologically relevant peptides, including a melanoma neoantigen (COL18A1) differing from its wild-type counterpart by a single amino acid substitution, and Alzheimer's-associated Abeta(17-27) variants [5].

Engineering of biological nanopores has been instrumental in these advances. Sun et al. systematically reviewed functionalization strategies-including chemical modification, site-directed mutagenesis, and metal-ion coordination-applied to aerolysin, FraC, MspA, and other pore types, demonstrating that each approach offers distinct mechanisms for enhancing the chemical sensitivity of the sensing zone [6].

## 2. Label-Assisted Nanopore Sensing Strategies

Label-assisted strategies circumvent the difficulty of directly reading amino acid sequences by attaching molecular tags-fluorescent labels, chemical moieties, or DNA barcodes-to proteins, converting protein identity into more readily decodable signals. Motone et al. systematically reviewed tag-based approaches for controlling protein translocation through nanopores, including molecular tags, DNA barcodes, and chemical modifications, concluding that such labeling strategies are essential for achieving the temporal resolution required for single-molecule protein analysis [7]. Cardozo et al. demonstrated multiplexed nanopore detection: each protein target was labeled with a unique DNA barcode that generated a characteristic current signal upon translocation, enabling simultaneous detection of up to nine proteins [7].

The rationale behind label-assisted approaches is that the chemical complexity of twenty amino acid side chains makes direct electrical discrimination fundamentally difficult, whereas DNA barcodes can be read with the maturity of existing nanopore DNA sequencing technology. This conversion strategy trades single-molecule amino acid resolution for multiplexing capability and robustness. While label-assisted methods cannot yet read de novo protein sequences, they provide a practical bridge

toward targeted protein detection in complex biological samples, particularly when combined with antibody-based capture or affinity enrichment. Future developments may integrate DNA barcode readout with enzymatic protein digestion to approach sequence-level information through labeled fragment identification.

## 3. Native Protein Direct Sensing via Electroosmotic Flow

The native protein sensing strategy aims to drive intact, unlabeled proteins through a nanopore, exploiting intrinsic physicochemical properties to generate current fingerprints. The central obstacle is that proteins carry heterogeneous charges-acidic residues are negative, basic residues positive, hydrophobic residues nearly neutral-making electrophoretic translocation highly unpredictable.

Sauciuc et al. (2024) achieved a critical breakthrough by engineering the CytK (Cytolysin K) nanopore from *Staphylococcus aureus* [9]. The wild-type pore has no net lumen charge and produces negligible electroosmotic flow (EOF). By introducing rings of negatively charged aspartate residues along the 5 nm beta-barrel lumen, the engineered 2E-4D-CytK mutant achieved a K+/Cl- ion selectivity ratio of 4.04, generating a strong cis-to-trans EOF at negative applied bias. Molecular dynamics simulations quantified the EOF force at approximately 20 pN at -100 mV and 50 pN at -250 mV-sufficient to overcome electrophoretic repulsion of negatively charged polypeptide segments.

This EOF-driven mechanism enabled unidirectional translocation of full-length proteins in the direction opposing electrophoretic force. Three natural proteins-malE219a (412 residues), H152A glucose-binding protein (341 residues), and DHFR variant (170 residues)-were successfully driven through the pore after chemical denaturation with urea [9]. Each protein produced distinctive, multi-level current signatures: malE219a showed a current exclusion of 70.5 +/- 0.6%, while H152A-GBP exhibited a two-step blockade pattern (70.6% and 89.1%) suggestive of stepwise unfolding. These results demonstrated that even without single-amino-acid resolution, EOF-driven translocation can generate protein-specific electrical fingerprints. Anion-selective control pores with engineered positive charges failed to drive translocation, revealing a mechanistic asymmetry that constrains the design space [9].

## 4. Post-Translational Modification Detection at Single-Residue Resolution

PTM detection is a critical application cutting across all three nanopore strategies. Nova et al. (2024) achieved single-residue phosphorylation detection using a hybrid approach that bridges label-assisted and direct sensing: peptides were covalently conjugated to DNA handles, and a Hel308 helicase DNA motor enzyme processively ratcheted the DNA-peptide conjugate through an MspA nanopore, controlling translocation speed to single-nucleotide steps [10]. This enzyme-controlled stepping enabled precise localization of phosphorylation sites along the peptide backbone at single-residue resolution-a capability complementary to MS-based phosphoproteomics, which often requires enrichment steps and suffers from incomplete site localization.

Cao et al. (2024) further advanced PTM analysis by combining an engineered K238A aerolysin nanopore with an LSTM-based deep learning pipeline for label-free, single-molecule detection of multiple PTM types on alpha-synuclein C-terminal peptides (alpha-syn 124-140) [8]. The system discriminated eight variant classes including phosphorylation (pY125, pS129), nitration (nY125, nY133, nY136), oxidation (oM127), and combinatorial modifications. Different PTM types produced measurably distinct translocation dynamics: phosphorylation at Y125 shortened dwell times to 0.55 +/- 0.08 ms (versus 2.58 +/- 0.4 ms for wild-type), while phosphorylation at the neighboring S129 lengthened them to 3.62 +/- 0.2 ms, and nitration at Y125 further extended them to 4.51 +/- 0.5 ms. Binary classification of wild-type versus nY125 achieved 94% accuracy. Under asymmetric salt conditions, the detection limit dropped to approximately 100 pM. These site-specific kinetic signatures, extractable by neural networks but invisible to threshold-based analysis, establish nanopore sensing as a uniquely capable platform for combinatorial PTM pattern detection-an unmet need that antibody-based assays cannot address, as neighboring modifications interfere with epitope recognition [8].

## 5. Challenges and Future Perspectives

Several fundamental challenges remain. Protein translocation control is foremost: although EOF-driven approaches have solved the directional transport problem [9], achieving stable stepwise translocation with single-amino-acid resolution across diverse protein species has not yet been demonstrated. The natural proteins tested required chemical denaturation and destabilizing mutations, limiting applicability to native proteomes.

Signal resolution presents a second challenge. Even with Cu(II)-functionalized MspA, several amino acid pairs (Lys/Arg, Met/Leu, Pro/Phe) exhibit overlapping blockade signals, requiring sophisticated Random Forest classifiers with 1,003 features for discrimination [5]. High-confidence filtering in the Zhang et al. study retained only 30.9% of signals at 99.1% accuracy, representing significant data loss [5]. PTM-induced current changes are even more subtle, demanding higher signal-to-noise ratios and more sensitive algorithms. Furthermore, the field lacks standardized reference databases and bioinformatic tools for converting raw nanopore current traces to amino acid sequences.

A comparative assessment of the three strategies reveals complementary strengths. Direct sequencing via aerolysin and Cu(II)-MspA achieves the finest amino acid discrimination but operates on individual residues or short peptides rather than intact proteins. Label-assisted approaches offer multiplexed protein detection without amino acid resolution, serving as a pragmatic bridge for targeted assays. EOF-driven native protein sensing handles full-length proteins but currently lacks single-residue resolution and requires chemical denaturation. The ideal platform would combine the amino acid specificity of direct sequencing, the robustness of label-assisted detection, and the full-length protein compatibility of EOF-driven translocation-a convergence that remains a central goal of the field.

Looking forward, enzyme-controlled translocation-combining EOF-driven transport with processive enzymes such as ClpX or carboxypeptidase Y-may achieve true single-residue reading [5, 10]. The integration of tunneling junctions with solid-state nanopores could combine molecular trapping with electronic single-molecule sensitivity. Deep learning architectures, including the LSTM networks demonstrated by Cao et al. [8], will play an increasingly central role in extracting information from complex current traces. Large-scale parallel nanopore arrays offer a pathway to proteome-wide throughput.

## Conclusion

Nanopore protein sequencing has progressed from conceptual aspiration to demonstrated capability for single-amino-acid discrimination. Three complementary strategies have achieved significant milestones: aerolysin and Cu(II)-MspA pores electrically recognize all twenty proteinogenic amino acids with up to 99.1% accuracy; EOF-driven CytK pores translocate full-length proteins against electrophoretic force;

and deep learning-enhanced aerolysin sensing detects combinatorial PTM patterns at picomolar concentrations. While challenges in translocation control, signal resolution, and bioinformatic infrastructure remain substantial, the accelerating pace of progress suggests that nanopore protein sequencing may eventually realize a transformative role in proteomics analogous to that of nanopore DNA sequencing in genomics.

## References


[1] Guo, T., Steen, J. A., & Mann, M. (2025). Mass-spectrometry-based proteomics: from single cells to clinical applications. *Nature*, 638(8052), 901-911.

[2] Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620), 347-355.

[3] Switzar, L., Giera, M., & Niessen, W. M. A. (2013). Protein Digestion: An Overview of the Available Techniques and Recent Developments. *Journal of Proteome Research*, 12(3), 1067-1077.

[4] Ouldali, H., et al. (2020). Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nature Biotechnology*, 38(2), 176-181.

[5] Zhang, M., et al. (2024). Real-time detection of 20 amino acids and discrimination of pathologically relevant peptides with functionalized nanopore. *Nature Methods*, 21, 609-618.

[6] Sun, K., et al. (2026). Functionalized protein nanopores: Strategies and applications. *Nano Research*, 19(2), 94908343.

[7] Motone, K., Cardozo, N., & Nivala, J. (2021). Herding cats: Label-based approaches in protein translocation through nanopore sensors. *iScience*, 24(9), 103032.

[8] Cao, C., et al. (2024). Deep Learning-Assisted Single-Molecule Detection of Protein Post-translational Modifications with a Biological Nanopore. *ACS Nano*, 18(2), 1504-1515.

[9] Sauciuc, A., et al. (2024). Translocation of linearized full-length proteins through an engineered nanopore under opposing electrophoretic force. *Nature Biotechnology*, 42, 1275-1281.

[10] Nova, I. C., et al. (2024). Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nature Biotechnology*, 42(5), 710-714.