

Unraveling proteins at the single molecule level using nanopores

Restrepo Perez, Laura

DOI

[10.4233/uuid:a431659a-da38-42a5-be17-05b8c241e355](https://doi.org/10.4233/uuid:a431659a-da38-42a5-be17-05b8c241e355)

Publication date

2019

Document Version

Final published version

Citation (APA)

Restrepo Perez, L. (2019). *Unraveling proteins at the single molecule level using nanopores*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:a431659a-da38-42a5-be17-05b8c241e355>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

**UNRAVELING PROTEINS AT THE SINGLE
MOLECULE LEVEL USING NANOPORES**

UNRAVELING PROTEINS AT THE SINGLE MOLECULE LEVEL USING NANOPORES

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates to be defended publicly on
Friday 10 May 2019 at 10:00 o'clock

by

Laura RESTREPO PEREZ

Master of Science in Nanoscience and Nanotechnology
Technische Universitat Dresden, Germany
Katholieke Universiteit Leuven, Belgium
Born in Caldas, Colombia.

This dissertation has been approved by the:
Promoters: Prof. dr. C. Dekker and dr. C. Joo

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. C. Dekker	Delft University of Technology, promotor
Dr. C. Joo	Delft University of Technology, promotor

Independent members:

Prof. dr. H. Balci	Kent State University
Prof. dr. S.W. Lee	Ewha Womans University
Dr. J. Alfaro	University of Edinburgh/University of Gdansk
Prof. dr. M. Dogterom	Delft University of Technology
Prof. dr. N. H. Dekker	Delft University of Technology, reserve

Other member:

Prof. dr. G. Maglia	University of Groningen
---------------------	-------------------------

 **NanoFront**



Keywords: Single-molecule protein sequencing, protein fingerprinting, proteins, nanopores, post-translational modifications.

Printed by: Gildeprint

Front & Back: Laura Restrepo

Copyright © 2019 by L. Restrepo Perez

Casimir PhD Series 2019-12

ISBN 978.90.8593.395.3

An electronic version of this dissertation is available at <http://repository.tudelft.nl/>

*A mi familia:
Mamá, papá y Juan*

Contents

General introduction	1
1.1 Proteins – the workhorses of the cell	2
1.2 Protein structure	3
1.3 Nanotech meets biotech	4
1.4 Nanopores	5
1.5 Nanopore meets proteins	6
1.6 Outline of this thesis	8
References	9
Paving the way to single-molecule protein sequencing	13
2.1 Introduction	14
2.2 Protein fingerprinting using fluorescence	18
2.3 Protein sequencing using tunnelling currents	21
2.4 Protein sequencing using nanopores	23
2.5 Outlook	28
References	31
Peptide analysis with the FraC nanopore:	39
3.1 Introduction	40
3.2 Results	42
3.3 Conclusions	51
3.4 Methods	52
3.5 Supplementary information	57
References	60
Resolving modifications in single amino acids using the FraC nanopore	65
4.1 Introduction	66
4.2 Results	67
4.3 Discussion and conclusion	77
4.4 Materials and Methods	77
4.5 Supplementary information	79
References	83

Label-free detection of post-translational modifications with a nanopore	85
5.1 Introduction	86
5.2 Results	87
5.3 Discussion and conclusions	94
5.4 Supplementary information	96
References	98
SDS-assisted protein transport through solid state nanopores	101
6.1 Introduction	102
6.2 Molecular dynamics simulations	104
6.3 Single-molecule experiments	108
6.4 Conclusions	113
6.5 Materials and methods	113
6.6 Supplementary information	117
References	125
Single-molecule protein analysis: From the lab to the market	129
7.1 Unmet market needs and opportunities	130
7.2 Current techniques for protein analysis	131
7.3 A solution based on single-molecule techniques	132
7.4 Potential applications	133
7.5 Market analysis	134
7.6 Conclusion	137
References	139
Summary	141
Samenvatting	145
Acknowledgements	151
Curriculum Vitae	159
List of publications	161

1

General introduction

1.1 Proteins – the workhorses of the cell

The human body is composed of approximately 37 trillion cells¹. Each cell can be seen as a micrometer-sized factory in which thousands of assembly lines are running in parallel in a tightly coordinated fashion. To fulfill its function, the cell relies on molecules such as DNA and proteins. The DNA is the code containing the complete set of instructions necessary for the cell to function. Proteins are the molecular machines (or operators) that follow the code to operate and control each of the assembly lines of the cell. Proteins are therefore, the main workhorses of the cell, performing nearly all its functions twenty-four hours a day and seven days a week for all our lifespan.

The diversity of functions that proteins perform in the cell is astonishing² (Figure 1.1). Proteins form membrane channels that control the passage of ions and other molecules across the different cellular compartments or to the cell exterior. One of the most complex membrane protein machineries is for example the nuclear pore complex (NPC), formed by approximately 30 different proteins, with a main function to tightly regulate the passage of molecules across the nuclear envelope³. Proteins act as antibodies playing a key role in

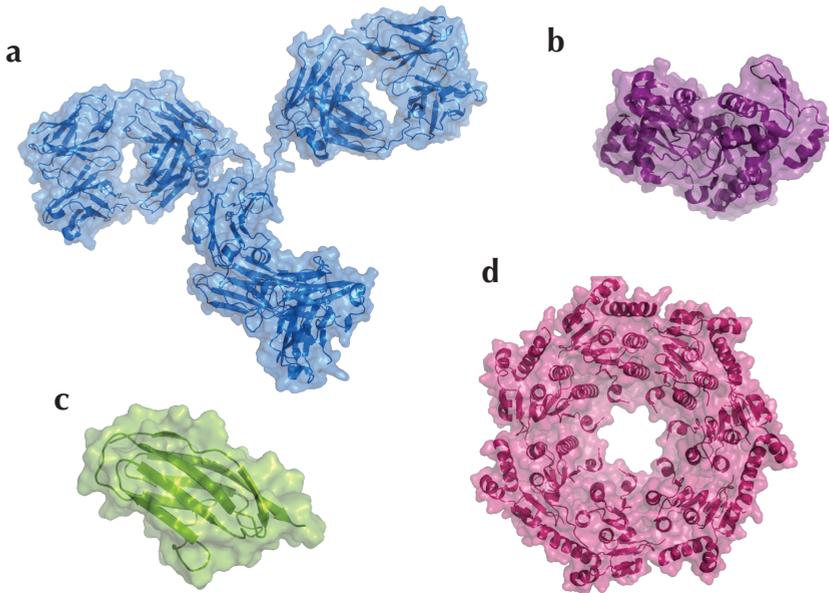


Figure 1.1: Crystal structure of several proteins showing diverse protein structures. (a) The Immunoglobulin G (IgG) antibody. Representation adapted from the PDB structure 1IGT. (b) Single actin monomer. Monomers form actin filaments, an essential component of the cell's cytoskeleton. Representation adapted from the PDB structure 1J6Z. (c) Titin I27 domain. A domain of titin, an important protein for muscle elasticity. Representation adapted from the PDB structure 1TIT. (d) ClpR. An enzyme that plays an important role in the degradation of misfolded proteins. Representation adapted from the PDB structure 4HNK.

our adaptive immune system. Antibodies bind to pathogens and other antigens, blocking their pathogenic function or targeting them for destruction by other cells of the immune system. Proteins also play crucial roles as enzymes, catalyzing multiple biochemical reactions in the cell⁴. In the context of DNA processing, different proteins are crucial for DNA replication, repair, and transcription. Finally, proteins are important structural elements in the cell. Protein filaments such as microtubules and actin filaments form the cytoskeleton, the scaffolds that provide rigidity to the cell and maintain its shape⁵.

1.2 Protein structure

Amino acids are the building blocks of proteins. Amino acids are small organic molecules consisting of an amine group, a carboxylic acid, and a variable side chain (R) that changes for each of the different amino acids (Figure 1.2a). Twenty different amino acids are proteogenic and are encoded in the genetic code. Amino acids are linked together via peptide bonds to form a polypeptide chain. A protein is composed of one such polypeptide chain.

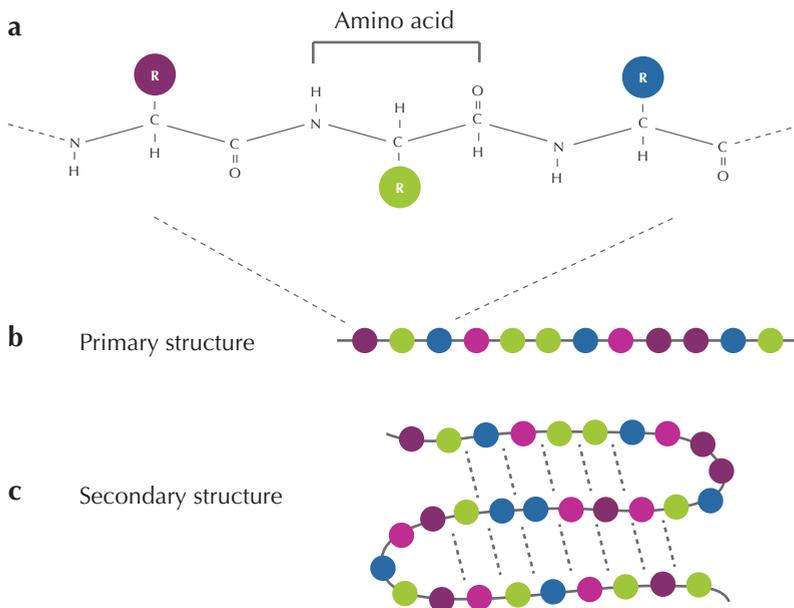


Figure 1.2: Amino acid and protein structure. (a) Schematic illustrating amino acid structure in the polypeptide chain forming the proteins primary sequence. (b) Representation of protein primary sequence. (c) Representation of protein secondary sequence. Figure inspired from reference 7.

The linear sequence of amino acids in a protein is referred to as its *primary structure* (Figure 1.2 b). The primary structure ultimately dictates the final three dimensional arrangement of the protein. A second level of structure emerges from local interactions in the polypeptide

1 backbone chain. Hydrogen bonds between amino and carboxyl groups of neighboring regions create local substructures known as alpha helices and beta sheets. These represent the *secondary structure* of the protein (Figure 1.2c). A single protein often contains multiple alpha helices and beta sheets that fold and arrange together to create the three dimensional structure of the protein, also known as *tertiary structure*. The structure is usually held together by interactions between amino acid residues, for example, hydrophobic interactions, disulfide bridges, and electrostatic interactions between charged residues. Protein complexes made out of multiple polypeptide chains also comprise a *quaternary structure*. The quaternary structure is the three dimensional structure that derives from the aggregation of two or more individual polypeptide chains^{6,7}.

1.3 Nanotech meets biotech

Nanotechnology is our ability to manipulate and control atoms and molecules at the nanometer scale. Already in the 16th century, medieval artisans used gold and silver nanoparticles to create vibrant red and yellow colors in stained glass⁸. Back then, however, none of them imagined they were manipulating materials at the nanoscale. It is only in the last few decades when single nanometer-sized objects have been fabricated and characterized in a controlled way. The vision for building and manipulating objects at the nanoscale can be for example pointed to the inspiring talk “there is plenty of room at the bottom” by Richard Feynman in 1959⁹.

Nanotechnology has since then developed into a multidisciplinary field, where materials and physics show characteristics never seen before. One of the main driving forces for the recent advances in nanotechnology is the need for ever smaller and cheaper electronic components. Moore's law, which states that the number of transistors in an integrated circuit should double every year, has brought a tremendous challenge to physicists and material scientists. As a consequence, there has been a constant search for more precise fabrication and characterization tools.

The continuous downscaling of electromechanical components to nanometer sizes has allowed them to reach the dimensions of biological entities such as cells, proteins, and DNA. The median length of a human protein is 375 amino acids¹⁰. A globular protein of that size has an approximate diameter of 3-5 nm. The diameter of a double helix of DNA is 2nm. Before the advent of nanotechnology, the idea of studying individual molecules at the nanometer scale remained elusive. With the precise and sensitive tools brought about by nanotechnology, biology can now be explored at the single-molecule level at the nanometer scale.

1.4 Nanopores

Many different nanotechnology tools have been explored for single-molecule studies of proteins and DNA^{11,12}. Among them, nanopores emerged as a powerful tool with unique characteristics such as real-time molecule sensing, label-free analysis, and high temporal resolution. In a nanopore sensor, an insulating membrane separates two compartments filled with an electrolyte. A nanometer-sized pore is made within the membrane connecting the two compartments. Solid-state pores are traditionally made using a transmission electron microscope (TEM) to drill a nanometer-sized hole in a synthetic membrane made of materials such as SiN or graphene (Figure 1.3a). Biological nanopores are pore-forming transmembrane proteins that can be reconstituted in free-standing lipid membranes (Figure 1.3b).

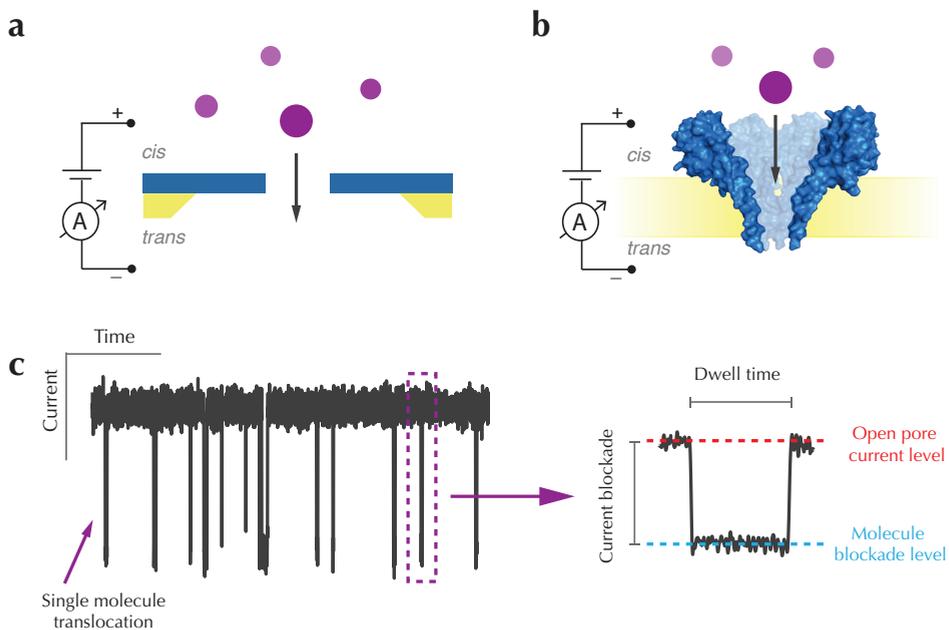


Figure 1.3: Schematic of the nanopore sensing principle (a) Schematic representation of molecules passing through a solid-state nanopore. An electrolyte solution is present both in the cis and trans compartments, which are connected only by the nanopore. (b) Schematic representation of a biological nanopore set up in which the FraC nanopore is inserted in a free-standing lipid membrane (yellow). The wild type FraC structure was created using the 4TSY structure from the PDB. (c) Typical current traces observed for the translocation of particles through a FraC nanopore. On the left panel, each current deep corresponds to a translocation event of a 30 aa peptide. A close up of one event is shown in the right. Event characteristics such as dwell time and current blockade are used to characterize the analyte.

When a voltage is applied across the membrane, an ionic current flows through the nanometer-sized aperture. Molecules passing or translocating through the pore modulate the ionic current, which provides the basic sensor signal. Figure 1.3 shows the basic sensing

mechanism.

The first steps towards single-biomolecule nanopore sensing can be traced back to the 1990s, with the early studies of alpha-hemolysin, a pore forming protein from *Staphylococcus aureus*^{13–16}. The field progressed fast, fueled by large investments aimed at achieving fast and cheap DNA sequencing with nanopores. All the early experiments were performed using biological nanopores. Because of their protein nature, biological pores have the advantage of being highly reproducible in their geometry and atomic configuration. Moreover, mutagenesis can be used to specifically engineer amino acids and binding sites in the pore. Their main limitation comes from the stability of the lipid membrane. Solid-state pores were later introduced as a more stable and robust alternative¹⁷. These advantages, however, come with a price, as it has proven hard to fabricate small solid-state pores with the same atomic precision and reproducibility of biological pores. Moreover, unspecific interactions between proteins and the synthetic membrane aren't uncommon.

1.5 Nanopore meets proteins

Nanopores have not been limited to DNA-sequencing applications. After the successful experiments with DNA, nanopore scientists started to explore other molecules, such as PEG polymers¹⁸, ssDNA, RNA^{19,20}, and ultimately, peptides and proteins^{21–23}. The latter is the focus of this thesis. Compared to DNA transport through nanopores, proteins bring additional challenges. For example, while DNA is transported through the nanopore via electrophoresis because of its high and uniform negative charge, proteins are often neutral or weakly charged, so not only electrophoresis, but also electroosmosis plays a key role in protein translocation²⁴.

Nanopores have been used for several aspects of protein analysis. Previous studies have reported the use of nanopores for the analysis of protein structure (protein size, shape, and folding state)^{25–30}, protein-protein interactions^{31–33}, and even some post-translational modifications^{34–37}. The ultimate goal for single-molecule protein analysis with a nanopore would be to identify and characterize all proteins present in a biological sample. First attempts have been shown in which a limited number of proteins (~10 proteins) and biomarkers are identified mainly by their current blockade and dwell time characteristics^{38,39}. In these studies, folded proteins of different size and shape were analyzed with a nanopore. This approach, however, rapidly runs into limitations as a handful of characteristics is not enough to differentiate the hundreds of thousands of different proteins present in real biological samples. An alternative way, in which comprehensive protein analysis can be done at the level of proteomics requires the realization of protein sequencing, i.e., reading off

the actual linear sequence of amino acids in the primary structure. In a protein sequencing approach with nanopores, the protein has to be unfolded and the amino acid sequence needs to be sequentially read as a polypeptide chain translocates through the pore - an approach that is similar to that currently applied in DNA sequencing.

There are many technical challenges for the realization of nanopore protein sequencing. One of them is the differentiation of 20 different amino acids during translocation. It was recently proposed, that a fingerprinting approach in which only a subset of amino acids is labeled and detected would allow for the identification of most proteins in the proteome^{40–43}. Figure 1.4 shows the working principle of protein fingerprinting using nanopores.

In this thesis we present important steps for the realization of protein fingerprinting using nanopores and address some of the challenges including post-translational modification (PTM) analysis and protein unfolding.

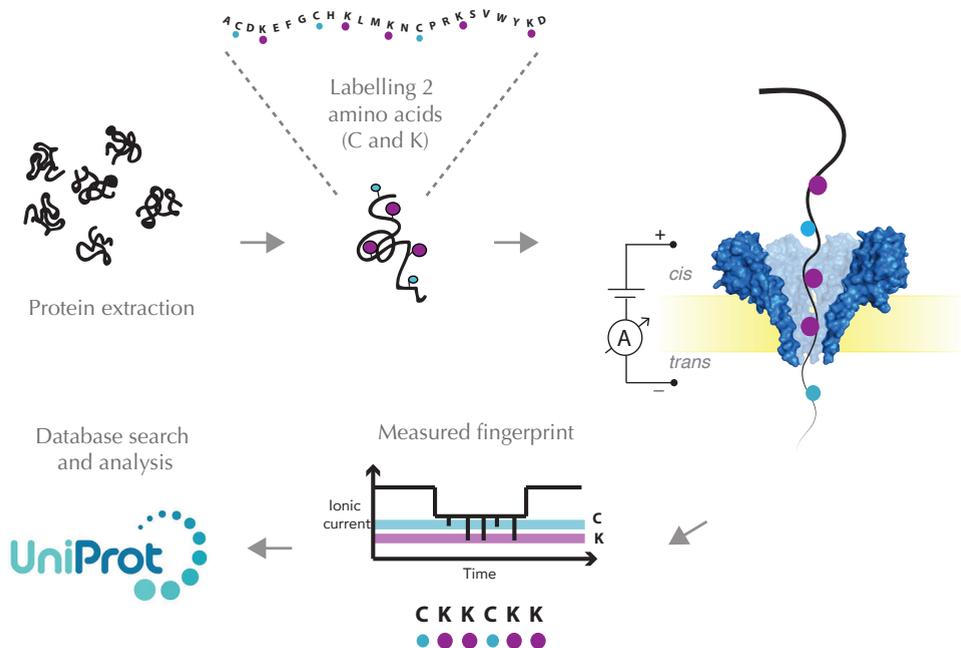


Figure 1.4: Representation of a protein fingerprinting method with nanopores. First, proteins are extracted and a subset of amino acids is labeled. Cysteine and lysine labeling has been proposed previously. Each protein is passed through a nanopore and the sequence of labelled amino acids read. The measured fingerprint is then compared to a protein database for protein identification.

1.6 Outline of this thesis

This thesis describes a number of developments and advances for the realization of protein fingerprinting and PTM (post-translational modification) detection using both biological and solid-state nanopores.

In Chapter 2 we present a review of the emerging field of single-molecule protein sequencing. The proposed methods are based on single-molecule techniques such as nanopores, fluorescence, and tunnelling currents across nanogaps. We describe the schemes proposed so far and discuss their advantages and drawbacks.

In Chapter 3 we explore the use of a bipolar peptide as a model peptide system. The opposite charges at the ends of the peptide generate forces to opposing directions, stalling the peptide while it translocates the FraC pore. In this chapter we propose to use this bipolar peptide as an electromechanical gate that can modulate nanopore conductance in a dynamic manner at the single-molecule level. The stalling mechanism observed with our model peptide allows for extended examination of the central region of the peptide, which is further explored in Chapter 4 and 5.

In Chapter 4, we investigate the effect of adding different chemical modifications to single amino acid side chains. This is important for a fingerprinting approach. We show that sensitive and reproducible detection of labels with a mass of 0.4-1.3 kDa can be obtained. Information about the position of the label along the peptide chain can be extracted from the individual event characteristics.

In Chapter 5 we explore the detection of two post-translational modifications: phosphorylation and glycosylation. We show that using our model peptide and a FraC nanopore, these modifications can be detected and distinguished in a label-free manner.

In Chapter 6 we move to the study of proteins. Here we combine molecular dynamics (MD) simulations with single-molecule experiments to investigate the utility of SDS (Sodium Dodecyl Sulfate) to unfold proteins for solid-state nanopore translocation. Our simulations and experiments prove that SDS-treated proteins show, as desired, a considerable loss of the protein structure during the nanopore translocation.

Finally, in Chapter 7, we analyze the most important aspects of bringing a technology for single-molecule protein sequencing from the lab to the market. We describe the unmet market needs and opportunities, compare this technology to current methods, and explore the potential applications and markets in which such a technology could enter.

References

1. Bianconi, E. et al. An estimation of the number of cells in the human body. *Ann. Hum. Biol.* 40, 463–471 (2013).
2. Alberts et al. *Molecular biology of the cell.* (Garland Science, 2002). doi:10.1091/mbc.E14-10-1437
3. Kabachinski, G. & Schwartz, T. U. The nuclear pore complex--structure and function at a glance. *J. Cell Sci.* 128, 423–9 (2015).
4. Schroeder, H. W., Cavacini, L. & Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* 125, S41-52 (2010).
5. Hoyt, M. A., Hyman, A. A. & Bähler, M. Motor proteins of the eukaryotic cytoskeleton. *Proc. Natl. Acad. Sci. U. S. A.* 94, 12747–8 (1997).
6. Alberts, B. et al. *The Shape and Structure of Proteins.* (2002).
7. O'Connor, C. M. & Adams, J. U. *Essentials of Cell Biology.* (Cambridge, MA: NPG Education, 2010., 2010).
8. Daw, R. Nanotechnology is ancient history. *The Guardian* (2012). Available at: <https://www.theguardian.com/nanotechnology-world/nanotechnology-is-ancient-history>. (Accessed: 22nd February 2019)
9. Feynman, R. P. There's plenty of room at the bottm. *Eng. Sceince* 23, 22–36 (1960).
10. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* 33, 3390–400 (2005).
11. *Single-molecule Studies of Proteins.* (Springer New York, 2013). doi:10.1007/978-1-4614-4921-8
12. Robison, A. D. & Finkelstein, I. J. High-throughput single-molecule studies of protein-DNA interactions. *FEBS Lett.* 588, 3539–3546 (2014).
13. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524 (2016).
14. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13770–3 (1996).
15. Song, L. et al. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274, 1859–66 (1996).
16. Baldarelli, R., Branton, D., Church, G., Deamer, D. W. & Kasianowicz, J. Characterization of Individual Polymer Molecules Based on Monomer-Interface Interactions. *US Pat.* (1998).
17. Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* 2, 209–215 (2007).
18. Sergey M. Bezrukov, *,†,‡, Igor Vodyanoy, †,§, Rafik A. Brutyan, †, and John J. Kasianowicz*. Dynamics and Free Energy of Polymers Partitioning into a Nanoscale Pore. (1996). doi:10.1021/MA960841J

19. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13770 (1996).
20. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys. J.* 77, 3227–33 (1999).
21. Movileanu, L., Schmittschmitt, J. P., Martin Scholtz, J. & Bayley, H. Interactions of peptides with a protein pore. *Biophys. J.* 89, (2005).
22. Sutherland, T. C. Structure of peptides investigated by nanopore analysis. *Nano Lett.* 4, (2004).
23. Stefureac, R., Long, Y.-T., Kraatz, H.-B., Howard, P. & Lee, J. S. Transport of alpha-Helical Peptides through alpha-Hemolysin and Aerolysin Pores †. *Biochemistry* 45, 9172–9179 (2006).
24. Firnkes, M., Pedone, D., Knezevic, J., Döblinger, M. & Rant, U. Electrically facilitated translocations of proteins through silicon nitride nanopores: Conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* 10, 2162–2167 (2010).
25. Yusko, E. C. et al. Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* 12, 360–367 (2017).
26. Restrepo-Pérez, L., John, S., Aksimentiev, A., Joo, C. & Dekker, C. SDS-assisted protein transport through solid-state nanopores. *Nanoscale* 9, 11685–11693 (2017).
27. Talaga, D. S. & Li, J. Single-Molecule Protein Unfolding in Solid State Nanopores. 9287–9297 (2009). doi:10.1021/ja901088b
28. Oukhaled, A. et al. Dynamics of Completely Unfolded and Native Proteins through Solid-State Nanopores as a Function of Electric Driving Force. *ACS Nano* 5, 3628–3638 (2011).
29. Li, J., Fologea, D., Rollings, R. & Ledden, B. Characterization of Protein Unfolding with Solid-state Nanopores. *Protein Pept. Lett.* 21, 256–265 (2014).
30. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* 8, 288–95 (2013).
31. Thakur, A. K. & Movileanu, L. Real-time measurement of protein–protein interactions at single-molecule resolution using a biological nanopore. *Nat. Biotechnol.* 37, 96–101 (2018).
32. Wei, R., Gatterdam, V., Wieneke, R., Tampé, R. & Rant, U. Stochastic sensing of proteins with receptor-modified solid-state nanopores. *Nat. Nanotechnol.* 7, 257–263 (2012).
33. Ying, Y.-L., Yu, R.-J., Hu, Y.-X., Gao, R. & Long, Y.-T. Single antibody–antigen interactions monitored via transient ionic current recording using nanopore sensors. *Chem.*

- Commun. 53, 8620–8623 (2017).
34. Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* 32, (2014).
 35. Wloka, C. et al. Label-Free and Real-Time Detection of Protein Ubiquitination with a Biological Nanopore. *ACS Nano* 11, 4387–4394 (2017).
 36. Fahie, M. A. & Chen, M. Electrostatic Interactions between OmpG Nanopore and Analyte Protein Surface Can Distinguish between Glycosylated Isoforms. *J. Phys. Chem. B* 119, 10198–10206 (2015).
 37. Harrington, L., Alexander, L. T., Knapp, S. & Bayley, H. Single-Molecule Protein Phosphorylation and Dephosphorylation by Nanopore Enzymology. *ACS Nano* 13, 633–641 (2019).
 38. Huang, G., Willems, K., Soskine, M., Wloka, C. & Maglia, G. Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores. *Nat. Commun.* 8, 935 (2017).
 39. Yusko, E. C. et al. Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* 12, 360–367 (2016).
 40. Yao, Y., Docter, M., Ginkel, J., Ridder, D. & Joo, C. Single-molecule protein sequencing through fingerprinting: computational assessment. *Phys. Biol.* 12, (2015).
 41. Swaminathan, J., Boulgakov, A. A. & Marcotte, E. M. A Theoretical justification for single molecule peptide sequencing. *PLOS Comput. Biol.* 11, (2015).
 42. Restrepo-Pérez, L., Joo, C. & Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* 13, 786–796 (2018).
 43. Stevens, B. et al. FRET-based identification of mRNAs undergoing translation. *PLoS One* 7, e38344 (2012).

2

Paving the way to single-molecule protein sequencing

Proteins are major building blocks of life. The protein content of a cell and an organism provides key information for the understanding of biological processes and disease. Despite the importance of protein analysis, only a handful of techniques are available to determine protein sequences, and these methods face limitations, e.g. requiring a sizable amount of sample. Single-molecule techniques would revolutionize proteomics research providing ultimate sensitivity for the detection of low-abundance proteins and the realization of single-cell proteomics. In recent years, novel single-molecule protein sequencing schemes have been proposed, using fluorescence, tunnelling currents, and nanopores. Here we present a review of these approaches, together with the first experimental efforts towards their realization. We discuss their advantages and drawbacks, and present our perspective in the development of single-molecule protein sequencing techniques.

This chapter has been published as: Laura Restrepo-Pérez, Chirlmin Joo and Cees Dekker, Paving the way to single-molecule protein sequencing, *Nature Nanotechnology*, 13, pages 786–796 (2018).

2.1 Introduction

2 Proteins are the workhorses in all living cells. Thousands of different proteins sustain all functions of the cell, from copying DNA and catalysing basic metabolism to producing cellular motion. Protein analysis can therefore provide key information for the understanding of biological processes and disease (Box 2.1). Compared to the impressive technical advances in DNA sequencing, the development of highly sensitive, high-throughput protein sequencing techniques lags severely behind. The only methods currently available for protein sequencing are Edman degradation, mass spectrometry, or their combination^{1,2,3} (see Box 2.2).

Box 2.1: Genomic, transcriptomic and proteomic analysis in diagnostics

When the human genome project was realized in 2003, sequencing an entire human genome would cost approximately US\$50 million and would require 100 machines working for ~2,500 h. Today, thanks to the tremendous advances in DNA sequencing technologies, a human genome can be sequenced for only US\$1,000 using one machine working for ~72 h (refs 100,101). DNA sequencing is thus becoming a routine technique in clinics, allowing the collection of genetic information from patients at reasonable time and cost.

The challenge ahead is the interpretation of the data gathered from DNA sequencing with respect to the health condition of patients. A large gap resides between genotype and phenotype. Transcriptomics studies are often used as a first bridge, which provides information about which genes are actively being expressed. However, the gap still persists as mRNAs levels do not simply correlate to protein levels due to factors such as the variability in translational efficiency of different mRNAs, and the difference between mRNA and protein lifetimes¹⁰². Moreover, protein post-translational modifications further influence the function and structure of proteins.

Proteome analysis is therefore key to understand biological processes and their dynamic nature^{103,104}. After all, proteins dictate most biological functions and are directly related to the phenotype of a cell. So, while genomics offers a quick glimpse, much like looking at the menu in a restaurant, proteomics brings you inside the heart of the kitchen, to closely examine what the food looks like.

The current gold standard for protein sequencing is mass spectrometry^{4,5,6,7}. The technique, however, has fundamental drawbacks in terms of its limit of detection and dynamic range⁸. Human samples are extremely complex, comprising a wide range of protein concentrations. In human plasma, for example, the concentration of proteins can vary from few picograms per millilitre (interleukin 6) to few milligrams per millilitre (albumin)^{9,10}. Therefore, an exceedingly high dynamic range ($\sim 10^9$) is necessary for comprehensive proteome analysis^{9,11}. State-of-the-art mass spectrometers are limited to a dynamic range of $\sim 10^4$ to 10^5 (refs^{9,11}). Another drawback of the instrument is its detection limit, which hinders biomarker discovery and translates into the need for large amounts of sample. If we consider a protein that is present in a cell in a low copy-number (less than 1,000 molecules per cell)¹², millions of cells are required to reach the limit of detection of the instrument (0.1 to 10 femtomole)^{13,14,15}. Mass spectrometry is thus far away from comprehensive single-cell analysis.

The spectacular advances in DNA sequencing technology, where even single DNA molecules can be sequenced, have inspired dreams of novel technologies for protein sequencing. However, the search for such protein sequencing methods is not trivial due to the complex nature of proteins. Proteins are built from 20 distinctive amino acids, while DNA is comprised of only four different bases. Independent of the read-out method of choice, the detection of 20 distinguishable signals is a tremendous challenge. Moreover, DNA samples with low concentrations of analyte can be amplified using polymerases, whereas protein sequencing platforms cannot benefit from such amplification since there is no polymerase chain reaction-like amplification method for proteins. Protein sequencing techniques that would read the exact sequence of individual proteins at the single-molecule level could bring a revolution to proteomics, providing the ultimate sensitivity for the detection of low-abundance proteins. Moreover, such a method would enable single-cell proteome studies with higher capabilities than current methods^{16,17,18,19,20}.

In this Review, we present an overview of the exciting nascent field of single-molecule protein sequencing. Several approaches for protein sequencing at the single-molecule level have emerged in the past few years. These new ideas run from renovating Edman degradation and mass spectrometry, through repurposing single-molecule DNA sequencing platforms for protein sequencing, to developing entirely new molecular devices. The proposed methods are based on single-molecule techniques such as nanopores, fluorescence and tunnelling currents across nanogaps (Fig. 2.1). We describe the schemes proposed so far and discuss their advantages and drawbacks. First experimental efforts and proof-of-principle experiments towards their realization are also discussed.

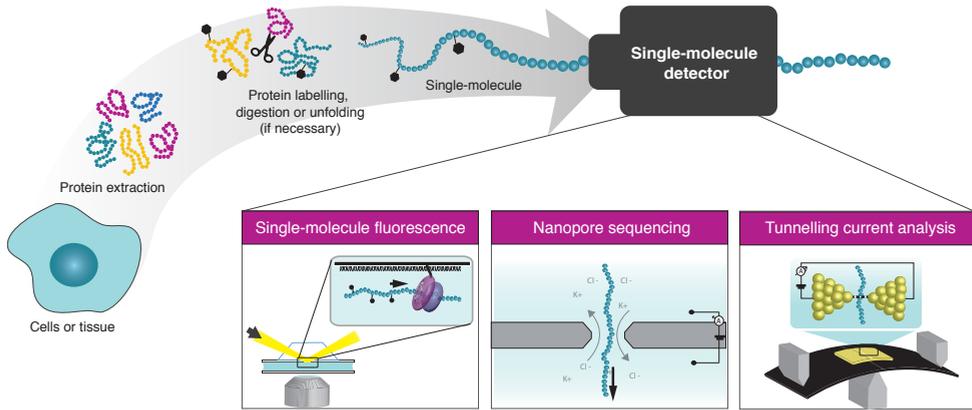
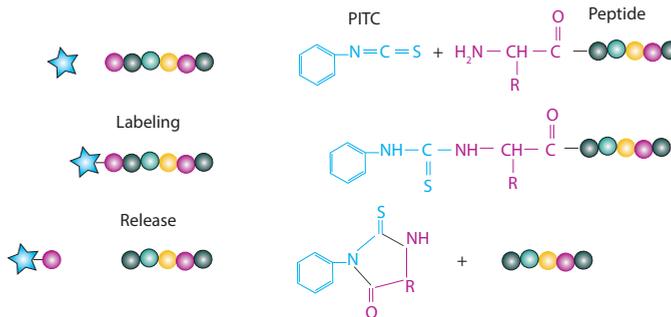


Figure 2.1: Schematic of the single-molecule protein sequencing workflow with fluorescence, nanopores or tunnelling currents. In a typical experiment, proteins are extracted from a biological sample or even a single cell, then labelled, unfolded and partly digested (if necessary), and finally, each molecule is sequenced with a single-molecule technique.

Box 2.2: Current protein sequencing methods

Edman degradation:

Invented by Pehr Edman in 1950, Edman degradation allows the ordered identification of the amino acid sequence in a protein from the N- to the C-terminus¹⁰⁵. It consists of cyclic chemical reactions that label, cleave and identify the amino acid at the terminus of a protein, one at the time (see figure below). In the first step of the reaction, the Edman reagent (phenylisothiocyanate, PITC) reacts with the amino group at the N-terminus of the protein under mild basic buffer conditions. The modified N-terminal amino acid is removed as a thiazolinone derivative under acidic conditions. This derivative is then identified using chromatography.



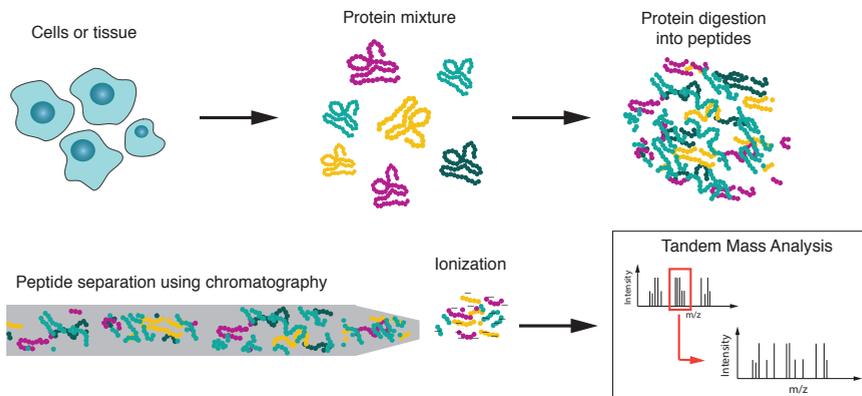
Schematic of Edman degradation reaction showing the process of labelling and cleavage of the amino acid in the N-terminus of the peptide

Box 2.2: Current protein sequencing methods

Edman degradation is a useful tool for sequencing, but it is limited to the analysis of purified peptides that are shorter than ~50 amino acids. It cannot be used for the analysis of complex protein mixtures, such as those present in most biological samples. Additionally, each degradation cycle can take approximately 45 min (ref. 106), making the process extremely time-consuming. N-terminus modifications can also interfere with the process. For example, if the N-terminus of the peptide is acetylated (a common post-translational modification), the reaction cannot take place, prohibiting protein sequencing.

Mass spectrometry:

Since the 1980s, with the discovery of new ionization techniques (matrix-assisted laser desorption/ionization and electron-spray ionization), mass spectrometry has evolved into an important analytical tool for the life sciences⁵. For deep protein analysis, the introduction of shotgun proteomics marked an important step for the study of samples containing protein mixtures¹⁰⁷. In a typical experiment, proteins are digested into peptides and separated according to hydrophobicity and charge using chromatography (see figure below). As peptides elute from the column, they are ionized and analysed according to their mass-to-charge ratio using tandem mass spectrometry.



Workflow of proteome analysis with mass spectrometry. Proteins are extracted from cells or tissues and digested into peptides. The peptide mixture is separated using chromatography. Peptides are ionized and analysed using tandem mass spectrometry.

2.2 Protein fingerprinting using fluorescence

2

Fluorescence techniques have been central for the development of high-throughput DNA sequencing devices. In systems such as those of Illumina²¹, Pacific Biosciences²² and Helicos²³, DNA is de novo sequenced by monitoring the incorporation of fluorescently labelled nucleotides during strand replication. The development of a de novo protein sequencing method based on fluorescence faces enormous challenges. Major constraints are the lack of organic fluorophores for the detection of 20 different amino acids without substantial signal crosstalk, and the absence of a suitable chemistry to specifically label all 20 amino acids²⁴.

Recently, simplified schemes, in which only a small subset of amino acids is fluorescently labelled and detected, have been proposed. If demonstrated, these could lead to the development of protein identification methods with single-molecule sensitivity^{25,26}. These approaches resemble optical mapping of DNA, where partial sequence information is sufficient to identify certain characteristics of a genome or to identify different pathogens²⁷. Similar to how optical mapping has served as a complementary lower-resolution technique to DNA sequencing, protein fingerprinting could constitute a complementary technique to de novo protein sequencing.

In 2015, Joo and colleagues proposed a fingerprinting scheme based on the detection of two types of amino acid²⁵. In their approach, the cysteine (C) and lysine (K) residues of a protein are labelled and sequentially detected. This sequence of cysteines and lysines (or CK sequence) can then be used to identify the protein of interest using a protein database (Fig. 2.2b). To read the CK sequence, an unfoldase called ClpX is immobilized on a single-molecule surface and used as a protein scanner. This molecular motor recognizes tagged polypeptides and unfolds them while translocating them through its internal cavity. If the enzyme is labelled with a donor fluorophore and the substrate contains acceptor dyes in its cysteines and lysines, fluorescence resonance energy transfer (FRET) occurs as each of these amino acids approaches the ClpX constriction, generating a CK read in a string of two different acceptor signals (Fig. 2.2a).

The feasibility of this CK fingerprinting approach was computationally assessed using a human protein database containing ~20,000 protein entries²⁵. CK sequences were generated computationally taking into consideration the most common errors expected during experimental readings. These generated CK sequences were compared to the database, and the probability of retrieving an original sequence was calculated based on the accuracy of the matches. Considering a 10% error level in the readings, approximately half of the protein sequences could be correctly retrieved. When additional parameters, such as the distance between cysteines and lysines, were considered (Fig. 2.2b, CK-dist read), the method

could accurately identify a major percentage (>70–80%) of proteins even when high error rates (20–30%) were considered (Fig. 2.2c).

A proof of concept was experimentally demonstrated by Joo et al. this year²⁸. Using a donor-labelled ClpP (the proteolytic chamber that binds ClpX), the authors sequentially read out FRET signals from acceptor-labelled substrates. They could fingerprint 29-, 40-, 51-amino acid long peptides, and a monomeric (119 amino acids) and a dimeric (210 amino acids) titin protein. The repurposed ClpXP showed a constant translocation speed and unidirectionality, features that are suitable for reliable fingerprinting. Note that a similar fingerprinting system was proposed and experimentally demonstrated using a labelled ribosome to monitor the production of specific proteins inside the cell as a way to gain information on protein expression location and levels^{29,30}.

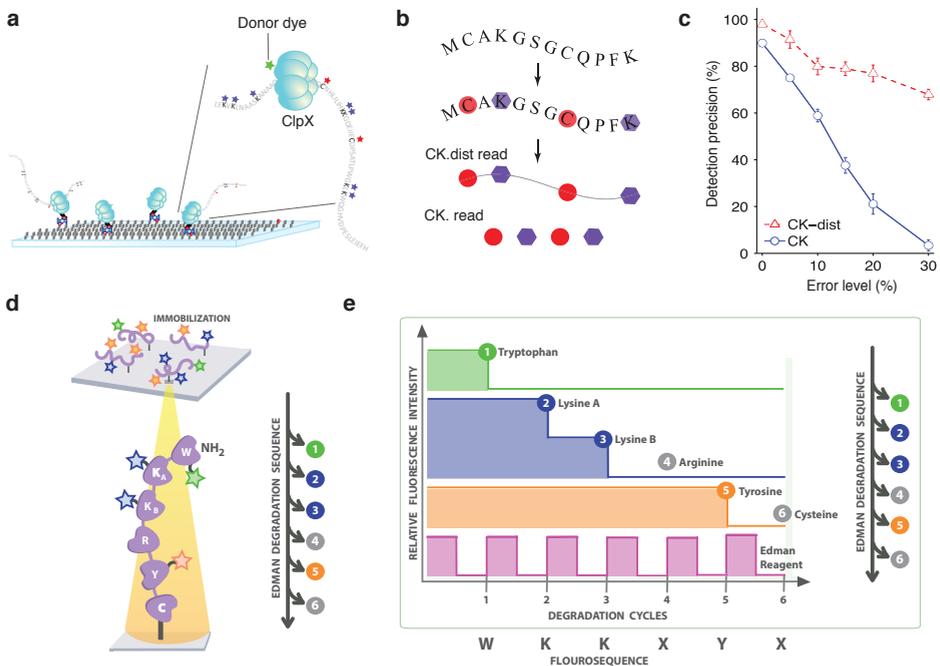


Figure 2.2: Protein fingerprinting schemes using fluorescence. a, Scheme proposed in ref. 25, in which a labelled unfoldase is immobilized on a surface and used to scan protein substrates. b, Cysteines and lysines of the protein substrate are labelled and FRET is detected upon the translocation of these residues. The CK sequence is then compared to a protein database. The CK read corresponds to the sequence of cysteine and lysine residues. The CK-dist read incorporates the distance between these amino acids. c, Graph of the detection precision (number of true positives divided by the number of read-outs returned by the algorithm) versus error level (number of errors divided by the fingerprint length). d, Scheme proposed in ref. 26. In this approach, labelled peptides are immobilized and subjected to sequential cycles of Edman degradation. e, The scheme shows the expected fluorescence intensity signal for the proposed peptide. The loss in fluorescence after each cycle is used to determine the sequence. Panels a, b, and c adapted from ref. 33 under a Creative Commons license (<https://creativecommons.org/licenses/by/3.0/>); panels d and e adapted from ref. 26.

2 A different method is pursued by Marcotte and colleagues, in which peptide fingerprinting is accomplished using a single-molecule version of Edman degradation²⁶. Unlike conventional Edman degradation methods, the single-molecule detection allows for analysis of mixed populations. In this approach, proteins are digested into peptide fragments (~10–30 amino acids long) and specific amino acids are labelled with fluorophores of distinguishable colours. The labelled peptides are immobilized on a surface, and fluorescence microscopy is used to monitor each cycle of Edman degradation at single-molecule resolution (Fig. 2.2d). Each degradation cycle removes the N-terminal amino acid of the peptide, so that the sequence of labelled amino acids can be detected by monitoring the change of the fluorescence intensity in each cycle. The decrease in fluorescence after a degradation cycle indicates that a labelled amino acid has been cleaved. The cleaved amino acid can be identified using spectral information (Fig. 2.2e).

Computer simulations were used to investigate the probability of detecting proteins from the identification of a unique peptide sequence using Marcotte's fingerprinting method²⁶. Different immobilization, labelling and cleavage strategies were evaluated, and it was determined that at least four different labelled amino acids are required to identify 98% of the human proteome²⁴.

The fingerprinting schemes proposed here take advantage of the fact that proteins can be identified using incomplete sequence information. The approach proposed by Joo and colleagues reads full-length proteins and therefore requires simple two-colour labelling of substrates. The main limitation of this approach is the requirement of a recognition tag in the N- or C-terminus of the substrate for unfoldase recognition. It seems possible to devise ligation schemes to add such a tag to all proteins in a mixture or to engineer the enzyme to allow recognition of any protein coming from cellular preparations and other biological samples. Marcotte's approach to fingerprinting benefits from an entirely chemical approach, which can be beneficial for commercialization purposes. However, the harsh conditions required for the Edman reaction demands for a careful selection of fluorophores, and a set of adaptations to a conventional total internal reflection fluorescence microscope³¹. A disadvantage of this method is that each cycle of Edman degradation can take approximately 45 min, making the sequencing process extremely slow. An alternative approach to Edman degradation is currently being explored in which an enzyme has been designed that is capable of cleaving off amino acids, one at the time, from the protein N-terminal³². The use of this enzyme, called Edmanase, may allow Edman degradation to proceed under physiological conditions, and potentially at a faster pace.

Fluorescence fingerprinting may play a crucial role in the development of fast techniques for parallel protein identification and analysis. Millions to thousands of millions of single

molecules can be immobilized and monitored together, opening the door to high-throughput assays. Single-molecule protein identification using fluorescence could complement *de novo* protein sequencing methods, improving the sensitivity of current bulk identification techniques such as antibody microarrays or mass-spectrometry protein identification based on peptide fingerprints. The improved sensitivity of these methods brings important advantages for applications such as biomarker detection for disease diagnosis.

2.3 Protein sequencing using tunnelling currents

The idea of using tunnelling currents to measure single molecules was first conceived in the 1970s³³. Tunnelling currents are measured between two metal electrodes separated by a gap that ranges from a few angstroms to a few nanometres (Fig. 2.3a,d). When individual molecules pass through the nanoscopic gap, a change in the tunnelling current is measured. This current modulation can be used to determine which molecule is transiently residing in the gap in real time. With the invention of the scanning tunnelling microscope (STM) in the 1980s, the possibility to realize this idea became clear and led to the development of a new field named molecular electronics^{34,35,36}. In recent years, this technique has evolved to study a variety of biomolecules aiming towards DNA and RNA sequencing^{37,38,39,40,41} (for a detailed review of these developments see refs 39,40). In a similar way, interest has emerged in the study of amino acids and peptides in an urge towards protein sequencing. In this section, we present a review of these developments.

In 2014, Lindsay et al. reported the first measurements of amino acids and short peptides using tunnelling currents^{42,43}. They demonstrated the sensitivity of their approach by analysing three sets of amino acids with minor structural differences: glycine versus its methylated form sarcosine, the enantiomers of asparagine (L- versus D-asparagine), and the isobaric amino acids leucine versus isoleucine. Their experimental set-up consisted of two palladium electrodes, separated by a gap of 2 nm. The electrodes were functionalized with a recognition molecule (4(5)-(2-mercaptoethyl)-1H-imidazole-2-carboxamide), which was covalently bound to the electrodes. The recognition molecule interacted temporarily with the analyte to orient the molecule and thus provided a better-defined current path (Fig. 3a). When amino acids were introduced, the transient interactions between each amino acid and the recognition molecule were detected as a train of current spikes (Fig. 2.3b). Using two-dimensional maps of the current amplitude and the spike shape, the amino acids analysed in each set could be discriminated with an accuracy of 80% or higher (Fig. 2.3c).

A subsequent study was reported by Kawai and colleagues in which all 20 amino acids and phosphotyrosine were tested using tunnelling currents with a different experimental

set-up⁴⁴. In their study, smaller gaps of 0.70 nm and 0.55 nm were created using gold break junctions. The small size of the gap allowed the detection of amino acids without a recognition molecule (Fig. 2.3d). The 0.70-nm gaps produced detectable signals for eight (Y, F, W, H, P, E, D, I) out of the 20 different amino acids, while smaller gaps of 0.55 nm produced signals for nine (P, H, E, D, I, K, C, L, M) amino acids. In total, 12 out of the 20 amino acids could be recognized; the rest did not produce a detectable signal. When one of the detectable amino acids was introduced in the measuring set-up, peaks in the current trace were observed, indicating the transient presence of an individual molecule between the electrodes (Fig. 2.3e). The amplitude and duration of each peak was used to characterize each amino acid as shown in the scatter diagram in Fig. 2.3f. Seven amino acids showed distinctive signals and show potential for their differentiation in complex mixtures; the remaining five produced indistinguishable signals. The detection of post-translational modifications was also demonstrated using 0.70-nm gaps. Tyrosine and phosphotyrosine produced distinctive signals, and mixtures of them yielded two populations in the amplitude histograms. Last, using the same approach, short peptides containing tyrosine and phosphotyrosine could be distinguished.

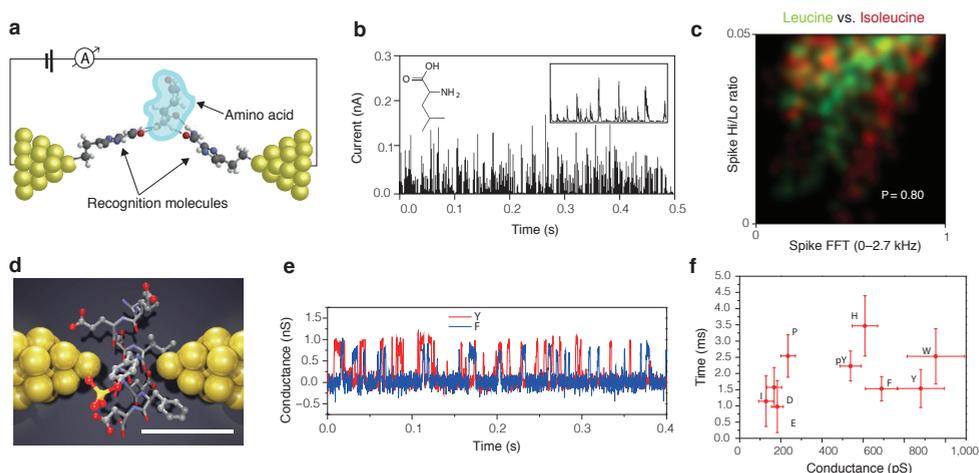


Figure 2.3: Amino acid and peptide characterization with tunnelling currents. **a**, Recognition-tunnelling scheme where STM-coupled palladium electrodes are functionalized with recognition molecules. **b**, Typical current versus time trace obtained for the measurement of an amino acid (here Leucine). **c**, Two-dimensional plot of probability density using two different fast Fourier transform (FFT) features for leucine (green) and isoleucine (red). P indicates the accuracy for calling single-molecule events. **d**, Schematic of the operating principle: a molecule is sandwiched between two gold nanogap electrodes created using mechanically controlled break junctions. Scale bar, 1 nm. **e**, Conductance versus time traces obtained for measurements of the amino acids Y and F. **f**, Scatter plot of time versus conductance for different amino acids measured in a 0.55-nm gap. Panel a adapted from and panels b and c reproduced from ref. 43, Springer Nature Ltd; panel e adapted from and panels d and f reproduced from ref. 44, Springer Nature Ltd.

The recognition tunnelling approach used by Lindsay and colleagues shows the remarkable sensitivity of quantum tunnelling currents. This technique can discriminate isomers and molecules with minor structural differences that are indistinguishable by other techniques such as mass spectrometry. The downside of this method is the non-trivial complexity of the data. Each molecule can orient in many different ways within the junction, and exhibits significant translational and rotational fluctuations, leading to considerably different current signals. Therefore, machine-learning algorithms may be necessary to distinguish each molecule, considering the multiple conformations that can be observed.

The study of Kawai and co-workers presented a systematic characterization of different amino acids and short peptides. Out of the 20 amino acids studied, 7 amino acids generated distinguishable signals. This represents a promising step towards amino acid discrimination for protein sequencing. Arrays containing junctions of different sizes might increase the number of amino acids that are detectable and increase the possibility to distinguish amino acids in a mixture. Technical improvements in the experimental set-ups and fabrication processes would facilitate this task. For example, it has recently been shown that extra coatings on the nanoelectrodes could bring improvements in terms of the signal-to-noise ratio and bandwidth of the measurements^{45,46}.

To make this proof-of-concept into a sequencing tool, measurements of tunnelling currents should be coupled with a mechanism that threads a polypeptide through the gap in a controlled way. An exopeptidase or other molecular motor could be adapted to translocate the polypeptide through such an electrode gap. Alternatively, electrophoresis, electro-osmosis or a pressure difference could be used as a driving mechanism for molecules if the tunnelling device is coupled to a nanopore. Several groups have reported first experimental efforts in this direction^{47,48,49,50}.

2.4 Protein sequencing using nanopores

In 2014, Oxford Nanopore Technologies announced the release of the first single-molecule DNA sequencing device based on nanopores^{51,52,53,54}. These pocket-sized devices are revolutionizing DNA sequencing by allowing extremely long reads and in situ detection at remote laboratories (even in outer space)^{51,55}. In a nanopore experiment, an insulating membrane containing a nanometre-sized pore is placed between two electrolyte-filled compartments. When a voltage is applied across the membrane, an ionic current flows through the nanopore. As individual molecules translocate through the pore, a modulation in ionic currents is observed, which provides structural information about the molecule of interest^{56,57,58}. Using this principle, biopolymers can be sequenced as each individual component of the chain sequentially transverses the nanopore constriction.

2 Nanopores have proven their potential for DNA sequencing^{54,59}. Exploiting nanopores for single-molecule protein sequencing is the next frontier. This is by no means an easy task, as numerous challenges need to be tackled to sequence a protein with a nanopore. First, amino acid residues vary widely in charge distribution, unlike DNA that is essentially uniformly charged. Electrophoresis-driven unidirectional translocation of polypeptides through nanopores thus cannot be simply employed. Second, most proteins are folded in their native state. Disruption of their secondary and tertiary structure is necessary to thread them through a nanopore. Third, protein sequencing requires distinction of 20 different amino acids, a fivefold larger number than the four bases in DNA sequencing.

First translocations of polypeptides through nanopores were performed using peptides of only 20 to 30 amino acids^{60,61,62,63,64}. Short peptides lack stable tertiary structure and can translocate without the need of denaturing agents. In these studies, peptides containing specific motifs such as beta-hairpins, alpha-helices or collagen-like helices were analysed using alpha-haemolysin and aerolysin nanopores. This research elucidated important aspects about the kinetics of polypeptide translocation and emphasized the crucial role of peptide–nanopore interactions during the passage of the molecule. In particular, the detailed work presented by Bayley et al. on helical peptides containing the (AAKAA)_n sequence provided key insights into the process of protein capture and partitioning into the nanopore⁶².

While the translocation of peptides continues to be a valuable model system to understand basic steps in the complex process of protein translocation^{65,66}, the final end of a nanopore-based protein sequencer is to read entire proteins, which requires protein denaturation. Multiple chemical and physical methods have been proposed for protein unfolding in nanopore analysis. Several groups have shown the successful unfolding and translocation of proteins through solid-state nanopores using strong denaturants such as urea, sodium dodecyl sulphate (SDS) or guanidine hydrochloride (GdnHCl)^{67,68,69}. Translocation of proteins through biological nanopores using denaturants has also been achieved^{70,71,72}. In this context, solid-state nanopores have an advantage over biological nanopores, displaying higher stability when exposed to extreme buffer conditions (8 M urea, 6 M GdnHCl or 1% SDS).

Biological channels are more susceptible to denaturing conditions than solid-state devices, but can remarkably withstand concentrations of up to 4 M urea and 1.5 M GdnHCl (ref. 73). These concentrations are sufficient to break the structure of some protein substrates and allow translocation. For example, pioneering work (Fig. 2.4a,b), which showed protein unfolding and translocation through alpha-haemolysin for the first time, was done using the maltose binding protein, which could be unfolded at low denaturant concentrations (0.8

M GdnHCl)⁷⁰.

Physical methods such as high temperature have been used to unfold proteins in both solid-state and biological nanopores^{74,75}. Pelta and colleagues studied the thermal denaturation of a maltose binding protein variant in a temperature range from 20 °C to 70 °C in both alpha-haemolysin and aerolysin nanopores⁷⁵. Temperature facilitates protein unfolding, but speeds up translocation dynamics, which makes sequencing more challenging. In a similar way, two research groups have shown that high voltages help stretch proteins during the movement through solid-state nanopores^{76,77,78}. These approaches are not compatible with biological nanopores due to the electroporation of the lipid bilayer at high voltages (~0.4 V), and also cause an increase in translocation speed.

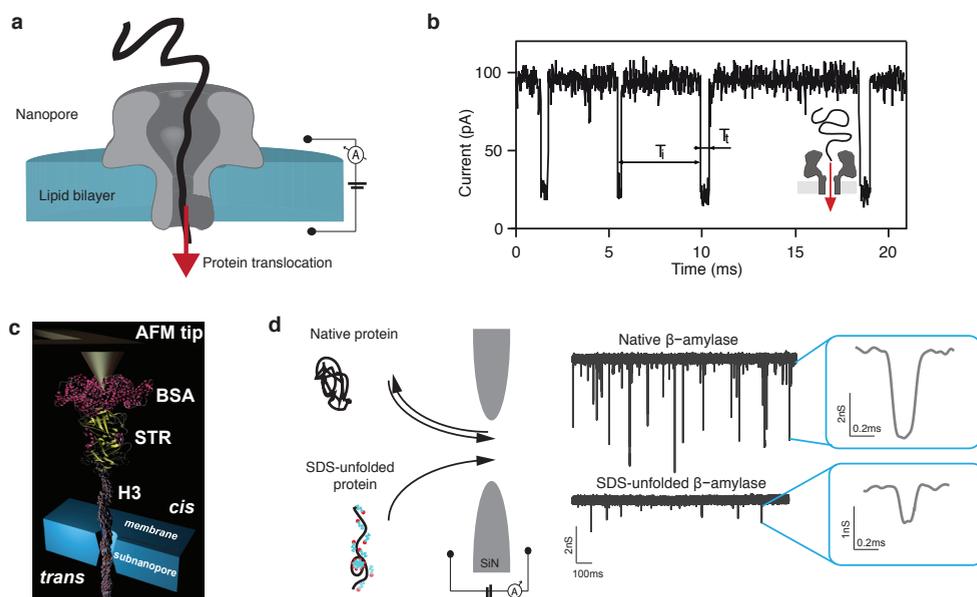


Figure 2.4: Translocation of peptides and unfolded proteins through nanopores. a, Schematic representation of a biological nanopore set-up. b, Representative current traces when GdnHCl was used for unfolding and translocation of a maltose binding protein through an alpha-haemolysin pore. T_i indicates the inter-event time, and T_t the translocation time. c, Schematic where a protein is immobilized at an atomic force microscope tip and translocated through a nanopore. BSA, bovine serum albumin; STR, streptavidin. d, Schematic of native and SDS-unfolded protein translocation through a solid-state nanopore including typical current traces of native and SDS-unfolded proteins. Panel b adapted from ref. 70, APS; panel c reproduced from ref. 82, American Chemical Society; and panel d reproduced from ref. 69 under a Creative Commons license (<https://creativecommons.org/licenses/by/3.0/>).

A major roadblock for the development of a protein sequencer with nanopores is the non-uniform charge distribution of amino acid residues. Unlike DNA that is uniformly charged and moves through a nanopore by electrophoretic forces, proteins carry different

2 local charges. It is therefore not well-defined if electrophoretic or electro-osmotic forces on the protein dominate the transport (unless it is set by the electro-osmotic force due to ions at the nanopore surface)^{79,80}. One way to address this issue is to use SDS as a denaturant. SDS not only unfolds proteins, but also wraps them around with a homogeneous negative charge given by the sulphate groups in the head of the detergent. SDS has been used to enforce proteins through pores with sub-nanometre diameters, hinting at the potential of using a nanopore for differentiating individual amino acids (Fig. 2.4c)^{81,82}. A more comprehensive understanding of the effect of SDS on protein unfolding and translocation was presented by our group⁶⁹. Experiments showed that SDS could unfold stably folded proteins such as titin and beta-amylase (Fig. 2.4d). Additionally, a consistent direction of translocation was induced by the electrophoretic force, thanks to the negative charge conveyed by SDS.

An alternative approach to control the direction of translocation is to attach an oligonucleotide strand to the N- or C-terminus of a protein. The negative charge carried by this lead sequence drags the polypeptide in the direction of the electrophoretic force^{83,84,85,86,87}. This principle was first used by Bayley and colleagues to study the translocation of thioredoxin through alpha-haemolysin^{83,84}. In their work, a 30-mer oligonucleotide was attached to the C-terminus of the protein and upon adding the substrate to the cis compartment, a repetitive pattern with multiple current levels was observed, which corresponded to the capture of the DNA tag, the local unfolding of the C-terminus and the unfolding of the remaining of the protein (Fig. 2.5a). The partially unfolded intermediate in which the C-terminus of the protein was locally unfolded and translocated through the constriction of the nanopore was further used to discriminate between unphosphorylated, monophosphorylated and diphosphorylated proteins⁸⁵. Other groups have also recently used this approach. Lindsay and colleagues developed a simple and effective click chemistry to facilitate the tagging reaction, while Pelta and colleagues used a DNA lead in a protein to present a direct proof of protein translocation using amplification by polymerase chain reaction^{86,87}.

In all the studies presented this far, the translocation of proteins occurs at timescales faster than 1 ms, which is too fast for sequencing purposes. Indeed, single-protein translocations characteristically occur very fast⁸⁸. Control of the translocation speed will be necessary to guarantee ample time for the accurate reading of different amino acids by a nanopore.

The controlled and unidirectional movement of DNA through a nanopore using helicases or polymerases marked a breakthrough in the development of a nanopore-based DNA sequencer. Akesson and colleagues proposed a similar approach for proteins^{89,90}. In their work, a motor enzyme, ClpX, unfolds and pulls the polypeptide chain in a controlled manner through alpha-haemolysin. ClpX translocates proteins at a speed slow enough for sequencing (80 amino acids per second), with defined step-sizes, and it generates a strong

enough force (~ 20 pN) to unfold proteins⁹¹. In their experimental scheme (Fig. 2.5b) a lipid bilayer containing alpha-haemolysin separates two compartments. The cis side contains a protein known as Smt3, which is modified with a 65-amino acid negatively charged extension and an *ssrA* tag. The *ssrA* tag is necessary for ClpX recognition and the 65-amino acid extension is used as an unstructured anchor that orients the protein and allows the *ssrA* tag to be exposed to the trans side, where ClpX is added. Time traces showed the process of substrate capture and translocation by ClpX. In a follow-up study⁹⁰, a machine-learning algorithm with three parameters (dwell time, average current amplitude and standard deviation of the current amplitude) was used to distinguish different domains as well as variants of those domains such as mutations or truncations.

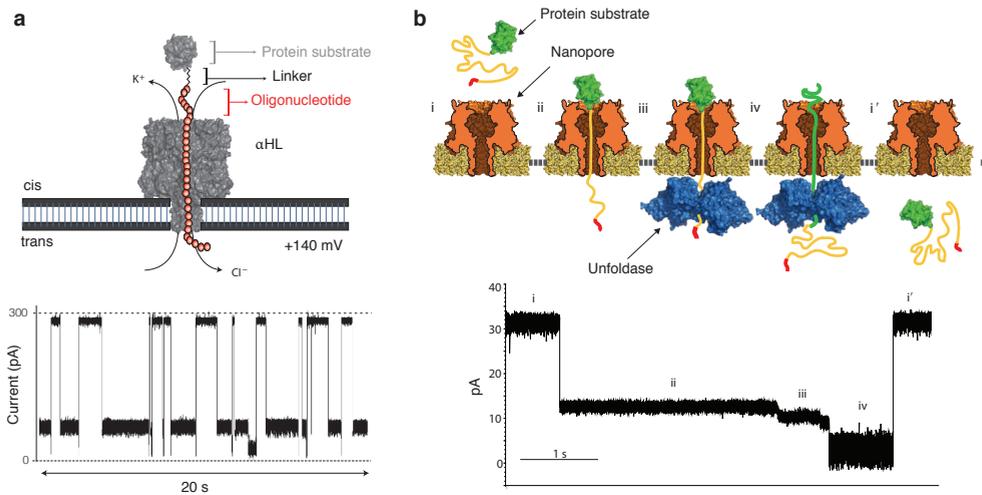


Figure 2.5: Translocation of unfolded proteins through nanopores using an oligonucleotide linker (left) or an unfoldase (right). a, Schematic in which a DNA strand is used as lead for protein unfolding and translocation (top). Current traces observed for the translocation of DNA-tagged proteins (bottom). b, Experimental set-up in which an unfoldase is used to unfold and pull the protein substrate (top). Typical current trace observed during a translocation and unfolding event (bottom). Panel a adapted from ref. 83, Springer Nature Ltd; panel b adapted from ref. 89, Springer Nature Ltd.

This approach overcomes two critical requirements for protein sequencing using nanopores: protein unfolding and controlled translocation of the substrate. The main drawback of this method is the need to add a polypeptide extension in the substrate. This could, however, be overcome by chemically attaching a polypeptide to the N-terminus of proteins. The high level of noise in the signal also needs to be improved. Other approaches have been proposed, but lack experimental proof^{92,93,94,95}. The use of a double pore system in which two nanopores are placed in series has been proposed⁹². As the polypeptide transverses the first pore, it is cleaved by an exopeptidase, and the amino acids released by the enzyme are then analysed with a second nanopore. The use of perpendicular nanochannels in which a protein is stretched in the longitudinal direction, while ionic current is recorded transversally,

2 has also been proposed⁹³, as has the use of graphene to control polypeptide translocation. Graphene and other two-dimensional materials are proposed as attractive nanopore membranes since they can be atomically thin, thereby improving the spatial resolution required to detect individual amino acids⁴⁹. Using molecular dynamics simulations, it was shown that proteins and peptides collapsed on top of a graphene membrane by the surface absorption of amino acids, leading to a slow stepwise motion of amino acids into a nanopore⁹⁴. There is also a noticeable attempt of repurposing nanopores for improving mass spectrometry. The use of solid-state nanopores to create a renewed version of a mass spectrometer, in which the electrospray ionization, conventionally done with micrometre-sized nozzles, is initiated from a nanopore has been proposed. This could potentially allow proteins to be sequenced if they are fragmented as they pass through the nanopore and individual amino acids are sequentially ionized and detected⁹⁵. For a more detailed description of efforts in improving the sensitivity of mass spectrometry, we refer to other reviews^{14,96}.

In summary, great advances have been presented with the nanopore approach towards sequencing peptides and proteins. It is an extremely active field of research, and therefore significant advances are anticipated for the development of a protein sequencer in the coming years. An advantage that a nanopore sequencer could provide is the possibility to perform long reads. Traditional sequencing methods such as Edman degradation and mass spectrometry rely on the digestion of proteins into short peptides, but nanopore devices would allow sequencing of full-length proteins. A major challenge is the control of the polypeptide translocation speed. Different approaches are being explored at the moment, and it is very likely that enzyme-assisted translocations will command this step, as was the case for DNA sequencing. Exploring a pool of unfoldases beyond ClpX will be a critical step to accomplish this aim.

2.5 Outlook

The human genome project opened the door to exciting years of genomic research. The coming years will see significant progress in other omics, especially proteomics. In this area, the development of single-molecule approaches will be key for achieving the sensitivity and dynamic range required for protein analysis. Colossal efforts are on-going in the fields of single-molecule fluorescence, tunnelling currents and nanopores. In this Review, we presented the main approaches proposed up to now for single-molecule protein sequencing, with their strengths and limitations. Table 2.1 summarizes the different schemes presented, taking into consideration relevant criteria for the development of a protein sequencer, such as read length, and the possibility to perform de novo sequencing.

We anticipate that first single-molecule protein identification systems may appear as soon

as within five years. First systems will most probably rely on a fingerprinting scheme such as those proposed by Marcotte et al. and Joo et al. Marcotte's approach has the advantage of relying entirely on chemical reactions, which could lead to a robust device for in situ analysis. Major disadvantages of this approach are its slow speed, and the fact that only short peptides can be analysed. Alternatively, the scheme proposed by Joo et al. can be used for the analysis of full-length proteins, but on the down side, unfoldase engineering or substrate pre-processing need to be worked out for substrate recognition. Both methods need to overcome the challenge of reading multiple fluorophores with minimal error.

	Method	Read length	Potential for de novo sequencing	Labelling required	Proof of concept
Fluorescence	FRET scanning using ClpX	Full length	No	Yes	Computational (ref. ²⁵) Peptide analysis (ref. ²⁸)
	Edman degradation	A few amino acids	No	Yes	Computational (ref. ²⁶)
Tunnelling current	Recognition tunnelling	Full length if coupled with a nanopore or enzyme	Yes	No	Single-molecule measurements (ref. ⁴³)
	Sub-nanometre break junctions	Full length if coupled with a nanopore or enzyme	Yes	No	Single-molecule measurements (ref. ⁴⁴)
Nanopore	Solid-state nanopore	Full length	Yes	No	Single-molecule measurements (refs ^{67,68,69,81})
	Graphene nanopore	Full length	Yes	No	Computational (ref. ⁹⁴)
	Biological nanopore	Full length	Yes	No	Single-molecule measurements (refs ^{83,84})
	Biological nanopore coupled with an enzyme	Full length	Yes	No	Single-molecule measurements showing controlled translocation (refs ^{89,90})

Table 2.1: Summary of single-molecule protein sequencing approaches.

Nanopore research is moving fast in the direction of protein analysis and protein sequencing. A nanopore-based protein sequencer has the potential to be commercialized in the next decade. The main challenges revolve around the controlled translocation of proteins through the nanopore and the read-out. Akesson's approach, in which a ClpX enzyme was used to translocate a polypeptide through an alpha-haemolysin nanopore, is currently the only system in which a protein is unfolded and transported in a controlled way through the nanopore. The large levels of noise observed in their signals, however, obstructed the identification of specific amino acids. As has become clear from high-resolution DNA sequencing^{51,97}, alternative configurations schemes and possibly different enzymes should be explored.

A remaining question is whether the measurement of ionic currents will provide the sufficient resolution for the identification of 20 amino acids using nanopores. The experimental results from Lindsay and Kawai indicate that tunnelling currents are extremely sensitive,

and can differentiate molecules with minor structural differences. Thereby, the integration of a nanopore system for controlled transport with the sensitive measurement of tunnelling currents is an attractive alternative that would potentially allow single-molecule de novo protein sequencing.

2 A major aim of a single-molecule protein sequencer would be the development of a tool for single-cell analysis. Current attempts to single-cell proteomics^{16,17,18,19}, such as mass cytometry²⁰, rely on labelled antibodies. The reduced availability of highly specific antibodies and distinguishable labels limits these techniques to the detection of 10 to 40 proteins per cell, a minute fraction of the proteome. Single-molecule detection methods will not require such a preparatory step, and could, in principle, detect thousands of proteins from individual cells. A critical aspect that needs to be resolved is the manipulation and extraction of proteins from single cells without substantial losses or biases¹⁷. Recent advances in microfluidic devices, where proteins from single cells have been extracted and labelled on chip¹², show first steps towards this goal.

The realization of a single-molecule protein sequencer is technically very challenging. If realized, however, it would revolutionize proteomics research by facilitating the identification of low abundance proteins and achievement of true single-cell proteomics. Low abundance proteins are crucial in biomedical research as they allow the identification of disease-specific biomarkers⁹⁸. Moreover, sensitivity from single-molecule detectors could allow access to the so-called human dark proteome. The dark proteome comprises approximately 3,000 human proteins that have never been directly identified, despite evidence of their existence in genetic or transcriptional information⁹⁹. Besides protein identification, the detection of low abundance proteins can be beneficial for the study of post-translational modifications, reducing the need of complex enrichment processes. Finally, the possibility to perform single-cell proteomic analysis opens the possibility for exciting proteomics research, allowing scientists to study the change in protein expression of individual cells under specific stimuli.

References

1. Miyashita, M. et al. Attomole level protein sequencing by Edman degradation coupled with accelerator mass spectrometry. *Proc. Natl Acad. Sci. USA* 98, 4403–8 (2001).
2. Shimonishi, Y. et al. Sequencing of peptide mixtures by Edman degradation and field desorption mass spectrometry. *Eur. J. Biochem.* 112, 251–264 (1980).
3. Bradley, C. V., Williams, D. H. & Hanley, M. R. Peptide sequencing using the combination of edman degradation, carboxypeptidase digestion and fast atom bombardment mass spectrometry. *Biochem. Biophys. Res. Commun.* 104, 1223–30 (1982).
4. Steen, H. & Mann, M. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 5, 699–711 (2004).
5. Yates, J. R. III A century of mass spectrometry: from atoms to proteomes. *Nat. Methods* 8, 633–637 (2011).
6. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355 (2016).
7. Walther, T. C. & Mann, M. Mass spectrometry-based proteomics in cell biology. *J. Cell Biol.* 190, 491–500 (2010).
8. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* 28, 710–721 (2010).
9. A cast of thousands. *Nat. Biotechnol.* 21, 213 (2003).
10. Anderson, N. L. The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867 (2002).
11. Zubarev, R. A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 13, 723–726 (2013).
12. Huang, B. et al. Counting low-copy number proteins in a single cell. *Science* 315, 81–84 (2007).
13. Ham, B. M. & MaHam, A. *Analytical Chemistry: A Chemist and Laboratory Technician's Toolkit* (Wiley, Hoboken, 2015).
14. Hawkridge, A. M. in *Quantitative Proteomics* (eds Eyers, C. E. & Gaskell, S.) 1–25 (RSC, Cambridge, 2014).
15. Pagel, O., Lorocho, S., Sickmann, A. & Zahedi, R. P. Current strategies and findings in clinically relevant post-translational modification-specific proteomics. *Expert Rev. Proteomics* 12, 235–253 (2015).
16. Heath, J. R., Ribas, A. & Mischel, P. S. Single-cell analysis tools for drug discovery and development. *Nat. Rev. Drug Discov.* 15, 204–216 (2015).
17. Prakadan, S. M., Shalek, A. K. & Weitz, D. A. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. *Nat. Rev. Genet.* 18, 345–361 (2017).
18. Su, Y., Shi, Q. & Wei, W. Single cell proteomics in biomedicine: High-dimensional

- data acquisition, visualization, and analysis. *Proteomics* 17, 1600267 (2017).
19. Lu, Y., Yang, L., Wei, W. & Shi, Q. Microchip-based single-cell functional proteomics for biomedical applications. *Lab Chip* 17, 1250–1263 (2017).
 20. Spitzer, M. H. & Nolan, G. P. Mass Cytometry: Single Cells, Many Features. *Cell* 165, 780–791 (2016).
 21. Bentley, D. R. et al. Accurate whole human genome sequencing using reversible terminator chemistry. - Supplement. *Nature* 456, 53–9 (2008).
 22. Eid, J. et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323, 133–138 (2009).
 23. Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S. R. Sequence information can be obtained from single DNA molecules. *Proc. Natl Acad. Sci. USA* 100, 3960–3964 (2003).
 24. Hernandez, E. T., Swaminathan, J., Marcotte, E. M. & Anslyn, E. V. 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, 462–469 (2017).
 25. Yao, Y., Docter, M., van Ginkel, J., de Ridder, D. & Joo, C. Single-molecule protein sequencing through fingerprinting: computational assessment. *Phys. Biol.* 12, 055003 (2015).
 26. Swaminathan, J., Boulgakov, A. A. & Marcotte, E. M. A Theoretical justification for single molecule peptide sequencing. *PLOS Comput. Biol.* 11, e1004080 (2015).
 27. Müller, V. & Westerlund, F. Optical DNA mapping in nanofluidic devices: principles and applications. *Lab Chip* 17, 579–590 (2017).
 28. van Ginkel, J. et al. Single-molecule peptide fingerprinting. *Proc. Natl Acad. Sci. USA* 115, 3338–3343 (2018).
 29. Preminger, M. & Smilansky, Z. Methods for evaluating ribonucleotide sequences. US patent 9,012,150 (2009).
 30. Stevens, B. et al. FRET-based identification of mRNAs undergoing translation. *PLoS One* 7, e38344 (2012).
 31. Swaminathan, J. Single Molecule Peptide Sequencing. PhD thesis, University of Texas at Austin (2015).
 32. Borgo, B. & Havranek, J. J. Computer-aided design of a catalyst for Edman degradation utilizing substrate-assisted catalysis. *Protein Sci.* 24, 571–579 (2015).
 33. Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* 29, 277–283 (1974).
 34. Dekker, C., Tans, S. J., Oberdorff, B., Meyer, R. & Venema, L. C. STM imaging and spectroscopy of single copperphthalocyanine molecules. *Synth. Met.* 84, 853–854 (1997).
 35. Reed, M. A. Conductance of a Molecular Junction. *Science* 278, 252–254 (1997).

36. Ratner, M. A brief history of molecular electronics. *Nat. Nanotech.* 8, 378–381 (2013).
37. Tsutsui, M., Taniguchi, M., Yokota, K. & Kawai, T. Identifying single nucleotides by tunnelling current. *Nat. Nanotech.* 5, 286–290 (2010).
38. Tanaka, H. & Kawai, T. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nat. Nanotech.* 4, 518–522 (2009).
39. Shapir, E. et al. Electronic structure of single DNA molecules resolved by transverse scanning tunnelling spectroscopy. *Nat. Mater.* 7, 68–74 (2008).
40. Chang, S. et al. Electronic signatures of all four DNA nucleosides in a tunneling gap. *Nano Lett.* 10, 1070–1075 (2010).
41. Feng, J. et al. Identification of single nucleotides in MoS₂ nanopores. *Nat. Nanotech.* 11, 117–126 (2015).
42. Lindsay, S. et al. Recognition tunneling. *Nanotechnology* 21, 262001 (2010).
43. Zhao, Y. et al. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotech.* 9, 466–73 (2014).
44. Ohshiro, T. et al. Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat. Nanotech.* 9, 835–840 (2014).
45. Morikawa, T., Yokota, K., Tsutsui, M. & Taniguchi, M. Fast and low-noise tunnelling current measurements for single-molecule detection in an electrolyte solution using insulator-protected nanoelectrodes. *Nanoscale* 9, 4076–4081 (2017).
46. Morikawa, T., Yokota, K., Tanimoto, S., Tsutsui, M. & Taniguchi, M. Detecting single-nucleotides by tunneling current measurements at sub-MHz temporal resolution. *Sensors* 17, 885–893 (2017).
47. Taniguchi, M., Tsutsui, M., Yokota, K. & Kawai, T. Fabrication of the gating nanopore device. *Appl. Phys. Lett.* 95, 123701 (2009).
48. Yokota, K., Tsutsui, M. & Taniguchi, M. Electrode-embedded nanopores for label-free single-molecule sequencing by electric currents. *RSC Adv.* 4, 15886–15899 (2014).
49. Heerema, S. J. & Dekker, C. Graphene nanodevices for DNA sequencing. *Nat. Nanotech.* 11, 127–136 (2016).
50. Ivanov, A. P. et al. DNA tunneling detector embedded in a nanopore. *Nano Lett.* 11, 279–285 (2011).
51. Lu, H., Giordano, F. & Ning, Z. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14, 265–279 (2016).
52. Jain, M. et al. Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* 12, 351–356 (2015).
53. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17, 239–249 (2016).

54. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524 (2016).
55. Gaskill, M. First DNA sequencing in space a game changer. NASA (2016); https://www.nasa.gov/mission_pages/station/research/news/dna_sequencing
56. Waduge, P. et al. Nanopore-based measurements of protein size, fluctuations, and conformational changes. *ACS Nano.* 11, 5706–5716 (2017).
57. Bell, N. A. W. & Keyser, U. F. Digitally encoded DNA nanostructures for multiplexed, single-molecule protein sensing with nanopores. *Nat. Nanotech.* 11, 645–651 (2016).
58. Plesa, C., Ruitenberg, J. W., Witteveen, M. J. & Dekker, C. Detection of individual proteins bound along DNA using solid-state nanopores. *Nano Lett.* 15, 3153–3158 (2015).
59. Venkatesan, B. M. & Bashir, R. Nanopore sensors for nucleic acid analysis. *Nat. Nanotech.* 6, 615–624 (2011).
60. Stefureac, R., Long, Y.-T., Kraatz, H.-B., Howard, P. & Lee, J. S. Transport of alpha-helical peptides through alpha-hemolysin and aerolysin pores. *Biochemistry* 45, 9172–9179 (2006).
61. Sutherland, T. C. et al. Structure of peptides investigated by nanopore analysis. *Nano Lett.* 4, 1273–1277 (2004).
62. Movileanu, L., Schmittschmitt, J. P., Martin Scholtz, J. & Bayley, H. Interactions of peptides with a protein pore. *Biophys. J.* 89, 1030–1045 (2005).
63. Goodrich, C. P. et al. Single-molecule electrophoresis of beta-hairpin peptides by electrical recordings and langevin dynamics simulations. *J. Phys. Chem. B* 111, 3332–3335 (2007).
64. Mohammad, M. M. & Movileanu, L. Excursion of a single polypeptide into a protein pore: simple physics, but complicated biology. *Eur. Biophys. J.* 37, 913–925 (2008).
65. Mahendran, K. R., Romero-Ruiz, M., Schlösinger, A., Winterhalter, M. & Nussberger, S. Protein translocation through Tom40: Kinetics of peptide release. *Biophys. J.* 102, 39–47 (2012).
66. Ji, Z. et al. Fingerprinting of peptides with a large channel of bacteriophage Phi29 DNA packaging motor. *Small* 12, 4572–4578 (2016).
67. Talaga, D. S & Li, J. Single-molecule protein unfolding in solid state nanopores. *J. Am. Chem. Soc.* 131, 9287–9297 (2009).
68. Li, J., Fologea, D., Rollings, R. & Ledden, B. Characterization of protein unfolding with solid-state nanopores. *Protein Pept. Lett.* 21, 256–265 (2014).
69. Restrepo-Pérez, L., John, S., Aksimentiev, A., Joo, C. & Dekker, C. SDS-assisted protein transport through solid-state nanopores. *Nanoscale* 9, 11685–11693 (2017).
70. Oukhaled, G. et al. Unfolding of proteins and long transient conformations detected by single nanopore recording. *Phys. Rev. Lett.* 98, 158101 (2007)

71. Pastoriza-Gallego, M. et al. Dynamics of unfolded protein transport through an aerolysin pore. *J. Am. Chem. Soc.* 133, 2923–2931 (2011).
72. Merstorff, C. et al. Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording. *ACS Chem. Biol.* 7, 652–658 (2012).
73. Pastoriza-Gallego, M. et al. Urea denaturation of alpha-hemolysin pore inserted in planar lipid bilayer detected by single nanopore recording: Loss of structural asymmetry. *FEBS Lett.* 581, 3371–3376 (2007).
74. Freedman, K. J. et al. Chemical, thermal, and electric field induced unfolding of single protein molecules studied using nanopores. *Anal. Chem.* 83, 5137–5144 (2011).
75. Payet, L. et al. Thermal unfolding of proteins probed at the single molecule level using nanopores. *Anal. Chem.* 84, 4071–4076 (2012).
76. Cressiot, B. et al. Protein transport through a narrow solid-state nanopore at high voltage: Experiments and theory. *ACS Nano* 6, 6236–6243 (2012).
77. Oukhaled, A. et al. Dynamics of completely unfolded and native proteins through solid-state nanopores as a function of electric driving force. *ACS Nano* 5, 3628–3638 (2011).
78. Freedman, K. J., Haq, S. R., Edel, J. B., Jemth, P. & Kim, M. J. Single molecule unfolding and stretching of protein domains inside a solid-state nanopore by electric field. *Sci. Rep.* 3, 1638 (2013).
79. Firnkes, M., Pedone, D., Knezevic, J., Döblinger, M. & Rant, U. Electrically facilitated translocations of proteins through silicon nitride nanopores: Conjoint and competitive action of diffusion, electrophoresis, and electroosmosis. *Nano Lett.* 10, 2162–2167 (2010).
80. Huang, G., Willems, K., Soskine, M., Wloka, C. & Maglia, G. Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores. *Nat. Commun.* 8, 935 (2017).
81. Kennedy, E., Dong, Z., Tennant, C. & Timp, G. Reading the primary structure of a protein with 0.07 nm³ resolution using a subnanometre-diameter pore. *Nat. Nanotech.* 11, 968–976 (2016).
82. Dong, Z., Kennedy, E., Hokmabadi, M. & Timp, G. Discriminating residue substitutions in a single protein molecule using a sub-nanopore. *ACS Nano* 11, 5440–5452 (2017).
83. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotech.* 8, 288–295 (2013).
84. Rodriguez-Larrea, D. & Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat. Commun.* 5, 4841 (2014).
85. Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* 32, 179–181 (2014).

86. Biswas, S., Song, W., Borges, C., Lindsay, S. & Zhang, P. Click addition of a DNA thread to the N-termini of peptides for their translocation through solid-state nanopores. *ACS Nano* 9, 9652–9664 (2015).
87. Pastoriza-Gallego, M. et al. Evidence of unfolded protein translocation through a protein nanopore. *ACS Nano* 8, 11350–11360 (2014).
88. Plesa, C. et al. Fast translocation of proteins through solid state nanopores. *Nano Lett.* 13, 658–63 (2013).
89. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an alpha-hemolysin nanopore. *Nat. Biotechnol.* 31, 247–250 (2013).
90. Nivala, J., Mulrone, L., LiG., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* 8, 12365–12375 (2014).
91. Aubin-Tam, M.-E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* 145, 257–67 (2011).
92. Sampath, G. Amino acid discrimination in a nanopore and the feasibility of sequencing peptides with a tandem cell and exopeptidase. *RSC Adv.* 5, 30694–30700 (2015).
93. Boynton, P. & Di Ventra, M. Sequencing proteins with transverse ionic transport in nanochannels. *Sci. Rep.* 6, 25232 (2016).
94. Wilson, J., Sloman, L., He, Z. & Aksimentiev, A. Graphene nanopores for protein sequencing. *Adv. Funct. Mater.* 26, 4830–4838 (2016).
95. Maulbetsch, W., Wiener, B., Poole, W., Bush, J. & Stein, D. Preserving the sequence of a biopolymer's monomers as they enter an electrospray mass spectrometer. *Phys. Rev. Appl.* 6, 054006 (2016).
96. Keifer, D. Z. & Jarrold, M. F. Single-molecule mass spectrometry. *Mass Spec. Rev.* 36, 715–733 (2016).
97. Manrao, E. A. et al. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* 30, 349–353 (2012).
98. Million, R. et al. High abundance proteins depletion vs low abundance proteins enrichment: Comparison of methods to reduce the plasma proteome complexity. *PLoS One* 6, e19603 (2011).
99. Baker, M. S. et al. Accelerating the search for the missing proteins in the human proteome. *Nat. Commun.* 8, 14271 (2017).
100. Wetterstrand, K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcostsdata (Accessed 2 July 2018).
101. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351 (2016).
102. Haider, S. & Pal, R. Integrated analysis of transcriptomic and proteomic data. *Curr.*

- Genomics 14, 91–110 (2013).
103. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* 405, 837–846 (2000).
 104. Tyers, M. & Mann, M. From genomics to proteomics. *Nature* 422, 193–197 (2003).
 105. Edman, P. Method for determination of the amino acid sequence in peptides. *Acta Chemica Scandinavica* 4, 283–293 (1950).
 106. Li, K. W. & Geraerts, W. P. M. in *Neuropeptide Protocols* (eds Irvine, G. B. & Williams, C. H.) 17–26 (Humana Press, New York, 1997).
 107. McCormack, A. L. et al. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* 69, 767–776 (1997).

3

Peptide analysis with the FraC nanopore: Using a bipolar peptide gate to modulate nanopore conductance

Peptides of different charge, size and structure have been previously studied with biological nanopores. Peptides often serve as a simplified model system for the study of protein-nanopore interactions. Here we explore the use of a bipolar peptide as model system. This peptide contains 10 negative amino acids in the N-terminus and 10 positive amino acids in the C-terminus. The opposite charges at the ends of the peptide generate force in opposing directions stalling the peptide while it translocates the FraC pore. This mechanism allows for extended examination of the central region of the peptide, which is further explored in Chapter 4 and 5. In this chapter we explore other interesting feature observed in this peptide-FraC system: the non-linear current voltage dependence observed for the system. Our experiments and molecular dynamics simulations show that the relative blockade caused by the peptide changes depending on the applied bias as a consequence of peptide stretching and compressing. In this chapter we propose to use this bipolar peptide as an electromechanical gate that can modulate nanopore conductance in a dynamic manner and at the single-molecule level.

This chapter has been published as: Zhao S.*, Restrepo-Pérez L.*, Soskine M. , Maglia G. , Joo C., Dekker C. and Aksimentiev A. Electro-mechanical conductance modulation of a nanopore using a removable gate. *ACS Nano*. 13 (2), 2398–2409 (2019).

3.1 Introduction

Lipid membranes play a crucial role in eukaryotic cells because they separate the cell's cytoplasm from the outside environment and from the interior of the intracellular organelles, leaving the job of transporting signals and nutrient across the membranes to membrane-embedded proteins called membrane channels.⁽¹⁾ One particular class of membrane channels, ion channels, facilitates transmembrane transport of specific ion species, enabling, among many other biological functions, neuronal activity,^(2,3) and the senses of sound, smell, sight, taste, and touch. A critical feature of an ion channel's function is gating, i.e., the ability of an ion channel to change its conductance in response to external stimuli.

The gating of a membrane channel usually involves the opening or closing of the transmembrane pore that connects the opposite sides of a membrane with a water-filled passage that allows for ion transport across it. Diverse mechanisms have evolved to regulate the membrane channel's conductance in response to external stimuli (Figure 3.1), which include a change of the transmembrane voltage, ligand binding, pH, light, temperature, and tension.⁽⁴⁾ A voltage-gated ion channel opens its transmembrane pore when the transmembrane voltage exceeds a threshold value;⁽⁵⁾ see Figure 3.1a. The closing of a voltage-gated channel can proceed as a reversal of its opening or involve an additional inactivation mechanism such as, for example, in sodium channels.⁽⁶⁾ In a ligand-gated channel, binding of a specific chemical compound to a binding site at the channel's surface results in a conformational transition that can either open or close the transmembrane pore; see Figure 3.1b. Mechanosensitive channels control the passage of ions and solutes through the cellular membrane in response to mechanical forces generated by other proteins or the membrane itself;^(7,8) see Figure 3.1c. With the exception of a few channel types,⁽⁹⁾ the gating of a biological membrane channel involves changes in the physical structure of the channel.⁽¹⁰⁾

Selective transport and gating of biological membrane channels has inspired biomimicry efforts.^(9,11–17) Using a focused, high-energy ion or electron beam, nanopores of controlled dimensions were produced in a variety of solid-state membranes.^(18,19) Because of their symmetric, uniformly charged and rigid structure, the vast majority of such nanopores exhibit a linear current–voltage dependence that nevertheless can sensitively depend on the chemical state of the nanopore surface and thus be exploited to demonstrate the pH,⁽²⁰⁾ light,⁽²¹⁾ and ligand⁽²²⁾ gating of ion conductance and rectification phenomena.⁽²³⁾ Nonlinear voltage-dependent phenomena can also occur in nanometer-diameter pores in ultrathin synthetic membranes,⁽²⁴⁾ where quantized ion conductance^(25,26) and water-compression gating⁽²⁷⁾ have been reported. Nanopores in multilayer membranes offer voltage gating functionality via additional electrodes that control the electrostatic potential within the membrane⁽²⁸⁾ and, thereby, the nanopore surface charge.⁽²⁹⁾ Ligand⁽³⁰⁾ and voltage^(31,32) gating was realized in synthetic nanopores made from DNA origami. Furthermore, biological

pores that were initially nongating were modified to incorporate synthetic covalently linked gating elements, such as DNA constructs⁽³³⁾ and proteins.⁽³⁴⁾

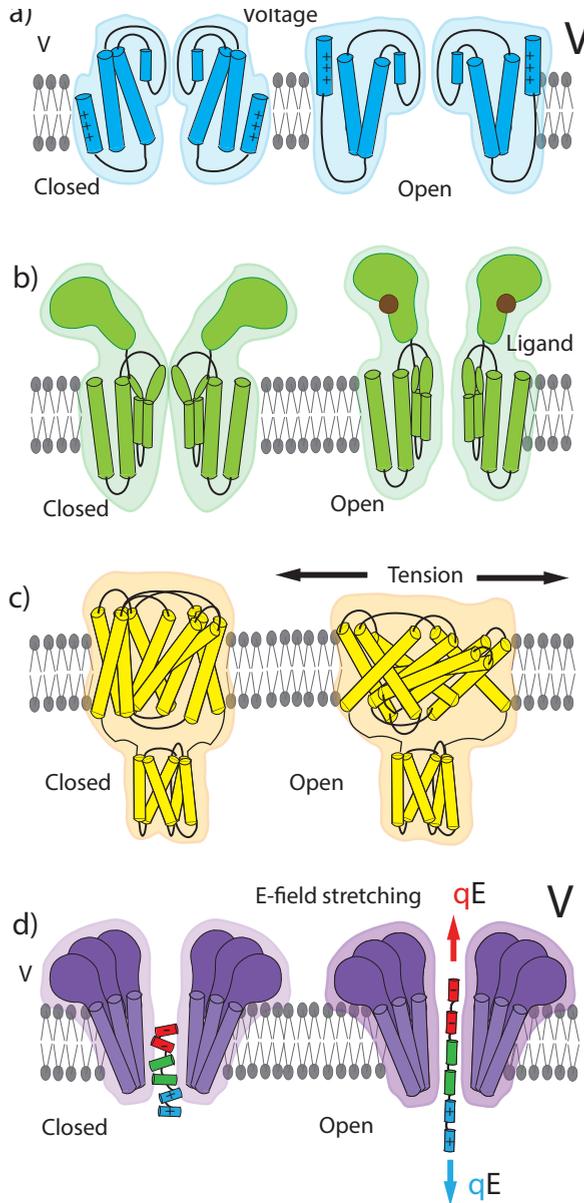


Figure 3.1: Mechanisms of membrane ion channel gating. (a) Voltage-gated ion channel changes its conformation in response to an increase of the transmembrane bias. (b) Ligand-gated channel opens or closes its transmembrane pore in response to ligand binding. (c) Mechanosensitive channel opens when the membrane tension exceeds a threshold value. (d) The electromechanical gating by an insertable peptide gate investigated in this work.

Here, we demonstrate another approach to modulation of a membrane channel's conductance: the inclusion of a removable gate that can be electromechanically stretched by an applied voltage; see Figure 3.1d. In contrast to the gating of a biological channel, which typically involves a change of the channel's conformation, or gating in a solid-state nanopore, which typically involves a change of the electrostatic environment, we modulate the conductance by first trapping a bipolar molecular gate⁽³⁵⁾ and then alter its conformation by varying the magnitude of the transmembrane bias. Our single-molecule electrophysiology recording and all-atom molecular dynamics (MD) simulations show that stretching of the molecular dipole can alter the ionic conductance of the assembly in a controllable and reproducible way. Our work sets the stage for the development of artificial nanopore systems with programmable voltage–current responses, which could, for example, be used to introduce an action potential-like response into a network of synthetic cells.⁽³⁶⁾

3.2 Results

To demonstrate the working principle of our removable gate, we used a wild-type fragaceatoxin C (FraC) nanopore as our model system (Figure 3.2a). As revealed by its crystal structure,⁽³⁷⁾ FraC is a conically shaped octameric transmembrane pore. At the trans opening of the pore, eight alpha-helices form a V-shaped channel that ends in a narrow pore constriction of 1.4 nm. Because of its narrow constriction and conical shape, FraC has been successfully used for the analysis of a wide range of peptides and proteins⁽³⁸⁾ as well as DNA.⁽³⁹⁾

Figure 3.2a shows our measurement set up. We first experimentally characterized the pore in its open state by measuring the current–voltage response of the channel (I–V curve) in the –120 to 120 mV range (Figure 3.2b). All of our measurements were performed in a buffer containing 1 M NaCl, 10 mM Tris, and 1 mM EDTA at pH 7.5. Our reference electrode was located in the cis compartment, while the working electrode was placed in the trans compartment. Figure 3.2b shows I–V recordings for nine FraC nanopores, all of which feature a persistent rectifying behavior. In addition to the current–voltage dependency, we also analyzed the conductance of individual channels in an extensive set of pores ($n = 93$). At an applied voltage of 50 mV, we observe a consistent conductance value of 1.76 ± 0.15 nS for positive bias and 2.39 ± 0.07 nS for negative bias; see Figure 3.2c.

Reproducing the experimental setup, we constructed an all-atom model of the FraC nanopore embedded in a lipid bilayer and solvated in 1 M NaCl solution; see Figure 3.2a and the Methods section for a detailed description of the structural model and simulation protocols. The resulting model was equilibrated in a 80 ns MD simulation, reaching a stable conformation characterized by the average root mean squared deviation from the crystal

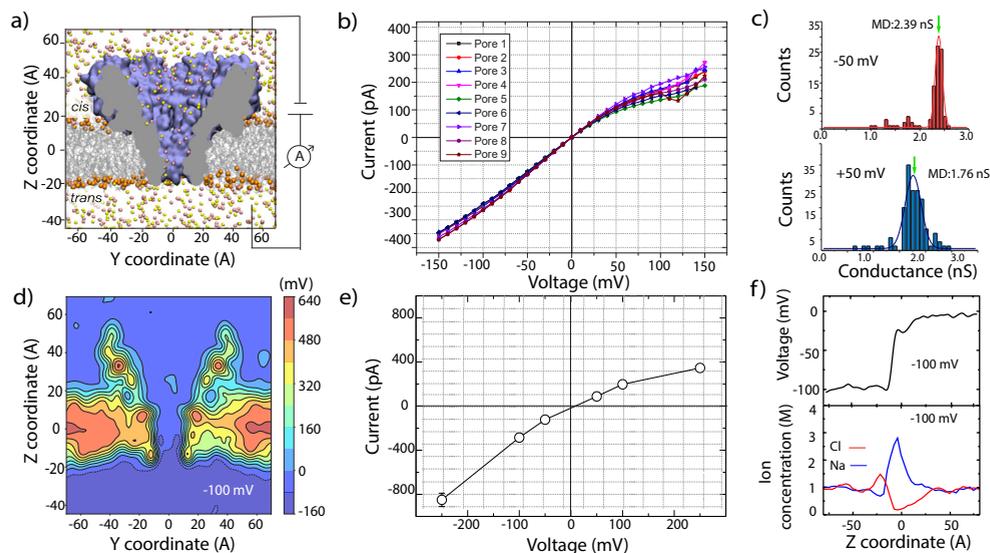


Figure 3.2: Ionic conductivity of the FraC nanopore. (a) All-atom model of a FraC nanopore (blue cut-away surface) embedded in a DPhPC membrane (gray lines and orange spheres) and submerged in 1 M NaCl solution (pink and yellow spheres representing Na⁺ and Cl⁻ ions, respectively). A bias of transmembrane electric potential is applied to produce a current of ions through the FraC nanopore. (b) Experimentally measured current–voltage dependence of nine FraC nanopores. (c) Experimentally determined distribution of the FraC conductance value under a negative (top) and positive (bottom) bias of 50 mV. Green arrows denote the conductance values obtained using the MD method. (d) Electrostatic potential map of FraC obtained from MD.⁽⁴⁰⁾ The potential map was produced by averaging the instantaneous distributions of the electrostatic potential over a 48 ns MD trajectory obtained at a –100 mV transmembrane bias. A 2D cross-section (parallel to the pore axis) of the 3D map is shown. (e) Simulated current–voltage dependence of the electric potential (top) and of the Na⁺ and Cl⁻ concentrations (bottom) along the central axis of the FraC nanopore. (f) Simulated profiles of the electric potential (top) and of the Na⁺ and Cl⁻ concentrations (bottom) along the central axis of the FraC nanopore. The data were obtained by averaging instantaneous configurations over a 48 ns MD trajectory.

lographic coordinates of about 3 Å; see Figure S3.1. The all-atom model of the open-pore FraC system was then simulated under a –100 mV transmembrane bias for 48 ns, which revealed the distribution of the electrostatic potential within and around the channel; see Figure 3.2d. Similar to electrostatic potential maps of alpha-hemolysin⁽⁴⁰⁾ and MspA,⁽⁴¹⁾ the transmembrane voltage is found to vary sharply within the constriction of the channel.

By repeating the simulation at several values of the transmembrane voltage and recording the resulting displacement of ions, we obtained the *in silico* I–V curve of the FraC nanopore; see Figure 3.2e. The simulated I–V curve reproduced many features of the experimentally determined I–V curve, in particular the ionic current rectification experimentally observed and its dependence on the voltage magnitude. Furthermore, the simulated ionic conductance values appear to be in quantitative agreement with the experimentally determined conductance values, with 1.76 nS at +50 mV and 2.39 nS at –50 mV (Figure 3.2c). The cur-

rent rectification in FraC is of inverse polarity to that previously observed for large conical solid-state and polymeric nanopores.^(42–45) Such an inversion of the current rectification was reported previously for nanopores with similar characteristics to FraC, i.e., a pore diameter that is comparable to the width of the electric double layer and strong ion selectivity.^(46,47) Indeed, under our simulation conditions, the current through FraC nanopore was cation-selective, with Na⁺ ions carrying 55 to 75% of the total current, in broad agreement with the ion selectivity of FraC established earlier for KCl.⁽³⁹⁾ The selective conductance produced an electro-osmotic effect: the direction and magnitude of the water flux through FraC followed the direction and magnitude of the current of Na⁺ species. The plot of the electrostatic potential along the symmetry axis of the nanopore (Figure 3.2f top panel) exhibits a sharp drop across the FraC constriction. Interestingly, the concentration of chloride ions within the FraC constriction is reduced, whereas the concentration of sodium ions is enhanced; see the Figure 3.2f bottom panel. The variation of the electrostatic potential is more gradual at a reverse bias polarity (Figure S3.2), whereas the concentration profiles remain largely unchanged. The observed distributions of the ionic species within FraC are explained well by the overall negative charge of the FraC constriction.

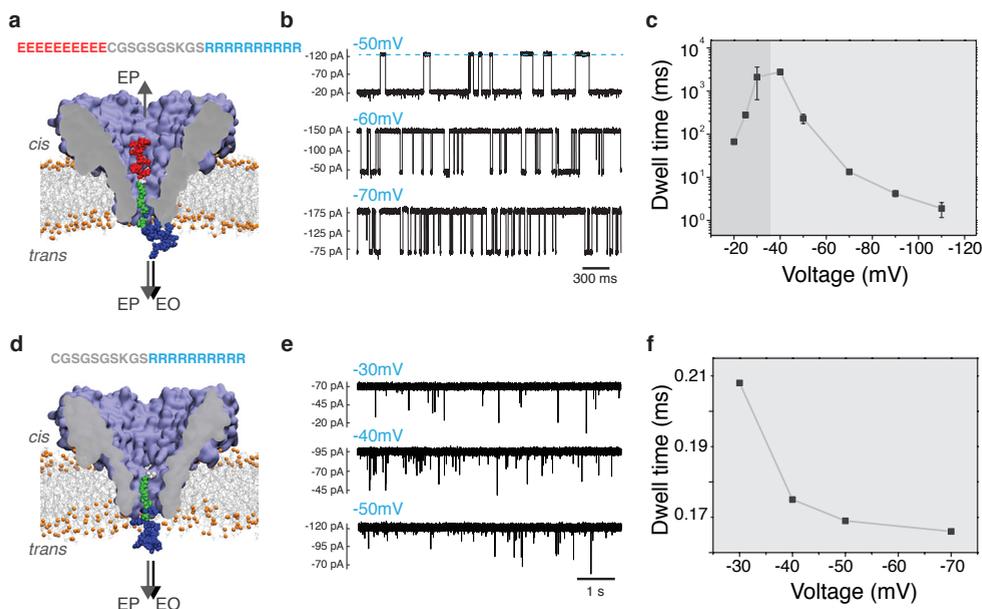


Figure 3.3: Experimental detection of the bipolar peptide translocation through FraC. (a) Schematic of the peptide translocation experiment. The sequence of the bipolar peptide is listed above the molecular graphics image of the nanopore. Negative bias is applied to the trans side of the nanopore. (b) Typical current traces observed after adding 0.4 μM of bipolar peptide to the cis compartment. Transient reductions of the current indicate interactions of the individual bipolar peptides with the FraC nanopore. (c) Average duration of the current blockade (dwell time) produced by the bipolar peptide as a function of the transmembrane bias. Below approximately -40 mV, the dwell time increases with voltage indicating transient entrapment of the peptide and, likely, subsequent escape through the cis entrance of the nanopore. At voltages higher by

magnitude than -40 mV, the dwell time decreases with the voltage indicating that the peptide exits the pore to the trans side. (d) Schematic representation of a control measurement performed using a truncated version of the bipolar peptide lacking the negatively charged segment (10 glutamate residues). (e) Typical current traces observed after adding 0.5 μM of the truncated peptide variant to the cis compartment. (f) Average residence time of the truncated peptide vs transmembrane bias.

Having established the ion conductance properties of the FraC nanopore, we experimentally characterized the effect of the electromechanical gate on the FraC conductance. Our electromechanical gate was a 30-amino acid bipolar peptide containing ten negatively charged amino acids at its N-terminus and ten positively charged amino acids at its C-terminus. A flexible linker region, consisting predominantly of glycine (G) and serine (S) residues, connected the two charged regions. The bipolar peptide gate was added to the cis compartment in concentrations of 0.3 to 0.5 μM . Upon applying a negative bias to the trans compartment, frequent and pronounced current blockades were observed at voltage biases in the -20 to -110 mV range; see Figure 3.3b. At a fixed bias voltage, the event rate was found to increase with the peptide concentration, whereas the dwell time remained constant (Figure S3.3), confirming our interpretation of current blockades as being produced by the interaction of individual peptide gates with the FraC nanopore.^(48,49) We suggest that the peptide traverses the pore with its positively charged end first, but after the negatively charged end enters the FraC vestibule, the two charged regions are pulled in opposite directions by the transmembrane bias. Together with the force exerted by the electro-osmotic flow (EOF), this creates a tug-of-war mechanism that stalls the peptide at the position where the forces on the probe balance each other. The exceptionally long translocation times (>1 s) observed at some voltages confirm this hypothesis.

The residence time of the bipolar gate in the FraC pore has a pronounced dependence on voltage (Figure 3.3c), spanning the range from a few milliseconds (1.88 ms at -110 mV) up to few seconds (2.7 s at -40 mV). At low applied potentials, with a magnitude below -40 mV, the dwell time is observed to increase with voltage, while at higher voltages, the dwell time decreases with voltage. This biphasic behavior of the dwell time has been previously reported by others and is typically associated with two different regimes of the analyte interaction with the nanopore.^(38,50–54) The first regime, where the dwell time increases with voltage, is generally attributed to events in which the molecule exits the pore through the cis opening. The second regime, where the dwell time decreases with voltage, is generally attributed to molecules that translocate the pore and exit through the trans side. We hypothesize that in our FraC system, at higher voltages (>40 mV by magnitude) the increased EOF exerts increasing forces to the peptide toward the trans side of the pore, thus facilitating translocations and decreasing its residence time in the pore.⁽⁵⁵⁾

The effect of the positive and negative charges of the peptide gate was further confirmed by carrying out nanopore translocation experiments using a modified version of the peptide

gate in which the polyanionic fragment of the peptide was removed. Upon the addition of the modified peptide to the cis compartment, we observed very fast translocations, with a translocation time of the order of hundreds of microseconds. At -30 mV, for example, the average residence time of the modified peptide was $208 \mu\text{s}$ compared to 2.12 s for the bipolar gate, a difference of 4 orders of magnitude. These results demonstrate that both charged ends are necessary to establish trapping of the peptide gate.

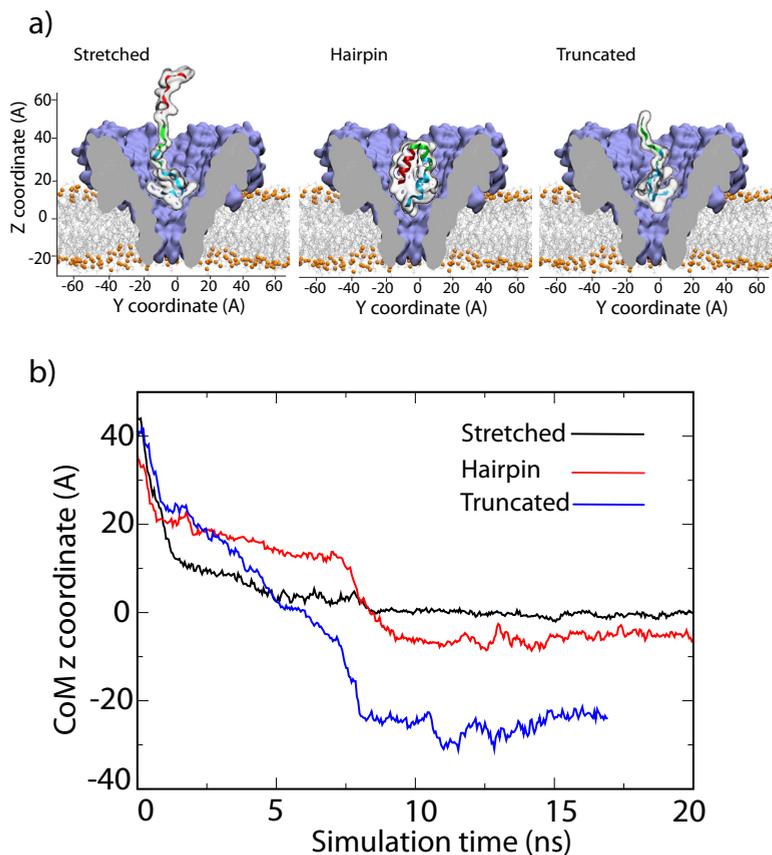


Figure 3.4. MD simulation of peptide translocation through the FraC nanopore. (a) Setup of MD simulations that examined the effect of the peptide sequence and conformation on electric field-driven translocation of model peptides. Molecular graphics images illustrate initial conformations of the peptides: two conformations of the bipolar peptide (stretched, left, and forming a hairpin, middle) and one conformation of the truncated peptide (right). In each simulation, one peptide construct was placed at the cis entrance of the FraC nanopore and simulated under a transmembrane bias of -1.2 V. (b) Center of mass coordinate of the “CGSGSGSKGS” central segment of the three peptides vs simulation time. The z coordinate is defined in panel a. After about 10 ns, the bipolar peptides remain trapped within the FraC nanopore regardless of their initial conformations, whereas the truncated peptide exits the FraC pore toward the trans side.

To obtain a microscopic interpretation of the peptide translocation experiment, we constructed three all-atom models of the FraC nanopore system differing by the structure and conformation of the peptide; see Figure 3.4a. In the first system, the bipolar peptide was placed in the FraC vestibule having its polycationic end proximal to the constriction. After initial energy minimization and 19 ns equilibration, the system was simulated under a -1.2 V bias, which produced rapid capture and transport of the polycationic tag through the FraC constriction. The translocation, however, halted after the polycationic tag passed through the constriction; see Figure 3.4b, which plots the z-coordinate of the peptide CFSFSFSKFS segment's center of mass (CoM) as a function of the simulation time. Figure 3.5a shows a representative conformation of the trapped peptide system. A qualitatively different outcome was observed when the simulation was repeated using a version of the peptide lacking the polyanionic part: the peptide completed the translocation in 10 ns leaving the FraC constriction; see Figure 3.4b. To investigate the robustness of the entrapment mechanism with regard to the initial conformation of the peptide, the peptide was simulated in 1 M NaCl solution for 50 ns in the absence of the FraC nanopore. The conformation of the peptide was observed to change from an initially extended, disordered structure into an alpha-helical hairpin, held together by electrostatic interactions between the polycationic and polyanionic parts. Next, we repeated the nanopore translocation simulation starting from the full length bipolar peptide in the hairpin conformation. Similar to the outcome of the previous simulation, the polycationic part of the bipolar peptide was seen to pass through the FraC constriction, unfolding the hairpin. Interestingly, the translocation halted when the peptide reached approximately the same location within the FraC nanopore as in the simulation that initiated starting from a stretched conformation (Figure 3.4b), which confirmed robustness of the trapping mechanism.

Next, we investigated the effect of the applied bias on the ionic current blockade. Figure 3.5b shows the scatter plot of relative blockade versus dwell time obtained experimentally for a range of voltages between -50 and -110 mV. The relative blockade is calculated as the current blockade (ΔI) divided by the open pore current (I). It is clear from the plot that the relative blockade decreases at increasing bias, indicating that the fraction of the pore that is blocked by the peptide gate changes depending on the applied bias. These results are confirmed by the current-voltage relation presented in Figure 3.5c: a superlinear behavior is observed for this I-V curve, indicating, a change in resistivity of the peptide/FraC system at different voltages.

Complementing these experiments, we probed the dependence of the nanopore ionic current blockade produced by the bipolar gate on the transmembrane voltage in MD simulations. The conformations of the trapped peptide obtained at the end of the nanopore capture and translocation simulations (Figure 3.4a,b) were used to initiate four independent

3

simulations, each lasting about 150 ns, at -250 , -100 , -50 , and 0 mV. The final state of the 0 V run was then used to run three independent simulations at -250 , -100 , and -50 mV to obtain at least two independent trajectories for each bias condition. Figure S3.4 illustrates the change in the location of the peptide's central fragment during these simulations. While the bipolar peptide remains trapped within the nanopore, water and ions can flow through the gate. Quantitative analysis of the MD trajectories confirms our earlier conjecture about the presence of the EOF (Figure 3.5d) and shows that the EOF magnitude indeed increases with the magnitude of the voltage bias, pushing the peptide, on average, toward the trans side. Figure 3.5e shows the simulated dependence of the blockade current on voltage: the current increases superlinearly with the voltage for both types of initial conditions, as in the experiments. The simulated relative blockade current is found to decrease with the bias, similar to the dependence seen experimentally. By plotting the peptide residue density inside the constriction region of FraC (defined by residues Ala5 and Leu23) as a function of voltage, shown in Figure 3.5f, we find a lower peptide density under higher voltage, a manifestation of the polar gate stretching in electric field. A set of simulations carried out at a fixed conformation of the gate peptide yielded a linear I-V dependence (Figure S3.5), validating our assertion that peptide stretching is responsible for the nonlinear behavior. Based on the results of simulation and experiment, we arrive at a model, where higher transmembrane bias stretches the removable gate, thereby occupying less volume inside the FraC constriction, which lets more ions to pass through, i.e., the conductance is electro-mechanically modulated.

The above results confirm the working mechanism of our electromechanical gate, where the fraction of the pore occupied by the gate can be modulated as the peptide is stretched to varying degrees at different applied voltages. However, the data presented in Figure 3.5b,c were obtained by averaging over hundreds of single-molecule experiments, where individual peptides were captured at a constant applied voltage. To demonstrate dynamic electromechanical modulation of the ionic current at the true single-molecule level, we trapped individual peptides, one at a time, in the FraC nanopore and subsequently modulated the conformation of the captured peptide by changing the magnitude of the applied voltage. Figure 3.6a shows a typical ionic current trace recorded from such measurements. First, we apply a low bias of -30 mV until a single peptide is captured by the FraC pore. Panel 1 represents the open pore current upon applying -30 mV, and panel 2 shows the moment when the peptide is captured. At this voltage, the relative blockade of the peptide is 0.77 ± 0.05 , indicating that a large portion of the pore current is obstructed. We subsequently apply steps of -50 , -70 , and -90 mV while the peptide remains trapped inside the nanopore. As expected, a current increase is observed after each voltage step, as can be observed in panels 3, 4, and 5. Interestingly, when the relative blockade is calculated for each of the voltage intervals, we find a relative blockade of 0.73 ± 0.04 for -50 mV, 0.61

± 0.02 for -70 mV, and 0.48 ± 0.02 for -90 mV. Thus, the relative blockade reduces with increasing bias for an individual molecule, indicating that the peptide is stretched further with each voltage step. As shown in the traces of Figure 3.6b, a trapped peptide can also be continuously stretched (top trace), or it can be first stretched and then compressed (bottom trace). Importantly, we observed well-defined, reproducible levels of blockade currents regardless of the direction of the voltage ramp.

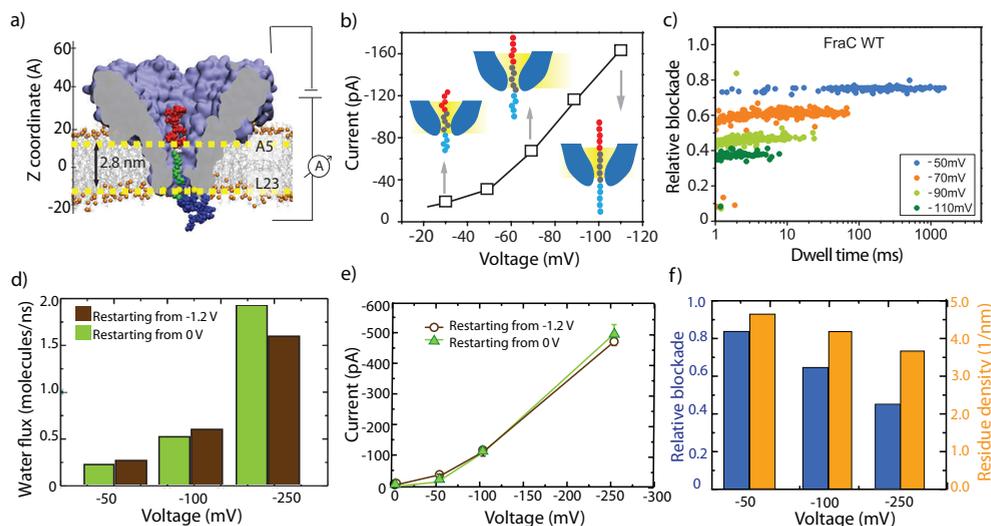


Figure 3.5. Effect of transmembrane bias on relative current blockade. (a) Representative microscopic configuration observed at the end of an MD simulation in which a bipolar peptide was captured and trapped by a transmembrane bias. The amino acids of the bipolar peptide are colored according to their charge: negative, positive, and neutral residues are shown in red, blue, and green, respectively. The constriction region of FraC is marked by horizontal dashed lines corresponding to the average z coordinate of the eight Ala5 and eight Leu23 residues. (b) Relative blockade current, I_b/I_0 , vs dwell time for bipolar peptides trapped within the FraC nanopore at several biases. (c) Average blockade current experimentally measured upon trapping of the bipolar peptides within the FraC nanopore as a function of transmembrane bias. Inset images illustrate the proposed voltage-induced stretching of the peptide trapped within the nanopore. (d) Average flux of water molecules through FraC nanopore blocked by the peptide gate. Data shown in different colors correspond to two independent simulations differing by the initial conformations of the bipolar peptides. (e) Average blockade currents measured from MD simulations of trapped bipolar peptides as a function of transmembrane bias. Data shown in black and green correspond to two independent simulations of the blockade currents differing by the initial conformations of the bipolar peptides. (f) The relation between peptide stretching and blockade current. Blue bars (left axis) show the simulated relative blockade, whereas orange bars (right axis) show the peptide density within the FraC constriction.

To verify the correspondence between current modulations observed at the single-molecule level and those measured previously from the ensemble measurements at constant bias conditions, we plot in Figure 3.6c the histograms of the relative blockade currents obtained for the two types of experiments. The peaks of the relative blockade current are seen to occur at the same values at the same voltage bias in both single-molecule and ensemble

measurements. Altogether, our results prove that our electromechanical gate can be used to modulate pore conductivity both dynamically and at a true single-molecule level.

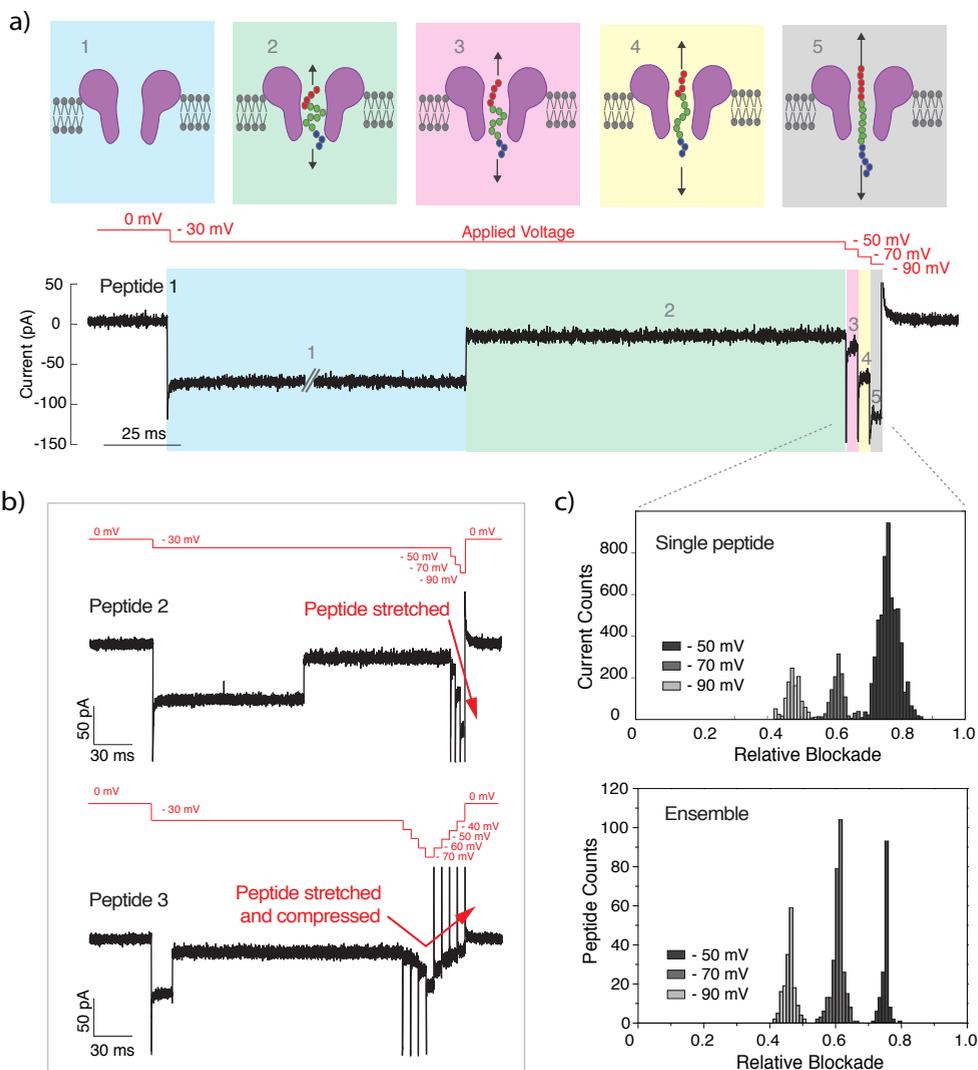


Figure 3.6. Electro-mechanical modulation of FraC conductance. (a) Schematic of the single-peptide capture and stretching. Level 1 represents the open pore current. Level 2 represents the peptide captured at low voltages (-30 mV). Levels 3, 4, and 5 represent peptide stretching at -50 , -70 , and -90 mV consecutively. (b) Current trace corresponding to the capture and stretching of a single peptide. The very first part of the recording shows capture of the FraC nanopore. (c) Current traces of individual peptides being stretched (top) or stretched and consequently compressed (bottom). (d) Example histogram of the current values corresponding to stretching a single peptide within FraC nanopore by -50 , -70 , and -90 mV voltages (top). Histogram of current blockade values obtained from multiple measurements of multiple peptides, each carried out at one of the following fixed value of the applied voltage: -50 , -70 , and -90 mV (bottom).

3.3 Conclusions

In this work, we combined ionic current measurements with MD simulations to demonstrate a mechanism for the modulation of a membrane channel conductance: electromechanical stretching of a removable gate. Our gate was a peptide containing a fragment of 10 negative amino acids at the N-terminus and 10 positive amino acids at the C-terminus. Subject to opposing electrophoretic forces pulling both ends of the peptide in different directions, the peptide is transiently stalled in the nanopore for a time interval that greatly exceeds typical peptide translocation time and can reach seconds. The magnitude of the ionic current flowing through the nanopore blocked by the removable gate was found to increase superlinearly with the magnitude of the applied bias. Our simulations determined such a peculiar conductance modulation to originate from differential stretching of the peptide under applied biases. The peptide can therefore act as a gate that modulates the pore conductance in a voltage-dependent manner. Additionally, true single-molecule experiments were performed, where an individual peptide was stalled within the pore and voltage steps were applied to gradually stretch and compress the peptide, yielding relative blockades that matched those obtained in ensemble measurements involving several hundreds of peptides.

This mechanism of channel conductance modulation could be used for several applications. For example, ion channel blockers or modulators are often used to control ion transport through channels for therapeutic purposes in diseases such as multiple sclerosis or epilepsy. A peptide gate could be modified with a recognition amino acid sequence to target a specific family of channels for therapeutic purposes. Additionally, this system could work as an interface between gene expression and ionic current, where the expression (and translation) of gate peptides affects membrane potential, which may be useful for creation of regulatory circuits in synthetic biology and for a realization of natural computing.⁽⁵⁶⁾ The properties of the gate can obviously be fine-tuned by changing the peptide sequence. Different peptide sequences would result in different elastic response and baseline current, potentially fulfilling any range. The initial hairpin conformation of the peptide can be exploited, such that a threshold bias is necessary for hairpin rupture. Finally, the long observation times offered by the dipolar peptide constructs may allow for extended examination of the central region of the peptide, which could be used for the analysis of amino acids and their post-translational modifications.

3.4 Methods

3.4.1 General MD Methods

All simulations were performed using the classical MD package NAMD,⁽⁵⁷⁾ periodic boundary conditions, and a 2 fs integration time step. The CHARMM36 force field⁽⁵⁸⁾ was used to describe proteins, dioctadecatrienoylphosphatidylcholine (DPhPC) phospholipids, TIP3P water, and ions along with the CUFIX corrections applied to improve the description of charge–charge interactions.⁽⁵⁹⁾ RATTLE⁽⁶⁰⁾ and SETTLE⁽⁶¹⁾ algorithms were applied to covalent bonds that involved hydrogen atoms in protein and water molecules, respectively. The particle mesh Ewald (PME)⁽⁶²⁾ algorithm was adopted to evaluate the long-range electrostatic interaction over a 1 Å-spaced grid. van der Waals interactions were evaluated using a smooth 10–12 Å cutoff. Langevin dynamics were used to maintain the temperature at 295 K. Multiple time stepping was used to calculate local interactions every time step and full electrostatics every three time steps. The Nose-Hoover Langevin piston pressure control⁽⁶³⁾ was used to maintain the pressure of the system at 1 atm by adjusting the system's dimension. Langevin thermostat⁽⁶⁴⁾ was applied to all the heavy atoms of the system with a damping coefficient of 0.1 ps⁻¹ to maintain the system temperature at 295 K.

3.4.2 MD Simulations of FraC Nanopores and Peptides

An all-atom model of the FraC protein was constructed starting from its crystallographic structure, Protein Data Bank entry 4TSY⁽³⁷⁾ taking into account the crystallographic symmetry of the structure. Atoms missing in the crystallographic structure were added using the psfgen tool of VMD.⁽⁶⁵⁾ The structure was then aligned to be coaxial with the z axis of our coordinate system. The protein was embedded in a pre-equilibrated 16 nm × 16 nm patch of dioctadecatrienoylphosphatidylcholine (DPhPC) bilayer. The lipid bilayer membrane was aligned with the x–y plane and shifted along the z axis to have the Trp112 residues of the protein located within the same plane as the head groups of the nearest lipid leaflet. Lipid and water molecules that overlapped with the protein were removed. One molar solution of NaCl was added on both sides of the membrane using the solvate and autoionize plugins of VMD, respectively, producing an electrically neutral system of 357 243 atoms. Following assembly, the system was minimized in 1200 steps using the conjugate gradient method and then equilibrated for 80 ns at a constant number of atoms, pressure, and temperature (NPT) ensemble performed while keeping the ratio of the system's size along the x and y axes constant. During the initial stage of equilibration, all alpha-carbon atoms of the protein were restrained to their initial coordinates using harmonic potentials; the spring constant (kSPRING) of the potentials was set to 1.0 kcal mol⁻¹ Å⁻² for the first 30 ns and then decreased to 0.8, 0.5, and 0.1 kcal mol⁻¹ Å⁻² in 5 ns steps, following which the system

was equilibrated in the absence of any restraints for 35 ns. All subsequent simulations of the FraC protein under the applied electric field were performed using the protein structure obtained at the end of last stage of the restrained ($k_{\text{SPRING}} = 0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) equilibration.

The simulations under a transmembrane bias were performed in a constant number of particle, volume, and temperature (NVT) ensemble, restraining the protein's alpha-carbon coordinates ($k_{\text{SPRING}} = 0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) to the crystallographic values. For the NVT simulations, the system's dimensions were set to the average dimensions observed within the last 5 ns of the restrained NPT equilibration. An external electric field, $E = -V/LZ$, was applied along the z axis (normal to the membrane) to produce a transmembrane bias V , where LZ is the dimension of the simulated system in the direction of the applied electric field.^(40,66)

The all-atom structure of the bipolar peptide was obtained by extracting a relatively straight 30-residue fragment from the FraC structure (reside 4 to 34) and mutating the amino acid sequence of the fragment to EEEEEEEECGSGSGSKGSRRRRRRRRRR using the psfgen tool. The peptide was solvated in a $7 \text{ nm} \times 7 \text{ nm} \times 6.5 \text{ nm}$ volume of 1 M NaCl (28 862 atoms) and equilibration for 50 ns in the NPT ensemble. A pair of microscopic conformations of the bipolar peptide (stretched and hairpin) were obtained from this MD trajectory by extracting the protein coordinates at 1.6 and 50 ns, respectively. The structure of the truncated peptide was obtained by truncating ten terminal glutamate residues from the stretched structure. Each peptide was placed at the cis entrance of the FraC protein, having the CoM of the terminal arginine residue located 15 \AA above (toward the cis side) of the CoM of the FraC's transmembrane part (defined as residues 4 to 29). Lipid and 1 M NaCl solution were added following the same protocols as above. Each system was minimized 1200 steps and then equilibrated under constant ratio NPT ensemble for 4.8 ns, having all alpha-carbon atoms of the protein and the peptide restrained to their initial coordinates ($k_{\text{SPRING}} = 0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). After equilibration, the systems were simulated under applied electric field in the NVT ensemble following the same protocol as described above.

3.4.3 Ion Current Calculation

Prior to ionic current calculations, frames from the MD trajectory were aligned using protein coordinates to correct for the drift in the x-y plane and lipid bilayer coordinates to correct for the drift along the z axis. The ionic current was calculated as:

$$I(t) = \frac{1}{\delta t l_z} \sum_{j=1}^N q_j \delta z_j(t)$$

where $\delta z_j(t)$ is the displacement of ion j along the z direction during the time interval $\delta t = 4.8$ ps and q_j is the charge of ion j . To minimize the effect of thermal noise, the current was calculated within an $l_z = 48$ Å thickness slab centered at the midplane of the lipid bilayer membrane (the slab spanned the entire simulation system in the x - y plane).

3.4.4 Calculation of Electrostatic Potential

To visualize the electrostatic potential in our systems, we averaged the instantaneous distributions of the electrostatic potential over the MD trajectory using a previously described method,⁽⁴⁰⁾ implemented in the PMEpot plugin of VMD. Each atom of the system was approximated by a spherical Gaussian:

$$\rho_i(\mathbf{r}) = q_i \left(\frac{\beta}{\sqrt{\pi}} \right)^3 e^{-\beta|\mathbf{r}-\mathbf{r}_i|^2}$$

where β was the Gaussians' width. The instantaneous distribution of the electrostatic potential corresponding to the instantaneous charge configuration was obtained by solving the Poisson equation:

$$\nabla^2 \phi(\mathbf{r}) = 4\pi \sum_i \rho_i(\mathbf{r})$$

To obtain the average distribution of the potential in a given MD simulation, instantaneous distributions of the potential were averaged over the entire MD trajectory. The three-dimensional (3D) electrostatic potential maps were obtained by averaging the last 48 ns fragments of MD trajectories; $\beta = 0.1$ Å⁻¹ was used for these calculations. One-dimensional profiles of the electrostatic potential through the nanopores were obtained by taking values from the 3D profiles along the z coordinate, which is also the nanopore axis in our coordinate system.

3.4.5 Calculations of Residue Density

The residue density of the peptides confined to the FraC constriction region was computed as the ratio of the number of peptide residues located within the FraC constriction to the height of the constriction, 2.8 nm. The constriction region was defined by the average z coordinate of residues Ala5 and Leu23. The number of peptide residues within the constriction was determined as the ratio of the number of peptide backbone atoms within the constriction to the number of backbone atoms in one peptide residue. The peptide density was averaged over the last 72 ns of the respective MD trajectory.

3.4.6 Peptide Design and Synthesis

The peptides used in this work were a peptide with sequence EEEEEEEECGSGSGSKGSR-RRRRRRRRR (high performance liquid chromatography, HPLC, purity of 95.8% and molecular weight, MW, of 3678.9 Da) and the truncated peptide with sequence SGSGCGSKGSR-RRRRRRRRR (HPLC purity of 99.1% and MW of 2387 Da). Peptides were synthesized by Biomatik Corporation (Cambridge, CA). The synthesis was performed using standard solid-phase methods and the peptides were further purified using reverse-phase HPLC and analyzed by mass spectrometry (Biomatik). Peptides were kept lyophilized or, when necessary, aliquoted to a final concentration of 10 mg/mL at -20°C .

3.4.7 FraC Expression and Purification

WT FraC was expressed and purified as described before.^(38,39) E. coli EXPRESS BL21(DE3) cells were transformed with the pT7-SC1 plasmid, containing the FraC gene with an N-terminus His6-tag. Transformed cells were moved into 200 ml fresh 2-YT media with 100 mg/L ampicillin. The cell culture was grown at 37°C with shaking at 220 rpm until it reached an optical density of 0.8 at 600 nm. A total of 0.5 mM IPTG was added to the culture to induced FraC expression, after which the growth was continued overnight at 25°C . Cells were harvested by centrifugation at 2000g for 30 min, and the pellets were stored at -80°C . The pellets (derived from 50 to 100 mL of bacterial culture) were thawed and resuspended in lysis buffer containing 15 mM Tris base at pH 7.5, 1 mM MgCl_2 , 4 M urea, 0.2 mg/mL lysozyme, and 0.05 units per milliliter of DNase. The culture was sonicated to fully disrupt the cells, and the crude lysate was then centrifuged at 5400g for 20 min at 4°C . The supernatant solution was mixed with 100 μL of NiNTA slurry (Qiagen) at room temperature for 1 h with gentle mixing. The mixture was spun down at 2000g for 5 min at 4°C . The pellet containing the resin and the protein that was bound were transferred to a spin column (BioRad). The beads were washed once and eluted with 300 mM imidazole. Protein concentration was estimated using NanoDrop. The monomers were stored at 4°C until oligomerized.

Sphingomyelin and DPhPC (Avanti Polar Lipids) were mixed in a 1:1 ratio and dissolved in 4 mL of pentane (Sigma-Aldrich) with 0.5% ethanol. The mixture was placed in a rounded flask and rotated slowly to evaporate the solvent and allow the lipid film to deposit in the walls of the flask. The lipid film was resuspended using a sonicator bath in a buffer containing 150 mM NaCl and 15 mM Tris-HCl (pH 7.5) to a final lipid concentration of 10 mg/mL. The liposomes were stored at -20°C .

Monomeric FraC was mixed with the liposomes in a lipid-to-protein ratio of 10:1. The

mixture was briefly sonicated and incubated for 30 min at 37 °C. The proteo-liposomes were solubilized with 0.6% LDAO and then diluted 20 times with buffer containing DDM (150 mM NaCl, 15 mM Tris base, pH 7.5, 0.02% DDM). A second round of purification was performed using Ni-NTA beads. Ni-NTA slurry was incubated with the protein/lipid mixture for 1 h with gentle shaking. Afterward, the mixture was loaded into a spin column, washed, and eluted using 200 mM EDTA, 75 mM NaCl, 7.5 mM Tris base, pH 8, 0.02% DDM. Oligomers were kept at 4 °C for several months.

3.4.8 Electrical Recording in Planar Lipid Membranes

Electrical recording was performed using planar lipid membranes (BLMs) as has been described before.^(67,68) Briefly, a 25 μm thick Teflon film (Goodfellow Corporation) containing an orifice of approximately 70 μm separates the cis and trans compartments. To form the membranes, 10 μL of 5% hexadecane in pentane is added to the Teflon film, and the pentane is allowed to evaporate. The reservoirs are filled with buffer and 10 μL of 10 mg/mL DPhPC in pentane. Membranes were spontaneously formed using the Montal–Mueller method. Ag/AgCl electrodes are placed in each compartment, with the ground electrode in the cis side. WT FraC oligomers are added to the cis side of the chamber. Upon pore insertion, the pore is characterized by measuring traces at different voltages and taking an I–V curve. For the single-channel conductance measurements, nanopores were measured at 0, –50, and 50 mV. The substrate was added to the cis side of the chamber and measured at multiple voltages.

3.4.9 Single-Peptide Stretching and Compressing Experiments

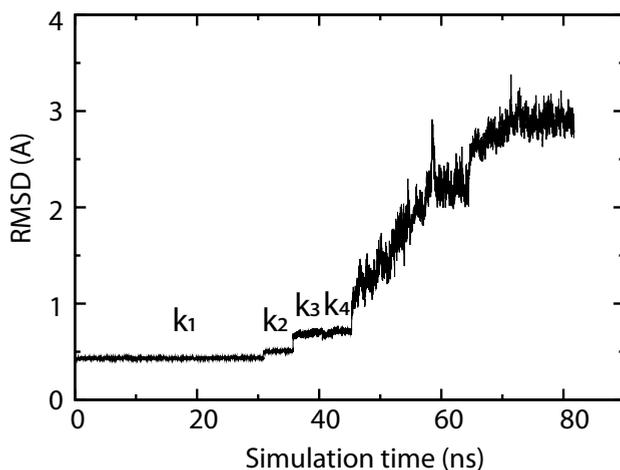
For the single-peptide stretching experiments, a protocol that generates steps at different voltages was created using the pCLAMP software from molecular devices. A pair of main modalities were recorded. The first one is a stretching protocol in which the voltage starts at 0 mV, decreases to –30 mV for 500 ms for peptide capture, and sequentially decreases to –50, –70, and –90 mV for 5 ms at each voltage. The second protocol was a stretching and compressing protocol in which the voltage starts at 0 mV, decreases to –30 mV for 500 ms, and then is sequentially decreased down to –70 mV and increased again to –30 mV in steps of 10 mV for 5 ms in each step. For the experiments, peptide was added to the cis side of the chamber, and both voltage protocols were applied and recorded.

3.4.10 Data Acquisition and Analysis

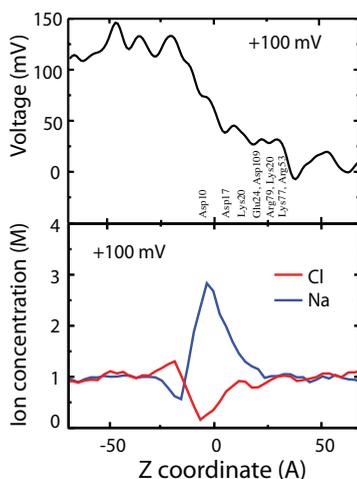
Nanopore recordings were collected using a patch-clamp amplifier (Axopatch 200B, Molecular Devices) at a filtering frequency of 100 kHz. The data were digitized using a

Digidata 1550B (Molecular Devices) at a sampling frequency of 500 kHz. The signal was low-pass filtered at 5 kHz and processed using the Clampfit software, a Matlab script, and the software package Transalyzer.(69)

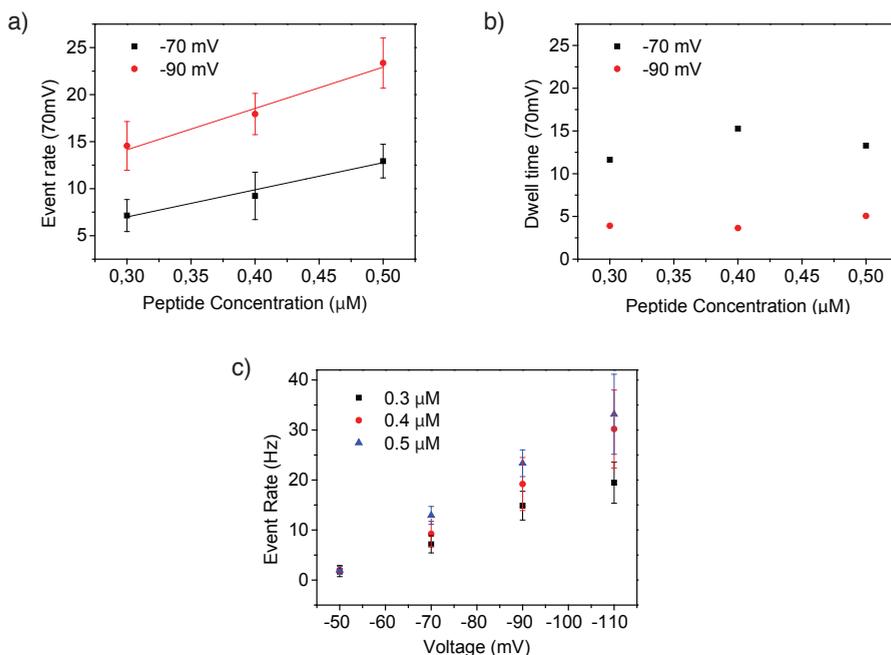
3.5 Supplementary information



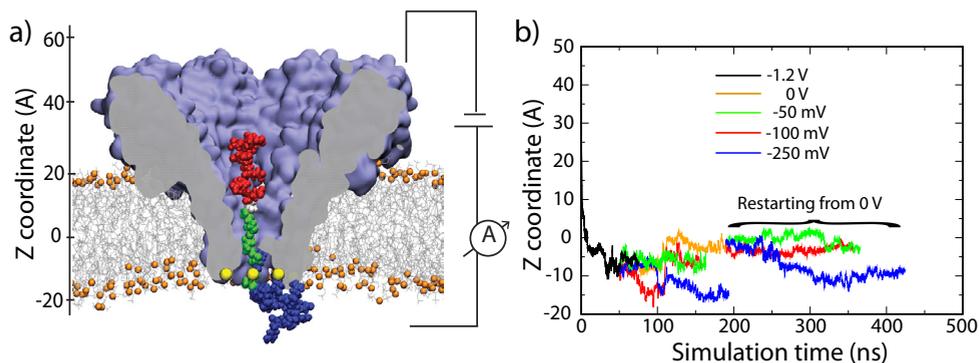
Supplementary Figure S3.1: Stability of a FraC nanopore in all-atom MD simulation. The root mean squared deviation (RMSD) of FraC's alpha carbon atoms from their crystallographic coordinates over the course of the equilibration trajectory. For the first 45 ns, the alpha-carbon atoms were restrained to their crystallographic coordinates using harmonic springs of the following magnitude: 1 (k1), 0.8 (k2), 0.5 (k3), and 0.1 (k4) kcal/(mol Å²). The last 35 ns were performed in the absence of any restraints. RMSD values of less than 4 Å generally signify a stable structure. To reduce uncertainty of ionic current determination associated with fluctuations of the protein structure, all subsequent simulations were performed using the structure obtained at the end of the last restrained simulation (k4), maintaining the same restraints.



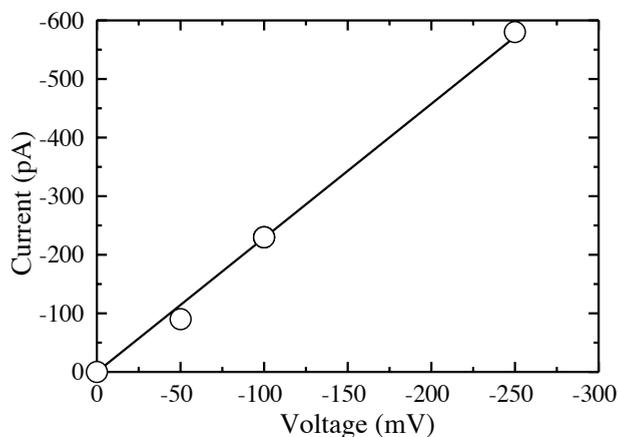
Supplementary Figure S3.2: Simulated profiles of the electric potential (top) and of the Na⁺ and Cl⁻ concentrations (bottom) along the central axis of the FraC nanopore at a +100 mV bias. The data were obtained by averaging instantaneous configurations over a 48 ns MD trajectory. Approximate locations of the charged residues of the FraC nanopore are indicated at the middle horizontal axis.



Supplementary Figure S3.3: Characterization of peptide trapping experiments. (a) Event rate versus peptide concentration. The event rate increases linearly with peptide concentration indicating a simple bimolecular interaction between the peptide and the pore. (b) Dwell time versus peptide concentration. Dwell time remains constant at different peptide concentrations. (c) Event rate versus voltage magnitude. The event rate increases linearly with applied voltage, as expected.



Supplementary Figure S3.4: MD simulation of peptide gate stretching by transmembrane voltage. (a) Representative configuration of the simulation system featuring a bipolar peptide that is lodged in the constriction of FraC. The amino acids of the bipolar peptide are colored according to their charge: negative, positive and neutral residues are shown in red, blue and green, respectively. (b) Center of mass coordinate of the “CGSGSGSKGS” segment (central part) of peptide versus simulation time under different biases condition. The z coordinate is defined in panel a. Using the conformation of the peptide observed in the second half of the peptide capture simulations under a -1.2V bias (Fig. 3.4 of the main text shows the first 20 ns of the 80 ns trajectory), four simulations were performed under voltage bias of -250 (blue), -100 (red), -50 (green), and 0 (orange) mV for approximately 120 ns each. The final state of the 0 V trajectory was used to initiate three additional simulations under -250 (blue), -100 (red), and -50 (green) mV, each lasting approximately 150 ns. These two sets of simulations were used to determine the dependence of the water flux and ionic current on the degree of peptide stretching, main text Figure 3.5d-e.



Supplementary Figure S3.5: Simulated current-voltage dependence of the FraC nanopore containing a fixed-conformation gate threaded through the FraC constriction. The gate conformation was taken from the last frame of the gate peptide capture simulation under -1.2 V. Each data point was obtained from a 50 ns MD trajectory.

References

1. Boyle, J.; Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P., *Molecular Biology of The Cell*, 5th edition Biochem. Mol. Biol. E d u c . 2008, 36, 317-318.
2. Hille, B., *Ion Channels of Excitable Membranes*. Sinauer Sunderland, MA: 2001; Vol. 507.
3. Camerino, D. C.; Tricarico, D.; Desaphy, J.-F., *Ion Channel Pharmacology. Neurotherapeutics* 2007, 4, 184-198.
4. Beckstein, O. *Principles of Gating Mechanisms of Ion Channels*. University of Oxford, 2004.
5. Schmidt, D.; MacKinnon, R., Voltage-Dependent K Channel Gating and Voltage Sensor Toxin Sensitivity Depend on the Mechanical State of the Lipid Membrane. *Proc. Natl. Acad. Sci.* 2008, 105, 19276-19281.
6. West, J. W.; Patton, D. E.; Scheuer, T.; Wang, Y.; Goldin, A. L.; Catterall, W. A., A Cluster of Hydrophobic Amino Acid Residues Required for Fast Na(+)- Channel Inactivation. *Proc. Natl. Acad. Sci.* 1992, 89, 10910-10914.
7. Khalili-Araghi, F.; Gumbart, J.; Wen, P.-C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K., Molecular Dynamics Simulations of Membrane Channels and Transporters. *Curr. Opin. Struct. Biol.* 2009, 19, 128-137.
8. Chang, G.; Spencer, R. H.; Lee, A. T.; Barclay, M. T.; Rees, D. C., Structure of the MscL Homolog from Mycobacterium Tuberculosis: A Gated Mechanosensitive Ion Channel. *Science* 1998, 282, 2220.
9. Zhu, F.; Hummer, G., Drying Transition in the Hydrophobic Gate of the GLIC Channel Blocks Ion Conduction. *Biophys. J.* 2012, 103, 219-227.
10. Watson, H., *Biological Membranes. Essays in biochemistry* 2015, 59, 43-69.
11. Siwy, Z. S.; Howorka, S., Engineered Voltage-Responsive Nanopores. *Chem. Soc. Rev.* 2010, 39, 1115-1132.
12. Siwy, Z.; Fuliski, A., Fabrication of a Synthetic Nanopore Ion Pump. *Phys. Rev. Lett.* 2002, 89, 198103.
13. Zhu, F.; Schulten, K., Water and Proton Conduction through Carbon Nanotubes as Models for Biological Channels. *Biophys. J.* 2003, 85, 236-244.
14. Mamad-Hemouch, H.; Ramoul, H.; Abou Taha, M.; Bacri, L.; Huin, C. c.; Przybylski, C. d.; Oukhaled, A.; Thiébot, B. n. d.; Patriarche, G.; Jarroux, N., Biomimetic Nanotubes Based on Cyclodextrins for Ion-Channel Applications. *Nano Lett.* 2015, 15, 7748-7754.
15. Ying, Y.-L.; Zhang, J.; Meng, F.-N.; Cao, C.; Yao, X.; Willner, I.; Tian, H.; Long, Y.-T., A Stimuli-Responsive Nanopore Based on a Photoresponsive Host-Guest System. *Sci. Rep.* 2013, 3, 1662.

16. Gao, R.; Ying, Y.-L.; Hu, Y.-X.; Li, Y.-J.; Long, Y.-T., Wireless Bipolar Nanopore Electrode for Single Small Molecule Detection. *Anal. Chem.* 2017, 89, 7382-7387.
17. Decker, K.; Page, M.; Aksimentiev, A., Nanoscale Ion Pump Derived from a Biological Water Channel. *J. Phys. Chem. B.* 2017, 121, 7899-7906.
18. Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M. J.; Golovchenko, J. A., Ion-Beam Sculpting at Nanometre Length Scales. *Nature* 2001, 412, 166.
19. Apel, P. Y.; Korchev, Y. E.; Siwy, Z.; Spohr, R.; Yoshida, M., Diode-Like Single-Ion Track Membrane Prepared by Electro-Stopping. *Nucl. Instrum. Methods Phys. Res.* 2001, 184, 337-346.
20. Yeh, L.-H.; Zhang, M.; Qian, S., Ion Transport in a pH-Regulated Nanopore. *Anal. Chem.* 2013, 85, 7527-7534.
21. Ma, T.; Walko, M.; Lepoitevin, M.; Janot, J. M.; Balanzat, E.; Kocer, A.; Balme, S., Combining Light Gated and pH Responsive Nanopore Based on PEG Spiropyran Functionalization. *Adv. Mater. Interfaces* 2018, 5, 1701051.
22. Powell, M. R.; Sullivan, M.; Vlassioux, I.; Constantin, D.; Sudre, O.; Martens, C. C.; Eisenberg, R. S.; Siwy, Z. S., Nanoprecipitation-Assisted Ion Current Oscillations *Nat. Nanotechnol.* 2008, 3, 51.
23. Vlassioux, I.; Siwy, Z. S., Nanofluidic Diode. *Nano Lett.* 2007, 7, 552-556.
24. He, Z.; Zhou, J.; Lu, X.; Corry, B., Bioinspired Graphene Nanopores with Voltage-Tunable Ion Selectivity for Na⁺ and K⁺. *ACS Nano* 2013, 7, 10148-10157.
25. Zwolak, M.; Lagerqvist, J.; Di Ventra, M., Quantized Ionic Conductance in Nanopores. *Phys. Rev. Lett.* 2009, 103, 128102.
26. Feng, J.; Liu, K.; Graf, M.; Dumcenco, D.; Kis, A.; Di Ventra, M.; Radenovic, A., Observation of Ionic Coulomb Blockade in Nanopores. *Nat. Mater.* 2016, 15, 850.
27. Wilson, J.; Aksimentiev, A., Water-Compression Gating of Nanopore Transport. *Phys. Rev. Lett.* 2018, 120, 268101.
28. Gracheva, M. E.; Vidal, J.; Leburton, J.-P., p-n Semiconductor Membrane for Electrically Tunable Ion Current Rectification and Filtering. *Nano Lett.* 2007, 7, 1717-1722.
29. Shankla, M.; Aksimentiev, A., Modulation of Molecular Flux Using a Graphene Nanopore Capacitor. *J. Phys. Chem. B.* 2017, 121, 3724-3733.
30. Fert, A.; Cros, V.; Sampaio, J., Skyrmions on the Track. *Nat. Nanotechnol.* 2013, 8, 152.
31. Plesa, C.; Ananth, A. N.; Linko, V.; Gülcher, C.; Katan, A. J.; Dietz, H.; Dekker, C., Ionic Permeability and Mechanical Properties of DNA Origami Nanoplates on Solid-State Nanopores. *ACS Nano* 2013, 8, 35-43.
32. Göpflich, K.; Zettl, T.; Meijering, A. E. C.; Hernández-Ainsa, S.; Kocabey, S.; Liedl, T.; Keyser, U. F., DNA-Tile Structures Induce Ionic Currents Through Lipid Membranes. *Nano Lett.* 2015, 15, 3134-3138.
33. Howorka, S.; Bayley, H., Probing Distance and Electrical Potential Within a Protein

- Pore with Tethered DNA. *Biophys. J.* 2002, 83, 3202-3210.
34. Movileanu, L.; Howorka, S.; Braha, O.; Bayley, H., Detecting Protein Analytes that Modulate Transmembrane Movement of a Polymer Chain Within a Single Protein Pore. *Nat. Biotechnol.* 2000, 18, 1091.
 35. Asandei, A.; Chinappi, M.; Kang, H.-K.; Seo, C. H.; Mereuta, L.; Park, Y.; Luchian, T., Acidity-Mediated, Electrostatic Tuning of Asymmetrically Charged Peptides Interactions with Protein Nanopores. *ACS Appl. Mater. Interfaces* 2015, 7, 16706-16714.
 36. Maglia, G.; Heron, A. J.; Hwang, W. L.; Holden, M. A.; Mikhailova, E.; Li, Q.; Chelley, S.; Bayley, H., Droplet Networks with Incorporated Protein Diodes Show Collective Properties. *Nat. Nanotechnol.* 2009, 4, 437.
 37. Tanaka, K.; Caaveiro, J. M.; Morante, K.; González-Mañas, J. M.; Tsumoto, K., Structural Basis for Self-Assembly of a Cytolytic Pore Lined by Protein and Lipid. *Nat. Commun.* 2015, 6, 6337.
 38. Huang, G.; Willems, K.; Soskine, M.; Wloka, C.; Maglia, G., Electro-Osmotic Capture and Ionic Discrimination of Peptide and Protein Biomarkers with FraC Nanopores. *Nat. Commun.* 2017, 8, 935.
 39. Wloka, C.; Mutter, N. L.; Soskine, M.; Maglia, G., Alpha-Helical Fragaceatoxin C Nanopore Engineered for Double-Stranded and Single-Stranded Nucleic Acid Analysis. *Angew. Chem.* 2016, 55, 12494-12498.
 40. Aksimentiev, A.; Schulten, K., Imaging alpha-Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability, and the Electrostatic Potential Map. *Biophys. J.* 2005, 88, 3745-3761.
 41. Bhattacharya, S.; Yoo, J.; Aksimentiev, A., Water Mediates Recognition of DNA Sequence via Ionic Current Blockade in a Biological Nanopore. *ACS Nano* 2016, 10, 4644-4651.
 42. Wei, C.; Bard, A. J.; Feldberg, S. W., Current Rectification at Quartz Nanopipet Electrodes. *Anal. Chem.* 1997, 69, 4627-4633.
 43. Siwy, Z., Ion Current Rectification in Nanopores and Nanotubes with Broken Symmetry. *Adv. Funct. Mater.* 2006, 16, 735-746.
 44. Siwy, Z.; Heins, E.; Harrell, C. C.; Kohli, P.; Martin, C. R., Conical-Nanotube Ion-Current Rectifiers: the Role of Surface Charge. *J. Am. Chem. Soc.* 2004, 126, 10850-10851.
 45. Yameen, B.; Ali, M.; Neumann, R.; Ensinger, W.; Knoll, W.; Azzaroni, O., Single Conical Nanopores Displaying pH-Tunable Rectifying Characteristics. Manipulating Ionic Transport with Zwitterionic Polymer Brushes. *J. Am. Chem. Soc.* 2009, 131, 2070-2071.
 46. Yan, Y.; Wang, L.; Xue, J.; Chang, H.-C., Ion Current Rectification Inversion in Conic Nanopores: Nonequilibrium Ion Transport Biased by Ion Selectivity and Spatial Asymmetry. *J. Chem. Phys.* 2013, 138, 044706.

47. Momotenko, D.; Cortés-Salazar, F.; Josserand, J.; Liu, S.; Shao, Y.; Girault, H. H., Ion Current Rectification and Rectification Inversion in Conical Nanopores: a Perm-Selective View. *Phys. Chem. Chem. Phys.* 2011, 13, 5430-5440.
48. Hughes, A. B., *Amino Acids, Peptides and Proteins in Organic Chemistry: Analysis and Function of Amino Acids and Peptides*. John Wiley & Sons: 2013; Vol. 5.
49. Pollard, T. D., *A Guide to Simple and Informative Binding Assays*. *Mol. Biol. Cell* 2010, 21, 4061-4067.
50. Soskine, M.; Biesemans, A.; Maglia, G., Single-Molecule Analyte Recognition with ClyA Nanopores Equipped with Internal Protein Adaptors. *J. Am. Chem. Soc.* 2015, 137, 5793-5797.
51. Movileanu, L.; Schmittschmitt, J. P.; Scholtz, J. M.; Bayley, H., Interactions of Peptides with a Protein Pore. *Biophys. J.* 2005, 89, 1030-1045.
52. Alibakhshi, M. A.; Halman, J. R.; Wilson, J.; Aksimentiev, A.; Afonin, K. A.; Wanunu, M., Picomolar Fingerprinting of Nucleic Acid Nanoparticles Using Solid-State Nanopores. *ACS Nano* 2017, 11, 9701-9710.
53. Biesemans, A.; Soskine, M.; Maglia, G., A Protein Rotaxane Controls the Translocation of Proteins Across a ClyA Nanopore. *Nano Lett.* 2015, 15, 6076-6081.
54. Asandei, A.; Chinappi, M.; Lee, J.-k.; Ho Seo, C.; Mereuta, L.; Park, Y.; Luchian, T., Placement of Oppositely Charged Amino Acids at a Polypeptide Termini Determines the Voltage-Controlled Braking of Polymer Transport Through Nanometer-Scale Pores. *Sci. Rep.* 2015, 5, 10419.
55. Boukhet, M.; Pigué, F.; Ouldali, H.; Pastoriza-Gallego, M.; Pelta, J.; Oukhaled, A., Probing Driving Forces in Aerolysin and alpha-Hemolysin Biological Nanopores: Electrophoresis Versus Electroosmosis. *Nanoscale* 2016, 8, 18352-18359.
56. Nunes de Castro, L., *Fundamentals of Natural Computing: An Overview*. *Phys. Life Rev.* 2007, 4, 1-36.
57. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 2005, 26, 1781-1802.
58. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* 2010, 31, 671-690.
59. Yoo, J.; Aksimentiev, A., New Tricks for Old Dogs: Improving the Accuracy of Biomolecular Force Fields by Pair-Specific Corrections to Non-Bonded Interactions. *Phys. Chem. Chem. Phys.* 2018, 20, 8432-8449.
60. Andersen, H. C., RATTLE: A 'Velocity' Version of the SHAKE Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* 1983, 52, 24-34.
61. Miyamoto, S.; Kollman, P. A., Settle: An Analytical Version of the SHAKE and RATTLE

- Algorithm for Rigid Water Models. *J. Comput. Chem.* 1992, 13, 952-962.
62. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 1993, 98, 10089-10092.
 63. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* 1994, 101, 4177-4189.
 64. Brünger, A. T., X-PLOR: Version 3.1: a System for X-Ray Crystallography and NMR. Yale University Press: 1992.
 65. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 1996, 14, 33-38.
 66. Gumbart, J.; Khalili-Araghi, F.; Sotomayor, M.; Roux, B., Constant Electric Field Simulations of the Membrane Potential Illustrated with Simple Systems. *Biochim. Biophys. Acta.* 2012, 1818, 294-302.
 67. Maglia, G.; Heron, A. J.; Stoddart, D.; Japrun, D.; Bayley, H., Analysis of Single Nucleic Acid Molecules with Protein Nanopores. In *Methods Enzymol.*, Elsevier: 2010; Vol. 475, pp 591-623.
 68. Gutschmann, T.; Heimburg, T.; Keyser, U.; Mahendran, K. R.; Winterhalter, M., Protein Reconstitution Into Freestanding Planar Lipid Membranes for Electrophysiological Characterization. *Nat. Protoc.* 2015, 10, 188-198.
 69. Calin, P.; Cees, D., Data Analysis Methods for Solid-State Nanopores. *Nanotechnology* 2015, 26, 084003.

4

Resolving modifications in single amino acids using the FraC nanopore

Unlike DNA sequencing, which is currently inexpensive and fast to run, protein sequencing remains a tremendous challenge. Nanopores have emerged as attractive tools for the development of a protein sequencer with single-molecule detection capabilities. The identification of 20 different amino acids, however, presents a daunting technical challenge. Simplified fingerprinting schemes, in which only a subset of amino acids is labelled and detected, is a more feasible approach. Nanopore protein fingerprinting requires the modification of amino acids with chemical structures that generate a distinct signal in the ionic current. Here we explore the feasibility of resolving chemical tags located on a specific amino acid within a protein chain. We use a model peptide and a Fragaceatoxin C (FraC) nanopore to characterize different chemical modifications. We show that sensitive and reproducible detection can be obtained for labels with a spectrum of different physicochemical properties such as mass, geometry, charges, and hydrophobicity. We further demonstrate that information about the position of the label along the peptide chain can be extracted from individual current-blockade event features. Our findings provide insights into the effects of different chemical labels on nanopore ionic current signals, which is important for protein sensing and fingerprinting.

This chapter is in preparation for publication as: Laura Restrepo-Pérez, Gang Huang, Peggy Bohlaender, Nathalie Worp, Rienk Eelkema, Giovanni Maglia, Chirlmin Joo *, Cees Dekker*. Resolving modifications in single amino acids using a biological nanopore.

4.1 Introduction

DNA, RNA, proteins, and metabolites form a complex network of interactions that determines the phenotype of cells^{1,2}. To obtain a broad understanding of a biological system, we need a comprehensive approach that integrates genomics, transcriptomics, proteomics, and metabolomics³⁻⁷. Recent technological developments have mostly focused on the study of genomes, making DNA sequencing fast, cheap, and ubiquitous^{8,9}. The study of other -omics, especially proteomics, however, still remains costly and time-consuming¹⁰⁻¹².

4 One of the main challenges in proteomics is the lack of sensitive techniques that allow for detection of proteins present in low abundance. As, unlike for DNA, there is no biochemical method to amplify proteins present in a sample^{13,14}. Single-molecule techniques offer ultimate sensitivity and hold great promise for single-cell protein analysis¹⁵. Nanopores, in particular, have demonstrated to be ultrasensitive biosensors¹⁶, capable of successfully sequencing biopolymers such as DNA^{17,18}.

In a nanopore sensor, an insulating membrane made of a lipid bilayer or a solid-state membrane separates two compartments filled with an electrolyte. A nanometer-sized pore is made within the membrane, by inserting a single protein-pore into a lipid bilayer or drilling a pore in a solid-state membrane using TEM. When a voltage is applied across the membrane, an ionic current flows through the nanometer-sized aperture. Molecules passing or translocating through the pore modulate the ionic current, which provides the basic sensor signal.

For DNA translocation, fine changes within the current blockade signal were found to correlate to the sequence of the DNA passing through the pore^{19,20} – opening up the way to DNA sequencing at the single-molecule level. In principle, the same mechanism can be applied for protein sequencing, but there are multiple challenges. While DNA is comprised of only 4 different bases, proteins contain 20 different amino acids, presenting protein sequencing with a significant technical hurdle. In recent years, the groups of Joo and Marcotte proposed an alternative simpler idea, namely, protein fingerprinting, in which proteins can be identified if a subset of amino acids is labelled and read^{21,22}. Joo and colleagues proposed protein identification through detection of the sequence of cysteines and lysines along the peptide chain. The lysine and cysteine sequence is then compared to a protein database for protein identification²¹. Nanopores offer an attractive technique for the implementation of protein fingerprinting, thanks to their high sensitivity and time resolution. For such a nanopore protein fingerprinting method, cysteines, lysines, or other amino acids should be modified with labels that can produce a distinct modulation in the current while the linearized protein traverses the nanopore. The identification of such chemical modifica-

tions has not been addressed before using nanopores.

Here, we study the detection of six potential sequencing tags on a model peptide, as measured with the FraC nanopore^{23,24}. We attach different chemical groups to a single cysteine in the central part of the peptide and measure their effect on the nanopore signals. We show that labels as small as 427 Da can be clearly and reproducibly detected. Information about the position of the label can be extracted from the individual event characteristics. The relative blockade current correlates with label properties such as geometry and molecular weight, and the translocation time is found to be proportional to the measured tag hydrophobicity. The results represent an important advance towards the development of a single-molecule protein-fingerprinting device with nanopores. Besides protein fingerprinting, the identification of amino acid modifications has other important applications as, for example, post-translational modifications (PTMs) could also be detected.

4.2 Results

4.2.1 Distinguishing labelled and unlabelled peptides using a FraC nanopore

Nanopore protein fingerprinting requires the electrical detection of chemical modifications on specific amino acids. We sought to characterize the blockade levels produced by different chemical labels in a well-defined manner. In order to avoid bandwidth-related distortions of the signal, we designed a method to obtain long translocation times. We used a 30 amino acid long peptide, containing 10 glutamates at the N-terminus, and 10 arginines at the C-terminus, which at neutral pH features a strongly negatively charged N-terminus and a strongly positively charged C-terminus. Upon applying a negative bias to the trans side of the nanopore set-up (Figure 4.1), the peptide is dragged into the nanopore with its positive end entering first. When the negative region subsequently enters the pore, the electrophoretic force pulls the negatively charged end of the peptide in the opposite direction, thus stretching the peptide and stalling the molecule at the point where the forces in both directions equilibrate. The “tug-of-war” created thus allows for long observation times where the centre part of the peptide is probed in the pore constriction²⁵. A similar peptide was previously used by Asandei et al. to study the effect of pH in peptide-nanopore interactions²⁶.

Figure 4.1 illustrates the typical nanopore experiments. All our experiments were performed using buffer containing 1M NaCl, 10mM TRIS, and 1mM EDTA at pH 7.5. We use a wild-type FraC nanopore, and the peptide described above as our substrate. Our model peptide is added to the cis compartment at concentrations between 0.1 μ M and 0.5 μ M. Negative voltages are applied to the trans compartment for all our measurements to avoid gating that

is observed in FraC under positive bias. The measurements presented here are performed under a voltage bias of -90mV, unless stated otherwise.

Well-defined translocation events are consistently observed when the peptide is present in the cis compartment (Figure 4.1a and Figure S4.1). Notably, the relative current blockade is well reproducible from pore to pore with an average relative blockade of 0.47 ± 0.03 pA at -90mV, as measured in three independent experiments. The relative blockade is defined as the ratio between the current blockade ($\Delta I = I_{\text{OpenPore}} - I_{\text{Blockade}}$) and the open pore current (I_{OpenPore}). Long translocation times of 4.2 ± 0.6 ms are observed at -90mV. The translocation time is observed to decrease with increasing bias (Figure S4.1), indicating that the peptide exits the pore to the trans side of the chamber^{27,28}. A detailed characterization of the translocation behaviour of this model peptide through a FraC nanopore can be found elsewhere²⁵.

To probe the effect of an added chemical label to the peptide, we first used maleimide chemistry to label the cysteine in position C11, i.e. near the N-terminus, with a fluorescein dye, (Figure 4.1b). Fluorescein maleimide is a small molecule of molecular weight 427 Da. Our labelled peptides were HPLC-purified and the labelling verified using mass spectrometry (MALDI-TOF) (Figure S4.2). We find that our samples are nearly 100% labelled. We performed measurements of the unlabelled and the fluorescein-labelled peptide as shown in Figure 4.1.

We find that the current levels from the unlabelled and labelled peptides are clearly separated (Figure 4.1a, 4.1b). The unlabelled peptide produced a relative blockade of 0.47 ± 0.01 while the labelled peptide produced a relative blockade of 0.57 ± 0.01 . The values presented here correspond to the mean and standard deviation derived from a Gaussian fit of the relative blockade histograms. The increase in the relative blockade upon labelling is expected due to the additional volume provided by the presence of the fluorescein tag. Additionally, control experiments, using a mixture of labelled and unlabelled peptides were performed, confirming the presence of two populations with blockade levels of 0.46 ± 0.01 and 0.56 ± 0.01 , very close to the values obtained from the independent measurements (Figure 4.1c). We conclude that the fluorescein-labelled peptides can be readily distinguished from their unlabelled counterpart. The difference in the blockade levels is so clear and reproducible that, while nanopore data typically rely on data comparison in stochastic scatter plots of hundreds of events, we can distinguish labelled from unlabelled molecules one by one from their current levels in individual events (Figure 4.1a-b).

4.2.2 Identifying the most sensitive region of the model peptide in the FraC pore

We observed events with a sizeable (>1 ms) translocation time. This indicates that the pep

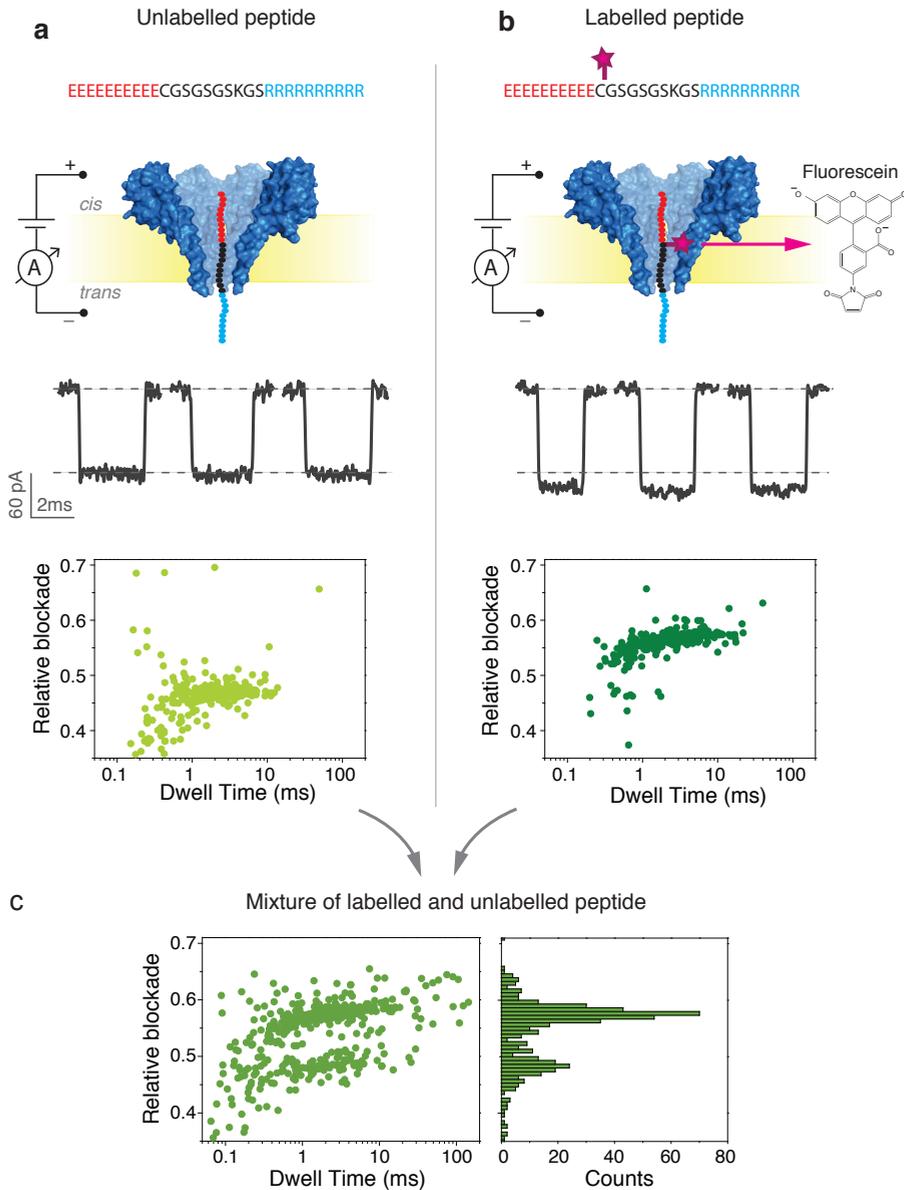


Figure 4.1: Analysis of labelled and unlabelled peptides with FraC. From top to bottom: Schematic representation, typical events, and scatter plot of relative blockade vs. dwell time of the unlabelled model peptide (a), and for the fluorescein-labelled model peptide (b) through a FraC nanopore. (c) Scatter plot of relative blockade vs. dwell time (left) and relative blockade histogram (right) of a mixture containing labelled and unlabelled peptide. All measurements were done using a buffer containing 1M NaCl, 10mM Tris and 1mM EDTA at pH 7.5. The peptides were added to the cis compartment, and recordings were performed at a bias of -90mV. Data was recorded at a sampling frequency of 500kHz and further low-pass filtered at 10kHz. FraC is shown as a surface representation from the PDB structure 4TSY.

4

tide is, as designed, stalled in the pore at the point where the forces pulling in both directions are equal. Our previous experiments and simulations confirmed the mechanism of such peptide stalling in the FraC nanopore²⁵. During this stalling, a particular region of the peptide will stay closest to the pore constriction at the applied voltage of -90mV. This portion of the peptide represents the most sensitive sensing region of our peptide-nanopore system, and its identification is therefore important for fingerprinting. Hence, we tested three different variants of our model peptide in which the cysteine is placed at different positions along the central part of the peptide, namely, at position 11, 15, or 20 from the N-terminus (Figure 4.2). Notice that a similar approach was used in the early days of DNA sequencing, where a DNA homopolymer was trapped in the alpha-hemolysin nanopore, and single bases were changed sequentially at different positions to find the region closest to the pore constriction¹⁹.

The three different peptide variants were labelled with fluorescein maleimide, HPLC-purified, and mass spectrometry-verified as shown in Figure S4.2. The three samples were measured in FraC as shown in Figure 4.2. The relative blockade observed for an unlabelled peptide is displayed in Figure 2a as a reference (derived from the scatter plots presented in Figure 4.1a). The unlabelled peptide produced relative current blockades of 0.47 ± 0.01 , while a larger relative blockade of 0.57 ± 0.01 was observed for the peptide labelled with fluorescein in position C11, as discussed in the previous section. In this case a consistent increase in the relative blockade can be detected with a well-defined current blockade level as observed in Figure 4.2b. With the peptide labelled with fluorescein in position C15, in contrast, current fluctuations are observed and events often contained a lower current level in the last fraction of the event (Figure 4.2c). As a consequence, a broad population is observed in the relative blockade histogram with a mean of 0.54 and a larger standard deviation of 0.05. Finally, when the label is placed in position C20 we observe a blockade level of 0.48 ± 0.02 , i.e., there is virtually no increase in the relative blockade compared to the unlabelled peptide and we merely observe a tail on the right hand side of the histogram. Most interestingly, while the current level of the event corresponds to that one of the unlabelled peptide for most of the event duration, a dip in the current is observed at the start of most of these events (81%) (Figure 4.2d).

The results suggest that when the label is placed at position C11, it remains in the proximity of the pore constriction during the entire duration of the event. As a consequence a well-defined increase in the blockade is observed. On the other hand, when the label is placed in position C15, fluctuations in the current level are observed, indicating that the label resides in the proximity of the pore constriction only for a fraction of the time of the event duration. Finally when the label is placed in position C20, it resides far from the nanopore constriction for most part of the event and is only temporarily observed at the beginning of the event

as the label moves fast through the nanopore. From the results for these three positions that were tested, we conclude that the label remains the closest to the nanopore constriction when it is placed in position C11, and hence this is the most sensitive region in our model peptide in the FraC system. Moreover, the results suggest that both the relative blockade and individual event characteristics can be used to extract information on the position of the label in such a peptide system.

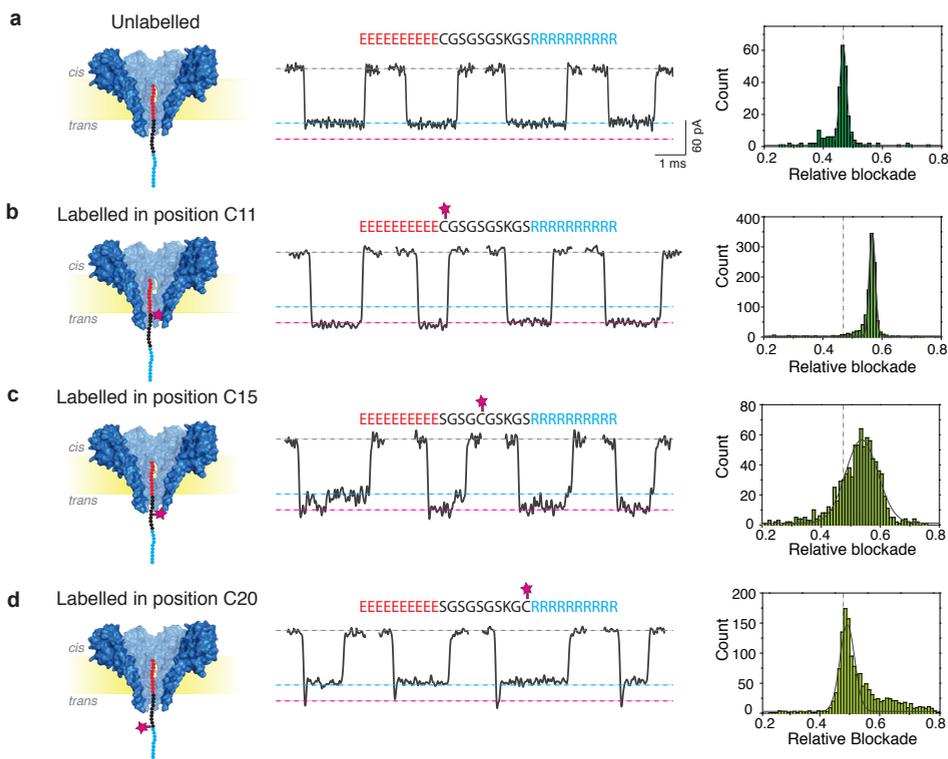


Figure 4.2: Peptide labelled at different positions through FraC. From left to right: Schematics, current traces, and relative blockade histograms for unlabelled peptide (a), and for fluorescein-labelled peptide at positions C11 (b), C15 (c), and C20 (d). All labelled peptide samples were HPLC purified and verified using mass spectrometry. Measurements were done in a buffer containing 1M NaCl, 10mM Tris and 1mM EDTA at pH 7.5. Peptides were added to the cis compartment and measured at -90mV .

4.2.3 Exploring diverse labels using the peptide-FraC system

Protein fingerprinting requires multiple different chemical tags. We proceeded to label the peptide in position C11 with a variety of labels as shown in Figure 4.3a. In ascending order according to their molecular weight, six labels studied were fluorescein (427 Da), Texas Red (728 Da), PEG11-biotin (922 Da), His6 (992 Da), Alexa633 (1089 Da), and 3polyA (1275 Da). We note that these are relatively large tags, 2-7 times larger than the largest

4 amino acid (tryptophan, 204 Da), with different physicochemical properties. Fluorescein, Texas Red and Alexa633 are fluorescent dyes. Their hydrophobic nature potentially allows specific binding modes to the pore, but might promote protein aggregation and reduce protein solubility in a fingerprinting approach, in which a protein should be labelled at every 5-10 amino acids. The hydrophilic tags 3polyA, His6, and PEG11-biotin were also tested. 3polyA is a small oligonucleotide containing 3 adenine bases and a maleimide coupling group at the 5'-end. Due to the three phosphate groups, this tag has a net negative charge. His6 is a stretch of 6 histidines containing a maleimide moiety. At the pH of our measurements (pH 7.5), and taking into consideration that the pKa of histidine has been reported to be higher than 6.5 when in a his tag, this tag is essentially neutral. By lowering the pH to values closer to the pKa of histidine the number of charges of this tag can be modified. Lastly, the PEG11-biotin structure, comprised of 11 PEG units, was tested. PEG is known to bind positive ions such as K⁺ and Na⁺. It has been established that approximately 8 PEG units are necessary to encompass a Na⁺ ion, and therefore the PEG11 biotin is likely to have a single positive charge^{29,30}.

Figure 4.3b shows the six relative blockade histograms observed for peptides labelled with each of the six tags mentioned above. Example event traces of each of the labels can be found in Figure S4.3. In each of these measurements, labelled and unlabelled peptides were measured as a mixture, where the unlabelled peptide was acting as a reference. Hence, two peaks can be observed in each of the histograms. The first peak near 0.45-0.50 corresponds to the unlabelled peptide, and the second peak, near 0.56-0.66 corresponds to the labelled peptide with each of the tags. We calculated the change in relative blockade caused by each label by calculating the difference in relative blockade between the labelled and unlabelled peptide. The result is shown in Figure 4.4a. Based on the change in relative blockade, the tags can be categorized into two groups: Texas Red, Alexa633 and 3polyA causing a large change in relative blockade of 0.17-0.19; fluorescein, His6 and Peg11-biotin causing a smaller change of 0.10-0.11 (Figure 4.4b). We also attempted to measure the free tags in solution, but for most of them, no conclusive results could be obtained, probably because translocation of these small groups occurred too fast for reliable detection.

Interestingly, the increase in relative blockade produced by a particular tag (Figure 4.4) does not correlate well with its molecular weight ($R^2 = 0.16$), as shown in Figure S4.4. For example, while Texas Red has a lower molecular weight than PEG11-biotin (728 Da vs. 922 Da, respectively), it generates a larger blockade. The poor correlation is potentially related to their different geometry. While PEG has a linear flexible structure, Texas Red is comprised of multiple aromatic rings packed tightly one next to another. One can imagine that a larger number of ions is blocked by Texas Red while sitting in the constriction as compared to the linear structure of PEG that most probably extends partially out of the constriction area. To

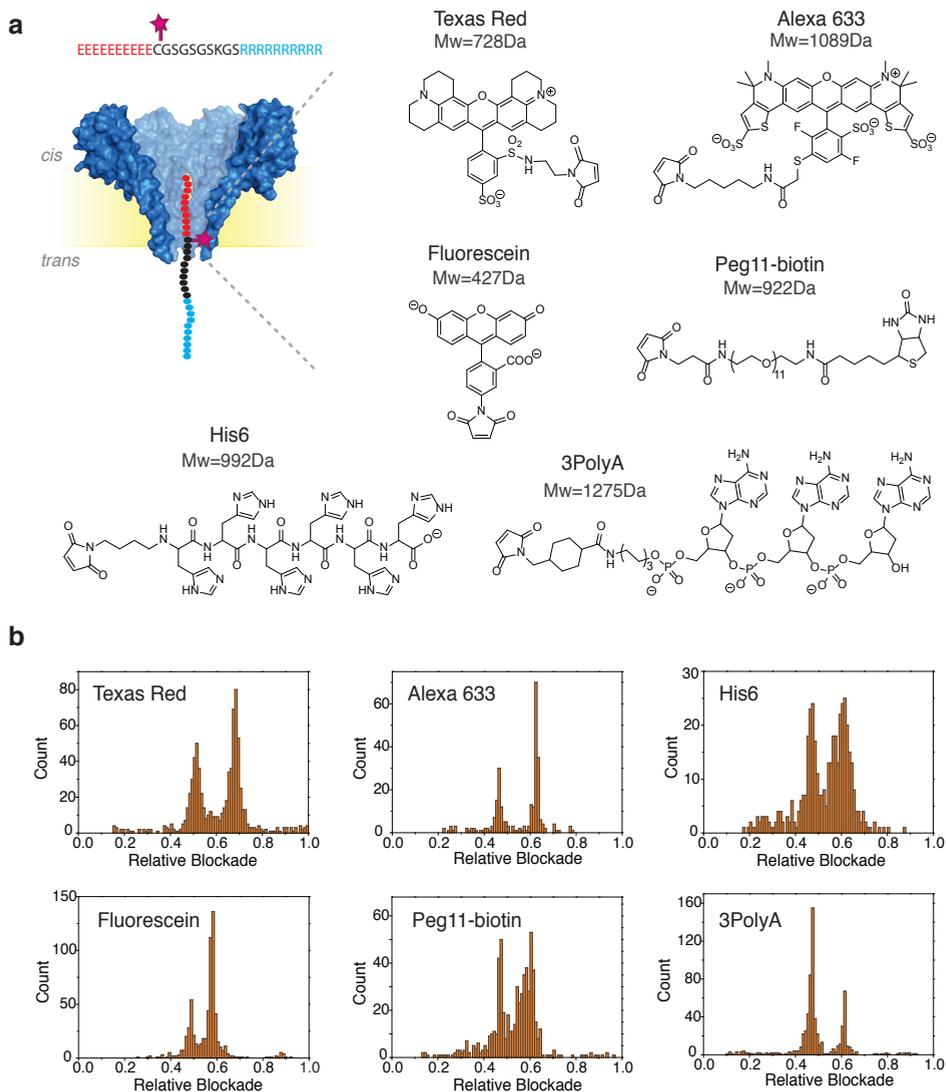


Figure 4.3: (a) Different chemical structures used to label the peptide at position C11. The six structures correspond to Texas Red (728 Da), Alexa633 (1089 Da), 3polyA (1275 Da), Fluorescein (427 Da), PEG11-biotin (922 Da), and His6 (992 Da). (b) Histograms of the relative of blockade of unlabelled and labelled peptide mixtures. Two peaks can be observed in each of the histograms. The first peak corresponds to the unlabelled peptide and the second peak corresponds to the labelled peptide with each of the tags. These measurements were performed in buffer containing 1M NaCl, 10mM Tris and 1mM EDTA at pH 7.5. Peptide mixtures were added in the cis chamber and measured at -90mV. Data was recorded at 500kHz and low-pass filtered with a Gaussian filter at 10kHz.

find an alternative characteristic that takes into consideration both the molecular weight and shape, we define a phenomenological parameter $P = S \times M$, where M is its molecular

weight of the label and S is a shape factor calculated as the ratio between the width of the molecule and its length. The length of the molecule is defined as the end-to-end distance of the molecule following the direction of the linker attached to the nitrogen of the maleimide group, and the distance measured in the perpendicular axis is the width (Figure 4.4c insert). Figure 4.4c shows the increase in relative blockade measured for each label vs. the calculated P value. A fair correlation ($R^2 = 0.84$) is observed between parameter P and the increase in relative blockade confirming that not only the molecular weight of the label, but also its geometry affects the amount of current blocked by a label.

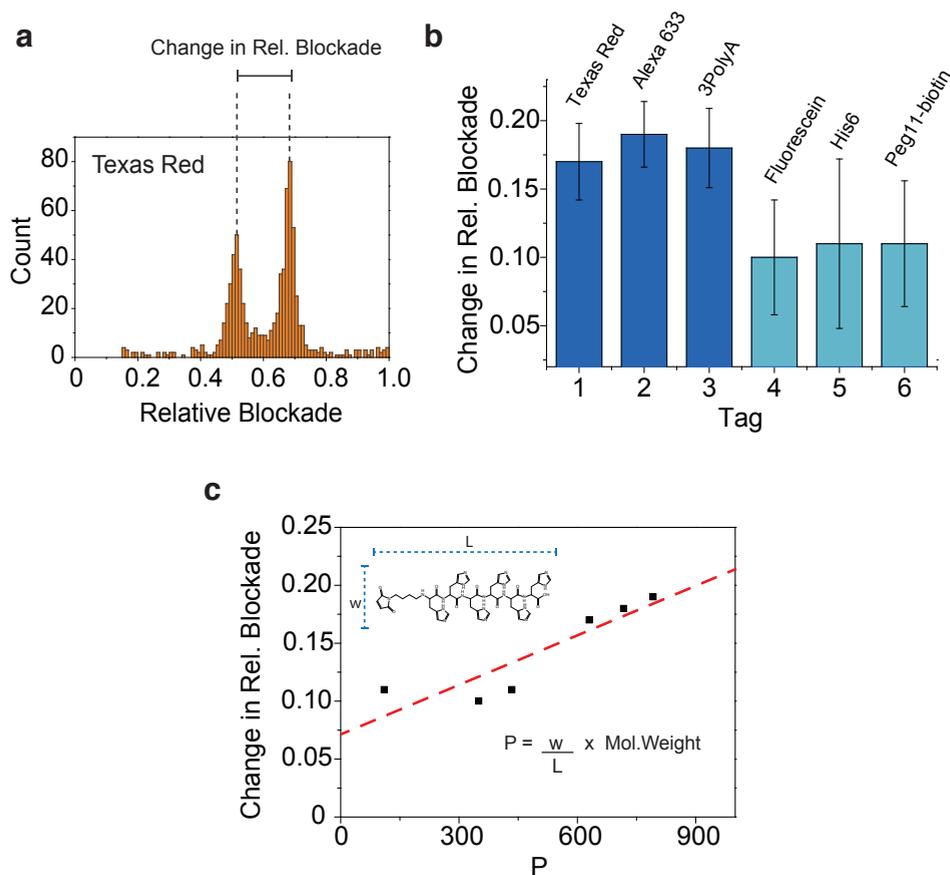


Figure 4.4: (a) Relative blockade histogram for the mixture of unlabelled peptide and peptide labelled with Texas Red showing the change in relative blockade. (b) Shift in relative blockade measured for each label. Texas Red, Alexa633 and 3polyA cause a larger change in relative blockade. (c) Correlation between the change in relative blockade and parameter P which characterizes the label's geometry and molecular weight ($R^2 = 0.84$).

Different chemical labels not only cause a measurably different increase in the relative blockade, but they also influence the translocation time of the peptide. See Figure 4.5,

where panels a and b compare the scatter plots of relative blockade vs. dwell time for the peptide labelled with His6 and with Alexa633. The dwell time is significantly different with a mean translocation time of 29 ms for Alexa633 versus 0.22 ms for His6 at a bias of -90 mV. We observed that two main properties of the tags have an effect on the translocation time: charge and hydrophobicity. Charges present in the labels can act together with the glutamates or arginines of the peptide increasing the electrophoretic force pulling towards the cis or the trans opening of the pore, thus increasing or decreasing the translocation time respectively. Figure 4.5c shows the scatter of dwell time vs. net charge for each of the labels. A correlation between these two characteristics is observed with $R^2 = 0.70$. We also performed control experiments with peptides containing a constant number of 10 positive charges in the C terminus, but varying number of negative charges in the N terminus (10, 12, or 14). As shown in Figure S4.5, the translocation dwell times increased with the increasing number of negative charges, as expected.

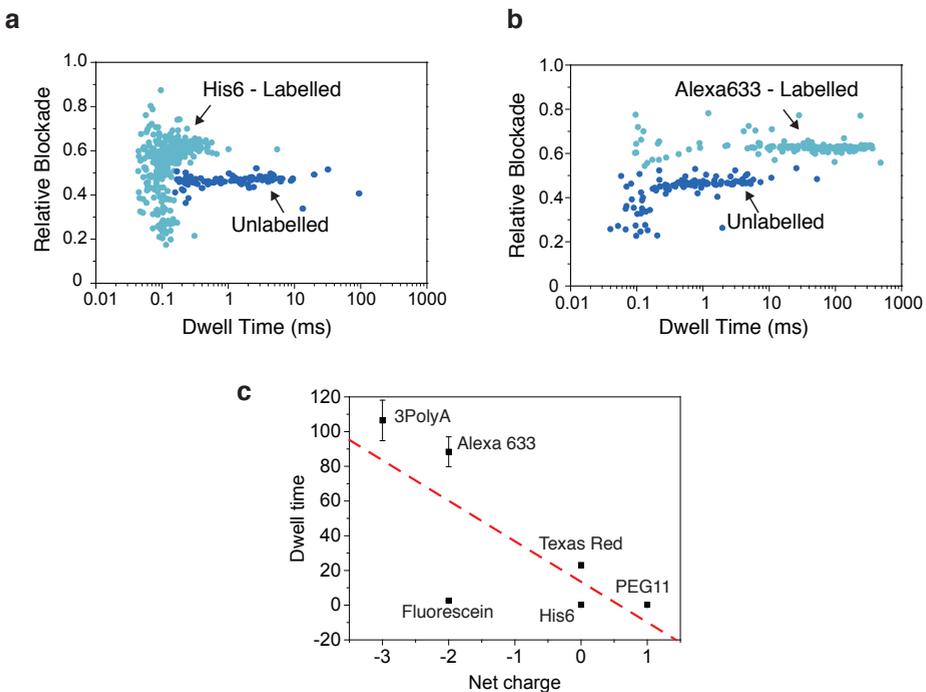


Figure 4.5: Scatter plots of the peptide labelled with (a) His6 and (b) Alexa633. Faster translocation times are observed in peptides labelled with His6 compared to Alexa633. (c) Plot of dwell time vs. net charge. The error bars represent the standard error of the fit. A correlation ($R^2 = 0.70$) is observed between the two parameters.

4.2.4 Moving towards a generalized protein fingerprinting scheme

Up to now, we focused on the detection of chemical labels in a model peptide in which a tug-of-war mechanism stalls the central part of the peptide close to the pore constriction. This represents an important step towards the well-resolved characterization of different labels. However, in a more generalized fingerprinting scheme, the peptide will be translocating at a higher speed and the label will most probably appear as a short dip in the current. This is indeed what we observed when we placed a label in position C20 (Figure 4.2d, Figure 4.6a). At this position the label moved fast through the pore and was only observed in the first fraction of the event. As these measurements provide useful information for a generalized fingerprinting scheme, we proceeded to characterize these events in more detail. Specifically, we quantified the number of peaks in each of the events using a MATLAB script. The results are shown in Figure 4.6b. From a total of 2427 events, 1978 events contained a clear single dip in the current. This corresponds to 81% of the events. Only 315 events (13%) contained no distinguishable peak and 134 events showed more than one peak (6%). The high percentage of events containing a single peak (81%) is encouraging for the realization of a generalized fingerprinting scheme. The small percentage of events in which no peak is visible is most likely due to fast translocations of the label through the pore, which thus occasionally escapes detection due to the finite time resolution. If an enzyme or other mechanism would be used to slow down the speed of a substrate translocating through the pore, this percentage will be reduced significantly. Event traces with more than one peak can be attributed to situations in which the label is read multiple times due to thermal motion of the peptide. Using a processing enzyme to control the translocation of the peptide through the pore would reduce the occurrence of these events.

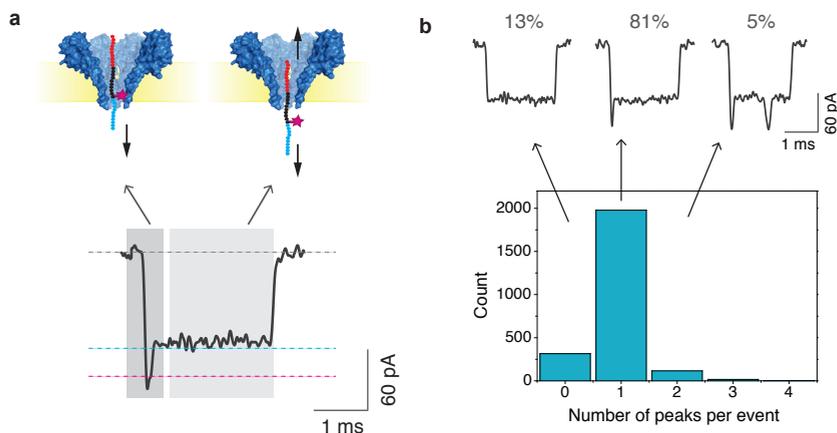


Figure 4.6: (a) Typical event observed for fluorescein-labelled peptide in position 20C. A decrease in current is observed in the first fraction of the event as the labelled portion of the peptide moves through the FraC pore. (b) Histogram of the number of peaks observed per event. Data was recorded at -90mV using a sampling frequency of 500kHz. Event traces were low-pass filtered at 5kHz.

4.3 Discussion and conclusion

Using the bipolar peptide and FraC nanopore as a model system, we studied the characteristics of different chemical labels and explored their potential for a nanopore fingerprinting approach. Six different labels were characterized in terms of their current blockade and translocation time. We observed a correlation between the translocation time of the peptide-tag system and the hydrophobic nature of the tag, as measured by reversed-phase chromatography. We can successfully predict the current blockade generated by a particular label if information about the geometry and molecular weight of the label is available. Our study suggests that it is possible to label amino acids with multiple distinguishable labels.

We explored different positions along the peptide sequence to find the most sensitive region and showed that information about the position of the dye can be derived from their relative blockade and event characteristics. Moreover, when the label is placed at the bottom C20 position of the peptide, a dip in the current is observed as expected in a fingerprinting scheme where a polypeptide sequentially translocates through the pore. Altogether, our results indicate that a protein fingerprinting approach in which different chemical tags are used to label and recognize different amino acids is feasible.

4.4 Materials and Methods

4.4.1 Peptide design and synthesis

The peptides used in this work were the model peptide C11 with sequence EEEEEEEECGSGSGSKGSRRRRRRRRRR (HPLC purity= 95.8%, MW= 3678.9 Da), model peptide with C15 with sequence EEEEEEEESGSGCGSKGSRRRRRRRRRR (HPLC purity= 95.1%, MW= 3678.9 Da), and model peptide with C20 with sequence EEEEEEEESGSGSGSKGCRRRRRRRRRR (HPLC purity= 95.7%, MW= 3678.9 Da). Peptides were synthesized by Biomatik Corporation (Cambridge, CA). The synthesis was performed using standard solid-phase methods and the peptides were further purified using reverse phase HPLC and analysed by mass spectrometry (Biomatik). Peptides were kept lyophilized or, when necessary, aliquoted to a final concentration of 10 mg/ml at -20°C.

4.4.2 Chemical tags containing a maleimide group

The maleimide-containing molecules used as tags were Fluorescein-5-maleimide (Thermo), Texas Red C2 maleimide (Thermo), Alexa Fluor 633 C5 maleimide (Thermo), EZ-link maleimide-PEG11-biotin (Thermo), Histidine6 maleimide (Biomatik), 5'maleimide 3PolyA

(Biosynthesis).

4.4.3 Peptide labelling and purification

Polypeptides containing a cysteine residue were labelled using maleimide chemistry. For labelling, the final peptide concentration was 1 mg/ml and an excess of tag from 10:1 was used. An exception was 3PolyA where a ratio of 2:1 was used due to the low amounts of tag available. Labelling proceeded at 4 °C overnight in 1x PBS at pH 7. Labelling was done with degassed buffers and under nitrogen to prevent cysteine oxidation. For synthetic peptides, no significant difference in labelling was observed if cysteines were reduced previous to labelling, and therefore TCEP (tris(2-carboxyethyl)phosphine, Sigma Aldrich) was not necessary. Labelled peptides were purified using reverse phase chromatography. For that an Agilent 1260 Infinity HPLC system was used with a Waters CSH C18 column as the stationary phase and, a mobile phase consisting of a gradient of acetonitrile and water with 0.1%TFA. Fractions were collected and analysed using MALDI-TOF (Autoflex Speed). The matrix consisted of 10 mg/mL alpha-Cyano-4-hydroxycinnamic acid and 0.2% TFA

4.4.4 Electrical recording in planar lipid membranes

Electrical recording were performed using planar lipid membranes (BLMs) as described before^{32,33}. Briefly, A 25 µm-thick Teflon film (Goodfellow Corporation, Pennsylvania USA) containing an orifice of approximately 70 µm separates the cis and trans compartments. To form the membranes, 10 µl of 5% hexadecane in pentane is added to the Teflon film and the pentane is allowed to evaporate. The reservoirs are filled with buffer and 10 µl of 10 mg/ml DPhPC (1,2-diphytanoyl-sn-glycero-3-phosphocholine, Avanti Polar lipids) in pentane. Membranes were spontaneously formed using the Montal-Mueller method³⁴. Ag/AgCl electrodes are placed in each compartment, with the ground electrode in the cis side. WT FraC oligomers are added to the cis side of the chamber^{23,24}. Upon pore insertion, the pore is characterized by measuring traces at different voltages and taking an IV curve. The substrate was added to the cis side of the chamber and measured at -90 mV.

4.4.5 Data acquisition and analysis

Nanopore recording were collected using a patch-clamp amplifier (Axopatch 200B, molecular devices, USA) at a filtering frequency of 100 kHz. The data was digitized using an Axon Digidata 1550B digitizer at a sampling frequency of 500 kHz. The signal was low-pass filtered at 10 kHz and processed using a Matlab script (Transalyzer)³⁵. Event traces were filtered at 5kHz for display.

4.5 Supplementary information

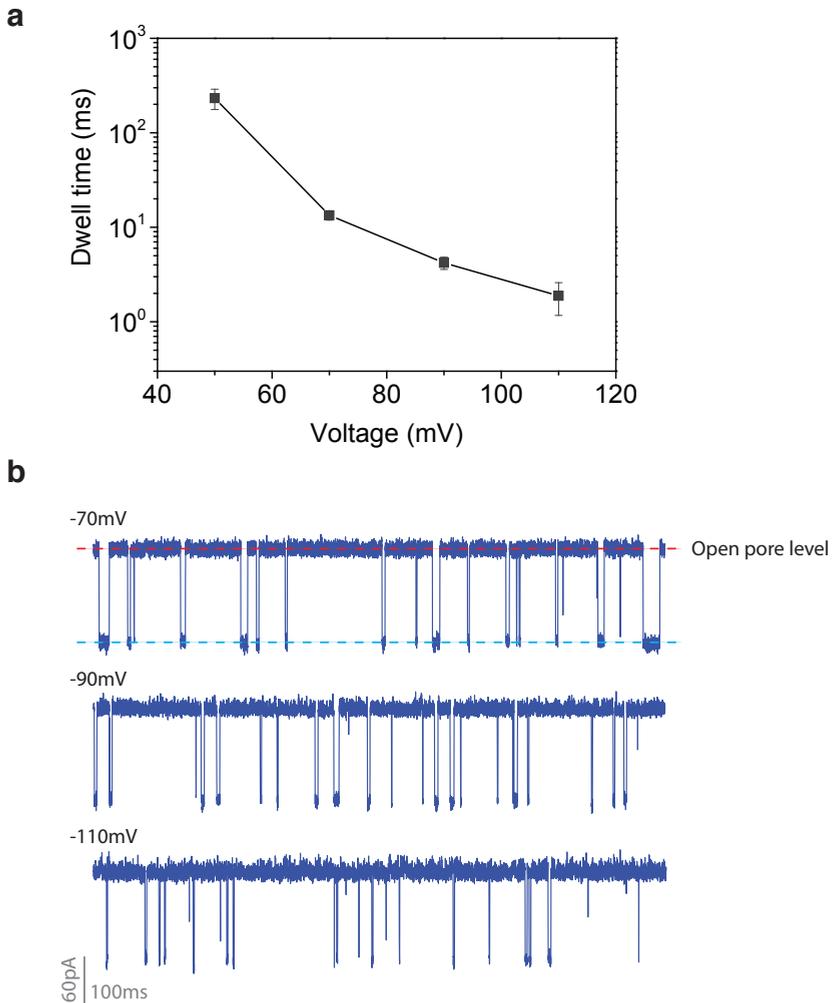


Figure S4.1. Translocation of the model peptide through a FraC pore at different voltages. (a) Average dwell time vs. applied bias for the unlabelled model peptide. The mean and standard deviation are calculated based on three independent measurements. (b) Typical current traces of the unlabelled peptide at -70mV, -90mV and -110mV. Measurements were done in buffer containing 1M NaCl, 10mM Tris and 1mM EDTA at pH 7.5. Peptide was added at the cis compartment.

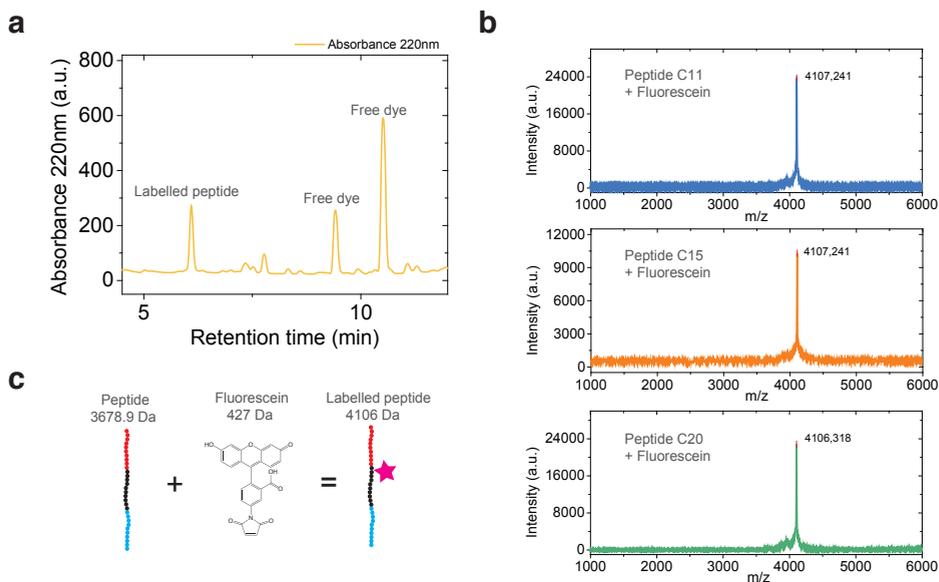


Figure S4.2. Purification and characterization of labelled peptides. (a) 220nm absorbance signal vs. retention time obtained for peptide labelled at position C11 during the purification using reversed-phase HPLC. The fraction with a retention time of 5.9 min corresponds to the labelled peptide. Fractions at times 9.4 and 10.5 correspond to free dye. (b) MALDI-TOF spectra of the fraction corresponding to labelled peptide in position C11, C15, and C20. The masses correspond to the labelled peptide as shown in the schematic of panel c.

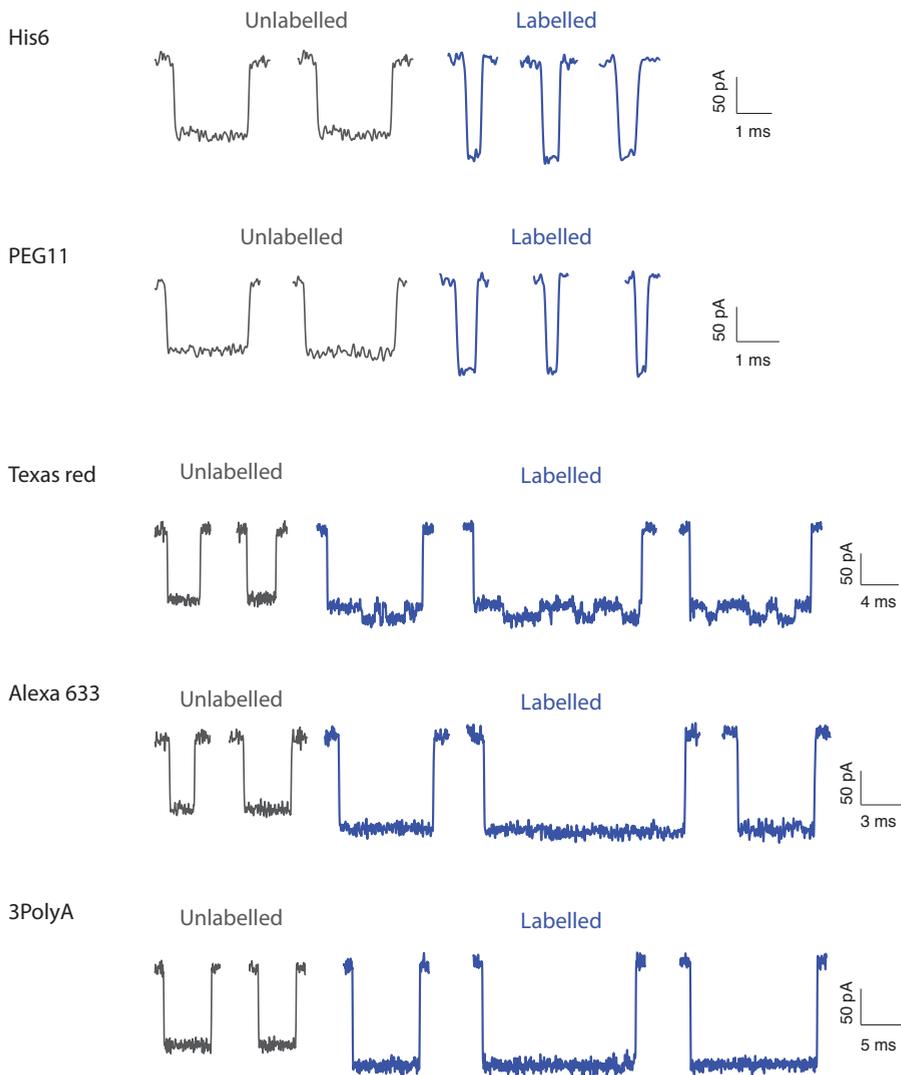


Figure S4.3. Example event traces of the peptide labelled in position 11 with the different tags. From top to bottom we present event traces of the peptide labelled with His6, PEG11, Texas Red, Alexa 633, and 3PolyA. Events of the unlabelled peptide are displayed on the left for reference. With Texas Red, two level fluctuations are observed in each event. The other labels produce a well-defined blockade level.

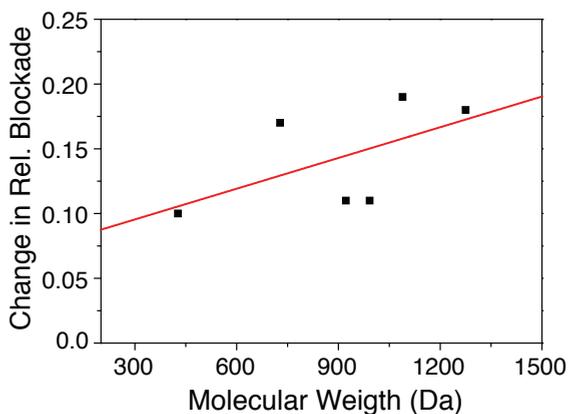


Figure S4.4. Scatter plot of the Change in relative blockade vs. molecular weight for the six different labels. Poor correlation ($R^2 = 0.16$) is observed between these parameters.

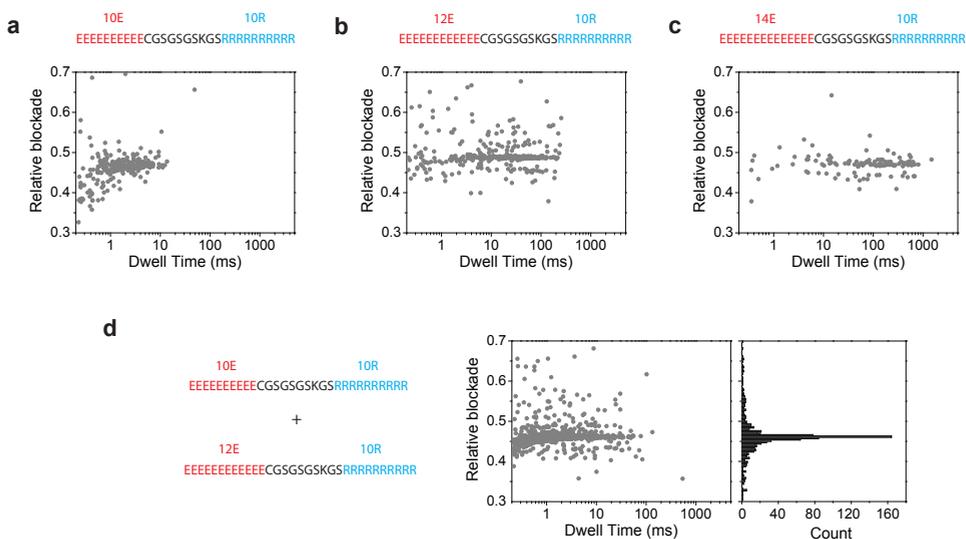


Figure S4.5. Control measurements of peptides with different number of negatively charged amino acids in the N-terminus. a) Measurement of the regular model peptide with 10 arginines and 10 glutamates. b) Measurement of a peptide with 10 arginines and 12 glutamates. c) Measurement of a peptide with 10 arginines and 14 glutamates. From the scatter plots a clear increase in dwell time can be observed for peptides with higher number of glutamates. d) mixture of regular model peptide and peptide with 12 glutamates and 10 arginines. No difference in relative blockade is observed in these peptides.

References

1. Kim, M.; Rai, N.; Zorraquino, V.; Tagkopoulos, I. *Nat. Commun.* 2016, 7, 13090.
2. Franklin, S.; Vondriska, T. M. *Circ. Cardiovasc. Genet.* 2011, 4 (5), 576.
3. Karczewski, K. J.; Snyder, M. P. *Nat. Rev. Genet.* 2018, 19 (5), 299–310.
4. Tyers, M.; Mann, M. *Nature* 2003, 422 (6928), 193–197.
5. Karahalil, B. *Curr. Med. Chem.* 2016, 23 (37), 4221–4230.
6. Haider, S.; Pal, R. *Curr. Genomics* 2013, 14 (2), 91–110.
7. Liu, Y.; Beyer, A.; Aebersold, R. *Cell* 2016, 165 (3), 535–550.
8. KA., W. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) www.genome.gov/sequencingcostsdata (accessed Jul 2, 2018).
9. Goodwin, S.; McPherson, J. D.; McCombie, W. R. *Nat. Rev. Genet.* 2016, 17 (6), 333–351.
10. Sidoli, S.; Kulej, K.; Garcia, B. A. J. *Cell Biol.* 2017, 216 (1), 21–24.
11. *Nat. Biotechnol.* 2003, 21 (3), 213.
12. Mitchell, P. *Nat. Biotechnol.* 2002, 21, 233–237.
13. Million, R.; Tolin, S.; Puricelli, L.; Sbrignadello, S.; Fadini, G. P.; Tessari, P.; Arrigoni, G. *PLoS One* 2011, 6 (5), e19603.
14. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R. *Chemical Reviews*. NIH Public Access April 10, 2013, pp 2343–2394.
15. Restrepo-Pérez, L.; Joo, C.; Dekker, C. *Nat. Nanotechnol.* 2018, 13 (9), 786–796.
16. Dekker, C. *Nat. Nanotechnol.* 2007, 2 (4), 209–215.
17. Jain, M.; Olsen, H. E.; Paten, B.; Akeson, M. *Genome Biol.* 2016, 17 (1), 239.
18. Deamer, D.; Akeson, M.; Branton, D. *Nat. Biotechnol.* 2016, 34 (5), 518–524.
19. Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. *Proc. Natl. Acad. Sci.* 2009, 106 (19), 7702–7707.
20. Manrao, E. A.; Derrington, I. M.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. *PLoS One* 2011, 6 (10), e25723.
21. Yao, Y.; Docter, M.; van Ginkel, J.; de Ridder, D.; Joo, C. *Phys. Biol.* 2015, 12 (5), 055003.
22. Swaminathan, J.; Boulgakov, A. A.; Marcotte, E. M. *PLOS Comput. Biol.* 2015, 11 (2), e1004080.
23. Huang, G.; Willems, K.; Soskine, M.; Wloka, C.; Maglia, G. *Nat. Commun.* 2017, 8 (1), 935.
24. Wloka, C.; Mutter, N. L.; Soskine, M.; Maglia, G. *Angew. Chemie - Int. Ed.* 2016, 55 (40), 12494–12498.
25. Zhao, S.; Restrepo-Pérez, L.; Soskine, M.; Maglia, G.; Joo, C.; Dekker, C.; Aksimentiev, A. Submitted.
26. Asandei, A.; Chinappi, M.; Kang, H. K.; Seo, C. H.; Mereuta, L.; Park, Y.; Luchian, T.

- ACS Appl. Mater. Interfaces 2015, 7 (30), 16706–16714.
27. Movileanu, L.; Schmittschmitt, J. P.; Martin Scholtz, J.; Bayley, H. *Biophys. J.* 2005, 89 (2), 1030–1045.
 28. Biesemans, A.; Soskine, M.; Maglia, G. *Nano Lett.* 2015, 15 (9), 6076–6081.
 29. Bogan, M. J.; Agnes, G. R. *J. Am. Soc. Mass Spectrom.* 2002, 13 (2), 177–186.
 30. Wong, S. F.; Meng, C. K.; Fenn, J. B. *J. Phys. Chem.* 1988, 92 (2), 546–550.
 31. Maglia, G.; Heron, A. J.; Stoddart, D.; Japrun, D.; Bayley, H. *Methods Enzymol.* 2010, 475 (C), 591–623.
 32. Gutsmann, T.; Heimbürg, T.; Keyser, U.; Mahendran, K. R.; Winterhalter, M. *Nat. Protoc.* 2015, 10 (1), 188–198.
 33. Montal, M.; Mueller, P. *Proc. Natl. Acad. Sci. U. S. A.* 1972, 69 (12), 3561–3566.
 34. Plesa, C.; Dekker, C. *Nanotechnology* 2015, 26 (084003), 1–7.

5

Label-free detection of post-translational modifications with a nanopore

Protein Post Translational Modifications, or PTMs, play key roles in cellular processes such as replication, cell-to-cell recognition, and apoptosis. Consequently, their detection and identification is important for elucidating complex cellular processes and disease. Current detection methods are in many cases unable to detect low-copy number variants within the vast mixture of proteins present in biological samples. Here we show an alternative method for PTM detection at the single-molecule level using biological nanopores. In particular, we focus our attention in phosphorylation and O-glycosylation. These two PTMs are highly dynamic and have an intricate regulatory process, where they mutually compete for protein modification sites. Their dysregulation has been implicated in pathogenic pathways for cancer, Alzheimer's disease, diabetes, and many more. We show that using a FraC nanopore the phosphorylated and non-phosphorylated peptide variants can be differentiated by their relative current blockade. The same was shown for an O-glycosylated peptide variant. Besides the effect of these modifications on relative blockade, their effect in dwell time was also exploited. Finally, conditions in which these modifications could be differentiated are shown, demonstrating the identification of phosphorylation and O-glycosylation in a label-free manner using nanopores. These results represent an important step for the single molecule identification of proteoforms or protein isoforms, which is of tremendous importance for disease diagnosis and cell biology.

This chapter is in preparation for publication as: Restrepo-Pérez L., Wong CH, Maglia G., Dekker C, Joo C. Label-free detection of post-translational modifications with a nanopore.

5.1 Introduction

5

In the pursuit of a comprehensive understanding of the molecular mechanisms at work in living organisms, one inevitably arrives at the study of proteins. These organic polymers are regarded as the molecular machines of the cell, that enable a spectacularly wide array of characteristics and functionalities. Cells employ proteins for a whole host of different processes including signalling, recognition, differentiation, gene regulation, structuring, and many more¹. To realize such a diverse set of functionalities, cells need a way to create a very diverse array of different proteins. Post-Translational Modifications, or PTMs, are one of the methods by which the protein pool of a cell can be expanded by several orders of magnitude^{2,3}. This collective term refers to all modifications occurring during and after the translational synthesis of proteins, and usually involves the addition of chemical groups to specific amino acids along the protein. PTMs play vital roles in protein (in)activation, stability, and recognition, among other things⁴⁻⁶. In the medical field in particular, PTMs have been found to play key roles in pathogenic pathways for cancer, Parkinson's disease, Alzheimer's disease and diabetes^{7,8}. Consequently, it is pivotal to detect protein PTM variants with high sensitivity.

Mass spectrometry is the current standard method for PTM detection^{9,10}. The technique, however, has its limitations, mainly brought about by the large difference in its dynamic range (10^4 - 10^5) compared to the immense dynamic range of protein concentrations that are present in *in vivo* samples (10^8 - 10^9)^{3,11-13}. Challenges arise when low-copy number proteoforms are of interest but get swamped in a sea of others, especially because a protein can have both different kinds and quantities of PTMs. In practice, extremely sparse proteins are invisible in the noise of a diverse mixture of abundant proteins¹¹. For this very reason, many biomarkers for cancer and other diseases escape timely detection¹⁴⁻¹⁶. To alleviate this problem, pre-purification steps based on chromatography or antibodies are often used to enrich samples of interest^{9,15}. Other approaches include the PTM-specific removal and subsequent chemical labelling, enabling the researcher to detect it through characteristic mass-shifts in a mass spectrum^{15,17}. These enrichment and labelling methods are, however, time-consuming and highly PTM specific, which discourages its use in the blanket analysis approaches that are sought after in clinical research.

Single-molecule techniques, such as nanopores, open the door for new approaches for PTM detection with higher sensitivity. In a nanopore set up, an insulating membrane with a nanometer-sized aperture separates two electrolyte-filled compartments. Upon applying an electric potential, an ionic current flows through the nanopore and a decrease in the current is measured when a protein translocates through the pore^{18,19}. Several groups have previously reported the successful study of peptides²⁰⁻²⁶ or unfolded proteins²⁷⁻³⁵ translocating

through a nanopore. Features in the protein structure, such as PTMs, can result in slightly different current blockage characteristics, allowing for label-free detection. In recent years, this principle was used for the detection of large PTMs such as a large N-glycosylation or ubiquitination^{36,37}. The detection of small PTMs (<500Da) has remained a challenge, and so far was only demonstrated for a phosphorylated protein³⁸.

In recent years, it has been demonstrated that phosphorylation and the O-glycosylation GlcNAc (N-acetylglucosamine) have an intricate connection and complex interplay within the cell. These PTMs compete for serine and threonine modification sites, and therefore mutually regulate each other. Malfunctions of this regulatory network have been associated to chronic diseases such as diabetes, cancer, and Alzheimer's^{7,8,10}. For example, in Alzheimer's disease, a decrease in the levels of O-GlcNAc in the tau protein, has been associated to its hyper-phosphorylated state and the formation of the intraneuronal tangles characteristic of the disease. Consequently, finding efficient mechanisms for the label-free detection of these PTMs is of tremendous medical relevance.

Here, we demonstrate the label-free detection of both phosphorylation and O-glycosylation, and their discrimination from the wild-type unmodified proteins. This is achieved using a biological nanopore (Fragaceatoxin C, or FraC) and a model peptide system in which a serine is modified with either of these modifications. We first show that phosphorylated and non-phosphorylated peptides can be distinguished by their difference in the relative current blockade. Subsequently, the same is done for glycosylated and non-glycosylated peptides. Finally, we demonstrate for the first time, at the single molecule level, that phosphorylated and O-glycosylated peptides can be differentiated in a label-free manner.

5.2 Results

For the detection of phosphorylation and O-glycosylation, we used a model peptide (N'-EEEEEEEEESGSGSGSKGSRRRRRRRRRR- C') in a FraC nanopore. Our approach is presented in Figure 5.1. This peptide contains a stretch of 10 negatively charged glutamic acid residues at the N-terminus and a stretch of 10 positively charged arginine residues at the C-terminus. The two charged regions are connected by a flexible sequence of glycine (G) and serine (S) residues. Upon applying a negative bias to the trans compartment, the positive R-stretch enters the pore first, thus orienting the peptide in a defined way. Subsequently, when the negative E-stretch enters the proximity of the pore constriction, the positive bias of the cis compartment exerts an opposing force. A similar peptide was used by Asandei et al to study peptide-nanopore interactions at different pH values³⁹. The peptide thus essentially gets stalled at an equilibrium position where the forces in both directions cancel out. This

tug-of-war mechanism gives rise to long read lengths (> 1 ms), enabling one to extensively probe a particular region of the peptide until it is thermally disturbed and translocated. Moreover, the pulling mechanism stretches the peptide, allowing the analysis of a linearized molecule.

Substantial evidence was found for the existence of such an equilibrium position by tagging this model peptide with large chemical moieties at various locations along the peptide (Chapter 4). These previous experiments reported that the peptide is stalled in the FraC nanopore at the position where amino acid in position 11 (from the N-terminus) is closest to the pore constriction. For PTM analysis, we therefore placed a serine in position 11, and evaluated variants of the peptide containing either a phosphorylation or an O-glycosylation in this position.

5

5.2.1 Identification of phosphorylated peptides with a FraC nanopore

Measurements were performed using a wild type FraC nanopore in a buffer containing 10 mM TRIS, 1 mM EDTA and either 1M NaCl or 0.8M NaCl as specified. Peptides were added to the cis side of the flow cell at concentrations between 100 – 300 nM. A negative bias of -90mV was applied to the trans compartment to avoid gating that is observed in FraC under positive bias.

After the addition of peptides containing no modifications, we observed well-defined and consistent current blockades. Figure 5.1b shows a scatter diagram of the relative current blockade versus the dwell time observed for $N=449$ translocation events, and the histogram for the relative blockade. We define the relative blockade as the current blockade ($\Delta I = I_{\text{blockade}} - I_{\text{open pore}}$) divided by the open pore current ($I_{\text{open pore}}$). For the control peptide without PTMs, a relative blockade of 0.48 ± 0.01 was observed with a mean translocation time of 1.38 ± 0.49 ms. Error bars represent the standard deviation over at least 3 independent experiments.

Subsequently, peptides containing a single phosphorylation (i.e. a single PO_3^{2-} group) in position 11 were added to the cis side of the chamber at a same concentration, thus leading to a 1 to 1 mixture of unlabeled and PTM-labelled peptides. As shown in Figure 5.1c, this led to two clear populations with comparable densities in the scatter plot of relative blockade vs. dwell time ($N=398$ for the phosphorylated, and $N=350$ for the control peptide). This indicates that the phosphorylation PTM can be clearly detected with the FraC nanopore. The well-defined difference in current blockade is also apparent in the histogram displayed in Figure 5.1c which shows two clearly separated peaks. The relative blockade of the bottom population (0.48 ± 0.02) corresponded to the unmodified control peptide,

while a larger relative blockade of 0.52 ± 0.01 was found for the phosphorylated peptide. This represents an 8.7 % increase compared to the control peptide.

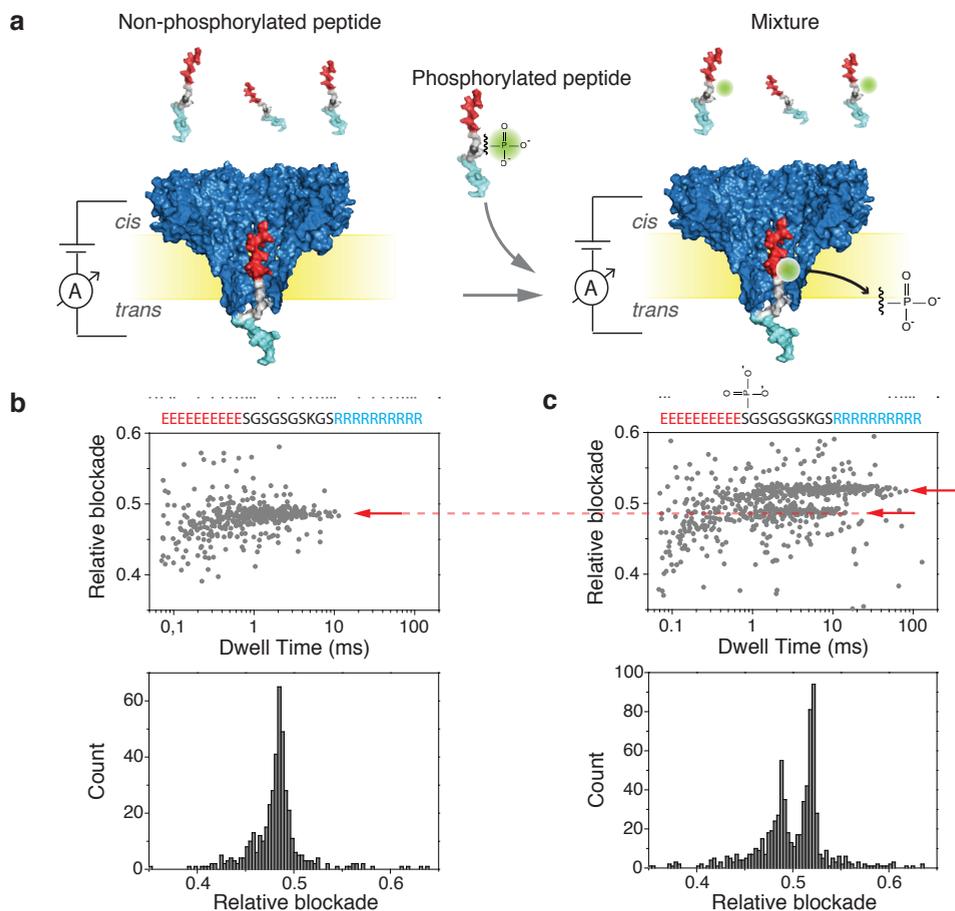


Figure 5.1: (a) Schematic representation of the measurement set-up, where peptides are driven through a FraC nanopore. Experiments were done with a non-phosphorylated peptide (left), and with an equimolar-concentration mixture of phosphorylated and non-phosphorylated peptides (right). (b) Scatter plot of relative blockade versus dwell time (top), and relative-blockade histogram (bottom) for the non-phosphorylated peptide. (c) Same for the mixture of phosphorylated and non-phosphorylated peptides. A second population with higher relative blockade is now visible, as shown by the arrows.

5.2.2 Identification of glycosylated peptides with a FraC nanopore

After the successful detection of the phosphorylated variant, we moved to the analysis of the O-glycosylated peptide. O-glycosylated peptides have the glycan group attached to the amino acid residue through an oxygen group. O-glycans such as GlcNAc (N-acetylglucosamine) occur naturally in serines and threonines and are of growing importance in

proteomics because of their close connection to phosphorylation regulation which plays a key role in several chronic diseases.

Here, we analyse a peptide containing an O-GlcNAc glycosylation in position 11. GlcNAc is a small PTM comprised of a single sugar moiety with a total mass of 203 Da. Figure 5.2 shows the result of the non-glycosylated peptide as well as that of a mixture containing the O-glycosylated peptide together with the control peptide at equimolar concentrations. In Figure 5.2c and 5.2e, two distinct populations are clearly discernible in their relative blockade values, leading us to conclude that the nanopore is able to distinguish O-GlcNAc PTMs from the unmodified peptide. Analysis of repeat measurements show that the O-GlcNAc variant has an average relative blockade of 0.52 ± 0.01 . This represents an 8.3% increase compared to the unmodified peptide.

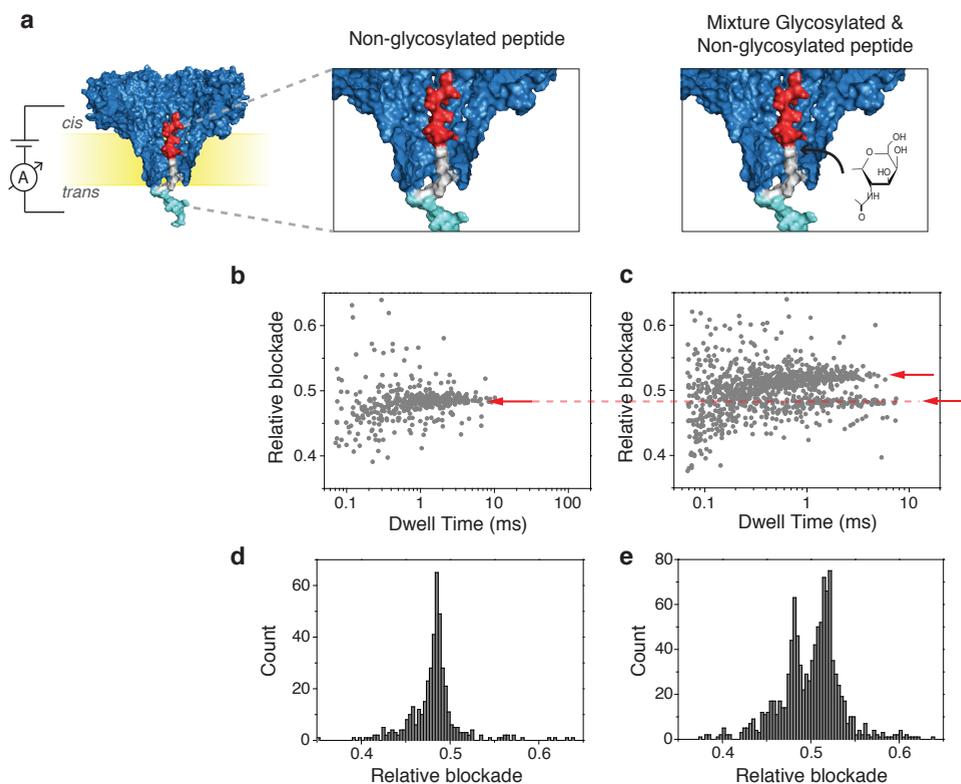


Figure 5.2: (a) Schematic representation of the measurements for the non-glycosylated peptides (left) and for an equimolar-concentration mixture of glycosylated and non-glycosylated peptides (right). (b) Scatter plot of relative blockade versus dwell time for the non-glycosylated peptide. (c) Same for the mixture of glycosylated and non-glycosylated peptides. A second population with higher relative blockade is visible as shown by the arrows. (d) Relative blockade histogram of the non-glycosylated peptide. (e) Relative blockade histogram of the mixture of non-glycosylated and glycosylated peptide.

We note that the increase in relative blockade observed by the phosphorylated and glycosylated variants are almost identical (8.7% vs 8.3%) despite their molecular weight difference (80 Da for phosphorylation versus 203 Da for glycosylation). We hypothesize that the negative charge of the phosphorylation group creates an Electrical double layer (EDL) of counter-ions that leads to an increased effective volume of the phosphoryl group.

5.2.3 Distinguishing different PTM modifications with a FraC nanopore

As observed in Figures 5.1 and 5.2, the phosphorylated and glycosylated peptides can not be mutually differentiated based on the relative blockade as measured at 1M NaCl, because the increase in relative blockade caused by the phosphoryl group (8.7%) and the O-GlcNAc group (8.3%) are nearly identical under these measurements conditions. This was also confirmed by measuring a mixture containing the three different peptides: the phosphorylated peptide, the O-glycosylated peptide, and the control peptide (SI; Figure S5.1), where only two populations were observed.

Despite the similarity in relative blockade observed for both PTMs, we noted a difference between the phosphorylated and glycosylated peptides in other translocation characteristics, notably the translocation time (Figure 5.3). For the phosphorylated peptide an average dwell time of 3.21 ± 2.10 ms is observed, indicating an increase in translocation time compared to the control peptide that has an average dwell time of 1.56 ± 1.1 ms. For the glycosylated peptide, on the other hand, a decrease in dwell time is observed leading to an average dwell time of 0.41 ± 0.24 ms. We attribute the increase in dwell time observed with the phosphorylated peptide to a charge imbalance in the construct caused by the PTM. The addition of a phosphoryl group with a net negative charge shifts the force equilibrium brought about by the positive and negative tail regions of peptide, and as a result, the dwell time increases. We hypothesize that the decrease in dwell time for the GlcNAc group is of entropic nature. Confining the O-GlcNAc moiety into the pore constriction limits its degrees of freedom. As a result, the equilibrium state of the peptide is less energetically favorable for the glycosylated variant than for the unmodified peptide. This increases the probability to thermally disturb the force equilibrium, leading to shorter dwell times. We have previously reported a similar effect when other hydrophilic moieties such as PEG11 were attached to the peptide in the same position (our other paper).

Although the average dwell time for the two PTMs is significantly different, the dwell time of each population is quite widely spread and therefore this characteristic is not ideal for a proper differentiation between the two PTM modifications. Hence, we looked for other ways to differentiate the two PTM variants. As mentioned, the relative blockades are nearly identical in 1M NaCl, but we devised a way to differentially alter the relative blockade of

each modification by changing the salt conditions. Unlike the O-GlcNAc group, which is essentially uncharged, the phosphorylation contains two negative charges at pH=7.5. The ionic strength of our buffer should therefore particularly affect the screening of the phosphoryl group, altering the extend of its electrical double layer (EDL) and its current blockade.

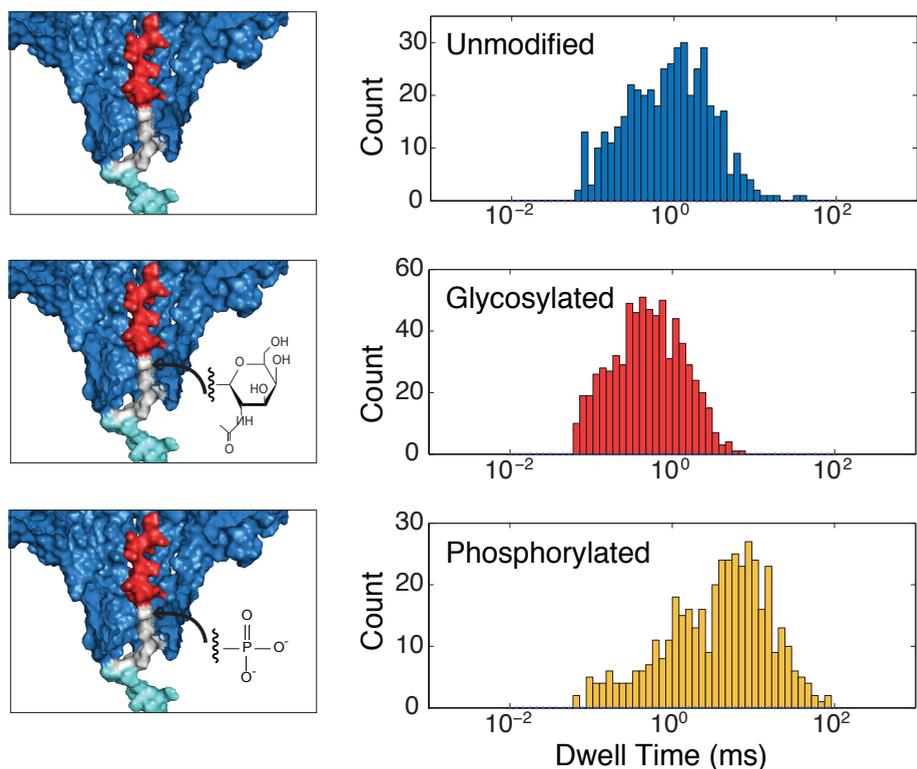


Figure 5.3: Dwell time histograms for the unmodified peptide (top) the glycosylated peptide (middle) and the phosphorylated peptide (bottom).

To test this, we measured a mixture containing the three peptide variants (the two modified peptides and the control) in equimolar concentrations in a buffer containing a lower salt concentration of 0.8M NaCl at pH 7.5. As shown in Figure 5.4, three different populations can be clearly observed in the scatter plot and histogram. Importantly, these conditions thus enable us to confidently detect and differentiate the three different peptide variants. The results were confirmed by independent measurements of each of the PTMs (Figure S5.2). From the data, we can establish that at 0.8M NaCl the increase in relative blockade of the glycosylated peptide remained essentially unchanged (7.9%), whereas for the phosphorylated peptide, an increase in relative blockade of 4.3% is observed at 0.8M NaCl, which is reduced as compared to 8.7% at 1M NaCl.

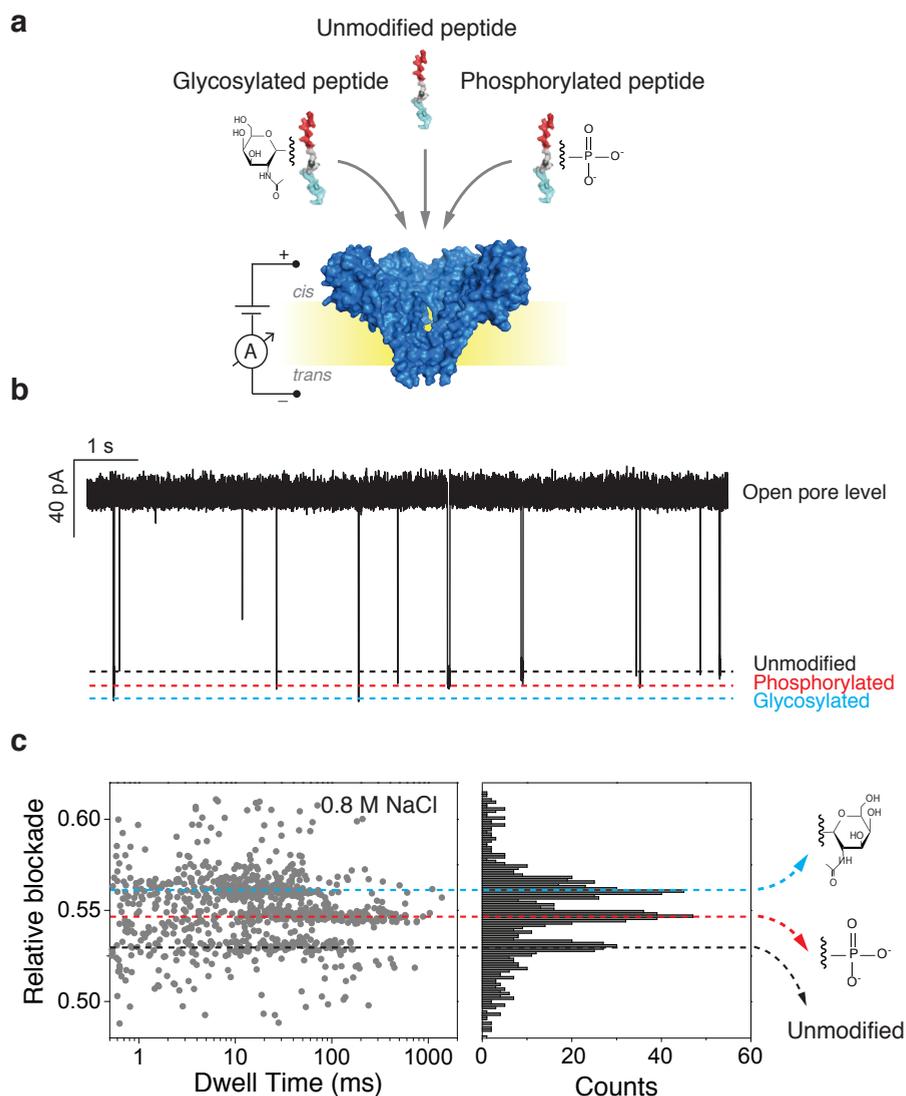


Figure 5.4: Schematic representation of the measurement approach (a) and an example current trace (b) obtained for a measurement on a mixture of the three peptides: the unmodified control peptide, the phosphorylated peptide and the glycosylated peptide. Data are taken at 0.8 M NaCl with pH 7.5. (c) Scatter plot of relative blockade vs. dwell time, and a relative blockade histogram of the mixture of the three peptides. Clearly three different current blockade levels are observed.

We hypothesize that the lower increase in relative blockade observed for the phosphoryl moiety at lower salt has the same origin as the inverted current rectification previously observed in FraC (Figure S5.3). When the diameter of a nanopore with strong ion selectivity is comparable to the Debye screening length λ_D , the polarity of the current rectification

is reversed^{40,41}. The exact mechanism is still a matter of debate, but it is speculated that it originates from the distribution of ions in the pore vestibule and at the *trans* region at the narrow tip of the pore. Accumulation or depletion of charges in this region has an important effect on the pore conductance. In line with this reasoning, the presence of a stronger negatively charged phosphoryl group in the FraC constriction has an effect on this mechanism leading to an increase in current, and thus a decrease in relative blockade compared to the 1M NaCl case.

5.3 Discussion and conclusions

5 Here we have shown that phosphorylation and O-glycosylation post-translational modifications can be detected using the FraC nanopore. We use a bipolar peptide that contains a stretch of negative charges at the N-terminus and positive charges at the C-terminus. Thus, gets stalled upon translocating the nanopore, allowing us to probe a particular region of the peptide with great sensitivity. Upon placing a phosphorylation or a small O-glycosylation group in this peptide region, both PTM peptide variants could be distinguished from the control peptide with an unmodified serine residue. This constitutes, to our knowledge, the first reported successful detection of an O-glycan PTM using nanopores. O-glycans play a crucial role for protein conformation, solubility, and stability, but, compared to N-glycans for example, their study is challenging in mass spectrometry¹⁰. The main reasons for the latter difficulties are the lack of a consensus amino acid sequence where the PTM is present, which hinders the determination of the PTM site, and the lack of universal enzymes to release O-glycans from the protein substrates. Nanopores thus present an attractive alternative for O-glycosylation characterization.

Next to the change in relative blockade detected for the PTM variants, we also observed changes in the translocation time of the peptides containing PTMs. Longer dwell times were observed for the phosphorylated peptide, presumably because of the negative charges in the phosphoryl group, which increase the forces in the peptide towards the *cis* compartment. Shorter dwell times were observed for the O-GlcNAc variant, which may be due to the entropic penalty associated with confining the sugar moiety in the pore constriction.

Upon changing the NaCl salt concentration from 1M to 0.8M, an interesting effect was observed for the phosphorylation. Upon decreasing the salt concentration of the buffer, one would expect an increase in the size of the electrical double layer, and thus an increase of the current blockade caused by the phosphoryl group. Our experiments showed the opposite, however, with a reduced blockade was observed at lower salt concentration. We hypothesize that this decrease has the same origin as the inversion of current rectification

phenomena observed in FraC: the presence of an effectively larger negative charge in the pore constriction generates an increased current.

These results show, for the first time, nanopore-based differentiation between phosphorylated and O-GlcNAc glycosylated peptide variants in a label-free manner. These results bear potential for PTM detection that is important for the study of diseases such as cancer or diabetes, where the intricate regulation of phosphorylation and O-glycosylation plays an important role in the pathogenesis of the disease.

In future work, the nanopore-based detection of PTMs on peptides shown here can be expanded to PTM detection in full proteins, upon using an enzyme or another mechanism to slow down the linear translocation of a full protein through the nanopore. Such PTM detection on proteins would allow the identification of individual protein isoforms in a sample, providing a radical improvement to current proteomics. Present-day proteomics methods typically rely on approaches in which proteins are digested and peptide fragments are analyzed with their PTMs, which is inherently problematic since proteoforms cannot be correctly identified from the fragmented information of these analytes – a problem often referred to as the protein inference problem. Nanopore methods that sequentially read full proteins could thus allow the sensitive detection of the proteoforms present in a sample, bringing important improvements for PTM-based biomarker detection and the correct identification of proteoforms. Our current results present a promising step into this direction.

5.4 Supplementary information

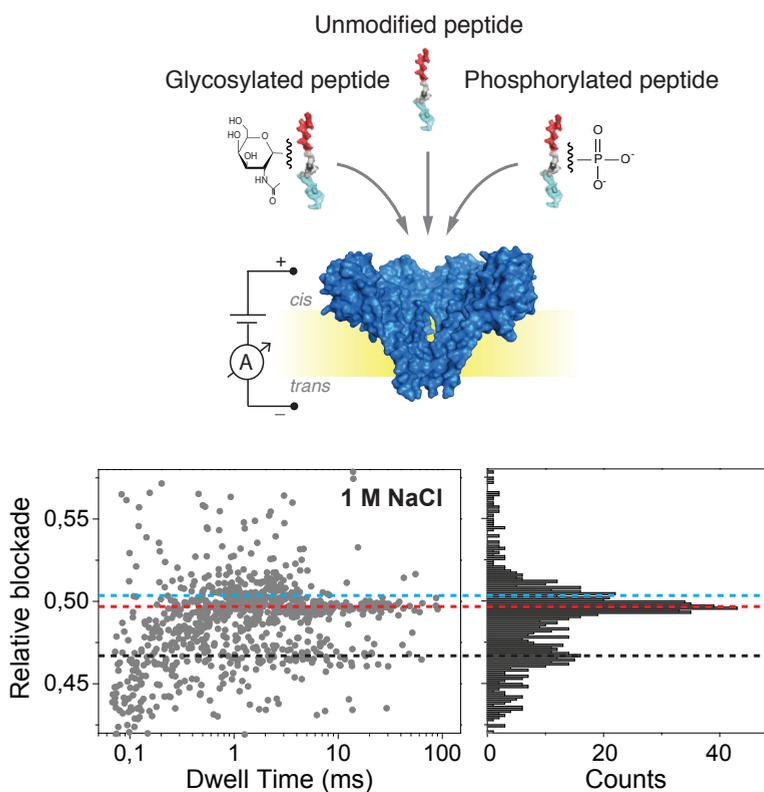


Figure S5.1. Translocation of a peptide mixture containing phosphorylated, glycosylated and unmodified peptide as a control. Left plot shows a scatter plot of relative blockade vs. dwell time. Right plot shows the relative blockade histogram. Measurements were done in buffer containing 1M NaCl, 10mM Tris and 1mM EDTA at pH 7.5. Peptides were added at the cis compartment. A clear distinction between the phosphorylated and glycosylated peptide is not possible under these experimental conditions.

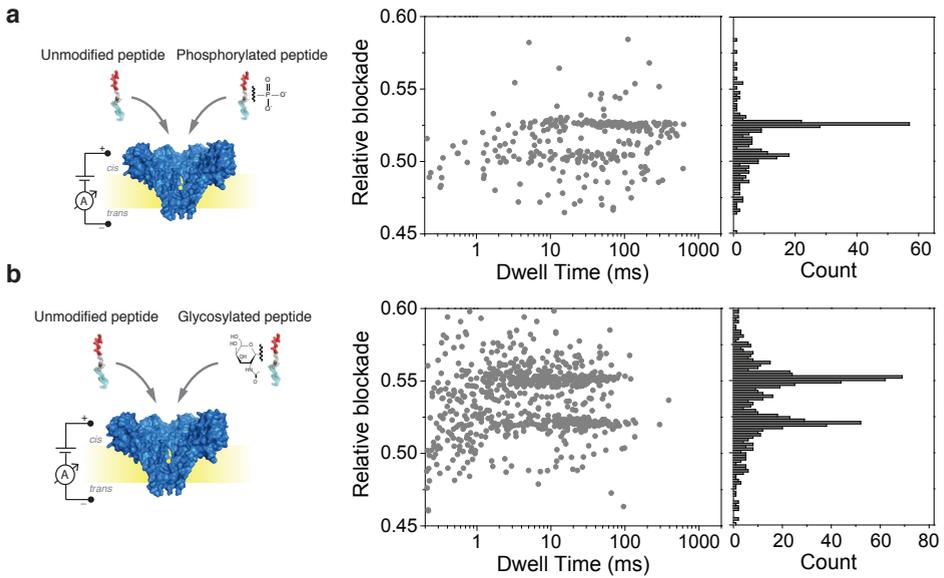


Figure S5.2. (a) Scatter plot of relative blockade vs. dwell time and relative blockade histogram for a mixture containing phosphorylated peptide and unmodified peptide as a reference. (b) Scatter plot of relative blockade vs. dwell time and relative blockade histogram for a mixture containing glycosylated peptide and unmodified peptide as a reference. Measurements were performed in a buffer containing 0.8M NaCl, 10 mM Tris and 1 mM EDTA at pH 7.5.

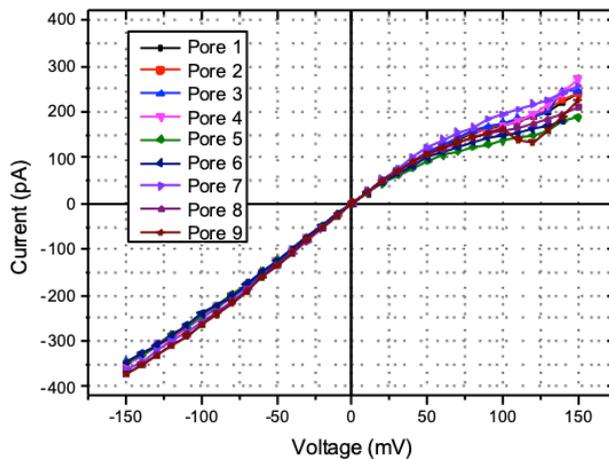


Figure S5.3. Plot of current vs. voltage for 9 wild-type FraC nanopores as measured in a voltage range between -150mV and 150mV, in buffer containing 1M NaCl, 10 mM Tris and 1 mM EDTA at pH 7.5

References

1. Alberts et al. *Molecular biology of the cell*. (Garland Science, 2002). doi:10.1091/mbc.E14-10-1437
2. Aebersold, R. et al. How many human proteoforms are there? *Nat. Chem. Biol.* 14, 206–214 (2018).
3. Ponomarenko, E. A. et al. The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* 2016, 1–6 (2016).
4. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Comput. Biol.* 11, e1004049 (2015).
5. Knorre, D. G., Kudryashova, N. V & Godovikova, T. S. Chemical and functional aspects of posttranslational modification of proteins. *Acta Naturae* 1, 29–51 (2009).
6. Doyle, H. A. & Mamula, M. J. Post-translational protein modifications in antigen recognition and autoimmunity. *Trends Immunol.* 22, 443–449 (2001).
7. Hart, G. W. & Akimoto, Y. *The O-GlcNAc Modification*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2009). doi:NBK1954 [bookaccession]
8. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. & Gong, C.-X. O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* 101, 10804–9 (2004).
9. Pagel, O., Loroch, S., Sickmann, A. & Zahedi, R. P. Current strategies and findings in clinically relevant post-translational modification-specific proteomics. *Expert Rev. Proteomics* 12, 235–253 (2015).
10. Mulagapati, S., Koppolu, V. & Shantha Raju, T. Decoding of O-Linked Glycosylation by Mass Spectrometry. (2017). doi:10.1021/acs.biochem.6b01244
11. Zubarev, R. A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 13, 723–726 (2013).
12. Anderson, N. L. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* 1, 845–867 (2002).
13. A cast of thousands. *Nat. Biotechnol.* 21, 213 (2003).
14. Rusling, J. F., Kumar, C. V, Gutkind, J. S. & Patel, V. Measurement of biomarker proteins for point-of-care early detection and monitoring of cancer. *Analyst* 135, 2496 (2010).
15. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* 452, 571–579 (2008).
16. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 24, 971–983 (2006).
17. Nørregaard Jensen, O. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 8, 33–41 (2004).

18. Dekker, C. Solid-state nanopores. *Nat. Nanotechnol.* 2, 209–215 (2007).
19. Albrecht, T., Gibb, T. & Nuttall, P. in *Engineered Nanopores for Bioanalytical Applications* (Elsevier Inc., 2013). doi:10.1016/B978-1-4377-3473-7.00001-7
20. Goodrich, C. P. et al. Single-Molecule Electrophoresis of beta-Hairpin Peptides by Electrical Recordings and Langevin Dynamics Simulations. *J. Phys. Chem. B* 111, 3332–3335 (2007).
21. Movileanu, L., Schmittschmitt, J. P., Martin Scholtz, J. & Bayley, H. Interactions of Peptides with a Protein Pore. *Biophys. J.* 89, 1030–1045 (2005).
22. Stefureac, R., Long, Y.-T., Kraatz, H.-B., Howard, P. & Lee, J. S. Transport of alpha-Helical Peptides through alpha-Hemolysin and Aerolysin Pores †. *Biochemistry* 45, 9172–9179 (2006).
23. Mohammad, M. M. & Movileanu, L. Excursion of a single polypeptide into a protein pore: simple physics, but complicated biology. *Eur. Biophys. J.* 37, 913–925 (2008).
24. Sutherland, T. C. et al. Structure of Peptides Investigated by Nanopore Analysis. *Nano Lett.* 4, 1273–1277 (2004).
25. Mahendran, K. R., Romero-Ruiz, M., Schlösinger, A., Winterhalter, M. & Nussberger, S. Protein Translocation through Tom40: Kinetics of Peptide Release. *Biophys. J.* 102, 39–47 (2012).
26. Ji, Z. et al. Fingerprinting of Peptides with a Large Channel of Bacteriophage Phi29 DNA Packaging Motor. *Small* 12, 4572–4578 (2016).
27. Payet, L. et al. Thermal unfolding of proteins probed at the single molecule level using nanopores. *Anal. Chem.* 84, 4071–4076 (2012).
28. Pastoriza-Gallego, M. et al. Dynamics of Unfolded Protein Transport through an Aerolysin Pore. *J. Am. Chem. Soc.* 133, 2923–2931 (2011).
29. Oukhaled, G. et al. Unfolding of Proteins and Long Transient Conformations Detected by Single Nanopore Recording. *Phys. Rev. Lett.* 98, 158101 (2007).
30. Restrepo-Pérez, L., John, S., Aksimentiev, A., Joo, C. & Dekker, C. SDS-assisted protein transport through solid-state nanopores. *Nanoscale* 9, 11685–11693 (2017).
31. Talaga, D. S. & Li, J. Single-Molecule Protein Unfolding in Solid State Nanopores. *J. Am. Chem. Soc.* 131, 9287–9297 (2009).
32. Li, J., Fologea, D., Rollings, R. & Ledden, B. Characterization of Protein Unfolding with Solid-state Nanopores. *Protein Pept. Lett.* 21, 256–265 (2014).
33. Freedman, K. J., Haq, S. R., Edel, J. B., Jemth, P. & Kim, M. J. Single molecule unfolding and stretching of protein domains inside a solid-state nanopore by electric field. *Sci. Rep.* 3, 1638 (2013).
34. Dong, Z., Kennedy, E., Hokmabadi, M. & Timp, G. Discriminating Residue Substitutions in a Single Protein Molecule Using a Sub-nanopore. *ACS Nano* 11, 5440–5452 (2017).
35. Kennedy, E., Dong, Z., Tennant, C. & Timp, G. Reading the primary structure of a

- protein with 0.07 nm³ resolution using a subnanometre-diameter pore. *Nat. Nanotechnol.* 11, 968–976 (2016).
36. Fahie, M. A. & Chen, M. Electrostatic Interactions between OmpG Nanopore and Analyte Protein Surface Can Distinguish between Glycosylated Isoforms. *J. Phys. Chem. B* 119, 10198–10206 (2015).
 37. Wloka, C. et al. Label-Free and Real-Time Detection of Protein Ubiquitination with a Biological Nanopore. *ACS Nano* 11, 4387–4394 (2017).
 38. Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* 32, 179–181 (2014).
 39. Asandei, A. et al. Acidity-Mediated, Electrostatic Tuning of Asymmetrically Charged Peptides Interactions with Protein Nanopores. *ACS Appl. Mater. Interfaces* 7, 16706–16714 (2015).
 40. Momotenko, D. et al. Ion current rectification and rectification inversion in conical nanopores: a perm-selective view. *Phys. Chem. Chem. Phys.* 13, 5430 (2011).
 41. Yan, Y., Wang, L., Xue, J. & Chang, H.-C. Ion current rectification inversion in conic nanopores: Nonequilibrium ion transport biased by ion selectivity and spatial asymmetry. *J. Chem. Phys.* 138, 044706 (2013).
 42. Maglia, G., Heron, A. J., Stoddart, D., Japrun, D. & Bayley, H. Analysis of Single Nucleic Acid Molecules with Protein Nanopores. *Methods Enzymol.* 475, 591–623 (2010).
 43. Gutschmann, T., Heimburg, T., Keyser, U., Mahendran, K. R. & Winterhalter, M. Protein reconstitution into freestanding planar lipid membranes for electrophysiological characterization. *Nat. Protoc.* 10, 188–198 (2015).
 44. Plesa, C. & Dekker, C. Data analysis methods for solid-state nanopores. *Nanotechnology* 26, 1–7 (2015).

6

SDS-assisted protein transport through solid state nanopores

Using nanopores for single-molecule sequencing of proteins – similar to nanopore-based sequencing of DNA – faces multiple challenges, including unfolding of the complex tertiary structure of the proteins and enforcing their unidirectional translocation through nanopores. Here, we combine molecular dynamics (MD) simulations with single-molecule experiments to investigate the utility of SDS (Sodium Dodecyl Sulfate) to unfold proteins for solid-state nanopore translocation, while simultaneously endowing them with a stronger electrical charge. Our simulations and experiments prove that SDS-treated proteins show a considerable loss of the protein structure during the nanopore translocation. Moreover, SDS-treated proteins translocate through the nanopore in the direction prescribed by the electrophoretic force due to the negative charge imparted by SDS. In summary, our results suggest that SDS causes protein unfolding while facilitating protein translocation in the direction of the electrophoretic force; both characteristics being advantageous for future protein sequencing applications using solid-state nanopores.

This chapter has been published as: Restrepo-Pérez, L., John, S., Aksimentiev, A., Joo, C. & Dekker, C. SDS-assisted protein transport through solid-state nanopores. *Nanoscale* 9, 11685–11693 (2017).

6.1 Introduction

The proteome contains essential information for the understanding of biological processes and diseases. Despite its importance, only a handful of methods are available for proteome analysis. Mass spectrometry is currently the gold standard for protein sequencing. However, analysis of complex biological samples remains a challenge for the mass spectrometry method, in particular, as many proteins are present in low-copy numbers.¹ Techniques operating at the single-molecule level could provide ultimate sensitivity, enabling detection of proteins of low abundance.

Nanopores have emerged as a platform for fast and label-free detection of single molecules.²⁻⁴ In a nanopore measurement, a thin membrane containing a nanopore is placed between two compartments filled with electrolytes. A voltage bias across the membrane produces a current of ions that flow from one compartment to the other. The presence of a biomolecule in a nanopore is sensed from transient blockades of the ionic current that the biomolecule produces. In the case of linear nucleic acids such as DNA or RNA, the ionic current signal carries information about the nucleotide content of the molecules.⁵⁻¹⁰

Nanopores have also been used to detect proteins. Biological and solid-state nanopores were utilized to detect the presence of proteins bound to DNA,¹¹⁻¹⁴ characterize proteins captured by the nanopore,¹⁵⁻¹⁹ and distinguish different states of proteins.^{16,20-22} A motor protein was used to produce unidirectional protein translocation.²³ First attempts to distinguish the type of individual amino acids in a nanopore were made both experimentally²⁴ and theoretically,²⁵ including detection of post-translational modifications.²⁶

Nanopore-based protein sequencing presents a number of challenges. Proteins, contrary to nucleic acids, do not adopt a random coil conformation but fold into complex tertiary structures. The local charge of a polypeptide chain is non-uniform, alternating between neutral and positive and negative values. The direction of the electrophoretic transport through the nanopore is influenced by both the charge of the protein and the electroosmotic flow in the nanopore.²⁷ As a result, a unidirectional and single-file transport of protein through a nanopore, which is a prerequisite for nanopore-based sequencing, is not ensured.

Detergent treatment of proteins may help achieve better control over the process of protein translocation. Polyacrylamide gel electrophoresis (PAGE) in the presence of sodium dodecyl sulfate (SDS) is a common laboratory technique used in molecular biology to separate proteins according to their size. SDS denatures the proteins and imparts a negative charge to the linearized peptide chains. When SDS-treated proteins are placed on a gel and an electric field is applied, proteins travel to the positive electrode due to the negative

charge imparted by the detergent. These two properties, linearization and uniform charge distribution, are promising to promote linear translocation of proteins through a nanopore. In a previous study, Li and colleagues made a survey of different denaturants that facilitate the translocation of unfolded proteins through nanopores.²⁸ However, no detailed study has been reported on the use of SDS for nanopore translocation of proteins. More recently, Kennedy et al. used SDS to translocate proteins through a sub-nanometer pore, however, no information on the detergent–protein interaction has been provided.²⁹

Here we employ all-atom molecular dynamics (MD) simulations and nanopore experiments to determine how SDS influences the protein transport through nanopores. Our integrated approach elucidates the microscopic conformation and the transport modality of the SDS–protein assemblies. We show that the transport of protein–SDS complexes is determined by electrophoresis, unlike that of native proteins, which is governed by the electroosmotic flow. This is consistent with a prevailing model that SDS imparts uniform negative charges on proteins. SDS-unfolded proteins exhibit a lower current blockade than native proteins do and the blockade duration increases with the molecular weight of the protein. We also explore the use of SDS at different concentrations and show that it can be used both above and below its Critical Micelle Concentration (CMC).

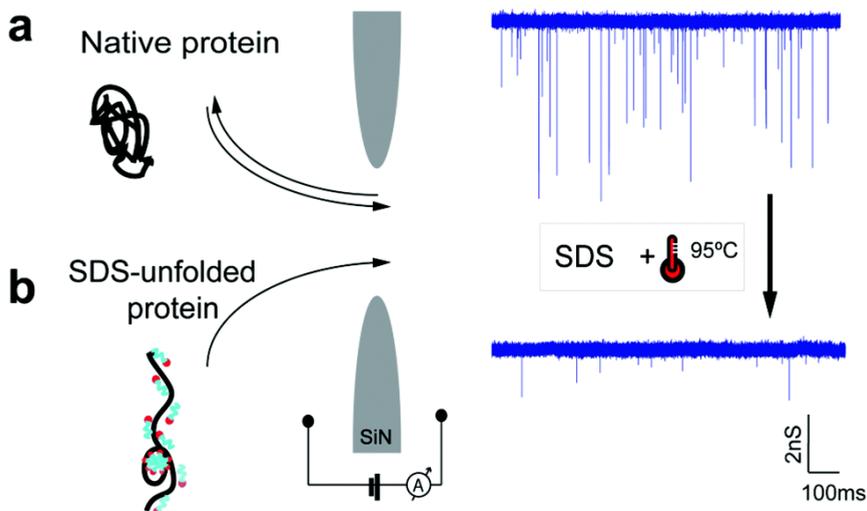


Figure 6.1: Schematic illustration and representative current traces of the translocation of native proteins (a) and SDS-unfolded proteins (b) through a solid-state nanopore. The traces correspond to the translocation of β -amylase through a 10 nm nanopore.

Fig. 6.1 illustrates the concept of protein translocation through a nanopore. We focused our investigation on three different proteins: titin I27 domain (13 kDa, a well-characterized protein often used in protein unfolding studies), titin I27 dimer (26 kDa), and β -amylase (200

kDa, a globular protein with a high molecular weight). We performed the majority of our measurements and simulations using silicon nitride (SiN) nanopores 6 or 10 nm in diameter and an electrolyte solution containing 0.4 M NaCl. Higher electrolyte concentrations or potassium-based buffers caused SDS precipitation in experiments.

6.2 Molecular dynamics simulations

We explored the effects of SDS treatment on proteins and its consequences for nanopore translocation. First, we obtained microscopic models of SDS/protein complexes by simulating spontaneous self-assembly of unfolded proteins with SDS molecules in the electrolyte solution (0.4 M NaCl). In a typical simulation (Fig. 6.2a), SDS molecules were observed to aggregate around an unfolded protein at boiling temperature (373 K) forming a complex that contains approximately one SDS molecule per two amino acids of the protein, in agreement with previous experimental findings.³⁰ The complex remained stable upon subsequent cooling to room temperature. Fig. 6.2b shows examples of final room temperature SDS/protein complexes obtained from such self-assembly simulations. The structures of such SDS/protein assemblies appear to be in accord with a decorated micelle model³⁰ in which the protein wraps around multiple micelle-like agglomeration of SDS molecules.

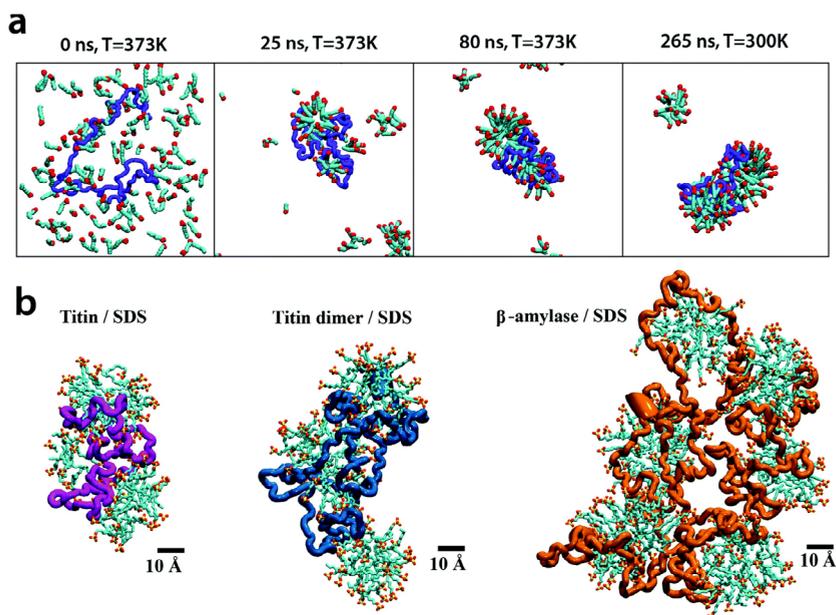


Figure 6.2: Microscopic models of SDS/protein complexes. (a) MD simulation of a titin/SDS complex self-assembly. The sequence of snapshots illustrates the microscopic state of the system during a 300 ns explicit solvent all-atom MD simulation. The protein backbone is shown in blue, SDS molecules in cyan and red, and water and 0.4 M NaCl are not shown. (b) Typical microscopic conformations of the SDS/protein assemblies

obtained at the end of the self-assembly simulations. The SDS molecules are shown using the molecular bond representation: carbon, sulfur and oxygen atoms are shown in cyan, yellow and red, respectively; hydrogen atoms are not shown. Each protein molecule is shown as a trace of the protein backbone.

To simulate nanopore translocation (Fig. 6.3), atomic models of different substrates (folded proteins, SDS micelles, and SDS–protein complexes) were combined with atomic models of solid-state nanopores; water and ions were added to produce a rectangular volume of NaCl solution. An external electric field was applied normal to the membrane to produce a transmembrane bias of desired magnitude. Subject to the transmembrane bias, biomolecules move through a nanopore altering the nanopore ionic current. SI Table S1† provides a summary of nanopore translocation simulations.

Our simulations of a folded protein (titin) translocation (Fig. 6.3a) demonstrate that a protein's encounter with a nanopore is hardly determined by the electrophoretic force exerted due to the transmembrane bias. In five independent simulations of the folded titin system (60 ns each; ± 500 mV bias), folded titin rarely translocated through a 6 nm wide nanopore (SI Fig. S6.1). We ascribe this behaviour to the low net charge of native titin at pH 7.5. In the only simulation where a protein translocation event was observed, the protein was seen moving through the nanopore in the direction set by the electro-osmotic flow.

In contrast, the translocation of SDS–protein assemblies was clearly governed by electrophoresis (Fig. 6.3b–d). A complex of SDS molecules and either monomeric or dimeric titin permeated through a 6.5 nm diameter nanopore multiple times, producing sharp spike-like blockades. The SDS–titin assembly maintained its integrity. During the translocation, the SDS–titin assemblies noticeably deformed in the center of the nanopore, particularly in the case of the titin dimer. The protein–SDS assembly relaxed back to a compact globular state after each translocation. Repeating the simulations starting from a differently orientated SDS–titin assembly (Fig. 6.3b and c) or at a lower transmembrane bias (SI Fig. S6.2) produced similar outcomes.

Pronounced deformations of the SDS–protein assemblies were observed during the simulations of the translocation of large SDS-unfolded β -amylase complexes through a 6.5 nm diameter nanopore (Fig. 6.3d and e). The translocation of the SDS– β -amylase assembly began when two or more peptide chains detached from the assembly, entering the nanopore in a “multi-chained” conformation. The peptide chains stretched through the nanopore, pulling the rest of the protein in a train-like fashion, reminiscent of DNA translocation. Fig. 6.3e illustrates one of the permeation trajectories. Despite such major conformational changes, the number of SDS molecules associated with the protein remained constant during the nanopore translocation. Note that the physical size of β -amylase in its native folded conformation (~ 10 nm) is greater than the 6.5 nm diameter of the simulated

nanopore.

SDS micelles alone also translocated through the nanopore in the direction prescribed by the electrophoretic force. Each permeation event produced a pronounced reduction of the ionic current (Fig. 6.3f). The shape of the ionic current blockades—a downward spike—was consistent with that produced by a spherical particle passing through a larger pore.³¹ During the translocation process, the micelle was observed to deform slightly, recovering its spherical shape after each passage. At all transmembrane biases tested, the SDS micelle maintained its integrity with no SDS molecules leaving the micelle.

Fig. 6.3g plots the average duration of the translocation events for different molecular assemblies and under several transmembrane bias conditions. At a given transmembrane bias, SDS micelles moved through the nanopore considerably faster than SDS–protein assemblies. Predictably, the translocation of an SDS–titin monomer complex occurred faster than that of an SDS–titin dimer complex. Both complexes moved through the nanopore considerably faster than the large SDS– β -amylase complex, despite a stronger bias used in the simulation of the latter system. These trends suggest that the translocation time increases with the molecular weight of the protein.

The translocation of different SDS–protein assemblies and SDS micelles through a 6 nm diameter nanopore produced conductance blockades of similar amplitudes (Fig. 6.3h). This result is expected since all molecular assemblies had a similar cross sectional area in the narrowest part of the pore when permeating through a 6 nm diameter nanopore. By contrast, considerably different blockade amplitudes were observed when the same molecular assemblies were placed at the center of a larger, 10 nm diameter nanopore (Fig. 6.3h and SI Fig. S6.3). This observation indicates that the amplitude of nanopore blockades produced by a compact object, such as a folded protein or an SDS micelle, depends on the nanopore size.³¹

Encouraged by our observation of SDS-treated β -amylase translocation through a nanopore that was smaller in cross-section than the folded β -amylase protein, we investigated if a titin–SDS complex could be electrophoretically driven through a 3 nm-diameter nanopore (System 3, see Materials and methods). The steric confinement of such a narrow nanopore was expected to produce single-file translocation of the SDS-treated protein. However, no fragment of the titin protein was observed to enter the 3 nm nanopore within multiple 80 ns simulations performed at 125 and 250 mV biases despite being placed in close proximity to the nanopore entrance. Increasing the transmembrane bias to 1 V resulted in stripping of the SDS molecules away from the titin–SDS assembly but no protein translocation through the nanopore. A similar outcome was observed when the temperature of the system was

increased to 373 K.

Summing up, our MD simulations show that SDS treatment provides proteins an overall negative charge, thereby enabling the electrophoretic transport through nanopores. Subject to the electric field and steric forces from the nanopore, the SDS-treated proteins deform and permeate through nanopores that are smaller in diameter than the unperturbed dimensions of the folded protein. The depth of ionic current blockades produced by the permeation of SDS–protein assemblies depends on both the diameter of the nanopore and the conformation of the protein. Realizing single-file translocation of a protein/SDS complex may require pre-stretching of the complex prior to reaching the nanopore entrance.

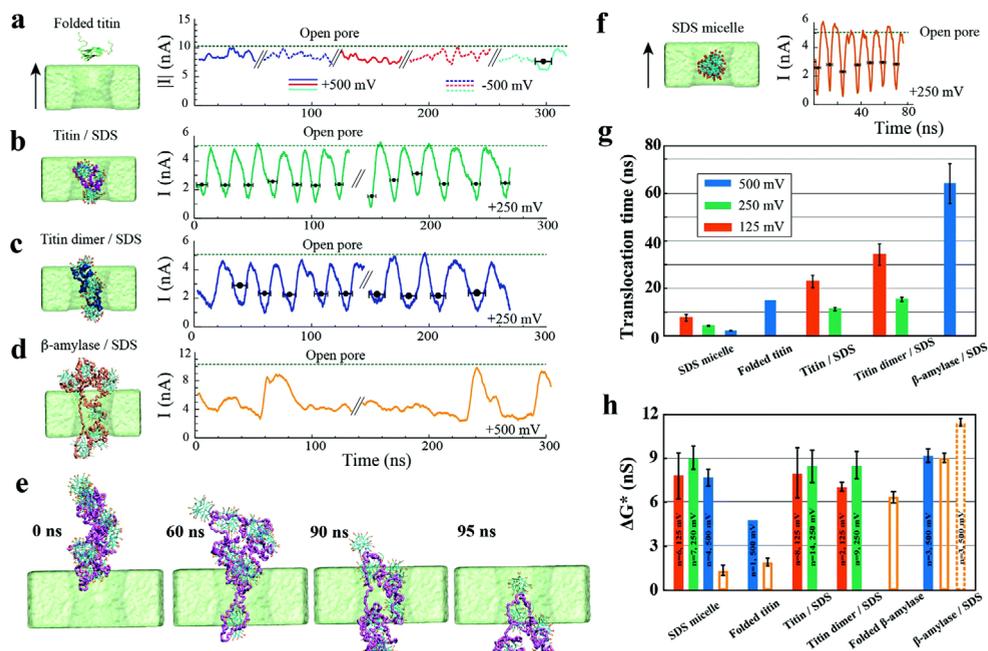


Figure 6.3: MD simulation of SDS/protein nanopore translocation. (a–d) Simulated ionic current blockades produced by translocation of various molecular species through a 6 nm diameter nanopore (System 1). The left column illustrates typical microscopic conformations observed during nanopore translocation simulations; the black arrow indicates the direction of the positive transmembrane bias. Protein conformation is depicted as a trace of the protein backbone; SDS molecules are shown as molecular bonds; water and ions are not shown for clarity. Right columns show the ionic current traces recorded from the translocation simulations and the open pore current level. Because of the periodic boundary conditions employed in our MD simulations, individual ionic current traces feature multiple translocation events. Ionic current traces from independent simulations are delineated by the “//” mark. Black circles and horizontal black bars indicate the average blockade current of individual blockade events and the duration of each event, respectively. With the exception of native titin, the blockade events were defined by the reduction of the nanopore current below 75% of the open pore value; the blockade event in the native titin simulation was characterized using the titin’s center of mass coordinates. (e) Sequence of snapshots illustrating the nanopore translocation of a β -amylase/SDS complex. (f) Same as in panels (a–d) but for an SDS micelle. (g) Average translocation time of the SDS/protein assemblies. The color of the bars indicates the transmembrane voltage. (h) The average conductance blockade amplitudes. To enable direct comparison with experiment,

the conductance blockades computed from MD simulations were scaled by the ratio of the experimental and simulation bulk conductivity of 0.4 NaCl (3.6/4.8). Open bars indicate conductance blockades obtained from the molecular assemblies placed at the center of a 10 nm diameter nanopore (System 2) at a 500 mV bias; the molecular configurations are defined in SI Fig. S6.3.

6.3 Single-molecule experiments

We experimentally measured the conductance blockades of native proteins and SDS-unfolded proteins (Fig. 6.1). We performed all our measurements using SiN nanopores with a diameter of 10 nm at 100 mV in an electrolyte solution containing 0.4 M NaCl, buffered to pH 7.5 with 10 mM Tris-HCl and 1 mM EDTA. SDS-denatured proteins were heated up to 95 °C for 5 min with 1 mM DTT. DTT was used to disrupt and prevent the formation of disulphide bonds. This procedure guaranteed protein denaturation facilitating SDS interaction. The denatured proteins were cooled down before measurement. Note that a denaturant is necessary to guarantee irreversible protein unfolding, which cannot be guaranteed by temperature on its own. Our experiments with native titin revealed that folded proteins translocate preferentially in the direction of the electro-osmotic flow, consistent with the MD simulations (Fig. 6.3a) and also reported by others.²⁷ SDS–protein complexes, however, move in the direction set by the electrophoretic force, directly verifying the simulations.

To elucidate the mechanism of the altered translocation direction caused by SDS, we analyzed the event rate of titin for translocations at several different SDS concentrations, Fig. 6.4a. In the regime of low SDS concentrations, below the CMC, we observed a high event rate in the direction of the electro-osmotic flow (Fig. 6.4b). At SDS concentrations above the CMC, however, the event rate associated with the electrophoretic force was dominant. This result is in agreement with the direction of translocation observed in our simulations, which were carried out at SDS concentrations above the CMC. Using dynamic light scattering, we estimated the CMC under our experimental conditions to be between 0.01% and 0.05% (SI Note S6.1 and Fig. S6.4).

Our observation that electrophoresis dominates at high SDS concentrations is consistent with previous studies on protein unfolding. Above the CMC, SDS forms micelles and cooperatively binds to the polypeptide chain.^{32–34} Cooperative binding of SDS above the CMC imparts a strong negative charge to the denatured protein, thus increasing the electrophoretic force acting on it. Below the CMC, SDS monomers individually interact with the polypeptide via hydrophobic and electrostatic interactions.

When measurements were performed above the CMC, the translocation of SDS micelles was observed. We performed control experiments in which we characterized these trans-

locations at different SDS concentrations in the absence of any protein (Fig. 6.4c). SDS micelles moved towards the positive electrode, as expected from their negative charge. Micelles produced well-defined conductance blockades of 3.5 ± 0.5 nS with low event rates (only ~ 2 events per s at 0.5% SDS, equivalent to 17 mM) and unusually long dwell times (~ 1.2 ms) (Fig. 6.5). From these characteristics, we speculate that the observed events are from those micelles that interact strongly with the SiN surface in the pore lumen, whereas micelles that do not interact with the pore surface pass through the nanopore much too quickly to be detected by the ionic current measurement. The low event rate (Fig. 6.4c) indicates that the SDS micelles alone contribute only to approximately 25% of the event rates presented in Fig. 6.4b.

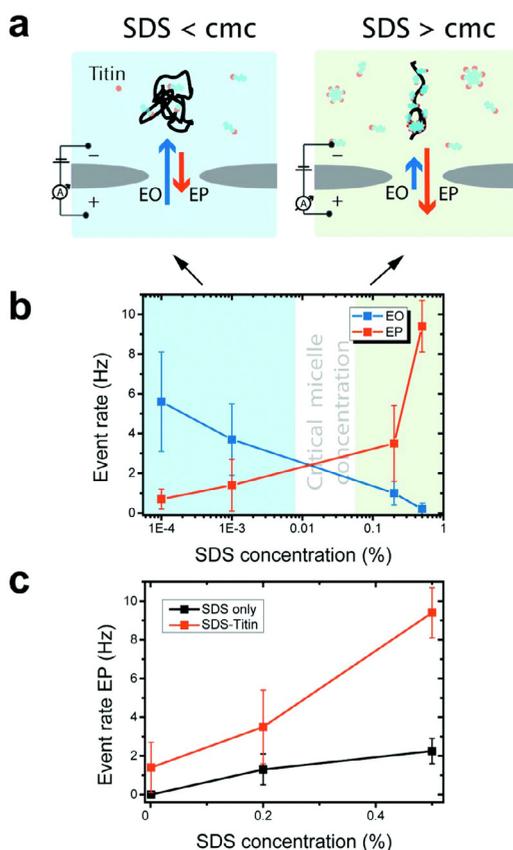


Figure 6.4: (a) Schematic of the dominant mechanisms of protein translocation at SDS concentrations below and above the CMC. (b) Event rates of the translocation of titin–SDS complexes through a 10 nm solid-state nanopore at different SDS concentrations. At SDS concentrations below the CMC (blue) electro-osmosis is the dominant transport mechanism of titin–SDS through solid-state nanopores. At SDS concentrations above the CMC (green) electrophoresis dominates. (c) Plot of the event rate vs. SDS concentration for SDS micelles alone and SDS-unfolded proteins.

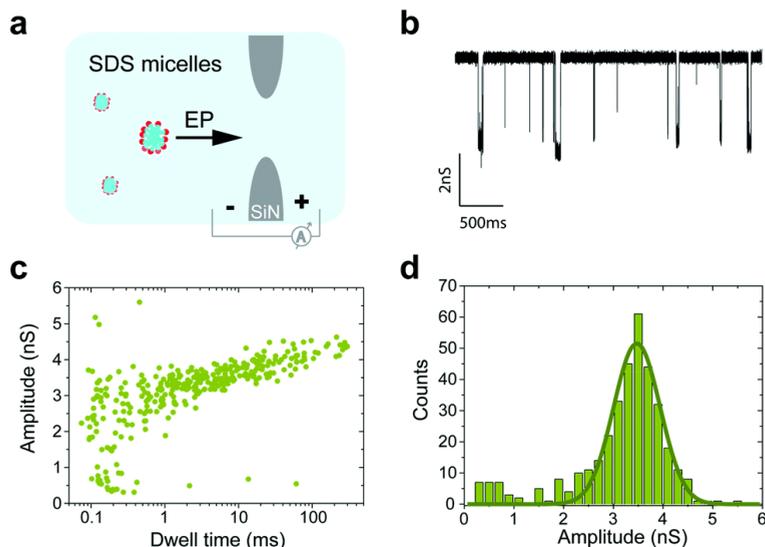


Figure 6.5: Translocation of SDS micelles through a 10 nm solid-state nanopore. (a) Schematic representation of the nanopore control experiment. (b) Example of a current trace measured showing the translocation of SDS micelles. (c) Dwell time vs. conductance blockade scatter plot for SDS-micelles and SDS-treated protein. (d) Conductance blockade histogram of the translocation of micelles.

Fig. 6.6 presents a key result of our study. It shows that we can distinguish folded and SDS-induced unfolded proteins in nanopore translocation experiments. Fig. 6.6a shows examples of translocations of β -amylase proteins, in which it is clear that SDS-treated proteins exhibit a much lower conductance blockade level than native proteins. This is evidently observed in a scatter plot of conductance blockade vs. dwell time for native and SDS-unfolded β -amylase (Fig. 6.6b). The translocation of native proteins was measured in the direction prescribed by the electro-osmotic flow, whereas SDS-treated proteins were found to translocate in the direction of the electrophoretic force. These measurements were performed at low SDS concentrations (0.005%) below the CMC to unambiguously ensure that the observed blockades are due to SDS-protein complexes, and not due to SDS micelles. We verified that SDS disrupts the structure of β -amylase even at SDS concentrations below the CMC using tryptophan fluorescence measurements (SI Note S6.2 and Fig. S6.5).

Our data suggest that SDS treatment causes significant protein unfolding, as also observed in our simulations. As shown in the scatter plot (Fig. 6.6b) and the histograms presented in Fig. 6.6c (left and middle), a pronounced decrease in the amplitude of the blockade was observed when β -amylase was treated with SDS, from 3.7 ± 1.7 nS (mean \pm standard deviation) for folded proteins to 0.53 ± 0.29 nS for SDS-unfolded proteins. This significant decrease in the conductance blockade is consistent with the protein adopting an elongated structure of a smaller cross section than its native folded conformation. Note that for the

folded β -amylase the conductance values are lower than expected, presumably due to the well-known bandwidth limitations when measuring protein translocations.¹⁸

We also carried out measurement of SDS-treated β -amylase at an SDS concentration above the CMC. As Fig. 6.6c (right panel) shows, a low conductance blockade peak (at ~ 0.4 nS) persisted, while an additional blockade peak appeared at 4.26 ± 0.37 nS. The presence of the first peak suggests that a population of SDS-associated proteins remains, which translocates in an elongated fashion without binding to micellar SDS. The additional blockade can be attributed to SDS micelles.

In our measurements with titin, no appreciable difference was observed between the native and the SDS-denatured protein, as both gave ~ 0.4 nS blockades (SI Fig. S6.6). This is explained by the small size of the protein, which even in the native conformation produces blockades of ~ 0.4 nS, a value that unfortunately coincides with the low value expected for the SDS-unfolded protein. Although the conductance blockade of titin did not change significantly upon treatment with SDS, the change in the direction of translocation (Fig. 6.4) clearly indicates the interaction of SDS with the protein.

The conductance blockades for SDS-unfolded β -amylase and SDS-unfolded titin proteins are comparable (0.40 ± 0.17 nS for titin and 0.49 ± 0.18 nS for β -amylase), suggesting that SDS-denatured proteins translocate through a nanopore with a similar cross-section regardless of the protein molecular weight. For the SDS-treated β -amylase, most of our simulations show a deeper conductance blockade (~ 9 nS) compared to our experiments. However, note that, in most of our simulations, the protein-SDS complex adopts a “decorated micelle” structure (see Fig. 6.2b) and, hence, multiple polypeptide chains simultaneously translocate through the nanopore when driven by a transmembrane bias. Because translocation of such decorated micelle structures is very fast (at the sub-microsecond time scale), they would likely escape experimental detection because of the bandwidth limitations.¹⁸ By contrast, events measured in the experiments suggest that the protein adopts an elongated conformation, given the low conductance blockade measured. Indeed, when a single polypeptide chain was simulated in a 10 nm pore (SI Fig. S6.7), a much lower conductance blockade value of ~ 2 nS was observed, regardless of whether SDS molecules were bound to the peptide chain or not. The simulated and experimentally measured ionic current blockades are in much better agreement in the case of pure SDS micelle systems, suggesting that experiment and simulation characterize similar molecular configurations of biomolecules within the nanopore. Thus, the low conductance blockades and the relatively slow translocation kinetics observed in the experiments are consistent with the notion that the translocation of protein/SDS complexes is affected by interactions with the nanopore surface.

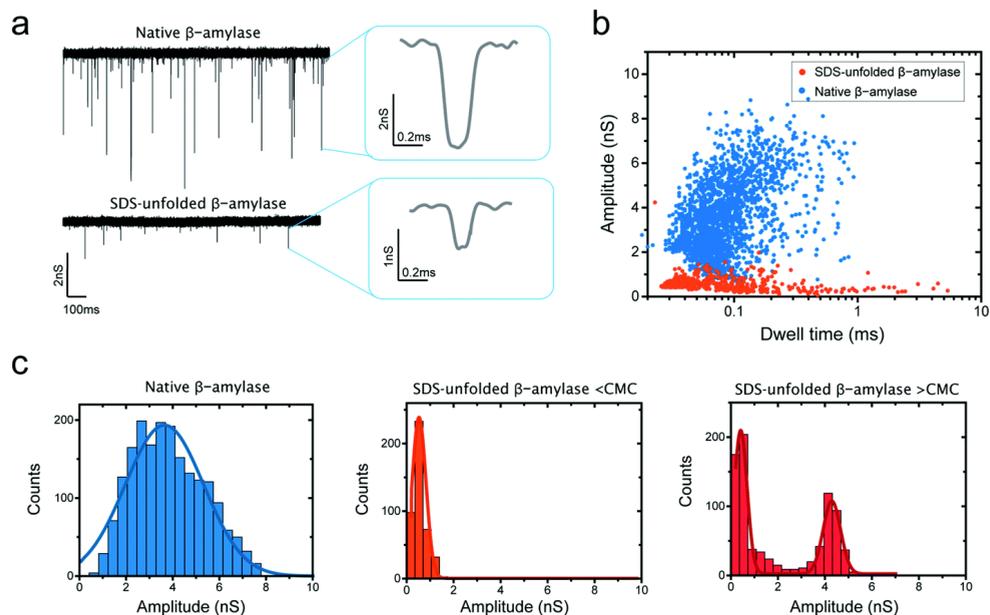


Figure 6.6: (a) Current traces produced by the translocation of native β -amylose and SDS-unfolded β -amylose through a 10 nm pore at 10 kHz bandwidth. Typical events are shown in the right panels. (b) Scatter plot of the conductance blockade vs. dwell time of native β -amylose and SDS-treated β -amylose. Native protein translocations are measured in the direction of the electro-osmotic flow, SDS-treated proteins are measured in the electrophoretic direction. (c) Histograms of conductance blockades of native protein (left), SDS-treated protein below the CMC (middle), and SDS-treated protein above the CMC (right). The conductance blockade of the native protein has a value of 3.7 ± 1.7 nS. The blockade for the SDS-treated protein below the CMC is 0.53 ± 0.29 nS. Above the CMC, two peaks are observed. The first has a value of 0.40 ± 0.27 nS, and the second one, presumably due to SDS micelles, has a value of 4.3 ± 0.4 nS.

To summarize, our experimental results suggest that SDS treatment of proteins has drastic effects on the process of protein translocation through solid-state nanopores: the treatment reverses the translocation direction and can reduce the conductance blockade. At high SDS concentrations, proteins preferentially translocate by electrophoretic forces as shown both experimentally and computationally. SDS causes protein unfolding, which can present the polypeptide chain in an elongated conformation to the nanopore volume. For example, the measured conductance blockade of SDS-unfolded β -amylose is clearly lower than the blockade produced by that protein in its folded conformation. The low conductance blockade level is consistent with the conductance blockade observed in our MD simulations of a single peptide chain threaded through the nanopore. This correspondence suggests that single-chain translocations are being experimentally detected in the case of SDS-treated proteins. Further studies are necessary to verify this hypothesis.

6.4 Conclusions

Through nanopore measurements combined with MD simulations, we have shown that upon SDS treatment, proteins can translocate through nanopores in an unfolded fashion. Additionally, our simulations show that the dwell time of the SDS-treated proteins in a nanopore increases with the protein molecular weight. These results suggest that a nanoscaled single-molecule version of SDS-PAGE, in which proteins are differentiated according to their translocation time, could be created using solid-state nanopores. Our study also shows that proteins move through the nanopore by electrophoresis, overcoming the electroosmotic flow that runs in the reverse direction. This can be attributed to the enhanced negative charge imparted to the protein by the SDS sulfate groups.

We also explored a wide range of SDS concentrations to show that measurements are feasible both above and below the CMC. Characterization of the micelles produced by SDS above the CMC is also presented.

Altogether, our results suggest that SDS causes protein unfolding while facilitating protein translocation in the direction of the electrophoretic force, both characteristics could be advantageous for future protein sequencing applications using solid-state nanopores.

6

6.5 Materials and methods

6.5.1 Nanopore fabrication

20 nm-thick freestanding silicon nitride membranes were fabricated using trans-chip illumination lithography as previously described by Janssen et al.³⁵ TEM is used to drill nanopores on the membranes and then the chips are manually PDMS painted to reduce the membrane capacitance and improve the signal-to-noise ratio.³⁶

6.5.2 Nanopore measurements

Before each measurement, the nanopores are cleaned using oxygen plasma and mounted in a PEEK flow cell. Buffer containing 0.4 M NaCl, 10 mM Tris pH 7.5 and 1 mM ETDA is used for all measurements. SDS is added at different concentrations, as indicated. Measurements are recorded at 100 kHz using an Axopatch 200B (Molecular Devices, LLC) and digitized at 500 kHz using a National instruments Digidata 1322A DAQ. The recorded signals were further low-pass filtered at 10 kHz and analyzed using the Transalyzer Matlab package.³⁷

6.5.3 Protein substrates

Titin I27 V13P containing an N-terminal His6-tag and a C-terminal SsrA tag was purified by affinity chromatography as follows. A 1-L culture of *E. coli* strain BL21ai containing the plasmid pRSETa was incubated at 37 °C in LB medium supplemented with ampicillin. At an OD₆₀₀ of ~0.8, arabinose (0.2%) was added and the culture was grown for an additional 4 h at 37 °C before harvesting the cells and freezing them at –80 °C. Pellets were re-suspended in lysis buffer (50 mM of sodium phosphate pH 8.0, 300 mM of NaCl, 50 mM of imidazole) and lysed by sonication. The lysate was centrifuged (17 000 rpm, 30 min) and titin was purified from the supernatant using Ni²⁺-NTA affinity resin and eluted with imidazole (200 mM). The sample was dialyzed overnight at 4 °C against PBS pH 7.4. β-Amylase from sweet potato was purchased from Sigma-Aldrich.

6.5.4 General MD methods

All MD simulations were performed using the NAMD2 software package,³⁸ periodic boundary conditions, a 2 fs integration time step and a multiple time-stepping integration scheme.³⁹ Proteins, SDS molecules, water and ions were described using the CHARMM36 force field.⁴⁰ A custom CHARMM-compatible force field was used for SiO₂⁴¹ along with custom NBFIX corrections to non-bonded interactions of sodium ions and the sulphate groups of SDS or acetate groups of the proteins.^{42,43} SETTLE⁴⁴ and RATTLE⁴⁵ algorithms were applied to the covalent bonds involving hydrogen atoms in water and protein, respectively. The van der Waals forces were cut off smoothly starting at 10 Å and cut off completely at 12 Å. Long range electrostatic interactions were evaluated every third time step using the particle mesh Ewald method⁴⁶ over a 0.11 nm-spaced grid. The temperature was controlled using a Langevin thermostat acting on the membrane's atoms with a damping constant of 0.5 ps⁻¹. In the NPT (i.e. constant number of particles, pressure, and temperature) simulations, Nosé–Hoover Langevin piston⁴⁷ maintained the pressure at 1 atm by adjusting the systems' dimensions along the direction normal to the membrane (z axis in our setup). In all production simulations of the nanopore systems, atoms of the inorganic membrane were harmonically restrained to their coordinates attained at the end of the annealing simulation with the force constant of 20 kcal mol⁻¹ Å⁻².

6.5.5 Atomic-scale models of SDS–protein assemblies

Computational models of SDS–protein assemblies were prepared by combining an unfolded conformation of a protein with randomly distributed SDS molecules in a 1 : 1 amino acid/SDS molecule ratio. The systems were submerged in 0.4 M NaCl solution. Following 10 000 steps of energy minimization and 1.5 ns of equilibration in an NPT ensemble, the

systems were simulated at 373 K over the course of 200 ns. Following that, the systems were cooled down to 300 K in eight simulation steps, decreasing the temperature of the system by 10 K between each step. Following that, the SDS–protein systems were simulated for another 50 ns at 300 K. SDS molecules that were not in contact with the protein at the end of the cooling phase were removed from the systems. The remaining SDS–protein assembly was used for the nanopore translocation simulations.

6.5.6 Atomic-scale models of nanopore systems

The atomic-scale models of solid-state nanopores were obtained by simulated annealing of amorphous silica subject to a shape-defining grid-based potential.⁴⁸ The initial models of the nanopores were made by removing atoms from a slab of a crystalline SiO₂ membrane. The systems were then simulated using the BKS⁴⁹ force field in a vacuum at 7000, 5000, 2000, and 300 K for 160, 160, 400, and 400 ps, respectively, and in the presence of a grid-based potential⁵⁰ that confined the atoms to a predefined volume. The nanopores produced using this method are electrically neutral but contain negatively charged oxygen atoms at the nanopore surface, which makes the nanopore ionic current slightly cation selective.⁵¹ Three nanopore models were built, differing by the dimensions of the SiO₂ membrane (15 × 15 nm² cross-section/6 nm thickness, System 1; 15 × 15 nm² cross-section/10 nm thickness, System 2 and 10 × 10 nm² cross section/5 nm thickness, System 3). Each membrane contained a single hourglass-shaped nanopore with the constriction/entrance diameters of 6/7.5 (System 1), 10/13 (System 2) or 3/4.5 (System 3) nm, respectively. The nanopore models produced by the annealing procedures were merged with the atomic-scale models of the biomolecules and solvated with pre-equilibrated volumes of water. Sodium and chloride ions were added to produce a 0.4 M solution; the precise number of ions was chosen to make the entire system electrically neutral.

Following the assembly, the systems were minimized using a conjugate gradient method and equilibrated in the NPT ensemble for 2 ns each at 1 atm pressure and 295 K temperature. During the equilibration, the protein and/or SDS atoms were restrained to their initial coordinates with a force constant of 10 kcal mol⁻¹ Å⁻² for 1 ns; then the atoms were restrained with a force constant of 5 kcal mol⁻¹ Å⁻² for another 1 ns. All subsequent simulations were performed in the presence of a grid-based potential that prevented a direct contact interaction between protein and SDS atoms with the nanopore surface,⁵⁰ thereby eliminating non-specific binding; the grid potential did not apply to water molecules or ions. All simulations of the nanopore systems under applied electric field conditions were carried out in the NVT ensemble (i.e. constant number of particles, volume, and temperature). The center of mass of the biomolecules was restrained to remain at the symmetry axis of the nanopore using steered molecular dynamics (SMD) features of NAMD2. Such

restraints allowed the SDS protein assembly to undergo considerable conformational transformations and move through the nanopore while maintaining its center of mass aligned with the geometrical center of the nanopore. A transmembrane potential V was induced by applying a constant electric field $E = -\Delta V/l_z$ normal to the membrane, where l_z is the length of the simulation system normal to the membrane. SI Table S6.1 provides a complete list of simulations performed. Open pore systems were built and simulated using the same procedure as described above. Atomic coordinates were recorded every 9.6 ps.

6.6 Supplementary information

Nanopore and membrane dimensions	Biomolecules	Number of atoms	Transmembrane bias (mV)	Simulation time (ns)	
System 1: Membrane cross section: 15 nm x 15 nm Membrane thickness: 6 nm Inner pore diameter: 6 nm Outer pore diameter: 7.5 nm	SDS micelle (70 SDS molecules)	322,601	125	130	
			250	80	
			500	20	
	Folded titin, conf. 1	330,898	500	60	
			-500	60	
	Folded titin, conf. 2	338,624	500	60	
			-500	60	
				500/-500	20/40
	SDS/titin, conf. 1 (76 SDS molecules)	340,937	125	160	
			250	130	
SDS/titin, conf. 2 (76 SDS molecules)	344,343	125	250		
		250	125		
SDS/titin dimer, conf. 1 (104 SDS molecules)	395,308	125	160		
		250	120		
SDS/titin dimer, conf. 2 (104 SDS molecules)	416,979	125	140		
		250	120		
		500	110		
SDS/ β -amylase, conf. 1 (249 SDS molecules) SDS/ β -amylase, conf. 2 (249 SDS molecules) SDS/ β -amylase, conf. 3 (249 SDS molecules)	568,521	500	120		
		500	160		
		500	160		
System 2: Membrane cross section: 15 nm x 15 nm Membrane thickness: 10 nm Inner pore diameter: 10 nm Outer pore diameter: 13 nm	SDS micelle (70 SDS molecules)	290,558	500	50	
	Folded titin	291,193	500	50	
	Folded β -amylase	290,155	500	50	
	SDS/ β -amylase, conf. 2 (249 SDS molecules)	351,622	500	50	
	SDS/ β -amylase, orient. 1 (249 SDS molecules)	411,720	500	50	
	SDS/ β -amylase, orient. 2 (249 SDS molecules)	389,398	500	50	
SDS/ β -amylase, orient. 3 (249 SDS molecules)	358,543	500	50		
System 3: 10x10 nm ² membrane 5 nm thickness Inner/outer pore diameter: 3/4.5 nm	SDS/titin, conf. 1 (76 SDS molecules)	163,621	125 250 1V	100 80 40	
	SDS/titin, conf. 2 (76 SDS molecules)	159,721	125 250	40 40	

Table S6.1: Summary of MD simulations of nanopore systems.

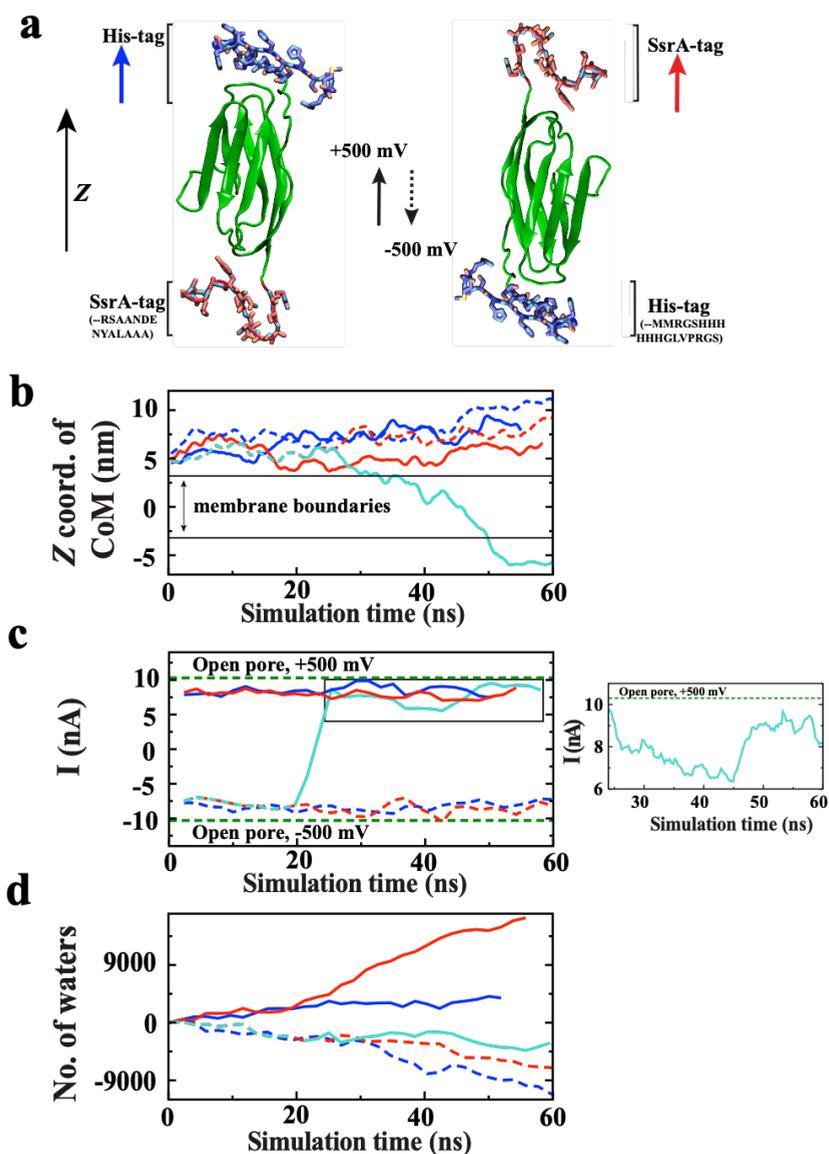


Figure S6.1: MD simulation of folded titin translocation. (a) Molecular structure of the folded titin monomer. To reproduce the constructs used in experiment, SsrA and His tags were added to the C and N terminals of the protein, respectively. The blue and red arrows define the initial orientation of the protein with respect to the z-axis of our setup (shown in black), which is also the direction of a positive transmembrane bias. (b-d) MD simulation of folded titin translocation through a 6 nm diameter nanopore (System 1). (b) The z coordinate of the titin's center of mass as a function of simulation time. The horizontal black lines indicate the location of the membrane. The color and the style (solid/dashed) of the lines indicate the initial orientation of the protein and the direction of the transmembrane bias, respectively, defined in panels a. In one simulation (gray), the bias was switched during the MD simulation. Each translocation trace shows a 1-ns running

average of 4.8 ps sampled data. (c) The nanopore ionic current recorded in the simulations of folded titin translocation. Each ionic current trace shows a 5 ns running average of 4.8 ps sampled instantaneous current. The inset on the right shows a zoomed-in view of the ionic current trace for the fragment of the MD trajectory where protein translocation was observed (black rectangle in the main plot). In the inset, the ionic current trace shows 1 ns running average of 4.8 ps sampled instantaneous currents. (d) The total number of water molecules that passed through the nanopore from the beginning of the applied bias simulation. In this panel, positive values indicate net water flux in the direction of the z axis. The direction of the water flux is correlated with the sign of the transmembrane bias, indicating an electro-osmotic flow produced by a negatively charged nanopore surface.

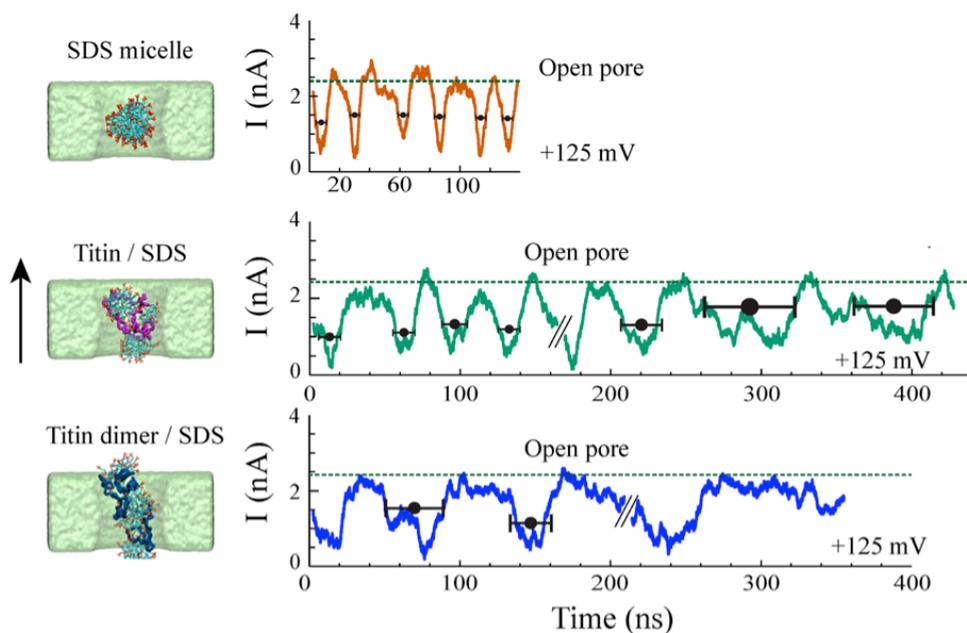


Figure S6.2: Simulated ionic current blockades produced by translocation of molecular species through a 6 nm diameter nanopore. Left column illustrates typical microscopic conformations observed during nanopore translocation simulations; the black arrow indicated the direction of the positive transmembrane bias. The conformation of a protein is depicted as a trace of the protein backbone; SDS molecules are shown as molecular bonds; water and ions are not shown for clarity. Right columns shows the ionic current traces recorded from the translocation simulations and the open pore current levels. Because of the periodic boundary conditions employed in our MD simulations, individual ionic current traces feature multiple translocation events. Ionic current traces from independent simulations are delineated by the “//” mark. Black circles and horizontal black bars indicate the average blockade current of individual blockade events and the duration of each event, respectively. The blockade events were defined by the reduction of the nanopore current below 75% of the open pore value.

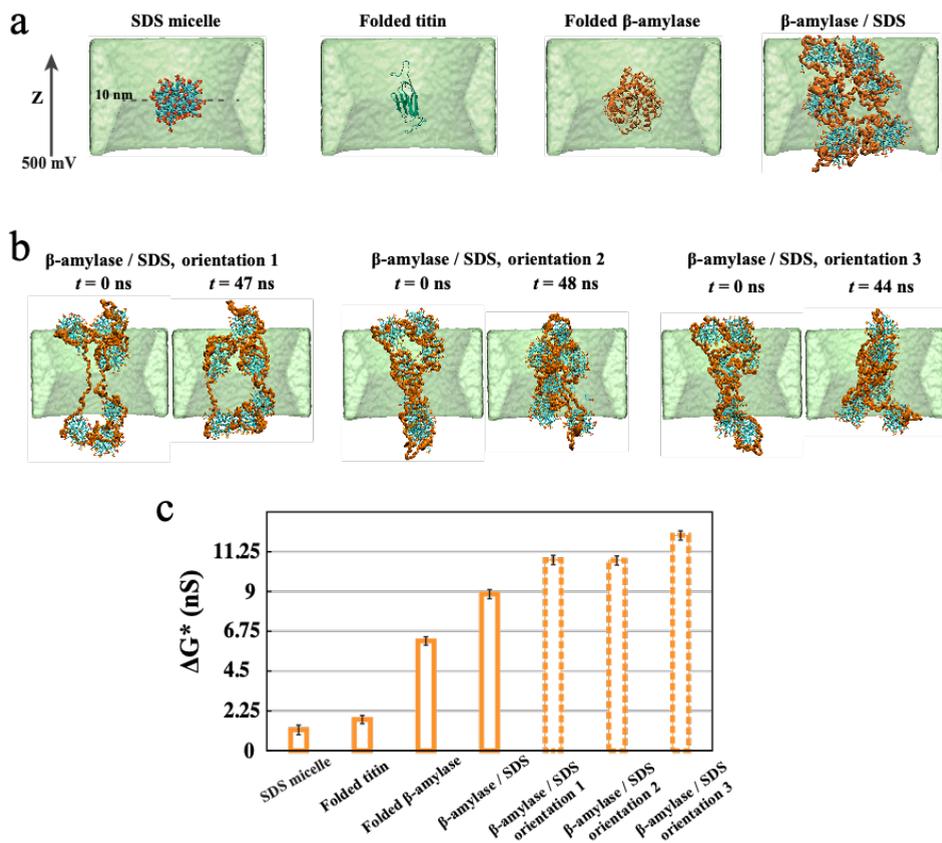


Figure S6.3: Simulated ionic current blockades in a 10 nm-diameter nanopore. (a) Molecular configurations used for the simulation of the ionic current blockades. The center of mass of each molecule or molecular assembly was restrained to the center of the nanopore. Water and ions are not shown for clarity. (b) Additional simulations of β -amylose/SDS blockade currents. The initial conformations of the β -amylose/SDS assemblies were taken from the translocation simulations performed using the 6 nm-diameter nanopore (System 1). In each 48-ns simulation, the center of mass of the β -amylose/SDS assembly was restrained to the center of the nanopore. (c) The average conductance blockade amplitudes. To enable direct comparison with experiment, the conductance blockades computed from MD simulations were multiplied by the ratio of the experimental and simulation bulk conductivities of 0.4 KCl (3.6/4.8). The average blockade conductance amplitudes were each computed from a 48 ns MD trajectory.

Note S1. Estimation of the critical micelle concentration of SDS using dynamic light scattering (DLS)

It is well known that SDS critical micelle concentration is dependent on the ionic strength of the solution. The formation of micelles depends on the interplay between the attractive forces driven by hydrophobic interactions of the hydrocarbon tails and the repulsive electrostatic interaction between the charged heads. At high ionic concentrations, charges are screened and the critical micelle concentration is reduced. To determine the CMC of SDS in the buffer used for our experiments, we used DLS to determine at which concentrations we notice the presence and absence of SDS micelles.

DLS measurements were performed using a Malvern Instruments Zetasizer Nano ZS. Samples containing different SDS concentration were prepared in a buffer containing 0.4M NaCl, 10mM TRIS and 1mM EDTA pH 7.5. The hydrodynamic radius of each sample was measured and is presented in Figure S6.4. For SDS concentrations higher than 0.05% micelles were observed with the expected hydrodynamic radius. No micelles were observed below 0.01% SDS. These results indicate a CMC for SDS in the range of 0.01% - 0.05%.

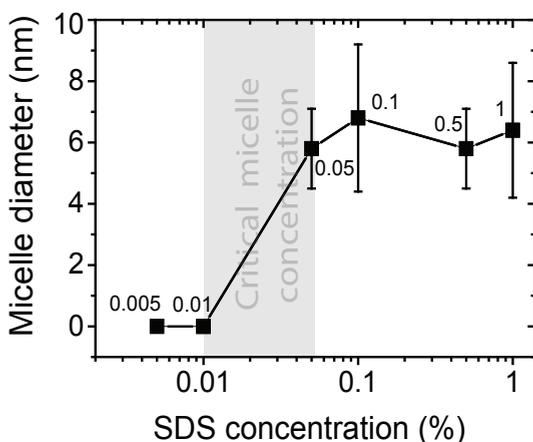


Figure S6.4: DLS measurements of micelle hydrodynamic radius at different SDS concentrations. The presence of micelles above 0.05% SDS and their absence below 0.01% SDS marks the region where the CMC is located.

Note S2. Tryptophan fluorescence bulk assay

Tryptophan fluorescence is routinely used to study protein conformational changes. Tryptophan fluorescence emission is highly dependent on the nature of its local environment, which can be used to infer the conformation of a protein. In a native or folded protein, most tryptophan residues are either partially or fully buried in the core of the protein due to their hydrophobic nature. Once the protein is unfolded with denaturing agents such as SDS or urea, tryptophan residues become partially or fully exposed and their fluorescent emission shifts. Here we used this principle to study the effect of SDS at different concentrations on one of our protein substrates.

All our measurements were done in a solution containing 0.4 M NaCl, 10mM Tris pH 7.5, 1mM EDTA, 1mM of DTT and different concentrations of SDS as shown in the following figure. Fluorescent measurements were done in a Cary eclipse fluorescence spectrophotometer with an excitation wavelength of 295nm and emission collected from 300nm – 450nm. The buffer background was measured and subtracted from each protein measurement.

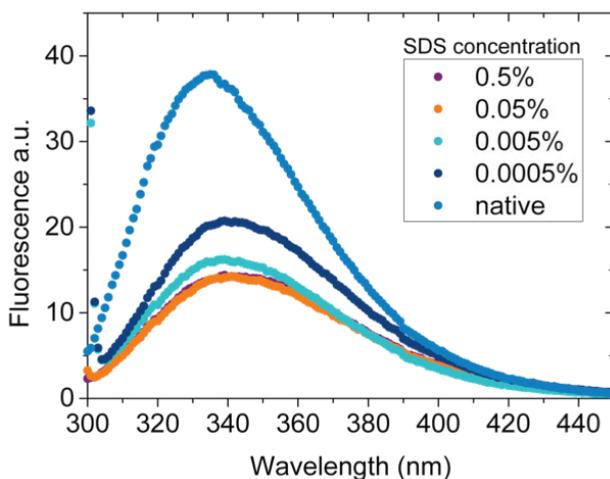


Figure S6.5: Tryptophan fluorescence bulk measurements of β -amylase at different SDS concentrations. A clear conformational change is observed in the protein even at SDS concentrations below CMC.

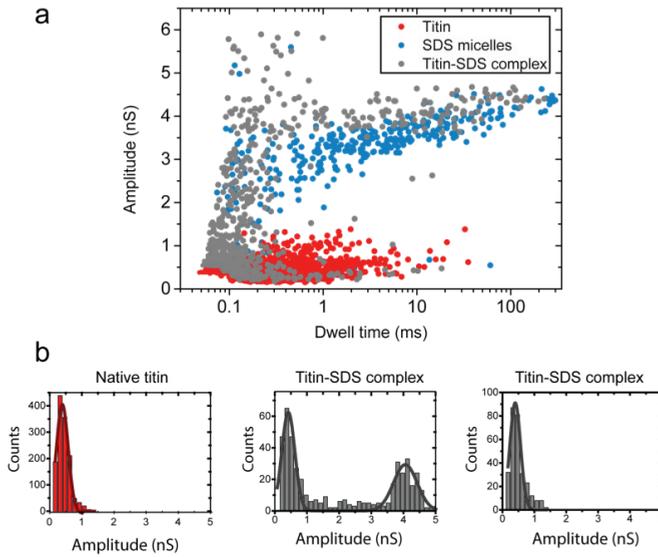


Figure S6.6: Translocation of native and SDS-unfolded titin. (a) Scatter plots of dwell time vs. conductance blockade for titin, SDS-titin and SDS micelles. (b) Histograms of the amplitude of the blockade for native titin (left), SDS-titin at an SDS concentration above CMC (middle) and SDS-titin below CMC (right).

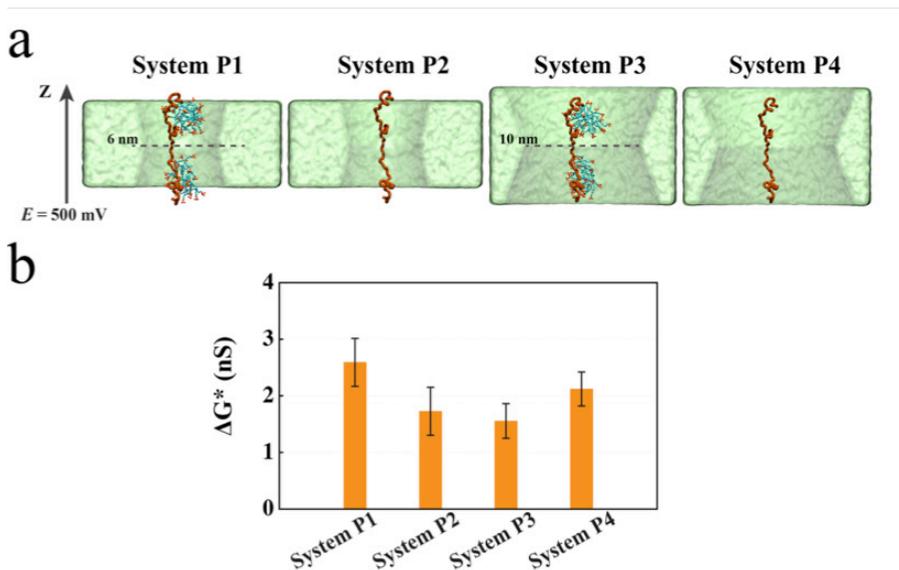


Figure S6.7: Simulated ionic current blockades produced by a stretched peptide chain. (a) Molecular configurations featuring a 60 amino acid fragment of β -amylase (residues 10 to 70) with (Systems P1 and P3) and without (Systems P2 and P4) attached SDS molecules threaded through either the 6 nm diameter nanopore (Systems P1 and P2) or the 10 nm diameter nanopore (Systems P3 and P4). The initial configuration of the peptide was taken from the simulation of the β -amylase/SDS assembly, orientation 3 at $t = 0 \text{ ns}$ (see Figure S6.3). The center of mass of the peptide chain/SDS assembly (Systems P1 and P3) or only of the peptide chain (Systems P2 and P4) was aligned with the geometrical center of each nanopore. During MD simulations of the ionic current, the backbone atoms of the peptide chain were harmonically restrained to their initial coordinates; SDS molecules were not restrained. At the end of the 60 ns simulations of Systems P1 and P3 under a 500 mV transmembrane bias, 20 and 1 SDS molecules detached from the peptide chain, respectively. (b) The average conductance blockade amplitudes. To enable direct comparison with experiment, the conductance blockades computed from MD simulations were multiplied by the ratio of the experimental and simulation bulk conductivities of 0.4 KCl (3.6/4.8). The average blockade conductance amplitudes for each system were computed from a 60 ns MD trajectory.

References

1. P. Mitchell *Nat. Biotechnol.*, 2002, 21 , 233 -237.
2. B. N. Miles , A. P. Ivanov , K. a. Wilson , F. Doan , D. Japrun and J. B. Edel , *Chem. Soc. Rev.*, 2013, 42 , 15 -28 RSC .
3. C. Dekker *Nat. Nanotechnol.*, 2007, 2 , 209 -215.
4. M. Wanunu *Phys. Life Rev.*, 2012, 9 , 125 -158.
5. J. J. Kasianowicz , E. Brandin , D. Branton and D. W. Deamer , *Proc. Natl. Acad. Sci. U. S. A.*, 1996, 93 , 13770 -13773.
6. A. Meller , L. Nivon , E. Brandin , J. Golovchenko and D. Branton , *Proc. Natl. Acad. Sci. U. S. A.*, 2000, 97 , 1079 -1084 .
7. M. Akeson , D. Branton , J. J. Kasianowicz , E. Brandin and D. W. Deamer , *Biophys. J.*, 1999, 77 , 3227 -3233.
8. G. M. Cherf , K. R. Lieberman , H. Rashid , C. E. Lam , K. Karplus and M. Akeson , *Nat. Biotechnol.*, 2012, 30 , 344 -348.
9. E. A. Manrao , I. M. Derrington , A. H. Laszlo , K. W. Langford , M. K. Hopper , N. Gillgren , M. Pavlenok , M. Niederweis and J. H. Gundlach , *Nat. Biotechnol.*, 2012, 30 , 349 -353.
10. A. H. Laszlo , I. M. Derrington , B. C. Ross , H. Brinkerhoff , A. Adey , I. C. Nova , J. M. Craig , K. W. Langford , J. M. Samson , R. Daza , K. Doering , J. Shendure and J. H. Gundlach , *Nat. Biotechnol.*, 2014, 32 , 829 -833.
11. B. Hornblower , A. Coombs , R. D. Whitaker , A. Kolomeisky , S. J. Picone , A. Meller and M. Akeson , *Nat. Methods*, 2007, 4 , 315 -317.
12. Q. Zhao , G. Sigalov , V. Dimitrov , B. Dorvel , U. Mirsaidov , S. Sligar , A. Aksimentiev and G. Timp , *Nano Lett.*, 2007, 7 , 1680 -1685.
13. S. W. Kowalczyk , A. R. Hall and C. Dekker , *Nano Lett.*, 2010, 10 , 324 -328 .
14. C. Plesa , J. W. Ruitenber , M. J. Witteveen and C. Dekker , *Nano Lett.*, 2015, 15 , 3153 -3158 .
15. M. M. Mohammad , S. Prakash , A. Matouschek and L. Movileanu , *J. Am. Chem. Soc.*, 2008, 130 , 4081 -4088 .
16. K. J. Freedman , S. R. Haq , J. B. Edel , P. Jemth and M. J. Kim , *Sci. Rep.*, 2013, 3 , 1638.
17. E. C. Yusko , J. M. Johnson , S. Majd , P. Prangkio , R. C. Rollings , J. Li , J. Yang and M. Mayer , *Nat. Nanotechnol.*, 2011, 6 , 253 -260.
18. C. Plesa , S. W. Kowalczyk , R. Zinsmeister , A. Y. Grosberg , Y. Rabin and C. Dekker , *Nano Lett.*, 2013, 13 , 658 -663.
19. E. C. Yusko , B. R. Bruhn , O. M. Eggenberger , J. Houghtaling , R. C. Rollings , N. C. Walsh , S. Nandivada , M. Pindrus , A. R. Hall , D. Sept , J. Li , D. S. Kalonia and M. Mayer , *Nat. Nanotechnol.*, 2016, 12 , 360 -367.

20. D. S. Talaga and J. Li , *J. Am. Chem. Soc.*, 2009, 131 , 9287 -9297.
21. D. Rodriguez-Larrea and H. Bayley , *Nat. Nanotechnol.*, 2013, 8 , 288 -295.
22. L. Movileanu , S. Howorka , O. Braha and H. Bayley , *Nat. Biotechnol.*, 2000, 18, 1091 -1095.
23. J. Nivala , D. B. Marks and M. Akeson , *Nat. Biotechnol.*, 2013, 31 , 247 -250.
24. Y. Zhao , B. Ashcroft , P. Zhang , H. Liu , S. Sen , W. Song , J. Im , B. Gyarfás , S. Man-na , S. Biswas , C. Borges and S. Lindsay , *Nat. Nanotechnol.*, 2014, 9 , 466 -473.
25. J. Wilson , L. Sloman , Z. He and A. Aksimentiev , *Adv. Funct. Mater.*, 2016, 26 , 4830 -4838.
26. C. B. Rosen , D. Rodriguez-Larrea and H. Bayley , *Nat. Biotechnol.*, 2014, 32 , 179-181.
27. M. Firnkes , D. Pedone , J. Knezevic , M. Döblinger and U. Rant , *Nano Lett.*, 2010, 10 , 2162 -2167.
28. J. Li , D. Fologea , R. Rollings and B. Ledden , *Protein Pept. Lett.*, 2014, 21 , 256-265.
29. E. Kennedy , Z. Dong , C. Tennant and G. Timp , *Nat. Nanotechnol.*, 2016, 11 , 968 -976.
30. D. Otzen *Biochim. Biophys. Acta, Proteins Proteomics*, 2011, 1814 , 562 -591.
31. L. Luo , S. R. German , W.-J. Lan , D. A. Holden , T. L. Mega and H. S. White , *Annu. Rev. Anal. Chem.*, 2014, 7 , 513 -535.
32. A. Lee , S. K. Y. Tang , C. R. MacE and G. M. Whitesides , *Langmuir*, 2011, 27 , 11560 -11574.
33. M. M. Nielsen , K. K. Andersen , P. Westh and D. E. Otzen , *Biophys. J.*, 2007, 92 , 3674 -3685.
34. D. E. Otzen *Biophys. J.*, 2002, 83 , 2219 -2230.
35. X. J. a. Janssen , M. P. Jonsson , C. Plesa , G. V. Soni , C. Dekker and N. H. Dekker , *Nanotechnology*, 2012, 23 , 475302.
36. V. Tabard-cossa , D. Trivedi , M. Wiggin , N. N. Jetha and A. Marziali , *Nanotechnology*, 2007, 18, 305505.
37. C. Plesa and C. Dekker , *Nanotechnology*, 2015, 26 , 1 -7.
38. J. C. Phillips , R. Braun , W. Wang , J. Gumbart , E. Tajkhorshid , E. Villa , C. Chipot , R. D. Skeel , L. Kalé and K. Schulten , *J. Comput. Chem.*, 2005, 26 , 1781 -1802.
39. P. F. Batcho , D. A. Case and T. Schlick , *J. Chem. Phys.*, 2001, 115 , 4003 -4018.
40. A. D. MacKerell , D. Bashford , M. Bellott , R. L. Dunbrack , J. D. Evanseck , M. J. Field , S. Fischer , J. Gao , H. Guo , S. Ha , D. Joseph-McCarthy , L. Kuchnir , K. Kuczera , F. T. K. Lau , C. Mattos , S. Michnick , T. Ngo , D. T. Nguyen , B. Prodhom , W. E. Reiher , B. Roux , M. Schlenkrich , J. C. Smith , R. Stote , J. Straub , M. Watanabe , J. Wiórkiewicz-Kuczera , D. Yin and M. Karplus , *J. Phys. Chem. B*, 1998, 102, 3586 -3616.
41. E. R. Cruz-chu , A. Aksimentiev and K. Schulten , *Society*, 2006, 110 , 21497 -21508.

42. J. Yoo and A. Aksimentiev , *J. Phys. Chem. Lett.*, 2012, 3 , 45 -50.
43. J. Yoo and A. Aksimentiev , *J. Chem. Theory Comput.*, 2016, 12 , 430 -443.
44. S. Miyamoto and P. A. Kollman , *J. Comput. Chem.*, 1992, 13 , 952 -962.
45. H. C. Andersen *J. Comput. Phys.*, 1983, 52 , 24 -34.
46. T. Darden , D. York and L. Pedersen , *J. Chem. Phys.*, 1993, 98 , 10089.
47. G. J. Martyna , D. J. Tobias and M. L. Klein , *J. Chem. Phys.*, 1994, 101 , 4177.
48. S. Banerjee , J. Wilson , J. Shim , M. Shankla , E. A. Corbin , A. Aksimentiev and R. Bashir , *Adv. Funct. Mater.*, 2015, 25 , 936 -946.
49. B. W. H. Van Beest , G. J. Kramer and R. A. Van Santen , *Phys. Rev. Lett.*, 1990, 64 , 1955 -1958.
50. D. B. Wells , V. Abramkina and A. Aksimentiev , *J. Chem. Phys.*, 2007, 127 , 125101.
51. M. Shankla and A. Aksimentiev , *J. Phys. Chem. B*, 2017, 121 , 3724 -3733.

7

Single-molecule protein analysis: From the lab to the market

Bridging the gap between scientific developments and their societal utilization is an important step for the impact of research and education in society. In this chapter we analyze the most important aspects of bringing a technology for single-molecule protein sequencing from the lab, to the market. We describe the unmet market needs and opportunities, compare this technology to current methods, and explore the potential applications and markets in which such a technology could enter.

This chapter was written in collaboration with S.Heerema, C. Joo and C. Dekker.

7.1 Unmet market needs and opportunities

Proteins are the major building blocks of life. A human cell contains millions of different protein species, each having their own unique role: to duplicate the genome, to catalyze chemical processes, and to defend the cell against viral attack, etc¹. The functions of proteins are vital to the cell, and the protein composition of a cell and tissue provides key information for the understanding of biological processes and diseases. Protein identification and characterization is therefore crucial for diagnostics, drug discovery, and fundamental sciences.

Despite the huge importance of protein analysis, current technologies are limiting when it comes to the level of detail that is needed². Most cost-effective technologies, such as immunoassays, probe for known 'biomarkers', proteins that are present in a patient's sample that mark for a specific disease. But the methods are applicable to only a limited number of diseases in which biomarkers have been identified and are abundant in a patient's sample. Mass spectrometry allows for the de novo detection of unknown proteins, but the technique is very costly, is time-consuming, requires a large amount of sample (e.g. blood volume), and needs highly skilled technicians.

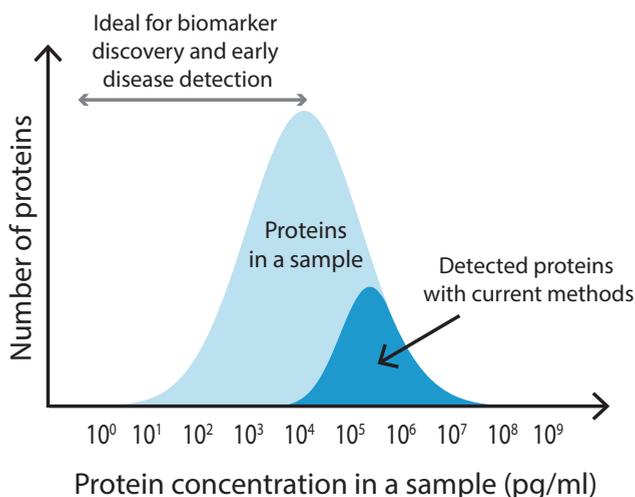


Figure 7.1: The bottleneck of protein analysis and biomarker discovery. Protein concentrations vary widely in a sample. Current methods can detect only the most abundant proteins, but new methods are necessary for low-abundance protein detection. Adapted from 3.

Radically new methods are necessary to achieve the sensitivity, dynamic range and cost-effectiveness needed for proteomics. One of the main challenges emerges from the fact that protein concentrations in a biological sample can vary by up to 9-10 orders of magnitude

(Fig 7.1)³⁻⁵. This large variability in concentration, combined with the lack of amplification methods for proteins, brings tremendous difficulties for the measurement of proteins present in low abundance.

7.2 Current techniques for protein analysis

Mass spectrometry is the current gold standard for protein sequencing. In Chapter 1 Box 2 we have described the working principle of the technique. A mass spectrometer easily costs 200.000 - 500.000 euros and requires a trained technician for its operation, as well as data analysis. Besides the costs and time factors, there are other fundamental limitations with the technique. Human samples are extremely complex, comprising a wide range of protein concentrations. In human plasma, for example, the concentration of proteins can vary over 9 orders of magnitude, from a few picograms per milliliter (Interleukin 6) to milligrams per milliliter (albumin)^{2,4}. Therefore, an exceedingly high dynamic range is necessary for comprehensive proteome analysis. In the most common approach to proteomics using mass spectrometry, proteins are digested into peptides and thousands of peptides are analyzed at once. Therefore, proteins present in low abundance cannot be distinguished from the background noise of the measurement. Thus, state-of-the-art mass spectrometers are limited to a dynamic range of $\sim 10^4$ to 10^5 . Where the most abundant proteins can be identified, but rare proteins that are present in low-copy numbers are mostly invisible⁶ (Fig 7.1). Many biomarkers for cancer and other diseases often escape timely detection for this very reason.

Mass spectroscopy performs an unbiased measurement, meaning that it analyses several proteins present in a sample. It has the highest dynamic range among the different protein analysis methods currently available. The main downside of the technique comes from its cost-effectiveness (Fig 7.2). Mass spectrometry is expensive, slow, and it requires a high-level trained user to handle and readout the machine. Moreover, concerns about its intra and inter-laboratory reproducibility have been raised⁷.

Because of the above, mass spectrometry is not commonly used in clinics and hospitals for routine diagnosis. Its use is limited to the discovery phase of proteomics, where samples of controls and diseased patients are analyzed in the search for disease specific biomarkers. Bringing mass spectrometry to the clinical chemistry lab has been a challenge. Only a percentage of hospitals can afford using mass spectrometry for routine clinical chemistry analysis. For that reason, clinical chemistry labs perform biomarker discovery with mass spectrometry, but move quickly to immunoassays for validation and adaptation of tests into the clinical chemistry lab.

In immunoassays, proteins are identified based on their affinity for particular antibodies. The analyte/antibody interaction is then detected using different methods including electrochemistry, fluorescence, luminescence, etc. Immunoassays, such as ELISA (enzyme-linked immunosorbent assay), are cost-effective and do not need a very skilled user. Making it an ideal tool for the clinical chemistry lab. However, it can only perform a biased read-out towards a predefined protein, limiting its usefulness for biomarker discovery. Moreover, its dependence on the specificity and availability of antibodies brings an inherent problem for multiplexing. Immunoassays can only detect tens or occasionally hundreds of different proteins at once, which makes their dynamic range low.

The fact that discovery and validation are often done with different techniques brings additional challenges to proteomics, as the findings made with one method not always translate smoothly to the other. This process brings an important challenge for the translation of a biomarker from research to clinical diagnosis. Techniques in which both biomarker discovery and validation can be done are of great need.

7.3 A solution based on single-molecule techniques: the competitive advantages of single-molecule analysis

Single molecule techniques, in which molecules are analyzed one at the time, have brought a revolution to biophysics, allowing for extremely sensitive analysis. Much like if happened for DNA sequencing, a breakthrough is expected for protein sequencing based on single-molecule approaches.

As presented here in Chapter 2, colossal efforts are undergoing now in the field of single-molecule protein sequencing in which different research laboratories around the world are using fluorescence, nanopores or tunneling currents to achieve sequencing readouts. Because of its scalability, sensitivity, and temporal resolution, we favor a nanopore method for protein sequencing or fingerprinting.

We envision that a protein sequencer based on nanopores, could become a cost-effective tool that can be used for biomarker discovery and validation. By detecting single-molecules one at the time, high sensitivity and dynamic range are expected. This would be particularly important for the detection of low-abundance proteins.

Finally, a nanopore method for protein sequencing would ideally read full-length proteins. This feature is important for the correct identification of proteoforms. The most common method for mass spectrometry-based protein analysis is bottom up proteomics, in which

proteins are digested and peptide digests are analyzed. A disadvantage of this approach is that several protein proteoforms share peptide segments, therefore identification of the exact proteoforms and their quantification is not possible⁸.

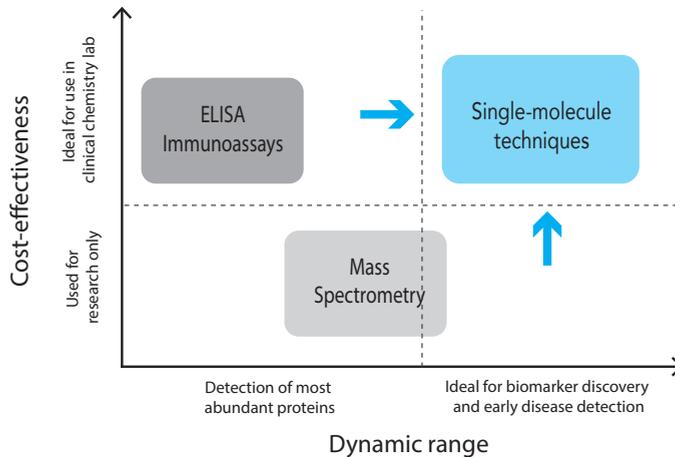


Figure 7.2: Cost-effectiveness vs. dynamic range of current methods used for protein analysis, compared to the expected performance of the single molecule techniques.

7.4 Potential applications

7.4.1 Biomarker discovery

Biomarker discovery is key for the development of personalized medicine. During the process of biomarker discovery, biological samples (blood, plasma, cerebrospinal fluid, etc) of control and diseased patients are compared to find the differences in protein expression profiles. High sensitivity is required for biomarker discovery, as proteins present in diseased tissues or tumors are often present in the blood in extremely low concentrations.

7.4.2 Biomarker detection for early disease diagnosis

Biomarkers that are identified as indicative of a disease should be detected with high sensitivity, selectivity and accuracy. High sensitivity is important for early disease detection. Biomarker detection in patients samples is normally performed in the clinical chemistry laboratory, therefore techniques for biomarker detection should be cost-effective and easy to use.

7.4.3 Quality control

Biotech and pharmaceutical companies produce a vast number of enzymes and antibodies for therapeutic purposes. The characterization of the produced proteins, including their

PTM profiles, is important to determine their activity and immunogenic response. In the same way food companies need to characterize the proteins present in products such as dairy.

7.4.4 Scientific research

The expression and purification of recombinant proteins is widely used in academic research in areas such as molecular biology, biochemistry and biophysics. Characterization of the expressed proteins is often limited to techniques such as SDS PAGE. A cost effective protein sequencing technique with higher sensitivity would be useful for a better characterization

7.5 Market analysis

A single-molecule protein analysis tool can enter both the protein identification (ID) and characterization market, or the larger proteomics market, depending on the particular application and capabilities of the technique. The protein ID market comprises the characterization and quantification of a particular protein or a limited set of proteins. For example, the characterization of the PTMs present in a protein. The larger proteomics market covers the analysis of the proteomes present in biological samples, often implying the analysis of thousands of different proteins.

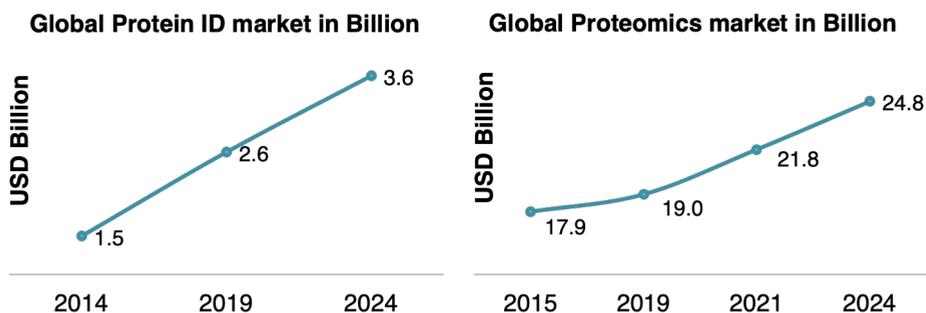


Figure 7.3: The Global Protein ID market (left) and the Global Proteomics market (right). Predictions are based on a conservative 7% yearly growth. Adapted from 9 and 10.

7.5.1 Protein identification and characterization market

Single-molecule detectors or sensors aimed at characterizing a particular set of proteins will enter this market. The protein ID market is expected to reach \$2.6B by 2019, and \$3.6B by 2024 (based on a conservative 7% yearly growth rate). The major driving forces behind this growth are the increased budgets for proteomics R&D in both the private and public sector,

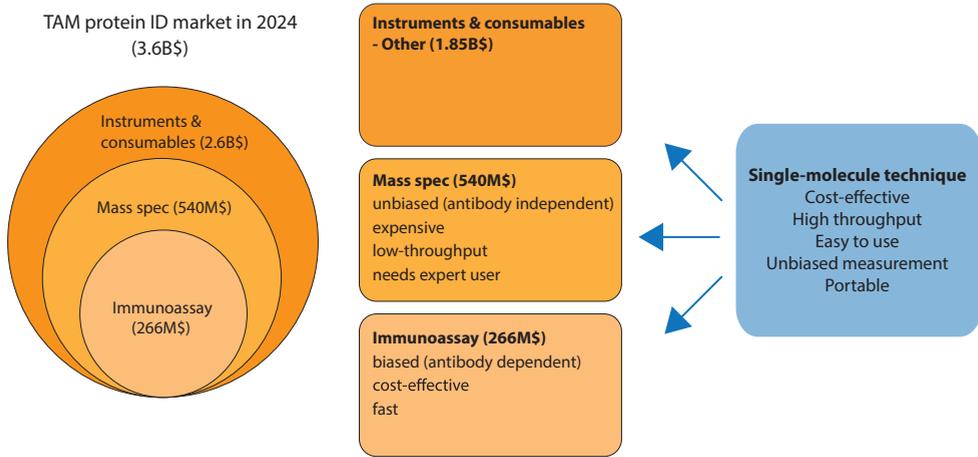


Figure 7.4: The Global Protein ID market Instruments segments as predicted for 2024.

and technological developments. The customers of the protein identification and characterization market can be segmented into the biopharmaceutical industry, academic research, and clinical diagnostics. The market's products are segmented into instruments & consumables, and services. Within the instruments & consumables segment, the mass spectrometry part is estimated to be \$540M, and immunoassay part to be \$266M in 2024 (Fig. 7.4).

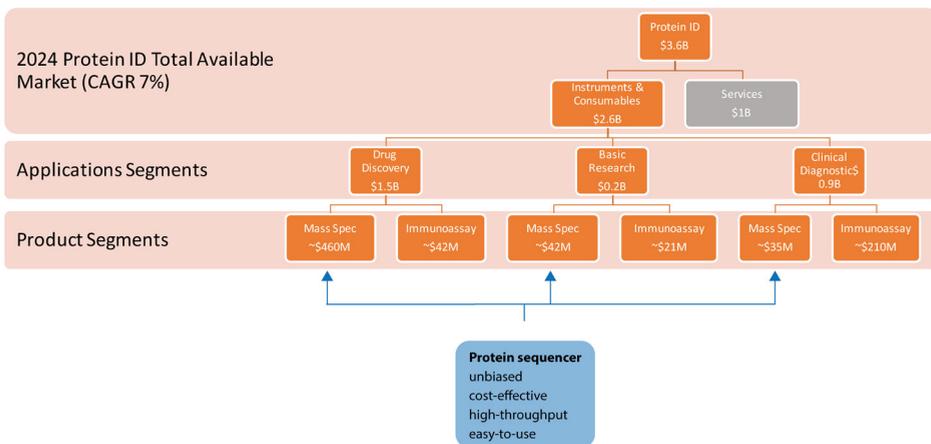


Figure 7.5: The Global Protein ID market segments as predicted for 2024.

The product segments were single-molecule technologies will most probably penetrate are mass spectroscopy and immunoassays. These are the most standard methods for protein analysis and are the methods that will compete with a single molecule detector directly. The two technologies have their advantages and limitations. Mass spectroscopy performs an

unbiased measurement, but has several downsides; i.e. it is expensive, slow, and it requires a high-level trained user to handle and readout the machine. In ELISA technology, one pre-selects for certain proteins in a sample, as it uses antibodies for the detection readout. ELISA is cost-effective, and does not need a very skilled user, but can only perform a biased read-out.

From both segments, the mass spectrometry market is of greater interest due to the lack of alternative cheap and easy to use technologies. The mass spectrometry market can be estimated to become \$540M by 2024 (taking a 7% yearly growth rate for the global market).

7.5.2 The larger proteomics market

Technologies able to perform protein sequencing at the single-molecule level in complex samples will enter the larger Proteomics market, where large amounts (thousands) of proteins are analyzed in biological samples.

The global proteomics market is expected to reach \$21.87B by 2021, and \$24.8B by 2024. The major driving forces behind this growth are the strongly increasing need for personalized medicine and increased budgets for proteomics R&D in both private and public sector. The most important barriers in the proteomics market are the high costs in tools and equipment, and the need for skilled researchers.

The proteomics market is ~7 times larger than the protein identification market. Here, clinical diagnosis is the largest application segment. This is expected, as proteomics is mainly done for clinical research and discovery of biomarkers. As a consequence clinical research centers such as hospitals will be the main costumers.

7.5.3 Obtainable market

To provide an estimate of the expected obtainable market, we can consider a device capable of performing single molecule protein PTM analysis of samples containing a limited number of proteins. Such a device would enter the Protein ID market in 2024.

As previously explained, the mass spectrometry segment is estimated to be \$540M, and the immunoassay segment to be \$266M in 2024. From both segments, the mass spectrometry market is of greater interest due to the lack of alternative cheap and easy to use technologies. Moreover, mass spectrometry is a common tool for PTM analysis.

If we look at the customer segments of the protein ID market (Figure 7.5), we find that the

market is divided into biopharmaceutical industry (42%), academic research (24%), and clinical diagnostics (34%). The biopharmaceutical industry is the larger segment and is an interesting potential target for a protein characterization product. Biopharmaceuticals are expected to sell \$120 B in monoclonal antibodies in 2020¹¹. Techniques able to characterize the quality of these antibodies in a fast and high-throughput manner are crucial. A device that can analyze the PTM content of intact proteins is not available at the moment, so it is expected to penetrate that market as an attractive alternative to current technologies. The academic research labs correspond to a smaller portion of the market (24%). However, academic researchers are likely to be early technology adaptors and can be important partners in the development process.

Based on this information, an attractive go-to market plan for such a device would consist on commercializing the device to academic researchers in the first 2-3 years, and then move towards the biopharmaceutical industry. A nanopore device can be considerably cheaper than mass spectrometry and easier to use. Therefore we can conservatively estimate to obtain 5% of the market in the first year (approx. \$2.1M) and 8% in the second year (approx. \$3.4M). After that, commercialization should be extended to the biopharmaceutical industry, where revenues can be ten times higher. Here, with an obtainable market of 5%, revenues would correspond to approx. \$23M.

7.6 Conclusion

In conclusion, the development and commercialization of a single-molecule protein fingerprinting platform offers good business opportunities with high impact for various industries:

- Medical healthcare - for early-stage disease detection
- Pharmaceutical industry - for quality control of proteins during drug development
- Agriculture - for quality control during food production
- Scientific research centers for biochemical research - to screen proteins for fundamental understanding on molecular processes in cells, and for biomarker detection, to sense proteins associated with specific diseases.

The technology can be disruptive in areas such as clinical research for biomarker discovery, or in disease diagnosis. Moreover, a technology able to read full length-proteins has significant competitive advantages compared to other methods.

The main limitations for the development of a product that can be commercialized are the technical challenges discussed in this thesis. Among them, controlled protein unfolding

and translocation is an important obstacle that needs to be overcome. Ultimately, for the realization of single-cell proteomics and low-abundance protein analysis, the challenges of sample preparation should now be overlooked.

References

1. Alberts et al. *Molecular biology of the cell*. (Garland Science, 2002). doi:10.1091/mbc.E14-10-1437
2. A cast of thousands. *Nat. Biotechnol.* 21, 213 (2003).
3. Eriksson, J. & Fenyő, D. Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs. *Nat. Biotechnol.* 25, 651–655 (2007).
4. Anderson, N. L. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* 1, 845–867 (2002).
5. Aebersold, R. et al. How many human proteoforms are there? *Nat. Chem. Biol.* 14, 206–214 (2018).
6. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* 452, 571–579 (2008).
7. Bell, A. W. et al. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat. Methods* 6, 423–430 (2009).
8. Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* 4, 1419–40 (2005).
9. Data obtained from Market&Markets research performed in 2014. The research assumed 10% CAGR between 2014-2019. We have extrapolated this to calculate the market and segment sizes in 2024 but projected a CAGR of 7% between 2019-2024 for a more conservative scenario.
10. Market&Markets Proteomics Market by Instrument, Reagent & Service (Spectroscopy, Chromatography, Electrophoresis, Immunoassay, HPLC, X-Ray Crystallography, Mass Spectrometry and Surface Plasmon Resonance), Application (Diagnostics, Drug Discovery) - Forecast to 2021. Available at : <https://www.marketsandmarkets.com/PressReleases/proteomics.asp>
11. Grilo, A. L. & Mantalaris, A. The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol.* (2018).

Summary

The function and phenotype of a cell is determined by a complex network of interactions between DNA, RNA, proteins and metabolites. Therefore, a comprehensive approach that integrates genomics, transcriptomics, proteomics, and metabolomics is necessary to achieve full understanding of biological processes and disease. Recent technological developments have mostly focused on the study of genomes. DNA sequencing has become fast, cheap, and ubiquitous. The study of other -omes, especially the proteome, remains expensive and time-consuming.

Nanopores have emerged as attractive tools for the development of a protein sequencer with single-molecule capabilities. In a nanopore sensor, an insulating membrane made of a lipid bilayer or a solid-state membrane separates two compartments filled with an electrolyte. A nanometer-sized pore is made within the membrane, by inserting a single protein-pore into a lipid bilayer or by drilling a pore in a solid-state membrane using transmission electron microscopy. When a voltage is applied across the membrane, an ionic current flows through the nanometer-sized aperture. Molecules passing or translocating through the pore modulate the ionic current, which provides the basic sensor signal. This principle has been successfully applied to DNA sequencing. Protein sequencing is the next frontier.

Several technical challenges need to be overcome for the realization of single-molecule protein sequencing with nanopores. For example, the possibility to distinguish all the 20 amino acids remains a tremendous challenge. Additionally, the protein needs to be unfolded and the full peptide chain moved across the pore in a controlled manner. In **chapter 2**, we present a summary of these challenges, together with the relevant developments reported so far. Besides nanopores, other single molecule methods such as fluorescence and tunnelling currents are currently explored for protein sequencing. A review of these developments is also presented in this chapter.

For a nanopore-based protein sequencer, the protein needs to be unfolded and the amino acid sequence sequentially read as the polypeptide chain translocates through the pore. Protein unfolding and controlled translocation across a nanopore remains a tremendous challenge. For that reason, peptides are often used as simplified model systems for proof of concept experiments. Peptides of different charge, size and structure have been previously studied with biological nanopores. In **chapter 3** we explore the use of a bipolar peptide

as model system. This peptide contains 10 negative amino acids in the N-terminus and 10 positive amino acids in the C-terminus. The opposite charges at the ends of the peptide generate force in opposing directions stalling the peptide while it translocates the FraC pore. This mechanism helps to linearize the peptide and allows for extended examination of the central region of the peptide, which is further explored in chapters 4 and 5. In chapter 3 we explore other interesting features observed in this peptide-FraC system: the non-linear current voltage dependence observed for the system. Our experiments and molecular dynamics simulations show that the relative blockade caused by the peptide changes depending on the applied bias as a consequence of peptide stretching and compressing. We propose to use this bipolar peptide as an electromechanical gate that can modulate nanopore conductance in a dynamic manner and at the single-molecule level.

For the realization of de novo protein sequencing, differentiating 20 different amino acids during nanopore translocation is a significant technical hurdle. In recent years, the groups of Joo and Marcotte proposed an alternative simpler idea, namely, protein fingerprinting, in which proteins can be identified if a subset of amino acids is labelled and read. Joo and colleagues proposed protein identification through detection of the sequence of cysteines and lysines along the peptide chain. The lysine and cysteine sequence is then compared to a protein database for protein identification. Nanopores offer an attractive technique for the implementation of protein fingerprinting, thanks to their high sensitivity and time resolution. For such a nanopore protein fingerprinting method, cysteines, lysines, or other amino acids should be modified with labels that can produce a distinct modulation in the current while the linearized protein traverses the nanopore. In **chapter 4**, we investigate the effect of adding different chemical modifications in single amino acid side chains within a polypeptide chain. We use the model peptide described in chapter 3 to characterize different chemical modifications. We show that sensitive and reproducible detection can be obtained for labels with a spectrum of different physicochemical properties such as mass, geometry, charges, and hydrophobicity. We further demonstrate that information about the position of the label along the peptide chain can be extracted from individual current-blockade event features. Our findings provide insights into the effects of different chemical labels on nanopore ionic current signals, which is important for protein sensing and fingerprinting.

The results presented in chapter 4 demonstrate the possibility to detect modifications in single amino acid side chains using the FraC nanopore. In nature, modifications in single amino acids side chains are abundant. Post-translational modifications or PTMs are modifications occurring after the synthesis of proteins, and usually involve the addition of chemical groups to specific amino acids along the protein. PTMs play important roles in protein (in)activation, stability, and recognition, among other things. Consequently, their detection

and identification is important for elucidating complex cellular processes and disease. In **chapter 5** we explore the detection of two post-translational modifications with nanopores: Phosphorylation and O-glycosylation. These two PTMs have an intricate and highly dynamic regulatory process, where they compete for protein modification sites. Their dysregulation has been implicated in the development of cancer, Alzheimer's disease, diabetes, and many more. We show that using a FraC nanopore the phosphorylated and non-phosphorylated peptide variants can be differentiated by their relative current blockade. The same was shown for an O-glycosylated peptide variant. Besides the effect of these modifications on relative blockade, the translocation dwell-times were also analyzed. Finally, conditions in which these modifications could be differentiated are shown, demonstrating the identification of phosphorylation and O-glycosylation in a label-free manner using nanopores. These results represent an important step for the single molecule identification of proteoforms or protein isoforms, which is of tremendous importance for disease diagnosis and cell biology.

In **chapter 6** we move to the study of full-length proteins. Here we combine molecular dynamics (MD) simulations with single-molecule experiments to investigate the utility of SDS (Sodium Dodecyl Sulfate) to unfold proteins for solid-state nanopore translocation while simultaneously endowing them with a stronger electrical charge. Our simulations and experiments prove that SDS-treated proteins show a considerable loss of the protein structure during the nanopore translocation. Moreover, SDS-treated proteins translocate through the nanopore in the direction prescribed by the electrophoretic force due to the negative charge imparted by SDS. In summary, our results suggest that SDS causes protein unfolding while facilitating protein translocation in the direction of the electrophoretic force; both characteristics being advantageous for future protein sequencing applications using solid-state nanopores. Despite successful protein unfolding, SDS denatured proteins translocate too fast through the nanopore and additional mechanisms to control translocation speed are necessary.

Finally, in **chapter 7**, we analyze the most important aspects of bringing a technology for single-molecule protein sequencing from the lab, to the market. We describe the market needs and opportunities, compare this technology to current methods, and explore the potential applications and markets in which such a technology could enter.

Samenvatting

De functies en fenotypes van cellen worden bepaald door een complex netwerk van interacties tussen DNA, RNA, eiwitten en metabolieten. Om een totaalbeeld van biologische processen en ziektes te krijgen is een brede aanpak nodig die genomica, transcriptomica, proteomica en metabolomica samenbrengt. Recente technologische ontwikkelingen hebben zich voornamelijk gericht op genomica, waardoor het bepalen van DNA sequenties nu snel, goedkoop en alomtegenwoordig is geworden. De studie van andere -oom velden, met name het proteoom, blijft echter duur en tijdrovend.

Nanoporiën zijn naar voren gekomen als aantrekkelijke hulpmiddelen voor de ontwikkeling van een eiwit-sequencer met enkel-molecuul capabilities. In een nanopore sensor worden twee compartimenten, die gevuld zijn met een elektrolyt, gescheiden door een isolerend dubbellaags lipidemembraan of door een 'vaste-stof' membraan. In het membraan wordt op nanometer schaal een porie gemaakt door een enkel porie-eiwit in het lipidemembraan te plaatsen of door met een transmissie-elektronenmicroscop een gat te boren in het 'vaste-stof' membraan. Als er een spanning over het membraan wordt gezet ontstaat er een ionenstroom door de nanoporie. Moleculen die de porie passeren zullen de ionenstroom veranderen, waardoor een meetbaar signaal ontstaat. Dit principe is succesvol toegepast voor het bepalen van DNA sequenties; de volgende stap is het bepalen van de aminozuursequenties van eiwitten.

Om een enkel-molecuul eiwit-sequencer te realiseren met nanoporiën, zullen verschillende technische uitdagingen overwonnen moeten worden. Het onderscheiden van alle 20 aminozuren blijft bijvoorbeeld een enorme uitdaging. Daarnaast moet het eiwit worden ontvouwen en de intacte peptideketen moet op een gecontroleerde manier door de porie worden verplaatst.

In **hoofdstuk 2** geven we een overzicht van deze uitdagingen samen met de relevante ontwikkelingen die tot nu toe zijn gepubliceerd. Naast nanoporiën worden er ook andere methodes gebruikt waarmee de sequentie van eiwitten per molecuul kan worden bepaald, zoals het gebruik van fluorescentie en van tunnelstromen. Een samenvatting van deze ontwikkelingen wordt ook beschreven in dit hoofdstuk.

Voor het bepalen van eiwitsequenties met nanoporiën, moet het eiwit eerst ontvouwen

worden en vervolgens moet de aminozuursequentie sequentieel worden afgelezen terwijl de polypeptideketen zich door de porie verplaatst. Het ontvouwen van het eiwit en de gecontroleerde translocatie door de porie blijven enorme uitdagingen. Daarom worden peptides vaak gebruikt als vereenvoudigd modelsysteem voor proof of concept experimenten. Peptides van verschillende lengtes, groottes and structuren zijn eerder al bestudeerd met biologische nanopores. In **hoofdstuk 3** kijken we naar een bipolaire peptide als modelsysteem. Dit peptide heeft 10 negatieve aminozuren aan de N-terminus en 10 positieve aminozuren aan de C-terminus. De tegenovergestelde ladingen aan beide uiteinden van het peptide genereren een kracht in de tegenovergestelde richting, wat ervoor zorgt dat het peptide vast blijft zitten wanneer het zich door de FraC-porie verplaatst. Dit mechanisme helpt om het peptide te ontvouwen en zorgt voor een uitgebreide detectie van de centrale regio van het peptide, waar in hoofdstuk 4 en 5 verder op in wordt gegaan. In hoofdstuk 3 bestuderen we andere interessante eigenschappen van dit peptide-FraC systeem: het niet-lineaire verband tussen stroom en spanningen dat geobserveerd is voor dit systeem. Onze experimenten en moleculaire dynamica simulaties laten zien dat de relatieve blokkade die veroorzaakt wordt door dit peptide verandert afhankelijk van het aangebrachte potentiaal, wat het resultaat is van het strekken en samendrukken van het peptide. Wij stellen voor om deze bipolaire peptide te gebruiken als een elektromechanische gate die de soortelijke geleiding van de nanopore op een dynamische manier en op moleculair niveau kan veranderen.

Een significante technische horde om de novo sequentiebepaling van eiwitten te realiseren is de identificatie van 20 verschillende aminozuren tijdens een translocatie door een nanopore. In de afgelopen jaren hebben de groepen van Joo en Marcotte een simpeler idee voorgesteld, namelijk het identificeren van eiwitten aan de hand van hun 'vingerafdrukken'. Bij deze methode wordt slechts een deel van aminozuren gelabeld en geïdentificeerd. Joo en collega's willen eiwitten identificeren door de sequentie van cysteïnes en lysines van een peptideketen uit te lezen. De cysteïne en lysine sequentie wordt dan vergeleken met een eiwit database waardoor het eiwit kan worden geïdentificeerd. Het gebruik van nanoporiën is, dankzij de hoge gevoeligheid en tijdsresolutie, een aantrekkelijke methode voor de implementatie van een eiwit-vingerafdruktechniek. Voor een dergelijke methode moeten cysteïnes, lysines of andere aminozuren worden gemodificeerd met labels die een duidelijke verandering in de stroom geven wanneer het ontvouwen eiwit de nanopore passeert.

In **hoofdstuk 4** bestuderen we het effect van het toevoegen van verschillende chemische modificaties aan zijgroepen van individuele aminozuren in de peptideketen. We gebruiken hetzelfde modelpeptide als beschreven in hoofdstuk 3 om deze modificaties te karakteriseren. Wij laten zien dat gevoelige en reproduceerbare detectie mogelijk is van labels

met een spectrum aan fysiochemische eigenschappen, zoals massa, geometrie, lading en hydrofobiciteit. Verder laten we zien dat informatie betreffende de positie van het label op de peptideketen uit individuele stroom-blokkades kan worden bepaald. Onze bevindingen geven een overzicht van de effecten die verschillende chemische labels hebben op de stroomsignalen van een nanoporie, wat erg belangrijk is voor eiwit-detectie en de vinger-afdruktechniek.

De resultaten van hoofdstuk 4 laten zien dat het mogelijk is om modificaties te detecteren in zijketens van individuele aminozuren met behulp van de FraC nanoporie. In de natuur komen de modificaties van zijketens van individuele aminozuren vaak voor. Posttranslatie-nele modificaties, oftewel PTMs, zijn modificaties die ontstaan na de synthese van eiwitten. Hierbij worden voornamelijk chemische groepen aan specifieke aminozuren in het eiwit toegevoegd. PTMs spelen onder andere een belangrijke rol in de (in)activatie, stabiliteit en herkenning van eiwitten. De detectie en identificatie van PTMs is daarom belangrijk om inzicht te krijgen in de werking van complexe cellulaire processen en ziektes.

In **hoofdstuk 5** bestuderen we de detectie van twee posttranslationele modificaties met nanoporiën: fosforylatie en O-glycosylering. Deze twee PTMs hebben een complex en dynamisch reguleringsproces, waar ze concurreren voor dezelfde modificatieplekken op eiwitten. Ontregeling van deze processen heeft implicaties voor de vorming van kanker, Alzheimer, diabetes en vele andere ziektes. Wij laten zien dat met behulp van een FraC nanoporie de gefosforyleerde en niet-gefosforyleerde peptide varianten onderscheiden kunnen worden aan de hand van hun relatieve stroom blokkade. Ditzelfde laten we zien voor een O-geglycosyleerde peptide variant. Naast het effect van deze modificaties op de relatieve blokkade, zijn de translocatietijden ook geanalyseerd. Ten slotte demonstreren we de condities waarmee deze modificaties van elkaar onderscheiden kunnen worden. Hiermee tonen we aan dat fosforylatie en O-glycosylering op een label-vrije manier geïdentificeerd kunnen worden met behulp van nanoporiën. Deze resultaten vertegenwoordigen een belangrijke stap voor de identificatie van proteovormen of eiwit isovormen op het niveau van individuele moleculen, welke voor diagnostiek van ziektes en celbiologie van bijzonder groot belang is.

In **hoofdstuk 6** richten we onze aandacht op het bestuderen van gehele, intacte eiwitten. Hier combineren wij simulaties van moleculaire dynamica (MD) met experimenten op het niveau van individuele moleculen om te onderzoeken of SDS (Natriumdodecylsulfaat) gebruikt kan worden om eiwitten te ontvouwen voor vaste-stof nanoporie translocaties terwijl ze tegelijkertijd een sterkere elektrische lading krijgen. Onze simulaties en experimenten bewijzen dat eiwitten behandeld met SDS een significant verlies van structuur ondergaan tijdens de translocatie door de nanoporie. Eiwitten die behandeld zijn met SDS translo-

ceren door een nanoporie in de richting van de elektroforetische kracht, dit komt door de negatieve lading die SDS aan het eiwit geeft. Om samen te vatten, suggereren onze resultaten dat SDS eiwit-ontvouwing veroorzaakt en dat SDS tegelijkertijd de translocatie van het eiwit faciliteert in de richting van de elektroforetische kracht; beide karakteristieken zijn gunstig voor toekomstige toepassingen van aminozuursequentie bepaling met behulp van vaste-stof nanoporiën. Ondanks dat eiwitten snel worden ontvouwen door SDS, transloceren deze SDS gedenatureerde eiwitten te snel door de nanoporie. Aanvullende methodes om de translocatie snelheid te reguleren zijn daarom nodig.

Tot slot, in **hoofdstuk 7** analyseren we de belangrijkste aspecten die nodig zijn om de technologie voor het bepalen van aminozuursequenties van individuele eiwitten, van het lab naar de markt te brengen. We beschrijven de marktbehoeften en -kansen. Ook vergelijken we deze technologie met huidige methodes, en verkennen we de potentiële toepassingen en markten waarin deze technologie kan worden geïntroduceerd.

Acknowledgements

Writing this section is the most tangible realization that my PhD is coming to an end! It's a strange feeling. On one hand, I feel nostalgic for the culmination of this unforgettable journey, but on the other hand I feel happy for having made it this far. Most of all, I feel immensely grateful for all the meaningful moments I have lived during the last 5 years. In every single step of this journey, I found myself surrounded by smart, fun, energetic, and kind people that inspired me and helped me in so many different ways. It's good to finally get a chance to thank you properly.

First of all, big big thanks to Chirlmin and Cees. To me, you were the best supervisory pair I could have dreamed of. The yin and yang of my PhD. Thank you for your continuous enthusiasm in the project, and for always being present.

Chirlmin, I admire your endless kindness and positivism. You are a very brilliant and creative scientist. Most importantly, you have showed me that top science can be done with humbleness and kindness. It's very unique to find this combination in the competitive nature of academia today. Please, never change that. Your character is also reflected in the way you run the lab. It's a very friendly and fun atmosphere. Thank you for being always supportive and for your immense sense of empathy. I truly feel that you care about us and our well-being.

Cees, your endless amount of energy and enthusiasm is remarkable. Despite of your busy schedule, you always find time to meet with us and discuss even the smallest technical details. You zoom out to see the big picture, but you also zoom in without much effort. It's very admirable and I have learned a lot from you. You have taught me more about leadership, time management, and communication than any graduate school course could ever do.

I also want to extend my gratitude to the committee members (Prof. Sang Wook Lee, Prof. Hamza Balci, Prof. Marileen Dogterom, Prof. Giovanni Maglia, Prof. Nynke Dekker, and Dr. Javier Alfaro). Thank you for taking part in my thesis defense. A special thank you goes to Giovanni. Giovanni, our collaboration brought great momentum to my PhD. Thank you (and your lab members Huang, Florian, Misha, Carsten, and Natalie) for welcoming me in your lab and introducing me to the world of biological nanopores. Your creativity never

ceases to surprise me. Aleksei, you were another very important collaborator for this thesis. It was always a pleasure to work with you. One of the sharpest minds I have ever seen. During our discussions, you could always cut through the noise and get to the main point in no time. Thanks to Shalini and Shidi for their simulations.

It was very enjoyable to have some “partners-in-crime” to discuss, brainstorm, and struggle with nanopores. Thanks to all the students that worked together with me in this project. Mees, Louis-Marie, Nathalie, Lois, and Chun. I was lucky to have very talented and motivated students next to me. Thank you for your help and contribution.

BN is a very unique department where you have the opportunity to both do cutting edge science and make lasting friendships. At the center of this special place are the very brilliant and enthusiastic PIs that keep the wheel spinning: Marileen, Cees, Chirlmin, Marie-Eve, Bertus, Greg, Stan, Anne, Christophe, Nynke, Martin, Timon, Arjen, Liedewij, Huyn, Dimphna, Andreas. Next to the PIs, all the support staff makes sure everything is running properly. Jolijn, Amanda, Diana, Emmylou, Esther, Chantal, Tracey, Nadine, Angela, Marije, Sussane, thank you for all your help. Sasha, its always a pleasure to run into you in the department because your first question is never: “how is your thesis going?”, but: “how are you? how is your family?”. I like your sense of priorities. Niels and Marek, thank you for all your help and answers about mass spec and HPLC. Jelle, thank you for all the flow cells. I love how you treat each of your creations as a small piece of art. Thanks as well to Jeremie, Roland, and Dimitri.

Very quickly after stating my PhD, the BN family welcomed me. In particular there were 7 suspects that made my first years unforgettable. Mal, I have missed you so much in the past 2 years! You were always like a PhD big sister to me. Thank you for being always in the best mood, for great life advice, and for all your help with science. I hope to see you very soon and meet Victor. Luuk, it was amazing to share all these PhD years together. Many good memories remain. I always admire the great discipline you put into everything you do. I hope Switzerland will be a very fruitful professional and personal experience. Mathi, you are one of the most positive and easy-going people I know. You have a kind hearth. I have never heard you say anything bad about anybody. It’s been a lot of fun to share a little bit of my culture with you. Moh, I will never forget that Colombian football match when you first tried aguardiente. It was always a lot of fun to have you around. I also enjoyed all the moments I spent with your family. I am happy you found a nice warm place to settle. I hope I will see you and your family soon. Pawel, holaaa, no experiment was boring with you. Thank you for teaching me so much about proteins and peptides. Especially thank you for all the fun lunches, dinners, drinks, and parties. Stanley, it was fun to share an office in the old building. Thanks for all the PhD advice. I am glad you found a nice job in Oxford.

I hope you are enjoying it there. Micky, it was so enjoyable to have a warm Italian around. Fun was guaranteed when you and Ciro were around. I am very happy that now you have a beautiful family. Hope to visit you in Spain.

Being part of two labs has been double fun! Unforgettable lab trips, conferences, dinners, coffees, and friendships. Mike, meneer, thanks for representing the sequencing team as my paranymp. It's a lot of fun to have you around. The sequencing project is in good hands with your creativity and hard work. I am looking forward to working together in the coming months. Just remember, the idea in itself is very simple. Thijsito, I like your good mood and easy-going attitude. I admire the way you have juggled all these different projects. Good luck in the final stretch! Viktorija, alias V, the last few years have been quite a rollercoaster for you! I have seen you change and grow a lot. I have had fun seeing you in all the different facets. I am happy to see the nice family you and Seb have made together <3. Sungc-hurl, alias beercules, its been fun to share an office together. We have talked about Korea, Colombia, science, politics, football, etc., etc. You are a very smart and creative scientist. I am sure you will have a very brilliant future in science. Ivo, you always carry a good smile with you. I am glad you decided to join the lab for your PhD. Thanks for all the amazing cakes! Iasonas, alias Greek guy, you seemed very serious when I first met you, but it soon became clear that you are a very funny guy. It's nice to have a theoretician in the group. Thanks for your tips on data analysis, etc. Ilja, you have been very brave fighting all the issues related to our building. I hope you will come back soon. We miss you! Cecilia, being in the sequencing project can be quite demanding, but you are doing a great job. I hope we can get to know each other and work together in the coming months. Margreet and Jan, thank you for keeping the lab running in such a smooth way. Thanks as well to former JooCie lab members Jetty and Anna. Jetty, you were the first sequencer. Thanks for introducing me to the sequencing world when I joined the lab. I am glad you have found a nice path to follow after your PhD. Anna, thanks for all your help in the lab. Looking forward to seeing Laura soon. Peggy and Carlos, you are not officially JooCie lab members, but everybody knows you definitely are! It's been so great to work together. Peggy, thanks for your efforts in making tags for me. We had a couple of crazy days in the lab together. I look forward to working together again. I extend my gratitude to your supervisors Rienk and Dick, and to all the FOM team for their input.

My introduction into the nanopore world went smoothly thanks to all the help from the amazing nanopore team. Adi, you were the first one to show me the tips and tricks of nanopores. It was fun to share the lab on the old building. There was always a story to talk about. I am glad you are still around for casual chats and laughs. Steph, I am very happy to have shared the PhD journey with you. During the PhD you were always a great inspiration. You persevered and managed to keep your cool while struggling with a very hard

project. Now we have started a new adventure, and its fun to see a different side of you, both in the business world and as a mom. I admire your drive, positivism, and relaxed attitude. I am learning a lot from you. Sergii, a man of many talents. Good dancer, photographer, smart scientist, and great friend. I am glad you are staying around Delft. Thanks for the nice breakfasts with Kseniya and Alex. Daniel V, your enthusiasm and cheerfulness is contagious. Calin and Magnus, the nanopore grandfathers, thank you for all the nice input and good discussions. Xin, you have proven that talent and hard work are a great combination. Wayne, our journey started a bit rough with the Cas9 project, but in the end it brought us closer. I really enjoyed getting to know you. Long days and nights with our nanopores allowed us to talk about so many different things. I love your humour and your instagram posts. Good luck with the last stretch! Sonja, I admire your determination and hard work. The sequencing project is in very good hands and I am sure you will make big steps. Yoones, my birthday partner. I enjoyed our parties together. Maybe one more party this year? Ale, 549 4eva, NPC meets Italy part 1. You work hard, but also play hard. Keep it up. Paola, NPC meets Italy part 2. I like the new bio perspective you bring to the team. You are full of nice ideas and enthusiasm. Sabina, we have shared nice moments in parties and dinners. You seem to have teamed the graphene beast. Good for you! Nils, you have a big challenge ahead, but you strike me as a smart and pragmatic guy, good luck!

With other CD lab members, we don't share our love for nanopores, but we share many other things. Michel, we clicked since we talked during your interview. Now I am lucky to have the best neighbours. Thanks for all the dinners and chats. You are the bravest of all. We are lucky that you brought Madi into our lives, and Madi is lucky to have such a great and caring mom. Siddharth, the plasma cleaner in the old building united us. You are always so cheerful, always with a big smile. Thanks for the dinner invitations with your family. We need to chat more before one of us leaves! Jorine, I like your pragmatic approach to life. It was fun to plan Sinterklaas parties together. It was nice to catch up in your last visits to NL. I am happy you found an exciting place for your postdoc. Anthony, your presence is so enjoyable. I am happy that you are keeping up the the social life of BN. Ganji, I enjoyed our conversations. You are such a resourceful scientist. Kuba, creative both in art and science. I hope you are enjoying Sweden. Fede, thanks for making music a big part of the BN parties. To all the other (former) CD lab members: Fabai, Felix, Greg, Yaron, Sung Hyun, Sandro, JK, Alberto, Eugene, Mitasha, Biswajit, and Kevin. Thanks for bringing a nice spirit to the lab. Jaco, Jacob, Allard, and Eli. If Cees is the DNA of the lab, you are the proteins. Jaco, it was always a pleasure to come by the wet lab and find you there. I enjoy our chats. Jacob, I admire you for living two parallel lives as an artist and scientist.

As if having two great labs is not enough. BN is full of cheerful people, with whom I have shared many special moments. I want to start with the cheerful ladies of our BN girls group:

Mathi, Becca, Alicia, Lisa, Helena, Nicole, Eve, Esengul, and Elena. Weather we meet on the dance floor or having a nice cosy dinner, fun is guaranteed. I am happy to have you around. I hope our meetings will transcend our PhDs. Becca, I like your chilled and fun personality. Ali, a conversation is never boring if you are around. Lisa, English channel?! You go get it! Helena, our Latino-french diva. Love our shared euphoria when those reggaeton songs play. Nicole, you are really the cool kid on the block. Eve, nobody beats you in the dance floor. Esi, chica, you make my days brighter with your big smile. Elena, welcome! nice to have you around. Thanks to all the other BN suspects that with a friendly and open mind, were always up for a chat: Sam, Humberto, Carsten, Benjamin, Mehran, David, Misha, Sumit, Philip, Richard, Louis, Nathalia, Da, Jochem, Cristian, Patrick, Boris, Jonas, Orkide, Seungkyu, Mariana, Andrew, Pauline, Dominik, Vanessa, Victor, Bojk, David D.

Outside of BN, many friends have made my life very gezellig. Getting settled in Delft was easy thanks to my amazing Balkan friends Ranko and Dejan. By the time I got here, they already had the whole system figured out (including the TVs ;)). Rankito, thank you for being my paranymp. Life gave us the chance to re-meet in Delft and I am happy for that. I love all the memories we have cherished over the years. Whenever we talk, I feel that I have an honest and kind friend listening. Thank you for so any special moments. Don't worry, I will not forget to thank you for the biggest gift of all: bringing Taylor into my life. I am forever in debt. Dejancito, kaisive (I have no idea how to write that!). We don't see each other as often as we should, however, every time we meet it feels as if we just saw each other. Being as smart and warm as you are is a unique combination. I am lucky to have you around.

My lovely girls Yildiz and Fahimeh. I am enormously grateful to Maria for introducing us. I feel lucky to have your support and company. Our conversations are full of laughter and inspiration. I appreciate that we can talk about essentially anything. Thanks for introducing me to Grazia and other inspiring women.

Vero y Gabi, chachorritas, son mis paranymps de la vida. Me siento muy afortunada de haberlas conocido. Con sus grandes corazones y alegria me recargan de energia cada vez que nos vemos. Ya son muchas las memorias que hemos construido y espero que sigamos siendo amigas hasta viejitas. Las quiero mucho.

Carmelitas! Ya son mas de 15 años de amistad. Todas tan diferentes, pero cuando estamos juntas (como en Puerto Rico), todo tiene sentido. Todo vuelve a ser como en el colegio. Me siento muy orgullosa de todas. Las quiero encantadoras. Las super mas: Karen, Isa, Nadia y Naty. Gracias por tantos años de amistad. Las extraño mucho. A los biomedicos y EMM nano, muchas gracias por tantos buenos recuerdos.

Mamá, papá y Juan. Ustedes son mi motor y motivación. Me siento inmensamente bendecida por tenerlos. Mami, tu eres una verdadera guerrera. El pilar de nuestro hogar. Gracias por el amor incondicional que nos muestras todos los días. Papi, gracias por que siempre nos has apoyado en todos nuestros proyectos. Gracias por alegrarnos los días con tu buen humor y serenidad. Los quiero mucho. Juanchito, mi parcerito de la vida, por donde empezar? Aunque no hablemos todos los días, yo se que siempre estamos ahí el uno para el otro. Cuente conmigo siempre, siempre. Lo admiro muchísimo por su tenacidad y constancia. Usted es un verdadero campeón. Lo quiero mucho.

And finally you Taylor, mi lindo. We met soon after starting our PhDs and since then we have been walking this path together. I have enjoyed every step of it. Thank you for being so kind and loving. With your sense of humor, you can make me laugh even in the hardest of the days. I have learned, and keep learning a lot from you. I would like to thank your family and friends as well, for welcoming me with arms wide open. I am proud of the home we have built together. It is my absolute favorite place on earth. I am looking forward to seeing what life brings for us next. I love you.

Delft, March 2019

List of publications

8. Zhao S.*, **Restrepo-Pérez L.***, Soskine M. , Maglia G., Joo C., Dekker C. and Aksimentiev A. Electro-mechanical conductance modulation of a nanopore using a removable gate. *ACS Nano*. 13 (2), 2398–2409 (2019).
7. **Restrepo-Pérez L.**, Huang g., Bohlaender P., Worp N., Eelkema R., Maglia G., Joo C., Dekker C. Resolving modifications in single amino acids using a biological nanopore. (In preparation).
6. **Restrepo-Pérez L.**, Wong C.H., Maglia G., Dekker C., Joo C. Label-free detection of post-translational modifications with a nanopore. (In preparation).
5. **Restrepo-Pérez L.**, Joo C. & Dekker C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* 13, 786–796 (2018).
4. Yang W., **Restrepo-Pérez L.**, Bengtson M., Heerema S.J., Birnie A., van der Torre J., Dekker C. Detection of CRISPR-dCas9 on DNA with Solid-State Nanopores. *Nano Lett.* 10, 6469-6474 (2018).
3. **Restrepo-Pérez L.***, John S.*, Aksimentiev A., Joo, C. & Dekker C. SDS-assisted protein transport through solid-state nanopores. *Nanoscale* 9, 11685–11693 (2017).
2. **Restrepo-Pérez L.**, Soler L., Martínez-Cisneros C., Sánchez S., Schmidt O.G. Bio-functionalized self-propelled micromotors as an alternative on-chip concentrating system. *Lab Chip*. 16, 2914-7 (2014).
1. **Restrepo-Pérez L.**, Soler L., Martínez-Cisneros C.S., Sánchez S., Schmidt O.G. Trapping self-propelled micromotors with microfabricated chevron and heart-shaped chips. *Lab Chip*. 9, 1515-8 (2014).

