University of Groningen

Engineering nanopore sensors for proteomic and metabolomic analysis

Lucas, Florian

DOI:
[10.33612/diss.196801819](https://doi.org/10.33612/diss.196801819)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Lucas, F. (2022). *Engineering nanopore sensors for proteomic and metabolomic analysis*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.196801819>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/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 (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Engineering Nanopore Sensors for Proteomic and Metabolomic Analysis

Florian Leonardus Rudolfus Lucas

The research described in this thesis was performed at the department of Chemical Biology 1, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands.

Printed and bound by Gildeprint

© Florian L. R. Lucas, The Netherlands, 2021

university of
 groningen

Engineering Nanopore Sensors for Proteomic and Metabolomic Analysis

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on
Tuesday 18 January 2022 at 12.45 hours

by

Florian Leonardus Rudolfus Lucas

born on 25 February 1994
in Apeldoorn

Supervisor

Prof. G. Maglia

Co-supervisor

Dr. C. Wloka

Assessment Committee

Prof. W.R. Browne

Prof. M.W. Fraaije

Prof. G.L. Corthals

Table of content

Chapter 1	3
1.1. Introduction	4
1.1.1. <i>Proteins, Peptides, and Proteomics</i>	5
1.1.2. <i>The Metabolome and Metabolomics</i>	6
1.1.3. <i>Single-Molecule Sequencing</i>	8
1.1.4. <i>Towards Nanopore Protein Sequencing</i>	9
1.2. Nanopores.....	9
1.2.1. <i>Biological Nanopores</i>	10
1.2.2. <i>Solid-State Nanopores</i>	12
1.2.3. <i>Biological Nanopores: Resistive Pulse Sensing</i>	13
1.2.4. <i>Biophysical Properties</i>	15
1.2.5. <i>Data Interpretation</i>	16
1.3. Targeted Engineering of the Channel	17
1.4. Protein Detection using Nanopore Spectrometry.....	20
1.5. Adapter-based Metabolite Detection using Nanopores	23
1.6. Scope of the Thesis.....	25
1.7. References	26
1.8. Supporting Information	34
Chapter 2	35
2.1. Introduction	36
2.2. Results and Discussion	37
2.2.1. <i>Fragaceatoxin C Mutant Screening</i>	37
2.2.2. <i>Fragaceatoxin C Mutant Characterization</i>	43
2.2.3. <i>Characterization using Molecular Dynamics Simulations</i>	45
2.3. Conclusions	50
2.4. Materials and Methods	51
2.5. References	58
2.6. Supporting Information	63
Chapter 3	73
3.1. Introduction	74
3.2. Results.....	75

3.2.1. Calibration of the G13F-FraC-T1 Nanopore.....	75
3.2.2. Detection of Tryptic Digested Lysozyme.....	77
3.2.3. Protein Profiling using G13F-FraC-T1	78
3.3. Discussion	79
3.4. Methods.....	81
3.5. References	89
3.6. Supporting Information	93
Supplementary References	103
Chapter 4	105
4.1. Introduction	106
4.2. Results and Discussion	107
4.2.1. Wild-type TbpA (WT-TbpA) Signal.....	107
4.2.2. The Engineering of the TbpA Sensor.....	108
4.2.3. Detection of Phosphorylated Derivatives	110
4.2.4. Automated Analysis.....	111
4.2.5. Automatic (real-time) Detection of Thiamine in Urine	112
4.3. Conclusion.....	113
4.4. Materials and Methods	114
4.5. References	118
4.6. Supporting Information	122
Chapter 5	131
5.1. Augmenting Nanopore Resolution for Peptide Detection	132
5.2. Nanopore-based Protein Identification	133
5.3. Real-time Quantification of Small Molecules.....	133
5.4. Outlook for Fully Folded Protein Analysis	134
5.5. References	135
Summary.....	137
Samenvatting.....	141
Acknowledgements.....	145
List of Publications	148
Lay Summary.....	149

Chapter 1

Introduction

1.1. Introduction

The quest to decipher the enigma of life has led to the development of numerous techniques and the emergence of the field of life-science. Our understanding of the underlying mechanisms in biology has evolved greatly since the discovery of microscopic organisms by Robert Hooke and Antoni van Leeuwenhoek in the age of enlightenment. With the relatively recent discovery of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), we have gained molecular understanding of—and ability to manipulate—organisms on the molecular scale. The potential of the information for life contained within these molecules did not go unnoticed, culminating in one of the most successful and ambitious projects in life-sciences, the Human Genome Project (1990-2003).¹ Next-generation DNA sequencing (NGS) techniques—such as dye-based fluorescence and nanopore resistive pulse sensing—have especially led to a massive reduction in required resources,² with the cost of whole exome sequencing dropping from US\$100 million in 2001 to under US\$1000 in 2018.³

The interplay between the rapid analysis time and high-throughput of NGS paved the way to incorporate a patient's cellular profile into treatment, allowing an increase in treatment effectiveness and reduction of adverse effects.^{4,5} Physiological information can be extracted from the genome, transcriptome, proteome and metabolome.⁶ The disciplines that study the interactions between and extraction of this information, are generally referred to as *-omics*. Especially genomics and transcriptomics have revolutionised the effort, due to their abundance of high-throughput and low-cost methods.⁷ The biopolymers that build the genome (composed of DNA) and transcriptome (composed of RNA) are easy targets for analysis due to their chemically similar subunits. The read-out of the proteome and metabolome is, however, more difficult due to the significant number of chemically dissimilar subunits and a large number of possible chemical functionalization, such as post-translational modifications on proteins or phosphorylation of metabolites.

The aim of this thesis is to establish a nanopore-based technology for the analysis of proteins, and to provide a new approach to rapidly, and automatically, extract metabolomic information using nanopores.

1.1.1. Proteins, Peptides, and Proteomics

The primary functional elements in carbon-based organisms—proteins—are essential complex biological entities comprising of (over) 20 *proteinogenic*ⁱ amino acids ranging in mass between 57 to 186 Daltonⁱⁱ (**Supplementary Figure 1**).^{8,9} Amino acids contain a wide range of chemical functionalities—allowing the formation of three-dimensional structures due to the wide array of possible electrostatic and hydrophobic interactions.^{8,9} The ambiguous terms protein and peptide both describe a polymeric chain of amino acids, however, generally a peptide refers to short polymeric chains without a well-defined three dimensional structure, while proteins are usually larger and specifically configured in all three dimensions. The chemical functionality of *proteinogenic* amino acids is dominated by the side chain connected to an almost universally shared *peptide backbone* which allows polymer formation (**Supplementary Figure 1**). The polymeric bond between *proteinogenic* amino acids—with the exception of prolineⁱⁱⁱ—is formed across the carboxyl and amino group on the *peptide backbone*;⁸ therefor, proteins are also commonly referred to as polypeptides.⁸ We write the amino acid sequence of (poly)peptides using either a three- or single letter code and, to avoid ambiguous interpretation, the order from left to right is always written from the amino-terminal (N-terminal) to the carboxyl-terminal (C-terminal).⁸

Functional proteins are (generally) encoded in genes and the entire set of genes is called the genome. Consequently, the entire set of proteins encoded by the genome is called the proteome. The study of proteomes (proteomics) generally aims to determine the relationship between (specific parts of) genes and cellular function as expressed on the protein level.¹⁰ By studying the proteome, information of the current phenotype of the host organism (*e.g.* healthy or affected) can be established that accounts for both the genotype and translational differences.

The *Homo sapiens* proteome consists of approximately 20,000 proteins ranging between a few hundred to over a million Dalton in mass.¹¹ The dynamic mass range alone makes direct detection of the full proteome nearly impossible, but is further complicated by translational differences and post-translational modifications (PTMs).

At the time of writing this thesis, the most common method for the exploration of the proteome is based on a combination of liquid chromatography (LC) and mass spectrometry (MS), together abbreviated as LC-MS. During LC, the (poly)peptides pass through a column containing a stationary phase, usually tightly packed silica beads with specific physico-chemical properties. The stationary

ⁱ Amino acids incorporated into proteins during translation.

ⁱⁱ A Dalton (Da) is equal to $1/12$ the mass of carbon-12 in its ground state.

ⁱⁱⁱ Proline does not contain an amino group, peptide bonds are formed between the carboxyl and imino group instead.

phase retains the (poly)peptides as they interact with it, thus they elute at different times. However, this separation is unable to resolve all (poly)peptides due to their large similarities. To identify the (poly)peptides, they are ionised and their mass over charge ratio (m/z) is detected using a MS, allowing the accurate determination of the total mass of the (poly)peptides.¹² Ionisation is required to attract the (poly)peptides towards the detector of the MS. However, it is important to notice that the efficiency of ionization is difficult to predict for a (poly)peptide, because it is dependent on its physical properties, rendering the technique inherently non-quantitative.

Usage of a single mass analyser allows the accurate determination of the mass of a (poly)peptide, however, the mass of a (poly)peptide is not unique due to the vast array of possible combinations between amino acids. To determine the sequence, two mass analysers are put in serial with a fragmentation chamber in between.¹³ The first mass analyser filters the ions by m/z and the second analyser detects the fragments resulting from fragmentation. This fragment spectrum can be directly used to determine the original (poly)peptides based on expected *in-silico* fragmentation, a process known as fingerprinting. Alternatively, it can be analysed and reconstructed to gain the original (poly)peptide sequence using (computationally expensive) *de novo* sequencing.

Two major methodologies for MS-based proteomics exist: top-down or bottom-up.¹⁴ The aim of the top-down approach is to measure intact proteins.¹⁵ MS requires the conversion of proteins into gas-phase ions, which is problematic for proteins, due to their size.¹⁶ However, it is shown that the 3-dimensional structural data of proteins may be recovered,^{17–19} which potentially makes it a powerful (future) method. The bottom-up approach utilizes a sequence-specific protease to proteolytically hydrolyse the peptide bonds of certain amino acids (digestion). After digestion, the protein fragments (peptides) are analysed and the protein is reconstructed.¹³ Peptides are smaller than proteins and many are slightly polar, making them relatively easy to ionise.¹² However, the polar character of many peptides limits the phase transition between aqueous and gas, making the approach non-comprehensive, *i.e.* not all peptides are detected equally.¹²

1.1.2. The Metabolome and Metabolomics

Endogenous and exogenous substances that are involved in, or produced from, metabolomic processes are called metabolites.²⁰ Because they are involved in (nearly) all biological processes, they are a prime targets for disease recognition and fundamental in *systems biology*^{iv}.²¹ Metabolites include nucleotides, lipids, amino acids and carbohydrates, however, can theoretically contain virtually any

^{iv} Field of study that focusses on interactions between molecular components.²¹

chemical property, making the development of a general method for analysis extraordinarily difficult.²¹ While the metabolome, collective name for all metabolites in an organism, is highly complex, it can also reveal an abundance of information on patient health. The field that studies the regulation and function of metabolites in an organism is called metabolomics.²⁰

Metabolite regulation is especially informative for patient health, it is therefore unsurprising that it was one of the first *-omics* techniques to be put into clinical practise.²¹ A good example of metabolomics in clinical practise is the screening of newborns (NBS),²² which was widely implemented, even before the field of metabolomics gained its name. In the case of NBS, the screening procedure might include the detection of the activity of phenylalanine hydroxylase (PAH), which is an enzyme that catalyses the hydroxylation of phenylalanine to tyrosine. Screening methods may target the enzyme PAH using an activity-based assay, *e.g.* observe conversion of phenylalanine to tyrosine or the direct detection of phenylalanine and tyrosine levels in blood.²³ While the former method does not detect the metabolites, it does provide important insights to the metabolomic pathway. Methods that target specific metabolites, or their known pathways, are referred to as first-generation methods.

The field has evolved to target a larger set of metabolites in high-throughput, using untargeted methods such as LC-MS and gas-chromatography mass spectrometry (GC-MS), or nuclear magnetic resonance (NMR) spectroscopy.^{21,24} MS based methods are usually single-stage,²⁵ and work similarly to those described in the preceding section, however, direct identification of compounds is complicated by the sheer number of metabolites.^{21,24,25} Nonetheless, a *m/z* profile can be correlated to certain disease markers due to up and down regulation of certain observed masses.²¹ The utilisation of either LC or GC separation prior to MS is dependent on the targeted metabolites, *e.g.* volatile compounds are separated using GC, and are not discussed separately.

As described in the previous section, the ionisation of certain compounds is complicated for polar molecules, which include, *e.g.* carbohydrates and organic acids. The structural determination of these types of (purified) compounds is quite usually performed using NMR spectroscopy.²⁶ This technique is highly complex to understand in detail and out of the scope of this thesis. However, in essence, compounds are characterised in minute detail based on their nuclear resonant frequencies with the magnetic spectrum.²⁶ NMR spectroscopy provides interesting opportunities for metabolomic profiling as it can be used to characterise, localise and obtain temporal information of metabolites *in vivo* using Magnetic Resonance Imaging (MRI).²⁷ Nonetheless, analysis of both NMR and MRI spectra of complex mixtures is difficult as the signals received are highly multiplexed.^{24–27}

1.1.3. Single-Molecule Sequencing

Single-molecule analysis is a new emerging field of science, allowing the detection of analytes on the smallest biologically-relevant scale.²⁸ Whereas most methods require an ensemble of molecules to interact with the detector, single-molecule techniques require theoretically only one.

In recent years, single-molecule analysis has excelled in oligonucleotide sequencing, maturing into commercial products capable of sequencing long stretches of oligonucleotides. These techniques are collectively known as third-generation (DNA) sequencing, not to be confused with NGS which also includes second-generation sequencing techniques. At the time of writing, two third-generation methods have reached the stage of commercialisation, namely: zero-mode waveguided sequencing by synthesis, and nanopore based sequencing.

The first technique, zero-mode waveguided sequencing by synthesis, utilises nanostructures smaller than the wavelength of light in order to reduce the detection volume to less than the diffraction limit of visible light.²⁹ A polymerase is immobilised in the nanostructure, which elongates single stranded DNA using fluorescently labelled nucleotides.³⁰ The nucleotides are added in 0.1 to 10 μM range, such that (on average) there is only a single nucleotide present in the sampling volume which is in the atto- to zeptoliter range (10^{-18} - 10^{-23} L). The nucleotides are excited using a laser, causing nucleotide emission which is optically detected. The time of emission is an indication of binding, as unbound nucleotides enter and exit the sampling volume more rapidly. The laser excitation is, however, the main limitation of the system, as this causes photo-oxidation of the polymerase, rendering it inactive over-time.

Contrary to other sequencing techniques, nanopores provide a polymerase- and label-free platform. The principle working of nanopore based sequencing is similar to nanometer-sized Coulter counters. They allow the differentiation of nucleotides (and other molecules)—due to ion displacement—based on the size and interaction of the analyte with the nanopore and its environment.³¹ For the purpose of oligonucleotide sequencing, an ATP-dependent helicase (*e.g.* Phi29) is brought on top of a nanopore (*e.g.* an engineered α -hemolysin or *Mycobacterium smegmatis* porin A nanopore).³² The nucleobases are fed through the pore using the helicase, controlling speed by ATP concentration, and the nucleobases are read as a decrease in ionic current. Nucleobases translocating through the nanopore are not measured one-by-one. At any given time, there will be several nucleobases excluding ionic current, such that a chain of nucleobases is measured. Sequencing is performed by re-constructing the most probable sequence of nucleotides causing the excluded current over time.

1.1.4. Towards Nanopore Protein Sequencing

The independence of polymerase and labels makes nanopores a promising technique for sequencing of proteins, analogously to DNA sequencing. However, the complexity of the nanopore signal is increased exponentially. DNA is composed from 4 nucleobases with roughly the same size and equal charge, which can be easily translocated through helicases. Proteins can consist of (usually) 20 amino acids with widely diverging sizes and charges. Thus, they are not easily translocated using motor-proteins. In fact, the uniform charge of DNA allows the strands to translocate even without the presence of a helicase, while proteins require exact conditions to give them a uniform charge.³³ As base calling is performed founded on the moving average, while an oligopolymer is translocated, the difficulty to extract sequence information changes from 4^n to 20^n , with n being the number of bases or potential proteinogenic amino acids in the sensing region of the pore. Thus, protein sequencing requires much more sensitive nanopores.³⁴

1.2. Nanopores

Figure 1. Schematic representation of analyte translocation through a biological nanopore. **A:** shows an isolating membrane, including a lipid bilayer into which a nanopore is inserted. **B:** shows the translocation of a particle through a (Fragaceatoxin C) nanopore. The resistance across the membrane is stable and dominated by the unblocked nanopore (I). During translocation (II), the observed resistance is increased. After translocation (III), the resistance across the membrane is re-stabilised to the resistance of the unblocked nanopore.

A detailed description of the experimental setup used throughout this thesis is described in **Section 2.4**. In brief, the resistance across an electrolyte-rich solution, separated by a non-conductive membrane, is measured (**Figure 1A**). The non-conductive membrane contains a nanometer-sized gap ($\varnothing \sim 10^{-9}$ m, *i.e.* nanopore), which allows ion flux to pass through. The resistance measured across the membrane is dominated by the nanopore (**Figure 1B-I**). When an analyte traverses the nanopore—it excludes or in other ways blocks ions, resulting that—a change in the ionic flux is observed (**Figure 1B-**

II). After translocation, the resistance is stabilised back to the resistance over the nanopore (**Figure 1B-III**). Nanopores can be divided into two classes: biological (**Section 2.1**) and solid-state (**Section 2.2**).

1.2.1. Biological Nanopores

Biological nanopores are (self-assembling) water-filled nano-scale apertures that are incorporated on an amphiphilic membrane.³⁵ These “nanopores” are formed using transmembrane proteins or peptides, however, can also be formed using folded DNA strands.^{36,37} The former class is especially used in present nanopore sensors due to their ease in construction and abundance of easily incorporable functionalities. The working of both protein and peptide based nanopores is similar. Usually, several identical subunits (monomers) self-assemble into tubular shapes, creating ring-like shapes with a hollow centre. The functional element of nanopore insertion is controlled by the transmembrane region made of a hydrophobic peptide strands that can traverse a lipid bilayer to form a channel. Peptide based nanopores are not discussed separately.

Figure 2. Biological nanopore interface of Fragaceatoxin C. The structure of Fragaceatoxin C (FraC, PDB: 4TSY) is shown as it is inserted in a lipid membrane. The lumen facing positions are indicated as facing towards the aqueous phase. The transmembrane domain is shown with an orange overlay. The soluble domain is shown with a blue overlay.

Hydrophobic proteins are generally not soluble in aqueous solutions. However, protein based biological nanopores generally contain a soluble domain attached on-top of the transmembrane region (**Figure 2**). These nanopores are often assembled (oligomerised) outside the measurement chamber, for example, on liposomes and extracted using a surfactant.^{38,39} After oligomerisation, these proteins

can generally be kept stable in solution by using a detergent to mimic the hydrophobicity of the cell membrane.⁴⁰ The residues that are oriented towards the interior are called the “lumen” (**Figure 2**). The lumen is of major importance to the functionality of the nanopores, as the narrowest part of the nanopore is determined by these residues. Modification of lumen facing residues is not trivial, however, a generic distinction can be made based on the secondary structure of the transmembrane region, either α -helical or β -barrel.

Biological nanopores with a β -barrel transmembrane region are commonly used in commercial sensors; as the strands that make this domain are packed to form a stable cylindrical shape. The inner diameter of the pores described in this section is in the range of 1-2 nm, making them ideal for the detection of single-stranded DNA.⁴¹ Especially, β -barrel nanopores such as α -hemolysin (α HL, **Figure 3A**) and *Mycobacterium smegmatis* porin A (MspA, **Figure 3B**), have proven useful for this purpose.^{42–45} Both α HL and aerolysin (AeL, **Figure 3C**) can also be utilized for the detection of peptides and proteins with high specificity.^{46–48}

Figure 3. β -barrel biological nanopores. **A:** Alpha hemolysin (α HL, PDB: 7AHL). **B:** *Mycobacterium smegmatis* porin A (MspA, PDB: 1UUN). **C:** Aerolysin (AeL, PDB: 5JZT)

The second class of protein based biological nanopores contain an α -helical transmembrane region. These helices are generally flexible and commonly consist of a V-shape in their transmembrane region.

Approximately every seventh amino acid in an α -helix faces the same position, with an average turn every 3.7 amino acids, or a distance of 5.4 Å (0.54 nm) between lumen facing residues.⁴⁹

There are few α -helical pores in use as nanopore sensors, when compared to their β -barrel counterpart; namely, cytolysin A (ClyA) and fragaceatoxin C (FraC). The ClyA nanopore contains a long tube-like soluble region on top the transmembrane helices, with a diameter of approximately 5.5 nm; while the transmembrane lumen are a mere 3.3 nm (**Figure 4A**).⁵⁰ The differences in these dimensions allow it to be used as a nanometer sized test-tube, with the ability to trap (small) folded proteins in its interior.^{51–53} The FraC nanopore, however, contains a compact soluble region with an entrance of approximately 6 nm, narrowing down in a V-shaped structure to 1.6 nm at its narrowest (**Figure 4B**).⁵⁴ This funnel-like design is shown to be suitable for analysis of smaller entities, such as DNA strands and peptides.^{33,52,55}

Figure 4. α -helical biological nanopores. A: Cytolysin A (ClyA, PDB: 2WCD). **B:** Fragaceatoxin C (FraC, PDB: 4TSY).

1.2.2. Solid-State Nanopores

Singular nanopores can be produced fastidiously in—non-conductive—rigid materials, appropriately named solid-state nanopores. These materials include, but are not limited to, silicon, aluminium oxide, graphene, molybdenum disulphide and hexagonal boron nitride.^{56–63} The last three are especially

useful, as these can produce membranes with a single-layer (or few-layers) of atoms.⁶⁴ These materials are also referred to as 2D materials as they can be represented as a single sheet with little to no three-dimensional structure. Nanopores in 2D materials are advantageous because they theoretically could allow single amino acid resolution, as the nanopore volume is determined by both the diameter of the pore and the thickness of the membrane.^{65,66}

Numerous techniques exist for the creation of solid-state nanopores. One approach is to “drill” a nanometer-sized hole using external force, for example, by utilizing transmission electron microscopy, helium ion microscopy or dielectric breakdown.⁶⁷ These approaches can create nanopores with a diameter of 1 nm or less.⁵⁷ However, they are hard to fabricate reproducibly. Instead, another common method is used that involves the (chemical) etching of the nanopore inside the material.^{67–71} For this method, a photoresist is applied and cured using lithography, and non-cured areas are etched away. After nanopore creation, the inner diameter can be shrunk down to several nanometers using atomic layer deposition.^{58,72,73}

Solid-state nanopores—contrary to their biological counterpart—are independent of biocompatible solvents as they can be produced to be chemically inert. This property makes them—in theory—useable in organic solvents, where biological nanopores would lose their functionality. They also provide superior resistance to physical perturbation when compared to biological nanopores as the membrane is constructed of rigid materials instead of loosely bound lipids. Solid-state nanopores, however, are currently hard to produce on the sub-nanometer scale, making reproducibility a major concern.⁷⁴ Another down-side to the use of solid-state nanopores is the difficulty to engineer chemical functionalities on the scale of biological nanopores (single amino acid).

1.2.3. Biological Nanopores: Resistive Pulse Sensing

Resistive-pulse sensing (RPS) is commonly used to measure molecules traversing (biological) nanopores in aqueous solutions, allowing the native detection of biological analytes.⁷⁶ A “standard” setup for planar lipid bilayer electrophysiology (**Figure 5A**) consists of a two-compartment chamber separated by a non-conductive hydrophobic membrane, usually made from polytetrafluoroethylene (PTFE) and clamped together using two screws. The membrane is punctured to generate a 100 μm aperture, over which a lipid bilayer can be formed.⁷⁵ Using the Montal-Meuller method,⁷⁷ a droplet of *n*-hexadecane, dissolved in a volatile solvent (*e.g.* pentane), is applied to the aperture of the non-conductive membrane, and the solvent is allowed to evaporate. Both compartments are filled with an aqueous, electrolyte rich, buffer solution and topped with a solution of protein-specific lipid in a volatile solvent, which is also allowed to evaporate. The aqueous solution is pipetted out of, and slowly

returned to, the chamber, covering the aperture with a lipid bilayer film (**Figure 5B**). This process is repeated to reform the bilayer. Silver/silver chloride (Ag/AgCl) electrodes are immersed into both compartments and a small potential is applied to measure the resistance across the chamber. Due to the low conductance between the chambers, direct emersion of the Ag/AgCl electrodes into the buffer is possible, however, salt-bridges must be used when the buffer solution reacts with the electrodes,⁷⁸ when the electrolyte concentrations between the chambers is different, or if the salt concentration is low. Salt bridges can be placed between the electrode and the buffer, they contain an inert solid phase (e.g. agarose) containing 3 M KCl to ensure a non-biased interface between the solution and the solid.

Figure 5. Drawing of a typical electrophysiology chamber. **A:** shows the drawing of a typical electrophysiology chamber and **B:** shows the formation of a planar lipid bilayer, based on Maglia et al.⁷⁵ All components are fabricated using polyoxymethylene unless specified differently. The dotted lines represent components fabricated inside the chamber. Components are shown once, however represent all components of the same time, with; **1:** Top view, **2:** Side view (XZ-axis), **3:** side view (YZ-axis), **4:** 20 µm polytetrafluoroethylene film, **5:** electrode hole, **6:** perfusion hole, **7:** compartment, **8:** screw hole, **9:** O-ring, **10:** aperture of 100 µm in diameter punctured into 4, **11:** lipid monolayer, **12:** *n*-hexadecane cone (oil), **13:** lipid bilayer.

The compartment connected to the working electrode is referred to as the *trans* side, while the compartment connected to the ground electrode is called the *cis* side. To perform single molecule nanopore RPS, a single nanopore must be inserted into the bilayer. A highly diluted solution of biological nanopores is added to the *cis* compartment to gain pores oriented with the electric field. An exact concentration cannot be given, as only a single molecule is required to form the nanopore. However, experimentally, the nanopore solution is diluted to gain nanopore insertion every few minutes, allowing abundant time to perfuse the compartment with fresh buffer solution.

1.2.4. Biophysical Properties

Unguided analyte translocation through a nanopore can be described, for the most part, as the conjunction of the electrophoretic force (EF) and the electro-osmotic flow (EOF). The EF is a direct result from the voltage that is applied over the membrane, which orients the analyte with the electric field that is formed inside the nanopore. Importantly, the electric field near the nanopore is very large due to the high resistance of the nanopore.⁷⁹ This will migrate the analyte towards, or away, from the nanopore entrance, which can be effectively used to capture and translocate uniformly charged analytes, such as DNA,^{55,80–82} while providing possible adverse effects on non-uniformly charged analytes such as proteins; *i.e.* some proteins will migrate towards the nanopore, while oppositely charged proteins will migrate away from the nanopore and neutral proteins will not be captured.

Figure 6. Schematic drawing of the electric double layer. Shows the migration of positively charged ions along a negatively charged surface under a large applied potential. The presence of positively charged ions is augmented near the surface. The electric double layer (EDL) is marked, equally the outer Helmholtz plane (OHP) is represented as a line through the center of the mobile ions.

The EOF, however, is a result of the Coulombic force of ions interacting with the nanopore lumen, which form an electric layer on this liquid-solid interface. Most nanopores display a number of charges on their channel walls, which augments the attraction of oppositely charged ions, resulting in a diffusive layer, also known as an electric double layer (EDL).⁸³ The net charge is forced to move when an electrical potential is applied, and the mobile layer will drag (water) molecules along its surface, resulting in a net force in the same direction as this diffusive layer.⁸⁴ The magnitude of the drag is proportional to thickness of the outer Helmholtz plane (OHP, see **Figure 6**), which is directly depended on the radius of hydration of ions in the diffusive layer and is inversely proportional to the Hoffmeister

series ($\text{Li}^+ > \text{Na}^+ > \text{K}^+$).⁸⁵ Advantageously, the EOF can be tuned in (biological) nanopores by mutation of charges on the nanopore surface (**section 3**),⁸⁶ or by changing the pH,⁵² or ionic composition of the buffer.⁸⁷ For example, if the pH is lowered, the negative charge on glutamic- and aspartic acid and cysteine can be reduced; similarly, when the pH is increased, the positive charge on lysine, arginine, and histidine can be diminished (**Figure 1**). Equally, substitution of the electrolyte can significantly alter the observed EOF as well as the usage of asymmetric salts.^{88,89}

1.2.5. Data Interpretation

The analysis of the resulting current traces from nanopore RPS is crucial. A representative signal involves a single nanopore baseline and closing event(s) as shown in **Figure 7A**. The first (pre-processing) step in trace analysis involves the detection of the blocking events (**Figure 7B**). Several methods can be used to detect these events, however, the resulting information is the same.³⁴

Figure 7. Representative trace analysis resulting from biological nanopore resistive pulse sensing. **A:** shows a part of biological nanopore resistive pulse sensing trace and the 0 current line. **B:** Shows the same trace as A with an ideal flat-top shape shown around, as well as the annotation for the baseline current of the nanopore (I_0) and the current of the event (I_B). **C:** Shows the same trace and annotation as B with additional annotation for the dwell time of the events and the difference between the I_B and I_0 (ΔI_B).

Several important parameters can be extracted; most notably the difference between I_B and I_0 (ΔI_B) and dwell time (**Figure 7C**). The ΔI_B is used to determine the relative blockade of the event with regard to the open pore, denoted as excluded current ($I_{ex}\%$), using **equation 1**. It is equally common to describe this characteristic as residual current ($I_{res}\%$), which is the complement to 100% of the $I_{ex}\%$. The $I_{ex}\%$ of an event is characteristic for the size of the traversing analyte, whilst the dwell time is related to the translocation speed. During non-controlled translocation, the dwell time provides information regarding the pore-analyte interaction.

$$I_{ex}\% = \left(\frac{\Delta I_B}{I_0} \right) * 100\% \quad (1)$$

In practical applications, such as DNA sequencing, additional information is extracted from the fluctuations observed within the event.³⁴ For example, during DNA sequencing, multiple bases are read at any given moment as discussed before. The interplay between the observed signals is used to reconstruct the original fragments. However, the number of possible current states during protein and metabolite detection is much lower than the number of theoretical proteins and metabolites. Therefore, the proteins and metabolites cannot be qualified using the same analogy. It has been proposed that additional information of independently translocating analytes can be extracted. This can be achieved by, *e.g.* relating current fluctuations in the current signal to the dipole and shape of the analyte or probing the driving force required for translocation^{90–92}. Nonetheless, the sample complexity makes direct identification extraordinarily difficult.

1.3. Targeted Engineering of the Channel

One of the earliest and possibly most influential nanopore, alpha hemolysin (α HL) from *Staphylococcus aureus*, had been studied for 10 years prior to the publication of its crystal structure.^{93,94} This pore is amiable for engineering, as shown by Cheley *et al.*, who showed that the transmembrane domain of α HL could be reversed while retaining membrane spanning properties.⁹⁵ Interestingly, by adjusting the internal charge of the nanopore, the electrostatic interactions can be tuned. In 2000, Gu *et al.* showed that the anion-selectivity could be augmented by non-covalently bound β -cyclodextrin.⁹⁶ As a rule of thumb, when the internal charge is modified to include a positive inner surface, the capture of negatively charged DNA is augmented.^{55,80–82} Similarly, when the internal charge is modified to contain a negative inner surface, the capture of positively charged proteins and peptides is increased.^{33,97–99}

The rational design of the transmembrane region of a nanopore, and addition of chemical functionality, requires the determination of the constriction site(s) which varies depending on the structure of the transmembrane domain. Movileanu *et al.* determined the constriction region experimentally, due to the absence of a crystal structure, with a combination of polymer penetration and cysteine mutagenesis.¹⁰⁰ Polyethylene glycol (PEG) of different sizes, and functionalized to interact with cysteines inside the pore lumen at different depths, were used to explore such constriction sites. Upon binding of PEG with cysteines in the nanopore, an increased dwell time and excluded current could be observed near the constriction. Thus, by introducing cysteine residues along the

transmembrane region of the nanopore, the constriction site(s) can be determined from the residence time of the (translocating) polymers.

Nowadays, many three-dimensional structures are available for study. It is therefore beneficial to localize the exact constriction site(s) by using systematic mutations near the narrowest region based on the crystal structure. As translocation of ligands mainly depends on the EF and EOF (**section 1.2.4**) near the constriction,^{79,97,99} mutagenesis of these residues to include a neutral, negative, and/or positive charged group gives a good indication of the sensing region. This method has been used to characterize, for example, the sensing region of the aerolysin nanopore.^{81,101}

While the structure of most nanopores is unique, a notable distinction can be made between β -barrel and α -helical nanopores. β -barrel pores can be subdivided into two types based on their three dimensional structure, either a long rigid tube or goblet-like structure. Similarly, α -helical pores can be subdivided by the structure of the α -helix, either a single strand traversing the membrane, or α -helices that loop down across the membrane and stretching back towards the soluble domain. In addition, nanopores can be classified by their pore forming mechanism, either oligomers formed from (identical or non-identical) monomeric subunits, as well as monomeric pores. While monomeric nanopores may have advantages for certain uses, oligomeric nanopores are, generally, more amiable for engineering.

The three-dimensional structure of the transmembrane domain of oligomeric β -barrel pores such as α HL and aerolysin contain a rigid beta-barrel tube with (near) equal resistance in each direction. This makes it that the constriction site is not exactly localised, rather, it is defined as a range of lumen-facing positions. Specifically, α HL is a heptameric nanopore that can be described as having an hourglass shape, with a vestibule on the *cis* entrance ($\varnothing \approx 2.5$ nm), a narrow constriction ($\varnothing \approx 1.4$ nm) in the center, followed by a rigid beta-barrel ($\varnothing \approx 2$ nm).¹⁰² The pore constriction sites, located at residues E111-K147-M113, are amiable for engineering. Mutagenesis of M113 to arginine augments the capture of DNA, whereas substitution to negatively charged aspartic acid on M113 removes DNA capture altogether.⁸² Additional mutagenesis to include V124L, G130S, N139G and I142L (α HL-M113R-RL2) allows efficient RNA sequencing; while a mutant containing E111N, K147N, and M113Y provides enhanced hydrogen bonding and aromatic stacking with translocating bases, increasing the nucleobase discrimination of the pore.¹⁰³

Another well studied β -barrel nanopore, Aerolysin (AeL) from *Aeromonas hydrophila*, is an heptameric nanopore containing two constriction sites at R282-R220 and K238-K242.^{81,104,105} The monomers form a heptameric structure with a vestibule on the *cis* entrance ($\varnothing \approx 3$ nm) and two concentric β -barrels

with an inner diameter of 1.37 nm and a length of ~10 nm.^{48,104,106} In 2018, Wang *et al.* used molecular dynamic (MD) simulations to determine the sensing spot of K238, and they showed that the translocation time of DNA could be augmented by introducing a glycine on this position which, in addition, allows the discrimination of methylated cytosines from their non-modified counterpart.¹⁰⁵ In 2019, Cao *et al.* utilized MD simulations in combination with an extensive amount of single-point mutations on the sensing regions of AeL and found several interesting properties.¹⁰⁴ Indeed, they found that mutagenesis of the K238 to glutamine allowed an increased dwell time of translocating analytes due to the reduced diameter in the constriction site. Additionally, when the sensing site R282 was modified to R282A, or if R220 was modified to an alanine (R220A) or tryptophan (R220W), analyte capture was augmented.

The second class of β -barrel pores with a thin constriction site, such as MspA are especially amiable for the sequencing of DNA. MspA pores consist of goblet-like shapes with a narrow constriction (1.2 nm and 1.3 nm respectively) defined by residues D90, D91, and D93. A modified MspA mutant (MspA-m2, D90N/D91N/D93N/D118R/D134R/E139K) allows the efficient capture and detection of DNA, while, the incorporation of gold ions inside MspA-m2 modified with D91M allows the discrimination of biothols.¹⁰⁷

Nanopores with a single traversing α -helix disallow the introduction of multiple constriction sites. The α -helix is located at the beginning (or end) of the sequence, making it amiable to a more flexible mutagenesis, compared to its β -barrel counterparts, as it mostly depends on the membrane spanning and α -helix forming properties. Advantageously, the α -helix makes a turn approximately every 3.7 amino acids such that residues that face the same direction are stacked 0.54 nm in the orthogonal plane.⁴⁹ The most studied α -helical pore of this type, fragaceatoxin C (FraC) from *Actinia fragacea*, consists of a multimeric nanopore, which allows the formation of nanopores with varying diameter. When the positions W112 and W116 of the FraC nanopore are adjusted to serine,³³ a large percentage of the resulting oligomers form as hexa- or heptamers. This reduces the recognition volume of the constriction to approximately 1 nm³ and 1.8 nm³ respectively, while the octameric form has a recognition volume of approximately 3.6 nm³. Early work with this pore focussed on the adaptability towards DNA recognition, and found that modification of residue D10 to arginine allowed such detection.⁵⁵ Surprisingly, the wild-type FraC nanopore enables the identification of peptides at a low pH, displaying a good correlation between the peptide mass and excluded current at the exact pH of 3.8.³³ In the same study, it was suggested that the recognition volume should be encompassing the residues D10 to D17. Advantageously, the single traversing α -helix allows mutagenesis to a biomimetic

protoxin by genetically fusing a globular protein with a protease specific cleavage site to the N-terminus. This removes the ability of FraC to form pores; however, when the globular protein is cleaved by a specific protease, the activity of the nanopore can be recovered.¹⁰⁸ Additionally, the C-terminus of FraC is also amiable for engineering, allowing the inclusion of a nanobody to this end for increased lipid association.¹⁰⁸ In addition, small molecules can be covalently attached to the nanopore. By inclusion of cysteine groups, especially on residues W112, E134, or Y138, a photo-switchable azobenzene compound could be coupled using a reactive chloroacetyl moiety. In a study by Mutter *et al.*, the selected cysteine residue Y138C was orientated towards the membrane, only enabling the formation of pores when irradiated with a wavelength of 365 nm.¹⁰⁹

The second class of α -helical nanopores have α -helices that loop back towards the soluble domain. Mutagenesis of the transmembrane region is less trivial than the single traversing counterpart, as it is more involved in protein folding of the monomeric chains. However, the small constriction provided by the transmembrane region, combined with a rigid α -helical tube attached on-top, allows it to be utilised as a nanometer sized test tube. This is especially useful when combined with binding proteins, which allows the single-molecule investigation of their binding mechanics, and allows their use as single-molecule biosensors. The most common pore of this type is cytolysin A (ClyA) from either *Staphylococcus aureus* or *Escherichia coli*. The *S. typhi* variant can be described as a 13 nm long tube ($\varnothing \approx 7$ nm) followed by a narrow transmembrane ($\varnothing \approx 3.8$ nm).⁵⁰ Interestingly, the two cysteines at positions C87 and C285 contained within this structure could be modified to serine, which makes this pore especially amiable to engineering.¹¹⁰ Incorporated cysteines may be problematic during protein purification, due to oxidation, and cysteine free mutants allow the incorporation of unnatural amino acids using this codon. Beneficially, small pieces of DNA that bind to specific ligands (aptamers) can be covalently attached through a reactive cysteine. This allows the pore to distinguish a mixture of different analytes with high specificity.¹¹⁰

1.4. Protein Detection using Nanopore Spectrometry

The approach taken by nanopore spectrometry^v is conceptually similar to mass spectrometry. During nanopore spectrometry a nanopore is utilised to estimate the excluded volume of traversing analytes. Robertson *et al.* noticed that the excluded current caused by translocation of different lengths of polyethylene glycol (PEG) can be directly correlated with the mass of the translocating analyte.^{111,112}

^v Named by convention, not based on interaction with the electromagnetic spectrum.

This method was referred to as nanopore mass spectrometry^{vi}. However, the excluded current is only related to the mass of the translocating PEG due to the homopolymeric nature of PEG, and the method has since been renamed as single-molecule nanopore spectrometry (SMNS).¹¹³

Proteins and peptides are not homopolymeric. Their size depends not only on the length of their chain, but also the three-dimensional structures they can form. Additionally, the diverging charges and chemical functionalities on these entities result in heterogenous capture by the nanopore. These properties make the extraction of single amino acid information difficult, however, allows individual protein characterization by their shape, volume, rotational diffusion coefficient, dipole and charge.^{90,91} The occupied volume of proteins can be readily estimated from the excluded current,⁹¹ and the residence time of a translocating protein gives an indication of charge.^{114,115} Interestingly, contributions from Mayer *et al.* showed that the dipole of proteins can be sampled from the fluctuation observed in the excluded current distribution when proteins are sampled in a canonical nanopore.^{90,91} In brief, they showed that the fluctuations observed in the blocked current of translocating proteins can be explained based on the shape of the translocating protein. As a protein tumbles inside the nanopore, it results in fluctuations in the residual current, as each conformation has a slightly different ion exclusion. These conformations can be roughly approximated by a spheroid, which can be used to explain the current fluctuations. While proteins are not spheroids, the shape is within the measurable error, *i.e.* experimentally the shape of a protein is indistinguishable from a spheroid.

The general consensus for the sequencing of proteins is to follow the strategy set by DNA sequencing. In brief, it is envisaged to control the transport of unfolded proteins across a nanopore and subsequently reading the individual (or few) amino acids as they translocate across the sensing region. Especially the diverging charges on proteins have proven problematic, resulting in stochastic capture. One method to homogenise the charges on proteins is to shield their exposed charges using sodium dodecyl sulphate (SDS).⁷⁶ While the chemical shielding and unfolding seems highly beneficial, the presence of a large SDS group (288 Da) clouds the chemical information of singular amino acids, making sequencing using this method infeasible. In another approach, linearised proteins are attached to a piece of DNA.¹¹⁶ The negatively charged DNA acts as a carrier to uniformly translocate itself and the attached proteins across the nanopore, thereby recognizing and discriminating the attached proteins. However, similarly to the SDS method, it clouds the chemical information of singular amino acids due to the presence of these bulky groups.

While the analysis of fully folded proteins seems overly complex, a similar approach can be taken as commonly used in bottom-up MS based proteomics. In brief, a protein is digested using a specific

^{vi} Nanopore mass spectrometry now refers to nanopores coupled to mass spectrometers.

protease, and the resulting peptide fragments are used to identify the original protein. The peptides resulting from digestion have a lower dynamic mass range and it is therefore likely that their volumes are similar as well. Early work by Lee *et al.* showed that homopolymeric peptides could be readily discriminated from one another using either, the α HL or the AeL nanopore.

The charges on naturally occurring peptides are, however, non-uniform, resulting in different capture mechanisms for each peptide. This has proven to be a difficult problem to solve, and many contributions have used unnatural (chemical) modifications to achieve controlled translocation while tuning the sensitivity of the nanopore.

The electric field can be used to retard the transport across a nanopore. Designed peptides containing homopolymeric stretches of charges at the C- and N-termini. Examples included the use of peptides with a negative glutamine or aspartic acid chain on the N-terminus and a positive arginine or lysine chain on the C-terminus.^{117–119} The introduced dipole balanced the peptide at the centre of the nanopore near the sensitive region of the nanopore and differences between singular amino acid substitutions, containing either a neutral, hydrophobic or hydrophilic side chain, were determined.^{120,121} It was also shown previously that such designed peptides allow the detection and localisation of post-translational modifications.¹¹⁸ Indeed, addition of a charged group to carry the peptide along the nanopore is beneficial, even if only a cationic carrier-arginine heptapeptide is used. Ouldali *et al.* showed that such a carrier peptide can discriminate 13 out of the 20 natural amino acids when translocating through the (overall negatively-charged) aerolysin nanopore.⁴⁶

Figure 8. Structure and pK_a of charged amino acids and effect on peptide translocation in Fragaceatoxin C. A: structure and pK_a of charged amino acids. The pK_1 refers to the carboxyl group and pK_2 to the amine group of free amino acids. **B:** effective force applied on a peptide translocating a Fragaceatoxin C nanopore with (1) the electro osmotic flow and (2) the electro osmotic force.

Naturally occurring peptides, as resulting from a proteolytic digest are, however, not modified and will contain diverging charges. Still, the charge on peptides can be adjusted in a generic way by reducing the pH of the buffer solution and balancing the applied potential.^{122,123} As the pH is lowered, the charge contained on the negative residues, glutamic ($pK_a = 4.25$) and aspartic acid ($pK_a = 3.65$), is reduced or completely eliminated, while the positive residues—lysine ($pK_a = 10.53$), arginine ($pK_a = 12.48$) and histidine ($pK_a = 6.00$)—maintain (or gain) their charge (**Figure 8A**). Even when no positive residues are contained within a peptides, the positive charge on the N-terminus can remain charged. The reduction of pH may not always be compatible with biological nanopores, as they would experience the same effects on their charges. It is therefore remarkable, that Huang *et al.* showed that FraC remains efficient at capturing peptides at low pH (**Figure 8B**), more so, they demonstrated that the current blockades observed from peptides scale proportionally with the mass of the peptide, regardless of the chemical functionality of these peptides.⁵²

1.5. Adapter-based Metabolite Detection using Nanopores

Biologically-relevant endogenous and exogenous molecules can be measured using nanopores by either, 1) direct analysis^{52,93,113,120,124–127} or, 2) adapter-ligand interactions.^{51,53,86,128–133} The direct detection of small analytes in biological samples is very challenging, due to the vast number of different analytes with widely diverging chemical and physical properties. It is unlikely that all small analytes can be detected directly in a complex sample, as the signal complexity scales exponentially with the number of observable analytes. However, adapters that are specific towards one, or few, analytes can be utilised to greatly increase the selectivity of measurement. This principle was shown by the group of Bayley, who showed that cyclodextrin and cyclic peptide adapters lodged inside an α HL nanopore could lend their ligand specificity for the analysis of several organic compounds.^{129,133,134} In recent years, substrate binding proteins have been used instead for their large selectivity,¹²⁸ which provides the possibility to detect a wide variety of biological molecules with precision. In essence, a substrate binding protein lends its selectivity to the measurement while the nanopore provides the single-molecule sensitivity. Advantageously, many binding proteins have been well studied due to their large utilisation in commercial sensing, collectively referred to as ligand binding assays (LBA).¹³⁵ This reduces the time to production, as off-target selectivity and binding mechanics are generally well studied. Thus, the relatively low limit of detection can be provided by the nanopore, while selectivity towards the analyte can be provided by the adapter.

The general principle of this technique is shown in **Figure 8**. A binding protein is trapped inside of a nanopore with a long test-tube-like funnel, *e.g.* a cysteine free *S. typhi* ClyA nanopore (ClyA-AS, **Figure 9A**). A ligand is added, and when the binding protein undergoes a conformational change,⁵³ or changes position in the nanopore,⁵¹ the current is reduced (or increased) and the events are characterised (**Figure 9B**). Yet, as shown in the preceding section, the orientation of proteins inside the nanopore can change over time, and the protein can translocate quickly through the nanopore. However, test-tube-like nanopores such as ClyA can sterically prevent protein translocation, lodging proteins inside its inner compartment. Additionally, Van Meervelt *et al.* showed that the orientation of binding proteins in *S. typhi* ClyA-AS can be controlled by addition of (single-point) charges orthogonal to the dipole of the protein.⁵¹

Figure 9. Principle of binding protein based metabolite detection using nanopore resistive pulse sensing. A: shows a binding protein (PDB: 2QRY) inside a ClyA nanopore (PDB: 2WCD) and an expected signal (1). **B:** Shows the same as A, however, an analyte binding to the protein (thiamine monophosphate) has been added, the binding causes the observed current (1) to change over time (2).

Adapter-based nanopore detection can be utilised to determine the characteristics of metabolites, or to perform single-molecule enzymology. Nonetheless, from a technical view they describe the same type of system. Recently, Galenkamp *et al.* showed that it is possible to simultaneously sense both glucose and asparagine by capturing and detecting the conformational changes of glucose-binding protein inside a ClyA-AS nanopore.¹³⁰ Moreover, they presented that ligand-induced conformational changes could be observed of dihydrofolate reductase (DHFR) upon binding with nicotinamide adenine dinucleotide phosphate (NADPH) and methotrexate (MTX).^{53,131} They showed that DHFR can bind in four ground state conformations, whereas not all conformations could be observed in bulk-ensemble experiments. These observations allow, not only metabolite detection, but concurrently provide insights to the specific binding pathway between the enzyme and ligands.

In a related contribution, Li *et al.* showed that the single-molecule binding of maltose variants could be investigated by utilising a ClyA nanopore originally extracted from *E. coli*.¹³⁶ For this, they immobilised a maltose-binding protein (MBP) inside the confines of the nanopore and detected different current levels for the binding of α -maltose and β -maltose. Moreover, the group of Maglia recently showed that at least nine substrate binding proteins can be utilised to detect their respective ligands, however, it is highly likely that more SBDs will be identified for this purpose in the near future.

In summary, this section has shown that there are many approaches available for the detection of metabolites with high specificity by utilisation of nanopores. One such approach is taken in chapter 4, where a thiamine substrate binding protein (TbpA) is used to sample the changes in the protein behaviour confined within ClyA-AS.

1.6. Scope of the Thesis

As illustrated in this chapter, nanopores can be used to probe diverse analytes ranging from small molecules to large and bulky proteins. In this thesis, the properties of the Fragaceatoxin C (FraC) are modified to augment the capture and resolution of peptides, providing a platform for protein fingerprinting. Additionally, this thesis extends on the concept of metabolite detection using ClyA-AS, which can be utilised for the rapid detection of thiamine when combined with an empirical Bayesian method to extract the concentration within one minute in biological fluids.

Chapter 1 provides a general background on the detection of proteins, peptides, and metabolites using conventional techniques and introduces the concept of single-molecule analysis for these analytes. A complete description of the nanopore setup is provided, as well as a global overview of the interpretation of the data. Furthermore, a review is given on related research for the targeted engineering of nanopores, and the concept of nanopore spectrometry is defined. In addition, strategies for the investigation of metabolites using nanopores is shown displaying different approaches that may be taken for such quantification.

Chapter 2 investigates the engineering of Fragaceatoxin C (FraC) for the analysis of peptides. To this effort, the FraC nanopore is probed using single-point mutations near the (presumed) recognition site of the pore. The effect of charge and aromatic interactions in single channels are elaborated resulting in a nanopore with exceptional resolution and capture efficiency for peptide detection.

Chapter 3 provides a proof-of-concept nanopore based protein fingerprinting utilising a modified FraC nanopore augmented for the detection of peptides. Herein, a model for the conversion of mass spectra

to corresponding nanopore spectra is provided, which shows that the two methods show a good correlation between intensity profiles. Additionally, it shows that protein spectra of ten different proteins can be distinguished, providing a platform for protein spectral matching.

Chapter 4 presents a thiamine binding protein coupled nanopore sensor that displays large changes in the ionic current upon binding with its specific ligand, thiamine. In this chapter, the direct detection of thiamine in urine is shown, and an empirical Bayesian method is presented for the quantification of thiamine within one minute of measurement time of a single nanopore.

Chapter 5 gives the general discussion and outlook of the presented contribution. Specifically, the potential impact and implications of the results presented herein are reviewed. In addition, an outlook is presented for future challenges to be solved and expected advances in the near future.

1.7. References

1. Hood, L. & Rowen, L. The human genome project: big science transforms biology and medicine. *Genome Med.* **5**, 79 (2013).
2. Xu, C. & Jackson, S. A. Machine learning and complex biological data. *Genome Biol.* **20**, 76 (2019).
3. Suwinski, P. *et al.* Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics. *Front. Genet.* **10**, (2019).
4. Azad, R. K. & Shulaev, V. Metabolomics technology and bioinformatics for precision medicine. *Brief. Bioinform.* **20**, 1957–1971 (2019).
5. Schleidgen, S., Klingler, C., Bertram, T., Rogowski, W. H. & Marckmann, G. What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC Med. Ethics* **14**, 55 (2013).
6. Chen, R. *et al.* Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. *Cell* **148**, 1293–1307 (2012).
7. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet.* **30**, 418–426 (2014).
8. Berg, J. M., Tymoczko, J. L. & Stryer, L. Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains. in *Biochemistry, Fifth Edition* 1100 (W. H. Freeman, 2002).
9. Then, A., Mácha, K., Ibrahim, B. & Schuster, S. A novel method for achieving an optimal classification of the proteinogenic amino acids. *Sci. Rep.* **10**, 15321 (2020).
10. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
11. Omenn, G. S. Reflections on the HUPO Human Proteome Project, the Flagship Project of the Human Proteome Organization, at 10 Years. *Mol. Cell. Proteomics* **20**, 100062 (2021).
12. Watson, J. T. & Sparkman, O. D. *Introduction to Mass Spectrometry*. (John Wiley & Sons, Ltd,

- 2007). doi:10.1002/9780470516898.
13. *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*. vol. 919 (Springer International Publishing, 2016).
 14. Moradian, A., Kalli, A., Sweredoski, M. J. & Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* **14**, 489–497 (2014).
 15. *Proteome Research*. (Springer Berlin Heidelberg, 2007). doi:10.1007/978-3-540-72910-5.
 16. Bludau, I. & Aebersold, R. Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nat. Rev. Mol. Cell Biol.* **21**, 327–340 (2020).
 17. Wu, H. *et al.* Rapid 3-dimensional shape determination of globular proteins by mobility capillary electrophoresis and native mass spectrometry. *Chem. Sci.* **11**, 4758–4765 (2020).
 18. Li, H., Nguyen, H. H., Ogorzalek Loo, R. R., Campuzano, I. D. G. & Loo, J. A. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **10**, 139–148 (2018).
 19. Brown, K. A., Melby, J. A., Roberts, D. S. & Ge, Y. Top-down proteomics: challenges, innovations, and applications in basic and clinical research. *Expert Rev. Proteomics* **17**, 719–733 (2020).
 20. Liu, X. & Locasale, J. W. Metabolomics: A Primer. *Trends Biochem. Sci.* **42**, 274–284 (2017).
 21. Klassen, A. *et al.* Metabolomics: Definitions and Significance in Systems Biology. in 3–17 (2017). doi:10.1007/978-3-319-47656-8_1.
 22. Bodamer, O. A., Scott, C. R. & Giugliani, R. Newborn Screening for Pompe Disease. *Pediatrics* **140**, S4–S13 (2017).
 23. Mitchell, J. J., Trakadis, Y. J. & Scriver, C. R. Phenylalanine hydroxylase deficiency. *Genet. Med.* **13**, 697–707 (2011).
 24. Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **40**, 387–426 (2011).
 25. Tavassoly, I., Goldfarb, J. & Iyengar, R. Systems biology primer: the basic methods and approaches. *Essays Biochem.* **62**, 487–500 (2018).
 26. Richards, S. A. & Hollerton, J. C. *Essential Practical NMR for Organic Chemistry*. (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470976401.
 27. Emwas, A.-H. *et al.* NMR Spectroscopy for Metabolomics Research. *Metabolites* **9**, 123 (2019).
 28. Gyurcsányi, R. E. Chemically-modified nanopores for sensing. *TrAC Trends Anal. Chem.* **27**, 627–639 (2008).
 29. Levene, M. J. Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations. *Science (80-.)*. **299**, 682–686 (2003).
 30. Eid, J. *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Science (80-.)*. **323**, 133–138 (2009).
 31. Wanunu, M. Nanopores: A journey towards DNA sequencing. *Phys. Life Rev.* **9**, 125–158 (2012).

32. 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).
33. Huang, G., Voet, A. & Maglia, G. FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution. *Nat. Commun.* **10**, 835 (2019).
34. Rang, F. J., Kloosterman, W. P. & de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **19**, 90 (2018).
35. Ma, L. & Cockroft, S. L. Biological Nanopores for Single-Molecule Biophysics. *ChemBioChem* **11**, 25–34 (2009).
36. Burns, J. & Howorka, S. Structural and Functional Stability of DNA Nanopores in Biological Media. *Nanomaterials* **9**, 490 (2019).
37. Burns, J. R., Stulz, E. & Howorka, S. Self-Assembled DNA Nanopores That Span Lipid Bilayers. *Nano Lett.* **13**, 2351–2356 (2013).
38. Galenkamp, N. S. *et al.* Preparation of Cytolysin A (ClyA) Nanopores. in 11–18 (2021). doi:10.1007/978-1-0716-0806-7_2.
39. Mutter, N. L. *et al.* Preparation of Fragaceatoxin C (FraC) Nanopores. in 3–10 (2021). doi:10.1007/978-1-0716-0806-7_1.
40. Martin, J. & Sawyer, A. Elucidating the structure of membrane proteins. *Biotechniques* **66**, 167–170 (2019).
41. Bezrukov, S. M. & Kasianowicz, J. J. Current noise reveals protonation kinetics and number of ionizable sites in an open protein ion channel. *Phys. Rev. Lett.* **70**, 2352–2355 (1993).
42. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
43. Astier, Y., Braha, O. & Bayley, H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J. Am. Chem. Soc.* **128**, 1705–10 (2006).
44. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl. Acad. Sci.* **106**, 7702–7707 (2009).
45. Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci.* **105**, 20647–20652 (2008).
46. Ouldali, H. *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
47. Piguet, F. *et al.* Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* **9**, 966 (2018).
48. Stefureac, R., Long, Y., Kraatz, H.-B., Howard, P. & Lee, J. S. Transport of α -Helical Peptides through α -Hemolysin and Aerolysin Pores. *Biochemistry* **45**, 9172–9179 (2006).
49. Haimov, B. & Srebnik, S. A closer look into the α -helix basin. *Sci. Rep.* **6**, 38341 (2016).
50. Soskine, M., Biesemans, A., De Maeyer, M. & Maglia, G. Tuning the Size and Properties of ClyA Nanopores Assisted by Directed Evolution. *J. Am. Chem. Soc.* **135**, 13456–13463 (2013).

51. Van Meervelt, V. *et al.* Real-Time Conformational Changes and Controlled Orientation of Native Proteins Inside a Protein Nanoreactor. *J. Am. Chem. Soc.* **139**, 18640–18646 (2017).
52. 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).
53. Galenkamp, N. S., Biesemans, A. & Maglia, G. Directional conformer exchange in dihydrofolate reductase revealed by single-molecule nanopore recordings. *Nat. Chem.* **12**, 481–488 (2020).
54. Tanaka, K., Caaveiro, J. M. 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.* **6**, 6337 (2015).
55. 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. Chemie Int. Ed.* **55**, 12494–12498 (2016).
56. Li, J. *et al.* Ion-beam sculpting at nanometre length scales. *Nature* **412**, 166–169 (2001).
57. Storm, A. J., Chen, J. H., Ling, X. S., Zandbergen, H. W. & Dekker, C. Fabrication of solid-state nanopores with single-nanometre precision. *Nat. Mater.* **2**, 537–540 (2003).
58. Venkatesan, B. M., Shah, A. B., Zuo, J.-M. & Bashir, R. DNA Sensing Using Nanocrystalline Surface-Enhanced Al₂O₃ Nanopore Sensors. *Adv. Funct. Mater.* **20**, 1266–1275 (2010).
59. Fischbein, M. D. & Drndić, M. Electron beam nanosculpting of suspended graphene sheets. *Appl. Phys. Lett.* **93**, 113107 (2008).
60. Garaj, S. *et al.* Graphene as a subnanometre trans-electrode membrane. *Nature* **467**, 190–193 (2010).
61. Feng, J. *et al.* Identification of single nucleotides in MoS₂ nanopores. *Nat. Nanotechnol.* **10**, 1070–1076 (2015).
62. Danda, G. *et al.* Monolayer WS₂ Nanopores for DNA Translocation with Light-Adjustable Sizes. *ACS Nano* **11**, 1937–1945 (2017).
63. Zhou, Z. *et al.* DNA Translocation through Hydrophilic Nanopore in Hexagonal Boron Nitride. *Sci. Rep.* **3**, 3287 (2013).
64. Macha, M., Marion, S., Nandigana, V. V. R. & Radenovic, A. 2D materials as an emerging platform for nanopore-based power generation. *Nat. Rev. Mater.* **4**, 588–605 (2019).
65. Wang, Y., Yang, Q. & Wang, Z. The evolution of nanopore sequencing. *Front. Genet.* **5**, (2015).
66. Ni, Z. H. *et al.* Graphene Thickness Determination Using Reflection and Contrast Spectroscopy. *Nano Lett.* **7**, 2758–2763 (2007).
67. Yuan, Z. *et al.* Solid-State Nanopore. *Nanoscale Res. Lett.* **13**, 56 (2018).
68. Hildreth, O. J., Brown, D. & Wong, C. P. 3D Out-of-Plane Rotational Etching with Pinned Catalysts in Metal-Assisted Chemical Etching of Silicon. *Adv. Funct. Mater.* **21**, 3119–3128 (2011).
69. Huang, Z., Geyer, N., Werner, P., de Boer, J. & Gösele, U. Metal-Assisted Chemical Etching of Silicon: A Review. *Adv. Mater.* **23**, 285–308 (2011).
70. Lee, W., Ji, R., Gösele, U. & Nielsch, K. Fast fabrication of long-range ordered porous alumina

- membranes by hard anodization. *Nat. Mater.* **5**, 741–747 (2006).
71. Park, S. R., Peng, H. & Ling, X. S. Fabrication of Nanopores in Silicon Chips Using Feedback Chemical Etching. *Small* **3**, 116–119 (2007).
 72. Nam, S.-W., Rooks, M. J., Kim, K.-B. & Rossnagel, S. M. Ionic Field Effect Transistors with Sub-10 nm Multiple Nanopores. *Nano Lett.* **9**, 2044–2048 (2009).
 73. Chen, P. *et al.* Atomic Layer Deposition to Fine-Tune the Surface Properties and Diameters of Fabricated Nanopores. *Nano Lett.* **4**, 1333–1337 (2004).
 74. Ayub, M. & Bayley, H. Engineered transmembrane pores. *Curr. Opin. Chem. Biol.* **34**, 117–126 (2016).
 75. Maglia, G., Heron, A. J., Stoddart, D., Japrun, D. & Bayley, H. Analysis of Single Nucleic Acid Molecules with Protein Nanopores. in 591–623 (2010). doi:10.1016/S0076-6879(10)75022-9.
 76. Restrepo-Pérez, L., John, S., Aksimentiev, A., Joo, C. & Dekker, C. SDS-assisted protein transport through solid-state nanopores. *Nanoscale* **9**, 11685–11693 (2017).
 77. Montal, M. & Mueller, P. Formation of Bimolecular Membranes from Lipid Monolayers and a Study of Their Electrical Properties. *Proc. Natl. Acad. Sci.* **69**, 3561–3566 (1972).
 78. Janz, G. J. & Ives, D. J. G. Silver, Silver Chloride Electrodes. *Ann. N. Y. Acad. Sci.* **148**, 210–221 (1968).
 79. Willems, K. *et al.* Accurate modeling of a biological nanopore with an extended continuum framework. *Nanoscale* **12**, 16775–16795 (2020).
 80. Franceschini, L., Brouns, T., Willems, K., Carlon, E. & Maglia, G. DNA Translocation through Nanopores at Physiological Ionic Strengths Requires Precise Nanoscale Engineering. *ACS Nano* **10**, 8394–8402 (2016).
 81. Cao, C. *et al.* Mapping the sensing spots of aerolysin for single oligonucleotides analysis. *Nat. Commun.* **9**, 2823 (2018).
 82. Maglia, G., Restrepo, M. R., Mikhailova, E. & Bayley, H. Enhanced translocation of single DNA molecules through α -hemolysin nanopores by manipulation of internal charge. *Proc. Natl. Acad. Sci.* **105**, 19720–19725 (2008).
 83. Kurzweil, P. Electrochemical Double-layer Capacitors. in *Electrochemical Energy Storage for Renewable Sources and Grid Balancing* 345–407 (Elsevier, 2015). doi:10.1016/B978-0-444-62616-5.00019-X.
 84. Tavares, M. F. M. & McGuffin, V. L. Theoretical Model of Electroosmotic Flow for Capillary Zone Electrophoresis. *Anal. Chem.* **67**, 3687–3696 (1995).
 85. Nihonyanagi, S., Yamaguchi, S. & Tahara, T. Counterion Effect on Interfacial Water at Charged Interfaces and Its Relevance to the Hofmeister Series. *J. Am. Chem. Soc.* **136**, 6155–6158 (2014).
 86. Van Meervelt, V., Soskine, M. & Maglia, G. Detection of Two Isomeric Binding Configurations in a Protein–Aptamer Complex with a Biological Nanopore. *ACS Nano* **8**, 12826–12835 (2014).
 87. Kowalczyk, S. W., Wells, D. B., Aksimentiev, A. & Dekker, C. Slowing down DNA Translocation through a Nanopore in Lithium Chloride. *Nano Lett.* **12**, 1038–1044 (2012).
 88. Piguet, F. *et al.* Electroosmosis through α -Hemolysin That Depends on Alkali Cation Type. *J. Phys. Chem. Lett.* **5**, 4362–4367 (2014).

89. Lim, C. Y., Lim, A. E. & Lam, Y. C. Ionic Origin of Electro-osmotic Flow Hysteresis. *Sci. Rep.* **6**, 22329 (2016).
90. Houghtaling, J. *et al.* Estimation of Shape, Volume, and Dipole Moment of Individual Proteins Freely Transiting a Synthetic Nanopore. *ACS Nano* **13**, 5231–5242 (2019).
91. Yusko, E. C. *et al.* Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* **12**, 360–367 (2017).
92. Li, M.-Y. *et al.* Revisiting the Origin of Nanopore Current Blockage for Volume Difference Sensing at the Atomic Level. *JACS Au* **1**, 967–976 (2021).
93. Menestrina, G. Ionic channels formed by *Staphylococcus aureus* alpha-toxin: Voltage-dependent inhibition by divalent and trivalent cations. *J. Membr. Biol.* **90**, 177–190 (1986).
94. Song, L. *et al.* Structure of Staphylococcal alpha -Hemolysin, a Heptameric Transmembrane Pore. *Science* (80-.). **274**, 1859–1865 (1996).
95. Cheley, S., Braha, O., Lu, X., Conlan, S. & Bayley, H. A functional protein pore with a “retro” transmembrane domain. *Protein Sci.* **8**, 1257–1267 (1999).
96. Gu, L.-Q. *et al.* Reversal of charge selectivity in transmembrane protein pores by using noncovalent molecular adapters. *Proc. Natl. Acad. Sci.* **97**, 3959–3964 (2000).
97. Huang, G. *et al.* Electro-Osmotic Vortices Promote the Capture of Folded Proteins by PlyAB Nanopores. *Nano Lett.* **20**, 3819–3827 (2020).
98. Mohammad, Prakash, S., Matouschek, A. & Movileanu, L. Controlling a Single Protein in a Nanopore through Electrostatic Traps. *J. Am. Chem. Soc.* **130**, 4081–4088 (2008).
99. Willems, K. *et al.* Engineering and Modeling the Electrophoretic Trapping of a Single Protein Inside a Nanopore. *ACS Nano* **13**, 9980–9992 (2019).
100. Movileanu, L., Cheley, S., Howorka, S., Braha, O. & Bayley, H. Location of a Constriction in the Lumen of a Transmembrane Pore by Targeted Covalent Attachment of Polymer Molecules. *J. Gen. Physiol.* **117**, 239–252 (2001).
101. Wang, Y.-Q. *et al.* Rationally Designed Sensing Selectivity and Sensitivity of an Aerolysin Nanopore via Site-Directed Mutagenesis. *ACS Sensors* **3**, 779–783 (2018).
102. Bonome, E. L., Cecconi, F. & Chinappi, M. Electroosmotic flow through an alpha -hemolysin nanopore. *Microfluid. Nanofluidics* **21**, 96 (2017).
103. Ayub, M. & Bayley, H. Individual RNA Base Recognition in Immobilized Oligonucleotides Using a Protein Nanopore. *Nano Lett.* **12**, 5637–5643 (2012).
104. Cao, C. *et al.* Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores. *Nat. Commun.* **10**, 4918 (2019).
105. Wang, Y.-Q. *et al.* Identification of Essential Sensitive Regions of the Aerolysin Nanopore for Single Oligonucleotide Analysis. *Anal. Chem.* **90**, 7790–7794 (2018).
106. Iacovache, I. *et al.* Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. *Nat. Commun.* **7**, 12062 (2016).
107. Cao, J. *et al.* Giant single molecule chemistry events observed from a tetrachloroaurate(III) embedded *Mycobacterium smegmatis* porin A nanopore. *Nat. Commun.* **10**, 5668 (2019).
108. Mutter, N. L. *et al.* Modular Pore-Forming Immunotoxins with Caged Cytotoxicity Tailored by

- Directed Evolution. *ACS Chem. Biol.* **13**, 3153–3160 (2018).
109. Mutter, N. L., Volarić, J., Szymanski, W., Feringa, B. L. & Maglia, G. Reversible Photocontrolled Nanopore Assembly. *J. Am. Chem. Soc.* **141**, 14356–14363 (2019).
 110. Soskine, M. *et al.* An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry. *Nano Lett.* **12**, 4895–4900 (2012).
 111. Robertson, J. W. F. *et al.* Single-molecule mass spectrometry in solution using a solitary nanopore. *Proc. Natl. Acad. Sci.* **104**, 8207–8211 (2007).
 112. Reiner, J. E., Kasianowicz, J. J., Nablo, B. J. & Robertson, J. W. F. Theory for polymer analysis using nanopore-based single-molecule mass spectrometry. *Proc. Natl. Acad. Sci.* **107**, 12080–12085 (2010).
 113. Chavis, A. E. *et al.* Single Molecule Nanopore Spectrometry for Peptide Detection. *ACS Sensors* **2**, 1319–1328 (2017).
 114. Fologea, D., Ledden, B., McNabb, D. S. & Li, J. Electrical characterization of protein molecules by a solid-state nanopore. *Appl. Phys. Lett.* **91**, 053901 (2007).
 115. Muthukumar, M. Communication: Charge, diffusion, and mobility of proteins through nanopores. *J. Chem. Phys.* **141**, 081104 (2014).
 116. Bell, N. A. W. & Keyser, U. F. Specific Protein Detection Using Designed DNA Carriers and Nanopores. *J. Am. Chem. Soc.* **137**, 2035–2041 (2015).
 117. Dragomir, I. S. *et al.* The Nanopore-Tweezing-Based, Targeted Detection of Nucleobases on Short Functionalized Peptide Nucleic Acid Sequences. *Polymers (Basel)*. **13**, 1210 (2021).
 118. Restrepo-Pérez, L. *et al.* Resolving Chemical Modifications to a Single Amino Acid within a Peptide Using a Biological Nanopore. *ACS Nano* **13**, 13668–13676 (2019).
 119. Restrepo-Pérez, L., Wong, C. H., Maglia, G., Dekker, C. & Joo, C. Label-Free Detection of Post-translational Modifications with a Nanopore. *Nano Lett.* **19**, 7957–7964 (2019).
 120. Asandei, A., Rossini, A. E., Chinappi, M., Park, Y. & Luchian, T. Protein Nanopore-Based Discrimination between Selected Neutral Amino Acids from Polypeptides. *Langmuir* **33**, 14451–14459 (2017).
 121. Asandei, A. *et al.* Single-Molecule Dynamics and Discrimination between Hydrophilic and Hydrophobic Amino Acids in Peptides, through Controllable, Stepwise Translocation across Nanopores. *Polymers (Basel)*. **10**, 885 (2018).
 122. Mereuta, L. *et al.* Slowing down single-molecule trafficking through a protein nanopore reveals intermediates for peptide translocation. *Sci. Rep.* **4**, 3885 (2015).
 123. Asandei, A. *et al.* Electroosmotic Trap Against the Electrophoretic Force Near a Protein Nanopore Reveals Peptide Dynamics During Capture and Translocation. *ACS Appl. Mater. Interfaces* **8**, 13166–13179 (2016).
 124. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci.* **93**, 13770–13773 (1996).
 125. Braha, O. *et al.* Simultaneous stochastic sensing of divalent metal ions. *Nat. Biotechnol.* **18**, 1005–1007 (2000).
 126. Boersma, A. J., Brain, K. L. & Bayley, H. Real-Time Stochastic Detection of Multiple

- Neurotransmitters with a Protein Nanopore. *ACS Nano* **6**, 5304–5308 (2012).
127. Boersma, A. J. & Bayley, H. Continuous Stochastic Detection of Amino Acid Enantiomers with a Protein Nanopore. *Angew. Chemie Int. Ed.* **51**, 9606–9609 (2012).
 128. Zernia, S., van der Heide, N. J., Galenkamp, N. S., Gouridis, G. & Maglia, G. Current Blockades of Proteins inside Nanopores for Real-Time Metabolome Analysis. *ACS Nano* **14**, 2296–2307 (2020).
 129. Sanchez-Quesada, J., Ghadiri, M. R., Bayley, H. & Braha, O. Cyclic Peptides as Molecular Adapters for a Pore-Forming Protein. *J. Am. Chem. Soc.* **122**, 11757–11766 (2000).
 130. Galenkamp, N. S., Soskine, M., Hermans, J., Wloka, C. & Maglia, G. Direct electrical quantification of glucose and asparagine from bodily fluids using nanopores. *Nat. Commun.* **9**, 4085 (2018).
 131. Galenkamp, N. S. & Maglia, G. Substrate binding and turnover modulate the affinity landscape of dihydrofolate reductase to increase its catalytic efficiency. *bioRxiv* 2020.04.14.040733 (2020) doi:10.1101/2020.04.14.040733.
 132. Soskine, M., Biesemans, A. & Maglia, G. Single-Molecule Analyte Recognition with ClyA Nanopores Equipped with Internal Protein Adapters. *J. Am. Chem. Soc.* **137**, 5793–5797 (2015).
 133. Gu, L.-Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature* **398**, 686–690 (1999).
 134. Li, M.-Y., Wang, Y.-Q., Ying, Y.-L. & Long, Y.-T. Revealing the transient conformations of a single flavin adenine dinucleotide using an aerolysin nanopore. *Chem. Sci.* **10**, 10400–10404 (2019).
 135. de Jong, L. A. A., Uges, D. R. A., Franke, J. P. & Bischoff, R. Receptor–ligand binding assays: Technologies and Applications. *J. Chromatogr. B* **829**, 1–25 (2005).
 136. Li, X., Lee, K. H., Shorkey, S., Chen, J. & Chen, M. Different Anomeric Sugar Bound States of Maltose Binding Protein Resolved by a Cytolysin A Nanopore Tweezer. *ACS Nano* **14**, 1727–1737 (2020).
 137. Rumble, J. R., Bruno, T. J. & Doa, M. J. *CRC handbook of chemistry and physics : a ready-reference book of chemical and physical data.* (2021).

1.8. Supporting Information

Supplementary figure 1. Peptide backbone and proteinogenic amino acids. Structure and dissociation constants for the peptide backbone and proteinogenic amino acids. R represents the side chain connected to the backbone. The pK_a and pK_b represent the dissociation constants for the $-COOH$ and $-NH_2$ groups, respectively. The pK_x is the dissociation constant for the functional group on the amino acid. The boxes shown in the corner left of each amino acid represent the single-letter code. Proteinogenic amino acids (at pH 7.5) with a positive charge are marked in blue, negative charges are represented in red, polar non-charged amino acids shown in purple, and non-polar amino acids are marked in green. The yellow marking represents the residual amino acids. All dissociation constants, and grouping (colours) were reproduced from the CRC handbook of chemistry and physics.¹³⁷

Chapter 2

The Manipulation of the Internal Hydrophobicity of FraC Nanopores Augments Peptide Capture and Recognition

The detection of analytes and the sequencing of DNA using biological nanopores has seen major advances over recent years. The analysis of proteins and peptides with nanopores, however, is complicated by the complex physico-chemical structure of polypeptides, and the lack of understanding of the mechanism of capture and recognition of polypeptides by nanopores. In this work, we show that introducing aromatic amino acids at precise positions within the lumen of α -helical fragaceatoxin C (FraC) nanopores increased the capture frequency of peptides and largely improved the discrimination among peptides of similar size. Molecular dynamic simulations determined the sensing region of the nanopore, elucidated the microscopic mechanism enabling accurate characterization of the peptides *via* ionic current blockades in FraC, and characterized the effect of the channel on peptide discrimination. This work provides insights to improve the recognition and augment the capture of peptides by nanopores, which is important for developing a real-time and single-molecule size-analyser for peptide recognition and identification.

This chapter was published as:

Lucas, F. L. R., Sarthak, K., Lenting, E. M., Coltan, D., van der Heide, N. H., Versloot, R. C. A., Aksimentiev, A., Maglia, G. (2021). The Manipulation of the Internal Hydrophobicity of FraC Nanopores Augments Peptide Capture and Recognition. *ACS Nano*, 15(6), 9600-9613

2.1. Introduction

Nanopores are potential candidates for developing low-cost and high-throughput portable detectors. Biological nanopores have been shown to be particularly suitable for the detection and discrimination of small molecules based on the current blockade generated when an analyte binds to or translocates through the nanopore.^{1–9} The amplitude of the analyte-induced current signal, however, cannot always be easily predicted. Molecules usually reduce the nanopore current by a value that is proportional to the volume of electrolyte displaced inside the nanopore.^{10–12} And for model analytes such as polyethylene glycol (PEG), it has been shown that the nanopore currents can size the polymer in a similar fashion as mass spectrometry.^{13–15} However, many other factors can influence the overall ionic current, such as the conformation of the molecule, its charge or dipole, its position or its shape.^{8,16–20} For more complex polymers such as DNA or proteins, the relationship between analyte mass and signal is more complicated. Most notably, the four DNA nucleotides²¹ or DNA homopolymers^{22,23} can induce four different ionic currents that are not related to the mass or the volume of the individual bases.

Recently, we have characterised actinoporin fragaceatoxin C (FraC) for nanopore analysis.²⁴ FraC monomers have a mass of approximately 20 kDa and form nanopores on membrane containing sphingomyelin.²⁵ The crystal structure of wild-type FraC (FraC) revealed an oligomeric pore formed from eight identical subunits (octameric).²⁶ Compared to the cylindrical shape of β -barrel nanopores commonly used in nanopore analysis, FraC forms a V-shaped α -helical nanopore with a ~ 5.5 nm *cis* entry and a ~ 1.5 nm narrow *trans* exit. In a previous work, we have shown that wild type FraC (hereafter FraC, see methods for the exact sequence) is capable of forming different oligomeric forms—most notably the octameric (T1) and heptameric (T2)—with a distinct pore volume and range of detectable peptides.²⁷ Intriguingly, peptides did not enter the nanopore at physiological pH values. Only when the pH of the solution is lowered to less than 6.5 peptide blockades are observed, reaching optimal capture frequencies at pH values lower than 4.5.²⁸ Although the detailed mechanism responsible for the pH dependence of capture is not understood,²⁹ it is likely that the shape and the highly negatively charged inner constriction of the FraC nanopore play an important role, most likely by influencing the electro-osmotic flow across the nanopore.^{9,28,30–36} Importantly, we also demonstrated that at the exact pH of 3.8 and 1 M KCl, the current observed from peptide translocation through FraC correlates with the mass of the peptide despite their chemical composition,²⁷ as also observed with α HL nanopores,¹⁵ making FraC a prime target for the development of single molecule *nanopore spectrometry* for peptides.

Although several properties of FraC—such as the electro-osmotic flow (EOF), the recognition volume and the ability to capture peptides^{27,28} or DNA²⁴—could be adjusted through protein engineering, the interactions between the nanopore and the analytes and the influence of such interactions on the analyte distinguishability remain poorly characterized. Furthermore, the duration of detectable peptide blockades is rather short—hundreds of microseconds (average dwell time for Angiotensin 1 is 0.15 ± 0.04 ms),²⁷ which implies that a sizable fraction of the translocation events occurs undetected and that some detected events could be inaccurately characterized because of the limited temporal resolution of the ionic current measurement. Previous work with α -hemolysin (α HL) nanopores revealed that a positive surface charge in the nanopore is important for increasing the frequency of DNA capture.^{37,38} In this work, we show that the capture and recognition of peptides in α -helical FraC is improved by the introduction of aromatic amino acids near the constriction. This finding enables development of nanopore systems for real-time identification of peptides according to their volume and is crucial for single-molecule sensing where the efficiency of peptide capture is paramount.

2.2. Results and Discussion

2.2.1. Fragaceatoxin C Mutant Screening

The engineering of α -helical nanopores—such as FraC—is not as straightforward as in the case for β -barrel nanopores—such as α -hemolysin³⁹—, because the side chains in α -helical transmembrane region has a complex interaction with the aqueous and lipid phases. In order to identify the variable region, we aligned the sequence of FraC with other actinoporins (**Figure 1A**), which have 60-80% common sequence identity.⁴⁰ We identified five non-conserved—lumen facing—positions: D10, G13, G15, D17 and K20 (numbering based on the wild type FraC sequence, **Figure 1B**). Notably, the mutation of D10 to arginine (R) in the lumen was reported previously for the preparation of nanopores amenable to DNA analysis.²⁴ We engineered these non-conserved positions into different functionalities. Each of the positions near the presumed recognition site (G13),²⁷ was modified to a residue of a positively (K, R or H) or negatively (D or E) charged group as well as a neutral (G or Q) or aromatic (W,Y,F,V) group. In FraC—a glycine residue is positioned at residue 15—while the most common amino acid in other actinoporins is a threonine. We reason that the additional methyl group pointing towards the bilayer (**Figure 1B**) might stabilize the baseline current relative to FraC (**Figure 1C**). In addition, the aligned sequence (**Figure 1A**) often contains a pair of opposite charges at positions 20 / 21, therefore, we constructed two mutants that have the same characteristics: T21D and the double-mutant K20D / T21K. For completion, we also included a change of charge on position 20 by introducing a glutamic acid (K20D).

Figure 1. Actinoporins common sequence alignment and wild-type Fragaceatoxin C. **A:** Common sequence alignment of known actinoporins, the dots represent the same amino acid as the common sequence, other amino acids are represented by their single-letter code. **B:** Artistic model of Fragaceatoxin C (PDB: 4TSY) inserted into a lipid bilayer, across which a voltage is applied. Several non-conserved positions are enlarged. **C:** Representative traces of the octameric (T1) and heptameric (T2) form of wild-type Fragaceatoxin C under an applied potential of -50 mV in 1 M KCl and 50 mM citric acid titrated with bis-tris propane to pH 3.8. Traces were collected at a sampling frequency of 50 kHz, using a 10 kHz Bessel filter and 5 kHz Gaussian filter.

At pH 3.8 and in 1 M KCl solutions, FraC exists in three oligomeric forms, presumably corresponding to octamers, heptamers and hexamers.²⁷ For initial screening, we only considered octameric pores (or type I pores, T1), with the exception FraC-T2, which corresponds to the heptameric type II pores. Octameric oligomers were identified as the nanopores with the highest conductance. Several mutations significantly reduced the open pore current (I_0) relative to WtFraC-T1 (95 ± 1 pA), some to an extent that the I_0 resembled by FraC-T2 (47 ± 3 pA, **Supplementary Table 1**). Notably, decreased I_0 were observed when residues with a larger volume were introduced, for example for the aromatic residues (W/F/Y) introduced on position 13 ($I_0 = 64 \pm 8$ pA, 77 ± 4 pA and 82 ± 3 pA, respectively). Neither mutation of residues 20 and 21 reduced nor increased the I_0 and the introduction of tryptophan residues at positions D10 and D17 resulted in nanopores that did not fold or did not insert into the 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine planar lipid bilayers under aforementioned conditions.

The signal-to-noise ratio ($I_0/\sigma(I_0)$, SNR, **Supplementary Table 1**) of most mutant pores was similar to WtFraC-T1 (35 ± 3 , -50 mV). While the introduction of a threonine residue on position 15 showed a 5% increased I_0 (100 ± 3 pA), but showed no significant difference in the SNR (39 ± 6 , -50 mV). Replacing the aspartic acid on position 10 by a neutral glycine reduced the SNR (17 ± 2 , -50 mV) significantly, as well as introducing a tryptophan residue at position 13 (G13W-FraC-T1, 21 ± 2 , -50 mV), a value comparable to FraC-T2 (22 ± 5 , -50 mV). However, the introduction of either phenylalanine (G13F-FraC-T1, 31 ± 4 , -50 mV) or tyrosine (G13Y-FraC-T1, 36 ± 3 , -50 mV) at position 13 did not significantly reduce the SNR compared to WtFraC-T1. Interestingly, we observe a rapid decrease in SNR when a valine residue was introduced on position 13 (G13V-FraC-T1, 5 ± 2 , -50 mV).

The observed baseline signals of all aromatic substitutions at position G13 contained current blockades without the presence of added analyte (gating), with an event frequency varying greatly between buffer preparations. These stochastically distributed blockades—with a median dwell time near 0.7 ± 0.4 ms and relative blocked current (excluded current) of 57 ± 10 %—reduced to 5 ± 2 events·s⁻¹ by additional filtering of the buffer solvent using activated charcoal. As this procedure especially reduces the presence of hydrophobic molecules in the final solution, we assume that the gating of these pores is caused by sensitivity towards hydrophobic molecules due to interactions with the hydrophobic patch that was introduced in the recognition area of the pore.

Peptides have different size and chemical properties. Therefore, the predicted range of detectable analytes and frequency of their capture is expected to differ between mutant pores, given their different physico-chemical properties. In order to compare between pores, we utilize a mixture of peptides that were generated from the non-specific tryptic digest of lysozyme (*Gallus-Gallus*). We used non-specific trypsin (not purified, containing additional proteases such as chymotrypsin) as it yields a more stochastic peptide mixture with a broad distribution of peptide mass (**Supplementary Table 3 and 4**), which is advantageous when sampling nanopores with different chemical composition. In addition, the mixture has a relatively uniform charge since trypsin cleaves at positively charged amino acids and the pH of the solution is set to 3.8. Therefore, most peptides will have a positive charge next to the zwitterionic charges on the peptide, yielding a positive net charge. All pores were tested with the same proteolytic mixture.

We classify the observed events by quantifying the flat-top shape fitted using a least-squared Levenberg-Marquardt method and a generalized flat-top normal distribution function (**see methods**). In brief, this fit results in a β value that can classify the events as either a spike with $\beta < 1$, a normal distribution $\beta = 1$ or flat-top distribution $\beta > 1$. Unless stated otherwise, we report events with $\beta > 1$. In addition, for each blockade we determine the excluded current ($I_{ex}\%$), which is the percentage of

the current that is blocked during a translocation event relative to the open pore current [$(I_o - I_b)/I_o$, where I_b is the average ionic current of the peptide event].

Figure 2. Electrophysiology recordings of (mutant) Fragaceatoxin C with trypsin digested lysozyme. A: Representative traces of Fragaceatoxin C nanopores after addition of an equal concentrations of trypsin-digested lysozyme under an applied potential of -50 mV. The dotted blue line represents the baseline current of the octameric (T1) form of wild-type Fragaceatoxin C, the dotted red line represents the baseline current of the heptameric (T2) form of wild-type Fragaceatoxin C. **B-D:** representative trace of octameric Fragaceatoxin C (T1, **B**), heptameric Fragaceatoxin C (T2, **C**), and Fragaceatoxin C mutant G13F (**D**). The grey line represents the recorded trace and the black superimposed line represents the fit after event detection. The block above the trace align with the length of the events. Traces were collected in 1 M KCl and 50 mM citric acid titrated with bis-tris propane to pH 3.8 at a sampling frequency of 50 kHz, using a 10 kHz Bessel filter and 5 kHz Gaussian filter.

Under an applied potential of -50 mV (+50 mV for D10R-FraC-T1) and in 1 M KCl at pH 3.8, we observe that the capture efficiency of the non-specific tryptic digest of lysozyme is affected by the mutations near the nanopore constriction (**Figure 2A, Supplementary Table 5**). When the charge at position 10 or 17 was removed (D10G-FraC-T1 or D17Q-FraC-T1 mutation) the capture frequency was reduced from 13 ± 2 events $\cdot$ s $^{-1}$ for WtFraC-T1 to 3.8 ± 0.7 events $\cdot$ s $^{-1}$ and 1.8 ± 0.5 events $\cdot$ s $^{-1}$ respectively. It is

important to consider that, even at pH 3.8, about half of the aspartic acid residues retain a negative charge. It has been shown that the EOF is critical for efficient capturing of peptides in the nanopore, and the strength and direction of the EOF is dependent on charges in the constriction site.²⁸ Therefore, the reduced capturing efficiency of peptides with a non-charged constriction can be attributed to the reduced EOF across the nanopore. Replacing the negative charge with a positive charge on these positions (D10R or D17K) significantly changed the behavior of the pore. D10R-FraC-T1 showed a destabilized baseline current under an applied bias of -50 mV, but stable under +50 mV (**Supplementary Figure 1**), thereby behaving opposite to FraC. In contrast, D17K-FraC-T1 showed an unstable baseline signal under an applied bias potential of +50 mV, however, stable when -50 mV was applied. Replacing the charge of K20 by introducing an aspartic acid increased the capture frequency only slightly ($18.5 \pm 0.4 \text{ events}\cdot\text{s}^{-1}$) compared to WtFraC-T1 ($13 \pm 2 \text{ events}\cdot\text{s}^{-1}$). As expected, the lipid facing mutation G15T did not significantly change the capture frequency of the pore ($12 \pm 2 \text{ events}\cdot\text{s}^{-1}$).

Interestingly, we find that the introduction of a neutral, hydrophilic asparagine residue on position 13 results in no observed peptide capture, whereas the introduction of an aromatic residue (Y, F or W) increases the capture frequency to 43 ± 8 , 41 ± 4 and $43 \pm 9 \text{ events}\cdot\text{s}^{-1}$ respectively, which is significantly different from WtFraC-T1 ($13 \pm 2 \text{ events}\cdot\text{s}^{-1}$) and FraC-T2 ($11 \pm 3 \text{ events}\cdot\text{s}^{-1}$). Most of the blockades in pores with an aromatic residue on G13 were flat-top shaped with relatively long dwell times ($0.32 \pm 0.06 \text{ ms}$, $0.18 \pm 0.03 \text{ ms}$ and $0.22 \pm 0.06 \text{ ms}$ for G13Y-FraC-T1, G13F-FraC-T1 and G13W-FraC-T1, respectively, compared to $0.09 \pm 0.06 \text{ ms}$ for WtFraC-T1 (**Figure 2B**) and $0.10 \pm 0.01 \text{ ms}$ for FraC-T2 (**Figure 2C**). In contrast, we observe that the substitution of glycine 13 to valine showed little significant difference in capture frequency (17 ± 0.1) or dwell time ($0.084 \pm 0.002 \text{ ms}$) when compared to WtFraC-T1. The reversal potential of WtFraC-T1 and G13F-FraC-T1 nanopores at pH 3.8 is the same and near zero (**Supplementary Table 6**), indicating that in both pores the EOF is weak.

In order to compare the different mutants, we constructed the excluded current spectrum (shown for 4 pores in **Figure 3A-D**) by creating a histogram of the excluded currents ($I_{ex}\%$) using all events with $\beta > 1$ (5 kHz Gaussian filter, **see methods**). We normalized the spectra and observe distinct patterns for WtFraC-T1 and T2 (**Figure 3A/B**) with sharp gaussian shaped peaks for G13F-FraC-T1 (**Figure 3C**), while G13N-FraC-T1 results in a seemingly stochastic spectrum (**Figure 3D**). We compared the excluded current spectra (**Supplementary Figure 3**) using a point-to-point spectral matching algorithm, using the excluded current spectrum where $40 \% < I_{ex}\% < 95 \%$ (**see methods**). This range was chosen as events with an I_{ex} lower than 40% are near the baseline and we are unable to identify them accurately and events with an I_{ex} larger than 95% are full blockades which result from the convoluted spectrum of large fragments, which individually don't contribute to the identification of proteins.

Figure 3. Event counts and signal correlation of (mutant) Fragaceatoxin C with trypsin digested lysozyme. A-D: Observed excluded current spectra from tryptic digest of lysozyme. A: octameric wild-type Fragaceatoxin C (T1), B: heptameric wild-type Fragaceatoxin C (T2), C: Fragaceatoxin C mutant G13F and D: Fragaceatoxin C mutant G13N. E: Squared first derivative Euclidean cosine correlation of residual current spectra of (mutant) Fragaceatoxin C combined with equal units of trypsin digested lysozyme. The black boxes surrounding multiple mutants represent similar signals. Traces were collected in 1 M KCl and 50 mM citric acid titrated with bis-tris propane to pH 3.8 at a sampling frequency of 50 kHz, using a 10 kHz Bessel filter and 5 kHz Gaussian filter. The external bias was -50 mV except for D10R⁺ and G13H⁺, which were tested at +50 mV.

To better represent the data, we perform hierarchical clustering using the Ward distance,⁴¹ which revealed three major clusters of nanopores (see methods, Figure 3E). The first cluster forms a group of nanopores with high similarity to the octameric WtFraC-T1 (Figure 3A), containing the mutations of T21D, G13V, G15T, K20D, K20D T21K and G15V, V22A. The majority of the mutations on positions 20/21 were included in this cluster, indicating that mutations on these residues did not significantly alter the recognition site of FraC. This observation is in agreement with previous results, where the constriction site was expected to be located between residue 10 and 17.²⁷ The second cluster comprised most of the positive nanopores (e.g. D17K -FraC-T1) and G13N-FraC-T1 (Figure 3D). We believe that the stochastic nature of these spectra is caused by the inability to correctly localize the $I_{ex}\%$ of the events, as most peptides translocate rapidly (faster than 100 μ s). The third group of octameric mutant nanopores shared a high similarity to the heptameric form of FraC (Figure 3B). This group includes the aromatic nanopores (G13Y/F/W-FraC-T1) and one positively charged pore (D10R-FraC-T1). Importantly, the D10R mutation prevented efficient capturing of peptides at negative applied potential, but the ability to capture peptides could be restored when a positive potential was applied instead.

2.2.2. Fragaceatoxin C Mutant Characterization

We selected five mutants that show interesting characteristics towards protein detection for further characterization, namely: G15T-FraC-T1, as it is comparable to WtFraC-T1 with a slightly increased I_0 , K20D-FraC-T1 as it had one of the higher SNRs and good capture frequency and the aromatic mutations at G13 (G13Y/F/W-FraC-T1) for their increased dwell times compared to FraC-T2 and capture frequency. For the characterization of these pores we used a mixture of well-defined, chemically similar peptides, in contrast to the unspecific lysozyme digest that was used before. The mixture consisted of four peptides:²⁷ Angiotensinogen (DRVYIHPFHLVIHN, 1758.9 Da, charge = +3.96), Angiotensin 1 (DRVYIHPFHL, 1296.5 Da, charge = +2.96), Angiotensin 3 (RVYIHPF, 931.1 Da, charge = +2.16) and Angiotensin 4 (VYIHPF, 774.9 Da, charge = +1.16) abbreviated as Pre-Ang, Ang-I, Ang-III and Ang-IV respectively. The resolution of the nanopores was quantified by measuring the separation between peptides using the difference between the peak centers and their mean standard deviation as shown in **Equation 1 and 2**.

$$\bar{\sigma} = \frac{(\sigma_1 + \sigma_2)}{2} \quad (1)$$

$$R_s = \frac{\mu_1 - \mu_2}{\bar{\sigma}} = \frac{2 \cdot (\mu_1 - \mu_2)}{\sigma_1 + \sigma_2} \quad (2)$$

Where R_s is resolution, μ_1 and μ_2 are the peak centers with standard deviation σ_1 and σ_2 respectively. If $R_s < 2$, the difference between the peak centers is less than twice the average standard deviation and no baseline separation is achieved. To achieve an overlap of less than 5%, a $R_s \geq 4$ is required, that is, the difference between the peak centers is equal or bigger than four times the average standard deviation of the peaks, thus we can consider them separated. Larger values of R_s indicate a better separation (**Table 1**).

MW: 1759 - 931	FraC-T1	FraC-T2	K20D- FraC-T1	G15T- WtFraC-T1	G13F-FraC-T1	G13Y- FraC-T1	G13W- FraC-T1
$\Delta I_{ex}\%$ (Ang-IV — Ang-III)	8.8 ± 0.7 %	18 ± 3 %	14 ± 6 %	12 ± 5 %	9.2 ± 0.3 %	9.1 ± 0.7 %	5.0 ± 0.3 %
$\Delta I_{ex}\%$ (Ang-III — Ang-I)	17 ± 2 %	12.3 ± 0.5 %	15 ± 1 %	17 ± 2 %	24 ± 1 %	22 ± 1 %	19.9 ± 0.2 %
$\Delta I_{ex}\%$ (Ang-I — Pre-Ang)	19.0 ± 0.2 %	9.3 ± 0.3 %	16.2 ± 0.4 %	19.0 ± 0.3 %	10 ± 1 %	6.1 ± 0.8 %	6.4 ± 0.2 %
R_s (Ang-IV — Ang-III)	2.1 ± 0.7	4.1 ± 1.2	2.6 ± 1.4	2.0 ± 0.5	4.6 ± 0.5	4.4 ± 1.1	3.6 ± 0.4
R_s (Ang-III — Ang-I)	3.5 ± 0.5	4.2 ± 0.5	2.3 ± 0.2	3.3 ± 0.4	12.1 ± 4.3	11.8 ± 2.9	19.1 ± 1.7
R_s (Ang-I — Pre-Ang)	4.1 ± 0.3	4.0 ± 0.3	3.2 ± 0.5	4.6 ± 0.2	6.1 ± 2.3	4.0 ± 0.7	7.2 ± 1.2
MW: 772 -556		FraC-T2	FraC-T3		G13F- FraC-T2		G13W-FraC-T2
$\Delta I_{ex}\%$ (Leu-enk — Ang-II (4-8))		N.O.	27.6 ± 0.8 %		19.1 ± 0.1 %		10.6 ± 0.8 %
$\Delta I_{ex}\%$ (Ang-II (4-8) — Kemptide)		N.O.	N.O.		6 ± 2 %		N.O.
$\Delta I_{ex}\%$ (Leu-enk — Ang-II (4-8))		N.O.	5 ± 1		11 ± 2		3 ± 2
R_s (Ang-II (4-8) — Kemptide)		N.O.	N.O.		3 ± 2		N.O.

Table 1 The differences between peptide peak centres ($\Delta I_{ex}\%$) and the observed baseline separation (R_s).

As observed before from the analysis of the proteolytic digest of lysozyme, both the G15T and K20D mutation did not significantly improve the resolution compared to WtFraC-T1. In fact, the mutation to K20D reduced the resolution by broadening the peaks of all peptides (**Table 1**). Comparing the dwell times of the peptides measured in FraC-T2 and the aromatic pores shows that the increased retention (**Figure 4**), which reduces the spread in the residual current, explains the increased resolution. All the aromatic pores revealed a similar trend in the resolution, however, the difference between peak centers is largest for the G13F mutation and lowest for the G13W mutation, while the standard deviation within the residual current follows a reverse trend, resulting in similar observed resolutions.

Figure 4. Peptide recognition of (mutant) Fragaceatoxin C. Mutations are shown in red on the lumen of Fragaceatoxin C modeled on PDB: 4TSY. The fit of the residual current is shown for Angiotensin IV (VYIHPF), Angiotensin III (RVYIHPF), Angiotensin I (DRVYIHPFHL) and Angiotensinogen (DRVYIHPFHLVIHN) each in 2.5 μ M concentration, recorded under an applied potential of -50 mV. Traces were collected in 1 M KCl and 50 mM citric acid titrated with bis-tris propane to pH 3.8 at a sampling frequency of 50 kHz, using a 10 kHz Bessel filter and 5 kHz Gaussian filter. Histograms were created from all events with a dwell time larger than 200 μ s, with the exception of G13F, G13Y, and G13W where the minimal dwell time was set to 1 ms (dashed black line). The dwell time set against the residual current shows all events with a dwell time larger than 200 μ s. The marker size of datapoints in the residual current set against the dwell time is adjusted for visualization.

We tested the resolution of aromatic heptameric (T2) nanopores, and compared to hexameric (T3) WtFraC-T3 and WtFraC-T2 nanopores using Leucine-enkephalin (Leu-enk, YGGFL, 555.6 Da), Angiotensin II (4-8) {Ang-II(4-8), YIHPF, 675.8 Da}, and Kemptide (LRRASLG, 771.9 Da). For WtFraC-T3 we use a FraC version with two altered membrane-interfacing modifications, W112S-W116S, which allowed the formation of hexameric nanopores²⁷. WtFraC-T2 showed no blockades (**Figure 5**), suggesting that the majority of peptides translocated through the pore undetected. FraC-T3 and G13W-FraC-T2 showed leucine-enkephalin and angiotensin II (4-8) blockades, while kemptide blockades were not observed. This is surprising, considering kemptide has higher molecular weight

than leucine-enkephalin and angiotensin II (4-8). Most likely, the two arginine residues in the kemptide induce a fast electrophoretic translocation across these nanopores. Interestingly, we found that kemptide induced blockades to G13F-FraC-T2. A likely explanation is that cation- π interactions between the aromatic ring of phenyl alanine residues and the two arginine residues are crucial to reduce the residence time of the peptide inside the nanopore.

Figure 5. Peptide recognition of heptameric and hexameric Fragaceatoxin C. The fit of the residual current is shown for Leucine-enkephalin (YGGFL), Angiotensin II (4-8) (YIHPF) and Kemptide (LRRASLG) each in 10 μ M concentration, recorded under an applied potential of -70 mV. Traces were collected in 1 M KCl and 50 mM citric acid titrated with bis-tris propane to pH 3.8 at a sampling frequency of 50 kHz, using a 10 kHz Bessel filter and 5 kHz Gaussian filter.

2.2.3. Characterization using Molecular Dynamics Simulations

To obtain a molecular level understanding of the ionic current distinguishability, we probed ion and peptide transport through the engineered FraC nanopores using the all-atom molecular dynamics (MD) method. From the experimentally studied FraC pores (**Figure 4**), we selected three pores, WtFraC-T1, FraC-T2, G13F-FraC-T1 which had the best distinguishability of the angiotensin peptides (**Figure 6A-C**). Each system was simulated at the experimental condition with a bias of -50 mV and a solution pH of 3.8 (**see methods**). To compare the current levels with experimental results, the resulting MD currents were scaled by the ratio of the experimental bulk conductivity of 1 M KCl (10.5 S/m) and the simulated bulk conductivity of 1 M KCl (16.6 S/m). The narrowest pore (FraC-T2) showed the lowest MD current value, which was in good quantitative agreement with experiment (**Figure 6D**). The highest MD current was obtained from the WtFraC-T1 pore, while the G13F-FraC-T1 pore showed an intermediate current value, in qualitative agreement with experiment (**Figure 6D**). The simulated current values for the T1 pores were higher in the MD simulations than in experiment, which we attribute to possible

local changes in the pore structure introduced by the mutations and the approximate treatment of the pH 3.8 conditions in the MD simulations. Although the lumen of the nanopores carry overall high positive charge at pH of 3.8, the nanopore current was found to be carried predominantly by the potassium ions (**Figure 6D**, **SI Figure 6**) translocating from the *cis* to the *trans* chamber, which agrees with our previous experimental ion selectivity measurements²⁷ and our MD simulations of the FraC nanopore at physiological pH and NaCl electrolyte.⁴² Accordingly, we find the water flux through the FraC nanopores to be directed from *cis* to *trans* compartment, as to facilitate the capture of neutral peptides, with the flux magnitudes being in the range from 3 to 6 water molecules per nanosecond (**SI Figure 6**).

The electrostatic potential profile along the symmetry axis of the pores revealed a barrier in the vicinity of the D10 residues, located at $z = -10 \text{ \AA}$, in our coordinate system (**Figure 6E**). The barrier was the highest for the FraC-T2 pore, lower for G13F-FraC-T1 and even lower for WtFraC-T1. Just above the barrier, each electrostatic potential profile also exhibits a minimum near residue K20, at $z = 4.5, 5$ and $5.5 \text{ \AA}$ for FraC-T2, WtFraC-T1 and the G13F-FraC-T1 pores, respectively. The potential minimum is the deepest for the FraC-T2 pore and is more shallow for G13F-FraC-T1 and even more shallow for WtFraC-T1. The local concentration of potassium ions inside each pore (**Figure 6F**) has a peak at $z = -10 \text{ \AA}$ with the peak height being the largest for FraC-T2, followed by G13F-FraC-T1 and then by WtFraC-T1, similar to the peak height of the electrostatic potential (**Figure 6E**). Similarly, the local concentration of chloride ions (**Figure 6G**) peaks at the location where the local electrostatic potential has a minimum. In our control simulations of WtFraC-T1 at pH 7, the maximum of the electrostatic potential near D10 and the local concentration of potassium ions increased in comparison to the pH 3.8 conditions, reflecting the higher negative charge of the D10 residues, whereas the electrostatic minimum near K20 remained unchanged (**SI Figure 7**). To summarize, all three pores show similar electrostatic potential profiles, with a minimum near K20 and a maximum near D10, which we attribute to the congregation of chloride and potassium ions, respectively, near those charged residues. This peculiar distribution of the electrostatic potential would favor transient arrest of a positively charged analyte entering the nanopores from the *cis* side, right before entering the nanopore constriction.

Figure 6: Molecular dynamics simulation of mutant Fragaceatoxin C nanopores. A-C: All-atom models of WtFraC-T1 (A), G13F-FraC-T1 (B) and WtFraC-T2 (C) nanopores. The protein is shown as a gray cutaway surface, embedded in a DPhPC lipid bilayer (blue). The G13F mutation site in panel b is shown in red. All systems contain 1 M KCl solution (potassium in orange and chloride in green, water not shown). The protonation states of the titratable residues are set to reflect pH of 3.8. The z axis is shown on the left for scale. Yellow and red horizontal lines show the position of the electrostatic minima and maxima respectively. D: Experimental and simulated open pore currents at -50 mV for the three systems. The simulated values reflect scaling of the raw MD current with the ratio of the experimental and simulated bulk conductivity of 1 M KCl. The error bars represent the standard error computed by splitting the MD trajectories into 10 ns fragments and considering each fragment as an independent measurement of the current. The contribution of the potassium ion to the MD current is specified at the bottom. E: Average electrostatic potential along the symmetry axis (z axis) of the three pores. F-G: Profiles of potassium (F) and chloride (G) ion concentration along the symmetry axis of each nanopore.

To identify the sensing regions of the FraC nanopore, we modelled the translocation of Ang-I through the three pores (FraC-T1, G13F-FraC-T1 and FraC-T2) using the Steered Molecular Dynamics (SMD) protocol⁴³ (Figure 7A, SI Movie 1-3). As the peptide was pulled through the pore, it experienced a steric barrier when entering the constriction (Figure 7B), followed by another barrier within the constriction, as indicated by the downward spikes in the plots of the SMD force (Figure 7C). The first barrier was of lower magnitude and located around $z = 10 \text{ \AA}$, near residue E24, whereas the second barrier was more pronounced, originating from the constriction of the pore at $z = -10 \text{ \AA}$, near residue D10. Both barriers were more pronounced in the FraC-T2 and G13F-FraC-T1 pores, reflecting the smaller constriction of the heptamer and the bulky phenylalanine substitution in the octamer. It is important to note that the electrostatic minima—as calculated from the open pore simulations—are located between the two steric barriers.

Using the Steric Exclusion Model⁴⁴ and the ensemble of conformations provided by the SMD simulations, we computed the relative excluded current [$I_{ex} = (I_o - I_b)/I_o$] as a function of Ang-I location within each of the three pores (**Figure 7D-E**). As the peptide approaches the constriction of the pores, I_{ex} values increase rapidly, reaching the value of 1 (complete blockade) when passing through the ring of D10 residues near $z = -10$ Å. After passing the constriction, I_{ex} returns to a value close to zero, *i.e.*, the open pore value. From the Ang-I SMD trajectory, we computationally reconstructed conformations of Ang-III and Ang-IV peptides using a previously described protocol.¹ **Figures 7D-F** show the computed excluded current for the three peptides in each pore vs the center of mass z coordinate of the common peptide segment (VYIHPF). For reference, we also plot the experimental I_{ex} values as dashed horizontal lines in the figures. First, we note that the simulated I_{ex} values substantially differ among the peptides only in the region where I_{ex} increases rapidly (around $z = 0$ Å), being either close to 0 or close to 1 away or when passing through the pore constriction, respectively. Averaging the simulated I_{ex} values over a 10 Å interval centered at $z = 0$ Å yielded the excluded current values in close agreement with experiment for the three pores and the three peptides (**Figure 7G-I**). Interestingly, the computationally derived sensing region, which corresponds well with the sensing region estimated experimentally,²⁷ is located near the electrostatic potential minima and the two steric barriers.

The ability of our SEM model to quantitatively describe the experimental blockade current data indicates that the origin of the observed current blockades is steric exclusion, similar to the mechanism enabling amino acid differentiation in the aerolysin nanopore.¹ Compared to the WtFraC-T1 pore, modifications in the G13-FraC-T1 and FraC-T2 pores increase the steric barrier that a peptide needs to overcome to complete the translocation while simultaneously deepening the electrostatic well that keeps the peptides tethered to the entrance of the nanopore constriction in the *cis* vestibule of FraC. According to our experiments, such tethering is further facilitated through hydrophobic interactions between the peptides and the hydrophobic side chains of FraC introduced through protein engineering.

Figure 7: MD simulation of peptide translocation through FraC pores. **A:** Steered MD simulation of a peptide translocation through a WtFraC-T1 pore. The pore is shown as a gray cutaway surface embedded in a DPhPC lipid bilayer (blue). Water and ions are not shown for clarity. The Ang-I peptide (DRVYIHPFHL) is placed at the *cis*-side rim of the pore (backbone shown in green). The peptide is pulled through the pore at a constant velocity v of 1 Å/ns using a spring attached to a template particle. Additional forces F are applied in the x - y plane to constrain the motion of the peptide to the symmetry axis of the pore (white dashed line). **B:** Conformation of Ang-I in the sensing region (orange) of the G13F-FraC-T1 pore. **C:** The force exerted by the SMD spring (running average, 0.5 Å) vs the z coordinate of Ang-I peptide in the three pores. Green vertical lines show the approximate locations of the two steric barriers; the yellow line shows the location of the electrostatic potential minima (Fig 5E). **D-F:** The relative excluded current through the three pores, calculated using SEM (running average, 2 Å), as a function of the center of mass z coordinate of the VYIHPF segment of the three peptides, Ang-I, Ang-III and Ang-IV (sequence in inset). The dashed lines show the average experimental excluded current fraction. The sensing region is highlighted in orange. **G-I:** Simulated vs experimental average excluded current fractions for the three peptides (colored as in D-F) in the three pores. The error bars show the standard deviation of I_{ex} values over the 50 to 60 peptide conformations used to compute each average value.

2.3. Conclusions

Nanopores are emerging as powerful single-molecule sensors for DNA and RNA sequencing devices. Recent advances in nanopore analysis revealed that peptides might be recognized by nanopore current.^{1,15,27,28,42,45–49} Moreover, it has been reported that the size and volume (which relate to the mass) of generic peptides might be measured using nanopores.^{27,29} Notably, by lowering the pH of the solution to 3.8, peptides can be captured despite their chemical composition.^{27,28} Furthermore, it has also been suggested that if a protease-unfoldase pair is coupled directly above the nanopore, the nanopore approach might allow single-molecule protein identification.²⁷

Two of the main challenges in nanopore peptide analysis include the ability to control the speed of peptide translocation, which is often too fast for accurate peptide analysis, and the rate at which the biopolymers are captured by the nanopore. Previous work with DNA revealed that an effective strategy to increase the frequency of polymer capture is to line the nanopore with positive charges in order to increase the electroosmotic flow and augment the electrostatic interactions between the DNA and the nanopore.³⁷ In this work we found that the introduction of either positive or negative charges did not improve peptide retention. Instead, the introduction of aromatic residues near the constriction region of the nanopores enhanced the capture frequency of peptides, and improved significantly the discrimination among peptides of similar size. The former is important in proteomic applications where the volume of peptides must be identified with high precision. The latter is important in single-molecule applications, where all peptides or amino acids cleaved by a peptidase must be captured in sequence.

We performed molecular dynamic simulations to improve our molecular understanding of how peptides are identified by the nanopore. We found that peptides are trapped between two electrostatic energy barriers near D10 at the constriction and K20 about 1 nm above the constriction. Interestingly, the introduction of aromatic residues at the energy minimum in the middle of the sensing region at position 13 increased both the capture frequency and the dwell time of peptides inside the nanopore. Since the neutral amino acid substitutions did not affect the electroosmotic properties of the nanopore, this effect is likely to be mediated by a more efficient trapping of peptides transiting the nanopore mediated by cation- π interaction between the nanopore and the positively charged peptides at pH 3.8. Importantly, the MD simulations revealed that the peptide ionic blockades scaled well with the excluded volume of the peptide inside the nanopore, indicating that this approach should be suitable to develop a nanopore peptide size identifier.

2.4. Materials and Methods

Chemicals. Sphingomyelin (Porcine brain, $\geq 99\%$, CAS# 383907-91-3) and diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC, $\geq 99\%$, CAS# 207131-40-6) were retrieved from Avanti Polar Lipids. Ni-NTA resin was obtained from Qiagen. Lysozyme (Albumin free for tryptic digest, CAS# 12650-88-3), Glucose ($\geq 99\%$, CAS# 50-99-7), Sodium chloride ($\geq 99.5\%$, CAS# 7647-14-5), Potassium chloride ($\geq 99\%$, CAS# 7447-40-7), Dithiothreitol (DTT, $\geq 99.0\%$, 3483-12-3), Trizma[®] HCl ($\geq 99\%$, CAS# 1185-53-1), Trizma[®] base ($\geq 99.9\%$, CAS# 77-86-1), Imidazole ($\geq 99\%$, CAS# 288-32-4), n-Dodecyl β -D-maltoside (DDM, $\geq 99\%$, CAS# 69227-93-6), Hydrochloric acid (1 M, CAS# 7647-01-0), Urea ($\geq 99.5\%$, CAS# 57-13-6), Magnesium chloride ($\geq 98.5\%$, CAS# 7786-30-3), LB Broth (Luria/Miller), Agar-agar and 2x YT Broth were obtained from Carl Roth. Ampicillin sodium salt (CAS# 69-52-3), Isopropyl β -D-1-thiogalactopyranoside (IPTG, $\geq 99\%$, CAS# 367-93-1), Ethanol ($\geq 99.8\%$, CAS# 64-17-5) and all enzymes were received from Fisher Scientific. Lysozyme from chicken egg white (for Lysis, CAS# 12650-88-3), *N,N*-Dimethyldodecylamine *N*-oxide (LDAO, $\geq 99.0\%$, CAS# 1643-20-5), Pentane ($\geq 99\%$, CAS# 109-66-0), Iodoacetamide (IAA, $\geq 99\%$, CAS# 144-48-9), Bis-tris propane ($\geq 99.0\%$, CAS# 64431-96-5) were bought from Sigma-Aldrich. n-Hexadecane (99 %, CAS# 544-76-3) and Citric acid (99.6 %, CAS# 77-92-9) were purchased from Acros. Trypsin (bovine pancreas, CAS# 9002-07-7) was obtained from Alfa Aesar.

Fragaceatoxin C (FraC) monomer expression and purification. pT7-SC1 vector containing His₆-tagged FraC plasmids (MA-FraC-GSAHHHHHH, hereafter FraC) were electrochemically inserted into *E. coli* BL21 (DE3) cells and grown overnight at 37 °C on LB agar plates supplemented with 100 mg/l ampicillin and 1% glucose. Colonies were used to inoculate 200 ml 2xYT medium supplemented with 100 mg/l ampicillin and grown at 37 °C until the optical density at 600 nm (OD₆₀₀) reached 0.6, after which expression was induced using 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG), allowing continued growth overnight at 21 °C. Cell pellets were collected by centrifugation (6,000g, 20 min, 4 °C) and stored at -80 °C for at least one hour. The pellets were resuspended in 10 ml lysis buffer per 50 ml culture, with a lysis buffer consisting of 150 mM NaCl, 15 mM Tris base solution at pH 7.5 supplemented with 1 mM MgCl₂, 2 M Urea, 20 mM imidazole, 0.2 mg/ml lysozyme and 0.2 units/ml DNase. The solution was mixed for 1 hour at room temperature (21°C) using a rotating mixer at 15 RPM. The cells were fully disrupted by sonification, applying 30 sweeps (duty cycle 30%, output control 3) three times using a Branson Sonifier 450. The lysate was centrifuged at 6000g for 20 minutes at 4 °C. The supernatant was incubated for 1 hour, while under constant rotation (15 RPM), with 100 μ L resuspended Ni-NTA resin (resuspended in 150 mM NaCl, 15 mM Tris base at pH 7.5 supplemented with 20 mM imidazole). The solution was loaded onto a prewashed Micro Bio-Spin column (Bio-Rad). The Ni-NTA beads were extensively washed with 20 ml WB (150 mM NaCl, 15 mM Tris base at pH 7.5

supplemented with 20 mM imidazole). The column was inserted into a microtube and spin-dried using a centrifuge (13,300g, 1 min) in order to remove residual wash buffer. 150 µl of 150 mM NaCl, 15 mM Tris base solution at pH 7.5 supplemented with 300 mM imidazole (EB) was added and left to incubate for 5 minutes before elution. This step was repeated four times to retrieve four fractions containing FraC monomers. The presence and purity of FraC monomers was estimated using SDS-PAGE. Pure fractions were pooled and stored at 4 °C. The concentration of FraC monomers was estimated using a Nano Drop 2000 UV-Vis Spectrophotometer (Thermo Scientific) using the elution buffer as blank.

Sphingomyelin-DPhPC liposomes preparation. 25 mg sphingomyelin (Brain, Porcine) was mixed with 25 mg 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC) and dissolved in 4 ml pentane containing 0.5 v/v% ethanol. The lipid mixture was evaporated while turning inside a round bottom flask by application of a hot air stream to create a thin lipid film over the surface of the flask. The film was reconstituted into 10 ml of Sdex buffer (150 mM NaCl, 15 mM tris, pH 7.5) using a sonication bath. The liposome solution (5 mg/ml) was frozen and stored at -20 °C.

Fragaceatoxin C oligomerisation. Liposomes were thawed and added to FraC monomers in a lipid to protein mass ratio of 10:1. The mixture was incubated for 30 minutes at 37 °C, after which *N,N*-Dimethyldodecylamine *N*-oxide (LDAO) was added to a final concentration of 0.6 v/v% to dissolve the liposomes. The solution was diluted 10-fold in 150 mM NaCl supplemented with 15 mM Tris (pH 7.5) and 0.02 v/v% n-Dodecyl β-D-maltoside (DDM). The diluted solution was combined with 100 µl of Ni-NTA, prewashed using WB2 (150 mM NaCl, 15 mM Tris base, pH 7.5 supplemented with 20 mM imidazole and 0.02 v/v% DDM). The mixture was left to incubate for 30 minutes while mixing under constant rotation (15 RPM). The solution was loaded onto a Micro Bio-Spin column (Bio-Rad), prewashed with 500 µl WB2. The Ni-NTA beads were washed extensively using 10 ml WB2. The column was spin-dried in a microtube using a centrifuge (13,300g, 1 min) to remove residual wash buffer. 150 µl elution buffer was added onto the column (150 mM NaCl, 15 mM Tris base supplemented with 1 M imidazole and 0.02 v/v% DDM) and left to stand for 10 minutes before elution into a clean microtube by centrifugation (13,300g, 2 min). The oligomers are stable for several months at 4 °C and can be frozen at -80 °C for long-term storage.

Construction of Fragaceatoxin C mutants. Fragaceatoxin C mutant DNA was prepared using the MEGAWHOP method.⁵⁰ The megaprimer was constructed using a forward primer synthesized by Integrated DNA Technologies (IDT, see **Supplementary Table 7**) and a T7 reverse primer (5'-GCTAGTTATTGCTCAGCGG-3'). Six reactions were performed per mutation—in order to receive enough DNA for the second PCR—using 25 µl REDTag® ReadyMix™ PCR Reaction Mix (Sigma-Aldrich) combined with 22 µl PCR grade water (Sigma-Aldrich), 1 µl of each forward and reverse primer and 1 µl His₆-

tagged Fragaceatoxin C template DNA. The PCR protocol consisted of a 90 second denaturation step at 95 °C followed by 30 cycles of denaturation at 95 °C (15 seconds), annealing at 55 °C (15 seconds) and extension at 72 °C (120 seconds). The six PCR reactions were combined and purified using a GeneJET PCR Purification Kit (Thermo Scientific). For the second PCR, 10 µl 5x Phire Buffer (Thermo Scientific) was combined with 1 µl template DNA, 1 µl dNTPs (10 mM), 2 µl megaprimer (first PCR), 35 µl PCR grade water (Sigma-Aldrich) and 1 µl Phire II Hot Start DNA Polymerase (Thermo Scientific). The PCR protocol consisted of an initial pre-denaturing step of 98 °C (30 seconds) followed by 25 cycles of denaturation at 98 °C (5 seconds) and extension at 72 °C (90 seconds). 5.7 µl 5x FD green buffer (Thermo Scientific) and 1 µl Dpn1 enzyme (Thermo Scientific) was added to the PCR mix and let to digest at 37 °C for 1-3 hours. 0.5 µl of the digested product was electrochemically transformed into 50 µl *E. coli* 10G® (Lucigen) competent cells and grown on LB agar plates containing 100 mg/l ampicillin and 1% glucose. Single colonies were enriched using a GeneJET Plasmid Miniprep Kit (Thermo Scientific) and the sequence was confirmed using the sequencing service of Macrogen Europe.

Protein sequence of His₆-tagged wild type Fragaceatoxin C. MASADVAGAVIDGAGLGFDVLKTVLE

ALGNVKRKIAVGIDNESGKTWTAMNTYFRSGTSDIVLPHKVAHGKALLYNGQKNRGPVATGVVGVIAYSMSDGNT
LAVLFSVPYDYNWYSNWWNVRYKGQKRADQRMYEELYHRSFPRGDNGWHSRGLGYGLKSRGFMNSSGHAI
LEIHVTKAGSAHHHHHH

Unspecific lysozyme digestion. Lysozyme (Carl Roth, From chicken egg white, free from albumin) was dissolved in 8 M urea supplemented with 15 mM Tris (pH 9.5) to a final concentration of 20 mg/ml and left to denature at 95 °C for 5 minutes. 200 µl denatured lysozyme solution was incubated for 30 minutes at 37 °C with 20 mM dithiothreitol (DTT), to reduce the cysteine residues. Iodoacetamide (IAA) was added to the mixture, to react with reduced cysteines, with a final concentration of 45 mM and incubated in the dark for 30 minutes at room temperature. The mixture was diluted 5x with 100 mM Tris (pH 8.5) and trypsin (Alfa Aesar™ Trypsin, bovine pancreas) was added in a ratio of 1:50 (trypsin:protein). The mixture was left to digest overnight (~18 hours) at 37 °C. In order to denature and deactivate any remaining trypsin, the next day, the final mix was denatured at 95 °C for 10 minutes and HCl was added to lower the pH (approximately pH 4). The mixture was then frozen at -20 °C until use.

Planar lipid bilayer electrophysiological recordings. The electrophysiology chamber consisted of two compartments separated by a 25 µm thick Teflon (Goodfellow Cambridge Ltd) membrane (**Supplementary Figure 4**). The Teflon membrane contained an aperture with a diameter of approximately 100-200 µm. Lipid membranes were formed by first applying 5 µl of 5% hexadecane (Sigma Aldrich) in pentane (Sigma Aldrich) to the Teflon membrane, near the aperture. The pentane

was left to dry and 400 µl of buffer (1 M KCl, 50 mM citric acid, titrated with bis-tris propane to pH 3.8) was added to both sides. 20 µl of a 6.25 mg/ml solution of DPhPC dissolved in pentane was added on top of the buffer on each side of the chamber. The chamber was left to dry for ~2 minutes to allow evaporation of pentane. Silver/silver chloride electrodes were attached to each compartment. The *cis* compartment was connected to the ground electrode and the *trans* was connected to the working electrode. Planar lipid bilayers were created using the Langmuir-Blodgett method described before.⁵¹ The orientation of FraC nanopores was determined by the asymmetry of the current-voltage relationship. A baseline of 2 minutes was recorded for each of the pores recorded. Analytes were added to the *cis* compartment of the chamber.

Ion permeability measurement. A single channel was obtained in a symmetrical buffer containing 2M KCl supplemented with 50 mM Citric acid titrated to pH 3.8 using bis-tris-propane, or 15 mM Tris at pH 7.5. The electrodes were connected via 2.5 % agarose salt-bridges containing 3M KCl in the agarose and liquid-solid interface. Upon insertion of a single nanopore, the *cis* and *trans* chambers were perfused three times using the 2M KCl buffer solution, in order to ensure the correct concentration. The electrodes were balanced and a current-voltage curve (IV-curve) was collected to ensure appropriate size and balancing. Afterwards, the *trans* solution was perfused with salt-free buffer to set the final salt concentration to 0.5 M KCl. The salt concentration was ensured by perfusion of *trans* using 0.5 M KCl buffer and perfusion of *cis* using 2 M KCl buffer. After equilibrium was reached, IV-curves between -100 and +100 mV were recorded with 1 mV steps. The zero current reversal potential (V_r) was determined using second-order polynomial regression over the IV-curve. The ion selectivity (P_{K^+}/P_{Cl^-}) was calculated using the Goldman-Hodgkin-Katz equation.

$$\frac{P_{K^+}}{P_{Cl^-}} = \frac{[\alpha_{Cl^-}]_{trans} - [\alpha_{Cl^-}]_{cis} e^{\frac{V_r F}{RT}}}{[\alpha_{K^+}]_{trans} e^{\frac{V_r F}{RT}} - [\alpha_{K^+}]_{cis}} \quad (3)$$

Where $[\alpha_{K^+/Cl^-}]_{cis/trans}$ are the molar ion activities of K⁺ or Cl⁻ in *cis* and *trans* as described previously²⁷. V_r is the reversal potential in J·C⁻¹. F is the Faraday constant in C·mol⁻¹. R is the gas constant in J·K⁻¹·mol⁻¹. T is the temperature in Kelvin.

Data recording. Recordings of ionic currents were obtained using an Axopatch 200B (Axon Instruments) combined with a Digidata 1550B A/D converter (Axon instruments), similar to preceding work.^{24,27,28,51} The sampling frequency was set at 50 kHz for analyte recordings, the analogue Bessel filter was set at 10 kHz. Data was recorded using Clampex 10 (Molecular Devices).

Data analysis. Data was analysed using Jupyter Notebook (version 5.5.0) running with Python 3.6.5 (64-bit), both within the Anaconda (version 5.2.0) environment. Additional packages were installed from PyPi using pip (version 10.0.1) unless stated otherwise. Axon Binary Format files were loaded and

converted into NumPy arrays (NumPy version 1.14.3) using neo (version 0.7.1) and an *ad-hoc* script, clustering was performed using SciPy (version 1.4.1.). The open pore current (I_o) of all traces was determined by calculating the mean current of 3 independent measurements, bootstrapped for 100 iterations of 10 second snippets for each measurement, similarly so, the standard deviation of the open pore current ($\sigma(I_o)$) was calculated. The error displayed in **Supplementary Table 1** was calculated over the 3 resulting values. For event detection, the baseline current and standard error of the recorded traces were determined from a full current histogram of the blank measurement. The value for the baseline was then used to determine the events when analyte was added. All data points above the baseline current and standard error that are separated by at least two times the sampling periods are detected as events. The excluded current ($I_{ex}\%$) of each event was calculated from the complement to 100% of the event signal divided by the median current of the preceding open pore current.

Impartial event detection. We found that short lived events—with a dwell time near the sampling frequency—tend to form a spike or Gaussian profile due to under sampling and filtering effects, while longer events follow a flat-top shape. Therefore, we introduced a parameter describing the shape of current blockades in order to impartially compare the performance of mutant pores. We assume that the profile of ionic current blockades can be described by a generalized flat-top normal distribution function (gNDF, **Equation 4**). Each observed block was fit to equation 3 using least-squares fitting, due to the non-polynomial nature of the function.

$$f(x) = \Delta I_B * \exp\left(-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)^\beta\right) + I_o \quad \text{for } \beta > 0 \quad (4)$$

Where μ is the events centre in the time domain with variance σ^2 and ΔI_B is the current difference (pA) between the baseline (I_o) and the event maximum. The variable β describes the shape of the function and can take any real number larger than zero (**supplementary Figure 5**). If β is less than one but larger than zero, the shape of the function is a spike (**supplementary Figure 5a**). If β is equal to one, the function is equal to the normal distribution function. When β is larger than one, the function starts to follow a rectangular—flat-top—profile. Advantageously, the variable β can also be used to assess the quality of individual events in the following way. Events with a $\beta < 1$ are mostly events that are too short-lived to accurately measure the ionic current blockade. Therefore, only those events with a $\beta \geq 1$ should be regarded as accurate measurements of peptides. Similarly, we distinguish events with a $\beta \geq 10$, since these events—having a flat-top shape—permit an accurate estimation of the blocked current. The gNDF fit also permits an estimation of the dwell time of an event by taking the full width at half maximum (FWHM) of the gNDF (**Equation 5**). Estimation of the dwell time using this equation is advantageous, because it allows the treatment of this parameters as continuous rather than discrete, which is the case if the number of data points are counted within the event.

$$FWHM = 2\sigma\sqrt{2^{\beta}\sqrt{\ln 2}} \quad (5)$$

Where σ equals the square root of the variance (σ^2 , **Equation 4**) and β describes the shape parameter.

Spectral matching. Several of the residual current spectra we obtain (**Supplementary Figure 3**) are expected to contain random events induced by factors other than the analyte (gating), so in order to reduce baseline sloping and to maintain high sensitivity, we utilize the squared first derivative Euclidean cosine correlation (**Equation 6**).⁵² This comparison is sensitive to the position of the peaks observed in the spectra, but not as sensitive to a shifting baseline.

$$Correlation = \frac{(\sum_i \Delta A_{1,i} \Delta A_{2,i})^2}{\sum_i \Delta A_{1,i}^2 \sum_i \Delta A_{2,i}^2} \quad (6)$$

Where A_1 and A_2 equal the vectors of excluded current counts and $A_{1,i}$ and $A_{2,i}$ represent the individual bins of the excluded current spectrum.⁵² In a more detailed description, we set A_1 and A_2 as the vector of counts we observe for each residual current bin (e.g. A_n = counts (40-41%), counts(41-42%), ..., counts(94-95%)). ΔA_n is the derivative of A_n (difference between bins). In the numerator, we multiply each element ΔA_n with the corresponding ΔA_n of the comparing spectrum and take the squared sum of all items. In the denominator, we take the squared sum of each element in ΔA_n and multiply that with the squared sum of each element in the spectrum we want to compare. So, if the two vectors A_1 and A_2 are equal, the correlation is 1, else it is less than 1, and because the derivative of A_1 and A_2 is taken, linear baseline sloping is less impactful.

We perform hierarchal clustering using the Ward distance as implemented in SciPy version 1.4.1. on the resulting correlation coefficients to determine which spectra are most similar.^{41,53} In essence, this metric orders the data in such a way that the variance between neighbours is minimal, therefore building a map of similar spectra.

MD Methods. All MD simulations were carried out using NAMD,⁵⁴ a 2fs integration timestep, periodic boundary conditions and the CHARMM36⁵⁵ force field. SETTLE and RATTLE algorithms were respectively used to fix all water and protein bonds containing hydrogen atoms.⁵⁶ Constant pressure (NPT) simulations used the Nosé-Hoover Langevin piston pressure control.⁵⁷ The temperature in the simulation system was maintained by coupling the non-hydrogen atoms of the lipids to a Langevin thermostat.⁵⁸ Van der Waal forces were calculated with a cut-off of 12 Å and a switching distance of 10 Å. The particle mesh Ewald summation was used for calculating long range electrostatics over a 1 Å grid.⁵⁹ Multiple time stepping was used to calculate local interactions every 2fs and the full electrostatics every 4fs.⁶⁰

All-atom model of FraC nanopores. The initial structural models of WtFraC-T1 and FraC-T2 nanopores were taken from the previous study.²⁷ The WtFraC-T1 is available as PDB ID 4TSY, the FraC-T2 was created in the previous work, from the monomer using Rosetta.⁶¹ The G13F-FraC-T1 variant was created by mutating the 13th residue of each WtFraC-T1 monomer from glycine to phenylalanine, retaining the backbone conformation. Each protein was arranged to have its nanopore axis aligned to the z axis of the simulation system and merged with a 14 nm × 14 nm pre-equilibrated patch of DPhPC lipid bilayer such that the centre of mass of residues 4-31 (pore) coincided with that of the bilayer. All lipid and water molecules overlapping with the protein were removed and the resulting system was solvated with pre-equilibrated TIP3P water,⁶² extending the systems size along the z axis to 16 nm. To emulate the pH 3.8 condition of the experiment, we used the Henderson-Hasselbach equation⁶³ to probabilistically assign fixed protonation states to the following titratable residues of the protein: aspartates (pKa = 3.8), glutamates (pKa = 4.5), histidines (pKa = 6.5) and lysines (pKa = 10.5). Selecting the locations of titratable groups at random, we protonated 50% of all aspartate residues, 83% of all glutamate residues and all of the histidines and lysines. Specific to FraC constriction, the D10 and D17 residues of both T1 pores had the following charge states (in the units of proton charge, clockwise, looking down from the *cis* side): -1, -1, -1, 0, -1, 0, 0 and 0, -1, -1, -1, 0, -1, 0, 0, respectively. The D10 and D17 charge distribution for the T2 pore was: -1, 0, -1, -1, 0, -1, -1 and 0, 0, 0, -1, 0, -1, -1, respectively. This initial assignment of the protonation states was maintained throughout the simulation. The total charge of type I and type II FraC nanopores was +177e and +155e, respectively, where *e* is the charge of a proton. To produce 1 M KCl solution, Cl⁻ ions were added first to neutralize the system, followed by the addition of approximately 1500 K⁺ and Cl⁻ ions. Each final system measured 14nm×14nm×16nm and contained about 330,000 atoms.

Each system underwent 1000 steps of energy minimization followed by equilibration in the constant-ratio NPT ensemble at 293 K and 1 atm for 6 ns, which was performed by restraining all non-hydrogen atoms of the protein to their initial coordinates. In the first nanosecond of the equilibration, water molecules were forced out from the interface between the protein and the lipid membrane using a custom *tlforces* script. During the equilibration, the system's size was allowed to fluctuate while maintaining the ratio of the unit cell's dimensions along the x and y axes constant. The final frame of the equilibration simulation was used to initiate electric field simulations, which were performed in the NVT ensemble. A constant electric field, *E* was applied normal to the lipid bilayer corresponding to a transmembrane bias of $V = -E \cdot L_z$,⁶⁴ where L_z was the length of the simulation unit cell along the z axis. The C α atoms of the nanopore were restrained to their initial coordinates using a harmonic spring with the spring constant of 6.95 pN/nm.

MD simulation of angiotensin peptide translocation. A model of the Ang-I peptide (sequence: DRVYIHPFHL) was created by mutating an alpha-helical homopeptide. The aspartate and histidine amino acids were protonated to reflect the 3.8 pH of the solution. The peptide was introduced into each equilibrated FraC nanopore system, placing the peptide at the center of the nanopore, at the level of the FraC's *cis* side rim. Water and ions overlapping with the peptide were removed, while keeping the systems electrically neutral. Each system was then equilibrated for 5 ns in the NPT ensemble while restraining all C α atoms of both FraC nanopore and the peptide. From the last frame of each equilibration trajectory, an SMD simulation⁵⁶ was carried out to translocate the Ang-I peptide through the pore. The C α atoms of the nanopore were kept harmonically restrained using a spring constant of 6.95 pN/nm. The SMD force was applied to the center of mass of all non-hydrogen atoms of the peptide using a spring constant of 4,865 pN/nm. The other end of the SMD spring was moved long along the nanopore axis with a constant velocity of 1 Å/ns. Additional forces were applied to the peptide using the colvars module of NAMD⁶⁴ to restrain the center of mass of the peptide in the XY plane to the symmetry axis of the pore. A constant electric field was also applied corresponding to a transmembrane potential difference of -50 mV.

SEM analysis of peptide translocation simulations. For each SMD trajectory, coordinates of all non-solvent atoms were saved every 0.2 ns and analyzed using the SEM approach⁵⁷ to obtain a nanopore ionic current value for each atomic configuration. The conformations for Ang-III (sequence: RVYIHPF) and Ang-IV (sequence: VYIHPF) were computationally created for each frame of the SMD trajectories by deleting terminal amino acids from the Ang-I peptide. Produced thereby sequence of conformation was analyzed using SEM as described previously¹.

2.5. References

- (1) Ouldali, H.; Sarthak, K.; Ensslen, T.; Piguet, F.; Manivet, P.; Pelta, J.; Behrends, J. C.; Aksimentiev, A.; Oukhaled, A. Electrical Recognition of the Twenty Proteinogenic Amino Acids Using an Aerolysin Nanopore. *Nat. Biotechnol.* **2020**, *38* (2), 176–181. <https://doi.org/10.1038/s41587-019-0345-2>.
- (2) Cao, C.; Ying, Y.-L.; Hu, Z.-L.; Liao, D.-F.; Tian, H.; Long, Y.-T. Discrimination of Oligonucleotides of Different Lengths with a Wild-Type Aerolysin Nanopore. *Nat. Nanotechnol.* **2016**, *11* (8), 713–718. <https://doi.org/10.1038/nnano.2016.66>.
- (3) Soskine, M.; Biesemans, A.; Maglia, G. Single-Molecule Analyte Recognition with ClyA Nanopores Equipped with Internal Protein Adaptors. *J Am Chem Soc* **2015**, *137* (17), 5793–5797. <https://doi.org/10.1021/jacs.5b01520>.
- (4) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proc. Natl. Acad. Sci.* **1996**, *93* (24), 13770–13773. <https://doi.org/10.1073/pnas.93.24.13770>.
- (5) Gu, L. Q.; Braha, O.; Conlan, S.; Cheley, S.; Bayley, H. Stochastic Sensing of Organic Analytes by

- a Pore-Forming Protein Containing a Molecular Adapter. *Nature* **1999**, 398 (6729), 686–690. <https://doi.org/10.1038/19491>.
- (6) 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 (10), 1091–1095. <https://doi.org/10.1038/80295>.
 - (7) Sutherland, T. C.; Long, Y.-T.; Stefureac, R.-I.; Bediako-Amoa, I.; Kraatz, H.-B.; Lee, J. S. Structure of Peptides Investigated by Nanopore Analysis. *Nano Lett.* **2004**, 4 (7), 1273–1277. <https://doi.org/10.1021/nl049413e>.
 - (8) Qiu, Y.; Hinkle, P.; Yang, C.; Bakker, H. E.; Schiel, M.; Wang, H.; Melnikov, D.; Gracheva, M.; Toimil-Molares, M. E.; Imhof, A.; Siwy, Z. S. Pores with Longitudinal Irregularities Distinguish Objects by Shape. *ACS Nano* **2015**, 9 (4), 4390–4397. <https://doi.org/10.1021/acs.nano.5b00877>.
 - (9) Asandei, A.; Schiopu, I.; Chinappi, M.; Seo, C. H.; Park, Y.; Luchian, T. Electroosmotic Trap against the Electrophoretic Force near a Protein Nanopore Reveals Peptide Dynamics during Capture and Translocation. *ACS Appl. Mater. Interfaces* **2016**, 8 (20), 13166–13179. <https://doi.org/10.1021/acsami.6b03697>.
 - (10) Bhattacharya, S.; Yoo, J.; Aksimentiev, A. Water Mediates Recognition of DNA Sequence via Ionic Current Blockade in a Biological Nanopore. *ACS Nano* **2016**, 10 (4), 4644–4651. <https://doi.org/10.1021/acs.nano.6b00940>.
 - (11) Luo, L.; German, S. R.; Lan, W.-J.; Holden, D. A.; Mega, T. L.; White, H. S. Resistive-Pulse Analysis of Nanoparticles. *Annu. Rev. Anal. Chem. (Palo Alto, Calif.)* **2014**, 7, 513–535. <https://doi.org/10.1146/annurev-anchem-071213-020107>.
 - (12) Kowalczyk, S. W.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Modeling the Conductance and DNA Blockade of Solid-State Nanopores. *Nanotechnology* **2011**, 22 (31), 315101. <https://doi.org/10.1088/0957-4484/22/31/315101>.
 - (13) Robertson, J. W. F.; Rodrigues, C. G.; Stanford, V. M.; Robinson, K. A.; Krasilnikov, O. V.; Kasianowicz, J. J. Single-Molecule Mass Spectrometry in Solution Using a Solitary Nanopore. *Proc. Natl. Acad. Sci.* **2007**, 104 (20), 8207–8211. <https://doi.org/10.1073/pnas.0611085104>.
 - (14) Baaken, G.; Ankri, N.; Schuler, A.-K.; Rühle, J.; Behrends, J. C. Nanopore-Based Single-Molecule Mass Spectrometry on a Lipid Membrane Microarray. *ACS Nano* **2011**, 5 (10), 8080–8088. <https://doi.org/10.1021/nn202670z>.
 - (15) Chavis, A. E.; Brady, K. T.; Hatmaker, G. A.; Angevine, C. E.; Kothalawala, N.; Dass, A.; Robertson, J. W. F.; Reiner, J. E. Single Molecule Nanopore Spectrometry for Peptide Detection. *ACS Sensors* **2017**, 2 (9), 1319–1328. <https://doi.org/10.1021/acssensors.7b00362>.
 - (16) Zernia, S.; van der Heide, N. J.; Galenkamp, N. S.; Gouridis, G.; Maglia, G. Current Blockades of Proteins inside Nanopores for Real-Time Metabolome Analysis. *ACS Nano* **2020**, 14 (2), 2296–2307. <https://doi.org/10.1021/acs.nano.9b09434>.
 - (17) Soskine, M.; Biesemans, A.; Moeyaert, B.; Cheley, S.; Bayley, H.; Maglia, G. An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry. *Nano Lett.* **2012**, 12 (9), 4895–4900. <https://doi.org/10.1021/nl3024438>.
 - (18) Soskine, M.; Biesemans, A.; De Maeyer, M.; Maglia, G. Tuning the Size and Properties of ClyA Nanopores Assisted by Directed Evolution. *J Am Chem Soc* **2013**, 135 (36), 13456–13463. <https://doi.org/10.1021/ja4053398>.

- (19) Yusko, E. C.; Bruhn, B. R.; Eggenberger, O. M.; Houghtaling, J.; Rollings, R. C.; Walsh, N. C.; Nandivada, S.; Pindrus, M.; Hall, A. R.; Sept, D.; Li, J.; Kalonia, D. S.; Mayer, M. Real-Time Shape Approximation and Fingerprinting of Single Proteins Using a Nanopore. *Nat. Nanotechnol.* **2017**, *12* (4), 360–367. <https://doi.org/10.1038/nnano.2016.267>.
- (20) Si, W.; Aksimentiev, A. Nanopore Sensing of Protein Folding. *ACS Nano* **2017**, *11* (7), 7091–7100. <https://doi.org/10.1021/acsnano.7b02718>.
- (21) Astier, Y.; Braha, O.; Bayley, H. Toward Single Molecule DNA Sequencing: Direct Identification of Ribonucleoside and Deoxyribonucleoside 5'-Monophosphates by Using an Engineered Protein Nanopore Equipped with a Molecular Adapter. *J. Am. Chem. Soc.* **2006**, *128* (5), 1705–1710. <https://doi.org/10.1021/ja057123+>.
- (22) Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. Single-Nucleotide Discrimination in Immobilized DNA Oligonucleotides with a Biological Nanopore. *Proc Natl Acad Sci U S A* **2009**, *106* (19), 7702–7707. <https://doi.org/10.1073/pnas.0901054106>.
- (23) Akeson, M.; Branton, D.; Kasianowicz, J. J.; Brandin, E.; Deamer, D. W. W. Microsecond Time-Scale Discrimination among Polycytidylic Acid, Polyadenylic Acid, and Polyuridylic Acid as Homopolymers or as Segments within Single RNA Molecules. *Biophys. J.* **1999**, *77* (6), 3227–3233. [https://doi.org/10.1016/S0006-3495\(99\)77153-5](https://doi.org/10.1016/S0006-3495(99)77153-5).
- (24) 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 Int Ed Engl* **2016**, *55* (40), 12494–12498. <https://doi.org/10.1002/anie.201606742>.
- (25) Anderluh, G.; Maček, P. Cytolytic Peptide and Protein Toxins from Sea Anemones (Anthozoa: Actiniaria). *Toxicon* **2002**, *40* (2), 111–124. [https://doi.org/https://doi.org/10.1016/S0041-0101\(01\)00191-X](https://doi.org/https://doi.org/10.1016/S0041-0101(01)00191-X).
- (26) Tanaka, K.; Caaveiro, J. M. 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* (1), 6337. <https://doi.org/10.1038/ncomms7337>.
- (27) Huang, G.; Voet, A.; Maglia, G. FraC Nanopores with Adjustable Diameter Identify the Mass of Opposite-Charge Peptides with 44 Dalton Resolution. *Nat. Commun.* **2019**, *10* (1), 835. <https://doi.org/10.1038/s41467-019-08761-6>.
- (28) 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* (1), 935. <https://doi.org/10.1038/s41467-017-01006-4>.
- (29) Chinappi, M.; Yamaji, M.; Kawano, R.; Cecconi, F. Analytical Model for Particle Capture in Nanopores Elucidates Competition among Electrophoresis, Electroosmosis, and Dielectrophoresis. *ACS Nano* **2020**, *14* (11), 15816–15828. <https://doi.org/10.1021/acsnano.0c06981>.
- (30) Laohakunakorn, N.; Keyser, U. F.; Nadanai, L.; Ulrich, F. K. Electroosmotic Flow Rectification in Conical Nanopores. *Nanotechnology* **2015**, *26* (27), 275202. <https://doi.org/10.1088/0957-4484/26/27/275202>.
- (31) Bonome, E. L.; Cecconi, F.; Chinappi, M. Electroosmotic Flow through an α -Hemolysin Nanopore. *Microfluid. Nanofluidics* **2017**, *21* (5), 96. <https://doi.org/10.1007/s10404-017-1928-1>.
- (32) Gu, L.-Q.; Cheley, S.; Bayley, H. Electroosmotic Enhancement of the Binding of a Neutral Molecule to a Transmembrane Pore. *Proc. Natl. Acad. Sci.* **2003**, *100* (26), 15498–15503.

<https://doi.org/10.1073/pnas.2531778100>.

- (33) Andreev, V. P. Cytoplasmic Electric Fields and Electroosmosis: Possible Solution for the Paradoxes of the Intracellular Transport of Biomolecules. *PLoS One* **2013**, *8* (4), e61884. <https://doi.org/10.1371/journal.pone.0061884>.
- (34) Firnkes, M.; Pedone, D.; Knezevic, J.; Döblinger, M.; Rant, U.; Dobliger, M.; Rant, U. Electrically Facilitated Translocations of Proteins through Silicon Nitride Nanopores: Conjoint and Competitive Action of Diffusion, Electrophoresis, and Electroosmosis. *Nano Lett.* **2010**, *10* (6), 2162–2167. <https://doi.org/10.1021/nl100861c>.
- (35) Boukhet, M.; Piguet, F.; Ouldali, H.; Pastoriza-Gallego, M.; Pelta, J.; Oukhaled, A. Probing Driving Forces in Aerolysin and α -Hemolysin Biological Nanopores: Electrophoresis versus Electroosmosis. *Nanoscale* **2016**, *8* (43), 18352–18359. <https://doi.org/10.1039/C6NR06936C>.
- (36) Luan, B.; Aksimentiev, A. Electro-Osmotic Screening of the DNA Charge in a Nanopore. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **2008**, *78* (2 Pt 1), 021912. <https://doi.org/10.1103/PhysRevE.78.021912>.
- (37) Maglia, G.; Restrepo, M. R.; Mikhailova, E.; Bayley, H. Enhanced Translocation of Single DNA Molecules through α -Hemolysin Nanopores by Manipulation of Internal Charge. *Proc. Natl. Acad. Sci.* **2008**, *105* (50), 19720–19725. <https://doi.org/10.1073/pnas.0808296105>.
- (38) Butler, T. Z.; Pavlenok, M.; Derrington, I. M.; Niederweis, M.; Gundlach, J. H. Single-Molecule DNA Detection with an Engineered MspA Protein Nanopore. *Proc. Natl. Acad. Sci.* **2008**, *105* (52), 20647–20652. <https://doi.org/10.1073/pnas.0807514106>.
- (39) Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of Staphylococcal α -Hemolysin, a Heptameric Transmembrane Pore. *Science* (80-.). **1996**, *274* (5294), 1859–1865. <https://doi.org/10.1126/science.274.5294.1859>.
- (40) García-Ortega, L.; Alegre-Cebollada, J.; García-Linares, S.; Bruix, M.; Martínez-del-Pozo, Á.; Gavilanes, J. G. The Behavior of Sea Anemone Actinoporins at the Water–Membrane Interface. *Biochim. Biophys. Acta - Biomembr.* **2011**, *1808* (9), 2275–2288. <https://doi.org/10.1016/j.bbamem.2011.05.012>.
- (41) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- (42) Zhao, S.; Restrepo-Pérez, L.; Soskine, M.; Maglia, G.; Joo, C.; Dekker, C.; Aksimentiev, A. Electro-Mechanical Conductance Modulation of a Nanopore Using a Removable Gate. *ACS Nano* **2019**, *13* (2), 2398–2409. <https://doi.org/10.1021/acsnano.8b09266>.
- (43) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 224–230. [https://doi.org/10.1016/S0959-440X\(00\)00194-9](https://doi.org/10.1016/S0959-440X(00)00194-9).
- (44) Wilson, J.; Sarthak, K.; Si, W.; Gao, L.; Aksimentiev, A. Rapid and Accurate Determination of Nanopore Ionic Current Using a Steric Exclusion Model. *ACS Sensors* **2019**, *4* (3), 634–644. <https://doi.org/10.1021/acssensors.8b01375>.
- (45) Piguet, F.; Ouldali, H.; Pastoriza-Gallego, M.; Manivet, P.; Pelta, J.; Oukhaled, A. Identification

- of Single Amino Acid Differences in Uniformly Charged Homopolymeric Peptides with Aerolysin Nanopore. *Nat. Commun.* **2018**, *9* (1), 966. <https://doi.org/10.1038/s41467-018-03418-2>.
- (46) Li, S.; Cao, C.; Yang, J.; Long, Y.-T. Detection of Peptides with Different Charges and Lengths by Using the Aerolysin Nanopore. *ChemElectroChem* **2019**, *6* (1), 126–129. <https://doi.org/10.1002/celec.201800288>.
- (47) Restrepo-Pérez, L.; Wong, C. H.; Maglia, G.; Dekker, C.; Joo, C. Label-Free Detection of Post-Translational Modifications with a Nanopore. *Nano Lett.* **2019**, *19* (11), 7957–7964. <https://doi.org/10.1021/acs.nanolett.9b03134>.
- (48) Restrepo-Pérez, L.; Huang, G.; Bohländer, P.; Worp, N.; Eelkema, R.; Maglia, G.; Joo, C.; Dekker, C.; Restrepo-Perez, L.; Huang, G.; Bohländer, P.; Worp, N.; Eelkema, R.; Maglia, G.; Joo, C.; Dekker, C. Resolving Chemical Modifications to a Single Amino Acid within a Peptide Using a Biological Nanopore. *ACS Nano* **2019**, *13* (12), 13668–13676. <https://doi.org/10.1021/acsnano.9b05156>.
- (49) Asandei, A.; Rossini, A. E.; Chinappi, M.; Park, Y.; Luchian, T. Protein Nanopore-Based Discrimination between Selected Neutral Amino Acids from Polypeptides. *Langmuir* **2017**, *33* (50), 14451–14459. <https://doi.org/10.1021/acs.langmuir.7b03163>.
- (50) Miyazaki, K. MEGAWHOP Cloning. In *Methods Enzymol*; 2011; Vol. 498, pp 399–406. <https://doi.org/10.1016/B978-0-12-385120-8.00017-6>.
- (51) Maglia, G.; Heron, A. J.; Stoddart, D.; Japrun, D.; Bayley, H. Analysis of Single Nucleic Acid Molecules with Protein Nanopores. In *Methods in Enzymology*; 2010; Vol. 475, pp 591–623. [https://doi.org/10.1016/S0076-6879\(10\)75022-9](https://doi.org/10.1016/S0076-6879(10)75022-9).
- (52) Li, J.; Hibbert, D. B.; Fuller, S.; Vaughn, G. A Comparative Study of Point-to-Point Algorithms for Matching Spectra. *Chemom. Intell. Lab. Syst.* **2006**, *82* (1–2), 50–58. <https://doi.org/10.1016/j.chemolab.2005.05.015>.
- (53) Müllner, D. Fastcluster : Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *J. Stat. Softw.* **2013**, *53* (9). <https://doi.org/10.18637/jss.v053.i09>.
- (54) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 044130. <https://doi.org/10.1063/5.0014475>.
- (55) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. <https://doi.org/10.1021/jp973084f>.
- (56) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952–962. <https://doi.org/10.1002/jcc.540130805>.
- (57) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177–4189. <https://doi.org/10.1063/1.467468>.
- (58) Brünger, A. T. *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR*; 1992.

- (59) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.
- (60) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97* (3), 1990–2001. <https://doi.org/10.1063/1.463137>.
- (61) André, I.; Bradley, P.; Wang, C.; Baker, D.; Andre, I.; Bradley, P.; Wang, C.; Baker, D. Prediction of the Structure of Symmetrical Protein Assemblies. *Proc. Natl. Acad. Sci.* **2007**, *104* (45), 17656–17661. <https://doi.org/10.1073/pnas.0702626104>.
- (62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (63) Henderson, L. J. Concerning the Relationship between the Strength of Acids and Their Capacity to Preserve Neutrality. *Am. J. Physiol. Content* **1908**, *21* (2), 173–179. <https://doi.org/10.1152/ajplegacy.1908.21.2.173>.
- (64) Fiorin, G.; Klein, M. L.; Hénin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111* (22–23), 3345–3362. <https://doi.org/10.1080/00268976.2013.813594>.

2.6. Supporting Information

Mutation	V (mV)	I_0 (pA)	$\sigma(I_0)$ (pA)	S_0 (nS)	SNR ($I_0/\sigma(I_0)$)	Gating (events·s ⁻¹)	Median dwell time (μs)
WtFraC-T1	-50 mV	95 ± 1	2.7 ± 0.2	1.92 ± 0.02	35 ± 3	-	90 ± 6
WtFraC-T2	-50 mV	47 ± 3	2.0 ± 0.4	0.93 ± 0.06	22 ± 5	-	101 ± 12
D10G	-50 mV	46 ± 2	2.6 ± 0.3	0.92 ± 0.03	17 ± 2	-	81 ± 2
D10R[†]	+50 mV	40 ± 1	2.4 ± 0.2	0.80 ± 0.03	15 ± 5	-	82 ± 23
G13E	-50 mV	89 ± 3	2.7 ± 0.1	1.78 ± 0.05	32 ± 1	-	78 ± 1
G13N	-50 mV	75 ± 4	2.5 ± 0.2	1.49 ± 0.08	30 ± 4	-	100 ± 25
G13H	-50 mV	69 ± 5	2.4 ± 0.2	1.38 ± 0.10	30 ± 3	-	86 ± 8
G13H[†]	+50 mV	67 ± 1	3 ± 9	1.39 ± 0.03	23 ± 10	-	78 ± 25
G13V	-50 mV	65 ± 4	12 ± 6	1.31 ± 0.08	5 ± 2	4 ± 2	84 ± 2
G13W	-50 mV	64 ± 8	3.0 ± 0.2	1.29 ± 0.17	21 ± 2	7.3 ± 3.9	321 ± 61
G13F	-50 mV	77 ± 4	2.3 ± 0.2	1.53 ± 0.08	31 ± 4	7.1 ± 0.7	182 ± 29
G13Y	-50 mV	82 ± 3	2.2 ± 0.1	1.63 ± 0.06	36 ± 3	2.3 ± 0.3	219 ± 56
D17K	-50 mV	104 ± 5	2.6 ± 0.2	2.08 ± 0.11	40 ± 5	-	94 ± 7
D17Q	-50 mV	84 ± 3	2.4 ± 0.1	1.68 ± 0.06	36 ± 3	-	80 ± 5
G15T	-50 mV	100 ± 3	2.4 ± 0.5	2.00 ± 0.06	39 ± 6	-	87 ± 8
G15V V22A	-50 mV	113 ± 9	2.9 ± 0.1	2.27 ± 0.19	39 ± 2	-	85 ± 3
K20D	-50 mV	96 ± 10	2.6 ± 0.3	1.91 ± 0.20	38 ± 1	-	101 ± 7
K20D T21K	-50 mV	101 ± 18	2.9 ± 0.2	2.01 ± 0.36	33 ± 5	-	98 ± 9
T21D	-50 mV	101 ± 10	2.6 ± 0.2	2.03 ± 0.20	38 ± 1	-	93 ± 2
K20Q	-50 mV	96 ± 4	2.6 ± 0.1	1.92 ± 0.08	37 ± 2	-	92 ± 13

Supplementary Table 1. Characteristics of Fragaecatoxin C mutant nanopores. The average baseline conductance I_0 is shown with the standard deviation error $\sigma(I_0)$. The average baseline conductance (S_0) is calculated from I_0 and the applied voltage (V). The signal-to-noise ratio (SNR) is calculated by dividing the baseline conductance I_0 by its standard deviation $\sigma(I_0)$. The intrinsic gating of pores is shown for all pores with significant events in the baseline. The median dwell time and standard deviation over 3 independent measurements is shown.

<i>Bias potential (mV)</i>	<i>WtFraC T1</i>	<i>WtFraC T2</i>	<i>D10G</i>	<i>D10R</i>	<i>G13E</i>	<i>G13N</i>	<i>G13H</i>	<i>G13W</i>	<i>G13V</i>	<i>G13F</i>
-150	-297 ± 8	-140 ± 1	-140 ± 12	-106 ± 5	-297 ± 1	-220 ± 23	-191 ± 10	-171 ± 32	-206 ± 5	-176 ± 75
-145	-287 ± 6	-135 ± 1	-150 ± 32	-91 ± 12	-286 ± 1	-213 ± 22	-185 ± 10	-165 ± 32	-199 ± 5	-220 ± 7
-140	-279 ± 7	-130 ± 1	-144 ± 30	-90 ± 9	-275 ± 1	-207 ± 20	-179 ± 9	-94 ± 59	-192 ± 5	-214 ± 7
-135	-269 ± 5	-126 ± 1	-140 ± 28	-87 ± 9	-262 ± 1	-200 ± 20	-173 ± 9	-153 ± 31	-185 ± 5	-160 ± 72
-130	-258 ± 5	-124 ± 1	-134 ± 26	-83 ± 7	-253 ± 1	-193 ± 19	-168 ± 9	-147 ± 31	-179 ± 4	-202 ± 7
-125	-248 ± 4	-117 ± 1	-139 ± 39	-82 ± 8	-242 ± 1	-187 ± 18	-162 ± 9	-141 ± 30	-171 ± 4	-196 ± 7
-120	-239 ± 5	-113 ± 1	-132 ± 35	-77 ± 7	-231 ± 1	-180 ± 16	-156 ± 9	-133 ± 27	-165 ± 4	-130 ± 89
-115	-229 ± 4	-108 ± 1	-117 ± 19	-72 ± 5	-221 ± 1	-173 ± 15	-150 ± 9	-127 ± 26	-157 ± 4	-183 ± 7
-110	-219 ± 4	-104 ± 1	-102 ± 4	-64 ± 10	-210 ± 1	-166 ± 14	-144 ± 8	-121 ± 26	-151 ± 4	-177 ± 7
-105	-209 ± 3	-99 ± 1	-97 ± 3	-64 ± 5	-200 ± 1	-139 ± 41	-138 ± 8	-115 ± 25	-143 ± 3	-170 ± 7
-100	-199 ± 3	-98 ± 1	-104 ± 19	-59 ± 7	-189 ± 2	-152 ± 12	-132 ± 8	-110 ± 24	-136 ± 3	-116 ± 72
-95	-189 ± 3	-90 ± 1	-87 ± 2	-58 ± 3	-178 ± 2	-144 ± 11	-126 ± 8	-104 ± 24	-129 ± 3	-109 ± 71
-90	-178 ± 2	-86 ± 1	-83 ± 2	-56 ± 4	-168 ± 1	-137 ± 10	-120 ± 7	-99 ± 25	-122 ± 3	-106 ± 68
-85	-168 ± 2	-82 ± 1	-78 ± 2	-52 ± 5	-158 ± 2	-130 ± 9	-114 ± 7	-93 ± 22	-114 ± 3	-142 ± 8
-80	-158 ± 2	-77 ± 1	-73 ± 2	-47 ± 6	-148 ± 2	-122 ± 8	-107 ± 8	-87 ± 21	-107 ± 2	-133 ± 7
-75	-147 ± 2	-72 ± 1	-69 ± 1	-48 ± 2	-137 ± 2	-114 ± 7	-101 ± 6	-81 ± 20	-100 ± 2	-125 ± 7
-70	-137 ± 1	-68 ± 1	-64 ± 1	-44 ± 2	-127 ± 2	-106 ± 6	-95 ± 6	-76 ± 19	-93 ± 2	-118 ± 7
-65	-127 ± 1	-63 ± 1	-60 ± 1	-41 ± 1	-117 ± 2	-98 ± 6	-88 ± 6	-70 ± 17	-85 ± 2	-110 ± 6
-60	-116 ± 1	-58 ± 1	-55 ± 1	-39 ± 1	-107 ± 2	-91 ± 5	-82 ± 5	-64 ± 16	-78 ± 1	-102 ± 6
-55	-106 ± 1	-53 ± 1	-50 ± 1	-35 ± 1	-97 ± 2	-82 ± 5	-76 ± 5	-59 ± 15	-71 ± 1	-65 ± 44
-50	-96 ± 1	-48 ± 1	-46 ± 1	-32 ± 1	-88 ± 2	-74 ± 4	-69 ± 5	-54 ± 14	-63 ± 1	-85 ± 5
-45	-85 ± 1	-43 ± 1	-41 ± 1	-27 ± 4	-78 ± 2	-66 ± 3	-62 ± 4	-48 ± 13	-56 ± 1	-77 ± 5
-40	-75 ± 1	-38 ± 1	-36 ± 1	-21 ± 9	-68 ± 2	-59 ± 2	-55 ± 4	-43 ± 11	-49 ± 1	-68 ± 5
-35	-65 ± 1	-33 ± 1	-32 ± 1	-21 ± 4	-59 ± 2	-51 ± 2	-49 ± 3	-37 ± 10	-43 ± 1	-59 ± 4
-30	-55 ± 1	-28 ± 1	-27 ± 1	-15 ± 7	-50 ± 2	-43 ± 1	-42 ± 3	-32 ± 8	-36 ± 1	-51 ± 4
-25	-45 ± 1	-23 ± 1	-23 ± 1	-12 ± 6	-41 ± 2	-36 ± 1	-35 ± 2	-26 ± 7	-29 ± 1	-42 ± 3
-20	-36 ± 1	-18 ± 1	-18 ± 1	-9 ± 5	-32 ± 1	-28 ± 1	-28 ± 2	-21 ± 5	-23 ± 1	-23 ± 15
-15	-26 ± 1	-14 ± 1	-13 ± 1	-7 ± 4	-24 ± 1	-21 ± 1	-21 ± 1	-16 ± 4	-17 ± 1	-11 ± 12
-10	-17 ± 1	-9 ± 1	-9 ± 1	-4 ± 2	-15 ± 1	-13 ± 1	-14 ± 1	-10 ± 2	-11 ± 1	-16 ± 2
-5	-8 ± 1	-4 ± 1	-4 ± 1	-2 ± 1	-7 ± 1	-6 ± 1	-6 ± 1	-5 ± 1	-5 ± 1	-7 ± 1
0	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1
5	8 ± 1	4 ± 1	4 ± 1	2 ± 1	8 ± 1	6 ± 1	7 ± 1	5 ± 2	5 ± 1	8 ± 1
10	17 ± 1	9 ± 1	8 ± 1	5 ± 2	15 ± 1	12 ± 1	14 ± 1	10 ± 4	10 ± 1	16 ± 1
15	25 ± 1	13 ± 1	13 ± 1	7 ± 4	23 ± 1	19 ± 1	21 ± 2	15 ± 6	15 ± 1	17 ± 8
20	33 ± 1	17 ± 1	17 ± 1	14 ± 1	30 ± 1	25 ± 1	29 ± 3	21 ± 8	20 ± 1	22 ± 13
25	41 ± 1	21 ± 1	21 ± 1	18 ± 1	37 ± 1	29 ± 2	36 ± 3	26 ± 10	25 ± 1	16 ± 22
30	48 ± 1	25 ± 1	26 ± 1	21 ± 1	44 ± 1	36 ± 1	44 ± 4	31 ± 11	29 ± 1	36 ± 25
35	56 ± 1	29 ± 1	30 ± 1	26 ± 1	51 ± 2	40 ± 3	51 ± 5	37 ± 13	34 ± 1	42 ± 30
40	63 ± 2	33 ± 1	35 ± 1	30 ± 1	57 ± 2	47 ± 2	59 ± 6	42 ± 15	38 ± 1	47 ± 31
45	71 ± 2	37 ± 1	39 ± 2	34 ± 1	64 ± 3	52 ± 2	66 ± 6	47 ± 17	42 ± 1	22 ± 8
50	78 ± 2	40 ± 1	44 ± 2	39 ± 1	71 ± 3	53 ± 7	74 ± 7	52 ± 19	46 ± 1	60 ± 40
55	85 ± 2	44 ± 1	48 ± 2	44 ± 1	77 ± 4	59 ± 6	82 ± 8	57 ± 22	50 ± 1	44 ± 30
60	92 ± 2	47 ± 1	52 ± 3	48 ± 2	83 ± 4	62 ± 8	90 ± 9	46 ± 2	54 ± 1	57 ± 35
65	99 ± 3	50 ± 1	57 ± 3	53 ± 2	89 ± 4	66 ± 10	98 ± 10	50 ± 3	58 ± 1	54 ± 34
70	106 ± 3	54 ± 1	61 ± 3	58 ± 2	94 ± 4	69 ± 13	106 ± 11	54 ± 4	61 ± 1	51 ± 35
75	114 ± 3	57 ± 1	66 ± 4	63 ± 2	101 ± 5	76 ± 10	114 ± 12	57 ± 4	65 ± 1	73 ± 56
80	121 ± 3	61 ± 1	70 ± 4	69 ± 1	108 ± 6	84 ± 6	122 ± 12	61 ± 5	69 ± 1	26 ± 10
85	128 ± 3	64 ± 1	75 ± 4	74 ± 1	114 ± 6	79 ± 18	131 ± 13	65 ± 5	72 ± 1	61 ± 76
90	135 ± 4	67 ± 1	80 ± 4	79 ± 2	120 ± 7	90 ± 10	140 ± 14	68 ± 4	76 ± 1	70 ± 77
95	142 ± 4	70 ± 1	84 ± 5	84 ± 2	127 ± 7	87 ± 20	149 ± 15	71 ± 5	79 ± 1	82 ± 83
100	149 ± 4	74 ± 1	89 ± 5	89 ± 2	131 ± 8	90 ± 21	158 ± 16	77 ± 8	82 ± 1	122 ± 93
105	157 ± 4	77 ± 1	93 ± 5	95 ± 2	139 ± 8	93 ± 25	167 ± 17	80 ± 9	87 ± 1	148 ± 85
110	164 ± 4	81 ± 1	98 ± 5	100 ± 2	145 ± 8	97 ± 26	176 ± 18	82 ± 7	91 ± 1	140 ± 63
115	172 ± 4	83 ± 1	103 ± 5	106 ± 2	152 ± 9	103 ± 25	186 ± 18	88 ± 11	94 ± 1	152 ± 51
120	179 ± 4	87 ± 1	82 ± 42	112 ± 2	105 ± 67	109 ± 24	195 ± 20	92 ± 12	98 ± 1	127 ± 93
125	144 ± 58	91 ± 1	84 ± 46	117 ± 2	166 ± 10	113 ± 25	206 ± 20	92 ± 9	102 ± 1	106 ± 95
130	193 ± 5	94 ± 1	88 ± 47	86 ± 49	172 ± 10	115 ± 30	216 ± 21	71 ± 41	106 ± 1	151 ± 50
135	155 ± 64	96 ± 1	92 ± 49	129 ± 3	180 ± 10	121 ± 29	227 ± 22	77 ± 39	110 ± 1	173 ± 63
140	211 ± 5	102 ± 1	96 ± 51	135 ± 3	187 ± 11	125 ± 31	238 ± 23	142 ± 64	115 ± 1	106 ± 67
145	219 ± 5	105 ± 1	100 ± 53	140 ± 3	194 ± 11	129 ± 33	249 ± 24	110 ± 15	118 ± 1	128 ± 62
150	228 ± 6	109 ± 1	104 ± 55	147 ± 3	200 ± 12	134 ± 35	260 ± 25	113 ± 16	93 ± 43	108 ± 79

Supplementary Table 2A. Voltage-current (IV) curves of Fragaceatoxin C mutants. Current (pA) at different applied bias potentials (mV) for Fragaceatoxin C mutants measured in 1 M KCl, 50 mM citric acid (pH 3.8).

<i>Bias potential (mV)</i>	<i>G13Y</i>	<i>G15T</i>	<i>D17K</i>	<i>D17Q</i>	<i>K20D</i>	<i>T21D</i>	<i>K20Q</i>	<i>K20D T21K</i>	<i>G15V V22A</i>
-150	-218 ± 31	-312 ± 18	-162 ± 4	-260 ± 12	-396 ± 24	-339 ± 8	-332 ± 78	-337 ± 6	-300 ± 4
-145	-211 ± 31	-302 ± 17	-159 ± 4	-252 ± 9	-382 ± 24	-327 ± 8	-320 ± 76	-325 ± 6	-291 ± 4
-140	-204 ± 31	-292 ± 17	-158 ± 4	-246 ± 8	-256 ± 96	-314 ± 7	-308 ± 73	-314 ± 5	-281 ± 3
-135	-197 ± 32	-283 ± 16	-154 ± 4	-239 ± 6	-354 ± 22	-302 ± 7	-298 ± 69	-302 ± 5	-271 ± 3
-130	-197 ± 37	-273 ± 15	-152 ± 4	-232 ± 5	-341 ± 22	-290 ± 6	-290 ± 61	-291 ± 5	-261 ± 2
-125	-183 ± 31	-263 ± 15	-149 ± 4	-226 ± 3	-326 ± 20	-278 ± 6	-277 ± 59	-279 ± 5	-251 ± 2
-120	-175 ± 32	-253 ± 14	-146 ± 4	-218 ± 1	-313 ± 21	-266 ± 5	-262 ± 61	-268 ± 5	-241 ± 2
-115	-168 ± 31	-242 ± 13	-143 ± 4	-211 ± 1	-300 ± 21	-254 ± 5	-252 ± 56	-257 ± 5	-231 ± 1
-110	-160 ± 31	-232 ± 13	-140 ± 4	-203 ± 1	-283 ± 19	-243 ± 5	-240 ± 54	-245 ± 5	-221 ± 1
-105	-153 ± 31	-222 ± 12	-137 ± 4	-195 ± 1	-269 ± 17	-231 ± 4	-231 ± 48	-234 ± 5	-211 ± 1
-100	-145 ± 31	-212 ± 12	-133 ± 4	-188 ± 1	-259 ± 16	-219 ± 3	-217 ± 47	-223 ± 4	-158 ± 59
-95	-138 ± 31	-201 ± 11	-129 ± 3	-180 ± 2	-241 ± 16	-207 ± 3	-205 ± 44	-211 ± 4	-190 ± 1
-90	-130 ± 30	-191 ± 10	-126 ± 3	-171 ± 3	-227 ± 16	-196 ± 2	-193 ± 41	-199 ± 4	-180 ± 1
-85	-122 ± 30	-180 ± 10	-121 ± 3	-163 ± 3	-212 ± 15	-184 ± 2	-182 ± 37	-188 ± 4	-170 ± 1
-80	-114 ± 29	-170 ± 9	-117 ± 3	-154 ± 3	-198 ± 14	-172 ± 2	-169 ± 35	-177 ± 3	-159 ± 2
-75	-106 ± 28	-159 ± 9	-112 ± 3	-146 ± 3	-185 ± 12	-160 ± 1	-158 ± 32	-165 ± 3	-149 ± 2
-70	-98 ± 27	-148 ± 8	-107 ± 3	-137 ± 3	-170 ± 12	-149 ± 1	-146 ± 29	-154 ± 3	-138 ± 2
-65	-90 ± 25	-137 ± 8	-102 ± 2	-128 ± 3	-156 ± 12	-137 ± 1	-134 ± 26	-142 ± 3	-128 ± 2
-60	-82 ± 24	-126 ± 7	-96 ± 2	-118 ± 3	-143 ± 11	-125 ± 1	-122 ± 23	-131 ± 3	-117 ± 2
-55	-75 ± 22	-115 ± 6	-90 ± 2	-109 ± 3	-129 ± 10	-114 ± 1	-111 ± 20	-119 ± 2	-107 ± 2
-50	-67 ± 20	-104 ± 6	-83 ± 2	-99 ± 3	-116 ± 9	-103 ± 1	-100 ± 17	-108 ± 2	-96 ± 3
-45	-56 ± 17	-93 ± 5	-76 ± 2	-90 ± 2	-103 ± 8	-91 ± 1	-88 ± 15	-97 ± 2	-86 ± 3
-40	-43 ± 17	-83 ± 5	-69 ± 2	-80 ± 2	-90 ± 7	-80 ± 1	-77 ± 13	-85 ± 2	-76 ± 3
-35	-44 ± 14	-72 ± 4	-62 ± 2	-70 ± 1	-77 ± 7	-69 ± 1	-67 ± 11	-74 ± 1	-66 ± 3
-30	-23 ± 19	-61 ± 3	-54 ± 1	-60 ± 1	-65 ± 5	-59 ± 1	-56 ± 9	-63 ± 1	-56 ± 2
-25	-20 ± 16	-50 ± 3	-46 ± 1	-50 ± 1	-53 ± 4	-48 ± 1	-46 ± 7	-52 ± 1	-46 ± 2
-20	-16 ± 12	-40 ± 2	-37 ± 1	-40 ± 1	-42 ± 4	-38 ± 1	-36 ± 5	-41 ± 1	-36 ± 2
-15	-11 ± 9	-30 ± 2	-28 ± 1	-30 ± 1	-31 ± 3	-28 ± 1	-26 ± 4	-31 ± 1	-26 ± 1
-10	-7 ± 6	-19 ± 1	-19 ± 1	-20 ± 1	-20 ± 2	-18 ± 1	-17 ± 2	-20 ± 1	-17 ± 1
-5	-2 ± 3	-9 ± 1	-9 ± 1	-10 ± 1	-9 ± 1	-9 ± 1	-8 ± 1	-10 ± 1	-8 ± 1
0	1 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1	0 ± 1
5	5 ± 2	9 ± 1	10 ± 1	9 ± 1	9 ± 1	9 ± 1	8 ± 1	9 ± 1	9 ± 1
10	4 ± 3	19 ± 1	20 ± 1	19 ± 1	18 ± 1	17 ± 1	16 ± 1	19 ± 1	17 ± 1
15	5 ± 4	28 ± 2	31 ± 2	29 ± 1	27 ± 2	26 ± 1	24 ± 2	29 ± 1	25 ± 2
20	7 ± 6	37 ± 2	42 ± 2	39 ± 1	35 ± 3	34 ± 1	31 ± 3	38 ± 1	33 ± 2
25	9 ± 8	46 ± 3	53 ± 3	49 ± 1	43 ± 4	42 ± 1	38 ± 3	47 ± 1	41 ± 3
30	10 ± 9	55 ± 4	64 ± 4	58 ± 1	51 ± 5	50 ± 2	45 ± 4	56 ± 1	49 ± 4
35	12 ± 10	63 ± 5	76 ± 5	68 ± 1	59 ± 6	57 ± 2	52 ± 4	65 ± 1	56 ± 4
40	12 ± 11	72 ± 5	88 ± 6	77 ± 1	65 ± 7	64 ± 3	58 ± 5	73 ± 1	64 ± 5
45	15 ± 12	80 ± 6	98 ± 9	86 ± 1	72 ± 8	71 ± 3	64 ± 5	82 ± 1	71 ± 6
50	18 ± 13	89 ± 7	112 ± 8	95 ± 1	79 ± 9	78 ± 4	70 ± 6	90 ± 2	78 ± 7
55	19 ± 15	97 ± 7	124 ± 9	104 ± 1	86 ± 9	85 ± 4	76 ± 6	98 ± 3	86 ± 8
60	19 ± 17	105 ± 8	136 ± 10	113 ± 2	92 ± 10	92 ± 5	82 ± 6	106 ± 3	93 ± 8
65	20 ± 21	113 ± 9	149 ± 11	122 ± 2	98 ± 11	99 ± 5	87 ± 7	113 ± 4	99 ± 9
70	24 ± 20	121 ± 9	162 ± 13	131 ± 2	104 ± 12	105 ± 6	93 ± 7	121 ± 5	106 ± 10
75	24 ± 19	129 ± 10	99 ± 40	139 ± 2	109 ± 13	112 ± 6	99 ± 7	128 ± 6	113 ± 11
80	20 ± 12	137 ± 11	142 ± 57	147 ± 3	116 ± 13	118 ± 7	104 ± 8	135 ± 7	120 ± 12
85	32 ± 25	145 ± 11	150 ± 62	155 ± 3	121 ± 14	125 ± 8	109 ± 8	143 ± 8	127 ± 13
90	32 ± 24	153 ± 12	161 ± 65	164 ± 4	128 ± 15	132 ± 8	114 ± 9	150 ± 9	134 ± 14
95	37 ± 29	161 ± 12	172 ± 68	172 ± 5	133 ± 16	138 ± 8	119 ± 10	157 ± 10	142 ± 14
100	33 ± 28	169 ± 13	181 ± 70	180 ± 7	137 ± 15	145 ± 9	124 ± 11	164 ± 11	149 ± 15
105	29 ± 20	177 ± 14	214 ± 53	188 ± 8	143 ± 15	151 ± 9	129 ± 11	171 ± 12	156 ± 15
110	33 ± 24	185 ± 15	203 ± 81	196 ± 9	147 ± 15	158 ± 10	134 ± 12	178 ± 14	163 ± 15
115	41 ± 35	194 ± 15	283 ± 24	204 ± 11	154 ± 16	166 ± 10	136 ± 12	185 ± 15	172 ± 15
120	22 ± 10	203 ± 15	298 ± 26	211 ± 12	157 ± 15	173 ± 11	144 ± 13	195 ± 16	180 ± 16
125	40 ± 33	211 ± 16	312 ± 28	218 ± 14	164 ± 18	180 ± 11	149 ± 14	200 ± 18	188 ± 18
130	44 ± 37	220 ± 16	328 ± 29	227 ± 16	169 ± 18	181 ± 5	154 ± 15	206 ± 20	196 ± 18
135	48 ± 41	229 ± 17	343 ± 31	235 ± 19	171 ± 16	194 ± 12	159 ± 16	214 ± 21	204 ± 19
140	45 ± 35	239 ± 19	358 ± 32	243 ± 21	173 ± 13	201 ± 12	164 ± 17	222 ± 23	212 ± 20
145	56 ± 43	245 ± 16	373 ± 33	251 ± 23	178 ± 13	209 ± 13	169 ± 18	222 ± 30	220 ± 20
150	59 ± 41	256 ± 18	389 ± 34	258 ± 26	192 ± 22	217 ± 13	172 ± 22	230 ± 33	174 ± 88

Supplementary Table 2B. Voltage-current (IV) curves of Fragaceatoxin C mutants. Current (pA) at different applied bias potentials (mV) for Fragaceatoxin C mutants measured in 1 M KCl, 50 mM citric acid (pH 3.8).

Sequence	Molecular weight (Da)	Charge at pH 3.8
NLCNIPCSALLSSDITASVNCAK	2336.12	+ 1.0
IVSDGNGMNAW	1162.51	0.0
GILQINSR	899.52	+ 1.2

NTDGSTDY	871.32	- 0.2
CELAAMK	835.40	+ 1.0
NTQATNR	803.39	+ 1.2
GTDVQAW	775.35	0.0
HGLDNY	717.31	+1.0
SLGNW	575.27	+ 0.2
CNDGR	563.21	+ 1.0
TPGSR	516.27	+ 1.2

Supplementary Table 3. Lysozyme peptides resulting from a non-specific trypsin/chymotrypsin digest.

Peptide sequence	Molecular weight (Da)	<i>m/z</i>	Charge	Relative abundance (%)	PTM
NLC(+57.02)NIPC(+57.02)SALLSSDITASVNC(+57.02)AK	2507.182	1254.602	2	100	Carbamidomethylation
FESNFTQATNR	1427.643	714.8295	2	85.8	
GILQINSR	899.5189	450.7684	2	79.8	
KIVSDGDGMNAW	1291.587	646.803	2	66.2	
WWC(+57.02)NDGR	992.3923	497.2051	2	65.3	Carbamidomethylation
NLC(+57.02)NIPC(+57.02)SALL	1273.616	637.8157	2	59.8	Carbamidomethylation
SLGNWVC(+57.02)AAK	1104.539	553.2784	2	59.1	Carbamidomethylation
SSDITASVNC(+57.02)AK	1251.577	626.7985	2	55.1	Carbamidomethylation
IVSDGDGMNAW	1163.492	582.7549	2	49.6	
NLC(+57.02)NIPC(+57.02)SALLSSDITASVN	2148.019	1075.021	2	37.1	Carbamidomethylation
LSSDITASVNC(+57.02)AK	1364.661	683.3383	2	34	Carbamidomethylation
NTDGSTDYGILQINSR	1752.828	877.4232	2	30.6	
FESN(+.98)FNTQATNR	1428.627	715.3217	2	30.3	Deamidation
SALLSSDITASVNC(+57.02)AK	1635.814	818.9156	2	29.1	Carbamidomethylation
NLC(+57.02)N(+.98)IPC(+57.02)SALLSSDITASVNC(+57.02)AK	2508.166	1255.093	2	28.1	Carbamidomethylation Deamidation
NLC(+57.02)NIPC(+57.02)SALLSSDITASVN(+.98)C(+57.02)AK	2508.166	1255.094	2	27.8	Carbamidomethylation Deamidation
YC(+57.02)AIVSAGQ(+.98)N(+.98)ALHFSSGSLDLAK	2523.231	842.0687	3	27	Carbamidomethylation Deamidation
AAALAGINSR	942.5247	472.2708	2	26.7	
N(+.98)LC(+57.02)NIPC(+57.02)SALLSSDITASVNC(+57.02)AK	2508.166	1255.093	2	26.6	Deamidation Carbamidomethylation
KIVSDGDGM(+15.99)NAW	1307.582	654.799	2	26.3	Oxidation
HGLDNYRGY	1093.494	547.7551	2	25.1	
ESNFTQATNR	1280.575	641.2963	2	23.9	
SSDITASVN(+.98)C(+57.02)AK	1252.561	627.2888	2	23.6	Deamidation Carbamidomethylation
SLGN(+.98)WVC(+57.02)AAK	1105.523	553.7689	2	23.2	Deamidation Carbamidomethylation
LGEDNINVVEGNEQF	1675.769	838.8934	2	23.2	
NLC(+57.02)NIPC(+57.02)SAL	1160.532	581.2738	2	23.1	Carbamidomethylation
NTDGSTDYGILQIN(+.98)SR	1753.812	877.9144	2	22.6	Deamidation
HGLDN(+.98)YR	874.3933	438.2048	2	21	Deamidation
WWC(+57.02)N(+.98)DGR	993.3763	497.6954	2	21	Carbamidomethylation Deamidation
SSDITASVNC(+57.02)AKK	1379.672	690.8423	2	20.8	Carbamidomethylation
KFESNFTQATNR	1555.738	778.8769	2	20.7	
SSGTSYDPVLK	1152.566	577.2905	2	20.4	
IVSDGDGM(+15.99)NAW	1179.487	590.7489	2	20.2	Oxidation
KIVSDGNGMNAW	1290.603	646.3096	2	17.3	
LGEDNINVVEGNEQFISASK	2162.049	721.6898	3	16.8	
GRC(+57.02)ELAAAMK	1105.537	553.7755	2	14.5	Carbamidomethylation
STALAN(+.98)QAAM	977.4488	489.7285	2	14.3	Deamidation
NTDGSTDYGILQI	1395.652	698.833	2	13.6	
NTDGSTDYGIL	1154.509	578.2621	2	13.4	
KIVSDGDGMN(+.98)AW	1292.571	647.2928	2	13.3	Deamidation
HGLDN(+.98)YRGY	1094.478	548.2425	2	12.8	Deamidation
GYSNGWVC(+57.02)AAK	1324.623	663.32	2	12.8	Carbamidomethylation
AAALAGIN(+.98)SR	943.5087	472.7613	2	12.7	Deamidation
RIVSDGDGMNAW	1319.593	660.8039	2	12.3	
SAYPGQITSNMF	1314.591	658.304	2	12.3	
NLC(+57.02)NIPC(+57.02)	889.3786	445.6983	2	11.6	Carbamidomethylation
WWC(+57.02)N(+.98)DGRTPGSR	1491.631	746.823	2	11.4	Carbamidomethylation Deamidation

ESN(+.98)FNTQATNR	1281.559	641.7872	2	11.1	Deamidation
LSSDITASVN(+.98)C(+57.02)AK	1365.645	683.8296	2	10.7	Deamidation
NLC(+57.02)NIPC(+57.02)SALLSSDITASVNC(+57.02)AKK	2635.277	1318.631	2	10.5	Carbamidomethylation
LSGNWIR	844.4555	423.2351	2	10.3	
LLSSDITASVNC(+57.02)AK	1477.745	739.8796	2	10.3	Carbamidomethylation
LC(+57.02)NIPC(+57.02)SALLSSDITASVNC(+57.02)AK	2393.139	1197.579	2	10.2	Carbamidomethylation

Supplementary Table 4. Lysozyme peptides resulting from a non-specific trypsin/chymotrypsin digest as measured by ESI-MS. The peptide sequence is shown, with mass modifications as well as the total mass of the peptide. The charge is calculated from the mass-over-charge ratio (m/z) and the relative abundance is calculated from the base peak (highest intensity). The final column shows the detection post translation modification (PTM). All spectra were analysed using PEAKS Studio and comparison with the *Gallus-Gallus* lysozyme (P00698|LYSC_CHICK) sequence.

	<i>Events·s⁻¹ where $\theta > 0$</i>	<i>Events·s⁻¹ where $\theta > 1$</i>	<i>Events·s⁻¹ where $\theta > 10$</i>
WtFraC-T1	13.5 ± 1.6	13.4 ± 1.6	7.4 ± 1
WtFraC-T2	11.4 ± 2.5	11.2 ± 2.6	7.7 ± 2
D10G	3.8 ± 0.7	3.8 ± 0.6	2.9 ± 0.5
D10R[‡]	9.9 ± 1.3	9.8 ± 1.3	7.5 ± 0.6
G13E	7.4 ± 0.8	7.1 ± 0.7	3.9 ± 0.4
G13N	17.4 ± 17.2	15.4 ± 15.4	10.5 ± 10.8
G13H	11.7 ± 2.2	11.2 ± 2	7.8 ± 2.1
G13H[‡]	3.4 ± 0.7	3.3 ± 0.7	2.4 ± 0.8
G13W	43 ± 8.8	41.9 ± 8.7	29.7 ± 7.6
G13V	17.6 ± 0.1	17 ± 0.1	11.6 ± 0.6
G13F	41.3 ± 3.5	40.9 ± 3.5	30.2 ± 1.6
G13Y	43.2 ± 8.1	42.6 ± 7.7	27.3 ± 3.8
D17K	2.8 ± 0.9	2.5 ± 0.8	1.8 ± 0.5
D17Q	1.8 ± 0.5	1.8 ± 0.5	1.2 ± 0.3
G15T	12.5 ± 1.5	12.4 ± 1.5	7.4 ± 0.4
G15V V22A	8.2 ± 0.4	8.1 ± 0.3	4.7 ± 0.5
K20D	18.5 ± 0.4	18 ± 0.3	10.2 ± 0.5
K20D T21K	11.7 ± 0.8	11.2 ± 0.7	6.3 ± 0.4
T21D	9.4 ± 1.6	9.2 ± 1.5	5.1 ± 1
K20Q	24.3 ± 10.9	22.6 ± 8.7	15.4 ± 8.6

Supplementary Table 5. Event rates resulting from the addition of unspecific lysozyme digest to FraC mutants. The event rate is shown for all events $\theta > 0$, events with $\theta > 1$, and events with $\theta > 10$.

	pH 7.5		pH 3.8	
	Reversal potential (mV)	P_{K^+}/P_{Cl^-}	Reversal potential (mV)	P_{K^+}/P_{Cl^-}
WtFraC-T1	17.2 ± 1.2*	3.6 ± 0.4*	1.0 ± 1.7	1.03 ± 0.04
WtFraC-T2	20.8 ± 1.6*	5.2 ± 0.9*	0.7 ± 4.9	1.1 ± 0.4
G13F-FraC-T1	17.0 ± 0.7	3.7 ± 0.2	0.0 ± 0.4	1.00 ± 0.03

*replicated from Huang *et al.* 2019

Supplementary Table 6. Reversal potentials of WtFraC-T1, WtFraC-T2, and G13F-FraC-T1 at pH 7.5 and pH 3.8.

Mutation	Forward Primer
G13E	5'-GGTGCGGAAGCTGGGCTTTGACGTACTG-3'
G13N	5'-CGGTGCGAATCTGGGCTTTGAC-3'
G13H	5'-GGTGCGCATCTGGGCTTTGACGTAC-3'
G13W	5'-GGTGCGTGGCTGGGCTTTGACGTACTG-3'
G13F	5'-GGTGCGTTCCTGGGCTTTGACGTACTG-3'
G13Y	5'-GGTGCGTATCTGGGCTTTGACGTACTG-3'
G15T	5'-CGGGTCTGACCTTTGACGTACTG-3'
G15V	5'-CGGGTCTGGTCTTTGACGTACTG-3'
D17K	5'-CTGGGCTTTAAGGTACTGAAAACCGTGC-3'
D17Q	5'-CTGGGCTTTCAAGTACTGAAAACCGTGC-3'
D17W	5'-CTGGGCTTTTGGGTACTGAAAACCGTGC-3'
G15V V22A	5'-CGGGTCTGGTCTTTGACGTACTG-3'
K20D	5'-CGTACTGGACACCGTGCTGGAG-3'
K20D T21K	5'-GACGTACTGGACAAGGTGCTGGAG-3'
T21D	5'-GTACTGAAAGACGTGCTGGAGGCC-3'
K20Q	5'-CGTACTGCAAACCGTGCTGGAG-3'
G15T K20D	5'-GTACTGGACACCGTGCTGGAG-3'

Supplementary Table 7. Forward primers used in conjunction with T7 reverse (5'-GCTAGTTATTGCTCAGCGG-3') for the mutation of wild-type Fragaetoxin C

Supplementary figure 1. Voltage-current (IV) curves of Fragaceatoxin C mutants. Current (pA) at different applied bias potentials (mV) for Fragaceatoxin C mutants measured in 1 M KCl, 50 mM citric acid (pH 3.8).

Supplementary figure 2. Baseline current behaviour of Fragaceatoxin C mutants at positive and negatively applied potential. Current (pA) behaviour at an positively applied bias of 100 mV (above middle line) and negatively applied bias of 100 mV (below middle line) for different Fragaceatoxin C mutants. Data was collected in 1 M KCl and 50 mM citric acid titrated with bis-tris propane to pH 3.8 at a sampling frequency of 50 kHz, using a 10 kHz Bessel filter and 5 kHz Gaussian filter.

Supplementary figure 3. Residual current histograms of (mutant) Fragaceatoxin C with trypsin digested lysozyme. **A:** Representative residual current histograms of (mutant) Fragaceatoxin C combined with equal units of trypsin digested lysozyme under an applied potential of -50 mV. **B:** Number of detected events with a β parameter larger than 100 relative to the number of events with a β larger than 1, representing the percentage of well characterizable events relative to the total number of events, for a tryptic digest of lysozyme in multiple Fragaceatoxin C mutants. Data was collected in 1 M KCl and 50 mM citric acid titrated with bis-tris propane to pH 3.8 at a sampling frequency of 50 kHz, using a 10 kHz Bessel filter and 5 kHz Gaussian filter.

Supplementary figure 4. Electrophysiology setup. The schematic shows a chamber consisting of two compartments made of Delrin, separated by a Teflon film containing a 100 μm hole. Both compartments were filled with buffer and a silver/silver-chloride electrode is attached to serve as a liquid-solid interface for the detector. A lipid membrane is formed over the hole inside the Teflon film using the Langmuir-Blodgett method and nanopore are added to the *cis* compartment to allow nanopore insertion.

Supplementary figure 5. Effect of the β parameter on the super-Gaussian function shape. Super-Gaussian function shape with $\mu=0$, $\Delta I_B=100$ and $\sigma=1$ with in **A**: $\beta = 0.5$, **B**: $\beta = 1$, **C**: $\beta = 10$.

Supplementary Figure 6: MD simulation of ion transport through mutant Fragaceatoxin C nanopores. **A:** All-atom model of WtFraC-T1 pore. The protein is shown as a gray cutaway surface, embedded in a DPhPC lipid bilayer (blue) and contains 1 M KCl solution (potassium in orange and chloride in green, water not shown). The protonation states of the titratable residues are set to reflect pH of 3.8. The z axis is shown on the left for scale. **B:** The charge carried by the chloride (blue) and potassium (red) ions through the pore as a function of the simulation time along with the total charge (black) **C:** The total number of water molecules that pass through the pore constriction as a function of simulation time. The average rate of transport is approximately 5.8 molecules/ns. **D-F:** Same as A-C but for the G13F-FraC-T1 pore. **G-I:** Same as A-C but for the Wt-FraC-T2 pore.

Supplementary Figure 7: Molecular dynamics simulation of Wt-FraC-T1 pore at pH 7. **A:** All-atom model of Wt-FraC-T1 pore. The protein is shown as a gray cutaway surface, embedded in a DPhPC lipid bilayer (blue) and contains 1 M KCl solution (potassium in orange and chloride in green, water not shown). The protonation states of the titratable residues are set to reflect pH of 7. The z axis is shown on the left for scale. **B:** Average electrostatic potential along the symmetry axis (z axis) of the pore, calculated from an 80 ns MD trajectory. **C:** Average local concentration profiles of potassium (blue) and chloride (red) ions along the symmetry axis of the nanopore. **D:** The total number of water molecules that pass through the pore constriction as a function of simulation time. The average rate of transport is approximately 14.3 molecules/ns. **E:** The charge carried by the chloride (blue) and potassium (red) ions through the pore as a function of the simulation time along with the total charge (black). Upon scaling of the raw MD current with the ratio of the experimental and simulated bulk conductivity of 1 M KCl, the average simulated ionic current through Wt-FraC-T1 pore at pH 7 and -50 mV is -187.5 +/- 6.9 pA a factor 1.5 higher than the corresponding experimental value.²¹

Supplementary Figure 8. Comparison of relative excluded currents in the sensing region. The difference in raw relative excluded current from MD simulation in the sensing region compared to the experimental differences.

Chapter 3

Protein Identification by Nanopore Peptide Profiling

Nanopores are single-molecule sensors used in nucleic acid analysis, whereas their applicability towards full protein identification has yet to be demonstrated. Here, we show that an engineered Fragaceatoxin C nanopore is capable of identifying individual proteins by measuring peptide spectra that are produced from hydrolysed proteins. Using model proteins, we show that the spectra resulting from nanopore experiments and mass spectrometry share similar profiles, hence allowing protein fingerprinting. The intensity of individual peaks provides information on the concentration of individual peptides, indicating that this approach is quantitative. Our work shows the potential of a low-cost, portable nanopore-based analyser for protein identification.

This chapter was published as:

Lucas, F. L. R., Versloot, R. C. A., Yakovlieva, L., Walvoort, M. T. C., Maglia, G. (2021).
Protein Identification by Nanopore Peptide Profiling. *Nature Communications* **12**, 5795

3.1. Introduction

The ever-increasing demand for high-throughput proteomic studies and personalized medicine requires the development of scalable and low-cost protein analysers.^{1–3} Modern proteomics relies heavily on tandem mass spectrometry (MS) for its high precision and capability to identify and quantify proteins in complex mixtures.^{3–5} However, most mass analysers are large, have a high cost of investment, are expensive to maintain, and require specialized operators to function.^{2,3,6}

In contrast to mass spectrometry devices, nanopore-based analysers provide a low-cost and high-throughput platform with the adaptability towards native environments.^{7–10} In nanopore analysis, analytes are measured as they disrupt an ionic current passing through individual nanopores under an applied potential. Importantly, the magnitude of the current blockade (I_b) is mainly proportional to the volume the analyte excludes, allowing size discrimination of chemically similar (bio)polymers such as PEG chains, DNA, proteins, and peptides.^{11–23} Furthermore, nanopores are capable of accurately detecting a variety of molecules, including proteins and DNA, with high precision at the single-molecule level.^{16,20,23–27} Peptides are of special interest for protein characterization, as they allow identification analogous to bottom-up MS-based proteomics.

Three types of Fragaceatoxin C (FraC) nanopores FraC-T1, FraC-T2 and FraC-T3, most likely corresponding to octameric, heptameric and hexameric pores, respectively, can be used in peptide nanopore analysis.²⁸ FraC nanopores have been shown to differentiate peptides with a resolution of ~40 Da,²⁸ while also enabling the detection of small chemical modifications.²⁹ Recently we have shown that at acidic pH values (less than pH 4.5)²² peptides are captured by the nanopore despite their composition, and they are most efficiently analysed at the exact pH of 3.8,²⁸ the condition under which the nanopore has no significant electroosmotic flow.¹⁵ Furthermore, we showed that the introduction of an aromatic residue in the sensing region of FraC (G13F-FraC containing a glycine to phenylalanine substitution at position 13) augmented the residence time of peptides inside the nanopore and their capture efficiency.¹⁵

In this contribution, we show that G13F-FraC-T1 can be used to directly sample proteins that are digested by a protease. The resulting collected peptides describe a spectrum that can then be used to identify proteins (**Figure 1**). A similar strategy relying heavily on the fingerprinting of (tryptic) peptides^{5,21} has been used in the early days of mass spectrometry for the detection and identification of proteins.^{30–32} The nanopore approach might then provide a low-cost and high-throughput approach to protein identification. Furthermore, if the protease is implemented directly above the nanopore,³³ this approach is amenable for single-molecule identification.

Figure 1. Graphical overview of the nanopore protein fingerprinting approach. Peptides are pre-hydrolysed by a specific protease (e.g. trypsin) and the resulting peptides are measured as they translocate the nanopore. Each peptide entering the nanopore reduces the open pore current (I_o) to the blocked pore current (I_B). The resulting excluded current ($\Delta I_B = I_o - I_B$) relates to the volume of the peptide. The subsequent histogram of the percent of excluded currents ($I_{ex}\% = \Delta I_B / I_o \%$) is used to identify the protein.

3.2. Results

3.2.1. Calibration of the G13F-FraC-T1 Nanopore

In order to characterize G13F-FraC-T1 for protein analysis, we first measured seven (synthetic) peptides (**Figure 2A**) with a mass between 500 and 1700 Da that would have resulted from the complete, *in-silico*, hydrolysis of *Gallus-gallus* lysozyme. We included a reduction/alkylation procedure using dithiothreitol (DTT) and iodoacetamide (IAA) prior to nanopore analysis, to prevent disulfide bond formation interfering with the protein hydrolysis. The peptide signals were measured in 1 M KCl at pH 3.8 under an applied potential of -70 mV and recorded at 50 kHz filtered to 10 kHz using an analog Bessel-filter, and further processed using a digital Gaussian filter at 5 kHz (**Figure 2B**). Under these conditions the peptides are expected to have one positive charge as the acidic residues should be mainly protonated at low pH. Notably, this is a recurring feature for trypsinated substrates, because trypsin cleaves preferentially after a lysine or arginine residue, thus most peptides will have a positive charge next to the C-terminus of the peptide, yielding an overall net charge of +1. Under these conditions, numerous peptide translocation events were observed, each with a specific current blockade (I_B) (**Figure 2B**). For each blockade, the percentage excluded current ($I_{ex}\%$) was calculated from the decrease in current observed during a blockade (ΔI_B) relative to the observed current of the open pore (I_o , **Figure 1**). We show the dwell time and excluded current as well as a histogram of the excluded current (excluded current spectrum, or $I_{ex}\%$ spectrum) of all seven synthetic peptides added cumulatively in equimolar concentrations (**Figure 2B**).

Figure 2 Lysozyme fingerprinting using FraC-G13F-T1 nanopores. **A.** Peptides expected from lysozyme's tryptic digest including their mass. The additional mass of 57.02 Da is expected for alkylated peptides at a cysteine position. **B.** Ionic current versus time (top), $I_{ex}(\%)$ versus dwell time (middle) and probability density (bottom) versus $I_{ex}(\%)$ as obtained from the measurement of an equimolar mixture of the peptides expected from *Gallus-gallus* lysozyme. The black lines in the top panel indicate the fitted event. **C.** Mass of model peptides from tryptic *Gallus-gallus* lysozyme (grey spheres) set against the measured excluded current (%), based on 3 individual measurements. The black dotted line represents a polynomial fit through the data. **D.** Excluded current spectrum density from the tryptic digest of *Gallus-gallus* lysozyme. The inset shows the dwell time versus $I_{ex}(\%)$ spectrum. **E.** Constructed $I_{ex}\%$ spectrum density from the same tryptic digest of *Gallus-gallus* lysozyme analysed by ESI-MS. Each peak indicates a peptide identified by ESI-MS, with $I_{ex}\%$ calculated using the calibrated exclusion current % in B. The height of the peak reflects the relative abundance of peptides measured by ESI-MS. The spread of the peak indicates an arbitrary standard deviation of 0.5 $I_{ex}\%$. All nanopore measurements were performed in 1 M KCl buffered to pH 3.8 using 50 mM citric acid titrated with bis-tris-propane under an applied potential of -70 mV. Recording was performed at 50 kHz using an analog Bessel-filter at 10 kHz and a digital Gaussian filter of 5 kHz.

The signals corresponding to the individual peptides were further confirmed by assessing the peptides individually. Work with alpha-hemolysin, aerolysin and FraC nanopores revealed that the relationship between the $I_{ex}\%$ and mass of the analyte might be complex. Although the electrical signal relates primarily on the volume excluded by the analytes,^{28,34} other factors such as the structure of the peptide, or the interactions of the peptides with the pore surface, or with the electrolyte and other buffer elements might play a role.^{16,19,28,29,35–41} In a prior contribution, utilizing a wild type FraC nanopore, we described the relationship between the blockade and the mass of the peptides as a second order polynomial.²⁸ Here we also found that a second order polynomial allowed a reasonable fit (**Figure 2C**), although an alternative fitting is possible (**Supplementary Figure 1**).³⁷ A better relationship will probably be obtained once the volume rather than the mass of the peptide will be accounted for.

3.2.2. Detection of Tryptic Digested Lysozyme

Next, we performed the experimental tryptic digest of lysozyme from *Gallus gallus*, and constructed an $I_{ex}\%$ spectrum. The sample was alkylated and reduced using dithiothreitol / iodoacetamide, and digested using mass spectrometry-grade trypsin (**Figure 2D**). We noticed that the smaller peptides, TPGSR (1), C(+57.02)ELAAMK (2), and HGLDNYR (3) were not clearly observed in the tryptic digest when measured using a nanopore. Electrospray ionization mass spectrometry (ESI-MS) confirmed that (1) and (2) were not present in the sample and we could therefore assume that this is a result of incomplete cleavage. Peptide (3) is detected less efficiently by the nanopore system compared to other peptides. Possibly, this is caused by the increased charge density caused by the presence of a histidine residue in the peptide sequence, which brings an additional positive charge (at pH 3.8) compared to other trypsinated peptides. In turn, for small peptides this might reduce the retention inside the nanopore by a stronger electrophoretic force.

In order to compare mass spectrometry and nanopore analysis, we converted the ESI-MS measurement into an expected $I_{ex,MS}\%$ spectrum using an *ad-hoc* algorithm (**Figure 2E**). Each peptide detected by ESI-MS was converted into a peak in the $I_{ex,MS}\%$ spectrum, of which the specific position on the x-axis was extrapolated from the $I_{ex}\%$ calibration curve from the individual lysozyme peptides (**Figure 2C**). The spread (width) of the peak was introduced using an arbitrary standard deviation (σ) of 0.5 $I_{ex}\%$, while the height of the peak was derived from the relative peptide abundance from the ESI-MS measurement. Notably, we observed a good correlation between the nanopore measurements and ESI-MS-based mapping (**Figure 2E**). As in the case of ESI-MS, it is expected that both positively charged and hydrophobic peptides are better captured by the nanopore than peptides lacking

chemical functionalities. This is because the electrophoretic migration is augmented in positively charged peptides, while the retention of hydrophobic peptides is increased by the interaction with the hydrophobic inner surface of the G13F-FraC-T1 nanopore.

3.2.3. Protein Profiling using G13F-FraC-T1

To further test the approach of protein fingerprinting using nanopores, we selected nine additional proteins with a molecular weight between 12.4 and 66.5 kDa: cytochrome C (12.4 kDa), elongation factor P (EF-P, 21.0 kDa),⁴² dihydrofolate reductase (DHFR, 19.1 kDa),⁴³ alpha casein (24.5 kDa), beta casein (25.1 kDa), bovine trypsin (trypsin, 23.3 kDa), C-terminal part of the high molecular weight adhesin 1 (HMW1ct, 34.6 kDa),⁴⁴ proteasome-activating nucleotidase (PAN, 49.6 kDa), and bovine serum albumin (BSA, 66.5 kDa). For each protein, we prepared a tryptic digest and subjected the resulting peptide mixture to measurement with the FraC nanopore. From these experiments, we constructed an $I_{ex}\%$ spectrum (**Figure 3A, Supplementary Figures 2, 3, and 4**) and related to the expected $I_{ex,MS}\%$ built from ESI-MS (**Supplementary Figure 5**) as well as the *in-silico* predicted peptide masses (**Supplementary Figure 6**). Rewardingly, we find that the reproducibility of the spectra is very high (**Supplementary Figure 7**). Comparison between the $I_{ex}\%$ spectrum and the $I_{ex,MS}\%$ revealed that the predicted $I_{ex,MS}\%$ of BSA, DHFR, EF-P, PAN, trypsin, and lysozyme correlated well with the experimental nanopore data, while other proteins such as cytochrome C and alpha casein showed a less accurate prediction displaying more peaks than observed in the nanopore. Finally, beta-casein and HMW1ct, were poorly resolved in the nanopore. We noticed that not all the peaks overlapped between the $I_{ex}\%$ and $I_{ex,MS}\%$, most likely because the $I_{ex,MS}\%$ spectra are produced considering the mass of the peptides, while current blockades relate to the volume of the analyte. Hence, nanopore analysis might require further improvements, for example by using nanopores with higher resolution or by reducing the sample complexity using chromatography devices upstream of the nanopore measurements, as is already the case in LC-MS analysis. In order to test whether nanopores are capable of distinguishing each protein, we analysed the $I_{ex}\%$ spectra using a spectral matching algorithm (**see methods**). Interestingly, we observed that all nine proteins and lysozyme are correctly assigned under these conditions (**Figure 3B**), revealing that the nanopore approach can be used to identify proteins.

Figure 3. Protein identification using nanopore spectrometry. **A.** Baseline-corrected excluded current spectra of 9 tryptic digested proteins. Based on 3 individual measurements. **B.** Leave-one-out spectral matching of the baseline corrected excluded current spectra using Euclidean cosine cross-correlation. With on the Y-axis the protein that is being matched using leave-one-out matching with the proteins on the X-axis. The black boxes represent the matches with the highest score. All nanopore measurements were performed in 1 M KCl buffered to pH 3.8 using 50 mM citric acid titrated with bis-tris-propane under an applied potential of -70 mV. Recording was performed at 50 kHz using an analogue Bessel-filter at 10 kHz and a digital Gaussian filter of 5 kHz.

3.3. Discussion

Proteins are commonly identified using bottom-up mass spectrometry (MS), where proteins are enzymatically digested at specific sites and the resulting peptides are identified.⁴⁵ MS analysis has revolutionized our understanding of proteins and almost single handedly initiated the emerging field of proteomics.^{3–5} Although MS can now identify thousands of proteins in mixtures, some peptides escape MS detection and a relatively large amount of material is required. Therefore, addressing post translational modifications where only a few proteins are modified, heterogeneous proteins or low abundance proteins is challenging. Furthermore, mass spectrometers, are expensive and complicated devices that have high operational costs, which is why in protein analysis they are most often operated in centralized facilities.^{2,3,6} In this work we introduce a method for protein analysis based on nanopores. Unlike MS, the electrical output signal arising from a nanopore can be easily interfaced with low-cost and portable devices. Nanopores have additional advantages as they detect single-molecules and they are amenable to incorporation into arrays of thousands of nanopores for high throughput analysis.

However, despite nanopores having been applied successfully to sequence DNA, their use in protein identification has yet to be proven.

In this contribution, we describe a nanopore approach for the identification of proteins. Using FraC nanopores, we show that the tryptic digest of proteins generates clustered events from individual trypsinated peptides. Calibration using synthetic peptides corresponding the expected complete hydrolysis of lysozyme (mass between 500 and 1700 Da) revealed that the nanopore captured all peptides and that an approximate relationship between the signal and the size of the peptide can be established (**Figure 2**). Comparison between ten protein substrates showed a reasonably good correlation with MS analysis (**Figure 2** and **Supplementary Figure 5**), and revealed that proteins can be recognized by spectral matching, which is similar to peptide mass fingerprinting (PMF), a technique used in early protein analysis. Although the proteins tested here are very different from each other, previous work showed that nanopore currents can detect differences in one single amino acid^{19,37,38,41} or single post-translational modifications.^{29,46} At present, we observe misassignments between peptide masses, which complicates the identification of unknown peptides and makes the recognition of unknown proteins challenging. However, nanopore currents report the volume of a peptide rather than its mass. Hence, we expect large improvements once calibration curves that take into consideration the exact volume of the peptide in solution will be used. We also found that peptide abundance measured by the nanopore matched relatively well that predicted by MS. Further research on the relationship between the capture frequency and the physico-chemical properties of the peptides will reveal the strength and limitation of nanopore analysis for quantification of peptides.

At the moment, the resolution of nanopore analysis does not yet match that of mass spectrometry. However, improvements can be envisaged. For example, upstream separation devices could be used, the interaction between the nanopore and the analytes can be improved, or nanopores with different sizes, shapes and physical properties can be used. Nonetheless, nanopores have distinguishable advantages: they can be integrated into inexpensive and portable devices,^{1,3} they work in solution and they can be directly interfaced with other analytical devices such as liquid chromatography or spectroscopy devices. Further, nanopores measure a fundamentally different property of peptides compared to MS (volume rather than mass/charge). Hence, nanopores might be used alongside MS for the identification and/or quantification of analytes (*e.g.* isobaric peptides) that are not easily studied by MS. Furthermore, arrays of thousands of nanopores are currently under development in commercial devices, suggesting that nanopore arrays have the potential to allow high-throughput protein analysis. Finally, nanopores are single-molecule sensors. Therefore, if a protease will be coupled to a nanopore directly,³³ this approach might be used for fingerprinting-based identification

of single molecules. The latter might find application in the characterization of low abundance proteins or the heterogeneity of protein samples.

3.4. Methods

Chemicals. Potassium chloride, sodium chloride, urea, imidazole, *N,N*-dimethyldodecylamine *N*-oxide (LDAO), *n*-dodecyl β -D-maltoside (DDM), and lysozyme were received from Carl-Roth. 2-Amino-2-(hydroxymethyl)-1,3-propanediol (Tris) was obtained from Roche; citric acid, bovine serum albumin (BSA), and *n*-hexadecane were purchased from Acros. Pentane, magnesium chloride, dithiothreitol, iodoacetamide, trypsin, beta-casein, cytochrome C, alpha-casein, sheep blood and peptides were purchased from Sigma-Aldrich (Merck), ethanol from (Boomlab). 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC), and sphingomyelin were received from Avanti Polar Lipids. Ni-NTA beads were obtained from Qiagen.

Protein digestion. 100 μ g of protein stock was taken and the volume was adjusted to 50 μ l using 20 mM Tris.HCl buffer (pH 7.5). A final concentration of 20 mM dithiothreitol (DTT) was added to reduce any disulphide bonds. The sample was incubated at 37 °C for 15 minutes followed by a denaturing step at 95 °C for 15 minutes. Afterwards, a 20 mM iodoacetamide (IAA) was added and the sample was left to incubate for 15 minutes at room temperature in the dark in order to alkylate the reduced cysteine residues. Finally, the total volume was adjusted to 100 μ l using 100 mM Tris Buffer (pH 8.5).

Tryptic digestion was performed using a kit purchased from Sigma-Aldrich, containing proteomics grade trypsin singles. 50 μ l of sample (containing 50 μ g of protein) was added to 1 μ g of mass-spec grade trypsin (1:50 enzyme:protein ratio) and the sample was subsequently incubated overnight at 37°C. Since large (> 2000 Da) peptides could clog the nanopore, the protein solution was passed through a centrifugal filter with a molecular weight cut-off of 3000 Da (Amicon). Filtered samples were stored in -20°C prior to use.

Fragaceatoxin C monomer purification. Fragaceatoxin C nanopores were expressed and purified as described previously.²⁵ In brief, a pT7-SC1 plasmid containing the G13F-*FraC* gene (Supplementary Table 1), was electrochemically transformed into BL21(DE3) competent *Escherichia coli* cells and grown overnight at 37°C on LB agar plates supplemented with 100 mg/L ampicillin and 1% glucose. On the next day, grown LB plates were solubilized into 200 mL 2xYT medium, supplemented with 100 mg/L ampicillin. Cultures were grown under constant shaking at 37 °C until an optical density (OD₆₀₀) of 0.6 was reached. Afterwards, 0.5 mM isopropyl β -D-1-thiogalactopyranoside was added for induction and growth continued overnight at 21°C. Bacterial cells were pelleted using centrifugation (6,000g, 20

minutes, 4 °C) and stored for at least one hour at -80 °C. 100 mL (original culture) cell pellets were resuspended using 20 ml lysis buffer, consisting of 150 mM NaCl buffered at pH 7.5 using 15 mM Tris base and supplemented with 1 mM MgCl₂, 2M Urea, 0.2 mg/mL lysozyme, 0.2 units/mL DNase and 20 mM imidazole. Solubilised pellets were vigorously shaken for 1 hour at room temperature. The lysate was disrupted fully by sonification for 60 seconds (duty cycle 30%, output control 3) using a Branson Sonifier 450. The lysate was centrifuged at 6000g for 20 minutes at 4 °C and the supernatant was transferred to a fresh 50 mL falcon tube. Subsequently, 200 µL pre-washed Ni-NTA beads (Qiagen, stored at 4 °C) was added and the falcon tube was incubated for 1 hour at room temperature (21 °C) under constant rotation. The supernatant, incubated with beads, was added to a pre-washed Micro Bio-Spin column (Bio-Rad) and, afterwards, washed extensively using a buffer consisting of 150 mM NaCl buffered at pH 7.5 using 15 mM Tris base supplemented with 20 mM imidazole. The column was dried by centrifugation (13,300g, 1 minute) in order to remove residual wash buffer. Finally, monomeric protein was eluted (13,300g, 2 minutes) after a 10 minute incubation (room temperature, 21 °C) with 150 µL of 150 mM NaCl buffered at pH 7.5 using 15 mM Tris base and supplemented with 300 mM imidazole.

Sphingomyelin-DPhPC liposomes preparation. An equal mixture of 25 mg 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC) and 25 mg sphingomyelin (Brain, Porcine) was dissolved in 4 mL pentane containing 0.5 v/v% ethanol. A film was formed on the side of a round bottom flask by application of heat under constant rotation, evaporating all solvents. The resulting film was dissolved in 10 mL of 150 mM NaCl, buffered at pH 7.5 using 15 mM Tris base. The resulting liposome solution (5 mg/mL) was frozen (-20 °C) and thawed multiple times.

Fragaceatoxin C oligomerisation. Freeze-thawed liposomes were added to purified monomers of fragaceatoxin C in a mass ratio of 10:1 (liposomes:protein). Liposomes were left to incubate for 30 minutes, at 37 °C, and later solubilised by the addition of 0.6 v/v% LDAO. Subsequently, the solution was diluted 20 times with 150 mM NaCl buffered at pH 7.5 using 15 mM Tris supplemented with 0.02% DDM. 200 µL washed regenerated Ni-NTA were added and incubated for 1 hour at room temperature (21 °C) under constant rotation.

Incubated Ni-NTA beads were transferred onto a pre-washed Micro Bio-Spin column (Bio-Rad) and washed extensively using a buffer consisting of 150 mM NaCl buffered at pH 7.5 using 15 mM Tris base supplemented with 20 mM imidazole and 0.02 v/v% DDM. The Micro Bio-Spin column was dried by centrifugation (13,300g, 1 minute) in order to remove residual wash buffer. Protein was eluted (13,300g, 2 minutes) after a 10 minute incubation (room temperature, 21 °C) with 150 µL of 150 mM NaCl buffered at pH 7.5 using 15 mM Tris base supplemented with 1 M imidazole and 0.02 v/v% DDM.

The oligomers were stored at -80 °C for long-term storage, thawed aliquots were kept at 4 °C and refreshed regularly.

Plasmid preparation. DHFR: A pT7-SC1 plasmid⁴⁷ containing the His₆-tagged *DHFR* gene (Supplementary Table 1), available from a previous study.¹² PAN: A synthetic gene containing His₆-tagged *PAN* (obtained from Integrated DNA Technologies, Supplementary Table 1) was ligated into a pT7-SC1 plasmid. HMW1ct: A pET45b plasmid harboring the *hmw1ct* gene (Supplementary Table 1) was constructed as described previously,⁴⁸ EF-P. A pBAD-His₆-SUMO plasmid harbouring the *efp* gene (from *Pseudomonas aeruginosa* PAO1, synthesized and cloned by GenScript, Supplementary Table 1).

Expression of dihydrofolate reductase (DHFR) and proteasome-activating nucleotidase (PAN).

Plasmids containing the gene of interest were electrochemically transformed into BL21(DE3) competent *Escherichia coli* cells. The cells were grown overnight at 37 °C on LB agar plates supplemented with 100 mg/L ampicillin and 1% glucose. On the next day, grown LB plates were solubilized into 200 mL 2xYT medium, supplemented with 100 mg/L ampicillin. Cultures were grown under constant shaking at 37 °C until an optical density (OD₆₀₀) of 0.6 was reached. Afterwards, 0.5 mM isopropyl β-D-1-thiogalactopyranoside was added for induction and growth was continued overnight at 21 °C. Cells were pelleted using centrifugation (6000g, 20 minutes, 4 °C) and stored for at -80 °C.

His₆-tag protein purification of DHFR and PAN. Cell pellets from 100 mL culture were resuspended using 20 mL lysis buffer, consisting of Sdex (150 mM NaCl, 15 mM Tris at pH 7.5) supplemented with 1 mM MgCl₂, 0.2 mg/mL lysozyme, 0.2 units/mL DNase and 20 mM imidazole. Solubilised pellets were vigorously shaken for 40 minutes at room temperature. The lysate was disrupted fully by sonification using a Branson Sonifier 450. The lysate was centrifuged at 6000g for 20 minutes at 4 °C and the supernatant was incubated with 200 μL pre-washed Ni-NTA beads (Qiagen) for 30 minutes at room temperature (21 °C) under constant rotation. The supernatant, incubated with beads, was added to a pre-washed Micro Bio-Spin column (Bio-Rad) and washed with Sdex supplemented with 20 mM imidazole. Finally, protein was eluted in steps of 150 μL with Sdex supplemented with 300 mM imidazole. The collected protein fractions were stored at -20°C until analysed.

Additional purification of PAN. Following His₆-tag protein purification, in order to increase the purity, an additional purification step was included. Four fractions (600 μl) with the highest protein concentration were combined and purified using the anion exchange chromatography purification described in these methods. The presence of PAN was demonstrated by SDS-PAGE and the fractions with highest protein concentration were combined and concentrated using a 10 kDa MWCO spin filter (Amicon).

Expression and purification of HMW1ct. A pET45b plasmid harboring the *hmw1ct* gene, constructed as described previously,⁴⁸ was used for expression of the HMW1ct protein (C-terminal fragment of *Haemophilus influenzae* high-molecular weight adhesin protein, residues 1205-1536). Cells were harvested by centrifugation (3220g), resuspended in ice-cold lysis buffer (50 mM HEPES, 100 mM NaCl, 10% glycerol, pH 7.5) and lysed by sonication (Branson Sonifier 450: 30% duty cycle, 2.5 minutes) in the presence of the protease inhibitors cocktail (Roche). Cell debris was removed by centrifugation and supernatant was used for Ni-affinity chromatography purification. Briefly, 3-4 mL of Ni-NTA resins (Qiagen) were applied on the gravity column, washed with water and equilibrated with the lysis buffer. Cell-free extract was then mixed with the resins for 1.5 h at 4 °C with gentle shaking. Afterwards, cell-free extract was allowed to flow through and resin-bound proteins were washed twice with washing buffer (50 mM HEPES, 300 mM NaCl, 5% glycerol, 15 mM imidazole, pH 7.5) and then eluted in three steps with elution buffer (50 mM HEPES, 300 mM NaCl, 5% glycerol, 400 mM imidazole, pH 7.5). Fractions containing protein of interest were collected and dialyzed using SnakeSkin dialysis system (MWCO 10kDa, Thermo Fischer Scientific) against storage buffer (50 mM HEPES, 100 mM NaCl, 10% glycerol, pH 7.5). After dialysis protein was aliquoted and stored at -80 °C until further use.

Expression and purification of EF-P. A pBAD-His₆-SUMO plasmid harbouring the *efp* gene was used to transform chemically competent *E. coli* TOP10 cells (standard heat-shock protocol) and plated on LB-agar plates containing ampicillin. A single colony was selected from the plate and used to prepare glycerol stock. To express the protein on large scale, a preculture (10 mL) in LB (100 µg/mL ampicillin) was prepared from the glycerol stock and grown at 37 °C with shaking (200 rpm) for 16-18 hours. The preculture was then used to inoculate 500 mL of Terrific Broth (TB) (100 µg/mL ampicillin) at 1:200 dilution ratio and incubated at 37 °C with shaking until OD₆₀₀ reached values of 0.6-0.7. Protein expression was induced by addition of 0.05% L-Ara (w/v, final concentration) and further incubation for 4 hours at 37 °C with shaking. Cells were harvested by centrifugation at 3220g for 15 minutes (Sorvall centrifuge, F-12 6x500 LEX fixed angle rotor, Thermo Scientific). The supernatant was discarded, and the cell pellet was resuspended in ice-cold lysis buffer (20 mM Tris, 500 mM NaCl, pH 8) in the presence of the protease inhibitor cocktail (Roche, complete, EDTA-free). Cells were lysed by sonication (Branson Sonifier 450, output control 30%, 2 minutes) and subsequently spun down at 6311g at 4 °C for 1 hour. For His₆-tag protein purification, the cell-free extract was incubated with Ni-NTA resin (Qiagen) for 1.5 hours at 4 °C with gentle shaking. The mixture was loaded on a gravity column and the lysate was allowed to flow through, followed by a washing step (twice) with washing buffer (20 mM Tris, 500 mM NaCl, 25 mM imidazole, pH 8). The protein of interest was eluted with elution buffer (20 mM Tris, 500 mM NaCl, 400 mM imidazole, pH 8) in three steps. Column fractions were analyzed by 12% SDS-PAGE analysis and the resulting gels were stained using Instant Blue protein

stain. Fractions containing protein of interest were pooled and desalted using midi PD-10 desalting columns (GE Healthcare). The target protein was routinely obtained in yields of 20-40 mg per 1 L of culture. To cleave the His₆-SUMO tag off, EF-P- His₆-SUMO was incubated with SUMO-protease (1 mg SUMO-protease for 12.5 mg of EF-P- His₆-SUMO) overnight at 4 °C. Next day, cleaved EF-P was purified with Ni-affinity chromatography. Briefly, reaction mixture was incubated with Ni-NTA resin for 1.5 hours at 4 °C with gentle shaking. The resulting suspension was allowed to pass through the gravity column and the flow through was collected. Next, the resin was washed with washing buffer 1 (20 mM Tris, 500 mM NaCl, 15 mM imidazole, pH 8), and washing buffer 2 (20 mM Tris, 500 mM NaCl, 30 mM imidazole, pH 8). This was followed by elution with elution buffer (20 mM Tris, 500 mM NaCl, 400 mM imidazole, pH 8). Analysis of the purification fractions with SDS-PAGE indicated that most of EF-P protein was present in the flow-through fraction. This fraction was concentrated and stored at -80 °C until further use.

Protein purification of bovine serum albumin (BSA). 10 mg lyophilised BSA (Arcos Organics) was dissolved in 1 mL of buffer A (50 mM Tris, pH 7.5) and purified using the anion exchange chromatography purification described in these methods. Presence of BSA was demonstrated by SDS-PAGE and the fractions with highest protein concentration were combined and concentrated using a 10 kDa MWCO spin filter (Amicon).

Anion exchange chromatography purification. BSA and PAN were selected for subsequent purification using ÄKTA pure chromatography (GE Healthcare Life Sciences), equipped with an HiTrap Q HP anion exchange column (GE Healthcare Life Sciences). Samples were loaded onto the column using a flow of 1 mL/minute buffer A (50 mM Tris, pH 7.5). Proteins were eluted using a flow of 1 mL/minute with a 40% gradient of buffer B (1 M NaCl, 50 mM Tris, pH 7.5) in 40 minutes.

Planar lipid bilayer electrophysiological recordings. A 25 µm thick Teflon membrane (Goodfellow Cambridge Ltd.) containing an aperture with a diameter of 100 µm was used to separate two compartments of a flow cell. 5 µL of a solution of 5% hexane in pentane (v/v) was applied near the aperture of the Teflon membrane. After an evaporation period of 1 minute, 400 µL buffer, consisting of 1 M KCl buffered at pH 3.8 using 50 mM citric acid with bis-tris-propane, was added to both compartments of the chamber. Afterwards, 20 µL of a 6.25 mg/mL solution of DPhPC in pentane was added on top of each compartment. The pentane was left to evaporate for approximately 2 minutes before mixing. A silver/silver chloride electrode was attached to each compartment. The Langmuir-Blodgett method, as described by Maglia *et al.*,²² was used to create planar lipid bilayers. Peptides and protein digests were always added to the *cis* compartment of the chamber.

ESI-MS experiments. For each trypsin-digested protein, approximately 10 μg is taken for mass-spectrometry analysis. The sample is analyzed with an LC system, EASY-nLC II (Thermo Scientific), connected to an LTQ Orbitrap XL (Thermo Scientific) using electron spray ionization (ESI). Peptides were separated in a reverse phase over an in-house packed C18 (ReproSil-Pur C18-AQ, 3 μm resin, Dr. Maisch) nano LC column (75 μm I.D., 15 cm, New Objective) under a gradient of solvent A: 2% acetonitrile, 0.1% Formic acid and solvent B: 0.1% formic acid in acetonitrile with 5% B to 28% B for 60 minutes, followed by 28% to 40% B for 10 minutes, 40% to 50% B for 2 minutes and 50% isocratic for 18 minutes at a flow of 200 nL/min and a column oven temperature of 60°C. Prior MS analysis, the sample was cleaned using C18 tips (Pierce) to remove salts and other contaminants. From the mass-spectrometry measurements we obtain the peak area of the detected peptides, which we use to generate an $I_{\text{ex}}\%$ spectrum.

Data recording. High impedance ionic current recordings were obtained using an Axopatch 200B amplifier combined with a Digidata 1440a or Digidata 1550B A/D converter (Molecular Devices), similar to preceding work.²² Data was recorded using Clampex 10 (Molecular Devices) at a sampling frequency of 50 kHz using an analogue Bessel filter of 10 kHz, unless stated otherwise.

Data analysis

***In-silico* digestion of lysozyme by trypsin.** *In-silico* protein trypsination was performed from the single-letter code sequence by cleaving each arginine (R) and lysine (K) residues unless they were followed by proline (P) using (python) regular expression "(?:(?![KR])(?P)).)". Peptides with a mass lower than 500 Da and larger than 1700 Da were excluded, as they are not observed by the nanopore. For *Gallus-gallus* lysozyme, this results in peptides as shown in Figure S6.

Event and $I_{\text{ex}}\%$ extraction. Peptide translocation events were extracted from the data using a threshold-search algorithm and characterized using a generalized flat-top normal distribution function (gNDF).

We use a threshold function to determine all events at 3σ from the baseline, and subsequently fit a generalized flat-top normal distribution. This distribution is a good descriptor for the events we observe, as it resembles a spike-like profile when $\beta < 1$, a Gaussian profile when $\beta = 1$, and a flat-top shape when $\beta > 1$. We filter events where $\beta < 1$, as these resemble a spike and the residual current (height) cannot be accurately estimated. Therefore, the events that we use to construct the residual current spectrum are all Gaussian or flat-top shaped, resulting in less variance between spectra.

This approach shares similarities with other methods to characterize events, such as MOSAIC.^{49,50} In both cases, an idealized event are represented and a filter effect is applied. Advantageously, the gNDF

describes both the filter effect and flat-top profile in a closed form. This allows the discrimination based on the event shape, however, it is slower for event recognition than using the algorithms implemented in MOSAIC.

$$f(x) = \Delta I_B * \exp\left(-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)^\beta\right) + I_0 \text{ for } \beta > 0 \quad (1)$$

Where μ is the events centre in the time domain with variance σ^2 and ΔI_B is the current difference (pA) between the baseline (I_0) and the event maximum. The variable β describes the shape of the function.

We utilize the generalized flat-top normal distribution, which is also used for the estimation of the dwell time. We estimate the events based on the full-width at half maximum of this distribution. Therefore, the dwell time is not skewed in short events. We discard spike-like events with a $\beta < 1$, however, we would not be able to determine the dwell time of these events in either algorithm.

$$FWHM = 2\sigma\sqrt{2^\beta\sqrt{\ln 2}} \quad (2)$$

Where σ equals the square root of the variance (σ^2) and β describes the shape parameter.

For each event, the excluded current was calculated by dividing the median current difference between the blockade (I_B) and the median open pore current (I_0) by I_0 .

$$I_{ex}\% = \frac{I_B - I_0}{I_0} * 100\% = \frac{\Delta I_B}{I_0} * 100\% \quad (3)$$

Where I_B is the blocked pore current, I_0 represents the open pore current, and $I_{ex}\%$ is the excluded current.

$I_{ex}\%$ spectrum construction and re-alignment. For each event, the $I_{ex}\%$ was calculated and a histogram between 0 and 100 $I_{ex}\%$ was constructed with a bin-width of 1 $I_{ex}\%$. The resulting histogram is called the excluded current spectrum, or abbreviated as the $I_{ex}\%$ spectrum.

In order to correct for shifting in the residual current due to experimental fluctuations (*e.g.* slight differences in salt concentration, temperature and instrument offset), we performed spectral re-alignment. We selected one spectrum (per protein) as a reference to which we re-align. Then, we subtracted bin-by-bin each additional spectrum to the reference spectrum and collected the residual sum of squares (as “error”). The x-axis of each bin of the new spectrum was moved in steps of 0.05 $I_{ex}\%$ and the error recorded. After performing 100 step-wise additions (total +5 $I_{ex}\%$) and 100 step-wise subtractions (total -5 $I_{ex}\%$) a plot of $\Delta I_{ex}\%$ versus the error was obtained showing the spectral offset where the error is minimal (**Supplementary Figure 3**).

Data analysis. All the data was analysed using Python 3.7 and is contained within a Jupyter notebook as an *ad-hoc* script and is available from the corresponding authors upon reasonable request.

Mapping mass to excluded current. When the average $I_{ex}\%$ is plotted versus the peptide mass, we obtain a non-linear relationship, instead of a linear curve that would be expected if the excluded current would only depend on the mass of the peptide. It is important to notice that the observed current is reduced because of the steric exclusion of ions induced by the analyte. Steric exclusion correlates to the mass of a peptide due to the sum of partial volume,⁵¹ and assumes no 3-dimensional structure, nor an interaction with the nanopore. We reasoned that the correlation between the molecular weight and $I_{ex}\%$ of peptides can be represented as a second order polynomial (Equation 4).²⁸ We have shown this relationship in a previous contribution, and also include the origin in the fit. We used the poly1d function as implemented in the Numpy library to find the ideal fit around the data. We found the following parameters: $b_2 = -1.33 \cdot 10^{-5}$, $b_1 = 7.23 \cdot 10^{-2}$, and $b_0 = 3.28$.

$$I_{ex,MS}(m) = b_0 + b_1 * m + b_2 * m^2 \quad (4)$$

Where b_0 , b_1 , and b_2 are the exponential terms. The input variable m represents the mass of the peptide.

$I_{ex}\%$ spectrum density from MS analysis. The mass of each peptide was converted to $I_{ex,MS}\%$ using the fit of Figure 2C. Each value was then used as the center (μ) a peak. The spread of each peak was made using an arbitrary standard deviation (σ) of 0.5 $I_{ex}\%$. The amplitude (a) of the peaks was matched to the area observed from ESI-MS measurements. Throughout this contribution, we utilized a ESI-MS containing an Orbitrap detector. The intensity of each ion is therefore measured as the amplitude of the free induction decay, which has a square relationship with the number of ions detected. Therefore, we had to take the square root of the intensity resulting in equation 5.

$$g(x) = \sqrt{\sum_{i=0}^n a_i * \exp\left(-\frac{(x - I_{ex}(m_i))^2}{2\sigma^2}\right)} \quad (5)$$

Where a_i is the area resulting from ESI-MS. m_i is the mass of the peptide and is mapped to the excluded current using equation 4. σ is the peak width.

Spectral matching. We analyzed the $I_{ex}\%$ *spectra* using a spectral matching algorithm incorporating the squared first derivate Euclidean cosine correlation (DEuc) (Equation 6); which is advantageous, as it corrects for baseline sloping.⁵² The DEuc is a direct result of the dot product equation and allows the

estimation of the angle between two vectors, which can be used as a measure of similarity. We use the normalized counts of the excluded current spectra which are represented as vectors, *e.g.* $\vec{A} = [a_1, a_2, \dots, a_n]$ and $\vec{B} = [b_1, b_2, \dots, b_n]$. We estimate the derivative of $\vec{A}$ and $\vec{B}$ by numerical differentiation, resulting in $\Delta\vec{A}$ and $\Delta\vec{B}$. Usage of the derivative is advantageous as this is less sensitive to background noise, which is usually stochastically distributed.

$$DEuc = \cos(\theta) = \frac{\Delta\vec{A} \cdot \Delta\vec{B}}{\|\Delta\vec{A}\| \|\Delta\vec{B}\|} = \frac{(\sum_{i=0}^N \Delta a_i \Delta b_i)^2}{\sum_{i=0}^N \Delta a_i^2 \sum_{i=0}^N \Delta b_i^2} \quad (6)$$

We chose to only consider the $I_{ex}\%$ *spectrum* between 50 and 98 $I_{ex}\%$, as noise below the limit-of-detection and fully blocked events may have skewed the comparison. Subsequently, we performed a leave-one-out comparison using the DEuc as a score, normalised to 100% for visualisation. The leave-one-out comparison compares the $I_{ex}\%$ spectrum of each measurement (sample), with the average $I_{ex}\%$ spectrum of each protein (database). For each comparison, we constructed a database that contains all measurements except for the sample. The score represented in Figure 3B of the main text are the average scores over all samples for each protein.

Data and code availability. All data and corresponding analysis generated in this study have been deposited in the Zenodo database under <https://doi.org/10.5281/zenodo.5205565>.

3.5. References

- (1) Baaken, G.; Ankri, N.; Schuler, A.-K.; Rhe, J.; Behrends, J. C. Nanopore-Based Single-Molecule Mass Spectrometry on a Lipid Membrane Microarray. *ACS Nano* **2011**, 5 (10), 8080–8088. <https://doi.org/10.1021/nn202670z>.
- (2) Suhre, K.; McCarthy, M. I.; Schwenk, J. M. Genetics Meets Proteomics: Perspectives for Large Population-Based Studies. *Nat. Rev. Genet.* **2020**. <https://doi.org/10.1038/s41576-020-0268-2>.
- (3) Timp, W.; Timp, G. Beyond Mass Spectrometry, the next Step in Proteomics. *Sci. Adv.* **2020**, 6 (2), eaax8978. <https://doi.org/10.1126/sciadv.aax8978>.
- (4) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, 422 (6928), 198–207. <https://doi.org/10.1038/nature01511>.
- (5) Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, 537 (7620), 347–355. <https://doi.org/10.1038/nature19949>.
- (6) Armah, S.; Ferruzzi, M. G.; Gletsu-Miller, N. Feasibility of Mass-Spectrometry to Lower Cost and Blood Volume Requirements for Assessment of B Vitamins in Patients Undergoing Bariatric Surgery. *Metabolites* **2020**, 10 (6), 240. <https://doi.org/10.3390/metabo10060240>.
- (7) Soskine, M.; Biesemans, A.; Moeyaert, B.; Cheley, S.; Bayley, H.; Maglia, G. An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry.

- Nano Lett.* **2012**, *12* (9), 4895–4900. <https://doi.org/10.1021/nl3024438>.
- (8) Stefureac, R.; Waldner, L.; Howard, P.; Lee, J. S. Nanopore Analysis of a Small 86-Residue Protein. *Small* **2008**, *4* (1), 59–63. <https://doi.org/10.1002/smll.200700402>.
 - (9) Si, W.; Aksimentiev, A. Nanopore Sensing of Protein Folding. *ACS Nano* **2017**, *11* (7), 7091–7100. <https://doi.org/10.1021/acsnano.7b02718>.
 - (10) Nivala, J.; Mulrone, L.; Li, G.; Schreiber, J.; Akeson, M. Discrimination among Protein Variants Using an Unfoldase-Coupled Nanopore. *ACS Nano* **2014**, *8* (12), 12365–12375. <https://doi.org/10.1021/nn5049987>.
 - (11) Zernia, S.; van der Heide, N. J.; Galenkamp, N. S.; Gouridis, G.; Maglia, G. Current Blockades of Proteins inside Nanopores for Real-Time Metabolome Analysis. *ACS Nano* **2020**, *14* (2), 2296–2307. <https://doi.org/10.1021/acsnano.9b09434>.
 - (12) Willems, K.; Ruić, D.; Biesemans, A.; Galenkamp, N. S. N.; Van Dorpe, P.; Maglia, G. Engineering and Modeling the Electrophoretic Trapping of a Single Protein Inside a Nanopore. *ACS Nano* **2019**, *13* (9), 9980–9992. <https://doi.org/10.1021/acsnano.8b09137>.
 - (13) Cao, C.; Ying, Y.-L.; Hu, Z.-L.; Liao, D.-F.; Tian, H.; Long, Y.-T. Discrimination of Oligonucleotides of Different Lengths with a Wild-Type Aerolysin Nanopore. *Nat. Nanotechnol.* **2016**, *11* (8), 713–718. <https://doi.org/10.1038/nnano.2016.66>.
 - (14) Baaken, G.; Halimeh, I.; Bacri, L.; Pelta, J.; Oukhaled, A.; Behrends, J. C. High-Resolution Size-Discrimination of Single Nonionic Synthetic Polymers with a Highly Charged Biological Nanopore. *ACS Nano* **2015**, *9* (6), 6443–6449. <https://doi.org/10.1021/acsnano.5b02096>.
 - (15) Lucas, F. L. R.; Sarthak, K.; Lenting, E. M.; Coltan, D.; van der Heide, N. J.; Versloot, R. C. A.; Aksimentiev, A.; Maglia, G. The Manipulation of the Internal Hydrophobicity of FraC Nanopores Augments Peptide Capture and Recognition. *ACS Nano* **2021**. <https://doi.org/10.1021/acsnano.0c09958>.
 - (16) Sutherland, T. C.; Long, Y.-T.; Stefureac, R.-I.; Bediako-Amoa, I.; Kraatz, H.-B.; Lee, J. S. Structure of Peptides Investigated by Nanopore Analysis. *Nano Lett.* **2004**, *4* (7), 1273–1277. <https://doi.org/10.1021/nl049413e>.
 - (17) Robertson, J. W. F.; Reiner, J. E. The Utility of Nanopore Technology for Protein and Peptide Sensing. *Proteomics* **2018**, *18*, 1–36. <https://doi.org/10.1002/pmic.201800026>.
 - (18) Robertson, J. W. F.; Rodrigues, C. G.; Stanford, V. M.; Robinson, K. A.; Krasilnikov, O. V.; Kasianowicz, J. J. Single-Molecule Mass Spectrometry in Solution Using a Solitary Nanopore. *Proc. Natl. Acad. Sci.* **2007**, *104* (20), 8207–8211. <https://doi.org/10.1073/pnas.0611085104>.
 - (19) Restrepo-Pérez, L.; Huang, G.; Bohländer, P. R.; Worp, N.; Eelkema, R.; Maglia, G.; Joo, C.; Dekker, C. Resolving Chemical Modifications to a Single Amino Acid within a Peptide Using a Biological Nanopore. *ACS Nano* **2019**, *13* (12), 13668–13676. <https://doi.org/10.1021/acsnano.9b05156>.
 - (20) Reiner, J. E.; Kasianowicz, J. J.; Nablo, B. J.; Robertson, J. W. Theory for Polymer Analysis Using Nanopore-Based Single-Molecule Mass Spectrometry. *Proc Natl Acad Sci U S A* **2010**, *107* (27), 12080–12085. <https://doi.org/10.1073/pnas.1002194107>.
 - (21) Li, Z.; He, B.; Feng, W. Evaluation of Bottom-up and Top-down Mass Spectrum Identifications with Different Customized Protein Sequences Databases. *Bioinformatics* **2020**, *36* (4), 1030–1036. <https://doi.org/10.1093/bioinformatics/btz733>.
 - (22) 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* (1), 935. <https://doi.org/10.1038/s41467-017-01006-4>.

- (23) Gu, L. Q.; Braha, O.; Conlan, S.; Cheley, S.; Bayley, H. Stochastic Sensing of Organic Analytes by a Pore-Forming Protein Containing a Molecular Adapter. *Nature* **1999**, 398 (6729), 686–690. <https://doi.org/10.1038/19491>.
- (24) Mohammad, M. M. M.; Movileanu, L. Excursion of a Single Polypeptide into a Protein Pore: Simple Physics, but Complicated Biology. *Eur Biophys J* **2008**, 37 (6), 913–925. <https://doi.org/10.1007/s00249-008-0309-9>.
- (25) Mutter, N. L.; Huang, G.; van der Heide, N. J.; Lucas, F. L. R.; Galenkamp, N. S.; Maglia, G.; Wloka, C. Preparation of Fragaceatoxin C (FraC) Nanopores; 2021; pp 3–10. https://doi.org/10.1007/978-1-0716-0806-7_1.
- (26) 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 (10), 1091–1095. <https://doi.org/10.1038/80295>.
- (27) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proc. Natl. Acad. Sci.* **1996**, 93 (24), 13770–13773. <https://doi.org/10.1073/pnas.93.24.13770>.
- (28) Huang, G.; Voet, A.; Maglia, G. FraC Nanopores with Adjustable Diameter Identify the Mass of Opposite-Charge Peptides with 44 Dalton Resolution. *Nat. Commun.* **2019**, 10 (1), 835. <https://doi.org/10.1038/s41467-019-08761-6>.
- (29) Restrepo-Pérez, L.; Wong, C. H.; Maglia, G.; Dekker, C.; Joo, C. Label-Free Detection of Post-Translational Modifications with a Nanopore. *Nano Lett.* **2019**, 19 (11), 7957–7964. <https://doi.org/10.1021/acs.nanolett.9b03134>.
- (30) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. Protein Identification by Mass Profile Fingerprinting. *Biochem. Biophys. Res. Commun.* **1993**, 195 (1), 58–64. <https://doi.org/10.1006/bbrc.1993.2009>.
- (31) Pappin, D. J. C.; Hojrup, P.; Bleasby, A. J. Rapid Identification of Proteins by Peptide-Mass Fingerprinting. *Curr. Biol.* **1993**, 3 (6), 327–332. [https://doi.org/10.1016/0960-9822\(93\)90195-T](https://doi.org/10.1016/0960-9822(93)90195-T).
- (32) Cottrell, J. S.; Sutton, C. W. The Identification of Electrophoretically Separated Proteins by Peptide Mass Fingerprinting. In *Protein and Peptide Analysis by Mass Spectrometry*; Humana Press: New Jersey; pp 67–82. <https://doi.org/10.1385/0-89603-345-7:67>.
- (33) Zhang, S.; Huang, G.; Versloot, R.; Herwig, B. M.; de Souza, P. C. T.; Marrink, S.-J.; Maglia, G. Bottom-up Fabrication of a Multi-Component Nanopore Sensor That Unfolds, Processes and Recognizes Single Proteins. *bioRxiv* **2020**, 2020.12.04.411884. <https://doi.org/10.1101/2020.12.04.411884>.
- (34) Chinappi, M.; Yamaji, M.; Kawano, R.; Cecconi, F. Analytical Model for Particle Capture in Nanopores Elucidates Competition among Electrophoresis, Electroosmosis, and Dielectrophoresis. *ACS Nano* **2020**, 14 (11), 15816–15828. <https://doi.org/10.1021/acsnano.0c06981>.
- (35) Li, S.; Cao, C.; Yang, J.; Long, Y.-T. Detection of Peptides with Different Charges and Lengths by Using the Aerolysin Nanopore. *ChemElectroChem* **2019**, 6 (1), 126–129. <https://doi.org/10.1002/celec.201800288>.
- (36) Angevine, C. E.; Robertson, J. W. F.; Dass, A.; Reiner, J. E. Laser-Based Temperature Control to Study the Roles of Entropy and Enthalpy in Polymer-Nanopore Interactions. *Sci. Adv.* **2021**, 7 (17), eabf5462. <https://doi.org/10.1126/sciadv.abf5462>.
- (37) Chavis, A. E.; Brady, K. T.; Hatmaker, G. A.; Angevine, C. E.; Kothalawala, N.; Dass, A.; Robertson,

- J. W. F.; Reiner, J. E. Single Molecule Nanopore Spectrometry for Peptide Detection. *ACS Sensors* **2017**, 2 (9), 1319–1328. <https://doi.org/10.1021/acssensors.7b00362>.
- (38) Asandei, A.; Rossini, A. E.; Chinappi, M.; Park, Y.; Luchian, T. Protein Nanopore-Based Discrimination between Selected Neutral Amino Acids from Polypeptides. *Langmuir* **2017**, 33 (50), 14451–14459. <https://doi.org/10.1021/acs.langmuir.7b03163>.
- (39) Wolfe, A. J.; Mohammad, M. M.; Cheley, S.; Bayley, H.; Movileanu, L. Catalyzing the Translocation of Polypeptides through Attractive Interactions. *J. Am. Chem. Soc.* **2007**, 129 (45), 14034–14041. <https://doi.org/10.1021/Ja0749340>.
- (40) Cao, C.; Cirauqui, N.; Marcaida, M. J.; Buglakova, E.; Duperrex, A.; Radenovic, A.; Dal Peraro, M. Single-Molecule Sensing of Peptides and Nucleic Acids by Engineered Aerolysin Nanopores. *Nat. Commun.* **2019**, 10 (1), 4918. <https://doi.org/10.1038/s41467-019-12690-9>.
- (41) Ouldali, H.; Sarthak, K.; Ensslen, T.; Piguet, F.; Manivet, P.; Pelta, J.; Behrends, J. C.; Aksimentiev, A.; Oukhaled, A. Electrical Recognition of the Twenty Proteinogenic Amino Acids Using an Aerolysin Nanopore. *Nat. Biotechnol.* **2020**, 38 (2), 176–181. <https://doi.org/10.1038/s41587-019-0345-2>.
- (42) Yakovlieva, L.; Wood, T. M.; Kemmink, J.; Kotsogianni, I.; Koller, F.; Lassak, J.; Martin, N. I.; Walvoort, M. T. C. A β -Hairpin Epitope as Novel Structural Requirement for Protein Arginine Rhamnosylation. *Chem. Sci.* **2021**. <https://doi.org/10.1039/D0SC05823H>.
- (43) Galenkamp, N. S.; Biesemans, A.; Maglia, G. Directional Conformer Exchange in Dihydrofolate Reductase Revealed by Single-Molecule Nanopore Recordings. *Nat. Chem.* **2020**, 12 (5), 481–488. <https://doi.org/10.1038/s41557-020-0437-0>.
- (44) Walvoort, M. T. C. C.; Testa, C.; Eilam, R.; Aharoni, R.; Nuti, F.; Rossi, G.; Real-Fernandez, F.; Lanzillo, R.; Brescia Morra, V.; Lolli, F.; Rovero, P.; Imperiali, B.; Papini, A. M. Antibodies from Multiple Sclerosis Patients Preferentially Recognize Hyperglucosylated Adhesin of Non-Typeable *Haemophilus Influenzae*. *Sci. Rep.* **2016**, 6 (1), 39430. <https://doi.org/10.1038/srep39430>.
- (45) Sinha, A.; Mann, M. A Beginner's Guide to Mass Spectrometry–Based Proteomics. *Biochem. (Lond.)* **2020**, 42 (5), 64–69. <https://doi.org/10.1042/BIO20200057>.
- (46) Rosen, C. B.; Rodriguez-Larrea, D.; Bayley, H. Single-Molecule Site-Specific Detection of Protein Phosphorylation with a Nanopore. *Nat. Biotechnol.* **2014**, 32 (2), 179–181. <https://doi.org/10.1038/nbt.2799>.
- (47) Cheley, S.; Malghani, M. S.; Song, L.; Hobaugh, M.; Gouaux, J. E.; Yang, J.; Bayley, H. Spontaneous Oligomerization of a Staphylococcal Alpha-Hemolysin Conformationally Constrained by Removal of Residues That Form the Transmembrane Beta-Barrel. *Protein Eng. Des. Sel.* **1997**, 10 (12), 1433–1443. <https://doi.org/10.1093/protein/10.12.1433>.
- (48) Walvoort, M. T. C.; Testa, C.; Eilam, R.; Aharoni, R.; Nuti, F.; Rossi, G.; Real-Fernandez, F.; Lanzillo, R.; Brescia Morra, V.; Lolli, F.; Rovero, P.; Imperiali, B.; Papini, A. M. Erratum: Antibodies from Multiple Sclerosis Patients Preferentially Recognize Hyperglucosylated Adhesin of Non-Typeable *Haemophilus Influenzae*. *Sci. Rep.* **2017**, 7 (1), 44969. <https://doi.org/10.1038/srep44969>.
- (49) Forstater, J. H.; Briggs, K.; Robertson, J. W. F.; Etedgui, J.; Marie-Rose, O.; Vaz, C.; Kasianowicz, J. J.; Tabard-Cossa, V.; Balijepalli, A. MOSAIC: A Modular Single-Molecule Analysis Interface for Decoding Multistate Nanopore Data. *Anal. Chem.* **2016**, 88 (23), 11900–11907. <https://doi.org/10.1021/acs.analchem.6b03725>.
- (50) Balijepalli, A.; Etedgui, J.; Cornio, A. T.; Robertson, J. W. F.; Cheung, K. P.; Kasianowicz, J. J.; Vaz, C. Quantifying Short-Lived Events in Multistate Ionic Current Measurements. *ACS Nano* **2014**, 8

(2), 1547–1553. <https://doi.org/10.1021/nn405761y>.

- (51) Harpaz, Y.; Gerstein, M.; Chothia, C. Volume Changes on Protein Folding. *Structure* **1994**, 2 (7), 641–649. [https://doi.org/10.1016/S0969-2126\(00\)00065-4](https://doi.org/10.1016/S0969-2126(00)00065-4).
- (52) Li, J.; Hibbert, D. B.; Fuller, S.; Vaughn, G. A Comparative Study of Point-to-Point Algorithms for Matching Spectra. *Chemom. Intell. Lab. Syst.* **2006**, 82 (1–2), 50–58. <https://doi.org/10.1016/j.chemolab.2005.05.015>.

3.6. Supporting Information

Supplementary Figure 1. Mapping of molecular weight to excluded current. A. Mass of model peptides from tryptic lysozyme (grey) set against the measured excluded current (%), based on 3 individual measurements. The dashed line represents a fit through the data based on the Flory exponent as described by Chavis et al. 2017:

$$I_{ex} \% = \left(1 - \left(1 - \left(\frac{M_p}{M_f} \right)^\nu \left(1 - a \left(\frac{M_p}{M_f} \right)^\gamma \left(1 - \left(\frac{M_p}{M_f} \right)^{1-\nu} \right)^{-1} \right) \right)^{-1} \right) * 100\% , \text{ with model parameters: } M_f =$$

1825.85 Da, $a = 0.98$, $\nu = 0.27$, $\gamma = 0.0$, where M_f is the mass of the smallest polymer that fills the pore volume completely, ν is the Flory exponent, a and γ are the free parameters describing ionic binding to the polymer.¹ We excluded the datapoint at 516 Da because such model was developed for polymers and it is expected to perform poorly for small peptides. Furthermore, the last data point has a large leverage on the data. In addition, it should be noticed that the model mentioned before was developed for cylindrical nanopores, rather than for conical nanopore, such as FraC. The dashed line stops at the values that are accessible experimentally. **B.** Mass of model peptides from tryptic lysozyme (grey) set against the measured excluded current (%), based on 3 individual measurements. The dashed line represents an empirical second-order polynomial fit ($y = b_0 + b_1 * x + b_2 * x^2$) through the data with $b_2 = -1.33 * 10^{-5}$, $b_1 = 7.23 * 10^{-2}$, $b_0 = 3.28$.

Supplementary Figure 2. Excluded current spectra without re-alignment. Resulting excluded current spectrum of 9 tryptic digested proteins and a control (“tryptic digest” of water) prior to spectral re-alignment, normalized to a total area of 100%. The solid black line represents the mean of 3 individual measurements of the same sample. The red dashed lines represent the standard deviation of 3 individual measurements. All measurements were performed in 1 M KCl buffered to pH 3.8 using 50 mM citric acid titrated with bis-tris-propane under an applied potential of -70 mV. Recording was performed at 50 kHz using an analog Bessel-filter at 10 kHz and a digital Gaussian filter of 5 kHz.

Supplementary Figure 3. Excluded current re-alignment error. Re-alignment error of 9 tryptic digested proteins and a control (“tryptic digest” of water). Each (colored) line represents a spectral alignment, the color of each line indicates a separate measurement.

Supplementary Figure 4. Excluded current spectra with re-alignment. Resulting excluded current spectrum of 9 tryptic digested proteins and a control (“tryptic digest” of water) after spectral re-alignment, normalized to a total area of 100%. The solid black line represents the mean of 3 individual measurements of the same sample. The red dashed lines represent the standard deviation of 3 individual measurements. All measurements were performed in 1 M KCl buffered to pH 3.8 using 50 mM citric acid titrated with bis-tris-propane under an applied potential of -70 mV. Recording was performed at 50 kHz using an analog Bessel-filter at 10 kHz and a digital Gaussian filter of 5 kHz.

Supplementary Figure 5. Excluded current set against the dwell time, excluded current spectra and constructed excluded current spectra for tryptic digested proteins. Each row represents the data for a protein as marked with a title on the central column. The first column shows the excluded current set against the dwell time for each event. The second column shows the excluded current spectrum as measured by nanopore electrophysiology. The third column shows the predicted excluded current spectrum from mass spectrometry based results.

Supplementary Figure 6. Protein identification using nanopore spectrometry. Resulting baseline corrected excluded current spectra of 9 tryptic digested proteins. The vertical lines represent the peptides as predicted *in-silico*, mapped to the $I_{ex}\%$.

Supplementary Figure 7. Reproducibility of nanopore protein spectra. Each row presents three independent repeats for each proteolytic digest. The left panel shows the excluded current histograms with a normalized area of 100% and the right panel shows the excluded current set against the dwell time (in seconds) for all events. All measurements were performed in 1 M KCl buffered to pH 3.8 using 50 mM citric acid titrated with bis-tris-propane under an applied potential of -70 mV. Recording was performed at 50 kHz using an analog Bessel-filter at 10 kHz and a digital Gaussian filter of 5 kHz.

Fragaceatoxin C (from *Actinia fragacea*) mutant G13F

G13F-FraC

ATGGCGAGCGCCGATGTCGCGGGTGC GGTAATCGACGGTGC GTTCCTGGGCTTTGACGTA CTGAAAACCGT
GCTGGAGGCCCTGGGCAACGTTAAACGCAAAATTGCGGTAGGGATTGATAACGAATCGGGCAAGACCTGG
ACAGCGATGAATACCTATTTCCGTTCTGGTACGAGTGATATTGTGCTCCACATAAGGTGGCGCATGGCAAG
GCGCTGCTGTATAACGGTCAAAAAATCGCGGTCTGTGCGACCGGCGTAGTGGGTGTGATTGCCTATAG
TATGTCTGATGGGAACACACTGGCGGTACTGTTCTCCGTGCCGTACGATTATAATTGGTATAGCAATTGGTG
GAATGTGCGTGTCTACAAAGGCCAGAAGCGTGCCGATCAGCGCATGTACGAGGAGCTGTACTATCATCGCT
CGCCGTTTCGCGGCGACAACGTTGGCATTCCCGGGGCTTAGGTTATGGACTCAAAGTCGCGGCTTTATG
AATAGTTCGGGCCACGCAATCCTGGAGATTACGTTACCAAAGCAGGCTCTGCGCATCATCACCACCATCAC
TGA

Dihydrofolate reductase (from *Escherichia coli*)

DHFR

ATGGCTACCAACCACCACCACCGTTCCGGCTATGATTTCTCTGATTGCGGCACTGGCTGTCGATCGTGTT
ATTGGTATGGAAAACGCTATGCCGTGGAATCTGCCGGCTGATCTGGCGTGGTTTAAACGTAACACCCTGGA
CAAGCCGGTCATTATGGGCCGCCATACGTGGGAAAGCATCGGTCGTCCGCTGCCGGGTCGAAAAATATTA
TCCTGAGCAGCCAGCCGGGACCGATGACCGTGTGACGTGGGTTAAGAGCGTCGATGAAGCAATTGCGGC
GGCAGGCGACGTGCCGGAATTATGGTTATCGGCGGTGGCCGCGTTTATGAACAGTTCCTGCCGAAAGCCC
AAAAGCTGTACCTGACCATATCGATGCAGAAGTCGAAGGTGATACGCACTTCCGGACTATGAACCGGAT
GACTGGGAAAGTGTGTTCTCCGAATTCACGACGCCGACGCTCAGAACAGCCACTCATACTCATTCGAAATC
CTGGAACGCCGTTGATAA

proteasome-activating nucleotidase (from *Methanocaldococcus jannaschii*)

PAN

ATGGCTCATCATCACCATCACCACGGGGGCGTTTTCGAAGAATTTATCTCTACTGAGTTGAAGAAGGAAAAG
AAGGCATTCACAGAGGAGTTCAAAGAAGAGAAGGAGATTAACGACAATAGCAACTTGAAGAACGATCTGT
TGAAAGAAGAATTACAAGAGAAGGCTCGTATTGCAGAGCTGGAATCACGCATTCTTAAGCTGGAGTTAGAG
AAAAAAGAGTTAGAGCGCGAAAAATTTGCAGCTTGCAAAAGAAAACGAGATTTTACGTCGTGAACCTTGACCG
TATGCGTGTTCCGCCTTTAATTGTAGGAACAGTGTTGGACAAAGTCGGAGAACGTAAAGTTGTGGTTAAAT
CCTCCACCGGACCATCTTTCTTAGTTAACGTGTCACATTTTGTGAACCCGGATGATCTGGCCCCTGGCAAGCG
TGTATGCTTGAATCAACAGACTCTTACGGTTGTTGATGTATTACCGGAAAATAAGGATTACCGTGCCAAGGC
GATGGAGGTGGATGAACGTCCAAATGTGCGTTACGAGGACATCGGTGGATTAGAGAAGCAAATGCAGGAA
ATCCGTGAAGTAGTCGAACTGCCGTTAAAGCACCTGAGTTATTCGAAAAGTTCGGAATTGAACCACCTAA
AGGGATTTTATTGTATGGACCCCTGGGACAGGCAAAACCCTGCTGGCCAAGGCAGTTGCGACGGAAACTA
ACGCAACCTTTATTGCGTGGTCGGTTCAGAGCTTGTTAAGAAGTTCATCGGAGAAGGAGCGAGCTTAGTG
AAGGATATTTTTAAGCTGGCCAAGGAGAAGGCTCCCTCGATTATCTTTATCGATGAGATCGATGCGATCGCA
GCCAAGCGTACGGATGCATTAACGGGTGGGGATCGTGAAGTACAACGCACACTGATGCAGCTTCTTGCCGA
AATGGATGGGTTTCGATGCACGTGGAGACGTCAAGATCATTGGCGCAACAAACCGTCCAGATATCCTTGATC
CAGCGATCCTTCGCCAGGTGCTTCGACCGCATTATCGAGGTCCCGGCACCGGACGAAAAGGGCCGCCTT
GAGATTTTGA AAATTCATACACGCAAAATGAACTTGGCAGAGGATGTAACTTGGAAGAAATTGCTAAGAT
GACTGAGGGGTGCGTAGGTGCAGAACTGAAGGCTATTTGCACGGAGGCCGGAATGAATGCAATCCGTGAA
CTTCGTGATTATGTAACGATGGACGATTTTCGAAAGCAGTTGAGAAAATCATGGAGAAAAGAAGGTTAA
GGTAAAGAGCCAGCCATCTGGACGTACTGTATCGTTGATAA

high molecular weight adhesin 1 (from *Haemophilus influenzae*)

HMW1ct

ATGGCACATCACCACCACCATCACGTGTGGACCGCAAACCTCAGGCGCACTGACCACGCTGGCGGGTTCTAC
CATCAAAGGCACGGAATCGGTTACGACCTCATCCAATCGGGTGACATCGGCGGCACCATTAGTGCGGTA
CGGTCGAAGTGAAAGCGACCGAATCCCTGACCACGCAGAGCAACTCTAAAATCAAAGCGACCACGGGCGA
AGCCAATGTTACCAGCGCAACCGGCACGATTGGCGGCACCATCTCTGGTAACACCGTTAATGTCACGGCAA
ACGCTGGTGATCTGACCGTGGGCAACGGTGCAGAAATTAATGCTACCGAAGGTGCAGCAACGCTGACCACC
AGCAGCGGTAACTGACCACGGAAGCCAGTTCCTCATATTACCTCAGCAAAAGGCCAGGTCAACCTGTCTGGC
TCAAGACGGCTCAGTGGCGGGTTCGATCAACGCAGCTAATGTTACCCTGAATACCACGGGCACGCTGACCA
CGGTCAAAGGTAGTAACATTAATGCCACCTCCGGTACGCTGGTGATCAACGCAAAAGATGCTGAACTGAAT
GGCGCGGCCCTGGGTAACCACACCGTGGTTAATGCGACGAACGCCAATGGCAGTGGTTCCTGTATTGCGAC
CACGTCATCGCGTGTTAACATCACCGGCGACCTGATTACGATCAACGGTCTGAACATCATCAGCAAAAACGG
CATCAACACCGTTCTGCTGAAAGGTGTTAAAATCGATGTCAAATACATCCAGCCGGGCATCGCTTCTGTGGA
CGAAGTTATTGAAGCGAAACGCATCCTGGAAAAAGTGAAAGATCTGAGTGACGAAGAACGTGAAGCGCTG
GCCAACTGGGTGTCAGCGCCGTGCGCTTTATTGAACCGAACAATACCATCACGGTTGATACCCAAAATGA
ATTCGCAACCCGTCCGCTGAGCCGCATTGTCATCTCTGAAGGCCGTGCCTGCTTTAGTA ACTCTGACGGTGC
TACGGTTTGCCTGAACATCGCTGATAATGGTCGT

elongation factor P (from *Pseudomonas aeruginosa*)

EF-P

ATGAAAACCGCTCAAGAGTTCGCGCCGGCCAGGTTGCCAACATCAATGGCGCTCCCTGGGTCATCCAGAA
GGCCGAGTTCAACAAGTCCGGCCGTAACGCTGCCGTCGTCAAGATGAAGCTGAAGAACCTGCTGACCGGC
GCCGGTACCGAGACCGTGTTCAAGGCCGACGACAAGCTGGAGCCGATCATCCTCGATCGCAAGGAAGTGA
CCTACTCCTACTTCGCCGACCCGCTGTACGTCTTCATGGACAGCGAGTTCAACCACTACGAGATCGAGAAAG
ACGATCTGGAAGGCGTGCTGACCTTCATGAAGACGGCATGACCGACATCTGCGAAGCCGTGTTCTACAAC
GACAAGGTGATCTCGGTAGAGCTGCCGACCACCATCGTTTCGCCAGATCGCCTACACCGAGCCGGCCGTCCG
CGGCGACACCTCGGGCAAGGTGATGAAGACCGCGCGCCTGAACAACGGCGCCGAGTTGCAGGTTTCCGCG
TTCTGCGAAATCGGCGACTCGATCGAGATCGATACCCGCACCGGCGAGTACAAGTCCCGCGTCAAGGCCTG
A

Supplementary Table 1. Oligonucleotide sequences of expressed proteins. This table contains the oligonucleotide sequences used to express fragaceatoxin C mutant G13F, dihydrofolate reductase, proteasome-activating nucleotidase, high molecular weight adhesin 1, and elongation factor P.

Supplementary References

- (1) Chavis, A. E.; Brady, K. T.; Hatmaker, G. A.; Angevine, C. E.; Kothalawala, N.; Dass, A.; Robertson, J. W. F.; Reiner, J. E. Single Molecule Nanopore Spectrometry for Peptide Detection. *ACS Sensors* **2017**, 2 (9), 1319–1328. <https://doi.org/10.1021/acssensors.7b00362>.

Chapter 4

Automated Electrical Quantification of Vitamin B1 in a Bodily Fluid using an Engineered Nanopore-Sensor

The ability to measure the concentration of metabolites in biological samples is important, both in the clinic and for home diagnostics. Here we present a nanopore-based biosensor and automated data analysis for quantification of thiamine in urine in less than a minute, without the need for recalibration. For this we use the Cytolysin A nanopore and equip it with an engineered periplasmic thiamine binding protein (TbpA). To allow fast measurements we tuned the affinity of TbpA for thiamine by redesigning the π - π stacking interactions between the thiazole group of thiamine and TbpA. This substitution resulted furthermore in a marked difference between unbound and bound state, allowing the reliable discrimination of thiamine from its two phosphorylated forms by residual current only. Using an array of nanopores, this allows the quantification within seconds, paving the way for next-generation single-molecule metabolite detection systems.

This chapter was published as:

Lucas, F. L. R., Piso, T. R. C., van der Heide, N. H., Galenkamp, N. S., Hermans, J., Wloka, C., Maglia, C. (2021). Automated Electrical Quantification of Vitamin B1 in a Bodily Fluid using an Engineered Nanopore-Sensor. *Angewandte Chemie (International ed. in English)*. 6 p., anie.202107807

Thiamine (ThOH) is an essential vitamin (vitamin B1), which is important for human metabolism and cognitive functions.^{1,2} Deficiency of thiamine can lead to a range of non-specific symptoms, delaying diagnosis of its underlying cause and thereby potentially leading to permanent neural degeneration. The clinical relevance is not only limited to beriberi-related diseases (*e.g.* Wernicke encephalopathy and Korsakoff syndrome),³⁻⁶ but research has also highlighted thiamine as a potential biomarker for certain neurodegenerative diseases.⁷⁻⁹ The detection of thiamine is commonly performed using liquid chromatography, which requires a relatively high level of expertise and is time-consuming. Therefore, a cheap, fast, and specific detection system, enabling the quantification of thiamine in easily accessible bodily fluids, such as urine, may satisfy an unmet medical need.

A

B

C

R=H, PO_3^{2-} , $\text{P}_2\text{O}_6^{3-}$

106

Here we use the periplasmic thiamine binding protein (TbpA, **Figure 1B**), which has a specific binding affinity to thiamine (**Figure 1C**).⁴⁵ We developed a mutant of TbpA that displays fast transitions between its *apo* and bound states, and allows distinguishing thiamine from its phosphorylated derivatives. Finally, we develop an algorithm to automatically quantify the concentration of thiamine in human urine in less than a minute using only a single nanopore.

4.2. Results and Discussion

4.2.1. Wild-type TbpA (WT-TbpA) Signal

The entry of WT-TbpA (15 nM, added to the *cis* compartment) to ClyA-AS nanopores (-45 mV, 150 mM NaCl, 15 mM Tris.HCl, pH 7.5) was observed as a decrease in the ionic current (I_B) in respect to the open pore current (I_o) with a dwell time (τ) of 3 ± 1 s ($N = 10$, $n = 573$, where N is the number of nanopores used and n the number of individual protein blockades analyzed). The resulting residual current percentage ($I_{res}\% = I_B / I_o \times 100$, where I_B is the current during an event - *e.g.* a protein blockade) of the unbound *apo*-protein [$I_{res}\%(apo)$] was 49.0 ± 0.6 % ($N = 3$, $n = 398$). In addition, we observed non-active blockades (*i.e.* blockades that did not respond to the addition of thiamine, see later) with an $I_{res}\%$ (non-active) of 59.1 ± 0.7 % ($N = 4$, $n = 160$) and more rarely other levels, which we attribute mainly to misfolded proteins and impurities (**Supplementary Figure 1**). The addition of 4 μ M thiamine to the *cis* compartment reduced the residual current of the protein blockades [$\Delta I_{res}\%(\text{thiamine}) = 3.1 \pm 0.6$ %, $N = 4$, $n = 114$]. Transitions were observed only rarely between the *apo* and thiamine-bound form of WT-TbpA, suggesting a tight binding of the ligands to the protein (**Supplementary Figure 2**). The number of events with residual currents corresponding to thiamine-bound levels relative to the number of events corresponding to *apo*-levels was directly proportional to the concentration of thiamine added to the *cis* compartment (**Figure 2A**). The apparent dissociation constant (K_d^{app}) for thiamine binding to WT-TbpA could then be calculated by fitting a Hill-Langmuir equation (**Equation 1**) to the concentration dependency of the fraction of $I_{res}\%(apo)$ blockades.

$$\theta([L]) = a + b \frac{[L]^n}{K_d^{app} + [L]^n} \quad (1)$$

Where θ is the bound fraction observed given the ligand concentration $[L]$, a is the bound fraction in absence of substrate and represents intrinsic closing, b the normalization factor,⁵⁵ K_d^{app} the apparent dissociation constant and n represents the Hill coefficient.

Interestingly, in the absence of thiamine 16 ± 3 % of WT-TbpA blockades showed a residual current corresponding to the bound configuration, indicating that, as observed for other substrate binding proteins,⁴⁵ TbpA intrinsically closes without the substrate. We calculated an apparent dissociation constant (K_d^{app}) for thiamine binding to WT-TbpA of 67 ± 5 nM, and found the Hill coefficient to be one, indicating that, as expected, there is no cooperative binding. The K_d^{app} was roughly ten times higher than previously reported in ensemble studies (K_d of 3.8 ± 0.3 nM).⁵⁶ Since the concentration of the protein adaptor (15 nM) is similar to the amount of thiamine added to the *cis* solution (2 nM – 4 μ M), most likely, the discrepancy between the two K_d values reflects the uncertainty on the concentration of the free ligand in the *cis* solution. Surprisingly, the addition of thiamine to the *trans* solutions did not elicit additional current blockades, suggesting that the diffusion of thiamine from *trans* to *cis* might be impaired by the position of the wild-type protein adaptor inside the nanopore.

4.2.2. The Engineering of the TbpA Sensor

To determine thiamine concentration quickly and reliably, hundreds of events in a short period of time need to be gathered. Hence, we sought to redesign the WT-TbpA adaptor to obtain a sensor that is capable of rapid switching between its unbound and bound state. The thiazole ring of thiamine monophosphate is sandwiched between residues Tyr27 and Trp197 of TbpA (**Figure 2B**).⁵⁶ Thus, we reasoned that substitution of the tyrosine at position 27 with an alanine (Y27A-TbpA) would reduce π - π stacking interactions between TbpA and thiamine, therefore increasing the k_{off} .

Y27A-TbpA (15 nM, *cis*, -45 mV, 150 mM NaCl, 15 mM Tris.HCl at pH 7.5) active blockades were markedly shallower than WT-TbpA active blockades, (Y27A-TbpA - $I_{res}\%(apo) = 55.3 \pm 0.7$ %, $\tau = 3 \pm 1$ s, $N = 3$, $n = 236$, **Figure 2C, Supplementary Figures 3 and 4**), possibly reflecting the binding of the protein to a higher residence site inside the nanopore. Two residence sites were observed previously also for thrombin.⁴³ Y27A-TbpA was intrinsically closed 8 ± 3 % of the time providing a limit of detection at 3σ of 4.62 ± 0.02 nM, calculated from the intrinsic closing at 3σ (17% closed). Upon addition of 4 μ M thiamine to the *cis* side, we observed further rapid current blockades raising from the $I_{res}\%(apo)$ level, showing a $\Delta I_{res}\%(\text{thiamine})$ of 6.9 ± 0.6 % ($N = 12$, $n = 363$, **Supplementary Figure 5**). The latter is more than twice the value measured for WT-TbpA. The closed state of Y27A-TbpA [$I_{res}\%(\text{thiamine}) = 48.4 \pm 0.9$ %] is similar to the closed state of WT-TbpA [$I_{res}\%(\text{thiamine}) = 45.9 \pm 0.8$ %], suggesting that, upon binding of thiamine, Y27A-TbpA might assume a lower position inside the ClyA-AS pore. Conveniently, also the addition of thiamine to the *trans* chamber induced thiamine events (**Supplementary Figure 6**), further indicating that *apo*-Y27A-TbpA is likely positioned at a higher binding site than WT-TbpA, that allows the diffusion of the thiamine through the nanopore. Hence, Y27A-TbpA allows measuring

concentration dependency from either side of the nanopore. We used an *ad hoc* algorithm (see **methods**) to detect all events. We report the K_d^{app} value as estimated from the Hill curve fitted to the bound fraction (**Equation 1, Supplementary Information 1**) and calculated the on-rate and off-rate from the linear correlation between thiamine concentration and event rate (**Supplementary Figure 7**), resulting in the binding constants: $K_d^{app}(cis) = 23.1 \pm 0.1$ nM, $k_{on}(cis) = 69 \pm 9$ $\mu\text{M}^{-1}\text{s}^{-1}$, $k_{off}(cis) = 1.8 \pm 0.3$ s^{-1} and $K_d^{app}(trans) = 91.1 \pm 0.4$ nM, $k_{on}(trans) = 14 \pm 3$ $\mu\text{M}^{-1}\text{s}^{-1}$, $k_{off}(trans) = 1.7 \pm 0.1$ s^{-1} and a Hill coefficient of 1.

Figure 2. Electrical detection of thiamine. (A) Representative traces (left and middle) for wild-type thiamine binding protein (WT-TbpA, 15 nM), in the presence of thiamine, where thiamine and protein were added to the *cis* compartment. The red dotted line represents the *apo* current and the blue dotted line represents the ligand-bound current. The right panel indicates the observed bound fraction set against the thiamine concentration. The solid black line represents a fit of the corrected Hill equation. The black dashed and red dashed line indicate the apparent dissociation constant (K_d^{app}) in bound fraction and thiamine concentration, respectively. (B) Binding site of WT-TbpA with thiamine monophosphate adapted from Soriano et al.^[56] (C) Representative traces for an expansion of a Y27A thiamine binding protein (Y27A-TbpA, 15 nM) blockade in absence and presence of 16 nM thiamine added to *cis*. The red dotted line represents the *apo* current and the blue dotted line represents the ligand bound current. The observed bound fraction set against the thiamine concentration is shown on the right. The solid black line represents a fit of the corrected Hill equation. The black dashed and red dashed line indicate the apparent dissociation constant (K_d^{app}) in bound fraction and thiamine concentration, respectively. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), at -45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter, all traces were digitally filtered to 100 Hz using a Gaussian filter.

The 4-fold difference in the apparent K_d and k_{on} rates between *cis* and *trans* can be explained partially by the different concentration of free thiamine in solution (*vide supra*) and by the positive charge of thiamine. Under an applied potential of -45 mV, the electrophoretic force created by the negative *trans* electrode promotes the entry of the positively charged thiamine from *cis* to *trans* while reducing translocation from *trans* to *cis*. In addition, the *trans* entry of the pore is smaller than the *cis* entry, reducing accessibility of the analyte towards the protein, in turn decreasing the apparent K_d . Furthermore, the orientation of the protein within the nanopore might also play a role. Overall, Y27A-TbpA is more suitable as a sensor than WT-TbpA because it allows faster measurement of thiamine, a larger dynamic range, and allows thiamine capture from *trans*.

4.2.3. Detection of Phosphorylated Derivatives

Phosphorylation of the hydroxy group of thiamine (ThOH) can form thiamine mono- and diphosphate derivatives.⁵⁷ The diphosphorylated form (ThDP), is an essential co-factor in many organisms.⁵⁸

In contrast, thiamine monophosphate (ThMP) is not active as a co-factor. However, it has been suggested that ThMP may facilitate transfer across cellular membranes.⁵⁹ Both ThMP and ThDP are synthesized from ThOH. It was previously reported that WT-TbpA can bind ThOH and its phosphorylated derivatives at near-equimolar K_d .⁵⁶ As both ThDP and ThMP are present in some biological fluids, we sought to detect these derivatives with the Y27A-TbpA. The binding of ThMP and ThDP induced specific current events to Y27A-TbpA blockades (**Figure 3**). Upon the addition of 0.1 μM ThMP to the *cis* side, we observed two kinds of bound events with a $\Delta I_{res}\%(\text{ThMP}_1)$ of $-4.8 \pm 0.2\%$ ($N = 3$, $n = 236$) and with a $k_{off}(\text{ThMP}_1)$ of $2.3 \pm 0.4\text{ s}^{-1}$ as well as $\Delta I_{res}\%(\text{ThMP}_2)$ of $-7.3 \pm 0.3\%$ ($N = 3$, $n = 395$) with a $k_{off}(\text{ThMP}_2)$ of $2.0 \pm 0.2\text{ s}^{-1}$. The capture frequency (an approximate k_{on}) for both events was $16 \pm 4\text{ }\mu\text{M}^{-1}\text{ s}^{-1}$ (0.1 μM ThMP). Similarly, at an approximated $K_d^{app}(\text{cis})$ of 2 μM ThDP, we observed one type of event with an $\Delta I_{res}\%(\text{ThDP}_1)$ of $-4.0 \pm 0.1\%$ ($N = 3$, $n = 720$) with a $k_{off}(\text{ThDP}_1)$ of $2.9 \pm 0.4\text{ s}^{-1}$, and a second type of event with an $\Delta I_{res}\%(\text{ThDP}_2)$ of $-7.2 \pm 0.1\%$ ($N = 3$ and $n = 853$) with a $k_{off}(\text{ThDP}_2)$ of $1.7 \pm 0.3\text{ s}^{-1}$. The capture frequency for both events was $6.5 \pm 3\text{ }\mu\text{M}^{-1}\text{ s}^{-1}$ (2 μM ThDP). The off-rates of both ThMP_{1-2} and ThDP_{1-2} were similar to the off-rate of thiamine. (**Figure 3**). Thus, we hypothesized that the additional events may correspond to thiamine rather than ThMP or ThDP. To exclude this possibility, we performed anion-exchange chromatography, directly prior to nanopore experiments as well as again after nanopore experiments, to make sure no hydrolysis had occurred and found no evidence of thiamine in the samples (**Supplementary Figure 7**). In addition, we performed the same experiments after letting the samples incubate overnight at 37°C and observed no significant hydrolysis (**Supplementary Figure 8**). Moreover, we subjected each sample to mass spectrometry, and we did not observe thiamine in the phosphorylated samples (**Supplementary Figure 9**). The most likely

explanation, therefore, is that while thiamine binds to a singular Y27A-TbpA conformation, ThMP and ThDP may bind to two conformations. It was also previously reported that the thiazole ring of ThOH has to turn in order to form π - π stacking interaction with Tyr27 and Trp197, while the phosphate groups on ThMP and ThDP disallow such a rearrangement.⁵⁶ The difference between the off-rates of ThMP₁ and ThDP₁ bound states is possibly due to hydrogen bonds formed between the ThDP β -phosphate group with the side chains of Asp28 and Ser29.⁵⁶ It might also be possible that the differences between the thiamine derivatives is caused by a different position Y27A-TbpA assumes in the nanopore as a consequence of the additional charges in the phosphorylated substrates. We previously observed that manipulation of the dipole of a confined binding protein in ClyA-AS, by means of an addition of a negatively charged point mutation, may cause a similar effect.⁴⁸ Hence, despite the analysis of the thiamine phosphate derivatives is challenging, our data shows that ThOH, ThMP and ThDP can be distinguished by nanopore currents.

Figure 3. Detection of thiamine (ThOH), thiamine monophosphate (ThMP) and thiamine diphosphate (ThDP) using Y27A-Thiamine binding protein (Y27A-TbpA). Representative blockades induced by (A) 16 nM ThOH added to *cis* (B) 0.1 μ M ThMP added to *cis* (C) 2 μ M ThDP added to *cis*. All data was recorded in 150 mM NaCl, 15 mM Tris.HCl, pH 7.5 using a sampling frequency of 10 kHz and an analogue Bessel-filter of 2 kHz and an applied bias -45 mV. Traces were filtered using a 100 Hz digital Gaussian filter.

4.2.4. Automated Analysis

Next, we sought to develop an automated nanopore sensor that, upon addition of the sample, can determine the concentration of thiamine without further human intervention. We selected human urine, because it does not require invasive sampling and has a very low protein background. The

primary form of thiamine found in urine is ThOH.⁴ For automated data analyses we utilized a recursive event detection algorithm, which recognizes the protein blockades and ligand-induced events in the data using a threshold search. The algorithm analyzes snippets of 10 seconds in real-time. From each snippet a full-point histogram is taken, and the open pore current identified by fitting a normal distribution around the expected value. Then, for events longer than 25 ms, a threshold search (3σ upper boundary of the I_o , **Figure 4A**) is performed in order to identify protein blockades (**Figure 4A**). Each protein blockade is compared to expected value of the *apo* level and used for subsequent threshold analysis if this is not significantly different. From the latter, the start and end times are tabulated and the ligand events within each of these blockades are detected using a second threshold analysis. Start and end times of each ligand event are then tabulated (**Figure 4A**).

From each tabulated snippet, protein blockades are extracted and combined. The resulting combined trace, which only contains the signal from protein blockades, is further divided into snippets of one second and a normalized full-point histogram of each snippet is created. Each bin of the normalized histograms is then multiplied with the prior probability of the ligand state and the prior probability of the *apo* state to create two histograms: one for the marginal likelihood of the *apo* state and one for the liganded state (**Figure 4B**). The fraction of the likelihood is then used as the bound fraction in each 1-second snippet. The estimated bound fractions for each 1-second snippet are weighted by the determinate error and averaged to obtain the bound fraction. Using the Hill equation, the bound fraction is used to calculate the concentration of thiamine. We show that this workflow will converge into the true bound fraction (**Supplementary Information 1**).

4.2.5. Automatic (real-time) Detection of Thiamine in Urine

In healthy human adults, ThOH is generally found in tens to hundreds of ng/mL (nanomolar range),⁶⁰ which is near the $K_d^{app}(cis)$ we measured for the Y27A-TbpA mutant. Hence, we used the automated real-time data analysis to measure the thiamine concentration in urine from a young adult male, using a palm-sized eOne HS amplifier. We added a final concentration of 3% (v/v) human urine samples to the *cis* solution containing Y27A-TbpA (15 nM), 150 mM NaCl, 15 mM Tris.HCl at pH 7.5 using a single dodecameric ClyA-AS nanopore. By using the same calibration curve described in **Figure 2C**, the automatic analysis established a thiamine concentration in the urine sample of 68 ± 4 ng/mL ($N = 3$), which is not significantly different from the HPLC-MS result at 3σ (57 ± 1 ng/mL). To test the limits of our system, we performed automatic (real-time) analysis by resampling obtained traces (**Figure 4C/D**). We show that the error converges after approximately 20 seconds of a concatenated protein blockade time, indicating that using a single nanopore an accurate measurement takes at least 20 seconds.

Taken together, we demonstrate a nanopore-sensor for the rapid, automated, and direct analysis of thiamine in samples of urine, without the need of sample preparation.

Figure 4. Real-time detection of the thiamine (ThOH) concentration in a urine sample using Y27A-Thiamine binding protein-equipped ClyA-AS nanopore. (A) Current trace showing the capture of Y27A-TbpA (blockades) by the ClyA-AS nanopore. The trace is divided into snippets of 10 seconds and analysed for blockades. The blockades are further analysed and divided into snippets of 1 second. For each blockade snippet a histogram is created and used for subsequent analysis. The dotted black line represents the baseline current, and the red dashed lines represent the 3σ thresholds. The vertical (grey) lines represent the snippets for blockade and bound fraction determination. (B) Workflow for real-time detection. A normalised histogram is created from 1 second (blockade) snippets and the bound fraction for each snippet is calculated. The ligand bound state is indicated in blue and the *apo* state is represented in red. The weighted mean of the bound fractions is calculated using the inverse slope in ligand concentration as weights. The thiamine concentration is determined from the Hill equation. (C) Estimated bound fraction of three independent measurements where 3% (cyan) or 5% (magenta/green) of urine were added to the *cis* compartment. The analysis time is the number of snippets used in B (1 s = 1 snippet). (D) The weighted mean thiamine concentration in nanomolar and ng/mL estimated from three independent measurements, see C.

4.3. Conclusion

Nanopore biosensors have emerged as potential low-cost candidates to rapidly detect biologically relevant molecules in biological samples. Recent contributions have shown the capabilities of nanopore-coupled binding proteins as a way to specifically detect small molecules in biological

fluids.^{44,45,47} In this work, we have shown that the affinity of *Escherichia coli* TbpA toward thiamine can be tuned by reducing π - π stacking interactions between the thiazole group of thiamine and TbpA, by substitution of a single residue, Y27A. This modification enhanced the nanopore signal, which allowed the quantification and detection of thiamine and its phosphorylated derivatives. We have also developed an algorithm that allowed to determine the concentration of thiamine in human urine without requiring an operator in a matter of tens of seconds. It also demonstrates that the sensor is not impaired by the thousands of different molecules present in urine.⁶¹ Because the protein sensor element (Y27A-TbpA, in *cis*) can be separated from the sample (*trans*), this technology may also be applicable to samples containing significant quantities of protein, such as blood. This is because large molecules cannot pass the narrow *trans* constriction and must overcome a comparable strong electroosmotic flow under negative applied potential. Compared to other amperometric sensors, this approach does not require calibration, because the signal from inactive proteins is ignored. Since hundreds of proteins exist that can detect molecules with high selectivity, nanopores fitted with internal protein sensors provide a starting point for new home diagnostic metabolite sensors.

4.4. Materials and Methods

Chemicals. Tris-(hydroxymethyl)-aminomethane hydrochloride (Tris.HCl), sodium chloride, and 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris.base), β -dodecylmaltoside (DDM), lysozyme, 2xYT-broth, LB-medium, TB-medium, ampicillin, agar-agar, agarose (broad-range), glucose, sucrose, glycerol, and imidazole were retrieved from Carl-Roth. DNase1, Phire II Hot Start Polymerase, Dpn1, isopropyl β -D-1-thiogalactopyranoside (IPTG) were retrieved from Fisher Scientific. Pentane and ethanol were acquired from Boomlab. 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC), ethylenediaminetetraacetic acid (EDTA), magnesium chloride ($MgCl_2$), N-hexadecane, thiamine hydrochloride (ThOH), thiamine-(4-methyl-¹³C-thiazol-5-yl-¹³C₃), thiamine monophosphate chloride dehydrate (ThMP) and thiamine pyrophosphate (ThDP) were received from Merck. Primers and synthetic genes were retrieved from Integrated DNA Technologies (IDT). Ni-NTA agarose beads were obtained from Qiagen. *E. coli* 10G and *E. coli* BL21 (DE3) cells were originally retrieved from Lucigen. Polyacrylamide gels (PAGE) and Blue native polyacrylamide gel electrophoresis (BN-PAGE) was obtained from Bio-Rad. All solutions were prepared in ultrapure water (Milli-Q®), unless stated otherwise.

Purification of ClyA-AS. ClyA-AS was purified as described previously.⁶² In brief: ClyA-AS was transformed into *E. coli* BL21 (DE3) using a pT7-SC1 plasmid with an ClyA-AS insert.^{44,62} Transformants were grown overnight at 37 °C on a LB-agar plate supplemented with 100 mg/L ampicillin and 1%

glucose. The (whole) LB-agar plate was used to inoculate 200 mL 2xYT medium supplemented with 100 mg/L ampicillin, which was grown at 37 °C under constant shaking (180 rpm) to OD₆₀₀=0.6, after which 0.5 mM IPTG was added and left to grow overnight at 25 °C under constant shaking (180 rpm). The cell culture was harvested using centrifugation for 20 min at 4 °C and 6000 rpm. The cell pellets were frozen for 1 hour at -80°C and solubilized in 20 mL lysis buffer (150 mM NaCl, 15 mM Tris.HCl, pH 7.5, 1 mM MgCl₂, 0.2 unit/mL DNase1, 10 mM imidazole, 10 µg/mL lysozyme). Lysis was performed by incubation for 20 minutes at room temperature and sonication (Branson). Pellet cell debris was removed by centrifugation (6500 rpm, 30 min, 4 °C). Monomers were purified using Ni-NTA affinity purification. ClyA-AS monomers were incubated for 30 min at 37 °C in 150 mM NaCl, 15 mM Tris.HCl (pH 7.5) and 0.02% DDM. The oligomeric forms and monomers were separated using 4-20% polyacrylamide gels and blue native polyacrylamide gel electrophoresis. The band corresponding to Type 1 ClyA-AS was cut out and stored in-band until used. Oligomer containing bands were extracted using 150 mM NaCl, 15 mM Tris.HCl (pH 7.5) supplemented with 0.2 % DDM and 10 mM EDTA.

Cloning of TbpA. A synthetic gene containing His₆-tagged TbpA was inserted into a pT7-SC1 vector using standard TOPO cloning.⁶³

Expression of TbpA. TbpA (wild-type and mutant) was expressed by *E. coli* BL21 (DE3) containing a pT7-SC1 plasmid with a His₆-tagged TbpA insert. The *E. coli* BL21 (DE3) cells were grown overnight on a LB-agar plate supplemented with 100 mg/L ampicillin and 1% glucose at 37 °C. A single colony was inoculated into 10 mL LB-medium supplemented with 100 mg/mL ampicillin for 4 hours at 37 °C while shaking (180 rpm). The 10 mL starter was used to inoculate 1 L TB-medium supplemented with 100 mg/L ampicillin. The TB-medium was left to grow to OD₆₀₀~0.6 at 37 °C while shaking (180 rpm), after which the temperature was lowered to 20 °C until OD₆₀₀~1.2. Upon reaching OD₆₀₀~1.2, the culture was induced using 0.5 mM IPTG and left to grow overnight.

Purification of TbpA. Overnight (induced) cell pellets were collected using centrifugation at 4000 rpm for 30 minutes at 4 °C. The cell pellets (1 L of growth) were resuspended in 200 mL ice-cold sucrose buffer, containing 20% Sucrose (w/v), 50 mM Tris.HCl pH 7.5 and 1 mM EDTA, and incubated for 30 minutes under rotation at room temperature. Washed pellets were collected using centrifugation at 7500 rpm for 10 minutes at 4 °C. The pellets were resuspended in 200 mL ice-cold 5 mM MgCl₂ and incubated for 1 hour at room temperature while under constant rotation. The solution was centrifuged at 7500 rpm for 30 minutes at 4 °C and the supernatant was loaded onto a Ni-NTA affinity column. The Ni-NTA affinity column was extensively washed with 100 mL of 50 mM Tris.HCl pH 7.5, 1M NaCl, 20 mM imidazole and 1% glycerol followed by a wash with 200 mL of 50 mM Tris.HCl pH 7.5, 150 mM NaCl, 20

mM imidazole and 1% glycerol. TbpA was eluted using 5 steps of 5 mL elution buffer containing 50 mM Tris.HCl pH 7.5, 150 mM NaCl, 300 mM imidazole and 1% glycerol. The purest fractions were combined and concentrated using 30 kDa cut-off Amicon Ultra Centrifugal Filters.

Construction of Y27A-TbpA. The Y27A point mutations was introduced to TbpA using site-directed mutagenesis.⁶⁴ The His₆-tagged TbpA plasmid was mutated using two complementary primers, forward (CCCGTGTGACTGTGTACA CGGCAGACAGT) and reverse (CCAGTCTTGCCCGAAACTG TCTGCCGTGTA) using Phire II Hot Start Polymerase with a final volume of 50 µL. Thermal cycle: 30 seconds 98 °C followed by 30 cycles of denaturation 98 °C for 5 seconds, annealing 59 °C for 30 seconds and elongation 72 °C for 180 seconds followed by an elongation step of 300 seconds at 72 °C. The plasmid template was removed by incubating the PCR product with Dpn1 (1 FDU) for 2 hours at 37 °C. The product was incorporated into electro competent E. cloni[®] 10G cells. Transformants were grown overnight at 37 °C on LB-agar plates supplemented with 100 mg/L ampicillin and 1% glucose. Single colonies were grown in 5 mL LB-medium supplemented with 100 mg/L ampicillin and DNA was extracted using QIAprep[®] Spin Miniprep. The incorporation of the mutation was confirmed by MacroGen EZ-seq standard direct sequencing.

Protein sequence of Y27A-TbpA.

MGSSHHHHHHSSGLVPRGSHMKPVLTVYTADSFAADWGP GPVVKKA FEADCNCELKLVALEDGVSLNRLRMEG
KNSKADVVLGLDNNLLDAASKTGLFAKSGVAADAVNVPGGWNNDTFVPFDYGYFAFVYDKNKLKNPPQSLKELVES
DQNRWVIYQDPRTSTPGLGLLLWMQKVYGDDAPQAWQKLAKKTVTVTKGWSEAYGLFLKGESDLVLSYTTSPAYHI
LEEKKDNYAAANFSEGHYLQVEVAARTAASKQPELAQKFLQFMVSPAFQNAIPTGNWMPVANVTLPAGFEKLT KP
ATTLEFTPAEVAAQRQAWISEWQRAVSR

DNA sequence of Y27A-TbpA.

ATGGGTTCTCCACCATCACCATCATTCGAGTGGACTGGTCCCCGCGGTTTCGCACATGAAGCCCGTGTGTA
CTGTGTACACGTATGACAGTTTTGCGGCAGACTGGGGTCCCGGGCCCGTAGTGAAAAAGGCTTTGAAGCGGAT
TGTAATTGTGAGTTGAAGTTGGTAGCCTTAGAAGATGGGGTATCACTTCTGAACCGCCTTCGTATGGAAGGTAAG
AACTCCAAGGCCGACGTTGTGTTGGGCTTGGATAATAATTTGCTTGATGCCGCGAGTAAGACTGGCCTGTTCGCT
AAGTCTGGAGTGGCCGCTGATGCAGTGAACGTACCCGGTGGATGGAACAACGACACCTTTGTGCCATTCGATTA
CGGTTACTTCGCCTTCGTGTACGACAAGAATAAGTTAAAGAATCCCCCAATCACTGAAAGAACTTGTTGAGTCT
GATCAGAACTGGCGTGTGATCTATCAGGACCCGCGCACCTCGACCCCGGGCTTGGGACTGTTGCTTTGGATGCA
GAAAGTGACGGTGACGATGCCCCGCAAGCATGGCAGAACTGGCTAAGAAAACCGTAAGTGTACAAAAGGG
TGGAGCGAAGCATACGGCTTATTCCTGAAGGGTGAGAGCGACCTTGTCCTTTCGTATACTACGTCTCCAGCATACC
ATATTTTGGAGGAAAAGAAAGACAACATATGCAGCGGCTAACTTCTCGGAAGGTCATTACCTTCAAGTGGAAGTT

GCTGCCCCGACAGCCGCGTCGAAGCAACCGGAGTTGGCTCAAAAATTTCTGCAGTTCATGGTTAGCCCCGCTTT
TCAGAACGCCATCCCCACTGGCAACTGGATGTACCCGGTGGCAAACGTTACCCTGCCTGCTGGATTGAGAAATT
GACAAAGCCAGCTACAACATTAGAATTTACCCAGCGGAGGTAGCAGCACAGCGTCAGGCCTGGATCTCAGAAAT
GGCAGCGTGCAGTGTCTCGTTGA

Black lipid membrane electrophysiological recordings. A two compartment recording chamber consisting of two polyoxymethylene cuvettes (600 μ L, described in an earlier study)⁶⁵ separated using a 25 μ m thick Teflon membrane, with an aperture diameter of approximately 100 μ m, was used. 5 μ L of 2.5% hexadecane in pentane was applied to both sides of the aperture and left to dry for 1 minute. 500 μ L of buffer (150 mM NaCl, 15 mM Tris.HCl, pH 7.5) was added to both compartments topped with 20 μ L of 12.5 mg/mL DPhPC in pentane, the pentane was left to evaporate for 2 minutes. A silver/silver chloride electrode was inserted into the *cis* and the *trans* compartment and connected to the ground and working electrode of the Axon 200b amplifier, respectively. Lipid bilayers were created at the aperture using the Langmuir-Blodgett method described by Maglia *et al.* .⁶⁵ The orientation of ClyA nanopores was determined by the asymmetry of the current-voltage relationship. 15 nM TbpA was added to the *cis* compartment and well mixed (unless stated otherwise). Thiamine was added to either the *cis* or *trans* compartment.

Data acquisition. Ionic current data was acquired using two instruments. **Setup A** (Figure 2 and 3): an Axopatch 200B amplifier combined with a Digidata 1550B A/D converted, both obtained from Axon instruments, was used. Similar to preceding work,^{10,65} recordings were made using the Clampex 10 software, recording at a sampling frequency of 10 kHz using an analogue Bessel filter of 2 kHz. The final bandwidth was reduced to 10 Hz for analysis using a digital Gaussian filter. **Setup B** (in Figure 4): an eOne XV (elements SRL) operating on hereby presented software was used at a sampling frequency of 20 kHz, digitally filtered down to 10 kHz. The final bandwidth was reduced to 10 Hz for analysis using a digital Gaussian filter.

Computational procedures.

Data analysis. Electrophysiology traces were analyzed using Jupyter Notebook (version 5.5.0), implemented in the Anaconda (version 5.2.0) environment supported by Python 3.6.5 (64-bit). Module packages not included in Anaconda (version 5.2.0) were installed from PyPi using pip (version 10.0.1) unless stated otherwise. Neo (version 0.7.1) was used to load Axon Binary Format traces into NumPy arrays (NumPy version 1.14.3). Events were detected, and the bound fraction was estimated using the algorithm as described in this contribution (Recursive Event Detection). Corrected Hill functions were fitted using curve fit optimizer as implemented in SciPy (version 1.1.0).

Thiamine determination using HPLC. Urine samples were stored at -25 °C, thawed on the day of analysis. Thawed samples were pre-treated by addition of formic acid to a final concentration of 0.1% (pH 3). Prior to analysis, the standard mixtures (sample + standard) were vortexed and centrifuged for 10 minutes at 20 000 rpm. The supernatant was used for analysis. The quantification of thiamine was performed using a Prominence LC-20AD (Shimadzu Europa GmbH, Duisburg, Germany) equipped with an analytical Discovery C18 column (2.1 x 100 mm, 5 µm, Supelco, Bellefonte, USA), connected to a TSQ Quantum Ultra™ Mass Spectrometer (Thermo Electron, San Jose, USA). The mobile phase consisted of 4.8 mM ammonium bicarbonate, pH 7.9 (solvent A) and acetonitrile (solvent B). An isocratic gradient of 50% solvent B for 10 minutes at a flow rate of 0.25 mL/minute and a sample injection volume of 2 µL was used. The thiamine concentration was determined using a thiamine-(4-methyl-¹³C-thiazol-5-yl-¹³C₃)-labelled internal standard at a concentration of 1 µM, and a calibration curve of ThOH between 0.05 µM to 10 µM, containing 0.1% formic acid in all samples.

4.5. References

1. Bender, D. A. Optimum nutrition: thiamin, biotin and pantothenate. *Proc. Nutr. Soc.* **58**, 427–433 (1999).
2. Frank, R. A. W., Leeper, F. J. & Luisi, B. F. Structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cell. Mol. Life Sci.* **64**, 892–905 (2007).
3. Wai, J. M. *et al.* Using clinical decision support through the electronic medical record to increase prescribing of high-dose parenteral thiamine in hospitalized patients with alcohol use disorder. *J. Subst. Abuse Treat.* **99**, 117–123 (2019).
4. Whitfield, K. C. *et al.* Thiamine deficiency disorders: diagnosis, prevalence, and a roadmap for global control programs. *Ann. N. Y. Acad. Sci.* **1430**, 3–43 (2018).
5. Salloum, S., Goenka, A. & Mezooff, A. Beriberi disease in an 11-year-old girl with total colectomy. *Oxford Med. Case Reports* **2018**, (2018).
6. Fattal-Valevski, A. Outbreak of Life-Threatening Thiamine Deficiency in Infants in Israel Caused by a Defective Soy-Based Formula. *Pediatrics* **115**, e233–e238 (2005).
7. Gibson, G. E. *et al.* Vitamin B1 (thiamine) and dementia. *Ann. N. Y. Acad. Sci.* **1367**, 21–30 (2016).
8. Rodríguez, J.-L., Qizilbash, N. & López-Arrieta, J. Thiamine for Alzheimer's disease. *Cochrane Database Syst. Rev.* (2001) doi:10.1002/14651858.CD001498.
9. Lu'ong, K. vinh quốc & Nguyễn, L. T. H. Role of Thiamine in Alzheimer's Disease. *Am. J. Alzheimer's Dis. Other Dementias* **26**, 588–598 (2011).
10. 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 Int Ed Engl* **55**, 12494–12498 (2016).
11. Bayley, H. Nanopore Sequencing: From Imagination to Reality. *Clin. Chem.* **61**, 25–31 (2015).

12. Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci.* **105**, 20647–20652 (2008).
13. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A* **106**, 7702–7707 (2009).
14. Maglia, G., Heron, A. J., Stoddart, D., Japrun, D. & Bayley, H. Analysis of Single Nucleic Acid Molecules with Protein Nanopores. in *Methods in Enzymology* vol. 475 591–623 (2010).
15. Bhattacharya, S., Yoo, J. & Aksimentiev, A. Water Mediates Recognition of DNA Sequence via Ionic Current Blockade in a Biological Nanopore. *ACS Nano* **10**, 4644–4651 (2016).
16. Cao, C. *et al.* Discrimination of oligonucleotides of different lengths with a wild-type aerolysin nanopore. *Nat. Nanotechnol.* **11**, 713–718 (2016).
17. Sutherland, T. C. *et al.* Structure of Peptides Investigated by Nanopore Analysis. *Nano Lett.* **4**, 1273–1277 (2004).
18. Movileanu, L., Schmittschmitt, J. P., Martin Scholtz, J. & Bayley, H. Interactions of Peptides with a Protein Pore. *Biophys. J.* **89**, 1030–1045 (2005).
19. Cao, C. *et al.* Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores. *Nat. Commun.* **10**, 4918 (2019).
20. Restrepo-Pérez, L., Wong, C. H., Maglia, G., Dekker, C. & Joo, C. Label-Free Detection of Post-translational Modifications with a Nanopore. *Nano Lett.* **19**, 7957–7964 (2019).
21. Huang, G., Voet, A. & Maglia, G. FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution. *Nat. Commun.* **10**, 835 (2019).
22. Li, S., Cao, C., Yang, J. & Long, Y.-T. Detection of Peptides with Different Charges and Lengths by Using the Aerolysin Nanopore. *ChemElectroChem* **6**, 126–129 (2019).
23. Restrepo-Pérez, L. *et al.* Resolving Chemical Modifications to a Single Amino Acid within a Peptide Using a Biological Nanopore. *ACS Nano* **13**, 13668–13676 (2019).
24. Ouldali, H. *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
25. Lucas, F. L. R. *et al.* The Manipulation of the Internal Hydrophobicity of FraC Nanopores Augments Peptide Capture and Recognition. *ACS Nano* (2021) doi:10.1021/acsnano.0c09958.
26. Wilson, J., Sarthak, K., Si, W., Gao, L. & Aksimentiev, A. Rapid and Accurate Determination of Nanopore Ionic Current Using a Steric Exclusion Model. *ACS Sensors* **4**, 634–644 (2019).
27. Si, W. & Aksimentiev, A. Nanopore Sensing of Protein Folding. *ACS Nano* **11**, 7091–7100 (2017).
28. Robertson, J. W. F. *et al.* Single-molecule mass spectrometry in solution using a solitary nanopore. *Proc. Natl. Acad. Sci.* **104**, 8207–8211 (2007).
29. Mohammad, M. 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).
30. Cao, C. *et al.* Discrimination of oligonucleotides of different lengths with a wild-type aerolysin nanopore. *Nat. Nanotechnol.* **11**, 713–718 (2016).
31. Asandei, A. *et al.* Electroosmotic Trap against the Electrophoretic Force near a Protein

- Nanopore Reveals Peptide Dynamics during Capture and Translocation. *ACS Appl. Mater. Interfaces* **8**, 13166–13179 (2016).
32. Chavis, A. E. *et al.* Single Molecule Nanopore Spectrometry for Peptide Detection. *ACS Sensors* **2**, 1319–1328 (2017).
 33. Asandei, A., Rossini, A. E., Chinappi, M., Park, Y. & Luchian, T. Protein Nanopore-Based Discrimination between Selected Neutral Amino Acids from Polypeptides. *Langmuir* **33**, 14451–14459 (2017).
 34. Robertson, J. W. F. & Reiner, J. E. The utility of nanopore technology for protein and peptide sensing. *Proteomics* **18**, 1–36 (2018).
 35. Piguet, F. *et al.* Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat. Commun.* **9**, 966 (2018).
 36. Gu, L. Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature* **398**, 686–690 (1999).
 37. Li, X., Lee, K. H., Shorkey, S., Chen, J. & Chen, M. Different Anomeric Sugar Bound States of Maltose Binding Protein Resolved by a Cytolysin A Nanopore Tweezer. *ACS Nano* **14**, 1727–1737 (2020).
 38. Sanchez-Quesada, J., Ghadiri, M. R., Bayley, H. & Braha, O. Cyclic Peptides as Molecular Adapters for a Pore-Forming Protein. *J. Am. Chem. Soc.* **122**, 11757–11766 (2000).
 39. Boersma, A. J., Brain, K. L. & Bayley, H. Real-Time Stochastic Detection of Multiple Neurotransmitters with a Protein Nanopore. *ACS Nano* **6**, 5304–5308 (2012).
 40. Boersma, A. J. & Bayley, H. Continuous Stochastic Detection of Amino Acid Enantiomers with a Protein Nanopore. *Angew. Chemie Int. Ed.* **51**, 9606–9609 (2012).
 41. Cheley, S., Gu, L.-Q. & Bayley, H. Stochastic Sensing of Nanomolar Inositol 1,4,5-Trisphosphate with an Engineered Pore. *Chem. Biol.* **9**, 829–838 (2002).
 42. Braha, O., Webb, J., Gu, L.-Q., Kim, K. & Bayley, H. Carriers versus Adapters in Stochastic Sensing. *ChemPhysChem* **6**, 889–892 (2005).
 43. Soskine, M., Biesemans, A., De Maeyer, M. & Maglia, G. Tuning the size and properties of ClyA nanopores assisted by directed evolution. *J Am Chem Soc* **135**, 13456–13463 (2013).
 44. Galenkamp, N. S., Soskine, M., Hermans, J., Wloka, C. & Maglia, G. Direct electrical quantification of glucose and asparagine from bodily fluids using nanopores. *Nat. Commun.* **9**, 4085 (2018).
 45. Zernia, S., van der Heide, N. J., Galenkamp, N. S., Gouridis, G. & Maglia, G. Current Blockades of Proteins inside Nanopores for Real-Time Metabolome Analysis. *ACS Nano* **14**, 2296–2307 (2020).
 46. Soskine, M. *et al.* An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry. *Nano Lett.* **12**, 4895–4900 (2012).
 47. Soskine, M., Biesemans, A. & Maglia, G. Single-Molecule Analyte Recognition with ClyA Nanopores Equipped with Internal Protein Adaptors. *J Am Chem Soc* **137**, 5793–5797 (2015).
 48. Van Meervelt, V. *et al.* Real-Time Conformational Changes and Controlled Orientation of Native Proteins Inside a Protein Nanoreactor. *J. Am. Chem. Soc.* **139**, 18640–18646 (2017).
 49. Willems, K. *et al.* Accurate modeling of a biological nanopore with an extended continuum

- framework. *Nanoscale* **12**, 16775–16795 (2020).
50. Willems, K. *et al.* Engineering and Modeling the Electrophoretic Trapping of a Single Protein Inside a Nanopore. *ACS Nano* **13**, 9980–9992 (2019).
 51. Galenkamp, N. S., Biesemans, A. & Maglia, G. Directional conformer exchange in dihydrofolate reductase revealed by single-molecule nanopore recordings. *Nat. Chem.* **12**, 481–488 (2020).
 52. Galenkamp, N. S. & Maglia, G. Substrate binding and turnover modulate the affinity landscape of dihydrofolate reductase to increase its catalytic efficiency. *bioRxiv* 2020.04.14.040733 (2020) doi:10.1101/2020.04.14.040733.
 53. Li, M.-Y. *et al.* Revisiting the Origin of Nanopore Current Blockage for Volume Difference Sensing at the Atomic Level. *JACS Au* jacsau.1c00109 (2021) doi:10.1021/jacsau.1c00109.
 54. Li, M.-Y., Wang, Y.-Q., Ying, Y.-L. & Long, Y.-T. Revealing the transient conformations of a single flavin adenine dinucleotide using an aerolysin nanopore. *Chem. Sci.* **10**, 10400–10404 (2019).
 55. Gesztelyi, R. *et al.* The Hill equation and the origin of quantitative pharmacology. *Arch. Hist. Exact Sci.* **66**, 427–438 (2012).
 56. Soriano, E. V. *et al.* Structural Similarities between Thiamin-Binding Protein and Thiaminase-I Suggest a Common Ancestor † , ‡. *Biochemistry* **47**, 1346–1357 (2008).
 57. Goyer, A. Thiamine in plants: Aspects of its metabolism and functions. *Phytochemistry* **71**, 1615–1624 (2010).
 58. Ajjawi, I., Rodriguez Milla, M. A., Cushman, J. & Shintani, D. K. Thiamin pyrophosphokinase is required for thiamin cofactor activation in Arabidopsis. *Plant Mol. Biol.* **65**, 151–162 (2007).
 59. Manzetti, S., Zhang, J. & van der Spoel, D. Thiamin Function, Metabolism, Uptake, and Transport. *Biochemistry* **53**, 821–835 (2014).
 60. *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline.* (National Academies Press, 1998). doi:10.17226/6015.
 61. Bouatra, S. *et al.* The Human Urine Metabolome. *PLoS One* **8**, e73076 (2013).
 62. Galenkamp, N. S. *et al.* Preparation of Cytolysin A (ClyA) Nanopores. in 11–18 (2021). doi:10.1007/978-1-0716-0806-7_2.
 63. Shuman, S. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J. Biol. Chem.* **269**, 32678–32684 (1994).
 64. Braman, J., Papworth, C. & Greener, A. Site-Directed Mutagenesis Using Double-Stranded Plasmid DNA Templates. in *In Vitro Mutagenesis Protocols* 31–44 (Humana Press). doi:10.1385/0-89603-332-5:31.
 65. Maglia, G., Heron, A. J., Stoddart, D., Japrun, D. & Bayley, H. Analysis of Single Nucleic Acid Molecules with Protein Nanopores. in *Methods Enzymol* vol. 475 591–623 (2010).
 66. Bishop, C. *Pattern Recognition and Machine Learning.* (Springer-Verlag New York, 2006).

4.6. Supporting Information

1. Determination of ligand bound fraction using empirical Bayesian inference

We determine the bound fraction of the *ligand-bound* level using empirical Bayesian inference⁶⁶, which allows the estimation of the bound fraction on sparse data.

$$\theta = a + b * \left(\frac{[L]^n}{K_d + [L]^n} \right), \quad [\text{Hill} - \text{Langmuir equation}] \quad (1)$$

Where θ is the bound fraction observed given the ligand concentration $[L]$. The constant a represents the offset to correct for intrinsic closing and b represents the normalization factor. K_d is the apparent dissociation constant and n represents the Hill coefficient. With the binding affinity is either positively ($n > 1$), negatively ($n < 1$) or not ($n = 1$) influenced by analyte binding.

We rewrite the Hill-Langmuir equation in terms of the normalized bound fraction, which results in **equation 2**, and insert the equation to calculate the dissociation constant from the on and off rate resulting in **equation 3**.

$$\tilde{\theta} = \frac{\theta - a}{b} = \left(\frac{[L]^n}{K_d + [L]^n} \right) \quad (2)$$

Replace K_d using the dissociation constant equation such that $K_d = \frac{k_{off}}{k_{on}}$,

$$\tilde{\theta} = \frac{k_{on}}{k_{on}} * \frac{[L]^n}{\left(\frac{k_{off}}{k_{on}} + [L]^n \right)} = \frac{[L]^n * k_{on}}{k_{off} + [L]^n * k_{on}} \quad (3)$$

We rewrite equation 3 in terms of time spend in the *apo* (**equation 4**) and ligand bound (**equation 5**) state. The resulting (**equation 6**) can be expressed as the probability of finding the liganded state $\{P(\text{ligand})\}$ relative to the total probability of finding the liganded and *apo* state $\{P(\text{apo})\}$. We determine the probability of finding the liganded state (**equation 7**) and probability of finding the *apo* state (**equation 8**) based on the data ($\mathcal{D}$) using Bayes theorem.

$$event_{time}(\text{apo}) = \frac{1}{[L]^n * k_{on}} \quad (4)$$

$$event_{time}(\text{ligand}) = \frac{1}{k_{off}} \quad (5)$$

$$\tilde{\theta} = \frac{event_{time}(\text{apo})^{-1}}{event_{time}(\text{ligand})^{-1} + event_{time}(\text{apo})^{-1}} = \frac{event_{time}(\text{ligand})}{event_{time}(\text{ligand}) + event_{time}(\text{apo})} = \frac{P(\text{ligand}|\mathcal{D})}{P(\text{ligand}|\mathcal{D}) + P(\text{apo}|\mathcal{D})} \quad (6)$$

$$P(\text{ligand}|\mathcal{D}) = \frac{P(\mathcal{D}|\text{ligand}) * P(\text{ligand})}{P(\mathcal{D})} \quad (7)$$

$$P(apo|\mathcal{D}) = \frac{P(\mathcal{D}|apo)*P(apo)}{P(\mathcal{D})} \quad (8)$$

Where $P(ligand|\mathcal{D})$ is the probability of the ligand state given the data, $P(\mathcal{D}|ligand)$ is the probability of the data given the ligand state, $P(ligand)$ is the probability of the ligand state and $P(\mathcal{D})$ is the probability of the data. Where $P(apo|\mathcal{D})$ is the probability of the *apo* state given the data, $P(\mathcal{D}|apo)$ is the probability of the data given the *apo* state, $P(apo)$ is the probability of the *apo* state and $P(\mathcal{D})$ is the probability of the data.

Prior to analysis of an unknown sample, we use the control (for *apo*) and ligand-saturated (for *ligand-bound*) traces to determine the prior probabilities for the *apo* $\{P(apo)\}$ and ligand-bound state $\{P(ligand)\}$. During real-time analysis, we determine the empirical likelihood $P(\mathcal{D}|apo, ligand)$ from the full point histogram of protein blockades that contain the I_{apo} . We take peak drifting into account by allowing the centers of the prior probabilities to shift an equal amount. The bound fraction is calculated as the fraction between the *apo* and ligand states (**equation 9.**)

$$\tilde{\theta} = \frac{P(ligand|\mathcal{D})}{P(ligand|\mathcal{D})+P(apo|\mathcal{D})} = \frac{P(\mathcal{D}|ligand)*P(ligand)}{P(\mathcal{D}|ligand)*P(ligand)+P(\mathcal{D}|apo)*P(apo)} \quad (9)$$

We estimate the bound fraction based on snippets of 1 second blocked time (protein capture into ClyA-AS). We calculate the weighted average over time as the error in the bound fraction is non-linear, *e.g.* the error is lower near the K_d . The derivative of the ligand concentration is proportional to the determinate error (**equations 10 and 11, Supplementary Figure 6**), thus we calculated the weighted average based on this error estimation (**equations 12**), as well as the variance (**equation 13**).

$$[L](\theta) = \left(\frac{(\theta-a)*K_d}{a+b-\theta} \right)^{\frac{1}{n}} \quad (10)$$

$$\frac{d}{d\theta} [L](\theta) = - \frac{b * \left(\frac{(\theta-a)*K_d}{a+b-\theta} \right)^{\frac{1}{n}}}{n(a-\theta)(a+b-\theta)} \quad (11)$$

$$\bar{\theta} = \frac{\sum_{i=1}^n \theta_i * \frac{d}{d\theta} [L](\theta_i)}{\sum_{i=1}^n \frac{d}{d\theta} [L](\theta_i)} \quad (12)$$

$$\sigma_{\bar{\theta}}^2 = \frac{\sum_{i=1}^n \frac{d}{d\theta} [L](\theta_i) * (\theta_i - \bar{\theta})^2}{\sum_{i=1}^n \frac{d}{d\theta} [L](\theta_i)} \quad (13)$$

Inverse slope of the ligand concentration calculated from the bound fraction based on the Y27A-TbpA (added to *cis*) calibration. The bound fraction plotted against the slope of the ligand concentration for Y27A-TbpA (added to *cis*) calibration. The slope of the ligand concentration is calculated using equation 11.

Supplementary Figure 1. Continuous trace of WT-TbpA in presence of thiamine Two parts of a continuous trace showing ionic current blockades elicited by WT-TbpA (15 nM, *cis*) to the ClyA nanopore displaying open and blocked pore currents. Thiamine-bound level (ThOH) and unbound (apo) level are denoted. Inactive blockades are marked with an asterisk. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), at -45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter, all traces were digitally filtered to 100 Hz using a Gaussian filter. 64 nM thiamine is present (near the apparent K_D).

Supplementary Figure 2. Representative signal snippets of WT-TbpA with thiamine added to the *cis* chamber. Representative traces for WT thiamine binding protein (WT-TbpA, 15 nM, *cis*) using a protocol expelling the protein every few seconds, in the presence of thiamine, where thiamine was added to the *cis* compartment. The red dotted line represents the *apo* current and the blue dotted line represents the ligand bound current. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), using a negative applied bias potential of 45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter, all traces were digitally filtered to 100 Hz using a Gaussian filter.

Supplementary Figure 3. Representative signal snippets and binding curves of Y27A-TbpA with thiamine added to *trans*. Representative traces for Y27A thiamine binding protein (Y27A-TbpA, 15 nM, *cis*), in the presence of thiamine, where thiamine was added to the *trans* compartment. The panel on the right shows the observed bound fraction set against the thiamine concentration, for Y27A-TbpA, addition of thiamine to the *trans* compartment. The dotted red line represents the K_d^{app} concentration. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), using a negative applied bias potential of 45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter, all traces were digitally filtered to 100 Hz using a Gaussian filter.

Supplementary Figure 4. Continuous trace of Y27A-TbpA in presence of thiamine Two parts of a continuous trace showing ionic current blockades elicited by Y27A-TbpA (15 nM, *cis*) to the ClyA nanopore. Thiamine-bound level (ThOH) and unbound (*apo*) level are denoted. Inactive blockades are marked with an asterisk. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), at -45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter, all traces were digitally filtered to 100 Hz using a Gaussian filter. 16 nM thiamine is present (near the apparent K_D).

Supplementary Figure 5. Representative signal snippets of Y27A-TbpA with thiamine added to the *cis* chamber. Representative traces for Y27A thiamine binding protein (Y27A-TbpA, 15 nM, *cis*), in the presence of thiamine, where thiamine was added to the *cis* compartment. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), using a negative applied bias potential of 45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter, all traces were digitally filtered to 100 Hz using a Gaussian filter.

Supplementary Figure 6. Representative signal snippets of Y27A-TbpA with thiamine added to the *trans* chamber. Representative traces for Y27A thiamine binding protein (Y27A-TbpA, 15 nM, *cis*), in the presence of thiamine, where thiamine was added to the *trans* compartment. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), using a negative applied bias potential of 45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter, all traces were digitally filtered to 100 Hz using a Gaussian filter.

Supplementary Figure 7. Apparent on- and off-rates of thiamine binding to Y27A-Thiamine binding protein. Opening and closing rates of thiamine binding Y27A-Thiamine binding protein (Y27A-TbpA, 15 nM, *cis*) relative to the concentration of thiamine added to **(A)** *cis* or **(B)** *trans*. The solid lines represent the best linear fit (1st order polynomial) through the concentration depend event rate. The rising (black) line represents the closing rate of Y27A-TbpA as the inverse event time of observed $I_{res}(apo)$. The decreasing (red) line represents the opening rate of Y27A-TbpA as the inverse event time of observed $I_{res}(thiamine)$. All measurements were performed on an Axon 200B amplifier with a 1550B digitizer in 150 mM NaCl supplemented with 15 mM Tris.HCl buffer (pH 7.5), using a negative applied bias potential of 45 mV at a sampling frequency of 10 kHz filtered to 2 kHz using a Bessel filter.

Supplementary Figure 8. FPLC result of thiamine and its phosphorylated derivatives. **(A)** thiamine, **(B)** thiamine monophosphate injected directly (solid line) and after incubation overnight at 37°C (dotted), **(C)** thiamine diphosphate injected directly (solid line) and after incubation overnight at 37°C (dotted). Experiments were performed using a HiTrap Q anion exchange column and a linear gradient between 0 M and 1 M NaCl buffered with 15 mM Tris to pH 7.5.

Supplementary Figure 9. LC/ESI-MS result of thiamine and its phosphorylated derivatives. The top-left figure shows the background control. The top-right figure shows thiamine elution at 13.7 minutes. The bottom-left figures show the thiamine monophosphate sample eluting at 1.1 minutes without thiamine diphosphate contamination and no thiamine near 13 minutes. The bottom-right figure shows the thiamine diphosphate eluting at 1.1 minutes without thiamine monophosphate contamination and no thiamine near 13 minutes.

Chapter 5

Concluding Discussion and Outlook

5.1. Augmenting Nanopore Resolution for Peptide Detection

In the second chapter of this thesis, we described the engineering of fragaceatoxin C (FraC) to enhance the quantifiability of peptides. We showed that the constriction site of FraC is located between residues D10 and D17, while modification of residue K20 did not influence the nanopore behaviour. Moreover, we found that the resolution of peptide detection at pH 3.8 was largely augmented by a phenylalanine residue in the constriction site of the pore, while introduction of hydrophilic residues removed the ability of the pore to capture peptides. We suggest that the augmentation for peptide identification is the result of cation- π interactions. It was previously observed that the removal of the arginine on position 220 of the beta-barrel nanopore aerolysin augmented analyte capture.¹ In addition, they observed that a positively charged peptide has an especially increased residence time when R220 is modified to tryptophan. While it is likely that the removed positive charge is the main benefactor for this result, it also seems beneficial to incorporate aromatic residues in the sensing region of nanopores for the detection of (small) charged peptides.

It may also be possible to increase the residence time of analytes by molecular interactions with the nanopore surface. We have shown that position K20 of FraC can be modified, potentially providing interesting opportunities for further enhancement as it can be envisioned that a binding interaction can be engineered at this position. By slowing down peptides close to the constriction site, it may become possible to identify smaller differences between translocating peptides.

Interestingly, we observed no significant ion selectivity at the exact pH of 3.8, in both the mutant G13F-FraC and wild-type FraC nanopores. This suggests that an electroosmotic flow (EOF) from *cis* to *trans* significantly decreases the residence time, while the analyte is captured by electrophoresis. We also noticed that reversal of the EOF (*trans* to *cis*) by introduction of a positively charged residue on position D10 is not optimal, as this pore is unstable under a negatively applied potential, which is a requirement for the translocation of positively charged peptides. Instead, we propose that the ionic composition of the buffer solution can be modified, to control the EOF by a difference in the relative mobility of the ions. The ion mobility of potassium and chloride is roughly equal. However, substitution of potassium with low mobility cations, such as lithium, should provide a net negative EOF (*trans* to *cis*). Indeed, Oukhaled and co-workers observed that the EOF is stronger in the presence of LiCl rather than KCl.² We expect that this effect is less pronounced when substituting the anion chloride with another halogen ion as the relative mobilities of these ions are similar. It may be possible to replace the anion with a divalent sulphate ion, however, we expect that this would complicate the system while producing minimal improvement.

5.2. Nanopore-based Protein Identification

Enhancement of peptide residence inside the nanopore provides interesting opportunities for the fingerprinting of proteins with a higher resolution, in a similar fashion as shown in chapter three. In this chapter we showed that proteins can be identified based on the fingerprint resulting from tryptic peptides translocating the nanopore. While this does not provide single-molecule information, each observed peak in the residual current spectrum represents a single type of peptide. As we described, the correlation between the peptide sequence and excluded current cannot be easily drawn and further experimental and computational work is required to accurately estimate the current that a peptide will exclude. The height of the observed peaks, however, seems to correlate with the observed abundance from electrospray ionisation mass spectrometry (ESI-MS). Thus, the G13F-FraC nanopore sensor can be used in combination with ESI-MS to detect differences in volume when *e.g.* two peptides are isobaric or if these are difficult to ionise. The latter example is especially interesting, as post-translational modifications (PTMs), such as glycosylation, can reduce the ionisation efficiency.³ Nanopores do not need to ionise analytes for detection and also they do not require a phase transition (liquid-gas), making them ideal candidates for the detection of these analytes.

PTMs such as phosphorylation and glycosylation can be detected as a shift in the residual current. Moreover, it was shown previously that PTMs can be detected and partially localised on peptides with controlled transport.^{4,5} While localisation of a PTM is complicated due to uncontrolled translocation in the system described in chapter three, it is possible to provide an accurate concentration of modified and unmodified proteins.

The resolution of the current generation of nanopores is reaching the minimal requirement to detect individual amino acids.⁶ However, the uncontrolled translocation of peptides and proteins across a nanopore remains a complex problem yet to be solved. Nonetheless, recently it was shown by Zhang *et al.* that an active proteasome can be attached on top of a nanopore.⁷ This proteasome can in theory “chop” proteins and “drop” the resulting peptides in the nanopore. This method called “chop-n-drop” has been shown feasible for the G13F-FraC nanopore. In the study by de Lannoy *et al.*, the G13F-FraC nanopore with a presumed resolution of 4 Da can identify of 97.1 % of the human proteome.⁸

5.3. Real-time Quantification of Small Molecules

Another possible application of nanopores is their utilisation for the direct detection and identification of small molecules.^{9,10} While nanopores have been able to detect small molecules directly for over 20

years,¹¹ it is likely too complex to identify individual molecules when they are present in a complex mixture. Especially when complex biological samples are measured, where abundant analytes may obfuscate the detection of less abundant molecules. Moreover, the complexity for identification of analytes scales exponentially with the number of analytes. Thus, it would not be feasible to extract qualifiable information from single-molecule translocation in a non-selective system. Nevertheless, there are possible ways to circumvent this problem by using an array of nanopores with diverging selectivity, or by utilising substrate binding proteins (SBPs).

In the fourth chapter we described a system comprising a SBP coupled with a nanopore sensor. The nanopore lends its sensitivity for detection, while the SBP provides selectivity towards the analyte. As the binding kinetics of the SBP are known, we were able to show that real-time detection using an empirical Bayesian method can greatly improve the speed of quantification. While it may seem that only a single analyte can be detected, the system is not limited to incorporation of only a single SBP. Rather, multiple SBPs can be captured and distinguished allowing the quantification of glucose and asparagine simultaneously in blood, sweat, urine, and saliva, while only utilising a low amount of sample and minimal preparation.¹² The empirical Bayesian method should be applicable to both the glucose and asparagine SBPs individually, and each SBP displays a characteristic behaviour inside the nanopore. Thus, we envision a sensor where multiple SBPs are added to the nanopore, which are characterised by their protein blockade, and subsequently the concentration of an analyte can be calculated in automatically and in real-time.

5.4. Outlook for Fully Folded Protein Analysis

An interesting opportunity for nanopore sensors is their use to detect the fully folded shape of a protein. As explained in the introductory chapter of this thesis, the functionality of a protein is largely dependent on its three-dimensional structure. Nanopores could provide structural information of (mis)folded proteins in (near) physiological conditions, which may potentially be used to gain important insights for cellular profiling.

As nanopores observe fluctuations in excluded ionic current, it is unlikely that all proteins can be identified based on just their current fluctuations. Rather, we envision that a database of known protein behaviours would be compared with the measured signal. It had been shown that the current fluctuations observed during nanopore electrophysiology of fully folded proteins are related to their shape, volume, rotational diffusion coefficient, dipole and charge.^{13,14} These properties depend less on the primary sequence of a protein, but are highly dependent on the tertiary structure. It was previously

shown that the behaviour inside the ClyA-AS nanopore may be modelled accurately.^{15,16} This would allow protein characterisation based on in-silico predictions, given that the set of expected proteins are known, *e.g.* a library of proteins could be created. If we could model each protein inside the nanopore, we should be able to qualify and quantify fully folded proteins. While proteins can theoretically consist of any number of amino acids, proteomic research revealed that the human proteome consist of approximately twenty thousand proteins.¹⁷ Until recently, this would require an enormous computational investment to model each protein using atomistic, or course-grained molecular dynamic simulations. However, recent advances in machine learning-based protein structure prediction, such as AlphaFold2, allow the rapid and accurate prediction of the three-dimensional structure of proteins.¹⁸ It is important to notice, however, that these approaches are fairly new and have until now not been able to model protein-ligand or protein-protein interactions. And even if we consider that the three-dimensional structure and subsequent current behaviour can be modelled accurately enough, the system is likely to remain too complex when utilising only a single nanopore. That is, the resolution of a single nanopore is likely not sufficient to distinguish twenty thousand proteins. Nonetheless, it might be feasible to detect proteins in multiple orthogonal nanopores, where each nanopore is slightly more selective towards certain characteristic behaviours.

5.5. References

1. Cao, C. *et al.* Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores. *Nat. Commun.* **10**, 4918 (2019).
2. Boukhet, M. *et al.* Probing driving forces in aerolysin and α -hemolysin biological nanopores: electrophoresis versus electroosmosis. *Nanoscale* **8**, 18352–18359 (2016).
3. Illiano, A. *et al.* Protein Glycosylation Investigated by Mass Spectrometry: An Overview. *Cells* **9**, 1986 (2020).
4. 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).
5. Restrepo-Pérez, L. *et al.* Resolving Chemical Modifications to a Single Amino Acid within a Peptide Using a Biological Nanopore. *ACS Nano* **13**, 13668–13676 (2019).
6. Ouldali, H. *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
7. Zhang, S. *et al.* Bottom-up fabrication of a multi-component nanopore sensor that unfolds, processes and recognizes single proteins. *bioRxiv* 2020.12.04.411884 (2020) doi:10.1101/2020.12.04.411884.
8. de Lannoy, C., Maglia, G. & de Ridder, D. Chop-n-drop: *in silico* assessment of a novel single-molecule protein fingerprinting method employing fragmentation and nanopore detection. *bioRxiv* 2021.05.27.445917 (2021) doi:10.1101/2021.05.27.445917.

9. Li, M.-Y. *et al.* Revisiting the Origin of Nanopore Current Blockage for Volume Difference Sensing at the Atomic Level. *JACS Au* **1**, 967–976 (2021).
10. Li, M.-Y., Wang, Y.-Q., Ying, Y.-L. & Long, Y.-T. Revealing the transient conformations of a single flavin adenine dinucleotide using an aerolysin nanopore. *Chem. Sci.* **10**, 10400–10404 (2019).
11. Gu, L.-Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature* **398**, 686–690 (1999).
12. Galenkamp, N. S., Soskine, M., Hermans, J., Wloka, C. & Maglia, G. Direct electrical quantification of glucose and asparagine from bodily fluids using nanopores. *Nat. Commun.* **9**, 4085 (2018).
13. Yusko, E. C. *et al.* Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nat. Nanotechnol.* **12**, 360–367 (2017).
14. Houghtaling, J. *et al.* Estimation of Shape, Volume, and Dipole Moment of Individual Proteins Freely Transiting a Synthetic Nanopore. *ACS Nano* **13**, 5231–5242 (2019).
15. Willems, K. *et al.* Accurate modeling of a biological nanopore with an extended continuum framework. *Nanoscale* **12**, 16775–16795 (2020).
16. Wilson, J., Sarthak, K., Si, W., Gao, L. & Aksimentiev, A. Rapid and Accurate Determination of Nanopore Ionic Current Using a Steric Exclusion Model. *ACS Sensors* **4**, 634–644 (2019).
17. Omenn, G. S. Reflections on the HUPO Human Proteome Project, the Flagship Project of the Human Proteome Organization, at 10 Years. *Mol. Cell. Proteomics* **20**, 100062 (2021).
18. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* (2021) doi:10.1038/s41586-021-03819-2.

Summary

Chapter 1 provides a general introduction on nanopore-based technologies for the analysis of proteins, and small metabolites. The chapter starts with a gentle introduction of polypeptides and a global overview of common techniques and methodologies used to extract proteomic and metabolomic information. Followed by an overview of single-molecule techniques that have successfully been employed to sequence deoxynucleic acids (DNA) and focussing on the key challenge towards nanopore based protein sequencing.

The second part of chapter one describes the nanopore technique in detail, with a focus towards biological nanopores. In essence, nanopore sensors consist of a non-conductive sheet with a nanometre sized aperture connecting two electrolyte filled compartments. The conductance across this aperture is reduced due to ion exclusion whenever a molecule translocates through it, which can be observed as a reduction in current when a potential is applied. In order to create these nanopores, we utilise membrane-spanning pore-forming proteins, which allows a high reproducibility and enables the engineering of nanopores with atomistic precision. The engineering of nanopores has seen many advances over recent years. This resulted in nanopores that are ideal for the detection of peptides, such as fragaceatoxin C (FraC); but also nanopores with an ideal shape to capture fully folded proteins such as cytolysin A (ClyA). These smaller pores resulted in a new technique called nanopore spectrometry, where peptides are characterized based on their excluded current. However, the detection and characterisation of peptides is complicated due to uncontrolled rapid translocation, which combined with low sensitivity did not yet yield a reliable platform to identify naturally occurring peptides. Nonetheless, smaller molecules such as metabolites can be detected with high specificity. By trapping a substrate binding protein inside a nanopore such as ClyA, conformational changes of the protein can be observed. The substrate binding protein acts as a transducer, which can cause large current fluctuations upon binding to a substrate.

Chapter 2 describes that FraC nanopores can be engineered to augment the quantifiability of peptides. The detection of peptides using nanopores has attracted interest from the scientific community. However, the controlled transport of peptides across nanopores remains a complex problem yet to be solved. We probed FraC using single-point mutations near the presumed recognition site of the pore, and residues that were most amiable for engineering were identified using homology search. By quantification of translocating peptides resulting from a stochastic tryptic digest at pH 3.8, we were able to determine the most favourable mutations on the nanopore. We showed evidence that the recognition site of FraC is located between residues D10 and D17, and that the charge on these residues is vital for the capture of peptides; while substitution to a negative charge at position K20

does not influence the nanopore characteristics. Interestingly, we found that addition of hydrophilic residues at position G13 removed the ability of FraC to accurately detect peptides, while introduction of aromatic residues on this position largely augmented the detection of these analytes. Especially, when the position G13 of octameric FraC was exchanged for a tryptophan or phenylalanine, a dramatic increase in resolution was observed. However, upon further investigation, we found that the heptameric form of FraC highly favours a phenylalanine on this position, allowing the detection of the previously unobserved heptapeptide kemptide. We expect that kemptide remains undetected in other types of FraC due to the high positive charge on this heptapeptide, which induces fast electrophoretic translocation. We also found that there is no significant ion selectivity for both the wild-type FraC and G13F-FraC nanopores at pH 3.8, explaining the importance of this exact acidity. This shows that the substitution of glycine with phenylalanine did not affect the electroosmotic properties of the nanopore, thus, we believe that the retention of peptides is augmented by cation- π interactions between the phenylalanine on position G13 and the charged N-terminal on the peptides at pH 3.8.

Chapter 3 follows the analogy set by the early days of mass spectrometry based proteomics, which identified proteins based on their peptide mass fingerprint. Herein, proteins are digested using trypsin and the resulting observed fragment masses are used to identify the initial protein. We observed a high resolution of peptide capture in the G13F-FraC mutant, which would, in theory, allow the identification of proteins based on their proteolytic peptides. We show that the G13F-FraC mutant displays a polynomial correlation between the molecular weight and excluded current of translocating (synthetic) peptides expected to result from tryptic digestion of lysozyme. By digesting a model protein lysozyme with trypsin, we obtained a mixture of peptides and subjected this sample to electrospray ionisation mass spectrometry (ESI-MS). Utilising the polynomial correlation, we were able to map the molecular weight observed from ESI-MS to the expected excluded current spectrum, and utilized the observed intensity of the ESI-MS to determine the peak height. Rewardingly, we found that the overlap between the expected excluded current spectrum and experimentally measured spectrum was high for lysozyme. We furthered this correlation by subjecting nine additional proteins to ESI-MS and nanopore electrophysiology, and observe that both the expected excluded current spectra and measured excluded current spectra share similar profiles. Additionally, we used a spectral matching algorithm to identify individual spectra and see that we can assign each individual spectrum to the correct protein. These findings show the potential of nanopores to be utilised as low-cost analysers for protein identification.

Chapter 4 presented a thiamine binding protein coupled with a nanopore sensor, which displayed large changes in the ionic current upon binding with thiamine (Vitamin B1). We showed that the wild-type periplasmic thiamine binding protein (TbpA) can be efficiently captured by a ClyA-AS nanopore, and that a clear difference between *apo* TbpA and thiamine-bound TbpA can be observed. However, we rarely observed unbinding of thiamine, which complicated the rapid quantification of this analyte, and bound-TbpA was only observed when the binding protein and analyte were added to the same compartment. This indicated that binding of thiamine to TbpA was an almost exclusive bulk-process. It was previously shown that the binding pocket of TbpA is highly dependent on the π - π stacking interaction of residue Y27 to the thiazole ring of thiamine. Thus, we engineered the binding region of TbpA to the thiazole ring of thiamine by substituting the tyrosine at position Y27 to an alanine (Y27A-TbpA). This resulted in a binding protein with rapid transitions of approximately 2 s^{-1} , and was able to also detect thiamine when this was added in the opposite compartment. Additionally, we probed the Y27A-TbpA using two phosphorylated derivatives of thiamine and found that the residual current was different, however, the compounds were not quantifiable. We also sought to automatically detect thiamine in Human urine, and developed an algorithm to rapidly detect the bound-fraction of thiamine bound to TbpA. We derived an empirical Bayesian method to provide *a priori* information, which to allowed us to detect an accurate thiamine concentration within a minute.

Samenvatting

Hoofdstuk 1 geeft een algemene introductie over nanoporie gebaseerde technologieën voor de analyse van eiwitten en kleine metabolieten. Het hoofdstuk start met een introductie van polypeptiden en geeft een globaal overzicht van standaard technieken en methodologieën in gebruik voor het extraheren van proteomische en metabolomische informatie. Waarna enkel-molecuul technieken worden toegelicht voor het uitlezen desoxyribonucleïnezuur (DNA) en een focus op de belangrijkste uitdaging om het nanoporie systeem te vertalen van het uitlezen van DNA naar eiwitten.

Het tweede deel van dit hoofdstuk beschrijft het nanoporie systeem in detail, met een focus op biologische nanoporiën. Nanoporie sensoren bestaan, in essentie, uit een membraan met zeer hoge weerstand waarin een apertuur met een nanometer grote diameter wordt toegebracht, welke twee elektrolyt gevulde compartimenten verbindt. De weerstand over de dit membraan wordt verhoogt door ion exclusie wanneer translocatie van een molecuul door de apertuur plaatsvindt. Wij maken gebruik van membraan spannende porie vormende eiwitten om de nanoporiën te creëren, welke met grote reproduceerbaarheid op atomistische schaal kunnen worden aangepast. Dit resulteerde in de afgelopen jaren tot kleine nanoporiën zoals fragaceatoxin C (FraC) met een ideale vorm voor de detectie van peptiden, maar ook grote nanoporiën zoals cytolysin A (ClyA) voor de analyse van gevouwen eiwitten. Van deze kleine nanoporiën vloeide een nieuwe techniek voort genaamd nanoporie spectrometrie, waarbij een peptide gekarakteriseerd wordt op basis van de uitgesloten stroom. De detectie en karakterisering van peptiden wordt echter bemoeilijkt door ongecontroleerde translocatie, welke gecombineerd met een lage sensitiviteit nog geen betrouwbaar platform heeft gebouwd voor de identificatie van natuurlijk voorkomende peptiden. Desalniettemin is het wel mogelijk om kleinere moleculen, zoals metabolieten, te detecteren met hoge specificiteit. Door een substraat bindend eiwit te meten in een grote nanoporie, zoals ClyA, kunnen conformatie verandering geobserveerd worden. Dit substraat bindend eiwit werkt als een omvormer, welke een grote verandering in de ion exclusie veroorzaakt wanneer het bindt met een substraat.

Hoofdstuk 2 beschrijft de aanpassing van FraC nanoporiën om de kwantificeerbaarheid van peptiden te vergroten. De detectie van peptiden met nanoporiën heeft de aandacht getrokken van de wetenschappelijke gemeenschap. Echter, blijft de ongecontroleerde translocatie van peptiden door nanoporiën een complex probleem. Wij maken gebruik van enkel-residu mutaties rond de herkenning regio van de nanoporie, waarbij de makkelijkst aanpasbare aminozuren gevonden werden door middel van het zoeken van homologie. Wij zijn erin geslaagd om de meest gunstige modificaties te bepalen door middel van de kwantificering van peptiden op pH 3.8, welke een resultaat waren van een stochastische tryptische ontleding. Wij laten bewijs zien dat de herkenning regio van FraC gelokaliseerd

is tussen de residuen D10 en D17, en dat de lading op deze residuen vitaal is voor de herkenning van peptiden. Daarnaast laten wij zien dat een negatieve lading op residu K20 geen invloed uitoefent op het gedrag van FraC. Een interessante bevinding was dat het vervangen van residu G13 naar een hydrofiel residu de mogelijk voor FraC om peptiden te herkennen verwijderde, terwijl de vervanging naar een aromatische functionaliteit de herkenning van peptiden verbeterde. In het bijzonder observeerde wij dat de resolutie van FraC werd drastisch verbeterd wanneer positie G13 van de FraC octameer werd aangepast naar een tryptofaan of fenylalanine. Echter, na verder onderzoek bleek dat de heptamerische vorm van FraC een grote voorkeur heeft voor een fenylalanine op positie G13, welke de mogelijkheid bied om een eerder verborgen zeer positief geladen heptapeptide, kemptide, te meten. Wij veronderstellen dat kemptide niet gedetecteerd kan worden door andere vormen van FraC door de hoge positieve lading op deze heptapeptide, welke door elektroforese snelle translocatie kan veroorzaken. Wij vonden geen significante ion selectiviteit op een pH van 3.8 voor zowel de originele als gemuteerde FraC, welke een verklaring geeft voor de voorkeur van deze specifieke zuurtegraad. Dit laat zien dat de vervanging van glycine met een fenylalanine geen invloed uitoefent op de elektroosmotische eigenschappen van de nanopore, daartoe geloven wij dat de retentie van peptiden verbeterd word door kation- π interacties tussen de fenylalanine op positie G13 en de positief geladen N-terminus van peptiden in een zuurtegraad van 3.8.

Hoofdstuk 3 volgt de analogie gezet door de vroegere dagen van massa spectrometrie gebaseerde proteomica, welke eiwitten herkend door middel van de peptiden massa. Hierbij worden eiwitten verteerd door trypsin, en de massa van de resulterende fragmenten worden gebruikt om het oorspronkelijke eiwit te bepalen. Wij zagen een hoge resolutie voor peptiden herkenning met de G13F-FraC mutant, welke, in theorie, de mogelijkheid bied om eiwitten te identificeren gebaard op hun proteolytische peptiden. Wij laten zien dat de G13F-FraC mutant een polynomische correlatie laat zien tussen het moleculair gewicht en tegengehouden stroom van synthetische peptiden (verwacht van proteolyse van lysozym) die translocatie ondergaan in de nanopore. Wij creëerde een mengsel van proteolytisch peptiden door ontleding van lysozym en observeerde de massa van de resulterende peptiden met de hulp van *electrospray* ionisatie massa spectrometrie (ESI-MS). Wij slaagden erin om het moleculaire gewicht te extrapoleren naar een histogram van de tegengehouden stroom (uitgesloten stroom spectrum) in de nanopore. Hierbij maakten wij gebruik van de intensiteit zoals deze werd geobserveerd in ESI-MS. Wij zagen dat het voorspelde uitgesloten stroom spectrum een gelijkwaardig profiel vormde als het experimenteel gemeten uitgesloten stroom spectrum voor lysozym. Wij hebben negen extra eiwitten gemeten om deze correlatie te verbreden. Hiertoe hebben wij het ESI-MS en nanopore spectrum gemeten en zagen wij dat het voorspelde en experimentele

uitgesloten stroom spectra gelijkwaardige profielen vertoonde. Om de mogelijkheid tot identificatie te laten zien, hebben wij de individuele spectra laten aanwijzen met de hulp van een spectraal koppeling algoritme, welke elk spectrum correct koppelde aan het verwachte eiwit. Deze resultaten laten zien dat nanoporiën gebruikt kunnen worden als een goedkoop platform voor eiwit identificatie.

Hoofdstuk 4 presenteert een thiamine bindend eiwit gekoppeld met een nanoporie sensor, welke grote stroom fluctuaties laat zien wanneer het bind met thiamine (vitamine B1). Wij laten zien dat het originele periplasmatische thiamine bindend eiwit (TbpA) efficiënt kan worden gemeten in de ClyA-AS nanoporie. Hierbij kan een duidelijk verschil worden gemeten tussen de niet gebonden *apo* TbpA en thiamine gebonden TbpA. Wij zagen maar zelden ontbinding van thiamine, welke de snelle kwantificatie van het analyt bemoeilijkte. Daarnaast zagen wij alleen gebonden eiwit wanneer thiamine en TbpA in het zelfde compartiment werden toegevoegd. Dit liet zien dat de binding van thiamine naar TbpA bijna exclusief gebeurd in het compartiment en niet in de nanoporie. Uit voorgaand onderzoek bleek dat de manier waarop TbpA bind aan thiamine hoogst afhankelijk is van een π - π stapeling interactie tussen residu Y27 en de thiazool ring van thiamine. Om deze interactie aan te passen, hebben wij de tyrosine op positie Y27 aangepast naar een alanine (Y27A-TbpA). Dit resulteerde in een eiwit met snelle transitie van 2 s^{-1} tussen gebonden en ongebonden thiamine, welke ook de mogelijkheid bied om thiamine te meten als deze word toegevoegd in het andere compartiment. Daarnaast hebben wij ook Y27A-TbpA gemeten met twee gefosforyleerde afgeleiden van thiamine en zagen dat de uitgesloten stroom ander was, echter, waren deze afgeleiden niet kwantificeerbaar. Wij hebben een algoritme ontwikkeld voor de snelle en automatische detectie van thiamine in menselijke urine. Hiertoe maken wij gebruik van een empirische Bayesiaanse methode, om *a priori* informatie te verschaffen. De combinatie van de Y27A-TbpA sensor en het algoritme resulteerde in de mogelijkheid om met grote precisie de concentratie van thiamine te meten in minder dan één minuut van analyse.

Acknowledgements

My PhD has been a four year long journey, and there are many people I'd like to thank for various reasons.

First and foremost, I would like to thank Giovanni for giving me the opportunity to join the Maglia lab for my promotion. Your critical and logical thinking has provided me with the opportunity to excel in our field. You always treated me with respect and while we did not always agree, it felt like a symbiotic partnership with this thesis as a result. While on the topic of this thesis, I would like to thank the members of my reading committee, professors W. R. Browne, M.W. Fraaije, and G.L. Corthals for taking the time to assess the thesis and provide valuable feedback. In addition, I'd like to express my gratitude to the opposition during the defence ceremony, and would like to thank my paranymphs Nieck and Amanda for your support.

I would like to thank Bea and Anmara for being the best secretaries I could've wished for. To my co-promoter, Carsten, you were there from the beginning of this journey and I would to thank you for everything you taught me. From the start you treated me with respect and you were always there when I had questions, both related to work and other.

Folkert, Petár, Nieck, and Tjemme I would to thank you for all the support you have given me over the past few years.

To Roderick and Lisa, we started our journey together and, while we can't be colleagues forever, I would like to thank you both for all the work and parties we shared.

I'd like to thank Matthijs and Bart for our very enjoyable conversations and collaborations. I would like to thank the other Maglia lab colleagues, Shengli, Sabine, Adina, Stefania, Nicole, Natalie, Huang, Veerle, Mariam, Kherim, Kaan, Oğuzhan, Laura, Edo, Kaori, Kasia, Xiangling, Yulia, Maaruthy, and the rest for being among the best colleagues I could have wished for. A special thanks to my students Jordy, Erica, Alberto, David, Reinier and Pieter for your hard work and dedication.

I would not have pursued a PhD (or MSc) without the courage and advise given by the following people (chronological order). Starting at the HAVO, a very special thank you to my biology teacher Sebastian for showing that you can succeed regardless of what other people say. Next I would like to thank Cor for giving me your wise advise and encouraging me to the leap from HBO to WO. Thank you Wesley for taking me under your wing when I made the transition from HBO to WO. I can't express in words how important your support has been on my journey ever since.

I would also like to thank Simon for our endless discussions about statistical models.

Ik wil graag mijn familie bedanken, met name mijn vader en moeder, dank voor de eindeloze gesprekken en adviezen die ik van jullie heb mogen ontvangen. Uiteraard wil ik graag mijn zus Amanda bedanken voor de gesprekken, adviezen, en verdere ondersteuning. Alsmede de ondersteuning verzorgt door onze viervoetige vrienden.

Daarnaast wil ik graag mijn ooms Gerard en Ruud bedanken voor alle steun. Een speciaal dankwoord voor mijn overleden grootmoeder, dank voor alle wijze adviezen. Ook wil ik graag mijn grootmoeder van moeders kant bedanken voor alle steun tijdens mijn studies en daar buiten.

If you are a friend, colleague, or family and I forgot to mention your name, please fill the dots.

I would like to thank.....

for your

and

List of Publications

Lucas FLR, Versloot RCA, Yakovlieva L, Walvoort MTC, Maglia G.

Protein identification with peptide nanopore fingerprinting.

Nature Communications (2021) doi: 10.1038/s41467-021-26046-9

de Lannoy C, Lucas FLR, Maglia G, de Ridder D.

Chop-n-drop: in silico assessment of a novel single-molecule protein fingerprinting method employing fragmentation and nanopore detection.

iScience. (2021) doi: 10.1016/j.isci.2021.103202

Lucas FLR, Piso TRC, van der Heide NJ, Galenkamp NS, Hermans J, Wloka C, Maglia G.

Automated Electrical Quantification of Vitamin B1 in urine using an Engineered Nanopore-Sensor.

Angewandte Chemie (2021) doi: 10.1002/anie.202107807

Lucas FLR, Sarthak K, Lenting EM, Coltan D, van der Heide NJ, Versloot RCA, Aksimentiev A, Maglia G.

The Manipulation of the Internal Hydrophobicity of FraC Nanopores Augments Peptide Capture and Recognition.

ACS Nano. (2021) doi: 10.1021/.acsnano.0c09958

Wloka C, Galenkamp NS, van der Heide NJ, Lucas FLR, Maglia G.

Strategies for enzymological studies and measurements of biological molecules with the cytolysin A nanopore.

Methods in Enzymology. (2021) doi: 10.1016/bs.mie.2021.01.007

Mutter NL, Huang G, van der Heide NJ, Lucas FLR, Galenkamp NS, Maglia G, Wloka C.

Preparation of Fragaceatoxin C (FraC) Nanopores.

Methods in Molecular Biology. (2021). doi: 10.1007/978-1-0716-0806-7_1

Willems K, Ruic D, Lucas F, Barman U, Hofkens J, Maglia G and Van Dorpe P.

Accurate modeling of a biological nanopore with an extended continuum framework.

Nanoscale (2020) doi: 10.1039/D0NR03114C

Lay Summary

This thesis describes experimental and computational analysis for the creation of (biological) nanopore based single-molecule sensors. In essence, nanopore sensors consist of a non-conductive sheet with a nanometre sized aperture connecting two electrolyte filled compartments (**Figure 1A**). The conductance across this aperture is reduced due to ion exclusion whenever a molecule translocates through it, which can be observed as a reduction in current when a potential is applied (**Figure 1B**).

Figure 1. Schematic representation of analyte translocation through a biological nanopore. **A:** shows an isolating membrane, including a lipid bilayer into which a nanopore is inserted. **B:** shows the translocation of a particle through a (Fragaceatoxin C) nanopore. The resistance across the membrane is stable and dominated by the unblocked nanopore (I). During translocation (II), the observed resistance is increased. After translocation (III), the resistance across the membrane is re-stabilised to the resistance of the unblocked nanopore.

Here, we engineered membrane spanning proteins for the creation of these nanometre sized apertures in order to observe and characterize peptides. By characterizing the current fluctuations over time, we were able to fingerprint trypsin digested proteins based on their translocating peptides.

In addition, we were able to trap substrate binding proteins inside a nanopore, which undergo a change in their ion exclusion when bound to an analyte (**Figure 2**). We sought to rapidly detect vitamin B1 in human urine, using a single nanopore in combination with a binding protein. We combined a physical understanding of the system with an empirical Bayesian approach, allowing the detection of thiamine in approximately 20 seconds.

Figure 2. Principle of binding protein based metabolite detection using nanopore resistive pulse sensing. **A:** shows a binding protein (PDB: 2QRY) inside a ClyA nanopore (PDB: 2WCD) and an expected signal (1). **B:** Shows the same as A, however, an analyte binding to the protein (thiamine monophosphate) has been added, the binding causes the observed current (1) to change over time (2).